

A forecasting model for wave heights based on a long short-term memory neural network

Song Gao^{1,2}, Juan Huang^{1,2}, Yaru Li^{1,2*}, Guiyan Liu^{1,2}, Fan Bi^{1,2}, Zhipeng Bai³

¹North China Sea Marine Forecasting Center of State Oceanic Administration, Qingdao 266061, China

²Shandong Provincial Key Laboratory of Marine Ecological Environment and Disaster Prevention and Mitigation, Qingdao 266061, China

³Mailbox 5111, Beijing 100094, China

Received 12 April 2020; accepted 2 June 2020

© Chinese Society for Oceanography and Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

To explore new operational forecasting methods of waves, a forecasting model for wave heights at three stations in the Bohai Sea has been developed. This model is based on long short-term memory (LSTM) neural network with sea surface wind and wave heights as training samples. The prediction performance of the model is evaluated, and the error analysis shows that when using the same set of numerically predicted sea surface wind as input, the prediction error produced by the proposed LSTM model at Sta. N01 is 20%, 18% and 23% lower than the conventional numerical wave models in terms of the total root mean square error (RMSE), scatter index (SI) and mean absolute error (MAE), respectively. Particularly, for significant wave height in the range of 3–5 m, the prediction accuracy of the LSTM model is improved the most remarkably, with RMSE, SI and MAE all decreasing by 24%. It is also evident that the numbers of hidden neurons, the numbers of buoys used and the time length of training samples all have impact on the prediction accuracy. However, the prediction does not necessarily improve with the increase of number of hidden neurons or number of buoys used. The experiment trained by data with the longest time length is found to perform the best overall compared to other experiments with a shorter time length for training. Overall, long short-term memory neural network was proved to be a very promising method for future development and applications in wave forecasting.

Key words: long short-term memory, marine forecast, neural network, significant wave height

Citation: Gao Song, Huang Juan, Li Yaru, Liu Guiyan, Bi Fan, Bai Zhipeng. 2021. A forecasting model for wave heights based on a long short-term memory neural network. *Acta Oceanologica Sinica*, 40(1): 62–69, doi: 10.1007/s13131-020-1680-3

1 Introduction

Wave disasters are the most frequent marine disasters in the world. Accurate predictions of ocean waves can effectively improve the safety of marine activities and the efficiency of ocean operations, as well as reducing marine accidents. Wave forecasting is therefore regarded as important and fundamental by most marine forecasting institutions across the world. At present, the most widely adopted forecasting method in the field of operational wave prediction is the conventional numerical models based on energy balance equation (EBE) theory (Duan et al., 2016). However, as the complex nonlinear physical processes of waves and their mechanisms are not clear, the ability of numerical models to resolve ocean waves with a higher accuracy is still largely limited.

Artificial neural networks are one of the nonlinear fitting methods in statistical methods. This method is a theoretical mathematical model of human brains and their activities, and it is a large-scale nonlinear adaptive system (Jiang, 2001). Research on the application of artificial neural networks in the field of meteorology and ocean forecasting has been carried out in the past decades (Tissot et al., 2001; Zhang et al., 2006; Liu and Wang, 2008; Chaudhari et al., 2008; Kuang et al., 2016; Deshmukh et al., 2016). However, due to its limited forecasting ability, it has not been widely applied in operational forecasting institutions. In re-

cent years, machine learning, especially deep learning technologies based on big data have been developed very rapidly (Hinton and Salakhutdinov, 2006), especially those based on the recursive neural networks (RNN) (Lipton, 2015), long short-term memory (LSTM) neural networks (Hochreiter and Schmidhuber, 1997) and other algorithms, which provide an effective solution for prediction and regression problems. LeCun et al. (2015) systematically expounded the connotation, principles, methods and future development of deep learning. At present, deep learning is widely used in speech recognition, visual recognition, target detection, finance, transportation and other industries, which has significantly promoted the technical development in the aforementioned fields. Deep learning methods have been applied preliminarily in meteorological and marine forecasting (Filippo et al., 2012; Gao et al., 2018). Shi et al. (2015) used the convolutional LSTM method to better capture the spatial and temporal correlation of precipitation in short-term precipitation predictions, the results of which are better than the operational nowcasting results by ROVER. Using self-organizing map networks, Vilibić et al. (2016) built a real-time current forecast system based on ground wave radar and meteorological numerical forecasts. By comparing with the regional ocean model ROMS, they found the accuracy of their method was comparable or even slightly better

Foundation item: The National Key R&D Program of China under contract No. 2016YFC1402103.

*Corresponding author, E-mail: liyaru@ncs.mnr.gov.cn

than the model results based on the ocean dynamic equations, which opened up new ideas on the quantitative prediction of coastal marine forecasting factors based on artificial intelligence. Gao et al. (2018) predicted typhoon paths using LSTM methods, and believed that the model established had certain ability to predict the short-term path of typhoons. Other recent works using neural networks include hybrid deep learning and empirical models developed for time series applications (Yang and Chen, 2019), machine learning models for wave forecast (James et al., 2018), sequential learning neural networks for ocean wave predictions (Kumar et al., 2017), hybrid extreme learning machine models for wind speed forecast (Zhang et al., 2019), etc. However, to the best knowledge of the authors, applying the LSTM neural network algorithm for the purpose of operational wave height forecasts is still rarely reported. In this paper, a LSTM neural network model is trained by the sea surface wind and significant wave height data continuously observed by three anchored buoys in the Bohai Sea. Forced by numerically predicted sea surface wind, the prediction results of significant wave heights by the LSTM model and the wave model SWAN are compared to measurements, respectively, to evaluate the performance of the LSMT model over the conventional numerical model.

2 Data

2.1 Observations

The observation data of three ocean buoys in the Bohai Sea are used for the training and analysis of the prediction model. The synchronous observation elements are hourly-recorded significant wave height, and the wind speed and direction at 10 m above the sea surface. The data duration is from January 1, 2017 to March 31, 2018. Station N01 is located off the Laizhou Bay, Sta. N02 is approximately in the middle of the Bohai Sea, and Sta. N03 is to the northeast of Sta. N02. The locations of the buoys are shown in Fig. 1.

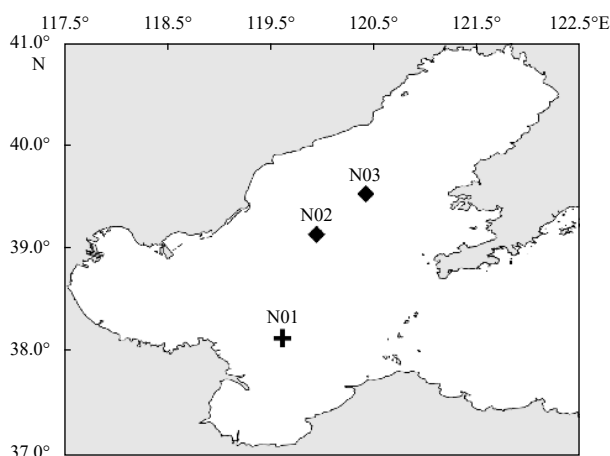


Fig. 1. Locations of the buoys.

2.2 Numerical forecast data of sea surface wind

The sea surface wind data are the results of the operational meteorological forecast model NMFC-WRF of the North China Sea Marine Forecasting Center of the State Oceanic Administration (Wu et al., 2015). The model is established using the weather and research forecasting (WRF) model (Skamarock et al., 2005), and the calculation domain covers the Bohai Sea and the Yellow

Sea (30°–41°N, 117°–130°E), and the mesh resolution is 10 km×10 km. The global forecasting system (GFS) data are used as the initial and external forcing fields. The model forecasts the next 120 h every day. In this paper, the 24-h forecast of wind speed and wind directions from January 1, 2018 to March 31, 2018 are used as the sample data.

2.3 Construction of sample data

The data in this study uses different units, which has a great impact on the modeling performance of LSTM algorithms. Therefore, before the sample sequences of predictors are input into the network, the ensemble samples need to be non-dimensionalized and standardized. The non-dimensionalized and normalized calculation method used in this paper is

$$x = \frac{X - X_{\text{mean}}}{\sigma_X},$$

where X is the sample, X_{mean} is the averaged value of the sample, σ_X is the sample variance, x is the dimensionless and standardized series. The calculation of X_{mean} and σ_X is performed over the training period.

The duration of the training sample is from January 1, 2017 to December 31, 2017, and that of the forecast sample is from January 1, 2018 to March 31, 2018. After the LSTM network calculation, the dimension and magnitude of the sample are retrieved by the corresponding reversed calculation.

3 Prediction model

3.1 SWAN

The operational wave forecast model NMFC-SWAN of the North China Sea Marine Forecasting Center is based on the third-generation wave model SWAN (Booij et al., 1999). Details regarding the evolution equation of the wave spectrum used in the model can be found in Hasselmann et al. (1973).

In cases of shallow water, the wave energy density changes due to the wind input, the triad wave-wave interaction, the quadruplet wave-wave interaction, the whitecapping, bottom friction, and depth-induced wave breaking are considered.

In the NMFC-SWAN model adopted here, the calculation domain covers the Bohai Sea and the Yellow Sea (30°–41°N, 117°–130°E), and the mesh resolution is 2 km×2 km. Results by NMFC-WRF are used as the meteorological forcing field in the model.

3.2 LSTM model

Feedforward neural networks (FNNs) assumes that each sample is independent of each other, i.e., it is irrelevant in time. LSTM is a cyclic neural network, which processes each element of the input sequence one by one, and transmits the information between the sequences selectively through the network structure to maintain the correlation between the sequence elements. The LSTM algorithm is developed based on the widely used algorithm RNN. Traditional RNN network can only directly affect its adjacent elements. The recurrent and error back propagation processes are retained in LSTM, with long and short-term memory cells used to replace the hidden neurons in a conventional RNN. With this introduction, it can save information in any length of time, and the vanishing gradient problem of RNN is solved.

Waves are generated by the wind above the ocean, including wind waves, swells, and nearshore waves. Sea surface wind plays

a leading role in the formation of waves. Therefore, sea surface wind speed and direction are considered as the predictors of the model, and the significant wave height is the dependent variable.

For predictors and dependent variables which last for continuous t time steps, they can be represented by sequences $\left(\left[\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)} \right]^T, \left[\mathbf{x}_1^{(2)}, \mathbf{x}_2^{(2)} \right]^T, \dots, \left[\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)} \right]^T \right)$ and $\left(\left[\mathbf{y}^{(1)} \right]^T, \left[\mathbf{y}^{(2)} \right]^T, \dots, \left[\mathbf{y}^{(t)} \right]^T \right)$, respectively, where $\mathbf{x}_1^{(t)}$ is sea surface wind speed, $\mathbf{x}_2^{(t)}$ is sea surface wind direction, and $\mathbf{y}^{(t)}$ is the significant wave height. The superscript t represents its index in the series, and each two neighboring elements have a time difference of Δt . One input sequence (predictors) and the corresponding output series (dependent variables) form a training sample.

An LSTM neural network for the prediction of significant wave heights is designed in this way with three layers, i.e., an input layer, a hidden layer, and an output layer. There are $2t$ neurons on the input layer, i.e., the dimension of each element in the input sequence is $2t$. The hidden layer consists of long short-term memory cells, the number of which can be adjusted according to the training performance. An illustration of a memory cell in LSTM is shown in Fig. 2. Here σ represents the sigmoid function, Π represents multiplication, the subscript c means that it is a single structure and all the solid arrows mean that the connection weight is one. s_c is a memory cell, which is a linear element used to store information to guarantee that information can be stored for a long time to retain the correlation among elements in a sequence. g_c is the input node, which denotes the comprehensive interaction of the input at time step t and the information of previous network status. Its value can be passed on to a memory cell through the control of an input gate i_c . If W is the weight, b is the threshold, then $g_c^{(t)} = \sigma(W_{hx}\mathbf{x}^{(t)} + W_{hh}\mathbf{h}^{(t-1)} + b)$. i_c is the input gate, which receive the input at time step t and the network status information at previous time steps, and pass the input value of node g_c into a memory cell s_c after the control of the sigmoid function. f_c is the forget gate, which determines whether the value of s_c is stored or not: if the weight is one, it is stored as it was, and if it is zero, it is cleared. o_c is the output gate, which receives the input at the time step t and the network status information at previous time step. It controls the output of s_c after the

sigmoid function. v_c is the output value. There are t neurons on the output layer, i.e., the dimension of each element in the output series is t . To denote the values of each layer using vectors \mathbf{x} and \mathbf{h} , then

$$\mathbf{g}^{(t)} = \tanh \left(W_{gx}\mathbf{x}^{(t)} + W_{gh}\mathbf{h}^{(t-1)} + b_g \right),$$

$$\mathbf{i}^{(t)} = \sigma \left(W_{ix}\mathbf{x}^{(t)} + W_{ih}\mathbf{h}^{(t-1)} + b_i \right),$$

$$\mathbf{f}^{(t)} = \sigma \left(W_{fx}\mathbf{x}^{(t)} + W_{fh}\mathbf{h}^{(t-1)} + b_f \right),$$

$$\mathbf{o}^{(t)} = \sigma \left(W_{ox}\mathbf{x}^{(t)} + W_{oh}\mathbf{h}^{(t-1)} + b_o \right),$$

$$\mathbf{s}^{(t)} = \mathbf{g}^{(t)} \cdot \mathbf{i}^{(t)} + \mathbf{s}^{(t-1)} \cdot \mathbf{f}^{(t)},$$

$$\mathbf{h}^{(t)} = \tanh(\mathbf{s}^{(t)}) \odot \mathbf{o}^{(t)},$$

where \odot is dot product (Gao et al., 2018).

All the network parameters are initialized as random numbers between 0 and 1 to train the network, and they are optimized using the training samples. For each training sample, each element in the series is read one by one by the LSTM neural network. An output vector is derived after the hidden layer and the output layer, which is then compared to labels and the errors are propagated backward by the BPTT algorithm. The forecast process follows the same way as the training process. In this paper, a one-direction network is employed in the LSTM model, with Adam (adaptive moment estimation) optimizer. The maximum number of epochs is 250, the gradient threshold is 1, the L_2 regularization (weight decay) factor is 0.0001. The initial learning rate is 0.005 and the learning rate schedule is piecewise. The learning rate drop period is 125, and learning rate drop factor is 0.2. The model is realized with MATLAB 2017b.

In order to evaluate the prediction capability of the model for different wave heights (h_s), the error analysis of wave heights are carried out in the following four ranges respectively: $0.3 \text{ m} < h_s \leq 1 \text{ m}$, $1 \text{ m} < h_s \leq 2 \text{ m}$, $2 \text{ m} < h_s \leq 3 \text{ m}$, $3 \text{ m} < h_s \leq 5 \text{ m}$.

In this paper, data at the three buoy stations in the Bohai Sea are used as examples for the LSTM model construction, and also in the feasibility experiment and sensitivity experiments. Firstly, the surface wind speed and significant wave height data at Station N01 is used in the feasibility experiment (Exp. 1-1) for model training, validating the feasibility of the LSTM model in predicting significant wave heights. The FNN model (Exp. 1-2) and Support vector machine regression (SVR) model (Exp. 1-3) are chosen to compare with LSTM. Then sensitivity experiments are carried out, including: Exp. 2, which examines the predictors' impact on prediction accuracy, and Exp. 3, which tests the model performance of prediction at Sta. N01 with varying numbers of stations being used. Additional experiments to Exp. 3 are also performed, where the significant wave height at Stas N02 and N03 are predicted respectively using the data from all the three buoys. After determining the input factors into the LSTM model, the model parameter tuning experiments are carried out, including Exp. 4 where the implicit neuron number for optimal outcome is tested and Exp. 5 where a proper time length of samples is examined to determine the optimal network structure of the

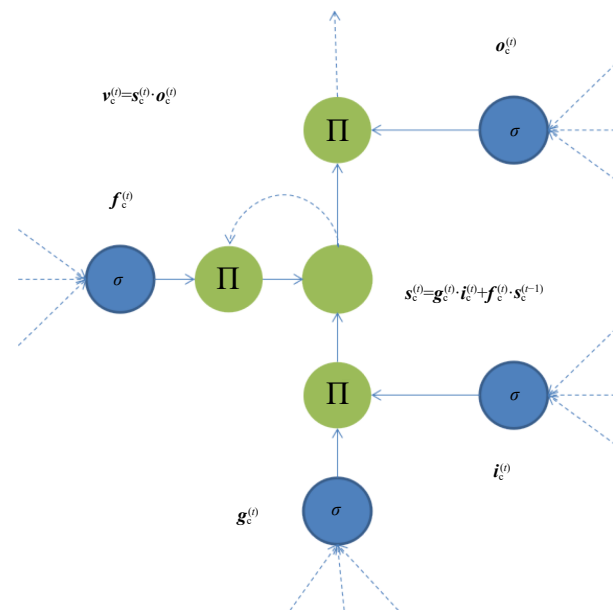


Fig. 2. A memory cell of LSTM.

LSTM model (Table 1). Please note that for all the experiments listed in Table 1, the dependent variable is significant wave height, the training data used are from 2017, and the prediction period is January 1, 2018 to March 31, 2018.

4 Results and discussion

4.1 Feasibility analysis

Figure 3 is the comparison between the model results and the measured values of the significant wave height in Exp. 1. It can be seen from the figure that the trending of the modeled results match well with the observed significant wave heights, and the correlation coefficient of the two series is 0.94, which is highly correlated. The total root mean square error (RMSE) is 0.34 m and total scatter index (SI) is 0.29 (Table 2). It can be seen that the prediction results can reflect the variation trending and amplitude of the observed significant wave height at Sta. N01, and the model performance shows desirable forecasting capability. In addition, the results of LSTM are significantly better than those of FNN and SVR models. The RMSE of FNN and SVR are approximately 130%–160% and 140%–400% times that of LSTM, respectively.

4.2 Analysis of the effects of wind direction

In Exp. 1, only wind speed is used as the input predictor, and in Exp. 2 both wind speed and wind direction is used as the predictors to conduct the LSTM model training, and the rest conditions of the model remain the same. As shown in Table 2, the sig-

nificant wave height prediction results in Exp. 2 have evident improvement regarding the evaluation indicators RMSE, SI, mean absolute error (MAE) and the correlation coefficient (r). The total RMSE decreases by 15% and the total MAE decreases by 20% (Table 2). Therefore, it is evident that wind direction plays a positive role in improving the forecasting results of the LSTM model. It is also a rational inference from the physical point of view that both the wind speed and direction are important to the development of waves.

4.3 Analysis of the influence of station number

Observation data at Sta. N01 are used as training sample data in Exp. 2 to train the LSTM model. On the basis of Exp. 2, the observation data at Stas N02 and N03 are added in Exp. 3 to train the LSTM model, and the other training conditions stay the same. Thus, the training sample data in Exp. 3 are wind speed, wind direction and significant wave height data at Stas N01, N02 and N03. As shown in Table 2, compared with the results in Exp. 2, the model performance indicated by the evaluation indicators RMSE, SI, MAE and the correlation coefficient r in Exp. 3 have been significantly improved. Compared with Exp. 2 (Exp. 1), the total RMSE in Exp. 3 decreases by 31% (41%), total SI decreases by 29% (41%), and the total MAE decreases by 25% (40%) (Table 2). In addition, the significant wave heights at Stas N02 and N03 are predicted respectively using the same methods and settings. The prediction result with wind speed and direction from the three buoys as the input is slightly better than using data from only one station as the input (Table 3), and the total RMSE of Sta. N02 (Sta.

Table 1. Key parameters in the training experiments

Experiment	Predictor	Station	Training data duration	Model	Number of hidden neurons
Exp. 1-1	w_spd	N01	Jan. 1 to Dec. 31	LSTM	200
Exp. 1-2	w_spd	N01	Jan. 1 to Dec. 31	FNN	10
Exp. 1-3	w_spd	N01	Jan. 1 to Dec. 31	SVR	not-applicable
Exp. 2	w_spd+w_dir	N01	Jan. 1 to Dec. 31	LSTM	200
Exp. 3	w_spd+w_dir	N01, N02, N03	Jan. 1 to Dec. 31	LSTM	200
Exp. 4-1	w_spd+w_dir	N01, N02, N03	Jan. 1 to Dec. 31	LSTM	300
Exp. 4-2	w_spd+w_dir	N01, N02, N03	Jan. 1 to Dec. 31	LSTM	250
Exp. 4-3	w_spd+w_dir	N01, N02, N03	Jan. 1 to Dec. 31	LSTM	150
Exp. 4-4	w_spd+w_dir	N01, N02, N03	Jan. 1 to Dec. 31	LSTM	100
Exp. 5-1	w_spd+w_dir	N01, N02, N03	Apr. 1 to Dec. 31	LSTM	200
Exp. 5-2	w_spd+w_dir	N01, N02, N03	Jul. 1 to Dec. 31	LSTM	200
Exp. 5-3	w_spd+w_dir	N01, N02, N03	Oct. 1 to Dec. 31	LSTM	200

Note: w_spd means wind speed, and w_dir means wind direction.

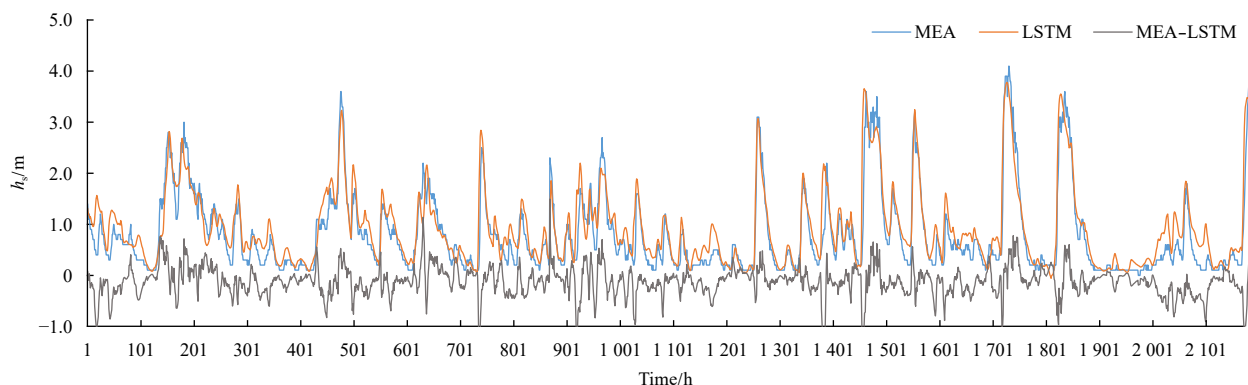


Fig. 3. Comparison between predicted and measured significant wave heights at Sta. N01 in Exp. 1 (blue line: measurements; red line: LSTM model; grey line: the difference between measurement and LSTM model result).

Table 2. Evaluation index of the prediction error in each experiment (Sta. N01)

Wave height	Indicator	Exp. 1-1	Exp. 1-2	Exp. 1-3	Exp. 2	Exp. 3	Exp. 4-1	Exp. 4-2	Exp. 4-3	Exp. 4-4	Exp. 5-1	Exp. 5-2	Exp. 5-3
0.3 m $\leq h_s \leq$ 5 m (SN: 1 372)	RMSE/m	0.34	0.50	0.76	0.29	0.20	0.21	0.22	0.21	0.21	0.23	0.23	0.23
	SI	0.29	0.43	0.65	0.24	0.17	0.18	0.19	0.18	0.18	0.21	0.21	0.21
	MAE/m	0.25	0.39	0.55	0.20	0.15	0.16	0.16	0.16	0.16	0.18	0.17	0.17
	r	0.94	0.86	0.82	0.95	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.97
0.3 m $< h_s \leq$ 1 m (SN: 817)	RMSE/m	0.23	0.34	0.32	0.17	0.14	0.14	0.14	0.14	0.17	0.17	0.17	0.18
	SI	0.36	0.52	0.50	0.27	0.22	0.22	0.23	0.22	0.26	0.26	0.27	0.29
	MAE/m	0.18	0.27	0.25	0.13	0.11	0.11	0.11	0.11	0.12	0.12	0.12	0.13
	r	0.72	0.42	0.42	0.81	0.86	0.88	0.86	0.87	0.87	0.85	0.84	0.84
1 m $< h_s \leq$ 2 m (SN: 353)	RMSE/m	0.45	0.67	0.73	0.38	0.24	0.27	0.27	0.26	0.24	0.27	0.31	0.29
	SI	0.31	0.47	0.51	0.27	0.17	0.19	0.19	0.19	0.17	0.19	0.22	0.20
	MAE/m	0.38	0.59	0.68	0.31	0.19	0.22	0.21	0.21	0.18	0.22	0.24	0.23
	r	0.51	0.41	0.45	0.55	0.74	0.68	0.66	0.75	0.69	0.67	0.66	0.60
2 m $< h_s \leq$ 3 m (SN: 133)	RMSE/m	0.47	0.62	1.37	0.46	0.27	0.31	0.28	0.26	0.24	0.33	0.32	0.30
	SI	0.19	0.25	0.54	0.18	0.11	0.12	0.11	0.10	0.10	0.13	0.13	0.12
	MAE/m	0.38	0.50	1.34	0.36	0.20	0.22	0.22	0.19	0.19	0.24	0.25	0.22
	r	0.44	0.38	0.38	0.64	0.75	0.67	0.70	0.71	0.75	0.73	0.67	0.67
3 m $< h_s \leq$ 5 m (SN: 69)	RMSE/m	0.49	0.80	1.99	0.56	0.33	0.25	0.35	0.36	0.32	0.40	0.23	0.27
	SI	0.14	0.24	0.59	0.17	0.10	0.07	0.10	0.11	0.09	0.12	0.07	0.08
	MAE/m	0.35	0.66	1.97	0.45	0.27	0.20	0.30	0.30	0.27	0.32	0.19	0.22
	r	0.55	-0.02	0.47	0.41	0.76	0.80	0.76	0.73	0.73	0.74	0.64	0.77

Note: SN is the number of the samples.

Table 3. Evaluation index of the prediction error using input data from one station and three stations, respectively

Wave height	Indicator	Model					
		Input data from one station			Input data from three stations		
		Sta. N01 (Exp. 2)	Sta. N02	Sta. N03	Sta. N01 (Exp. 3)	Sta. N02	Sta. N03
0.3 m $\leq h_s \leq$ 5 m (SN: N01: 1 372; N02: 1 732; N03: 1 732)	RMSE/m	0.29	0.24	0.25	0.20	0.23	0.22
	SI	0.24	0.21	0.23	0.17	0.20	0.21
	MAE/m	0.20	0.19	0.19	0.15	0.18	0.16
	r	0.95	0.95	0.95	0.98	0.95	0.95
0.3 m $\leq h_s \leq$ 1 m (SN: N01: 817; N02: 1 066; N03: 1 096)	RMSE/m	0.17	0.24	0.24	0.14	0.20	0.17
	SI	0.27	0.35	0.37	0.22	0.29	0.27
	MAE/m	0.13	0.18	0.18	0.11	0.15	0.14
	r	0.81	0.73	0.74	0.86	0.78	0.76
1 m $< h_s \leq$ 2 m (SN: N01: 353; N02: 503; N03: 449)	RMSE/m	0.38	0.24	0.27	0.24	0.24	0.23
	SI	0.27	0.16	0.18	0.17	0.16	0.15
	MAE/m	0.31	0.18	0.20	0.19	0.19	0.17
	r	0.55	0.72	0.69	0.74	0.69	0.71
2 m $< h_s \leq$ 3 m (SN: N01: 133; N02: 150; N03: 137)	RMSE/m	0.46	0.26	0.29	0.27	0.30	0.38
	SI	0.18	0.10	0.11	0.11	0.12	0.15
	MAE/m	0.36	0.20	0.22	0.20	0.25	0.30
	r	0.64	0.68	0.75	0.75	0.73	0.51
3 m $< h_s \leq$ 5 m (SN: N01: 69; N02: 55; N03: 50)	RMSE/m	0.56	0.25	0.38	0.33	0.44	0.47
	SI	0.17	0.07	0.29	0.10	0.13	0.14
	MAE/m	0.45	0.22	0.08	0.27	0.34	0.38
	r	0.41	0.60	0.26	0.76	0.37	0.33

N03) prediction is reduced by approximately 4% (12%). For the RMSE, the improvement in RMSE for significant wave height in the range of 0.3–1 m is more evident, which is reduced by approximately 17% to 29%; the prediction accuracy for significant wave height in the range of 2–3 m and 3–5 m is decreased, with the RMSE for waves in the range of 2–3 m at Sta. N02 (Sta. N03) increasing from 26 cm (29 cm) to about 30 cm (38 cm), and the RMSE for waves in the range of 3–5 m at Sta. N02 (Sta. N03) increasing from 25 cm (38 cm) to about 44 cm (47 cm). It

can be seen that an increase in the number of observation stations used as input data has a positive effect on the improvement of the prediction accuracy of the LSTM model. Increasing the number of stations as input, the prediction accuracy of significant wave height between 0.3 m and 2 m is improved at all the three stations. When the significant wave height is between 2 m and 5 m, the prediction accuracy of Sta. N01 is improved, while those of Stas N02 and N03 are worsened. In addition to the number of data input, it is also related to the locations of the buoys.

Since the prediction period is in winter when northerly wind prevails, wind information at Stas N02 and N03 impacts on the waves at Sta. N01. However, the wind information at Sta. N01 in the south is less relevant to the wave conditions at Stas N02 and N03. Therefore, for wave height greater than 2 m which is more sensitive to wind, the addition of less relevant or irrelevant wind information at other locations decreases the prediction accuracy. Thus, in addition to the number of buoys used as input data, the relative location of wind information to be fed into the model needs careful consideration for the improvement of prediction accuracy.

4.4 Analysis of the influence of hidden neuron number and sample time length

On the basis of Exp. 3, an experiment on the number of hidden neurons and an experiment on the time length of samples are designed and called Exp. 4 and Exp. 5 respectively. Under the condition of Exp. 3, experiments with 100 hidden neurons (Exp. 4-4), 150 neurons (Exp. 4-3), 250 neurons (Exp. 4-2) and 300 neurons (Exp. 4-1) respectively are carried out. As shown in Table 2, when there are 200 hidden neurons (Exp. 3), the overall performance is the best, outperforming other experiments with either more or fewer hidden neurons. This is also true for wave heights under 2 m. However, for the wave heights in the range of 2–3 m, Exp. 4-4 performs the best with 100 hidden neurons; for the wave heights in the range of 3–5 m, Exp. 4-1 with 300 hidden neurons is the most accurate whilst Exp. 4-4 is also better than Exp. 3. It is also worth mentioning that the percentage of large waves greater than 2 m only accounts for 15% in the whole dataset. Therefore, from the overall point of view, a larger number of hidden neurons does not necessarily guarantee a better prediction; and for different wave height ranges, the number of hidden neurons required for an accurate prediction may vary, not necessarily consistent with the optimal hidden neuron numbers for the same particular location under any wave conditions. This finding suggests the necessity and feasibility to predict different wave conditions using various model setting, depending on the study purpose. The underlying mechanism would also require future studies.

Under the same condition as Exp. 3, experiments with training sample time length being 3-month (Exp. 5-3), 6-month (Exp. 5-2) and 9-month (Exp. 5-1) long are conducted. The prediction results show that the model performance peaks when the training sample length is 12-month, however, the prediction error of this configuration is not the lowest in all levels of the significant wave height ranges (Table 2). For example, for the wave height in the range of 3–5 m, Exp. 5-2 with the training data from July to December 2017 outperforms other experiments. It is understandable that for large waves which relies more on the wind field, the percentage of relevant or similar wind information involved in the training sample is accountable for the prediction accuracy. Whereas for the overall wave prediction, training the model with a full year cycle (Exp. 3) has its advantages.

4.5 Model predictions

The network structure and parameter setting of Exp. 3, which has the best performance in all the experiments, are adopted in the prediction model LSTM-WAVE, and the wind speed and wind direction at the three buoy stations predicted by NMFC-WRF are taken as input to predict the significant wave height at Sta. N01. The results are then compared with the predicted significant wave height by NMFC-SWAN. The prediction period is from January 1, 2018 to March 31, 2018 with hourly output. As shown in Fig. 4, the results of LSTM-WAVE and NMFC-SWAN agree with

the development of significant wave height observed by the buoy at Sta. N01. The errors are shown in Table 4. For Sta. N01, it can be seen that the prediction error of the LSTM-WAVE model is superior to that of NMFC-SWAN in terms of both the overall behavior and the performance of the four evaluation indicators. The prediction error of the LSTM-WAVE model is 20%, 18% and 23% lower than the total RMSE, SI and MAE of the NMFC-SWAN prediction, respectively. Particularly, when the significant wave height is within the range of 3–5 m, the prediction accuracy of the LSTM-WAVE model is improved the most remarkably, with RMSE, SI and MAE all decreasing by 24%. The LSTM-WAVE results at Stas N02 and N03 are slightly better than those of NMFC-SWAN, with the total RMSE of the Sta. N02 (Sta. N03) reduced by approximately 7% (3%). However, the RMSE for significant wave height between 1 m and 3 m at Sta. N02 is reduced by approximately 26%, the RMSE for 0.3–1 m and 3–5 m is increased by approximately 7% and 30% respectively. For Sta. N03, the RMSE for significant wave height in the range of 1–5 m, and 0.3–1 m is increased by 15% to 23%, and 24%, respectively.

5 Conclusions

A LSTM model WAVE-LSTM is established for the significant wave height prediction, and the work shows conclusions as follows.

(1) The overall prediction accuracy of significant wave height by the WAVE-LSTM model is better than the conventional numerical model NMFC-SWAN, especially for the wave prediction with significant wave height in the range of 1–3 m. For significant wave height between 0.3 m and 1 m, the WAVE-LSTM prediction at Sta. N01 is better than the NMFC-SWAN prediction, while the contrary is true for Stas N02 and N03. For significant wave height in the range of 3–5 m, the WAVE-LSTM prediction at Stas N01 and N03 is better than the NMFC-SWAN prediction, while it is the opposite at Sta. N02.

(2) The sea surface wind speed and wind direction are the dominant factors for the prediction of significant wave height. Wind speed is a major factor for wave height prediction, and the inclusion of wind direction in the prediction model leads to the decrease of prediction error by 15%.

(3) It can be seen that a significant increase in the number of observation stations used as input data has a positive effect on the improvement of the prediction accuracy of the LSTM model. The increase of the input station numbers results in an improvement of the prediction accuracy at the three stations when the significant wave height is in the range of 0.3–2 m. For significant wave height in the range of 2–5 m, the prediction accuracy at the Sta. N01 station is improved, while those at Stas N02 and N03 stations are declined.

(4) In the training process of the LSTM network, the number of hidden neurons and the time length of training samples have influences on the prediction accuracy. It is necessary to determine the optimal network structure by testing various experimental settings.

(5) The results of LSTM model are significantly better than those of FNN and SVR models.

It can be seen that the significant wave height prediction technology based on LSTM network can sufficiently exploit the important information of sea surface wind and significant wave height, based on which prediction models can be established, and operational applications can be realized. This method has opened up a new field for marine forecasting with a wide development and application prospects. Meanwhile, it is essential to carry out the selection of the observation stations, and the inde-

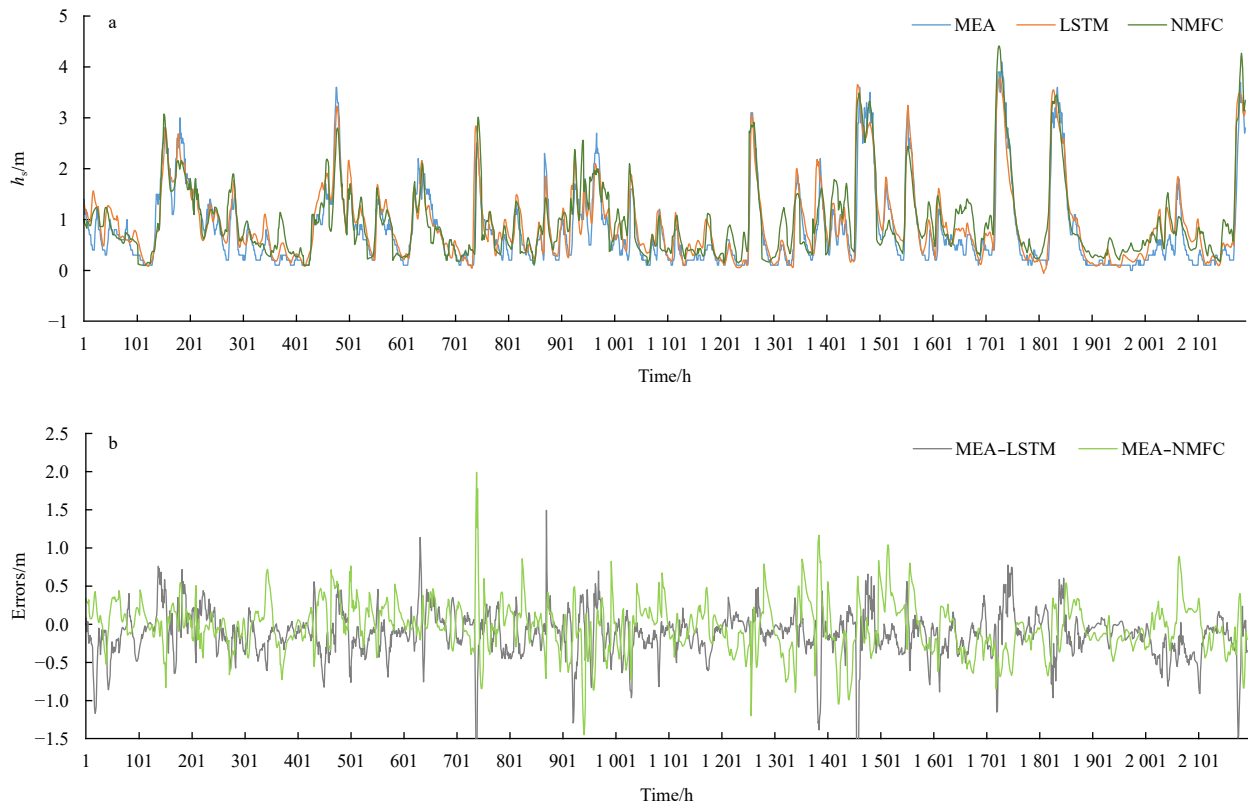


Fig. 4. Comparison between the predicted significant wave heights of LSTM-WAVE model and NMFC-SWAN model. a. Significant wave heights (blue line: measurements; red line: LSTM; green line: NMFC); and b. difference between measurement and LSTM result (grey line) and NMFC result (green line).

Table 4. Evaluation indexes of the prediction errors

Wave height	Indicator	Model					
		WAVE-LSTM			NMFC-SWAN		
		N01	N02	N03	N01	N02	N03
$0.3 \text{ m} \leq h_s \leq 5 \text{ m}$ (SN: N01: 1 372; N02: 1 732; N03: 1 732)	RMSE/m	0.32	0.35	0.38	0.40	0.37	0.39
	SI	0.40	0.32	0.33	0.49	0.35	0.34
	MAE/m	0.23	0.27	0.28	0.30	0.27	0.29
	r	0.93	0.90	0.89	0.89	0.89	0.88
$0.3 \text{ m} \leq h_s \leq 1 \text{ m}$ (SN: N01: 817; N02: 1 066; N03: 1 096)	RMSE/m	0.30	0.33	0.35	0.38	0.31	0.28
	SI	0.72	0.52	0.51	0.91	0.48	0.41
	MAE/m	0.21	0.25	0.26	0.28	0.22	0.21
$1 \text{ m} < h_s \leq 2 \text{ m}$ (SN: N01: 353; N02: 503; N03: 449)	RMSE/m	0.35	0.31	0.39	0.42	0.43	0.45
	SI	0.24	0.21	0.26	0.29	0.29	0.31
	MAE/m	0.25	0.25	0.30	0.30	0.33	0.38
$2 \text{ m} < h_s \leq 3 \text{ m}$ (SN: N01: 133; N02: 150; N03: 137)	RMSE/m	0.45	0.41	0.46	0.46	0.56	0.60
	SI	0.18	0.16	0.19	0.18	0.22	0.24
	MAE/m	0.36	0.34	0.37	0.38	0.46	0.48
$3 \text{ m} < h_s \leq 5 \text{ m}$ (SN: N01: 69; N02: 55; N03: 50)	RMSE/m	0.34	0.72	0.50	0.45	0.55	0.64
	SI	0.10	0.21	0.15	0.13	0.16	0.19
	MAE/m	0.28	0.65	0.43	0.37	0.45	0.55

pendent/dependent variables based on the principles of the ocean studies and the physical background for further improvement, as well as the extension of single-station forecast to regional forecasts.

References

Booij N, Ris R C, Holthuijsen L H. 1999. A third-generation wave

model for coastal regions: 1. Model description and validation. *Journal of Geophysical Research: Oceans*, 104(C4): 7649–7666, doi: [10.1029/98JC02622](https://doi.org/10.1029/98JC02622)

Chaudhari S, Balasubramanian R, Gangopadhyay A. 2008. Upwelling detection in AVHRR sea surface temperature (SST) images using neural-network framework. In: *Proceedings of IGARSS 2008–2008 IEEE International Geoscience and Remote Sensing Symposium*. Boston, MA, USA: IEEE

- Deshmukh A N, Deo M C, Bhaskaran P K, et al. 2016. Neural-network-based data assimilation to improve numerical ocean wave forecast. *IEEE Journal of Oceanic Engineering*, 41(4): 944–953, doi: [10.1109/JOE.2016.2521222](https://doi.org/10.1109/JOE.2016.2521222)
- Duan Wenyang, Huang Limin, Han Yang, et al. 2016. A hybrid EMD-AR model for nonlinear and non-stationary wave forecasting. *Journal of Zhejiang University-Science A*, 17(2): 115–129, doi: [10.1631/jzus.A1500164](https://doi.org/10.1631/jzus.A1500164)
- Filippo A, Torres Jr A R, Kjerfve B, et al. 2012. Application of artificial neural network (ANN) to improve forecasting of sea level. *Ocean & Coastal Management*, 55: 101–110, doi: [10.1016/j.ocecoaman.2011.09.007](https://doi.org/10.1016/j.ocecoaman.2011.09.007)
- Gao Song, Zhao Peng, Pan Bin, et al. 2018. A nowcasting model for the prediction of typhoon tracks based on a long short term memory neural network. *Acta Oceanologica Sinica*, 37(5): 8–12, doi: [10.1007/s13131-018-1219-z](https://doi.org/10.1007/s13131-018-1219-z)
- Hasselmann K, Barnett T P, Bouws E, et al. 1973. Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Erganzungsheft zur Deutschen Hydrographischen Zeitschrift Reihe A*, 1–95
- Hinton G E, Salakhutdinov R R. 2006. Reducing the dimensionality of data with neural networks. *Science*, 313(5786): 504–507, doi: [10.1126/science.1127647](https://doi.org/10.1126/science.1127647)
- Hochreiter S, Schmidhuber J. 1997. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)
- James S C, Zhang Yushan, O'Donncha F. 2018. A machine learning framework to forecast wave conditions. *Coastal Engineering*, 137: 1–10, doi: [10.1016/j.coastaleng.2018.03.004](https://doi.org/10.1016/j.coastaleng.2018.03.004)
- Jiang Zongli. 2001. *Introduction to Artificial Neural Networks* (in Chinese). Beijing: Higher Education Press
- Kuang Xiaodi, Wang Zhaoyi, Zhang Miaoyin, et al. 2016. An interpretation scheme of numerical near-shore sea-water temperature forecast based on BPNN. *Oceanologia et Limnologia Sinica* (in Chinese), 47(6): 1107–1115
- Kumar N K, Savitha R, Al Mamun A. 2017. Regional ocean wave height prediction using sequential learning neural networks. *Ocean Engineering*, 129: 605–612, doi: [2016.10.033](https://doi.org/10.1016/j.oceaneng.2016.10.033)
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature*, 521(7553): 436–444, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)
- Lipton Z C, Berkowitz J, Elkan C. 2015. A critical review of recurrent neural networks for sequence learning. arXiv preprint, arXiv: 1506.00019
- Liu Hui, Wang Jing. 2008. Estimation of ocean upper mixed layer depth using artificial neural network. *Journal of Tropical Oceanography* (in Chinese), 27(3): 9–13
- Skamarock W C, Klemp J B, Dudhia J, et al. 2005. A description of the advanced research WRF version 2 (No. NCAR/TN-468+STR). University Corporation for Atmospheric Research. doi: [10.5065/D6DZ069T](https://doi.org/10.5065/D6DZ069T)
- Tissot P E, Cox D T, Michaud P. 2001. Neural network forecasting of storm surges along the gulf of Mexico. In: *Fourth International Symposium on Ocean Wave Measurement and Analysis*. San Francisco, California, United States: American Society of Civil Engineers, 1535–1544, doi: [10.1061/40604\(273\)155](https://doi.org/10.1061/40604(273)155)
- Vilibić I, Šepić J, Mihanović H, et al. 2016. Self-organizing maps-based ocean currents forecasting system. *Scientific Reports*, 6(1): 22924, doi: [10.1038/srep22924](https://doi.org/10.1038/srep22924)
- Wu Lingjuan, Gao Song, Liu Aichao, et al. 2015. Operational system of monitoring, forecasting and warning on marine disaster for Shandong Province. *Journal of Institute of Disaster Prevention* (in Chinese), 17(2): 61–69
- Yang Haofan, Chen Y P P. 2019. Hybrid deep learning and empirical mode decomposition model for time series applications. *Expert Systems with Applications*, 120: 128–138, doi: [10.1016/j.eswa.2018.11.019](https://doi.org/10.1016/j.eswa.2018.11.019)
- Zhang Wenxiao, Gao Guodong, Mu Guangyu. 2006. Study on the model of salinity based on back-propagation artificial neural network. *Ocean Technology* (in Chinese), 25(4): 39–41
- Zhang Dan, Peng Xiangang, Pan Keda, et al. 2019. A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine. *Energy Conversion and Management*, 180: 338–357, doi: [10.1016/j.enconman.2018.10.089](https://doi.org/10.1016/j.enconman.2018.10.089)