2-1-2023

Towards carbon Neutrality: Prediction of wave energy based on improved GRU in Maritime transportation

Zhihan Lv
Qingdao Institute of Bioenergy and Bioprocess Technology

Nana Wang
Qingdao Institute of Bioenergy and Bioprocess Technology

Ranran Lou
Qingdao University

Yajun Tian
Qingdao Institute of Bioenergy and Bioprocess Technology

Mohsen Guizani
Mohamed Bin Zayed University of Artificial Intelligence

Recommended Citation

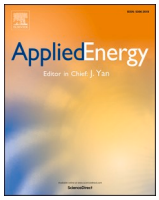
Z. Lv, N. Wang, R. Lou, Y. Tian, and M. Guizani, Towards carbon Neutrality: Prediction of wave energy based on improved GRU in Maritime transportation, in *Applied Energy*, vol 331, Feb 2023, doi:10.1016/j.apenergy.2022.120394

Applied Energy 331 (2023) 120394

Contents lists available at ScienceDirect

Applied Energy

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



Towards carbon Neutrality: Prediction of wave energy based on improved GRU in Maritime transportation

Zhihan Lv [a,b,c,1,*], Nana Wang [a,d,e,1], Ranran Lou [f], Yajun Tian [a,d,e,*], Mohsen Guizani [g]

[a] Extended Energy Big Data and Strategy Research Center, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, 266101, China
[b] Department of Game Design, Faculty of Arts, Uppsala University, Uppsala, 75375, Sweden
[c] State Key Laboratory of Media Convergence Production Technology and Systems, Beijing, 100803, China
[d] Shandong Energy Institute, Qingdao, 266101, China
[e] Qingdao New Energy Shandong Laboratory, Qingdao, 266101, China
[f] College of Computer Science and Technology, Qingdao University, Qingdao, 266071, China
[g] Machine Learning Department, Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, UAE

HIGHLIGHTS

- use wave energy as the power for transportation ships.
- A novel wave energy prediction method based on GRU.
- The features of the model are assigned different weights.

ARTICLE INFO

Keywords:
Carbon Neutrality
Maritime Transportation
Wave Energy Prediction
Gated Recurrent Unit Network
Bayesian optimization algorithm

ABSTRACT

Efficient use of renewable energy is one of the critical measures to achieve carbon neutrality. Countries have introduced policies to put carbon neutrality on the agenda to achieve relatively zero emissions of greenhouse gases and to cope with the crisis brought about by global warming. This work analyzes the wave energy with high energy density and wide distribution based on understanding of various renewable energy sources. This study provides a wave energy prediction model for energy harvesting. At the same time, the Gated Recurrent Unit network (GRU), Bayesian optimization algorithm, and attention mechanism are introduced to improve the model's performance. Bayesian optimization methods are used to optimize hyperparameters throughout the model training, and attention mechanisms are used to assign different weights to features to increase the prediction accuracy. Finally, the 1-hour and 6-hour forecasts are made using the data from China's NJI and BSG observatories, and the system performance is analyzed. The results show that, compared with mainstream prediction algorithms, GRU based on Bayesian optimization and attention mechanism has the highest prediction accuracy, with the lowest MAE of 0.3686 and 0.8204, and the highest R^2 of 0.9127 and 0.6436, respectively. Therefore, the prediction model proposed here can provide support and reference for the navigation of ships powered by wave energy.

1. Introduction

Maritime transportation is the most widely used mode of transport in international trade. Ships are means of maritime transportation. At present, ships are powered mainly by burning fossil fuels. However, with the increase of demand, the exhaustion of fossil energy is inevitable [1]. Because of fuel consumption, a large amount of nitrogen oxide, carbon dioxide and other greenhouse gases will be produced in the process of shipping [2]. Greenhouse gases such as carbon dioxide generated by large-scale energy consumption are one of the main causes of current global climate change [3 4]. In recent years, more and more countries have taken carbon neutrality as an important task in the future [5 6].

* Corresponding authors.
E-mail addresses: lvzhihan@gmail.com (Z. Lv), tianyajun@qibebt.ac.cn (Y. Tian).
[1] Both authors are the first authors.

https://doi.org/10.1016/j.apenergy.2022.120394
Received 26 June 2022; Received in revised form 28 September 2022; Accepted 18 November 2022
Available online 28 November 2022

Carbon neutrality refers to the gradual offset of people's carbon dioxide emissions by various mitigation [7]. By 2060, the Chinese government wants to be carbon neutral [8]. Therefore, it is very important to seek a new type of ship power energy that can replace traditional fossil fuels to reduce pollutant emissions and promote carbon neutrality [9].

The ocean is abundant in resources and has enormous development potential [10]. According to the report published by the International Energy Organization (IEA), different marine energy technologies can meet the current global electricity demand of nearly 20,000 TWh [11]. Ocean waves contain huge energy, belonging to renewable energy. The wave energy can be converted into electric energy through the wave energy converter (WEC) [12]. Ocean wave power generation has four advantages compared with traditional power generation methods based on analyzing the economic feasibility and components of wave power generation. (1) Energy density is higher [13]; (2) WEC has less impact on the environment during operation; (3) the wave loses little energy during propagation [14]; (4) the power generation efficiency is higher. Statistics show that the power generation rate of wave power generation devices is as high as 90 % [15]. Therefore, the application of wave energy on ships has unique advantages, is economically feasible, and has a low cost. At the same time, using wave energy as the power source of ships can greatly reduce the pollutant emissions of ships during transportation, thereby realizing green ocean transportation [16 17]. In 2006, Norway proposed the E/S Orcelle ship model, which uses wind, solar and wave energy as the power source [18]. In 2007, the American company Liquid Robotics developed a water vehicle powered by wave energy, which can sail for a long time on the water surface [19]. In 2011, Boston University built a wave-powered ship with autonomous sailing capabilities.

In order to effectively utilize the wave energy in shipping lines, this paper proposes to use wave energy as the power source of ships. The contributions and innovations of this research are as follows.

- Currently, there is a large amount of nitrogen oxides, carbon dioxide, and other greenhouse gases during shipping, increasing carbon emissions. To this end, a wave energy prediction model based on Gated Recurrent Unit network is proposed to achieve relatively zero greenhouse gas emissions.
- In order to solve the problem of difficult selection of model hyperparameters, we use the Bayesian optimization algorithm to optimize the hyperparameters. Through this algorithm, the best parameter combination can be obtained in a short time.
- During the training process of the model, we use the attention mechanism to assign different weights to the features to achieve a more accurate prediction effect.
- In the 1-hour and 6-hour prediction, we compare the model proposed in this article with the current popular algorithms, and the experimental findings demonstrate that our proposed model outperforms them.

The rest of this paper is structured as follows. Section 2 introduces the recent related researches on wave energy prediction by related scholars and highlights the significance of this research by analyzing its advantages and disadvantages. Section 3 introduces the principle of gated recurrent units, Bayesian optimization and attention mechanisms. Section 4 describes the proposed wave energy prediction method, including the GRU wave energy prediction model based on Bayesian optimization and attention mechanism, the dataset used in the test, the data preprocessing process, the hyperparameters and evaluation metrics. Section 5 evaluates the model's prediction accuracy and discuss the prediction results. Section 6 is the conclusion of this paper.

2. Recent related research

Historically, numerical models were used to forecast wave elements. This method establishes an energy balance equation by simulating the wave evolution process generated by the wind field acting on the ocean surface, so as to achieve relatively satisfactory forecast results [20]. At the same time, the numerical models have the disadvantages of complex implementation, many inputs, and long processing time, which is not conducive to the accurate and rapid prediction of waves [21]. In recent years, in order to solve the shortcomings of numerical models, researchers have begun to use artificial intelligence algorithms to study the marine environment. In 2018, James *et al.* proposed a Multi-Layer Perception (MLP) model that can replace the numerical model SWAN based on the ocean data of Monterey Bay [22]. Although the MLP has better performance than the numerical models, the pros and cons of the MLP model are too dependent on the sample data of the training process. The research of Mahjoobi and Mosabbeb believes that support vector machine (SVM) has better generalization performance [23]. In addition, Mahjoobi and Etemad-Shahidi's research believes that decision tree has better interpretability than MLP, and decision tree determines the relative importance of parameters by inputting branches, which is more suitable for wave element prediction [24]. The long short-term memory (LSTM) network improved from the recurrent neural network (RNN) has a unique chain structure, it is ideal for processing marine time series data [25]. In 2020, Fan *et al.* used the LSTM network to forecast the wave height of ten stations with different environmental conditions, and compared it with the results of algorithms such as random forest (RF), SVM and MLP, demonstrated the superiority of LSTM in wave height prediction [26]. Ni and Ma combined LSTM with principal component analysis (PCA) for continuous prediction of wave height under polar conditions [27]. The Gated Recurrent Unit (GRU) based on the improved LSTM simplifies the structure and computation of neurons [28]. Currently, GRU has been shown to outperform LSTM in several applied studies [29]. In 2022, Li *et al.* used the GRU network to forecast wave height for 6 stations along the coast of China [30]. Although the performance of GRU network is better than that of LSTM, a single neural network cannot select a small part of useful information from a large number of inputs to focus on processing. By combining an attention mechanism with a neural network, the model can be made to pay more attention to useful information [31]. In addition, the selection of hyperparameters is an important link in determining the prediction accuracy of the model. The same algorithm with different hyperparameters will bring different results. In the past, hyperparameter selection mostly relied on manual search and empirical setting, but this method is time-consuming, and the results obtained may not be optimal hyperparameters. Bayesian optimization algorithms can obtain optimal hyperparameter combinations in fewer iterations [32 33].

Although wave energy shows many advantages over other renewable energy sources, it is difficult to characterize and predict because of its randomness. According to existing research, wave energy can be expressed by the equation $F = 0.49 \bullet H^2 \bullet T$, where H is the wave height and T is the wave period [34]. Therefore, accurate prediction of wave height and wave period is an important prerequisite for wave energy power prediction.

3. Relevant technical theoretical knowledge

3.1. Gated recurrent unit network (GRU)

The GRU network is improved from RNN. By associating neurons between layers in the network, RNN solves the problem that the front and rear inputs in the traditional neural networks are independent of each other. Therefore, RNN offers certain advantages in learning the sequence's nonlinear features, making it more suitable for dealing with time problems. Natural language processing, time series forecasting, and other domains make extensive use of RNN. In 1991, Hochreiter discovered that RNN has a long-term dependence problem, which means that while learning a long sequence, the network would exhibit gradient disappearance and gradient explosion, making it unable to understand the nonlinear relationship of long time span [35]. In order to solve the

long-term dependency problem, improved neural networks based on RNNs continue to emerge, including LSTM and GRU.

Hochreiter and Schmidhuber proposed LSTM network in 1997 [36]. LSTM controls the transmission of information in the network through three gate devices (forget gate, input gate, and output gate). A sigmoid function (σ) and a dot product operation are included in each gate. σ outputs a number between 0 and 1, indicating how much information may travel through, 0 indicates no information is permitted to pass through, 1 means information is allowed to pass through, and the calculation equation is shown in equation (1). In contrast to the RNN's recursive calculation for the system state, the three gates form a self-loop to the LSTM unit's internal state. The input gate determines the current time step's input as well as the update of the internal state of the previous time step's system state; the forget gate determines the update of the internal state of the previous time step to the internal state of the current time step; the output gate determines the internal state to update the system state. The structure of LSTM is shown in Fig. 1 [37].

$$\sigma_{(x)} = \frac{1}{1+e^{-x}} \tag{1}$$

Google's testing suggests that three gates in an LSTM contribute differentially to its learning abilities, with the forgetting gate being the most essential, followed by the input gate, and lastly the output gate [38]. As a result, removing the gate with a tiny contribution and its related weight can simplify the neural network structure and increase learning efficiency. Based on the above concepts, Cho et al. proposed GRU in 2014 [39]. Only update gates and reset gates are included in the GRU. The update gate is analogous to the LSTM's forget gate and output gate. It is used to govern how much of the preceding moment's state information is carried into the present state. The greater the value of the update gate, the more prior state information is brought in. The reset gate is comparable to the LSTM's input gate in that it influences how fresh input information is integrated with old memory; the smaller the reset gate, the less information from the previous state gets recorded. The structure of GRU is shown in Fig. 2. The r_t in the update gate and the z_t in the reset gate are obtained by equation (2) and equation (3), respectively. Among them, U and W are weight parameters.

$$R_t = \sigma(W_r x_t + U_r h_{t-1}) \tag{2}$$

$$Z_t = \sigma(W_z x_t + U_z h_{t-1})\# \tag{3}$$

The current hidden state h_t is obtained by Equation (4), where the calculation process of the candidate set $\widetilde{h}_t$ is shown in Equation (5). The tanh is a hyperbolic tangent function whose expression is shown in equation (6).

$$H_t = (1-z_t)h_{t-1} + z_t\widetilde{h}_t\ \# \tag{4}$$

$$\sigma_{(x)} = \frac{1}{1+e^{-x}}$$

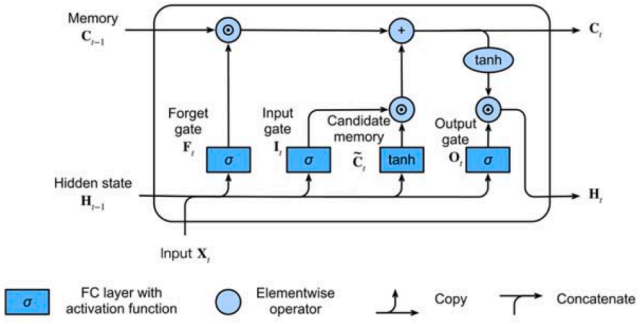


Fig. 1. LSTM network unit structure.

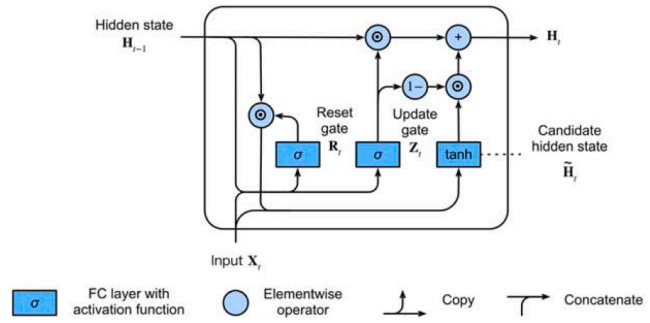


Fig. 2. GRU network unit structure.

$$\widetilde{H}_t = tanh(W_h x_t + U_h(r_t \odot h_{t-1})) \tag{5}$$

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{6}$$

3.2. Bayesian optimization

The optimization of the hyperparameters of the model is one of the important factors influencing the final prediction effect. Currently, the commonly used hyperparameter optimization methods in research are grid search, random search and Bayesian optimization. Grid search is time-consuming because it needs to iterate through all combinations of candidate hyperparameter values. Random search is similar to grid search, but unlike grid search, which traverses all parameter value combinations, random search randomly selects a fixed number of hyperparameter value combinations within a given parameter value range to find the optimal parameter value or an approximation of the optimal parameter value for the purpose. Random search has a faster search speed, but the resulting hyperparameter values may not be optimal. The difference between random search and grid search is shown in Fig. 3 [40].

The Bayesian parameter tuning method was proposed by Snoek et al. in 2012 [41]. Its optimization strategy is to obtain the posterior distribution of the given objective function through the Gaussian process for the parameter value combination selected by sampling. After that, the following parameter value combinations are continuously selected according to the posterior distribution of the previous parameter value combination until the posterior distribution matches the real distribution. For the search space X_n, the optimal solution x_{best} of Bayesian optimization can be expressed by equation (7), where f is the objective function. Compared with grid search and random search, the Bayesian optimization method has fewer iterations, faster speed, and more robust performance. And the Bayesian optimization method can continuously update the prior through the Gaussian process, using the historical parameters The combination of values makes decisions about the next choice.

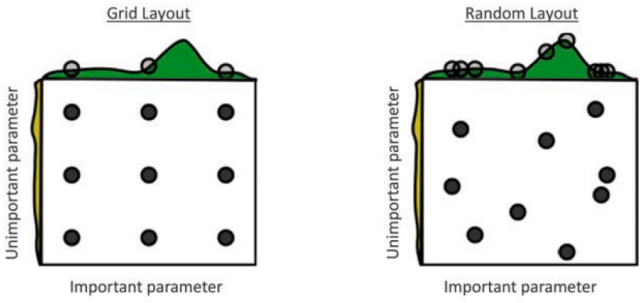


Fig. 3. Comparison of random search and grid search.

$$x_{best} = argmin_{X_n} f(X_n) \tag{7}$$

The Gaussian process of Bayesian optimization consists of mean and covariance functions, as shown in Equation (8), where μ is the mean and $k(x, x')$ is the covariance function. For dataset $D = \{(x_1, f(x_1)), (x_2, f(x_2)), \cdots, (x_t, f(x_t))\}$, the Gaussian distribution is shown in equation (9).

$$f(x)\ gp(\mu, k(x, x'))\# \tag{8}$$

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_t) \end{bmatrix} gp\left(\mu, \begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_t) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_t) \\ \vdots & \vdots & \vdots & \vdots \\ k(x_t,x_1) & k(x_t,x_2) & \cdots & k(x_t,x_t) \end{bmatrix}\right)\# \tag{9}$$

For the new sample x_{t+1}, the Gaussian distribution is shown in equation (10). The posterior probability distribution of f_{t+1} is shown in equation (13).

$$\begin{bmatrix} f_{1:t} \\ f_{t+1} \end{bmatrix} gp\left(\mu, \begin{bmatrix} K & k^T \\ k & k(x_{t+1}, x_{t+1}) \end{bmatrix}\right)\# \tag{10}$$

$$K = \begin{bmatrix} k(x_1,x_1) & \cdots & k(x_1,x_t) \\ \vdots & \ddots & \vdots \\ k(x_t,x_1) & \cdots & k(x_t,x_t) \end{bmatrix}\# \tag{11}$$

$$k = [((x_{t+1}, x_1), (x_{t+1}, x_2) \cdots (x_{t+1}, x_t))]\# \tag{12}$$

$$P(f_{t+1}|D, x_{t+1}) = gp\left(u(x_{t+1}), \delta^2(x_{t+1})\right)\# \tag{13}$$

$$u(x_{t+1}) = kK^{-1}f_{1:t}\# \tag{14}$$

$$\delta^2(x_{t+1}) = k(x_{t+1}, x_{t+1}) - kK^{-1}k^T\# \tag{15}$$

where u, δ^2 represents obeying x_i $N(u, \delta^2)$. The Bayesian optimization procedure is as follows:

1) Randomly initialize a set of hyperparameter value combinations in the search space, and calculate the value of the objective optimization function.
2) Continue to randomly select the hyperparameter combination, calculate the objective function value, and save the point if the value is better than the best value obtained in history.
3) Repeat step 2 until the set number of iterations is reached.

3.3. Attention mechanism

The attention mechanism stems from the research of human vision. Since the bottleneck of information processing, people will selectively focus on the part of the information they wish to view while disregarding other observable information; this mechanism is known as the attention mechanism in cognitive research [42,43]. 其中 , Fig. 4 illustrates the principle of attention mechanism applied to visual information processing.

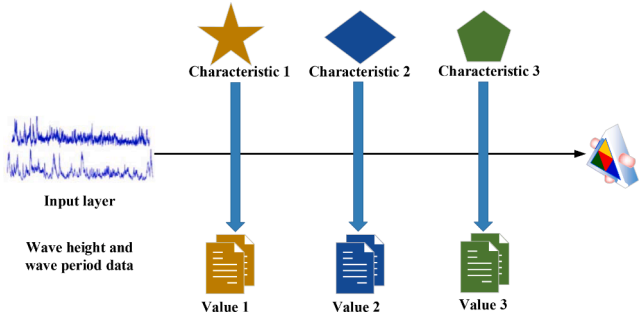


Fig. 4. The principle of attention mechanism applied to visual information processing.

In Fig. 4, the input data consists of features and corresponding quantized values. The similarity between the input data and each feature is calculated as the weight coefficient of the input data information on each feature. Finally, the attention score is obtained by weighting and summing the weight coefficients on all features. Therefore, the essence of the attention mechanism is the weighted summation of the weights of different features.

Nowadays, attention mechanism is widely used in the field of artificial intelligence, such as image recognition and natural language processing. In neural networks, the attention mechanism is the focus on the assignment of input weights. The attention mechanism can assign weights to elements according on their relevance, focusing on crucial information with high weights and disregarding unimportant information with low weights. In addition, it can continuously adjust the weights, so that important information can also be selected in different situations, so it has higher scalability and robustness [44]. In the time series prediction problem, the attention mechanism can prevent important features from being ignored with increasing of time step. The weight allocation method can be expressed by equations (16) and (17), where h_t is the state vector of the hidden layer in the neural network at time t, e_t is the attention probability distribution value, a_t is the attention score, u_a and W_a are the attention weight vectors, b_a is the attention bias vector.

$$a_t = \frac{exp(e_t)}{\sum_{k=1}^{t} e_k}\# \tag{16}$$

$$e_t = u_a \tanh(W_a h_t + b_a)\# \tag{17}$$

4. Improved GRU wave energy prediction model

4.1. Model structure

This study analyzes wave energy with high energy density and wide distribution to achieve relatively zero emissions of greenhouse gases to cope with the crisis brought about by global warming. In addition, the GRU algorithm, Bayesian optimization algorithm, and attention mechanism are adopted to provides a model for predicting wave energy that can be used to harvest energy. The structure of the GRU wave energy prediction model based on the Bayesian optimization and the attention mechanism is shown in Fig. 5.

The specific prediction process of the model in Fig. 5 is as follows. First, the input features of the predictive model are determined. Second, the Bayesian optimization algorithm is used to determine the hyperparameters of the model. In the hidden layer of the model, different weights are assigned to the features through the attention mechanism. Then, after training with a large amount of data, the wave height and wave period prediction models are obtained respectively. We use the test set to compare the prediction results of the model with the observed values to determine whether the optimization end condition of the Bayesian optimization algorithm is reached. If yes, we use the models to forecast the wave height and the wave period separately; if not, we continue the hyperparameter optimization. Finally, we use the wave energy conversion equation to convert the predicted wave height and wave period into the predicted value of wave energy.

This work chooses the mean square error (*MSE*) as the loss function during the model training process, as given in equation (18), where n is the number of samples, y_i is the observed value, and x_i is the predicted value. The update of the weight parameters is done by the Adam optimizer. The Adam optimizer combines the advantages of RMSProp and AdaGrad algorithms that are good at dealing with sparse gradients and non-stationary objectives, and it can achieve good results at a fast speed.

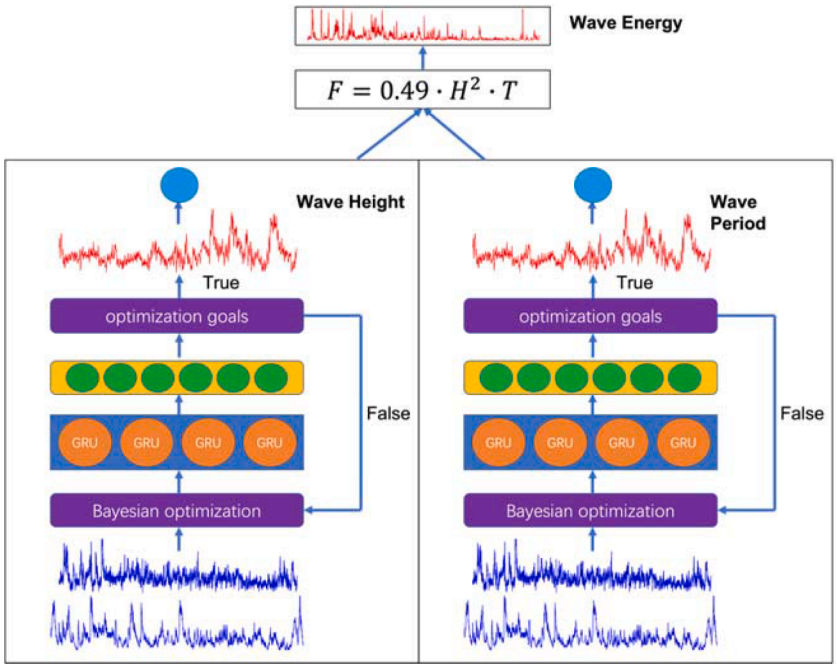


Fig. 5. The structure of the GRU wave energy prediction model based on the Bayesian optimization and the attention mechanism.

To prevent overfitting during model training, we adopted the Early Stopping algorithm, which stops training if the error on the validation set increases as the training rounds increase.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \# \tag{18}$$

4.2. Experimental environment

The system was built on the Matlab network simulation platform to verify the performance of the GRU wave energy prediction model based on Bayesian optimization and attention mechanism constructed here. In the experiment, the neural network was built using the Tensorflow framework open source by Google. This framework is a machine learning and deep learning programming framework based on vector flow graph. The matrix operation was completed by using Numpy and Pandas open source toolkit. Pandas library provides excellent assistance for data cleaning and data preprocessing in data analysis. In the software, the operating system is Linux 64bit, the Python version is Python 3.6.1, and the development platform is PyCharm. In hardware, the CPU is Intel core i7-7700@4. 2 GHz 8 cores, the memory is Kingston DDR4 2400 MHz 16G, and the GPU is Nvidia GeForce 1060 8G.

4.3. Ocean observation data

This paper selects the observation data of two observation stations in the coastal waters of China to achieve accurate prediction of wave energy. The data comes from Marine Professional Knowledge Service System (https://ocean.ckcest.cn/). The time interval of the observation data is 1 h and the precision is 0.1. The selected dataset contains observations for all time frames of the year. As a result, the model's prediction ability under various environmental situations may be tested. Table 1 contains information on the two stations, and Fig. 6 depicts their geographic distribution.

4.4. Data preprocessing

4.4.1. Missing value padding

The observation station is not a perfect ocean monitoring system. Since factors such as the design life of the equipment and the natural wear and tear of the instruments, observation interruption and missing data are common, and the original observation data has a large number of missing values. In order to improve data quality and reduce the impact of missing values on model prediction accuracy, this paper fills missing values in the data set using the before and after average value filling method, which takes the average value of the attribute value at the moment before the missing value and the attribute value at the moment after the missing value is taken as the filling value at the missing moment. When multiple consecutive values are missing, the average value of the two adjacent non-null values is used to fill in.

4.4.2. Feature selection

Since ocean waves are waves of sea water caused by the action of wind, there is a close relationship between wind and ocean waves. Previous studies have also shown that wind speed and direction are important factors affecting ocean waves [45,46]. Based on past research, in order to train a model with high prediction accuracy without consuming many computing resources, the wind speed and wind direction data within 2 h of history are selected as the characteristics of the prediction model. In addition, the wave height and wave period within 2 h of history are also added. Therefore, the 8 features of the model are historical 1-hour wind speed, historical 1-hour wind direction, historical 1-hour wave height, historical 1-hour wave period, historical 2-hour wind speed, historical 2-hour wind direction, historical 2-hour wind direction Wave height, historical 2-hour wave period.

Table 1
Details of selected site data.

Station	Latitude	longitude	Time	Maximum wind speed (m/s)	Maximum wave height (m)	Data size
NJI	27.5 N	121.1E	2018.07.01–2019.07.31	23.7	7.5	9504
BSG	26.7 N	120.3E	2019.08.01–2020.08.31	21.6	4.5	9504

Fig. 6. Geographical location of selected sites.

4.4.3. Feature normalization

In a model with multiple features, different units of measurement of features will lead to different calculation results. Large-scale features will play a decisive role, while small-scale features may be ignored. In order to reduce the impact of the measurement unit and scale differences between different features, this paper adopts zero-mean normalization to process feature data. This method can speed up the speed of gradient descent to find the optimal solution. The mean of the standardized data is 0 and the standard deviation is 1, which follows a standard normal distribution. Its calculation equation is shown in (19), where n is the sample size, X^* is the processed data, X is the original data, $\overline{X}$ is the original data's mean, and δ is the original data's standard deviation. The standard deviation is calculated as equation (20).

$$X^* = \frac{X - \overline{X}}{\delta} \# \tag{19}$$

$$\delta = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} \# \tag{20}$$

4.5. Model hyperparameters

In this section, the hyperparameters of the GRU wave energy prediction model and the hyperparameters of the other three comparison algorithms are optimized based on the Bayesian optimization algorithm, and the number of optimization iterations is 30. The value range and final value of the hyperparameters to be optimized are shown in Table 2 and Table 3, among them, time_step is the time step, and its value range is (2, 128); units is the number of neurons, and its value range is (2, 128); dense is the number of fully connected layer nodes, and its value range is (2, 128); the number of n estimators trees, and its value range is (10 200); max depth is the maximum depth of the tree, and its value range is (5, 10). Moreover, the hyperparameters of the GRU model and LSTM model constructed here include time_step, units, dense; the hyperparameters of MLP model include units; the hyperparameters of the RF model include N estimators and Max depth. In addition, the learning rates of the neural networks GRU, LSTM, and MLP are all 0.001, and the training rounds are all 100. The activation function of GRU and LSTM is tanh, as shown in equation (4), and the activation function of MLP is linear rectification function (ReLU), as shown in equation (21).

$$ReLU(x) = \max(0, x) \tag{21}$$

4.6. Model evaluation index

In order to test the model's prediction performance completely, this study uses *MSE*, root mean square error (*RMSE*), mean absolute error (*MAE*), mean absolute percentage error (*MAPE*), Pearson correlation coefficient (R) and coefficient of determination (R^2) as evaluation indices. We can easily assess the performance of the prediction model in the test set using these evaluation indicators, which include the difference between the observed and predicted values as well as the degree of correlation between the observed and predicted values. The evaluation indices are represented by equations (18), (22), (23), (24), (25) and (26), where n is the number of samples, y_i is the observed value, x_i is the predicted value, $\overline{y_i}$ is the mean of y_i, and $\overline{x_i}$ is the mean of x_i.

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i| \# \tag{23}$$

$$MAPE = \sum_{i=1}^{n}\left|\frac{y_i - x_i}{y_i}\right| \times \frac{100}{n} \# \tag{24}$$

$$R = \frac{\sum_{i=1}^{n}(y_i - \overline{y_i})(x_i - \overline{x_i})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y_i})^2 \sum_{i=1}^{n}(x_i - \overline{x_i})^2}} \# \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \# \tag{26}$$

Table 2
Hyperparameter optimization values of NJI site model.

Algorithm	Hyperparameter	Optimization range	1-hour wave height	1-hour wave period	6-hour wave height	6-hour wave period
GRU	time_step	(2, 128)	47	21	4	19
	units	(2, 128)	33	19	128	22
	dense	(2, 128)	37	6	100	57
LSTM	time step	(2, 128)	12	26	40	85
	units	(2, 128)	24	11	108	61
	dense	(2, 128)	7	3	82	89
MLP	units	(2, 128)	53	37	110	40
RF	n estimators	(10, 200)	64	181	137	31
	max depth	(5, 10)	6	5	5	8

Table 3
Hyperparameter optimization values of BSG site model.

Algorithm	Hyperparameter	Optimization range	1-hour wave height	1-hour wave period	6-hour wave height	6-hour wave period
GRU	time_step	(2, 128)	79	33	69	6
	units	(2, 128)	34	11	35	31
	dense	(2, 128)	127	16	127	38
LSTM	time step	(2, 128)	25	48	68	10
	units	(2, 128)	16	37	13	103
	dense	(2, 128)	86	12	34	73
MLP	units	(2, 128)	21	48	99	3
RF	n estimators	(10, 200)	53	13	34	31
	max depth	(5, 10)	6	7	8	5

5. Results and discussion

5.1. Analysis of 1-hour prediction results of different algorithms

Table 4 shows the 1-hour wave height prediction results of the wave height prediction model at the two stations after the training of the four algorithms is completed. The best outcomes are shown in bold. The results show that because both LSTM and GRU are improved from RNN, they can effectively learn historical information. So, their prediction performance is better than MLP and RF. Among them, all the evaluation indicators based on the improved GRU proposed in this paper are optimal at the two stations. The prediction accuracy of the MLP is higher than that of the RF. It can be seen that in the prediction of the wave height, the prediction effect of the neural network is better than that of the traditional machine learning algorithm. Compared with the LSTM algorithm, in the wave height prediction of the NJI station, the *MSE* of the GRU based on the Bayesian optimization and attention mechanism is reduced by approximately 8.3 %, the *RMSE* is reduced by approximately 3.8 %, the *MAE* is reduced by approximately 10.9 %, the *MAPE* is reduced by approximately 12.4 %, the *R* is improved by approximately 12.4 %, and the R^2 is improved by approximately 0.5 %.

Observation data for a period of time at the two stations is compared with the prediction data of the four algorithms to reflect the model's 1-hour wave height prediction impact clearly, as shown in Fig. 7.

As shown in Fig. 7, the prediction data of the four algorithms are compared with the observation data of NJI and BSG stations for a period. It can be seen that the fitting effect of the GRU algorithm based on Bayesian optimization and attention mechanism outperforms other algorithms. Among them, the performance of LSTM and GRU is similar, and the prediction effect is satisfactory. In contrast, the prediction curve of MLP and RF algorithms fluctuates dramatically. In particular, the prediction effect of RF is not as good as that of GRU and LSTM, which may be related to the simpler model structure. Therefore, the GRU algorithm based on Bayesian optimization and attention mechanism reported here can predict the wave height more accurately.

Table 5 presents the 1-hour wave period prediction results of the four different algorithms at two different stations, with the best results highlighted in bold. Similar to the wave height prediction, the evaluation indicators of the GRU algorithm based on the Bayesian optimization and attention mechanism proposed in this article are the best, followed by LSTM and MLP, and RF is the worst. Compared with the LSTM algorithm, in the wave period prediction of the NJI station, the *MSE* of the GRU based on Bayesian optimization and attention mechanism is reduced by approximately 3.4 %, the *RMSE* is reduced by approximately 1.8 %, the *MAE* is reduced by approximately 0.6 %, the *MAPE* is reduced by approximately 0.5 %, the *R* is improved by approximately 0.2 %, and the R^2 is improved by approximately 0.4 %.

Fig. 8 depicts a comparison of the predicted and observed values of the four algorithms. From the NJI station, we can see that when the wave period changes smoothly, the prediction gap between the four algorithms is not large; from the BSG station, when the wave cycle fluctuates frequently, the deviation between the prediction curve of MLP and RF and the observation curve is large. The prediction accuracy has dropped. Since the GRU and the LSTM can make decisions in the future according to the changing laws of historical time series information, they can better fit the observations.

Table 6 displays the 1-hour wave energy forecast performance of the four algorithms using various evaluation indices, with the best results highlighted in bold. From the comparison of algorithms, the four algorithms have shown satisfactory results in the 1-hour wave energy prediction, and their *R* are all greater than 91 %. The GRU based on the Bayesian optimization and attention mechanism has the best performance in all evaluation metrics. In the NJI station, the *MAE* is 0.5555, and the R^2 is 91.27 %. Compared with the wave height and the wave period prediction, the prediction results of LSTM and MLP are similar, and there is no obvious difference. The above results verify that the four algorithms GRU, LSTM, MLP, and RF all have high prediction accuracy in 1-hour wave energy prediction, and the improved GRU proposed in this article is superior in 1-hour wave energy prediction.

Fig. 9 depicts a comparison of the predicted values and observed values of wave energy for the four algorithms in the 1-hour prediction. It can be seen from Fig. 8 that the wave energy in the NJI observatory shows a trend of first increasing and then decreasing with time, while the wave energy in the BSG observatory shows a trend of first decreasing and then basically stable. The prediction effect of each algorithm is similar to that of wave height prediction. The prediction value of GRU and LSTM is the best, which is closer to the observed value; the prediction value of MLP and RF is poor and tends to fluctuate and large errors.

Table 4
1-hour wave height forecast results.

Station	Algorithm	*MSE*	*RMSE*	*MAE*	*MAPE*	*R*	R^2
NJI	GRU	**0.0100**	**0.1002**	**0.0667**	**0.0658**	**0.9695**	**0.9397**
	LSTM	0.0109	0.1042	0.0749	0.0751	0.9676	0.9347
	MLP	0.0116	0.1078	0.0778	0.0802	0.9644	0.9301
	RF	0.0139	0.1180	0.0865	0.0868	0.9577	0.9163
BSG	GRU	**0.0085**	**0.0925**	**0.0612**	**0.0824**	**0.9583**	**0.9179**
	LSTM	0.0090	0.0949	0.0649	0.0905	0.9564	0.9135
	MLP	0.0099	0.0997	0.0692	0.0944	0.9536	0.9045
	RF	0.0101	0.1005	0.0725	0.1011	0.9506	0.9029

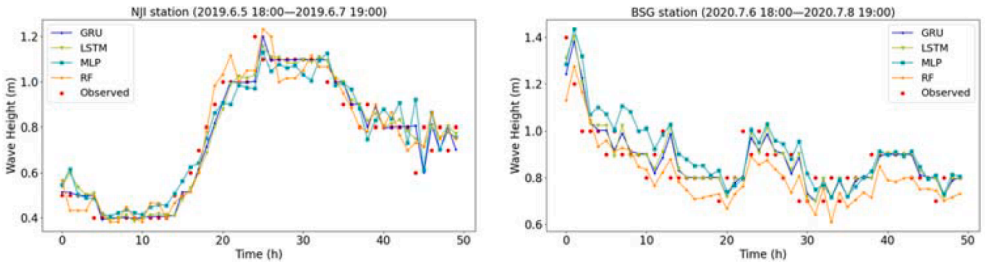


Fig. 7. 1-hour wave height forecast comparison.

Table 5
1-hour wave period forecast results.

Station	Algorithm	*MSE*	*RMSE*	*MAE*	*MAPE*	*R*	R^2
NJI	GRU	**0.1457**	**0.3816**	**0.2422**	**0.0440**	**0.9484**	**0.8993**
	LSTM	0.1508	0.3884	0.2436	0.0442	0.9465	0.8957
	MLP	0.1545	0.3931	0.2557	0.0466	0.9452	0.8932
	RF	0.1510	0.3886	0.2534	0.0463	0.9468	0.8956
BSG	GRU	**0.1233**	**0.3512**	**0.2643**	**0.0504**	**0.9355**	**0.8737**
	LSTM	0.1290	0.3591	0.2690	0.0509	0.9326	0.8679
	MLP	0.1405	0.3748	0.2825	0.0540	0.9276	0.8561
	RF	0.1442	0.3797	0.2820	0.0535	0.9256	0.8523

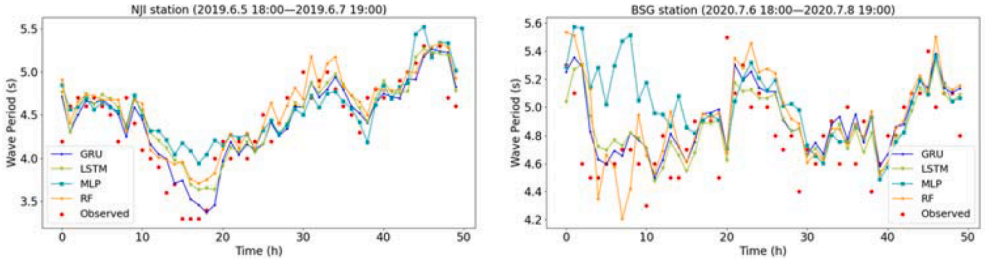


Fig. 8. 1-hour wave period forecast comparison.

Table 6
1-hour wave energy forecast results.

Station	Algorithm	*MSE*	*RMSE*	*MAE*	*MAPE*	*R*	R^2
NJI	GRU	1.1170	1.0569	**0.5555**	**0.1552**	**0.9554**	**0.9127**
	LSTM	1.2459	1.1162	0.5882	0.1651	0.9522	0.9027
	MLP	1.1537	1.0741	0.5943	0.1794	0.9540	0.9099
	RF	1.5407	1.2412	0.6694	0.1920	0.9389	0.8796
BSG	GRU	**1.3733**	**0.1719**	**0.3686**	**0.1876**	**0.9180**	**0.8422**
	LSTM	1.4531	1.2055	0.3805	0.2001	0.9161	0.8330
	MLP	1.4534	1.2056	0.4143	0.2157	0.9177	0.8330
	RF	1.4579	1.2074	0.3973	0.2161	0.9166	0.8324

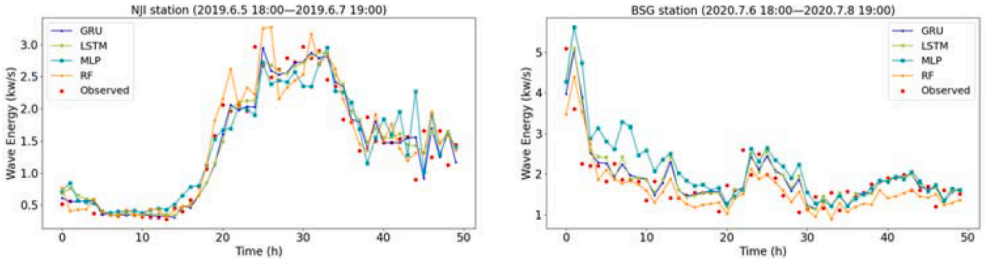


Fig. 9. 1-hour wave energy forecast comparison.

5.2. Analysis of 6-hour prediction results of different model algorithms

Table 7 summarizes the 6-hour wave height forecast results from the four algorithms at the two stations, with the best findings in bold. The table shows that when the predict time interval increases, the prediction accuracy of each method diminishes. Taking the GRU based on the

Table 7
6-hour wave height forecast results.

Station	Algorithm	*MSE*	*RMSE*	*MAE*	*MAPE*	*R*	R^2
NJI	GRU	**0.0509**	**0.2255**	**0.1619**	**0.1604**	**0.8354**	**0.6957**
	LSTM	0.0516	0.2271	0.1680	0.1732	0.8323	0.6915
	MLP	0.0522	0.2284	0.1668	0.1681	0.8298	0.6879
	RF	0.0557	0.2361	0.1731	0.1723	0.8264	0.6667
BSG	GRU	**0.0392**	**0.1980**	**0.1383**	**0.1880**	**0.8107**	**0.6274**
	LSTM	0.0414	0.2035	0.1463	0.2176	0.7956	0.6067
	MLP	0.0474	0.2178	0.1644	0.2520	0.7772	0.5492
	RF	0.0463	0.2151	0.1605	0.2450	0.7823	0.5603

Bayesian optimization and the attention mechanism proposed in this article as an example, at the NJI station, compared with the 1-hour wave height prediction, its *MSE* increased by about 409 %, *RMSE* increased by about 125 %, *MAE* increased is about 143 %, *MAPE* is increased by about 144 %, *R* is decreased by about 13.8 %, and R^2 is decreased by about 26 %. Even so, the performance of the GRU based on the Bayesian optimization and attention mechanism proposed in this article is still the best in all evaluation metrics. The prediction accuracy of LSTM is similar to that of the GRU.

Fig. 10 depicts a comparison of the predicted values of wave heights and the observed values of the four algorithms in the 6-hour prediction. The figure shows that as compared to the 1-hour prediction, the predicted value curve of each algorithm has a relatively obvious deviation compared with the observed value curve, and the deviation shows a slight hysteresis. When the wave height fluctuates, the forecast deviation is more serious. When multiple algorithms are compared, it is clear that the predicted value curve of GRU is closer to the observed value, which is notably noticeable at the BSG station.

Table 8 summarizes the 6-hour wave period prediction results from the four algorithms at the two stations, with the best results in bold. The table shows that, as with the 6-hour wave height prediction, increasing the forecast time interval reduces the forecasting accuracy of each method, but the forecasting accuracy remains within the acceptable range. The performance of the GRU based on Bayesian optimization and attention mechanism proposed in this article is still the best in all evaluation indicators. Compared with the 1-hour wave period prediction, at the NJI station, the *MSE* of the GRU increased by about 240 %, the *RMSE* increased by about 84.5 %, the *MAE* increased by about 115 %, the *MAPE* increased by about 111.6 %, the R decreased by about 13.9 %, and the R^2 decreased by about 27.1 %.

Fig. 11 compares the predicted and observed curves for the four methods for the 6-hour wave period. Compared with Fig. 8, the prediction deviation of each algorithm increases, and the RF algorithm is the most serious. In the curve graph of the NJI station, although the observed value of the wave period has been decreasing, the numerical fluctuation is small, and the deviation between the predicted curve and the observed value curve of each algorithm is also small, especially the GRU, the fitting effect is better. In the curve graph of the BSG station, the observed value of the wave period fluctuates frequently up and down. In this case, the deviation between the predicted curve of each algorithm and the observed value curve is also larger. As a result, the model's prediction accuracy under numerical fluctuation still has to be improved.

Table 9 presents the 6-hour wave energy forecast results of the four algorithms at two stations using the wave height, period, and power conversion equation, with the optimum result highlighted in bold. As seen in the table, the accuracy of each algorithm has dropped when compared to Table 6. Because the GRU based on the Bayesian optimization and attention mechanism proposed in this article is the best in the prediction of 6-hour wave height and period, its prediction accuracy is still the highest in wave energy prediction.

Fig. 12 depicts a comparison of the four algorithms' 6-hour wave energy predicted values with the observed values. From the NJI station, since numerical fluctuations, the predicted values of each algorithm have obvious deviations compared with the observed values. At the BSG station, since the wave energy changes relatively smoothly, the prediction effect of each algorithm is better. In summary, in the case of stable numerical fluctuations, the prediction accuracy of the algorithm is higher.

5.3. Discussion

In this study, the data of two selected NJI and BSG observatories were experimentally analyzed to evaluate the performance of the GRU wave energy prediction model based on Bayesian optimization and attention mechanism. The results indicate that in the one-hour wave height and wave period prediction, the model can well fit the actual observed values. This is because the Bayesian optimization algorithm is used to optimize the hyperparameters of the model. Besides, the attention mechanism assigns different weights to the model features, so that the model can obtain excellent prediction results after a lot of training. Similarly, the 6-hour wave height and wave prediction curve of the model reported here can fit the observed value curve better than models such as LSTM, MLP and RF. Therefore, the 1-hour and 6-hour forecast result of the model reported here is compared with the current popular algorithms. The experimental results suggest that this model outperforms them. This fully reflects the hyperparameter optimization effect of Bayesian optimization algorithm and the improvement of model prediction performance after the introduction of attention mechanism, which is consistent with the research of Zeng et al. (2022) [47]. Therefore, this study can provide a reference for the realization of carbon neutrality in the shipping process through the prediction of wave energy.

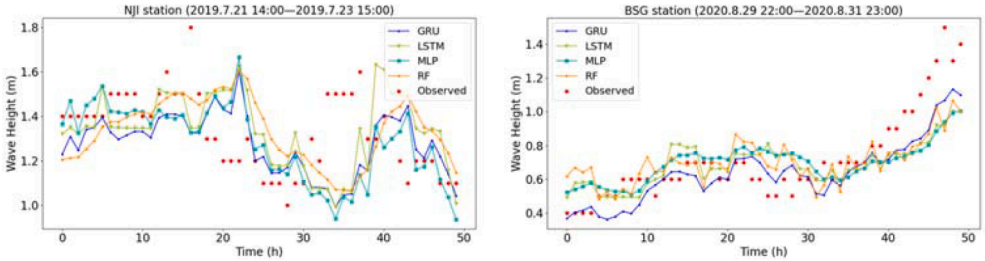


Fig. 10. 6-hour wave height forecast comparison.

Table 8
6-hour wave period forecast results.

Station	Algorithm	*MSE*	*RMSE*	*MAE*	*MAPE*	*R*	R^2
NJI	GRU	**0.4958**	**0.7041**	**0.5218**	**0.0931**	0.8170	**0.6556**
	LSTM	0.5039	0.7099	0.5243	0.0960	0.8163	0.6499
	MLP	0.5028	0.7091	0.5260	0.0951	**0.8183**	0.6507
	RF	0.5679	0.7536	0.5544	0.1000	0.7910	0.6055
BSG	GRU	**0.4411**	**0.6642**	**0.5011**	**0.0957**	**0.7511**	**0.5554**
	LSTM	0.4774	0.6910	0.5145	0.0963	0.7359	0.5188
	MLP	0.4597	0.6780	0.5164	0.0999	0.7441	0.5367
	RF	0.5122	0.7157	0.5505	0.1063	0.7141	0.4838

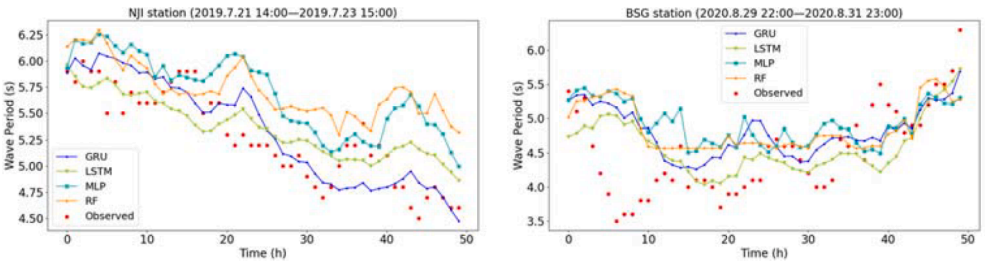


Fig. 11. 6-hour wave period forecast comparison.

Table 9
6-hour wave energy forecast results.

Station	Algorithm	*MSE*	*RMSE*	*MAE*	*MAPE*	*R*	R^2
NJI	GRU	4.6245	2.1505	**1.2011**	**0.3561**	**0.8045**	**0.6436**
	LSTM	4.6835	2.1641	1.2490	0.4088	0.7997	0.6390
	MLP	4.8384	2.1996	1.2631	0.3891	0.7932	0.6271
	RF	5.4742	2.3397	1.3616	0.4093	0.7722	0.5781
BSG	GRU	**5.7110**	**2.3898**	**0.8204**	**0.4272**	**0.6653**	**0.3549**
	LSTM	8.1342	2.8520	0.8533	0.5028	0.5600	0.0812
	MLP	6.1702	2.4840	0.9070	0.6285	0.6471	0.3031
	RF	6.7340	2.5950	0.9249	0.6176	0.6112	0.2394

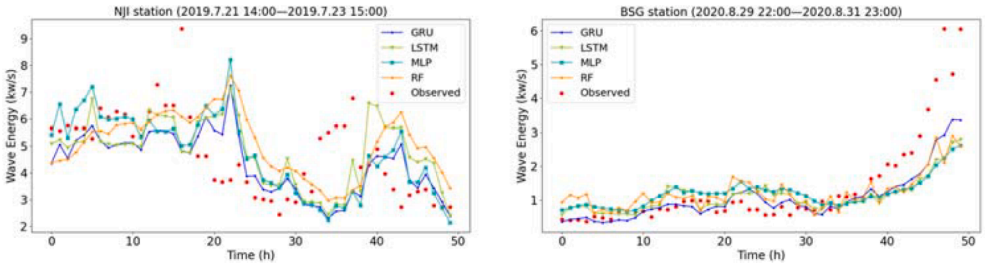


Fig. 12. 6-hour wave energy forecast comparison.

6. Conclusion

In recent years, many countries have proposed the goal of achieving carbon neutrality. Greenhouse gas emissions from the combustion of fossil fuels by transport ships place a significant pressure on the environment. It is crucial to develop a clean ship power source that can replace conventional fuels, which could contribute to carbon neutrality. The wave energy is one of the most important clean energy sources in the ocean and has many advantages over other energy sources. Therefore, this study combines wave energy prediction with ship driving. A Bayesian optimization algorithm is added to the original GRU to optimize the hyperparameters. At the same time, an attention mechanism is introduced to construct a GRU wave energy prediction model optimized by Bayesian optimization and attention mechanism. Finally, this study selects data from two Chinese stations to analyze the performance of the model. It is found that, the model reported here has the highest accuracy in predicting wave height, wave period, and wave energy compared with the three mainstream algorithms of LSTM, MLP and RF. In addition, in the 1-hour and 6-hour wave energy forecasts of the two stations, the minimum values are 0.3686 and 0.8204, and the maximum values are 0.9127 and 0.6436, respectively. The GRU based on Bayesian optimization and attention mechanism reported here can achieve accurate prediction of 1-hour and 6-hour wave energy power, providing strong support for ships to navigate using wave energy.

Still, this study has some shortcomings. For example, this study selected the GRU algorithm as one of the core algorithms, which is a relatively basic algorithm. However, more advanced improved or combined algorithms will appear as computer science advances. Therefore, the follow-up research will improve the performance of the core algorithm in the model. Secondly, the two observation stations selected here are specific locations in the South Sea of China, so it is unclear whether the Bohai Sea or the Yellow Sea will have an impact on the results.

Therefore, the number and location of observation stations can be further increased in the future to achieve the practical application of the constructed model as soon as possible.

CRediT authorship contribution statement

Zhihan Lv: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Nana Wang:** Methodology, Investigation. **Ranran Lou:** Methodology, Software, Investigation, Writing – original draft. **Yajun Tian:** Writing – original draft, Writing – review & editing, Supervision. **Mohsen Guizani:** Writing – original draft, Writing – review & editing, Supervision.

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.

Acknowledgements

Funding: This work was supported by the National Natural Science Foundation of China [grant number 61902203]; the open project of the Xinhua News Agency State Key Laboratory of Media Convergence Production Technology and System [grant number SKLMCPTS202103011].

References

[1] Zhou J, Zhang Y, Zhang Y, et al. Parameters identification of photovoltaic models using a differential evolution algorithm based on elite and obsolete dynamic learning[J]. Appl Energy 2022;314:118877.
[2] Bi H, Shang WL, Chen Y, et al. GIS aided sustainable urban road management with a unifying queueing and neural network model[J]. Appl Energy 2021;291:116818.
[3] Von Wald G, Sundar K, Sherwin E, et al. Optimal gas-electric energy system decarbonization planning[J]. Adv Appl Energy 2022;6:100086.
[4] He W, King M, Luo X, et al. Technologies and economics of electric energy storages in power systems: review and perspective[J]. Adv Appl Energy 2021;4:100060.
[5] Zhao X, Ma X, Chen B, et al. Challenges toward carbon neutrality in China: Strategies and countermeasures[J]. Resour Conserv Recycl 2022;176:105959.
[6] Martínez-Gordón R, Sánchez-Diéguez M, Fattahi A, et al. Modelling a highly decarbonised North Sea energy system in 2050: A multinational approach[J]. Adv Appl Energy 2022;5:100080.
[7] Zhao F, Liu X, Zhang H, et al. Automobile Industry under China's Carbon Peaking and Carbon Neutrality Goals: Challenges, Opportunities, and Coping Strategies[J]. J Adv Transp 2022.
[8] Weng Y, Cai W, Wang C. Evaluating the use of BECCS and afforestation under China's carbon-neutral target for 2060[J]. Appl Energy 2021;299:117263.
[9] Gray N, McDonagh S, O'Shea R, et al. Decarbonising ships, planes and trucks: An analysis of suitable low-carbon fuels for the maritime, aviation and haulage sectors [J]. Adv Appl Energy 2021;1:100008.
[10] Lira-Loarca A, Ferrari F, Mazzino A, et al. Future wind and wave energy resources and exploitability in the Mediterranean Sea by 2100[J]. Appl Energy 2021;302:117492.
[11] Melikoglu M. Current status and future of ocean energy sources: A global review [J]. Ocean Eng 2018;148:563–73.
[12] Choupin O, Têtu A, Del Río-Gamero B, et al. Premises for an annual energy production and capacity factor improvement towards a few optimised wave energy converters configurations and resources pairs[J]. Appl Energy 2022;312:118716.
[13] Akpınar A, Kömürcü Mİ. Assessment of wave energy resource of the Black Sea based on 15-year numerical hindcast data[J]. Appl Energy 2013;101:502–12.
[14] Blackledge J, Coyle E, Kearney D, et al. Estimation of wave energy from wind velocity[J]. Eng Lett 2013;21(4):158–70.
[15] López I, Andreu J, Ceballos S, et al. Review of wave energy technologies and the necessary power-equipment[J]. Renew Sustain Energy Rev 2013;27:413–34.
[16] Pan P, Sun Y, Yuan C, et al. Research progress on ship power systems integrated with new energy sources: a review[J]. Renew Sustain Energy Rev 2021;144:111048.
[17] Leung SF, Fu HC, Zhang M, et al. Blue energy fuels: converting ocean wave energy to carbon-based liquid fuels via CO 2 reduction[J]. Energ Environ Sci 2020;13(5):1300–8.
[18] Cotorcea A, Özkaynak S, Nicolae F, et al. Present and future of renewable energy sources onboard ships. case study: solar-thermal systems[J]. Scientific Bulletin" Mircea cel Batran" Naval Academy 2014;17(1):35.
[19] L, Robotics Energy harvesting ocean robot[J]. 2017-12-04)[2018-01-20]. https://www. liquid-robotics. com/platform/how-it-works.
[20] Xu X, Robertson B, Buckham B. A techno-economic approach to wave energy resource assessment and development site identification[J]. Appl Energy 2020;260:114317.
[21] Lou R, Lv Z, Dang S, et al. Application of machine learning in ocean data[J]. Multimedia Syst 2021:1–10.
[22] James SC, Zhang Y, O'Donncha F. A machine learning framework to forecast wave conditions[J]. Coast Eng 2018;137:1–10.
[23] Mahjoobi J, Mosabbeb EA. Prediction of significant wave height using regressive support vector machines[J]. Ocean Eng 2009;36(5):339–47.
[24] Mahjoobi J, Etemad-Shahidi A. An alternative approach for the prediction of significant wave heights based on classification and regression trees[J]. Appl Ocean Res 2008;30(3):172–7.
[25] Gao S, Huang J, Li Y, et al. A forecasting model for wave heights based on a long short-term memory neural network[J]. Acta Oceanol Sin 2021;40(1):62–9.
[26] Fan S, Xiao N, Dong S. A novel model to predict significant wave height based on long short-term memory network[J]. Ocean Eng 2020;205:107298.
[27] Ni C, Ma X. An integrated long-short term memory algorithm for predicting polar westerlies wave height[J]. Ocean Eng 2020;215:107715.
[28] Xiang L, Yang X, Hu A, et al. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks[J]. Appl Energy 2022;305:117925.
[29] Tuttle JF, Blackburn LD, Andersson K, et al. A systematic comparison of machine learning methods for modeling of dynamic processes applied to combustion emission rate modeling[J]. Appl Energy 2021;292:116886.
[30] Li X, Cao J, Guo J, et al. Multi-step forecasting of ocean wave height using gate recurrent unit networks with multivariate time series[J]. Ocean Eng 2022;248:110689.
[31] Niu D, Yu M, Sun L, et al. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism[J]. Appl Energy 2022;313:118801.
[32] Yang D, Guo J, Sun S, et al. An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting[J]. Appl Energy 2022;306:117992.
[33] He F, Zhou J, Feng Z, et al. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm[J]. Appl Energy 2019;237:103–16.
[34] Fernández JC, Salcedo-Sanz S, Gutiérrez PA, et al. Significant wave height and energy flux range forecast with machine learning classifiers[J]. Eng Appl Artif Intel 2015;43:44–53.
[35] Hochreiter S. Untersuchungen zu dynamischen neuronalen Netzen[J]. Diploma, Technische Universität München 1991;91(1).
[36] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Comput 1997;9(8):1735–80.
[37] A, Zhang, Z C, Lipton M, Li et al. Dive into deep learning[J]. arXiv preprint arXiv:2106.11342, 2021.
[38] M, Sundermeyer, R, Schlüter Ney H. LSTM neural networks for language modeling [C]//Thirteenth annual conference of the international speech communication association. 2012.
[39] K, Cho, B, Van Merriënboer, C, Gulcehre et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.
[40] Bergstra J, Bengio Y. Random search for hyper-parameter optimization[J]. J Mach Learn Res 2012;13(2).
[41] Snoek J, Larochelle H. Adams mation pR P. Practical bayesian optimization of machine learning algorithms[J]. Adv neural inforrocessing systems 2012:25.
[42] Zhu X, Guo K, Ren S, et al. Lightweight image super-resolution with expectation-maximization attention mechanism[J]. IEEE Trans Circuits Syst Video Technol 2021;32(3):1273–84.
[43] Duan Y, Li H, He M, Zhao D. A BiGRU autoencoder remaining useful life prediction scheme with attention mechanism and skip connection. IEEE Sens J 2021;21(9):10905–14.
[44] Hao S, Lee DH, Zhao D. Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system[J]. Trans Res Part C: Emerging Technol 2019;107:287–300.
[45] Mahjoobi J, Etemad-Shahidi A, Kazeminezhad MH. Hindcasting of wave parameters using different soft computing methods. Appl Ocean Res 2008;30(1):28–36.
[46] Nikoo MR, Kerachian R, Alizadeh MR. A fuzzy KNN-based model for significant wave height prediction in large lakes[J]. Oceanologia 2018;60(2):153–68.
[47] Zeng L, Ren W, Shan L, et al. Well logging prediction and uncertainty analysis based on recurrent neural network with attention mechanism and Bayesian theory [J]. J Pet Sci Eng 2022;208:109458.