Acta Oceanol. Sin., 2024, Vol. 43, No. 12, P. 13–33
https://doi.org/10.1007/s13131-024-2450-4
http://www.aosocean.com
E-mail: ocean2@hyxb.org.cn

A significant wave height prediction method with ocean characteristics fusion and spatiotemporal dynamic graph modeling

Xiao Yin[1], Taoxing Wu[1], Jie Yu[1], Xiaoyu He[2], Lingyu Xu[1]*

[1] Department of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

[2] School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China

Received 5 February 2024; accepted 8 September 2024

Abstract

Accurate significant wave height (SWH) prediction is essential for the development and utilization of wave energy. Deep learning methods such as recurrent and convolutional neural networks have achieved good results in SWH forecasting. However, these methods do not adapt well to dynamic seasonal variations in wave data. In this study, we propose a novel method—the spatiotemporal dynamic graph (STDG) neural network. This method predicts the SWH of multiple nodes based on dynamic graph modeling and multi-characteristic fusion. First, considering the dynamic seasonal variations in the wave direction over time, the network models wave dynamic spatial dependencies from long- and short-term pattern perspectives. Second, to correlate multiple characteristics with SWH, the network introduces a cross-characteristic transformer to effectively fuse multiple characteristics. Finally, we conducted experiments on two datasets from the South China Sea and East China Sea to validate the proposed method and compared it with five prediction methods in the three categories. The experimental results show that the proposed method achieves the best performance at all predictive scales and has greater advantages for extreme value prediction. Furthermore, an analysis of the dynamic graph shows that the proposed method captures the seasonal variation mechanism of the waves.

Key words: significant wave height forecasting, dynamic seasonal variation, dynamic graph neural networks

Citation: Yin Xiao, Wu Taoxing, Yu Jie, He Xiaoyu, Xu Lingyu. 2024. A significant wave height prediction method with ocean characteristics fusion and spatiotemporal dynamic graph modeling. Acta Oceanologica Sinica, 43(12): 13–33, doi: 10.1007/s13131-024-2450-4

1 Introduction

Consumption of nonrenewable energy sources is a major factor leading to global climate change and rising carbon dioxide levels. To mitigate the impact of climate change on human society, it is essential to develop renewable energy sources, such as solar (Ma et al., 2022), wind (Ren et al., 2023), and ocean energy (Gao and Xiao, 2021). Among them, wind and solar energy have been widely adopted because of their technological maturity and economic feasibility, whereas wave energy has better security, consistency, and certainty than wind energy (Reikard et al., 2011). However, the dynamic characteristics of waves affect their reliability, which in turn affects energy absorption and conversion by wave-energy converters (WECs).

Significant wave height (SWH) is an important indicator for describing wave qualities and is also one of the key parameters for wave energy generation (Ali and Prasad, 2019). Therefore, effective SWH prediction can assist decision-makers and provide references for engineers, project developers, and grid operators. Currently, there are two mainstream forecasting strategies: one is data-driven forecasting based on time series, and the other is wave forecasting based on wind field-driven wave fields. This study combines these two forecasting methods on a data-driven basis. It considers the driving influence of historical wind and wave information on waves.

One of the most direct and reliable methods of obtaining wave height is *in situ* measurement. However, *in situ* measurements are not always available in most locations. This is due to the high cost of specialized equipment such as buoys and collection vessels and the spatial sparsity of collection sites. Traditional wave prediction methods based on numerical models (Umair et al., 2019) are becoming alternatives for wave prediction and forecasting. This method achieves relatively satisfactory prediction results by establishing an energy balance equation (Xu et al., 2020). The wave simulation of the numerical model is driven by the spatial-temporal wind field and the wave field with boundary conditions to describe the basic mechanisms of wave generation and propagation. However, because numerical models use mathematical and physical models to simulate wave evolution, they consume considerable computational resources and time (Mahjoobi and Mosabbeb, 2009).

Considering the problems existing in numerical models, many researchers have begun to study data-driven methods for SWH prediction. These methods do not rely on explicit relationships between input and output variables but use historical data to learn wave patterns and features. Classical time-series methods, such as autoregressive (Soares and Cunha, 2000) and autoregressive and moving average models (Ge and Kerrigan, 2016), have been used for wave height prediction. However, these methods have a common limitation in that they are based on linearity and stationarity assumptions (Duan et al., 2016). This makes it difficult to capture nonstationary and nonlinear features in waves.

In recent decades, machine learning has achieved great success in fields such as solar radiation (Zhang et al., 2022) and wind

Foundation item: The National Key R&D Program of China under contract No. 2021YFC3101604.
*Corresponding author, E-mail: xly@shu.edu.cn

speed (WSD) prediction (Khosravi et al., 2018), because it can perform nonlinear modeling without these assumptions. Wave height prediction has been studied and verified using traditional machine learning algorithms, such as artificial neural networks (Feng et al., 2020), support vector machines (Mahjoobi and Mosabbeb, 2009) and decision trees (Mahjoobi and Etemad-Shahidi, 2008), which can provide acceptable wave height prediction results. However, these methods are easily affected by noise, which limits wave height prediction reliability (Chen et al., 2021).

Deep learning is a machine learning subclass that has been successfully applied to wave height prediction (Song et al., 2021; Ti et al., 2022) owing to its powerful feature extraction ability. Considering wave height data are represented as a time series, early studies have primarily focused on exploring temporal dependencies in the data. Mandal and Prabaharan (2006) established a recurrent neural network with a resilient backpropagation algorithm that outperformed artificial neural networks using short-term memory of early input wave-height data. However, recurrent neural networks (RNNs) suffer from gradient vanishing and explosions; therefore, they are rarely applied to practical engineering tasks. To address these problems, Pirhooshyaran and Snyder (2020) used a long short-term memory (LSTM) network (Hochreiter and Schmidhuber, 1997) to predict wave height, which effectively alleviated these problems and showed excellent prediction performance. The gated recurrent unit (GRU) (Cho et al., 2014) simplifies the LSTM network structure (Xiang et al., 2022), with fewer parameters and ease of training, and many application studies have also proved that the GRU is superior to LSTM in terms of accuracy and reliability (Tuttle et al., 2021). In wave height forecasting, Li et al. (2022) used a GRU to predict the wave height of six different coastal stations in China, and they also found that a GRU showed higher robustness than LSTM in long-term forecasting.

Wave height data from multiple nodes in the same sea area exhibit temporal and spatial dependencies. These are multiple time-series data points collected from different locations (longitude and latitude coordinates). However, most studies only used data from a single node to predict the future nodal wave height, ignoring the influence of data from other nodes on the prediction results.

Some studies have used data from multiple nodes; however, the potential for multi-node wave height prediction has not been fully explored. For example, Tsai et al. (2002) proposed a method that used data from two adjacent nodes within Taichung Harbor, Taiwan, to forecast the wave height at a third node, which improved the prediction accuracy compared with single-node methods. However, their method did not consider the spatial dependencies among nodes, which may have affected prediction performance. Wang et al. (2018) developed a method to use data from multiple nodes in two different geographical regions (China's Bohai Sea and Yellow Sea) to predict wave height, which captured regional wave energy characteristics. However, their method treated six nodes in the same sea area as one node to predict the wave height in that sea area, which may have resulted in some spatial information loss.

In the process of developing and using wave energy, it is not only necessary to predict the wave height of a single node but also the wave height of multiple nodes in the sea area simultaneously to understand the distribution of wave energy in that sea area and adjust the layout of WECs. It is especially important that WECs must be arranged in arrays on or below the sea surface, which can not only reduce the cost of fixing WEC nodes but also take advantage of the constructive interference between WEC nodes to generate more electricity (De Andrés et al., 2014; Yang et al., 2022). Therefore, multi-node wave height prediction is essential for WEC array layout design and optimization.

There are two primary types of multi-node SWH prediction methods: one based on ocean images and video data and the other based on transforming multi-node data into two-dimensional data. Quach et al. (2021) used synthetic aperture radar and satellite data to predict wave height. Song et al. (2021) used wave-monitoring video data near the coast to predict wave height. Zhou et al. (2021) used interpolated two-dimensional reanalysis data to establish a convolutional LSTM network that achieved high-precision and high-efficiency wave forecasting in the coastal waters of China. However, these methods lack interpretability because they ignore multi-node data spatial dependence.

Graph neural networks (GNNs) have received considerable attention from researchers in recent years, as they can represent data as graph structures and effectively model the dependency relationships in non-Euclidean spaces. Zhang et al. (2023) proposed a multi-characteristic, multi-node (MCMN) graph neural network that can predict SWH of multiple nodes in the South China Sea and East China Sea. This method integrates the influence of multiple characteristics and uses multiple characteristic fusion to improve prediction accuracy. However, this method also has some limitations in predicting SWH, as it does not fully account for the formation mechanism and propagation pattern of waves, which may result in some discrepancy between the predicted and observed values.

Wave data not only need to consider temporal and spatial dependence but also need to consider the driving influence of wind and wave information on waves. However, the traditional third-generation wave model [such as WAVEWATCH Ⅲ (Tolman, 2009)] uses wave action conservation theory to describe the driving process of wind and waves to predict the wave height in space. This method suffers from the same limitations as numerical models, which are very time-consuming and computationally expensive. Therefore, an efficient and accurate alternative to traditional wave models can be provided by using deep learning technology to study the driving influence of wind and wave information on waves.

The formation and propagation of waves in space is a complex process, and wave height is influenced by both long- and short-term patterns, resulting in dynamic seasonal wave variation. Figure 1 shows the monthly average of the mean wave directions of multiple nodes in the South China Sea in 2020; the data comes from the ERA5 monthly averaged data on single levels (Hersbach et al., 2023b). As shown in Figs 1a–d, the multi-node mean wave direction was from northeast to southwest, whereas in Figs 1f–h, the multi-node mean wave direction was south to north, and in Figs 1j–l, the multi-node mean wave direction was primarily northeast to southwest. This indicates that multi-node mean wave directions in the South China Sea vary dynamically in space, and this variation can be divided into two patterns: short- and long-term (Kitagawa and Gersch, 1984; Oreshkin et al., 2020). The short-term pattern refers to waves that are primarily driven by wind; therefore, wave height is affected by WSD and duration in the short term. The long-term pattern refers to the existence of stable or periodic fluctuations in the ocean, such as tides and internal waves, which are affected by celestial gravity, uneven seawater density distribution, and other long-term characteristics. Therefore, when modeling dynamic seasonal variations in waves, it is important to consider not only the effects of

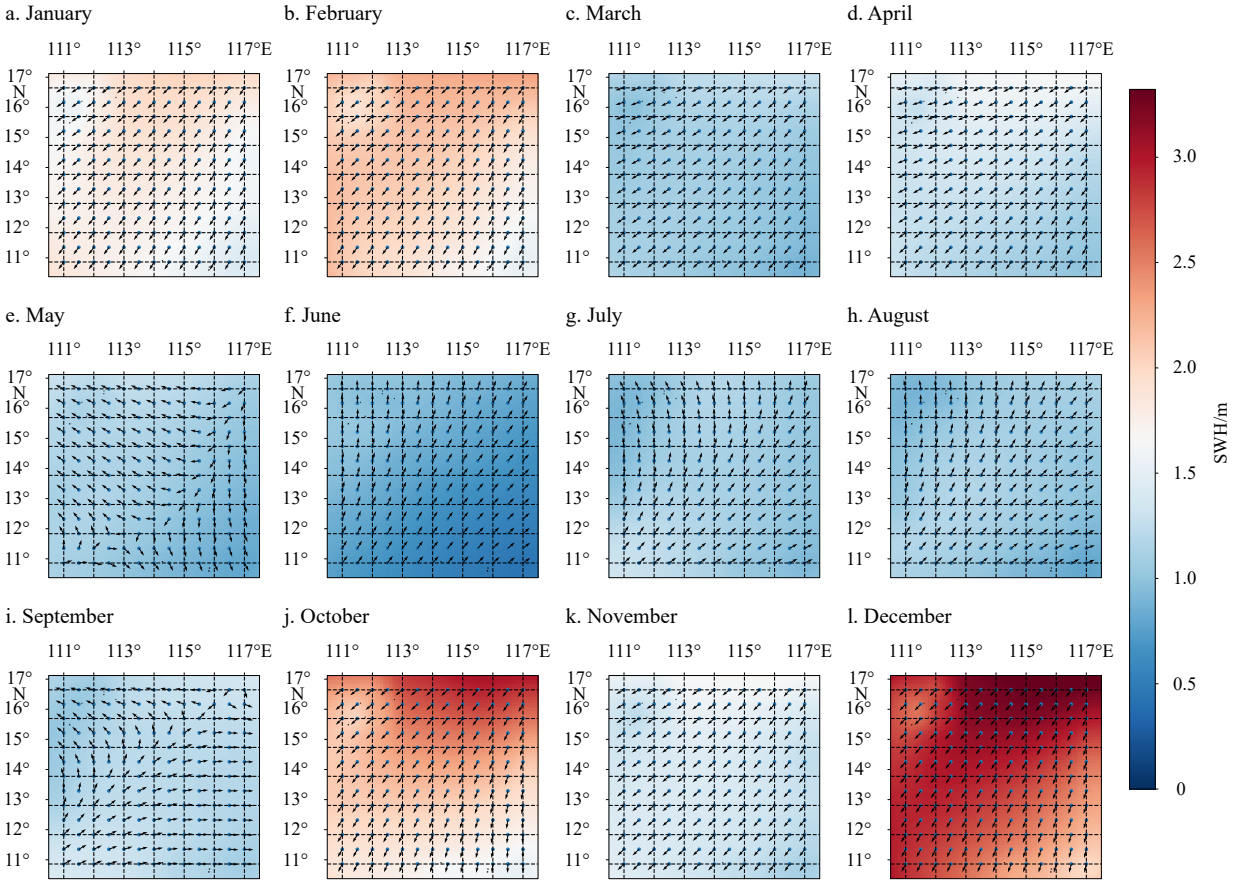

Fig. 1. Mean wave direction in the South China Sea for the 2020 monthly average; the data comes from the ERA5 monthly averaged data on single levels. The color bar represents monthly average significant wave heights (SWHs), and each arrow represents the mean wave direction at the longitude and latitude coordinates in the data.

long-term patterns but also the effects of short-term patterns of dynamic wave drivers (historical wind and wave effects).

To achieve the above objectives, this study proposes a new method, namely spatiotemporal dynamic graph (STDG) neural networks, which can predict multiple node SWH in a long-term and robust manner. The primary contributions of this study are as follows.

(1) A GNN based on dynamic graph modeling is proposed, which considers seasonal wave direction variations over time and models the spatial dependence of waves from long- and short-term pattern perspectives. In addition, a cross-characteristic transformer was introduced to effectively combine multiple characteristics that influence SWH.

(2) A spatial dependency graph construction (SDGC) module was designed that comprehensively considers the long-term pattern and dynamic wind and wave driving characteristics [wind direction (DWI), mean wave direction, and WSD]. By introducing GRUs and a self-attention mechanism (Vaswani et al., 2017), a dynamic graph consistent with the seasonal wave variation mechanism was constructed.

(3) The practicability and robustness of the proposed method were verified on the South China Sea and East China Sea datasets and compared with five prediction methods in three categories. The results show that the proposed method achieved the best performance at all prediction scales and has greater advantages for extreme-value prediction.

The remainder of this study is organized as follows: Section 2 introduces the proposed SWH prediction method. Section 3 introduces the South China Sea and East China Sea datasets used in this study, as well as the data preprocessing and key parameter settings. Section 4 presents the experimental analysis, evaluation, and discussion. Finally, Section 5 summarizes the conclusions and possible directions for future research.

2 Method

2.1 *Problem formulation*

The multi-node and multi-characteristic ocean data can be represented as $X \in \mathbb{R}^{T\times N\times M}$, where T is the number of time steps, N is the number of nodes, each of which is defined by a coordinate point of longitude and latitude, and M is the number of ocean characteristics. For each time step $t \in \{1, \ldots, T\}$, each node $i \in \{1, \ldots, N\}$ is associated with a feature vector of M ocean characteristics $F_i^{(t)} \in \mathbb{R}^M$. For time step t, let $X_t = \{F_1^{(t)}, \ldots, F_N^{(t)}\}$ be the snapshot record of M characteristics for all nodes.

In this study, the spatial dependency is represented by a graph $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_N\}$ is the node set consisting of N nodes and E is the edge set. As wave propagation has a certain spatial directionality, the spatial dependence is represented by the adjacency matrix of a directed graph. Specifically, the graph adjacency matrix $A_{\text{adj}} \in \mathbb{R}^{N\times N}$ of the directed graph indicates the spatial dependency between nodes, and the value of $A_{\text{adj}}^{i,j}$ indicates the strength of the connection. For any two points $v_i, v_j \in V$, $A_{\text{adj}}^{i,j} > 0$ indicates that there is a directed connection from point v_i to v_j, and $A_{\text{adj}}^{i,j} = 0$ indicates that there is no connection between the two points.

This study aims to train a model f with parameters θ to predict the future Q steps of SWH $\widehat{X}^{\text{swh}}_{(t+1):(t+Q)}$, given P historical steps of SWH data $X^{\text{swh}}_{(t-P+1):t}$, auxiliary characteristics $X^{\text{aux}}_{(t-P+1):t}$, driving characteristics $X^{\text{drive}}_{(t-P+1):t}$ and spatial dependency graph G. The multi-node SWH prediction problem can be formulated as follows:

$$\widehat{X}^{\text{swh}}_{(t+1):(t+Q)} = f_\theta\left(X^{\text{swh}}_{(t-P+1):t}, X^{\text{aux}}_{(t-P+1):t}, X^{\text{drive}}_{(t-P+1):t}, G\right), \quad (1)$$

where SWH is the target variable to be predicted; auxiliary characteristics include the mean wave period (MWP), sea surface temperature (SST), and mean sea level pressure (MSL), which affect wave characteristics and changes; and the driving characteristics include WSD, DWI, and mean wave direction (MWD), which determine wave formation and propagation.

2.2 *Model structure*

The proposed model overall framework is shown in Fig. 2. The STDG model comprises four primary components: SDGC, temporal convolution, graph convolution, and cross-characteristic transformer fusion (CCTF) modules.

The STDG model must control three types of characteristics: SWH data X^{swh}, auxiliary X^{aux} and driving X^{drive} characteristics. The SWH data and three auxiliary characteristics were fed into the primary module and each auxiliary module, respectively, where they passed through L layers of temporal and graph convolutions to extract spatiotemporal features. In each auxiliary module layer, multiple cross-characteristic transformers were concatenated to fuse the information from the three auxiliary modules, thereby enhancing SWH representation by using the relevant information from the auxiliary modules. In addition, the SDGC module generates spatial dependencies for long-term patterns without relying on any priori graph structure and then fuses them with the driving characteristics to construct short-term patterns to capture the dynamic associations between different nodes. Subsequently, the graph convolution module integrated the constructed short- and long-term pattern spatial dependencies to further refine the SWH spatiotemporal features and auxiliary characteristics. Finally, the hidden feature channel and time dimensions were projected onto the desired output dimension to produce the prediction outcome.

2.3 *SDGC module*

Considering the original multi-node wave time-series data do not provide any spatial information, such as the adjacency matrix, the spatial dependencies must be learned by the neural network itself. To this end, this study employed node embedding to train long-term patterns adaptively (Gao et al., 2022).

(1) Long-term pattern learning:

Two node embeddings $E_1, E_2 \in \mathbb{R}^{N\times dim}$ with trainable parameters were randomly initialized, where N is the number of nodes and dim is the embedding dimension. The long-term pattern $\widetilde{A}_{\text{long}}$ are then computed by Eq. (2). Among them, the ReLU activation function is used to eliminate weak connections in the weights, and the softmax activation function is used to normalize the weights.

$$\widetilde{A}_{\text{long}} = \text{softmax}\left(\text{ReLU}\left(E_1 E_2^T\right)\right). \quad (2)$$

Considering the wave short-term pattern fluctuates around the long-term pattern, this fluctuation is primarily related to the dynamic changes in the wave (Zhang et al., 2023), and the short-term pattern spatial dependencies must consider both the long-term pattern and dynamic changes in the wave. Previous studies have demonstrated that the GRU can effectively fuse the current input data and retain long-term memory information through a gating mechanism, thus exhibiting advantages in capturing short-term patterns, and multi-head self-attention (Vaswani et al., 2017) can effectively capture the internal data correlations, making the model more attentive to useful information (Niu et al., 2022). Therefore, this study uses a method combining GRU and multi-head self-attention to fuse information on the dynamic potential correlation between wind and waves based on the long-term pattern, thereby constructing a short-term pattern.

(2) Dynamic information fusion of wind and waves

The key to constructing short-term patterns is to consider the impact of wind and wave driving characteristics on future SWH in space and time. This is because waves in offshore areas are a combination of locally generated wind waves and waves propagating from a distance (Goda, 2010). Therefore, the driving characteristics (wind direction, mean wave direction, and wind speed) are first dimensioned using Eq. (3). Among them, MLP$(\cdot)$ represents a linear layer, and LN$(\cdot)$ is layer normalization (Ba et al., 2016).

$$X^d = \text{LN}\left(\text{MLP}\left(X^{\text{drive}}\right)\right). \quad (3)$$

where d represents the time dimension of the characteristic.

In this study, GRU is used to perform information fusion on the two node embeddings E_i $(i = 1, 2)$ of the long-term pattern, respectively. The driving characteristics of the processed winds and waves are fused with the node embedding of the long-term pattern to form new information using Eq. (4). And a certain percentage of new information is saved as new node embedding using Eqs (5) and (7). The information fusion method is similar to the wave numerical model (WAVEWATCH Ⅲ), and the modeling also considers the driving influence of locally generated wind and wave information.

$$r_i = \sigma\left(W_{r_i}E_i + U_{r_i}X^d + b_{r_i}\right), \quad (4)$$

$$z_i = \sigma\left(W_{z_i}E_i + U_{z_i}X^d + b_{z_i}\right), \quad (5)$$

$$\widehat{h}_i = \tanh\left(W_{h_i}X^d + r_i \odot \left(U_{h_i}E_i + b_{h_i}\right)\right), \quad (6)$$

$$\widetilde{H}_i = (1 - z_i) \odot \widehat{h}_i + z_i \odot \widehat{h}_i, \quad (7)$$

where z_i and r_i are the update and reset gates, respectively; $\widehat{h}_i$ is the hidden state calculated based on the reset gate. $\sigma(\cdot)$ is a sigmoid activation function, $\odot$ denotes the element-wise multiplication operator, and $W_{(\cdot)}$, $U_{(\cdot)}$, $b_{(\cdot)}$ are learnable parameters.

However, considering only the long-term pattern and the fact that the dynamic input cannot fully utilize the information of the nodes, the internal correlation of the nodes is equally important (Zhang et al., 2023). Therefore, multi-head attention was introduced in this study to capture the correlations within nodes to compute short-term patterns.

(3) Short-term pattern learning

Node Embedding $\widetilde{H}_i \in \mathbb{R}^{B\times N\times dim}$ $(i = 1, 2)$ first performs layer normalizing to obtain H_i. The result $head^j_k$ of the j-th data of the k-th self-attention head is then computed using Eq. (8), followed by splicing of the results of the s self-attentions using Eq.

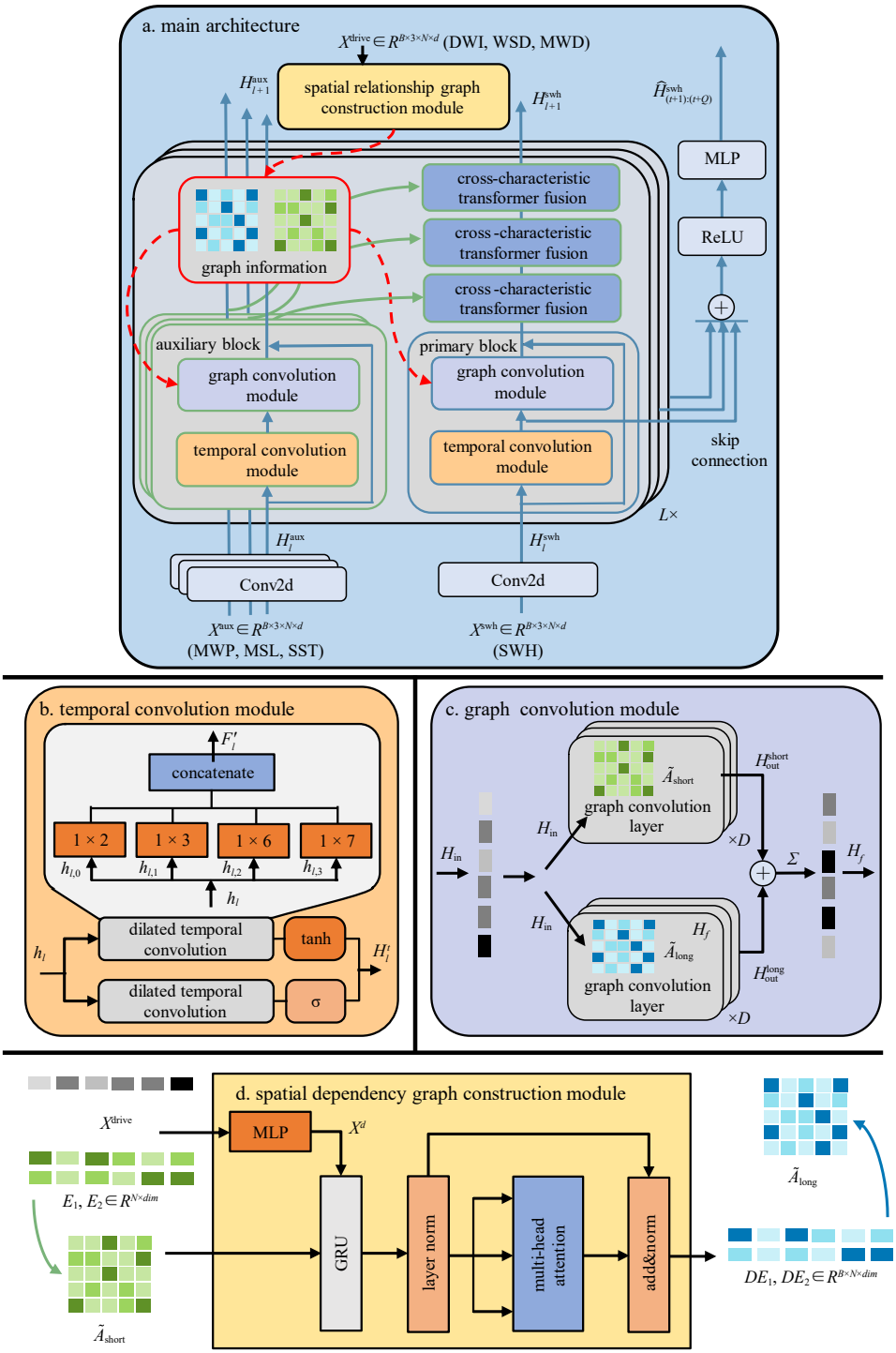

Fig. 2. Detailed STDG structure. a. Main STDG architecture, where the primary and three auxiliary modules run in parallel. b. An example of the temporal convolution module, which uses two dilated temporal convolutions with different dilation rates to capture multi-scale temporal features. c. A graph convolution module with depth $D = 2$, which contains two graph convolution layers that process the graph information of long- and short-term patterns, respectively. d. The detailed structure of the SDGC module, which fuses the dynamic input X^{drive} and the node embeddings of long-term patterns E_1, E_2, to obtain the final dynamic node embeddings DE_1, DE_2.

(9), which results in two embeddings OE_i $(i = 1, 2)$.

$$head_k^j = \text{softmax}\left(\frac{H_i^j W_k^Q \left(H_i^j W_k^K\right)^T}{\sqrt{d}}\right) H_i^j W_k^V, k = 1, \ldots, s, \quad (8)$$

$$OE_i[j, :, :] = \text{concat}\left(head_1^j, \ldots, head_s^j\right) \boldsymbol{W^O}, j = 1, \ldots, B, \quad (9)$$

where s is the number of self-attention heads, $W_s^Q, W_s^K, W_s^V \in \mathbb{R}^{d\times d}$ and the output matrix $\boldsymbol{W^O}$ is the learnable parameter matrix. Then, to prevent vanishing gradients, the two dynamic node embedding matrices are computed using Eq. (10), while the scaling

dot product operation is performed using Eq. (11). Among them, the ReLU activation function is used to eliminate weak connections in the weights, and the softmax activation function is used to normalize the weights.

$$DE_i = \mathrm{LN}\left(H_i + OE_i\right), \tag{10}$$

$$\widetilde{A}_{\mathrm{short}} = \mathrm{softmax}\left(\mathrm{ReLU}\left(\frac{DE_1(DE_2)^T}{\sqrt{dim}}\right)\right). \tag{11}$$

2.4 *Temporal convolution module*

Future significant wave heights have a certain correlation with historical data. To capture the temporal dependencies of nodes, this section introduces a method for processing wave data from a temporal perspective.

Considering training speed and computing resources, this study adopted a dilated causal convolution to extract high-level temporal features (Yu and Koltun, 2016), which can process long-distance sequences in parallel computing, thereby improving computing efficiency and alleviating the gradient explosion problem. Mathematically, given a node historical sequence input $h \in \mathbb{R}^T$ and a filter $f_{1\times k} \in \mathbb{R}^k$ of size k, the dilated causal convolution operation of h and $f_{1\times k}$ at the t-th step as follows:

$$h \star f_{1\times k}(t) = \sum_{s=0}^{K-1} h\left(t - a \times s\right) f_{1\times k}(s), \tag{12}$$

where a is the dilation factor used to control the jumping distance, and $\star$ is the dilated convolution operator.

To capture multiple temporal cycles in the historical data, such as 24 hours, 1 week, 12 months. This study uses dilated temporal convolution (Wu et al., 2020) to capture multi-scale temporal features in historical data. Different temporal patterns are discovered simultaneously by applying convolution different convolution filter sizes (that is, 1×2, 1×3, 1×6, 1×7) on the data using Eq. (13).

$$F_l' = \mathrm{concat}\left(h_{l,0} \star f_{1\times2}, h_{l,1} \star f_{1\times3}, h_{l,2} \star f_{1\times6}, h_{l,3} \star f_{1\times7}\right). \tag{13}$$

Simultaneously, the gating mechanism has been proven to effectively control information flow through the temporal convolution network layers (Dauphin et al., 2017). Therefore, the capacity of the model is improved by using Eq. (14) for temporal convolution.

$$H_l^t = \sigma\left(F_l'\right) \odot \tanh\left(F_l'\right). \tag{14}$$

2.5 *Graph convolution module*

To deal with spatial dependencies, this study used a graph convolution module to fuse the information of the nodes in the spatial dependencies with the information of their neighboring sites. The graph convolution module consists of two graph convolution layers that process the spatial dependencies of the short-term pattern $\widetilde{A}_{\mathrm{short}}$ and the long-term pattern $\widetilde{A}_{\mathrm{long}}$, as shown in Fig. 2b.

The spatial dependencies in this study are represented by the adjacency matrix of the directed graph. For the i-th node of the adjacency matrix, the row and column represent the information inflow and outflow, respectively. Therefore, two mix-hop propagation layers (Abu-El-Haija et al., 2019) were used as the graph convolution layer, which processed information inflow and outflow through each node. The structure of the mix-hop propagation layer is shown in Fig. 3a. It mainly performs deeper multi-node information fusion on the adjacency matrix. The mix-hop propagation layer is defined as follows:

$$H_k = \beta H_{\mathrm{in}} + (1-\beta) A_{\mathrm{adj}} H_{k-1}, \tag{15}$$

$$H = \mathrm{concat}(H_0, H_1, \ldots, H_K) W. \tag{16}$$

As shown in Fig. 3b, two mix-hop propagation layers process inflow and outflow information, respectively, and are finally added to obtain the output of the graph convolution layer. In addition, since the spatial dependence contains long-term and short-term patterns, two graph convolution layers are used in this study to perform information fusion respectively. As shown in Fig. 2c, the input features are calculated with the long-term pattern and the short-term pattern to obtain two outputs. As a result, the output of the graph convolution module is finally obtained by Eq. (17).

$$H_f = H_{\mathrm{out}}^{\mathrm{long}} + H_{\mathrm{out}}^{\mathrm{short}}. \tag{17}$$

2.6 *CCTF module*

Significant wave height is affected by multiple characteristics

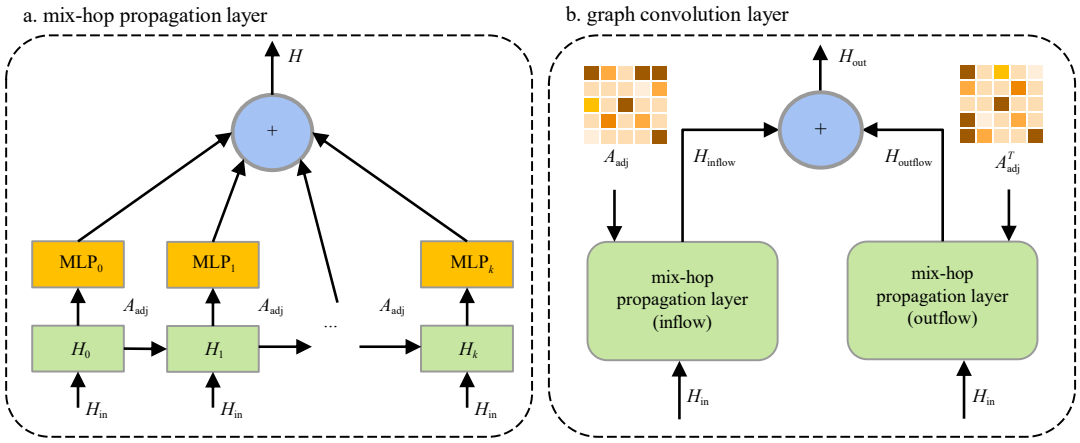

Fig. 3. Detailed graph convolution layer structure.

such as SST, air pressure, and wind speed. Considering multiple characteristics is crucial to improving the prediction accuracy of SWH. For example, some characteristics contain early indications of SWH, which can help improve the prediction accuracy of extreme sea conditions (Zhang et al., 2023). In some sea conditions, when the data quality of a single characteristic is low, multiple characteristics complement each other to improve the overall data quality and maintain a robust prediction effect (Luo et al., 2022). Therefore, this study proposes a CCTF module, as shown in Fig. 4, in which each CCTF learns the attention between the characteristics, thus updating the SWH sequence using auxiliary characteristics.

As shown in Fig. 4a, assume there are two characteristics a and b, represented as $X_a, X_b \in \mathbb{R}^{B\times N\times d}$, respectively, where B is the batch size of the data, and N represents the number of nodes. Inspired by a multi-modal transformer (Tsai et al., 2019), this study assumes that a good way to fuse multiple characteristics is to provide cross-characteristic latent adaptation, that is, b to a. The query is defined as $X_a W_a^Q$, key as $X_b W_b^K$, and values as $X_b W_b^V$, where $W_a^Q, W_b^K, W_b^V \in \mathbb{R}^{d\times d}$ are learnable parameters, the latent adaptation from b to a is expressed as cross-characteristic attention $\mathrm{CA}_{b\to a}(X_a, X_b) \in \mathbb{R}^{B\times N\times d}$:

$$\mathrm{CA}_{b\to a}(X_a, X_b) = \mathrm{softmax}\left(\frac{X_a W_a^Q \left(X_b W_b^K\right)^T}{\sqrt{d}}\right) X_b W_b^V. \tag{18}$$

Subsequently, based on cross-characteristic attention, this study designed a cross-characteristic transformer to enable the SWH to receive information from other characteristics, the structure of which is shown in Fig. 4b. Using SST transference to SWH as an example, denoted as "SST → SWH", each cross-characteristic transformer consists of K cross-characteristic attention blocks. The i-th cross-characteristic attention is

$$\begin{aligned}\widehat{X}^i_{\mathrm{SST}\to\mathrm{SWH}} =& \mathrm{CA}^i_{\mathrm{SST}\to\mathrm{SWH}}\left(\mathrm{LN}\left(X^{i-1}_{\mathrm{SST}\to\mathrm{SWH}}\right), \mathrm{LN}\left(X^0_{\mathrm{SST}\to\mathrm{SWH}}\right)\right) + \\ & \mathrm{LN}\left(X^{i-1}_{\mathrm{SST}\to\mathrm{SWH}}\right).\end{aligned} \tag{19}$$

Then, a feed-forward layer was applied, resulting in the output of the i-th cross-characteristic attention block $X^i_{\mathrm{SST}\to\mathrm{SWH}}$, as shown in Eq. (20):

$$X^i_{\mathrm{SST}\to\mathrm{SWH}} = \begin{cases} \mathrm{FF}\left(\mathrm{LN}\left(X^{i-1}_{\mathrm{SST}\to\mathrm{SWH}}\right)\right) + \mathrm{LN}\left(X^{i-1}_{\mathrm{SST}\to\mathrm{SWH}}\right), i \neq 0 \\ X^0_{\mathrm{SWH}}, i = 0 \end{cases}, \tag{20}$$

where $\mathrm{FF}(\cdot)$ denotes the feed-forward layer, and $\mathrm{CA}^i_{\mathrm{SST}\to\mathrm{SWH}}(\cdot)$ represents the multi-head self-attention of the i-th layer. As shown in Fig. 2c, the STDG model fuses multiple characteristics by concatenating multiple cross-characteristic transformers. During the multi-characteristic fusion process, the SWH continuously updates its own sequence with other characteristics from the multi-head cross-characteristic attention module. In each layer of the cross-characteristic attention module, low-level signals from the auxiliary characteristics are transformed into different key-value pairs to fuse with the SWH signals.

3 Experiment

3.1 *Study area and data*

The dataset used in this study was obtained from the ERA5 hourly data on single levels (Hersbach et al., 2023a) of the European Center for Medium-Range Weather Forecasts, which has provided hourly snapshots of atmospheric, land surface, and wave variables since 1940.

Although *in situ* data collected by buoys is also a source of SWH data, it has higher accuracy and better data quality. However, due to the high cost of equipment and the sparse distribution of collection buoys, data are not always available in most areas. Although the accuracy of the ERA5 reanalysis data is inaccurate, some studies (Wang et al., 2021; Wang and Wang, 2022) show that SWH in ERA5 data correlate well with *in-situ* data from multiple sites. In addition, the ERA5 data have the advantages of wide spatial-temporal coverage, multiple spatial-temporal resolution options, and rich characteristic types. This dataset has been widely used in ocean research in recent years (Ding et al., 2023; Zhang et al., 2023; Wang et al., 2018, 2022).

Considering the computational resources and running time, this study selected multiple nodes and characteristics from different geographical regions (the South China Sea and East China Sea) for research. The time resolution was six hours, and the time range was from 0:00 on January 1, 2003, to 0:00 on August 14,

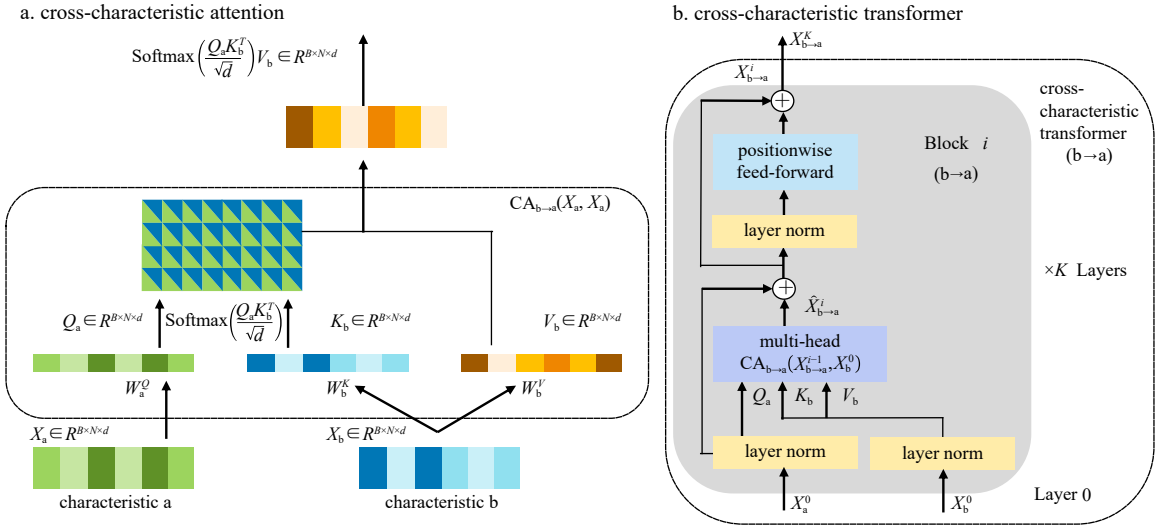

Fig. 4. Cross-characteristic attention and transformer.

2023. The dataset is described in detail as follows. It should be noted that the timestamps of the South China Sea and East China Sea datasets are consistent, and both datasets contain 30 120 timestamps.

The sea area ranges for the two datasets are shown in Fig. 5. In the SCS observation area, 91 points were selected from 11°–17°N and 111°–117°E, with a spatial resolution of 1° × 0.5°. In the East China Sea observation area, 90 nodes were selected from 25°–33°N and 123°–127.5°E, with a spatial resolution of 1° × 0.5°.

According to previous studies (Hashim et al., 2016; Li et al., 2022; Luo et al., 2022), wave height, wave period, WSD, DWI, pressure, and SST play dominant roles in SWH prediction. Therefore, the following seven characteristics were selected from the ERA5 reanalysis data: SWH, MWP, SST, MSL, 10-m wind U component, 10-m wind V component, and MWD.

The 10-m wind U component is the component along the latitudinal direction, and the 10-m wind V component is the component along the longitudinal direction, which cannot directly reflect DWI and WSD (Wang et al., 2018). Therefore, according to Eqs (21) and (22), we recalculated two new characteristics, DWI and WSD, replacing them with U_{10} and V_{10}:

$$\mathrm{DWI} = \mathrm{mod}\left(180 + \frac{180}{\pi}\mathrm{arctan2}\left(U_{10}, V_{10}\right), 360\right), \tag{21}$$

$$\mathrm{WSD} = \sqrt{U_{10}^2 + V_{10}^2}, \tag{22}$$

where mod $(\cdot)$ represents the modulo operation, and arctan2 $(\cdot)$ is the function to calculate the azimuth angle.

Finally, this study classified the above seven characteristics into the following three categories according to the STDG model input: (1) the main characteristic, SWH, which is the target variable to be predicted; (2) auxiliary characteristics, including MWP, SST, and MSL, which affect wave characteristics and changes; and (3) driving characteristics, including WSD, DWI, and MWD, which determine wave formation and propagation.

3.2 *Data preprocessing*

Different characteristics have different standards and magnitudes; to transform multiple characteristics into a unified scale and eliminate the influence of units and scale differences, this study adopted the Z-score standardization method to preprocess the data, which was defined as follows:

$$x' = \frac{x - \mu}{\sigma}, \tag{23}$$

where x' is the standardized characteristic, x is the original characteristic, and μ and σ are the mean and standard deviation of the original characteristic, respectively.

To achieve multistep prediction, this study used a sliding window to select the data. Specifically, given the multi-node multi-characteristic ocean data $X \in \mathbb{R}^{T\times N\times M}$, where T is the total number of time steps, N is the number of nodes, and M is the number of characteristics, assuming that the SWH in the next Q hours is predicted based on the multi-characteristic wave data in the previous P hours, the input $w_1 \in \mathbb{R}^{P\times N\times M}$ and output windows $w_2 \in \mathbb{R}^{Q\times N\times 1}$ were used to determine model input and output data, and these two windows were cut along the time dimension step by step, thus obtaining the STDG model input $X_{\mathrm{in}} \in \mathbb{R}^{T'\times P\times N\times M}$ and actual output sets $Y_{\mathrm{out}} \in \mathbb{R}^{T'\times Q\times N\times 1}$, where T' is the number of samples after cutting.

This study divided the dataset into three categories according to the number of samples T': the first 70% of the dataset was used as training data, the next 20% was used for validation, and the remaining 10% was used as test data. In particular, the original datasets of the South China Sea and the East China Sea encompass 30 120 timestamps, respectively. Subsequently, the original dataset was partitioned by the sliding window, resulting in a division into 30 081 samples. Each sample comprises input data with a window size of 28 and output data with a window size of 12. The data samples were classified into three categories: a training set with 21 057 samples, a validation set with 6 016 samples, and a

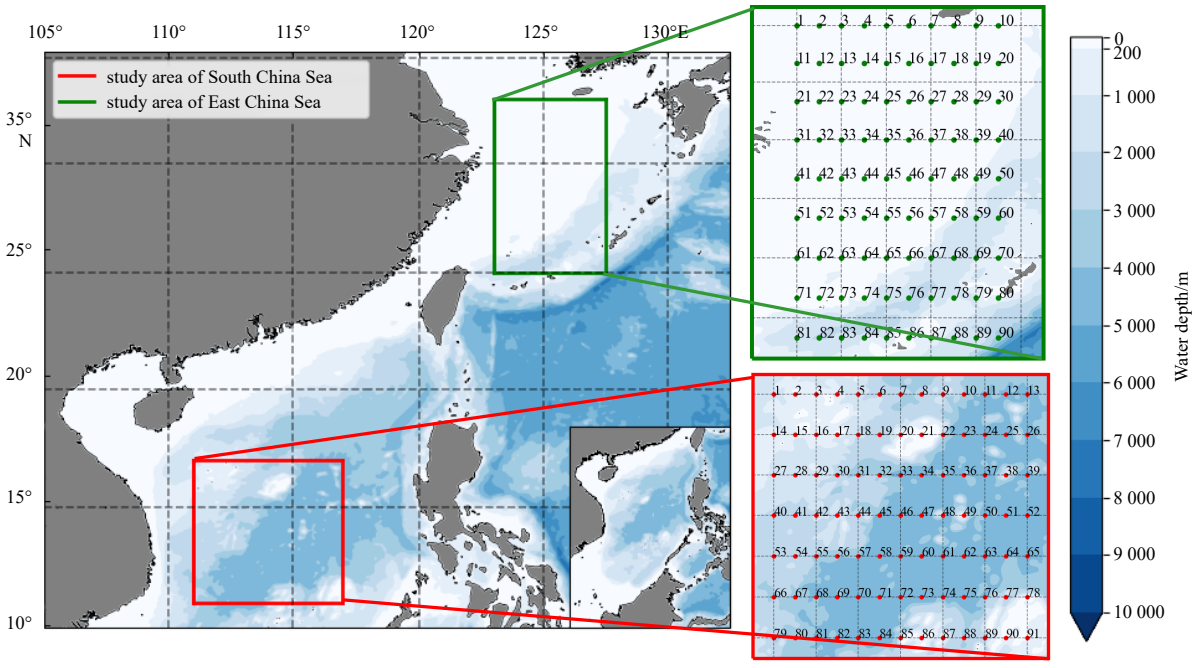

Fig. 5. The sea area range of the two datasets.

test set with 3 008 samples.

3.3 *Model hyperparameter setting and statistical metrics*

The proposed method was implemented using Python 3.8 and Pytorch 1.10.0 and successfully executed and tested on a Windows 10 platform equipped with a 12th Gen Intel(R) Core (TM) i5-12400F CPU@2.50 GHz, an NVIDIA Geforce RTX 3080 (PCI-E) 10 GB GPU card, and 16 GB DDR4 memory.

This study used mean absolute error (MAE) as the loss function and trained the STDG model using the Adam optimizer (Kingma and Ba, 2015). The hyperparameters and best model were determined based on the performance of the validation set. In the SDGC module, we set the node embedding dimension *dim* to 40, the dropout to 0.1, and the number of heads of the multi-head self-attention to 4. For the temporal convolution module, we set the dilation factor to 1. For the graph convolution module, we set the dropout to 0.3, the depth D of the graph convolution to 2, and the retention coefficient of the initial information in each graph convolution operation to 0.05. For the CCTF module, we set the number of cross-characteristic attention blocks K to 4, batch size to 64, learning rate to 0.000 5, and training epochs to 500, and stopped training when the validation loss did not decrease for 30 consecutive epochs to prevent model overfitting. This study used 28 historical time steps to predict SWH for the following 12-time steps.

As no single evaluation indicator can reflect all aspects of model prediction (Luo et al., 2022), this study used three different indicators to evaluate the model. These are the root mean square error (RMSE), MAE, and coefficient of determination (R-square, R^2). Their definitions are

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2}, \tag{24}$$

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\widehat{y}_i\right|, \tag{25}$$

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\overline{y}_i)^2}, \tag{26}$$

where n is the data length, y_i is the i-th sample actual value, $\widehat{y}_i$ is the i-th sample predicted value, and $\overline{y}_i$ is the mean of the observed values, the RMSE is very sensitive to outliers in the data and can be used to comprehensively evaluate the prediction process. The MAE is the average of the absolute errors between the actual and predicted values, which can reflect the overall prediction performance. The smaller the RMSE and MAE values, the higher is the model prediction accuracy (Geng et al., 2021), the R^2 describes the proportion of the variance of the dependent variable to the total variance explained by the model. The larger the R^2, the stronger is the correlation between the independent and dependent variables, and the better the model performance.

4 Results and discussion

4.1 *Experimental results and analysis*

Considering that the use of numerical models for wave prediction is time-consuming and computationally expensive, this study compared the performance of STDG with mainstream deep learning methods (LSTM, GRU, convolutional long-term time series network (CLTS-Net), multivariate time series forecasting with graph neural network (MTGNN), and MCMN) in multi-node SWH prediction under different prediction scales. These methods are centered on RNN, convolutional neural network (CNN), and GNN, covering three dimensions of deep learning technology, as listed in Table 1. This study will further compare and analyze these deep learning methods through the comparison of these three dimensions. The objective is to illustrate the importance of considering temporal dependencies, spatial dependencies, and explicit modeling in the task of SWH prediction.

The specific descriptions of these methods are as follows:

LSTM (Pirhooshyaran and Snyder, 2020): This method is based on an LSTM network and a sequence-to-sequence framework and performs well in SWH prediction.

GRU (Li et al., 2022): This method is based on a GRU that learns the long-term pattern between multivariate sequence data and achieves robust short- and long-term wave height forecasts.

CLTS-Net (Li et al., 2021): This method uses a CNN to discover the local dependency patterns between multidimensional input variables and uses an LSTM layer and skip connections to capture complex long-term patterns. Finally, it uses a parallel combination of a traditional autoregressive linear model and a nonlinear layer to obtain more robust prediction results.

MTGNN (Wu et al., 2020): This method adapts to learn the spatial dependency relationship of multiple nodes by constructing a static graph of long-term patterns, and then aggregates features using temporal and graph convolutions.

MCMN (Zhang et al., 2023): This method proposes a spatiotemporal GNN, that performs multi-characteristic fusion according to the lag effect between ocean characteristics. In addition, the model learns the spatial dependency of wave data by fusing the long- and short-term patterns between nodes.

In this study, experiments were conducted using STDG and five comparative methods. First, to verify the generalization and robustness of the STDG method, this study uses different sea area datasets (South China Sea and East China Sea) to train, verify, and test the method. Secondly, this study continuously predicts the SWH for 6 h, 12 h, 24 h (1 d), 48 h (2 d), and 72 h (3 d) on the test datasets of these two sea areas. The results are shown in Tables 2 and 3. The boldface in the table indicates the best prediction and the underline indicates the second-best prediction.

The comparison of Table 2 and Table 3 shows that, firstly, these six methods have lower MAE and RMSE and higher R^2 on the South China Sea dataset than on the East China Sea dataset, indicating that all the methods have better performance on the South China Sea dataset. This may be due to differences in

Table 1. Comparison method classification

Classification	Method	Specific description
RNN	LSTM, GRU	These methods model temporal dependencies.
CNN	CLTS-Net	This method models temporal and spatial dependencies.
GNN	MTGNN, MCMN	These methods explicitly model multiple node spatial dependencies.

Table 2. The average prediction results of forecasting methods for 91 nodes in the South China Sea

Time step	Metrics	LSTM	GRU	CLTS-Net	MTGNN	MCMN	STDG
6 h	MAE	0.096 3	0.094 3	0.157 3	0.094 0	0.093 5	0.090 4
	RMSE	0.159 9	0.152 4	0.221 7	0.150 2	0.149 8	0.143 1
	R^2	0.972 8	0.975 3	0.947 7	0.976 0	0.976 9	0.978 2
12 h	MAE	0.173 1	0.171 8	0.207 5	0.145 8	0.143 0	0.139 3
	RMSE	0.279 0	0.272 9	0.306 3	0.233 5	0.230 2	0.221 9
	R^2	0.917 2	0.920 8	0.900 2	0.942 0	0.945 3	0.947 6
18 h	MAE	0.239 4	0.238 2	0.265 0	0.193 6	0.190 9	0.182 5
	RMSE	0.381 9	0.375 8	0.398 0	0.312 8	0.312 6	0.291 5
	R^2	0.844 9	0.849 7	0.831 5	0.895 9	0.896 1	0.909 6
1 d	MAE	0.295 4	0.294 8	0.315 4	0.239 1	0.231 7	0.220 1
	RMSE	0.465 6	0.461 1	0.478 7	0.385 7	0.364 2	0.352 5
	R^2	0.769 3	0.773 8	0.756 2	0.841 7	0.855 6	0.867 8
2 d	MAE	0.472 5	0.471 6	0.478 1	0.422 3	0.417 3	0.381 3
	RMSE	0.708 8	0.704 5	0.712 3	0.646 2	0.632 3	0.583 3
	R^2	0.465 4	0.471 8	0.460 1	0.555 7	0.577 4	0.637 9
3 d	MAE	0.561 9	0.559 0	0.562 4	0.537 9	0.526 1	0.502 4
	RMSE	0.809 2	0.802 8	0.804 3	0.783 1	0.776 9	0.731 0
	R^2	0.302 9	0.313 8	0.311 2	0.347 1	0.350 7	0.431 0

Table 3. The average prediction results of forecasting methods for 90 nodes in the East China Sea

Time step	Metrics	LSTM	GRU	CLTS-Net	MTGNN	MCMN	STDG
6 h	MAE	0.153 8	0.154 5	0.173 1	0.142 5	0.142 2	0.141 8
	RMSE	0.266 6	0.268 6	0.295 9	0.231 0	0.229 7	0.226 8
	R^2	0.915 0	0.913 7	0.895 3	0.936 2	0.937 3	0.938 5
12 h	MAE	0.272 0	0.271 2	0.284 8	0.228 8	0.228 5	0.228 0
	RMSE	0.457 8	0.459 4	0.473 5	0.375 3	0.373 2	0.363 4
	R^2	0.749 3	0.747 5	0.731 8	0.831 5	0.836 1	0.842 0
18 h	MAE	0.368 4	0.367 2	0.378 4	0.308 8	0.307 6	0.299 3
	RMSE	0.602 7	0.605 4	0.617 6	0.507 5	0.505 6	0.480 9
	R^2	0.565 5	0.561 6	0.543 6	0.691 8	0.694 2	0.723 4
1 d	MAE	0.436 4	0.435 1	0.446 1	0.373 5	0.369 5	0.350 5
	RMSE	0.702 5	0.703 3	0.720 2	0.610 1	0.605 9	0.573 1
	R^2	0.409 7	0.408 3	0.379 4	0.554 7	0.560 8	0.607 1
2 d	MAE	0.562 6	0.562 7	0.573 2	0.517 8	0.516 8	0.498 9
	RMSE	0.874 0	0.869 3	0.886 0	0.806 7	0.794 1	0.763 9
	R^2	0.086 0	0.095 9	0.060 8	0.221 4	0.245 5	0.318 6
3 d	MAE	0.587 8	0.587 5	0.597 9	0.559 7	0.559 7	0.551 9
	RMSE	0.917 3	0.908 2	0.910 0	0.866 8	0.866 3	0.845 6
	R^2	0.006 9	0.013 0	0.009 1	0.101 0	0.102 0	0.144 5

oceanography, climate, and geographical conditions between the two datasets. It can also be seen from Fig. 5 that, in addition to the different longitude and latitude of the two selected sea areas, the South China Sea is deeper than the East China Sea in most areas. These different factors may make the ocean dynamic process of the South China Sea more complicated. For the model, a good data distribution is easier for the model to learn, and a data set with more noise is more likely to affect the prediction accuracy of the model, which leads to the results of the South China Sea being better than those of the East China Sea.

Secondly, these six methods have achieved high prediction accuracy in the short-term prediction of both data sets, and as the prediction step size increases, the change trend of the prediction results is consistent. However, the difference is that STDG has lower MAE and RMSE and higher *R*-square compared to other methods. Therefore, in terms of capturing the spatial-temporal characteristics of short-term waves and making long-term predictions, the prediction performance and generalization of STDG are better than those of other methods.

Figure 6 intuitively compares the prediction performances of different methods, where the horizontal axis is the prediction duration and the vertical axis is the error value. Different curve colors represent different methods. It can be seen from Fig. 6 that STDG performs well over different time ranges.

The STDG builds a spatial dependency graph to fuse dynamically changing input data, which can effectively capture the spatial dependency between multiple nodes. CLTS-Net, LSTM, and GRU do not consider the spatial dependency between multiple nodes; therefore, their prediction performances are inferior to those of the MTGNN, MCMN, and STDG. The MCMN and MTGNN employ graphs to describe the information propagation direction between multiple nodes over time. Compared with other methods, they achieved better results. However, the spatial dependence modeled by the MTGNN is fixed and cannot flexibly handle dynamically changing spatial dependencies. The MCMN improves performance significantly by fusing multiple characteristics using the time lag between them and by constructing a dynamic graph using a self-attention mechanism. However, compared with the STDG, it still has some limitations because it does not construct a dynamic graph from the wind-wave relationship but directly uses the fused characteristics to construct a dynamic graph, which reduces model interpretability. In addition, although the accuracy of ERA5 data is not very high as reanalysis data, the STDG model still shows good prediction performance on this dataset by considering multi-characteristic fusion and dynamic graph modeling.

4.2 *Visual comparison of prediction results*

To visually compare the prediction results, this study selected GRU and MCMN, two excellent comparative methods, as representatives and visualized the prediction results on the test set of the South China Sea and East China Sea data together with

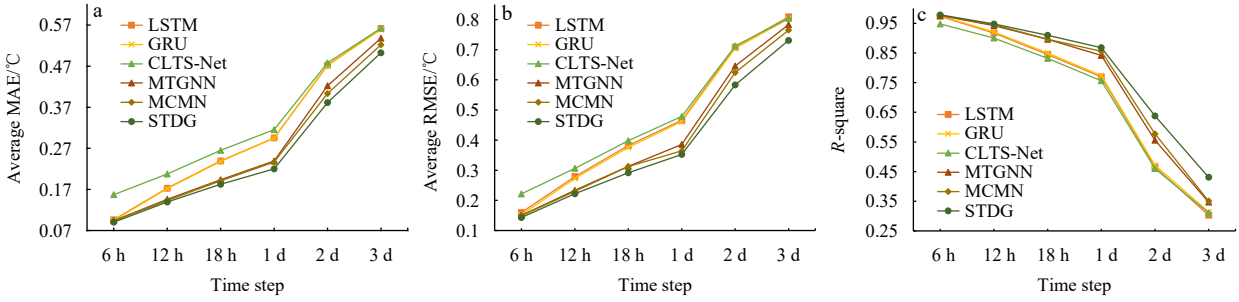

Fig. 6. The average MAE, RMSE, and *R*-square of the prediction methods in the South China Sea.

STDG.

4.2.1 *Prediction effect of SWH in the sea area*

This study conducted 6-hour multi-node predictions on three methods on test data sets in different sea areas. The prediction results are shown in Figs 7 and 8.

It can be seen from Figs 7 and 8 that (1) as time changes, the SWH distribution in the same area is different, and the three

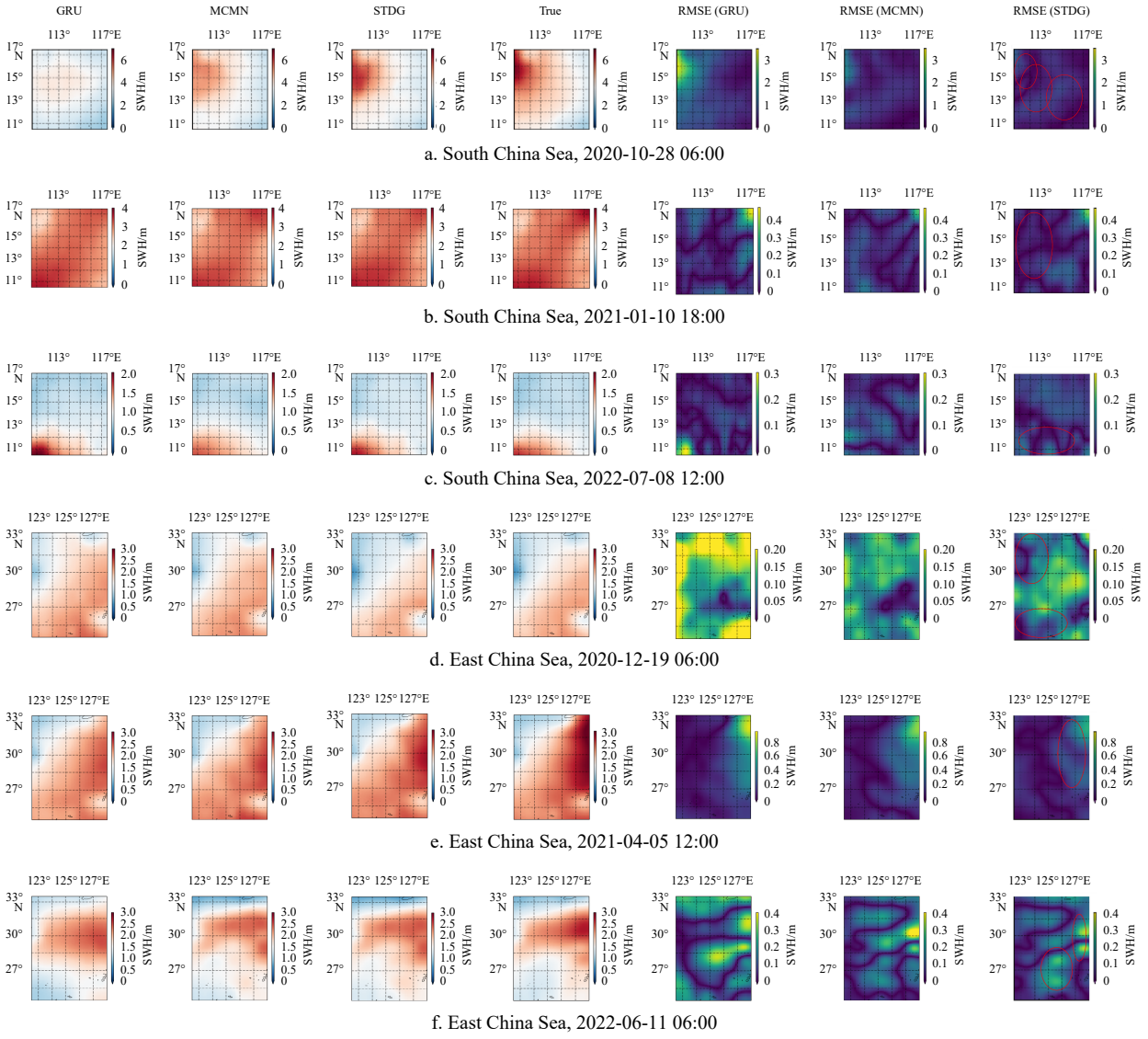

Fig. 7. Predicted values, true values, and RMSE for three methods.

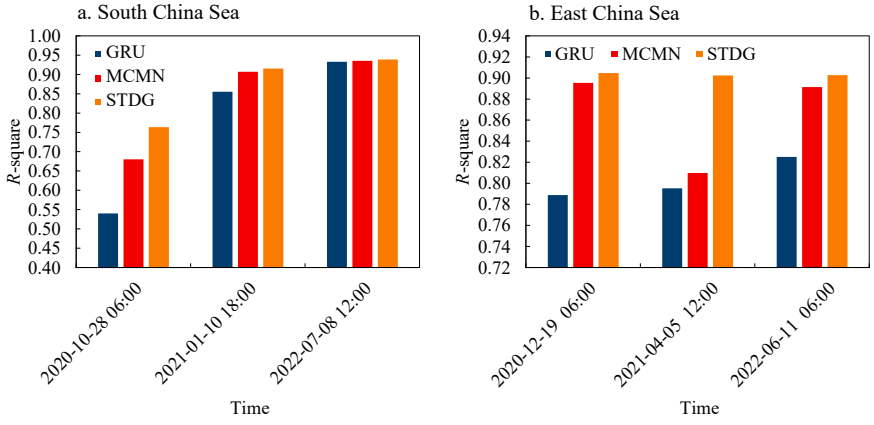

Fig. 8. Coefficient of determination of three methods.

methods can effectively capture the dynamic changes in the SWH distribution. (2) Compared with GRU, the RMSE of MCMN and STDG is lower, which illustrates the effectiveness of considering spatial dependencies. (3) For large-value areas of SWH, the RMSE of GRU and MCMN is higher than that of STDG, which shows that the prediction capabilities of GRU and MCMN are limited for large-value areas of SWH. (4) In different sea areas, the coefficient of determination of STDG is higher than that of MCMN and GRU, indicating that the correlation between the predicted value of STDG and the true value is higher. In summary, the predictive performance of STDG is better than other methods.

4.2.2 *Prediction results of SWH at a single node*

This study compared the prediction effects of the three methods on a single node with the actual values. Figure 9 shows the

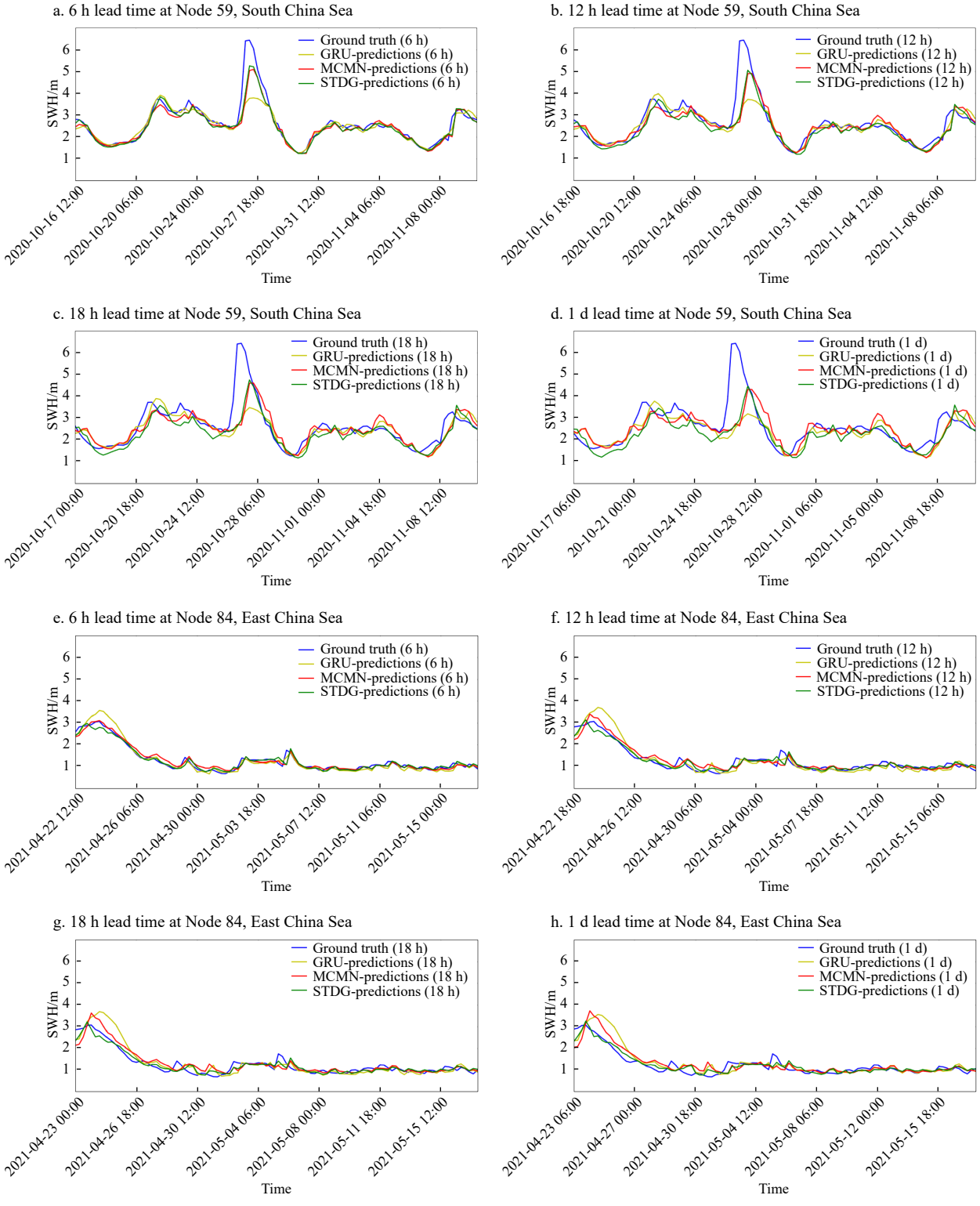

Fig. 9. Significant wave height prediction effect of single node.

prediction results for 6 h, 12 h, 18 h, and 1 d, where the horizontal axis is the time of the test sample, the vertical axis is the corresponding SWH, the blue curve represents the actual SWH, and the yellow, red, and green curves represent the SWH prediction values obtained by GRU, MCMN, and STDG, respectively.

It can be seen from Fig. 9: (1) In different sea areas, the SWH oscillation amplitude of Node 59 in the South China Sea is high, while the SWH of node 84 in the East China Sea is relatively stable. All methods tend to give more robust predictions in the stationary series. In the 6 to 18 h forecast, MCMN and STDG predict extreme values better than GRU, showing that the consideration of spatial dependency positively affects the prediction of extreme values. In the 1 d forecast, the forecast results of all methods deviate from the actual results due to the increase of the forecast time and the uncertainty of the wave motion, but when dealing with the decreasing phase of the curve, the forecast results of STDG are more accurate. (2) In the same sea area, the longer the forecast lead time, the lower the forecast accuracy of each model. Although all methods showed varying degrees of offset, the performance of GRU and MCMN was most affected, while STDG was able to maintain relatively accurate forecasts within 18 h and remained more robust over longer forecast periods. (3) In short, STDG can provide more robust forecasts within the 12 h forecast range, is more effective at predicting extreme values, and can also show some advantages at longer forecast times.

4.3 *Ablation experiment*

To evaluate the effectiveness of each component module of STDG, an ablation study was conducted on the two test datasets, and three variants were designed: STDG w/o SDGC, STDG w/o dynamic graph convolution (DGC), and STDG w/o CCTF.

(1) STDG w/o SDGC: This variant removed the SDGC module and retained only the temporal convolution module to verify the importance of considering spatial dependencies.

(2) STDG w/o DGC: This variant removed the DGC of the short-term pattern and retained only the graph convolution of the long-term pattern to verify the importance of considering the short-term pattern.

(3) STDG w/o CCTF: This variant removed the CCTF and processed only the SWH to verify the importance of fusing auxiliary characteristics.

The results of the ablation experiment are listed in Table 4,

Table 4. The prediction results of the ablation study on the South China Sea dataset

Time step	Metrics	STDG	STDG w/o CCTF	STDG w/o DGC	STDG w/o SDGC
6 h	MAE	0.090 4	0.096 2	0.093 0	0.093 3
	RMSE	0.143 1	0.147 6	0.146 9	0.146 1
	R^2	0.978 2	0.976 8	0.977 0	0.977 3
12 h	MAE	0.139 3	0.146 9	0.141 9	0.146 4
	RMSE	0.221 9	0.222 7	0.222 2	0.228 2
	R^2	0.947 6	0.947 2	0.947 0	0.944 6
1 d	MAE	0.220 1	0.229 7	0.229 4	0.237 8
	RMSE	0.352 5	0.352 5	0.357 3	0.369 3
	R^2	0.867 8	0.867 8	0.864 2	0.854 9
2 d	MAE	0.381 3	0.391 5	0.392 3	0.417 1
	RMSE	0.583 3	0.588 1	0.594 1	0.620 6
	R^2	0.637 9	0.632 0	0.624 4	0.590 2
3 d	MAE	0.502 4	0.509 0	0.511 1	0.534 0
	RMSE	0.731 0	0.736 0	0.746 5	0.759 7
	R^2	0.431 0	0.423 2	0.406 6	0.385 5

which shows that prediction accuracy decreased when each STDG module was removed. The STDG results in the East China Sea also verify the effectiveness of the proposed module, as the experimental results of the ablation experiments in the East China Sea dataset are similar to those in the South China Sea. Limited by the length of the article, this study shows only the South China Sea experimental results.

4.4 *Discussion*

4.4.1 *Predicted results for different wave levels and seasons*

To evaluate the prediction effect of STDG, this study visualized the STDG 6 h future forecast results for different wave levels and different seasons, as shown in Figs 10 and 11. It is worth noting that this study divides waves into three levels based on the maximum SWH that occurs in the sea area and refers to the Douglas wave classification scale (Owens, 1982): small waves (below 2 m), medium waves (2–4 m), and large waves (above 4 m). In addition, this study also analyzes the prediction performance of STDG from a seasonal perspective based on the seasonal variations of the ocean (Ti et al., 2022). Among them, the seasons corresponding to each row of subplots in Figs 10 and 11 are "winter" "spring" "summer" and "autumn" respectively.

It can be seen from the different seasonal perspectives in Fig. 10 that (1) in the South China Sea, the SWH maximum from spring to early summer (Figs 10c–e) is relatively small throughout the season, and STDG has a low RMSE; therefore, STDG has the best prediction performance at this time. (2) In summer and autumn (Figs 10f and g), due to the influence of monsoon changes, the wind and wave directions in the central part of the South China Sea are relatively disordered. The RMSE of the STDG in the large value range is less than 0.4 m and remains low in most areas. (3) In early winter (Fig. 10a) and late autumn (Fig. 10h), large waves are common in the South China Sea due to the influence of typhoons. The RMSE of STDG in the large value range is between 0.4–0.7 m, but STDG still has a low RMSE in most areas. In short, STDG can make reasonable predictions with seasonal variations, has high accuracy and stability, and can effectively predict SWH in different seasons.

In the wave predictions of different levels in Fig. 10, the prediction results of STDG have achieved high accuracy. The RMSE of the medium and small waves (Figs 10b–f) is less than 0.3 m, while in the large wave areas (Fig. 10a and Figs 10g–h), the RMSE is less than 0.3 m in most areas, but there are also obvious overestimations or underestimations in individual locations, with the RMSE reaching more than 0.4 m, which may be caused by unstable noise in the predictions. Since STDG uses past wind directions and wave characteristics to make predictions, a reasonable explanation is that the basic wind-wave relationship at these locations is weak, and waves are affected not only by wind and season, but also by local water depth and typhoons. Despite this, STDG can still provide robust and good predictions of SWH for most of the South China Sea.

In Fig. 11, similar conclusions can be obtained from the analysis of different seasons and different levels: (1) In the East China Sea, the maximum value of SWH from winter to early summer (Figs 11a–e) is lower than that at other times, with SWH below 3 m. The wave levels are medium and small waves, and the RMSE of STDG remains below 0.5 m, indicating that STDG has the best prediction performance at this time. (2) From mid-summer to the entire autumn (Figs 11f–h), with the appearance of large waves, the prediction performance of STDG decreases slightly, but it can stay below 0.8 m. Therefore, STDG can provide robust and good

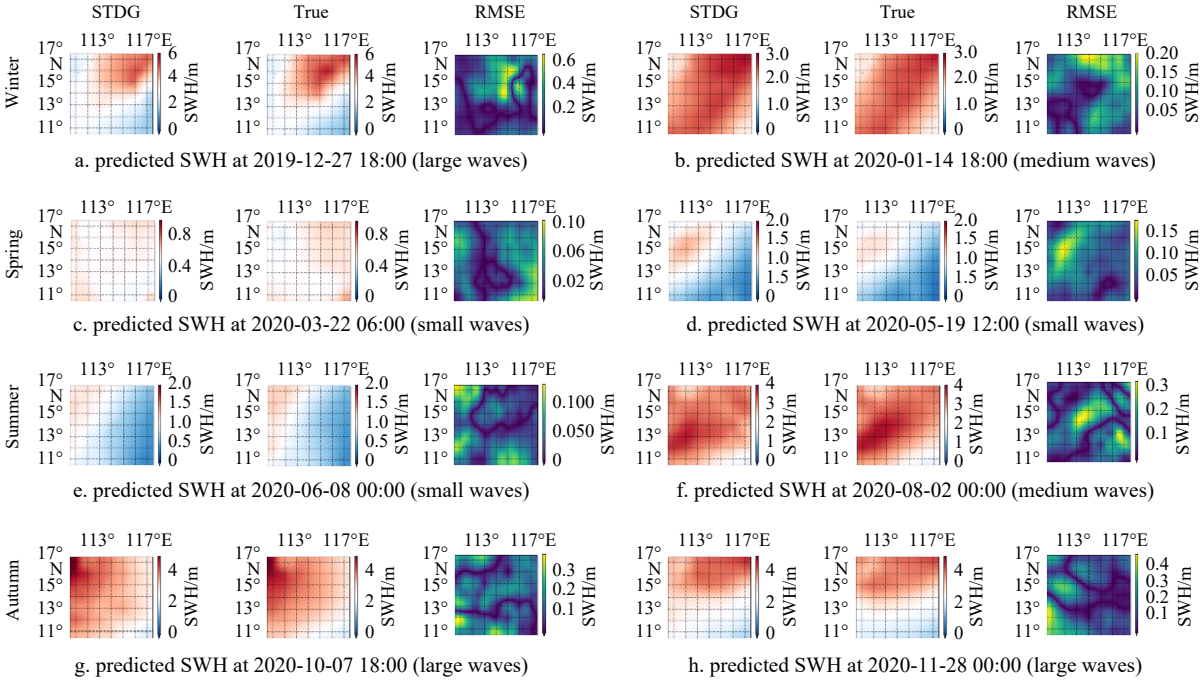

Fig. 10. Predicted value, true value, and RMSE of STDG in different seasons and levels in the South China Sea.

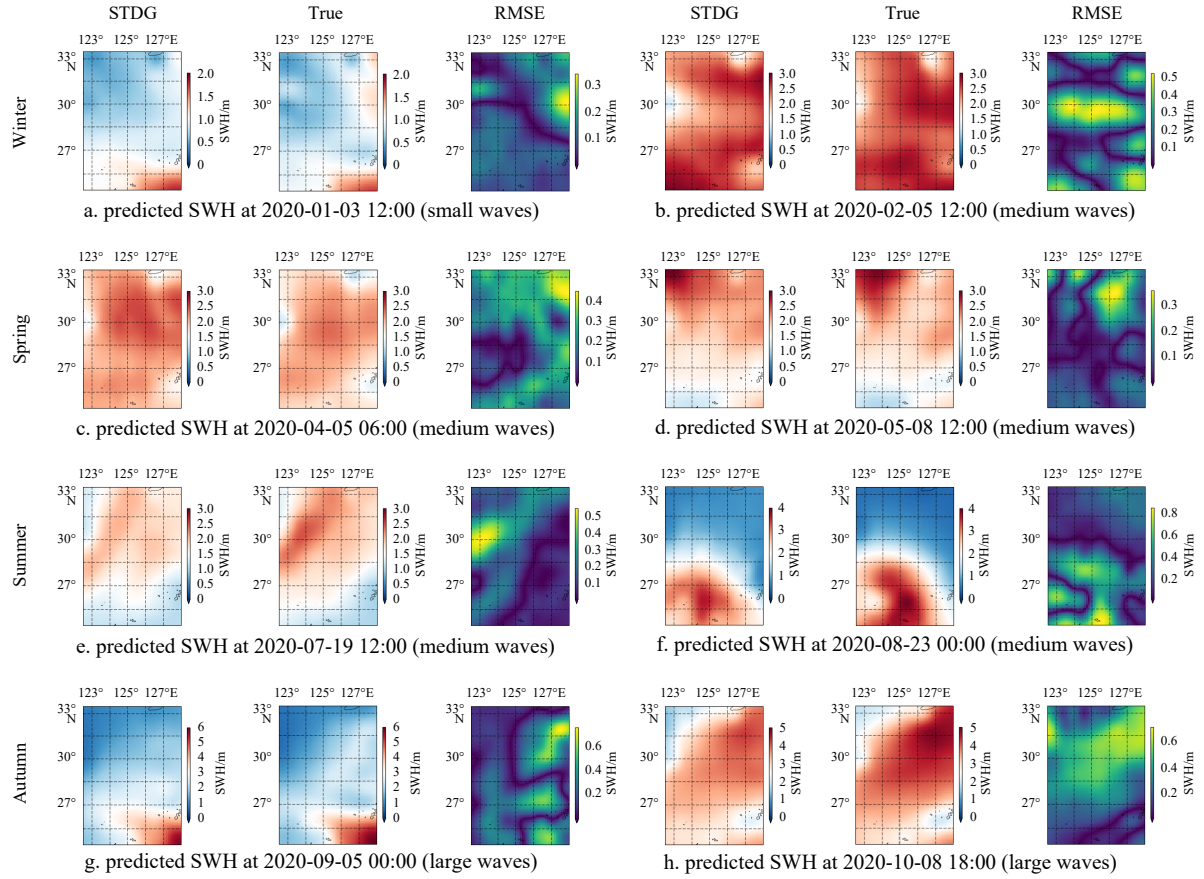

Fig. 11. Predicted value, true value, and RMSE of STDG in different seasons and levels in the East China Sea.

predictions of SWH in most areas of the East China Sea.

Considering that large waves are more harmful, this study further analyzed the prediction results of large waves. Since large waves appear randomly in different locations and at different times, this study first screened out the prediction results of large waves from the test data of the two sea areas (the time range was from June 2019 to August 2023) and counted the true value and predicted value of large waves at each node. On this basis, this

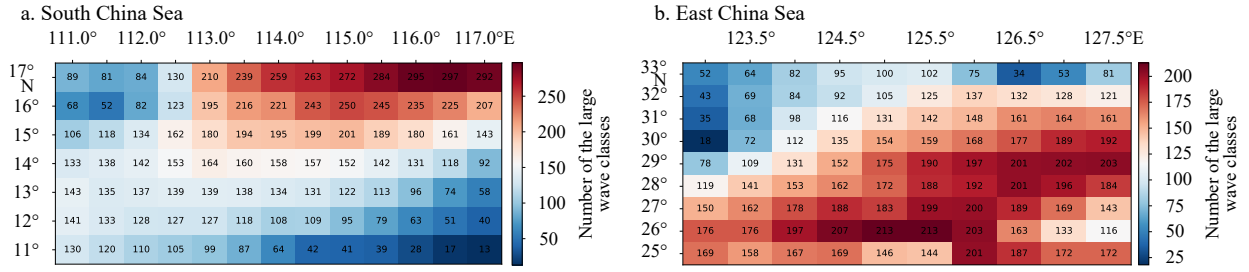

Fig. 12. Distribution of the number of large wave classes at different nodes in the test data.

study calculated the frequency of large waves at each node, as shown in Fig. 12, and calculated the MAE, RMSE, and *R*-square of large waves in the test data of STDG, as shown in Fig. 13. Each square represents a node in the sea area, and the coordinate axis represents the longitude and latitude of each node. The lower the MAE and RMSE, the more accurate the model. It is important to note that the *R*-square can take on negative values, ranging from negative infinity to zero. Negative values indicate that the model under consideration is less effective than a simple average, whereas positive values indicate that the model under consideration is more effective than a simple average. It is important to note that due to the inconsistency in the number and timestamps of the large waves recorded at each node, the figures (Figs 12 and 13) only reflect the geographical relationship between the nodes and are unable to provide any information regarding node associations.

As shown in Fig. 12, the frequency of large waves in the test data set is relatively low in the northwest corner and southeast part of the South China Sea, the northwest of the East China Sea, and its northern boundary. In contrast, the frequency of large waves is high in other areas. Concomitantly, the frequency of large waves is reflected in the evaluation indicators. As shown in Fig. 13, in regions with a relatively high frequency of large waves, the MAE and RMSE of the nodes are lower, and the *R*-square is higher, indicating that the STDG model can predict the occurrence of large waves with greater reliability for the majority of nodes. Nevertheless, for certain nodes with a low frequency of large waves, the predictive performance of the STDG model was suboptimal.

Many factors affect the prediction performance of nodes. First, as shown in Fig. 5, nodes with poor prediction effects are usually situated in the boundaries of the sea area and are relatively close to the land. The STDG model is based on graph structure modeling, which is insufficient for capturing the spatiotem-

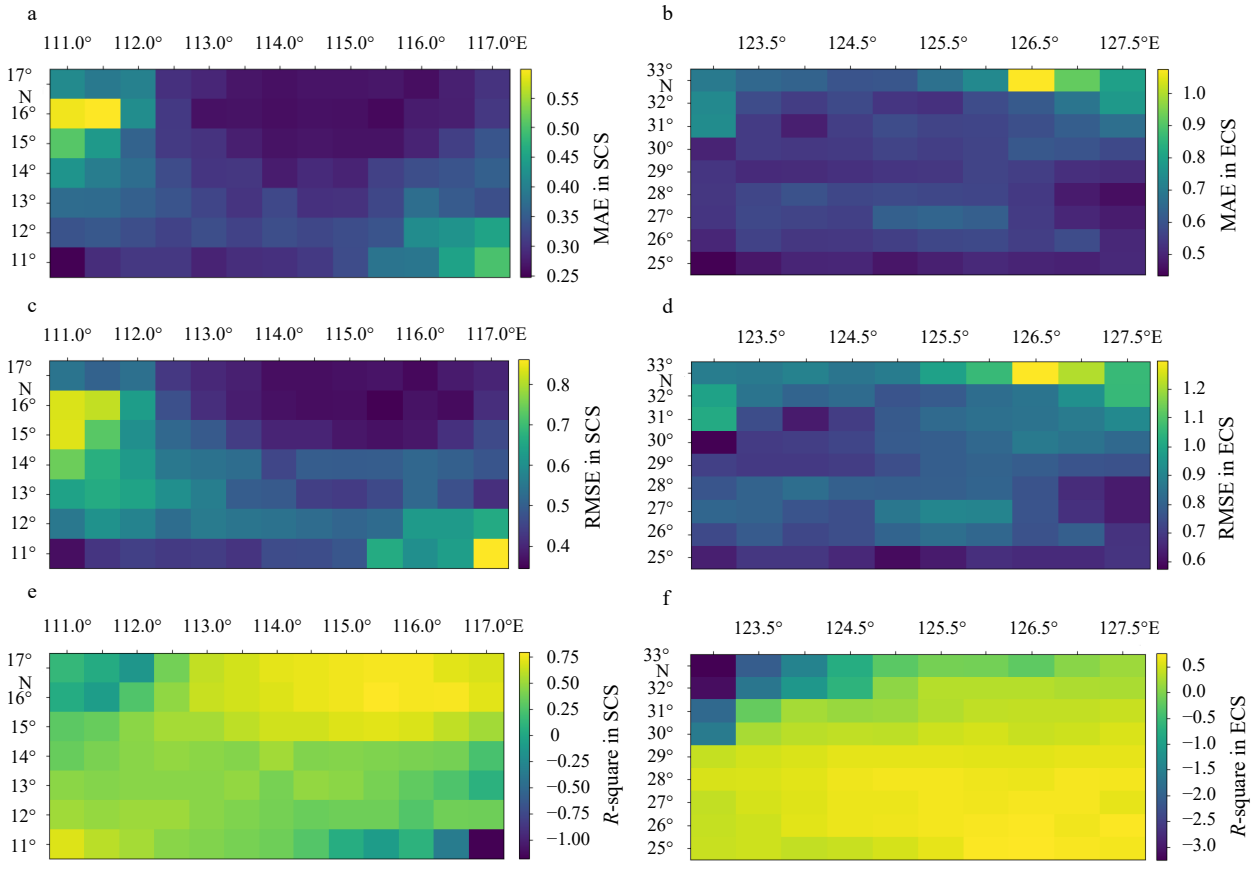

Fig. 13. The MAE, RMSE, and *R*-square of STDG model for large wave prediction in the South China Sea (SCS) and East China Sea (ECS) test data.

poral features of these boundary nodes. Secondly, the uncertainty and low frequency of large waves make the STDG model insensitive to the occurrence of large waves, thereby limiting its prediction capabilities. Consequently, in future studies, it is recommended to introduce more spatiotemporal features and environmental factors to enhance the sensitivity of the model to large waves, thereby improving the prediction performance of boundary nodes.

4.4.2 *The dynamic information propagation mechanism of waves*

To analyze the dynamic information propagation mechanism of waves in multiple nodes in STDG, this study selected multi-characteristic data in each month of 2022 from the test dataset of East China Sea and South China Sea, input it into the trained model, and thus constructed the graph adjacency matrix. Since the adjacency matrix is not easy to observe, this study converts the adjacency matrix into an actual sea area map for analysis. Limited by the length of the article, only the adjacency matrix visualization of the SCS is given here, as shown in Fig. 14.

It can be observed from Fig. 14 that the wave propagation in different months changed dynamically and could be divided into two different modes. Among them, the heat maps from January-May (Figs 14a–e) and October-December (Figs 14j–l) have similar propagation patterns, whereas the heat maps from June-August (Figs 14f–h) show a different pattern. Specifically, in the two periods of May-June (Figs 14e and f) and September-October (Figs 14i and j), the wave direction will change significantly. This phenomenon is more intuitive, as shown in Fig. 15.

Since the adjacency matrix is not easy to observe, for better analysis, this study converts the adjacency matrix into node information propagation direction and intensity and plots it on the latitude-longitude grid as shown in Figs 15 and 16.

As shown in Fig. 15, wave information propagation is a dynamic process. Specifically, from January to May (Figs 15a–e) and October to December (Figs 15j–l), the information propagation from the northeast dominated, whereas from June to August (Figs 15f–h), the information propagation primarily originated from the southwest. This seasonal variation in wave propagation can be explained by the direction of wind patterns in the South China Sea.

As the DWI in the South China Sea is primarily affected by the Australian (Asian monsoon) in this area (Mirzaei et al., 2015), the monsoon system comprises alternating summer (southwesterly) and winter (northeasterly) monsoons. During the summer monsoon (June-August), the prevailing southwesterly wind over the South China Sea (Zhai et al., 2021) causes waves to follow a southwestward direction. During the transition period of the summer monsoon (September-November), as the Siberian high-latitude airflow influence gradually increases (Dando, 2005), the wave direction of the South China Sea gradually turns southwestward and southward along the Malay Peninsula. From approximately November, the Siberian high-latitude airflow drives the air

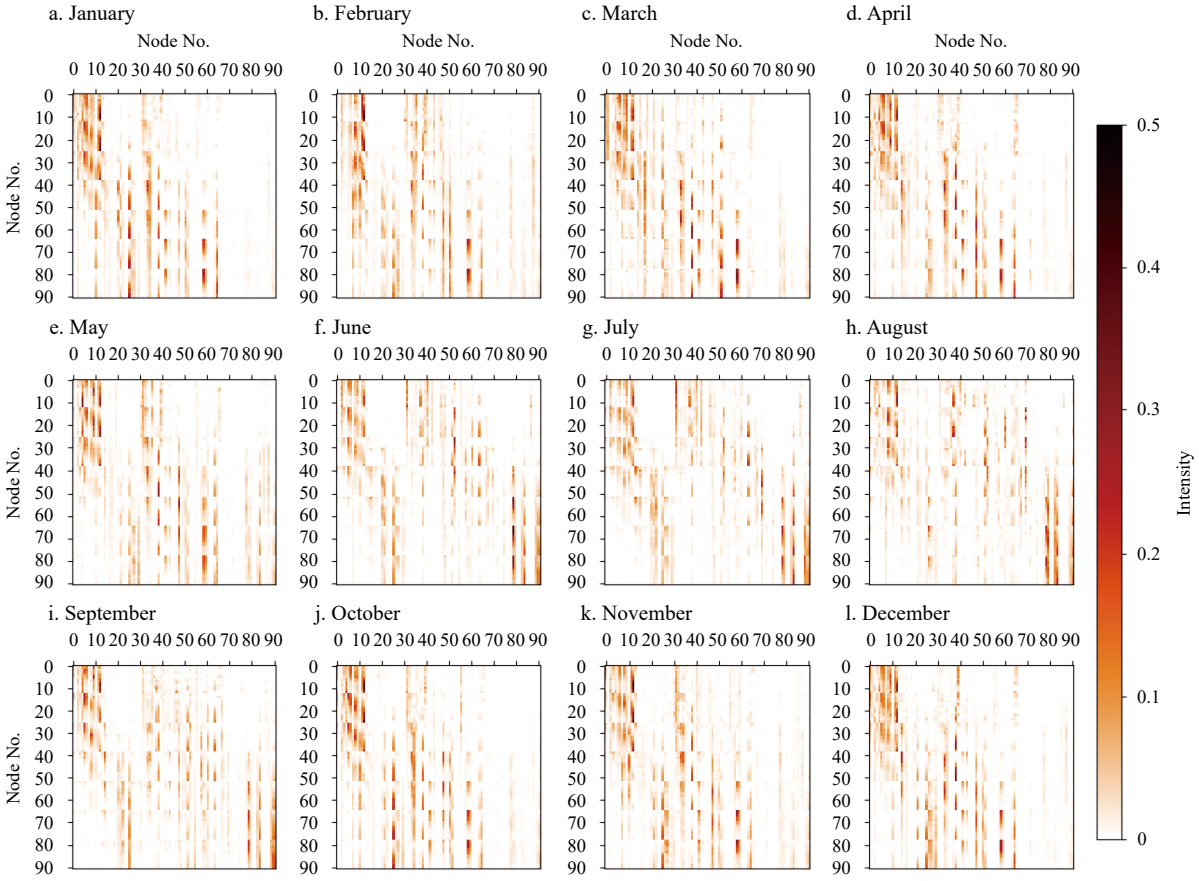

Fig. 14. Presents a visual representation of the adjacency matrix, which was derived from the data collected in each month of 2022 on the South China Sea test set. The adjacency matrix is employed to describe the direction and intensity of information propagation among the 91 nodes in the South China Sea. Each value in the adjacency matrix represents the propagation of a specific piece of information from a node on the horizontal axis to a node on the vertical axis. The color of the value indicates the extent of wave information propagation between the two nodes. The darker the color, the greater the intensity of information spread.

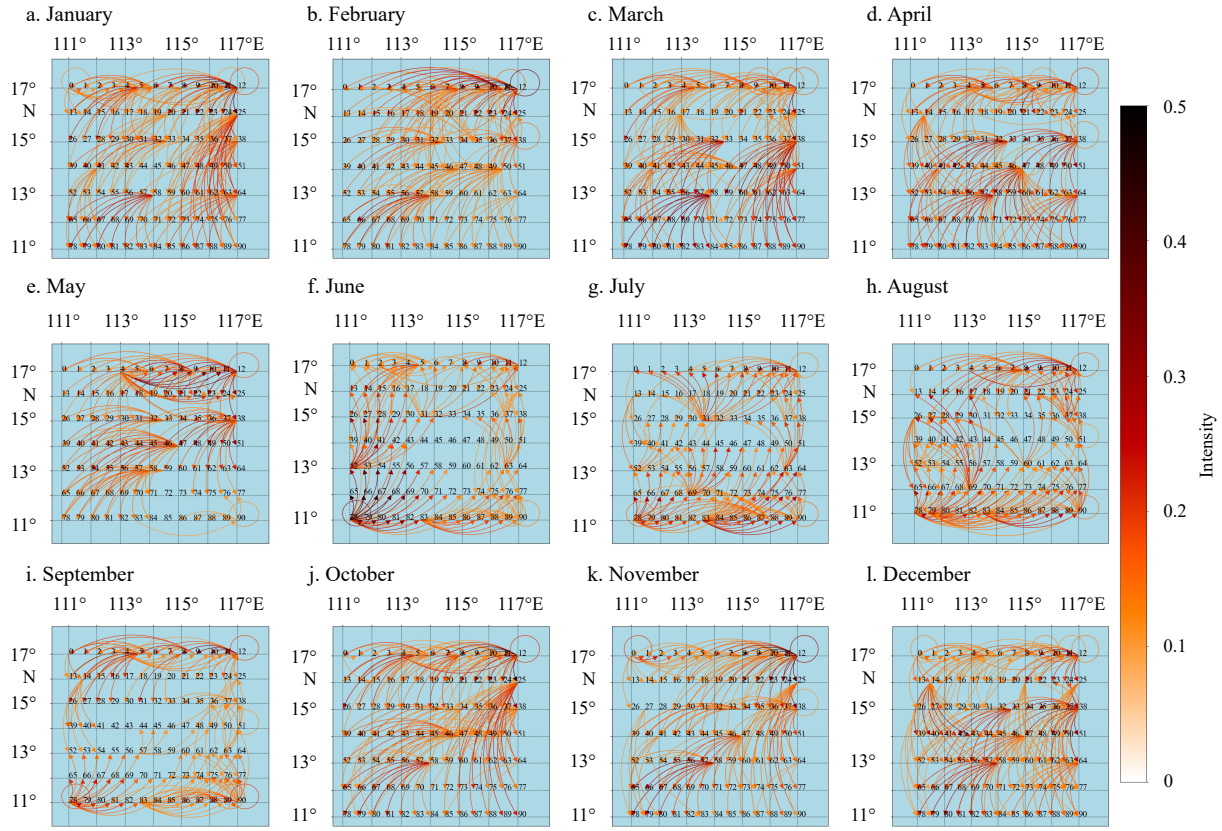

Fig. 15. Geographic visualization of the adjacency matrix for each month of 2022 on the South China Sea test set. The arrows represent the direction of wave propagation between nodes. The darker the color of the curve, the more information is transmitted between nodes.

around the anticyclone, which carries the air from Siberia across China and blows out from the South China Sea over the Nansha Islands. Thus, during the winter monsoon (December-May of the following year), the prevailing northeasterly wind over the South China Sea causes the wave direction to follow the northeastward direction.

The results for the East China Sea in Fig. 16 also show the dynamic seasonal variation mechanism. From January to March (Figs 16a–c) and October to December (Figs 16j–l), the information propagation in the north dominates; from April to September (Figs 16d–i), the information propagation originates from the southeast direction, which is also consistent with the results of related studies (Hisaki, 2023; Wang et al., 2016).

Furthermore, this study employed the test data from 2020 to investigate the dynamic seasonal variation mechanism across different years. Specifically, this study employed the test data to generate the 2020 adjacency matrix, which was then compared and analyzed with the 2022 adjacency matrix. The results are shown in Fig. 17. Each year is comprised of 12-month adjacency matrices. These adjacency matrices reflect the mode of information propagation of the nodes under consideration during a specified month. To ascertain the degree of similarity between the adjacency matrices of any two months in different years, this study employs the use of cosine similarity, which is a measure of the similarity of the two matrices in terms of their information propagation mode. The cosine similarity calculation results range between –1 and 1. When the result is closer to 1, the information propagation mode of the two matrices is more similar. To illustrate, in Fig. 17a, the cosine similarity between December 2020 and December 2022 exceeds 0.8, indicating a high degree of similarity between the two. Through the visualization of the adjacency matrix, it can be seen in Fig. 17a that the information propagation mode of the two adjacency matrices is from the northeast.

It can be seen in Fig. 17: (1) The diagonal of the matrix exhibits a high degree of similarity, indicating that the corresponding months of 2020 and 2022 have similar adjacency matrices. This similarity is reflected in the information propagation direction of the node. In the South China Sea (Fig. 17a), the information propagation in December 2020 and 2022 is primarily from the northeast direction. In April in the East China Sea (Fig. 17b), the information propagation in 2020 and 2022 is primarily sourced from the east. (2) In the South China Sea (Fig. 17a), the months with a cosine similarity greater than 0.5 in the matrix appear in a cyclical pattern across different regions. From the perspective of the months in question, this can be divided into two distinct modes. One mode tends to emerge from January to April and from October to December, while the other mode emerges from May to September. This is consistent with the dynamic seasonal variation mechanism of the South China Sea. (3) The matrix of the East China Sea (Fig. 17b) also exhibits two similar patterns. The adjacency matrix structure similarity is higher from January to March and from October to December, while the similarity is higher from April to September. This is consistent with the conclusion of dynamic seasonal variations in the East China Sea.

Therefore, the dynamic graph constructed by the STDG shows that the waves have a dynamic seasonal variation mechanism in space. This finding indicates that it is reasonable and ef-

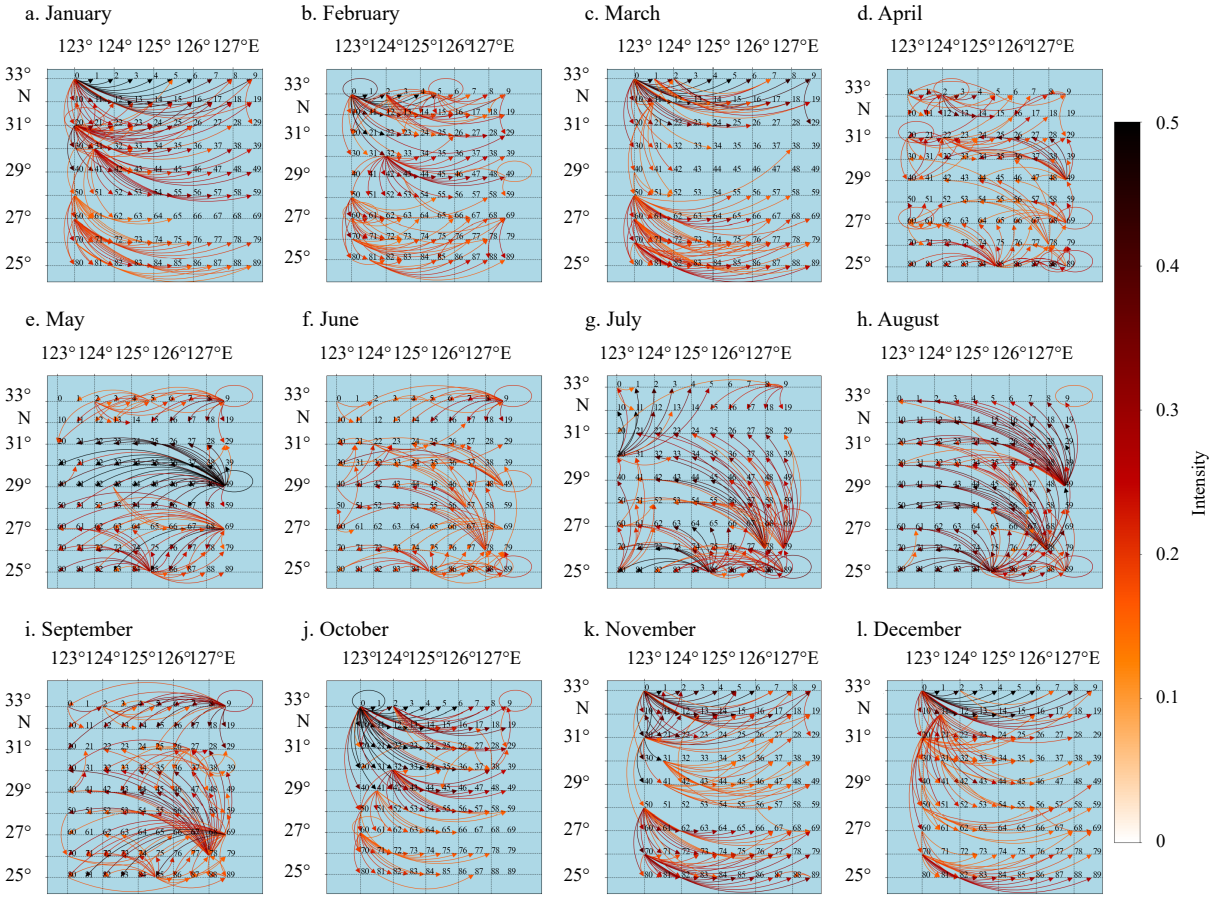

Fig. 16. Geographic visualization of the adjacency matrix for each month of 2022 on the East China Sea test set. The arrows represent the direction of wave propagation between nodes. The darker the color of the curve, the more information is transmitted between nodes.

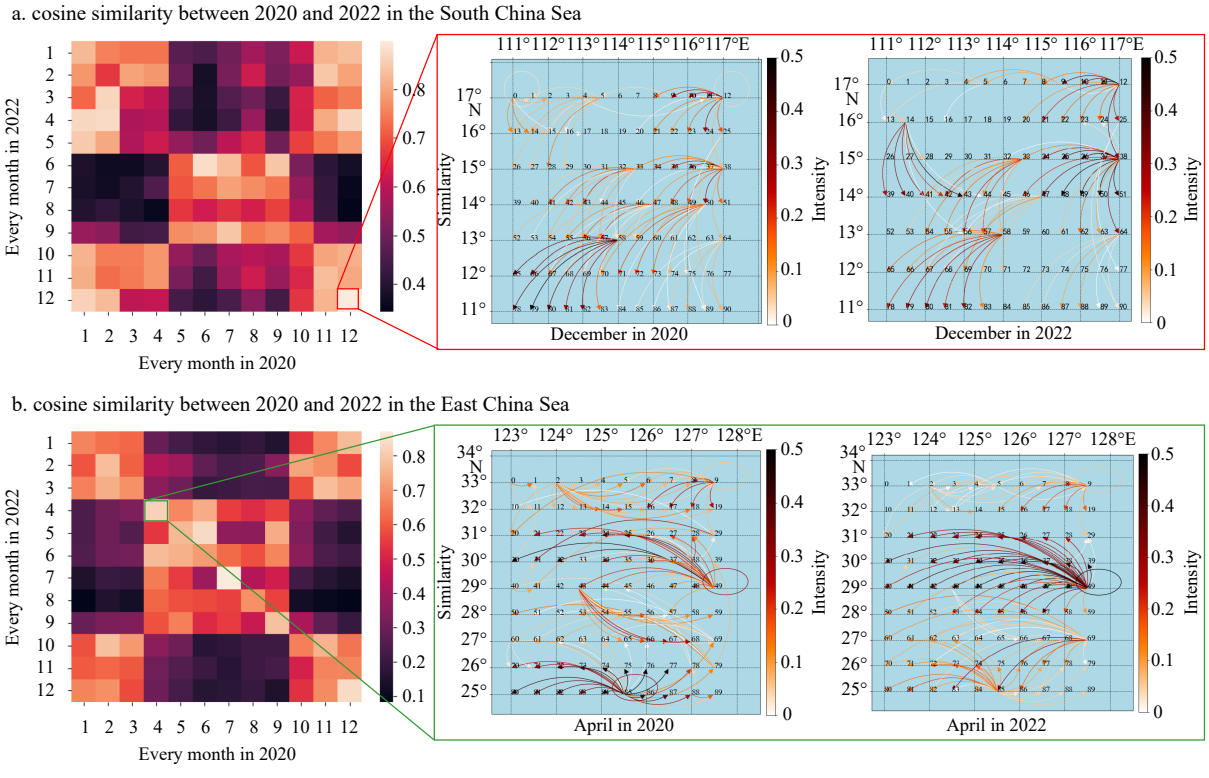

Fig. 17. Comparison of cosine similarity matrix results for each month in different years.

fective to fuse the dynamic driving characteristics of waves in the construction of dynamic graphs, which not only improves the predictive ability of the SWH of multiple nodes but also ensures that the constructed dynamic graph is consistent with the seasonal variation mechanism of waves. Furthermore, the dynamic graph constructed by the STDG can provide a reference for the design and use of WEC arrays to improve the power generation efficiency of wave energy.

5 Conclusions and future work

This study proposes a multi-node SWH prediction method based on dynamic graph modeling and multi-characteristic fusion, namely STDG. This method employs a unified graph structure comprising the coordinate points of different latitudes and longitudes in the sea area, with dynamic graphs used to capture the evolving graph structure over time. In consideration of the seasonal variations in the wave direction over time, which are the joint effects of long-term and short-term patterns, this study employs node embedding to adaptively learn fixed and static long-term patterns in data. Based on the long-term patterns, this study designed an SDGC module that can generate short-term patterns based on long-term patterns and dynamic inputs. These two patterns are then input into their respective graph convolution modules for further calculations. In addition, for the correlation of multiple characteristics with the SWH, this study designed a CCTF module that can effectively fuse multiple characteristics and enhance the accuracy of SWH prediction.

To verify the performance of the STDG, seven ocean characteristics were selected as input at multiple nodes in two different sea areas of the South China Sea and East China Sea. These characteristics were significant wave height, mean wave period, sea surface temperature, mean sea level pressure, 10 m wind U component, 10 m wind V component, and mean wave direction. This study compared the STDG with five methods, namely LSTM, GRU, CLTS-Net, MTGNN, and MCMN, and employed MAE, RMSE, and R-square as the evaluation indicators for the prediction of SWH over the next 72 h.

In the South China Sea and East China Sea datasets, STDG demonstrated the best performance in the prediction results over different time steps. Especially when predicting SWH in the next 24 h, the STDG showed a higher prediction accuracy in dealing with extreme values and curve decline stages. In addition, the dynamic graph analysis results showed that when predicting multi-nodal SWH for different months, the dynamic graph constructed by STDG showed seasonal variations; this is consistent with the conclusions of the existing research.

However, this study has some limitations. For example, it is unable to predict multiple ocean characteristics simultaneously and requires further verification with *in situ* data. Additionally, consideration of the differences between different locations is necessary. Therefore, in future research, we plan to combine existing observational data and bathymetry information and construct a multi-characteristic ocean prediction model to simultaneously predict multiple ocean characteristics and further exploration of model interpretability.

References

Abu-El-Haija S, Perozzi B, Kapoor A, et al. 2019. Mixhop: Higher-order graph convolutional architectures via sparsified neighborhood mixing. In: Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 21–29

Ali M, Prasad R. 2019. Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. Renewable and Sustainable Energy Reviews, 104: 281–295, doi: 10.1016/j.rser.2019.01.014

Ba J L, Kiros J R, Hinton G E. 2016. Layer normalization. USA: arXiv preprint arXiv: 1607. 06450, https://doi.org/10.48550/arXiv.1607.06450[2016-07-21/2023-09-01]

Chen Delong, Liu Fan, Zhang Zheqi, et al. 2021. Significant wave height prediction based on wavelet graph neural network. In: 2021 IEEE 4th International Conference on Big Data and Artificial Intelligence (BDAI). Qingdao: IEEE, 80–85

Cho K, Van Merriënboer B, Gulcehre C, et al. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: Association for Computational Linguistics, 1724–1734

Dando W A. 2005. Asia, climates of Siberia, Central and East Asia. In: Oliver J E, ed. Encyclopedia of World Climatology. Dordrecht: Springer, 102–114

Dauphin Y N, Fan A, Auli M, et al. 2017. Language modeling with gated convolutional networks. In: Proceedings of the 34th International Conference on Machine Learning. Sydney: JMLR. org, 933-941

De Andrés A D, Guanche R, Meneses L, et al. 2014. Factors that influence array layout on wave energy farms. Ocean Engineering, 82: 32–41, doi: 10.1016/j.oceaneng.2014.02.027

Ding Jie, Deng Fangyu, Liu Qi, et al. 2023. Regional forecasting of significant wave height and mean wave period using EOF-EEMD-SCINet hybrid model. Applied Ocean Research, 136: 103582, doi: 10.1016/j.apor.2023.103582

Duan Wenyang, Han Yang, Huang Limin, et al. 2016. A hybrid EMD-SVR model for the short-term prediction of significant wave height. Ocean Engineering, 124: 54–73, doi: 10.1016/j.oceaneng.2016.05.049

Feng Xi, Ma Gangfeng, Su S F, et al. 2020. A multi-layer perceptron approach for accelerated wave forecasting in Lake Michigan. Ocean Engineering, 211: 107526, doi: 10.1016/j.oceaneng.2020.107526

Gao Yuan, Miyata S, Akashi Y. 2022. Interpretable deep learning models for hourly solar radiation prediction based on graph neural network and attention. Applied Energy, 321: 119288, doi: 10.1016/j.apenergy.2022.119288

Gao Hong, Xiao Jie. 2021. Effects of power take-off parameters and harvester shape on wave energy extraction and output of a hydraulic conversion system. Applied Energy, 299: 117278, doi: 10.1016/j.apenergy.2021.117278

Ge Ming, Kerrigan E C. 2016. Short-term ocean wave forecasting using an autoregressive moving average model. In: 2016 UKACC 11th International Conference on Control (CONTROL). Belfast: IEEE, 1–6

Geng Xiulin, Xu Lingyu, He Xiaoyu, et al. 2021. Graph optimization neural network with spatio-temporal correlation learning for multi-node offshore wind speed forecasting. Renewable Energy, 180: 1014–1025, doi: 10.1016/j.renene.2021.08.066

Goda Y. 2010. Random seas and design of maritime structures. New Jersey: World Scientific Publishing Company, 3–10

Hashim R, Roy C, Motamedi S, et al. 2016. Selection of climatic parameters affecting wave height prediction using an enhanced Takagi-Sugeno-based fuzzy methodology. Renewable and Sustainable Energy Reviews, 60: 246–257, doi: 10.1016/j.rser.2016.01.098

Hersbach H, Bell B, Berrisford P, et al. 2023a. ERA5 hourly data on single levels from 1940 to present. United Kingdom: Copernicus Climate Change Service (C3S), Climate Data Store (CDS), doi: 10.24381/cds.adbb2d47, https://doi.org/10.24381/cds.bd0915c6 [2018-06-14/2023-09-01]

Hersbach H, Bell B, Berrisford P, et al. 2023b. ERA5 monthly averaged data on single levels from 1940 to present. United Kingdom: Copernicus Climate Change Service (C3S), Climate Data Store (CDS), doi: 10.24381/cds.f17050d7, https://doi.org/10.24381/cds.f17050d7 [2019-04-18/2023-09-01]

Hisaki Y. 2023. Swell and wind-wave height variability in the East

China Sea. Ocean Dynamics, 73(8): 493–515, doi: 10.1007/s10236-023-01552-0
Hochreiter S, Schmidhuber J. 1997. Long short-term memory. Neural Computation, 9(8): 1735–1780, doi: 10.1162/neco.1997.9.8.1735
Khosravi A, Machado L, Nunes R O. 2018. Time-series prediction of wind speed using machine learning algorithms: a case study Osorio wind farm, Brazil. Applied Energy, 224: 550–566, doi: 10.1016/j.apenergy.2018.05.043
Kingma D P, Ba J. 2015. Adam: a method for stochastic optimization. In: 3rd International Conference on Learning Representations. San Diego: ICLR
Kitagawa G, Gersch W. 1984. A smoothness priors-state space modeling of time series with trend and seasonality. Journal of the American Statistical Association, 79(386): 378–389
Li Xinfang, Cao Jinfeng, Guo Jihong, et al. 2022. Multi-step forecasting of ocean wave height using gate recurrent unit networks with multivariate time series. Ocean Engineering, 248: 110689, doi: 10.1016/j.oceaneng.2022.110689
Li Shuang, Hao Peng, Yu Chengcheng, et al. 2021. CLTS-net: a more accurate and universal method for the long-term prediction of significant wave height. Journal of Marine Science and Engineering, 9(12): 1464, doi: 10.3390/jmse9121464
Luo Qinrui, Xu Hang, Bai Longhu. 2022. Prediction of significant wave height in hurricane area of the Atlantic Ocean using the Bi-LSTM with attention model. Ocean Engineering, 266: 112747, doi: 10.1016/j.oceaneng.2022.112747
Ma Qijie, Wang Peijun, Fan Jianhua, et al. 2022. Underground solar energy storage via energy piles: an experimental study. Applied Energy, 306: 118042, doi: 10.1016/j.apenergy.2021.118042
Mahjoobi J, Etemad-Shahidi A. 2008. An alternative approach for the prediction of significant wave heights based on classification and regression trees. Applied Ocean Research, 30(3): 172–177, doi: 10.1016/j.apor.2008.11.001
Mahjoobi J, Mosabbeb E A. 2009. Prediction of significant wave height using regressive support vector machines. Ocean Engineering, 36(5): 339–347, doi: 10.1016/j.oceaneng.2009.01.001
Mandal S, Prabaharan N. 2006. Ocean wave forecasting using recurrent neural networks. Ocean Engineering, 33(10): 1401–1410, doi: 10.1016/j.oceaneng.2005.08.007
Mirzaei A, Tangang F, Juneng L. 2015. Wave energy potential assessment in the central and southern regions of the South China Sea. Renewable Energy, 80: 454–470, doi: 10.1016/j.renene.2015.02.005
Niu Dongxiao, Yu Min, Sun Lijie, et al. 2022. Short-term multi-energy load forecasting for integrated energy systems based on CNN-BiGRU optimized by attention mechanism. Applied Energy, 313: 118801, doi: 10.1016/j.apenergy.2022.118801
Oreshkin B N, Carpov D, Chapados N, et al. 2020. N-BEATS: neural basis expansion analysis for interpretable time series forecasting. In: 8th International Conference on Learning Representations. Addis Ababa: ICLR
Owens E H. 1982. Sea conditions. In: Schwartz M, ed. Beaches and Coastal Geology. New York: Springer, 722
Pirhooshyaran M, Snyder L V. 2020. Forecasting, hindcasting and feature selection of ocean waves via recurrent and sequence-to-sequence networks. Ocean Engineering, 207: 107424, doi: 10.1016/j.oceaneng.2020.107424
Quach B, Glaser Y, Stopa J E, et al. 2021. Deep learning for predicting significant wave height from synthetic aperture radar. IEEE Transactions on Geoscience and Remote Sensing, 59(3): 1859–1867, doi: 10.1109/TGRS.2020.3003839
Reikard G, Pinson P, Bidlot J R. 2011. Forecasting ocean wave energy: the ECMWF wave model and time series methods. Ocean Engineering, 38(10): 1089–1099, doi: 10.1016/j.oceaneng.2011.04.009
Ren Yuting, Li Zhuolin, Xu Lingyu, et al. 2023. The data-based adaptive graph learning network for analysis and prediction of offshore wind speed. Energy, 267: 126590, doi: 10.1016/j.energy.2022.126590
Soares C G, Cunha C. 2000. Bivariate autoregressive models for the time series of significant wave height and mean period. Coastal Engineering, 40(4): 297–311, doi: 10.1016/S0378-3839(00)00015-6
Song Wei, Li Qichao, He Qi, et al. 2021. Determining wave height from nearshore videos based on multi-level spatiotemporal feature fusion. In: 2021 International Joint Conference on Neural Networks (IJCNN). Shenzhen: IEEE, 1–8
Ti Zilong, Song Yubing, Deng Xiaowei. 2022. Spatial-temporal wave height forecast using deep learning and public reanalysis dataset. Applied Energy, 326: 120027, doi: 10.1016/j.apenergy.2022.120027
Tolman H L. 2009. User manual and system documentation of WAVEWATCH III TM version 3.14. Technical note, MMAB Contribution, 276: 220.
Tsai Y H H, Bai Shaojie, Liang P P, et al. 2019. Multimodal transformer for unaligned multimodal language sequences. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: Association for Computational Linguistics, 6558–6569
Tsai C P, Lin Chang, Shen Jia-N. 2002. Neural network for wave forecasting among multi-stations. Ocean Engineering, 29(13): 1683–1695, doi: 10.1016/S0029-8018(01)00112-3
Tuttle J F, Blackburn L D, Andersson K, et al. 2021. A systematic comparison of machine learning methods for modeling of dynamic processes applied to combustion emission rate modeling. Applied Energy, 292: 116886, doi: 10.1016/j.apenergy.2021.116886
Umair M, Hashmani M A, Hasan M H B. 2019. Survey of sea wave parameters classification and prediction using machine learning models. In: 2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS). Ipoh: IEEE, 1–6
Vaswani A, Shazeer N, Parmar N, et al. 2017. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 6000-6010
Wang Jin, Dong Changming, He Yijun. 2016. Wave climatological analysis in the East China Sea. Continental Shelf Research, 120: 26–40, doi: 10.1016/j.csr.2016.03.010
Wang Jichao, Liu Jincan, Wang Yue, et al. 2021. Spatiotemporal variations and extreme value analysis of significant wave height in the South China Sea based on 71-year long ERA5 wave reanalysis. Applied Ocean Research, 113: 102750, doi: 10.1016/j.apor.2021.102750
Wang Jichao, Sun Peidong, Liao Zhihong, et al. 2022. Long-term trend analysis of wave characteristics in the Bohai Sea based on interpolated ERA5 wave reanalysis from 1950 to 2020. Acta Oceanologica Sinica, 41(7): 97–112, doi: 10.1007/s13131-021-1974-0
Wang Wenxu, Tang Ruichun, Li Cheng, et al. 2018. A BP neural network model optimized by mind evolutionary algorithm for predicting the ocean wave heights. Ocean Engineering, 162: 98–107, doi: 10.1016/j.oceaneng.2018.04.039
Wang Jichao, Wang Yue. 2022. Evaluation of the ERA5 significant wave height against NDBC buoy data from 1979 to 2019. Marine Geodesy, 45(2): 151–165, doi: 10.1080/01490419.2021.2011502
Wu Zonghan, Pan Shirui, Long Guodong, et al. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Association for Computing Machinery, 753–763
Xiang Ling, Yang Xin, Hu Aijun, et al. 2022. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. Applied Energy, 305: 117925, doi: 10.1016/j.apenergy.2021.117925
Xu Xinxin, Robertson B, Buckham B. 2020. A techno-economic approach to wave energy resource assessment and development site identification. Applied Energy, 260: 114317, doi: 10.1016/j.apenergy.2019.114317
Yang Bo, Wu Shaocong, Zhang Hao, et al. 2022. Wave energy converter array layout optimization: a critical and comprehensive overview. Renewable and Sustainable Energy Reviews, 167:

112668, doi: 10.1016/j.rser.2022.112668

Yu F, Koltun V. 2016. Multi-scale context aggregation by dilated convolutions. In: 4th International Conference on Learning Representations. San Juan: ICLR

Zhai Fangguo, Wu Wenfan, Gu Yanzhen, et al. 2021. Dynamics of the seasonal wave height variability in the South China Sea. International Journal of Climatology, 41(2): 934–951, doi: 10.1002/joc.6707

Zhang Chu, Hua Lei, Ji Chunlei, et al. 2022. An evolutionary robust solar radiation prediction model based on WT-CEEMDAN and IASO-optimized outlier robust extreme learning machine. Applied Energy, 322: 119518, doi: 10.1016/j.apenergy.2022.119518

Zhang Yao, Xu Lingyu, Yu Jie. 2023. Significant wave height prediction based on dynamic graph neural network with fusion of ocean characteristics. Dynamics of Atmospheres and Oceans, 103: 101388, doi: 10.1016/j.dynatmoce.2023.101388

Zhou Shuyi, Xie Wenhong, Lu Yuxiang, et al. 2021. ConvLSTM-based wave forecasts in the South and East China seas. Frontiers in Marine Science, 8: 680079, doi: 10.3389/fmars.2021.680079