A hybrid model of empirical wavelet transform and extreme learning machine for dissolved oxygen forecasting

Juan Huan*, Weijian Cao and Yuwan Gu

Changzhou University,
Changzhou, Jiangsu, China
Email: huanjuan@cczu.edu.cn
Email: caoweijian320623@outlook.com
Email: gywss0724@126.com
*Corresponding author

Yilin Qin

Department of Creative Information,
Changzhou Technical Institute of Tourism and Commerce University,
Changzhou, Jiangsu, China
Email: qinyilin101@163.com

Abstract: The accurate predicting trend of dissolved oxygen (DO) can reduce the risks to aquaculture, so a combined nonlinear prediction model based on empirical wavelet transform (EWT) and extreme learning machine (ELM) optimised by adaptive disturbance particle swarm optimisation (ADPSO) is proposed. First of all, DO series are decomposed into a term of relatively subsequence by EWT, secondly, the decomposed components are reconstructed using the C-C method, and thirdly an ELM prediction model of every component is established. At last, the predicted values of DO datasets are calculated by using RBF to reconstruct the forecasting values of all components. This model is tested in the special aquaculture farm in Liyang City, Jiangsu Province. Results indicate that the proposed prediction model of EWT-ELM has better performance than WD-ELM, EMD-ELM, ELM and EWT-BP. The research shows that the combined forecasting model can effectively extract the sequence characteristics, and can provide a basis for decision-making management of water quality, which has certain application value.

Keywords: water quality prediction; hybrid model; dissolved oxygen; empirical wavelet transform; EWT.

Reference to this paper should be made as follows: Huan, J., Cao, W., Gu, Y. and Qin, Y. (2020) 'A hybrid model of empirical wavelet transform and extreme learning machine for dissolved oxygen forecasting', *Int. J. Embedded Systems*, Vol. 13, No. 1, pp.9–17.

Biographical notes: Juan Huan received her BS and MS degrees in Computer Science from Jiangsu University, Zhenjiang, China in 2004 and 2007, respectively. Currently, she is an Associate Professor of Computer Science with School of Information Science and Engineering, Changzhou University. She is a member of China Computer Federation (CCF). Her research interests are intelligent prediction algorithms.

Weijian Cao is a Master student at the School of Information, Changzhou University, his research interest is image processing. He received his Bachelor's degree in Electronic Information Engineering from Changzhou University in 2016.

Yuwan Gu received her BS and MS degrees in Computer Science from Changzhou University, Changzhou, China in 2006 and 2010, respectively, and PhD in Agricultural Electrification and Automation from Jiangsu University, Zhenjiang, China in 2016. Currently, she is a Lecturer of Computer Science with School of Information Science and Engineering, Changzhou University. She is a member of China Computer Federation (CCF). Her research interests are intelligent algorithm, distributed computing and parallel parsing.

Yilin Qin is a Professor of Changzhou Technical Institute of Tourism and Commerce University. He is a member of China Electrical Engineering Committee. His current research interests in internet of things.

This paper is a revised and expanded version of a paper entitled 'A hybrid model of empirical wavelet transform and extreme learning machine for dissolved oxygen forecasting' presented at 2018 IEEE International Conference on Cyber, Physical and Social Computing (CPSCom 2018), Halifax, Canada, 30 July–3 August 2018.

1 Introduction

The boom in big data and cloud computing has also promoted the development of modern agriculture (Wang et al., 2017b). Water quality deterioration is the primary factor that induces the outbreak of aquatic disease and even large quantities of death. The concentration of dissolved oxygen (DO) has an important impact on the yield of aquaculture (Jinyou et al., 2015). Photosynthesis, weather, metabolic rate, and energy expenditure are affected by DO concentration. When the dissolved oxygen is less than 3 mg/L, it will adversely affect the ingestion, digestion and health of aquatic products (Huan et al., 2018). As a result, it is necessary and important for the development of modern aquaculture to create a high-precision, practical DO forecasting model (Li et al., 2018).

In recent years, many researchers at home and abroad have put forward some methods for DO prediction and achieved good developments. Liu et al. (2012) proposed the optimisation methods for least squares support vector regression (LSSVR) parameters. The model was used for water quality prediction and achieved good results. Usman and Caterina (2017) established fuzzy neural network based on possibility theory for dissolved oxygen prediction. Wu et al. (2018) used the PSO-BPANN algorithm combined with kriging interpolation to propose a three-dimensional prediction model for DO content. Wu et al. (2018) proposed a DO forecasting model for water quality monitoring system based on BP neural network.

The previous researchers' methods can be broadly divided into single forecasting model and assorted forecasting model. In the assorted prediction model, they introduced empirical mode decomposition (EMD) or wavelet decomposition to remove noise. However, wavelet analysis is a kind of non-adaptive decomposition method which needs to set the basis function in advance and relies on subjective experience. EMD has a serious mode aliasing phenomenon, which is easy to generate false modal components. Moreover, the number of decomposition components and the calculation scale is large, which increases the calculation amount of the combined forecasting method (Longqin et al., 2016). Empirical wavelet decomposition is a signal-adaptive analysis method that combines the scientificity of wavelet transform with the adaptive advantages of EMD (Gilles, 2013). Empirical wavelet transform (EWT) is an adaptive decomposition method that established under the wavelet framework. Compared with EMD, it has the advantages of strong theory, small computation quantity, small number of decomposed modes, and no unexplainable false modal components. Li et al. (2017) put forward a model based on EWT, kernel mini max probability machine regression to forecast short-term photovoltaic power. The study of Hui et al. (2017) proposed an EWT and synchroextracting transform model, which was used to analyse time-frequency for hydrocarbon detection.

The rest of this paper is organised as follows: Section 2 describes EWT, extreme learning machine (ELM), adaptive disturbance particle swarm optimisation (ADPSO) and reconstruction of RBF neural network. Section 3 presents the detailed modelling steps of model. Section 4 gives the actual data and compares the performance of the proposed model with other prediction models. Section 5 discusses the conclusions of this article.

2 Materials and methodology

2.1 Empirical wavelet transform

EWT is a data analysis method combining EMD adaptive decomposition and wavelet transform theory. Compared with EMD method, EWT method has better effect and avoids the problem of mode aliasing (Thirumala et al., 2015). The hypothesis says that the Fourier spectrum is divided into N consecutive sections and every section is given as $\Lambda_n = [\omega_{n-1}, \omega_n]$. Where $\omega_n(\omega_0 = 0, \omega_n = \pi)$ is the midpoint between the two large phase-adjacent points in the Fourier spectrum, $2\tau_n$ is the width of the transition region. τ_n has several possible choices, i.e., $\tau_n = \gamma\omega_n(0 < \gamma < 1)$, therefore, the expressions for scaling function $\theta_n'(\omega)$ and the empirical wavelets $\eta_n'(\omega)$ are defined as (Thirumala et al., 2015):

$$\theta_n'(\omega) = \begin{cases} 1 & |\omega| \le (1-\gamma)\omega_n \\ \cos\left[\frac{\pi}{2}\delta\left(\frac{1}{2\gamma\omega_n}(|\omega| - (1-\gamma)\omega_n)\right)\right] & (1-\gamma)\omega_n \le |\omega| \le (1-\gamma)\omega_n \\ 0 & \text{other} \end{cases} \quad (1)$$

$$\eta_n'(\omega) = \begin{cases} 1 & (1+\gamma)\omega_n \le |\omega| \le (1-\gamma)\omega_{n+1} \\ \cos\left[\frac{\pi}{2}\delta\left(\frac{1}{2\gamma\omega_{n+1}}(|\omega| - (1-\gamma)\omega_{n+1})\right)\right] & (1-\gamma)\omega_{n+1} \le |\omega| \le (1+\gamma)\omega_{n+1} \\ \sin\left[\frac{\pi}{2}\delta\left(\frac{1}{2\gamma\omega_n}(|\omega| - (1-\gamma)\omega_n)\right)\right] & (1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n \\ 0 & \text{other} \end{cases} \quad (2)$$

where the function $\delta(x)$ is as follows:

$$\delta(x) = x^4(35 - 84x + 70x^2 - 20x^3) \quad (3)$$

The specific steps of EWT decomposition are as follows:

Step 1 The FFT is applied to the analysis signal $f(t)$ in order to obtain the spectrum $F(\omega)$.

Step 2 According to estimating the relative best number N of modes, guarantee the largest N–1 maxima in $|F(\omega)|$, and sort them in descending order.

Step 3 Calculate the centre between two consecutive vertices and use it as the boundary ω_n, then, get N–1 boundaries except 0 and π.

Step 4 After constructing a set of empirical wavelets, the detailed coefficients $W_f^{\varepsilon}(n,t)$ and approximate coefficients $W_f^{\varepsilon}(0,t)$ of $f(t)$ in the EWT are defined by using a method similar to the traditional wavelet transform. The detailed coefficient is generated by the empirical wavelet function and signal inner product, the approximate coefficient is generated by the inner product of scale function and signal inner product. The reconstruction can be given by:

$$f(t) = W_f^{\varepsilon}(0,t) * \theta_1(t) + \sum_{n=1}^{N} W_f^{\varepsilon}(n,t) * \eta_n(t) \tag{4}$$

where * expresses the convolution operation.

Step 5 By the above formalism, the empirical mode $f_k(t)$ can be expressed as:

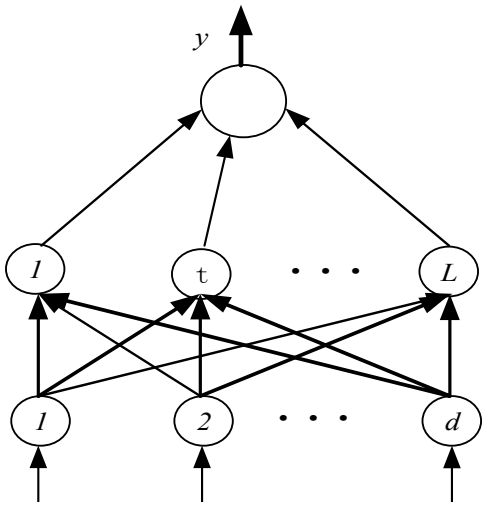
$$\begin{aligned} f_0(t) &= W_f^{\varepsilon}(0,t) * \theta_1(t) \\ f_k(t) &= W_f^{\varepsilon}(k,t) * \eta_k(t) \end{aligned} \tag{5}$$

2.2 Extreme learning machine

Given the training set $\{x_i, t_i\}$, i = 1, 2,…, N, where x_i is the i^{th} input vector and t_i is the i^{th} target. When the excitation function is $f(x)$, this means that (Di et al., 2017):

$$H\beta = T \tag{6}$$

Figure 1 ELM framework



The three layer structure network is effective for any function approximation problem. The ELM network is made up of three independent layers, namely, input layer, hidden layer and output layer. The typical structure of the network is shown in Figure 1, where the hidden layer output matrix is defined as $H = \{h_{ij}\}$ (i = 1,…, N and j = 1,…, M), $h_{ij} = g(w_j^T \cdot X_i + b_j)$ denotes the output of the hidden node j with respect to x_i, $g(\cdot)$ is activation function, w_j is the weight vector for connecting hidden and input nodes, and b_j is the deviation of hidden node j; β is output weight matrix and T denotes the matrix of targets.

ELM works as follows. At first, it randomly generated input weights and hidden biases. Secondly, it determines the output weight, connects the hidden layer with the output layer, and finds the least squares solution of the linear system given by equation (6) simply. This solution is as follows:

$$\beta^* = H^+ T \tag{7}$$

where H^+ is Moore-Penrose generalised inverse of hidden layer output matrix H.

2.3 Phase space reconstruction

Phase space reconstruction (PSR) is the premise and foundation of analysing chaotic time series, which can transform one-dimensional time series into multi-dimensional phase space, and transform chaotic time series prediction into short-term evolution problem of phase space, so as to determine the input and output of the model reasonably, which is beneficial to improve the accuracy of model prediction.

According to the Takens embedding delay theorem, if the embedding dimension m and time delay τ are properly selected, the reconstruction of phase space is equivalent to the original system and has the same topology.

Set the time series of one-dimensional traffic flow with length N as $x = [x_i | \ i = 1, 2, \ldots, N]$, so the reconstructed phase space is

$$X = \left\{ X_i \mid X_i = \left[x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau} \right], i = 1, 2, \ldots M \right\} \tag{8}$$

where $M = N - (m - 1)\tau$ is the number of phase. In the reconstructed phase space, each phase point includes m elements, and the interval between each two adjacent elements is τ.

The selection of embedding dimension m and time delay τ determines the quality of PSR. The traditional view holds that m and τ are mutually independent and need to be solved individually or sequentially. Recent studies have shown that m and τ are closely related and the relationship between the two parameters can be established through the delay window:

$$\tau_w = (m-1)\tau \tag{9}$$

The C-C algorithm uses the correlation integral to form the statistics, and obtains τ_w and τ simultaneously through the relation between the statistics and time delay, and then gets the embedded dimension. The correlation integral is defined as follows:

$$C(m, r, t) = \frac{2}{M(M-1)} \sum_{1 \le i \le j \le M} H\left(r - d_{ij}\right) \tag{10}$$

where M is phase number; r is neighbourhood radius; d_{ij} is the Euclidean distance between two points in phase space, $d_{ij} = \|X_i - X_j\|$, $H(z)$ is the Heaviside step function, the definition for:

$$H(z) = \begin{cases} 1, & z > 0 \\ 0, & z \le 0 \end{cases} \tag{11}$$

The time series $\{x(t), t = 1, 2, \ldots, N\}$ is divided into disjoint sub-sequences, and for each sub-sequence, there is

$$s(m, N, r, t) = \frac{1}{t} \sum_{s=1}^{t} \left[C_s\left(n, N/t, r, t\right) - C_s^m\left(1, N/t, r, t\right) \right] \tag{12}$$

You take the average of all $S(m, N, r, t)$:

$$\overline{S} = \frac{1}{MJ} \sum_{m=1}^{M} \sum_{j=1}^{J} S\left(m, r_j, t\right) \tag{13}$$

where M is the number of embedded dimensions, J is the number of r, and $\Delta S(m, t)$ is defined as the maximum deviation of r, then

$$\Delta S(m, t) = \max\left\{S\left(m, r_j, t\right)\right\} - \min\left\{S\left(m, r_j, t\right)\right\} \tag{14}$$

According to the statistical conclusions about several important asymptotic distributions in literature (Ray, 1993), $m = 2, 3, 4, 5$, $r_i = i\sigma/2$, $i = 1, 2, 3, 4$, is often taken in practice.

Where is the standard deviation of the sequence, the following three statistics are calculated respectively:

$$\overline{S}(t) = \frac{1}{16} \sum_{m=2}^{5} \sum_{j=1}^{4} S\left(m, r_j, t\right) \tag{15}$$

$$\Delta\overline{S}(t) = \frac{1}{4} \sum_{m=2}^{5} S(m, t) \tag{16}$$

$$S_{cor}(t) = \Delta\overline{S}(t) + \left|\overline{S}(t)\right| \tag{17}$$

Graph $\overline{S}(t)$ and t, $\Delta\overline{S}(t)$ and t, $S_{cor}(t)$ and t, where the time t corresponding to the first zero of $\overline{S}(t)$ or the first minimum point of $\Delta\overline{S}(t)$ is the optimal time delay τ, the time t corresponding to the global minimum point of $S_{cor}(t)$ is the delay window τ_w, and the optimal embedding dimension $m = \tau_w/\tau + 1$.

2.4 Adaptive disturbance particle swarm optimisation

PSO method initialises a group of random particles and finds the optimal solutions through iteration in the first place (Fang et al., 2017). In each iteration process, the particles update them selves by tracking the two extreme values, one of which is particle optimal value named *pbest*, $p_{i,i} = (p_{i,1}\ p_{i,2} \cdots p_{i,d})$, and the other is group optimal value named *gbest*, P_g, when finding the two optimal values, particles update their location and speed according to the following formulas:

$$\begin{aligned} v_{i,j}(t+1) = {} & w v_{i,j}(t) + c_1 r_1 \left[p_{i,j} - x_{i,j}(t) \right] \\ & + c_2 r_2 \left[p_{g,j} - x_{i,j}(t) \right] \end{aligned} \tag{18}$$

$$x_{i,j}(t+1) = x_{i,j}(t) + v_{i,j}(t+1) \tag{19}$$

where w is inertia factor which is not less than 0, r_1, r_2 are random numbers obeying uniform distribution between 0 and 1, c_1, c_2 are learning factors usually both assigned as 2, x is a current position, and v is the velocity of the particle.

In order to avoid the premature convergence of the particle swarm optimisation itself, the paper used ADPSO (Zhao et al., 2010). During the whole process, all the particles are looking for the optimal solution. When approaching the extremum, the velocity of the particle is reduced to zero. At this time, the algorithm falls into the local extremum. Therefore, according to the location of the particles in the population, it can judge whether the population falls into local optimal solution. The population fitness variance d^2 can be defined as follows:

$$d^2 = \sum_{t=1}^{M} \left(\frac{f_e - f_{avg}}{f} \right)^2 \tag{20}$$

where M is population size, f_e is the fitness of particle e and f_{avg} is the current group average fitness. It uses a normalised scaling factor f to limit the variance of fitness. f can be defined as follows:

$$f = \begin{cases} \max\left|f_e - f_{avg}\right| & \max\left|f_e - f_{avg}\right| > 1 \\ 1 & \text{other} \end{cases} \tag{21}$$

The population fitness variance d^2 is going to get smaller and smaller because the fitness of different particles will be closer as the number of iterations increases. The smaller d^2 indicates that the group of particles aggregated closely and the closer the group is to convergence. When d^2 is less than the given threshold η, the population will fall into local optimum. At this point, it is to disturb the particles, let it continue to search for new locations. Modified the formula for updating the particle position can be defined as follows:

$$x_{i,d}^{t+1} = x_{i,d}^{t} - \theta x_{i,d}^{t+1} \tag{22}$$

where θ is the disturbance factor between 0 and 1 randomly.

3 Model and experiments

3.1 Dissolved oxygen analysis and the modelling process

In this paper, experimental data came from the special aquaculture farms in Liyang City, Jiangsu province, China. The experiment area is equipped with a wireless monitoring system for aquaculture. The data of dissolved oxygen in this

paper is from the database of this system. The system is equipped with dissolved oxygen sensor, which can collect data in real time and send it to the monitoring centre. The monitoring centre will collect the dissolved oxygen values and send them to the server through GPRS, so that users can monitor them in real time with their mobile phones and computers. In addition, it is equipped with oxygen pump and other equipment, so as to achieve water quality control (Djamila et al., 2018).

Test data is collected once every half hour and a total of 1,728 data are selected. 1,680 data, from 25 May 2017 to 28 June 2017 were selected as the training set, and 48 data on 29 June 2017 were selected as the test set. The total dissolved oxygen time sequence of ten days is displayed in Figure 1.

Figure 2 DO time series (see online version for colours)

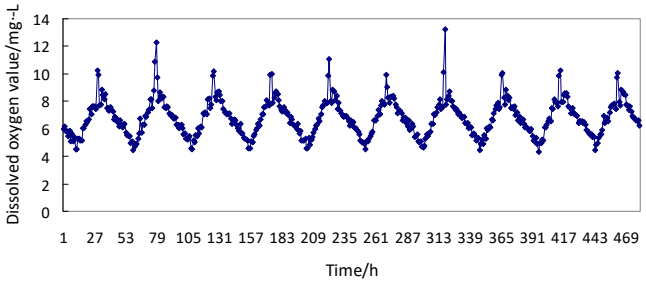
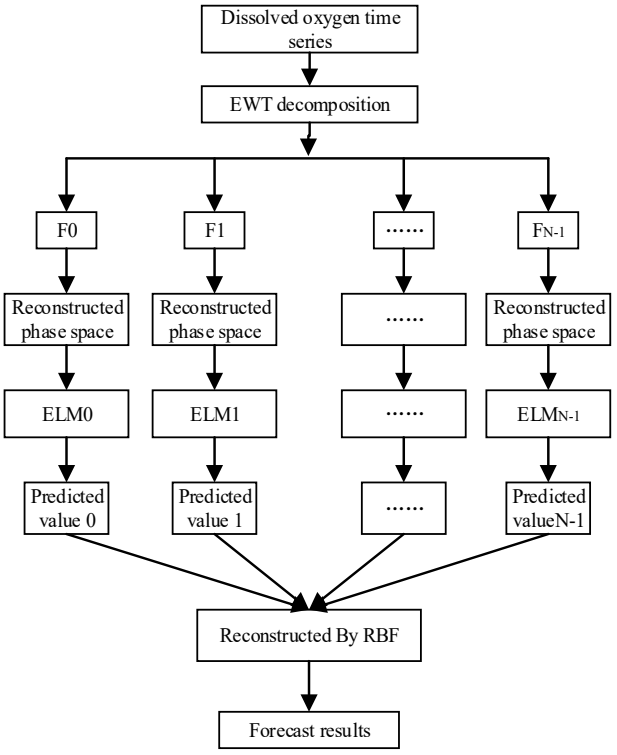


Figure 3 Predicting idea of EWT-ELM method



These man-made or natural reasons mainly include the work of oxygen pumps and the increase or decrease of solar radiation. The change of dissolved oxygen level in ponds caused by the above reasons is reflected in the whole dissolved oxygen time series. The dissolved oxygen level changes regularly with time. As shown in Figure 2, the ten-day time series of DO has periodic and nonlinear characteristics. The range of light intensity varies greatly, reaching tens of thousands of lux during the day and only 0 lux at night, so they are less correlated with dissolved oxygen. In addition, environmental factors such as chlorophyll, turbidity, respiration of animals and plants are very different, not with dissolved oxygen is highly correlated. Therefore, using these factors as input to the model will reduce the accuracy of the model prediction.

From Figure 2, since the dissolved oxygen system is typically nonlinear and non-stationary, series prediction has insurmountable difficulties. For this reason, the EWT-ELM algorithm is adopted, according to the principle of decomposition and integration. The framework of a EWT-ELM model is in Figure 3.

3.2 EWT decomposition

Dissolved oxygen sequence is decomposed into F0, F1, and F2 component by EWT. The results of the decomposition are shown in Figure 4. There are obvious characteristic differences among the decomposition components of EWT. First of all, F0 (experiential scale component) is a stable component, and almost 100% accuracy can be obtained. The fluctuation of F1 is relatively regular, and high prediction accuracy can be obtained. The change of component F2 is more drastic, which will generate certain prediction error, but its amplitude is small, which will not bring too large cumulative error to the final prediction result.

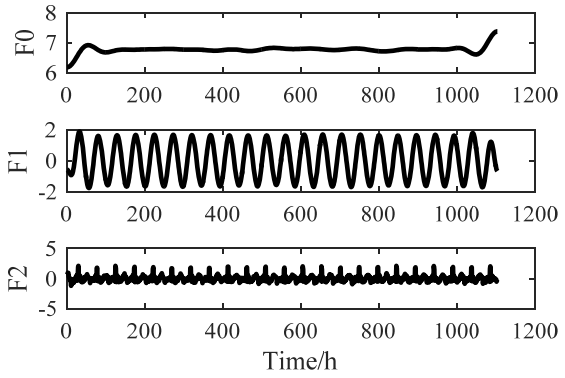
Step 1 EWT decomposition of dissolved oxygen original sequence. EWT was decomposed for the original time series according to the relevant steps, which was decomposed into *x* intrinsic mode components with different characteristics, F1~Fx.

Step 2 PSR was performed on the input vector according to the periodic determination of the embedding dimension and the delay time.

Step 3 Modelling ELM. ELM was selected for modelling training and prediction.

Step 4 Prediction result reconstruction. RBF network is used to perform nonlinear superposition.

Figure 4 The result of EWT decomposition

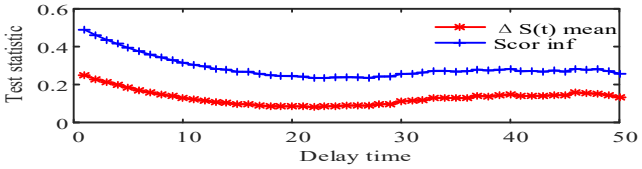


3.3 ELM prediction

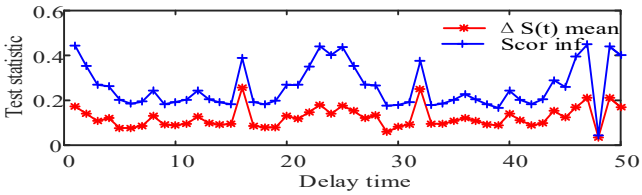
Each component after decomposition is reconstructed in phase space by C-C method. After PSR, the corresponding ELM prediction model is established for each component.

Parameter estimation of F0, F1 and F2 is carried out by using C-C method (See the literature 20 for specific steps). $\Delta\overline{S}(t)$, $S_{cor}(t)$ curves are shown in Figure 5.

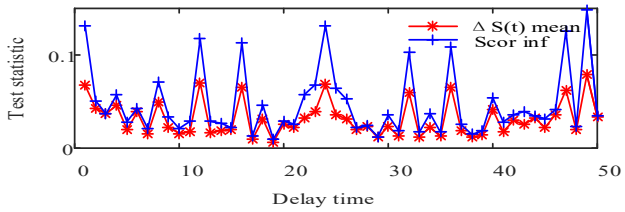
Figure 5 (a) Delay time and embedding dimension of F0 calculated by C-C method (b) Delay time and embedding dimension of F1 calculated by C-C method (c) Delay time and embedding dimension of F2 calculated by C-C method (see online version for colours)



(a)



(b)



(c)

$\Delta\overline{S}(t)$ and $S_{cor}(t)$ are the statistics of C-C method. The optimal delay time is the first local minimum point of the curve of $\Delta\overline{S}(t)$ and the global minimum value of the curve $S_{cor}(t)$ is the optimal window time delay.

To observe Figure 5, the first local minimum of F0's $\Delta\overline{S}(t)$ is 22, so F0's τ = 22, and the global minimum of F0's $S_{cor}(t)$ is 22, so F0's τ_w = 22. From this, we can also know that F1's τ = 3, F1's τ_w = 48 and F2's τ = 3, F2's τ_w = 19. According to formula (9), the embedded dimension can be obtained (Gu and Liao, 2017). Their results show in Table 1.

In this paper, to optimise the input weight and threshold of ELM, ADPSO is used in this paper. Taking the input weight and threshold of ELM as particles of ADPSO, the training sample mean square error is taken as the adaptive value function of ADPSO. The precision of predictive value increases with the decrease of adaptive value, so that the input weight and threshold value are better (Wang et al., 2017a).

The algorithm steps of ELM optimised by ADPSO are as below:

Step 1 Set up the ELM topology. Then initialise and the weights and thresholds are obtained by random training, and the reference weights and thresholds are taken as the optimal range of particle velocity and position.

Step 2 Select the appropriate parameters. Population size M = 50, maximum number of iterations T = 200, axial weight w = 0.6, and learning factor c_1 = 2, c_2 = 2.

Step 3 Determine the fitness function to calculate the fitness of each particle, and then find the individual and global extremum of each particle.

Step 4 Update the velocity and position of the particle.

Step 5 The algorithm can only be stopped when the precision or number of iterations has reached the desired value; otherwise, Step 2 should be skipped and continued until the demand is reached.

Step 6 Solve the model to get weights and thresholds and train the neural network.

Table 1 Results of PSR based on C-C method

Component	*Embedded dimension*	*Delay time*	*Hidden layer neurons*
F0	2	22	19
F1	17	3	15
F2	8	3	14

3.4 The reconstruction of the RBF neural network

After each component is predicted, the predicted results need to be reconstructed. RBF neural network is a feed forward network with a three-tier structure, namely input layer, hidden layer and output layer (Fei and Wu, 2017; Deng et al., 2018). The input of RBF neural network consists of predicted values of each component.

3.5 Error evaluation

To evaluate the forecasting accuracy, the four error indicators are selected, namely root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE) and decision coefficient (R2). The formulas are as follows.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - y_i^*\right)^2} \tag{23}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - y_i^*}{y_i}\right| \tag{24}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - y_i^*\right| \tag{25}$$

$$R^2 = \frac{\left(N\sum_{i=1}^{N} y_i y_i^* - \sum_{i=1}^{N} y_i^*\right)^2}{\left[N\sum_{i=1}^{N} y_i^{*2} - \left(\sum_{i=1}^{N} y_i^*\right)^2\right]\left[\left[N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right]\right]} \tag{26}$$

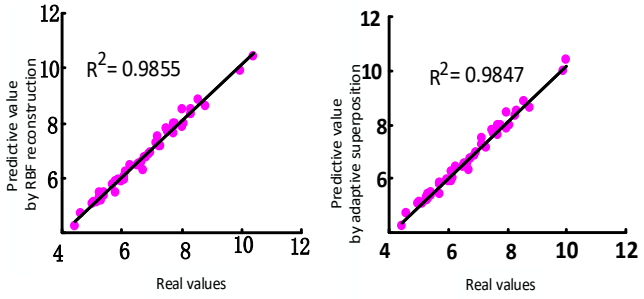
where y_i^* is predictive value, y_i is the real value, and the total number is defined as N.

4 Results and discussion

4.1 Comparison of reconstruction of results

The nonlinear superposition effect of RBF is tested by the determining coefficient. The comparison between the forecasting values of direct stacking and the real value are shown in Figure 6(a), and the comparison between the forecasting values of RBF are displayed in Figure 6(b).

Figure 6 (a) Predicted results by RBF reconstruction (b) Predicted results by adaptive superposition (see online version for colours)



(a) (b)

We can see from Figure 5 that the determining coefficients of predicted values, which are obtained by nonlinear superposition of RBF neural network, are higher than those obtained by adaptive superposition. So the effectiveness of the proposed RBF nonlinear superposition is also demonstrated.

4.2 Comparison of the EWT-ELM, WD-ELM, EMD-ELM, ELM and EWT-BP models

In order to verify the effective of the EWT-ELM prediction model, wavelet decomposition-extreme learning machine (WD-ELM), empirical modal decomposition – extreme learning machine (EMD-ELM), ELM and EWT-BP was compared with EWT-ELM.

Comparison method 1, the WD-ELM model, is that wavelet decomposition divides the dissolved oxygen sequence into four sub-sequences, each sub-sequence is established by ELM, and finally superimposes and the results of each subsequence are superimposed. The number of hidden layer neurons in four sub-sequences is 20, 14, 13 and 15, respectively.

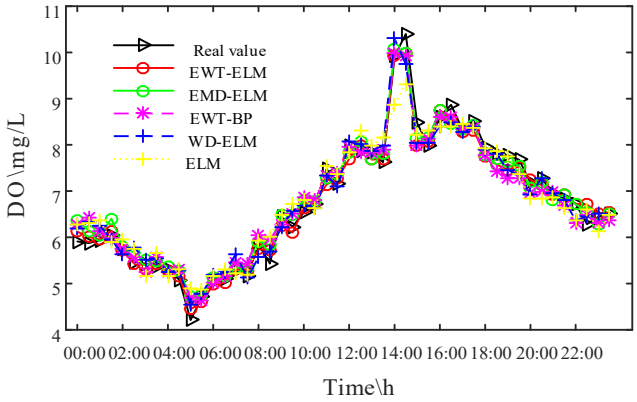
Comparison method 2, EMD-ELM model, can decompose the dissolved oxygen sequence with EMD, and get seven sub-sequences, and then model each sub-sequence. The number of hidden layer neurons in seven sub-sequences is 18, 14, 16, 19, 15, 20 and 20.

Comparison method 3, ELM model, does not deal with the de-noising of dissolved oxygen sequence and directly establishes the prediction model. Its number of hidden layer neurons is 28.

Comparison method 4, EWT-BP model, is that the dissolved oxygen sequence is decomposed into three components by EWT, and BP prediction model was established for each component. The number of hidden layer nodes of three components respectively is 3, 3 and 4.

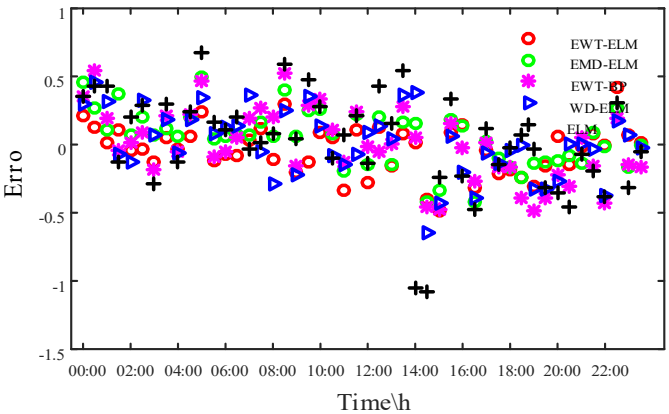
The predicted results are displayed in Figure 7.

Figure 7 Forecasting curve of five prediction models (see online version for colours)



As seen in Figure 6, it is obvious that the prediction curve of EWT-ELM is the closest to the real value curve, followed by WD-ELM, EMD-ELM and EWT-BP, while the prediction curve of single ELM is the furthest. Therefore, the proposed EWT-ELM model is better than the other models, as the EWT-ELM model has the best effect.

Figure 8 Error in forecasting results (see online version for colours)



In order to observe the change of error more directly, the prediction error of five models is listed as shown in Figure 8. The error fluctuation of EWT-ELM is smaller than

that of the other four models, which shows the high prediction accuracy of the model proposed in this paper.

Table 2 Error index of prediction results

Prediction model	R^2	*RMSE*	*MAE*	*MAPE*
EWT-ELM	0.9855	0.1782	0.1415	0.0035
WD-ELM	0.9782	0.2335	0.1887	0.0008
EMD-ELM	0.9824	0.2114	0.1732	0.0047
EWT-BP	0.9706	0.2487	0.2014	0.0023
ELM	0.9684	0.2532	0.2097	0.0016

In this paper, RMSE, MAPE, MAE and R^2 of the five prediction models are calculated, as shown in Table 2. By calculating the error index, the evaluation of the model is more objective and accurate.

- In order to verify the effect of denoising by decomposition, EWT-ELM and undecomposed ELM are compared. R^2 rose from 0.9684 to 0.9855 and RMSE dropped from 0.2532 to 0.1782. It shows that EWT decomposition is effective and has good effect on noise reduction.
- By comparing EWT-ELM with WD-ELM and EMD-ELM, the denoising effect of EWT was verified. The RMSE values are 0.1782, 0.2335 and 0.2114 for EWT-ELM, WD-ELM and EMD-ELM, respectively. The MAPE values of EWT-ELM, WD-ELM and EMD-ELM are 0.0035, 0.0008 and 0.0047, respectively. The MAE and R^2 for these models are 0.1415, 0.1887, 0.1732 and, 0.9855, 0.9782, 0.9824, respectively. The R^2 of EWT-ELM is higher than the other two models and With the exception of MAPE, RMSE index and MAE index of EWT-ELM is lower than the other two models. This shows that EWT can overcome the shortcomings of EMD and WD and has a good effect.
- To verify the predictive performance of ELM, EWT-ELM and EWT-BP were compared. RMSE, MAPE, MAE and R^2 of EWT-ELM are 0.1782, 0.0035, 0.1415, 0.9855, respectively and these four indexes of EWT-BP are 0.2487, 0.0023, 0.2014, and 0.9706, respectively. Therefore the prediction effect of ELM surpasses of BP neural network.

The MAE and RMSE of the model proposed in this paper is the smallest and R^2 is the biggest, which can be shown that the predicted value is closer to the true value. But our proposed model's MAPE is smaller than WD-ELM's R^2, which indicates that the predicted value of the WD-ELM model is closer to the true value than the predicted value of the proposed model. In brief, EWT-ELM model shows more obvious advantages, and it has the best prediction precision and generalisation performance, which is very suitable for nonlinear prediction of dissolved oxygen in small sample and non-stable aquaculture water.

5 Conclusions

Water quality environment plays an important role in the healthy growth and quality of aquatic products. In this paper, a hybrid model, which combines EWT, PSR and ELM optimised by PSO, is proposed. Furthermore, by comparing EMD-ELM, WD-ELM, ELM and EWT-BP, we find that the EWT-ELM model is more suitable for forecasting. In this paper, EWT is used to decompose the dissolved oxygen time series in multi-scale, which not only avoids the problem of mode aliasing, but also reduces the mutual interference between the features of the original sequence. The nonlinear superposition of RBF is used to replace the simple adaptive superposition, and the better fitting effect is obtained.

Based on the predicted results, the system alerts the manager to emergency measures before the sudden change of the specific contingency plan. The system also automatically turns on the oxygenator, presetting the oxygenator's operating time based on the trend of dissolved oxygen.

There are still some shortcomings in this method. For example, EWT needs to set the number of split modes when decomposing the signal

In the future researches, we can apply the methods of big data and cloud computing to water quality prediction (Wang et al., 2017c, 2016; Fan et al., 2016). In addition, the monitoring system (Zhang et al., 2018) can monitor many kinds of data,and then water temperature, pH, wind, weather, seasons and other factors can be taken into account into the prediction model, and it is a further research problem to explore a new multidimensional signal processing and machine learning method to realise the self-adaptive multi-resolution feature extraction of water quality parameters and improve the accuracy of water quality prediction.

Acknowledgements

This study is funded by the Chinese National Natural Science Foundation (61803050)(61772090) and the 2016 annual Liyang key research and development project (modern agriculture) (LB2016003).

References

Deng, H., Li, H., Chang, S., Xu, J. and Li, H. (2018) 'Feature binding pulse-coupled neural network model using a double colour space', *International Journal of Computational Science and Engineering*, Vol. 16, No. 2, pp.95–102.

Di, P., Yuan, X., Wang, Y., Geng, Z. and Zhu, Q. (2015) 'Soft-sensing in complex chemical process based on a sample clustering extreme learning machine model', *IFAC-Papers On Line*, Vol. 48, No. 8, pp.801–806.

Djamila, B., Lynda, M. and Hafid, H. (2018) 'Exploiting node mobility for fault management in RPL-based wireless sensor networks', *International Journal of High Performance Computing and Networking*, Vol. 12, No. 1, pp.26–38.

Fan, W., Zhang, H., Li, K.C., Zhang, S. and Jiang, H. (2016) 'An efficient algorithm for modelling and dynamic prediction of network traffic', *International Journal of Computational Science and Engineering*, Vol. 16, No. 3, pp.311–320.

Fang, J., Zhang, X. and Wang, T. (2017) 'A node localisation approach based on mobile beacon using particle swarm optimisation in wireless sensor networks', *International Journal of Embedded Systems*, Vol. 9, No. 2, p.112.

Fei, J. and Wu, D. (2017) 'Adaptive control of mems gyroscope using fully tuned rbf neural network', *Neural Computing and Applications*, Vol. 28, No. 4, pp.695–702.

Gilles, J. (2013). 'Empirical wavelet transform', *IEEE Transactions on Signal Processing*, Vol. 61, No. 16, pp.3999–4010.

Gu, X. and Liao, Z. (2017) 'Short-term load forecasting based on phase space reconstruction and Gaussian process regression', *Power System Protection & Control*, Vol. 45, No.5, pp.73–79.

Huan, J., Cao, W. and Qin, Y. (2018) 'Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework', *Computers & Electronics in Agriculture*, No. 150, pp.257–265.

Hui, C., Jiaxing, K., Yuanchun, C., Dan, X. and Ying, H. (2017) 'An improved time-frequency analysis method for hydrocarbon detection based on ewt and set', *Energies*, Vol. 10, No. 8, p.1090.

Jinyou, H., Jingjie, W., Xiaoshuan, Z., Zetian, F. and University, C.A. (2015) 'Research status and development trends of information technologies in aquacultures', *Transactions of the Chinese Society for Agricultural Machinery*, Vol. 46, No. 7, pp.251–263.

Li, C., Li, Z. and Wu, J. (2018) 'A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features', *Information Processing in Agriculture*, Vol. 5, No. 1, pp.11–20.

Li, Q., Sun, Y., Yu, Y., Wang, C. and Ma, T. (2017) 'Short-term photovoltaic power forecasting for photovoltaic power station based on EWT-KMPMR', *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 33, No. 20, pp.265–273.

Liu, S., Xu, L., Li, D. and Zeng, L. (2012) 'Dissolved oxygen prediction model of eriocheir sinensis culture based on least squares support vector regression optimized by ant colony algorithm', *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*, Vol. 28, No. 23, pp.167–175.

Longqin, X., Qianchuan, L., Shuangyin, L., Daoliang, L. and University, G.O. (2016) 'Prediction of ph value in industrialized aquaculture based on ensemble empirical mode decomposition and improved artificial bee colony algorithm', *Transactions of the Chinese Society of Agricultural Engineering*, Vol. 32, No. 3, pp.202–209.

Ray, D. (1993) 'Nonlinear dynamics, chaos and instability statistical theory and economic evidence', *Journal of the Operational Research Society*, Vol. 44, No. 2, pp.202–203.

Thirumala, K., Umarikar, A.C. and Jain, T. (2015) 'Estimation of single-phase and three-phase power-quality indices using empirical wavelet transform', *IEEE Transactions on Power Delivery*, Vol. 30, No. 1, pp.445–454.

Usman, K. and Caterina, V. (2017) 'Optimising fuzzy neural network architecture for dissolved oxygen prediction and risk analysis', *Water*, Vol. 9, No. 6, pp.381–393.

Wang, X., Yang, L.T., Feng, J., Chen, X. and Deen, M.J. (2016) 'A tensor-based big service framework for enhanced living environments', *IEEE Cloud Computing*, Vol. 3, No. 6, pp.36–43.

Wang, B., Wang, J. and Hu, G. (2017a) 'College english classroom teaching evaluation based on particle swarm optimization-extreme learning machine model', *International Journal of Emerging Technologies in Learning*, Vol. 12, No. 5, p.82.

Wang, X., Yang, L.T., Liu, H. and Deen, M.J. (2017b) 'A big data-as-a-service framework: state-of-the-art and perspectives', *IEEE Transactions on Big Data*, Vol. 4, No. 3, pp.325–340.

Wang, X., Yang, L.T., Xie, X., Jin, J. and Deen, M.J. (2017c) 'A cloud-edge computing framework for cyber-physical-social services', *IEEE Communications Magazine*, Vol. 55, No. 11, pp.80–85.

Wu, J., Li, Z., Zhu, L., Li, G., Niu, B. and Peng, F. (2018) 'Optimized BP neural network for dissolved oxygen prediction', *IFAC-Papers On Line*, Vol. 51, No. 17, pp.596–601.

Zhang, L., Wang, Z., Yi, M. and Hu, Y. (2018) 'Fully secure hierarchical inner product encryption for privacy preserving keyword searching in cloud', *International Journal of High Performance Computing & Networking*, Vol. 11, No. 1, p.45.

Zhao, C., Liu, X. and Ding, F. (2010) 'Melt index prediction based on adaptive particle swarm optimization algorithm-optimized radial basis function neural networks', *Chemical Engineering and Technology*, Vol. 33, No. 11, pp.1909–1916.