Computing and Informatics, Vol. 41, 2022, 885–904, doi: 10.31577/cai_2022_3_885

PREDICTION OF SIGNIFICANT WAVE HEIGHT BASED ON GATED RECURRENT UNIT AND SEQUENCE-TO-SEQUENCE NETWORKS IN THE TAIWAN STRAIT

Yindong Zeng

College of Oceanic and Atmospheric Sciences
Ocean University of China
Qingdao 266071, China
&
Marine Forecasting Center of Fujian Province
Fuzhou 350003, China
e-mail: zydzyd100@163.com

Jinshan Ma

Institute of Computer Applications
Fujian Agriculture and Forestry University
Fuzhou 350000, China
e-mail: JinshanMa@fafu.edu.cn

Xinping Chen*

National Marine Hazard Mitigation Service
Ministry of Natural Resources of China
Beijing 100194, China
e-mail: xinp.chen@foxmail.com

Minquan Guo, Zaichang Ren

Marine Forecasting Center of Fujian Province
Fuzhou 350003, China
e-mail: guominquan@163.com, yu02241129@126.com

Yanqi Jiang

College of Harbour, Coastal and Offshore Engineering
Hohai University
Nanjing 210098, China
e-mail: Jiangyq1230@foxmail.com

Zhenchang Zhang

Institute of Computer Applications
Fujian Agriculture and Forestry University
Fuzhou 350000, China
e-mail: stdin@fafu.edu.cn

Abstract. Wave forecasting approaches based on deep learning techniques have recently made a great progress. In this study, we developed a deep learning model based on Gated Recurrent Unit (GRU) and sequence-to-sequence neural networks (GRUS), to improve the forecasting accuracy of significant wave heights for the Taiwan Strait, where ocean waves and winds own their unique characteristics. The performances of our proposed GRUS model and the other deep learning models based on WaveNet and Long Short-Term Memory (LSTM) were compared by means of wind and wave observations at three buoys in the study area. Model parameters were optimized by means of various model experiments. Performance comparison illustrates that our proposed GRUS model outperforms the other models in 24-hour H_s forecasting, while the GRUS has extraordinary ability for short-term prediction (prediction horizon is less than 6 h). Moreover, for high wave states prediction (e.g., wave height over 4 m), the GRUS has the strongest prediction ability among the models, in which forecasted wave heights are mostly lower than the corresponding observations.

Keywords: Wave forecasting, significant wave height, gated recurrent unit, long short-term memory, sequence-to-sequence

1 INTRODUCTION

Wave forecasting is of great importance to various maritime activities and coastal engineering. Since there are many factors affecting wave height forecasting, the

* Corresponding author

accuracy of wave height prediction was, and still is, a big challenge for marine scientific research. With the development of Artificial Intelligence (AI) technology, data-driven, model-free approaches have become more and more popular in recent years [1].

Neural networks are among most powerful tools in AI techniques, which are able to approximate almost any complex nonlinear process to investigate possible relationships and dependencies to predict uncertain future events such as wave forecasting [1, 2, 3]. Deo and Sridhar Naidu [4] and Londhe and Panchang [5] explored the value of Artificial Neural Networks (ANN) in predicting significant wave heights based on measured data, but the prediction time was relatively short. Zamani et al. [6] used neural networks to establish a non-linear relationship between winds and waves to realize the prediction of significant wave heights in the Caspian Sea. Mahjoobi and Adeli Mosabbeb [7] proposed a method of Support Vector Machine (SVM) for ocean wave prediction, showing that SVM can be used to forecast significant wave heights. Nitsure et al. [8] added the wind information from measured data in a hindcast study for significant wave heights along the North American and Indian Ocean coasts. An ensemble numerical and ANN approach was introduced by Dixit and Londhe in [9] to predict the next 24-hour wave heights at different buoys along the Indian coast. Berbić et al. [10] studied the application of ANN and SVM for significant wave height prediction. From these studies, one can see that these approaches have made great efforts in improving the accuracy of wave forecasting while increasing the prediction horizon in time for the future. However, with the increase of the forecast time, the forecast accuracy inevitably decreases, and the forecasting horizon in time for most approaches was still relatively short.

The previous studies indicate the fact that ocean wave forecasting approaches require to consider the different geographical environment. The study area of the present study is the Taiwan Strait (Figure 1). The region has its own characteristics in terms of wind variation, wind function on waves and wave propagation patterns. In recent years, efforts have been made to study the prediction of ocean waves in this region. For example, Wang et al. [11] employed Gated Recurrent Unit (GRU) network to forecast the significant wave height in the Taiwan Strait and its Adjacent Waters. Ma et al. [12] developed a forecasting model based on the convolution operation, LSTM, and full connect networks to study the prediction of significant wave height in the Taiwan Strait.

Based on previous studies, this study develops a forecasting model based on GRU and sequence-to-sequence neural networks. In this study, we selected three typical buoys with high data quality, located in the northern, central and southern waters near the midline of the Taiwan Strait from north to the south (Figure 1). By means of the observational wave and wind data, we applied the forecasting model proposed in this paper for the prediction of significant wave heights in the Taiwan Strait, and compared the model results with those of the WaveNet, LSTM models to show the performances of these models.

The remainder of this paper is organized as follows: Section 2 introduces the related work of this study. Section 3 describes the forecasting model with GRU and sequence-to-sequence networks in detail. Section 4 presents the experimental setup. In Section 5, we investigate the model performance and comparisons. Finally, conclusions are presented in Section 6.

2 RELATED WORK

In recent years, the implementation of deep learning models for wave height prediction has become more and more popular. For example, a comparison between ANN, Bayesian Networks (BN), SVM, Adaptive Neuro-Fuzzy Interference System for predicting wave heights from wind speeds was presented by Malekmohamadi et al. [13], leading to the conclusion that BN and SVM can provide useful information on the reliability of input and output data. Sinha and Basu [14] used Genetic Algorithm for wave height prediction in the Bay of Benga. Nikoo and Kerachian [15] proposed to use the Artificial Immune Recognition System for the prediction of wave heights in Lake Superior in the northern USA, and its prediction results were better than those of ANN, BN and SVM. James et al. [16] introduced two highly efficient machine learning models of Multilayer Perceptron and SVM for regression analysis of wave heights and classification analysis of ocean wave periods. Kumar et al. [3] proposed an ensemble of Extreme Learning Machine (ELM) to predict the daily wave height. Moreover, Ali and Prasad [17] used an improved ELM model to predict wave heights with high model accuracy, by considering the influence of winds on ocean waves and using the series of the historical wave heights as a predictor of the model. In the study of Zhang and Dai [18], the restricted Boltzmann machine in the classical deep belief network was substituted with the conditional restricted Boltzmann machine containing temporal information to predict significant wave heights. Ni and Ma [19] used the Long Short-Term Memory (LSTM) algorithm [20] to study the prediction of polar westerlies wave heights, indicating that the LSTM model is feasible for the prediction of significant wave heights. Pirhooshyaran and Snyder [1] introduced Recurrent Neural Network (RNN) frameworks, integrated with Bayesian hyperparameter optimization and Elastic Net methods, to explore the concepts of ocean wave forecasting, hindcasting and feature selection.

In addition, some studies have established wave height prediction models by combining wave decomposition with soft computing approaches. For example, Deka and Prahlada [21] studied the prediction of significant wave heights from the perspective of combining wave decomposition and neural networks. Prahlada and Deka [22] utilized a hybrid model combining neural networks with wavelets (WLNN) to predict the wave heights of the 48-hour into the future. The work of Duan et al. [23] aims at predicting significant wave heights, by means of the test value decomposition and SVM. Ali and Prasad [17] investigated a prediction model for wave heights, based on the test value decomposition method combined with the ELM. Ni and Ma [19] used the principal component analysis to predict wave heights for the polar region. These studies indicate that: the decomposition of ocean waves illustrates that different factors can affect the future trend of wave heights, however, as the parameters of the ocean wave decomposition methods are not unique, the pros and cons of the decomposition methods might affect the prediction accuracy of wave heights.

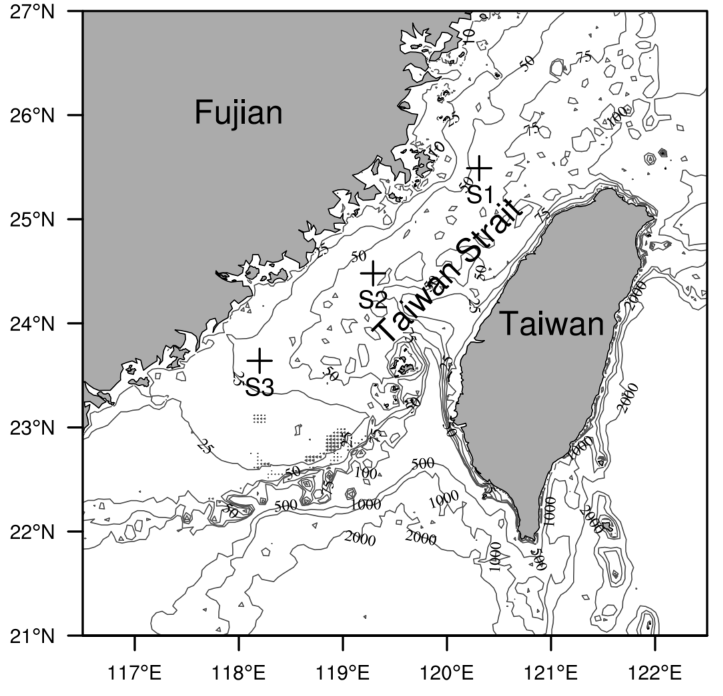

Figure 1. Bathymetry contours (m) for the study area together with locations of the wind and wave measurement buoys in black crosses (S1–S3) in the Taiwan Strait

3 METHODOLOGY

In this study, we developed a deep learning model based on GRU and sequence-to-sequence neural networks, hereafter referred to GRUS model. Figure 2 displays the framework of the proposed GRUS model. Sequence-to-sequence (Seq2Seq) networks are the main architecture of the GRUS, using two separate GRU networks. The first network maps (encodes) an input sequence to a fixed-sized vector representation, and the other decodes the representation into the output sequence. The GRUS model is described in detail in this section.

3.1 GRU

GRU is a gating mechanism in recurrent neural networks [24]. In this study, GRUs are introduced in the proposed GRUS networks. In the encoder network, the historical observations of wind speed, wind direction and H_s are used as the input data into the GRU, and an update gate and a reset gate are used to obtain the characteristics of the historical wind fields and H_s. The implementation of an update gate can be expressed as follows:

$$\mu_t = \delta(W_\mu \cdot [V_t, D_t, S_t, \varphi_{t-1}]) \tag{1}$$

where μ_t is the output of the update gate, V_t, D_t, S_t respectively represent wind speed, wind direction and H_s as well as φ_{t-1} is the previous feature state of wind speeds and directions as well as H_s, W_μ is the weight matrices that are learned by the update gate, δ is the logistic sigmoid function (see Section 3.3 for details) that makes the μ_t values in the range of 0–1.

Similarly, the implementation of a reset gate can be expressed as follows:

$$\gamma_t = \delta(W_\gamma \cdot [V_t, D_t, S_t, \varphi_{t-1}]), \tag{2}$$

in which γ_t is the output of the reset gate, and W_γ is the weight matrices that are learned by the reset gate.

After getting γ_t, the GRU network will further filter the information of φ_{t-1}, to obtain a candidate state vector, $\widetilde{\varphi}_t$, which can be calculated by:

$$\widetilde{\varphi}_t = tanh(W_{\widetilde{\varphi}_t} \cdot [V_t, D_t, S_t, \gamma_t * \varphi_{t-1}]) \tag{3}$$

where *tanh* is the activation function, and $W_{\widetilde{\varphi}_t}$ is the weight matrices which are learned.

Finally, we can calculate the output of the GRU network at any time, i.e., φ_t, by the following expression:

$$\varphi_{t-1} = (1 - \mu_t) * \varphi_{t-1} + \mu_t * \widetilde{\varphi}_t, \tag{4}$$

in which "$*$" is the element-wise operator.

3.2 Sequence-to-Sequence Neural Networks

Cho et al. [24] and Sutskever et al. [25] independently proposed similar two-part deep learning architectures consisting of two RNN, namely encoder and decoder. In the present study, we used the historical observations of wind fields and H_s as the input sequence into the encoder, and the output of the encoder and the forecasted 24-hour wind fields (based on the WRF model) are given to the decoder. Since the size of the encoder output vector is different from that of the input vector of the decoder, the encoder output vector needs to be processed accordingly. In this study, we used a layer of one-dimensional convolutional network (CNN1D) to compress the encoder output vector, by taking the following equation:

$$C_{out}^k = \sum_{i=1}^{c} \sum_{j=1}^{n} \omega_{(i,\dots,j)}^k \cdot \mathbf{E}_{(i,\dots,j)} \tag{5}$$

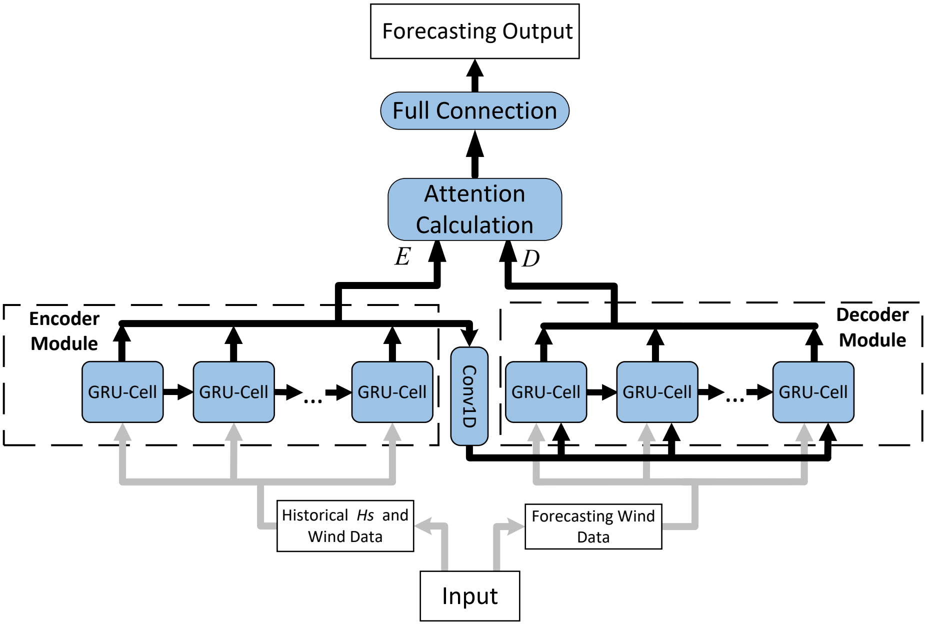

Figure 2. Framework of the proposed GRUS based on the encoder-decoder architecture with GRU and attention mechanism

where $\mathbf{E}_{(i,\dots,j)}$ represents the encoder output vector, c is the number of features of $\mathbf{E}_{(i,\dots,j)}$ at each moment, $\omega_{(i,\dots,j)}$ represents the convolution kernel, n is the size of the convolution kernel, C^k_{out} is the output of the convolution calculation, k is the number of the convolution kernel and also the number of features in the output of the convolution.

In the present study, we use the C_{out} and the forecasted wind speed and direction data (respectively represented by U and F) as the input data given to the decoder, the GRU calculation process in the decoder network is consistent with that in the encoder introduced above (see Equations (1), (2), (3) and (4)), where U, F and C_{out} are used instead of V, D and S. Then, the GRUS model uses attention calculation and a full connection (FC) network to achieve the prediction for H_s.

3.3 Activation and Loss Function

In this study, the PReLU, tanh activation functions are used in the GRUS for model training, while the ReLU is used for other models also tested in this study for model comparison (see Section 4.3 for details). The ReLU is a piecewise linear function that outputs zero if its input (x) is negative, which can be expressed as:

$$f(x) = \max(0, x). \tag{6}$$

The parametric rectified linear unit, i.e., PReLU, is common to use as follows:

$$f(\alpha, x) = \begin{cases} x, & x \geqslant 0 \\ \alpha x, & x < 0 \end{cases} []. \tag{7}$$

In addition, the loss function used for training in this study is the mean square error (MSE). The MSE has the advantages of convenient calculation, accurate measurement error, and good convergence effect, and can be expressed as:

$$Loss = MSE = \frac{1}{m}\sum_{i=1}^{m}(y^i - \hat{y}^i), \tag{8}$$

in which y and $\hat{y}$ represent observation and prediction values, respectively.

4 EXPERIMENTAL SETUP

4.1 Datasets

The region of interest in the study is the Taiwan Strait, which extends from 21.5 °N–26.0 °N and 117.5 °E–121.0 °E (Figure 1). Three buoys acquired by Marine Forecasting Center of Fujian Province of China are available for the region. The geographic locations of these buoys are given in Figure 1.

The collected buoy data adapted in the present study were provided hourly, consisting of wave and wind parameters, including H_s, wind speed and direction at 10 m height above the sea surface. The data covering period for each buoy used in this study is listed in Table 1. Some data were not recorded over a span of several days.

In addition to buoy measurement wind data, forecasting wind speed and direction data, interpolated to the buoy locations, are also used in the wave forecasting models of the present study for model testing. The forecasting wind data were taken from an operational atmospheric model developed by Marine Forecasting Center of Fujian Province (MFCF) of China, which was based on the WRF (Weather Research and Forecasting) model (WRF-Based model). The model domain extends from 15 °N–45 °N, 95 °E–150 °E, with the spatial resolution of 5 km$\times$5 km. The model results were carefully validated, by the comparison with observational data.

Three models for H_s forecasting (see Section 4.4 for details) were trained and tested in this study, to compare the model performances. We used a year data (Table 1) including previous 72-hour H_s and wind observations as well as next 24-hour predicted wind data, to train the prediction models. By means of the training models, the forecasting models were established for next 24-hour H_s prediction. Another one-year data, including observed H_s and wind data as well as predicted wind data from the WRF-Based model, were used for testing the models.

Buoy	Year		V_{mean}	V_{max}	$(H_s)_{mean}$	$(H_s)_{max}$	No. of Samples	
	Training	Testing					Training	Testing
S1	2016	2017	8.6	29.5	1.7	6.7	7 192	8 736
S2	2017	2019	7.7	23.5	1.4	6.8	8 715	8 664
S3	2016	2017	7.9	23.0	1.6	6.0	8 425	8 736

Table 1. List of measurement buoys in the Taiwan Strait region and some fundamental mean wind and wave characteristics as well as information of the data covering period and number for training and testing samples. Main wind and wave characteristics were statistically calculated, including wind speed, V (m/s); significant wave height, H_s (m).

4.2 Comparison Algorithm

In the present study, we compared the results of the GRUS with those of the other models based on WaveNet and LSTM algorithms, to illustrate the performance of the proposed GRUS model.

The WaveNet is a fully convolutional neural network, which has also been used for wave height prediction. For example, Liu et al. [26] implemented a WaveNet model to process the obtained sensor data and predict wave height and period. In this study, we also performed this algorithm to predict H_s, following the study of Liu et al. [26].

The LSTM is a variant of RNN aimed at avoiding the vanishing gradient problem by gated regulators [27], which has been used for wave state prediction (e.g., [19]).

4.3 Parameter Settings

In this study, all deep learning models used Adam optimizer with the learning rate (g) of 0.001, while the training iteration (p) for all models is 500. The model parameter settings are listed in Table 2. We use the greedy strategy to optimize the parameters of the deep-learning model.

WaveNet model parameter settings: Following Liu et al. [26], we also used a 1D convolutional neural network with the same parameter settings, except that $h_1 = 48$ and $h_2 = 24$ in the present study.

LSTM model parameter settings: First, we performed three sets of experiments to obtain the optimal number of LSTM layers. The number of the LSTM layers, l, for the three sets of experiments is, respectively, set to 2, 3 and 4 (denoted as LSTM-2, LSTM-3, LSTM-4), and the number of the neurons, h, in each LSTM layer is set to 24, while a FC layer also with 24 neurons is added to each of the LSTM models. Then, we carried out the other 21 experiments with $h = 26$, 28, 30, 32, 34, 36 and 38 neurons for LSTM-2, LSTM-3 and LSTM-4, respectively, to select the optimal number of neurons. The activation functions are tanh and PReLU. These experimental results show that the LSTM-2 with 24 neurons has the best results (hereafter referred to LSTM-2).

LSTM model parameter settings: In this study, a set of experiments were carried out, to obtain the optimal number of GRU neurons in the encoder and decoder, by means of 12, 16, 18, 22, 24, 26, 30, 34, 38, 42, and 46 GRU neurons. The activation functions are tanh and PReLU. Finally, these model experiments demonstrate that the model can provide the best results when the number of GRU neurons is 24.

Model	Component	Parameter Settings	
LSTM-2	LSTM	$l = 2$, $h_1 = 24$, $h_2 = 24$, activation = *tanh*	
	FC	$l = 1$, $h_1 = 24$, $h_2 = 24$, activation= PReLU	
WaveNet	Conv1D	$k_1 = 3$, $n_1 = 16$, $k_2 = 5$, $n_2 = 32$, $k_3 = 11$, $n_3 = 64$	$p = 500$
	FC	$l = 2$, $h_1 = 48$, $h_2 = 24$, activation= ReLU	$g = 0.001$
GRUS	GRU	$l = 1$, $h_1 = 24$, activation = *tanh*	
	FC	$l = 1$, $h_1 = 24$, activation= PReLU	

Table 2. Model parameter settings. l represents the number of the model hidden layers, h_i is the neuron number in each layer, n_i represents the number of convolution operations in the i^{th} layer, k_i represents the length of the convolution operations in the i^{th} layer, g is the learning rate, and p represents the training number.

4.4 Evaluation Measures

In this study, the performances of the proposed model, i.e., GRUS model, are compared to the WaveNet and LSTM forecasting models, at the three buoy stations in the study area. To objectively evaluate the forecasting models, we carried out some quantitative evaluations, using the estimates of root mean square error (RMSE), the correlation coefficient (R) and Mean Absolute Error (MAE).

RMSE is a metric showing the average distance between the predicted values from the model and the actual values in the dataset, which can be calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y^i - \hat{y}^i)} \tag{9}$$

where n is the sample size, y^i is the predicted value for the i^{th} observation in the dataset, and $\hat{y}^i$ is the observed value for the i^{th} observation in the dataset.

Correlation coefficient is used to measure the strength of the relationship between two variables, and is defined as:

$$R = \frac{Cov(y, \hat{y})}{\sqrt{Var(y)\,Var(\hat{y})}} \tag{10}$$

where $Cov(\cdot)$ and $Var(\cdot)$ refer to the covariance and variance operator, respectively.

MAE is the average difference between the observations and model predictions, which can be calculated by:

$$MAE = \frac{1}{n}\sum_{i=1}^{N}|y^i - \hat{y}^i|. \tag{11}$$

5 EXPERIMENTAL ANALYSIS

5.1 Wind and Wave States

In this study, we analyzed the wind and wave states in terms of the observed data for the year 2017. The time series of hourly wind speeds based on the buoy data in 2017 is plotted by Figure 3. It can be observed that wind speeds in the study area were mostly smaller than 20 m/s. Relatively high wind speeds of larger than 20 m/s happened occasionally, mostly occurring with extreme weather events such as tropical cyclones. Wind direction and wind speed for different seasons at the buoys is presented in terms of a wind rose figure (Figure 4). In spring, autumn and winter, winds were mostly from NE, followed by N and NNW; in summer, the prevailing wind direction was SW, while occasionally from NE.

Hourly H_s at the buoys considered in this study is also plotted in Figure 3. We can see from this figure that H_s was smaller than 4 m, most of the time, while relatively larger H_s values exceeding 5 m over the year are found only in some rare cases due to relatively strong winds. The correlations between the wind speeds and H_s at the stations are considerably high.

The fundamental characteristics of wind speeds and H_s, in terms of the annual mean and maximum wind speed (i.e., V_{mean} and V_{max}), annual mean and maximum H_s (i.e., $(H_s)_{mean}$ and $(H_s)_{max}$) are summarized in Table 1. In the study area, V_{mean} and $(H_s)_{mean}$ in 2017 were over 7.7 m/s and 1.4 m, respectively. The maximum values of wind speeds and H_s at the buoys are 29.5 m/s and 6.8 m, respectively.

Overall, the Taiwan Strait region displays its own unique characteristics in winds and H_s states, which is of importance for forecasting models to provide reasonable forecasting results.

5.2 Model Results

In this section, we compared the model forecasting results with buoy observations, to objectively study the performance of the forecasting models for 24-hour H_s prediction. Figure 5 shows the time series of the hourly observed and forecasted H_s at the buoy stations for the testing period (see Table 1). It is seen that the forecasted H_s variations of the models are generally in agreement with observations. The correlations between the model results and observations of H_s at each buoy for the testing period are displayed by the scatter diagram (Figure 6) with the statistics of the RMSE and correlation coefficient (R). Figure 6 exhibits that the R values are all

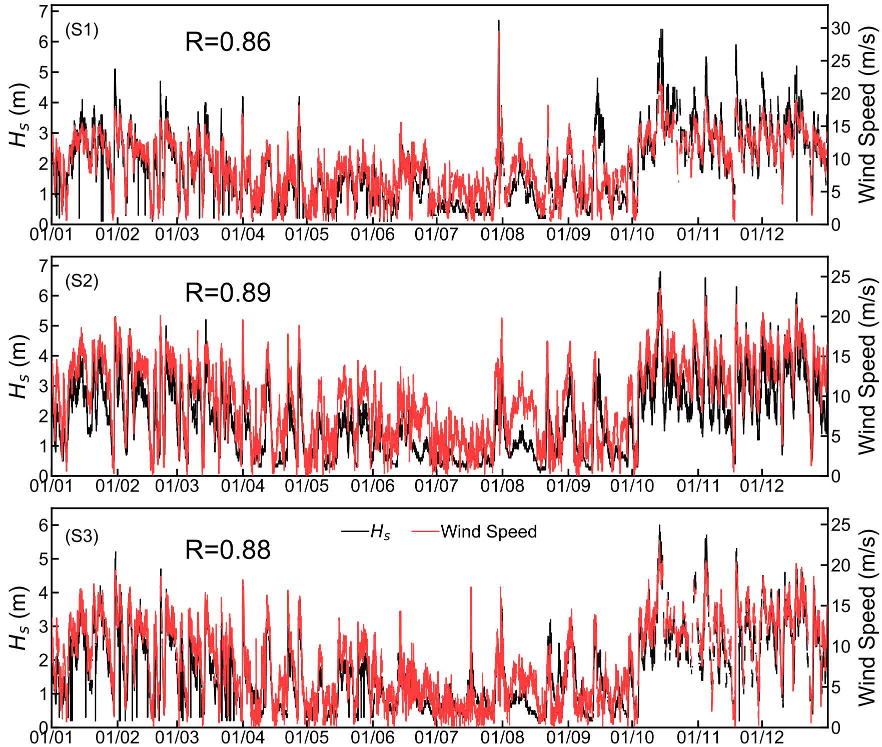

Figure 3. Time series of hourly wind speed (m/s) and significant wave height (H_s in m) at the buoys (S1–S3) shown in Figure 1 for the year 2017

statistically significant ($R > 0.78$), indicating that the models can provide generally reasonable results of H_s variability in 24-hour prediction for the study region.

The comparison between the model results shows that the WaveNet provides the worst model results, with lowest R while largest RMSE values among all model results (Figure 6). The RMSE values are all larger than 0.5 m for the WaveNet model, while smaller than 0.5 m for the other models. The correlation coefficients of the LSTM-2 model are over 0.9 at the buoy stations. Most importantly, the comparison between the model results exhibits that our proposed GRUS model outperforms the other methods in H_s forecasting. It is apparent that all indicators of GRUS are better than those of the other models. The RMSE values of the GRUS at S1–S3 are, respectively, 0.42, 0.38 and 0.38, while the lowest values of the other models are 0.50, 0.46, and 0.45, respectively (Table 3).

For high wave state prediction (e.g., $H_s > 4$ m), it is observed from Figure 6 that the WaveNet, LSTM-2 and GRUS models are insufficient, especially for extreme waves. For the wave states of $H_s > 4$ m, most of the forecasted H_s are smaller than the observed H_s; this insufficiency is most evident for the WaveNet model, while, in

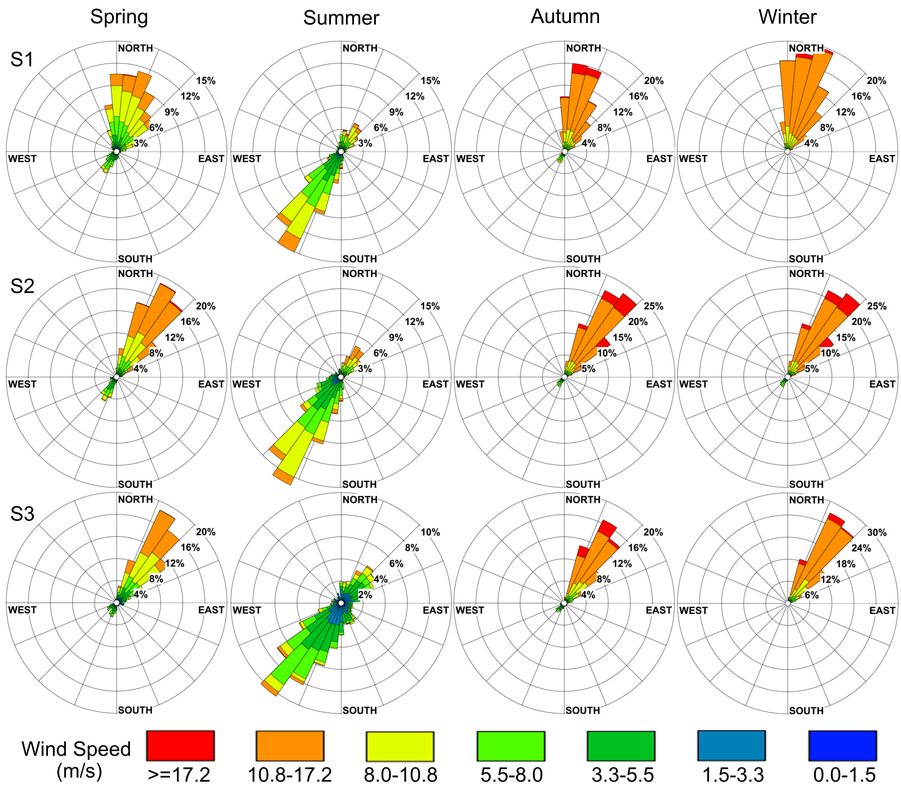

Figure 4. Seasonal wind speed and direction roses based on the measurement for the buoys S1–S3 for the year 2017

contrast, of all the models, the GRUS provides the best results that are considerable close to the observations.

Prediction horizon in time is an important aspect of the ability of forecasting methods, especially for long-term prediction. In general, the longer the prediction horizon is, the weaker the correlation in the data series is. In this study, we further analysed the trend of the forecasting accuracy of the modelled H_s for next X-hour. X is between 1–24 in the present study, meaning the prediction horizon in time is between 1–24 (h). The RMSE values of different prediction horizons are shown in Figure 7. It can be seen from the figure that the RMSEs of the proposed GRUS are smallest within 24-hour prediction, indicating, again, that the performance of the GRUS model is the best among these models. With prediction horizon increasing (i.e., X increasing), prediction errors of the models show overall increasing trends, with different increasing rates. Furthermore, it is noticed from Figure 7 that the GRUS RMSEs are significantly low for less than 6-hour prediction, illustrating that

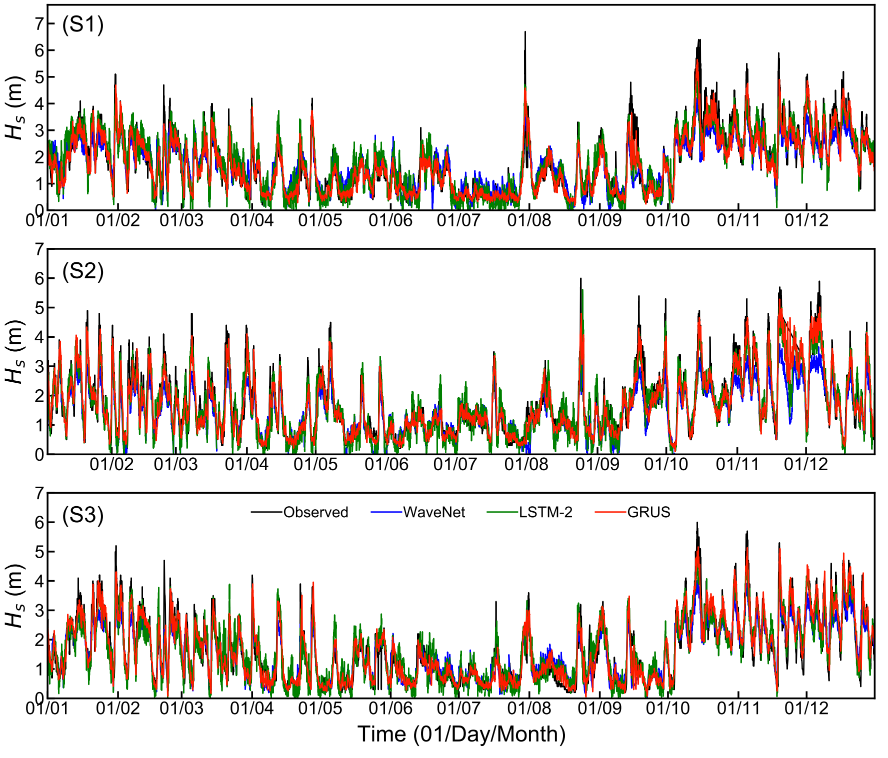

Figure 5. Comparison between modelled and observed H_s within 24-hour prediction at S1–S3

the GRUS has remarkable ability for short-term prediction. For example, for 3-h prediction, the RMSE values of the GRUS are reduced by over 29 %, 31 % and 22 % at S1–S3, respectively, compared to other model results. Overall, the performance of the GRUS developed in this study provides the best forecasting results for H_s, especially for short-term prediction.

6 CONCLUSIONS

In recent years, approaches based on deep learning techniques have become popular in wave forecasting. The goal of this study is to develop the forecasting techniques to improve the wave forecasting accuracy for the Taiwan Strait, where the wave states own its unique characteristics. Three buoy observations in wind and wave states were selected to represent the spatial feature of the study region. The hourly variability in H_s is highly correlated to that of wind speeds. Seasonality can be observed in H_s, wind speed and directions. NE winds prevail in winter time, while SW winds mostly occur in summer.

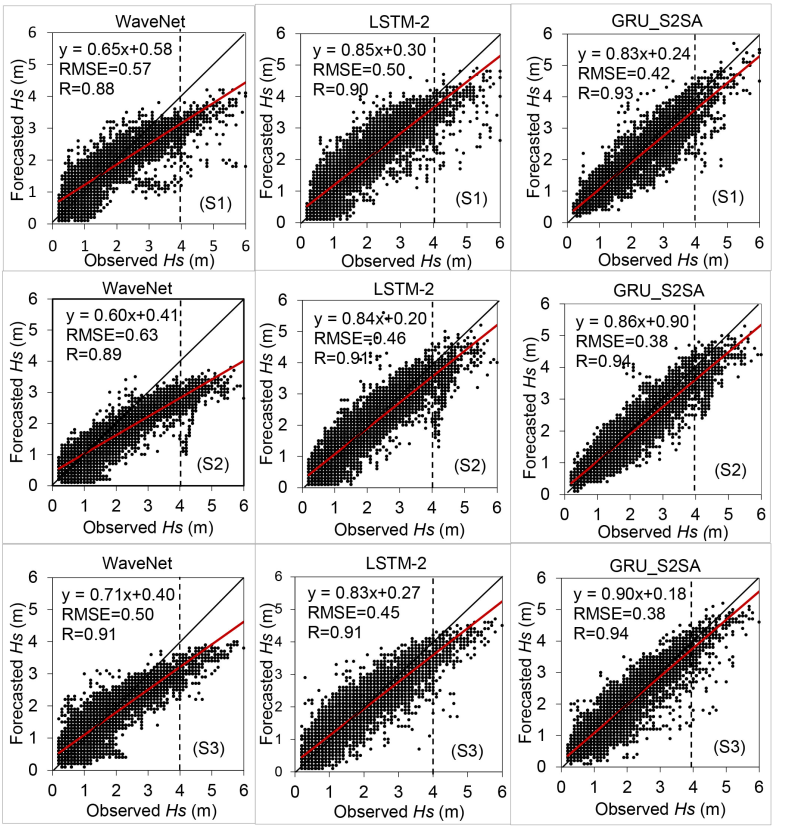

Figure 6. Comparison between modelled and observed H_s within 24-hour prediction at S1–S3. R represents correlation coefficient.

Station	Error Index	WaveNet	LSTM-2	CLSF	GRUS
	RMSE (m)	0.57	0.50	0.46	**0.42**
S1	MAE (m)	0.42	0.36	0.32	**0.28**
	R	0.88	0.90	0.92	**0.93**
	RMSE (m)	0.63	0.46	0.46	**0.38**
S2	MAE (m)	0.45	0.34	0.33	**0.27**
	R	0.89	0.91	0.92	**0.94**
	RMSE (m)	0.50	0.45	0.39	**0.38**
S3	MAE (m)	0.37	0.34	0.29	**0.27**
	R	0.91	0.91	0.94	**0.94**

Table 3. Error indices between forecasted next 24-hour H_s with corresponding observations

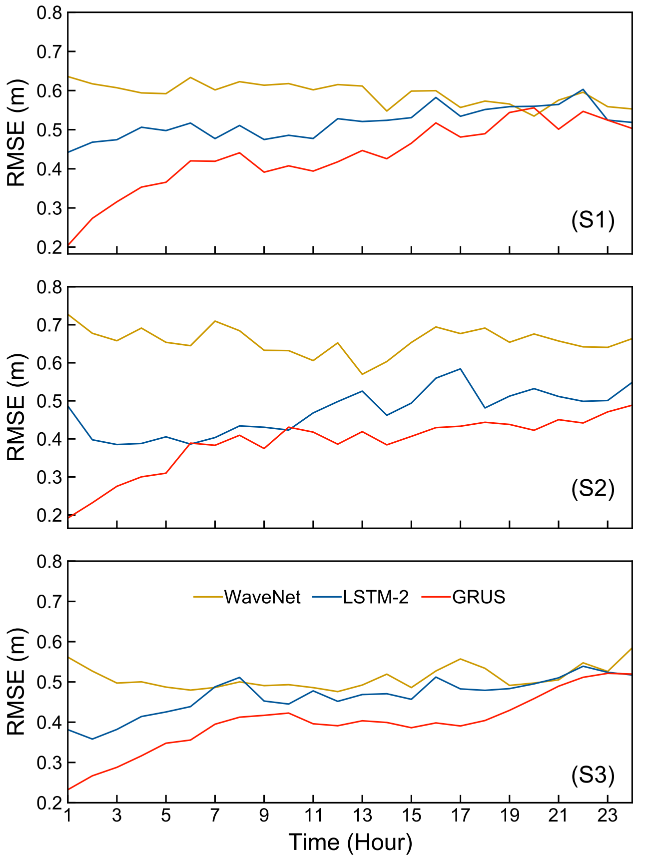

Figure 7. RMSE of forecasted H_s for next X-hour compared with observations. X represents the time on X-axis.

In this study, we developed a deep learning model by means of a GRU-based encoder-decoder architecture used in Seq2Seq networks, i.e., GRUS. The implementation of the model performance was carried out by means of the comparison between model results and buoy observations. We also compared the performances of the other deep learning models of the WaveNet and LSTM-2 from previous studies. We used one year data to train the models and another year data to test the models. Model parameters were optimized by means of various model experiments. Comparison results show that the proposed GRUS model outperforms the other models in H_s forecasting, while the GRUS has remarkable ability for short-term prediction (prediction horizon is less than 6 h). Moreover, the models are insufficient in prediction of high wave states (e.g., $H_s > 4$ m), while the GRUS model provides the best results among the models.

Acknowledgement

This study was funded by the Fujian Marine Economic Development Grant Fund Project as "Research and Application of Fine Observation, Forecasting and Intelligent Service System in the Key Sea Areas" (Project No. ZHHY-2019-2).

REFERENCES

[1] Pirhooshyaran, M.—Snyder, L. V.: Forecasting, Hindcasting and Feature Selection of Ocean Waves via Recurrent and Sequence-to-Sequence Networks. Ocean Engineering, Vol. 207, 2020, Art. No. 107424, doi: 10.1016/j.oceaneng.2020.107424.

[2] Malekmohamadi, I.—Ghiassi, R.—Yazdanpanah, M. J.: Wave Hindcasting by Coupling Numerical Model and Artificial Neural Networks. Ocean Engineering, Vol. 35, 2008, No. 3-4, pp. 417–425, doi: 10.1016/j.oceaneng.2007.09.003.

[3] Kumar, N. K.—Savitha, R.—Mamun, A. A.: Regional Ocean Wave Height Prediction Using Sequential Learning Neural Networks. Ocean Engineering, Vol. 129, 2017, pp. 605–612, doi: 10.1016/j.oceaneng.2016.10.033.

[4] Deo, M. C.—Sridhar Naidu, C.: Real Time Wave Forecasting Using Neural Networks. Ocean Engineering, Vol. 26, 1998, No. 3, pp. 191–203, doi: 10.1016/S0029-8018(97)10025-7.

[5] Londhe, S.—Panchang, V.: One Day Wave Forecasts Based on Artificial Neural Networks. Journal of Atmospheric and Oceanic Technology, Vol. 23, 2006, No. 11, pp. 1593–1603, doi: 10.1175/JTECH1932.1.

[6] Zamani, A.—Solomatine, D.—Azimian, A.—Heemink, A.: Learning from Data for Wind-Wave Forecasting. Ocean Engineering, Vol. 35, 2008, No. 10, pp. 953–962, doi: 10.1016/j.oceaneng.2008.03.007.

[7] Mahjoobi, J.—Adeli Mosabbeb, E.: Prediction of Significant Wave Height Using Regressive Support Vector Machines. Ocean Engineering, Vol. 36, 2009, No. 5, pp. 339–347, doi: 10.1016/j.oceaneng.2009.01.001.

[8] Nitsure, S. P.—Londhe, S. N.—Khare, K. C.: Wave Forecasts Using Wind Information and Genetic Programming. Ocean Engineering, Vol. 54, 2012, pp. 61–69, doi: 10.1016/j.oceaneng.2012.07.017.

[9] Dixit, P.—Londhe, S.: Prediction of Extreme Wave Heights Using Neuro Wavelet Technique. Applied Ocean Research, Vol. 58, 2016, pp. 241–252, doi: 10.1016/j.apor.2016.04.011.

[10] Berbić, J.—Ocvirk, E.—Carević, D.—Lončar, G.: Application of Neural Networks and Support Vector Machine for Significant Wave Height Prediction. Oceanologia, Vol. 59, 2017, No. 3, pp. 331–349, doi: 10.1016/j.oceano.2017.03.007.

[11] 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, Vol. 13, 2021, No. 1, Art. No. 86, doi: 10.3390/w13010086.

[12] Ma, J.—Xue, H.—Zeng, Y.—Zhang, Z.—Wang, Q.: Significant Wave Height Forecasting Using WRF-CLSF Model in Taiwan Strait. Engineering Applications

of Computational Fluid Mechanics, Vol. 15, 2021, No. 1, pp. 1400–1419, doi: 10.1080/19942060.2021.1974947.

[13] Malekmohamadi, I.—Bazargan-Lari, M. R.—Kerachian, R.—Nikoo, M. R.—Fallahnia, M.: Evaluating the Efficacy of SVMs, BNs, ANNs and ANFIS in Wave Height Prediction. Ocean Engineering, Vol. 38, 2011, No. 2-3, pp. 487–497, doi: 10.1016/j.oceaneng.2010.11.020.

[14] Sinha, M.—Rao, A. D.—Basu, S.: Forecasting Space-Time Variability of Wave Heights in the Bay of Bengal: A Genetic Algorithm Approach. Journal of Oceanography, Vol. 69, 2013, No. 1, pp. 117–128, doi: 10.1007/s10872-012-0154-4.

[15] Nikoo, M. R.—Kerachian, R.: Wave Height Prediction Using Artificial Immune Recognition Systems (AIRS) and Some Other Data Mining Techniques. Iranian Journal of Science and Technology, Transactions of Civil Engineering, Vol. 41, 2017, No. 3, pp. 329–344, doi: 10.1007/s40996-017-0067-y.

[16] James, S. C.—Zhang, Y.—O'Donncha, F.: A Machine Learning Framework to Forecast Wave Conditions. Coastal Engineering, Vol. 137, 2018, pp. 1–10, doi: 10.1016/j.coastaleng.2018.03.004.

[17] Ali, M.—Prasad, R.: Significant Wave Height Forecasting via an Extreme Learning Machine Model Integrated with Improved Complete Ensemble Empirical Mode Decomposition. Renewable and Sustainable Energy Reviews, Vol. 104, 2019, pp. 281–295, doi: 10.1016/j.rser.2019.01.014.

[18] Zhang, X.—Dai, H.: Significant Wave Height Prediction with the CRBM-DBN Model. Journal of Atmospheric and Oceanic Technology, Vol. 36, 2019, No. 3, pp. 333–351, doi: 10.1175/JTECH-D-18-0141.1.

[19] Ni, C.—Ma, X.: An Integrated Long-Short Term Memory Algorithm for Predicting Polar Westerlies Wave Height. Ocean Engineering, Vol. 215, 2020, Art. No. 107715, doi: 10.1016/j.oceaneng.2020.107715.

[20] Ushiku, Y.: Long Short-Term Memory. In: Ikeuchi, K. (Ed.): Computer Vision: A Reference Guide. 2^{nd} Edition. Springer, Cham, 2021, pp. 768–773, doi: 10.1007/978-3-030-63416-2_856.

[21] Deka, P. C.—Prahlada, R.: Discrete Wavelet Neural Network Approach in Significant Wave Height Forecasting for Multistep Lead Time. Ocean Engineering, Vol. 43, 2012, pp. 32–42, doi: 10.1016/j.oceaneng.2012.01.017.

[22] Prahlada, R.—Deka, P. C.: Forecasting of Time Series Significant Wave Height Using Wavelet Decomposed Neural Network. Aquatic Procedia, Vol. 4, 2015, pp. 540–547, doi: 10.1016/j.aqpro.2015.02.070.

[23] 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 Engineering, Vol. 124, 2016, pp. 54–73, doi: 10.1016/j.oceaneng.2016.05.049.

[24] Cho, K.—Merriënboer, B.—Gulcehre, C.—Bahdanau, D.—Bougares, F.—Schwenk, H.—Bengio, Y.: Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1724–1734, doi: 10.3115/v1/D14-1179.

[25] Sutskever, I.—Vinyals, O.—Le, Q. V.: Sequence to Sequence Learning with

Neural Networks. Advances in Neural Information Processing Systems 27 (NIPS 2014), 2014, pp. 3104–3112.

[26] Liu, T.—Zhang, Y.—Qi, L.—Dong, J.—Lv, M.—Wen, Q.: WaveNet: Learning to Predict Wave Height and Period from Accelerometer Data Using Convolutional Neural Network. IOP Conference Series: Earth and Environmental Science, Vol. 369, 2019, Art. No. 012001, doi: 10.1088/1755-1315/369/1/012001.

[27] Hochreiter, S.—Schmidhuber, J.: Long Short-Term Memory. Neural Computation, Vol. 9, 1997, No. 8, pp. 1735–1780, doi: 10.1162/neco.1997.9.8.1735.

Yindong Zeng is Senior Researcher at the Marine Forecasting Center of Fujian Province, China. He is pursuing his Ph.D. degree in physical oceanography at the College of Oceanic and Atmospheric Sciences, Ocean University of China. He is responsible for research and application of the operational marine numerical forecast. His current research interests include ocean numerical model, machine learning, and the application of artificial intelligence.

Jinshan Ma is the postgraduate student of College of Computer and Information Sciences, Fujian Agriculture and Forestry University. His research focuses on artificial intelligence.

Xinping Chen is Senior Researcher at the National Marine Hazard Mitigation Service, China. He received his Ph.D. degree in ocean science from University of Hamburg, Germany. His main research interests include ocean engineering, marine environment forecasting, marine disaster reduction and big data analysis. He is Project Leader of various industrial, national and international projects in the domain of marine environment forecasting, marine disaster prevention and hazard assessment.

Minquan Guo received his M.Sc. degree in environment science in 2010 at the College of Marine and Environment, Xiamen University, Xiamen, China. He is currently Senior Engineer of the Marine Forecasting Center of Fujian Province. His current research interest includes numerical simulation of ocean currents and intelligent data analysis.

Zaichang Ren received his B.A. degree in marine science and technology in 2008 at College of Oceanic and Atmospheric Sciences, Ocean University of China. He is currently engineer of Marine Forecasting Center of Fujian Province. His current research interest includes intelligent data analysis and marine forecasts.

Yanqi Jiang is a master student of College of Harbour, Coastal and Offshore Engineering, Hohai University. Her research focuses on ocean wave mechanism studies and forecasting technology based on AI.

Zhenchang Zhang is Associate Professor of the College of Computer and Information Sciences, Fujian Agriculture and Forestry University, China. He obtained his B.Sc., M.A. and Ph.D. degrees from the Xiamen University. During 2017, he worked at the University of California, Los Angeles as Visiting Scholar. His research focuses on artificial intelligence, 4DVar data assimilation, parallel computing. He has extensive academic and industry experience in computer sciences. His work has been published in prestigious academic journals and conferences.