Response to comments of Reviewers

Interactive comment on "Hybrid improved EMD-BPNN model for the prediction of sea surface temperature" by Zhiyuan Wu et al.

Huang (Referee)
huanglimin@hrbeu.edu.cn
Received and published: 28 January 2019

Dr. Huang knows the topic very well and his/her comments are indeed helpful in improving the quality of this MS. We are grateful to Dr. Huang for a careful checking and comments on the MS. All comments are addressed point by point, each starting with an original comment and followed by a response in italic, as follows.

This paper proposed a hybrid EMD-BPNN model for SST prediction. The research work is very interesting and important. However, in my opinion, the paper needs minor revision before acceptance. You can find my questions and suggestions bellow.

Response: *Thank you for these comments. The positive comments in our solid professional skills are good encouragement to us.*

1. Why the simple EMD algorithm is not compared to the EEMD and CEEMD? I suggest the authors to provide comparison results of the EMD-BPNN.

Response: *Thank you for the professional comment. Empirical Mode Decomposition (EMD) is a state-of-the-art signal processing method proposed by Huang et al. This method can decompose the signal data of different frequencies step by step according to the characteristics of the data and obtain several periodic and trending signals orthogonal to each other, the method can decompose the stronger nonlinear and non-stationary signals into weaker nonlinear and non-stationary signals. The variation of SST is a deterministic non-linear dynamic system and a non-stationary time series data with intermittent signals. Once the intermittent signal is present in the actual signal, the frequency aliasing phenomenon occurs in the decomposition method of EMD, also called Mode Mixing Problem. The specific manifestation of this problem is that there are multiple scale components in one IMF component, or one scale component exists in multiple IMF components. Therefore, we carry out this research based on EEMD and CEEMD methods.*

2. Mode mixing is the motivation that the authors applied the EEMD technique in the hybrid SST prediction model. Therefore, it is very important to demonstrate the mode mixing problem in decomposing the studied SST time series. But this is not given in this paper. I suggest the authors to provide discussions on the mode mixing problem in the present study.

Response: *Thank you for your suggestion, and it is indeed a very important issue. We added the following statement to the revised manuscript.*

The variation of SST is a deterministic non-linear dynamic system and a non-stationary time series data with intermittent signals. Empirical Mode Decomposition (EMD) method can *decompose the signal data of different frequencies step by step according to the characteristics of the data and obtain several periodic and trending signals orthogonal to each other, the method can decompose the stronger nonlinear and non-stationary signals into weaker nonlinear and non-stationary signals.*

However, we know that once an intermittent signal appears in the actual signal, the EMD decomposition method will produce a Mode Mixing Problem. The Mode Mixing Problem causes the essential modal function to lose its physical meaning. In addition, the Mode Mixing Problem will also make the algorithm of Empirical Mode Decomposition unstable, and any disturbance may generate a new intrinsic mode function. In order to solve this problem, scholars have proposed the use of noise-assisted processing methods, Ensemble empirical mode decomposition (EEMD) and Complementary Ensemble Empirical Mode Decomposition (CEEMD). The white noise has been added to the original signal to change the extreme point distribution of the signal in the EEMD method, while in the CEEMD method, a set of noise signals have been added to the original signal to change the extreme point distribution of the signal.

3. Line 55, "Consequently, parameters such as mean and variance also do not change over time." In this sentence, I think it will be better to revise "parameters" as "statistical parameters".

Response: *Thank you for your comment. It has been modified in the revised manuscript.*

4. Lines 59-62, "This method can decompose the signal data of different frequencies step by step according to the characteristics of the data and obtain several periodic and trending signals orthogonal to each other, which can decompose the stronger nonlinear and non-stationary signals into weaker nonlinear and non-stationary signals". This sentence needs to make some corrections. As we know, the IMFs are orthogonal components, but the trending component is not orthogonal to any IMF component. Therefore, the above descriptions are not accurate. Besides, the sentence of "which can decompose the stronger nonlinear and non-stationary signals into weaker nonlinear and non-stationary signals" is ambiguous and makes no sense. Accurately, the EMD technique decomposes a non-stationary time series into several stationary subcomponent and a trend. But it is not easy to say the nonlinearity becomes weaker. So, I suggest the authors to make the sentence more accurate.

Response: *Thank you for the valuable criticism. We modified these sentences in the revised manuscript. "This method can decompose the signal data of different frequencies step by step according to the characteristics of the data and obtain several periodic and trending signals orthogonal to each other, which can decompose the stronger nonlinear and non-stationary signals. The EMD method is powerful and adaptive in analyzing nonlinear and non-stationary data sets. It provides an effective approach for decomposing a signal into a collection of so-called intrinsic mode functions (IMFs), which can be treated as empirical basis functions (Duan et al., 2016)."*

5. Lines 117-119, "The purpose of this study is to combine the EEMD algorithm and the CEEMD decomposition algorithm respectively with the BP neural network algorithm to establish a new prediction model, an improved hybrid EMD-BPNN model." Accurately, the models of EEMD-BPNN and CEEMD-BPNN themselves are not new. Various works about

this models in different problems have already carried out in the last ten years. Therefore, I suggest the authors not to over emphasis “new” or “improved” here. Just simply descript them as “hybrid models”.

Response: *Thank you for the suggestion. We modified these statements in the revised manuscript. “The purpose of this study is to combine the EEMD algorithm and the CEEMD decomposition algorithm respectively with the BP neural network algorithm to establish a prediction model, a hybrid EMD-BPNN model.”*

6. Lines 294-295, “This paper presents a novel SST predicting method based on the hybrid improved EMD algorithms and BP neural network method to process the SST data with strong nonlinearity and non-stationarity.” I suggest the authors to delete the word of “novel” here (and the same in the highlight part). Becomes the hybrid models have already explored extensively in various prediction problems. Besides, the authors argue that “the SST data with strong nonlinearity and non-stationarity.”, what is the standard of weak or strong nonlinearity and non-stationarity? Therefore, this sentence need to be corrected.

Response: *Thank you for the suggestion. We modified these statements in the revised manuscript. “This paper presents an SST predicting method based on the hybrid EMD algorithms and BP neural network method to process the SST data with nonlinearity and non-stationarity.”*

References:

Wu Z, Schneider E K, Kirtman B P, et al. The modulated annual cycle: an alternative reference frame for climate anomalies[J]. Climate Dynamics, 2008, 31(7-8): 823-841.

Wu Z, Huang N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method[J]. Advances in adaptive data analysis, 2009, 1(01): 1-41.

Duan W, Huang L, Han Y, et al. A hybrid EMD-AR model for nonlinear and non-stationary wave forecasting[J]. Journal of Zhejiang University-SCIENCE A, 2016, 17(2): 115-129.

Hybrid improved EMD-BPNN model for the prediction of sea surface temperature

Zhiyuan Wu [a,b,c], Changbo Jiang [a,c,*], Mack Conde [d], Bin Deng [a,c], Jie Chen [a,c]

a. School of Hydraulic Engineering, Changsha University of Science & Technology, Changsha, 410004, China;

b. School for Marine Science and Technology, University of Massachusetts Dartmouth, New Bedford, MA 02744, USA;

c. Key Laboratory of Water-Sediment Sciences and Water Disaster Prevention of Hunan Province, Changsha, 410004, China;

d. Department of Mathematics, University of Massachusetts Dartmouth, North Dartmouth, MA 02747, USA.

Highlights

- A novel SST predicting method based on the hybrid ~~improved~~ EMD algorithms and BP neural network method are proposed in this paper.
- SST prediction results based on the hybrid EEMD-BPNN and CEEMD-BPNN models are compared and discussed.
- Cases study of SST in the North Pacific shows that the proposed hybrid CEEMD-BPNN model can effectively predict the time-series SST.

Abstract: Sea surface temperature (SST) is the major factor that affects the ocean-atmosphere interaction, and in turn the accurate prediction of SST is the key to ocean dynamic prediction. In this paper, an SST predicting method based on ~~improved~~ empirical mode decomposition (EMD) algorithms and back-propagation neural network (BPNN) is proposed. Two different EMD algorithms have been applied extensively for analyzing time-series SST data and some nonlinear stochastic signals. Ensemble empirical mode decomposition (EEMD) algorithm and Complementary Ensemble Empirical Mode Decomposition (CEEMD) algorithm are two improved algorithms of EMD, which can effectively handle the mode-mixing problem and decompose the original data into more stationary signals with different frequencies. Each Intrinsic Mode Function (IMF) has been taken as an input data to the back-propagation neural network model. The final predicted SST data is obtained by aggregating the predicted data of individual IMF. A case study, of the monthly mean sea surface temperature anomaly (SSTA) in the northeastern region of the North Pacific, shows that the proposed hybrid CEEMD-BPNN model is much more accurate than the hybrid EEMD-BPNN model, and the prediction accuracy based on BP neural network is improved by the CEEMD method. Statistical analysis of the case study demonstrates that applying the proposed hybrid CEEMD-BPNN model

is effective for the SST prediction.

Keywords.

Sea Surface Temperature; Back-Propagation Neural Network; Empirical Mode Decomposition; Prediction; Machine Learning Algorithms.

1 Introduction

The Sea Surface Temperature (SST) is a main factor in the interaction between the ocean and the atmosphere (Wiedermann et al., 2017; He et al., 2017; Wu et al., 2019a), and it characterizes the combined results of ocean heat (Buckley et al., 2014; Griffies et al., 2015; Wu et al., 2019b), dynamic processes (Takakura et al., 2018). It is a very important parameter for climate change and ocean dynamics process, reflects sea-air heat and water vapor exchange. Small changes in sea temperature can have a huge impact on the global climate. The well-known El Niño and La Niña phenomena are caused by abnormal changes in SST (Chen et al., 2016a; Zheng et al., 2016).

Therefore, scholars have begun to observe the SST in recent years, the observation of the SST is important (Kumar et al., 2017; Sukresno et al., 2018). Accurate observation and effective prediction of the SST are very important (Hudson et al., 2010). Predicting the SST in advance can enable people to take appropriate measures to reduce the impact on daily life and reduce unnecessary losses. However, due to the high randomness of the monthly mean sea surface temperature anomaly (SSTA), the nonlinear and non-stationary characteristics are obvious. At present, there is no clear and feasible method with high accuracy to effectively predict the SST (Zhu et al., 2015; Chen et al., 2016b; Khan et al., 2017).

In mathematics and science, a nonlinear system is a system in which the change of the output is not proportional to the change of the input. Nonlinear dynamical systems, describing changes in variables over time, may appear chaotic, unpredictable, or counterintuitive, contrasting with much simpler linear systems. A stationary process is a stochastic process whose unconditional joint probability distribution does not change when shifted in time. Consequently, statistical parameters such as mean and variance also do not change over time. The variation of SST is a deterministic non-linear dynamic system and a non-stationary time series data. ~~The observation sequence at a certain point contains not only the information of this point, but also the information of other relevant points.~~ Empirical Mode Decomposition (EMD) is a state-of-the-art signal processing method proposed by Huang et al. (1998). This method can decompose the signal data of different frequencies step by step according to the characteristics of the data and obtain several periodic and trending signals orthogonal to each other, ~~which~~ the method can decompose the stronger nonlinear and non-stationary signals ~~into weaker nonlinear and non-stationary signals~~ (Wang et al., 2015; Amezquita-Sanchez and Adeli,2015; Wang et al., 2016; Kim and Cho, 2016). The empirical mode decomposition (EMD) method is powerful and adaptive in analyzing nonlinear and non-stationary data sets. It provides an effective approach for decomposing a signal into a collection of so-called intrinsic mode functions (IMFs), which can be treated as empirical basis functions (Duan et al., 2016). However, there were some problems of the EMD method, such as mode mixing (Huang and Wu, 2008; Wu et al., 2008; Wu and Huang, 2009).

Once an intermittent signal appears in the actual signal, the EMD decomposition method will produce a Mode Mixing Problem. The Mode Mixing Problem causes the essential modal function to lose its physical meaning. In addition, the Mode Mixing Problem will also make the algorithm of Empirical Mode Decomposition unstable, and any disturbance may generate a new intrinsic mode function. In order to solve this problem, scholars have proposed the use of noise-assisted processing methods, Ensemble empirical mode decomposition (EEMD) and Complementary Ensemble Empirical Mode Decomposition (CEEMD). The white noise has been added to the original signal to change the extreme point distribution of the signal in the EEMD method, while in the CEEMD method, a set of noise signals have been added to the original signal to change the extreme point distribution of the signal.

To solve this problem, Wu and Huang (2009) proposed the Ensemble Empirical Mode Decomposition (EEMD) method by adding different white noise in each ensemble member to suppress mode mixing. Yeh et al. (2010) added two opposite-signal white noises to the time-series data sequence, and proposed an improved algorithm for EEMD, Complete Ensemble Empirical Mode Decomposition (CEEMD). The decomposition effect is equivalent to EEMD, and the reconstruction error caused by adding white noise is reduced (Tang et al., 2015). At present, the EMD model and its improved algorithms had been widely used in many fields on ocean science, such as storm surge and sea level rise (Wu et al., 2011; Lee, 2013; Ezer and Atkinson, 2014), tidal amplitude (Cheng et al., 2017; Pan et al., 2018) and wave height (Duan et al., 2016; Sadeghifar et al., 2017; López et al., 2017). These studies and applications reflected that the EMD model and its improved algorithms can effectively reduce the non-stationarity of the time-series data, which helps further analysis and processing.

For nonlinear prediction, the more commonly used methods are curve fitting (Motulsky and Ransnas, 1987), gray-box model (Pearson and Pottmann, 2000), homogenization function model (Monteiro et al., 2008), neural network (Deo et al., 2001; Wang et al, 2015; Kim et al., 2016) and so on. Among them, Back-Propagation Neural Network (BPNN) (Lee, 2004; Jain and Deo, 2006; Savitha and Al, 2017; Wang et al., 2018) has certain advantages in dealing with nonlinear problems, it is a basic machine learning algorithm and its principle is simple and operability is strong, so in ocean science and engineering it has been widely used.

In view of non-stationary and nonlinear monthly mean SST, the EEMD, CEEMD and BP neural network will be used here to study how to improve the accuracy of SST prediction. The ~~improved~~ hybrid EMD-BPNN models will be established for the prediction of SSTA in the northeastern region of the Pacific Ocean.

2 Data collection

The SST time-series data in this study is from NOAA Optimum Interpolation Sea Surface Temperature (OISST) official website (Reynolds et al., 2007; Banzon et al., 2016; https://www.ncdc.noaa.gov/oisst/data-access). The NOAA 1/4°daily OISST is an analysis constructed by combining observations from different platforms (satellites, ships, buoys) on a regular global grid. There are two kinds of OISST, named after the relevant satellite SST sensors. These are the Advanced Very High Resolution Radiometer (AVHRR) and Advanced Microwave Scanning Radiometer on the Earth Observing System (AMSR-E); the AVHRR dataset is used in this study. The average annual sea surface temperature in North Pacific (0°N-60°N, 100°E-100°W) from~~during~~ January 1982 to December 2016 is shown in Fig.1.

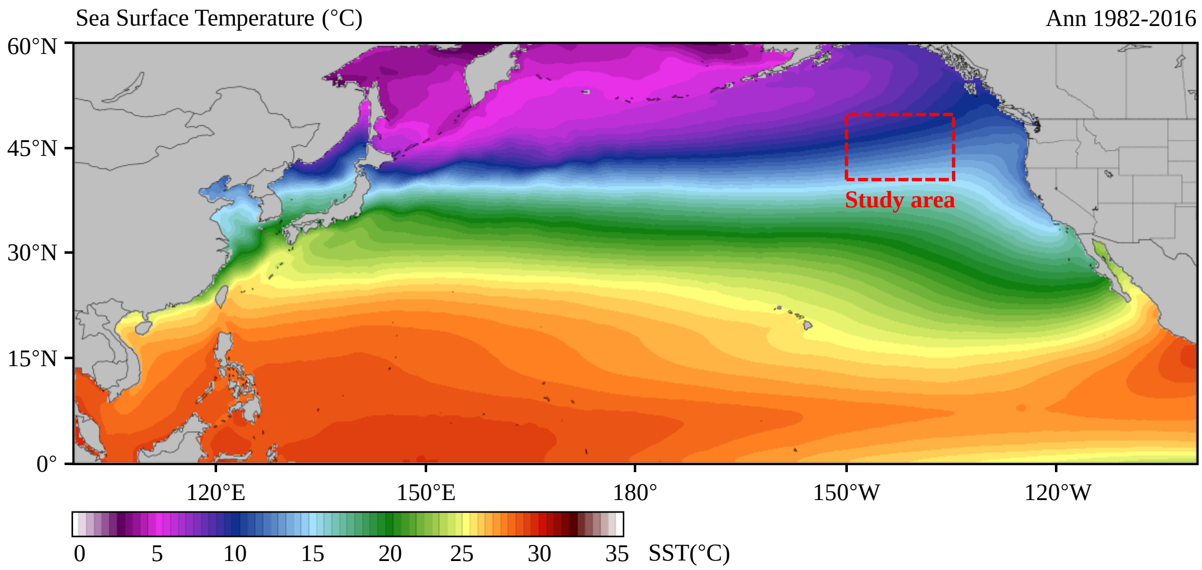

Fig.1 Average annual sea surface temperature in North Pacific during Jan 1982 to Dec 2016 (35-years).

It has been shown that the sea surface temperature anomaly in the northeastern Pacific in the ten years 2006-2016 was 2.0°C warmer than in the previous ten years 1996-2006. ~~It had been shown that the sea surface temperature anomaly in the northeastern Pacific is much hotter 2.0 °C than that in previous years from the observations in recent ten years (2006-2016).~~ Previous studies (Bond et al., 2015) showed that in the spring and summer of 2014, the high SST area of the northeastern Pacific had expanded to coastal ocean waters, which affected the weather in coastal areas and the lives of fishermen, and even affected the temperature in Washington, USA, causing interference to daily life.~~the daily life had been caused interference.~~

In this study, we select the northeastern region of the North Pacific Ocean (in Fig.1, 40°N-50°N, 150°W-135°W) to measure sea surface temperature. The time-series data of SST for the study area from January 1982 to December 2016 with a data length of 420 months was obtained from OISST-V2 (Fig. 2). The monthly mean sea surface temperature anomaly (SSTA) was used in the analysis and calculation. As shown in Fig. 2(a), it can be found the overall time-series data is very messy, nonlinear and random from the perspective of the image.

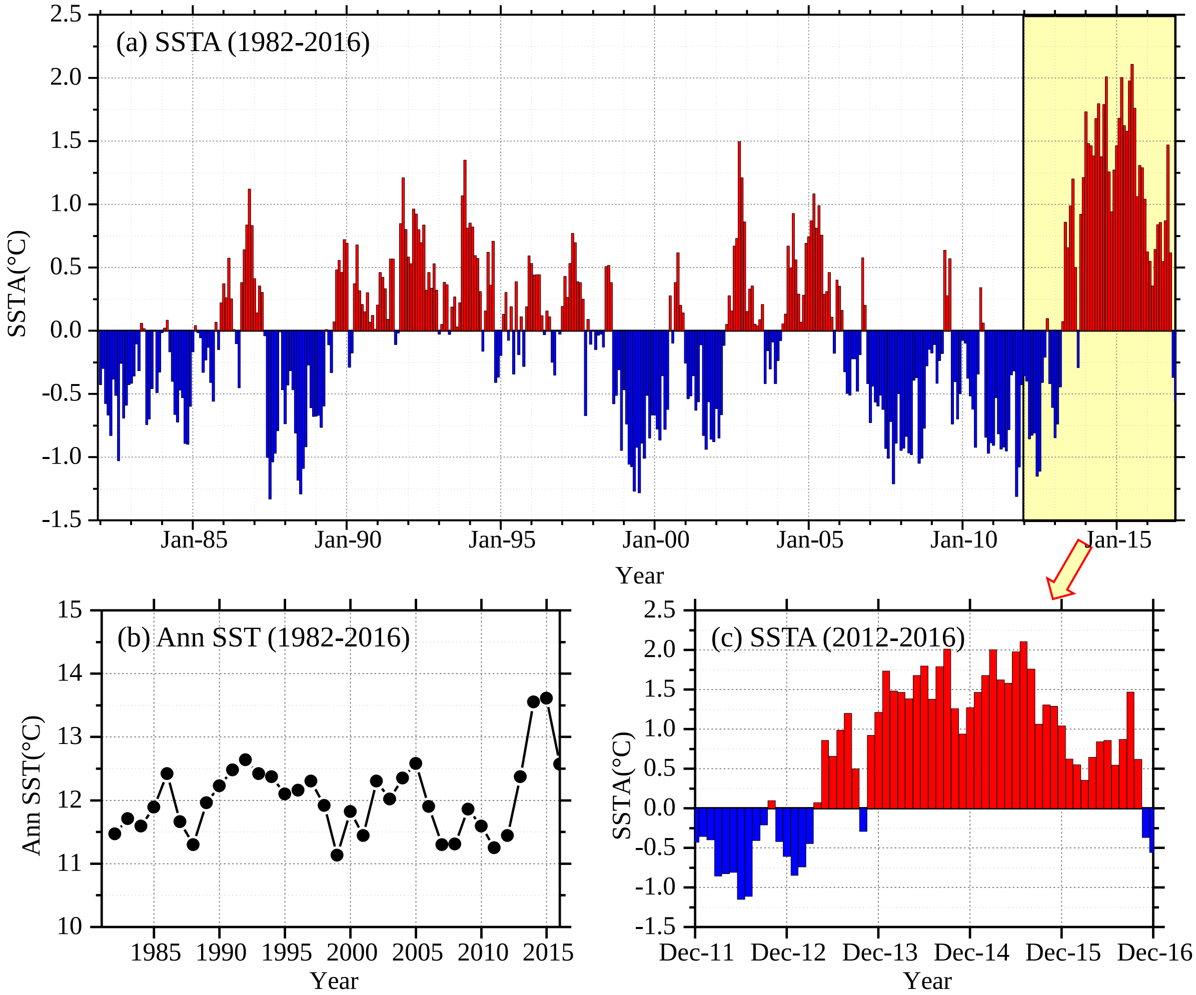

Fig.2 The time-series of sea surface temperature in the study area. (a) SST anomaly (1982-2016, 35 years); (b) Annual SST (1982-2016, 35 years); (c) SST anomaly (2012-2016, 5 years).

3 Decomposition of SSTA

The purpose of this study is to combine the EEMD algorithm and the CEEMD decomposition algorithm respectively with the BP neural network algorithm to establish a ~~new~~ prediction model, a~~n improved~~ hybrid EMD-BPNN model. The EEMD and CEEMD algorithms are performed on the monthly mean SSTA data to obtain a series of intrinsic mode functions (IMFi). Each IMFi is predicted by a BP neural network and then each IMFi is reconstructed to obtain the predicted value of SSTA.

3.1 Decomposition by the EEMD algorithm

The SSTA in Fig. 2(a) has been decomposed based on the ensemble empirical mode decomposition (EEMD algorithm), and seven IMF components and a residual component RES (Residue) are obtained as shown in Fig. 3.

~~It can be seen from Fig. 3 that the first three intrinsic mode function components IMF1, IMF2, and IMF3 still exhibit strong nonlinearity and non-stationarity. The IMF4 to IMF7 and the final trend term RES have some periodicity and relatively regular volatility, and the non-stationary and nonlinear properties are less than the first three components. The trend term RES reflects that the overall trend of SSTA has gradually increased since 1982. As the non-stationarity of each IMFi is gradually reduced, the EEMD algorithm will reduce the influence of non-stationarity on prediction. The absolute error (ERR) of the decomposition can been calculated by the following Formula (1).~~

$$\cancel{a(t)=\left|S(t)-\left[\sum_{i=1}^{7}I_i(t)+R(t)\right]\right| \qquad (1)}$$

~~where, $a(t)$ is the absolute error (ERR), $S(t)$ the original SSTA observation data, $I_i(t)$ the i-th component of the IMF (IMFi), and $R(t)$ the trend term (RES).~~

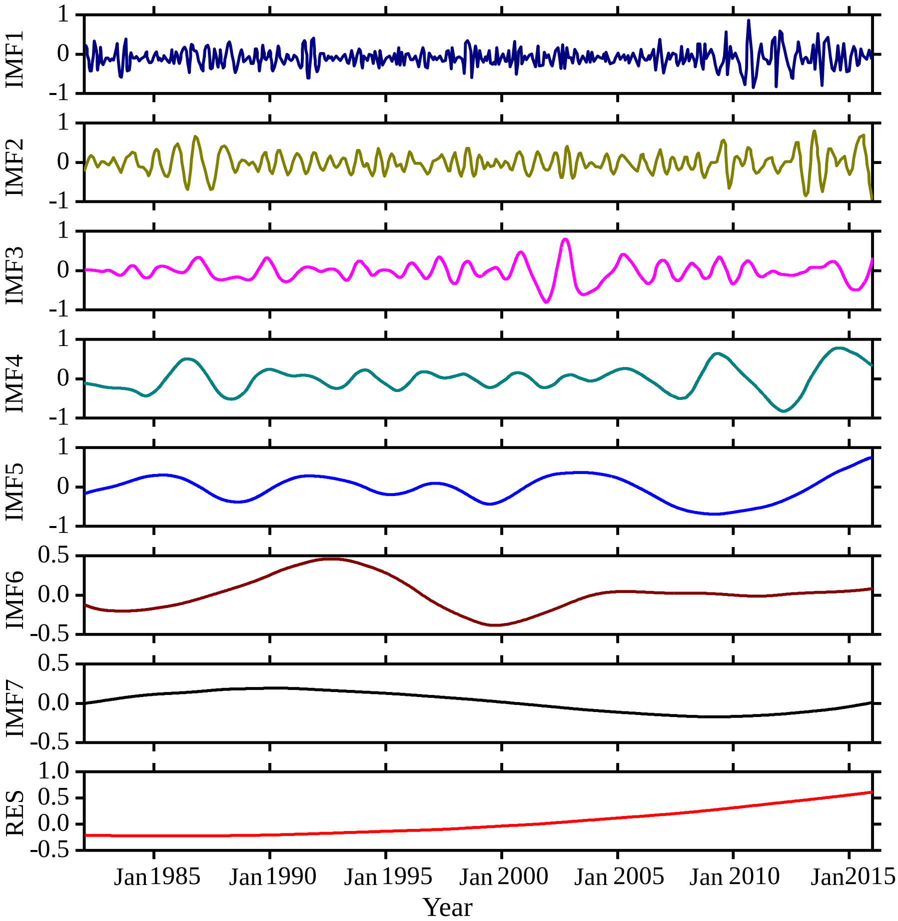

Fig.3 IMF components and the trend item RES of monthly mean SSTA over the study area based on the EEMD algorithm during 1982-2016.

It can be seen from Fig. 3 that the first three intrinsic mode function components IMF1, IMF2, and IMF3 still exhibit strong nonlinearity and non-stationarity. The IMF4 to IMF7 and the final trend term RES have some periodicity and relatively regular fluctuation, and the non-stationary and nonlinear properties are less than the first three components. The trend term RES reflects that the overall trend of SSTA has gradually increased since 1982. As the non-stationarity of each IMFi is gradually reduced, the EEMD algorithm will reduce the influence of non-stationarity on prediction. The absolute error (ERR) of the decomposition can been calculated by the following Formula (1).

$$a(t)=\left|S(t)-\left[\sum_{i=1}^{7} I_i(t)+R(t)\right]\right| \qquad (1)$$

where, $a(t)$ is the absolute error (ERR), $S(t)$ the original SSTA observation data, $I_i(t)$ the i-th component of the IMF (IMFi), and $R(t)$ the trend term (RES).

The absolute error (ERR) based on the EEMD algorithm ~~was~~ is shown in Fig. 4. It can be seen from the figure that the ERR of 420 months after decomposition is basically below 0.01 °C, and the ERR exceeds 0.01 °C in five months: June 1989, September 1993, July 1998, May 1999 and March 2010.

In addition to June 1989, the other four monthly data with a large ERR occurred during the El Niño period. The maximum error is in March 2010, the actual value is -0.1204 °C, the result based on EEMD algorithm is -0.1325 °C, the ERR of decomposition is 0.0121 °C; the minimum error, in April 1987, is 1.73×10^{-5} °C. The overall mean ERR based on the EEMD algorithm is 0.0035 °C and the order of magnitude is 10^{-3}.

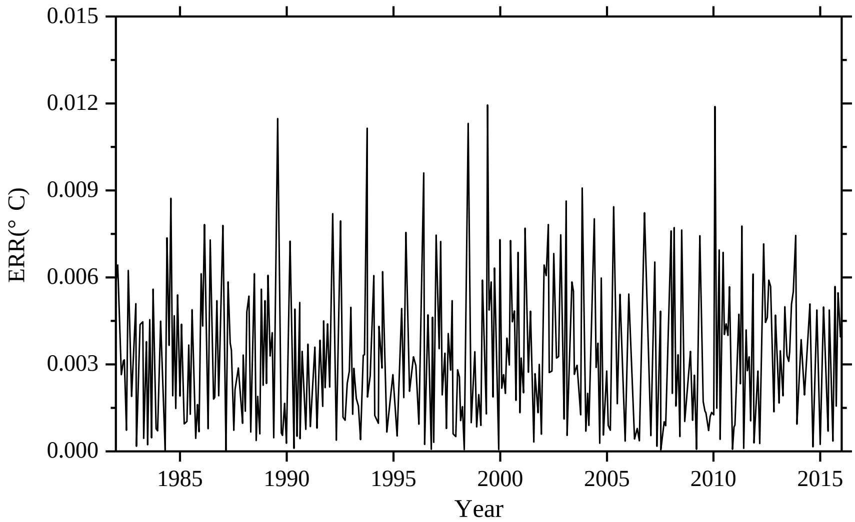

Fig. 4 The ERR of monthly mean SSTA over the study area based on the EEMD algorithm during 1982-2016.

~~In addition to June 1989, the other four monthly data with a large ERR occurred during the El Niño period. The maximum error is in March 2010, the actual value is -0.1204 °C, the result based on EEMD algorithm is -0.1325 °C, the ERR of decomposition is 0.0121 °C; the minimum error is in April 1987, which is 1.73×10^{-5} °C. The overall mean ERR based on EEMD algorithm is 0.0035 °C and the order of magnitude is 10^{-3}.~~

3.2 Decomposition by the CEEMD algorithm

The SSTA has been decomposed based on the complementary ensemble empirical mode decomposition (CEEMD algorithm) and seven IMF components and a residual component RES (Residue) are obtained as shown in Fig. 5. It can be seen when comparing the decomposition results based on EEMD and CEEMD algorithms that although the mode components decomposed by CEEMD algorithm are different from the corresponding results decomposed by EEMD, the nonlinearities and non-stationarities of the eight modes decomposed by the two decomposition algorithms are gradually decreasing, and the final trend term RES is an upward trend. Both decomposition algorithms confirm the characteristic of a gradual increase in~~for~~ the overall trend of the data series.

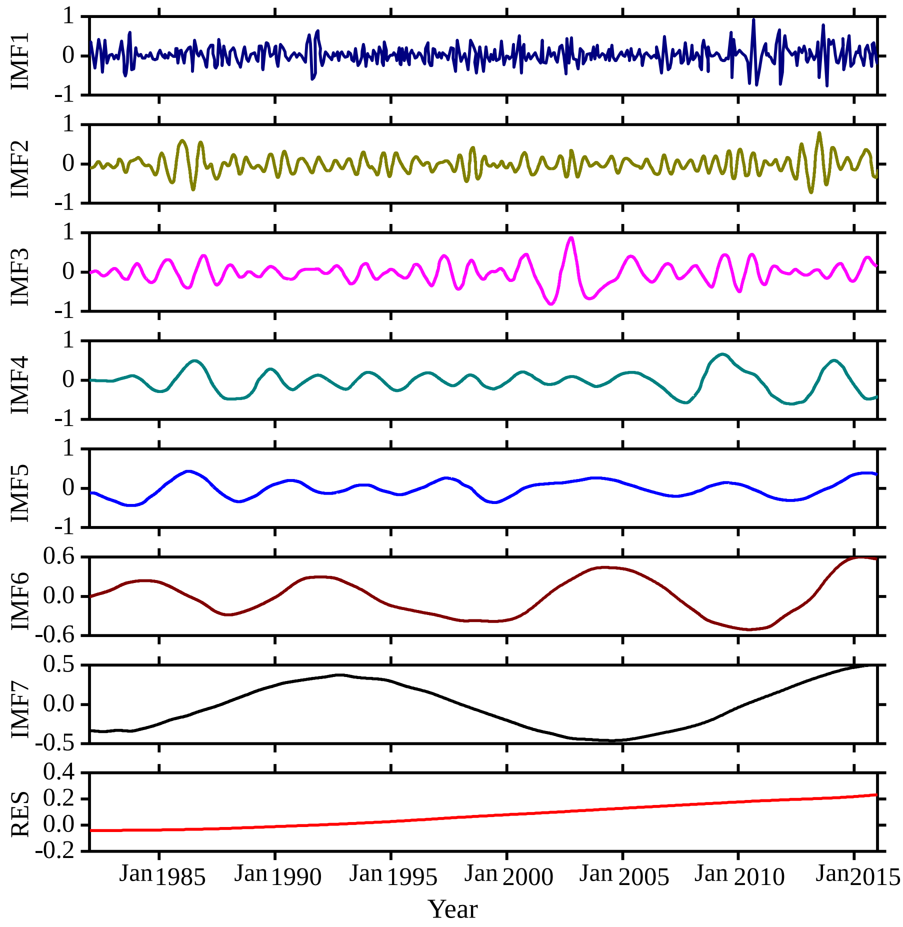

Fig.5 IMF components and the trend item RES of monthly mean SSTA over the study area based on the CEEMD algorithm during 1982-2016.

The absolute error (ERR) obtained based on the CEEMD algorithm is shown in Fig. 6. It can be seen from the figure that the ERR of 420 months data after decomposition is less than 5×10^{-16} °C, and the accuracy is very better. The maximum error is 4.48×10^{-16} °C in March 2016; the minimum error is zero. The overall mean ERR based on CEEMD algorithm is 6.10×10^{-17} °C and the order of magnitude is 10^{-17}. By comparing the results and errors of the above two decomposition algorithms, it can be seen that the error based on the improved algorithm (CEEMD) is much smaller than the error based on EEMD algorithm. Because more white noise with the opposite sign had been added in CEEMD algorithm, the reconstruction error caused by the white noise has been reduced over it in EEMD algorithm.

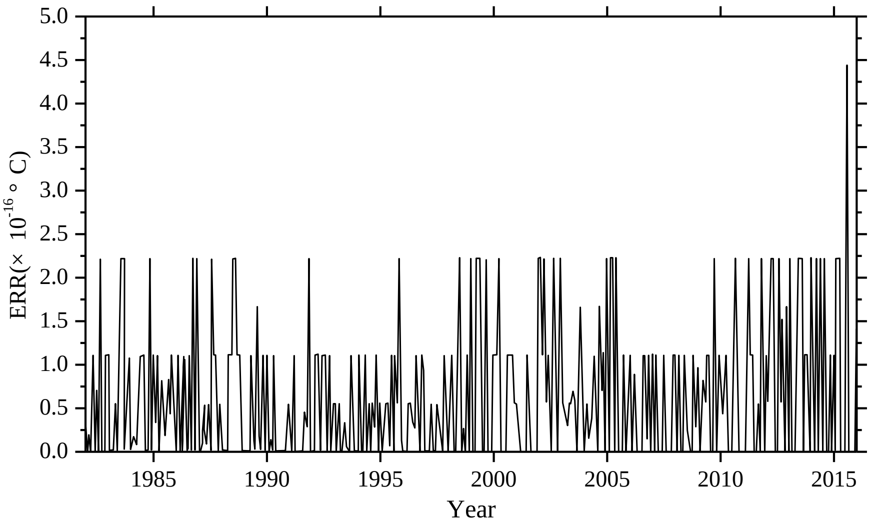

Fig. 6 The ERR of monthly mean SSTA over the study area based on the CEEMD algorithm during 1982-2016.

~~The absolute error (ERR) obtained based on the CEEMD algorithm is shown in Fig. 6. It can be seen from the figure that the ERR of 420 months data after decomposition is less than 5×10^{-16} °C, and the accuracy is very better. The maximum error is 4.48×10^{-16} °C in March 2016; the minimum error is zero. The overall mean ERR based on CEEMD algorithm is 6.10×10^{-17} °C and the order of magnitude is 10^{-17}. By comparing the results and errors of the above two decomposition algorithms, it can be seen that the error based on the improved algorithm (CEEMD) is much smaller than the error based on EEMD algorithm. This is because more white noise in CEEMD algorithm had been added than that in EEMD algorithm, so reducing the reconstruction error caused by white noise when the decomposition effect is equivalent to EEMD algorithm.~~

4 SSTA prediction model

4.1 The BP neural network

Artificial Neural Network (ANN) is an information processing approach based on the biological neural network (López et al., 2015; Kim et al., 2016). In theory, ANN can simulate any complex nonlinear relationship through nonlinear units (neurons) and has been widely used in the prediction area, such as wave height and storm surge. The most basic structure of ANN consists of input layers, hidden layers and output layers. One of the most widely used ANN models is the back propagation neural network (BPNN, Wang et al., 2018) algorithm based on the BP algorithm.

The BPNN algorithm is a multi-layer feedforward network trained according to the error back propagation algorithm and is one of the most widely used deep learning algorithms. The BP network can be used to learn and store a large number of mappings of input and output models without the need to publicly describe the mathematical equations of these mapping relationships. The learning rule is to use the steepest descent method. When applied to SST predicting, the input data are monthly mean SST in previous months and the output data are predicted SST time-series data. The desired data for comparison is the observed actual SST.

4.2 SSTA prediction model based on hybrid improved EMD-BPNN algorithm

The proposed monthly mean sea surface temperature anomaly (SSTA) predicting model includes three steps as follows. First, original SST datasets are decomposed into certain more stationary signals with different frequencies by EEMD. Second, the BP neural network is used to predict each IMF and the residue RES. A rolling forecasting process is studied. The prediction is made using the previous data for one step ahead. Finally, the prediction results of each IMF and the residue RES are aggregated to obtain the final SST prediction results. The flowchart of the SST prediction model based on hybrid improved empirical mode decomposition algorithm (improved EMD algorithm) and back-propagation neural network (BPNN)is shown in Fig. 7. The SST prediction model has been abbreviated as a hybrid improved EMD-BPNN model in the following article.

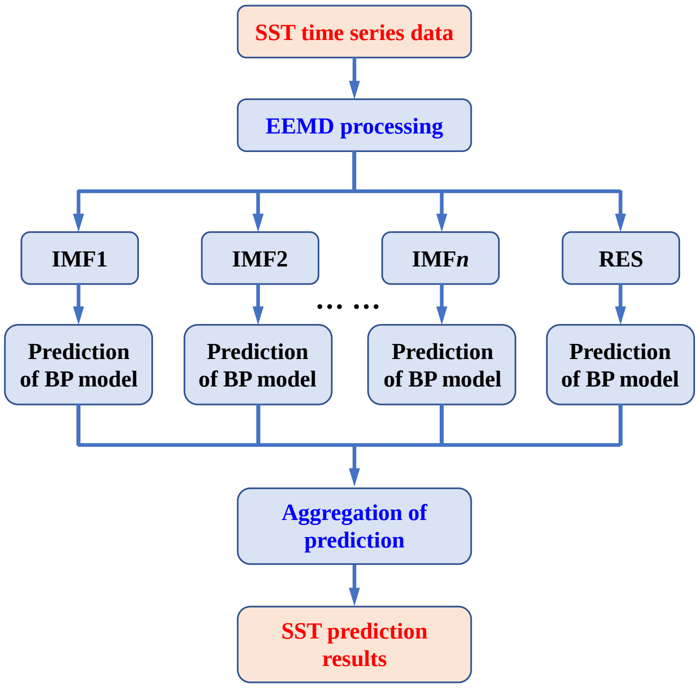

Fig.7 The flowchart of SST prediction model based on hybrid improved empirical mode decomposition algorithm (improved EMD algorithm) and back-propagation neural network (BPNN).

5 Case study: SSTA prediction based on the hybrid improved EMD-BPNN models

In order to study the effects of the two improved EMD algorithms (EEMD and CEEMD) on the prediction results, and to analyze the prediction ability of BP neural network, the following experiments were carried out. Predict SSTA results in 2017 and analyze the prediction abilities of different mode decomposition data based on EEMD and CEEMD algorithms. The experiment content is as follows: the BP neural network is trained with the decomposition data of each mode from 1982 to 2016, and the SSTA in 2017 is predicted by the trained neural network, and the observation results of 12 months in 2017 are~~is~~ used to compare and analyze with the prediction results.

Since the nonlinearity of the IMF1 to IMF3 is still relatively strong, a three-layer BP neural network structure has been chosen and independently analyze and predict each month. For the IMF4 and subsequent modes, since the nonlinearity and non-stationarity have been degraded relative to the first three modes, a BP neural network with 12 nodes at input layer and output layer has been used to train and predict SSTA.

The prediction results of each mode decomposition component based on the EEMD algorithm are shown in Fig. 8. The absolute errors of the predicted value and the actual value are shown in Table 1. ~~Root mean~~

~~square error (RMSE) is used as metrics to access the performance of the two different models.~~

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x_n - y_n\right)^2} \quad \cancel{(2)}$$

~~where, x_n and y_n are the observed and the predicted values respectively, N is the number of data used for the performance evaluation. Results are shown in Table 1.~~

~~It can be seen from Fig. 8 and Table 1 that the maximum absolute error (Max ERR) of the first decomposition component IMF1 based on the hybrid EEMD-BPNN model is 0.2197 °C in January. The minimum absolute error (Min ERR) is 0.0014 °C, which is in August. The prediction ability of the second mode decomposition component IMF2 is roughly equivalent to the IMF1, and the mean absolute error (Mean ERR) of the first three intrinsic mode function components IMF1, IMF2, and IMF3 are between 0.10 °C and 0.15 °C. The mean absolute errors of the IMF4 and IMF5 are 0.0663 °C and 0.0089 °C, respectively, and the prediction accuracy based on the hybrid EEMD-BPNN model is roughly equivalent to the decomposition accuracy of the EEMD algorithm. The prediction errors of the last two intrinsic mode function components and the residue RES are on the order of 10^{-4}. It can be seen that as the nonlinearity and non-stationarity of the series data decreases, the error of the prediction results becomes smaller and smaller.~~

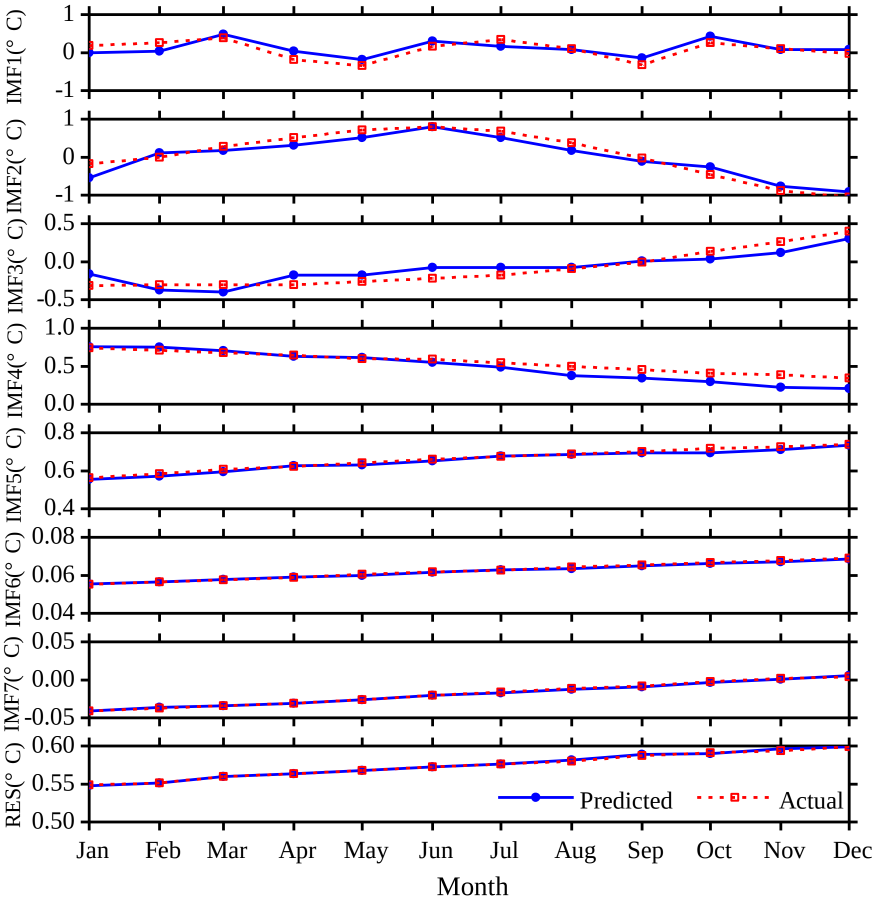

Fig. 8 SSTA prediction results based on the hybrid EEMD-BPNN model of each individual component in 2017.

Root mean square error (RMSE) is used as metrics to access the performance of the two different models.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x_n - y_n\right)^2} \tag{2}$$

where, x_n and y_n are the observed and the predicted values respectively, N is the number of data used for the performance evaluation. Results are shown in Table 1.

Table 1. The absolute errors ERRs of the SSTA prediction results of each individual component based on the hybrid EEMD-BPNN model (unit: °C).

	Max ERR	Min ERR	Mean ERR	RMSE
IMF1	0.2197	0.0014	0.1424	0.1486
IMF2	0.2166	0.0323	0.1297	0.1673
IMF3	0.1872	0.0051	0.1070	0.1245

IMF4	0.1602	1.~~6876~~6869 $\times 10^{-4}$	0.0663	0.0857
IMF5	0.0158	0.0010	0.0089	0.0104
IMF6	3.~~8778~~8766 $\times 10^{-4}$	1.9752 $\times 10^{-4}$	2.7221 $\times 10^{-4}$	0.0003
IMF7	5.2662 $\times 10^{-4}$	1.~~6396~~6387 $\times 10^{-4}$	1.~~7917~~7907 $\times 10^{-4}$	0.0002
RES	5.~~4864~~4859 $\times 10^{-4}$	2.~~2312~~2308 $\times 10^{-4}$	2.~~4774~~4766 $\times 10^{-4}$	0.0002

It can be seen from Fig. 8 and Table 1 that the maximum absolute error (Max ERR) of the first decomposition component IMF1 based on the hybrid EEMD-BPNN model is 0.2197 °C in January. The minimum absolute error (Min ERR) is 0.0014 °C, which is in August. The prediction ability of the second mode decomposition component IMF2 is roughly equivalent to the IMF1, and the mean absolute error (Mean ERR) of the first three intrinsic mode function components IMF1, IMF2, and IMF3 are between 0.10 °C and 0.15 °C. The mean absolute errors of the IMF4 and IMF5 are 0.0663 °C and 0.0089 °C, respectively, and the prediction accuracy based on the hybrid EEMD-BPNN model is roughly equivalent to the decomposition accuracy of the EEMD algorithm. The prediction errors of the last two intrinsic mode function components and the residue RES are on the order of 10^{-4}. It can be seen that as the nonlinearity and non-stationarity of the series data decreases, the error of the prediction results becomes smaller and smaller.

According to the same method, the eight mode components decomposed by CEEMD algorithm have been analyzed and predicted. The prediction results and error analysis have been shown in Fig. 9 and Table 2. It can be seen from Fig. 9 and Table 2 that the maximum error of the first decomposition component IMF1 based on the hybrid CEEMD-BPNN model is 0.1779 °C in May. The minimum error is 0.0068 °C, which is in June.

The prediction ability of the second mode decomposition component IMF2 is roughly equivalent to the IMF1. Except for the four months of May, September, October, and November, the accuracies of prediction results of other months are satisfactory. The prediction results of the first three intrinsic mode function components IMF1, IMF2, and IMF3 are basically the same as the actual data. In the prediction results of the fourth mode component IMF4, except for slight error in December, the prediction ability is better. The predicted results of the last three intrinsic mode function components IMF5, IMF6, IMF7 and the residue RES are basically consistent with the observation results.

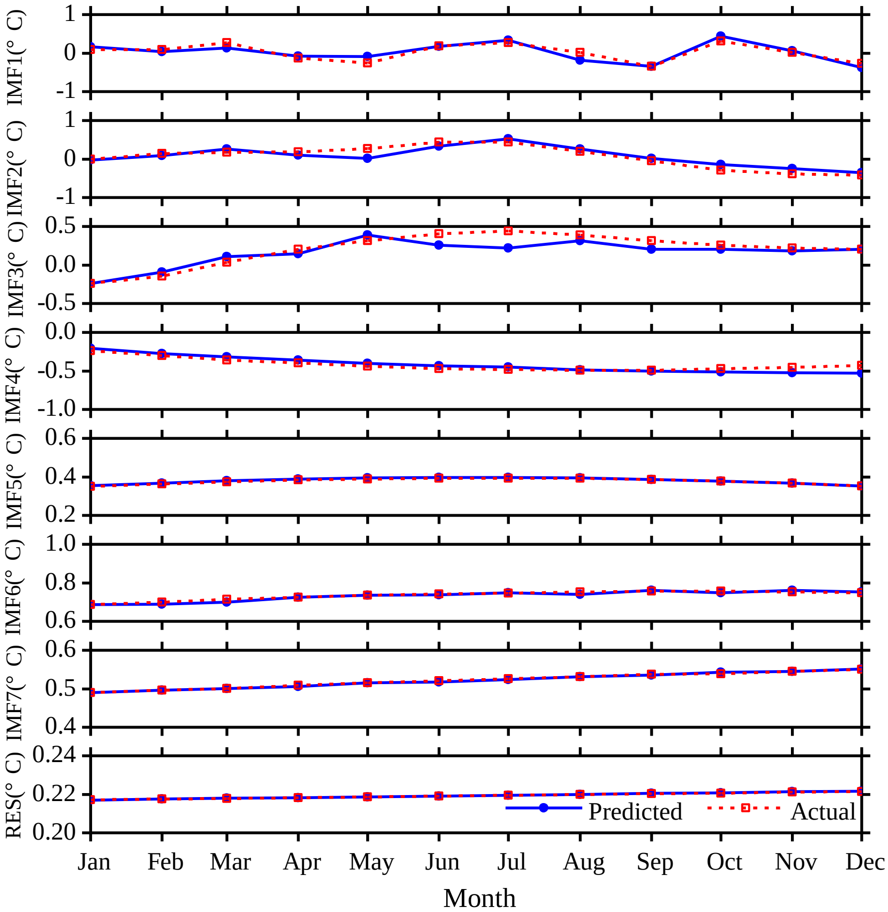

Fig. 9 SSTA prediction results based on the hybrid CEEMD-BPNN model of each individual component in 2017.

Table 2. The absolute errors ERRs of the SSTA prediction results of each individual component based on the hybrid CEEMD-BPNN model (unit: °C).

	Max ERR	Min ERR	Mean ERR	RMSE
IMF1	0.1779	0.0068	0.0827	0.0987
IMF2	0.1643	0.0413	0.0811	0.1124
IMF3	0.1521	0.0160	0.0713	0.1006
IMF4	0.0851	0.0211	0.0324	0.0427
IMF5	0.0052	8.7694×10^{-5}	0.0021	0.0029
IMF6	0.0103	5.~~7757~~7748×10^{-5}	0.0043	0.0056
IMF7	0.0017	3.~~6036~~026×10^{-5}	9.1374×10^{-4}	0.0010
RES	3.0342×10^{-5}	2.0163×10^{-6}	1.1572×10^{-5}	1.~~5025~~017×10^{-5}

~~The prediction ability of the second mode decomposition component IMF2 is roughly equivalent to the IMF1. Except for the four months of May, September, October, and November, the accuracies of prediction results of other months are satisfactory. The prediction results of the first three intrinsic mode function components IMF1, IMF2, and IMF3 are basically the same as the actual data. In the prediction results of the fourth mode component IMF4, except for slight error in December, the prediction ability is better. The predicted results of the last three intrinsic mode function components IMF5, IMF6, IMF7 and the residue RES are basically consistent with the observation results.~~

The prediction results of the monthly mean SSTA in 2017 are obtained by reconstructing the mode decomposition components (Fig. 10) and the absolute error (ERR) of prediction results have~~s~~ been shown in Table 3. It can be seen from the figure and table that the prediction results based on the EEMD-BPNN model have larger ERRs in January and ~~-~~August, exceeding 0.3 °C, and the accuracies of prediction results in other months are satisfactory (the ERR is less than 0.3). The prediction accuracy based on the CEEMD-BPNN model is satisfactory, except for the ERR exceeding 0.1 °C in October, and the prediction ability based on the CEEMD-BPNN model is generally better than that of the EEMD-BPNN model.

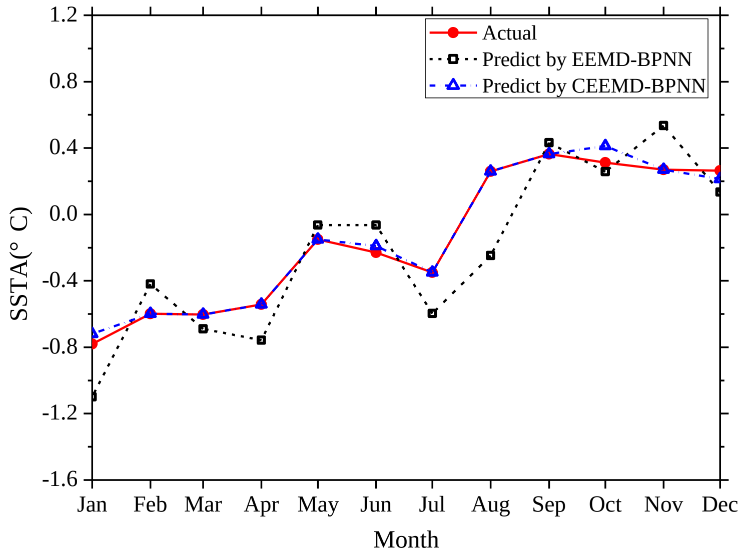

Fig. 10 Monthly SSTA prediction results based on the hybrid improved EMD-BPNN models in 2017.

Table 3. The absolute errors ERRs of the SSTA prediction results based on the two different hybrid improved EMD-BPNN models (unit: °C).

	EEMD-BPNN model	CEEMD-BPNN model		EEMD-BPNN model	CEEMD-BPNN model
Jan	0.3188	0.0623	Sep	0.0687	0.0132
Feb	0.1780	0.0103	Oct	0.0545	0.1607
Mar	0.0867	0.0063	Nov	0.2651	0.0101
Apr	0.2153	0.0137	Dec	0.1290	0.0183
May	0.0854	0.0102	**Min ERR**	**0.0545**	**0.0063**
Jun	0.1662	0.0224	**Max ERR**	**0.5068**	**0.1607**
Jul	0.2474	0.0077	**Mean ERR**	**0.1935**	**0.0289**
Aug	0.5068	0.0112	**RMSE**	**0.2299**	**0.0512**

~~The prediction values based on the CEEMD-BPNN model and the observation values at the significance level of 0.001, the correlation coefficient reached 0.97~~Correlation coefficient between the prediction values based on the CEEMD-BPNN model and observations is shown that the value of the correlation coefficient that indicates a significance level of 0.001 and the correlation coefficient reached 0.97.~~,~~ The result ~~which~~ indicates that SSTA in 2017 had been predicted accurately by the CEEMD-BPNN model. As can be seen from the above discussions, the ERR of decomposition components based on the EEMD and CEEMD algorithms will affect the accuracy of the final prediction results. Table 3 shows that predicting results of the hybrid CEEMD and BPNN model are ameliorated a lot as compared to the EEMD-BPNN direct predicting model. This is because after CEEMD, the original unsteady and nonlinear data are changed into certain components that have fixed frequency and periodicity. The CEEMD algorithm with less decomposition error has less error in the final prediction results, which proves that the CEEMD method has more advantages in data decomposition than the EEMD method. At the same time, we can find that the final prediction error of the two prediction models mainly comes from the first three mode decomposition components, and the error of the last five components has little effect on the accuracy of the final prediction results.

6 Conclusions

This paper presents a ~~novel~~ SST predicting method based on the hybrid ~~improved~~ EMD algorithms and BP neural network method to process the SST data with ~~strong~~ nonlinearity and non-stationarity. Through EEMD and CEEMD algorithms, SSTA time-series data are decomposed into different IMFs and a residue RES. BP neural network is applied to predict individual IMFs and the residue RES. Final results can be obtained by adding the predicting results of individual IMFs and RES.

In order to illustrate the effectiveness of the proposed approach, a case study was carried out. SSTA ~~predicton~~prediction results based on the hybrid EEMD-BPNN model and the hybrid CEEMD-BPNN model are discussed respectively. In comparison, the proposed hybrid CEEMD-BPNN model is much better and its prediction results are more accurate.

From the absolute error of the prediction results of each component IMF and the absolute error of the predicted SSTA, the prediction error of SSTA mainly comes from the prediction of the first three mode decomposition component (IMF1, IMF2 and IMF3), because the first three mode components still have strong nonlinearity and non-stationarity. As the nonlinearity gradually decreases, the absolute error of the prediction results gradually decreases.

SST prediction has been only preliminary carried out based on the two improved EMD algorithms and BP neural network in this paper. The results show that the hybrid CEEMD-BPNN model is more accurate in predicting SST. This work can provide a reference for predicting SST and El Niño in the future. In the follow-up study, how to improve the forecast duration is the focus of this work.

Acknowledgement

This work was supported by National Natural Science Foundation of China (Grant Nos. 51809023, 51879015, 51839002, 51809021 and 51509023).

References:

Amezquita-Sanchez, J. P. and Adeli, H.: A new music-empirical wavelet transform methodology for time–frequency analysis of noisy nonlinear and non-stationary signals, Digit. Signal Process., 45, 55-68, https://doi.org/10.1016/j.dsp.2015.06.013, 2015.

Banzon, V., Smith, T. M., Chin, T. M., Liu, C., and Hankins, W.: A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies, Earth Syst. Sci. Data, 8, 165-176, https://doi.org/10.5194/essd-8-165-2016, 2016.

Bond, N. A., Cronin, M. F., Freeland, H., and Mantua N.: Causes and impacts of the 2014 warm anomaly in the NE Pacific. Geophys. Res. Lett., 42, 3414-3420, https://doi.org/10.1002/2015GL063306, 2015.

Buckley, M. W., Ponte, R. M., Forget, G., and Heimbach, P.: Low-frequency SST and upper-ocean heat content variability in the North Atlantic, J. Climate, 27, 4996-5018, https://doi.org/10.1175/JCLI-D-13-00316.1, 2014.

Chen, C., Cane, M. A., Henderson, N., Lee, D. E., Chapman, D., Kondrashov D., and Chekroun, M. D.: Diversity, nonlinearity, seasonality, and memory effect in ENSO simulation and prediction using empirical model reduction, J. Climate, 29: 1809-1830, https://doi.org/10.1175/JCLI-D-15-0372.1, 2016b.

Chen, Z., Wen, Z., Wu, R., Lin X., and Wang J.: Relative importance of tropical SST anomalies in maintaining the Western North Pacific anomalous anticyclone during El Niño to La Niña transition years, Clim. Dynam., 46, 1027-1041, https://doi.org/10.1007/s00382-015-2630-1, 2016a.

Cheng, Y., Ezer, T., Atkinson, L. P., and Xu, Q.: Analysis of tidal amplitude changes using the EMD method, Cont. Shelf Res., 148: 44-52, https://doi.org/10.1016/j.csr.2017.09.009, 2017.

Deo, M. C., Jha, A., Chaphekar, A. S., and Ravikant, K.: Neural networks for wave forecasting, Ocean Eng., 28: 889-898, https://doi.org/10.1016/S0029-8018(00)00027-5, 2001.

Duan, W. Y., Han, Y., Huang, L. M., Zhao, B. B., and Wang, M. H.: A hybrid EMD-SVR model for the short-term prediction of significant wave height, Ocean Eng., 124, 54-73, https://doi.org/10.1016/j.oceaneng.2016.05.049, 2016.

Duan, W., Huang, L., Han Y., and Huang D.: A hybrid EMD-AR model for nonlinear and non-stationary wave forecasting, J Zhejiang Univ-Sc A, 17(2): 115-129, https://doi.org/10.1631/jzus.A1500164, 2016.

Ezer, T. and Atkinson, L. P.: Accelerated flooding along the US East Coast: on the impact of sea‐level rise, tides, storms, the Gulf Stream, and the North Atlantic oscillations, Earths Future, 2, 362-382, https://doi.org/10.1002/2014EF000252, 2014.

Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., Dunne, J. P., Goddard, P., Morrison, A. K., Rosati, A., Wittenberg, A. T., Yin, J., and Zhang, R.: Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. J. Climate, 28, 952-977, https://doi.org/10.1175/JCLI-D-14-00353.1, 2015.

He, J., Deser, C., and Soden, B. J.: Atmospheric and oceanic origins of tropical precipitation variability. J. Climate, 30, 3197-3217, https://doi.org/10.1175/JCLI-D-16-0714.1, 2017.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N., Tung, C. C., and Liu, H. H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, P. Roy. Soc. A-Math. Phy., 454, 903-995. https://doi.org/10.1098/rspa.1998.0193, 1998.

Huang, N. E. and Wu, Z.: A review on Hilbert‐Huang transform: Method and its applications to geophysical studies, Rev. Geophys., 46, RG2006, https://doi.org/10.1029/2007RG000228, 2008.

Hudson, D., Alves, O., Hendon, H. H., Wang, G.: The impact of atmospheric initialisation on seasonal prediction of tropical Pacific SST, Clim. Dynam., 36, 1155-1171, https://doi.org/10.1007/s00382-010-0763-9, 2011.

Jain, P. and Deo, M. C.: Neural networks in ocean engineering, Ships Offshore Struc., 1, 25-35, https://doi.org/10.1533/saos.2004.0005, 2006.

Khan, M. Z. K., Sharma, A., and Mehrotra, R.: Global seasonal precipitation forecasts using improved sea surface temperature predictions, J Geophys. Res. -Atmos., 122, 4773-4785, https://doi.org/10.1002/2016JD025953, 2017,

Kim, Y., Kim, H., and Ahn, I. G.: A study on the fatigue damage model for Gaussian wideband process of two peaks by an artificial neural network, Ocean Eng., 111, 310-322, https://doi.org/10.1016/j.oceaneng.2015.11.008, 2016.

Kumar, M., Parmar, C., Chaudhary, V., Kumar, A., and SST-1 team.: Observation of plasma shift in SST-1 using optical imaging diagnostics, J Phys. Conf. Ser., 823, 012056, https://doi.org/10.1088/1742-6596/823/1/012056, 2017.

Lee, H. S.: Estimation of extreme sea levels along the Bangladesh coast due to storm surge and sea level rise using EEMD and EVA, J Geophys. Res.-Oceans, 118, 4273-4285, https://doi.org/10.1002/jgrc.20310, 2013,

Lee, T. L.: Back-propagation neural network for long-term tidal predictions, Ocean Eng., 31, 225-238, https://doi.org/10.1016/S0029-8018(03)00115-X, 2004.

López, I., Aragonés, L., Villacampa, Y., and Serra, J. C.: Neural network for determining the characteristic points of the bars, Ocean Eng., 136: 141-151, https://doi.org/10.1016/j.oceaneng.2017.03.033, 2017.

Monteiro, E., Yvonnet, J., He, Q. C.: Computational homogenization for nonlinear conduction in heterogeneous materials using model reduction. Comp. Mater. Sci., 42, 704-712, https://doi.org/10.1016/j.commatsci.2007.11.001, 2008.

Motulsky, H. J. and Ransnas, L. A.: Fitting curves to data using nonlinear regression: a practical and nonmathematical review, Faseb J., 1, 365-374. https://doi.org/10.1096/fasebj.1.5.3315805, 1987.

Pan, H., Guo, Z., Wang, Y., and Lv, X.: Application of the EMD method to river tides, J. Atmos. Ocean. Tech., 35, 809-819, https://doi.org/10.1175/JTECH-D-17-0185.1, 2018.

Pearson, R. K. and Pottmann, M.: Gray-box identification of block-oriented nonlinear models, J. Process Contr., 10, 301-315, https://doi.org/10.1016/S0959-1524(99)00055-4, 2000.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax., M. G.: Daily high-resolution-blended analyses for sea surface temperature, J. Climate, 20, 5473-5496, https://doi.org/10.1175/2007JCLI1824.1, 2007.

Sadeghifar, T., Motlagh, M. N., Azad, M. T., and Mahdizadeh, M. M.: Coastal wave height prediction using Recurrent Neural Networks (RNNs) in the south Caspian Sea, Mar. Geod., 40, 454-465, https://doi.org/10.1080/01490419.2017.1359220, 2017.

Savitha, R. and Mamun, A. A,: Regional ocean wave height prediction using sequential learning neural networks, Ocean Eng., 129: 605-612, https://doi.org/10.1016/j.oceaneng.2016.10.033, 2017.

Sukresno, B., Hanintyo, R., Kusuma, D. W., Jatisworo, D., and Murdimanto., A.: Three-way error analysis of sea surface temperature (SST) between HIMAWARI-8, buoy, and mur SST in SAVU Sea, Int. J. Remote Sens. Earth Sci., 15, 25-36, https://doi.org/10.30536/j.ijreses.2018.v15.a2855, 2018,

Takakura, T., Kawamura, R., Kawano, T., Ichiyanagi, K., Tanoue, M., and Yoshimura, K.: An estimation of water origins in the vicinity of a tropical cyclone's center and associated dynamic processes, Clim. Dynam., 50, 555-569, https://doi.org/10.1007/s00382-017-3626-9, 2018.

Tang, L., Dai, W., Yu, L., and Wang, S.: A novel CEEMD-based EELM ensemble learning paradigm for crude oil price forecasting, Int. J. Inf. Tech. Decis., 14, 141-169, https://doi.org/10.1142/S0219622015400015, 2015.

Wang, S., Zhang, N., Wu, L., and Wang, Y.: Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method, Renew. Energ., 94, 629-636, https://doi.org/10.1016/j.renene.2016.03.103, 2016.

Wang, W., Chau, K., Xu, D., and Chen, X.: Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition, Water Resour. Manag., 29, 2655-2675, https://doi.org/10.1007/s11269-015-0962-6, 2015.

Wang, W., Tang, R., Li, C., Liu, P., and Luo, L.: A BP neural network model optimized by Mind Evolutionary Algorithm for predicting the ocean wave heights, Ocean Eng., 162, 98-107, https://doi.org/10.1016/j.oceaneng.2018.04.039, 2018.

Wang, Y., Wilson, P. A., Zhang, M., and Liu, X.: Adaptive neural network-based backstepping fault tolerant control for underwater vehicles with thruster fault, Ocean Eng., 110, 15-24, https://doi.org/10.1016/j.oceaneng.2015.09.035, 2015.

Wiedermann, M., Donges, J. F., Handorf, D., Kurths, J., and Donner, R. V.: Hierarchical structures in Northern Hemispheric extratropical winter ocean–atmosphere interactions, Int. J. Climatol., 37, 3821-3836, https://doi.org/10.1002/joc.4956, 2017.

Wu, L. C., Kao, C. C., Hsu, T. W., Jao K. C. and Wang, Y. F.: Ensemble empirical mode decomposition on storm surge separation from sea level data, Coast. Eng. J., 53, 223-243, https://doi.org/10.1142/S0578563411002343, 2011.

Wu Z., Schneider E. K. and Kirtman B. P.: The modulated annual cycle: an alternative reference frame for climate anomalies, Clim. Dyna., 31(7-8): 823-841, https://doi.org/10.1007/s00382-008-0437-z, 2008.

Wu, Z. and Huang, N. E.: Ensemble empirical mode decomposition: a noise-assisted data analysis method, Adv. Adap. Data Anal., 1, 1-41, https://doi.org/10.1142/S1793536909000047, 2009.

Wu Z., Jiang C., Chen J., Long Y., Deng B. and Liu X.: Three-Dimensional Temperature Field Change in the South China Sea during Typhoon Kai-Tak (1213) Based on a Fully Coupled Atmosphere–Wave–Ocean Model, Water, 11(1): 140, https://doi.org/10.3390/w11010140, 2019a.

Wu Z., Jiang C., Deng B., Chen J., Long Y., Qu K. and Liu X.: Numerical investigation of Typhoon Kai-tak (1213) using a mesoscale coupled WRF-ROMS model, Ocean Eng., 175: 1-15.

https://doi.org/10.1016/j.oceaneng.2019.01.053, 2019b.

Yeh, J. R., Shieh, J. S., and Huang, N. E.: Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method, Adv. Adap. Data Anal., 2, 135-156, https://doi.org/10.1142/S1793536910000422, 2010.

Zheng, X. T., Xie, S. P., Lv, L. H., and Zhou, Z. Q.: Intermodel uncertainty in ENSO amplitude change tied to Pacific Ocean warming pattern, J. Climate, 29, 7265-7279, https://doi.org/10.1175/JCLI-D-16-0039.1, 2016.

Zhu, J., Huang, B., Kumar, A., and Kinter, J. L.: Seasonality in prediction skill and predictable pattern of tropical Indian Ocean SST, J. Climate, 28, 7962-7984, https://doi.org/10.1175/JCLI-D-15-0067.1, 2015.