

Article

# Prediction of Significant Wave Height in Offshore China Based on the Machine Learning Method

Zhijie Feng<sup>1</sup>, Po Hu<sup>2,3,4,\*</sup>, Shuiqing Li<sup>2,3,4</sup> and Dongxue Mo<sup>2,3</sup>

<sup>1</sup> College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266580, China; zhijie\_sd@qdio.ac.cn

<sup>2</sup> Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Nanhai Road, 7, Qingdao 266071, China; lishuiqing@qdio.ac.cn (S.L.); dongxuemo@qdio.ac.cn (D.M.)

<sup>3</sup> Laboratory for Ocean Dynamics and Climate, Pilot National Laboratory for Marine Science and Technology (Qingdao), Wenhai Road, 1, Qingdao 266237, China

<sup>4</sup> University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, China

\* Correspondence: hupo@qdio.ac.cn; Tel.: +86-0532-82898516

**Abstract:** Accurate wave prediction can help avoid disasters. In this study, the significant wave height (SWH) prediction performances of the recurrent neural network (RNN), long short-term memory network (LSTM), and gated recurrent unit network (GRU) were compared. The 10 m u-component of wind (U10), 10 m v-component of wind (V10), and SWH of the previous 24 h were used as input parameters to predict the SWHs of the future 1, 3, 6, 12, and 24 h. The SWH prediction model was established at three different sites located in the Bohai Sea, the East China Sea, and the South China Sea, separately. The experimental results show that the performance of LSTM and GRU networks based on the gating mechanism was better than that of traditional RNNs, and the performances of the LSTM and GRU networks were comparable. The EMD method was found to be useful in the improvement of the LSTM network to forecast the significant wave heights of 12 and 24 h.

**Keywords:** wave height; recurrent neural network; long short-term memory network; GRU; EMD



**Citation:** Feng, Z.; Hu, P.; Li, S.; Mo, D. Prediction of Significant Wave Height in Offshore China Based on the Machine Learning Method. *J. Mar. Sci. Eng.* **2022**, *10*, 836. <https://doi.org/10.3390/jmse10060836>

Academic Editor: Alberto Ribotti

Received: 28 May 2022

Accepted: 16 June 2022

Published: 20 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Sea surface wind waves can change the course and speed of ships and even produce hull resonance, which can fracture the hull; they can damage ports, wharves, underwater engineering, and coastal protection engineering; and they can also affect the use of radar, the takeoff and landing of seaplanes and carrier-borne aircraft, mine laying, mine clearance, replenishment at sea, the use of shipborne weapons, and salvage at sea. Therefore, the study of wave conditions, especially the prediction of significant wave height, is of great significance for offshore operations.

Generally, wave prediction models can be divided into data-driven or physical-driven frameworks. The physical-driven method is based on the wave spectrum energy balance equation. Although the prediction of the numerical model is effective in large-scale space and time ranges, the disadvantage is that the costs in terms of computing resources and time are high [1]. The data-driven approach uses machine learning techniques to predict uncertain future times by analyzing possible relationships and dependencies between large amounts of data. Fuzzy systems (FSs) [2], evolutionary algorithms (EAs) [3], support vector machines (SVMs) [4], deep neural networks [5], and artificial neural networks (ANNs) [6] are all methods based on machine learning that are effective in constructing wave prediction models. Among the methods based on machine learning, it is effective to construct the prediction model of waves based on the recurrent neural network (RNN). This is because a wave is a continuous process, and the RNN is a type of recursive neural network that takes the sequence data as input, carries out recursion in the evolution direction of the sequence, and links all nodes in a chain. In the RNN, neurons can receive

information not only from other neurons but also from themselves, forming a network structure with loops, thus possessing short-term memory ability. However, when the input sequence is long, there will be gradient explosion and message problems [7]. The long short-term memory (LSTM) network [8] was proposed to solve the problem of gradient disappearance and gradient explosion in the RNN. Since its proposal, it has been widely applied. Theodoropoulos et al. [9] showed that LSTM models appeared to be suitable for propulsion power prediction, providing both great sensitivity and precision. Gao et al. [10] used the LSTM method to predict the path of typhoons and believed that the established model has a certain forecasting ability for the short-term path of typhoons. At the same time, Gao et al. [11] established significant wave height (SWH) prediction models for three stations in the Bohai Sea based on the LSTM method and believed that LSTM has broad application prospects in wave prediction. Fan et al. [12] conducted a comparative study on the prediction performance of six networks, such as a random forest, LSTM network, and SVM, in predicting significant wave height, and believed that the LSTM network has better prediction performance. However, the LSTM network has the problem of too many parameters, which leads to a long model training time. Therefore, the gated recurrent unit (GRU) network was developed. The GRU network can be understood as the simplified version of the LSTM network. It combines the forgetting gate and input gate of the LSTM network into a renewal gate, which reduces the training time of the model and improves the computing efficiency. The GRU network has been applied in many fields, such as traffic flow prediction [13,14] and short-term wind speed prediction [15]. Wang et al. [16] established a significant wave height (SWH) prediction model based on GRU, SVM, ELM, and other networks in the sea areas around Taiwan. The results show that the GRU network had the best performance in time series prediction.

Although these methods have achieved good results in the prediction of significant wave height, they are not effective over long time ranges, such as 12 or 24 h of significant wave height prediction. This is because waves are characterized by nonstationary characteristics. Although machine learning models can deal with nonlinear problems well, the modeling process of nonstationary data sets may be difficult without the preprocessing of input [17,18]. Therefore, Ozger et al. [19] added wavelet transform to the fuzzy logic method. Moreover, Deka et al. [18] combined wavelet changes and a neural network to predict the significant wave height. Their specific approach uses the wavelet technology to decompose the time series of the significant wave height and then takes the decomposed components as the input of the model. The prediction results are better than the single neural network prediction results. However, the neural network combined with wavelet analysis has limitations, because wavelet analysis needs to select a suitable mother wavelet and set up feasible decomposition layers.

Empirical mode decomposition (EMD) [20] does not need to set any basis function in advance; it only needs to decompose according to the time scale of the data. Since it was proposed, it has been widely used. Rios et al. [21] used the method based on EMD to conduct time series modeling and confirmed that such a method can be used as an effective tool to improve prediction accuracy. Duan et al. [22] used the EMD method to decompose wave time series to obtain each component and then used SVR to forecast each component. Finally, the prediction results of each component were summarized to obtain the prediction results. Huang et al. [23] evaluated the performance of this method by combining EMD and LSTM networks at three buoy stations on the east coast of the United States. The results show that this method was superior to the independent LSTM network, especially for the prediction of a longer time range of significant wave height. The combination of EMD and machine learning has become a new way to predict waves with higher accuracy than models using machine learning alone.

China is among the countries that has been most seriously affected by marine disasters. In recent years, the rapid development of the marine economy has led to the increasingly prominent risk of marine disasters in coastal areas, and the form of marine disaster prevention and mitigation is a serious concern. According to the 2020 Bulletin on China's

Marine Disasters, China's marine disasters are mainly storm surges and wave disasters. Various marine disasters have had many adverse effects on China's coastal economic and social development and marine ecology, causing total direct economic losses of CNY 832 million. Therefore, the study of wave conditions, especially the prediction of the significant wave height, is of great significance to the safety of people's lives and properties. The use of machine learning to predict the significant wave height of waves is a novel method. Compared with the traditional numerical model, the machine learning model has the characteristics of a lower computational cost and higher efficiency. For the long-term prediction of wave heights, instead of using a single machine learning model, a combination of an LSTM network and an EMD model is applied to wave height prediction. It provides a new research idea for wave forecasting method and has great reference significance.

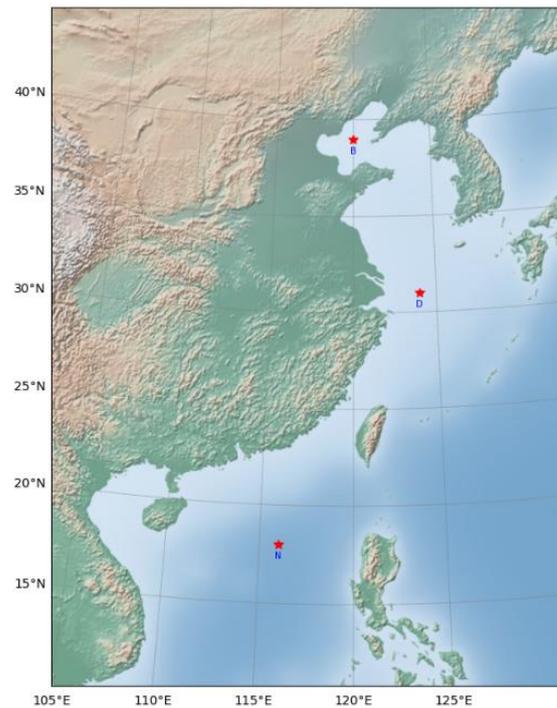
In this study, we evaluated the performance of the RNN, LSTM, and GRU, and the effect of EMD for wave height prediction in China's coastal seas, to derive the most efficient way to build a data-driven model for wave height prediction at different time ranges. We first compared the prediction performance of three networks for significant wave heights in China's offshore waters. The three networks were tested using numerical wave simulation results from three sites, located in the Bohai Sea, the East China Sea, and the South China Sea. As the connection between data decreased gradually with the increase in the prediction time range, the combination of EMD and the LSTM network was adopted to improve the prediction accuracy in the long time range prediction of waves. The remainder of the paper is structured as follows: Section 2 introduces the data and methods used in this study. Section 3 introduces the error index and prediction result of evaluating the performance of the prediction model. Then, the discussion and conclusion of this paper are given in the Sections 4 and 5, respectively.

## 2. Materials and Methods

### 2.1. Materials

The data used in this study were taken from ERA5 ("<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form>" (accessed on 1 November 2021)). ERA5 is the fifth generation of the European Centre for Medium-Range Weather Forecast (ECMWF) global climate reanalysis data warehouse. Reanalysis combines model data and observations from around the world into a global data set. Verification results of many researchers show that the ERA5 data set is consistent with buoy data in China's coastal seas [24,25]. Considering the complexity of wave formation in offshore China, representative locations in the Bohai Sea, the East China Sea, and the South China Sea were selected to compare the prediction performance of the three networks in offshore China. Figure 1 shows the exact location of the experimental site on a map. Table 1 gives detailed information about the selected sites, including the exact location of each site, the data period, and the total amount of data available. The wave prediction method based on machine learning has the characteristics of a higher computation time and lower cost compared with the traditional numerical wave prediction model. Therefore, the use of machine learning to predict waves has broad prospects, but machine learning training requires sufficient training samples to achieve excellent prediction results. Therefore, we selected ten years of site data to ensure sufficient data volume. Then, a statistical analysis was performed by drawing the effective wave height distribution histogram of each site; it was found that the data can well satisfy the requirement of machine learning data in terms of diversity. The main distribution interval of the significant wave height at position B was 0–4 m (Figure 2). The significant wave heights of D and N were mainly distributed in the interval of 0–6 m. The maximum significant wave height of position D reached more than 10 m, and that of position N is 8.510 m. According to statistics, there were 21 cases in which the significant wave height of position D was higher than 8 m, which may have resulted in the inferior prediction effect of position D compared to the other two points. This is because the machine learning model can only learn the relationship between the given input and output. When the samples between the given input and output were insufficient

or there was no relevant training sample at all, incorrect prediction results would be given, or the model would underestimate the wave height.

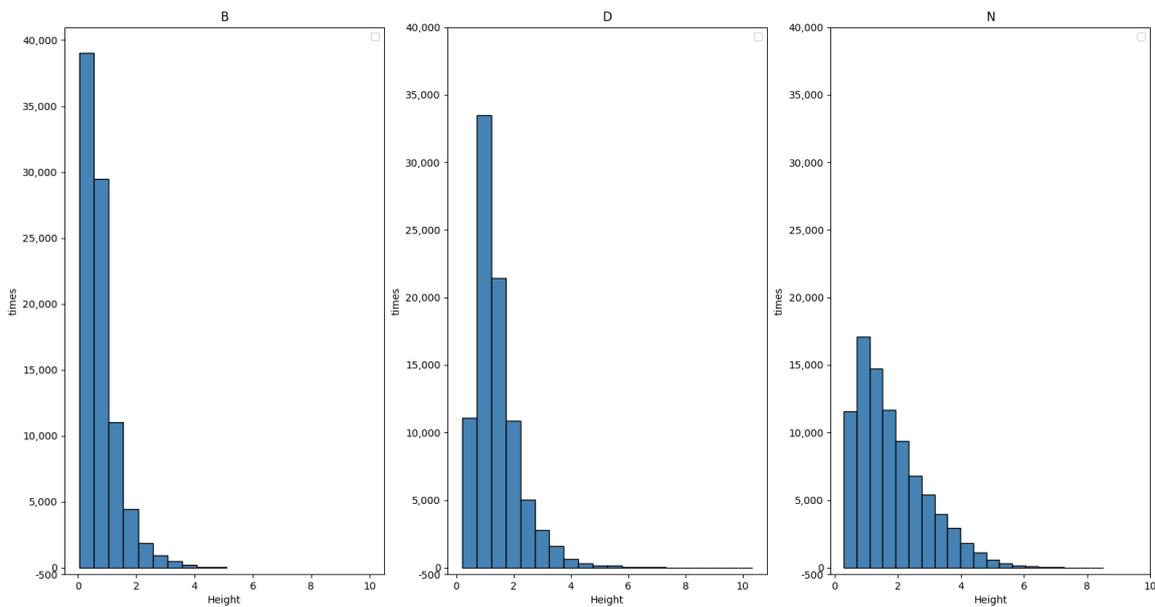


**Figure 1.** The specific location.

**Table 1.** Specific information about the selected location.

Location	Latitude (°N)	Longitude (°E)	Data Period	Total Number of Data Points
B	39.00	120.00	1 January 2011–31 December 2020	87,672
D	31.00	124.00	1 January 2011–31 December 2020	87,672
N	18.00	116.00	1 January 2011–31 December 2020	87,672

Wind speed has been identified as a major factor in wave generation [26]. In addition, as wave generation is a continuous process, this study considered the state of waves under the influence of wind speed and took the 10 m u-component of wind (U10), the 10 m v-component of wind (V10), and the historically significant wave height (SWH) as the inputs of the model. The interval of wind wave data used in this study was 1 h, and the data at each position were divided into two groups: the training validation set (80%) and the test set (20%). In general, the longer the forecast time horizon, the more challenging and less accurate it will be. Considering the above reasons, this study proposed to predict the significant wave heights in the following five periods: 1, 3, 6, 12, and 24 h.



**Figure 2.** Histogram of significant wave height distribution. The three subgraphs represent the significant wave height distribution histograms of the B, D, and N points from left to right.

2.2. Methods

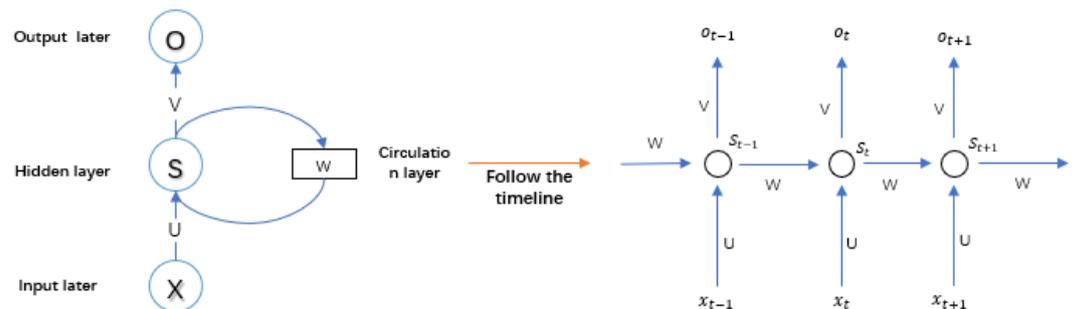
2.2.1. RNN

The RNN is a kind of neural network with short-term memory ability, and is widely used to mine temporal sequence information in data. The basic structure of the RNN is an input layer, hidden layer, and output layer. The hidden layer is used to learn and optimize parameters. The structural expansion of RNN is shown in Figure 3. The calculation method of the RNN is as follows:

$$O_t = g(V \cdot S_t) \tag{1}$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1}) \tag{2}$$

where  $X$  is sequence data, and  $X$  is a matrix consisting of three column vectors  $U_{10}$ ,  $V_{10}$ , and  $SWH$ ;  $U$  is the weight matrix from the input layer to the hidden layer;  $V$  is the weight matrix from the hidden layer to the output layer;  $W$  is the last value of the hidden layer as the weight matrix of this input;  $S$  is a vector representing the value of the hidden layer; and  $O$  represents the value of the output layer.



**Figure 3.** RNN timeline expansion diagram.

2.2.2. LSTM Network

RNNs are robust in modeling nonlinear time series, but they cannot avoid the problems of gradient disappearance and gradient explosion, and the accuracy decreases with the increase in the time span. To effectively avoid these problems, the LSTM [11] model was developed. The LSTM model uses three gating mechanisms based on the traditional RNN:

the forgetting gate, input gate, and output gate. Through these three gating mechanisms, the previous input information can be added or forgotten. The LSTM unit structure is shown in Figure 4.  $f_t$  represents the output result of the forgetting gate, whose function is to forget unimportant information and retain important information. In LSTM, a forgetting gate structure performs the following operations:

$$f_t = \sigma(W_f[x_t, h_{t-1}] + b_f) \tag{3}$$

where  $W_f$  is the weight matrix used to control the forgetting gate behavior,  $x_t$  is the sequence data,  $h_{t-1}$  is the hidden state at the last moment, and  $b_f$  is a bias vector. The output of the forgetting gate ( $f_t$  multiplies the corresponding element with the state value of the previous cell. Thus, if a value in  $f_t$  is 0 or close to 0, the corresponding information for the previous cell  $c_{t-1}$  will be discarded, and if the value in  $f_t$  is 1, the corresponding information will be retained.  $i_t$  represents the output result of updating the gate. The basic operation of updating the gate can be expressed as follows:

$$i_t = \sigma(W_f[h_{t-1}, x_t]) + b_i \tag{4}$$

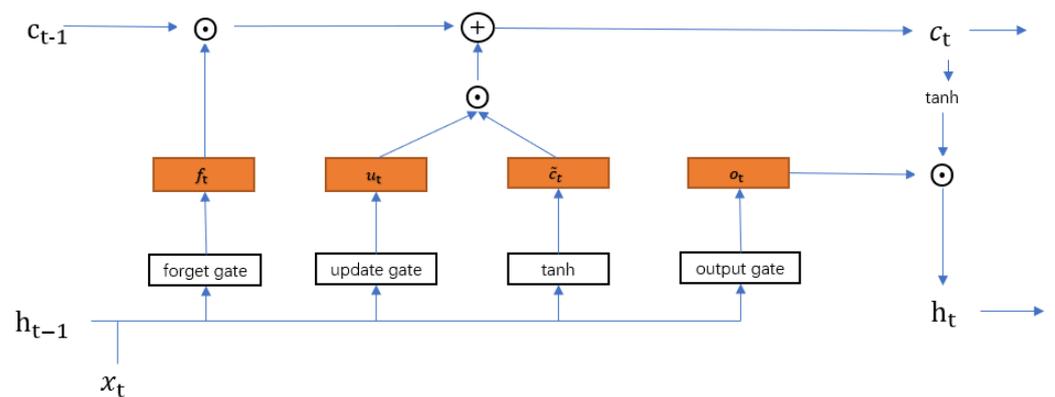


Figure 4. LSTM unit structure.

The output of the update gate ( $i_t$ ) is also a vector with values in the range [0, 1]. To calculate the new status information ( $c_t$ ) the output of the new gate will be multiplied among the elements of  $\tilde{c}_t$ . At the same time, we also need to update the unit state value ( $c$ ) passed between sequences, and the update process is as follows:

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{5}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{6}$$

where  $o_t$  is the output result of the output gate. The output value ( $o_t$  of the current unit and the hidden state ( $h_t$  passed to the next unit can be obtained from the output gate. The specific calculation process is as follows:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \odot \tanh(c_t) \tag{8}$$

### 2.2.3. GRU

GRU [27] is a variant of LSTM that optimizes the structure of LSTM while maintaining the performance of LSTM. It also solves the problems of gradient explosion and gradient disappearance of the standard RNN. The GRU network has only two gate structures: the update gate and reset gate. The update gate is equivalent to the combination of the

forgetting gate and input gate in LSTM, so the GRU network has fewer parameters and a faster training speed. The internal structure of the GRU structural unit is shown in Figure 5.  $x_t$  and  $h_{t-1}$  are sequence data and historical states, respectively, and the resetting gate ( $r_t$ ) is used to control whether the calculation of the candidate state ( $\tilde{h}_t$ ) depends on  $h_{t-1}$  of the previous municipality. The operations are as follows:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{9}$$

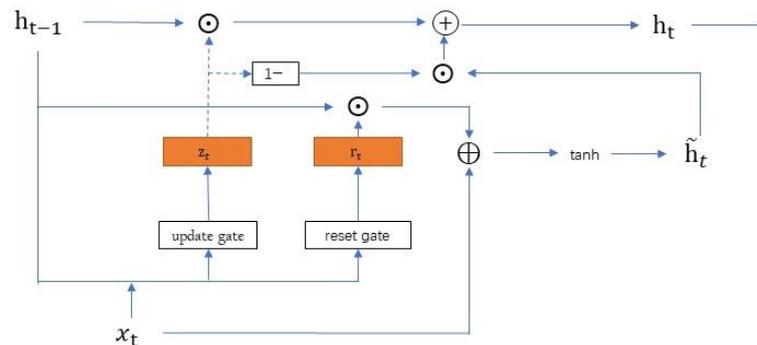


Figure 5. GRU unit structure.

The candidate status of the current moment is:

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \tag{10}$$

The update gate ( $z_t$ ) is used to control how much information is saved in the current state ( $h_t$ ) from the historical state ( $h_{t-1}$ ) and how much new information it needs to receive from the candidate state ( $\tilde{h}_t$ ).  $b_r$ ,  $b_h$  above and  $b_z$  below are all bias vectors. The calculation formula of  $z_t$  is:

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{11}$$

The hidden state ( $h_t$ ) is then calculated as:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \tag{12}$$

#### 2.2.4. Empirical Mode Decomposition

EMD, proposed in 1998, is a relatively new method for processing nonstationary signals [19], and works based on the time scale characteristics of a piece of data without setting any basis function in advance. This point is fundamentally different from Fourier decomposition and wavelet decomposition, which are based on the harmonic basis function and wavelet basis function, respectively. Therefore, the EMD method can be applied to any type of signal decomposition in theory and has obvious advantages in dealing with nonstationary nonlinear data. The time series of wave characteristics is composed of different oscillation scales and is a kind of complex nonlinear nonstationary signal. The prediction of the machine learning model with multiple oscillating scales is difficult, so proper signal preprocessing technology is needed to improve its performance.

Intrinsic mode functions (IMFs) are the signal components of each layer after element signals are decomposed by EMD. Any signal can be divided into the sum of several connotative modal components. Intrinsic mode functions (IMFs) have two constraints:

1. In the whole data segment, the number of local extreme value points and the number of zero crossing points must be equal or differ by a maximum of one.
2. At any time, the average value of the upper envelope formed by the local maximum point and the lower envelope formed by the local minimum point is zero; that is, the upper and lower envelope are locally symmetric with respect to the time axis.

Given time series  $x(t)$ , the EMD decomposition steps and flowchart (Figure 6) are as follows:

1. The upper and lower envelope lines ( $u_t$  and  $l_t$  respectively) are drawn according to spline interpolation among all the local maxima and the local minima of  $x(t)$ .
2. Find the mean of the upper and lower envelope and plot the mean envelope  $m(t) = [(l(t) + u(t))]/2$ .
3. Subtract the mean envelope from the original signal  $x(t)$  to obtain the intermediate signal  $f(t)$ .
4. Determine whether  $f(t)$  meets the two conditions of IMF. If so,  $f(t)$  is an IMF1; let us call it  $f_1(t)$ . If not, the analysis of (1)–(4) is repeated on the basis of  $f(t)$  until the two IMF conditions are met.
5. After the first IMF is obtained using the above method, the original signal is subtracted from IMF1 as the new original signal, and then IMF2 can be obtained through the analysis of (1)–(4) to complete EMD decomposition. Finally, the signal that does not satisfy the decomposition condition is denoted  $r(t)$ .
6. Through EMD algorithm, signals can be decomposed into:

$$x(t) = \sum_{i=1}^n f_i(t) + r(t) \tag{13}$$

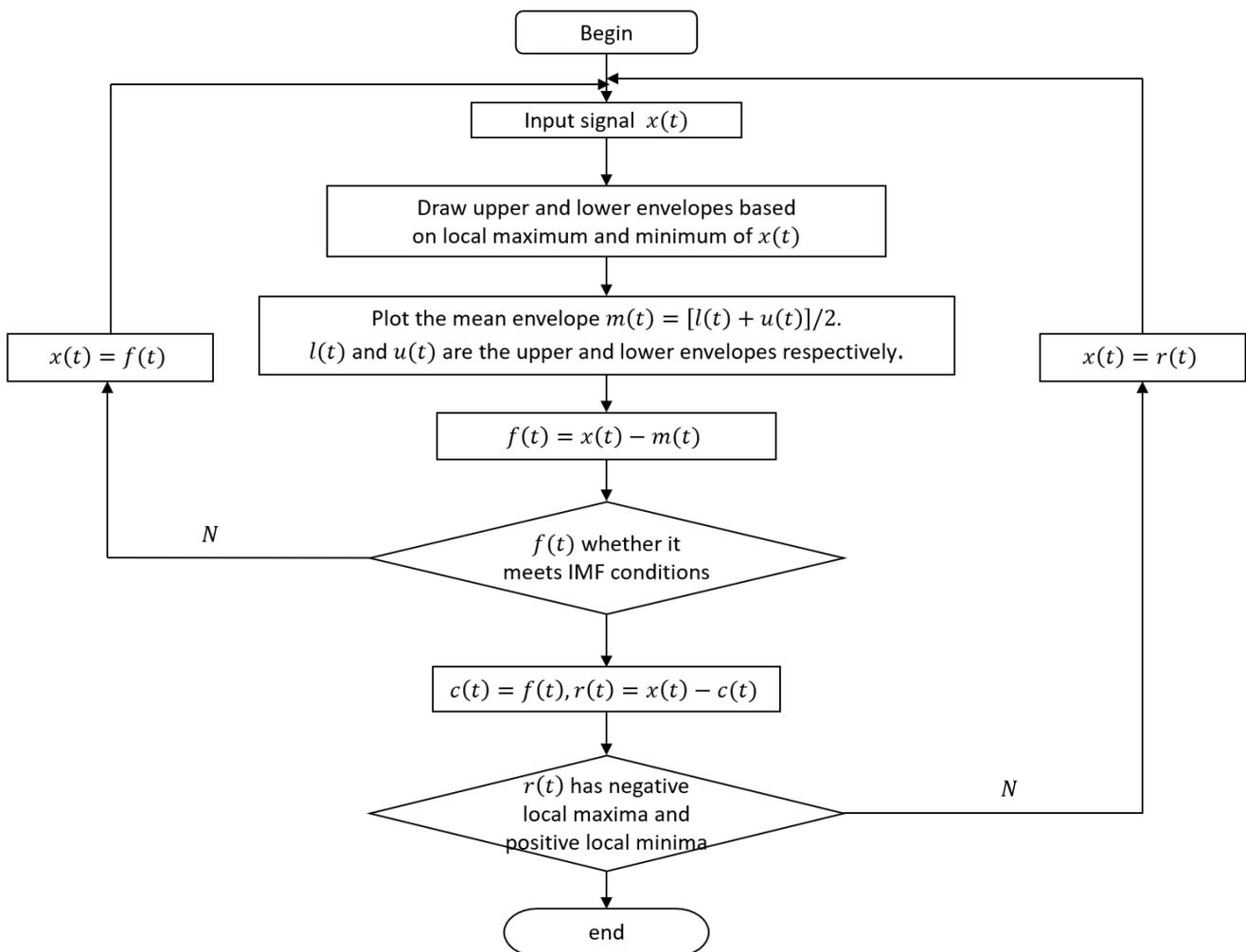


Figure 6. The EMD flowchart.

### 3. Results

#### 3.1. Error Metrics

In this study, wave prediction models for 1, 3, 6, 12, and 24 h were established at three locations with U10, U10, and SWH as inputs based on the RNN, LSTM, and GRU methods. The mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (R) were selected as the evaluation indexes of the model to evaluate the accuracy of the model, which are defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{14}$$

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

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \tag{16}$$

where  $n$  represents the total number of samples,  $y_i$  represents the tag value,  $\bar{y}$  is the average of the tag values,  $\hat{y}_i$  is the predicted value of the model, and  $\bar{\hat{y}}$  is the average of the predicted values. Meters are the units used for MAE and RMSE in this paper.

Based on the fact that wind is the relevant factor in the generation of waves, U10, V10, and SWH at historical moments were taken as input parameters, and the significant wave heights at 1, 3, 6, 12, and 24 h were taken as the label values. Using RNN, LSTM, and GRU networks to predict the significant wave heights in different forecast periods, we compared the predictions from the three machine learning algorithms.

To fairly compare the prediction performance of the three networks for significant wave heights in different sea areas, we set roughly the same total number of parameters for each model so that the calculation efficiency of the three models was roughly the same. Specifically, we referred to the method used by Chung et al. [28]. The loss function in the machine learning model is used to evaluate the gap between the predicted value and the label value, so as to guide the next training in the right direction. Since the output of the model is a specific value, the mean squared error (MSE) is used as the loss function. MSE is defined as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{17}$$

where  $y_i$  represents the tag value, and  $\hat{y}_i$  is the predicted value of the model.

The role of the optimizer is to optimize the weights in the network model so that the output of the model reaches the optimal value. The widely used Adam optimizer was used as the optimizer. The iteration period was chosen to be 15, because the loss function no longer decreases after the number of iterations exceeds 15 when training the model. Model size details are shown in Table 2.

**Table 2.** Dimensions of each model. The specific meaning of First layer is the number of neurons in the first layer. The specific meaning of Second layer is the number of neurons in the second layer.

	First Layer	Second Layer	Total Parameter
RNN	95	180	59,990
LSTM	60	80	60,885
GRU	70	90	59,945

#### 3.2. Results

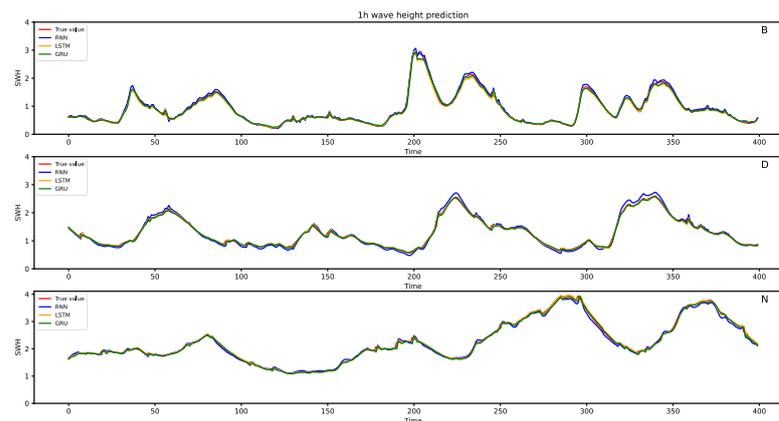
The error metrics of three different locations selected by the three algorithms in the Bohai Sea (B), the East China Sea (D), and the South China Sea (N) are summarized in Table 3. The results of 1 h prediction show that the three networks were excellent in the 1 h prediction of waves. However, in the three sea areas, the prediction accuracy of LSTM

and GRU was better than that of RNN. Of these, LSTM and GRU networks had basically the same ability to predict significant wave heights at identical locations. For example, the 1 h MAE of the two algorithms for position B was 0.031 and 0.029, and R values were consistent. The absolute errors of the MAE and RMSE of the LSTM and GRU networks at position D were 0.002 and 0.006, respectively, and the other error indicators were basically consistent. The possible reason why the RNN network lags behind is that LSTM and GRU networks can make full use of the information of the previous time period without the problems of gradient explosion and gradient disappearance.

**Table 3.** Error of different methods at each point. The bold text in the table represents the optimal forecast value for different locations at the same forecast time.

Time Span	RNN			LSTM			GRU			EMD-LSTM			
	MAE	RMSE	R	MAE	RMSE	R	MAE	RMSE	R	MAE	RMSE	R	
B	1	0.036	0.051	<b>0.998</b>	0.031	0.049	<b>0.998</b>	<b>0.029</b>	<b>0.046</b>	<b>0.998</b>			
	3	0.087	0.131	0.976	<b>0.072</b>	0.123	0.979	0.073	<b>0.120</b>	<b>0.980</b>			
	6	0.182	0.272	0.891	<b>0.156</b>	0.258	0.900	0.159	<b>0.256</b>	<b>0.901</b>			
	12	0.308	0.447	0.659	<b>0.283</b>	0.434	0.680	0.284	<b>0.431</b>	<b>0.683</b>	0.132	0.197	0.944
	24	0.388	<b>0.554</b>	0.345	<b>0.372</b>	<b>0.552</b>	0.371	0.378	<b>0.552</b>	0.373	0.223	0.335	0.825
D	1	0.058	0.081	0.997	<b>0.027</b>	0.059	0.997	0.029	<b>0.053</b>	<b>0.998</b>			
	3	0.090	0.140	0.986	<b>0.068</b>	<b>0.120</b>	<b>0.988</b>	0.071	0.121	0.987			
	6	0.149	0.243	0.951	<b>0.133</b>	<b>0.224</b>	<b>0.956</b>	0.135	0.226	0.955			
	12	0.253	0.409	0.843	0.250	0.399	0.855	<b>0.245</b>	<b>0.394</b>	<b>0.857</b>	0.159	0.229	0.954
	24	<b>0.407</b>	0.617	0.586	0.421	<b>0.622</b>	0.602	0.410	<b>0.602</b>	0.610	0.268	0.396	0.853
N	1	0.050	0.516	<b>0.999</b>	<b>0.027</b>	<b>0.043</b>	<b>0.999</b>	<b>0.027</b>	0.044	<b>0.999</b>			
	3	0.075	0.106	0.997	0.058	0.088	0.997	<b>0.055</b>	<b>0.085</b>	<b>0.997</b>			
	6	0.121	0.169	0.990	0.107	0.160	0.990	<b>0.104</b>	<b>0.154</b>	<b>0.991</b>			
	12	0.216	0.297	0.966	0.201	0.292	0.966	<b>0.195</b>	<b>0.283</b>	<b>0.968</b>	0.124	0.171	0.992
	24	0.386	0.516	0.894	<b>0.360</b>	0.510	0.892	0.369	<b>0.499</b>	<b>0.896</b>	0.176	0.263	0.974

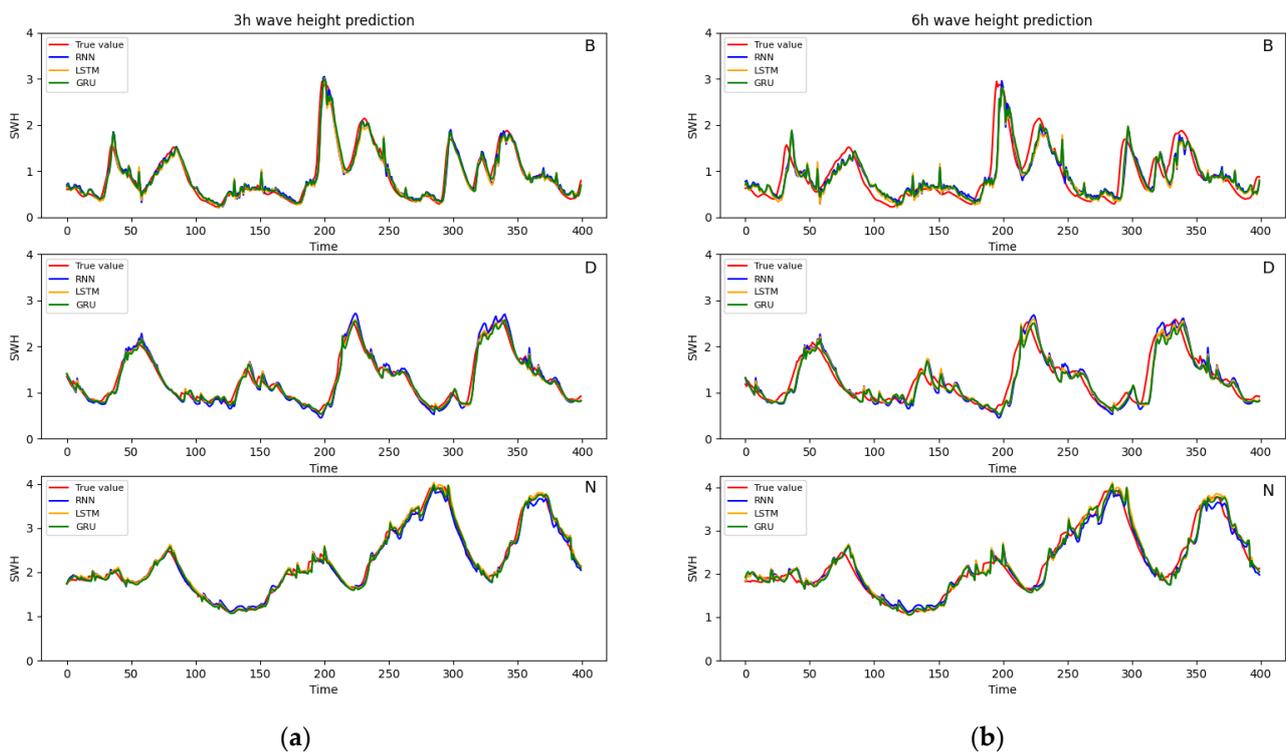
To clearly reveal the difference in the 1 h prediction performance of the wave height of the three networks in different regions, we randomly selected 400 consecutive hours of predictions for each location. The results are shown in Figure 7. From top to bottom, the three subgraphs show the comparison of the real value and the predicted value of different networks at sites B, D, and N. The red line represents the true value, the blue line represents the predicted value of the RNN network, the orange line represents the predicted value of the LSTM network, and the green line represents the predicted value of the GRU network. A careful observation of Figure 6 shows that the variation trend of the GRU network and the LSTM network at different positions was basically the same as that of the real value, whereas the volatility of RNN was slightly greater. The reason for the fluctuation may be that the traditional recurrent neural network misses important information or memorizes a large amount of unimportant information, whereas the LSTM and GRU networks control the accumulation speed of information by introducing a gating mechanism, including selectively adding new information and selectively forgetting the previously accumulated information. Thus, these networks have a stronger ability to mine information and have better stability.



**Figure 7.** One-hour forecast results; the abscissa represents the time, the ordinate represents the wave height, the red line represents the real wave height, the blue represents the predicted value of the RNN, the yellow represents the predicted value of LSTM, and the green represents the predicted value of GRU.

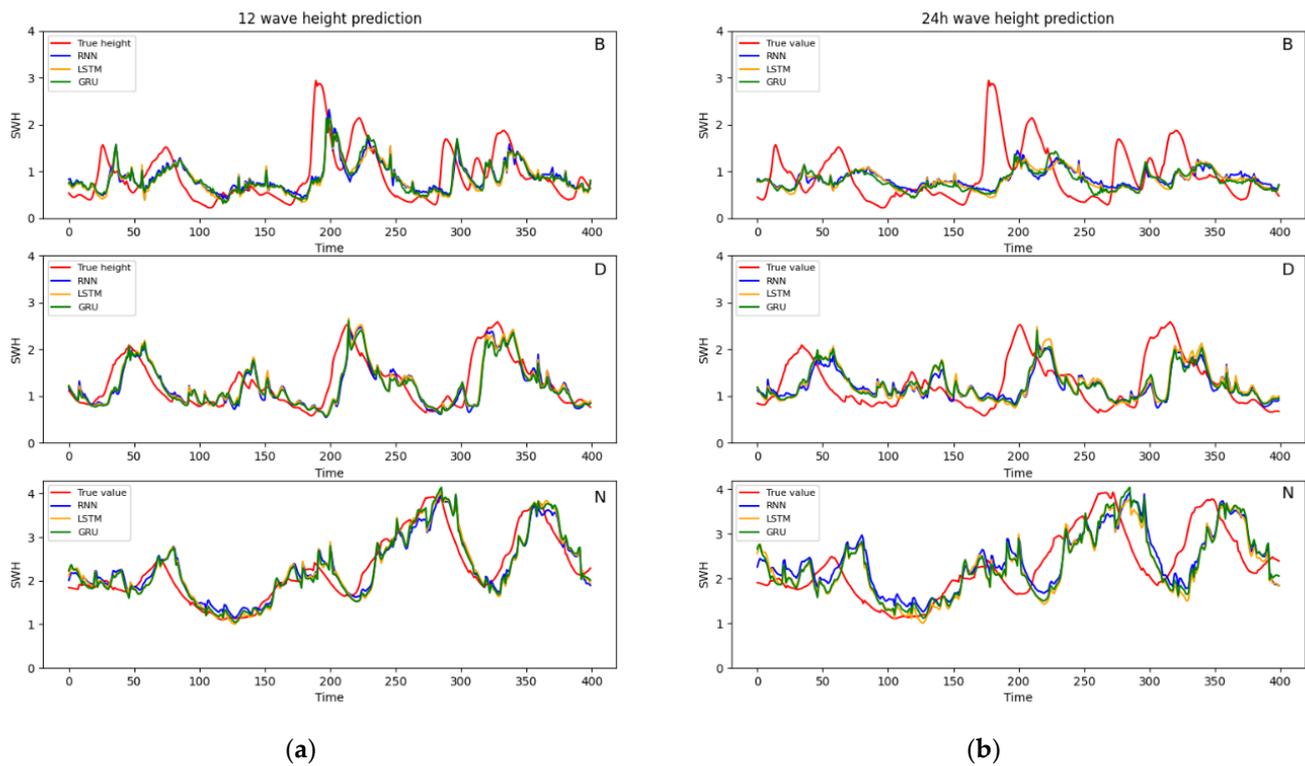
The 3 h and 6 h forecast results in Table 3 show that, with the increase in the prediction time interval, the prediction accuracy decreased as the MAE and the RMSE of each point gradually increased, and the correlation coefficient  $R$  gradually decreased. For example, the MAE of GRU network at N site for 1, 3, and 6 h was 0.027, 0.055, and 0.104, respectively. The RMSE was 0.044, 0.085, and 0.154.  $R$  was 0.999, 0.997, and 0.991. It can also be seen clearly that among the three locations B, D, and N, the forecasting ability of location N was still superior to that of locations B and D. The MAE and RMSE of the 3 h significant wave height of the RNN network at point N were 0.075 and 0.106, respectively, which are better than those of B. The MAE of the 3 h significant wave height of N-point LSTM network increased by 19.4% and 14.7% compared with B and D, respectively. Furthermore, the MAE of the 3 h significant wave height of the N-point GRU network increased by 24.7% and 22.5% compared with B and D, respectively. Similarly, the prediction performance of the LSTM and GRU networks at 3 and 6 h was still better than that of the traditional RNN. For example, the MAE of the 3 h RNN networks at position N was 29.3% and 36.4% higher than that of LSTM and GRU networks, respectively. Compared with the LSTM and GRU networks, the MAE of the 6 h RNN network at position B was increased by 16.7% and 14.5%. The 3 and 6 h predictive performances of LSTM and GRU networks were similar, and the maximum absolute error of MAE for the same location was only 0.003.

Figure 8 shows the comparison between the predicted results and the real values of the three networks at different locations in 3 and 6 h (400 time points were randomly selected). With the increase in prediction time, the connection in the time series grew weaker, leading to the increase in prediction accuracy. The variation trend in the predicted results of the three models at different locations within 3 h was basically consistent with that of the real value. The 6 h prediction results can be observed as follows: some peak points at B and D, and the predicted values lag. The variation trend in the predicted value and the real value at position N was basically the same, and there was basically no lag phenomenon. The likely reason for this is that the data distribution of the significant wave height at position N was more uniform, and the fitting effect was better.



**Figure 8.** Three-hour and 6 h prediction results with more instructions. The abscissa represents the time, the ordinate represents the significant wave height, the red line represents the real significant wave height, the blue line represents the predicted value of the RNN, the yellow line represents the predicted value of LSTM, and the green line represents the predicted value of GRU. (a) represents the 3 h forecast for different locations; (b) represents the 6 h forecast for different locations.

By checking the significant wave height prediction results at 12 and 24 h given in Table 3, we can see that both the forecast results of 12 h and the forecast error results of 24 h showed large errors compared with the real value, because the large increase in forecast time caused a sharp drop in the links in the data series, leading to worse forecast results. In the 12 h forecast, the MAE of the worst forecast error of the three networks at different locations was 0.308 and the corresponding RMSE was 0.447, whereas the MAE of the optimal forecast error was 0.195 and the corresponding RMSE was 0.283. The worst MAE and RMSE of the 24 h forecast were 0.421 and 0.622, respectively. The MAE of the optimal forecast error was 0.360, and the corresponding RMSE was 0.510. It can be seen that the RNN, LSTM, and GRU networks continued to maintain good performance in the long-term prediction of waves. Regarding the long-term prediction of significant wave heights, the LSTM and GRU networks still had better and more stable prediction performance than the RNN network. To more intuitively show the difference between the significant wave heights of different networks at 12 and 24 h, the prediction results of 400 moments randomly selected for each location are shown in Figure 9. As shown, the prediction results of 12 and 24 h had an obvious lag or advance trend, but the overall trend was still roughly consistent, indicating that RNN, LSTM, and GRU networks still have a certain reference value for the prediction of the long-term significant wave height. At the same time, the 12 and 24 h forecast results show that the forecasts of the three networks at the local maximum point were all underpredicted, especially at point B.



**Figure 9.** Twelve-hour and 24 h prediction results; the abscissa represents the time, the ordinate represents the significant wave height, the red line represents the real significant wave height, the blue line represents the predicted value of the RNN, the yellow line represents the predicted value of LSTM, and the green line represents the predicted value of GRU. (a) represents the 12 h forecast for different locations; (b) represents the 24 h forecast for different locations.

With the increase in the prediction time interval to 12 or 24 h, the prediction performance of the three networks degraded significantly. To improve the prediction accuracy at these long prediction time intervals, the EMD method was specifically used in the establishment of the LSTM network. The EMD method decomposes the signal to obtain IMF and residual components. Then, with U10 and V10 combined as the input data set. Model training, validation, and testing were performed at three locations: B, D, and N.

The time series of significant wave heights decomposed by EMD is shown in Figures 10–12. The significant wave height time series at positions B and N are decomposed into 15 IMFs and one residual by the EMD method, whereas the significant wave height time series at position D is decomposed into 16 IMFs and one residual.

Table 3 shows the prediction error results of three positions using the EMD method. The error results show that EMD-LSTM was significantly improved compared to the LSTM network in predicting significant wave heights at different positions. The 12 h correction results at point N show that the MAE and RMSE of EMD-LSTM were 0.124 and 0.171, respectively, which are 36.4% and 39.6% higher than those of the LSTM network, respectively. Moreover, the correlation coefficient reached 0.988. The MAE and RMSE of point D based on EMD-LSTM were 0.159 and 0.229, respectively, which are 36.4% and 42.6% smaller than those of LSTM, respectively. The correlation coefficient also increased from 0.855 to 0.954. Similarly, the errors of the other two points were significantly corrected, indicating that the EMD-LSTM method still performed well in the 24 h correction results.

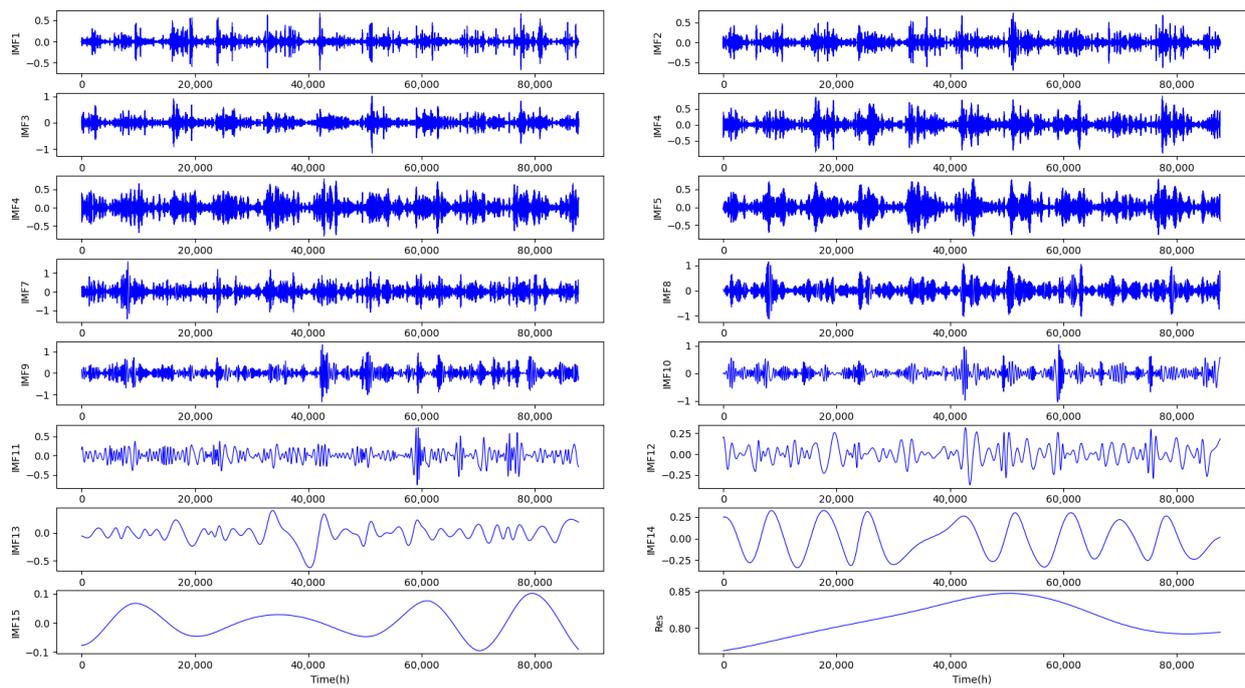


Figure 10. The EMD method was used to decompose the significant wave height results at position B.

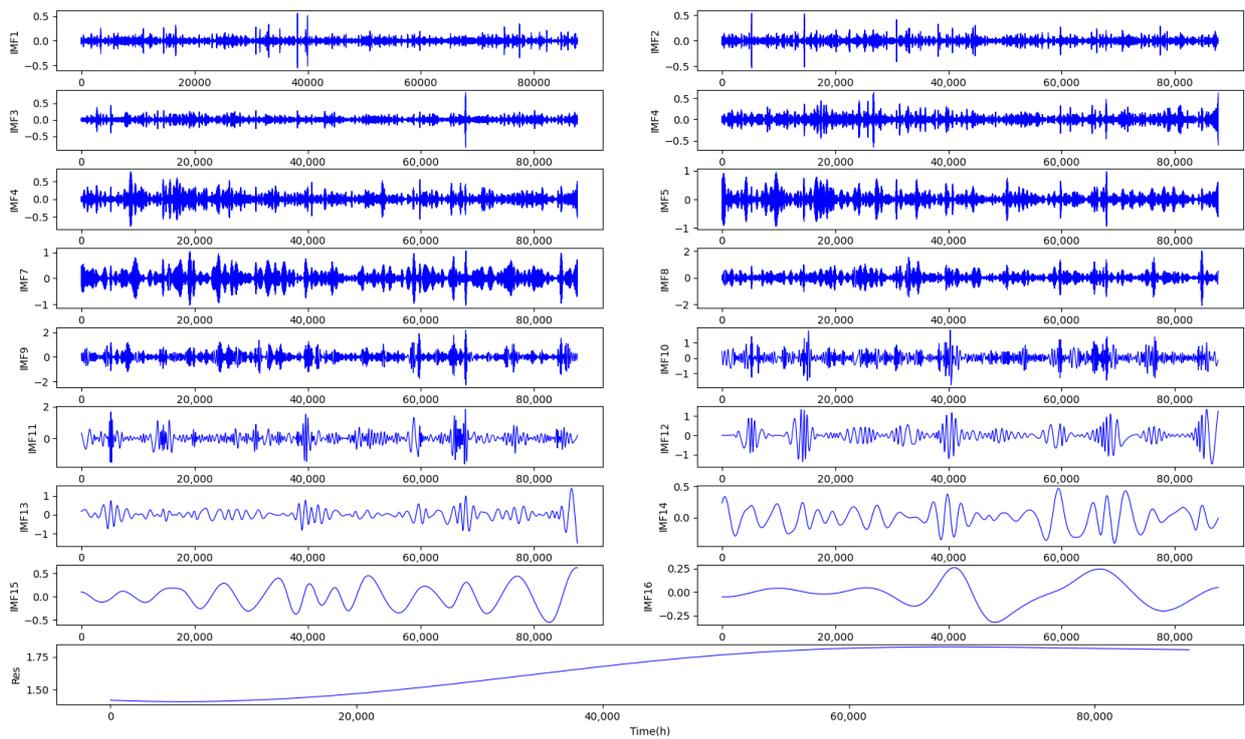
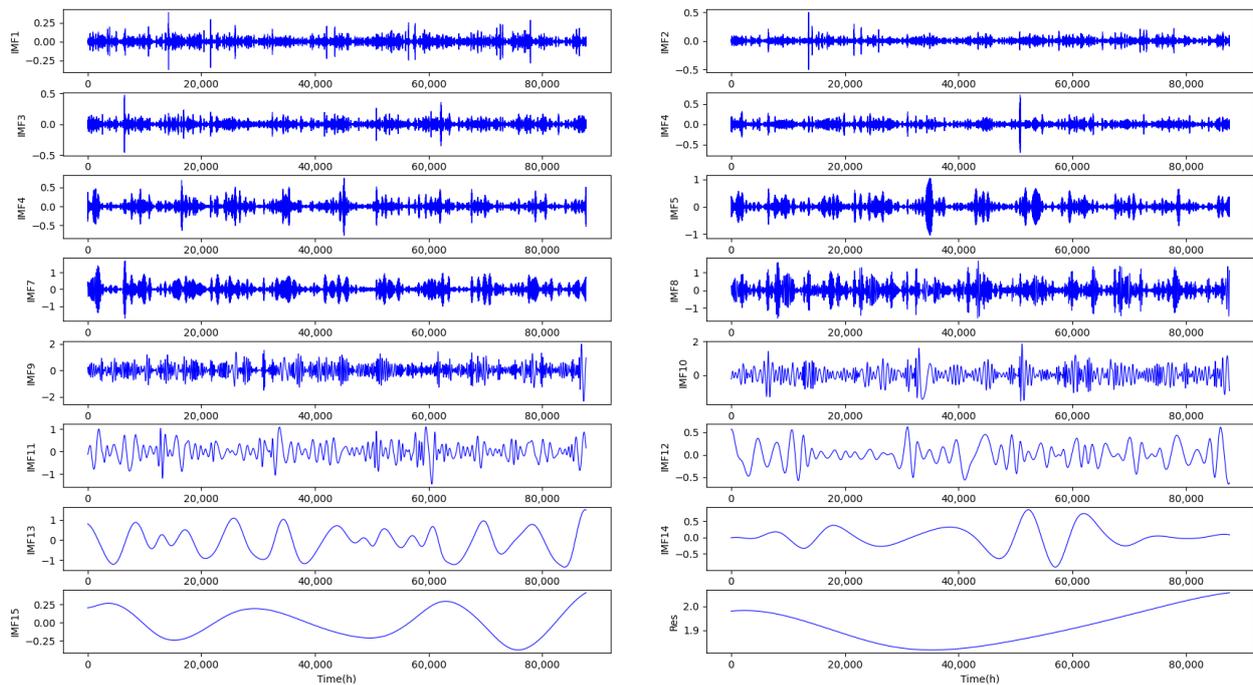
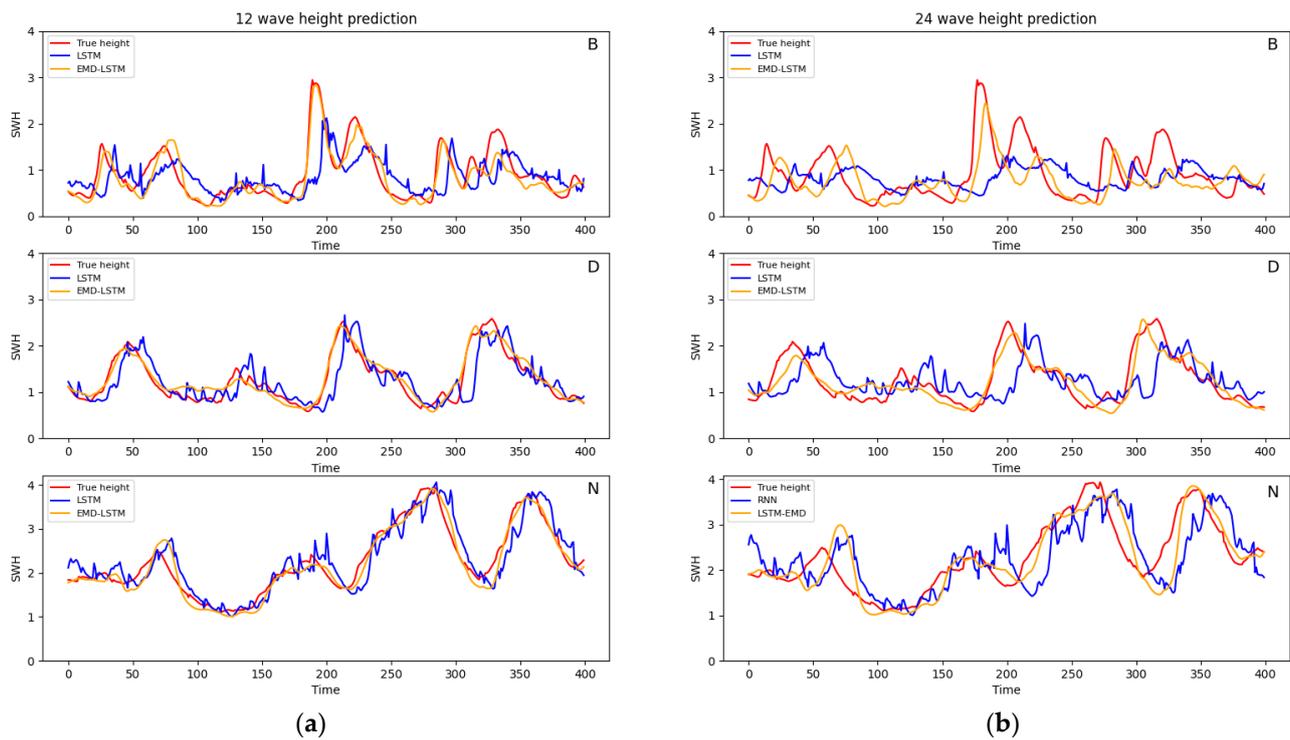


Figure 11. The EMD method was used to decompose the significant wave height results at position D.



**Figure 12.** The EMD method was used to decompose the significant wave height results at position N.

To intuitively display the effect of the EMD method on wave state modification at 12 and 24 h, 400 prediction points at random were selected from the test data set, as shown in Figure 13. The red line represents the true wave height, the blue line represents the forecast value of the LSTM network, and the orange line represents the forecast value of the EMD-LSTM network. The locations represented by the three subgraphs from top to bottom are B, D, and N. The 12 h forecast results show that the EMD-LSTM forecast value was more in line with the trend in the real value, and effectively alleviated the phenomenon of the LSTM forecast method at the local maximum point of the forecast and hysteresis. The most likely reason that the EMD method can effectively correct the forecast results is that the time series of significant wave heights is nonlinear and nonstationary, and there are large seasonal and regional differences in the wave characteristics in the offshore China seas. Furthermore, there is a close relationship between wave characteristics and monsoons. EMD has inherent advantages in analyzing nonlinear and nonstationary sequences. Using the EMD method to decompose the time series of significant wave height may effectively reveal the changing rule of the significant wave height and achieve a better forecast effect. EMD-LSTM still performed better than the LSTM network in predicting the 24 h wave height, which was especially obvious at the peak point of position B. Overall, the use of the EMD method significantly improved the prediction of 12 and 24 h significant wave heights.

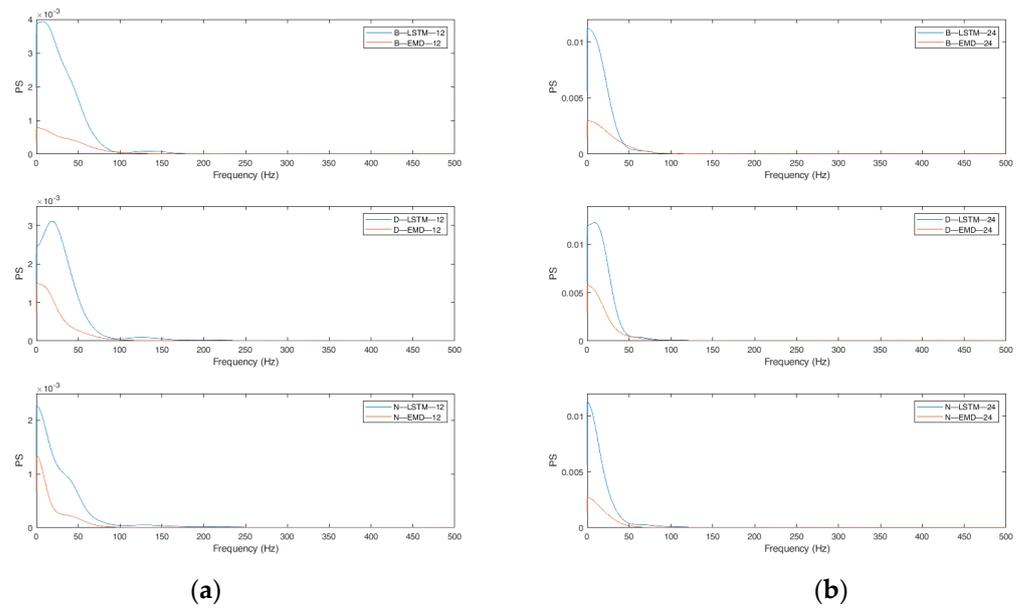


**Figure 13.** Twelve-hour and 24 h correction results. The abscissa represents time, the ordinate represents significant wave height, the red line represents the true significant wave height, the blue line represents the predicted value of LSTM, and the yellow line represents the predicted value of EMD-LSTM. (a) represents the 12 h forecast for different locations; (b) represents the 24 h forecast for different locations.

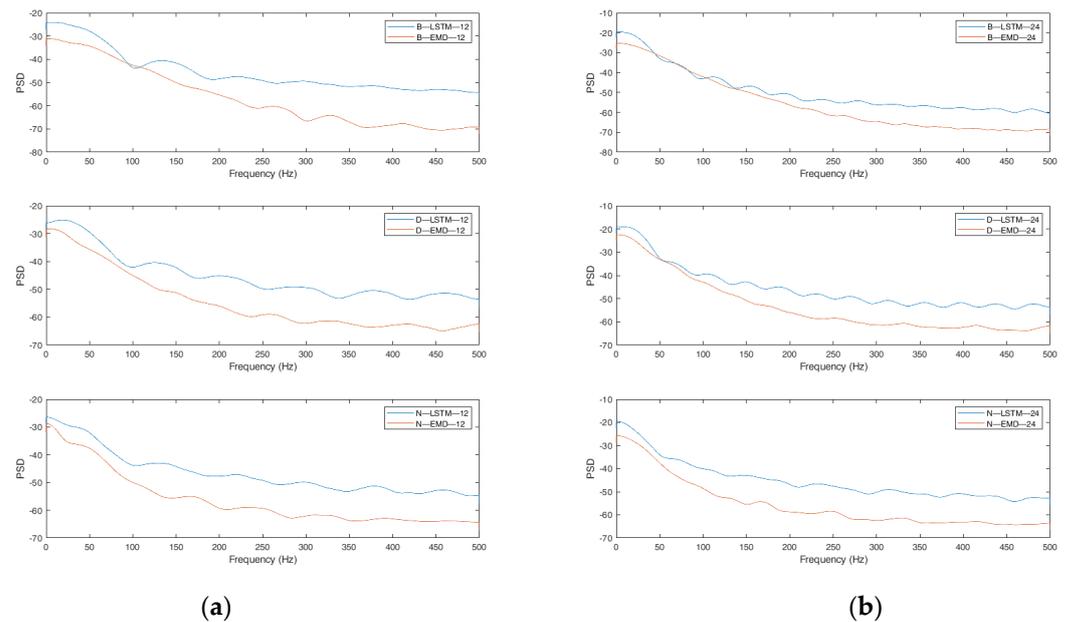
#### 4. Discussion

The combination of EMD and the LSTM network to predict the future 12 and 24 h significant wave heights performed better than the LSTM network alone. To test if the EMD method can effectively correct the prediction results, we first calculated the time series,  $n(t)$ , of relative errors between the predicted values and the real values of the two methods.  $n(t)$  consists of the predicted value minus the true value at each moment. Then, we plotted the power spectrum of  $n(t)$ . The power spectrum is defined as the signal power within a unit frequency band. It represents the change in signal power with frequency, that is, the distribution of signal power in the frequency domain. In Figure 14a, the three subplots are the power spectra of the Bohai Sea (B), the East China Sea (D), and the South China Sea (N) with a forecast time of 12 h  $n(t)$ , respectively. In these plots, the blue line and the orange line represent the power spectrum of  $n(t)$  using the LSTM method and the EMD-LSTM method, respectively. In the power spectrum, Figure 14b is the 24 h power spectrum; the abscissa is the frequency, and the ordinate is the power. Obviously, except for the first subgraph on the right, the power spectra of 12 and 24 h of the EMD-LSTM relative error time series at all three locations show that the power spectra of EMD-LSTM relative error time series were always below those of the LSTM network, within the frequency range of 0–100 Hz. This indicates that the EMD method had an obvious correction effect on wave height prediction at lower frequencies. However, for the high-frequency interval, there was no difference between the two methods from the results shown in Figure 13. Therefore, the power corresponding to each frequency was logarithmically processed; that is,  $PSD = 10 \times \log_{10} PD$ , where  $PD$  represents the power value, and  $PSD$  represents the power value after logarithmic operation. Figure 15 shows the coordinate system after transformation. It is obvious that, in the frequency range of 100–500 Hz, the orange line in the subgraph of different locations of the two methods is always at the bottom; that is, the EMD-LSTM method obviously played a correction role. The EMD method had a

significant correction effect on the prediction of 12 and 24 h significant wave heights in both low-frequency and high-frequency regions. The method of combining EMD and LSTM network decomposes the significant wave height into the IMFs with different frequencies, and then sends them together with U10 and V10 into the LSTM network as input values. It is precisely because of the decomposition into IMFs with different frequencies that the LSTM network could better capture the trend in data change, and thus produce a better forecast effect.



**Figure 14.** Power spectra of 12 and 24 h significant wave heights. The blue line represents the power value and frequency change curve of the relative error of the LSTM method. The orange line represents the power and frequency variation curve of the relative error of the EMD method. (a) represent the 12 h change curves of positions B, D, and N from top to bottom; (b) represent the 24 h change curves of positions B, D, and N from top to bottom.



**Figure 15.** Power spectra of 12 and 24 h significant wave height. (a) represent the 12 h change curves of positions B, D, and N from top to bottom; (b) represent the 24 h change curves of positions B, D, and N from top to bottom.

## 5. Conclusions

The use of machine learning to predict the significant wave height of waves is a novel method. Compared with the traditional numerical model, the machine learning model has the characteristics of a lower computational cost and higher efficiency. As a branch of machine learning, RNNs can process sequence data well and reveal potential connections between data. As the problem of gradient disappearance and gradient explosion occurs in the processing of long time series data in the ordinary RNN, a good solution is to introduce a gating mechanism, thus spawning the LSTM and GRU networks. This study mainly examined the prediction performance of the circulating neural network, LSTM network, and GRU network for 1, 3, 6, 12, and 24 h wave heights in different sea areas. The EMD method was used to correct the phase lag in the prediction results of 12 and 24 h from the LSTM network.

The RNN, LSTM, and GRU networks all performed well in the prediction of the significant wave height, and particularly in the prediction of the significant wave height at 1, 3, and 6 h. The prediction results of the three networks basically accurately captured the trend in data changes. Among these, 1 h prediction results at different locations show that the MAE of the best prediction results of RNN, LSTM, and GRU networks was 0.036, 0.027, and 0.027, respectively. The MAE of the best prediction results of 3 h networks was 0.075, 0.058, and 0.055, and the MAE of 6 h networks was 0.121, 0.107, and 0.104. Overall, the performance of the LSTM and GRU networks was better than that of the RNN network, although no specific conclusion was drawn about the superiority of LSTM and GRU networks. At the same time, we found that the prediction effect of point N was better than that of the other two points. Through analyzing the sample data, the most likely reason for this is that the data set of position N was more diverse. For the 12 and 24 h of significant wave height prediction, the accuracy of the prediction from the three networks obviously increased compared to those of the 1, 3, and 6 h of significant wave height prediction. The GRU network at the N position had the best prediction effect on the significant wave height for 12 h, and the relevant error indicators were 0.195 (MAE), 0.283 (RMSE), and 0.968 (R). The 24 h prediction results further increased significantly, and the worst MAE was 0.421 and the corresponding R was only 0.602. It was found that the prediction results of the three networks at the peak points are ahead or lagging. To correct this phenomenon, the LSTM was modified using the EMD method, and then the SWH of 12 and 24 h was predicted. The forecast results show that the EMD method can effectively improve the forecast accuracy and had a significant correction effect on the phenomenon of forecasting and hysteresis.

In general, the LSTM and GRU networks have a better ability to predict wave characteristics than RNN networks. The prediction performance of the gating-based LSTM and GRU networks is comparable. The EMD-LSTM method has high accuracy for the prediction of waves in a longer time range. Waves are random, so it is challenging to predict them accurately. The machine learning method provides new ideas for ocean wave forecasting and has broad development and application prospects. Much research on the prediction of SWH has taken a univariate as the input feature; that is, the historical wave height is used to predict the significant wave height at the future time. Although there is a strong correlation between the wave height at a later moment and the wave height at an earlier moment, there is no doubt that the wind is crucial to wave formation. Therefore, the historical data of U10, V10, and SWH were studied to predict the significant wave height in the future. The formation of waves has a complex physical mechanism, which is not only affected by wind, but also related to the water depth, sea surface temperature, air humidity, and other climatic factors [2]. Accurately selecting the input parameters for wave prediction remains a challenge. The next step is to compare the effects of different factors on wave height prediction by incorporating other parameters into the input characteristics.

**Author Contributions:** Conceptualization, Z.F. and P.H.; methodology, Z.F.; software, Z.F.; validation, Z.F., S.L., D.M. and P.H.; data curation, S.L.; writing—original draft preparation, Z.F. and P.H.;

writing—review and editing, Z.F., S.L., D.M. and P.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences, grant numbers XDA19060202; the National Natural Science Foundation of China, grant numbers 42076214, 42006027 and U1806227.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data used in this study are available from the ERA5 at <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form> (accessed on 1 November 2021).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Güner, H.A.A.; Yüksel, Y.; Çevik, E.Ö. Estimation of wave parameters based on nearshore wind–wave correlations. *Ocean Eng.* **2013**, *63*, 52–62. [[CrossRef](#)]
- Hashim, R.; Roy, C.; Motamedi, S.; Shamsirband, S.; Petković, D. Selection of climatic parameters affecting wave height prediction using an enhanced Takagi-Sugeno-based fuzzy methodology. *Renew. Sustain. Energy Rev.* **2016**, *60*, 246–257. [[CrossRef](#)]
- Cornejo-Bueno, L.; Nieto-Borge, J.C.; García-Díaz, P.; Rodríguez, G.; Salcedo-Sanz, S. Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm—Extreme Learning Machine approach. *Renew. Energy* **2016**, *97*, 380–389. [[CrossRef](#)]
- Mahjoobi, J.; Mosabbeb, E.A. Prediction of significant wave height using regressive support vector machines. *Ocean Eng.* **2009**, *36*, 339–347. [[CrossRef](#)]
- Abhigna, P.; Jerritta, S.; Srinivasan, R.; Rajendran, V. Analysis of feed forward and recurrent neural networks in predicting the significant wave height at the moored buoys in Bay of Bengal. In Proceedings of the 2017 International Conference on Communication and Signal Processing (ICCSPP), Chennai, India, 6–8 April 2017; pp. 1856–1860.
- Makarynsky, O.; Pires-Silva, A.A.; Makarynska, D.; Ventura-Soares, C. Artificial neural networks in wave predictions at the west coast of Portugal. *Comput. Geosci.* **2005**, *31*, 415–424. [[CrossRef](#)]
- Bengio, Y. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **2002**, *5*, 157–166. [[CrossRef](#)]
- Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
- Theodoropoulos, P.; Spandonidis, C.C.; Themelis, N.; Giordamli, C.; Fassois, S. Evaluation of Different Deep-Learning Models for the Prediction of a Ship’s Propulsion Power. *J. Mar. Sci. Eng.* **2021**, *9*, 116. [[CrossRef](#)]
- Gao, S.; Zhao, P.; Pan, B.; Li, Y.; Zhou, M.; Xu, J.; Zhong, S.; Shi, Z. A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network. *Acta Oceanol. Sin.* **2018**, *37*, 12–16. [[CrossRef](#)]
- Gao, S.; Huang, J.; Li, Y.; Liu, G.; Bi, F.; Bai, Z. A forecasting model for wave heights based on a long short-term memory neural network. *Acta Oceanol. Sin.* **2021**, *40*, 62–69. [[CrossRef](#)]
- Fan, S.; Xiao, N.; Dong, S. A novel model to predict significant wave height based on long short-term memory network. *Ocean Eng.* **2020**, *205*, 107298. [[CrossRef](#)]
- Zhang, D.; Kabuka, M.R. Combining Weather Condition Data to Predict Traffic Flow: A GRU Based Deep Learning Approach. *IET Intell. Transp. Syst.* **2018**, *12*, 578–585. [[CrossRef](#)]
- Dai, G.; Ma, C.; Xu, X. Short-term Traffic Flow Prediction Method for Urban Road Sections Based on Space-time Analysis and GRU. *IEEE Access* **2019**, *7*, 143025–143035. [[CrossRef](#)]
- Liu, H.; Mi, X.; Li, Y.; Duan, Z.; Xu, Y. Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional Gated Recurrent Unit network and Support Vector Regression. *Renew. Energy* **2019**, *143*, 842–854. [[CrossRef](#)]
- Wang, J.; Wang, Y.; Yang, J. Forecasting of significant wave height based on gated recurrent unit network in the Taiwan strait and its adjacent waters. *Water* **2021**, *13*, 86. [[CrossRef](#)]
- Sias, S.G. Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Phys. Chem. Earth Parts A/B/C* **2006**, *31*, 1164–1171.
- Deka, P.C.; Prahlada, R. Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. *Ocean Eng.* **2012**, *43*, 32–42. [[CrossRef](#)]
- Oezger, M. Significant wave height forecasting using wavelet fuzzy logic approach. *Ocean Eng.* **2010**, *37*, 1443–1451. [[CrossRef](#)]
- Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]

21. Rios, R.A.; De Mello, R.F. Improving time series modeling by decomposing and analyzing stochastic and deterministic influences. *Signal Process.* **2013**, *93*, 3001–3013. [[CrossRef](#)]
22. Duan, W.Y.; Han, Y.; Huang, L.M.; Zhao, B.B.; Wang, M.H. A hybrid EMD-SVR model for the short-term prediction of significant wave height. *Ocean Eng.* **2016**, *124*, 54–73. [[CrossRef](#)]
23. Huang, W.; Dong, S. Improved short-term prediction of significant wave height by decomposing deterministic and stochastic components. *Renew. Energy* **2021**, *177*, 743–758. [[CrossRef](#)]
24. Li, Z.; Li, S.; Hou, Y.; Mo, D.; Li, J.; Yin, B. Typhoon-induced wind waves in the northern East China Sea during two typhoon events: The impact of wind field and wave-current interaction. *J. Oceanol. Limnol.* **2022**, *40*, 934–949. [[CrossRef](#)]
25. Li, A.; Liu, Z.; Hong, X.; Hou, Y.; Guan, S. Applicability of the ERA5 reanalysis data to China adjacent Sea under typhoon condition. *Mar. Sci.* **2021**, *45*, 10, (In Chinese with English abstract).
26. Mahjoobi, J.; Etemad-Shahidi, A.; Kazeminezhad, M.H. Hindcasting of wave parameters using different soft computing methods. *Appl. Ocean Res.* **2008**, *30*, 28–36. [[CrossRef](#)]
27. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. In Proceedings of the SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, 25 October 2014; pp. 103–111.
28. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv* **2014**, arXiv:1412.3555.