Research
Artificial Intelligence—Article

A Hybrid Neural Network Model for Marine Dissolved Oxygen Concentrations Time-Series Forecasting Based on Multi-Factor Analysis and a Multi-Model Ensemble

Hui Liu [*], Rui Yang, Zhu Duan, Haiping Wu

Institute of Artificial Intelligence and Robotics (IAIR), Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

ARTICLE INFO

ABSTRACT

Dissolved oxygen (DO) is an important indicator of aquaculture, and its accurate forecasting can effectively improve the quality of aquatic products. In this paper, a new DO hybrid forecasting model is proposed that includes three stages: multi-factor analysis, adaptive decomposition, and an optimization-based ensemble. First, considering the complex factors affecting DO, the grey relational (GR) degree method is used to screen out the environmental factors most closely related to DO. The consideration of multiple factors makes model fusion more effective. Second, the series of DO, water temperature, salinity, and oxygen saturation are decomposed adaptively into sub-series by means of the empirical wavelet transform (EWT) method. Then, five benchmark models are utilized to forecast the sub-series of EWT decomposition. The ensemble weights of these five sub-forecasting models are calculated by particle swarm optimization and gravitational search algorithm (PSOGSA). Finally, a multi-factor ensemble model for DO is obtained by weighted allocation. The performance of the proposed model is verified by time-series data collected by the pacific islands ocean observing system (PacIOOS) from the WQB04 station at Hilo. The evaluation indicators involved in the experiment include the Nash–Sutcliffe efficiency (NSE), Kling–Gupta efficiency (KGE), mean absolute percent error (MAPE), standard deviation of error (SDE), and coefficient of determination (R^2). Example analysis demonstrates that: ① The proposed model can obtain excellent DO forecasting results; ② the proposed model is superior to other comparison models; and ③ the forecasting model can be used to analyze the trend of DO and enable managers to make better management decisions.

1. Introduction

Dissolved oxygen (DO) plays an important role in the sustainable development of aquaculture and marine ecology in coastal cities. Nowadays, with the rapid development of industry, large volumes of greenhouse gas emissions are increasingly intensifying global warming [1]. As a result, the seawater temperature is rising and DO is escaping from seawater [2]. If the DO concentrations in seawater are too low, marine organisms such as fish and shrimp will be deprived of oxygen, leading to slow growth and even suffocation [3]. For offshore aquaculture farmers, the massive scale of the death of aquatic products will bring huge economic losses [4]. More

seriously, it is difficult for DO to diffuse into the depths of the ocean, and a lack of oxygen will force aquatic organisms to change their habitats [5]. The marine ecosystem could change significantly as a result. Once the virtuous cycle of the ecosystem is broken, the survival of marine organisms will be greatly threatened [6]. Accurate DO content forecasting is very necessary for marine ecological balance and for the aquaculture development of coastal cities. It can provide a trend analysis of future changes in DO and facilitate the management of decision-makers by enabling them to take effective measures in advance [7,8]. Influenced by factors such as nutrients, climate, ecological environment, and the life activities of aquatic organisms, the DO content in seawater is nonlinear and has a considerable time delay. This characteristic makes DO difficult to predict. In addition, it is easy for data obtained by sensors to be lost during transmission, and such data contains uncertain outliers, posing more challenges to the accuracy of forecasting [9].

* Corresponding author.
 E-mail address: csuliuhui@csu.edu.cn (H. Liu).

1.1. Related works

In recent years, the rise of artificial intelligence (AI) technology has spawned a wave of AI models. On this basis, DO forecasting models can be divided into two types: physics-based models and data-driven models. Data-driven models include AI-based models and statistics-based models. To comprehensively evaluate state-of-the-art efforts in DO models, we summarize the literature on statistics-based models and AI-based models.

Statistics-based models have the advantage of easy calculation in application. In addition, a statistics-based model can predict an unknown situation by mining the potential relationship in the DO series. In practical applications, DO series are nonlinear and time-delayed. Faruk [10] used the autoregressive integrated moving average (ARIMA) model for DO forecasting to deal with the instability and nonlinearity of time series. In the experiments of Li et al. [11], the grey model was proposed to predict the trend term of DO. Huan et al. [12] obtained the optimal parameter values of the model through a Bayesian evidence framework. Khan et al. [13] compared the new Bayesian regression model with the new autoregressive modified fuzzy linear regression method. The comparison results show that the method based on fuzzy numbers can better capture DO changes in an urban river environment. Khan et al. [14] predicted changes in DO concentration by constructing multiple linear regression (MLR) models, in which the uncertainty of DO change was effectively characterized and propagated. In later research by Kisi and Parmar [15], the MLR model was used as an experimental comparison to explore the performance of other prediction models. In general, these statistics-based models have obtained the ideal forecasting effect for DO. However, forecasting accuracy remains to be improved.

Compared with statistics-based models, AI-component models have much better forecasting performance. The core of this section focuses on describing models with AI components. Commonly used AI algorithms include the multi-layer perceptron (MLP) [16], radial basis neural network (RBNN) [17], back propagation neural network (BPNN), extreme learning machine (ELM) [18], least square support vector machine (LSSVM) [19], and fuzzy neural network (FNN) [20,21], among others [22]. Since the selection of different parameters has a huge impact on the performance of an AI model, it is necessary to determine the optimal parameters by means of an optimization algorithm. Ren et al. [23] used a genetic algorithm (GA) to optimize the parameters of FNN. In addition, particle swarm optimization (PSO) [24], cauchy particle swarm optimization (CPSO) [25], the firefly algorithm (FFA) [26], and other optimization algorithms are used for parameter selection for intelligent models [27]. With its rapid development, deep learning has been widely studied for its excellent adaptive ability. Ma et al. [28] solved the problem of a sparse matrix by deep matrix factorization. Ren et al. [29] used a deep brief network (DBN)'s powerful feature extraction and functional representation capabilities to deal with highly complex nonlinear DO time-series data. Nevertheless, given the complex laws of DO changes and many other factors, simple machine learning basic predictors and parameter optimization methods are insufficient to allow accurate DO prediction. To further improve the forecasting performance of AI models, an increasing number of scholars are using feature selection, decomposition, and ensemble methods for the state-of-the-art research of DO forecasting. Further details about such research include the following:

(1) The change in DO concentration in the ocean and in freshwater environments is not isolated, as it is affected by many factors including temperature, turbidity, pH, chlorophyll, and specific conductance [30]. Changes in these factors have seemingly invisible effects on DO concentration. To effectively use these factors, the specific effects need to be studied. Shi et al. [22] used a clustering method to segment water quality time series, and finally improved the forecasting accuracy of DO. Ren et al. [23] arranged the factors that are positively and negatively correlated with DO through correlation analysis. It was found that the model showed a better forecasting performance when factors strongly related to DO were entered into the model [31]. From the perspective of the generation and consumption of DO, these influencing factors should be organically incorporated rather than directly and crudely input into the prediction model. This is not a comprehensive consideration. The direct superposition of a multi-factor forecasting series may prove to reduce the DO forecasting accuracy of a model. The weight allocation of each influencing factor needs to be analyzed, and the optimal combination method needs to be identified [32]. Therefore, exploring methods involving the multi-factor analysis of DO forecasting has great research value [33].

(2) The nonstationary properties of the original DO series are not favorable for forecasting. The extreme volatility of the DO series makes it difficult to predict. The birth of the decomposition method largely solved this problem. A decomposed sub-series has independent oscillation components and is more stable, making it easier to predict [34,35]. To apply this theoretical effect to practical problems, many data preprocessing decomposition methods for AI models have been proposed. Widely used decomposition algorithms include ensemble empirical mode decomposition (EEMD) [12], discrete wavelet transform (DWT), and variational mode decomposition (VMD) [36]. DO series are more predictable after decomposition. However, the current mainstream decomposition methods require users to determine the number of decomposition layers based on experience. This practice inevitably adds human error [37]. To improve the forecasting model performance, a decomposition algorithm that can adaptively determine the decomposition level should be investigated.

(3) After years of research and development, some ensemble models can be used for the intelligent forecasting of DO. An ensemble model can combine the advantages of multiple forecasting models to achieve better forecasting performance [35]. By means of a layered ensemble, Zhu et al. [38] successfully solved the problem of it being impossible to simultaneously achieve the accuracy of high and low concentration forecasting. Iterated stepwise multiple linear regression (ISMLR) is used to preprocess the data. An artificial neural network (ANN) and MLP are used to run layered forecasting. Finally, compromise programming (CP) is used to evaluate and select the best results. Kisi et al. [39] proposed a new bayesian model averaging (BMA) method as an ensemble model. Through comparison experiments with ELM, ANNs, an adaptive neuro-fuzzy inference system (ANFIS), a classification and regression tree (CART), and MLR, it was found that the BMA ensemble method is very effective for DO prediction. The essence of an ensemble model is to improve the forecasting accuracy and scope of model application by making use of the complementary advantages of multiple models. Unfortunately, the ensemble methods described above only use multiple models for hierarchical prediction, and then simply superimpose the results. A direct ensemble of multiple models will greatly increase the complexity of the ensemble model, which can easily lead to overfitting. At the same time, the shortcomings of a single model cannot be compensated for by other models. Therefore, it is very important to select a reasonable ensemble of models. Different machine learning prediction models have individual characteristics and can cope with time series in different states. If they can be combined organically, the ensemble model will have complementary advantages. Appropriate weight allocation can cover up the prediction defects of benchmark models and enhance the advantages of benchmark models. Therefore, it is necessary to explore the scientific ensemble method of multiple models.

The DO forecasting models discussed above have been demonstrated to be effective. Still, there are gaps in scientific research

that need to be filled in order to further improve DO forecasting performance. Table 1 [10–19,21–30,36,38–40] summarizes the reviewed state-of-the-art works using data-driven methods in DO forecasting.

1.2. The novelty of this study

To summarize the above literature, it can be seen that there are few comparative studies on multi-factor analysis, adaptive decomposition analysis, and scientific ensemble analysis in DO forecasting research. In view of the above limitations, a novel forecasting model is herein proposed for DO forecasting: the MF-RNNs-EWT-BEGOE model. MF represents the proposed multi-factor analysis

method. RNNs represents the replicator neural networks outlier detection method, EWT stands for empirical wavelet transform, BEGOE stands for BFGS-ENN-GRNN-ORELM-ELM. This model can be divided into the three stages of multi-factor analysis, adaptive decomposition, and optimization-based ensemble. Each sub-layer under the four factors uses the same type of benchmark model to predict; finally, the forecasting results of the model are obtained through a combination of the results. In the optimization ensemble phase, the ensemble weights for the Broyden–Fletcher–Goldfarb–Shanno (BFGS) model, elman neural network (ENN) model, general regression neural network (GRNN) model, outlier-robust extreme learning machine (ORELM) model, and ELM model are optimized by particle swarm optimization and gravitational search algorithm

Table 1
A summary of the reviewed state-of-the-art data-driven DO forecasting methods.

Category	Research content	Reference	Year of publication	Contribution
Statistics-based models	ARIMA model	[10]	2010	The ARIMA method was applied to study stationary and nonstationary time series
	Grey model	[11]	2018	The grey model was proposed to forecast the trend term of DO
	Bayesian model	[12]	2018	A bayesian evidence framework was applied to obtain optimal parameter values
		[13]	2017	The performance of the new bayesian regression model and the new autoregressive modified fuzzy linear regression method in DO prediction were compared
	MLR model	[14]	2013	The MLR model was used to characterize and propagate the uncertainty of DO changes
		[15]	2016	The MLR model was utilized to evaluate the performance of multiple models, such as LSSVM, multivariate adaptive regression splines (MARS), and the M5 model tree (M5Tree)
AI-component models	Basic prediction models	[21]	2016	The improved fuzzy neural network method was used to predict low DO events
		[18]	2020	The prediction performance of ELM and MLP on daily DO concentration was compared
		[17]	2012	The MLP and RBNN models based on ANNs were used as a comparison with MLR models based on statistical methods
		[30]	2013	ANNs, ANFIS, and gene expression programming (GEP) were used to compare the prediction performance of DO concentration
		[16]	2014	The k-means clustering method and MLP were combined to improve the prediction accuracy of the model
		[19]	2018	Several data-driven methods were used for modeling a comparison of daily DO concentration prediction, such as LSSVM, MARS, and M5Tree
	Optimization method for neural networks	[23]	2018	The GA algorithm was adopted to optimize the center and width of the FNN
		[24]	2018	The best parameters of the BPNN were determined by the PSO algorithm
		[27]	2019	The dual-scale DO soft-sensor modeling method was applied to improve forecasting performance
		[25]	2014	The CPSO was employed to optimize the kernel parameter and the regularization parameter of the LSSVR model
		[26]	2017	The FFA was used to optimize the three parameters of MLP
	Deep learning methods	[28]	2020	Deep matrix factorization was adopted for DO forecasting
		[29]	2020	A deep belief network was applied to improve forecasting performance
	Optimization method for decomposition algorithms	[40]	2019	The decomposed series was reconstructed by the sample entropy (SE) method to make forecasting easier
	Ensemble methods	[38]	2018	ISMLR + MLR + ANN was used to hierarchically predict high and low concentration data
		[39]	2020	By a comparison with multiple models such as ELM, ANNs, ANFIS, CART, and MLR, the application potential of the BMA ensemble model was verified
	Feature-selection methods	[22]	2019	k-medoids clustering based on segmentation was applied to select features
		[23]	2018	Correlation analysis was adopted for feature selection
	Decomposition methods	[36]	2019	DWT and VMD were adopted to decompose the original data
		[12]	2018	EEMD was utilized to improve DO forecasting performance

(PSOGSA) to obtain the BEGOE model. Section 2 provides the details of the proposed model. To fully understand the role of each part of the model, we performed numerous comparative experiments.

The innovations and contributions of this study are described below.

(1) A hybrid three-stage ocean DO forecasting ensemble model is designed. The model considers multi-factor analysis, adaptive decomposition analysis, and scientific ensemble analysis, thereby filling three of the research gaps described in Section 1.1. The multi-factor analysis reasonably considers the effects of temperature, salinity, turbidity, chlorophyll, and oxygen saturation on the concentration of DO in the ocean. EWT decomposition is utilized to adaptively improve the predictability of time series. The non-stationarity of DO series is reduced. The PSOGSA optimization method is used to optimize the weights of multiple benchmark models. The complementation of the benchmark model further improves the performance of the hybrid model. These three modules work together to achieve accurate forecasting of DO.

(2) Several factors that affect the concentration of DO in the ocean are reasonably considered. The multi-factor analysis method not only considers the influence of single and multiple factors on model accuracy, but also considers correlations among multiple factors. In this way, a hybrid model with better fusion performance can be obtained, avoiding the unilateral influence and local disadvantage of a single factor. This method also makes up for the shortcomings of direct multi-factor input, as such input may have a negative effect. This approach can take into account the reasons for a change in DO concentration from various aspects.

(3) The adaptive data preprocessing method has achieved ideal results in regard to the decomposition effect. Compared with the mainstream decomposition method, adaptive decomposition avoids the error caused by the artificial selection of decomposition layers. In this way, the decomposition layer can be predicted more easily. In addition, the EWT decomposition method discards the residual signal of the decomposition mode to extract meaningful

information from the ocean DO time series. The EWT method also compensates for the sensitivity of some decomposition algorithms to noise and sampling to a certain extent. The support of mathematical theory allows the machine learning algorithm performance to develop effectively.

(4) The organic combination of multiple benchmark models fills the research gap for the ensemble model. A meta-heuristic optimization algorithm is used to assign weights to BFGS, ENN, GRNN, ORELM, and ELM. The ensemble approach can achieve the advantages of multiple benchmark models while reducing the adverse effects of a single model. The ensemble model solves the poor robustness problem of some forecasting models and fully demonstrates the superiority of the complementary performance. Therefore, a multi-model complementary ensemble framework based on meta-heuristic optimization algorithm has great value.

2. The proposed multi-factor forecasting model

As mentioned earlier, the modeling process of the proposed MF-RNNs-EWT-BEGOE model can be divided into three stages: multi-factor feature extraction, EWT decomposition, and multi-model optimization ensemble optimization ensemble. Fig. 1 shows the modeling flow. It is worth mentioning that the proposed model uses the multi-input multi-output strategy (MIMOS), which is commonly used in the field of time-series forecasting [41]. Compared with recursive strategy (RS), the MIMOS has a smaller cumulative error [42]. Fig. 2 shows the schematic diagram of the RS and MIMOS.

2.1. Stage 1: The multi-factor analysis method

2.1.1. Substage 1.1: Replicator neural networks outlier detection

Outlier rejection is an important process for DO data. Preprocessing makes the data look more orderly and makes it more easily learned by the forecasting model, without paying attention to the sources of the outliers. It is worth noting that only the training sets
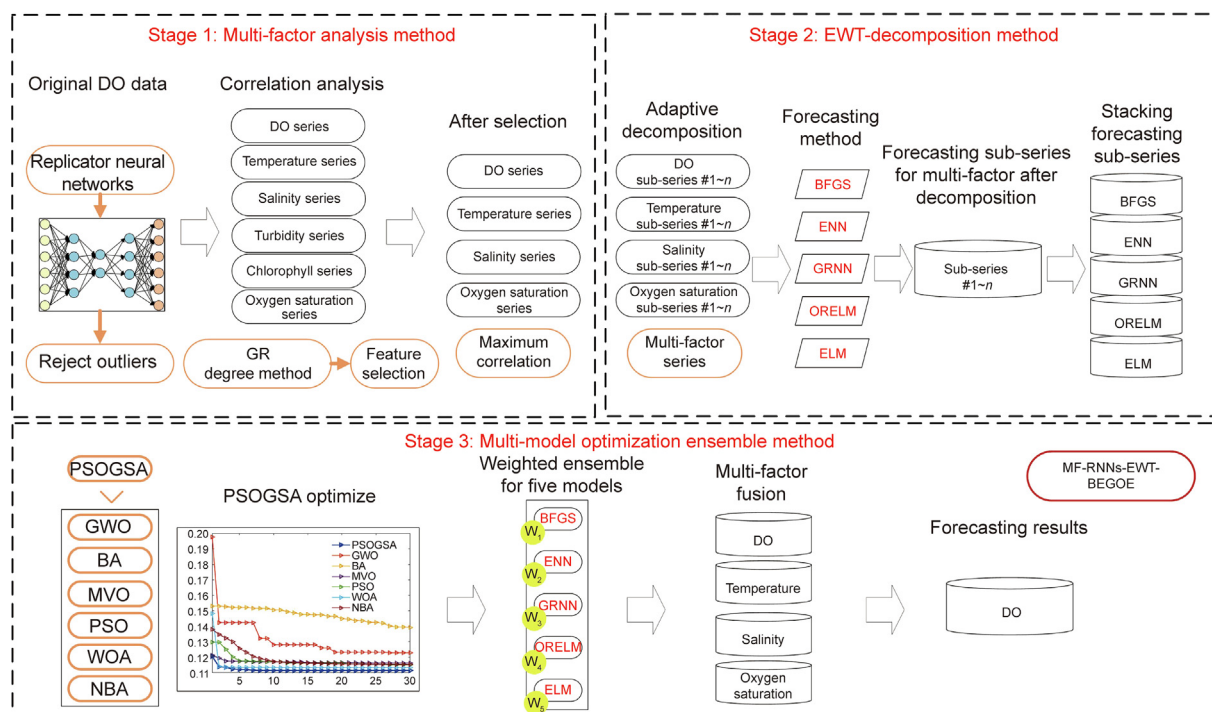

Fig. 1. The modeling flow of the proposed three-stage model. GR: grey relational; GWO: grey wolf optimizer; BA: bat algorithm; MVO: multi-verse optimization algorithm; WOA: whale optimization algorithm; NBA: new bat algorithm.
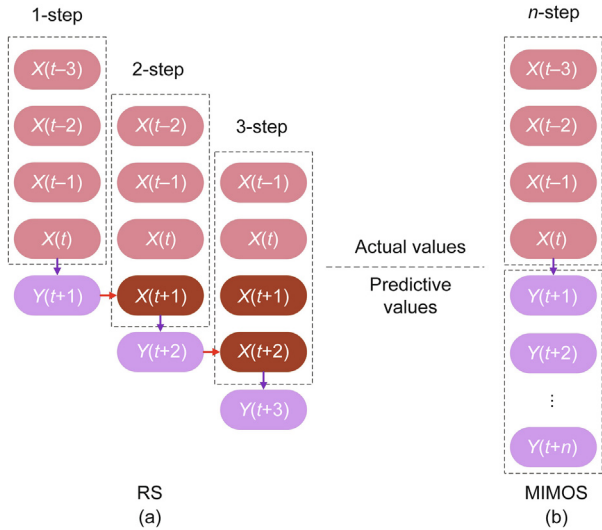
Fig. 2. Schematic diagram of multi-step forecasting strategies.

are corrected, while the test sets remain the same. This is a more convincing way to verify the validity of the model. RNNs is used to detect outliers of the original ocean DO series. This is a multi-layer feed-forward neural network [43].

Assuming that the output of the ith neuron in the k layer of the RNNs is $S_k(I_{ki})$, I_{ki} represents the input of the ith neuron in the k layer, and S_k represents the activation function used in the k layer. The input of the neuron can be expressed as follows:

$$\theta = I_{ki} = \sum_{j=0}^{L_{k-1}} w_{kij} Z_{(k-1)j} \tag{1}$$

where W_{kij} is the weight that connects ith neuron in the k layer and jth neuron in the $k-1$ layer, Z_{kj} is the output of the jth neuron in the k layer, and L_k is the number of neurons in the k layer. In this example, when $k = 2, 4$, the activation function can be expressed as follows:

$$S_k(\theta) = \tanh(a_k\theta) \quad (k = 2, 4) \tag{2}$$

where a_k is set to 1. For the intermediate layer where $k = 3$, its activation function can be expressed as a class step-like function. The activation function can be expressed as follows:

$$S_3(\theta) = \frac{1}{2} + \frac{1}{4} \sum_{j=1}^{N-1} \tanh\left[a_3\left(\theta - \frac{j}{N}\right)\right] \tag{3}$$

where N represents the number of steps and a_3 represents the rate of promotion to the next level.

In general, continuous data entering the third layer is converted into a batch of discrete values through the S_3 stepwise activation function of the RNNs. This is equivalent to mapping the series sample to N clusters. Finally, the RNNs can calculate a single outlier and a small cluster of outliers. This completes the outlier detection of marine data at the WQB04 site. Algorithm S1 in the Appendix A shows the pseudo-code of the RNNs outlier detection method.

2.1.2. Substage 1.2: Multi-factor analysis of the grey relational (GR) degree method

To screen out the most favorable factors for DO forecasting, the GR degree method is used to select the features of the original multi-factor series. In this example, the GR degree method is used to calculate the correlation of environmental factors such as temperature, salinity, turbidity, chlorophyll, and oxygen saturation. The greater the degree of grey correlation between time series, the closer the distance between the two series. Feature selection

is used to determine the series that are most associated with DO. The specific calculation method of the GR degree is presented as follows [44,45]:

It is assumed that $R_0 = (r_0(1), r_0(2), r_0(3), ..., r_0(n))$ is a series group of system characteristic behaviors, and $R_i = (r_i(1), r_i(2), r_i(3), ..., r_i(n))$ $(1 < i < N)$ is a series of related behaviors. Thus, the correlation coefficient between the two series can be obtained by the following calculation method:

$$\alpha_i(k) = \frac{\min_i\min_k|r_0(k) - r_i(k)| + \beta\max_i\max_k|r_0(k) - r_i(k)|}{|r_0(k) - r_i(k)| + \beta\max_i\max_k|r_0(k) - r_i(k)|} \tag{4}$$

where the absolute difference between $r_0(k)$ and $r_i(k)$ is point k, $\min_i\min_k|r_0(k) - r_i(k)|$ is the smallest difference between two layers, $\max_i\max_k|r_0(k) - r_i(k)|$ is the biggest absolute difference between the two layers, and β (the resolution factor) is usually 0.5.

The relationship ρ between the series of r_0 and r_i is shown as follows:

$$\rho(r_0, r_i) = \frac{\sum_{k=1}^{n} \alpha_i(k)}{n}, \quad (i = 1, 2, 3, ..., n) \tag{5}$$

2.2. Stage 2: The EWT-decomposition method

After the multi-factor analysis method, EWT decomposition is utilized to adaptively decompose the DO, water temperature, salinity, and oxygen saturation series. The flag corresponding to the sth predictor is #s. Fig. 3 shows the decomposition scheme of the time-series data. In the training phase, the decomposition is carried out in total $[X_1{}^s, ..., X_{1247}{}^s, X_{1248}{}^s]$. X represents the DO data sample point. In the forecasting phase, the data is divided into validation sets $[X_{1249}{}^s, ..., X_{1343}{}^s, X_{1344}{}^s]$ and test sets $[X_{1345}{}^s, ..., X_{1439}{}^s, X_{1440}{}^s]$. The validation set is used for ensemble model optimization, and the test set is used to test the performance of the proposed model. In this process, the BFGS, ENN, GRNN, ORELM, and ELM models are used to predict the sub-series, and the forecasting phase of the 1249th–1440th series can be obtained. The decomposition process includes the forecasting results of 1-step–3-step. Algorithm S2 in the Appendix A shows the pseudo-code of the EWT-decomposition method.

2.3. Stage 3: The multi-model optimization ensemble method

To further improve the forecasting performance, the BFGS, ENN, GRNN, ORELM, and ELM are combined to obtain the BEGOE model. The ensemble weights are optimized by the PSOGSA. Fig. 4 shows the weighted optimization scheme. The optimization objective function of PSOGSA is minimizing the mean absolute percent error (MAPE). The MAPE refers to the average absolute percentage error between different samples' actual and forecasting values. The objective function is calculated in the validation data. The specific objective function is shown as follows:

$$\text{Object}_{\text{MAPE}} = \min\left\{\sum_{p=1}^{P} E_t\left[\left(E_D[\hat{y}_p(t, D)] - E_\varepsilon[Y(t, \varepsilon)]\right)^2\right]/P\right\}$$

$$\text{s.t.} \quad \hat{y}_p(t, D) = \frac{\sum_{a=1}^{5}\left(\hat{X}_p^a(t, D) \times w_a\right)}{(w_1 + w_2 + w_3 + w_4 + w_5)} \tag{6}$$

$$0 \leq w_a \leq 1$$

where $Y(t, \varepsilon)$ is the actual DO series at time t with noise ε and $\hat{y}(t, D)$ represents the forecasting results of BEGOE at time t, where BFGS, ENN, GRNN, ORELM, and ELM are trained with the samples D. E_t, E_D, and E_ε represent the expectation over time, for training samples with noise ε. As for E_ε, $E_\varepsilon[Y(t, \varepsilon)]$ is equal to the actual DO series without noise. Nevertheless, the noise of the actual series
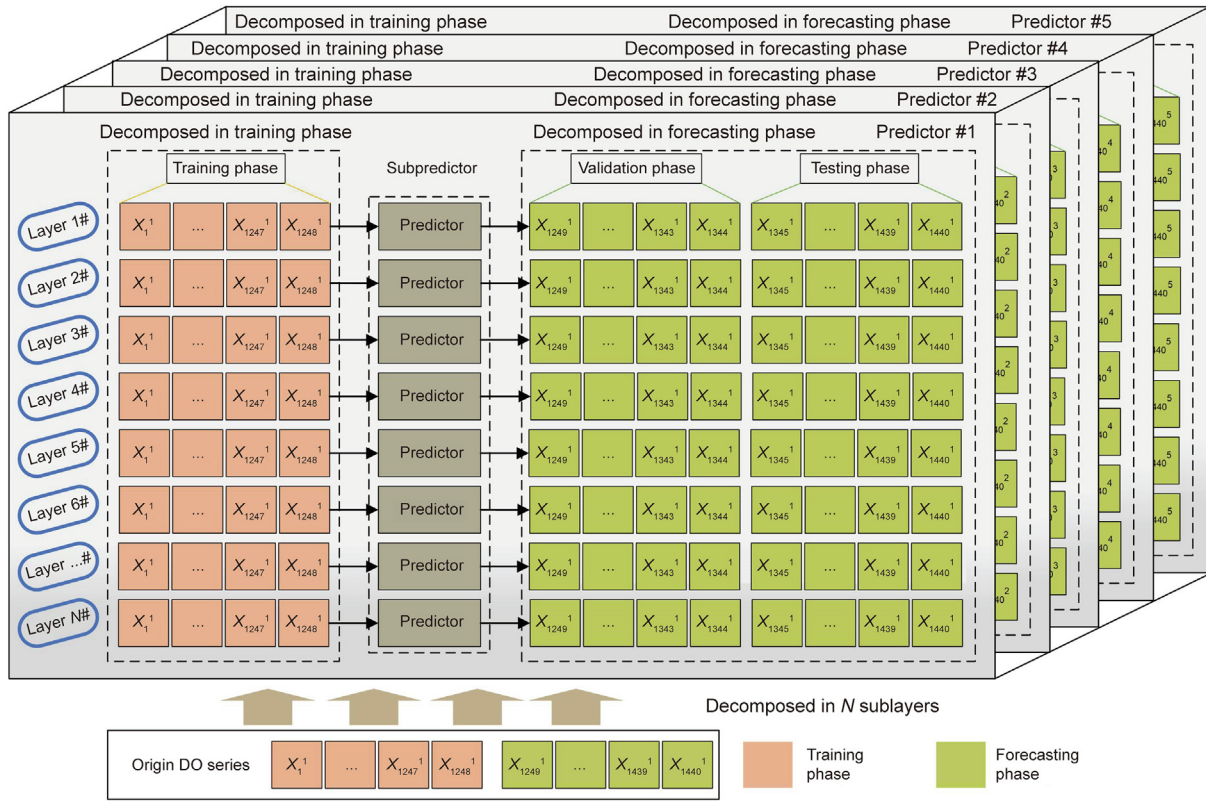
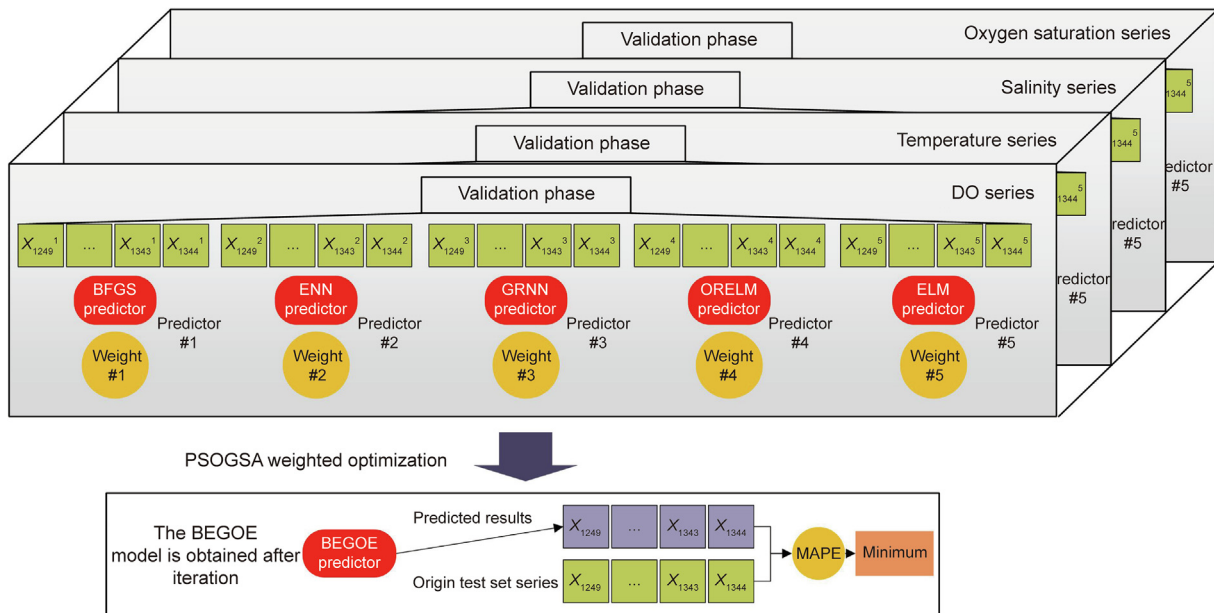**Fig. 3.** EWT-decomposed and layer-by-layer forecasting.

Fig. 4. The PSOGSA optimization ensemble process under multi-factor analysis.

cannot be eliminated, because the actual DO series is not an artificial series. To approximate $E_\varepsilon[Y(t,\varepsilon)]$, $E_\varepsilon[Y(t,\varepsilon)]$ is considered to be equal to $Y(t,\varepsilon)$ [46]. In addition, P is the maximum forecasting step, $[w_1, w_2, w_3, w_4, w_5]$ are the ensemble weights, and $\hat{X}_p^1$, $\hat{X}_p^2$, $\hat{X}_p^3$, $\hat{X}_p^4$, $\hat{X}_p^5$, $\hat{y}_p$ are the p-step forecasting results

of BFGS, ENN, GRNN, ORELM, ELM, and BEGOE for the validation data.

Considering that the DO, water temperature, salinity, and oxygen saturation series are involved in this example, the input of the BEGOE model are four series, and the output is a DO series.

By minimizing the objective function MAPE, appropriate ensemble weights can be obtained. The effects of water temperature, salinity, and oxygen saturation are added to the DO series. Finally, the forecasting results of the BEGOE can be obtained. Algorithm S3 of the Appendix A shows the pseudo-code of the multi-model optimization ensemble method.

3. Experiment and analysis

In this section, multiple comparative experiments are carried out to verify the effectiveness of the proposed model. In Section 3.1, six real-time series sets collected by the ocean observing system are introduced. In Section 3.2, the performance indices for evaluating the forecasting results are demonstrated. In Section 3.3, the implementation details and the contributions of each stage are explained. In Section 3.4, two experiments are set up. Experiment I compares the contributions of the model components horizontally, and Experiment II compares several state-of-the-art models to demonstrate the superiority of the proposed model.

3.1. Data description

In this section, time-series data such as DO, temperature, salinity, turbidity, chlorophyll, and oxygen saturation collected by the Pacific Islands Ocean Observing System from the WQB04 station at Hilo are used for experimental research. The time interval of the six time-series data is 15 min. To effectively build the subsequent models, the dataset is divided into three parts: a training set, validation set, and test set. The data corresponding to the six factors were all selected from the time period of 1–30 December 2016. Among them, the data for 16–30 December is regarded as Dataset #1, while the data for 1–15 December is regarded as Dataset #2. The total length of each series corresponding to each factor is 1440. The training set contains the 1st–1248th data, the validation set contains the 1249th–1344th data, and the test set contains the 1345th–1440th data. Fig. S1 and Table S1 in the Appendix A present the dataset information after RNN outlier processing.

3.2. Performance evaluation

An index set is applied to evaluate the deterministic forecasting performance, including the coefficient of determination (R^2), Nash–Sutcliffe Efficiency (NSE) [47] index, Kling–Gupta fficiency (KGE) [48], standard deviation of error (SDE), and MAPE. The expressions of these five indices are shown as follows:

$$R^2 = \left(\frac{\sum_{t=1}^{N_Y}\left(\left(Y_t - \overline{Y_t}\right)\left(\widehat{Y_t} - \overline{\widehat{Y_t}}\right)\right)}{\sqrt{\sum_{t=1}^{N_Y}\left(Y_t - \overline{Y_t}\right)^2 \sum_{t=1}^{N_Y}\left(\widehat{Y_t} - \overline{\widehat{Y_t}}\right)^2}} \right)^2 \tag{7}$$

$$\text{NSE} = 1 - \frac{\sum_{t=1}^{N_Y}\left(Y_t - \widehat{Y_t}\right)^2}{\sum_{t=1}^{N_Y}\left(Y_t - \overline{Y_t}\right)^2} \tag{8}$$

$$\text{KGE} = 1 - \sqrt{(R-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}$$

$$\alpha = \sqrt{\frac{\sum_{t=1}^{N_Y}\left(\widehat{Y_t} - \overline{\widehat{Y_t}}\right)^2}{\sum_{t=1}^{N_Y}\left(Y_t - \overline{Y_t}\right)^2}} \tag{9}$$

$$\beta = \frac{\overline{\widehat{Y_t}}}{\overline{Y_t}}$$

$$\text{SDE} = \sqrt{\frac{1}{N_Y}\sum_{t=1}^{N_Y}\left(\left(Y_t - \widehat{Y_t}\right) - \left(\overline{Y_t - \widehat{Y_t}}\right)\right)^2} \tag{10}$$

$$\text{MAPE} = \frac{\sum_{t=1}^{N_Y}\left|\dfrac{Y_t - \widehat{Y_t}}{Y_t}\right|}{N_Y} \tag{11}$$

where $\widehat{Y}$ is the forecasting results, Y is the actual data, and N_Y is the number of the actual data.

3.3. Modeling analysis

3.3.1. Multi-factor analysis

The DO in the ocean is affected by many factors. In this paper, a variety of factors affecting the increase and decrease of DO concentrations are analyzed. Fig. 5 shows the system dynamics model.

As shown in Fig. 5, many factors affect DO concentrations. Usually, atmospheric diffusion, water exchange, mechanical aerobics, aquatic photosynthesis, and the like are the main sources of DO. Under natural conditions, the DO content in the ocean is closely related to the partial pressure of oxygen in the air and the temperature of the water. When the rate of oxygen dissolved into the water is equal to the rate of oxygen escaping from the water, the dissolution reaches a dynamic equilibrium. The equilibrium of DO in the ocean is also affected by salinity, turbidity, pH, and other factors. In addition, chemical oxidation reactions, organic decomposition, breathing by phytoplankton at night, and aquatic respiration are the main factors of marine DO consumption. In short, there are many factors affecting changes in DO content, and the law of distribution variation is difficult to explain by a simple mechanism.

In this experiment, the relationship between temperature, salinity, turbidity, chlorophyll, oxygen saturation, and DO concentrations is calculated by using the GR degree method. Table 2 shows the specific results. According to Table 2, temperature, salinity, and oxygen saturation have the highest correlation degree with DO. The correlation indicators reach 0.61, 0.68, and 0.87, respectively, and all exceed the set threshold of 0.5. It is worth mentioning that salinity contains nutrients, and nutrients play an important role in DO concentration. Considering that high concentrations of turbidity will influence the photosynthesis of plants in water and affect oxygen production, turbidity was also input into the model as a DO-related factor. The experiment adopts the form of multi-factor variable input and single-factor variable output. In an actual comparison of repeated experiments, it was found that the forecasting accuracy is the best when temperature, salinity, and oxygen saturation are input into the model as the influential factors of DO. The experimental results show that if low-correlation factors such as turbidity, chlorophyll, and pH are input into the model, the forecasting accuracy decreases accordingly. In this case, the index of chlorophyll always maintains a value of -8.8×10^{-7}. Therefore, it does not show a correlation with the change in DO, so the correlation coefficient between chlorophyll and DO is not a number.

3.3.2. Decomposition model analysis

In this section, the EWT is used to reduce the non-stationarity of the DO series, the water temperature series, the salinity series, and the oxygen saturation series, all of which are adaptively decomposed. It is worth mentioning that only the training set data is decomposed. After the original DO series and the highly correlated series are decomposed, the sub-series having more stable characteristics are easier to predict. The primary predictor simply concentrates on a certain frequency band within these sub-series. Eventually, the final result after data preprocessing can be obtained by superimposing the prediction result of the sub-predictor. This
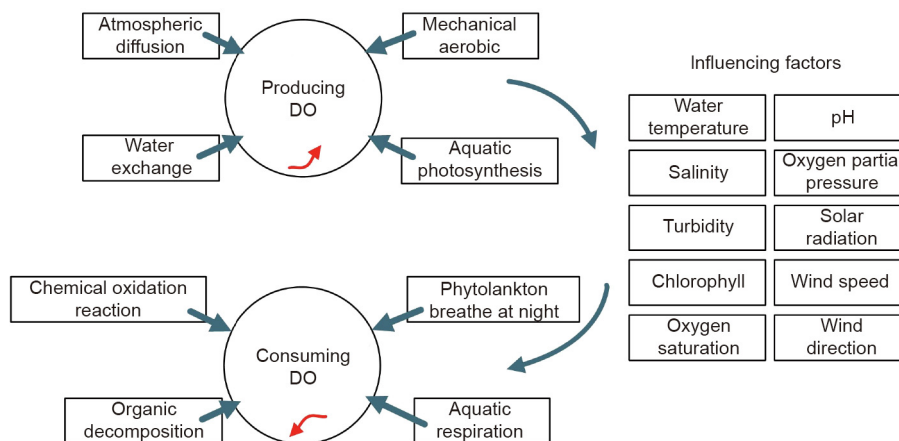
Fig. 5. System dynamics model diagram of dissolved oxygen change.

Table 2
Correlation coefficient between environmental factors and dissolved oxygen in the WQB04 dataset.

Indicator	Correlation coefficients				
	Temperature	Salinity	Turbidity	Chlorophyll	Oxygen saturation
Dissolved oxygen	0.61	0.68	0.32	NaN	0.87

NaN: not a number.

section compares the performance changes of the forecasting models by decomposing the input series.

3.3.3. Optimization method analysis

In this section, the PSOGSA is compared with multiple alternative methods, including the grey wolf optimizer (GWO), bat algorithm (BA), multi-verse optimization algorithm (MVO), PSO, whale optimization algorithm (WOA), and new bat algorithm (NBA). Most of these optimization algorithms are based on heuristic bionic algorithms, and it is difficult to clearly explain which is more suitable from a theoretical point of view. Therefore, experiments are needed to verify the DO prediction effect of each optimization algorithm in this situation.

3.4. Comparison study

3.4.1. Experiment I: Comparisons with each component of the proposed model

In Experiment I, five baseline models based on the components of the proposed model are constructed and used to evaluate the proposed model. Table 3 lists their specific expressions. To compare the effects of environmental parameters related to DO in actual calculations, we established a multi-factor analysis experiment. Models 1–3 are trained without multi-factor analysis. In the same situation, Models 4 and 5 and the proposed model are added with multi-factor analysis. In other words, Models 4 and 5 and the proposed model are trained with environmental factors

Table 3
Experimental setup of baseline models and the proposed model.

Model	Experimental setups
Model 1	RNNs-BFGS/ENN/GRNN/ORELM/ELM
Model 2	RNNs-BEGOE
Model 3	RNNs-EWT-BEGOE
Model 4	MF-RNNs-BFGS/ENN/GRNN/ORELM/ELM
Model 5	MF-RNNs-BEGOE
Proposed model	MF-RNNs-EWT-BEGOE

including DO series, temperature series, salinity series, and oxygen saturation series. To compare the effects of the optimization ensemble, a multi-model ensemble comparison experiment is established. Compared with Models 1 and 4, Models 2 and 5 are optimized and combined with five benchmark models. To verify the superiority of the EWT decomposition method, a set of experimental comparisons with or without data preprocessing are also implemented. More specifically, Model 3 and the proposed model add a module for EWT data preprocessing to reduce the instability of the original series. Tables 4 and 5 list the DO testing dataset forecasting results. In addition, in order to reflect the impact of optimization algorithms, the seven optimization algorithms described in Section 3.3.3 are used in the multi-model optimization ensemble module. Fig. 6 shows testing dataset forecasting results in 3-step of the PSOGSA and other optimization algorithms.

The conclusion can be summarized from Tables 4 and 5, and Fig. 6 as follows:

(1) Model 2 (RNNs-BEGOE) is significantly better than Model 1 (RNNs-BFGS/ENN/GRNN/ORELM/ELM) regardless of whether in Dataset #1 or Dataset #2, which shows that the multi-model ensemble method is effective. This method combines the advantages of five benchmark models. The multiple iterations of the optimization algorithm complete the combination of each model through the optimal weight allocation. The ensemble model incorporates the advantages of each benchmark model while discarding their negative performance. Taking the 1-step forecasting results of Dataset #1 as an example, the MAPE of RNNs-BFGS, RNNs-ENN, RNNs-GRNN, RNNs-ORELM, RNNs-ELM, and RNNs-BEGOE are 1.76%, 1.51%, 2.74%, 1.49%, 1.61%, and 1.32%, respectively. Other evaluation indicators such as NSE, KGE, SDE, and R^2 all show that Model 2 is better than Model 1 in forecasting performance.

(2) Model 3 (RNNs-EWT-BEGOE) is significantly better than Model 2 (RNNs-BEGOE), which shows that the decomposition-based data preprocessing method can have a positive effect on improving model prediction performance. The EWT adaptively decomposes the original series into multiple sub-series. The decomposed series has independent oscillation components and strong stability. The EWT has strong mathematical support while

Table 4
Evaluation indices of the compared models in Experiment I (Dataset #1).

Horizon	Model	NSE	KGE	MAPE (%)	SDE (mg·L^{-1})	R^2
1-step	RNNs-BFGS	0.47	0.69	1.76	0.16	0.50
	MF-RNNs-BFGS	0.61	0.77	1.45	0.14	0.62
	RNNs-ENN	0.55	0.69	1.51	0.15	0.56
	MF-RNNs-ENN	0.62	0.75	1.41	0.14	0.63
	RNNs-GRNN	−0.24	0.19	2.74	0.19	0.35
	MF-RNNs-GRNN	0.52	0.56	1.63	0.15	0.53
	RNNs-ORELM	0.58	0.77	1.49	0.14	0.62
	MF-RNNs-ORELM	0.62	0.77	1.41	0.14	0.63
	RNNs-ELM	0.53	0.71	1.61	0.15	0.55
	MF-RNNs-ELM	0.60	0.76	1.44	0.14	0.61
	RNNs-BEGOE	0.70	0.75	1.32	0.10	0.83
	MF-RNNs-BEGOE	0.91	0.90	0.67	0.07	0.91
	RNNs-EWT-BEGOE	0.97	0.96	0.38	0.04	0.97
	MF-RNNs-EWT-BEGOE	1.00	1.00	0.11	0.01	1.00
2-step	RNNs-BFGS	0.28	0.57	2.00	0.19	0.36
	MF-RNNs-BFGS	0.43	0.66	1.78	0.17	0.47
	RNNs-ENN	0.38	0.59	1.75	0.17	0.41
	MF-RNNs-ENN	0.47	0.63	1.72	0.16	0.48
	RNNs-GRNN	−0.33	0.15	2.84	0.20	0.28
	MF-RNNs-GRNN	0.32	0.41	1.99	0.18	0.33
	RNNs-ORELM	0.36	0.67	1.91	0.17	0.48
	MF-RNNs-ORELM	0.45	0.66	1.76	0.16	0.48
	RNNs-ELM	0.39	0.58	1.85	0.17	0.41
	MF-RNNs-ELM	0.46	0.65	1.72	0.16	0.48
	RNNs-BEGOE	0.53	0.65	1.69	0.13	0.68
	MF-RNNs-BEGOE	0.81	0.82	1.03	0.09	0.82
	RNNs-EWT-BEGOE	0.92	0.92	0.67	0.06	0.92
	MF-RNNs-EWT-BEGOE	0.99	0.99	0.14	0.02	1.00
3-step	RNNs-BFGS	0.02	0.46	2.28	0.22	0.22
	MF-RNNs-BFGS	0.24	0.54	2.06	0.19	0.32
	RNNs-ENN	0.23	0.47	1.99	0.19	0.29
	MF-RNNs-ENN	0.29	0.51	1.99	0.19	0.33
	RNNs-GRNN	−0.41	0.11	2.92	0.20	0.23
	MF-RNNs-GRNN	0.21	0.34	2.15	0.19	0.24
	RNNs-ORELM	0.09	0.56	2.31	0.19	0.33
	MF-RNNs-ORELM	0.27	0.54	2.04	0.19	0.33
	RNNs-ELM	0.22	0.46	2.12	0.20	0.27
	MF-RNNs-ELM	0.27	0.53	1.99	0.19	0.33
	RNNs-BEGOE	0.47	0.60	1.80	0.14	0.64
	MF-RNNs-BEGOE	0.71	0.76	1.24	0.11	0.75
	RNNs-EWT-BEGOE	0.80	0.88	1.07	0.10	0.80
	MF-RNNs-EWT-BEGOE	0.98	0.98	0.17	0.04	0.99

eliminating the interaction between different sub-series. The combination of these advantages makes EWT an excellent candidate for data preprocessing. Taking the 1-step forecasting results of Dataset #1 as an example, the NSE, KGE, MAPE, SDE, and R^2 are 0.70, 0.75, 1.32%, 0.10 mg·L^{-1}, and 0.83 for Model 2, and 0.97, 0.96, 0.38%, 0.04 mg·L^{-1}, and 0.97 for Model 3, respectively.

(3) Models 4 (MF-RNNs-BFGS/ENN/GRNN/ORELM/ELM) and 5 (MF-RNNs-BEGOE) and the proposed model with multi-factor analysis performed significantly better in comparison with the corresponding Models 1–3 without multi-factor analysis. This work shows that environmental factors such as temperature, salinity, and oxygen saturation have an impact on DO concentration. There is a complex nonlinear relationship between them, and the consideration of a single DO factor is not comprehensive. Taking the 1-step forecasting results of Dataset #1 as an example, the NSE, KGE, MAPE, SDE, and R^2 are 0.70, 0.75, 1.32%, 0.10 mg·L^{-1}, and 0.83 for Model 2, and 0.91, 0.90, 0.67%, 0.07 mg·L^{-1}, and 0.91 for Model 5, respectively. The experimental results show that the multi-factor consideration is correct. When multiple factors are reasonably input into the forecasting model, a more accurate and more scientific forecasting result can be obtained.

(4) Considering all the experimental datasets, forecasting steps, and different model evaluation indicators, the proposed hybrid model has the most outstanding forecasting performance. The excellent robustness and accuracy of the proposed model make

its prediction result very close to the actual DO value. Taking the 1-step forecasting results of Dataset #1 as an example, the NSE, KGE, MAPE, SDE, and R^2 of the proposed model are 1.00, 1.00, 0.11%, 0.01 mg·L^{-1}, and 1.00, respectively. Taking the 1-step forecasting results of Dataset #2 as an example, the NSE, KGE, MAPE, SDE, and R^2 of the proposed model are 1.00, 0.98, 0.18%, 0.01 mg·L^{-1}, and 1.00, respectively. Both sets of experiments show that the proposed model has good durability.

In summary, multi-factor analysis, adaptive decomposition analysis, and scientific ensemble model analysis are indispensable for improving the performance of the proposed hybrid model. The framework of the hybrid model has strong interpretability, and it provides a direction for the improvement of DO forecasting.

3.4.2. Experiment II: Comparison with existing models

In recent years, many AI-based models have been proposed for DO prediction. In Experiment II, we reproduced three state-of-the-art models that were published in 2018–2020 to compare them with our proposed hybrid model in order to verify the DO forecasting performance. These models are relatively complex, and basic machine learning models can also achieve satisfactory results. To complete a reasonable and scientific model evaluation, we added the support vector machines (SVM) model based on ANN, and the DBN model based on deep learning for comparison. The specific model introduction is elaborated as follows:

Table 5
Evaluation indices of the compared models in Experiment I (Dataset #2).

Horizon	Model	NSE	KGE	MAPE (%)	SDE (mg·L^{-1})	R^2
1-step	RNNs-BFGS	0.77	0.85	1.20	0.12	0.78
	MF-RNNs-BFGS	0.80	0.89	1.12	0.11	0.80
	RNNs-ENN	0.75	0.76	1.39	0.12	0.75
	MF-RNNs-ENN	0.79	0.85	1.17	0.11	0.79
	RNNs-GRNN	0.07	0.35	2.84	0.19	0.46
	MF-RNNs-GRNN	0.39	0.63	2.31	0.19	0.45
	RNNs-ORELM	0.79	0.89	1.15	0.11	0.80
	MF-RNNs-ORELM	0.80	0.86	1.13	0.11	0.80
	RNNs-ELM	0.76	0.83	1.26	0.12	0.76
	MF-RNNs-ELM	0.78	0.82	1.24	0.12	0.78
	RNNs-BEGOE	0.87	0.78	1.04	0.09	0.90
	MF-RNNs-BEGOE	0.95	0.94	0.60	0.06	0.95
	RNNs-EWT-BEGOE	0.98	0.99	0.35	0.03	0.98
	MF-RNNs-EWT-BEGOE	1.00	0.98	0.18	0.01	1.00
2-step	RNNs-BFGS	0.60	0.74	1.68	0.15	0.63
	MF-RNNs-BFGS	0.64	0.81	1.57	0.15	0.66
	RNNs-ENN	0.58	0.63	1.86	0.16	0.59
	MF-RNNs-ENN	0.64	0.75	1.65	0.15	0.64
	RNNs-GRNN	0.00	0.33	2.95	0.20	0.40
	MF-RNNs-GRNN	0.23	0.56	2.61	0.21	0.34
	RNNs-ORELM	0.62	0.81	1.68	0.15	0.66
	MF-RNNs-ORELM	0.66	0.76	1.59	0.14	0.66
	RNNs-ELM	0.60	0.72	1.73	0.16	0.60
	MF-RNNs-ELM	0.62	0.69	1.78	0.15	0.62
	RNNs-BEGOE	0.78	0.71	1.36	0.12	0.81
	MF-RNNs-BEGOE	0.88	0.91	0.95	0.08	0.89
	RNNs-EWT-BEGOE	0.94	0.95	0.62	0.06	0.95
	MF-RNNs-EWT-BEGOE	0.99	0.97	0.28	0.02	0.99
3-step	RNNs-BFGS	0.46	0.67	1.96	0.17	0.51
	MF-RNNs-BFGS	0.47	0.73	1.95	0.18	0.53
	RNNs-ENN	0.47	0.54	2.14	0.18	0.49
	MF-RNNs-ENN	0.50	0.65	1.99	0.17	0.51
	RNNs-GRNN	−0.06	0.30	3.06	0.20	0.35
	MF-RNNs-GRNN	0.18	0.54	2.71	0.22	0.31
	RNNs-ORELM	0.44	0.72	2.13	0.18	0.53
	MF-RNNs-ORELM	0.53	0.67	1.90	0.17	0.54
	RNNs-ELM	0.47	0.64	2.02	0.18	0.49
	MF-RNNs-ELM	0.49	0.61	2.07	0.17	0.50
	RNNs-BEGOE	0.69	0.65	1.60	0.14	0.73
	MF-RNNs-BEGOE	0.77	0.84	1.39	0.12	0.78
	RNNs-EWT-BEGOE	0.87	0.91	0.95	0.09	0.89
	MF-RNNs-EWT-BEGOE	0.96	0.94	0.58	0.04	0.98

Li et al. [31] used a recurrent neural network, long short-term memory network (LSTM), and gate recurrent I unit (GRU) to construct three DO forecasting models, respectively, in order to determine the most suitable one. In the experiment, the correlation coefficients between pH, turbidity, temperature, NH, and DO are calculated, and then these series are uniformly input into the forecasting model. Compared with Li's direct inputting of all parameters, we set thresholds and experimental screening methods to finally determine the factors that can have a positive effect on DO forecasting. Li's approach risks reducing accuracy. In fact, the correlation coefficient of the impact factors can only be used as a reference indicator, and the true relationship between these environmental factors and DO is very complicated.

Huan et al. [12] combined the EEMD data preprocessing method and LSSVM to improve the model's performance for DO forecasting. EEMD is used to decompose the DO series, and then LSSVM is used to predict these intrinsic mode function (IMF) components separately. The synthesized prediction result is obtained by superposition. Before the final DO forecasting results are generated, BPNN is used to reconstruct the forecast results, thereby eliminating the wrong analysis. It is a pity that EEMD lacks sufficient theoretical support in terms of mathematical definition, which results in decomposition flaws. We use a novel signal processing tool EWT to overcome this shortcoming. At the same time, the adaptive decomposition characteristics of EWT can determine the optimal

number of decomposition layers to maximize the decomposition effect.

Ren et al. [23] used GA to optimize the center and width of the center layer of the FNN to determine the best combination and improve DO prediction performance. In the proposed model, we apply the optimization algorithm to the ensemble of multiple benchmark models. The advantage of this approach is to reasonably combine the superior performance of multiple AI models while discarding bad performance. Compared with a single parameter-optimized predictor, the complementarity of multiple models can make the model more mature and allow it to obtain more adequate training. Among the models used, the reasonable weight allocation obtained by multiple iterations plays a key role.

More specifically, Tables 6 and 7, and Figs. 7 and 8 show the DO testing dataset forecasting results of the proposed model and state-of-the-art models.

The conclusion can be summarized from Figs. 7 and 8, and Tables 6 and 7, as follows:

Considering all datasets, forecasting steps, and different time-series evaluation indicators, the proposed model has the best forecasting performance in comparison with the other state-of-the-art models. Taking the 1-step forecasting results of Dataset #1 as an example, the MAPE of SVM, DBN, Li's model, Huan's model, Ren's model, and the proposed model are 1.41%, 1.44%, 0.52%, 0.56%, 0.51%, and 0.11%, respectively. The proposed model performs well
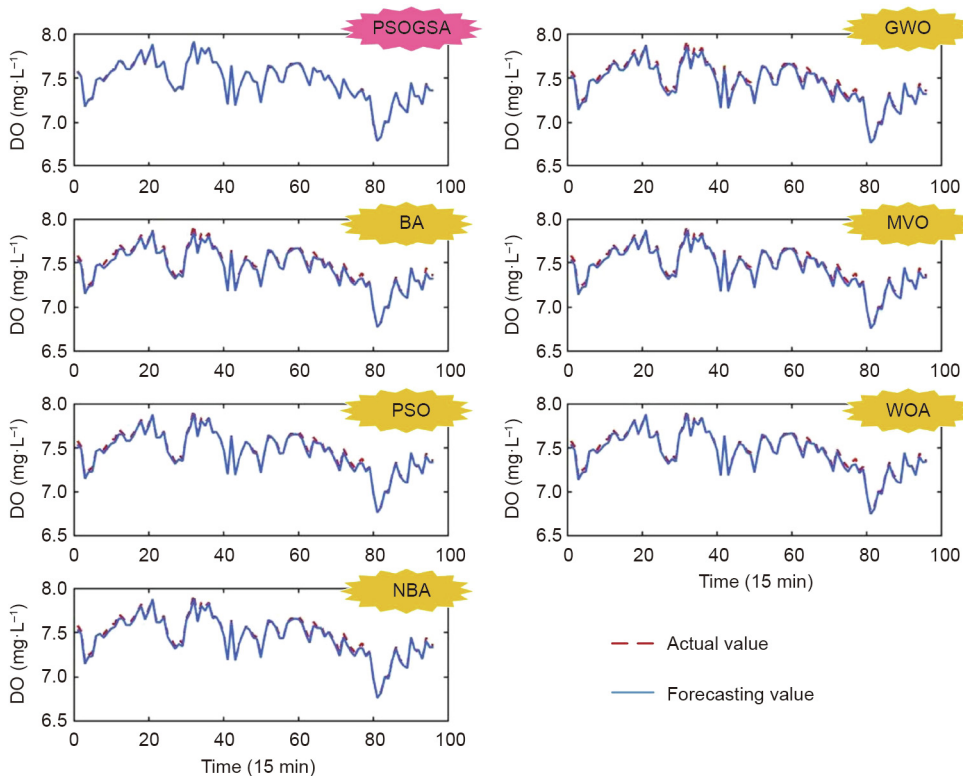
Fig. 6. Testing dataset forecasting results in 3-step of the PSOGSA and other optimization algorithms (the above-mentioned different optimization methods are all based on the multi-factor-EWT-BEGOE ensemble model to conduct control variable comparison experiments) (Dataset #1). Time (15 min) indicates that the sampling interval of data is 15 minutes.

Table 6
The DO testing dataset forecasting results of the proposed model and state-of-the-art models (Dataset #1).

Horizon	Model	NSE	KGE	MAPE (%)	SDE (mg·L^{-1})	R^2
1-step	SVM	0.62	0.75	1.41	0.14	0.63
	DBN	0.61	0.74	1.44	0.14	0.62
	Li's model [31]	0.95	0.94	0.52	0.05	0.95
	Huan's model [12]	0.95	0.96	0.56	0.05	0.95
	Ren's model [23]	0.95	0.96	0.51	0.05	0.95
	Proposed model	1.00	1.00	0.11	0.01	1.00
2-step	SVM	0.47	0.63	1.72	0.16	0.48
	DBN	0.45	0.61	1.75	0.16	0.48
	Li's model [31]	0.84	0.83	0.90	0.09	0.85
	Huan's model [12]	0.87	0.92	0.83	0.08	0.87
	Ren's model [23]	0.86	0.91	0.87	0.08	0.86
	Proposed model	0.99	0.99	0.14	0.02	1.00
3-step	SVM	0.29	0.51	1.99	0.19	0.33
	DBN	0.26	0.47	2.06	0.18	0.32
	Li's model [31]	0.79	0.76	1.05	0.10	0.80
	Huan's model [12]	0.85	0.91	0.89	0.08	0.86
	Ren's model [23]	0.73	0.85	1.21	0.12	0.74
	Proposed model	0.98	0.98	0.17	0.04	0.99

in forecasting robustness and can still perform satisfactorily even in higher prediction steps. Taking the 3-step forecasting results of Dataset #1 as an example, the MAPE of SVM, DBN, Li's model, Huan's model, Ren's model, and the proposed model are 1.99%, 2.06%, 1.05%, 0.89%, 1.21%, and 0.17%, respectively. Based on other persistence indices, the performance of the proposed model is very outstanding. Taking the 3-step forecasting results of Dataset #1 as an example, the NSE, KGE, MAPE, SDE, and R^2 of the proposed model are 0.98, 0.98, 0.17, 0.04 mg·L^{-1}, and 0.99, respectively. Taking the 3-step forecasting results of Dataset #2 as an example,

the NSE, KGE, MAPE, SDE, and R^2 of the proposed model are 0.96, 0.94, 0.58%, 0.04 mg·L^{-1}, and 0.98, respectively.

To sum up, the forecasting results of different datasets show that the proposed hybrid forecasting model has a reasonable framework and achieves a satisfactory forecasting performance. This is mainly because the proposed model hybridizes multi-factor analysis, adaptive decomposition analysis, and the scientific ensemble model analysis method. This combination is very helpful in improving the forecasting accuracy and robustness of the proposed model.

Table 7
DO testing dataset forecasting results of the proposed model and state-of-the-art models (Dataset #2).

| Horizon	Model	NSE	KGE	MAPE (%)	SDE (mg·L^{-1})	R^2
1-step	SVM	0.79	0.85	1.17	0.11	0.79
	DBN	0.79	0.89	1.14	0.11	0.80
	Li's model [31]	0.96	0.96	0.51	0.04	0.97
	Huan's model [12]	0.97	0.96	0.50	0.04	0.97
	Ren's model [23]	0.96	0.96	0.56	0.05	0.96
	Proposed model	1.00	0.98	0.18	0.01	1.00
2-step	SVM	0.64	0.75	1.65	0.15	0.64
	DBN	0.63	0.81	1.61	0.15	0.66
	Li's model [31]	0.91	0.91	0.81	0.07	0.92
	Huan's model [12]	0.95	0.94	0.62	0.06	0.95
	Ren's model [23]	0.92	0.94	0.77	0.07	0.92
	Proposed model	0.99	0.97	0.28	0.02	0.99
3-step	SVM	0.50	0.65	1.99	0.17	0.51
	DBN	0.46	0.72	1.95	0.18	0.53
	Li's model [31]	0.86	0.85	1.06	0.09	0.86
	Huan's model [12]	0.94	0.94	0.68	0.06	0.94
	Ren's model [23]	0.82	0.88	1.19	0.11	0.82
	Proposed model	0.96	0.94	0.58	0.04	0.98
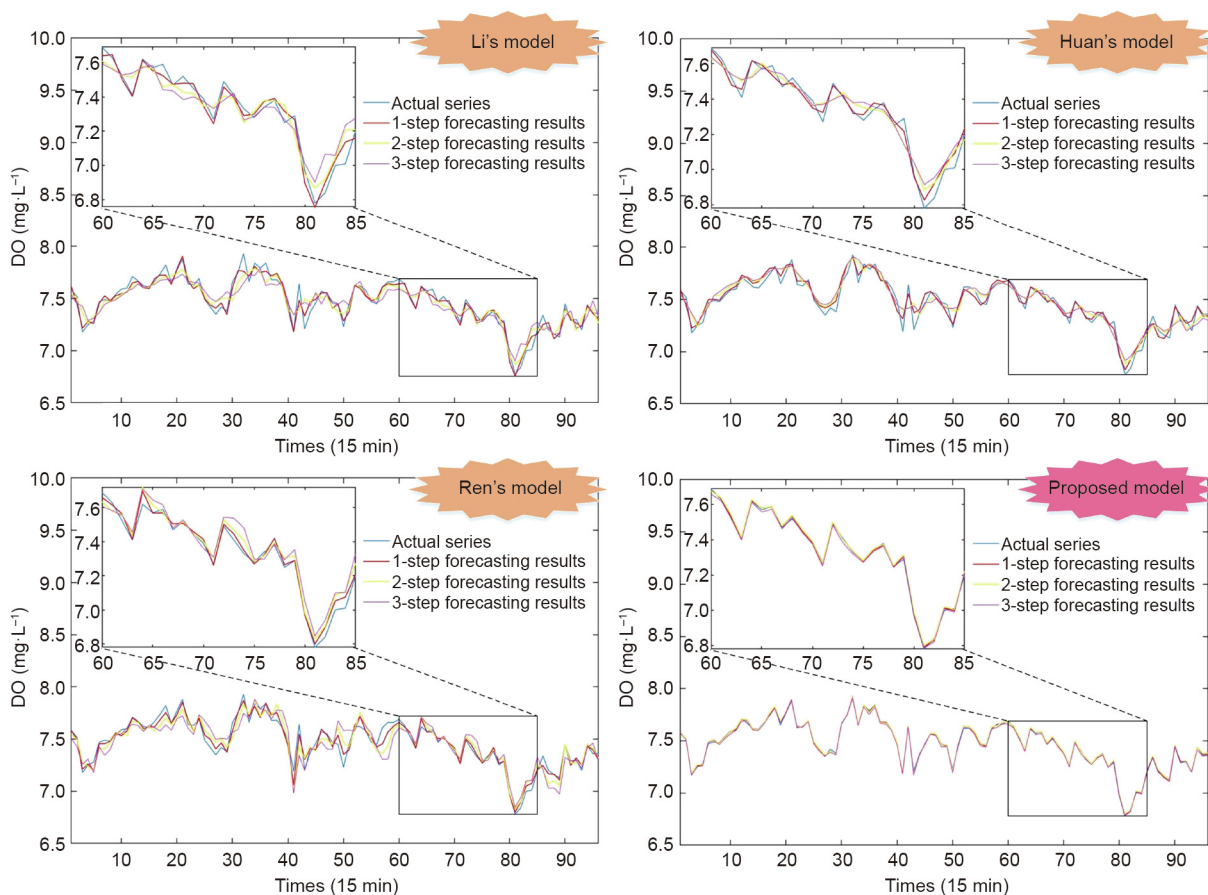
Fig. 7. DO testing dataset forecasting results of the proposed model (MF-RNNs-EWT-BEGOE) and other models (Dataset #1).

4. Application potential

The level of DO concentration in a marine environment is caused by two effects: the oxygen consumption effect, which decreases DO concentration; and the oxygen recovery effect, which increases DO concentration. The life activities of aquatic organisms consume a certain amount of oxygen, thereby reducing the DO concentration. The photosynthesis of aquatic plants and the diffusion of oxygen in the air will increase the DO concentration.

Fig. 9(a) shows the changes of actual DO time series data and 1-step, 2-step, and 3-step forecasting results. The following facts are illustrated in Fig. 9: ① As the temperature increases, the DO concentration decreases, and as the temperature decreases, the DO concentration increases. This is because higher temperatures will force DO in the water to escape to the surface (Fig. 9(b)). ② Salinity and DO are negatively correlated, due to oxygen consumption. Sufficient nutrients will promote the life activities of aquatic organisms, resulting in increased breathing and oxygen
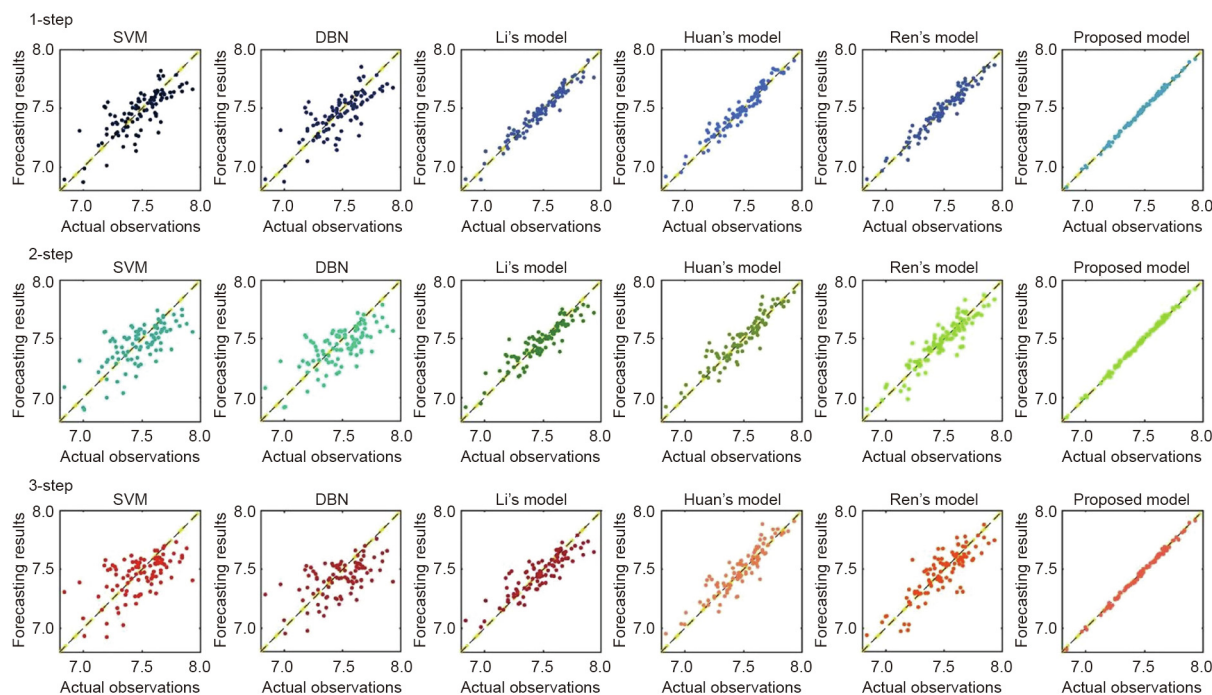
Fig. 8. Scatter plots of actual observations and corresponding forecasting results generated by the proposed model and other models.

consumption. An increase in salinity will also promote DO to a level closer to saturation (Fig. 9(c)). ③ High turbidity will reduce the photosynthesis of aquatic plants and affect the production of DO, so turbidity has a certain inhibitory relationship with the increase of DO (Fig. 9(d)). ④ The oxygen content dissolved in water is also known as the oxygen saturation. When the oxygen exchange between water and the atmosphere is in equilibrium, the concentration of DO in the water is most similar to the fluctuation of oxygen saturation (Fig. 9(e)). Mathematical modeling combined with AI technology can make a relatively scientific prediction of the changes in DO concentration. This prediction can provide useful guidance to people in related industries.

This study proposes a novel hybrid model that can generate reliable DO forecasting results. The model takes temperature, salinity, turbidity, oxygen saturation, and many other factors closely related to DO as input variables, and finally outputs a DO forecasting result. Fig. 9(f) shows the division of DO data and influence factors data. The specific multiple-input–single-output forecasting results are shown in Fig. 9. Potential applications for users of the proposed model are presented as follows:

- Aquaculture farmers can utilize the proposed DO forecasting model to determine future DO concentration changes in ponds and offshore aquaculture areas. They can then take corresponding emergency measures under guidance to reduce their economic losses.
- Fishery department managers can fully understand the situation of and trends in DO by means of accurate forecasting. This can provide a scientific decision-making basis for improving the living environment of marine life, which is conducive to the restoration and regulation of ecological water quality.

5. Conclusions

The development of DO forecasting technology has brought more possibilities to seafood farmers and related industries. Accurate real-time forecasting of DO can provide managers with scientific guidance for decision-making. Unfortunately, the instability of

the marine DO series and various other potential factors present challenges to DO forecasting. In our study, a novel hybrid model is proposed for DO forecasting: the MF-RNNs-EWT-BEGOE. The main characteristics of this model include multi-factor analysis, adaptive decomposition, and a multi-model ensemble. After the analysis of the experimental results in the case study, the following conclusions can be drawn:

(1) Considering that many factors affect DO concentration in the ocean, a multi-factor analysis method is proposed to incorporate the effects of these factors on DO concentration. The influence of abundant physical, chemical, and biological systems makes the DO fluctuation mechanism more complicated. Reasonable multi-factor consideration will make model forecasting more accurate.

(2) An adaptive data preprocessing decomposition method is used to reduce the instability of marine DO series, making sub-series easier to predict. The experimental results demonstrate that this method of forecasting after decomposing the original series is effective.

(3) A novel multi-model ensemble method is proposed to complete the combination of multiple models, which can absorb the excellent quality of each benchmark model to make the ensemble model more mature. A comparison with the experimental results of multiple benchmark models shows that the complementary advantages of multiple models can indeed cause the performance of the ensemble model to develop in a better direction.

(4) Each component of the proposed hybrid model can significantly improve the model's forecasting performance for DO. By detecting and eliminating outliers through RNNs, the negative impact of extreme outliers on the model can be reduced. The environmental factors that affect the change of DO can be considered more comprehensively through the multi-factor correlation analysis by GR. EWT adaptive decomposition effectively improves the forecasting accuracy of the model and reduces the impact of series non-stationarity. The BEGOE ensemble model combines the advantages of multiple benchmark models to maximize the model's forecasting performance. Thus, the proposed hybrid model is significantly better than the other models used for comparison.
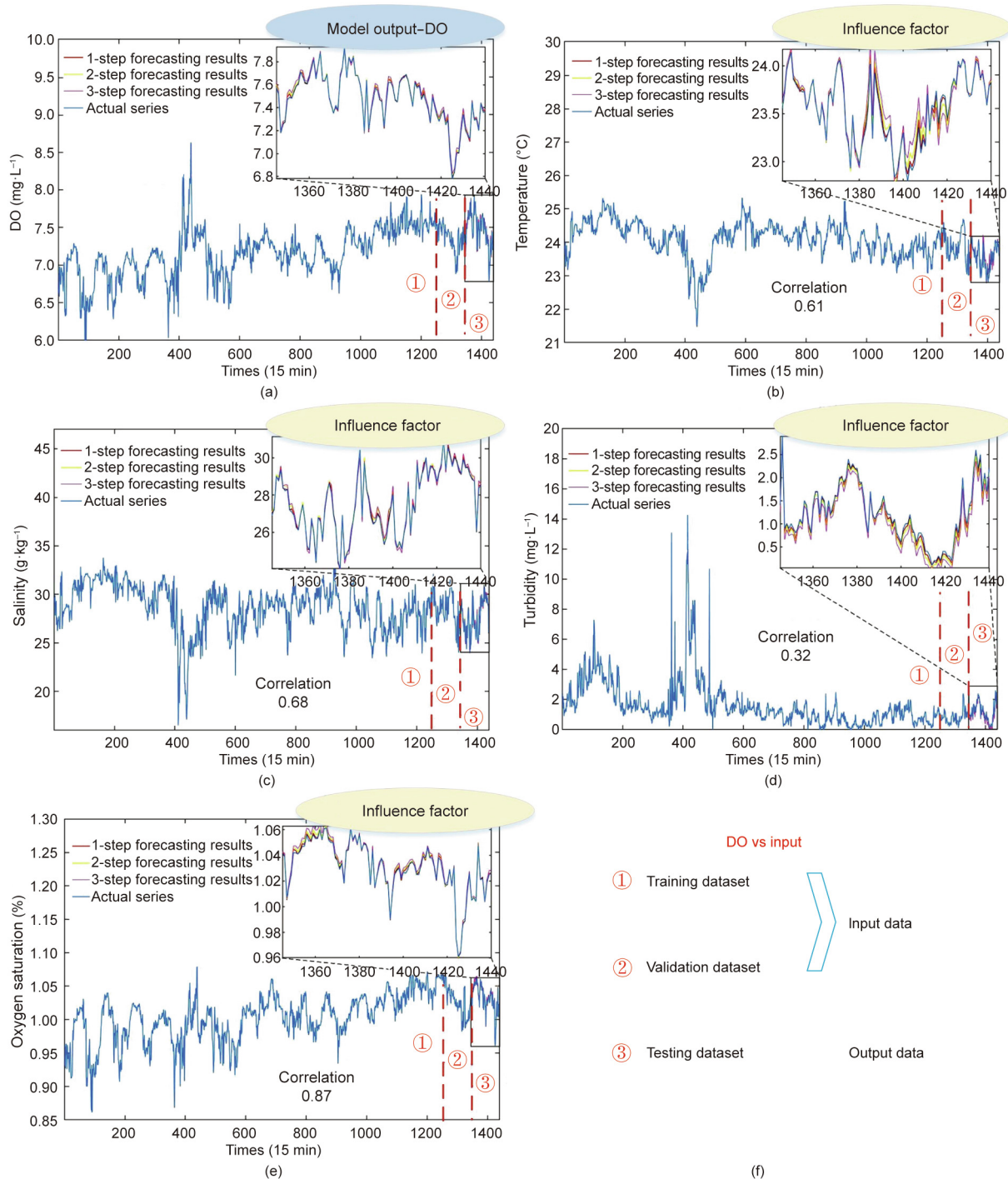
Fig. 9. DO vs input data change trend chart of the multiple–input–single–output model.

Given the potential for application described in Section 4, it is clear that accurate DO prediction can provide scientific guidance enabling aquaculture farmers to reduce their economic losses. Accurate DO forecasting can also provide a scientific basis for fishery department managers to help them improve the living environment of ponds and marine life. It can also help decision-makers to regulate and restore ecological water quality. The proposed hybrid model has excellent multi-step forecasting performance. The MIMOS mechanism provides sufficient information about future DO fluctuations for relevant industry personnel, thereby allowing effective management.

Acknowledgments

Compliance with ethics guidelines

Hui Liu, Rui Yang, Zhu Duan, and Haiping Wu declare that they have no conflict of interest or financial conflicts to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2020.10.023.

References

[1] Yang Z. Watershed ecology and its applications. Engineering 2018;4(5):582–3.
[2] Hoogakker BAA, Lu Z, Umling N, Jones L, Zhou X, Rickaby RE, et al. Glacial expansion of oxygen-depleted seawater in the eastern tropical Pacific. Nature 2018;562(7727):410–3.
[3] McClanahan TR, Ateweberhan M, Muhando CA, Maina J, Mohammed MS. Effects of climate and seawater temperature variation on coral bleaching and mortality. Ecol Monogr 2007;77(4):503–25.
[4] Gimpel A, Stelzenmüller V, Grote B, Buck BH, Floeter J, Núñez-Riboni I, et al. A GIS modelling framework to evaluate marine spatial planning scenarios: co-location of offshore wind farms and aquaculture in the German EEZ. Mar Policy 2015;55:102–15.
[5] Keller AA, Ciannelli L, Wakefield WW, Simon V, Barth JA, Pierce SD. Occurrence of demersal fishes in relation to near-bottom oxygen levels within the California Current large marine ecosystem. Fish Oceanogr 2015;24(2):162–76.
[6] Addanki SC, Venkataraman H. Greening the economy: a review of urban sustainability measures for developing new cities. Sustainable Cities Soc 2017;32:1–8.
[7] Schmidtko S, Stramma L, Visbeck M. Decline in global oceanic oxygen content during the past five decades. Nature 2017;542(7641):335–9.
[8] Culberson SD, Piedrahita RH. Aquaculture pond ecosystem model: temperature and dissolved oxygen prediction—mechanism and application. Ecol Modell 1996;89(1–3):231–58.
[9] Liu S, Xu L, Li D, Li Q, Jiang Yu, Tai H, et al. Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. Comput Electron Agric 2013;95:82–91.
[10] Faruk DÖ. A hybrid neural network and ARIMA model for water quality time series prediction. Eng Appl Artif Intell 2010;23(4):586–94.
[11] Li C, Li Z, Wu J, Zhu L, Yue J. A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features. Inf Process Agric 2018;5(1):11–20.
[12] Huan J, Cao W, Qin Y. Prediction of dissolved oxygen in aquaculture based on EEMD and LSSVM optimized by the Bayesian evidence framework. Comput Electron Agric 2018;150:257–65.
[13] Khan U, Valeo C, Comparing A. Bayesian and fuzzy number approach to uncertainty quantification in short-term dissolved oxygen prediction. J Environ Inform 2017;30(1):1–16.
[14] Khan UT, Valeo C, He J. Non-linear fuzzy-set based uncertainty propagation for improved DO prediction using multiple-linear regression. Stochastic Environ Res Risk Assess 2013;27(3):599–616.
[15] Kisi O, Parmar KS. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. J Hydrol 2016;534:104–12.
[16] Ay M, Kisi O. Modelling of chemical oxygen demand by using ANNs, ANFIS and *k*-means clustering techniques. J Hydrol 2014;511:279–89.
[17] Ay M, Kisi O. Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado. J Environ Eng 2012;138(6):654–62.
[18] Zhu S, Heddam S. Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN). Water Quality Res J 2020;55(1):106–18.
[19] Heddam S, Kisi O. Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. J Hydrol 2018;559:499–509.
[20] Ay M, Kişi Ö. Estimation of dissolved oxygen by using neural networks and neuro fuzzy computing techniques. KSCE J Civ Eng 2017;21(5):1631–9.
[21] Khan UT, Valeo C. Dissolved oxygen prediction using a possibility theory based fuzzy neural network. Hydrol Earth Syst Sci 2016;20(6):2267–93.
[22] Shi P, Li G, Yuan Y, Huang G, Kuang L. Prediction of dissolved oxygen content in aquaculture using Clustering-based Softplus Extreme Learning Machine. Comput Electron Agric 2019;157:329–38.
[23] Ren Q, Zhang L, Wei Y, Li D. A method for predicting dissolved oxygen in aquaculture water in an aquaponics system. Comput Electron Agric 2018;151:384–91.
[24] Wu J, Li Z, Zhu L, Li G, Niu B, Peng F. Optimized BP neural network for dissolved oxygen prediction. IFAC-PapersOnLine 2018;51(17):596–601.
[25] Liu S, Xu L, Jiang Yu, Li D, Chen Y, Li Z. A hybrid WA–CPSO-LSSVR model for dissolved oxygen content prediction in crab culture. Eng Appl Artif Intell 2014;29:114–24.
[26] Raheli B, Aalami MT, El-Shafie A, Ghorbani MA, Deo RC. Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: a case study of Langat River. Environ Earth Sci 2017;76(14):503.
[27] Yang H, Csukás B, Varga M, Kucska B, Szabó T, Li D. A quick condition adaptive soft sensor model with dual scale structure for dissolved oxygen simulation of recirculation aquaculture system. Comput Electron Agric 2019;162:807–24.
[28] Ma J, Ding Y, Cheng JCP, Jiang F, Xu Z. Soft detection of 5-day BOD with sparse matrix in city harbor water using deep learning techniques. Water Res 2020;170:115350.
[29] Ren Q, Wang X, Li W, Wei Y, An D. Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network. Aquacult Eng 2020;90:102085.
[30] Kisi O, Akbari N, Sanatipour M, Hashemi A, Teimourzadeh K, Shiri J. Modeling of dissolved oxygen in river water using artificial intelligence techniques. J Environ Inform 2013;22(2):92–101.
[31] Li W, Wu H, Zhu N, Jiang Y, Tan J, Guo Y. Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). Inf Process Agric 2020;8(1):185–93.
[32] Ta X, Wei Y. Research on a dissolved oxygen prediction method for recirculating aquaculture systems based on a convolution neural network. Comput Electron Agric 2018;145:302–10.
[33] Mandal S, Debnath M, Ray S, Ghosh PB, Roy M, Ray S. Dynamic modelling of dissolved oxygen in the creeks of Sagar island, Hooghly–Matla estuarine system, West Bengal, India. Appl Math Model 2012;36(12):5952–63.
[34] Liu H, Yang R, Duan Z. Wind speed forecasting using a new multi-factor fusion and multi-resolution ensemble model with real-time decomposition and adaptive error correction. Energy Convers Manage 2020;217:112995.
[35] Liu H, Yang R, Wang T, Zhang L. A hybrid neural network model for short-term wind speed forecasting based on decomposition, multi-learner ensemble, and adaptive multiple error corrections. Renew Energy 2021;165:573–94.
[36] Zounemat-Kermani M, Seo Y, Kim S, Ghorbani MA, Samadianfard S, Naghshara S, et al. Can decomposition approaches always enhance soft computing models? Predicting the dissolved oxygen concentration in the St. Johns River, Florida. Appl Sci 2019;9(12):2534.
[37] Gilles J. Empirical wavelet transform. IEEE Trans Signal Process 2013;61(16):3999–4010.
[38] Zhu JJ, Kang L, Anderson PR. Predicting influent biochemical oxygen demand: balancing energy demand and risk management. Water Res 2018;128:304–13.
[39] Kisi O, Alizamir M, Gorgij AD. Dissolved oxygen prediction using a new ensemble method. Environ Sci Pollut Res Int 2020;27(9):9589–603.
[40] Cao W, Huan J, Liu C, Qin Y, Wu F. A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multi-scale feature extraction. Aquacult Eng 2019;84:50–9.
[41] Chatfield C, Weigend AS. Time series prediction: forecasting the future and understanding the past: Neil A. Gershenfeld and Andreas S. Weigend, 1994, 'The future of time series', in: A.S. Weigend and N.A. Gershenfeld, eds., (Addison-Wesley, Reading, MA), 1-70. Int J Forecast 1994;10(1):161–3.
[42] Chatfield C. The future of the time-series forecasting. Int J Forecast 1988;4(3):411–9.
[43] Hawkins S, He H, Williams G, Baxter R. Outlier detection using replicator neural networks. In: Proceedings of International Conference on Data Warehousing and Knowledge Discovery; Kinsdale, Ireland. Berlin: Springer; 2002. p. 170–80.
[44] Hu J, Wang J, Zhang X, Fu Z. Research status and development trends of information technologies in aquacultures. Nongye Jixie Xuebao 2015;46(7):251–63.
[45] Ip WC, Hu BQ, Wong H, Xia J. Applications of grey relational method to river environment quality evaluation in China. J Hydrol 2009;379(3–4):284–90.
[46] Valentini G, Dietterich TG. Bias-variance analysis of support vector machines for the development of SVM-based ensemble methods. J Mach Learn Res 2004;5:725–75.
[47] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I—a discussion of principles. J Hydrol 1970;10(3):282–90.
[48] Knoben WJM, Freer JE, Woods RA. Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrol Earth Syst Sci 2019;23(10):4323–31.