Multistep-Ahead Prediction of Ocean SSTA Based on Hybrid Empirical Mode Decomposition and Gated Recurrent Unit Model

Xiaoyin Liu, Ning Li, Jun Guo, Zhongyong Fan, Xiaoping Lu, Weifeng Liu, Senior Member, IEEE, and Baodi Liu, Member, IEEE

Abstract—The prediction of sea surface temperature anomalies (SSTA) is vital to the study of marine ecosystems and global climate. The SSTA can be accurately forecasted one step ahead by numerical and statistical methods. However, multistep-ahead forecasting for SSTA is greatly challenging since the nonlinearity and nonstationarity of SSTA and the lag problem of prediction. Therefore, in this article, a multistep-ahead SSTA forecasting method based on the hybrid empirical mode decomposition (EMD) and gated recurrent unit (GRU) model is proposed, considering that EMD has the advantage of reducing data complexity and GRU is good at long-term prediction of data. First, the EMD algorithm is applied to obtain several intrinsic modal functions that are more stationary than the original data. Then, GRU is used for multistep forecasting, in which three multistep forecasting strategies (recursive, direct, and multioutput) are compared. The proposed hybrid model is validated by multistep forecasting for monthly average SSTA at the Niño3.4 region. The experimental results show that using reconstruction error as part of the prediction effectively improves the prediction accuracy of the EEMD-GRU model and outperforms other EMD algorithms combined with GRU. Compared with traditional models, the EMD-GRU model can better predict future multilong-term trends of SSTA and effectively alleviate the problem of prediction lag of the traditional model. Finally, this model is applied to the Niño3.4 regional SSTA prediction, and the results show that this model can provide a reference for ocean research.

Index Terms—Empirical mode decomposition, gated recurrent unit neural network, multistep-ahead prediction, prediction lags, sea surface temperature anomalies.

Manuscript received 11 May 2022; revised 11 July 2022; accepted 18 August 2022. Date of publication 24 August 2022; date of current version 13 September 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61671480, in part by the Major Scientific and Technological Projects of CNPC under Grant ZD2019-183-008, in part by the Natural Science Foundation of Shandong Province under Grant ZR2019MP073, and in part by the Fundamental Research Funds for the Central Universities, China University of Petroleum (East China) under Grant 20CX05001A. (Corresponding authors: Xiaoping Lu; Weifeng Liu.)

Xiaoyin Liu, Ning Li, Jun Guo, and Zhongyong Fan are with the College of Science, China University of Petroleum (East China), Qingdao 266580, China (e-mail: liuxyxt@163.com; ningl5210@gmail.com; 1653311514@qq.com; 1351140963@qq.com). Xiaoping Lu is with the COSMO Industrial Intelligence Institute, Qingdao 266500, China (e-mail: luxiaoping@haier.com).

Weifeng Liu and Baodi Liu are with the College of Control Science and Engineering, China University of Petroleum (East China), Qingdao 266580, China (e-mail: liuwf@upc.edu.cn; thu.liubaodi@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2022.3201228

I. INTRODUCTION

Sea surface temperature anomalies (SSTA) have profound effects on marine ecosystems and climate change [1]. In recent years, as ocean temperatures continue to rise, extreme weather has become more frequent [2]. Therefore, many scholars have studied SSTA forecasting as a way to study the prediction of ocean weather [3]. However, the nonlinear and nonstationary characteristics of SSTA pose a great challenge to the prediction of SSTA [4].

At present, the main SSTA forecasting methods include numerical methods and statistical methods. Chiodi [5] considered the laws of ocean and atmospheric dynamics and used ocean general circulation model accurately predict the development of SSTA. Although numerical forecasting is the main method of SSTA forecasting at this stage, there are problems such as large amounts of calculations and sensitivity to initial conditions [6]. Statistical forecasting methods, when the amount of data in the sample is large, can establish a forecast model in the statistical sense without considering the physical laws of the forecast object [7]. Statistical forecasting methods used for SSTA forecasting include support vector machines [8], autoregressive integrated moving average (ARIMA) [9], regression model [10], and neural networks [11], [12], [13]. However, the nonlinear and nonstationary data reduces prediction accuracy. Therefore, some signal processing methods, such as wavelet analysis [14] and EMD methods [15], are used to decrease the data complexity to improve forecasting effects.

Among statistical methods, neural networks have strong nonlinear expression capability and are now widely used in the prediction of complex time-series data [16]. Wu et al. [17] used the back propagation (BP) method to establish a nonlinear SSTA forecasting model that improves prediction accuracy compared to traditional linear model. In recent years, the long short-term memory neural network (LSTM) has been widely used in SSTA prediction [18], [19], [20]. Recurrent neural network (RNN) was proposed by Elman [21], which showed unique advantages for time-series data prediction. However, RNN suffers from the disadvantages of gradient disappearance and explosion, making them only have short-term memory [22]. To address this problem, Schmidhuber and Hochreiter [23] improved the RNN structure in 1997 and proposed LSTM, but it has the disadvantage of slow and time-consuming training. Chung

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
et al. [24] proposed GRU neural network, and compared to LSTM, GRU has a simple structure, which greatly improves the training efficiency. Su et al. [25] used bidirectional LSTM (Bi-LSTM) for predicting SSTA, and proved Bi-LSTM is more advantageous in time-series modeling. Sun et al. [26] used Internet of Things technology and temporal convolutional network to predict the monthly SST. Zhang et al. [27] applied GRU neural network to predict SSTA and proved the GRU performs better than LSTM in terms of stability and accuracy.

Among signal processing methods, empirical modal decomposition (EMD) has been widely used in the data processing and prediction of climate, such as the ocean, meteorology, and rainfall [28], [29], [30], [31]. These findings suggest that EMD and its improved models can effectively improve prediction accuracy. EMD was proposed by Huang et al. [32] to process and analyze the signal, which can analyze the physical characteristics of the various data, thus reducing the complexity of nonlinear time-series data. However, EMD has the problem of modal aliasing. Wu and Huang [33] proposed the ensemble empirical mode decomposition (EEMD) method by adding white noise to the original signal, which effectively solves the modal aliasing problem. However, the EEMD decomposition is computationally intensive and the added white noise cannot be completely neutralized. Yeh et al. [34] proposed a complete ensemble empirical modal decomposition (CEEMD) method by adding complementary white noise to the original signal, which effectively reduces the reconstruction error caused by white noise. Torres et al. [35] proposed a CEEMD with adaptive noise (CEEMDAN) method to obtain a component with smaller error and stronger physical meaning. Currently, a combination of signal processing and neural network methods has been applied to temperature prediction. Zhang et al. [36] used the EEMD-LSTM model to forecast land surface temperature and proved the hybrid model is superior to the signal neural network. Shao et al. [37] combined the empirical orthogonal function of multivariate, CEEMD, and multilayer perceptron to forecast SST considering the correlation of different variables in the real marine environment. Wu et al. [38] used the CEEMD-BP model to forecast SST and proved that the prediction result of the CEEMD-BP outperforms EEMD-BP, but this conclusion is not absolute.

Ocean forecasting studies often require data for multiple months into the future, and most of the previous studies are one-step forecasting. However, multistep forecasting of ocean data is rarely studied because it often suffers from low prediction accuracy and lag. Meanwhile, the complexity of SSTA data poses a great challenge to multistep prediction [39]. Niu et al. [40] built optimal feature selection and an artificial neural network model for multistep wind prediction. Wu et al. [41] used a decomposition-ANFIS model to predict wave conditions multistep-ahead. These studies demonstrated that the prediction error increases greatly with prediction steps increasing. However, the prediction results of these models have lagging problems. Therefore, the main purpose of this article is to build a multistep-ahead model for SSTA prediction and alleviate the problem of the multistep forecasting lag of the previous methods.

Considering the advantages of empirical modal decomposition to reduce the nonlinear characteristics of data and the high prediction accuracy of deep neural networks, this article builds a hybrid EEMD-GRU model for predicting SSTA. The contributions of this article are as follows:

1) The EMD algorithm is used to reduce the complexity of the data and a hybrid model of EEMD and GRU is established.
2) The SSTA multistep-ahead prediction model is built based on the hybrid EEMD-GRU model.
3) The reconstruction error of EEMD decomposition is used as a component for its prediction, which effectively improves the accuracy of the prediction.

The experiments show that one-step-ahead forecasting of EEMD-GRU performs better than other EMD algorithms combined with GRU, and the hybrid EEMD and GRU models outperform the single traditional model and alleviate the lag problem of traditional multistep-ahead prediction.

The rest of this article is organized as follows. Section II gives the methodology for SSTA multistep prediction, where the hybrid EMD and GRU model is explained. Section III describes the data and study area, and introduces relevant parameters and evaluation metrics of the experiment. Section IV analyzes and discusses the different EMD decomposition results, one-step forecasting results, and multistep forecasting results. Finally, Section V concludes this article.

II. METHODOLOGY

The nonstationary of the time-series data makes prediction extremely difficult, so this article combines signal processing methods and neural network models to build a multistep prediction model. This article first uses modal empirical decomposition signal processing to reduce the complexity of the data and then uses a neural network prediction method to achieve multistep prediction.

A. Empirical Mode Decomposition

Compared with other signal processing methods, EMD overcomes no self-adaptation problems of the basis functions. Various improved EMD algorithms (EEMD, CEEMD, CEEMDAN) are currently used to reduce the reconstruction error of the algorithm decomposition by adding different forms of white noise to the signal [33], [34], [35], so only the principle of EMD decomposition is presented here. Fig. 1 illustrates the EMD decomposition process [32], with the specific steps of the decomposition as follows:

Step 1: Calculate the upper and lower extreme value points of the original signal $x(t)$ to obtain the upper and lower envelopes. Calculate the mean value of the upper and lower envelopes to obtain the mean envelope $a(t)$. Subtract the mean envelope from the original signal to obtain the intermediate signal $c_{ik}(t)$.

Step 2: Determine whether the signal $c_{ik}(t)$ meets the two conditions of IMF: 1) The difference between the number of extreme value points and the number of crossing zero points must be equal to 0 or 1; 2) The upper and lower envelopes are locally symmetrical with respect to the time axis. If it is
Step 3: Subtract the IMF obtained from Step 2 from the original signal and use the result as the new original signal, repeat Step 1 and Step 2 until the IMF is a monotonic function, and end the EMD decomposition.

B. GRU Neural Network

GRU and LSTM neural networks are both variations of recurrent networks (RNN), both of which solve gradient explosion in RNNs. The GRU has been widely used in the prediction of nonlinear data, with only two gate structures, update gate, and reset gate (see Fig. 2), which is one less gate than the LSTM network structure and can effectively prevent overfitting.

The GRU network structure includes four main layers: the input, the hidden, the fully connected, and the output (see Fig. 2). The hidden layer (GRU cell) is the one with the memory function. In the GRU, $r_t$ is the reset gate. The larger its value the less information is ignored in the previous moment, which is calculated as follows [24]:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$$  

where $W_r$ represents the weight of the reset gate, $x_t$ is the input, $\sigma$ is the sigmoid function, and $h_{t-1}$ is the output of previous moment. $z_t$ is the update gate. The larger its value the more information is retained from the previous moment, which is computed by

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$  

where $W_z$ is the weight of the update gate, and the excitation $h_t$ of the GRU network at time $t$ is the linear interpolation of the excitation $h_{t-1}$ at the previous moment and the candidate excitation $\hat{h}_t$

$$h_t = (1 - z_t) \cdot \Theta h_{t-1} + z_t \cdot \hat{h}_t$$  

where $\Theta$ represents the multiplication of the corresponding point elements of the matrix, and the candidate excitation $\hat{h}_t$ is calculated as follows:

$$\hat{h}_t = \tanh(W \cdot [r_t \Theta h_{t-1}, x_t])$$  

where $W$ is the weight matrix corresponding to $\hat{h}_t$ and $\tanh$ is the hyperbolic tangent function.

C. One-Step-Ahead Prediction Model

For a one-step-ahead prediction model, signal processing, and neural network methods are used in SSTA prediction. Fig. 3 gives the EMD-GRU model forecasting steps applied in SSTA prediction [33]. First, the SSTA time series data are decomposed by EMD to obtain intrinsic modal function modes (IMFs), then each IMF component is forecasted separately via GRU neural network, and then the predictions of each component are reconstructed to obtain the final SSTA prediction results.

D. Multistep-Ahead Prediction Model

In marine atmospheric studies, especially El Niño and La Niña predictions, data for multiple months in the future are required and one-step predictions cannot meet the needs of the study, so multistep prediction results are essential for analyzing marine studies. Multistep prediction refers to using the historical time series $[x_1, \ldots, x_N]$ to predict the subsequent $h$-step series $[x_{N+1}, \ldots, x_{N+h}]$

Taieb et al. [39] reviewed the time series multistep prediction strategies. The current methods of multistep prediction are recursive, direct, and multioutput strategies. In (5)–(7), $g$ and $G$ represent the model of prediction and represent the functional relationship between the known and forecasted data, $d$ is the number of function input variables, and $w$ denotes the error term.

$$x_{t+1} = g(x_t, \ldots, x_{t-d+1}) + w$$  

$$[x_{t+h}, \ldots, x_{t+1}] = G(x_t, \ldots, x_{t-d+1}) + w$$  

$$x_{t+h} = g_h(x_t, \ldots, x_{t-d+1}) + w.$$  

Recursive strategy is to use the output of the last one-step forecast as the input of the next forecast, the forecasting model is shown in (5), but there is the disadvantage of error accumulation as the predictions are used to determine subsequent predictions; Multioutput strategy is a direct many-to-many mapping relationship to predict future multistep data at one time (6). It is a unique advantage of deep neural networks, but the number of features required for the data is particularly high and the accuracy of the results is difficult to guarantee.

The direct strategy is to build $h$ models to predict the future data from 1 to $h$ months, respectively (7). The advantage of the direct strategy is that it does not rely on the prediction results of the previous step, which overcomes the disadvantage of error accumulation and has relatively high accuracy for short-term prediction. Since the $h$ models are learned independently, the conditions of model predictions are independent, which will affect the prediction accuracy to some extent.
III. EXPERIMENTS SETUP

A. Study Area and Dataset

El Niño (La Niña) is defined as +0.5 °C (−0.5 °C) above (below) the mean temperature in the Niño3.4 region [see Fig. 4(b)] for five consecutive months. The SSTA refers to the magnitude of the deviation from the mean temperature of previous years. The SSTA data in this article are derived from the extended reconstructed sea surface temperature dataset with a resolution of 2° × 2° from the international integrated ocean-atmosphere dataset, which can be used for long-term global and basin-wide studies [42]. Fig. 4(a) shows the global SSTA distribution in July 2021, which can be concluded as anomalously low SST in the east-central Pacific. Since many extreme kinds of weather are closely connected with SSTA in the equatorial central-east Pacific, the Niño 3.4 region (170 °W to 120 °W, 5 °S to 5 °N) is selected as the experimental area in this article.

From Fig. 4(c), it can be concluded that the SSTA are nonstationary, nonlinear, and stochastic. The large range of fluctuations makes prediction extremely difficult, and the SSTA near the equator is more nonlinear and more difficult to predict than other regions. The graph also shows that the top three time periods on the timeline where SSTA peaked are 1982–1983, 1997–1998, and 2015–2016, consistent with strong El Niño years, which indicates that SSTA can reflect the El Niño phenomenon to some extent. If the trend of SSTA in the coming months can be predicted, it will have important implications for the prediction of El Niño-Southern Oscillation events.

B. Data Processing and Network Training Strategies

A diagram of the multistep prediction model training and prediction is given in Fig. 5. The first part of the SSTA data is the training set, the latter part is used to test prediction accuracy, and finally, the last \( N \) data forecast the SSTA for the next \( h \) months. The network parameters are set as follows: SSTA data are processed using minimum and maximum normalization, the input to the dataset is 12 consecutive months of time data, the target output is the latter month of the input, the number of hidden layers is 2, the number of neurons is 10, data training iterations is 100, the error is judged using mean squared deviation, and the optimizer is Adam.

Fig. 6 illustrates the process of direct strategy training and prediction [40]. In the training process, a total of \( h \) models are trained, where the blue boxes represent the input data and the red, grey, and yellow boxes represent the output data of the different models. \( \text{Model}_h \) represents the model predicted \( h \) steps ahead. In the testing process, a total of \( K - N - h \) test samples are applied to each of the \( h \) models, where the blue boxes represent the input data and the grey boxes in red font are the predicted data. The final prediction results for each model are aggregated together and used to assess the accuracy of the predictions.

C. Evaluation Metrics

To quantitatively evaluate and compare the predictive ability of different models, mean absolute error (MAE), maximum error (MRE), root square error (RMSE), and concordance index (CI) are selected as evaluation indicators, and the formulas are as follows:

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |x_i - y_i| \quad (8)
\]

\[
\text{MRE} = \max \left( |x_i - y_i| \right) \quad (9)
\]
Fig. 4. Visualization of study area and dataset. (a) Global SSTA distribution in July 2021. (b) SSTA situation of Niño 3.4 and surrounding areas in July 2021. (c) SSTA in the Niño 3.4 region from January 1980 to December 2020.

Fig. 5. Schematic diagram of multistep prediction training and testing of SSTA. The previous N SSTA data are used to predict the future h SSTA data in the future prediction phase.

IV. RESULTS AND ANALYSIS

A. Decomposition Results

The decomposition effects of different EMD algorithms are discussed in this section. Fig. 7 gives the results of the decomposition of SSTA at the 0 °N, 150 °W location by the different EMD algorithms. From Fig. 7, it can be concluded that the final decomposition of the IMF components differs for different EMD algorithms. For all EMD algorithms, the first eigenmodal function component has a high frequency and exhibits a certain nonsmoothness, and the nonsmoothness of IMFi decreases as i increases. The last few IMF components have a certain periodicity and relatively regular fluctuations.

In addition, it can be concluded from the figure that the amplitude of the second component of the EEMD decomposition is smaller than the results of the other EMD decompositions. The small amplitude value can reduce the final prediction error to some extent, because the larger the fluctuation of the component, the larger the prediction error. In Fig. 7, RERES represents the reconstruction errors for different EMD decompositions, where the values of both EMD, CEEMD, and CEEMDAN are at the $10^{-15}$ level and EEMD are at the $10^{-2}$ level. The decomposition effect of the EMD algorithm can be reflected by the magnitude

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2}
\]

\[
\text{CI} = 1 - \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (|x_i - \bar{x}| + |y_i - \bar{y}|)^2}
\]

where \(x_i\) and \(y_i\) are the true and predicted value, respectively, \(\bar{x}\) and \(\bar{y}\) are the average value of the true and predicted value, respectively, and \(n\) is the number of samples.
Fig. 6. Architecture of the multistep-ahead forecasting direct strategy. The $h$ models are individually trained to predict the data in the next $h$ steps. Model 1 uses the first $N$ data as input and the first data after as output, and model $h$ uses the first $N$ data as input and the $h$ data after as output.

Fig. 7. IMF components and RES of monthly SSTA at 0°N 150°W during 1980–2020 based on the (a) EMD, (b) EEMD, (c) CEEMD, and (d) CEEMDAN algorithm. The EEMD and CEEMD decomposition structures only show the first 7 IMF components. RERES stands for reconstruction errors, which is the difference between the sum of all decompositions and the original data.

of the final reconstruction error, which is calculated as follows:

$$RERES = x(t) - \sum_{i=1}^{n} \text{IMFi}$$  \hspace{1cm} (12)

where $x(t)$ is the original signal and IMFi is the decomposed component. The reconstruction error results of different EMD algorithms are given in Table I, where the reconstruction errors of EMD, CEEMD, and CEEMDAN are close to 0, and EEMD is close to 0.02. It can be found that the reconstruction error of EEMD is larger than that of other EMD algorithms. This is mainly due to the fact that the white noise added in the original signal cannot be canceled by EEMD decomposition, which may lead to a larger reconstruction error of EEMD. However, the
CEEMD and CEEMDAN algorithms add complementary white noise to the original signal, and the added white noise is basically canceled, effectively reducing reconstruction errors caused by white noise.

### B. One-Step-Ahead Prediction Results

Table II shows the GRU prediction error results for SSTA data at the 0 °N, 150 °W location after different EMD decompositions. It can be concluded that, for different EMD algorithms, the errors in the predictions of the first two components are large, mainly caused by the high frequency of the first two components. The predicted mean squared deviations of the first components are EEMD, CEEMDAN, CEEMD, EMD in descending order of magnitude. All components of the EEMD predict indicators that are mostly better than the CEEMD and CEEMDAN. After original EMD decomposition, although the component IMF3–IMF6 predictors are better than those predicted by EEMD decomposition, the errors of components IMF1-IMF2 are much larger than those of EEMD. Therefore, the EEMD decomposition prediction is better than the other EMD decomposition predictions.

It is also found experimentally that it is not the case that the better the decomposition of the EMD algorithm, the higher the accuracy of the prediction in combination with the GRU model. The prediction effectiveness of the GRU is closely related to the large amplitude and frequency of the decomposition components of the different EMD algorithms. The larger the frequency, the larger the nonlinearity, and the larger the prediction error of the GRU. The variance of the data can reflect the magnitude of the data nonlinearity to some extent, but not completely. Table III gives the variance of the first two components of different EMD decompositions, and it can be concluded that the first two components of EEMD are smaller than other EMD algorithms. This indicates to some extent that the nonlinearity of EEMD decomposition is weaker and the GRU prediction ability is better.
TABLE IV
PREDICTION RESULTS OF EEMD-GRU WITH AND WITHOUT RECONSTRUCTION ERROR

| Position     | Variance | RERES       |               |               |               |               |
|--------------|----------|-------------|---------------|---------------|---------------|---------------|
|              |          | Max Error   | Mean Error    | Considering RERES | Not Considering RERES |
| (50S,150W)  | 0.20849  | 0.15680     | 0.06059       | 0.17253       | 0.94093       | 0.19774       | 0.92305       |
| (30S,150W)  | 0.38931  | 0.28731     | 0.10412       | 0.20699       | 0.93398       | 0.38373       | 0.77003       |
| (10S,150W)  | 0.252571 | 0.25669     | 0.10796       | 0.13708       | 0.97771       | 0.35007       | 0.85325       |
| (10N,150W)  | 0.179899 | 0.33463     | 0.14042       | 0.10246       | 0.96020       | 0.17677       | 0.88109       |
| (30N,150W)  | 0.483399 | 0.64904     | 0.25823       | 0.29802       | 0.94094       | 0.74416       | 0.63450       |
| (50N,150W)  | 0.66396  | 0.15911     | 0.05032       | 0.40732       | 0.91112       | 0.41014       | 0.91206       |

The bold entities show that the larger the reconstruction error, the more obvious it is to improve the prediction accuracy by using it as part of the prediction.

TABLE V
PREDICTION RESULTS OF DIFFERENT COMBINED EMD AND GRU MODELS FOR DIFFERENT LOCATIONS

| Position     | Variance | EMD          |               | EEMD          |               | CEEMD         |               | CEEMDAN       |               |
|--------------|----------|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|              |          | RMSE | CI   | RMSE | CI   | RMSE | CI   | RMSE | CI   |
| (50S,150W)  | 0.20849  | 0.21773 | 0.90000 | 0.17253 | 0.94093 | 0.18335 | 0.93111 | 0.21965 | 0.89919 |
| (30S,150W)  | 0.38931  | 0.22943 | 0.91669 | 0.20699 | 0.93398 | 0.21149 | 0.92975 | 0.29666 | 0.86097 |
| (10S,150W)  | 0.252571 | 0.19003 | 0.95339 | 0.13708 | 0.97771 | 0.14798 | 0.97299 | 0.18571 | 0.95608 |
| (10N,150W)  | 0.179899 | 0.11841 | 0.94642 | 0.10246 | 0.96020 | 0.10512 | 0.95804 | 0.12112 | 0.94311 |
| (30N,150W)  | 0.483399 | 0.36839 | 0.90741 | 0.29802 | 0.94094 | 0.31558 | 0.93289 | 0.37391 | 0.90557 |
| (50N,150W)  | 0.66396  | 0.45294 | 0.89162 | 0.40732 | 0.91112 | 0.41419 | 0.91002 | 0.45569 | 0.88134 |

In addition, the reconstruction error is large in the EEMD decomposition, and if the effect of this part on the prediction is ignored, it must reduce the final prediction effect. Therefore, the sequence composed of reconstruction errors is also used as part of the final prediction. Since the EMD, CEEMD, and CEEMDAN reconstruction errors are close to zero, the effect of reconstruction errors need not be considered. Table IV gives the result of EEMD-GRU final predictions in six different locations considering the reconstruction error and not considering the reconstruction error. The number of EMD performance with added noise is set to 100. Due to the zero-mean property of white noise, the larger the number, the smaller the reconstruction error of EEMD decomposition, but the computational cost also rises. From the table, it can be concluded that the larger the mean value of the reconstruction error, the greater the impact on the final prediction, and using it as part of the prediction can greatly improve the accuracy of the EEMD-GRU prediction.

To further validate that the EEMD-GRU model outperforms other EMD algorithms combined with GRU, 492 months of SSTA data from 1980 to 2020 are selected for six other locations. The prediction results (RMSE and CI) of different EMD and GRU combination models for different locations are given in Table V. It shows that the EEMD-GRU model outperforms the other models for SSTA prediction in six different locations. Moreover, the effectiveness of the prediction is closely connected to the data variance: the larger the data variance (the less smooth the data), the larger the mean squared error of the prediction, and the lower the consistency index.

C. Results Analysis of Multistep Prediction Strategies

To compare the prediction effects of different kinds of multistep-ahead forecasting strategies, the SSTA data from 1980 to 2014 are used as the training set to calculate forecasting errors for each month from 2015 to 2020 when the lead time is 1–6, respectively. One of the advantages of neural networks is that they can achieve multiple outputs (multiple output strategy), which means that a multiple output model can be trained to forecast continuous future data at one time. However, the prediction error of the multistep strategy is larger for the time-series data of a single SSTA [see Fig. 8(a)], the mean error is about 0.76 °C when the output dimension is 6, and the maximum error reaches 2 °C, which is due to the fewer features of single data. The ARIMA model is commonly used for multistep forecasting, but it has too large a prediction error and the maximum error is also close to 2 °C [see Fig. 8(b)].

Fig. 8(c) shows the recursive strategy forecasting results with different lead times, from which it can be concluded that the prediction error is small in one-step-ahead, but from the second step onward, the prediction accuracy decreases and the maximum error reaches 4 °C. This is mainly due to the problem of error accumulation and has no error feedback mechanism.
If the error in the previous step is large, the error in the next step will show a dispersion trend, and the final prediction value may tend to a constant value when the lead time is large. The direct strategy prediction effect has a common problem with the previous strategy: the error gradually increases as the number of steps increases. However, it overcomes the problem of error accumulation, the stability of prediction is higher, and the prediction error is also smaller than the previous strategies. Therefore, this article uses the direct strategy for the multistep prediction of SSTA.

D. Error Analysis of Different Model Predictions

This section compares the prediction effects of single models and combined models and that of different EMD algorithms with GRU combined models. Fig. 9 shows the results of the error metrics (mean squared error, maximum error, average error, and consistency index) for the single model (GRU, LSTM, BP, ARIMA) and the combined model (EEMD-GRU) with 1–6 steps ahead prediction. From the figure, it can be concluded that the error of each model increases with the lead time. However, the error metrics results of the combined model are much better than those of the single model. The mean squared error of the combined model with 6 steps ahead is smaller than that of the BP with 1 step ahead. The consistency indices of the combined model are all above 0.9, while the minimum values of the single model are all around 0.5. This indicates that the forecasting ability of the combined model outperforms the single model.

Fig. 10 shows the comparison of the multistep-ahead forecasting results of different EMD algorithms with the GRU combination, where the data are SSTA sequences at the position of 0 °N 150 °W. It can be concluded that when the number of prediction steps is 6, the mean squared error and average error of the combined model of EEMD and GRU are about 0.1 lower and the consistency index is about 0.05 higher than that of other EMD algorithms with GRU models. Therefore, the EEMD-GRU model predicts better than the other combined models, from which the previous conclusion is further verified: it is not the case that the lower the reconstruction error of the improved EMD algorithm, the better the prediction effects, which also relays on characteristics of the data itself.

E. Lag Analysis of Different Model Predictions

Multistep prediction models often suffer from lags, so the prediction lags of different models are investigated here. Fig. 11 gives the 1–6 multistep forecasting results for different models, from which the lags of the combined model and the single model predictions are compared. From Fig. 11(a), it can be concluded that when the EEMD-GRU model is used for multistep forecasting, the forecasting accuracy is highest in one step ahead, and the prediction accuracy gradually decreases with steps increasing. Due to the EEMD decomposition, the predicted trend curve is relatively smooth, resulting in a larger error at the crest and trough. However, the model forecasts the future trend of the data well and solves the problem of prediction lag in the traditional model.

Fig. 11(b) shows the multistep forecasting results of the single GRU model, and it can be concluded that as prediction steps increases, the lag phenomenon becomes more serious. In terms of prediction error and the degree of lag, the prediction effect of the LSTM model is slightly inferior to the GRU model [see Fig. 11(c)], and the ARIMA multistep prediction has the most severe lag, with the predicted values for the next few
Fig. 9. Analysis results of forecast errors of combined model and single model in advance of 1–6 months. (a) RMSE. (b) Max error. (c) Mean error. (d) Concordance index.

Fig. 10. Analysis results of forecast errors of different EMD algorithms with the GRU combination in advance of 1–6 months. (a) RMSE. (b) Max error. (c) Mean error. (d) Concordance index.
Fig. 11. Time series of SSTA for 1–6 months lead prediction during 2014–2020 by the hybrid EEMD-GRU model (a) and the single GRU (b), LSTM (c), and ARIMA (d). The black curve represents actual SSTA observations, other colors represent the results of the 1–6 months lead forecast.
months being essentially constant [see Fig. 11(d)]. Therefore, for the SSTA multistep forecasting, the single GRU, LSTM, and ARIMA models have serious lagging problems, which can be effectively mitigated by using a hybrid EMD and GRU model.

**F. Model Application**

To validate the hybrid model prediction effect in practical application, the EEMD-GRU model is applied to the Niño 3.4 regional SSTA study for multistep prediction of the Nino3.4 index (average of the Niño 3.4 regional SSTA). Fig. 12(a) and (b) shows the prediction results of the 15-year El Niño and 20-year La Niña events, respectively. The 12-month Nino3.4 index (light blue area) is used to forecast the change of the Nino3.4 index (red and blue areas) for the next 6 months. From the figure, it can be concluded that the last three steps have larger prediction errors, but the first three steps of the prediction are highly accurate and the trend of the Nino3.4 index is consistent with the real change trend. The average Nino3.4 index is close to 1.5 °C in the second half of 2015 and close to −1 °C in the end of 2020. According to the measures of La Niña and El Niño, it can be concluded from the prediction results that a strong El Niño will be formed in the second half of 2015 and a moderate La Niña will be formed from the end of 2020 to the beginning of 2021, both of which are consistent with the reality. Therefore, although the prediction errors of the later steps of the EEMD-GRU model are slightly large, the trend of the nino3.4 index can be predicted many months in advance, thus providing some reference for the analysis and forecasting of extreme weather.

**V. CONCLUSION**

In this article, a multistep-ahead prediction model based on hybrid EMD and GRU was proposed to predict SSTA data multiple months in advance. SSTA data for the Niño3.4 region from January 1980 to December 2020 were used as an example for the article. The main findings of this article are as follows:

1) The improved EMD algorithm was used to reduce the nonstationary of the SSTA. Four statistical evaluation metrics (MAE, MRE, RMSE, and CI) were used to assess the performance of different EMD algorithms for decomposing forecasts. It was concluded that the reconstruction error of the EEMD decomposition has a significant impact on the prediction, and using it as part of the prediction can effectively improve the prediction accuracy, and the prediction result is better than other EMD combination models.

2) The multistep forecasting strategies of recursive, direct, and multioutput prediction errors were compared, and it was concluded that the direct strategy has higher prediction accuracy for SSTA, but still suffers from the problem that the error increases with increasing lead time.

3) Comparing the forecasting accuracy of the hybrid EEMD-GRU model with that of the single GRU, LSTM, and ARIMA models, it was concluded the combined model has better forecasting effects and can better predict the SSTA development trend in advance, alleviating the problem of multistep-ahead forecasting lag of the traditional models.

4) The EEMD-GRU model was applied to the Niño3.4 regional SSTA prediction, and the results show that the forecasting results can better describe the regional SSTA change process.

In a follow-up study, how to overcome the increase in error with increasing forecast lead time is the focus. Due to the small number of SSTA data features, it may be difficult to further improve the accuracy of multistep prediction by using statistical methods alone. Future research can consider ocean dynamics and combine numerical methods with statistical methods to build multistep prediction models.
ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive and helpful suggestions.

REFERENCES

[1] S. L. Yao, J. J. Luo, G. Huang, and P. Wang, "Distinct global warming rates tied to multiple ocean surface temperature changes," Nat. Climate Change, vol. 7, no. 7, pp. 486–491, 2017, doi: 10.1038/nclimate3304.

[2] X. T. Zheng, S. P. Xie, L. H. Lv, and Z. Q. Zhou, "Intermodel uncertainty in ENSO amplitude change tied to Pacific Ocean warming pattern," J. Climate, vol. 29, no. 20, pp. 7265–7279, Oct. 2016, doi: 10.1175/JCLI-D-16-0083.1.

[3] S. W. Fang and J. Y. Yu, "Contrasting transition complexity between El Niño and La Niña: Observations and CMIP5/6 models," Geophysical Res. Lett., vol. 47, no. 16, Jul. 2020, Art. no. e2020GL089926, doi: 10.1029/2020GL089926.

[4] J. López-Parages, B. Rodríguez-Fonseca, D. Dommenget, and C. Frauen, "Eddies Influence on the Atlantic-European climate: A non-linear and non-stationary approach," Climate Dyn., vol. 47, no. 7, pp. 2071–2084, Dec. 2015, doi: 10.1007/s00382-015-2951-0.

[5] A. M. Chiogi, "Diagnosing and predicting ENSO SST development from moored buoy and scatterometer winds," J. Climate, vol. 32, no. 24, pp. 8755–8770, Dec. 2019, doi: 10.1175/JCLI-D-19-0183.1.

[6] T. L. Hao et al., "Research on SST forecasting method based on EOF-NAR neural network hybrid model," presented at Chin. Proc. Ocean Aca.-Int. Biennial Conf. Chin. Soc. Oceanogr., Sanya, Hainan, China, Oct. 2019, pp. 31–45.

[7] J. M. Wallace and D. S. Gutzler, "Teleconnections in the geopotential height field during the Northern Hemisphere winter," Monthly Weather Rev., vol. 109, no. 4, pp. 784–812, Apr. 1981, doi: 10.1175/1520-0493.

[8] I. D. Linos, M. Araujo, M. Moura, M. A. Silva, and E. L. Drogue, "Prediction of sea surface temperature in the tropical Atlantic by support vector machines," Comput. Statist. Data Anal., vol. 61, pp. 187–198, 2013.

[9] R. Rosmiati, S. Alihasi, B. Tjasyono, and T. R. Ramalis, "Development of ARIMA technique in determining the ocean climate prediction skills for pre-service teacher," J. Phys. Conf. Ser., vol. 1731, no. 1, 2021, Art. no. 012072.

[10] T. J. Laepple, "Five year ahead prediction of sea surface temperature in the tropical Atlantic: A comparison between IPCC climate models and simple statistical methods," 2007, arXiv preprint physics/0701165, doi: 10.48550/arXiv.physics/0701165.

[11] D. C. R. Novitasari et al., "Forecasting sea surface temperature in java sea using generalized regression neural networks," in Proc. Smart Trends Commun., Proc. SmartCom, Springer, Singapore, 2021, pp. 207–218.

[12] H. Cui, L. Zhu, J. Li, and L. Nie, "Scalable deep hashing for large-scale social image retrieval," IEEE Trans. Image Process., vol. 29, no. 24, Nov. 2021, pp. 7265–7279, Oct. 2016, doi: 10.1109/TIP.2016.2547541.

[13] J. Liu, T. Huang, J. Gao, F. Ye, and L. You, "Deep residual convolutional neural network combining dropout and transfer learning for ENSO forecasting," Geophysical Res. Lett., vol. 48, no. 24, Nov. 2021, doi: 10.1029/2021GL093531.

[14] L. Lin, Y. Jiang, H. Xiao, and Z. Zhou, "Crude oil price forecasting based on a novel hybrid long memory GARCH-M and wavelet analysis model," Ocean Eng., vol. 205, 2020, doi: 10.1016/j.oceaneng.2020.104502.

[15] S. Yang, D. Chen, H. G. Zhong, and X. Sun, "Prediction of sea surface temperature using long short-term memory neural networks," Remote Sens., vol. 12, no. 17, pp. 34–39, Jul. 2020, doi: 10.3390/rs12172697.

[16] M. S. Pravallika, S. Vasavi, and S. P. Vigneshwar, "Prediction of temperature anomaly in Indian Ocean based on autoregressive long short-term memory neural network," Neural Comput. Appl., vol. 34, no. 10, pp. 7537–7545, 2022, doi: 10.1007/s00521-021-06078-8.

[17] A. Wu, W. W. Hsieh, and B. Tang, "Neural network forecasts of the tropical Pacific sea surface temperatures," Neural Netw., vol. 19, no. 2, pp. 145–154, Mar. 2006, doi: 10.1016/neunet.2006.01.004.

[18] L. Wei, L. Guan, L. Qu, and D. Guo, "Prediction of sea surface temperature in the China seas based on long short-term memory neural networks," Remote Sens., vol. 12, no. 17, pp. 34–39, Jul. 2020, doi: 10.3390/rs12172697.
[39] S. B. Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa, “A review and comparison of strategies for multi-step ahead time series forecasting based on the NNS forecasting competition,” Expert Syst. Appl., vol. 39, no. 8, pp. 7076–7083, Jun. 2012, doi: 10.1016/j.eswa.2012.01.039.

[40] T. Niu, J. Wang, K. Zhang, and P. Du, “Multi-step-ahead wind speed forecasting based on optimal feature selection and a modified bat algorithm with the cognition strategy,” Renewable Energy, vol. 118, pp. 213–229, Apr. 2018, doi: 10.1016/j.renene.2017.10.075.

[41] M. Wu, C. Stefanakos, Z. Gao, and S. Haver, “Prediction of short-term wind and wave conditions for marine operations using a multi-step-ahead decomposition-ANFIS model and quantification of its uncertainty,” Ocean Eng., vol. 188, Sep. 2019, Art. no. 106300, doi: 10.1016/j.oceaneng.2019.106300.

[42] B. Huang et al., “NOAA extended reconstructed sea surface temperature (ERSST), version 5,” NOAA Nat. Centers Environ. Inf., vol. 30, pp. 8179–8205, 2017.

Xiaoyin Liu is currently working toward the bachelor’s degree in automation with the China University of Petroleum (East China), Qingdao, China. He is currently a student of automation with the China University of Petroleum (East China). His research interests include complex system analysis and control, optimal control, and computer application technology.

Ning Li is currently working toward the B.Sc. degree in data science and big data technology with the China University of Petroleum (East China), Qingdao, China. He is currently a student of Data Science and Big Data Technology with the China University of Petroleum (East China). His research interests include artificial intelligence, big data analysis, and computer application technology.

Jun Guo was born in Cangzhou City, China, in 2000. He is currently working toward the B.Eng. degree in automation with the China University of Petroleum (East China), Qingdao, China. His current research interests include pattern recognition, computer vision, and intelligent control.

Zhongyong Fan is currently working toward the B.Eng. degree in automation with the China University of Petroleum (East China), Qingdao, China. He is currently a student of automation with the China University of Petroleum (East China). His research interests include pattern recognition and intelligent control.

Xiaoping Lu received the Ph.D. degree in mechanical and electronic engineering from Zhejiang University, Hangzhou, China, in 2010. He is currently the Technical Director with the COSMO Industrial Intelligence Institute, Qingdao, China. His current research interests include industrial Internet and industrial intelligent products.

Weifong Liu (Senior Member, IEEE) received the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2007. He is currently a Full Professor with the College of Control Science and Engineering, China University of Petroleum (East China), Qingdao, China. His research interests include computer vision, pattern recognition, and machine learning.

Baodi Liu (Member, IEEE) received the Ph.D. degree in information and communication engineering with the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2013. He is currently an Associate Professor with the College of Control Science and Engineering, China University of Petroleum (East China), Qingdao, China. He was a Visiting Scholar with the University of California, Merced, from 2019 to 2020. His research interests include image processing, computer vision, and machine learning.