

Significant wave height forecasts integrating ensemble empirical mode decomposition with sequence-to-sequence model

Lina Wang^{1,2}, Yu Cao¹, Xilin Deng¹, Huitao Liu¹, Changming Dong^{2,3*}

¹ School of Artificial Intelligence (School of Future Technology), Nanjing University of Information Science & Technology, Nanjing 210044, China

² Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519080, China

³ School of Marine Sciences, Nanjing University of Information Science & Technology, Nanjing 210044, China

Received 10 June 2023; accepted 15 August 2023

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

Abstract

As wave height is an important parameter in marine climate measurement, its accurate prediction is crucial in ocean engineering. It also plays an important role in marine disaster early warning and ship design, etc. However, challenges in the large demand for computing resources and the improvement of accuracy are currently encountered. To resolve the above mentioned problems, sequence-to-sequence deep learning model (Seq-to-Seq) is applied to intelligently explore the internal law between the continuous wave height data output by the model, so as to realize fast and accurate predictions on wave height data. Simultaneously, ensemble empirical mode decomposition (EEMD) is adopted to reduce the non-stationarity of wave height data and solve the problem of modal aliasing caused by empirical mode decomposition (EMD), and then improves the prediction accuracy. A significant wave height forecast method integrating EEMD with the Seq-to-Seq model (EEMD-Seq-to-Seq) is proposed in this paper, and the prediction models under different time spans are established. Compared with the long short-term memory model, the novel method demonstrates increased continuity for long-term prediction and reduces prediction errors. The experiments of wave height prediction on four buoys show that the EEMD-Seq-to-Seq algorithm effectively improves the prediction accuracy in short-term (3-h, 6-h, 12-h and 24-h forecast horizon) and long-term (48-h and 72-h forecast horizon) predictions.

Key words: significant wave height, wave forecasting, ensemble empirical mode decomposition (EEMD), Seq-to-Seq, long short-term memory

Citation: Wang Lina, Cao Yu, Deng Xilin, Liu Huitao, Dong Changming. 2023. Significant wave height forecasts integrating ensemble empirical mode decomposition with sequence-to-sequence model. *Acta Oceanologica Sinica*, 42(10): 54–66, doi: 10.1007/s13131-023-2246-y

1 Introduction

The significant wave height (SWH) is the actual wave height calculated in a certain way. SWH is an important parameter used to measure the marine climate and plays a crucial role in marine disaster prediction (Ardhuin et al., 2019), marine engineering construction (Vanem, 2016) and ship design (Caires and Sterl, 2005). The accurate and quick prediction of the SWH has become a considerable problem in the development of modern ocean technology. Currently, the SWH forecasting can be divided into three categories: numerical wave, machine learning and deep learning models. Given appropriate SWH forecasting models, the accuracy of prediction depends on the forecast time span. The longer the prediction time is, the lower the accuracy of prediction is. With the rapid development of computer science and technology, more attention has been paid to the numerical wave models. The third-generation numerical wave prediction models (e.g., Wave Modeling, Wave Watch III Model, and Simulating Waves Nearshore Model) are the most widely used numerical wave models. The machine learning methods can fit complex nonlinear processes and solve complex nonlinear problems of the physical mechanism without prior knowledge of the system. Therefore, these methods are widely applied to SWH predic-

tion. One is the single prediction model, such as artificial neural network (Deo and Naidu, 1998), support vector machine (Mahjoobi and Mosabbeeb, 2009), M5' model tree (Etemad-Shahidi and Mahjoobi, 2009); and the other is the composite prediction model, such as hybrid empirical mode decomposition support vector regression model (Duan et al., 2016), a fuzzy KNN-based model (Nikoo et al., 2018). Here, traditional machine learning methods usually require manual feature engineering. If important features cannot be learned, then the prediction accuracy of the model will be significantly reduced. As a result, manually selecting feature is a laborious task. The development of computer technology in recent years realizes the possible application of complex models. Deep learning algorithms such as the simulating waves nearshore (SWAN)-long short-term memory (LSTM) (Fan et al., 2020), convolutional neural network (Yang et al., 2021), convolutional LSTM network (ConvLSTM) (Zhou et al., 2021a) and convolutional neural network-bidirectional LSTM-attention mechanism model (CNN-BiLSTM-attention) (Wang et al., 2022), can automatically learn data features and achieve successful results in the field of ocean prediction. However, the above methods do not fully mine the signal characteristics of the SWH. The original data sequence is decomposed into a set of in-

Foundation item: The Project Supported by Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) under contract No.SML2020SP007; the National Natural Science Foundation of China under contract Nos 42192562 and 62072249.

*Corresponding author, E-mail: cmdong@nuist.edu.cn

trinsic mode functions (IMFs) by signal decomposition technology to reduce the non-stationarity of wave height data, and then the model is constructed for prediction at each intrinsic mode function sub-sequence. [Oh and Suh \(2018\)](#) proposed a hybrid model combining the empirical orthogonal function and wavelet analysis with the neural network algorithm (EOFWNN) and predicted the wave height values of eight wave observation stations along the coast of the Sea of Japan. Compared with the wavelet and neural network hybrid models, EOFWNN performed effectively despite the decomposition level of wavelet analysis. [Zhou et al. \(2021b\)](#) proposed the forecasting algorithm using a joint empirical mode decomposition LSTM (EMD-LSTM) model, in which the EMD algorithm is applied to decompose the SWH, and the decomposed IMF sub-sequences are trained by the LSTM network. Effective results are achieved in the SWH from two buoys in the Atlantic Ocean, east of the Bahamas. [Raj and Brown \(2021\)](#) jointly adopted a hybrid Boruta random forecast ensemble EMD (EEMD) bidirectional LSTM algorithm to predict SWH over 24 h along the coastal areas of Queensland, Australia. Considering the prediction of different forecast windows, [Sutskever et al. \(2014\)](#) proposed a sequence-to-sequence deep learning model (Seq-to-Seq) based on encoder–decoder, in which one recurrent neural network (RNN) network encodes the input information and the other RNN network decodes the encoded information, which can demonstrate increased continuity for long-term prediction. The Seq-to-Seq has been applied to long-term prediction problems in different fields, such as ocean waves ([Pirhooshyaran and Snyder, 2020](#)), wind power ([Zhang et al., 2020](#)), power load ([Gong et al., 2019](#)) and semantic trajectories ([Karatzoglou et al., 2018](#)), and achieved high prediction accuracy.

The accuracy of SWH at the long forecast windows is low. Thus, SWH prediction integrating EEMD with Seq-to-Seq (EEMD-Seq-to-Seq) model is proposed. Noise characteristics are applied to EEMD, which can effectively suppress modal aliasing caused by EMD decomposition, and EEMD improves prediction accuracy whilst reducing the difficulty of modeling ([Wu and Huang, 2009](#)). The Seq-to-Seq module applies to long-term pre-

dition, which comprises the encoder and the decoder. The encoder and decoder adopt LSTM-based multi-layer network structures ([Graves, 2012](#)), and an attention module is added between them ([Vaswani et al., 2017](#)). The attention mechanism can avoid the information compression problem caused by the Seq-to-Seq model in source sequence encoding. Therefore, the decoder can review the entire source sequence at each step of decoding. Consequently, predicted performance of the Seq-to-Seq model is improved. An improved algorithm of the EEMD-Seq-to-Seq is proposed in this paper based on EMD-LSTM model. The proposed algorithm can decompose the original wave height signal and train the decomposed stationary mode signal to obtain different time-window prediction models. The algorithm is then applied to establish SWH forecasting models under multiple forecast windows, and the experimental results on different buoys verify the effectiveness of the method.

2 Materials and methods

2.1 Materials

2.1.1 Buoy data

Observations of significant wave height are acquired by National Data Buoy Center from four buoys deployed in the Atlantic Ocean, east of the Bahamas. These buoys are as follows: Station 41040 (14.542°N, 53.341°W), Station 41044 (21.582°N, 58.630°W), Station 41046 (23.822°N, 68.384°W) and Station 41047 (27.514°N, 71.494°W), as shown in [Fig. 1](#). The acquired data range from 2019 to 2020 and are provided at 1-h resolution.

2.1.2 Data preprocessing

Uncertain factors in the process of data acquisition lead to data loss. EEMD algorithm needs a data stream without missing values; thus, spline interpolation is performed to supplement the missing values from the time series before entering the EEMD module. All buoy statistics, relate to geographical positions, water depth and the number of observations before and after interpolation, are shown in [Table 1](#).

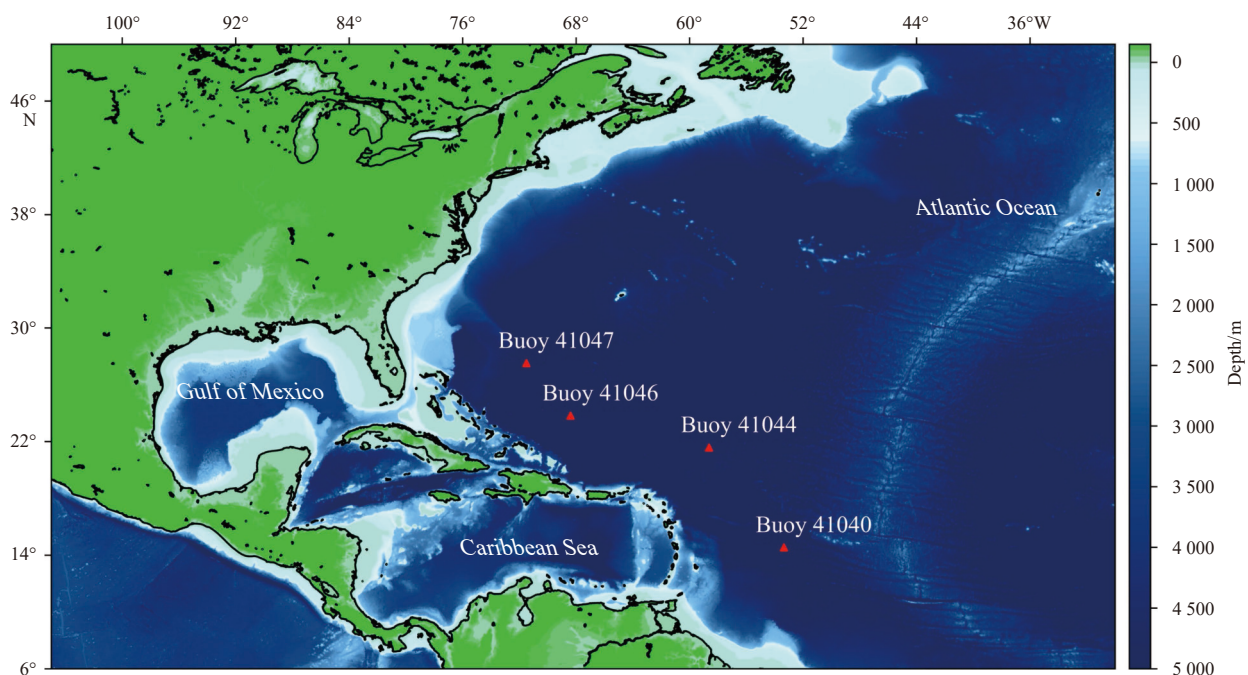


Fig. 1. Location of Buoys 41040, 41044, 41046 and 41047 (data acquired from National Data Buoy Center).

Table 1. Data statistics of the selected buoys from January 1, 2019 to December 31, 2020 (data acquired from National Data Buoy Center)

Buoy ID	Latitude	Longitude	Water depth/m	Number of observations (before interpolation)	Number of observations (after interpolation)
41040	14.542°N	53.341°W	5 159	17 273	17 520
41044	21.582°N	58.630°W	5 419	17 280	17 520
41046	23.822°N	68.384°W	5 549	16 924	17 520
41047	27.514°N	71.494°W	5 321	17 234	17 520

2.2 EEMD algorithm

EMD is an adaptive time-frequency data analysis method which can decompose complex signals into a series of IMFs according to the signal characteristics. The decomposed IMF contains local features of the original signal at different time scales. Applied in marine forecasting, EMD converts a non-stationary wave height sequence into wave height components with a definite pattern, which are predictable and can significantly improve prediction accuracy. The algorithm steps are as follows.

(1) Identify the local extreme values for the original signal $x(t)$. Interpolate the local extreme values by a cubic spline function to obtain the upper envelope $e_{\max}(t)$ and the lower envelope $e_{\min}(t)$.

(2) The mean of the upper and lower envelopes $m(t)$ is:

$$m(t) = \frac{1}{2}(e_{\max}(t) + e_{\min}(t)). \quad (1)$$

(3) Subtract the mean value $m(t)$ from $x(t)$ to obtain an IMF candidate $h(t)$:

$$h(t) = x(t) - m(t). \quad (2)$$

(4) Judge whether $h(t)$ meets the IMF conditions by the value of standard deviation (SD). The SD is calculated as follows:

$$SD = \sum_{t=1}^T \frac{|h_{j-1}(t) - h_j(t)|^2}{h_{j-1}^2(t)}, \quad (3)$$

where T is the time vector; $h_{j-1}(t)$ and $h_j(t)$ represent the signals after shifting $j-1$ and j times, respectively. If the value of SD is between 0.2 and 0.3, an IMF component $h_j(t)$ is selected. Namely, $h_j(t)$ is defined as $C_j(t)$.

(5) To obtain a new signal $r(t)$, subtract $C_j(t)$ from $x(t)$:

$$r(t) = x(t) - C_j(t). \quad (4)$$

(6) Repeat Steps (1)–(5) until $r_n(t)$ cannot be further decomposed into IMFS. The residual of the original signal is given by $r_n(t)$. The original signal $x(t)$ is finally decomposed into a series of IMFs and a residual.

$$x(t) = \sum_{j=1}^n C_j(t) + r_n(t), \quad (5)$$

where n is the number of IMFs, and $r_n(t)$ is the residual of the signal $x(t)$.

Modal aliasing exists in the conventional EMD method. Modal aliasing reduces the decomposition accuracy when the signal distribution is uneven. Huang et al. (1998) proposed an improved algorithm, that is, EEMD, to solve modal aliasing. De-

pending on the uniform frequency distribution characteristics of the white noise spectrum, the problem of modal aliasing is solved by adding white noise in the EEMD method to separate a signal to its adapted reference scale automatically. Due to the nature of zero mean noise, white noise is added to the calculation procedure of the signal data (Yang and Wang, 2021). The final IMF signal in EEMD decomposition is the mean value of IMFs obtained in the noise test. The specific decomposition process is as follows: (1) input the preprocessed wave height signal, initialize the number of tests M and set $m=1$; (2) in the m th test, white noise sequence $n_m(t)$ is added to the original signal x_m , $x_m = x(t) + n_m(t)$; (3) the newly obtained signal $x_m(t)$ is decomposed by the EMD model, and the obtained IMF is denoted as $C_{j,m}$; (4) repeat Steps (2) and (3). Each test uses new white noise sequences until the number of tests M is achieved. The integrated IMF signals are estimated by averaging IMFs obtained with trials of new noises, as shown in Eq. (6).

$$IMF_j(t) = \frac{1}{M} \sum_{m=1}^M C_{j,m}, \quad j = 1, 2, \dots, n. \quad (6)$$

The white noise sequences added in the tests cancel each other out in the corresponding integrated IMF due to the nature of zero mean noise (Bokde et al., 2020). The addition of a white noise signal provides signal continuity at different scales, changes the characteristics of the signal extreme point and promotes anti-aliasing decomposition. The mean value of the obtained IMFs is taken as the final result, and the flow chart of the EEMD algorithm is shown in Fig. 2.

2.3 Seq-to-Seq prediction model

2.3.1 Model structure

The sequence-to-sequence (Seq-to-Seq) prediction model can map input sequences to output sequences with variable lengths. This model has been widely applied to machine translation, sequence prediction and other fields, and mainly comprises an encoder and a decoder. The encoder reads the input sequence to generate a vector with a fixed dimension as the hidden state of the input sequence, whilst the decoder decodes the state into a predicted result. However, the encoder will lose valuable information when the input sequence is excessively long (Ye et al., 2022). Bahdanau et al. (2014) proposed a Seq-to-Seq model with an attention mechanism and added an attention module into the model to solve this problem. The attention mechanism enables the decoder to select a subset of the input sequence adaptively for subsequent prediction by weighting the encoded input source data. Figure 3 shows the structure of the Seq-to-Seq model with the attention module and gives the detailed design of the model framework.

The prediction model includes three parts: encoder layer, decoder layer and attention layer, as shown in Fig. 3. The encoder learns its hidden features from the input sequence x_{t-l+1} ,

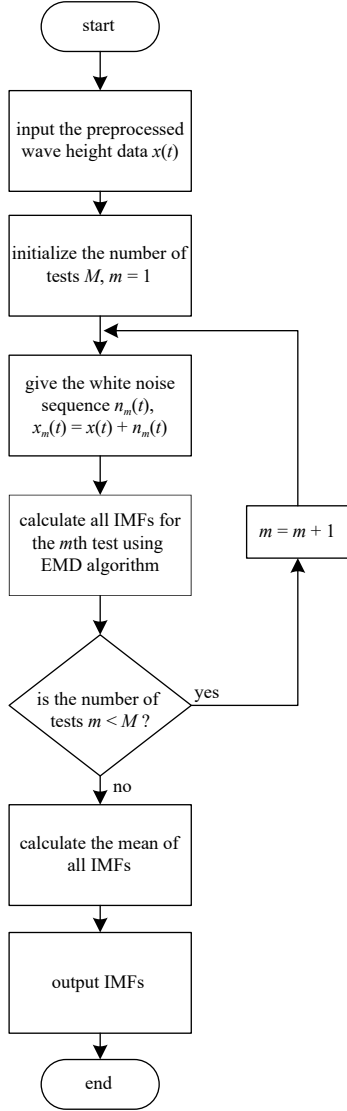


Fig. 2. The flow chart of the ensemble empirical mode decomposition algorithm. m is the mean of the upper and lower envelope; IMF: intrinsic mode function.

x_{t-1+2}, \dots, x_t through LSTM network. And construct historical sequence through context generation vector and attention vector, that is, encoded attention vector CA_t . The decoder is based on the encoded attention vector and decode through LSTM networks. Finally, the prediction target is attained.

2.3.2 LSTM neural network

LSTM network has been proved to be the most effective model for learning dependence characteristics of time series data. LSTM is better at processing sequential signal data than the basic RNN by adding forgetting gate, input gate, and output gate, i.e., deciding how much information should be retained by a sigmoid function. In the structure of LSTM neuron, it is first decided by a forgetting gate whether old information from the previous time step should be lost or not. Then it is determined which new information will be stored in the cell state, and through an input gate, a candidate vector activated by tanh and the state vector of the previous layer determine the current cell state. The final output cell state and the vector pass through the output gate to determine. The structure of LSTM neuron is shown in Fig. 4.

The various functions and gates within LSTM are calculated as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (7)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (8)$$

$$C_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \quad (9)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t, \quad (10)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (11)$$

$$h_t = o_t \circ \tanh(C_t), \quad (12)$$

where $\sigma(\cdot)$ and $\tanh(\cdot)$ are sigmoid and tanh activation functions, respectively; x_t is the input for the current time step; f_t is the output of the forget gate; i_t is the output of the input gate, which determines the acceptance of the candidate state \tilde{C}_t ; o_t is the output of the output gate; C_t is the state vector of the current time step, and obtained by the dot product; h_{t-1} is the hidden state vector of the output at time $t-1$; W_f , W_i , W_C , and W_o are the weights; b_f , b_i , b_C , and b_o are the bias; \circ is the Hadamard product operator. The final output result h_t is obtained by the Hadamard product of o_t and $\tanh(C_t)$.

2.3.3 Attention mechanism

The attention mechanism simulates the human brain focusing on a particular area at a particular time, selectively acquiring more useful information, and ignoring useless information. It can strengthen the influence of key information and enhance the accuracy of model judgment by assigning different weight values to the hidden layer units of a neural network. Adopting the attention mechanism enables the model to learn reasonable vector representations and makes the key information dominate the prediction process, thereby improving the prediction accuracy of the model. h_t and C_t are respectively the hidden state and coding context vector at time t . CA_i is the encoded attention vector at time i , which is calculated by the weighted average value of the hidden state vector h_i . f is a nonlinear mapping function (Keneshloo et al., 2020).

$$h_t, C_t = f(x_t, h_{t-1}, C_{t-1}), \quad (13)$$

$$CA_i = \sum_{j=1}^l a_{ij} h_j. \quad (14)$$

In the attention part, a small neural network is added and activation function of the output layer is softmax, which is a normalized exponential function. The weight corresponding to the hidden state of each encoder is calculated as follows.

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})}, \quad (15)$$

where e_{ij} is a function of correlation calculation, which mainly in-

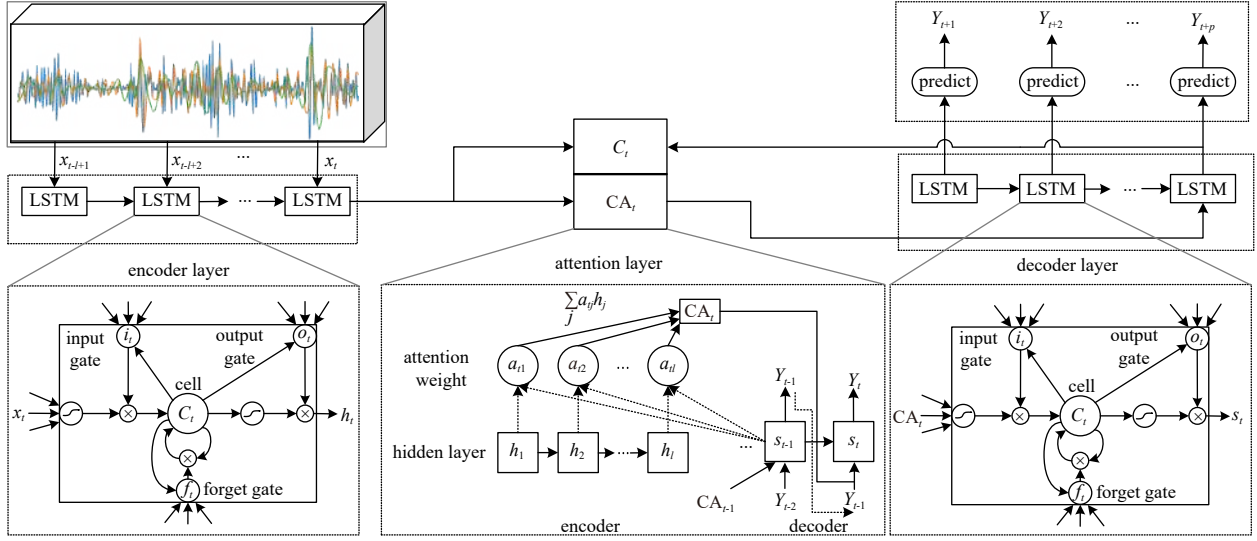


Fig. 3. The structure of the sequence-to-sequence prediction model with attention mechanism. LSTM: long short-term memory; CA: encoded attention vector. The meanings of symbols refer to formula in the text.

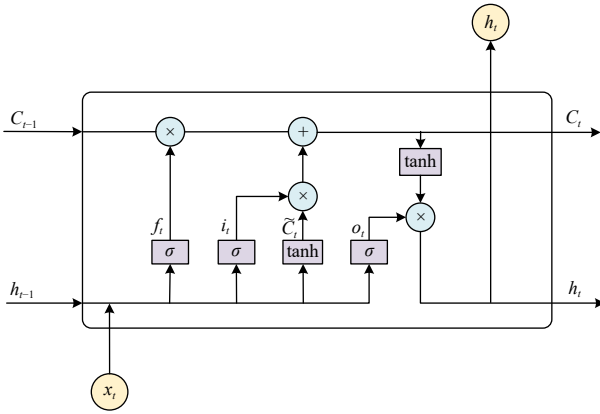


Fig. 4. The structure of long short-term memory neuron.

cludes dot product, bilinear product, connection and multi-layer perception (MLP). Herein, MLP is selected. It mainly calculates the hidden state of the decoder at the previous time, and the hidden state of the encoder at the current time by matrix multiplication with their respective parameters.

$$e_{ij} = v \times \tanh(w_1 s_{i-1} + w_2 h_j), \quad (16)$$

where v , w_1 and w_2 are the trainable parameters; s_{i-1} is the hidden state of the decoder at time $i-1$; h_j is the hidden state of the encoder at time j .

The decoder with LSTM network is based on the output of the target sequence, Y_{t-1} , the hidden state vector of the decoder at time $t-1$, s_{t-1} , and the encoded attention vector, CA_t , for decoding computing. g is a nonlinear mapping function. The hidden state vector of the decoder s_t at time t is calculated by Eq. (17).

$$s_t = g(Y_{t-1}, s_{t-1}, CA_t). \quad (17)$$

Finally, output the predicted result Y_t through a fully connected network layer.

$$Y_t = \sigma(s_t, CA_t). \quad (18)$$

3 EEMD-Seq-to-Seq model

3.1 Model structure

The Seq-to-Seq SWH prediction model based on EEMD decomposition is shown in Fig. 5. Firstly, the signal data are divided into multiple stationary sub-sequences (multiple IMFs and the residual) through the EEMD module, and then the decomposed signal at each IMF and the residual is sent to the Seq-to-Seq prediction module, and then the model is obtained through training. The encoder and decoder in the Seq-to-Seq module respectively adopt multi-layer LSTM network structure, and an attention module is added between them. In the decoding stage, the attention mechanism assigns different weights to each hidden vector of the source sequence. Therefore, the decoder can obtain the important information of the source sequence to produce im-

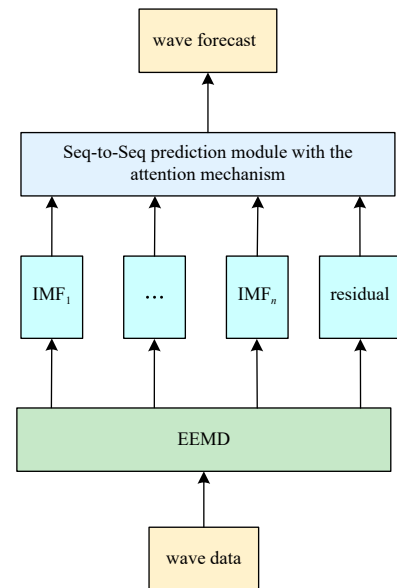


Fig. 5. The structure of the ensemble empirical mode decomposition sequence-to-sequence (EEMD-Seq-to-Seq) prediction model. IMFs: intrinsic mode functions.

proved decoding results. The prediction values at each IMF and the residual are summed to obtain the final prediction results. The time step was set to six as the maximum allowed time step as established by Fan et al. (2020), and the SWHs of 3-h, 6-h, 12-h, 24-h, 48-h and 72-h forecast windows are taken as the output. The training set is trained by the EEMD-Seq-to-Seq model, and the prediction performance is verified in the test set.

3.2 Algorithm flow

Firstly, the wave data are divided into training and test sets, and then are preprocessed. Next, the normalized signal data in training and test sets are separately divided into multiple stationary IMFs and the residual by the EEMD algorithm. In the training phase, training set is applied for modeling, and the decomposed signals at each IMF and the residual are organized as the training and the predicted data separately. Then, the training data at each IMF and the residual are sent to the encoder, and the encoder reads the input sequence to generate a vector with a fixed dimension as the hidden state of the input sequence, whilst the decoder decodes the state into a predicted result. Furthermore, the hidden states are assigned different weights combined with an attention mechanism through the encoder coding for hidden state vector. In the following decoding phase, the hidden state vector is decoded, and SWHs of different forecast windows are outputted. During the training process, the dropout layer is used to discard characteristics randomly to improve the robustness of the model, and the predicted data at each IMF and the residual are used to calculate the value of loss function for obtaining the optimal parameters of the prediction model. So the Seq-to-Seq prediction model at each IMF and the residual are obtained. In the test phase, the test set is applied for testing the prediction performance, and the decomposed signals at each IMF and the

residual are organized as the test data separately. The obtained predicted values at each IMF and the residual in the test set are outputted, and then add them to get the final forecast value. Finally, the error analysis between the predictive and observation values is provided. The flowchart of the model is shown in Fig. 6.

4 Experimental analysis

4.1 Experimental setup

The SWH data of four Buoys 41040, 41044, 41046 and 41047 from January 1, 2019 to December 31, 2020 are selected as the input dataset. The experimental environment is Windows 11 operating system, Intel Core i7-11800H CPU, 16 GB memory and NVIDIA RTX 3060 graphics card. Python 3.9 is selected as the development language, and the experiment is run on the basis of the Tensorflow-GPU 2.8 framework. LSTM, a forecasting algorithm using EMD-LSTM, EMD-Seq-to-Seq and EEMD-Seq-to-Seq are used to predict the SWH. The number of epochs at the stage of model training is set to 50.

The buoy data in 2019 are selected as the training set and the buoy data in 2020 as the test set. The training set is normalized before model training, as shown below, to improve the model performance.

$$X_N = \frac{X - \min(X)}{\max(X) - \min(X)}, \quad (19)$$

where X_N is the standardized data; X is the observed training data; $\min(X)$ is the minimum of the training data and $\max(X)$ is the maximum of the training data. And the test set is also needed to normalize with the same rule. Normalized data at Buoy 41044 are decomposed into 13 IMFs through EEMD, including 12 IMFs and a residual, and normalized data at the three other buoys are

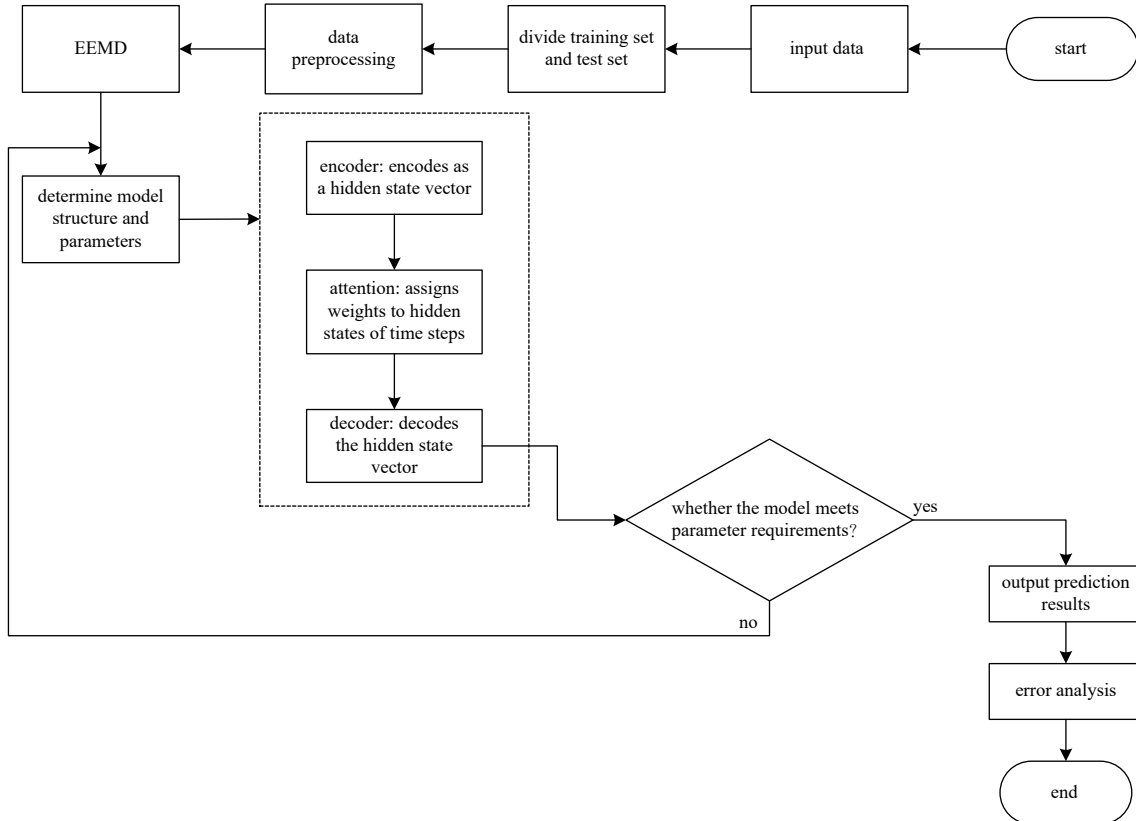


Fig. 6. The flowchart of ensemble empirical mode decomposition sequence-to-sequence (EEMD-Seq-to-Seq) prediction model.

decomposed into 14 IMFs through EEMD, including 13 IMFs and a residual. Compared with EMD decomposition, the EEMD algorithm solves the problem of modal aliasing problem due to EMD decomposition, and the decomposed stationary signal is suitable for prediction.

4.2 Loss function, optimizer and evaluation index

The loss function of the model is the mean square error (MSE) as presented by Eq. (20).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2, \quad (20)$$

where n is the total number of cases, x_i is the i th observed SWH and y_i is the i th predicted SWH.

The gradient of the loss function is optimized by the adaptive moment estimation (Adam) algorithm, which uses the adjustable attenuation of the adjustable learning rate to realize the robust convergence of the optimization through back propagation calculation.

The different models are evaluated by using mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and Pearson correlation coefficient (R) in the test set to assess the error and deviation between the prediction and the observed values and measure the linear correlation between these values. The specific expressions are as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i|, \quad (21)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}, \quad (22)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| \times 100\%, \quad (23)$$

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \times \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (24)$$

where x_i and \bar{x} respectively refer to the observed and mean values of SWH; y_i and \bar{y} are respectively the predicted and corresponding mean values of SWH; and n is the number of test cases. The prediction error increases with RMSE, MAE, and MAPE. R represents the correlation between the predicted and observed values.

4.3 Experimental results

The prediction advantages of integrating EEMD algorithm and the Seq-to-Seq model are compared in this section. LSTM, EMD-LSTM, EMD-Seq-to-Seq and EEMD-Seq-to-Seq models are adopted to predict the SWHs of four buoys. Tables 2–5 respectively demonstrate the prediction performances of different models for Buoys 41040, 41044, 41046 and 41047 at different forecast windows.

The tables below reveal that the evaluation indices of the EEMD-Seq-to-Seq model are superior to those of other algorithms. The advantages of different algorithms are as follows. (1) The EMD decomposition algorithm reduces the non-stationarity of data, and prediction performance of the EMD-LSTM is significantly improved compared with the LSTM algorithm; (2) the interconnection of the output sequence is adopted in the Seq-to-Seq model, which is superior to the LSTM network at a long-time-window forecast, and prediction performance of the EMD-Seq-to-Seq is further improved compared with the EMD-

Table 2. Comparisons of error statistics among four algorithms at the 3-h, 6-h, 12-h, 24-h, 48-h and 72-h forecast windows for Buoy 41040

Time span	LSTM				EMD-LSTM				EMD-Seq-to-Seq				EEMD-Seq-to-Seq			
	RMSE/m	MAE/m	MAPE/%	R	RMSE/m	MAE/m	MAPE/%	R	RMSE/m	MAE/m	MAPE/%	R	RMSE/m	MAE/m	MAPE/%	R
3 h	0.17	0.12	6.55	0.95	0.08	0.06	3.25	0.98	0.08	0.06	3.25	0.99	0.08	0.06	3.18	0.99
6 h	0.22	0.15	8.12	0.92	0.10	0.07	3.9224	0.98	0.10	0.07	3.79	0.98	0.09	0.06	3.33	0.99
12 h	0.29	0.21	11.08	0.84	0.14	0.10	5.32	0.97	0.14	0.10	5.18	0.93	0.11	0.08	4.32	0.98
24 h	0.39	0.28	14.93	0.69	0.21	0.15	7.93	0.92	0.20	0.14	7.45	0.93	0.16	0.12	6.04	0.96
48 h	0.48	0.34	18.43	0.47	0.31	0.21	11.50	0.83	0.31	0.22	11.66	0.83	0.27	0.18	9.32	0.87
72 h	0.51	0.37	20.02	0.36	0.38	0.26	15.24	0.74	0.38	0.29	15.28	0.70	0.38	0.29	14.89	0.70

Note: RMSE: root mean square error; MAE: mean absolute error; MAPE: mean absolute percentage error; R : Pearson correlation coefficient; LSTM: long short-term memory; EEMD: ensemble empirical mode decomposition; Seq-to-Seq: sequence-to-sequence deep learning model.

Table 3. Comparisons of error statistics among four algorithms at the 3-h, 6-h, 12-h, 24-h, 48-h and 72-h forecast windows for Buoy 41044

Time span	LSTM				EMD-LSTM				EMD-Seq-to-Seq				EEMD-Seq-to-Seq			
	RMSE/m	MAE/m	MAPE/%	R	RMSE/m	MAE/m	MAPE/%	R	RMSE/m	MAE/m	MAPE/%	R	RMSE/m	MAE/m	MAPE/%	R
3 h	0.21	0.13	7.25	0.95	0.11	0.07	3.59	0.99	0.11	0.08	4.24	0.99	0.08	0.05	2.93	0.99
6 h	0.27	0.17	9.23	0.92	0.14	0.08	4.42	0.98	0.14	0.09	4.98	0.98	0.09	0.06	3.63	0.99
12 h	0.38	0.24	13.22	0.82	0.21	0.12	6.33	0.97	0.20	0.13	7.08	0.96	0.13	0.09	4.99	0.98
24 h	0.54	0.34	18.90	0.52	0.33	0.20	10.95	0.88	0.33	0.20	14.61	0.89	0.21	0.14	7.44	0.91
48 h	0.65	0.42	23.93	0.31	0.51	0.31	18.32	0.72	0.47	0.30	15.16	0.75	0.32	0.21	10.85	0.88
72 h	0.68	0.44	25.73	0.16	0.53	0.32	18.79	0.72	0.51	0.33	17.14	0.73	0.48	0.32	16.87	0.72

Note: RMSE: root mean square error; MAE: mean absolute error; MAPE: mean absolute percentage error; R : Pearson correlation coefficient; LSTM: long short-term memory; EEMD: ensemble empirical mode decomposition; Seq-to-Seq: sequence-to-sequence deep learning model.

Table 4. Comparisons of error statistics among four algorithms at the 3-h, 6-h, 12-h, 24-h, 48-h and 72-h forecast windows for Buoy 41046

Time span	LSTM				EMD-LSTM				EMD-Seq-to-Seq				EEMD-Seq-to-Seq			
	RMSE/m	MAE/m	MAPE/%	<i>R</i>	RMSE/m	MAE/m	MAPE/%	<i>R</i>	RMSE/m	MAE/m	MAPE/%	<i>R</i>	RMSE/m	MAE/m	MAPE/%	<i>R</i>
3 h	0.23	0.15	8.59	0.95	0.11	0.07	4.03	0.99	0.12	0.08	4.61	0.99	0.08	0.06	3.13	0.99
6 h	0.29	0.19	11.12	0.91	0.13	0.09	5.09	0.98	0.13	0.09	5.13	0.98	0.09	0.06	3.44	0.99
12 h	0.41	0.26	15.92	0.83	0.19	0.13	7.31	0.96	0.18	0.12	6.68	0.97	0.12	0.09	4.81	0.99
24 h	0.55	0.37	22.64	0.66	0.30	0.20	11.38	0.91	0.27	0.18	10.42	0.93	0.21	0.15	8.09	0.96
48 h	0.67	0.46	28.75	0.41	0.42	0.31	16.84	0.83	0.38	0.27	15.32	0.85	0.32	0.22	12.35	0.90
72 h	0.71	0.49	30.60	0.30	0.47	0.34	18.48	0.77	0.46	0.33	18.66	0.77	0.42	0.30	16.54	0.83

Note: RMSE: root mean square error; MAE: mean absolute error; MAPE: mean absolute percentage error; *R*: Pearson correlation coefficient; LSTM: long short-term memory; EEMD: ensemble empirical mode decomposition; Seq-to-Seq: sequence-to-sequence deep learning model.

Table 5. Comparisons of error statistics among four algorithms at the 3-h, 6-h, 12-h, 24-h, 48-h and 72-h forecast windows for Buoy 41047

Time span	LSTM				EMD-LSTM				EMD-Seq-to-Seq				EEMD-Seq-to-Seq			
	RMSE/m	MAE/m	MAPE/%	<i>R</i>	RMSE/m	MAE/m	MAPE/%	<i>R</i>	RMSE/m	MAE/m	MAPE/%	<i>R</i>	RMSE/m	MAE/m	MAPE/%	<i>R</i>
3 h	0.25	0.16	9.39	0.96	0.13	0.07	3.97	0.99	0.11	0.07	4.12	0.99	0.10	0.06	3.91	0.99
6 h	0.33	0.21	12.43	0.93	0.13	0.09	4.95	0.98	0.12	0.08	4.74	0.99	0.11	0.07	4.56	0.99
12 h	0.45	0.30	18.13	0.85	0.21	0.13	7.35	0.97	0.19	0.12	6.90	0.97	0.15	0.11	6.16	0.98
24 h	0.63	0.43	26.43	0.68	0.39	0.26	13.01	0.91	0.37	0.24	12.53	0.91	0.25	0.17	10.03	0.96
48 h	0.79	0.55	34.01	0.40	0.58	0.38	20.85	0.79	0.55	0.38	22.03	0.78	0.39	0.27	16.81	0.89
72 h	0.83	0.59	36.52	0.25	0.62	0.43	23.76	0.71	0.60	0.42	24.95	0.72	0.49	0.36	19.85	0.82

Note: RMSE: root mean square error; MAE: mean absolute error; MAPE: mean absolute percentage error; *R*: Pearson correlation coefficient; LSTM: long short-term memory; EEMD: ensemble empirical mode decomposition; Seq-to-Seq: sequence-to-sequence deep learning model.

LSTM algorithm; (3) compared with EMD, the solution of the EEMD algorithm to the EMD decomposition-induced modal aliasing has superior performance, and prediction performance of the EEMD-Seq-to-Seq is further improved compared with the EMD-Seq-to-Seq algorithm. Prediction indices calculated by the EEMD-Seq-to-Seq algorithm for four different buoys are close over the 3-h, 6-h, 12-h, 24-h, 48-h and 72-h forecast window, and obtain the best prediction performance. The orders of prediction performance of EMD-Seq-to-Seq and EMD-LSTM algorithms are 2nd and 3rd, respectively, and the LSTM algorithm performs worst. At the 3-h forecast window, the predicted RMSE for four buoys is most in range of 0.08–0.10 m, 0.11–0.12 m, 0.11–0.13 m and 0.17–0.25 m in the EEMD-Seq-to-Seq, EMD-Seq-to-Seq, EMD-LSTM and LSTM models, respectively. Similarly, MAE in the EEMD-Seq-to-Seq model is 0.06 m and the MAPE is less than or equal to 3.91% at all buoys, which is better than the forecast indices of other algorithms. This finding indicates that the EEMD-Seq-to-Seq algorithm performs best. As the forecast window increases, RMSE, MAE and MAPE in different algorithms increase whilst *R* decreases. It shows that the prediction performance is deteriorating. At the 72-h forecast window, the EEMD-Seq-to-Seq model achieved the lowest RMSE, MAE and MAPE among all the algorithms for the four buoys and the highest correlation coefficient. Thus, the EEMD-Seq-to-Seq algorithm achieves the optimum predictive value.

The predicted SWHs of different models from September 10, 2020 to October 10, 2020 are shown in Figs 7–10. Figures 7–10 respectively show the SWH forecasts of Buoys 41040, 41044, 41046 and 41047 at 3-h, 6-h, 12-h, and 24-h windows.

Figures 7–10 reveal that predicted values in all the models in the area where waves change gently (as shown in Fig. 7, from September 20 to September 30, 2020) are close to the observation values. However, as the prediction interval is increased, the deviation of SWH obtained by LSTM algorithm occurs when wave height changes significantly, as shown in Fig. 7c and d, respectively. The EMD-Seq-to-Seq algorithm performs better than

LSTM and EMD-LSTM for predicted wave heights close to the observed values in most cases. The EMD-Seq-to-Seq and EEMD-Seq-to-Seq algorithms obtain similar prediction results at the four buoys. However, the stability of the EMD-Seq-to-Seq algorithm is worse than that of the EEMD-Seq-to-Seq algorithm, especially for the prediction of a high SWH value that rapidly changes. Figure 7d shows that the predicted value of EMD-Seq-to-Seq near September 17, 2020 has a considerable deviation from the observation values. However, the EEMD-Seq-to-Seq algorithm is substantially accurate for SWH forecasts on high-value points. The deviation appears with the increase of forecast windows when the EMD-LSTM algorithm is used to predict the high value of SWH, further demonstrating lower prediction accuracy. The LSTM algorithm is the worst because different levels of deviation between the predicted and observed values occur at all forecast windows. The predicted values obtained by the EEMD-Seq-to-Seq model for all buoys are optimal among all the algorithms mentioned.

The interconnection of the output sequence is adopted in the Seq-to-Seq model, which is superior to the LSTM network at a long-time forecast window. EMD-LSTM and EMD-Seq-to-Seq models are selected to compare the LSTM and Seq-to-Seq models and analyze the advantages of the Seq-to-Seq model. The Seq-to-Seq model provides the dependence of output sequence data on each other through the iterative output of a decoder, whilst the LSTM model does not capture the dependence between the output information. Therefore, the predicted error of the Seq-to-Seq model is smaller. The histograms of prediction errors of the EMD-LSTM and EMD-Seq-to-Seq models at Buoy 41047 are plotted in Fig. 11. This figure reveals in the test set, the total frequency of the occurrence in an error of a particular magnitude between the predicted value from the EMD-LSTM and EMD-Seq-to-Seq models and the observed value. The numbers of test data in all error magnitudes conform to a normal distribution, and the vast majority of test data are distributed at around the 0 m mark. As the absolute value of error increases, the num-

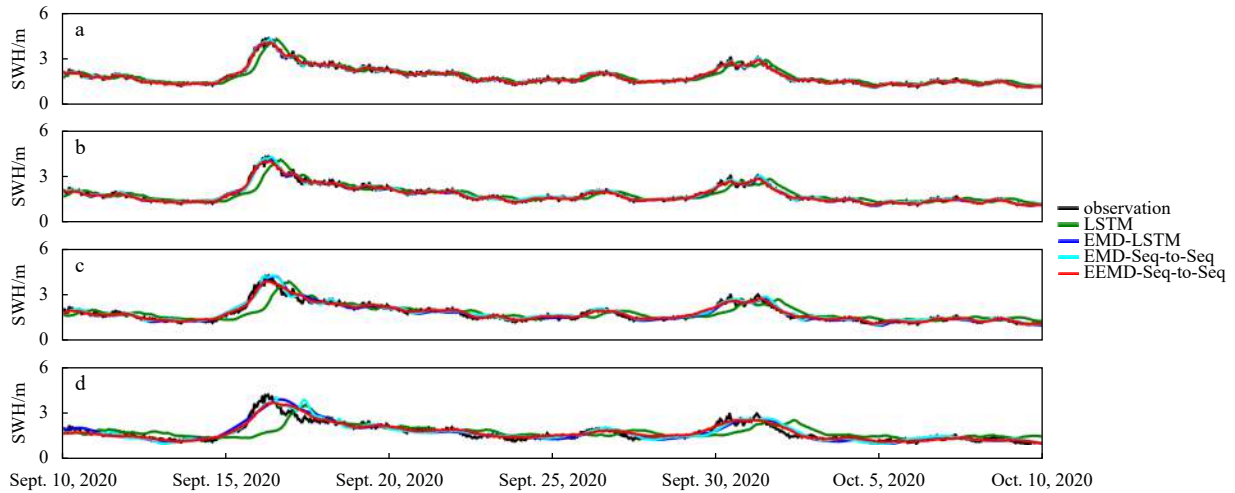


Fig. 7. Comparison of significant wave height (SWH) forecasts of different models for Buoy 41040 at the 3-h (a), 6-h (b), 12-h (c) and 24-h (d) windows. LSTM: long short-term memory; EEMD-Seq-to-Seq: ensemble empirical mode decomposition with the sequence-to-sequence model.

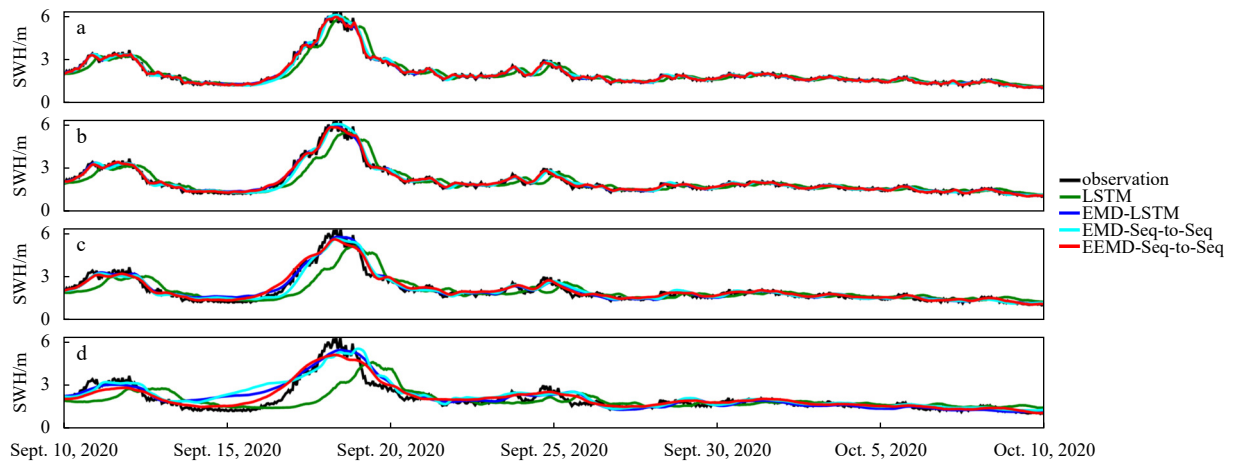


Fig. 8. Comparison of significant wave height (SWH) forecasts of different models for Buoy 41044 at the 3-h (a), 6-h (b), 12-h (c) and 24-h (d) windows. LSTM: long short-term memory; EEMD-Seq-to-Seq: ensemble empirical mode decomposition with the sequence-to-sequence model.

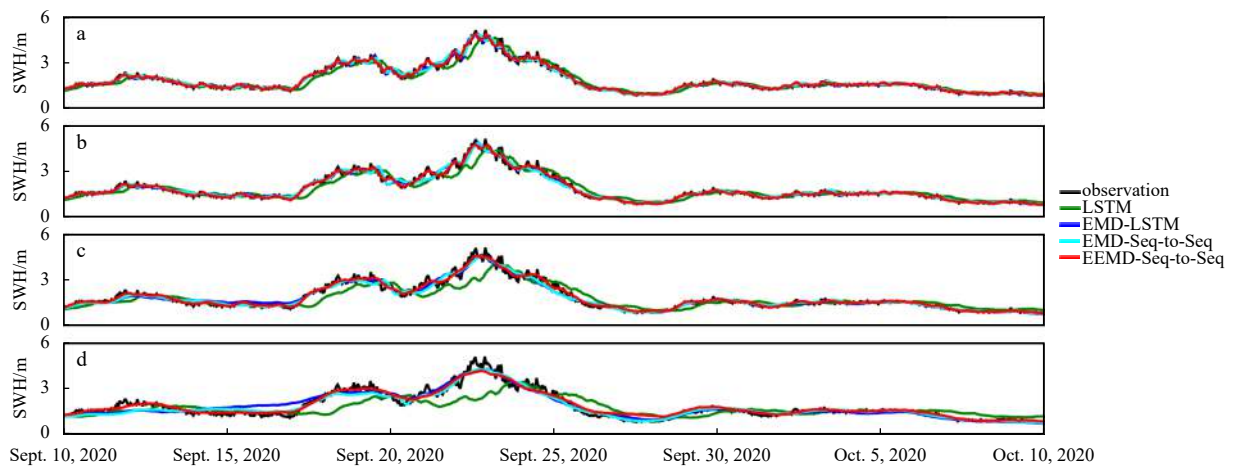


Fig. 9. Comparison of significant wave height (SWH) forecasts of different models for Buoy 41046 at the 3-h (a), 6-h (b), 12-h (c) and 24-h (d) windows. LSTM: long short-term memory; EEMD-Seq-to-Seq: ensemble empirical mode decomposition with the sequence-to-sequence model.

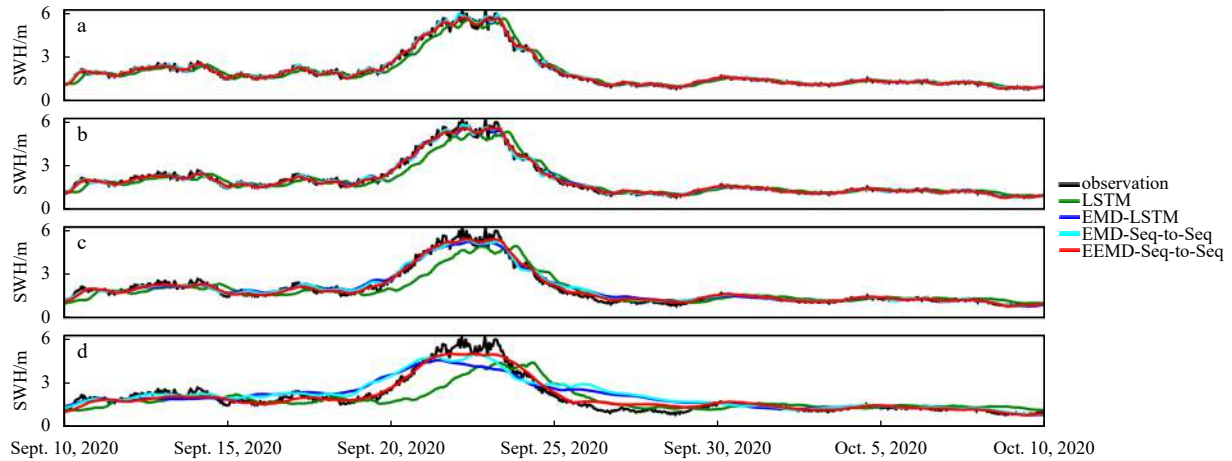


Fig. 10. Comparison of significant wave height (SWH) forecasts of different models for Buoy 41047 at the 3-h (a), 6-h (b), 12-h (c) and 24-h (d) windows. LSTM: long short-term memory; EEMD-Seq-to-Seq: ensemble empirical mode decomposition with the sequence-to-sequence model.

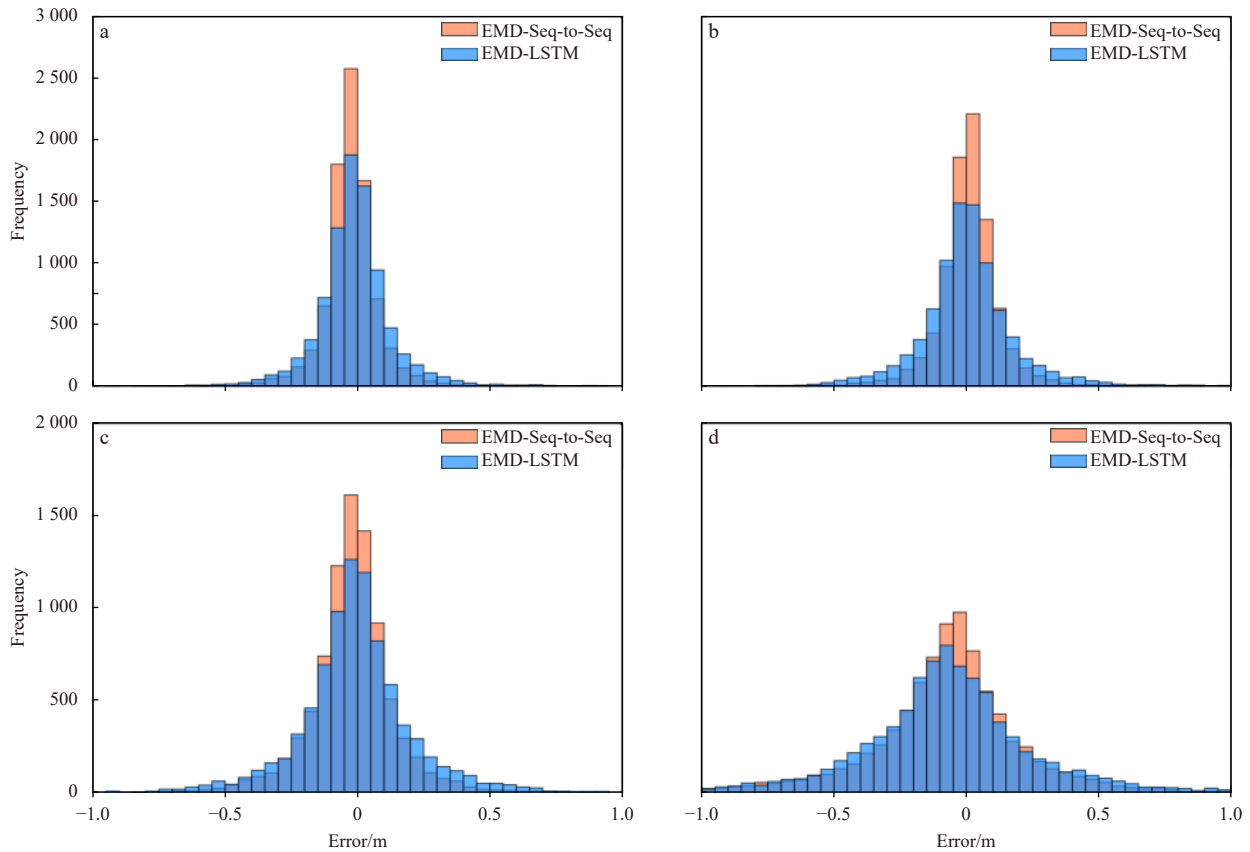


Fig. 11. Comparison of empirical mode decomposition-long short-term memory (EMD-LSTM) and EMD-sequence-to-sequence significant wave height forecast errors at the 3-h (a), 6-h (b), 12-h (c) and 24-h (d) forecast windows for Buoy 41047.

ber of test data distributed at around the 0 m mark decreases at the 3-h, 6-h, 12-h and 24-h forecast windows, as shown in Figs 11a–d. By contrast, EMD-Seq-to-Seq forecast errors are unevenly distributed at around the 0 m mark. In most cases, the frequency of absolute error greater than 0.15 m is lower than EMD-LSTM error frequencies at the 3-h, 6-h and 12-h forecast windows. In Fig. 11d, at the 24-h forecast window, EMD-Seq-to-Seq forecast errors are unevenly distributed at around the 0 m mark, whilst the frequency of error between 0.1 m and 0.3 m is slightly higher than that of EMD-LSTM model. Therefore, the EMD-Seq-to-Seq

maintains good robustness at the different forecast windows. This finding shows that the series of the Seq-to-Seq method not only has higher prediction accuracy than the LSTM model, but also avoids the important influence of time delay on prediction accuracy.

Modal aliasing exists in signal decomposition with discontinuities for EMD. Intermittent signals are high-frequency signals with small amplitudes at a certain time or within a remarkably small time interval. The obtained IMF is meaningless in the presence of mode aliasing. The EEMD algorithm adds white

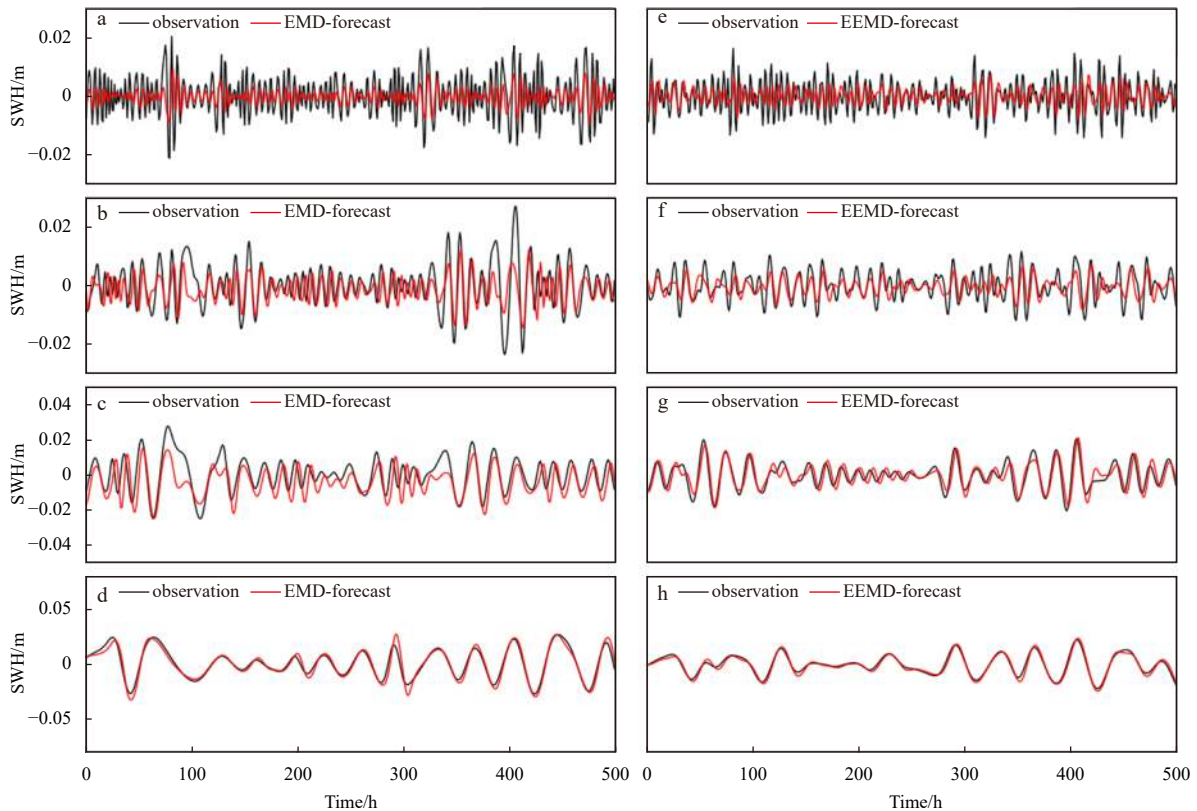


Fig. 12. Comparison of significant wave heights (SWHs) of the 2nd, 3rd, 4th and 5th intrinsic mode functions through empirical mode decomposition model (a–d) and ensemble empirical mode decomposition model (e–h).

noise to improve the modal aliasing caused by EMD decomposition and avoids high-frequency signals with small amplitudes after the decomposition of EEMD. The corresponding decomposed sequence signal is relatively stable, and the prediction accuracy of the algorithm is further improved. Since noise data are added into EEMD algorithm, the IMF values after the decomposition of EMD and EEMD algorithms are different. EMD-Seq-to-Seq and EEMD-Seq-to-Seq models for Buoy 41040 are selected to analyze the prediction performance at different IMFs sub-sequences through the EMD and EEMD algorithms, as shown in Fig. 12.

Figures 12a and e depict the prediction performance of the 2nd high-frequency IMFs decomposed by EMD and EEMD algorithms, respectively. Figure 12e shows that the predicted values obtained by the Seq-to-Seq model are close to the observation values through EEMD decomposition. Meanwhile, the Seq-to-Seq model can only capture major trends after EMD decomposition, ignoring important details and failing to predict observations, as shown in Fig. 12a. As the decomposed signal frequency gradually decreases, the fitting between the predicted and the observed values on the corresponding IMF sub-sequences continues to improve, and the prediction performance is significantly enhanced, as shown in Figs 12b–d (EMD algorithm) and f–h (EEMD algorithm). The prediction of the third IMF through EEMD decomposition (Fig. 12f) is superior to the corresponding prediction through EMD decomposition (Fig. 12b), whilst the absolute predicted value is significantly less than the observed extreme point value around the 100 h and 400 h after EMD decomposition, leading to the decrease in prediction accuracy. Figure 12h shows that the prediction on SWH of the 5th IMF decomposed by the EEMD algorithm is close to the

observation and better than that of the 5th IMF decomposed by the EMD algorithm. By contrast, the prediction performance through EEMD decomposition is further improved.

To compare the fitting degree between predicted and observed values in different models, the scatter diagrams of predicted and observed values at the 3-h, 6-h, 12-h, and 24-h forecast windows are depicted by taking Buoy 41040 as an example, as shown in Fig. 13. Among the four models, the EEMD-Seq-to-Seq model demonstrates the best performance under the same forecast window. The predicted values of each algorithm begin to spread outwards with the increase of the forecast window, but the dispersion of the EEMD-Seq-to-Seq model is the least. Consequently, the prediction performances of EMD-Seq-to-Seq and EMD-LSTM algorithms are respectively the 2nd and 3rd, and the LSTM algorithm has the most significant dispersion between predicted and the observed values, and demonstrates the worst performance. At the 24-h forecast window, the predicted values obtained by the EEMD-Seq-to-Seq model fit well with the observed values, and the predicted values obtained by other models have varying degrees of dispersion. The predicted values of SWH above 2 m are scattered in the EEMD-Seq-to-Seq model, whilst those around 1 m are dispersed for the three other models. For the EEMD-Seq-to-Seq model, the fitting line is optimal among all forecast windows, and its value of Pearson correlation coefficient is highest (larger than 0.95) among all forecast windows.

5 Conclusions

The SWH forecast is of considerable importance for oceanography and its application. Owing to the limitation of the prediction performance of current algorithms, the prediction performance of SWH is improved by studying the symbiotic relationship

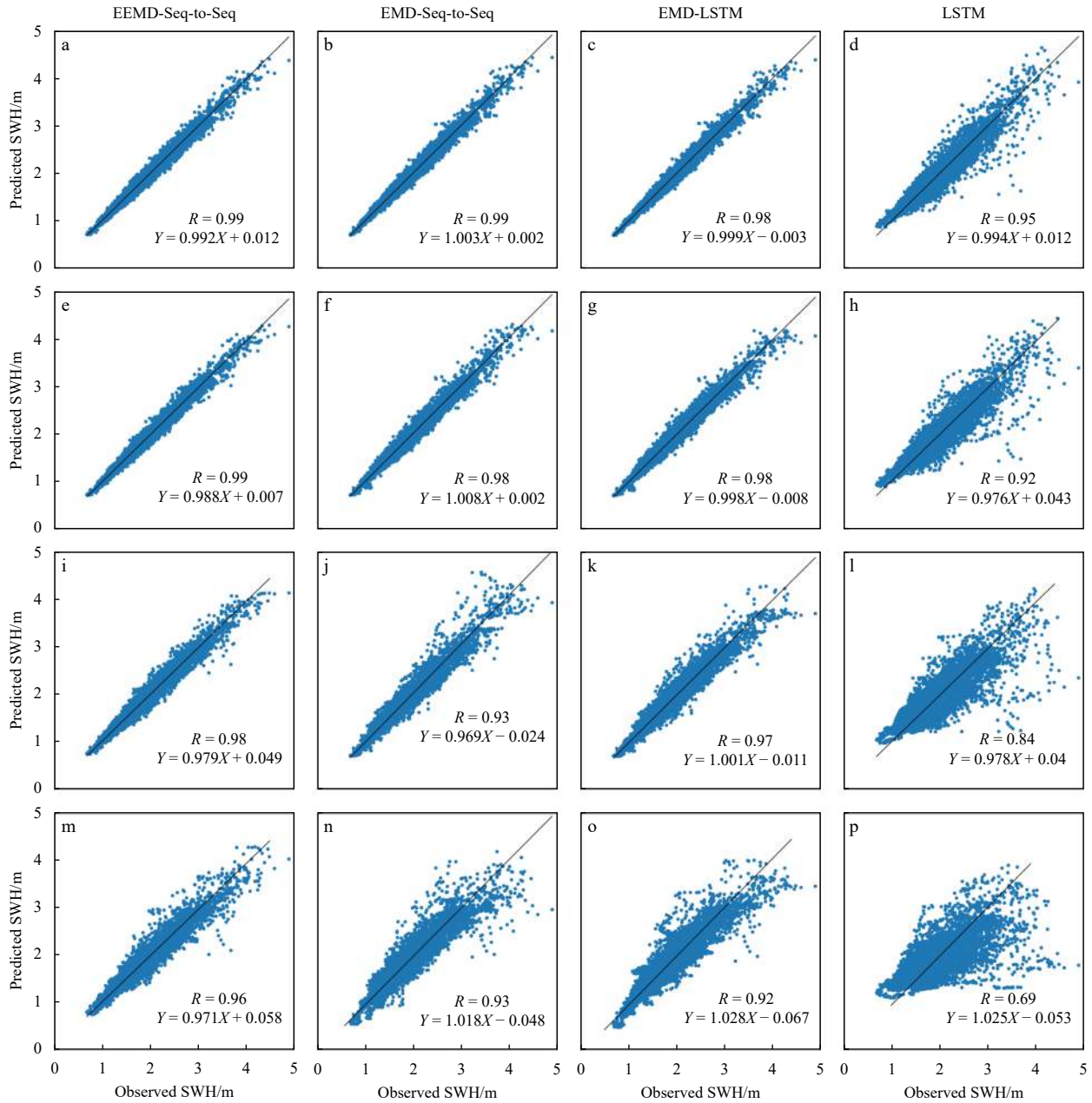


Fig. 13. Scatter diagram of the observed and predicted significant wave height (SWHs) obtained by different algorithms at Buoy 41040. a–d for 3-h forecast window, e–h for 6-h forecast window, i–l for 12-h forecast window, m–p for 24-h forecast window.

and effectiveness of the EEMD algorithm and the Seq-to-Seq network. EEMD adopts noise characteristics based on EMD decomposition to suppress modal aliasing caused by EMD decomposition effectively. Compared with the LSTM model, the Seq-to-Seq model comprises an encoder and a decoder. Multi-layer LSTM networks exist in an encoder and a decoder. Based on LSTM, the dependence between continuous predicted values increases in the Seq-to-Seq model, and the attention module is also considered. The attention mechanism allocates different weights to the source sequence in the decoding stage. Therefore, the decoder can maximize the important information of the source sequence to obtain better prediction performance than the LSTM model at multiple forecast windows. The results indicate that the Seq-to-Seq model integrated EEMD is superior to EMD-Seq-to-Seq, EMD-LSTM and LSTM considering prediction accuracy. As EEMD can decompose the original nonlinear SWH signal into a

series of stationary IMFs; thus, the Seq-to-Seq model can effectively capture the changes with the dependent trend. The proposed EEMD-Seq-to-Seq model is optimal from the perspectives of the predictive evaluation indices (RMSE, MAE, MAPE and Pearson coefficient).

The EEMD-Seq-to-Seq model significantly improves the prediction accuracy under short-term prediction. However, the prediction accuracy at 48-h and 72-h forecast windows must be further improved due to prediction errors probably caused by instrumentation (i.e., buoys) and other noise. The additional error may be due to the use of only wave height to predict waves and the effect of ocean currents on waves generated by wind (wind speed and direction) or modulation, which is an important barrier to further improvement. These factors not only improve the prediction accuracy of SWH but also enhance the response speed to extreme wave height. In future studies, addition information

about the local climate may be used to improve the prediction accuracy.

References

- Ardhuin F, Stopa J E, Chapron B, et al. 2019. Observing sea states. *Frontiers in Marine Science*, 6: 124, doi: [10.3389/fmars.2019.00124](https://doi.org/10.3389/fmars.2019.00124)
- Bahdanau D, Cho K, Bengio Y. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv: 1409.0473
- Bokde N, Feijóo A, Al-Ansari N, et al. 2020. The hybridization of ensemble empirical mode decomposition with forecasting models: application of short-term wind speed and power modeling. *Energies*, 13(7): 1666, doi: [10.3390/en13071666](https://doi.org/10.3390/en13071666)
- Caires S, Sterl A. 2005. 100-year return value estimates for ocean wind speed and significant wave height from the ERA-40 data. *Journal of Climate*, 18(7): 1032–1048, doi: [10.1175/JCLI-3312.1](https://doi.org/10.1175/JCLI-3312.1)
- Deo M C, Naidu C S. 1998. Real time wave forecasting using neural networks. *Ocean Engineering*, 26(3): 191–203, doi: [10.1016/S0029-8018\(97\)10025-7](https://doi.org/10.1016/S0029-8018(97)10025-7)
- Duan W Y, Han Y, Huang L M, et al. 2016. A hybrid EMD-SVR model for the short-term prediction of significant wave height. *Ocean Engineering*, 124: 54–73, doi: [10.1016/j.oceaneng.2016.05.049](https://doi.org/10.1016/j.oceaneng.2016.05.049)
- Etemad-Shahidi A, Mahjoobi J. 2009. Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior. *Ocean Engineering*, 36(15–16): 1175–1181, doi: [10.1016/j.oceaneng.2009.08.008](https://doi.org/10.1016/j.oceaneng.2009.08.008)
- Fan Shuntao, Xiao Nianhao, Dong Sheng. 2020. A novel model to predict significant wave height based on long short-term memory network. *Ocean Engineering*, 205: 107298, doi: [10.1016/j.oceaneng.2020.107298](https://doi.org/10.1016/j.oceaneng.2020.107298)
- Gong Gangjun, An Xiaonan, Mahato N K, et al. 2019. Research on short-term load prediction based on Seq2seq model. *Energies*, 12(16): 3199, doi: [10.3390/en12163199](https://doi.org/10.3390/en12163199)
- Graves A. 2012. Supervised Sequence Labelling with Recurrent Neural Networks. Heidelberg: Springer Berlin, 37–45
- Huang N E, Shen Zheng, Long S R, et al. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971): 903–995
- Karatzoglou A, Jablonski A, Beigl M. 2018. A Seq2Seq learning approach for modeling semantic trajectories and predicting the next location. In: *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Washington, Seattle: ACM
- Keneshloo Y, Shi Tian, Ramakrishnan N, et al. 2020. Deep reinforcement learning for sequence-to-sequence models. *IEEE Transactions on Neural Networks and Learning Systems*, 31(7): 2469–2489
- Mahjoobi J, Mosabbeq E A. 2009. Prediction of significant wave height using regressive support vector machines. *Ocean Engineering*, 36(5): 339–347, doi: [10.1016/j.oceaneng.2009.01.001](https://doi.org/10.1016/j.oceaneng.2009.01.001)
- Nikoo M R, Kerachian R, Alizadeh M R. 2018. A fuzzy KNN-based model for significant wave height prediction in large lakes. *Oceanologia*, 60(2): 153–168, doi: [10.1016/j.oceano.2017.09.003](https://doi.org/10.1016/j.oceano.2017.09.003)
- Oh J, Suh K D. 2018. Real-time forecasting of wave heights using EOF-wavelet-neural network hybrid model. *Ocean Engineering*, 150: 48–59, doi: [10.1016/j.oceaneng.2017.12.044](https://doi.org/10.1016/j.oceaneng.2017.12.044)
- Pirhooshyaran M, Snyder L V. 2020. Forecasting, hindcasting and feature selection of ocean waves via recurrent and sequence-to-sequence networks. *Ocean Engineering*, 207: 107424, doi: [10.1016/j.oceaneng.2020.107424](https://doi.org/10.1016/j.oceaneng.2020.107424)
- Raj N, Brown J. 2021. An EEMD-BiLSTM algorithm integrated with Boruta random forest optimiser for significant wave height forecasting along coastal areas of Queensland, Australia. *Remote Sensing*, 13(8): 1456, doi: [10.3390/rs13081456](https://doi.org/10.3390/rs13081456)
- Sutskever I, Vinyals O, Le Q V. 2014. Sequence to sequence learning with neural networks. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Cambridge: MIT Press
- Vanem E. 2016. Joint statistical models for significant wave height and wave period in a changing climate. *Marine Structures*, 49: 180–205, doi: [10.1016/j.marstruc.2016.06.001](https://doi.org/10.1016/j.marstruc.2016.06.001)
- Vaswani A, Shazeer N, Parmar N, et al. 2017. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook: Curran Associates Inc
- Wang Lina, Deng Xilin, Ge Peng, et al. 2022. CNN-BiLSTM-attention model in forecasting wave height over South-East China Seas. *Computers, Materials & Continua*, 73(1): 2151–2168
- Wu Zhaohua, Huang N E. 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1): 1–41, doi: [10.1142/S1793536909000047](https://doi.org/10.1142/S1793536909000047)
- Yang Shaobo, Deng Zegui, Li Xingfei, et al. 2021. A novel hybrid model based on STL decomposition and one-dimensional convolutional neural networks with positional encoding for significant wave height forecast. *Renewable Energy*, 173: 531–543, doi: [10.1016/j.renene.2021.04.010](https://doi.org/10.1016/j.renene.2021.04.010)
- Yang Yu, Wang Jun. 2021. Forecasting wavelet neural hybrid network with financial ensemble empirical mode decomposition and MCID evaluation. *Expert Systems with Applications*, 166: 114097, doi: [10.1016/j.eswa.2020.114097](https://doi.org/10.1016/j.eswa.2020.114097)
- Ye Lin, Dai Binhua, Pei Ming, et al. 2022. Combined approach for short-term wind power forecasting based on wave division and Seq2Seq model using deep learning. *IEEE Transactions on Industry Applications*, 58(2): 2586–2596, doi: [10.1109/TIA.2022.3146224](https://doi.org/10.1109/TIA.2022.3146224)
- Zhang Yu, Li Yanting, Zhang Guangyao. 2020. Short-term wind power forecasting approach based on Seq2Seq model using NWP data. *Energy*, 213: 118371, doi: [10.1016/j.energy.2020.118371](https://doi.org/10.1016/j.energy.2020.118371)
- Zhou Shuyi, Bethel B J, Sun Wenjin, et al. 2021. Improving significant wave height forecasts using a joint empirical mode decomposition-long short-term memory network. *Journal of Marine Science and Engineering*, 9(7): 744, doi: [10.3390/jmse9070744](https://doi.org/10.3390/jmse9070744)
- Zhou Shuyi, Xie Wenhong, Lu Yuxiang, et al. 2021. ConvLSTM-based wave forecasts in the South and East China Seas. *Frontiers in Marine Science*, 8: 680079, doi: [10.3389/fmars.2021.680079](https://doi.org/10.3389/fmars.2021.680079)