

Prediction of seawater pH by bidirectional gated recurrent neural network with attention under phase space reconstruction: case study of the coastal waters of Beihai, China

Chongxuan Xu¹, Ying Chen^{1*}, Xueliang Zhao^{1,2}, Wenyang Song¹, Xiao Li¹

¹ Hebei Province Key Laboratory of Test/Measurement Technology and Instrument, School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China

² Center for Hydrogeology and Environmental Geology, China Geological Survey, Geological Environment Monitoring Engineering Technology Innovation Center of the Ministry of Natural Resources, Baoding 071051, China

Received 23 August 2022; accepted 2 February 2023

© Chinese Society for Oceanography and Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Marine life is very sensitive to changes in pH. Even slight changes can cause ecosystems to collapse. Therefore, understanding the future pH of seawater is of great significance for the protection of the marine environment. At present, the monitoring method of seawater pH has been matured. However, how to accurately predict future changes has been lacking effective solutions. Based on this, the model of bidirectional gated recurrent neural network with multi-headed self-attention based on improved complete ensemble empirical mode decomposition with adaptive noise combined with phase space reconstruction (ICPBGA) is proposed to achieve seawater pH prediction. To verify the validity of this model, pH data of two monitoring sites in the coastal sea area of Beihai, China are selected to verify the effect. At the same time, the ICPBGA model is compared with other excellent models for predicting chaotic time series, and root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) are used as performance evaluation indicators. The R^2 of the ICPBGA model at Sites 1 and 2 are above 0.9, and the prediction errors are also the smallest. The results show that the ICPBGA model has a wide range of applicability and the most satisfactory prediction effect. The prediction method in this paper can be further expanded and used to predict other marine environmental indicators.

Key words: seawater pH prediction, Bi-gated recurrent neural (GRU) model, phase space reconstruction, attention mechanism, improved complete ensemble empirical mode decomposition with adaptive noise

Citation: Xu Chongxuan, Chen Ying, Zhao Xueliang, Song Wenyang, Li Xiao. 2023. Prediction of seawater pH by bidirectional gated recurrent neural network with attention under phase space reconstruction: case study of the coastal waters of Beihai, China. *Acta Oceanologica Sinica*, 42(10): 1–11, doi: 10.1007/s13131-023-2149-y

1 Introduction

In recent years, as carbon emissions increase, especially the burning of fossil fuels, more and more CO₂ is absorbed by the oceans. While ocean uptake of CO₂ can reduce the extent of global warming, it also directly contributes to ocean acidification (Zhao et al., 2020). Seawater pH closely links ocean climate, redox, and biological activity (Halevy and Bachan, 2017). Even slight changes in seawater pH can be lethal for some species. Therefore, the accurate measurement and prediction of seawater pH are research hotspots. To accurately grasp the changes in seawater pH, people establish marine monitoring stations in some sea areas, and relatively mature seawater pH measurement methods are used to achieve dynamic monitoring (Tilbrook et al., 2019). However, how to use the existing data to accurately predict future changes in seawater pH lacks effective solutions.

The pH data measured by the marine monitoring station are a time series sampled at equal time intervals. Time series prediction is a research method that assumes that future values are directly related to past measurements (Singla et al., 2022b).

However, unlike ordinary time series predicting methods, the measured pH data are complex due to the complex and changeable marine environment. Currently, the prediction methods for hydrological data are mainly divided into two categories (Wang et al., 2017). One is the numerical prediction method, and the other is the data-driven prediction method. Numerical methods require more specialized knowledge of the hydrology and environment, so it is more complicated. However, the data-driven method is less difficult to predict and can obtain higher resolution prediction results, so it is widely studied (Jiang et al., 2021). Data-driven predicting approach mainly includes statistical principles and machine learning. Hydrological prediction models based on statistical principles have been around for a long time. Krzysztofowicz summarized the use of statistical principles and probabilistic processes to solve forecasting problems in hydrology (Krzysztofowicz, 2001). Yang et al. (2007) established a hydrological model based on statistical principles and Bayesian reasoning for the Chaohe Basin in China. This method overcomes the seasonal factors of hydrological prediction and im-

Foundation item: The National Natural Science Foundation of China under contract No. 62275228; the S&T Program of Hebei under contract Nos 19273901D and 20373301D; the Hebei Natural Science Foundation under contract No. F2020203066.

*Corresponding author, E-mail: chenying@ysu.edu.cn

proves the analysis effect of the model. Malik et al. (2019) successfully predicted drought events in Uttarakhand State, India, using a multi-scalar streamflow drought index combined with a multiple linear regression model.

With the development of the artificial neural network (ANN) method, it receives extensive attention due to its strong generalization ability and high computational efficiency (Gao et al., 2021). Because hydrological data, especially marine data, are complex and have apparent nonlinearity, it is very suitable to use ANN for research. Lee (2004) used the back-propagation neural network (BP) to achieve long-term prediction of ocean tide data and used data from Taichung Port in Taiwan Province to verify the validity of the model. Kajiyama et al. (2011) used multi-layer perceptron (MLP) to predict the execution time of a parallel Monte Carlo radiative transfer simulation code for marine color applications, exploiting the performance benefits of MLP to substantially reduce the overall execution time of the simulation job. Patil and Deo (2017) used a wavelet neural network (WNN) to predict daily sea surface temperature and compared traditional neural networks with WNN. It is found that WNN has satisfactory results for the prediction of sea surface temperature. The gated recurrent neural network (GRU), as a variant of Long-Short Term Memory (LSTM), is widely used in the prediction of marine data due to its high computational efficiency and strong temporal dependence on captured data. Xie et al. (2022) proposed a convolutional GRU neural network with MLP to predict the temperature of the sea surface, enabling it to capture both the temporal and spatial effects of the sea area. The model is verified with the sea surface temperature data of the Bohai Sea and the South China Sea, which proves the validity of the model. Singla et al. (2022a) proposed an integrated model utilizing the extended range of wavelet transform and bidirectional LSTM (Bi-LSTM) deep learning network to predict the first 24 h of global horizontal solar irradiance compared to the direct use of the reference model, independent LSTM, and Bi-LSTM. The prediction effect has been significantly improved. Li et al. (2022) achieved wave height prediction for multivariate time series by using GRU combined with feature selection, which is more robust than LSTM.

Due to the complexity of marine data and the inclusion of noise, some signal decomposition methods to extract the main features of the signal are widely used to process marine data. Alizadeh and Kavianpour (2015) combined wavelet decomposition and ANN to develop a wavelet-ANN model to predict the water quality parameters of seawater and obtained better prediction results and lower prediction costs. Duan et al. (2016) applied empirical mode decomposition (EMD) to ocean wave data processing and combined it with support vector regression to successfully predict significant wave heights. Compared with the model without EMD, it proves that EMD can effectively extract nonlinearity and non-stationarity in data. Wu et al. (2019) combined Complementary Ensemble Empirical Mode Decomposition (CEEMD), an improved method based on Ensemble Empirical Mode Decomposition (EEMD), and BP to predict the sea surface temperature. The effect is better than the combination of EEMD and BP.

The pH data of seawater has typical nonlinear and obvious chaotic characteristics. Traditional time domain analysis is difficult to obtain complete information on seawater pH data. Therefore, this paper uses chaos theory to predict the data in phase space. Prediction methods based on chaotic time series are widely used in many fields. Niu et al. (2017) applied phase space reconstruction (PSR) to the prediction problem of PM_{2.5} and combined it with the least squares support vector machine,

which obtained a better prediction effect than the baseline model. Peng and Xiang (2020) combined PSR with BP optimized by the genetic algorithm to predict traffic flow, which captured the complex characteristics of traffic flow, and obtained a satisfactory prediction effect.

Based on the above analysis, and considering the nonlinear, non-stationary, and chaotic characteristics of seawater pH data, this paper proposes the model of bidirectional gated recurrent neural network with multi-headed self-attention based on improved complete ensemble empirical mode decomposition with adaptive noise combined with PSR (ICEEMDAN-PSR-Bi-GRU with multi-headed self-attention). After this, it is referred to as the improved complete ensemble empirical mode decomposition with adaptive noise combined with phase space reconstruction (ICPBGA) model. This paper is organized as follows: Section 2 introduces the materials and basic methods used in this paper. Section 3 presents the complete model framework. Section 4 introduces preprocesses the seawater pH data. Data prediction and model comparison is carried out in Section 5. Conclusions based on the experiments are shown in Section 6.

2 Materials and methods

2.1 Acquisition of seawater pH data

This paper takes the actual seawater pH value measured by the marine monitoring station as the research object. The data comes from monitoring sites in the coastal waters of Beihai, China. In this article, two buoys with a distance of 50 m between Monitoring Sites 1 and 2 were used, both within a range of 2 km from the coast. The original data used in this paper is the whole year of 2019, and the sampling time is 30 min. However, due to some outliers in the data and considering that the change of the data in a short period in the virtual environment are not obvious, this paper averages the data every 4 h. The total amount of data obtained from a single monitoring point is 1 460.

The main equipment used in the process of data acquisition in this study is shown in Fig. 1. Figure 1a shows the EMM700 buoy produced by YSI (USA). Figure 1b shows the multi-parameter water quality analyzer mounted on the buoy, Model 6600V2 manufactured by YSI. The pH sensor is provided by the multi-parameter water quality analyzer with a glass electrode. The electrode calibration method is the two-point method, and the calibration period is three months. The data acquisition process is to place the calibrated buoy equipped with the pH sensor

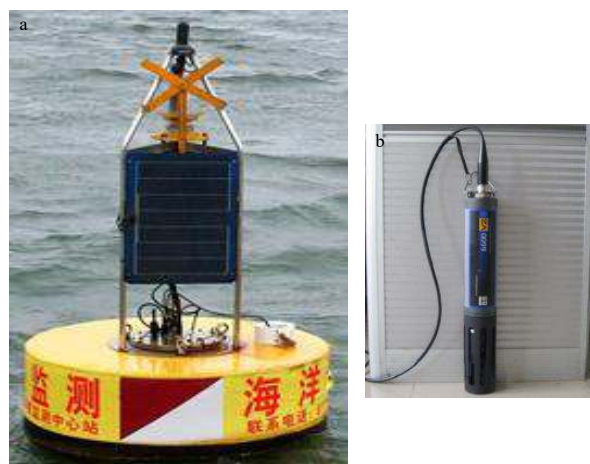


Fig. 1. Data acquisition equipment.

in a fixed sea area to monitor pH changes and then use the GPRS wireless transmission mode installed on the buoy for data transmission. The data acquisition software is EcoWatch. The data acquisition standards are adopted in Technical Specification for Environmental Monitoring of Coastal Waters (HJ442-2020) and Technical Specification for Automatic Monitoring of Water Quality of Coastal Waters (HJ731-2014) issued by the Ministry of Environmental Protection of China.

2.2 Improved complete ensemble empirical mode decomposition with adaptive noise

The CEEMDAN is an adaptive decomposition method that removes modal effects by adding adaptive noise. It has significant advantages over some common modal decomposition algorithms, such as EMD, EEMD, and CEEMD (Zhang et al., 2020). At the same time, it can decompose the sequence according to the time scale of the data and does not need to set any basis function. Therefore, it is suitable for the feature analysis of nonlinear and non-stationary time series. When the one-dimensional time series is decomposed by CEEMDAN, let the original time series be $x(t)$, and the decomposition result is shown in Eq. (1), where IMF_K is K intrinsic mode function (IMF), and $Res(t)$ is one residual (Res).

$$x(t) = \sum_{i=1}^K IMF_K + Res(t). \quad (1)$$

The sequence is composed of multiple IMFs and one residual after CEEMDAN.

Although the CEEMDAN effectively reduces the mode aliasing, it still cannot eliminate the influence of redundant and spurious components. This interferes with principal component screening. Therefore, to effectively select the principal component, the ICEEMDAN is proposed. When improving CEEMDAN, most scholars adopted the Pearson correlation coefficient to screen the optimal IMF (Cerrada et al., 2018). However, this method ignores the features existing in IMF, which makes it difficult to extract effective information completely. Therefore, this paper proposes to use permutation entropy (PE) to select the optimal IMF to improve CEEMDAN. PE can represent the complexity of the time series. The PE is a measurement method based on information theory, which estimates the complexity of the time series by calculating the arrangement information of adjacent values (Bandt and Pompe, 2002). Generally speaking, the larger PE is, the more noise it contains. However, if the noise term is removed directly, the feature will be lost. Therefore, we conducted a secondary reconstruction of the IMF component according to the size of PE and obtained three groups of new components.

The PE normalization formula is shown in Eq. (2). The normalized PE value is between 0 and 1. When the value of PE is closer to 1, it indicates that the complexity of the series is higher. When the value of PE is more relative to 0, it suggests that the complexity of the series is lower.

$$PE = \frac{PE(m)}{\ln(m!)} = \frac{-\sum_{i=1}^{m!} P_g \ln P_g}{\ln(m!)}, \quad (2)$$

where m is the embedded dimension, and P_g is the probability of the reconstructed component appearing in a specific column after the matrix is reconstructed. Its value is related to the choice

of the embedded dimension m and time delay, where $m = 3, 4, \dots, 7$, and the time delay is generally greater than or equal to 1 (Moscato-Lopez et al., 2019). Therefore, when calculating the PE, m and time delay are set to 3 and 1, respectively.

2.3 Phase space reconstruction

Chaos is a common phenomenon in production and life and has a certain system of irregularity and randomness (Sun and Wang, 2018). Chaotic time series is a typical representative of chaotic phenomena (Packard et al., 1980). Studying chaotic time series directly in the time domain is often easy to cause the loss of features, and it is difficult to obtain better research results. The embedding theorem proposed by Takens points out that it is possible to reconstruct a phase space from the one-dimensional chaotic time series that is the same as the motive power system in the topological sense. And the research on chaotic time series is carried out in the phase space (Takens, 1981). Therefore, the PSR is the key to studying the chaotic time series problem. There are two critical parameters in PSR, time delay τ and embedding dimension m . For the time series $\{x_i\} = \{x_1, x_2, \dots, x_N\}$, the reconstructed X and the output Y in the phase space can be expressed as

$$X = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(m-1)\tau} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N-m+1} & x_{N-m+1+\tau} & \cdots & x_{N-m+1+(m-1)\tau} \end{bmatrix}, \quad (3)$$

$$Y = \begin{bmatrix} x_{2+(m-1)\tau} \\ x_{3+(m-1)\tau} \\ \vdots \\ x_{N+1} \end{bmatrix}. \quad (4)$$

Analyzing the characteristics of time series, this paper employs the mutual information method to determine the time delay τ and Cao to determine the embedding dimension m . The mutual information method employs the mutual information function to express the dependency between the front and rear points in the time series and selects the time value corresponding to the first minimum value of the function as the τ (Fraser and Swinney, 1986). This method is adopted to determine the τ due to its simple calculation and is suitable for nonlinear and high-dimensional data.

The Cao overcomes the problem of sensitivity to data noise of the False Nearest Neighbors (Cao, 1997). When using the Cao, only the τ needs to be known. And the m can be calculated with a small amount of data. Therefore, this method is adopted.

2.4 Bidirectional gated recurrent neural network with multi-headed self-attention

2.4.1 Bidirectional gated recurrent neural network

In practical applications, time series usually show non-stationarity and nonlinearity, but traditional time series prediction methods are difficult to effectively capture its characteristics and achieve accurate prediction (Chen et al., 2023). The proposal of ANN provides a more effective strategy for solving time series predicting problems (Xu et al., 2021). As a method that can capture time dependence, the GRU neural network has the characteristics of high computational efficiency and accurate grasp of time characteristics, so it is suitable to solve the prediction problem of complex time series such as seawater pH value. The GRU neural network can only predict the output of the next moment based on the information of the previous moment (Cho et al.,

2014). However, in practice, the output of the current moment is not only affected by the previous moment but also may be related to the state of the future moment (Liu et al., 2020). Therefore, this paper proposes to employ the bidirectional GRU (Bi-GRU) neural network to predict time series. Bi-GRU neural network can divide neurons into forward layer direction and backward layer direction. The forward layer fetches the past information while backward layer fetches the future information of the input sequential data. Therefore, Bi-GRU makes predictions from the whole time series. The structure diagram of the Bi-GRU neural network is shown in Fig. 2, and the structure of the GRU is drawn in the figure. The state of hidden layers of Bi-GRU at present time t can be expressed as:

$$\text{Forward layer : } h_{ft} = \text{GRU}_f(X_t, H_{t-1}), \quad (5)$$

$$\text{Backward layer : } h_{bt} = \text{GRU}_b(X_t, H_{t-1}), \quad (6)$$

$$\text{Output layer : } h_t = h_{ft} + h_{bt}, \quad (7)$$

where H represents two candidate hidden states in Bi-GRU with opposite directions: candidate hidden state is an updated hidden state of a candidate. It is calculated based on the current input and the output of the reset gate. The candidate hidden state captures the correlation between the current input and past hidden states, and to some extent updates the hidden state.

2.4.2 Multi-head self-attention

For the prediction of seawater pH value, the value closer to the prediction time point has a more significant greater impact on the prediction result and should be given a higher weight. Therefore, this paper proposes to employ multi-head self-attention as a layer of the Bi-GRU neural network for prediction, to improve the prediction effect and enhance the rationality and interpretability of the model. The structure diagram of multi-head self-attention is shown in Fig. 3.

The multi-head self-attention is composed of self-attention, and the part within the dotted line in Fig. 3 is composed of self-attention modules with different weights (Vaswani et al., 2017), when Q is the query matrix, K is the key matrix, and V is the value matrix. For an input sequence of length N and dimension D_K , the formula for multi-head self-attention is

$$Q = XW^Q, \quad (8)$$

$$K = XW^K, \quad (9)$$

$$V = XW^V, \quad (10)$$

$$\text{self-attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{D_K}}\right)V, \quad (11)$$

$$\text{head}_i = \text{self-attention}(QW_i^Q, KW_i^K, VW_i^V), \quad (12)$$

$$\text{multi-headed self-attention}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^0, \quad (13)$$

where X is the input matrix. W^Q , W^K , and W^V are the weight matrices of Q , K , and V , respectively. Softmax represents the normalization function to improve learning efficiency. W_i^Q , W_i^K , and W_i^V are the weight matrices of i transformations. W^0 is the weight matrix after multi-head self-attention matrix splicing.

3 The framework of the model

The framework of the ICPBGA model is shown in Fig. 4. This gives new weights. Figure 4b is PSR. The new component decomposed by ICEEMDAN performs PSR, extracts the chaotic characteristics in the data, and constructs the input of the neural network. Figure 4c is Bi-GRU with multi-headed self-attention. This part predicts the new components after PSR. Using global prediction and the efficient ability of Bi-GRU to capture time dependencies to obtain prediction results of new components and add multi-headed self-attention to the network makes the model more reasonable in interpretability and prediction ability. The final prediction result of the ICPBGA model for seawater pH can be obtained by fusing the predictions of each component.

4 Data preprocessing

4.1 Decomposition of seawater pH data

Affected by the actual monitoring environment, the pH value data of seawater often have obvious nonlinearity and noise. Therefore, this paper proposes the ICEEMDAN to reduce the nonlinearity in the data and improve the signal-to-noise ratio. First, use CEEMDAN to decompose the original pH data of monitoring Site 1. When CEEMDAN is used, the PyEMD decomposition module (the function library called when using modal decomposition in Python programs) is used. The test times of adding noise are 20, the power threshold is 0.05, the amplitude threshold is 0.01, the white noise amplitude is 1, and other parameters are the initial parameters of the module. The decomposition result is shown in Fig. 5. The original data is decomposed into 7 IMFs, denoted as IMF_K , $K=1, 2, \dots, 7$, and one residual. Then, the PE of each IMFs is calculated, and the calculation results are

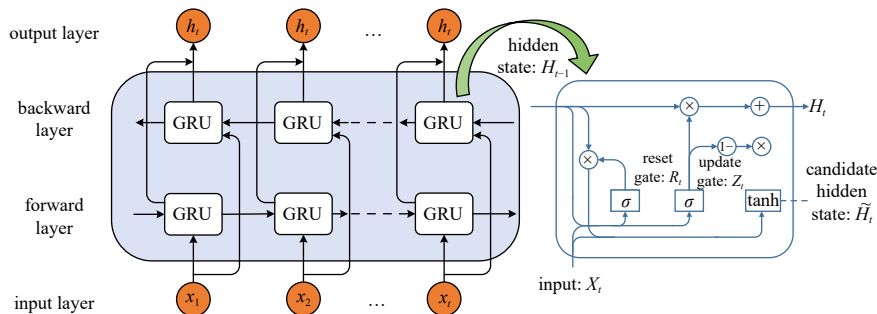


Fig. 2. The structure diagram of the bidirectional GRU neural network. σ represents the sigmoid function, which can convert data into values in the range of 0–1.

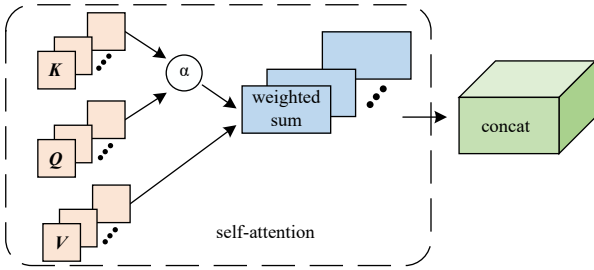


Fig. 3. The structure diagram of multi-head self-attention. α represents the degree of correlation between K and Q .

shown in [Table 1](#).

By calculating the average value of the original components of PE, the original components can be recombined into new components (n-coms) according to the complexity of the sequence, and the recombination results of the new components are shown in [Table 2](#). Visualize the new components and obtain the decomposed components of ICEEMDAN, as shown in [Fig. 6](#). According to the information in [Fig. 5](#), the n-com3 already contains most of the features of the time series and less noise. The tendency of the actual with time is the same. Although n-com1 and n-com2 contain more noise, they contain high-frequency components and have less impact on the prediction result. After that, ICEEMDAN's results and CEEMDAN's Res are correlated

with the original data. The contribution to the original data is calculated respectively. Contribution represents the amount of information that the decomposed data are contained in the original data. The greater the contribution, the greater the amount of information contained in the original data, and the more important the data is to the original data. Contribution is calculated as follows:

$$C_i = \frac{\text{Corr}_i}{\sum_{i=1}^n \text{Corr}_i}, \quad (14)$$

where C_i is the contribution of the i th component. Corr_i is the correlation coefficient with raw data of the i th component.

The calculation results are shown in [Table 3](#). According to the results, it can be found that the sum of contribution of the n-coms exceeds 99% and contains almost all the information. The contribution of Res is less than 1%, which can be regarded as noise mixed with the original data. So Res can be ignored when making predictions.

4.2 Phase space reconstruction

When reconstructing the phase space of the new component, the mutual information method is used to determine the τ first. The changing trend of the delay mutual information of the new

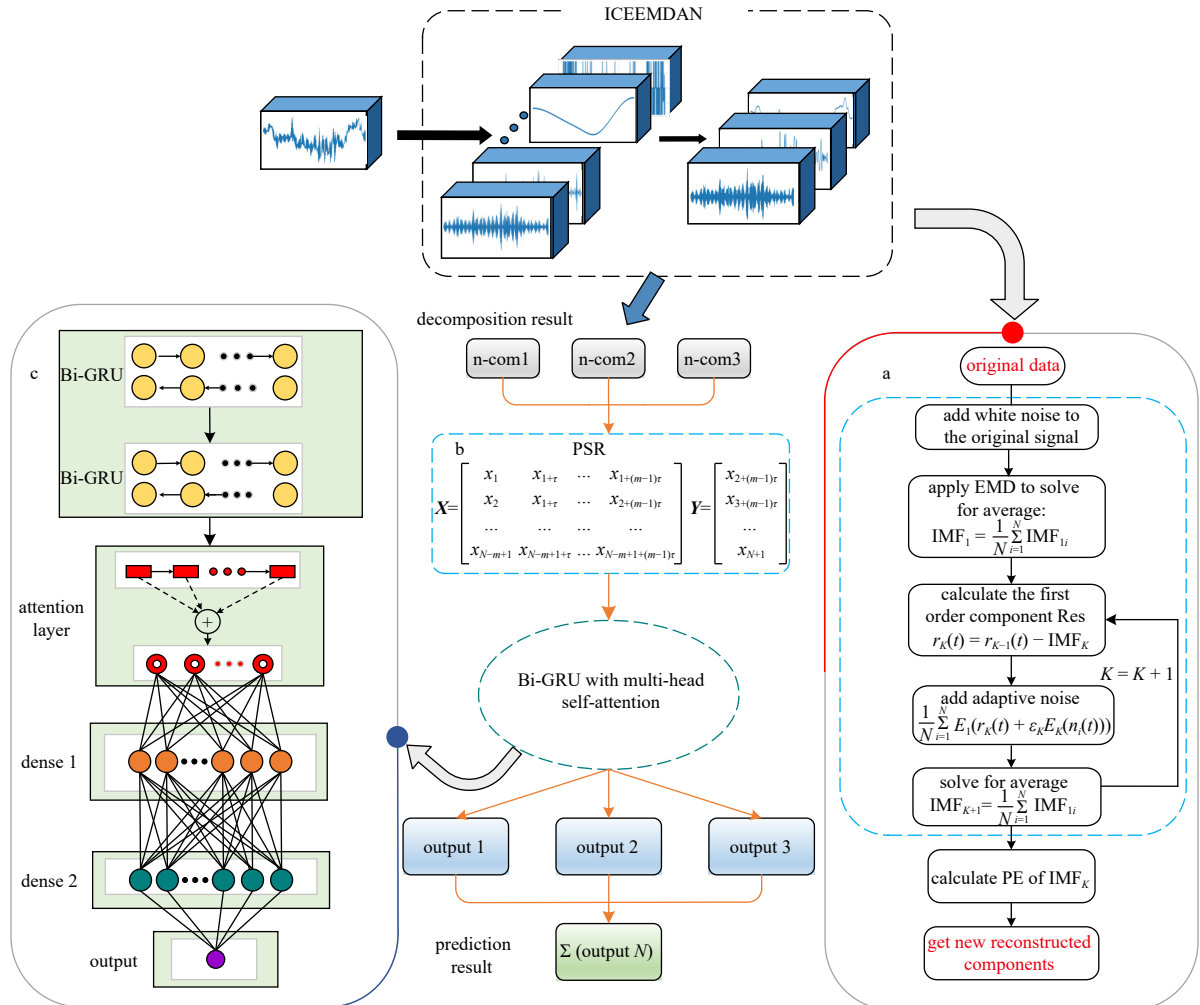


Fig. 4. The framework of the ICPBGA model.

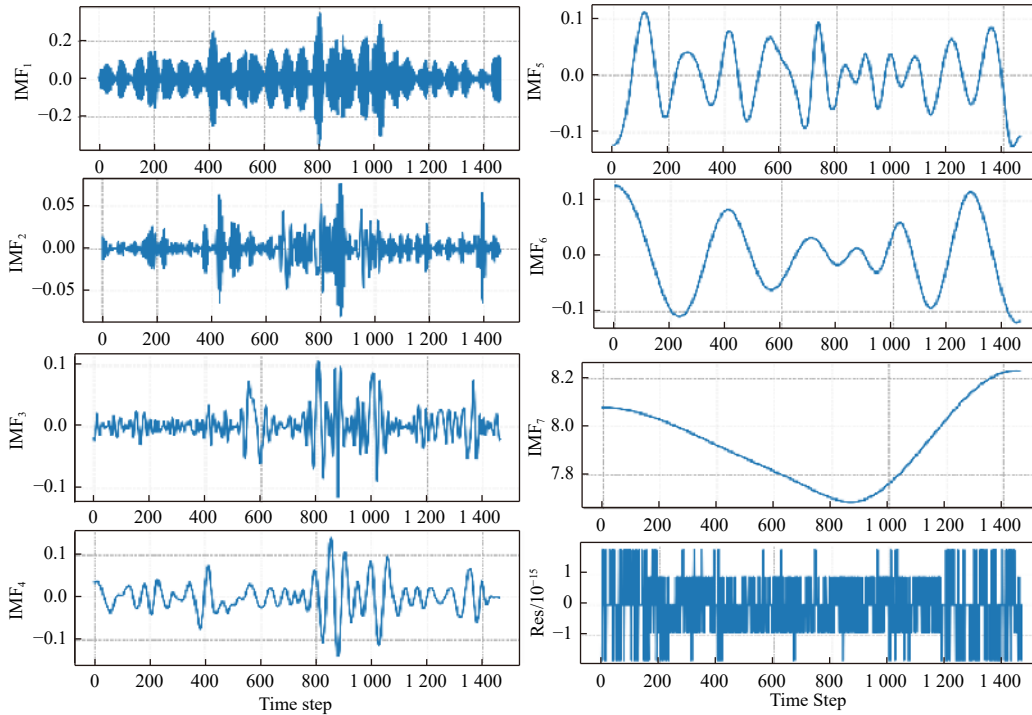


Fig. 5. The decomposition result of CEEMDAN.

Table 1. The PE value of CEEMDAN components

Component	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅	IMF ₆	IMF ₇
Value	0.980 1	0.808 4	0.610 3	0.496 3	0.437 9	0.411 4	0.380 7

Table 2. The new components result of ICEEMDAN

Index	n-com1	n-com2	n-com3
Result	IMF ₁ , IMF ₂	IMF ₃	IMF ₄ , IMF ₅ , IMF ₆ and IMF ₇

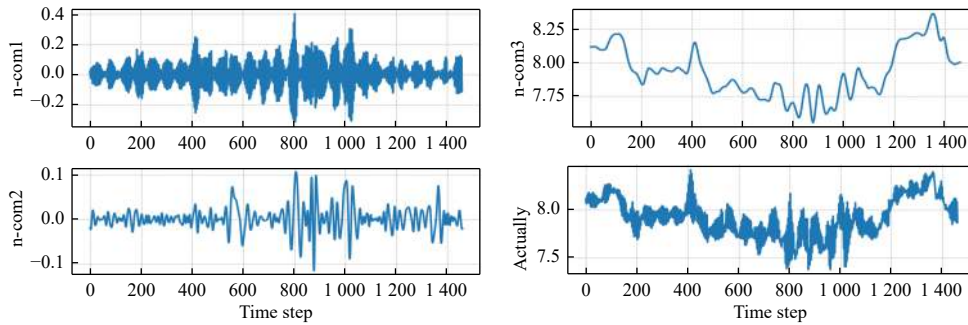


Fig. 6. The decomposition result of ICEEMDAN.

Table 3. The new components result of ICEEMDAN

ICEEMDAN component	Corr _i	C _i
n-com1	0.153	10.8%
n-com2	0.298	21.0%
n-com3	0.955	67.4%
Res	0.011	0.8%

components with the τ is shown in Fig. 7. The τ corresponding to the first minimum value of delay mutual information is the time delay of new components. After that, employ the Cao to determine the m . Substitute the τ calculated by the mutual information method into the Cao, and obtain the variation trend of $E(d)$ with the d as shown in Fig. 8. When $E_2(d)$ is not constant at 1. $E_1(d)$ increases with d , and when d rises to a certain value, $E_1(d)$ does not change, $d+1$ is the m currently. The time delay τ and the embedding dimension m of the new component are shown in Table 4.

5 Predict results and discussion

5.1 Model performance evaluation

To evaluate the performance of the prediction model and select a better model, this paper adopts root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) as the model performance metric. RMSE can reflect the precision of the prediction model. MAE can avoid the error canceling each other, to accurately reflect the size of the actual prediction error. MAPE can reduce the impact of the absolute error caused by a separate outlier. R^2 reflects the proportion of the total variation of the dependent variable that can be explained by the independent variable through the regression relationship (Yaseen, 2021). It can be known that the smaller the RMSE, MAE, and MAPE, the better the prediction effect is. The closer the R^2 is to 1, the higher the prediction accuracy is. The four metrics are defined by

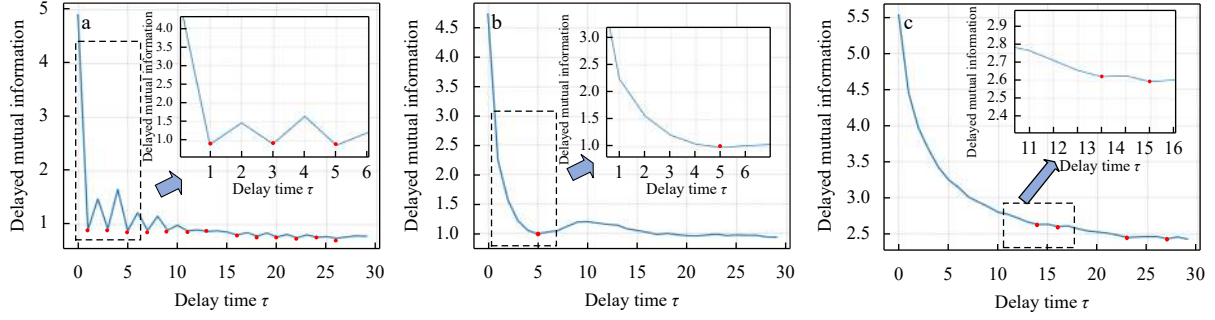


Fig. 7. Determine the delay time τ by the mutual information method. a, b, and c are the change trends of the delay mutual information of new components (n-com1, n-com2, and n-com3) with time delay τ .

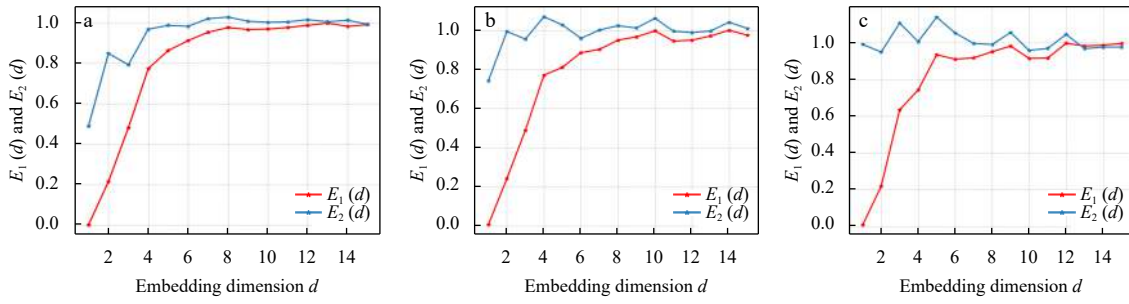


Fig. 8. Determine embedding dimension m by the Cao. a, b, and c are the change trends of the $E(d)$ of new components (n-com1, n-com2, and n-com3) with embedding dimension d .

Table 4. The parameters for new components (n-coms)

Index	n-com1	n-com2	n-com3
τ	1	5	14
m	8	9	6

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (15)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (16)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{\hat{y}_i} \right| \times 100\%, \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{\sum_{i=1}^n (y_i - \bar{y})}, \quad (18)$$

where \hat{y}_i is the predicted value, \bar{y} is the average of the actual values, y_i is the actual value, and n is the number of samples.

5.2 Hyperparameter settings

In this paper, Bi-GRU with multi-head self-attention is established for the components of PSR, and the hyperparameters of the prediction model of all components are shared. When dividing the training set and the test set, this paper selects 75% of the data as the training set and 25% as the test set. This partitioning choice combines training data volume and conventional partitioning methods (Keshan et al., 2015). This paper employs multi-head

self-attention as a layer of the neural network after the Bi-GRU layer. In this paper, a neural network with four hidden layers is built. The first two layers are Bi-GRU layers, and the number of neurons is 100. The third layer is the attention layer. The last layer is the dense layer, and the number of neurons is 50. The advantage of four hidden layers is that any smooth map with any precision can be fitted. The activation function of each layer adopts relu. At the same time, to prevent the model from overfitting, a dropout layer is added between the layers of the network, and the inactivation rate of the dropout layer is set to 0.2 (Chen et al., 2019). The input dimension of the neural network is the embedding dimension m of phase space reconstruction, and the number of output variables is 1. The optimizer selects the Adam, adopts the initial learning rate, and uses the average of the sum of squared errors (MSE) as the loss function. MSE is a common loss function. Moreover, as the error decreases, so does the gradient, which is good for convergence. The number of the epoch is 100.

5.3 Model comparison study

5.3.1 Ablation study

An ablation study has important implications for proving causality in complex models (Chaudhuri et al., 2021). The ablated variants of the ICPBGA model are GRU, Bi-GRU with attention, and CEEMDAN-Bi-GRU-attention (CBGA). The ablated variants are in phase space. The results of the ablation studies are shown in Fig. 9a, while the relationship of the actual, ablated variants and the ICPBGA model are plotted in Fig. 9b. From Fig. 9a, it can be found that compared with the ablated variants, the prediction effect of the ICPBGA model has an advantage, and the prediction effect of reducing any part can decrease. At the same time, compared with the prediction model without ICEEMDAN, the prediction model can effectively extract the main features of the seawater pH data and reduce the cumulative error during

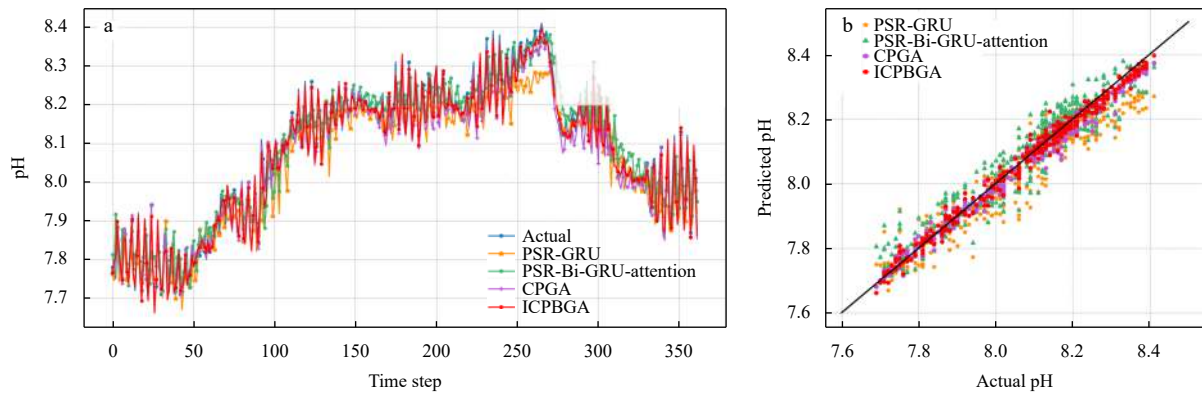


Fig. 9. Ablation study results. a is the prediction result graph of each model; b is the relationship between the actual value and the predicted value. CPGA: CEEMDAN-PSR-GRU with multi-headed self-attention.

prediction. From Fig. 9b, it can be found that the scatter distribution of the ICPBGA model is closer to the straight line, which also proves the advantages of the prediction model.

To quantitatively evaluate the effect of the prediction model, the performance evaluation metrics of the above models are calculated in this paper, as shown in Table 5. It can be seen from the calculation results that the RMSE, MAE, and MAPE of the ICPBGA model are smaller than other ablated variants, and R^2 is the closest to 1. Therefore, the model can effectively extract the chaotic characteristics of the pH data of seawater and achieve accurate prediction.

As a visual evaluation method of model effect that can comprehensively respond, the Taylor diagram is suitable for model evaluation in the field of hydrology (Sadeghifar et al., 2022).

Table 5. The performance evaluation metrics of models

Model	Performance evaluation metric			
	RMSE	MAE	MAPE/ 10^{-3}	R^2
PSR-GRU	0.063	0.037	6.077	0.848
PSR-Bi-GRU-attention	0.055	0.093	5.501	0.898
CPGA	0.027	0.017	2.568	0.975
ICPBGA	0.021	0.013	2.001	0.985

Note: CPGA: CEEMDAN-PSR-GRU with multi-headed self-attention.

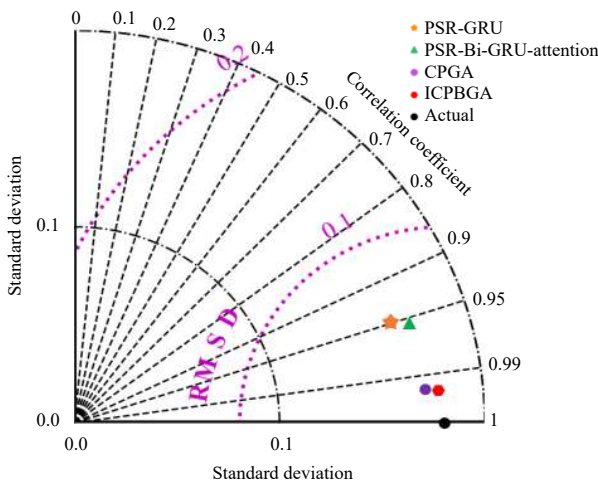


Fig. 10. Taylor diagram of the ablation experiment. CPGA: CEEMDAN-PSR-GRU with multi-headed self-attention; RMSD: root mean square difference.

Therefore, this paper draws the Taylor diagram of the ablation experiment, as shown in Fig. 10. It can be seen from Fig. 9 that the ICPBGA model has the highest prediction accuracy, and the prediction effect is better than other ablated variants.

5.3.2 Comparison of predicting the effect of monitoring sites

By fusing the prediction results reconstructed in the phase space of different components, the prediction results of the pH data of the seawater at Monitoring Site 1 can be obtained. To illustrate the predictive advantage of the combined model in this paper, the ICPBGA model is compared with other excellent models in handling chaotic time series. The comparison models in this paper are PSR and back propagation neural network (PSR-BP) (Peng et al., 2020), PSR and LSTM neural network (PSR-LSTM, Hu et al., 2022), PSR and temporal convolutional neural network (PSR-TCN, Dai et al., 2021), and PSR and GRU neural network (PSR-GRU, Du et al., 2021). The comparison results of each model are shown in Fig. 11. Through the analysis of the prediction results in Fig. 11a, the following conclusions are drawn. Compared with other models, the combined prediction model

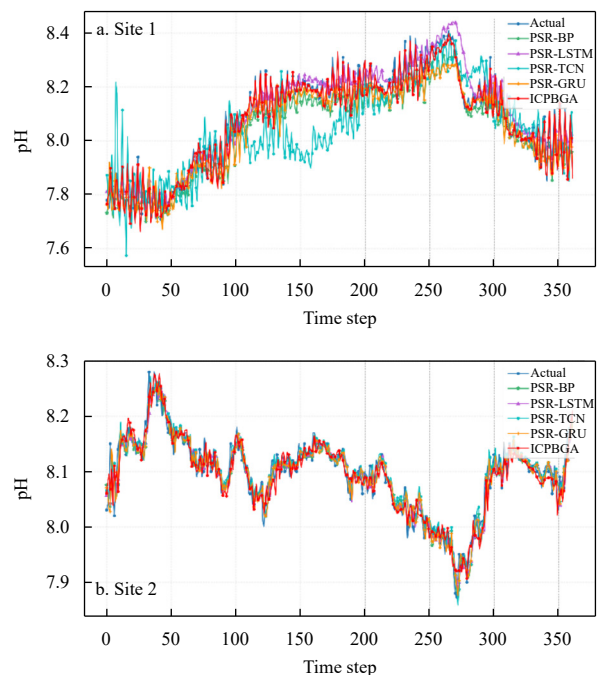


Fig. 11. Comparison of prediction results of various models.

proposed is more suitable for predicting the actual pH data of seawater and has a better prediction effect. Among them, the prediction effect of PSR-TCN is the worst, probably because TCN cannot effectively capture the data in phase space. The structure of BP is simple and cannot adapt to complex data in phase space, so it is difficult to obtain better prediction results. The prediction effect of LSTM and GRU is similar, but the structure of GRU is simpler, so the prediction efficiency can be better.

At the same time, to verify the applicability of the model, the seawater pH data of Monitoring Site 2 are introduced. The data preprocessing method for Monitoring Site 2 is the same as that for Monitoring Site 1. Therefore, it is not described here. The comparison diagram of the final prediction results is shown in

Table 6. The performance evaluation metrics of models

Model	Performance evaluation metric				
	RMSE	MAE	MAPE/ 10^{-3}	R^2	
Monitoring Site 1	PSR-BP	0.070	0.047	6.918	0.816
	PSR-LSTM	0.063	0.044	6.301	0.881
	PSR-TCN	0.117	0.075	11.278	0.465
	PSR-GRU	0.063	0.037	6.078	0.848
	ICPBGA	0.021	0.013	2.001	0.985
Monitoring Site 2	PSR-BP	0.024	0.015	2.257	0.871
	PSR-LSTM	0.023	0.014	2.197	0.880
	PSR-TCN	0.025	0.015	2.358	0.865
	PSR-GRU	0.023	0.014	2.272	0.875
	ICPBGA	0.019	0.010	1.176	0.921

Fig. 11b. The results in the figure further illustrate that the ICPBGA model has a good prediction effect and wide applicability.

To quantitatively evaluate the effect of the prediction model, the performance evaluation metrics of the above models are calculated, and the calculation results of Monitoring Sites 1 and 2 are shown in Table 6. It should be noted that the result of MAPE calculated in this paper is not a percentage but 1×10^{-3} . From the data in the table, the RMSE, MAE, and MAPE of the ICPBGA model at Monitoring Site 1 are 0.021, 0.013, and 0.002, respectively, and the RMSE, MAE, and MAPE at Monitoring Site 2 are 0.019, 0.010, and 0.001, respectively. Therefore, the model has the best prediction effect. At the same time, the R^2 at Monitoring Site 1 is 0.985, and the R^2 at Monitoring Site 2 is 0.921, both of which are the closest to 1. This indicates that the model has the best prediction accuracy. Compared with the worst performing PSR-TCN model, the prediction effect of the ICPBGA model at Monitoring Site 1 is improved by more than 5 times, and the prediction effect at Monitoring Site 2 is improved by more than 1 time. To better show how close the predicted value is to the actual value, the relationship between them is drawn in Fig. 12. The real value in the figure is a straight line with a slope of 1 passing through the origin. The more concentrated the distribution of the predicted value is on the straight line, the better the prediction effect, so the ICPBGA model has the best effect.

At the same time, this paper draws the prediction error violin plot of Monitoring Sites 1 and 2 to realize the visualization of the prediction error. The violin plot, as a combination of the box plot and kernel density plot, can effectively describe the distribution of data (Shan et al., 2022). The violin plots of different models are shown in Fig. 13. The results in Fig. 13 show that the ICPBGA

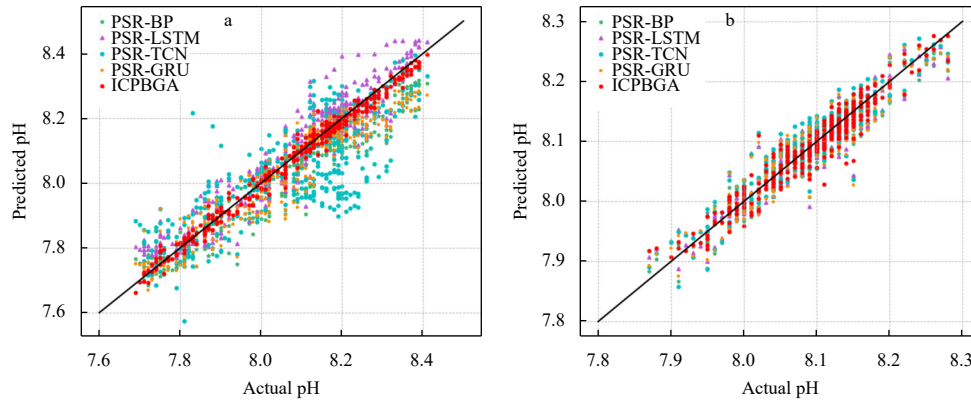


Fig. 12. The relationship between the actual value and the predicted value. a. The result of Monitoring Site 1; b. the result of Monitoring Site 2.

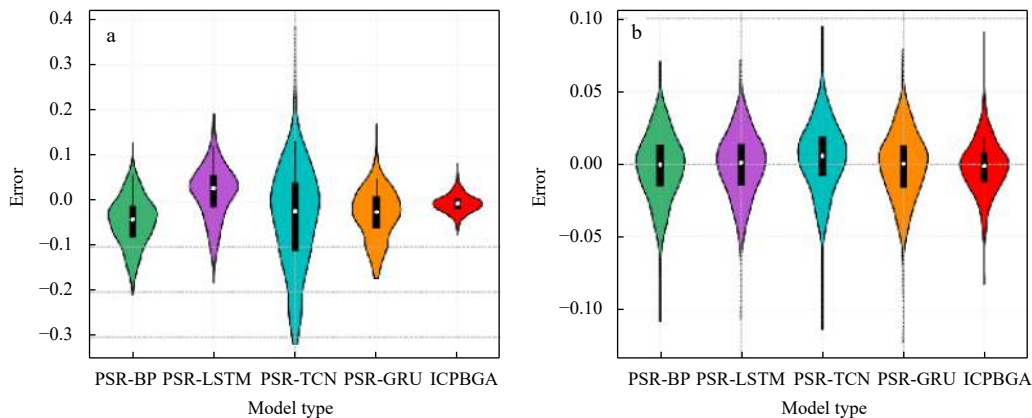


Fig. 13. The prediction error violin plot. a. The result of Monitoring Site 1; b. the result of Monitoring Site 2.

model has the most minor error dispersion at both Monitoring Sites 1 and 2, as well as the lowest upper edge and the smallest quartile.

6 Conclusions

Accurate prediction of seawater pH is essential to prevent seawater acidification and protect the marine environment and marine life. An innovative mixed prediction model of seawater pH data is proposed in this paper. The model consists of ICEEM-DAN, PSR, and the Bi-GRU neural network with multi-headed self-attention. Due to the influence of the actual measurement environment, the pH data of seawater are nonlinear, chaotic, and mixed with noise. Therefore, this paper uses the ICEEMDAN to process the original data, extract the main features of the original signal, and reduce the cumulative error and noise interference generated during prediction, first. Then, PSR is used to more effectively analyze the chaotic properties. Finally, the data reconstructed in the phase space is substituted into the Bi-GRU neural network with multi-headed self-attention for prediction to obtain the prediction result. To verify the applicability and effectiveness of the model, the seawater pH data of two monitoring sites are substituted into the model of this paper and compared with other excellent models in chaotic time series predicting. The validation results show that the model has satisfactory prediction accuracy. At the same time, the result of ablation studies shows that the prediction model of this paper is better than any of its ablated variants.

Although the prediction model proposed in this paper has achieved good results, it still has some limitations. First of all, the data used in this study is small. Although ICEEMDAN is used to reduce the interference of noise to the prediction effect, the prediction effect in the event of major pollution still lacks verification. Longer historical data can be used for verification in future studies. Secondly, this study mainly draws on some references and determines these hyperparameters by manual methods based on experience. In the follow-up research, some intelligent optimization algorithms can be considered to automatically search the optimal hyperparameters. The research object of this paper can not only the pH data of seawater, but also the model can be further extended to predict other marine environmental indicators, such as seawater temperature, salinity, etc., to understand the trend of ocean changes from a more comprehensive perspective.

References

- Alizadeh M J, Kavianpour M R. 2015. Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Marine Pollution Bulletin*, 98(1–2): 171–178
- Bandt C, Pompe B. 2002. Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88(17): 174102, doi: [10.1103/PhysRevLett.88.174102](https://doi.org/10.1103/PhysRevLett.88.174102)
- Cao Liangyue. 1997. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Non-linear Phenomena*, 110(1–2): 43–50
- Cerrada M, Sánchez R V, Li Chuan, et al. 2018. A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, 99: 169–196, doi: [10.1016/j.ymssp.2017.06.012](https://doi.org/10.1016/j.ymssp.2017.06.012)
- Chaudhuri T, Wu Min, Zhang Yu, et al. 2021. An attention-based deep sequential GRU model for sensor drift compensation. *IEEE Sensors Journal*, 21(6): 7908–7917, doi: [10.1109/JSEN.2020.3044388](https://doi.org/10.1109/JSEN.2020.3044388)
- Chen Guangyong, Chen Pengfei, Shi Yujun, et al. 2019. Rethinking the usage of batch normalization and dropout in the training of deep neural networks. arXiv preprint arXiv: 1905.05928. <https://arxiv.org/abs/1905.05928>[2019-05-14/2023-03-27]
- Chen Ying, Xu Chongxuan, Zhao Xueliang. 2023. Research on soft compensation of the potential drift signal of a pH electrode based on a gated recurrent neural network. *Measurement Science and Technology*, 34(2): 025107, doi: [10.1088/1361-6501/ac9ad2](https://doi.org/10.1088/1361-6501/ac9ad2)
- Cho K, Van Merriënboer B, Gulcehre C, et al. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv: 1406.1078. <https://arxiv.org/abs/1406.1078>[2014-06-03/2022-05-18].
- Dai Sicheng, Liu Yiru, Meng Jun. 2021. Sunspot forecast using Temporal Convolutional Neural (TCN) network based on phase space reconstruction. In: 2021 33rd Chinese Control and Decision Conference (CCDC). Kunming, China: IEEE, 2895–2900
- Du Liuqing, Li Baochuan, Guo Jiu hao, et al. 2021. Prediction of machine tool's accuracy degradation based on chaotic phase space reconstruction and depth GRU. In: 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI). Kunming, China: IEEE, 157–161
- Duan W Y, Han Y, Huang L M, et al. 2016. A hybrid EMD-SVR model for the short-term prediction of significant wave height. *Ocean Engineering*, 124: 54–73, doi: [10.1016/j.oceaneng.2016.05.049](https://doi.org/10.1016/j.oceaneng.2016.05.049)
- Fraser A M, Swinney H L. 1986. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33(2): 1134–1140, doi: [10.1103/PhysRevA.33.1134](https://doi.org/10.1103/PhysRevA.33.1134)
- Gao Song, Huang Juan, Li Yaru, et al. 2021. A forecasting model for wave heights based on a long short-term memory neural network. *Acta Oceanologica Sinica*, 40(1): 62–69, doi: [10.1007/s13131-020-1680-3](https://doi.org/10.1007/s13131-020-1680-3)
- Halevy I, Bachan A. 2017. The geologic history of seawater pH. *Science*, 355(6329): 1069–1071, doi: [10.1126/science.aal4151](https://doi.org/10.1126/science.aal4151)
- Hu Likun, Su Hao, Cui Ruyao, et al. 2022. KPI anomaly detection based on LSTM with phase space. In: 2022 IEEE 5th International Conference on Big Data and Artificial Intelligence (BDAl). Fuzhou, China: IEEE, 130–135
- Jiang Yuchen, Yin Shen, Dong Jingwei, et al. 2021. A review on soft sensors for monitoring, control, and optimization of industrial processes. *IEEE Sensors Journal*, 21(11): 12868–12881, doi: [10.1109/JSEN.2020.3033153](https://doi.org/10.1109/JSEN.2020.3033153)
- Kajiyama T, D'Alimonte D, Cunha J C. 2011. Performance prediction of ocean color Monte Carlo simulations using multi-layer perceptron neural networks. *Procedia Computer Science*, 4: 2186–2195, doi: [10.1016/j.procs.2011.04.239](https://doi.org/10.1016/j.procs.2011.04.239)
- Keshan N, Parimi P V, Bichindaritz I. 2015. Machine learning for stress detection from ECG signals in automobile drivers. In: 2015 IEEE International Conference on Big Data (Big Data). Santa Clara, CA, USA: IEEE, 2661–2669
- Krzysztofowicz R. 2001. The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, 249(1–4): 2–9
- Lee T L. 2004. Back-propagation neural network for long-term tidal predictions. *Ocean Engineering*, 31(2): 225–238, doi: [10.1016/S0029-8018\(03\)00115-X](https://doi.org/10.1016/S0029-8018(03)00115-X)
- Li Xinfang, Cao Jinfeng, Guo Jihong, et al. 2022. Multi-step forecasting of ocean wave height using gate recurrent unit networks with multivariate time series. *Ocean Engineering*, 248: 110689, doi: [10.1016/j.oceaneng.2022.110689](https://doi.org/10.1016/j.oceaneng.2022.110689)
- Liu Fagui, Zheng Jingzhong, Zheng Lailei, et al. 2020. Combining attention-based bidirectional gated recurrent neural network and two-dimensional convolutional neural network for document-level sentiment classification. *Neurocomputing*, 371: 39–50, doi: [10.1016/j.neucom.2019.09.012](https://doi.org/10.1016/j.neucom.2019.09.012)
- Malik A, Kumar A, Singh R P. 2019. Application of heuristic approaches for prediction of hydrological drought using multi-scalar streamflow drought index. *Water Resources Management*, 33(11): 3985–4006, doi: [10.1007/s11269-019-02350-4](https://doi.org/10.1007/s11269-019-02350-4)
- Moscoso-Lopez J A, Ruiz-Aguilar J J, Gonzalez-Enrique J, et al. 2019. Ro-Ro freight prediction using a hybrid approach based on empirical mode decomposition, permutation entropy and artificial neural networks. In: 14th International Conference on Hybrid Artificial Intelligence Systems. León, Spain: Springer, 563–574

- Niu Mingfei, Gan Kai, Sun Shaolong, et al. 2017. Application of decomposition-ensemble learning paradigm with phase space reconstruction for day-ahead $PM_{2.5}$ concentration forecasting. *Journal of Environmental Management*, 196: 110–118
- Packard N H, Crutchfield J P, Farmer J D, et al. 1980. Geometry from a time series. *Physical Review Letters*, 45(9): 712–716, doi: [10.1103/PhysRevLett.45.712](https://doi.org/10.1103/PhysRevLett.45.712)
- Patil K, Deo M C. 2017. Prediction of daily sea surface temperature using efficient neural networks. *Ocean Dynamics*, 67(3–4): 357–368
- Peng Yanni, Xiang Wanli. 2020. Short-term traffic volume prediction using GA-BP based on wavelet denoising and phase space reconstruction. *Physica A: Statistical Mechanics and its Applications*, 549: 123913, doi: [10.1016/j.physa.2019.123913](https://doi.org/10.1016/j.physa.2019.123913)
- Sadeghifar T, Lama G F C, Sihag P, et al. 2022. Wave height predictions in complex sea flows through soft-computing models: Case study of Persian Gulf. *Ocean Engineering*, 245: 110467, doi: [10.1016/j.oceaneng.2021.110467](https://doi.org/10.1016/j.oceaneng.2021.110467)
- Shan Kun, Ouyang Tian, Wang Xiaoxiao, et al. 2022. Temporal prediction of algal parameters in Three Gorges Reservoir based on highly time-resolved monitoring and long short-term memory network. *Journal of Hydrology*, 605: 127304, doi: [10.1016/j.jhydrol.2021.127304](https://doi.org/10.1016/j.jhydrol.2021.127304)
- Singla P, Duhan M, Saroha S. 2022a. An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. *Earth Science Informatics*, 15(1): 291–306, doi: [10.1007/s12145-021-00723-1](https://doi.org/10.1007/s12145-021-00723-1)
- Singla P, Duhan M, Saroha S. 2022b. A dual decomposition with error correction strategy based improved hybrid deep learning model to forecast solar irradiance. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 44(1): 1583–1607
- Sun Wei, Wang Yuwei. 2018. Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. *Energy Conversion and Management*, 157: 1–12, doi: [10.1016/j.enconman.2017.11.067](https://doi.org/10.1016/j.enconman.2017.11.067)
- Takens F. 1981. Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence*, Warwick 1980. Berlin, Heidelberg: Springer, 366–381
- Tilbrook B, Jewett E B, DeGrandpre M D, et al. 2019. An enhanced ocean acidification observing network: from people to technology to data synthesis and information exchange. *Frontiers in Marine Science*, 6: 337, doi: [10.3389/fmars.2019.00337](https://doi.org/10.3389/fmars.2019.00337)
- Vaswani A, Shazeer N, Parmar N, et al. 2017. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, California, USA: Curran Associates Inc., 6000–6010
- Wang Jianjin, Shi Peng, Jiang Peng, et al. 2017. Application of BP neural network algorithm in traditional hydrological model for flood forecasting. *Water*, 9(1): 48, doi: [10.3390/w9010048](https://doi.org/10.3390/w9010048)
- Wu Zhiyuan, Jiang Changbo, Conde M, et al. 2019. Hybrid improved empirical mode decomposition and BP neural network model for the prediction of sea surface temperature. *Ocean Science*, 15(2): 349–360, doi: [10.5194/os-15-349-2019](https://doi.org/10.5194/os-15-349-2019)
- Xie Jiang, Ouyang Jiaming, Zhang Jiyuan, et al. 2022. An evolving sea surface temperature predicting method based on multidimensional spatiotemporal influences. *IEEE Geoscience and Remote Sensing Letters*, 19: 1502005
- Xu Jianlong, Wang Kun, Lin Che, et al. 2021. FM-GRU: A time series prediction method for water quality based on seq2seq framework. *Water*, 13(8): 1031, doi: [10.3390/w13081031](https://doi.org/10.3390/w13081031)
- Yang Jing, Reichert P, Abbaspour K C, et al. 2007. Hydrological modelling of the Chaohe Basin in China: Statistical model formulation and Bayesian inference. *Journal of Hydrology*, 340(3–4): 167–182
- Yaseen Z M. 2021. An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: Review, challenges and solutions. *Chemosphere*, 277: 130126, doi: [10.1016/j.chemosphere.2021.130126](https://doi.org/10.1016/j.chemosphere.2021.130126)
- Zhang Zichen, Ding Shifei, Sun Yuting. 2020. A support vector regression model hybridized with chaotic krill herd algorithm and empirical mode decomposition for regression task. *Neurocomputing*, 410: 185–201, doi: [10.1016/j.neucom.2020.05.075](https://doi.org/10.1016/j.neucom.2020.05.075)
- Zhao Xinguo, Han Yu, Chen Bijuan, et al. 2020. CO_2 -driven ocean acidification weakens mussel shell defense capacity and induces global molecular compensatory responses. *Chemosphere*, 243: 125415, doi: [10.1016/j.chemosphere.2019.125415](https://doi.org/10.1016/j.chemosphere.2019.125415)