Expert Systems With Applications 239 (2024) 122329

Contents lists available at ScienceDirect

Expert Systems With Applications

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

# DHESN: A deep hierarchical echo state network approach for algal bloom prediction

Bo Hu [a,c,d], Huiyan Zhang [a,e,*], Xiaoyi Wang [a,b,*], Li Wang [a,e], Jiping Xu [a,e], Qian Sun [a,e], Zhiyao Zhao [a,e], Lei Zhang [f]

[a] School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China
[b] Beijing Institute of Fashion Technology, Beijing 100029, China
[c] Institute of Robotics and Automatic Information System (IRAIS), Nankai University, Tianjin 300350, China
[d] Tianjin Key Laboratory of Intelligent Robotic (tjKLIR), Nankai University, Tianjin 300350, China
[e] China Light Industry Key Laboratory of Industrial Internet and Big Data, Beijing Technology and Business University, Beijing 100048, China
[f] Daxing Water Environment Monitoring Center, Beijing 102600, China

ARTICLE INFO

*Keywords:*
Algal bloom prediction
Deep echo state network
Hierarchical reservoir
Elastic regularization

ABSTRACT

A deep hierarchical echo state network (DHESN) is designed for rectifying the shortcomings of the shallow coupled structure with less reservoir dynamics. This design is with reference to algal bloom which is a complex ecological phenomenon. Accurate prediction of algal bloom can reduce the environmental impact and economic loss. Since the formation of algal bloom has chaotic characteristics, the ESN has been employed to realize its prediction function. First, the candidate variables with strong causal relationship have been screened by transfer entropy, and the redundant variables is eliminated. Then, a hierarchical reservoir structure is established that is inspired by the hierarchical characteristics from the brain. The hierarchical reservoir has realized the connection between the representative nodes of each subreservoir, and improved the information processing ability of the reservoir. Finally, the pruning and compression of the output weights have been realized by the elastic regularization method, which improves the robustness of the prediction model. The simulation results demonstrate that the DHESN has appreciable prediction accuracy in both the chaotic and the public algal bloom datasets. The DHESN contains richer dynamic characteristics, and can realize the self-organization of the network structure. It provides a novel idea to realize the prediction model of algal bloom with a high accuracy and low complexity.

## 1. Introduction

Owing to the extensive consumption of environmental resources, industrial wastewater from production activities and sewage have been accumulated in the lake through various ways. When nitrogen, phosphorus, and other nutrients in the lake and reservoir increase, eutrophication of water becomes hazardous (Sun et al., 2022; Vinçon-Leite & Casenave, 2019; Lin et al., 2021). In the eutrophication state of lakes and reservoirs, the over-propagated algae in the water results in algal bloom (Cui et al., 2021). The outbreak of algal bloom further consumes the dissolved oxygen in the water body, leading to the death of underlying aquatic organisms in a large area, which further deteriorates the water quality, and seriously degrades the water ecosystem (Bae & Seo, 2021). Particularly, eutrophication and algal bloom are possible to occur owing to poor mobility. Water quality conditions are the basis for algal bloom, and the nutrients such as nitrogen and phosphorus are considered as the important factors (Beretta-Blanco & Carrasco-Letelier, 2021). With respect to the algal bloom outbreak, the current management measures include the controlling of the nutrients, physical methods, and chemical drugs (Huang et al., 2018). It is important to monitor and predict algal bloom accurately for improving the efficiency of the long-term management and emergency treatment of algal bloom. Therefore, it constitutes an important step for the water quality management to design an effective and reliable prediction model.

According to previous studies (Liu et al., 2022a; Yu et al., 2021), the evolution of algal bloom has the typical nonlinear and temporal characteristics, and the extensively employed recurrent neural network forms an effective tool in dealing with such problems (Wang et al., 2020;

* Corresponding authors.
*E-mail addresses:* crosscshb@163.com (B. Hu), zhanghuiyan369@126.com (H. Zhang), wangxy@btbu.edu.cn (X. Wang), wangli@th.btbu.edu.cn (L. Wang), xujp@th.btbu.edu.cn (J. Xu), sunqian0403@hotmail.com (Q. Sun), zhaozy@btbu.edu.cn (Z. Zhao), 15115468@qq.com (L. Zhang).

https://doi.org/10.1016/j.eswa.2023.122329
Received 7 October 2022; Received in revised form 9 September 2023; Accepted 23 October 2023
Available online 30 October 2023
0957-4174/

Kim et al., 2022). Recently, as a data-driven model with simple structure, fast training speed, and high prediction accuracy, echo state network (ESN) has been applied in several fields (Chitsazan et al., 2019; Yao et al., 2019; Ribeiro et al., 2021). ESN can effectively model and predict nonlinear signals, especially for chaotic time series (Ferreira et al., 2013). Considering the above advantages of the ESN, previous studies have also attempted the prediction of algal bloom and obtained effective prediction performance (Zhang et al., 2021; Zhang et al., 2022). Deep learning provides a richer perspective for the time series prediction tasks (Goodfellow et al., 2016). Owing to the complexity of algal bloom prediction, the structural design and training methods in deep learning provide further ideas for improvement of ESN. By the special processing of input data, such as introducing attention mechanism into the ESN, the prediction accuracy has further been improved (Liu et al., 2022b). Inspired by the above work, from the perspective of data-driven, the structure and training method of the ESN will be optimized to achieve satisfactory prediction requirements.

Owing to the problem of predicting algal bloom with abundant water quality data and complex evolution mechanism, the single-layer reservoir structure shows the shortcoming of the insufficient information processing ability, which indicates that the applicability of shallow ESN in complex application scenarios is limited. Compared with single-layer neural networks, deep neural networks can deal with the complex system modeling more effectively (Thakkar & Chaudhari, 2021). Therefore, it is necessary to consider the deep expansion of the reservoir structure. By stacking the reservoir layer, deep ESN can effectively improve its ability to extract the hidden features (Gallicchio et al., 2017). For the actual algal bloom prediction problem, the deep ESN can be employed for a better learning of the nonlinear relationship in water quality data, for realizing the effective prediction of algal bloom. In the process of constructing the network above, the design of its structure and the output weights solution are still the key factors for improving the performance (Kaviani & Sohn, 2021).

The modular reservoir with multiple sub-reservoirs is an effective method for designing the ESN. Complete decoupling between the sub-reservoirs is an ideal structure. However, from the perspective of structural bionics, the topological structure of human brain is connected in a hierarchical and modular manner (Jarvis et al., 2010). The information processing ability of the fully decoupled sub-reservoirs has been weakened owing to the lack of transmission of time sequence information. Considering that the submodules of the brain network need to have certain information transmission, the brain network shows a complex hierarchical structure and information processing ability (Jarvis et al., 2010). The hierarchical structure contains the modular function partition, and realizes the information transfer between each module. Similarly, simulating this topology can improve the overall dynamics of the reservoir while the coupling between neurons will not be increased in excess. This improves the prediction performance and the stability.

The hierarchical design of the reservoir can strengthen the internal information transmission. However, the random design of the reservoir will still generate redundant neurons, thus leading to the ill-posed problem of solving the output weights and limiting the prediction performance (Shen et al., 2018). Regularization is a kind of effective method to solve the output weight, which can alleviate the ill-conditioned solution problem (Yildirim & Ozkale, 2019). Currently, the widely employed regularization methods include the sparse regularization and ridge regression (Dutoit et al., 2009; Yang et al., 2019). Sparse regularization can realize the sparsity of output weights. Ridge regression can compress the output weights and improve the robustness of the ESN. To exploit the advantages of the two regularization methods, the elastic regularization method integrates the two methods, which can solve the problem that the large magnitude of the output weights, besides triming the output weights, thus realizing the self-organization of the network structure (Xu et al., 2018).

With respect to the above analysis, this paper proposes a deep hierarchical ESN (DHESN) model and applies it for the algal bloom prediction. First, the transfer entropy has been employed for screening the appropriate input variables of DHESN and eliminate redundant candidate variables (Schreiber, 2000). Then, the DHESN model based on the hierarchical reservoir structure has been constructed. Compared with the traditional deep ESN model, the DHESN model adopts a hierarchical structure for each layer to improve the internal dynamic characteristics, and its output weights have been solved by the elastic regularization method. Therefore, the DHESN model is close to the real brain network with respect to the structure, and can realize the self-organization of the weight that can improve the prediction performance. We have employed two kinds of standard chaotic time series data and actual water quality data to verify the effectiveness of DHESN in the prediction task.

The structure of this paper is given in the following. Section 2 introduces the basic structure and theoretical analysis of DHESN. Section 3 shows the experimental results of the three datasets and verifies the prediction performance of DHESN. Section 4 discusses the theory and application advantages of DHESN. Finally, Section 5 draws the conclusions.

## 2. Methodology

### 2.1. Transfer entropy

Transfer entropy is a method to describe the coupling degree between variables. It measures the asymmetry between the variables to find the transfer relationship, so as to measure the causal relationship (Schreiber, 2000). Let $x_t\ y_t\ _{t\ 1}^{n}$ be a set of time series consisting of variables $X$ and $Y$. Given the embedding dimension $m$ and delay time , the state space reconstruction vectors at time $t$ are $x_t$ $x_t\ x_t$ $x_{t\ \ m\ 1}$ and $y_t$ $y_t\ y_t$ $y_{t\ \ m\ 1}$, respectively. The transfer entropy $TE_{X\ \ Y}$ from $X$ to $Y$ is defined as Eq. (1): (Schreiber, 2000).

(1)

where $H\ x$ represents the Shannon entropy of variable $X$.

Therefore, this paper firstly utilizes transfer entropy to screen the variables with a strong causal relationship and delete the redundant variables, for simplifying the complexity of the proposed model. Further, the retained variables are input into the subsequent model for training and prediction.

### 2.2. Echo sate network

ESN is a recurrent neural network consisting of three layers, viz., input, reservoir, and output layers. The input, reservoir, and output layers have $K$, $M$, and $L$ neurons, respectively. Without considering the feedback from the output layer to the reservoir, the calculation process of ESN is given as Eqs. (2)-(3) (Jaeger et al., 2007).

(2)

(3)

where $u\ t\quad 1\quad {}^{K\ 1}$ is the input at time $t\quad 1$, $x\ t\quad 1\quad {}^{M\ 1}$ is the state of the reservoir at time $t\quad 1$, $y\ t\quad 1\quad {}^{L\ 1}$ is the output at time $t\quad 1$, $f$ is the activation function of the neurons in the reservoir, $W^{in}$ $^{M\ K}$ is the input weight, $W^{res}$ $^{M\ M}$ is the reservoir weight, and $W^{out}$ $^{L\ M}$ is the output weight. In the original ESN, $W^{in}$ and $W^{res}$ are fixed after the random initialization, and only $W^{out}$ needs to be trained. In order to overcome the effect of the initial transient, it is assumed that the internal state matrix $X_o$ $x\ t$ $1$ $x\ t$ $^{T}$ is collected from time $t$ $1$, and its corresponding expected output matrix is $Y_t$ $y_t\ t$ $1$ $y_t\ t$ $^{T}$, where $y_t$ is the expected output. Then the output weight matrix can be solved by the least square method, and its

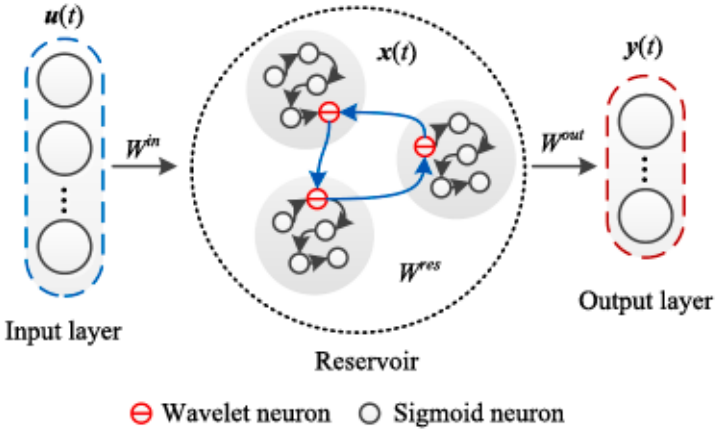


**Fig. 1.** The topology of the hierarchical reservoir.

objective function is described as Eq. (4):

$$\min E(\psi) = \min_{W^{out}} \|X_o W^{out} - T\|_2^2 \tag{4}$$

The solution for the output weight is described as Eq. (5) (Gallicchio & Micheli, 2011)

$$W^{out} = X_o^{\dagger} T = \left(X_o^T X_o\right)^{-1} X_o^T Y \tag{5}$$

where $X_o^{\dagger}$ represents the generalized inverse of $X_o$. According to the training method, the least square method needs the collected internal state matrix and expected output matrix while solving for the output weights. Therefore, when the training dataset $\{(u(t), y_t(t)) \mid t = t_{\min}+1, \cdots, t_{\max}\}$ is provided, the specific training steps of the original ESN are displayed in Algorithm 1.

Algorithm 1Step 1: Randomly generate the reservoir weight $W^{res}$ according to the reservoir size $M$, spectral radius $\rho$ and sparsity $D$;

Step 2: According to the input size $K$ and reservoir size $M$, the input weight $W^{in}$ is randomly initialized, and the reservoir state $x(0)$ is initialized.

Step 3: Input the data to the ESN. According to Eq. (2), collect the reservoir state at time $t_{\min}+1$ to obtain $X_o$;

Step 4: Calculate the output weight $W^{out}$ according to Eq. (5);

Step 5: Test the trained ESN.

### 2.3. Hierarchical reservoir

Each reservoir layer of the DHESN that is proposed in this paper is a hierarchical reservoir. The structure of a single-layer hierarchical reservoir has been designed in this section. According to Fig. 1, the design of the hierarchical reservoir consists of two parts, viz., one is the modular subreservoir, and the other part is the weight connection between the representative nodes in each subreservoir.

Here, the weight of the subreservoir has been generated by SVD decomposition (Qiao et al., 2016). That is, the fully connected subreservoir $W_p^{subres}(1 < p < N)$ has been randomly generated according to the given singular value distribution, and the size of each subreservoir is $n_{sub}$ (Qiao et al., 2016). After $N$ subreservoirs with size $n_{sub}$ have been randomly generated, the appropriate representative neurons need to be selected for the connection. We have selected the neuron with the largest connection strength in the subreservoir is selected as the representative neurons is described in Eqs.(6)-(7):

$$j_{rep} = \arg \max_{1<j_{rep}<n_{sub}} \{sc_1, sc_2, \cdots, sc_{n_{sub}}\} \tag{6}$$

where

$$sc_j = \sum_{z \in \varphi_j^{in}} |s_{jz}| + \sum_{z \in \varphi_j^{out}} |s_{zj}| \tag{7}$$

$\varphi_j^{in}$ represents the input weight set of the $j$ th neuron, $\varphi_j^{out}$ represents the output weight set of the $j$ th neuron, $s_{jz}$ represents the connection weight from $j$ th neuron to $z$ th neuron, and $s_{zj}$ represents the connection weight from $z$ th neuron to $j$ th neuron. After the representative nodes are determined, the activation function of the representative node in Eq. (8) is set as the wavelet activation function $\psi_{d_{j_{rep}}, t_{j_{rep}}}(x)$ (Cui et al., 2014).

$$\psi_{d_{j_{rep}}, t_{j_{rep}}}(x) = 2^{d_{j_{rep}}/2}\left(2^{d_{j_{rep}}} x - t_{j_{rep}}\right) e^{-\left(2^{d_{j_{rep}}} x - t_{j_{rep}}\right)^2/2} \tag{8}$$

where $d_{j_{rep}} = j_{rep}/N$, $t_{j_{rep}} = j_{rep}/N - 0.5$, $\left(j_{rep} = 1, 2, \cdots, N\right)$. In this work, the connection between the representative nodes is realized by the construction mechanism of the improved small-world network (Song & Feng, 2010). The improved small-world network employs the edge probability $p_w$ as the connection weight between each nodes is shown in Eq. (9).

$$p_w = \alpha e^{-\beta dist(u,v)} \tag{9}$$

where $\alpha$ is the density sensitivity coefficient, $\beta$ is the distance sensitivity coefficient, and $dist(u, v)$ is the Euclidean distance between nodes $u$ and $v$. When the distance between nodes is larger, the connection weight is smaller. Particularly, the unilateral or bidirectional connections can be realized between the representative nodes to improve the information transfer ability of the reservoir. The specific construction process of the representative node can be seen in Step 4 of Algorithm 2.

Therefore, the model of HESN is described in Eqs.(10)-(11).

$$x(n+1) = f\left(W^{in} u(n+1) + W^{hres} x(n)\right) \tag{10}$$

$$y(n+1) = W^{out} x(n+1) \tag{11}$$

where $W^{hres} \in \mathbb{R}^{(N \times n_{sub}) \times (N \times n_{sub})}$ is the connection weight of the hierarchical reservoir, and $f = \left[f_1, f_2, \cdots, f_{N \times n_{sub}}\right]$ is the activation function of the hierarchical reservoir. Particularly, if the neuron is not a representative node, the activation function of the node is the sigmoid function. If this node is a representative node, the activation function of this node is the wavelet function.

### 2.4. DHESN

The DHESN is composed of the multi-layer hierarchical reservoirs, and its network structure is displayed in Fig. 2. Assuming that each

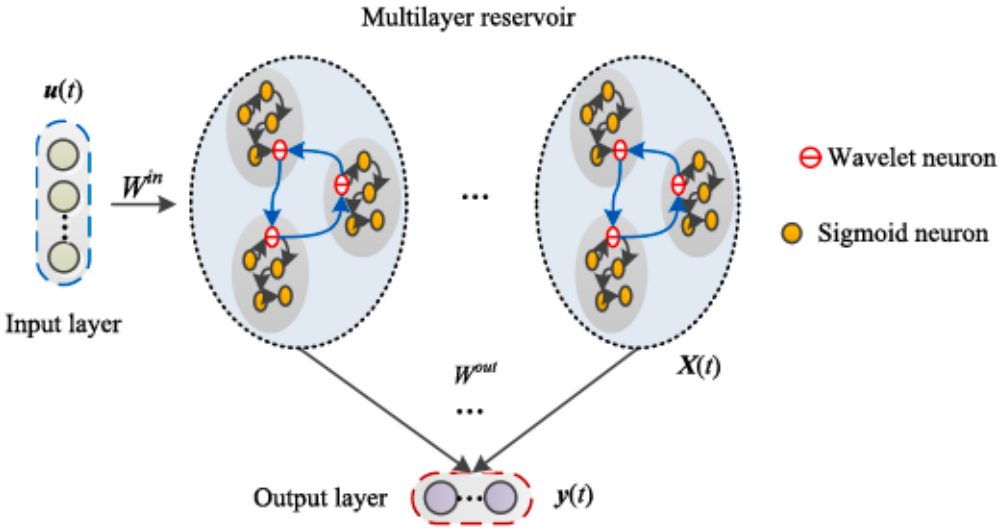


**Fig. 2.** The topology of DHESN.

hierarchical reservoir has the same number and size of subreservoirs, viz., $N$ and $n_{sub}$, respectively, then the DHESN can be described as Eqs. (12)-(13):

$$x^{(l)}(n+1) = f^{(l)}\left(W_{in}^{(l)}x_{in}^{(l)}(n+1) + W_{inter}^{(l)}x^{(l)}(n)\right)$$
$$\begin{cases} x_{in}^{(l)}(n+1) = u(n+1), l = 1 \\ x_{in}^{(l)}(n+1) = x^{(l-1)}(n+1), l > 1 \end{cases} \tag{12}$$

$$y(n+1) = W^{out}X(n+1) \tag{13}$$

where $W^{out} \in \mathbb{R}^{1\times(N\times n_{sub}\times l_{max})}$, $X(n+1) = \left[x^{(1)}(n+1), x^{(2)}(n+1), \cdots, x^{(l_{max})}(n+1)\right]$. $l_{max}$ is the maximum number of layers.

The input weights and the reservoir weights remain unchanged after initialization, similar to the training process of the original ESN, and only the output weights need to be trained. We have employed the elastic regularization with $L_2$ and $L_{1/2}$ to solve the output weights (Liang et al., 2013). Elastic regularization can compress and reduce the collinearity of the output weights. Moreover, the output weights can be trimmed automatically, which reduces the complexity. The elastic regularization objective function is shown as Eq. (14).

$$E(W^{out}) = \|Y - W^{out}X\|_2^2 + \lambda_2\|W^{out}\|_2^2 + \lambda_{1/2}\|W^{out}\|_{1/2}^{1/2} \tag{14}$$

where $X \in \mathbb{R}^{(N\times n_{sub}\times l_{max})\times T_{train}}$ is the state matrix of the reservoir, $Y \in \mathbb{R}^{1\times T_{train}}$ is the output matrix, $T_{train}$ is the length of the training set, $\lambda_2$ is the $L_2$ regularization coefficient, and $\lambda_{1/2}$ is the $L_{1/2}$ regularization coefficient. Concomitantly, the state matrix $X$ and the target output vector $Y$ have been supplemented to obtain the following augmented output matrix $X^*$ and the augmented target output vector $Y^*$ are described in Eq. (15).

$$X^* = \frac{1}{\sqrt{1+\lambda_2}}\begin{pmatrix} X \\ \sqrt{\lambda_2}I \end{pmatrix}, Y^* = \begin{pmatrix} Y \\ 0 \end{pmatrix} \tag{15}$$

Let $\lambda_h = \lambda_{1/2}/\sqrt{1+\lambda_2}$, and $W^{out^*} = \sqrt{1+\lambda_2}W^{out}$, then the objective function can be transformed into Eq. (16).

$$\begin{aligned} E(W^{out}) &= \|Y - W^{out}X\|_2^2 + \lambda_2\|W^{out}\|_2^2 + \lambda_{1/2}\|W^{out}\|_{1/2}^{1/2} \\ &= \left\|\begin{pmatrix} Y \\ 0 \end{pmatrix} - \frac{1}{\sqrt{1+\lambda_2}}\begin{pmatrix} X \\ \sqrt{\lambda_2}I \end{pmatrix}\sqrt{1+\lambda_2}W^{out}\right\|_2^2 \\ &\quad + \frac{\lambda_{1/2}}{\sqrt{1+\lambda_2}}\left\|\sqrt{1+\lambda_2}W^{out}\right\|_{1/2}^{1/2} \\ &= \|Y^* - W^{out^*}X^*\|_2^2 + \lambda_h\|W^{out^*}\|_{1/2}^{1/2} \end{aligned} \tag{16}$$

The original elastic regularization problem has been transformed into a generalized $L_{1/2}$ regularization problem. We have used the coordinate descent method is used to solve $W^{out^*}$ in Eq. (16) (Liang et al., 2013). In the coordinate descent method, only one parameter, i.e., the output weight component $W_k^{out^*}$, $k = 1, 2, \cdots, N\times n_{sub}\times l_{max}$, has been updated during the solution process, whereas other parameters remain unchanged. Then the objective function of the above generalized $L_{1/2}$ regularization problem has been expanded in terms of components as Eq. (17).

$$\begin{aligned} E_k\left(W_k^{out^*}\right) &= \sum_{t=1}^{T_{train}}\left(y(t) - W_k^{out^*}x_k(t) - \sum_{i\neq k}^{N\times n_{sub}\times l_{max}}\overline{W}_i^{out^*}x_i(t)\right)^2 \\ &\quad + \lambda_h\left(\left|W_k^{out^*}\right|^{\frac{1}{2}} + \sum_{i\neq k}^{N\times n_{sub}\times l_{max}}\left|\overline{W}_i^{out^*}\right|^{\frac{1}{2}}\right) \end{aligned} \tag{17}$$

where $\overline{W}_i^{out^*}$ represents the fixed output weight component in the previous step. After sorting out Eq. (17), we obtain

$$\begin{aligned} E_k\left(W_k^{out^*}\right) &= \sum_{t=1}^{T_{train}}\left(W_k^{out^*}x_k(t)\right)^2 + \sum_{t=1}^{T_{train}}\left(y(t) - \sum_{i\neq k}^{N\times n_{sub}\times l_{max}}\overline{W}_i^{out^*}x_i(t)\right)^2 \\ &\quad - 2\sum_{t=1}^{T_{train}}\left(y(t) - \sum_{i\neq k}^{N\times n_{sub}\times l_{max}}\overline{W}_i^{out^*}x_i(t)\right)W_k^{out^*}x_k(t) \\ &\quad + \lambda_h\left|W_k^{out^*}\right|^{\frac{1}{2}} + \lambda_h\sum_{i\neq k}^{N\times n_{sub}\times l_{max}}\left|\overline{W}_i^{out^*}\right|^{\frac{1}{2}} \end{aligned} \tag{18}$$

Assuming that $\sum_{t=1}^{T_{train}}x_k^2(t) \neq 0$, the minimization of the objective function in Eq. (18) is equivalent to the minimization in Eq. (19).

(19)

Let $C_k$ $\sum_{t=1}^{T_{train}} y(t)$ $\sum_{i=k}^{N+n_{sub}-l} \overline{W}_i^{out*} x_i(t)$ $x_k(t)$ $\sum_{t=1}^{T_{train}} x_k^2(t)$, $hk$ $h$ $\sum_{t=1}^{L_{train}} x_k^2(t)$, the Eq. (19) has been simplified as Eq. (20).

(20)

Taking the derivative of Eq. (19) with respect to $W_k^{out*}$, we obtain Eq. (21).

(21)

where *sign* is the sign function. According to (Liang et al., 2013), we solve $E_k$ $W_k^{out*}$ 0 to obtain Eq. (22).

(22)

where $C_k$ $_{hk}$ $C_k$ 3 $\frac{3}{2}$ 8 , and $W^{out}$ $W^{out*}$ 1 $_2$.

Based on the above analysis, the detailed training algorithm of DHESN can be given as Algorithm 2.Algorithm 2 Step 1: Calculate the transfer entropy of the candidate variable to the predicted variable and set the screening threshold $_{TE}$. The candidate variable is retained if the transfer entropy exceeds or is equal to the $_{TE}$. If it is less than $_{TE}$, the candidate variable is deleted.

**Step 2:** Set the number of reservoir layers $l$ and the input weight range of each layer. Each layer randomly generates the subreserveoirs through SVD method, and each subreservoir contains $n_{stub}$ neurons.

**Step 3:** Select the representative nodes in each subreservoir according to Eq. (8), and set the wavelet activation function of each representative node according to Eq. (7).

**Step 4:** Set a two-dimensional plane with both the ranges of x-axis and y-axis as 0 1 . Set $N_s$ seed nodes randomly among the representative nodes, and set the coordinate of seed nodes as $\frac{j_s}{N_c}$ 1 $\frac{j_s}{N_c}$ 1 $j_s$ 1 $N_c$ . According to Eq. (8), make all seed nodes have at least a unidirectional connection. Then, a new representative node with random coordinates has been added each time. Further, the unidirectional or bidirectional connection with the existing representative node in the plane is realized according to Eq. (8), where the connection distance of each representative node is set as the connection weights between each representative node. After all the weights are connected, the spectral radius of each hierarchical reservoir needs to be adjusted to ensure the echo state property.

**Step 5:** Set the regularization parameters and initialize the output weight $W_o^{out*}$ 0. Let $W^{out*}$ $W_o^{out*}$, and iteratively calculate the component $W_k^{out*}$ $k$ 1 2 $l$ $N$ $n_{stub}$ of $W^{out*}$ according to Eq. (21). If the maximum number of iterations is satisfied, then output $W^{out}$ $W^{out*}$ 1 $_2$. Otherwise, let $W_o^{out*}$ $W^{out*}$ and continue the process until the condition is satisfied.

**Step 6:** Test the trained DHESN.

### 2.5. Theoretical analysis

(1) Spectral radius

To ensure the echo state property of the DHESN, the spectral radius of each hierarchical reservoir has been updated after the representative nodes are connected as Eq. (23).

(23)

where $W_0^{hres\ l}$ is the hierarchical reservoir after the representative nodes are connected, and $W^{hres\ l}$ is the weight of the updated hierarchical reservoir at $l$ th layer. Each reservoir layer of DHESN only needs to set the random range of the singular value without setting the spectral radius in advance (Qiao et al., 2016).(2) Sparsity

After the representative nodes are connected, the sparsity of the hierarchical reservoir changes. The sparsity of the $l$ th reservoir layer (fully decoupled subreservoir) before completing the representative node connection completed is described in Eq. (24).

(24)

Since the weight connection between representative nodes is realized according to the small-world network method, its connection is random. The sparsity of the hierarchical reservoir, after the node connection is divided into two boundary cases. When all representative nodes are connected at least, the sparsity of the hierarchical reservoir is the minimum, and the minimum sparsity of the hierarchical reservoir at $l$ th layer is shown in Eq. (25).

(25)

When all the representative nodes are bidirectional connected, the sparsity of the hierarchical reserve pool reaches the maximum, and the maximum sparsity of the $l$ th hierarchical reservoir layer is shown in Eq. (26).

(26)

Therefore, the sparsity of the each hierarchical reservoir ranges from $\frac{n_{stub}^{l\ 2}\ 1}{N^l\ n_{stub}^{l\ 2}}$ $\frac{N^l\ n_{stub}^{l\ 2}\ 1}{N^l\ n_{stub}^{l\ 2}}$ . Similar to the spectral radius, the sparsity of each reservoir layer of the DHESN does not need to be set in advance.(3) Average state entropy (ASE)

From the information theory, the instantaneous state entropy of the reservoir output can be employed to measure the dynamic richness of the ESN effectively. For each reservoir layer $l$ and each time $t$, its instantaneous state entropy $H^l$ $t$ can be defined as Eq. (27): (Ozturk et al., 2007).

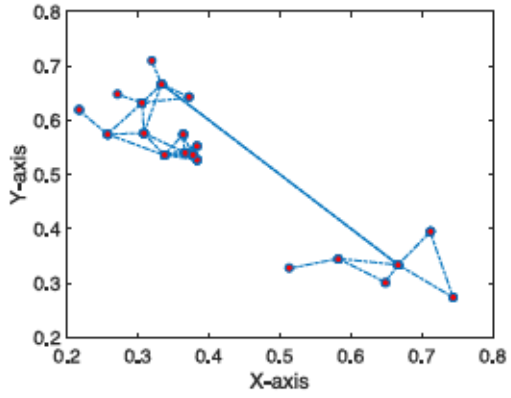


**Fig. 3.** The topology of the representative node connections in Henon map prediction.

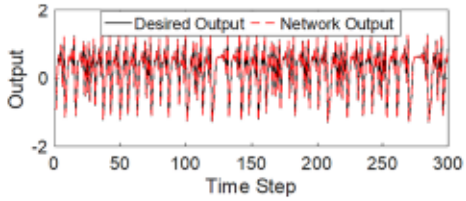


**Fig. 4.** Prediction results of DHESN in Henon map prediction.

$$H^{(l)}(t) = -\log\left(\frac{1}{\left(N^{(l)}n_{sub}^{(l)}\right)^2}\sum_{j=1}^{N^{(l)}n_{sub}^{(l)}}\left(\sum_{i=1}^{N^{(l)}n_{sub}^{(l)}}\kappa\left(x_j^{(l)}(t)-x_i^{(l)}(t)\right)\right)\right) \tag{27}$$

where $x_j^{(l)}(t)$ represents the state of the $j$ th neuron in $l$ th reservoir layer at time $t$, and $\kappa(\cdot)$ is Gaussian kernel function (Ozturk et al., 2007). The ASE in the time length $T$ is defined as Eq. (28).

$$ASE^{(l)} = \frac{1}{T}\sum_{t=1}^{T}H^{(l)}(t) \tag{28}$$

ASE can quantify the dynamic characteristics of each reservoir layer. It can analyze the influences of the reservoir structure on the dynamics from the perspective of single layer, and the deep reservoir structure on the dynamics from the perspective of the number of reservoir layers.

## 3. Experiment and result

Two typical chaotic time series datasets have been employed for simulation experiments, including Henon mapping prediction (Hénon, 1976) and Mackey-Glass time series prediction (Jaeger & Haas, 2004), for verifying the prediction performance of the DHESN, by considering the chaotic property of algal bloom. Then, the water quality dataset of Lake Mendota in Wisconsin, USA, has been employed to verify the effectiveness of DHESN in the actual algal bloom prediction (Lead et al., 2020). The compared neural network models include extreme learning machine (ELM) (Liang et al., 2006), original ESN (OESN) (Jaeger et al., 2007), GESN (Qiao et al., 2016), RESN (Dutoit et al., 2009), DBN (Hinton et al., 2006), $L_2$ regularized deep ESN (DeepESN-$L_2$) (Gallicchio et al., 2017), and $L_{1/2}$ regularized deep ESN (DeepESN-$L_{1/2}$). In all the experiments, with $\alpha = 0.5$ and $\beta = 40$, the number of subreservoir at each layer is 20, and each subreservoir contains five fully connected neurons. Further, the input weight range of each layer is $[-1,1]$, and the number of seed nodes at each hierarchical reservoir layer is $N_c = 0.1N$. The grid search method has been employed to determine the optimal parameters for the number of layers and the two regularization parameters. The maximum number of layers vary from two to five, and the selection range of the two regularization parameters is $(10^{-1}, 10^{-2}, \cdots, 10^{-10})$. Each experiment has been tested 50 times, and the experimental results were taken as the average of 50 experiments.

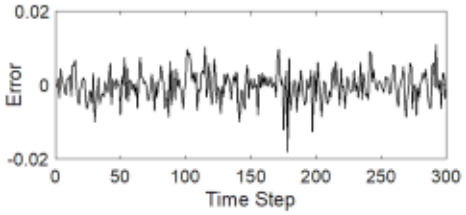


**Fig. 5.** Prediction error of DHESN in Henon map prediction.

**Table 1**
Comparison results of Henon map prediction performance based on different methods.

| Method | Structure | NRMSE Ave. | Std. |
|---|---|---|---|
| ELM | 400 | 0.0591 | 0.0009 |
| OESN | 400 | 0.0268 | 0.0123 |
| GESN | 400 | 0.0121 | 0.0006 |
| RESN | 400 | 0.0078 | 0.0004 |
| DBN | 50–50–50–50 | 0.0074 | 0.0011 |
| DeepESN-$L_2$ | 100–100–100–100 | 0.0064 | 0.0005 |
| DeepESN-$L_{1/2}$ | 100–100–100–100 | 0.0080 | 0.0016 |
| DHESN | 100–100–100–100 | 0.0057 | 0.0005 |

To test the prediction performance of DHESN, normalized root mean square error (NRMSE) has been selected as Eq. (29).

$$\mathrm{NRMSE} = \sqrt{\frac{\sum_{n=1}^{L_{test}}(d(n)-y(n))}{L_{test}\sigma_d^2}} \tag{29}$$

where $L_{test}$ is the size of the testing set, $d(n)$ is the actual value, $y(n)$ is the predicted value (the network output), and $\sigma_d^2$ is the variance of the actual value. A smaller NRMSE yields a better the prediction performance of the network.

### 3.1. Henon map prediction

Henon map is a classical chaotic discrete-time dynamic system, which is extensively employed as a benchmark function for the time series prediction problems to verify the effectiveness of the model. The formula for the Henon mapping system is described as Eq. (30).:

$$\begin{cases} u_1(t+1) = u_2(t) - au_1^2(t) + 1 \\ u_2(t+1) = bu_1(t) \end{cases} \tag{30}$$

where $a = 1.4, b = 0.3$, the initial value $u_1(1) = u_2(1) = 0$, and $u_2$ is the predicted value. The length of the dataset is 1000, among which the training length is 700 and the testing length is 30. Furthermore, the reservoir state is collected following a set of 50 training data to overcome the impact of the initial transient. In this experiment, the number of reservoir layer is 4, with 100 neurons in each layer. The $L_{1/2}$ regularization parameter is 1e-5, and the $L_2$ regularization parameter is 1e-4.

Fig. 3 shows the topology of the connections between the representative nodes, which exhibits the typical clustering characteristics of a small-world network. And each representative node can also realize the bidirectional connection and self-feedback connection to render a hierarchical structure for the improved reservoir.

Fig. 4 and Fig. 5 show the prediction results and error of DHESN in Henon map prediction. DHESN shows a good prediction effect, and its error values fall in the range $[-0.02, 0.02]$. Table 1 specifically shows the comparison of the prediction accuracy between different methods. The

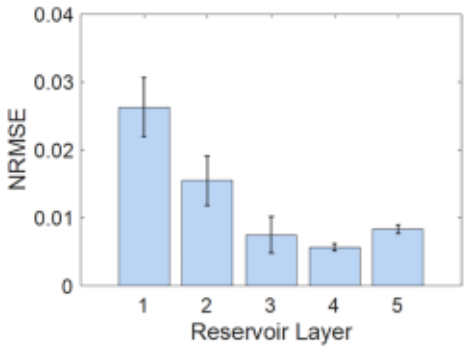


**Fig. 6.** Influence of different number of reservoir layers on the prediction performance of Henon map prediction.

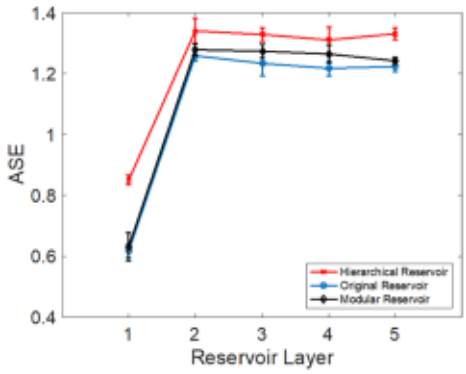


**Fig. 7.** Influence of different reservoir structure on the information content in Henon map prediction.

prediction accuracy of DHESN is significantly better than the shallow ESN, and slightly better than DeepESN. The prediction accuracy with elastic regularization exceeds that of DeepESN-$L_2$ and DeepESN-$L_{1/2}$. Although the standard deviation of DHESN exceeds that of RESN, the mean NRMSE of DHESN is the minimum of all models. It shows good prediction performance and stability.

Fig. 6 displays the influence of different numbers of the reservoir layers on the prediction performance. Owing to the increase of the number of reservoir layers, the NRMSE of DHESN decreases first and then increases, showing the best prediction effect when the number of layers is four. The prediction performance may not be better for a higher number of reservoir layers. We determine the appropriate number of reservoir layers according to specific problems.

To further analyze the reasons for the good prediction performance of DHESN, Fig. 7 displays the influence of different network structures on ASE in Henon map prediction. Accordingly, owing to the increase of layers, the ASE increases significantly in the second layer irrespective of the kind of reservoir structure, and is basically stable in the later layers. This indicates that the number of reservoir layers can enrich the dynamic characteristics of the reservoir state. However, the ASE of different reservoir structures is different. Among them, the ASE of the original reservoir structure is low, and when the fully decoupled modular reservoir is adopted, its dynamics is enhanced modestly, owing to the fully decoupled sub-reservoir having different dynamics, which can improve the information of the reservoir state. The ASE of the hierarchical reservoir exceeds that of the other two structures in each layer. By connecting the representative nodes with the wavelet activation function, the ASE of the hierarchical reservoir can effectively enrich the information of the reservoir. The deep features can be appreciably extracted, so that DHESN can have a better prediction performance.

Furthermore, elastic regularization is also an important way to improve the prediction accuracy of DHESN. According to Fig. 8, the

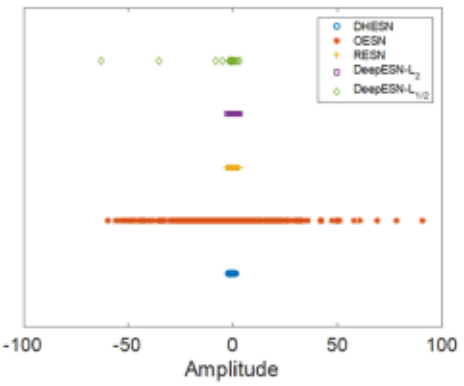


**Fig. 8.** Influence of different regularization methods on output weights in Henon map prediction.

output weight amplitude of DHESN exceeds that of OESN without regularization, and its output weight amplitude is compressed to $[-2.3037, 1.1892]$, which is also slightly better than other models based on $L_2$ regularization method. Concomitantly, $L_{1/2}$ regularization also prunes the output weight components of DHESN. Among the 400 neurons in layer four, 76 output weights have been pruned. Therefore, elastic regularization can effectively solve the approximate collinearity problem and weight redundancy problem of DHESN, thus improving the prediction effect of DHESN.

### 3.2. Mackey-Glass system prediction

The Mackey-Glass system is also a standard model for verifying network performance, which is shown in Eq. (31).

$$\frac{dx(t)}{dt} = \frac{ax(t-\tau)}{1+x^n(t-\tau)} + bx(t) \tag{31}$$

where $n = 10, a = 0.2, b = -0.1, \tau = 17$. The fourth-order Runge-Kutta method has been employed to generate 2000 sets of data, and one-step prediction has been performed on them, including 1000 training set and 1000 testing set. The reservoir state has been collected after 100 sets of data input. In this experiment, the number of layers in the reservoir is three, with 100 neurons in each layer. The $L_{1/2}$ regularization parameter is 1e-6, and the $L_2$ regularization parameter is 1e-5.

Fig. 9 and Fig. 10 respectively show the prediction results and error of DHESN, respectively, in Mackey-Glass system prediction, and the prediction error range of DHESN is between $[-0.01, 0.01]$. Table 2 lists the comparison between the prediction accuracy of different methods. After 50 experiments, the mean NRMSE of DHESN is better than the prediction accuracy of shallow ESN and deep ESN models.

Fig. 11 displays the influence of different reservoir layers on network performance in Mackey-Glass system prediction. Owing to the increase in the number of layers, DHESN shows a better prediction performance for three layers.

Fig. 12 further shows the impact of different reservoir structures on ASE in Mackey-Glass system prediction. Owing to the increase in the number of reservoir layers, the ASE increases significantly in the second layer, and is basically stable in the later layers. From the perspective of

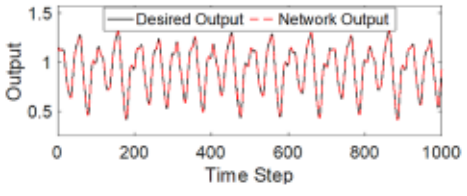


**Fig. 9.** Prediction results of DHESN in Mackey-Glass system prediction.

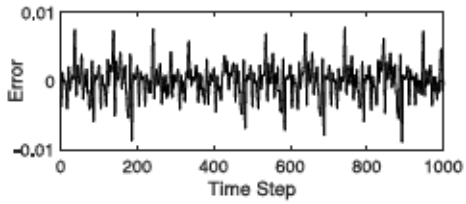


**Fig. 10.** Prediction error of DHESN in Mackey-Glass system prediction.

**Table 2**
Comparison results of prediction performance in Mackey-Glass system.

| Method | Structure | NRMSE Ave. | Std. |
|---|---|---|---|
| ELM | 300 | 0.0444 | 0.0013 |
| OESN | 300 | 0.0338 | 0.0120 |
| GESN | 300 | 0.0226 | 0.0002 |
| RESN | 300 | 0.0224 | 0.0001 |
| DBN | 50–50–50 | 0.0238 | 0.0038 |
| DeepESN-$L_2$ | 100–100–100 | 0.0126 | 0.0003 |
| DeepESN-$L_{1/2}$ | 100–100–100 | 0.0178 | 0.0022 |
| DHESN | 100–100–100 | 0.0098 | 0.0033 |

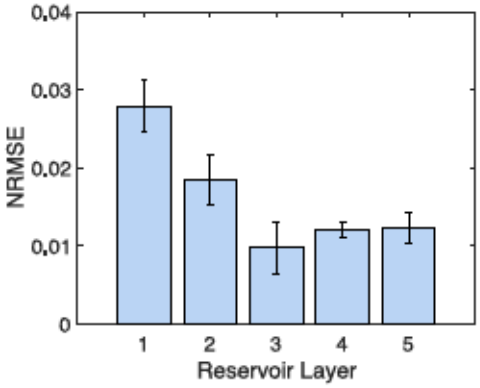


**Fig. 11.** Influence of different reservoir layers on the prediction performance of Mackey-Glass system.

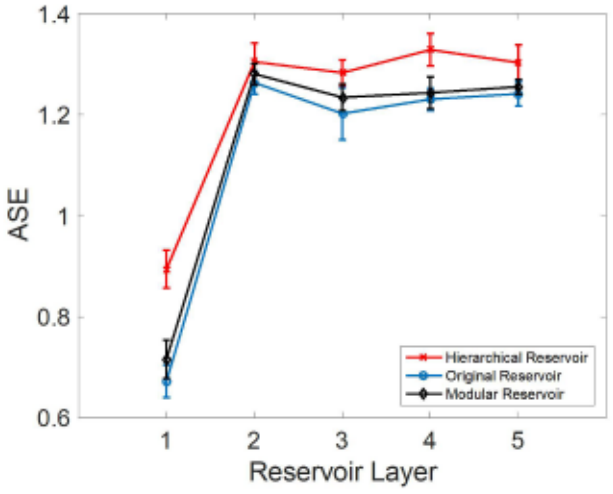


**Fig. 12.** Influence of different reservoir structure on the information content in Mackey-Glass system prediction.

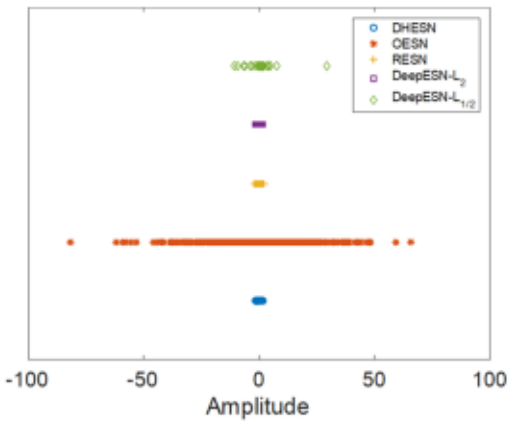


**Fig. 13.** Influence of different regularization methods on output weights in Mackey-Glass system prediction.

the reserve pool structure, the information content of the modular reservoir slightly increases, whereas the ASE of the hierarchical reservoir exceeds that of the other two structures in each layer, which has richer dynamic characteristics of the reservoir state.

Fig. 13 shows the influence of different regularization methods on the output weights in the Mackey-Glass system prediction. The output weight amplitude of DHESN is significantly smaller than that of the ESN model without $L_2$ regularization, and its output weight amplitude has been compressed to $[-1.5459, 1.6731]$. The pruning effect of DHESN is better than that of $L_{1/2}$ regularized deep ESN. 36 weights of DHESN have been pruned, whereas 23 output weight components of $L_{1/2}$ regularized ESN are zero.

### 3.3. Algal bloom prediction

The prediction of algal bloom is primarily based on the concentration of chlorophyll-a as the characteristic variable. In this experiment, the water quality and meteorological data collected in Lake Mendota, Wisconsin, USA has been used to verify the effectiveness of DHESN. Lake Mendota has generated 1489 sets of data from July 2016 to September 1, 2016. The data sampling interval has been for one hour. The name, units, and statistical indicators of all variables in the dataset are listed in Table 3.

The length of the dataset is 1489, of which the length of the training set is 1036 and the length of the testing set is 453. Furthermore, the experiment initially collects the reservoir state after 100 times. Owing to the large number of variables in the dataset, there may be redundant ones. Therefore, this experiment quantifies the causal relationship between the candidate variables and chlorophyll-a by transferring entropy.

**Table 3**
Detail information of dataset variables in Lake Mendota.

| Variable | Unit | Max. | Min. | Ave. | Std. |
|---|---|---|---|---|---|
| Air temperature | ℃ | 32.39 | 14.92 | 23.295 | 2.9668 |
| Relative humidity | % | 99.9 | 43.79 | 76.0086 | 12.0879 |
| Wind speed | m/s | 10.65 | 0.03 | 3.6585 | 1.8665 |
| Wind direction | ° | 345.39 | 20.59 | 204.7823 | 82.5804 |
| Phycocyanin | mg/L | 1329.5 | 434.81 | 654.1126 | 98.8418 |
| Photosynthetically-active radiation | μmol/($m^2$s) | 1996.7 | 0 | 487.663 | 602.7265 |
| Subsurface photosynthetically-active radiation | μmol/($m^2$s) | 216.86 | 0 | 20.035 | 33.3408 |
| Water temperature | ℃ | 29.35 | 23.52 | 25.348 | 1.0503 |
| Oxygen saturation | % | 213.02 | 19.19 | 99.8384 | 20.4296 |
| Dissolved oxygen | mg/L | 17.18 | 1.53 | 8.1964 | 1.6664 |
| Partial pressure of carbon dioxide | ppm | 410.12 | 198.87 | 271.6291 | 34.6809 |
| Chlorophyll-a | mg/L | 3396.2 | 853.95 | 1605.8 | 348.1447 |

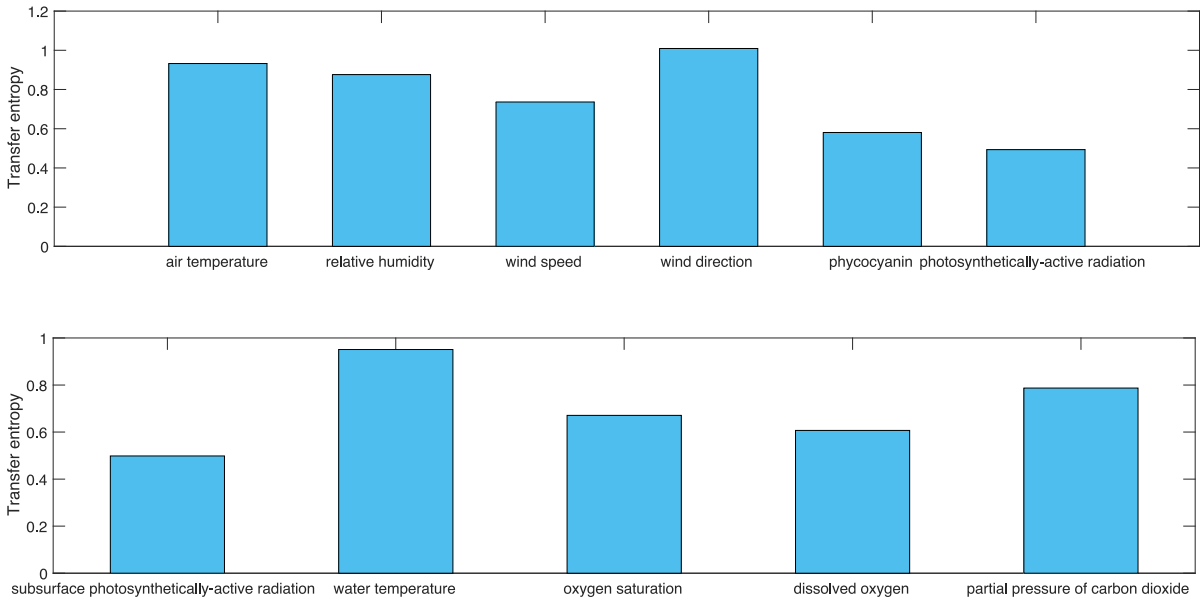


**Fig. 14.** The transfer entropy of candidate variables to chlorophyll-a.

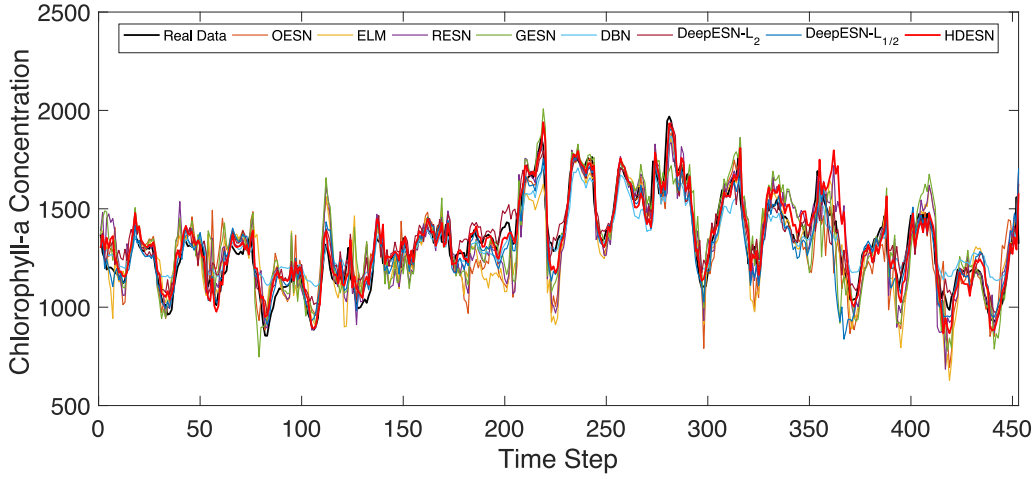


**Fig. 15.** Comparison of prediction results of chlorophyll-a concentration.

According to Fig. 14, the screening threshold has been set in this experiment, hence all the variables except photosynthetically-active radiation and subsurface photosynthetically-active radiation have been used as the input of DHESN for the chlorophyll-a prediction. Furthermore, the number of layers in this experiment is four, with 100 neurons in each layer. The $L_{1/2}$ regularization parameter is 1e-2, and the $L_2$ regularization parameter is 1e-5.

Fig. 15, Fig. 16 and Fig. 17 show the prediction results, prediction errors, and regression plots of DHESN, respectively, in this experiment. Compared with the alternate ESN models, the prediction results of DHESN are proximate to the actual value, and the prediction error range is in the interval 400 300 . Table 4 shows the comparison between the different methods. The mean NRMSE and standard deviation NRMSE of DHESN are better than other variant ESN, which realizes the effective prediction of actual algal bloom.

Fig. 18 displays the influence of different reservoir layers on the network performance in chlorophyll-a concentration prediction. DHESN shows better prediction performance for four layers. Fig. 19 shows the effect of different reservoir structures on ASE. ASE increases significantly in the second layer, but the number of later layers does not change significantly. The ASE of the hierarchical reservoir exceeds that of the other two structures in each layer, which verifies that the hierarchical reservoir has richer information and improves the ability of feature extraction with the increase in the number of layers.

Fig. 20 depicts the influence of different regularization methods on the output weights in the experiment of chlorophyll-a concentration prediction. The output weight amplitude of DHESN is smaller than that of other types of regularized ESN models, and compressed to 1 4518 0 6296 , which has better robustness. The pruning effect of DHESN is also better than that of $L_{1/2}$ regularized deep ESN in which 306 weights of DHESN are pruned, whereas 142 output weight components of $L_{1/2}$ regularized deep ESN are zero. DHESN can better reduce the complexity of the prediction model. By the mechanism design of hierarchical reservoir, the feature extraction of deep reservoirs, and the integration of regularization methods, the effective feature extraction and the self-organization of the structure have been achieved. Therefore, it can be applied to the actual lake and reservoirs for an effective prediction of algal bloom.

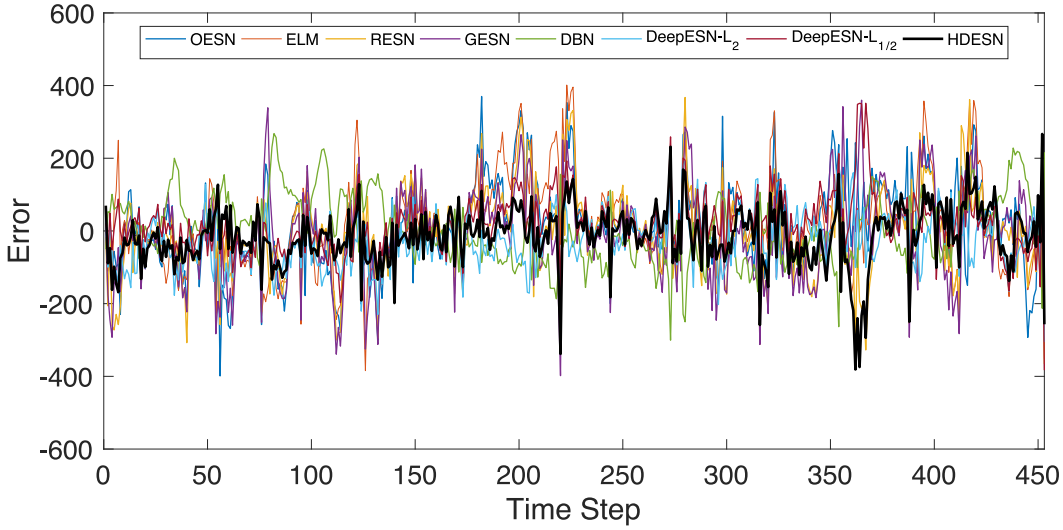


**Fig. 16.** Comparison of prediction errors of chlorophyll-a concentration with different methods.

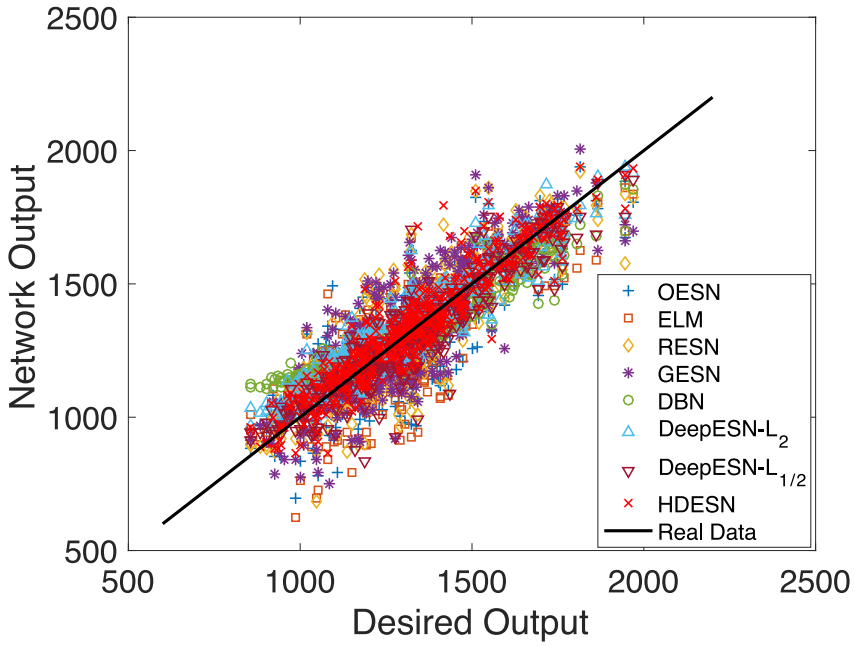


**Fig. 17.** Regression plots of predicted and actual values of chlorophyll-a concentration with different methods.

**Table 4**
Comparison results of prediction performance of chlorophyll-a concentration based on different methods.

| Method | Structure | NRMSE | |
|---|---|---|---|
| | | Ave. | Std. |
| ELM | 400 | 0.5906 | 0.0096 |
| OESN | 400 | 0.5976 | 0.0269 |
| GESN | 400 | 0.5826 | 0.0891 |
| RESN | 400 | 0.5712 | 0.0188 |
| DBN | 50 50 50 50 | 0.4321 | 0.0105 |
| DeepESN-$L_2$ | 100 100 100 100 | 0.4042 | 0.0226 |
| DeepESN-$L_{1/2}$ | 100 100 100 100 | 0.4071 | 0.0283 |
| DHESN | 100 100 100 100 | 0.3758 | 0.0085 |

## 4. Discussion

According to the previous references and the water quality data from Lake Mendota, the evolution of chlorophyll-a concentration is a mixture of nonlinearity and periodicity. Combined with Table 4, the shallow ESN model can achieve good prediction accuracy, whereas the deep reservoir model can improve the prediction performance. Combined with the above three experiments, the proposed model exploits the advantages of deep neural network structure, and finds application in algal bloom prediction.

DHESN adopts a hierarchical structure and mixed activation function for each reservoir layer. These highlights improve the information processing ability of deep ESN. According to Fig. 7, Fig. 12 and Fig. 19, the information content of the reservoir neurons is gradually increased with the further modularization and hierarchical design of the neuron-coupled reservoir, which helps in the improvement of the prediction

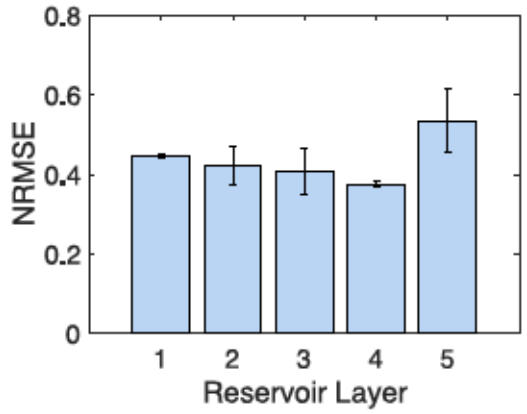


**Fig. 18.** Effects of different reservoir layers on the prediction performance of chlorophyll-a concentration.

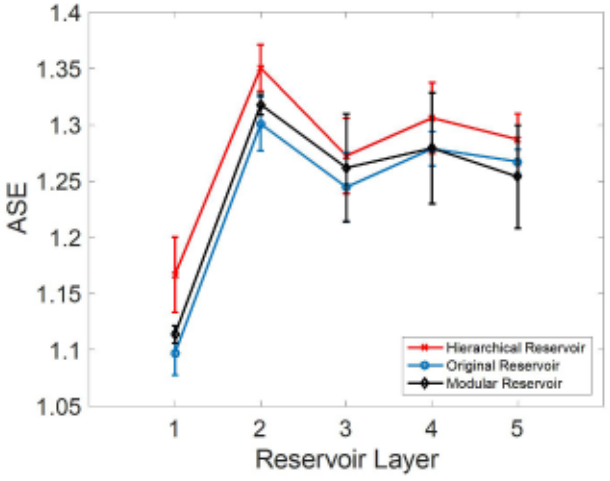


**Fig. 19.** Effects of different reservoir structure on the information content in chlorophyll-a prediction.

accuracy. The structural improvement of the reservoir is consistent with the actual prediction results, which reflects the effectiveness of the reservoir improvement.

According to the above experimental results, deep neural network architecture is an effective way to improve the performance when compared with shallow architecture. The increase of the number of layers does not significantly increase the complexity of the output weights, which is also the training advantage of the ESN model, particularly for the ESN model, when compared with the general multi-layer perceptron model. Combined with the elastic regularization method, the robustness of DHESN has been enhanced and the complexity reduced, which theoretically ensures the reliability and compactness of DHESN.

The theoretical analysis demonstrates the invariance of the hyper-parameter settings complexity before and after the improvement of the reservoir structure, i.e., two important hyperparameters, viz., spectral radius and sparsity, do not need to be pre-set. However, the sub-reservoir size, the number of sub-reservoirs, the number of reservoir layers, and the regularization parameters also need to be considered. Any further improvement in the efficiency of hyperparameter optimization is to be considered in the subsequent research.

## 5. Conclusion

The evolution process of algal blooms is characterized by nonlinear and time-varying characteristics. To overcome the lack of the ability of

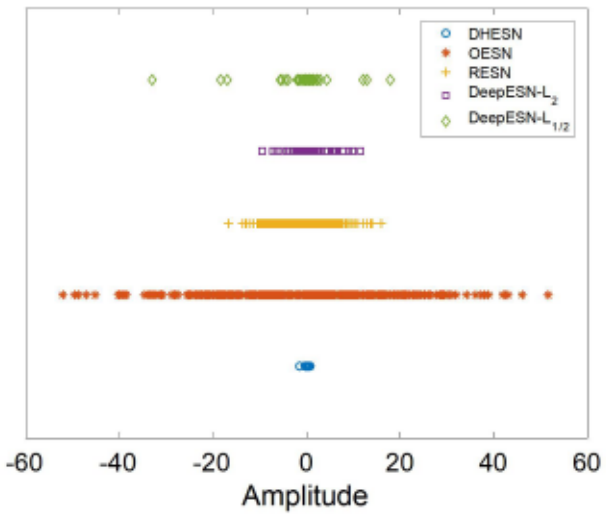


**Fig. 20.** Influence of different regularization methods on output weights in of chlorophyll-a prediction.

shallow neural networks to deal with nonlinear problems, a DHESN in this paper is proposed to predict the concentration of chlorophyll-a, the characteristic variable of algal bloom. Combined with the experiment results of henon mapping prediction, mackey–glass prediction and algal bloom prediction, the application advantages of DHESN are given:

1). Compared with variant ESNs and other typical networks, it is found that DHESN has lower average NRMSE and higher prediction accuracy. In the actual algal bloom dataset, the NRMSE of DHESN decreased by at least 34.21 % compared to shallow network, and 7.03 % compared to deep reservoir network.

2). Compared with other reservoir structures, DHESN has higher ASE, which can be used as an effective reservoir structure for further optimization.

3). Compared with other regularized ESN model, DHESN have a prunable and compressible output weight training process to obtain a compact structure. In the actual algal bloom dataset, compared with other pruning regularization methods, the output weight can be reduced by more than one times.

Therefore, DHESN can achieve high precision, which also indicates that this data-driven model is suitable for actual algal bloom prediction in lake and reservoir. In future research, we expect to optimize the topological structure between different reservoir layers and further explore the fusion of time series features in the reservoir network. It is expected that these modifications will further improve the prediction performance of the reservoir method in the application of algal bloom in lake reservoir.

## Declaration of interests

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

## CRediT authorship contribution statement

**Bo Hu:** Methodology, Writing – original draft. **Huiyan Zhang:** Conceptualization, Writing – review & editing. **Xiaoyi Wang:** Funding acquisition, Supervision. **Li Wang:** Software. **Jiping Xu:** Supervision. **Qian Sun:** Funding acquisition. **Zhiyao Zhao:** Investigation. **Lei Zhang:** Investigation.

## Declaration of Competing Interest

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

## Data availability

The data that has been used is confidential.

## Acknowledgement

This research was financially supported by National Social Science Fund of China (19BGL184), National Natural Science Foundation of China (61903008, 61802010, 61703008), Beijing Municipal Natural Science Foundation (4194074),Beijing excellent talent training support project for young top-notch team (2018000026833TD01), Beijing Institute of Fashion Technology High Level Talent Project Project (BIFTGCC202301) and the 2020 Graduate Research Capacity Improvement Program of Beijing Technology and Business University. Those supports are gratefully acknowledged.

## References

Bae, S., & Seo, D. (2021). Changes in algal bloom dynamics in a regulated large river in response to eutrophic status. *Ecological Modelling, 454*, 109590. https://doi.org/10.1016/j.ecolmodel.2021.109590

Beretta-Blanco, A., & Carrasco-Letelier, L. (2021). Relevant factors in the eutrophication of the Uruguay River and the Río Negro. *Science of the Total Environment, 761*, 143299. https://doi.org/10.1016/j.scitotenv.2020.143299

Chitsazan, M. A., Fadali, M. S., & Trzynadlowski, A. M. (2019). Wind speed and wind direction forecasting using echo state network with nonlinear functions. *Renewable Energy, 131*, 879 889. https://doi.org/10.1016/j.renene.2018.07.060

Cui, H., Feng, C., Chai, Y., Liu, R. P., & Liu, Y. (2014). Effect of hybrid circle reservoir injected with wavelet-neurons on performance of echo state network. *Neural Networks, 57*, 141 151. https://doi.org/10.1016/j.neunet.2014.05.013

Cui, J., Jin, Z., Wang, Y., Gao, S., Fu, Z., Yang, Y., & Wang, Y. (2021). Mechanism of eutrophication process during algal decomposition at the water/sediment interface. *Journal of Cleaner Production, 309*, 127175. https://doi.org/10.1016/j.jclepro.2021.127175

Dutoit, X., Schrauwen, B., Campenhout, J. V., Stroobandt, D., Brussel, H. V., & Nuttin, M. (2009). Pruning and regularization in reservoir computing. *Neurocomputing, 72*(7 9), 1534 1546. https://doi.org/10.1016/j.neucom.2008.12.020

Ferreira, A. A., Ludermir, T. B., & De Aquino, R. R. (2013). An approach to reservoir computing design and training. *Expert systems with applications, 40*(10), 4172 4182. https://doi.org/10.1016/j.eswa.2013.01.029

Gallicchio, C., & Micheli, A. (2011). Architectural and Markovian factors of echo state networks. *Neural Networks, 24*(5), 440 456. https://doi.org/10.1016/j.neunet.2011.02.002

Gallicchio, C., Micheli, A., & Pedrelli, L. (2017). Deep Reservoir Computing: A Critical Experimental Analysis. *Neurocomputing, 268*, 87 99. https://doi.org/10.1016/j.neucom.2016.12.089

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation, 18*(7), 1527 1554. https://doi.org/10.1162/neco.2006.18.7.1527

Henon, M. (1976). A two-dimensional mapping with a strange attractor. *Communications in Mathematical Physics, 50*, 69 77. https://doi.org/10.1007/BF01608556

Huang, J., Zhang, Y., Huang, Q., & Gao, J. (2018). When and where to reduce nutrient for controlling harmful algal blooms in large eutrophic lake Chaohu, China? *Ecological Indicators, 89*, 808 817. https://doi.org/10.1016/j.ecolind.2018.01.056

Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science, 304*(5667), 78 80. https://doi.org/10.1126/science.1091277

Jaeger, H., Maass, W., & Principe, J. (2007). Special issue on echo state networks and liquid state machines. *Neural Networks, 20*(3), 287 289. https://doi.org/10.1016/j.neunet.2007.04.001

Jarvis, S., Rotter, S., & Egert, U. (2010). Extending Stability Through Hierarchical Clusters in Echo State Networks. *Frontiers in Neuroinformatics, 4*, 1 10. https://doi.org/10.3389/fninf.2010.00011

Kaviani, S., & Sohn, I. (2021). Application of complex systems topologies in artificial neural networks optimization: An overview. *Expert Systems with Applications, 180*, 115073. https://doi.org/10.1016/j.eswa.2021.115073

Kim, T., Shin, J., Lee, D., Kim, Y., Na, E., Park, J. H., Cha, Y. (2022). Simultaneous feature engineering and interpretation: Forecasting harmful algal blooms using a deep learning approach. *Water Research, 215*, 118289. https://doi.org/10.1016/j.watres.2022.118289

Lead, N., Magnuson, J., Carpenter, S., & Stanley, E. (2020). North Temperate Lakes LTER: High Frequency Data: Meteorological, Dissolved Oxygen, Chlorophyll, Phycocyanin - Lake Mendota Buoy 2006 - current ver 31. *Environmental Data Initiative*. https://doi.org/10.6073/pasta/c03b39550e79d002d82a2281f8546c78

Liang, N. Y., Huang, G. B., Saratchandran, P., & Sundararajan, N. (2006). A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on Neural Networks, 17*(6), 1411 1423. https://doi.org/10.1109/TNN.2006.880583

Liang, Y., Liu, C., Luan, X. Z., Leung, K. S., Chan, T. M., Xu, Z. B., & Zhang, H. (2013). Sparse logistic regression with a L1/2 penalty for gene selection in cancer classification. *BMC Bioinformatics, 14*(1), 1 12. https://doi.org/10.1186/1471-2105-14-198

Lin, S. S., Shen, S. L., Zhou, A., & Lyu, H. M. (2021). Assessment and management of lake eutrophication: A case study in Lake Erhai. *China. Science of the Total Environment, 751*, 141618. https://doi.org/10.1016/j.scitotenv.2020.141618

Liu, M., He, J., Huang, Y., Tang, T., Hu, J., & Xiao, X. (2022a). Algal bloom forecasting with time-frequency analysis: A hybrid deep learning approach. *Water Research, 118591*. https://doi.org/10.1016/j.watres.2022.118591

Liu, Y., Shan, L., Yu, D., Zeng, L., & Yang, M. (2022b). An echo state network with attention mechanism for production prediction in reservoirs. *Journal of Petroleum Science and Engineering, 209*, 109920. https://doi.org/10.1016/j.petrol.2021.109920

Ozturk, M. C., Xu, D., & Principe, J. C. (2007). Analysis and design of echo state networks. *Neural computation, 19*(1), 111 138. https://doi.org/10.1162/neco.2007.19.1.111

Qiao, J., Li, F., Han, H., & Li, W. (2016). Growing echo-state network with multiple subreservoirs. *IEEE Transactions on Neural Networks and Learning Systems, 28*(2), 391 404. https://doi.org/10.1109/TNNLS.2016.2514275

Ribeiro, G. T., Santos, A. A. P., Mariani, V. C., & dos Santos Coelho, L. (2021). Novel hybrid model based on echo state neural network applied to the prediction of stock price return volatility. *Expert Systems with Applications, 184*, 115490. https://doi.org/10.1016/j.eswa.2021.115490

Schreiber, T. (2000). Measuring information transfer. *Physical review letters, 85*(2), 461 464. https://doi.org/10.1103/PhysRevLett.85.461

Shen, L., Chen, J., Zeng, Z., Yang, J., & Jin, J. (2018). A novel echo state network for multivariate and nonlinear time series prediction. *Applied Soft Computing, 62*, 524 535. https://doi.org/10.1016/j.asoc.2017.10.038

Song, Q., & Feng, Z. (2010). Effects of connectivity structure of complex echo state network on its prediction performance for nonlinear time series. *Neurocomputing, 73*(10 12), 2177 2185. https://doi.org/10.1016/j.neucom.2010.01.015

Sun, C., Wang, S., Wang, H., Hu, X., Yang, F., Tang, M., Zhong, J. (2022). Internal nitrogen and phosphorus loading in a seasonally stratified reservoir: Implications for eutrophication management of deep-water ecosystems. *Journal of Environmental Management, 319*, 115681. https://doi.org/10.1016/j.jenvman.2022.115681

Thakkar, A., & Chaudhari, K. (2021). A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions. *Expert Systems with Applications, 177*(2), 114800. https://doi.org/10.1016/j.eswa.2021.114800

Vinçon-Leite, B., & Casenave, C. (2019). Modelling eutrophication in lake ecosystems: A review. *Science of the Total Environment, 651*, 2985 3001. https://doi.org/10.1016/j.scitotenv.2018.09.320

Wang, L., Zhang, T., Jin, X., Xu, J., Wang, X., Zhang, H., Xie, Y. (2020). An approach of recursive timing deep belief network for algal bloom forecasting. *Neural Computing and Applications, 32*(1), 163 171. https://doi.org/10.1007/s00521-018-3790-9

Xu, M., Han, M., Qiu, T., & Lin, H. (2018). Hybrid regularized echo state network for multivariate chaotic time series prediction. *IEEE transactions on Cybernetics, 49*(6), 2305 2315. https://doi.org/10.1109/TCYB.2018.2825253

Yang, C., Qiao, J., Wang, L., & Zhu, X. (2019). Dynamical regularized echo state network for time series prediction. *Neural Computing and Applications, 31*(10), 6781 6794. https://doi.org/10.1007/s00521-018-3488-z

Yao, X., Wang, Z., & Zhang, H. (2019). A novel photovoltaic power forecasting model based on echo state network. *Neurocomputing, 325*, 182 189. https://doi.org/10.1016/j.neucom.2018.10.022

Yildirim, H., & Ozkale, M. R. (2019). The performance of ELM based ridge regression via the regularization parameters. *Expert Systems with Applications, 134*, 225 233. https://doi.org/10.1016/j.eswa.2019.05.039

Yu, P., Gao, R., Zhang, D., & Liu, Z. P. (2021). Predicting coastal algal blooms with environmental factors by machine learning methods. *Ecological Indicators, 123*, 107334.

Zhang, H., Hu, B., Wang, X., Xu, J., Wang, L., Sun, Q., & Wang, Z. (2021). Self-organizing deep belief modular echo state network for time series prediction. *Knowledge-Based Systems, 222*, 107007. https://doi.org/10.1016/j.knosys.2021.107007

Zhang, H., Hu, B., Wang, X., Wang, L., Xu, J., Sun, Q., & Zhao, Z. (2022). An echo state network based adaptive dynamic programming approach for time-varying parameters optimization with application in algal bloom prediction. *Applied Soft Computing, 122*, 108796. https://doi.org/10.1016/j.asoc.2022.108796