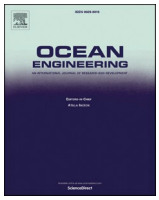
Ocean Engineering 267 (2023) 113218

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng



Instantaneous prediction of irregular ocean surface wave based on deep learning

Gang Xu [a,b], Siwen Zhang [a], Weichao Shi [c,*]

[a] School of Naval Architecture and Ocean Engineering, Jiangsu University of Science and Technology, Zhenjiang, 212100, China
[b] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, 519000, China
[c] Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde, Glasgow, UK

ARTICLE INFO

Keywords:
Deep learning
CNN-LSTM
Irregular wave
Instantaneous prediction

ABSTRACT

Ocean surface wave environment due to its randomness has mostly been analyzed and understood from a statistical way. Accurate and reliable wave prediction is the basic guarantee for ocean transportation and offshore operation. However, the significant nonlinearity of the ocean surface wave in real sea will result in a simulation and prediction challenge. To solve the problem, this paper proposes a Deep Learning instantaneous prediction model based on a Convolutional Neural Network-Long Short Term Memory model (CNN-LSTM) and wavelet function proven as the activation function of the neural network model to strengthen the fitting of the nonlinear mapping relationship between waves appearing in two successive different time periods. And the results shows that the wavelet activation function has better generalization ability than the traditional activation function for irregular wave data. In the numerical verification test, the two-parameter spectrum of International Towing Tank Conference (ITTC) and the combined model of extreme wave and random wave are used to simulate the irregular wave. The standardized time series of free surface elevation is made into a data set for model training and verification. In order to improve the performance of the model as accurate as possible, main parameters of the model are investigated in this paper. The optimal parameters are obtained and the model tuning suggestion is developed for future studies.

1. Introduction

Ocean surface wave environment is highly random and complex. It is dominated by wind but also affected by ocean current, ocean terrain and so on oceanographic elements. It adversely affects ships and offshore structures with life threatening risks. Therefore, an accurate prediction of ocean wave environment has critical socioeconomic significance.

However, traditional ocean wave analysis and study focuses on statistical analysis trying to understand the waves from the frequency domain. It provides useful and accurate representation of ocean waves statistically over long period. However, it lacks the ability to instantaneously predict the upcoming waves which are critical for ships and offshore structures to react in order to avoid damages. Therefore, how to accurately grasp the rapidly changing wave environment needs further study.

Nowadays, we are in the era of big data and artificial intelligence. Deep learning has already been successfully applied to many marine research. Artificial neural network (ANN) in the field of deep learning could be a solution for instantaneous wave prediction. ANN is derived from the simulation of the human brain nervous system and has strong self-learning and adaptive capabilities. ANN is widely used in the study of nonlinear systems. It can predict the regularity of network traffic (Zhao et al., 2019), and it can also predict crude oil production (Chao et al., 2019). Accurate ocean current forecasting can be accomplished by using ANN (Zhang et al., 2018), and a neural network with wavelet activation function can achieve effective forecasting of ship motion in waves (Huang, 2019). The analysis of the mooring system and the riser system has been carried out based on ANN (De Pina et al., 2013). However, in terms of intelligent ships and marine environment forecasting based on deep learning, it still needs to be developed (Li et al., 2021; Peng et al., 2020; Wang et al., 2021).

For the wave prediction, ANN was successfully applied to predict the significant wave height in the east coast of India (Deo et al., 2001), as well as the prediction of significant wave height and zero-up-crossing wave period for four different warning time horizons (Makarynskyy et al., 2005). Moreover, ANN has also been applied to the buoy

* Corresponding author.
E-mail address: weichao.shi@strath.ac.uk (W. Shi).

https://doi.org/10.1016/j.oceaneng.2022.113218
Received 3 May 2022; Received in revised form 4 November 2022; Accepted 16 November 2022

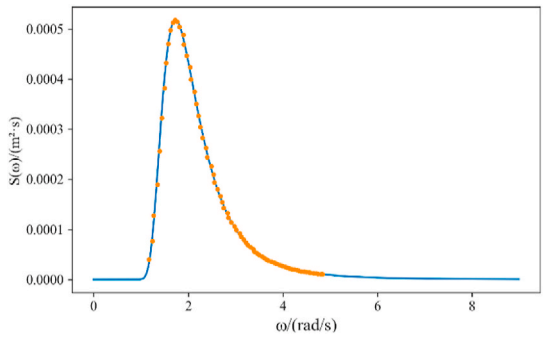
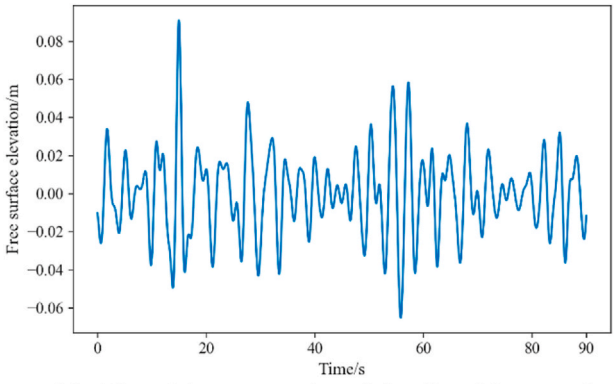
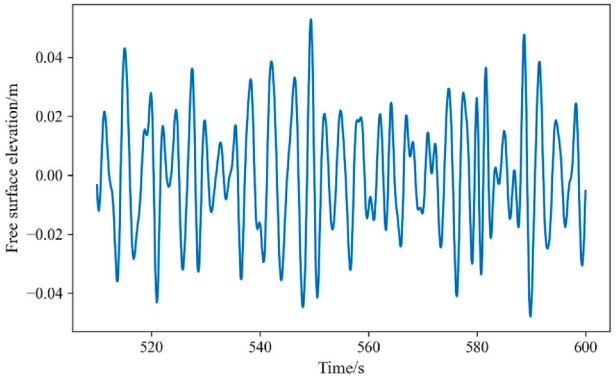


Fig. 1. ITTC two-parameter spectrum.



(a) Time-history results of the first 90 seconds



(b) Time-history results of the last 90 seconds

Fig. 2. Free surface elevation of the irregular wave.

prediction problem in offshore wave power generation (Castro et al., 2014), which can also be used to predict offshore wind and wave (Abhigna et al., 2017). When analyzing historical time series data, recurrent neural networks are generally the better choice (Sadeghifar et al., 2017). There are also many research results in the literature on the prediction of energy flow in the marine environment (Sanchez et al., 2018). Furthermore, in terms of short-term and long-term prediction, ANN also have an excellent prediction effect (Prahlada, 2012; Gómez-Orellana et al., 2022; Guijo-Rubio et al., 2020; Wang et al., 2018; Weilisi and Kojima, 2022).

The ANN takes historical data as input and self-adjusts the network parameters of ANN through continuous training and learning, and then inputs the test data to make predictions. The ANN prediction method finds out the inherent relationship between data samples through its structure learning, which saves a lot of time in analyzing data and modeling, and is to be operated easily. ANN can analyze any observed sequent data, but the ANN method also has some inherent shortcomings, like some network structure parameters need to be set empirically and the input dimension of ANN has the problem of "dimension curse".

In this work, an ANN model with multiple inputs and multiple outputs based on CNN-LSTM and wavelet activation function is designed and used to predict the time series of irregular ocean surface wave elevation. The calculation process from the waveform of the first few seconds to the waveform of the next few seconds is nonlinear, which is one of the fundamental reasons for using ANN, a model that is good at fitting nonlinear mapping relationships. Moreover, in mathematics experiments in this paper, the CNN-LSTM has a stronger fitting ability than classical model architectures such as traditional LSTM (the main reason is placed in Appendix 4), so this architecture was selected. In order to achieve the prediction effectively, the important parameters of the model will be discussed and analyzed in the study. The main contents of this paper are below: Section 2 introduces the theoretical background and the mathematics experiment process of this work, including the mathematical model of the irregular wave used to simulate ocean surface waves, the proposed ANN model based on CNN-LSTM and wavelet activation function. Section 3 compares and analyzes the prediction performance of the model with different main parameters, and in order to more objectively evaluate the prediction performance of the model, a comparison with the classic model is also set. Section 4 presents a brief conclusion based on numerical results.

2. Models and methods

2.1. Mathematical model of wave

The data used in this study comes from the mathematical model of the ocean surface wave. Firstly, the frequency equipartition method is used to select a suitable number of circular frequency-power pairs from the wave spectrum. Then, the selected circular frequency-power pairs are substituted into the combined model of extreme wave and random wave to obtain the irregular free surface elevation in time domain with significant nonlinearity, which is exported as the final required input and validation data.

2.1.1. Wave spectrum

The ITTC two-parameter spectrum used in this paper is a widely used wave power spectrum. It is a semi-theoretical and semi-empirical wave spectrum derived based on the observation of actual wave, which can describe the wave inherent distribution of the energy versus frequency. It is often used to approximate the real irregular wave environment in the mathematics experiment.

The spectral density function of the ITTC two-parameter spectrum is as follows:

$$S(\omega) = \frac{173H_{1/3}^2}{T^4\omega^5}\exp\left(-\frac{691}{T^4\omega^4}\right) \tag{1}$$

According to Eq. (1), when the significant wave height $H_{1/3}$ and the spectral peak period T are determined, the circular frequency ω and its corresponding spectral density $S(\omega)$ can be obtained. Among them, the circular frequency ω and wave number k satisfy the dispersion equation:

$$\omega^2 = gk\tanh(kh_0) \tag{2}$$

Using the sixth-order Pade approximation, an accurate approximate solution of the dispersion equation can be obtained (Hunt, 1979):

$$kh_0 = \sqrt{y^2 + \frac{y}{1+\sum_{n=1}^{6} c_n y^n}} \tag{3}$$

Here, h_0 is the depth of still water, which is set to 3 m in this study. $y=\omega^2 d/g$, $c_1 = 0.6666666666$, $c_2=0.3553553553$, $c_3=0.1608465608$, $c_4=0.0632098765$, $c_5=0.0217540484$, $c_6=0.0065407983$. Under any water depth, the relative errors of Eq. (3) are less than 0.2%. According

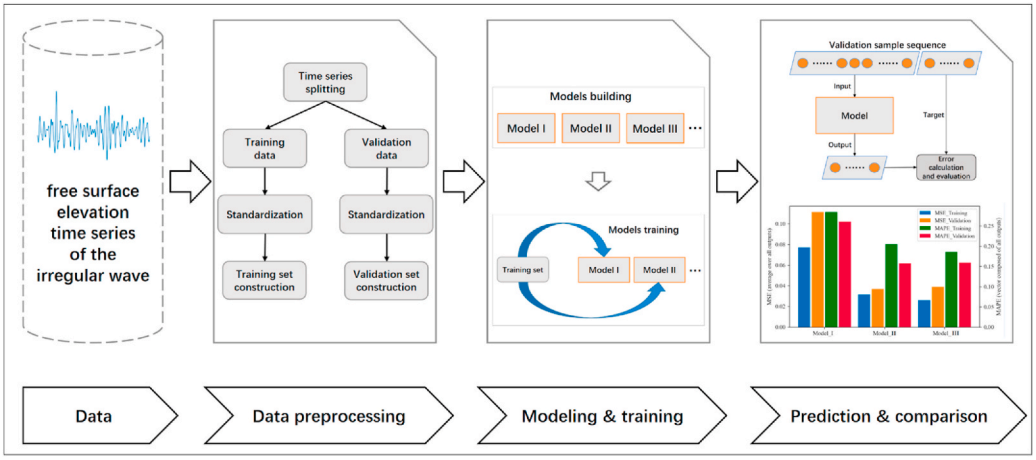


Fig. 3. The technical route of the research in this paper.

Table 1
The structure and parameters of the neural network.

Layers	Initial values of related parameters
Conv1d_1	filters: N_KERNEL_CONV, kernel_size: 3, strides: 1
Batch_normalization	None
Activation	Mortlet
Dropout	rate: RATE_DROPOUT
Conv1d_2	filters: N_KERNEL_CONV, kernel_size: 3, strides: 1
Batch_normalization	None
Activation	Mortlet
Conv1d_3	filters: N_KERNEL_CONV, kernel_size: 3, strides: 1
Batch_normalization	None
Activation	Mortlet
Concatenate	input: top-level input and the above three activation outputs
LSTM_1	units: N_KERNEL_LSTM1, return_sequences: True
Dropout	rate: RATE_DROPOUT
LSTM_2	units: N_KERNEL_LSTM2
Dropout	rate: RATE_DROPOUT
Activation	Mexican_hat
Batch_normalization	None
Conv1d_4	filters: 1, kernel_size: 3, strides: int(557/(N_KERNEL_LSTM2-1)), input: top-level input
Flatten	None
Batch_normalization	None
Add	input: outputs of the last two batch_normalization above
Dropout	rate: RATE_DROPOUT
Dense	units: 140
Activation	Gaussian
Dense	units: 280

Table 2
Default values of key parameters.

Names	Values
N_KERNEL_CONV	5
N_KERNEL_LSTM1	128
N_KERNEL_LSTM2	64
RATE_DROPOUT	0.2

to Eq. (3), the wave number k can be obtained, which is the foundation for the following work on simulating irregular wave.

Based on the ITTC two-parameter spectrum in Eq. (1), a program code is written to calculate the wave spectrum, as shown in Fig. 1. The significant wave height and peak spectral period are set to 0.1 m and 2.8 s. The parameter settings can be achieved in the physical tank, which is

Table 3
Replacement of activation functions.

Wavelet functions	Classical functions in neural network
Mortlet	Elu
Mexican_hat	Elu
Gaussian	Tanh

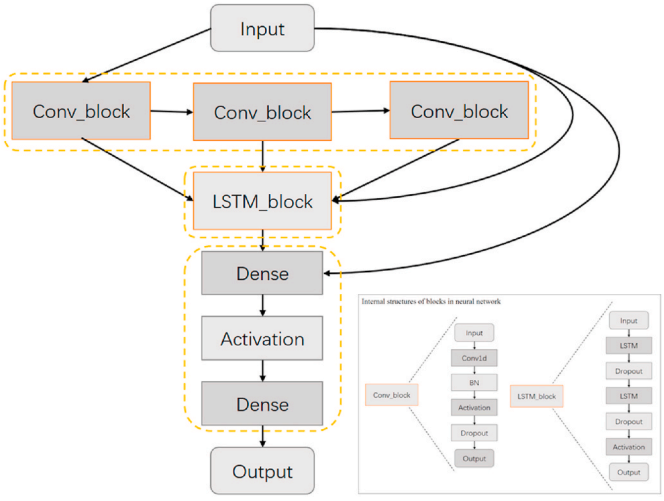


Fig. 4. Neural network structure diagram.

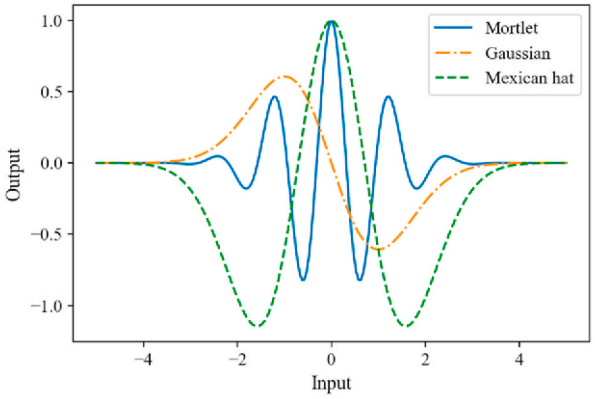


Fig. 5. Classical wavelet functions.

Table 4
Program parameters configuration.

Parameters	Values
TensorFlow random seed	521
Loss function	Mean Squared Error about mean of all output steps
Optimizer for model training	Adam
Training learning rate	0.001
Default batch size	64
Training epochs	80

Table 5
Values of the parameters.

Model names	N_KERNEL_CONV
Model_I1	5
Model_I2	16
Model_I3	32

also to facilitate the validation of the conclusions of this paper in a hydrodynamic testing facility, i.e. a towing tank or an ocean basin in the near future. In addition, regardless of how these parameters are set, this has no effect on the research method and corresponding conclusions in the paper. In order to improve the simulation efficiency, 75 combined waves with spectral density above 10^{-5} are selected, and the frequency range of energy concentration is obtained. The orange dot in Fig. 1 is the frequency corresponding to the wave components based on the wave spectrum.

2.1.2. Combination wave model

In the simulation of real ocean wave, the extreme wave and random

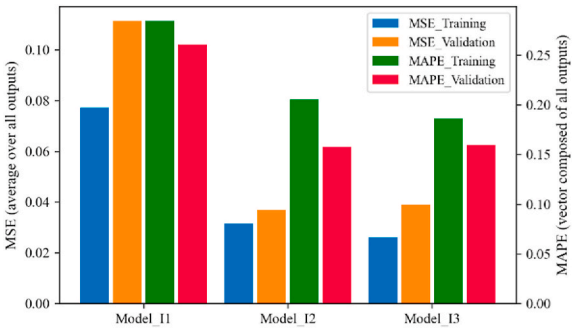


Fig. 7. Model prediction error comparison.

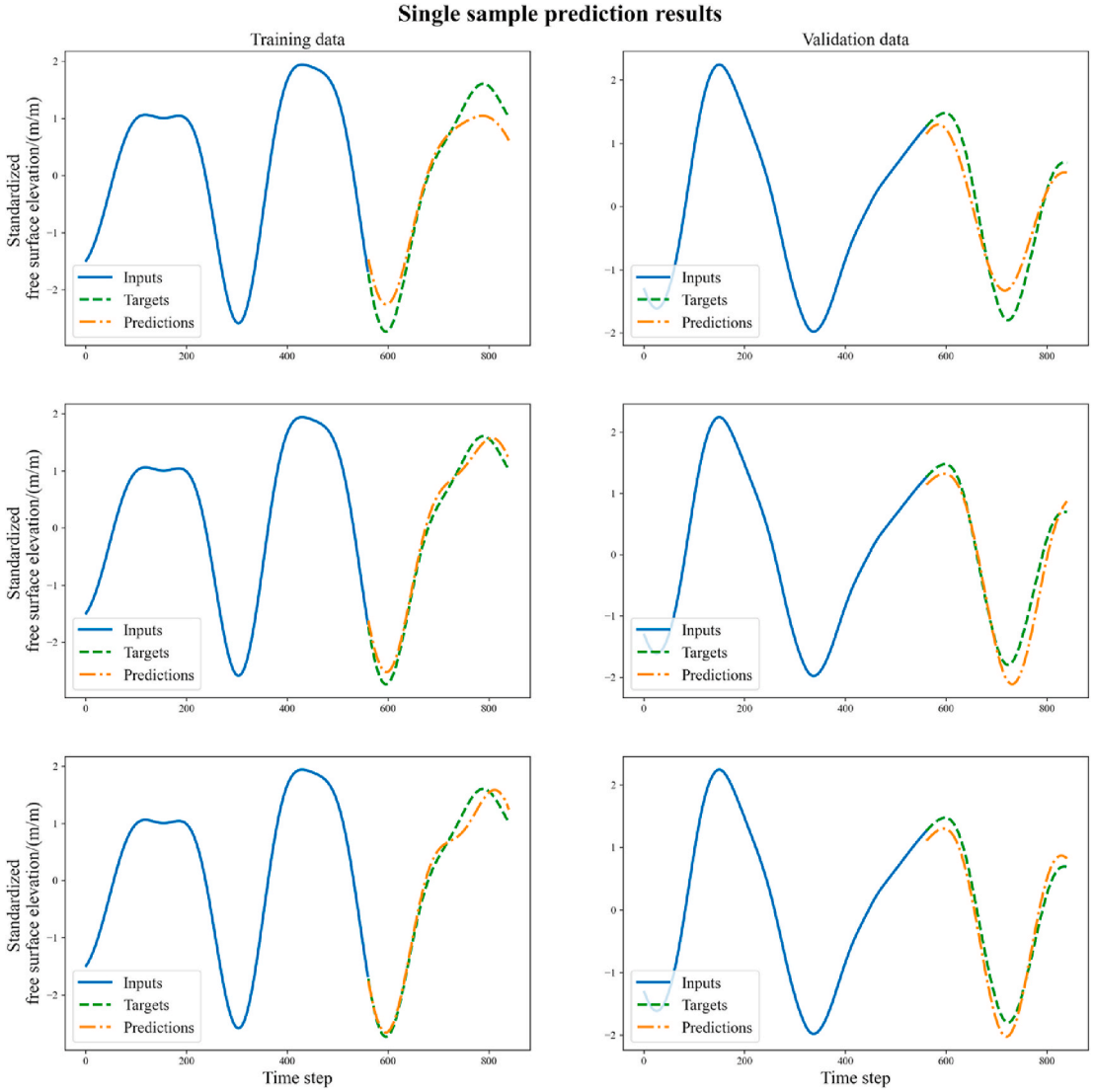


Fig. 6. Single sample prediction results.

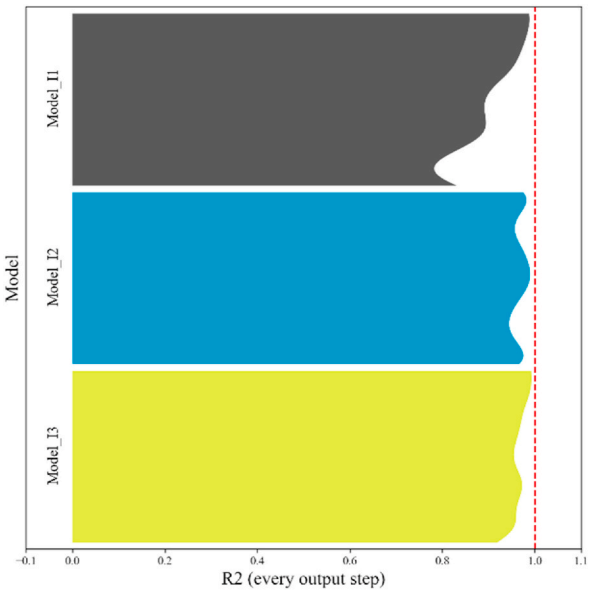


Fig. 8. R2 expanded by time step.

Table 6
Values of the parameters discussed.

Model names	N_KERNEL_LSTM1	N_KERNEL_LSTM2
Model_II1	64	32
Model_II2	128	64
Model_II3	256	128

wave combination model is adopted in this paper (Zhao et al., 2009). Based on the principle of linear superposition, the model can satisfy the mathematical description of irregular wave required in this work:

$$\eta(x,t) = E_p \sum_{i=1}^{N_f} a_i \cos\left(k_i(x-x_p) - \omega_i(t-t_p) + \varepsilon_i\right) + (1-E_p)\sum_{i=1}^{N_f} a_i \cos\left(k_i(x-x_p) - \omega_i(t-t_p)\right) \tag{4}$$

In Eq. (4), a_i, k_i and ω_i are the amplitude, wave number and circular frequency of the *ith* wave. The ε_i is the random phase of the *ith* wave, which is the random number in (0, 2π). E_p is the energy ratio coefficient, which is used to adjust the degree of energy focus of the generated wave, and it is set to 0.7 in this paper. x_p and t_p is the location and time of the largest peak value, which are taken as 10 m and 15 s in this paper, and the position variable x will also be taken as 10 m.

Among them, the parameter E_p that can be flexibly adjusted, which allows us to represent a variety of waves conveniently. The flexible and diverse representation of the wave will be conducive to further validation of the developed prediction model. Besides, in the combination wave model, the degree of wave energy focus can be increased by reducing the E_p (experiments show that freak wave can be generated when it is generally less than 0.6). Increasing the degree of wave energy focus as much as possible can increase the difficulty of the model prediction task, which is more beneficial for validating the prediction performance and generalization of the model. But the work of the paper is not a prediction of freak wave, therefore, the E_p is finally set to 0.7.

The wave amplitude in Eq. (4) can be obtained by wave spectrum:

$$a_i = \sqrt{2S(\omega_i)\Delta\omega_i} \tag{5}$$

For $\Delta\omega_i$, it is the frequency interval in the frequency equipartition method.

The time series of final irregular wave is shown in Fig. 2. The total duration of the simulated wave is 10 min, and the time step is 0.01 s. For ease of presentation, only the first 90 s and the last 90 s of data are shown in Fig. 2.

2.2. Data processing and ANN modeling

In this part, the specific technical map of this paper will be described. Fig. 3 shows the technical route of the study in this paper. In the study of this paper, the time series of free surface elevation for irregular wave needs to be split firstly. One part is used for subsequent model training, and the other part is used for the validation of the final model prediction performance. After splitting the time series, standardize the two parts of the data, separately. It is an essential data preprocessing procedure. ANN is more sensitive to standardized distributed data and easier to capture the inherent laws in the sample. The training set and the validation set can be constructed separately from the two standardized time series. Then build several groups of ANN models with different parameters, and use the training set to train these models. Finally, use these trained models to predict the validation set one by one, and calculate the performance metrics. After a systematic comparison and analysis of these models, the optimal combined parameter of the model will be obtained, and corresponding model tuning concept will be proposed. The main steps in the mathematics experiment procedure are described in details below.

2.2.1. Data standardization

Data standardization can transform the data into a standard distribution with a mean value of 0 and a variable value of 1. The neural network is more sensitive to values distributed around 0, so standardization is beneficial for the neural network to fit the data, allowing the neural network to better capture the internal laws of the sample. The data standardization formula is as follows:

$$x_i^* = \frac{x_i - \overline{x}}{s} \tag{6}$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7}$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{8}$$

Where x_i^* represents the standardized result of the original data x_i, $\overline{x}$ is the mean value of the original data, and s is the standard deviation of the original data.

2.2.2. Production of data sets

This part describes specific methods for making time series into datasets.

Assuming that the given complete time series $L = \{x_1, x_2, \cdots, x_N\}$ has N nodes, for the prediction model, each sample is a time series of length $k + l$. k time nodes $\{x_1, x_2, \cdots, x_k\}$ are inputs, and the predicted value of the subsequent l time nodes $\{x_{k+1}, x_{k+2}, \cdots, x_{k+l}\}$ are outputs, the prediction model will fit the mapping between the them. Therefore, $N - (k+l) + 1$ time series samples constructed from L are as follows:

$$\{x_1, x_2, \cdots, x_{k+l}\}, \{x_2, x_3, \cdots, x_{k+l+1}\}, \cdots, \{x_{N+1-(k+l)}, \cdots, x_N\}$$

The prediction strategy in this paper is "two periods of free surface elevation data are used as input to predict the free surface elevation data of the following one period". Besides, the peak period of wave spectrum is 2.8s and the time step size is 0.01s, so the "one periods" corresponds to 280 time steps. Therefore, the number of model input steps should be set to 560, and the number of model output steps should be set to 280.

Since the total duration of the irregular wave is 10 min, and the time-series length of the training data is 49,300 (about the first 82% of the numerical wave, and the validation set is about the last 18%), the

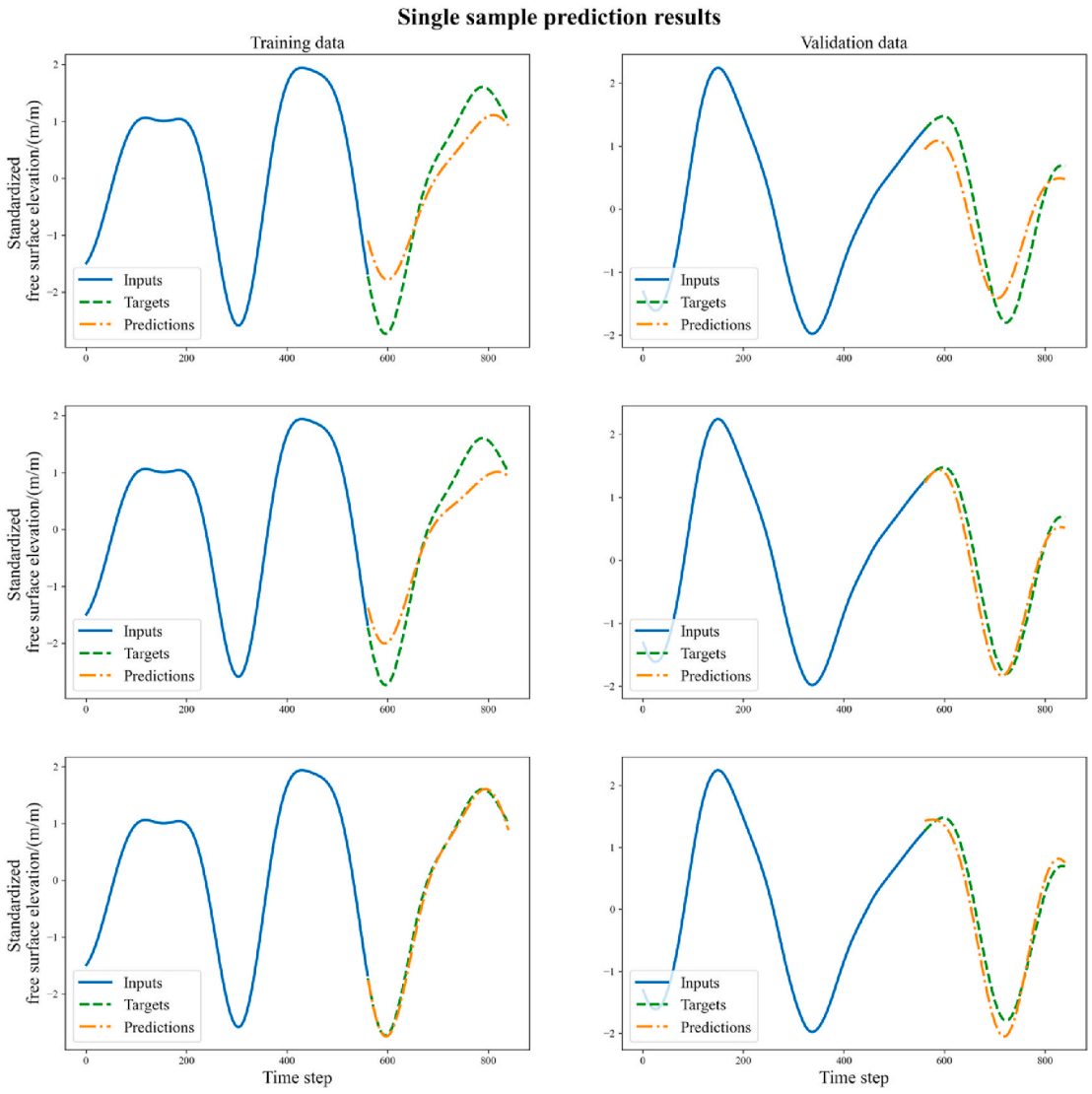


Fig. 9. Single sample prediction results.

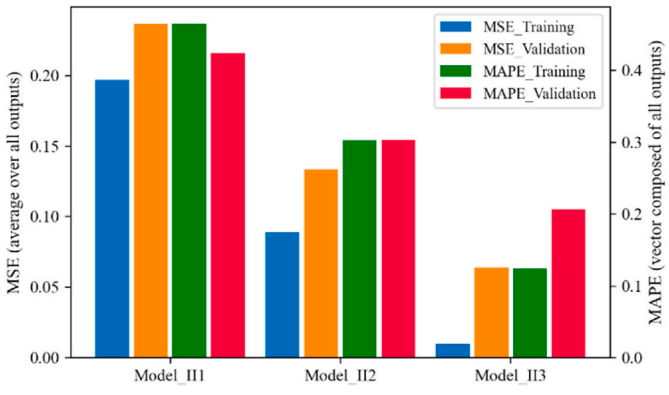


Fig. 10. Model prediction error comparison.

training set and validation set contain 48461 and 9861 samples, respectively.

2.2.3. Model structure and initial parameters

Table 1 is the overall structure of the neural network and the initial setting of related parameters. There are several constants represented by names, which are the key research parameters in this paper. Their values will be discussed and determined in this paper, and the initial default values are shown in Table 2. Table 1 contains three activation functions, which are wavelet functions. In the following study of activation functions, wavelet functions are replaced by activation functions, which are commonly used in neural network models. The replacement relationship of functions is shown in Table 3.

In order to show the structure of the neural network more clearly, Fig. 4 uses a simple form to describe the main structure of the neural network and the internal structures of the blocks in the neural network.

2.2.4. Wavelet function

Three classical wavelet functions are used in the process of establishing the ANN model. It has shock waves with fast decay and limited length, which are often used in wavelet analysis. It should be used as activation functions of the neural network to further improve the ability of the neural network to fit nonlinear data mapping. Fig. 5 shows the waveform of the three wavelet functions. It can be seen that the wavelet functions have quickly convergent characteristics in the direction of the horizontal axis. The expressions of the three wavelet functions are as follows:

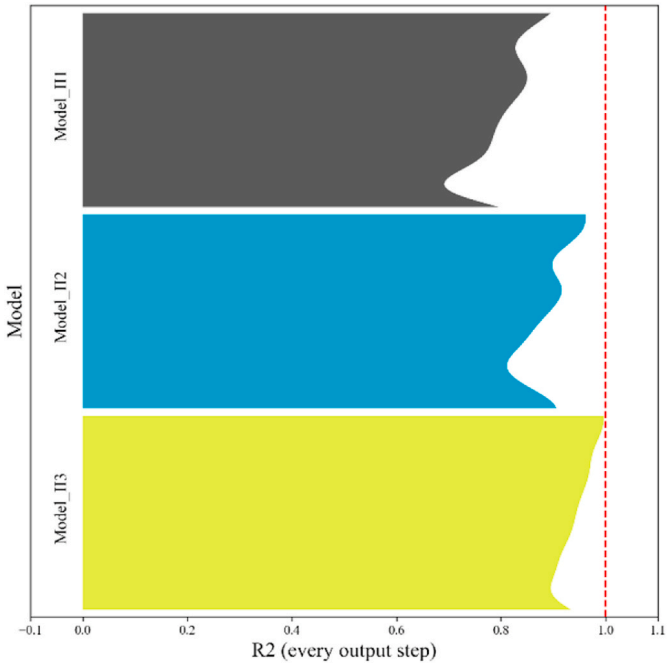


Fig. 11. R2 expanded by time step.

Table 7
Values of the parameters discussed.

Model names	RATE_DROPOUT
Model_III1	0.2
Model_III2	0.3
Model_III3	0.4

$$\text{Mortlet wavelet } \psi(x) = \cos(5x)\cdot\exp\left(\frac{-x^2}{2}\right) \tag{9}$$

$$\text{Gaussian wavelet } \psi(x) = -x\cdot\exp\left(\frac{-x^2}{2}\right) \tag{10}$$

$$\text{Mexican Hat wavelet } \psi(x) = \left(1-2x^2\right)\cdot\exp\left(\frac{-x^2}{2}\right) \tag{11}$$

2.2.5. Performance metrics

The performance metric is the important basis for quantitative evaluation of model prediction. The metrics used in this paper are R Squared (R2), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE), which describe the model prediction bias. It will represent the accuracy of model prediction from different angles.

The R2, also known as the coefficient of determination, is a classical model performance metric in the fields of statistics and deep learning. It indicates how much the predictive value explains the distribution of the true value, which is used to measure the overall fitting degree of the predictive value to the true value. Given true values y_i of all samples and the corresponding predictive value $\hat{y}_i$, the R2 is defined as:

$$R2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{12}$$

In Eq. (12), $\bar{y}$ is the average of the true value of all samples, N is the number of samples. The value range of R2 is $(-\infty, 1]$, which is close to 1, indicating that the model fits the data better. In this paper, R2 will be evaluated separately for each output step of the model.

MSE and MAPE represent the distance between two points in space based on the Euclidean distance and the Manhattan distance, respectively. Given true values y_i of all samples and the corresponding predictive value $\hat{y}_i$, MSE and MAPE are defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{13}$$

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{14}$$

The both values are greater than or equal to 0. The closer to 0, the smaller the prediction error and the better the performance of the model. In this paper, MSE evaluates the mean of all output steps, and MAPE evaluates the vectors in the high-dimensional space composed of all output steps.

2.2.6. Program parameters

In the process of the programming and the model training, it has important parameters that need to be determined. The configuration of these parameters in this study is shown in Table 4.

3. Results and discussion

In this part, the important parameters of the ANN model established in this paper will be discussed. These parameters are the hyper parameters of the model, which have a great impact on the prediction performance of the model, but the parameter settings lack systematic theoretical guidance, and often require multiple attempts based on experience to make the model have ideal prediction performance.

Sections 3.1 to 3.4 will discuss the parameters of the network structure, and section 3.5 will discuss the parameters of the model training.

3.1. Comparisons of the convolution kernel number

In this section, models based on different convolution kernel number will be compared and the parameters will be investigated, and the name of this parameter is specified as N_KERNEL_CONV. The values are shown in Table 5.

A sample is randomly selected from the training set and the validation set, respectively, and the prediction results of the models are shown in Fig. 6. The first column in Fig. 6 is the prediction of one sample in the training set, and the second column is the prediction of one sample in the validation set, from top to bottom are the predictions of the Model_I1 to the Model_I3. The first 560 steps are the input data of the models, and the last 280 steps are the prediction data output by the models and the true value of the prediction target.

Fig. 7 shows the MSE and MAPE of the models on the training set and validation set. And Fig. 8 shows the R2 values of the models at all time steps, and the 280 prediction time steps are arranged from top to bottom.

In Fig. 6, the three models have good prediction effects on randomly selected samples. The overall trend of the waveform and the appearance positions of the peak and trough have been effectively predicted, and the prediction bias of the phase and frequency is small. Among them, the Model_I2 and the Model_I3 have a relatively high degree of fitting to the data.

In Fig. 7, there is a phenomenon in all three models: the training error described by MSE is smaller than the validation error, and the validation error described by MAPE is smaller than the training error. In the comparison among the models, there is a decreasing trend of the four error metrics, among which the Model_I3 has the smallest error, and the error decreasing trend of the Model_I2 to the Model_I3 is significantly smaller than the error decreasing trend of the Model_I1 to the Model_I2, and the prediction accuracy tends to converge.

In Fig. 8, the R2 distribution over time steps of the Model_I2 and the

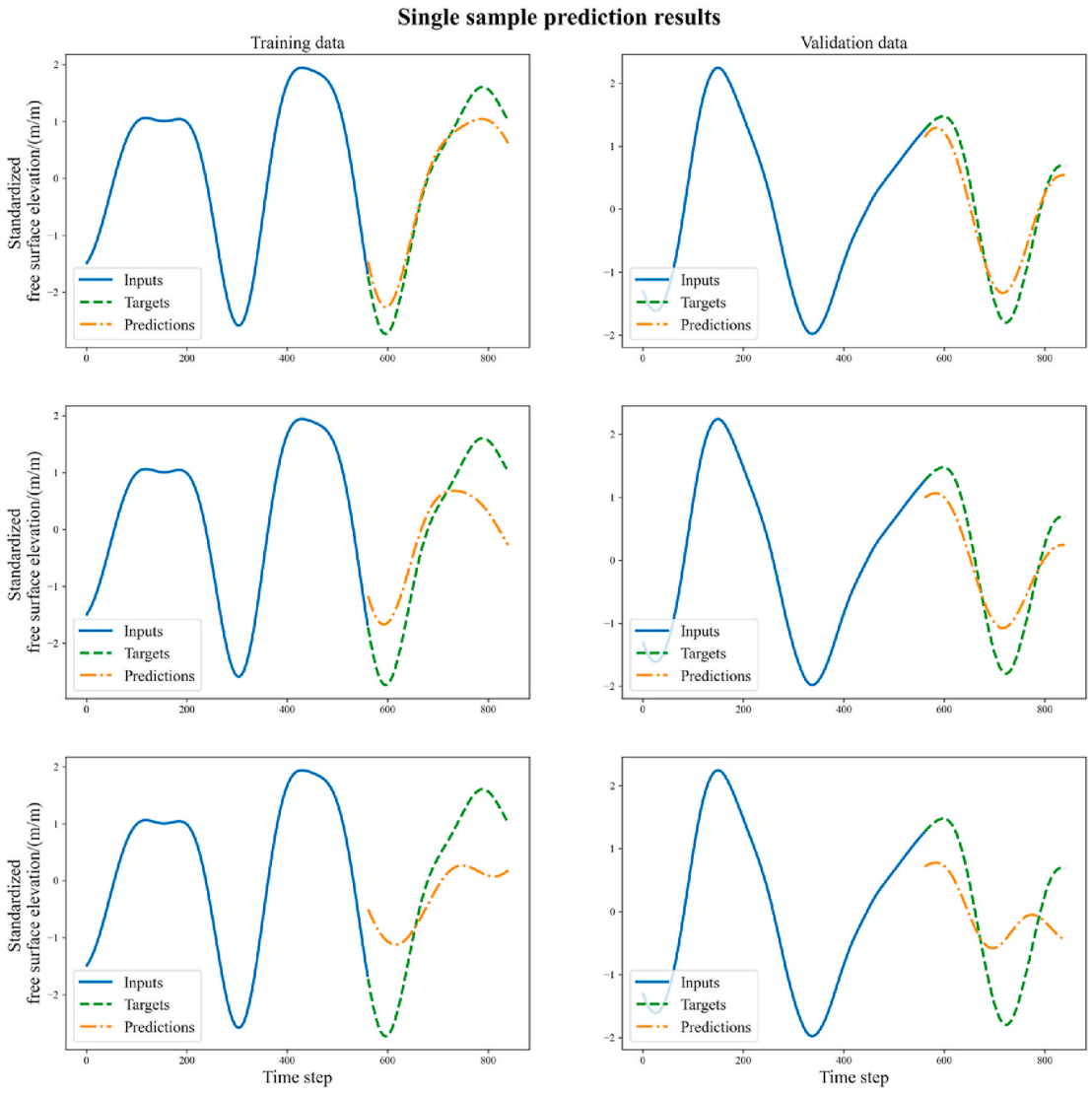


Fig. 12. Single sample prediction results.

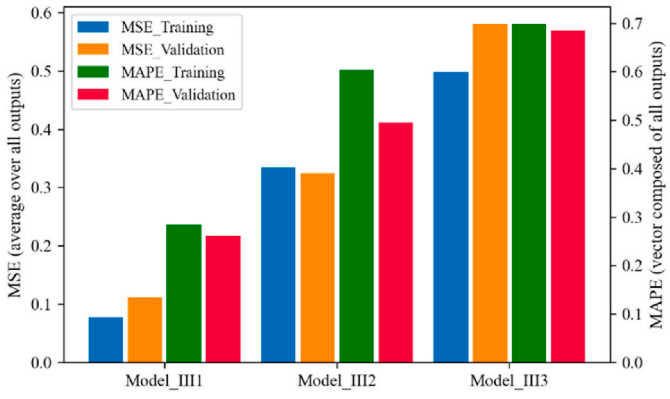


Fig. 13. Model prediction error comparison.

Model_I3 are uniform and similar, which confirms the convergence of prediction accuracy. The R2 of The Model_I1 obviously decreases in the direction of increasing time steps, that is, the relatively later time step has a lower fitting degree.

According to the comparison of the performance of the models on the training set and the validation set, it can be seen that none of the three models has overfitting. According to the distribution of R2 on the time step, it can be seen that the three models do not have a too low fitting degree for some time steps. On the whole, the prediction accuracy of the model has basically converged when the value of the convolution kernel number parameter is 16. If the value of this parameter continues to increase, the model capacity will be increased, on the one hand, the model operation speed will decrease, on the other hand, the model will be more prone to overfitting. In addition, an appropriate increase of this parameter can significantly improve the prediction ability of the model for later time steps.

The size of the number of convolution kernels directly affects the capacity of the model. The larger the parameter, the greater the capacity of the model, and the lower the probability of underfitting the training data. While larger models may generalize better, they also have a greater risk of overfitting. In summary, the number of convolution kernels is set to 16 is the most appropriate. When this parameter is set to 5, the fitting ability is insufficient due to the reduction of the model capacity, and the errors are obviously larger than the value of 16. When this parameter is set to 32, the model validation error is almost the same as when it is set to 16, or even slightly increases, which indicates that the model capacity is increased at this time, which not only does not bring the benefit of reducing the error, but increases the calculation amount and the risk of

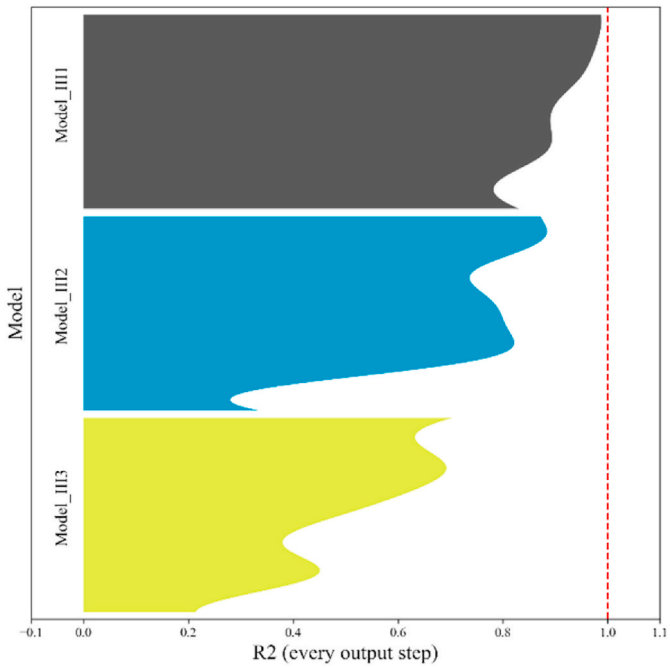


Fig. 14. R2 expanded by time step.

overfitting.

3.2. Comparisons of the unit number in the LSTM layer

In this section, models based on different unit number parameters in the LSTM layer will be compared, and the names of these parameters are specified as N_KERNEL_LSTM1 and N_KERNEL_LSTM2. The names of the models and their corresponding parameter values are shown in Table 6.

In Fig. 9, the three models have good prediction effects on randomly selected samples, the overall trend of the waveform and the appearance positions of the peak and trough have been predicted effectively, and the prediction bias of the phase and frequency is small. The Model_II3 fits the data best.

In Fig. 10, in all three models, the difference between the validation error and the training error increases with the increase of the parameters, but the increase does not have an excessive impact on the prediction performance of the model. In the comparison, All the four-error metrics have a significant decreasing trend, among which the Model_II3 has the smallest error, and the decreasing trend of the error has no obvious convergent phenomenon.

In Fig. 11, the R2 distributions of the Model_II1 and the Model_II2 are similar in the time step, and the three models all have a slight decrease in the fitting degree of the data at the later time steps. In the comparison, the R2 has an overall significant increasing trend, and the Model_II3 has the best fitting effect.

According to the comparison of the performance of the models on the training set and the validation set, it can be seen that there is no obvious overfitting phenomenon for the three models. According to the

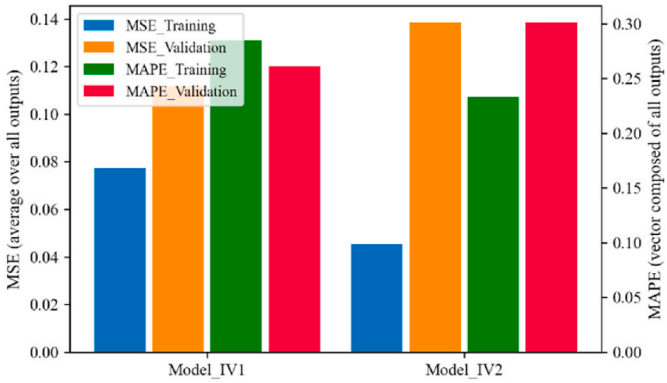


Fig. 16. Model prediction error comparison.

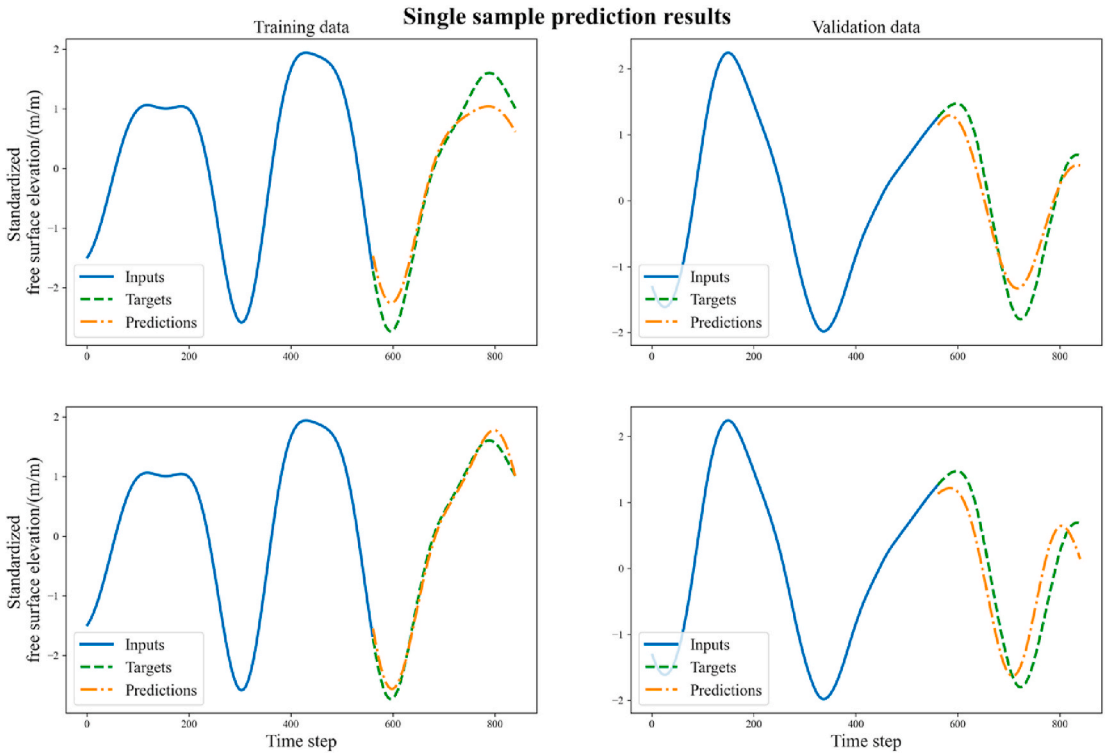


Fig. 15. Single sample prediction results.

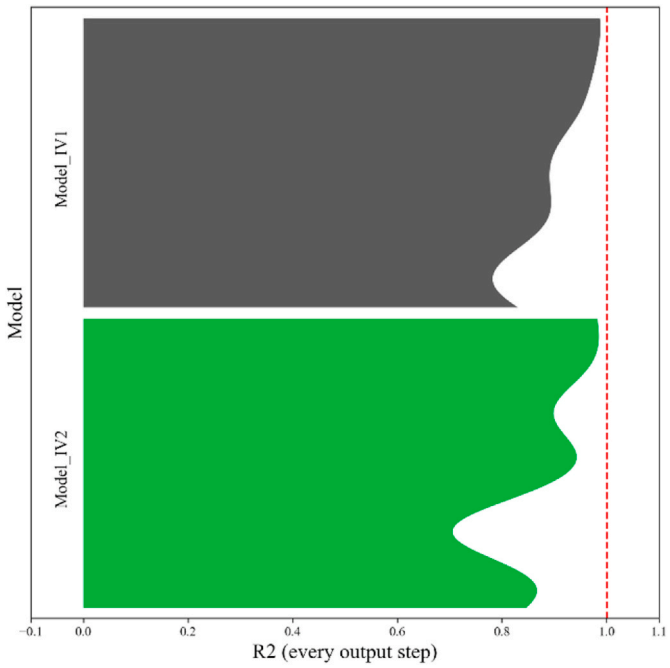


Fig. 17. R2 expanded by time step.

Table 8
Values of the parameters discussed.

Model names	Batch size
Model_V1	16
Model_V2	32
Model_V3	64
Model_V4	128
Model_V5	256

distribution of R2 on the time step, it can be seen that the three models have a lower degree of fitting to the relatively later time steps. In general, the model has the best prediction performance when the two-unit parameters in the LSTM layer are 256 and 128, respectively. At this point, the accuracy of the model is good enough.

The size of the numbers of units in the dual LSTM layer also directly affects the capacity of the model. When this parameter reaches the upper limit set in the mathematics experiment, the error of the model is still significantly reduced, because the capacity improvement of the model brings about the improvement of the model fitting ability, and does not reach the level of overfitting the model. To sum up, the numbers of units in the dual LSTM layer can be greater than or equal to 256 and 128 respectively.

3.3. Comparisons of the dropout rate in the dropout layer

In this section, models based on different dropout rate parameters in the Dropout layer will be compared, and the name of this parameter is specified at RATE_DROPOUT. The names of the models and their corresponding parameters are shown in Table 7.

In Fig. 12, only the Model_III1 has a good prediction effect on the randomly selected samples. The overall trend of the waveform and the appearance position of the peak and trough have been predicted effectively, and the prediction bias of the phase and frequency is small. The predictions of the other two models are not good.

In Fig. 13, the validation errors of the three models are not very different from the training errors. In the comparison, the four-error metrics all have a significant increasing trend, and only the error of Model_III1 is within a reasonable range.

In Fig. 14, the R2 values of the Model_III2 and the Model_III3 are obviously insufficient at some time steps, and the three models have a reduced degree of fitting to the data at the later time steps. Overall, in the comparison of the models, there is a significant decrease in R2, and only the Model_III1 has an ideal fitting effect.

According to the comparison of the performance of the models on the training set and the validation set, it can be seen that there is no obvious overfitting phenomenon for the three models. According to the distribution of R2 on the time step, it can be seen that the three models have a lower degree of fitting to the relatively later time steps. Overall, the model has the best prediction performance when the dropout rate parameter in the Dropout layer is 0.2. The function of the Dropout layer is to improve the overfitting phenomenon of the model, because there exists randomness when it works, as well as a greater number of the Dropout layer and a larger dropout rate will weaken the stability of the model to a certain extent. Therefore, reducing the number of Dropout layers and removing Dropout layers in structures other than Dense layers may improve the prediction performance of the model.

In summary, the model is more sensitive to the dropout rate of the Dropout layer. Since increasing this parameter, the model performance drops significantly from a good level, which indicates that the capacity of the model is appropriate under the characteristics of the data studied in this paper, and there is no need to take too many measures to suppress overfitting, otherwise it will underfit the model. In addition, in the follow-up more complex research, in order to improve the generalization ability of the model, we can try to increase the capacity of the model and at the same time increase the dropout rate of the Dropout layer to achieve a balance.

3.4. Comparisons of the activation function

In this section, two models are built, the Model_IV1 uses the wavelet functions as the activation functions, the Model_IV2 only uses the traditional activation functions.

In Fig. 15, both models have good prediction effects on randomly selected samples. The overall trend of the waveform and the appearance positions of the peak and trough have been predicted effectively, and the prediction bias of the phase and frequency is small.

In Fig. 16, according to the difference between the validation error and the training error, it can be seen that the Model_IV2 has some degree of overfitting, while the Model_IV1 has no obvious overfitting. In addition, the validation error of the Model_IV2 is larger than the Model_IV1, and its generalization ability is slightly lower than the Model_IV1.

In Fig. 17, the fitting degree of the two models to the data at the later time steps has decreased, and the fitting of the Model_IV2 is slightly insufficient at some time steps. Overall, the fitting effect of the Model_IV1 is slightly better than the Model_IV2.

In general, the wavelet activation function has better data mapping ability than the traditional activation function, has better prediction performance for data with nonlinear mapping relationship, and is more suitable for predicting the irregular ocean surface wave elevation in time domain.

In the neural network model of this paper, the action mechanism of the wavelet activation function is: first, the front-end network extracts the features of the input wave data, and then uses the extracted features as the parameters of the wavelet function. Finally, the wavelet functions with these different parameters are superimposed to generate the predicted wave. From the test results, compared with the traditional nonlinear activation function, the wavelet activation function is more effective for the instantaneous prediction of waves.

3.5. Further optimization on the training parameters

After the above discussion of model structure parameters, the model will be further optimized in this section. A key training parameter of the model will be discussed, which is the training batch size. If the

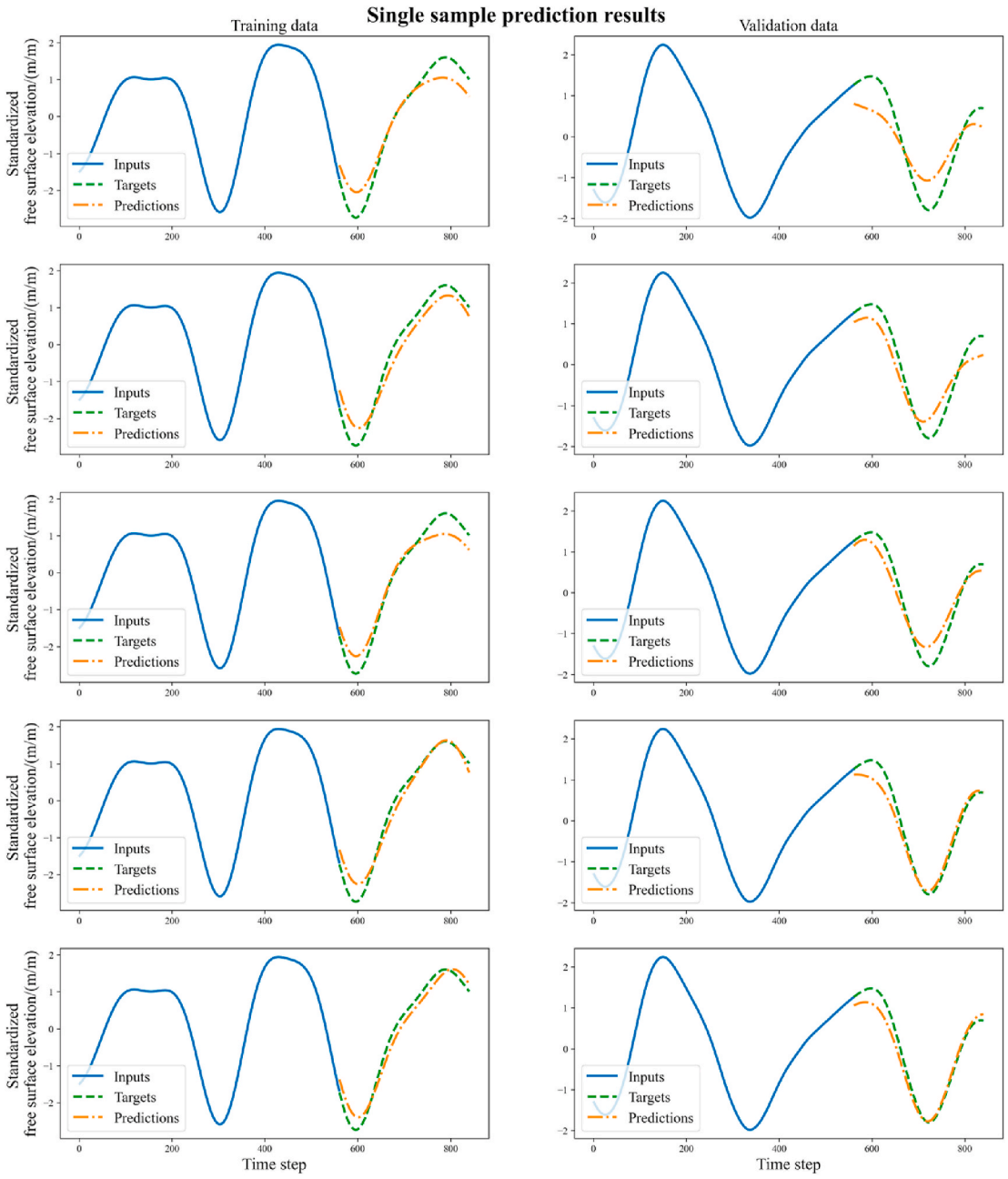


Fig. 18. Single sample prediction results.

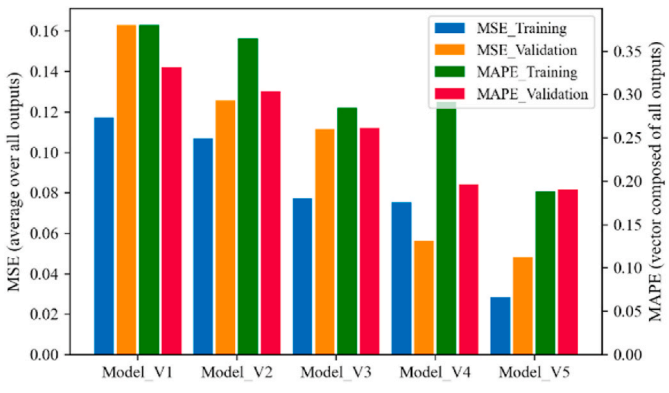


Fig. 19. Model prediction error comparison.

parameter is too small, the model will lose the iterative accuracy, and if it is too large, the steps per training epoch will be reduced and the model will lose generalization. The names of the models and their corresponding parameters are shown in Table 8.

In Fig. 18, the five models have good prediction effects on randomly selected samples. The overall trend of the waveform and the appearance positions of the peak and trough have been effectively predicted, and the prediction bias of the phase and frequency is small. On the validation samples, the model with a larger batch size predicts better, but this performance improvement is not too significant.

In Fig. 19, the validation errors within each of the five models are not very different from the training errors. In the comparison, the four-error metrics have a decreasing trend. The validation errors from the Model_V3 to the Model_V4 decrease the most seriously, and the validation errors from the Model_V4 to the Model_V5 decrease the least, tending to converge.

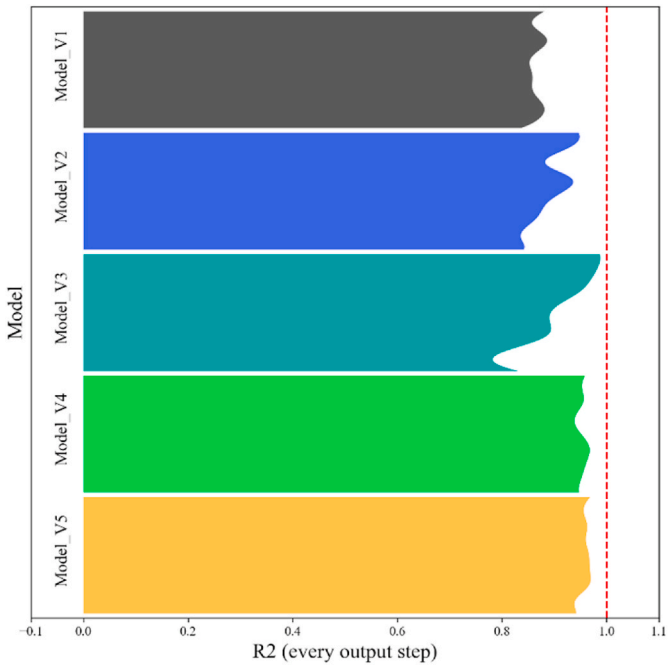


Fig. 20. R2 expanded by time step.

In Fig. 20, the R2 of the Model_V2 and the Model_V3 have an insufficient fit to the data at the later time steps, while the R2 of the other models are more uniformly distributed over time steps. Overall, R2 has a certain increasing trend, and the trend from the Model_V3 to the Model_V4 is the largest. The R2 distributions of the Model_V4 and the Model_V5 are very similar, and there are obvious signs of convergence.

According to the comparison of the performance of the models on the training set and the validation set, it can be seen that none of the five models has obvious overfitting. According to the distribution of R2 on the time step, only the Model_V2 and the Model_V3 have a low degree of fitting for the relatively later time steps. On the whole, the optimal value of the parameter batch size is 128. If the parameter is too small, the training samples of the model will be too miscellaneous, which will easily make the training of the model unstable. If the parameter is too large, it will greatly consume computing resources.

In summary, the batch size has a significant impact on the model performance. Generally speaking, the larger the value, the better the model performance, but there are also bottlenecks, and the excessively large batch size has extremely high requirements on the hardware. In the research data of this paper, when the batch size is 256, the model generalization is not significantly improved compared to 128, and it will consume more computing resources, so 128 is the most reasonable value.

3.6. More comparison of results

The above mainly expounds the details of model architecture design, parameter tuning and prediction performance comparison. In order to more objectively demonstrate the effectiveness and superiority of our prediction model, this section will use the model with the optimal parameters above (Model_best is the Model_I2) against a benchmark and the seq2seq which is a classic deep learning time series prediction model. Among them, the benchmark is defined as: taking the value of the last time step of the input ground-truth sequence for the prediction model as the value of all the time steps to be predicted, as shown in Fig. 21.

In Fig. 22, according to the comparison, it can be clearly seen that the errors of our model are much smaller than the benchmark and seq2seq. In Fig. 23, the benchmark values of R2 are almost all less than 0, which means that using the value at the last time step of the input sequence as the prediction value of the subsequent time steps has extremely poor effect, and this group of benchmark values is not meaningful for

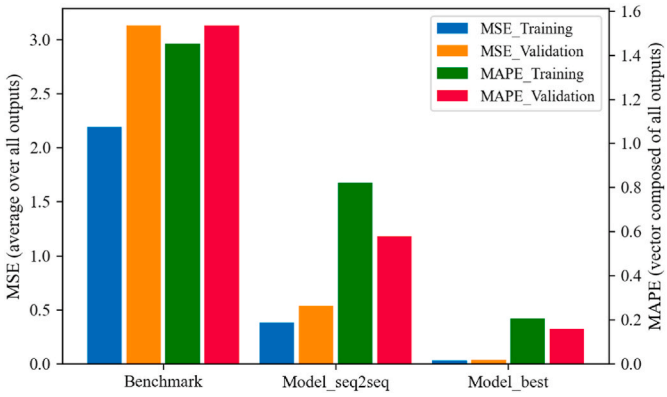


Fig. 22. Prediction error.

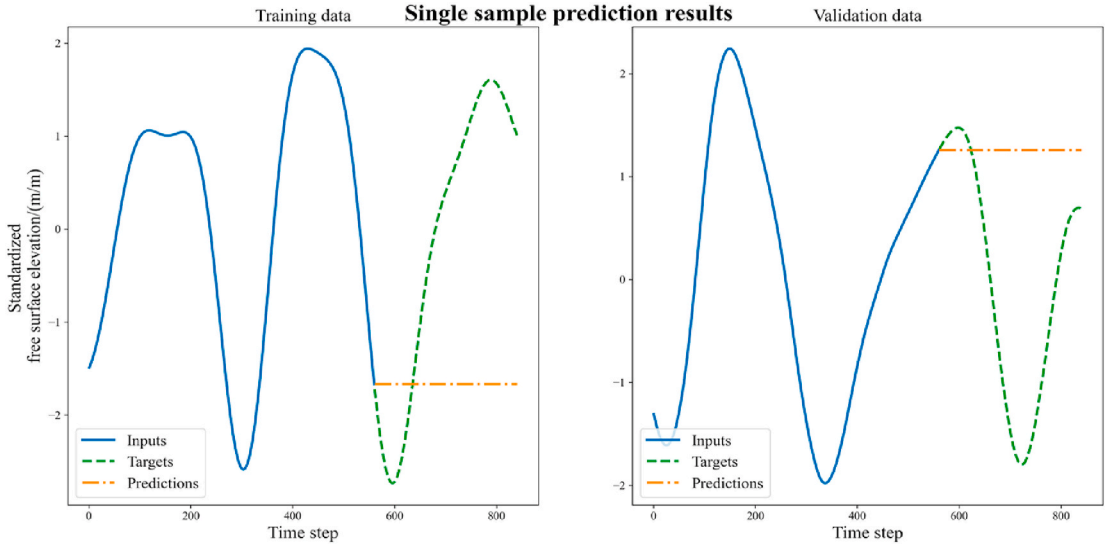


Fig. 21. Benchmark example.

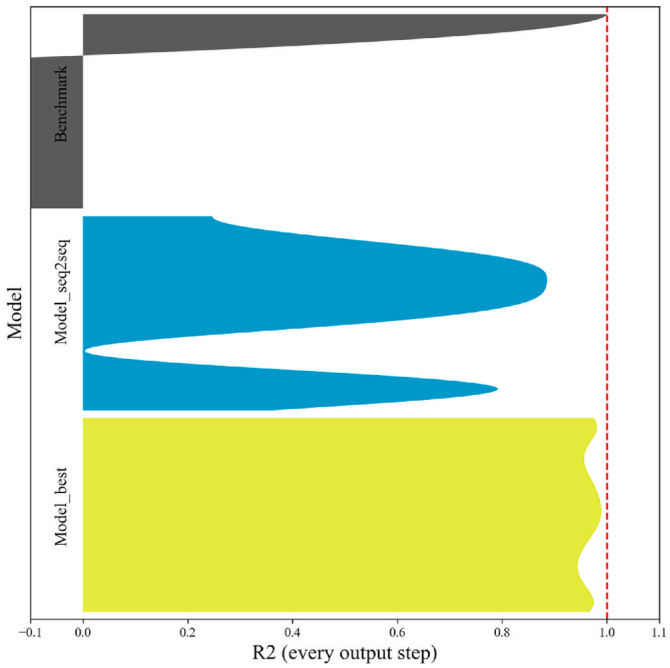


Fig. 23. R2 expanded by time step.

comparison. Compared with the R2 coefficient of seq2seq, the R2 of our model is maintained at a high level at each time step, while seq2seq only has a certain prediction ability for a specific part of the time steps, and the R2 coefficients are all smaller than our model.

4. Conclusions

Accurate wave prediction can effectively ensure the safety and economic benefits of marine transportations and ocean platform operations. In this study, an instantaneous prediction model of free surface elevation based on CNN-LSTM and wavelet function was established and discussed. In the case of inputting the time series of free-surface elevation of two spectral peak periods of the irregular wave, the high precision prediction of the time-series of the subsequent one spectral peak period is effectively achieved. The main conclusions of this study are summarized as follows:

(1) In terms of reducing the model prediction error and improving the R2 value, the most reasonable value of the number of convolution kernels is 16.
(2) A larger numbers of units in the dual LSTM layer has a positive effect on reducing the model prediction error and improving the R2 value, and the number of units in the dual LSTM layer needs to be greater than or equal to 256 and 128 respectively.
(3) The larger the dropout rate of the Dropout layer, the more obvious the model is under-fitting and the worse the prediction performance. The value should be less than or equal to 0.2.
(4) The wavelet activation function may make the model have a stronger fitting ability on the validation data, and significantly weaken the overfitting phenomenon of the model. It indicates that the wavelet function as an activation function has more advantages in the prediction of the irregular ocean surface waves.
(5) On the premise of ensuring prediction performance, and considering the rational use of computing resources, the batch size of 128 is the most appropriate value.
(6) Most mathematics experimental models have a lower degree of fitting to the relatively later time steps, and certain measures should be taken to improve this situation in subsequent work, such as using weighted loss.

In general, this study proposes a designed model which can perform high precision instantaneous wave prediction of the irregular ocean surface wave in time domain, gives the optimal selection of important parameters of the model, and provides model design ideas for the related work.

CRediT authorship contribution statement

Gang Xu: Writing, Conceptualization, Funding, Supervision. **Siwen Zhang:** Writing – original draft, Methodology, Software. **Weichao Shi:** Conceptualization, Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

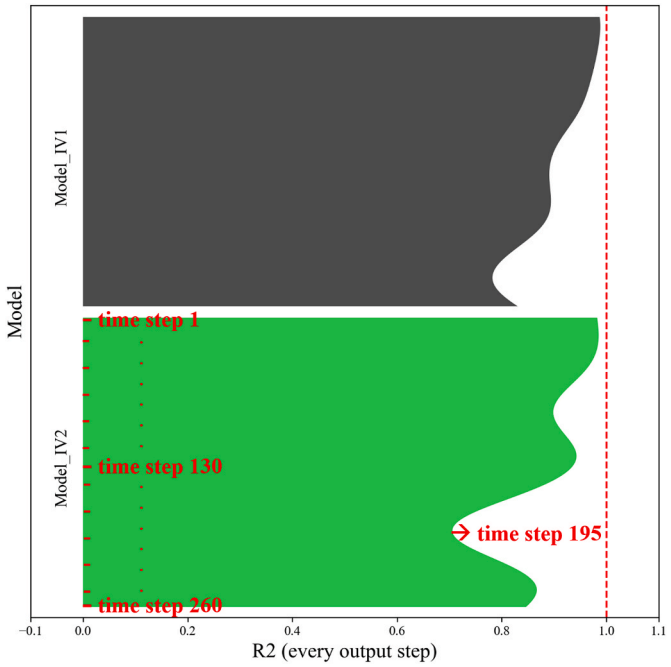
Data will be made available on request.

Acknowledgment

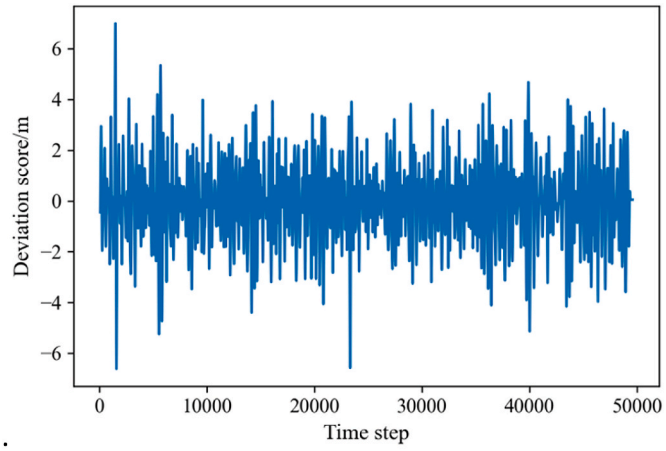
This work is supported by the National Natural Science Foundation of China (Grant No. 51879125), Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No.311021013) , Key University Science Research Project of Jiangsu (18KJA130001), Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX20_1490、SJCX20_1482、KYCX21_3439、KYCX21_3440、KYCX21_3499).

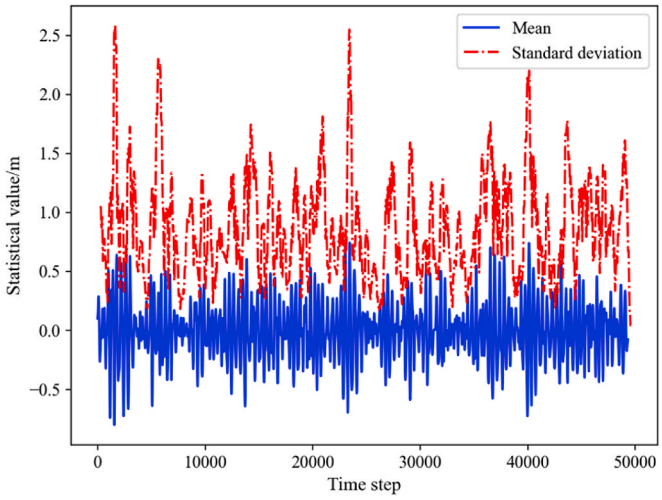
Appendix

1. This figure adds an explanation to the R2 figure in this paper. Different color blocks represent different models, and inside a color block, the R2 coefficients at each output time corresponding to the model are arranged from top to bottom (R2 based on the validation dataset). This figure not only shows the overall performance of the model, but also shows the distribution of the R2 coefficient at the model output time steps.



2. The following are the first-order difference data map and the mean standard deviation map (time window size is 260 time steps) of the entire wave time series after standardization. The irregular trend change of the first-order difference value indicates that the time series has significant nonlinearity, and the unsteady and irregular mean and standard deviation prove that the time series has obvious non-stationarity. It can be seen from the mean and standard deviation map that the moving mean of the wave series has large irregular fluctuations, and the changing trend of the moving standard deviation is more severe and irregular, which indicates that the series has strong non-stationarity. For non-stationary time series, difference operation is required before forecasting, and convolution operation can meet this requirement, which is one of the reasons for using convolution layer in this paper.

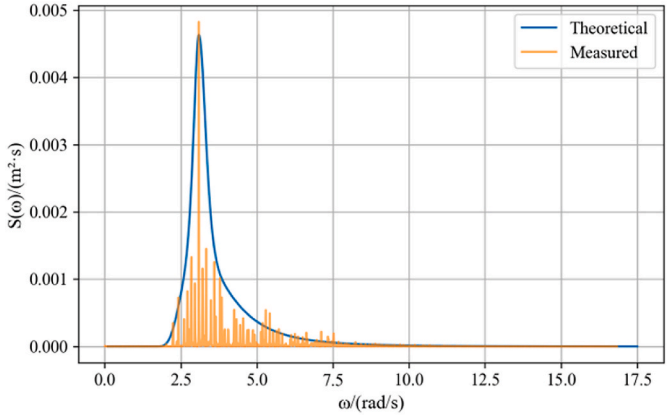




3. In order to further validate the prediction ability of the algorithm in this paper to actual irregular waves, we test the model on the wave free surface elevation data of irregular waves in a towing tank.

The wave spectrum used in the wave-making in this tank is the JONSWAP spectrum, with a significant wave height of 0.105 m and a spectral peak period of 2.03 s (So the number of time steps for model input and output is 406 and 203, respectively. In addition, the model architecture selects the optimal architecture in the conclusions. The dataset is divided into 5:1.), the spectral peak factor is 3.3. Under the condition that the ship model scale ratio is 1:100, the wave has a return period of 10-years.

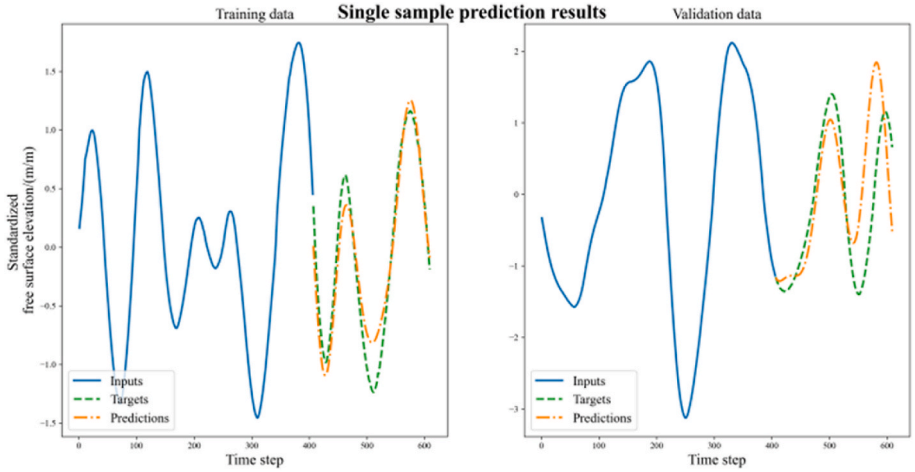
The sampling time of the wave is about 8 min 53 s, the sampling frequency is 100 Hz, and the total length of the data is 53287. To verify the accuracy of the waves generated by the wave maker, a wave altimeter was placed 5 m in front of the wave maker. Compare the actual wave with the target wave. Comparing the actual wave with the target wave, the measured and theoretical spectrum of irregular waves are compared as follows.

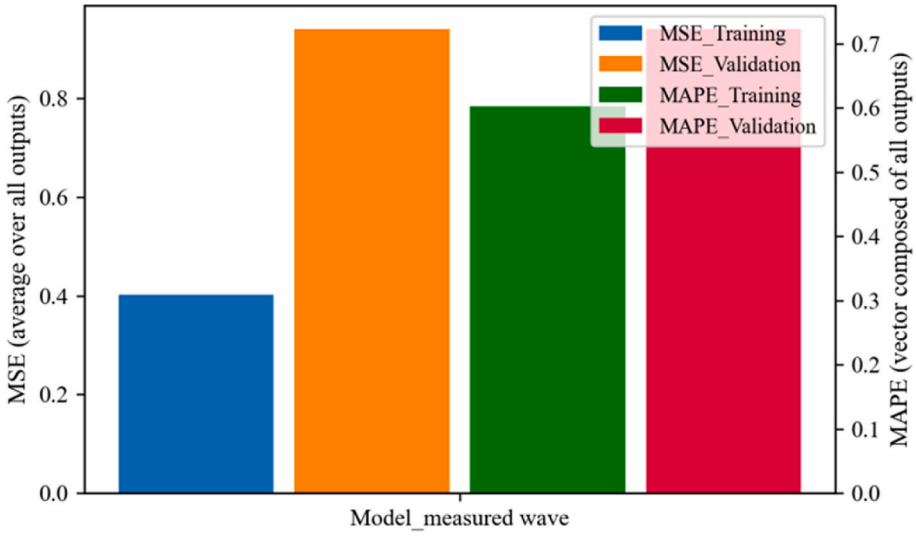


It can be seen from the above figure that the actual measured

irregular waves and the preset waves have a good degree of agreement with the power spectrum distribution, which indicates that the irregular waves caused by the towing tank in the experiment are effective and reliable.

The final prediction effect is shown in these figures below. It can be seen from these figures that the prediction performance of the algorithm on real water wave data is still ideal in the case of untargeted debugging. This shows that the algorithm has the ability to predict real irregular waves to a certain extent.





4. The note on the model architecture and its effectiveness:

1) In terms of model capacity, the CNN-LSTM in this paper has a deeper number of layers than the classical LSTM, which has an advantage in the ability to fit the data.
2) Further, regarding the specific model architecture, not only the CNN module is added to the work, but the advantages reflected in the final prediction result are contributed by the CNN and the residual architecture (originally from the ResNet) as a whole. The mechanism of action of this architecture may be related to the theory of Taylor's formula (convolution operation can realize numerical differentiation operation, and the result obtained by adding up multiple orders of differentiation is used as a feature extracted by the model, which is conducive to increasing the fitting ability of the model). In addition, CNNs can retain historical information to a large extent, and there is no problem of sequence dependence, which can be efficiently parallelized.
3) The wavelet transform is a new mathematical transformation method based on Fourier analysis, which overcomes the limitations of Fourier transform and the disadvantage of Fourier transform window invariance. The wavelet transform mainly realizes multi-scale refinement through expansion and translation, highlights the details of the problem to be handled, and effectively extracts local information. The wavelet activation function has the advantages of wavelet transformation, which avoids blindness in network design architecture, and accelerates the convergence of neural networks and improves model accuracy.

References

Abhigna, P., Jerritta, S., Srinivasan, R., Rajendran, V., 2017. Analysis of feed forward and recurrent neural networks in predicting the significant wave height at the moored buoys in bay of bengal. In: J. International Conference on Communication and Signal Processing (ICCSP), IEEE., 23, p. 1856e1860. https://doi.org/10.1109/iccsp.2017.8286717.

Castro, A., Carballo, R., Iglesias, G., Rabu~nal, J., 2014. Performance of artificial neural networks in nearshore wave power prediction. J. Appl. Soft Comput. 23, 194e201. https://doi.org/10.1016/j.asoc.2014.06.031.

Chao, Yan, Yishi, Qiu, Yongqiong, Zhu, 2019. Predict oil production with lstm neural network. In: C. Proceedings of the 9th International Conference on Computer Engineering andNetworks(CENet).

De Pina, A.C., de Pina, A.A., Albrecht, C.H., Pires de Lima, Leite, Jacob, B.S., P, B., 2013. Ann-based surrogate models for the analysis of mooring lines and risers. J. Applied Ocean Research. 41, 76–86. https://doi.org/10.1016/j.apor.2013.03.003.

Deo, M., Jha, A., Chaphekar, A., Ravikant, K., 2001. Neural networks for wave forecasting. J. Ocean Eng. 28 (7), 889e898. https://doi.org/10.1016/S0029-8018(00)00027-5.

Gómez-Orellana, A.M., Guijo-Rubio, D., Gutiérrez, P.A., Hervás-Martínez, C., 2022. Simultaneous short-term significant wave height and energy flux prediction using zonal multi-task evolutionary artificial neural networks. J. Renewable Energy. 184, 975–989. https://doi.org/10.1016/j.renene.2021.11.122.

Guijo-Rubio, D., Gomez-Orellana, A.M., Gutierrez, P.A., Hervas-Martínez, C., 2020. Short-and long-term energy flux prediction using multi-task evolutionary artificial neural networks. J. Ocean Eng. 216, 108089 https://doi.org/10.1016/j.oceaneng.2020.108089.

Huang, Baigang, 2019. On - Line Modeling and Forecasting of Ship Motion in Wave Based on Wavelet Neural Network. D.

Hunt, J.N., 1979. Direct solution of wave dispersion equation. J. Journal of Waterway, Port, Coastal and Ocean Engineering 105 (4), 457–459.

Li, Yongjie, Zhang, Rui, Muheng, Wei, Zhang, Yu, 2021. Research status and prospect of key technologies for autonomous navigation of ships. J. Research on Chinese Ships 16, 13. https://doi.org/10.19693/j.issn.1673-3185.01958.

Makarynskyy, O., Pires-Silva, A., Makarynska, D., Ventura-Soares, C., 2005. Artificial neural networks in wave predictions at the west coast of Portugal. J. Comput. Geosci. 31 (4), 415e424. https://doi.org/10.1016/j.cageo.2004.10.005.

Peng, Han, Li, Yuhang, Meng, Revealing, 2020. Present situation and prospect of global ocean circulation forecasting system. J. oceanic forecast 37, 8. CNKI:SUN:HYYB.0.2020-03-012.

Deka Paresh Chandra, Prahlada, R., 2012. Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. J. Ocean Eng. 43, 32–42. https://doi.org/10.1016/j.oceaneng.2012.01.017.

Sadeghifar, T., Nouri Motlagh, M., Torabi Azad, M., Mohammad Mahdizadeh, M., 2017. Coastal wave height prediction using recurrent neural networks (rnns) in the south caspian sea. J. Mar. Geodes. 40 (6), 454e465. https://doi.org/10.1080/01490419.2017.1359220.

Sanchez, A.S., Rodrigues, D.A., Fontes, R.M., Martins, M.F., Kalid, R.d.A., Torres, E.A., 2018. Wave resource characterization through in-situ measurement followed by artificial neural networks' modeling. J. Renew. Energy 115, 1055e1066. https://doi.org/10.1016/j.renene.2017.09.032.

Wang, Wenxu, Tang, Ruichun, Cheng, Li, Liu, Peishun, Luo, Liang, 2018. A bp neural network model optimized by mind evolutionary algorithm for predicting the ocean wave heights. J. Ocean Engineering. 162, 98–107. https://doi.org/10.1016/j.oceaneng.2018.04.039.

Wang, Yuanyuan, Liu, Jialun, Ma, Feng, Wang, Xingping, Yan, Xinping, 2021. Research status and trend of intelligent ship remote driving control technology. J. Research on Chinese Ships 16, 14. https://doi.org/10.19693/j.issn.1673-3185.01939.

Weilisi, Kojima, Toshiharu, 2022. Investigation of hyperparameter setting of a long short-term memory model applied for imputation of missing discharge data of the daihachiga river. J. Water. 14 https://doi.org/10.3390/w14020213.

Zhang, Feng, Wang, Qi, Lumei, Shi, Weiyong, Zhang, Junbiao, 2018. Research on ocean current prediction based on artificial neural network. J. oceanic forecast 35, 6. https://doi.org/10.11737/j.issn.1003-0239.2018.04.006.

Zhao, W., Yang, H., Li, J., Shang, L., Fu, Q., 2019. Network traffic prediction in network security based on emd and lstm. C. In: Proceedings of the 9th International Conference on Computer Engineering and Networks.

Zhao, X.Z., Zhao, S.C., Liang, S.X., 2009. Efficient focusing models for generation of freak waves. J. Chin Ocean 23 (3), 429–440.