Graph-Based Memory Recall Recurrent Neural Network for Mid-Term Sea-Surface Height Anomaly Forecasting

Yuan Zhou, *Senior Member, IEEE*, Tian Ren, Keran Chen, Le Gao, *Member, IEEE*, and Xiaofeng Li, *Fellow, IEEE*

***Abstract*—Sea surface height anomaly (SSHA) plays a pivotal role in ocean dynamics and climate systems. This article develops a graph-based memory recall recurrent neural network (GMR-Net) to achieve accurate and reliable mid-term spatiotemporal prediction of the SSHA field. The proposed method designs a newly developed long-term memory recall cell as the building block of the network, which utilizes the proposed memory store recall (MSR) module to learn and capture the mid- and long-term temporal dependencies of the SSHA field. The MSR module can efficiently recall memories stored in the memory bank across multiple timestamps through the proposed graph representation mechanism even after long periods of disturbance. The mid-term SSHA forecasting is performed with a 30-day ahead, and our proposed GMR-Net model achieves high prediction accuracy in different geographical regions: the Tropical Western Pacific and the South China Sea, yielding an RMSE of 0.026 and 0.035 m, respectively. Compared with advanced prediction models, our proposed GMR-Net model exhibits high reliability and superior performance in mid-term SSHA forecasting. Moreover, marine phenomena, such as Rossby waves, which can cause dramatic changes in sea-surface height, are successfully observed from our forecast data, further verifying the effectiveness of our prediction method.**

***Index Terms*—Sea-surface height anomaly (SSHA), deep learning (DL), spatiotemporal prediction, Rossby waves.**

I. Introduction

SEA-SURFACE height anomaly (SSHA) is a critical parameter for emergency response to marine disasters [1], [2], [3]. It is defined as the disparity between the satellite-observed sea-surface height (SSH) and the mean SSH. The variations in mean sea level driven by intraseasonal to interannual climate processes significantly impact the occurrence of extreme sea-level events, such as the El Niño/Southern Oscillation, the Southern Annular Mode, the Indian Ocean Dipole, and the Madden–Julian Oscillation [4], [5]. Therefore, mid- and long-term SSHA forecasting can provide more valuable information for understanding extreme weather or climate, facilitating the effective implementation of management policies and strategies to mitigate the risks of life and infrastructure damage. The development of ocean observation technology, especially remote sensing technology, provides high-precision satellite altimetry data to study SSHA [6]. Many methods have been proposed for SSHA forecasting using satellite altimetry data, which can be categorized into two main types based on the adopted model: 1) physics-based numerical models and 2) data-driven models.

Traditionally, physics-based numerical models utilize a series of complex physics equations to estimate sea-level changes. Chen et al. [7] employed a continental hydrological model to forecast annual and semiannual fluctuations in global mean sea-level data obtained from the TOPEX/Poseidon altimeter. Gregory and Lowe [8] undertook a comparison of future regional and global sea-level changes predicted by two different versions (HadCM2 and HadCM3) of the Hadley Centre coupled atmosphere–ocean general circulation model. Miles et al. [9] proposed a dynamical coupled ocean–atmosphere model that exhibited the capacity to forecast global seasonal sea-level anomalies seven months in advance. However, despite the prevalence of numerical forecasting as the primary method for SSHA forecasting, it is subject to certain limitations, such as high computational requirements and sensitivity to initial conditions.

The data-driven models can develop a forecasting model in the statistical sense without taking into account the physical principles of the predicted object when the amount of sample data is sufficiently large. Niedzielski et al. [10] employed a polynomial-harmonic model based on the least-squares (LS) method for mid- and long-term forecasting of global mean SSHA. Ezer et al. [11] devised a novel approach based on empirical mode decomposition (EMD) to extract long-term trends of sea-level variations from oscillatory patterns along the mid-Atlantic coast of the United States. Fu et al. [12] designed a hybrid model that integrates empirical model decomposition, singular spectrum analysis, and LS to forecast long-term sea-surface anomalies in the South China Sea (SCS). Imani et al. [13] combined empirical orthogonal function and support vector

Manuscript received 1 December 2023; revised 8 January 2024; accepted 9 February 2024. Date of publication 1 March 2024; date of current version 19 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant U2006211 and 42376175, in part by the Strategic Priority Research Program of the Chinese Academy of Sciences (CAS) under Grant XDA19060101 and Grant XDB42040401, in part by the Key Research and Development Project of Shandong Province under Grant 2019JZZY010102, and in part by the CAS program under Grant Y9KY04101L. *(Corresponding authors: Le Gao; Xiaofeng Li.)*

Yuan Zhou, Tian Ren, and Keran Chen are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: zhouyuan@tju.edu.cn; rentian990413@tju.edu.cn; erichen@tju.edu.cn).

Le Gao and Xiaofeng Li are with the Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China (e-mail: gaole@qdio.ac.cn; xiaofeng.li@ieee.org).

Digital Object Identifier 10.1109/JSTARS.2024.3368766

regression to forecast long-term sea-level anomalies derived from satellite altimetry in the tropical Pacific Ocean. Cui et al. [14] presented a new composite analysis-based random forest (RF) approach for high-precision mid-term prediction of the daily area-averaged SSHA in the SCS and the Western North Pacific subtropical region. Although data-driven models have the potential to forecast future SSHA based on historical statistical data, the nonlinearity and stochastic nature of SSHA data and the inherent uncertainties of the ocean make it challenging to use mathematical or statistical models for SSHA spatiotemporal prediction.

Deep learning (DL) is a data-driven method with a strong nonlinear expressive ability and DL-based models [15], [16] have achieved advanced performance in the field of ocean prediction. DL has also been widely employed for forecasting complex SSHA data. Braakmann-Folgmann et al. [17] combined recurrent neural network (RNN) with convolutional neural network (CNN) to analyze the evolution of SSH in both temporal and spatial dimensions and predict SSHA. Sun et al. [18] utilized long short-term memory (LSTM) models and the seasonal autoregressive integrated moving average to enhance the prediction performance of SSHA in the China Sea. Liu et al. [19] designed an attention-based LSTM network [20] that achieved highly reliable SSH forecasting. Song et al. [21] demonstrated that merged-LSTM could predict SSHA values for the next five days with acceptable accuracy. Zhou et al. [22] proposed a multilayer fusion RNN (MLFrnn) that combined global and local spatiotemporal features of SSHA to predict SSHA regions.

The DL-based studies mentioned above underscore the potential of DL techniques in marine environment prediction. However, longer-term spatiotemporal forecasting of the SSHA field is a less commonly studied area due to the relatively low prediction accuracy that often accompanies it. Meanwhile, the complexity inherent in SSHA data presents a significant challenge for achieving accurate longer-term forecasting. Therefore, it is imperative to identify a precise and reliable DL approach to facilitate longer-term spatiotemporal SSHA forecasting.

The longer-term spatiotemporal forecasting of SSHA necessitates capturing the mid- and long-term temporal dependencies of the SSHA field. The abovementioned DL methods have chiefly relied on nearby phenomena in time and the ongoing short-term motion to establish frame relationships. Consequently, these methods have struggled to achieve mid-term SSHA forecasting due to the lack of mid- and long-term motion information.

In order to solve this problem, this article proposes a mid-term SSHA forecasting method with a long-term memory recall (LMR) cell, which can effectively capture the mid- and long-term temporal dependencies of the SSHA field by recalling memories stored in the memory bank through a proposed graph representation mechanism.

Furthermore, the satellite data-driven DL models [23] have been proven to be effective in revealing complex ocean phenomena. Since temporal and spatial changes in sea level are the result of large-scale changes in ocean circulation, average ocean density, and ocean barometric changes [24], SSHA can be used to reveal various critical large-scale oceanic phenomena, such as Rossby waves. Rossby waves, the large-scale fluctuation in the ocean interior caused by the rotation and shape of the Earth [25], [26], are characterized as passing through many ocean basins from east to west at a speed of a few centimeters per second [27], [28]. Rossby waves regulate tropical cyclone activity over tropical oceans and play a crucial role in the timely adjustment of ocean circulation to large-scale atmospheric changes [29], [30]. The accurate mid-term SSHA forecasting can provide valuable data support for the observations of Rossby waves in advance. This article has also conducted corresponding research on this.

To summarize, this study proposes a graph-based memory recall recurrent neural network (GMR-Net) to effectively capture the mid- and long-term temporal dependencies of the SSHA field, improving the mid-term prediction performance. Moreover, Rossby waves, which cause dramatic changes in SSH, are successfully observed from our forecast data. The main contributions of this article are as follows.

1) A novel LMR cell is designed to learn mid- and long-term temporal dependencies of the SSHA field through the proposed memory store recall (MSR) module. The MSR module can store SSHA high-dimensional features of previous timestamps and utilize the proposed graph representation mechanism to recall stored features across multiple timestamps, effectively mining the mid- and long-term memory context features of SSHA and improving the mid-term SSHA forecasting performance.
2) A novel GMR-Net model is proposed to achieve end-to-end mid-term spatiotemporal forecasting of the SSHA field. Previous studies on mid- and long-term SSHA forecasting have only focused on predicting the SSHA values for a single site or the area-averaged SSHA values for an entire region. Our proposed GMR-Net model can simultaneously capture the spatial variation and temporal dynamics of the mid-term SSHA field, enabling the prediction of the SSHA map. Furthermore, we have proposed a Fast Track mechanism to alleviate the problem of cross-layer motion information attenuation in the recurrent predictor, resulting in more accurate mid-term spatiotemporal SSHA prediction.
3) This article is the first to observe Rossby waves in advance from the SSHA data predicted by a DL model. This finding further illustrates the validity of our proposed prediction method. Meanwhile, this study also provides a new perspective and approach for the study of Rossby waves.

The rest of this article is organized as follows. Section II introduces the preliminary related to our work. Section III elaborates on the specifics of our proposed model, including the network architecture, LMR cell, and MSR module. Section IV describes the study area and dataset, model setting, implementation details, analysis of the experimental results, and a case study on Rossby wave observations. Finally, Section V concludes this article.

II. Preliminary

A. Graph Structure

Graph is a data structure in non-Euclidean space, composed of nodes and edges. Let $G(V, E)$ represent a graph, where V

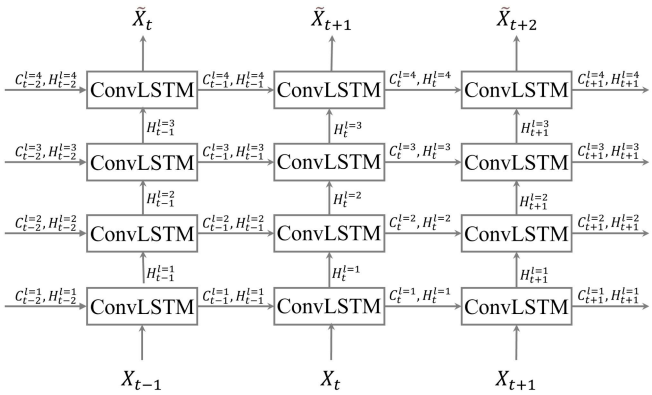


Fig. 1. *Four-layer ConvLSTM network.* $\mathcal{C}$ represents the standard temporal memory state, which is passed in the horizontal direction. $\mathcal{H}$ stands for hidden state, which is passed in both horizontal and vertical directions.

denotes the node set of the graph with $v_i \in V$, and E indicates the edge set with $< v_i, v_j > \in E$. Graph is a common data structure for elucidating a connection between disparate entities, whereby nodes denote individual entities, and edges represent the relationship between any two entities. Edges can be directed or undirected, and may possess weights or labels. The graph structure can be represented by an adjacency matrix A. For a weighted graph, A_{ij} can represent the weight of the edge from v_i to v_j. If it is an unweighted graph, A_{ij} can be set to 1 to indicate the existence of edges, and set to 0 to indicate the absence of edges.

By learning and analyzing the graph structure, the deep structural information of the data can be effectively explored to comprehend the interaction patterns among nodes [31], [32]. The graph structure employs edges to represent the relationships between nodes, thereby facilitating the capture of the intricate interactions and influences among nodes. For mid-term SSHA forecasting, by representing the information at different timestamps with nodes, a graph-based approach can better capture mid- and long-term temporal dependencies of the SSHA field.

B. ConvLSTM Network

In order to achieve precise mid-term SSHA forecasting, it is crucial to capture both spatial features and temporal dynamics in the model, which has been a challenge for previous LSTM-based methods. To address this limitation, Wang et al. [33] proposed ConvLSTM based on LSTM, which is designed to simultaneously capture the spatiotemporal information hidden in the sequence to realize the prediction of natural phenomena.

As shown in Fig. 1, a four-layer spatiotemporal prediction network is obtained by stacking ConvLSTM cells, which are characterized by the incorporation of convolution operations to capture spatial features and LSTM structure to process temporal dependencies. Compared with LSTM, ConvLSTM provides greater capability to model short-term spatiotemporal information by utilizing convolution operators. ConvLSTM can naturally deal with the spatiotemporal relationship in sequence data, avoiding the defects of separate processing of spatial and temporal information in traditional CNN and LSTM models.

However, ConvLSTM cannot capture mid- and long-term temporal dependencies of the SSHA field since the older standard temporal memory states $\mathcal{C}$ will be discarded immediately when the standard temporal memory states are refreshed. Moreover, features obtained by high-layer cells will lose feature information extracted by low-layer cells in the ConvLSTM network with stacked multilayer cells, which makes ConvLSTM unable to capture the complete spatial motion information of SSHA and result in ambiguous predictions. Our model improves on ConvLSTM to address the above issues.

III. Methodology

As previously mentioned, the problems with ConvLSTM are its inability to capture mid- and long-term temporal dependencies and the ambiguity of its predictions. Our proposed GMR-Net model is designed to address these problems and applied specifically to mid-term SSHA forecasting. The LMR cell captures the mid- and long-term temporal dependencies of the SSHA field through our proposed MSR module and extracts the spatiotemporal features of complex SSHA sequences using 3D convolution. In addition, the Fast Track mechanism is proposed in the network to provide another fast path for the motion context information and effectively avoid the motion vanishing. The network architecture, LMR cell, and MSR module are explained in detail as follows.

A. Network Architecture

Fig. 2 shows the architecture of our proposed GMR-Net model, which comprises three subnetworks: 1) encoder, 2) recurrent predictor, and 3) decoder. The encoder and decoder are composed of several layers of 3-D convolution. The recurrent predictor consists of the proposed LMR cell and Fast Track mechanism. To improve the model's expressivity, we stacked four-layer LMR cells to construct a recurrent predictor subnetwork. Meanwhile, the Fast Track mechanism is connected with four-layer LMR cells, which transfers and fuses the hidden states generated by the four-layer LMR cells.

Let $\mathcal{X}_t \in \mathbb{R}^{H\times W\times C}$ denotes the SSHA field at time step t, where $H, W,$ and C indicate height, weight, and channels of the input frame, respectively. In order to capture the mid- and long-term temporal dependencies, we enlarge the temporal receptive field of the model at each time step. Specifically, at each time step in our GMR-Net model, n consecutive input frames $\mathcal{X}_{t-n+1:t}$ are first encoded by an encoder to obtain high-dimension feature maps in $\mathbb{R}^{n\times H\times W\times C}$ tensors, where n indicates temporal length. Then, the obtained convolutional features are incorporated into the recurrent predictor to generate recurrent features. Finally, the recurrent features are decoded using a 3-D convolutional decoder to obtain the predicted map of the future frame $\tilde{\mathcal{X}}_{t+1}$. The predicted frame $\tilde{\mathcal{X}}_{t+1}$ can be represented as follows:

$$\tilde{\mathcal{X}}_{t+1} = \mathrm{Dec}[P(\mathrm{Enc}\,(\mathcal{X}_{t-n+1:t})] \tag{1}$$

where $\mathrm{Enc}(\cdot)$ denotes the encoder which is utilized to extract deep features from the input, $P(\cdot)$ denotes the proposed recurrent predictor, and $\mathrm{Dec}(\cdot)$ denotes the decoder which is utilized to map the predicted features to the frames.

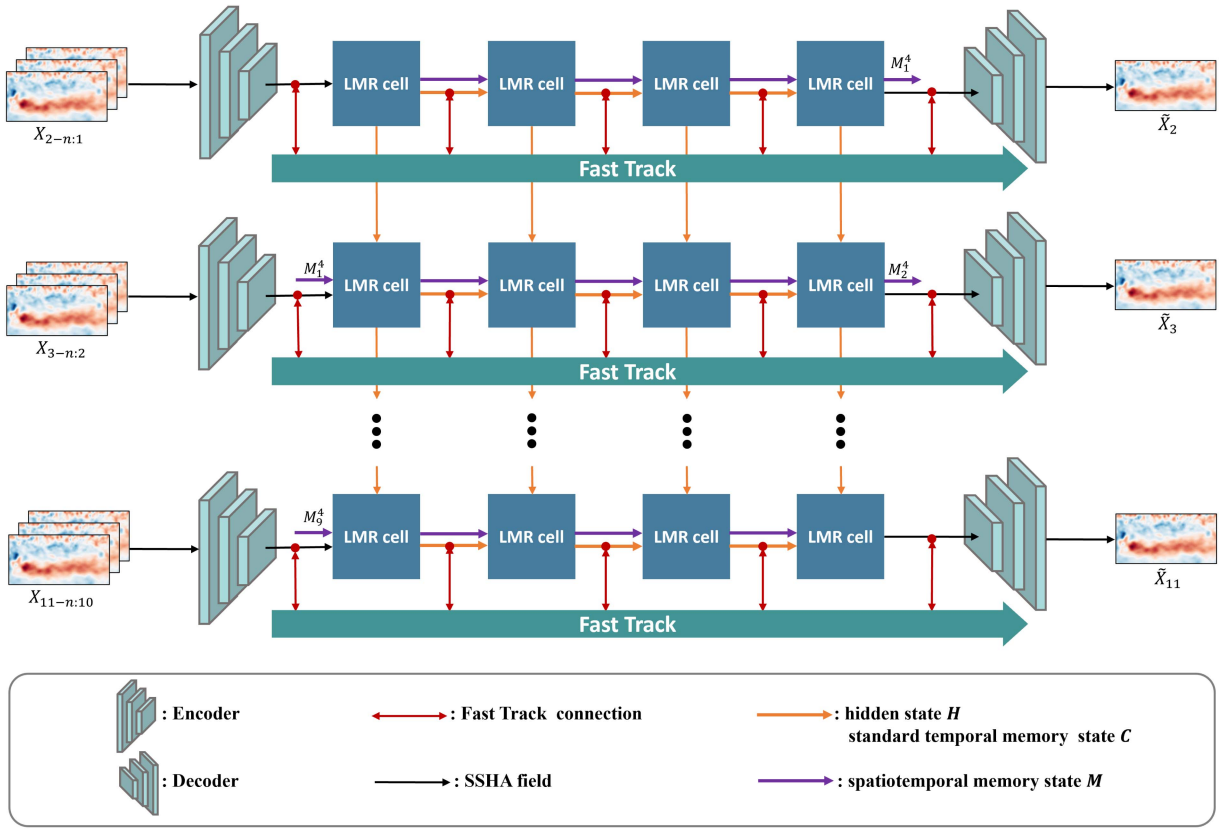


Fig. 2. *Proposed GMR-Net model for mid-term SSHA forecasting.* At each time step in the model, the SSHA maps of n consecutive moments are first input into the encoder to obtain high-dimension features. The obtained convolutional features pass through the recurrent structure of the four-layer LMR cells and then pass through the decoder to obtain the SSHA forecasting of the next moment. Meanwhile, the Fast Track mechanism is connected with four-layer LMR cells, which provide another fast path for the motion context information. The orange arrows indicate the direction of information flow for the hidden state $\mathcal{H}$ and standard temporal memory state $\mathcal{C}$. The purple arrows denote the direction of information flow for the spatiotemporal memory state $\mathcal{M}$. The red arrows represent the transfer of information flow between the Fast Track mechanism and LMR cells.

To alleviate the challenge of cross-layer motion information attenuation, we proposed the Fast Track mechanism in the network (see Fig. 2), which provides a rapid alternative pathway for the motion context information. To further improve the information flow between layers, we proposed a transfer pattern, that is, transferring the information of any layer to all subsequent layers. For each layer, fast track concatenates the hidden states of the previous layers and passes them to the current layer as input. Specifically, the lth layer receives the hidden states of all preceding layers, $\mathcal{H}_t^0, \mathcal{H}_t^1, \cdots, \mathcal{H}_t^{l-1}$, as input. Then, the hidden state output by the current layer is transmitted to all subsequent layers by the Fast Track mechanism.

To ensure maximum information flow between layers in the network, we regulate the information flow by a gating mechanism learned in the LMR cell. Due to this gating mechanism, the valuable information of all preceding layers can flow across several layers without attenuation. Specifically, we reuse the output gate O_t in the LMR cell to expose the desired unchanging content information. The output gate O_t determines which part of the cell states needs to be output in the LMR cell. We refer to $1 - O_t$ as the carry gate, indicating how much output to produce by carrying previous hidden state information. This Fast Track mechanism provides additional details to the current hidden states through the carry gate and balances the invariant and varying motion parts. When $O_t = 1$, no extra information is provided between layers in the network. When $O_t \in (0, 1)$, the previous hidden states are used to update the current hidden state. The fast track equation is shown as follows:

$$\mathcal{H}_t^l = \mathcal{H}_t^l + (1 - O_t) \odot W_{1\times1\times1}[\mathcal{H}_t^0, \mathcal{H}_t^1, \cdots, \mathcal{H}_t^{l-1}] \quad (2)$$

where $\odot$ is the Hadamard product. $[\mathcal{H}_t^0, \mathcal{H}_t^1, \cdots, \mathcal{H}_t^{l-1}]$ refers to the concatenation of the hidden states produced in all preceding layers $0, 1, \cdots, l-1$. $\mathcal{H}_t^l$ is the hidden state of the current layer l. O_t is the output gate of the LMR cell in layer l at time t. $W_{1\times1\times1}$ is the $1 \times 1 \times 1$ convolution. The dimensionality of $\mathcal{H}_t^l$, O_t, and $W_{1\times1\times1}[\mathcal{H}_t^0, \mathcal{H}_t^1, \cdots, \mathcal{H}_t^{l-1}]$ must be the same for (2) to be valid.

B. Long-Term Memory Recall Cell

The forecasting of future time steps can be enhanced by pertinent features from previous time steps, which motivates us to design LMR cell with memory capability. The LMR cell introduces MSR module with memory function based on ConvLSTM, as illustrated in Fig. 3. The MSR module is designed to capture mid- and long-term dependencies in the time domain, and the details of the storage and recall of temporal memory states in the MSR module are described in Section III-C.

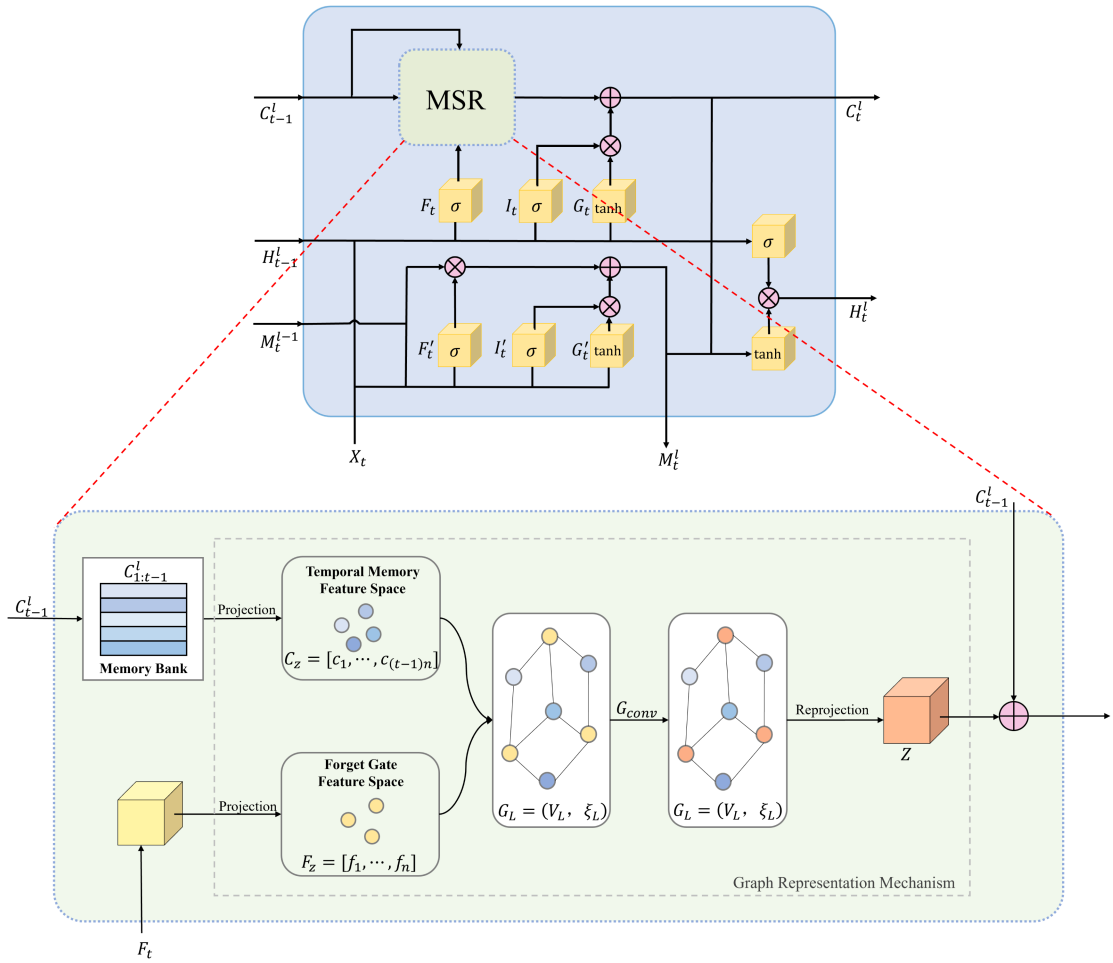


Fig. 3. *Long-term Memory Recall (LMR) cell.* To alleviate the deficiency of ConvLSTM in effectively capturing mid- and long-term temporal dependencies, we introduced an MSR module endowed with memory functionality. The MSR module first stores the standard temporal memory states of the previous time steps through the memory bank, and then recalls the mid- and long-term memory context features from the memory bank through the graph representation mechanism.

At time step t in layer l, the LMR cell receives four inputs: $\mathcal{X}_t$, the 3-D-Conv feature map from encoder or hidden state of the previous layer; $\mathcal{H}_{t-1}^l$, the hidden state from the previous timestamp; $\mathcal{C}_{t-1}^l$, the standard temporal memory state from the previous timestamp; and $\mathcal{M}_t^{l-1}$, the spatiotemporal memory state of the previous layer. The LMR cell generates three outputs: $\mathcal{H}_t^l$, the hidden state from the current timestamp; $\mathcal{C}_t^l$, the standard temporal memory state from the current timestamp; and $\mathcal{M}_t^l$, the spatiotemporal memory state of the current layer from the current timestamp.

In addition, unlike the 2-D convolution operation in ConvLSTM, the LMR cell integrates 3-D convolution into recurrent state transitions to mine the spatiotemporal features for complex SSHA sequences. The overall equations of the LMR cell are shown as follows:

$$I_t = \sigma(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1}^l + b_i) \tag{3}$$

$$G_t = \tanh(W_{xg} * \mathcal{X}_t + W_{hg} * \mathcal{H}_{t-1}^l + b_g) \tag{4}$$

$$F_t = \sigma(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1}^l + b_f) \tag{5}$$

$$\mathcal{C}_t^l = I_t \odot G_t + \mathrm{MSR}(F_t, \mathcal{C}_{t-1}^l) \tag{6}$$

$$I_t' = \sigma(W_{xi}' * \mathcal{X}_t + W_{mi} * \mathcal{M}_t^{l-1} + b_i') \tag{7}$$

$$G_t' = \tanh(W_{xg}' * \mathcal{X}_t + W_{mg} * \mathcal{M}_t^{l-1} + b_g') \tag{8}$$

$$F_t' = \sigma(W_{xf}' * \mathcal{X}_t + W_{mf} * \mathcal{M}_t^{l-1} + b_f') \tag{9}$$

$$\mathcal{M}_t^l = I_t' \odot G_t' + F_t' \odot \mathcal{M}_t^{l-1} \tag{10}$$

$$O_t = \sigma(W_{xo} * \mathcal{X}_t + W_{ho} * \mathcal{H}_{t-1}^l + W_{co} * \mathcal{C}_t^l + W_{mo} * \mathcal{M}_t^l + b_o) \tag{11}$$

$$\mathcal{H}_t^l = O_t \odot \tanh(W_{1\times1\times1} * [\mathcal{C}_t^l, \mathcal{M}_t^l]) \tag{12}$$

where σ is the sigmoid function, $*$ is the 3-D-Conv operation, $W_{1\times1\times1}$ is the $1 \times 1 \times 1$ convolution for the transformation of the channel number to match the dimensions. MSR$(\cdot)$ represents the operations in our proposed MSR module. The use of the input gate I_t, input-modulation gate G_t, forget gate F_t, and output gate O_t control temporal information flow across the memory state $\mathcal{C}_t^l$. The use of the input gate I_t', input-modulation gate G_t', forget gate F_t', and output gate O_t control spatial information flow across the memory state $\mathcal{M}_t^l$. At last, the final hidden state $\mathcal{H}_t^l$ of this node relies on the fused spatiotemporal memory.

C. Memory Store Recall Module

The bottom half of Fig. 3 shows the structure of the proposed MSR module, which comprises two primary parts: 1) memory bank and 2) graph representation mechanism. At time step t in layer l, two inputs will be fed into the MSR: the standard temporal memory state $\mathcal{C}_{t-1}^{l}$ from previous time step and the forget gate feature F_t from the current time step.

1) Memory Bank: The SSHA value at the current moment is typically related to the SSHA at multiple moments in the past. To provide mid- and long-term dependency information for predicting future SSHA, we utilize a memory bank to efficiently store the high-dimensional features of SSHA from the previous time steps, which are standard temporal memory states in LMR cells. The memory bank is a module to store previous information, which is widely applied in temporal sequential signal processing [34], [35]. As shown in Fig. 3, for time step t in layer l, the standard temporal memory state $\mathcal{C}_{t-1}^{l} \in \mathbb{R}^{n\times H\times W\times C}$ is entered into the MSR and is concatenated along the time dimension with the memory states stored in the memory bank at the previous time steps to form a sequence tensor, which is stored in the memory bank. As a result, all the standard temporal memory states stored in the memory bank constitute a sequence tensor $\mathcal{C}_{1:t-1}^{l} \in \mathbb{R}^{(t-1)n\times H\times W\times C}$ indicating that there are $t-1$ historical memory states that are attended.

2) Graph Representation Mechanism: In order to learn the importance of different historical temporal memory states in the memory bank, we designed a recall method called graph representation mechanism. The graph representation mechanism is based on a novel graph structure and a flexible graph affinity representation, aiming to achieve fast message updates and capture mid- and long-term temporal dependencies of SSHA.

In the graph representation mechanism, the forget gate feature F_t from the current time step and standard temporal memory states $\mathcal{C}_{1:t-1}^{l}$ stored in memory bank are used to recall mid- and long-term memory features. The original forget gate feature $F_t \in \mathbb{R}^{n\times H\times W\times C}$ and stored standard temporal memory states $\mathcal{C}_{1:t-1}^{l} \in \mathbb{R}^{(t-1)n\times H\times W\times C}$ are, respectively, projected into $F_t \in \mathbb{R}^{n\times m}$ and $\mathcal{C}_{1:t-1}^{l} \in \mathbb{R}^{(t-1)n\times m}$, where $m = H\times W\times C$. Let $\mathcal{C}_z = [c_1, \cdots, c_{(t-1)n}]$ represent standard temporal memory features after projection, where $c_i \in \mathbb{R}^{1\times m}$ is an m-channel temporal memory feature vector and i indicates the spatial location of the feature vector. At the same time, let $F_z = [f_1, \cdots, f_n]$ represent a set of forget gate features after projection, where $f_i \in \mathbb{R}^{1\times m}$ is an m-channel forget gate feature vector.

We then use each temporal memory feature vector c_i and forget gate feature vector f_i as a node to build an augmented graph structure $G_L = (V_L, \xi_L)$ with $(t-1)n+n$ nodes, where $V_L = V_c \cup V_f$. Here, V_c represents the nodes aligned with the temporal memory features and V_f denotes the nodes correlated with the forget gate features. Furthermore, we also define the graph connectivity by graph edges ξ_L which denote the connections between V_c and V_f. In particular, every node in V_c is connected to every node in V_f and nodes of the same type are not connected.

We make use of graph convolution G_{conv} to propagate information on the graph structure. Specifically, we update the forget gate features by transmitting messages from the temporal memory features in the V_c nodes to V_f nodes. The updated forget gate features can capture the mid- and long-term memory information from the historical temporal memory states, thereby more accurately deciding which information is forgotten and which important information should be retained. We construct the graph affinity matrix by considering the similarity of neighboring node features. Constructing the adjacency matrix in this manner promotes stronger connectivity among similar nodes, facilitating efficient information transfer and updates within the graph structure. The message passing process is shown as follows:

$$z_i = \sum_{j=1}^{(t-1)n} \frac{\exp(d(f_i, c_j))}{\sum_{k=1}^{(t-1)n}\exp(d(f_i, c_k))} W_z^T c_j^T \quad 1 \le i \le n \tag{13}$$

$$d(f_i, c_j) = f_i c_j^T \tag{14}$$

where z_i represents the updated feature representation at node in V_f. $d(\cdot,\cdot)$ computes similarity scores between two feature vectors f_i and c_j. $\exp(\cdot)/\sum\exp(\cdot)$ denotes softmax function which is used to aggregate features along each row. $W_z \in \mathbb{R}^{m\times m}$ is the weight matrix defining a linear mapping to encode the information from nodes in V_c. It is worth noting that we can express the updating (13) in a matrix form as follows:

$$Z = A\left(F_z\ \mathcal{C}_z\right)\mathcal{C}_z W_z \tag{15}$$

where $A(F_z, \mathcal{C}_z) \in \mathbb{R}^{n\times(t-1)n}$ represents the affinity matrix of the graph in which $A_{i,j} = \frac{\exp(d(f_i,c_j))}{\sum_{k=1}^{(t-1)n}\exp(d(f_i,c_k))}$. Finally, a mid- and long-term memory feature is obtained by reprojecting $Z \in \mathbb{R}^{n\times m}$ into $\mathbb{R}^{n\times H\times W\times C}$.

Given the mid- and long-term memory feature Z, we employ a straightforward feature augmentation approach to integrate the memory representation with the original temporal memory features. Specifically, we employ a residual connection, followed by layer normalization to obtain the final mid- and long-term memory

$$\text{MSR}(F_t, \mathcal{C}_{t-1}^{l}) = \text{LayerNorm}(\mathcal{C}_{t-1}^{l} + Z). \tag{16}$$

IV. Experiments

First, the study area, dataset, model setting, and implementation details are described. Second, our proposed GMR-Net model is compared with existing methods on two study areas, and the prediction performance of the model for mid-term SSHA is demonstrated. Third, the effects of LMR cells, the Fast Track mechanism, and the number of stacked LMR cell layers on the mid-term SSHA forecasting are discussed. Finally, the successful observations of Rossby waves from the model-predicted SSHA data are analyzed.

A. Study Area and Dataset

The Pacific Ocean is the largest ocean on Earth, covering over one-third of the Earth's surface. It is one of the most crucial sea trade routes in the world and an essential component of the global climate system. The sea-level changes in the Pacific Ocean have

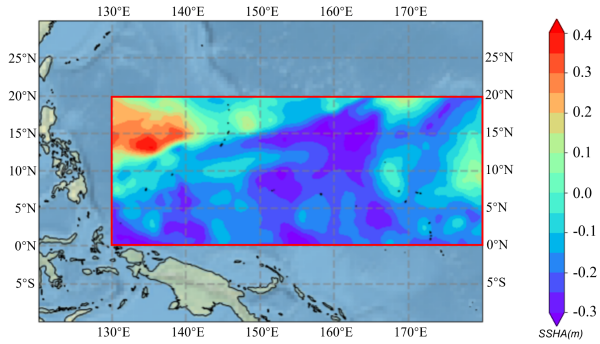


Fig. 4. Study area located in the Tropical Western Pacific (TWP) (boxed in red).

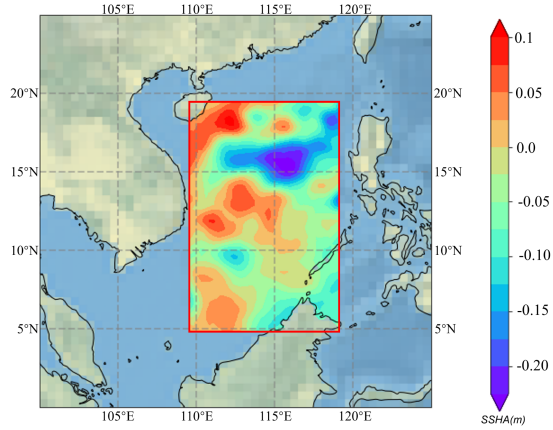


Fig. 5. Study area located in the South China Sea (SCS) (boxed in red).

far-reaching impacts on global climate and environment [36]. Therefore, mid-term SSHA forecasting in the Pacific Ocean is of great significance for predicting and mitigating climate change. In this article, two subregions of the Pacific Ocean are selected as the study areas, and experiments are carried out on these two study areas to verify the proposed GMR-Net model.

1) Tropical Western Pacific (TWP): One of the study areas is located in the TWP. The TWP is one of the regions with complex circulation systems in the world, which has a significant influence on global ocean and climate systems [37], [38], [39]. Furthermore, Rossby waves in this region exert a substantial impact on sea level [40], [41], [42], which in turn affects global climate change and ocean phenomena movements [43], [44], [45]. Therefore, studying the fluctuating properties of SSHA on TWP provides an opportunity to verify the ability of the proposed GMR-Net model to predict SSHA and observe Rossby waves. The study area was confined to a particular subregion of the TWP, which spans from 0.125°N to 19.875°N and from 130.125°E to 179.875°E (see Fig. 4). This area will allow us to better understand the dynamics of SSHA and Rossby waves on TWP.

The SSHA data on TWP provide a gridded daily global estimate of SSHA based on satellite altimetry measurements, which are produced by archiving, validating, and interpreting satellite oceanographic and distributed by the Copernicus Marine Environment Monitoring Service (CMEMS). The data encompass the time period from January 1, 1993, to December 31, 2020, and consist of daily average SSHA data within the area of 0.125 °N −19.875° N, 130.125°E −179.875° E, with a spatial resolution of 1/4°latitude by 1/4°longitude. In the experiment, the data are divided into a training set, consisting of SSHA fields from January 1, 1993, to July 31, 2012, and a test set, comprising SSHA fields from August 1, 2012, to December 31, 2020. For the training set, a sliding window with a width of 15 frames is used to slice the contiguous images, and the time interval between the frames is one day. Therefore, each sequence comprises 15 consecutive frames, with 10 input frames and 5 prediction frames. For the test set, a sliding window with a width of 40 frames is used to slice the successive images. Therefore, each sequence comprises 40 consecutive frames, with 10 input frames and 30 prediction frames. In total, there are 10 174 sequences, with 7138 sequences used for training and 3036 for testing.

2) South China Sea (SCS): Another study area is located in the SCS. The SCS is the largest marginal sea in the Pacific Ocean. Understanding the patterns of SSHA and predicting future SSHA in this area is not only of scientific significance but also of extensive practical value to offshore economic, military, engineering, and other fields. In addition, due to the influence of monsoons, tidal currents, topography, and other factors, mesoscale eddies and storm surges occur frequently in this area, which has a strong influence on the sea surface [46], [47], [48]. Therefore, the fluctuating characteristics of the SSHA on SCS make it suitable for testing the prediction performance of the proposed model. In this study, we choose a subarea of the SCS as our study area, which covers the spatial extent of 4.875°N −19.625° N and 109.875°E −119.625° E (see Fig. 5).

The SSHA data on SCS also come from CMEMS, which provides the gridded daily global estimate of SSHA covering the spatial region of 0.125°N −25.125° N, 100.125°E −125.125° E. The spatial coverage of the data is larger than our selected study area, which provides sufficient spatial features for accurate SSHA forecasting within the study area. In this experiment, the training set contains SSHA data from January 1, 2001 to May 1, 2016, and the test set contains SSHA data from May 2, 2016 to May 13, 2019. We follow the same data processing as the data on TWP, resulting in 6654 sequences, of which 5586 are used for training and 1068 for testing.

B. Model Setting

We represented the daily observed SSHA field by a 3-D tensor $\mathcal{X}_t$ of size $H \times W \times C$. The sequence $\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_T$ formed by the observations of T time steps is utilized as the input sequence of the model. At each time step t, n consecutive tensors from the input sequence are fed into the model. In our experiment, we set $n = 2, H = 80, W = 200, C = 1$, and $T = 10$. We applied a normalization step to the tensors before feeding them into the proposed model to ensure that their values fell within the range of [0, 1]. This normalization step helped to center the data, resulting in faster model training and convergence.

At the beginning of the training phase, we set the initial state values of the hidden state $\mathcal{H}^n_{t=0}$, standard temporal memory state $\mathcal{C}^n_{t=0}$, and spatiotemporal memory state $\mathcal{M}^n_{t=0}$ to 0. The 3-D-Conv operators integrated into the LMR cells have a kernel

size of $2 \times 3 \times 3$ (time $\times$ height $\times$ width). The number of hidden state channels of each LMR cell is 64. To extract features from the input frames, we used a single 3-D-Conv layer as the encoder. Similarly, a single 3-D-Conv layer was used as the decoder to map motion-aware hidden states to output frames.

C. Implementation Details

During the training phase, all models were trained to forecast the subsequent 5 frames, utilizing the first 10 frames as input. Choosing a specific number of frames as input and output is not the only option, and it can be adjusted based on the specific research and dataset. Training the model with 10 frames of SSHA images as input is aimed at providing sufficient historical data to capture patterns and trends in the spatiotemporal sequences. By leveraging 10 frames as input, the model can learn more contextual information and gain a better understanding of how past states influence the future. Although increasing the number of input time steps can slightly improve performance, it is important to note that the influence of historical frames on the forecasted frames diminishes as the forecast interval increases. Therefore, adding too many historical frames does not result in a significant improvement in prediction accuracy. In addition, using longer input sequences also increases computational complexity. Selecting 10 frames as input can balance the relationship between prediction accuracy and computational efficiency.

During the testing phase, all models were applied recursively. At the first recursive step, each model generated a prediction for the subsequent 5 frames based on the first 10 frames of input data. At the second recursive step, we used the predicted 5 frames from the first step and the previous 5 frames as input to generate a prediction for the next 5 frames. Similarly, in the third recursive step, we used the predicted 10 frames from the first two steps to generate a prediction for the next 5 frames. Recursively, we continued to make the SSHA forecasting at the fourth, fifth, sixth recursive steps, until future 30 frames were predicted. Therefore, all models were tested to predict the subsequent 30 frames based on the first 10 frames of input data.

The training process terminated after completing 80 000 iterations, with mse as the loss function and Adam [49] as the optimizer. The minibatch size of each iteration was set to 8. The base learning rate was initially set at 0.0001 and reduced by a factor of 0.9 every 2500 iterations [50]. The model was trained on a single NVIDIA GeForce RTX 3090 GPU utilizing the PyTorch library.

D. Compared Methods and Evaluation Metrics

The proposed GMR-Net model was compared with three types of advanced prediction models. The first category is RNN-based spatiotemporal prediction algorithms, including ConvLSTM [33], MAU [51], and MotionRNN [52]. The second category contains two prediction algorithms for predicting other ocean elements, such as sea surface temperature, which are SA-JSTN [53] and EEMD-GRU [54]. The third category is prediction algorithms for SSHA forecasting, including MLFrnn [22] and RF [14].

1) ConvLSTM [33]: This method incorporates the convolutional structure into LSTM, which overcomes the limitations of LSTM in handling spatiotemporal data. Unlike our GMR-Net model that employs an MSR module to facilitate information retention, the cell states in ConvLSTM network are refreshed immediately at each time step.

2) MAU [51]: This model improves spatiotemporal prediction performance by enlarging the temporal receptive field of prediction cells. Different from our GMR-Net model which focuses on capturing long-distance information, MAU pays more attention to capturing reliable interframe motion information.

3) MotionRNN [52]: This method achieves prediction by decomposing motion into global motion trends and instantaneous changes, with a focus on short-term prediction of complex spatiotemporal motions.

4) SA-JSTN [53]: An improved version of ConvLSTM has been developed for temperature prediction. Compared with ConvLSTM, it adds a self-attention module to increase the ability of the model to capture spatial information.

5) EEMD-GRU [54]: A hybrid model of EMD and gated recurrent unit (GRU) is proposed as a prediction method for sea surface temperature anomaly (SSTA). In contrast to our model that directly predicts the SSHA map for a specific region, it focuses on predicting the average SSTA values of the entire region.

6) MLFrnn [22]: This method achieves spatiotemporal prediction of the SSHA field by fusion cells to fully integrate spatiotemporal features at each time step.

7) RF [14]: A powerful predictor is constructed from a group of decision trees, similar to EEMD-GRU, which also enables mid-term prediction of area-averaged SSHA values.

In the experiment, we apply EEMD-GRU and RF to separately predict the SSHA value of each grid point in the whole study area. The other comparison models keep the same experimental settings as our GMR-Net model.

In order to assess the performance of the model, we utilized two commonly employed evaluation metrics, namely, the root mean square error (RMSE) and the mean absolute error (MAE). The formulas of RMSE and MAE are, respectively, shown as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - x_i)^2}{N}} \tag{17}$$

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - x_i| \tag{18}$$

where x_i is the actual observed value at the ith sample, y_i is the predicted value at the corresponding sample, and N denotes the total number of test samples. In (17), RMSE is highly sensitive to large or small errors and can effectively reflect the precision of measurement. In (18), MAE offers a more accurate reflection of the predicted value error. It is worth noting that lower values of both RMSE and MAE correspond to better prediction performance.

E. Experiment Results on TWP

1) Comparison of Daily and Average RMSE and MAE: To compare the performance of different models in forecasting the SSHA field, we evaluated the daily RMSE and MAE, as

TABLE I
DAILY AND AVERAGE RMSE COMPARISON RESULTS OF DIFFERENT METHODS DURING THE 30-D FORECAST PERIOD ON TWP

MODEL	1	7	13	19	25	30	Average
ConvLSTM	0.0037±0.00037	0.020±0.0024	0.037±0.0062	0.054±0.0108	0.070±0.0156	0.083±0.0193	0.044±0.0085
MAU (2021)	0.0032±0.00017	0.017±0.0015	0.030±0.0035	0.040±0.0053	0.049±0.0070	0.056±0.0083	0.032±0.0039
MotionRNN (2021)	**0.0030±0.00016**	0.015±0.0014	0.026±0.0035	0.036±0.0066	0.046±0.0109	0.055±0.0146	0.030±0.0051
SA-JSTN (2021)	0.0032±0.00024	0.015±0.0015	0.027±0.0039	0.044±0.0121	0.088±0.0423	0.156±0.0837	0.049±0.0162
EEMD-GRU (2022)	0.0043±0.00026	0.024±0.0022	0.041±0.0049	0.052±0.0070	0.059±0.0085	0.063±0.0094	0.041±0.0051
MLFrnn (2022)	0.0038±0.00018	0.015±0.0015	0.028±0.0046	0.037±0.0073	0.058±0.0083	0.054±0.0102	0.033±0.0050
RF (2023)	0.0044±0.00018	0.026±0.0020	0.045±0.0045	0.057±0.0066	0.065±0.0081	0.069±0.0090	0.045±0.0048
GMR-Net (proposed)	0.0030±0.00017	**0.014±0.0014**	**0.024±0.0031**	**0.032±0.0046**	**0.038±0.0058**	**0.043±0.0066**	**0.026±0.0032**

The bold entities represent the best prediction results.

TABLE II
DAILY AND AVERAGE MAE COMPARISON RESULTS OF DIFFERENT METHODS DURING THE 30-D FORECAST PERIOD ON TWP

MODEL	1	7	13	19	25	30	Average
ConvLSTM	0.0029±0.00033	0.015±0.0020	0.030±0.0054	0.044±0.0097	0.058±0.0141	0.069±0.0177	0.036±0.0076
MAU (2021)	0.0025±0.00012	0.013±0.0011	0.023±0.0027	0.031±0.0042	0.038±0.0057	0.043±0.0070	0.025±0.0031
MotionRNN (2021)	**0.0023±0.00011**	0.011±0.0010	0.020±0.0027	0.028±0.0057	0.037±0.0099	0.043±0.0136	0.023±0.0045
SA-JSTN (2021)	0.0025±0.00019	0.011±0.0010	0.021±0.0032	0.036±0.0120	0.080±0.0436	0.148±0.0858	0.043±0.0164
EEMD-GRU (2022)	0.0033±0.00019	0.019±0.0016	0.031±0.0035	0.039±0.0052	0.045±0.0064	0.048±0.0073	0.031±0.0038
MLFrnn (2022)	0.0031±0.00013	0.012±0.0011	0.022±0.0043	0.028±0.0065	0.051±0.0075	0.046±0.0104	0.027±0.0049
RF (2023)	0.0035±0.00013	0.020±0.0015	0.035±0.0033	0.044±0.0049	0.049±0.0061	0.053±0.0069	0.035±0.0036
GMR-Net (proposed)	0.0023±0.00012	**0.011±0.0009**	**0.018±0.0023**	**0.024±0.0036**	**0.029±0.0046**	**0.033±0.0053**	**0.020±0.0025**

The bold entities represent the best prediction results.

well as the average RMSE and MAE, during the 30-d forecast period. Table I shows RMSE comparison results with different methods on TWP, where the prediction results are in the form of mean plus standard deviation . Our proposed GMR-Net model exhibited superior performance in terms of daily RMSE and average RMSE compared to other models. Specifically, GMR-Net is much better than ConvLSTM, which illustrates the effectiveness of our proposed MSR module with memory storage function for mid-term SSHA forecasting. MAU focuses on interframe motion information and achieves better prediction results than ConvLSTM, but still cannot outperform our GMR-Net model. This demonstrates the importance of long-range motion information in mid-term SSHA forecasting. MotionRNN focuses on complex spatiotemporal motion and further improves prediction performance. But its prediction performance is still inferior to GMR-Net, which once again indicates that long-term motion information cannot be ignored in mid-term SSHA forecasting. The performance of the SA-JSTN model is acceptable for short-term forecasting, but it sharply declines for mid-term forecasting. This may be because mid-term SSHA forecasting needs to consider longer-term impacts and changes, while the SA-JSTN model mainly focuses on spatial dynamics, making it difficult to effectively capture complex temporal trends. Our model outperforms SA-JSTN, which indicates that focusing on mid- and long-term information in the temporal dimension is more important than focusing on global spatial information in mid-term SSHA forecasting. MLFrnn exhibits extremely unstable prediction performance in mid-term SSHA forecasting, probably because it does not consider the complex spatiotemporal dynamic relationship in longer time spans. The overall prediction results of EEMD-GRU and RF are slightly worse, indicating the spatiotemporal prediction of SSHA needs to capture both temporal and spatial information. The prediction of a single site cannot capture the spatial information of the SSHA field, resulting in poor prediction results. Table II shows MAE comparison results with different methods on TWP, which exhibit similar results to those observed for RMSE in Table I.

2) Comparison of Frame-by-Frame RMSE and MAE: The frame-by-frame quantitative comparison of RMSE and MAE for different models is shown in Fig. 6. A lower value corresponds to superior prediction performance. Fig. 6(a) and (b), respectively, show the RMSE and MAE comparison on TWP. From Fig. 6(a) and (b), we can see that our proposed GMR-Net model consistently outperforms other models. The prediction performance of SA-JSTN in the first 18 d is relatively good, but as the prediction lead increases, its prediction error accumulates sharply, which makes its performance unacceptable when predicting a longer time span. As the forecast horizon extends beyond the 18th d, the prediction performance of the MLFrnn model is extremely unstable. Owing to the cumulative effect of errors, the prediction performance of all models typically exhibits a decline as the number of predicted frames increases. Our model accumulated errors at a slower rate compared to other models, highlighting its superior and stable mid-term prediction performance.

3) Comparison of Spatial Distributions of RMSE: The spatial distributions of RMSE for different methods were calculated on a grid-by-grid basis using the forecasting errors of all samples during the testing period. Fig. 7 shows the RMSE spatial distribution of different methods on TWP. It can be intuitively found that the discrepancy between the predicted results generated by our proposed GMR-Net model and the observed values is very small and acceptable. Compared with other methods, the GMR-Net model yields prediction results with the best visual qualities at each frame. The prediction results of EEMD-GRU and RF exhibit a considerable number of noisy points, which may be due to the lack of spatiotemporal feature modeling because they make predictions for individual sites separately. SA-JSTN and ConvLSTM perform adequately in short-term forecasting but exhibit poor results in mid-term forecasting. This may be due to the inability of the model to capture mid- and long-term temporal dependencies. Although the MAU and MotionRNN frameworks generate more precise prediction outcomes, their capacity to accurately predict SSHA of edge locations and certain details is limited, likely due to the loss of mid- and long-term motion

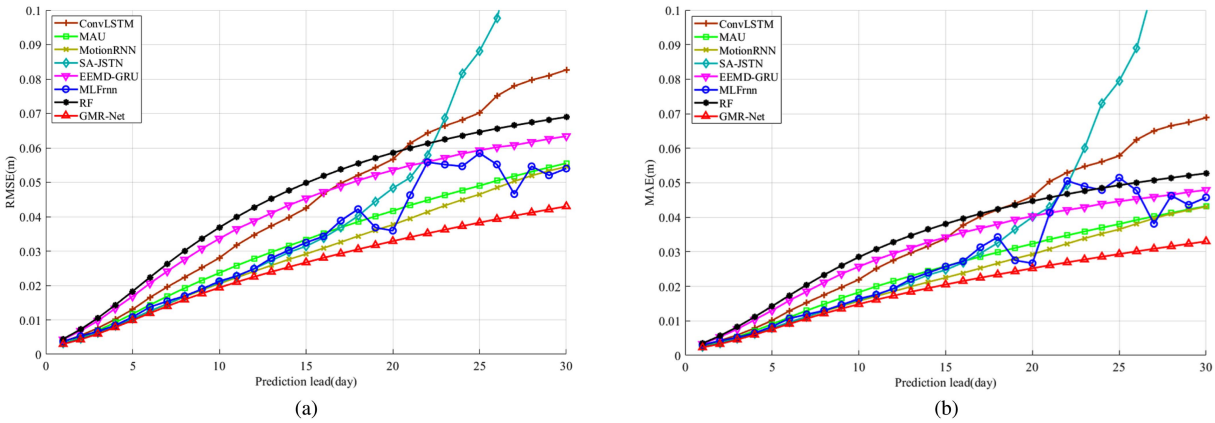


Fig. 6. *Framewise RMSE and MAE comparisons of different models during the 30-d forecast period on TWP.* (a) Framewise RMSE comparisons of different models. (b) Framewise MAE comparisons of different models. The lower the value of RMSE and MAE, the better the prediction accuracy. Compared with other models, GMR-Net has the lowest RMSE and MAE on every prediction lead.

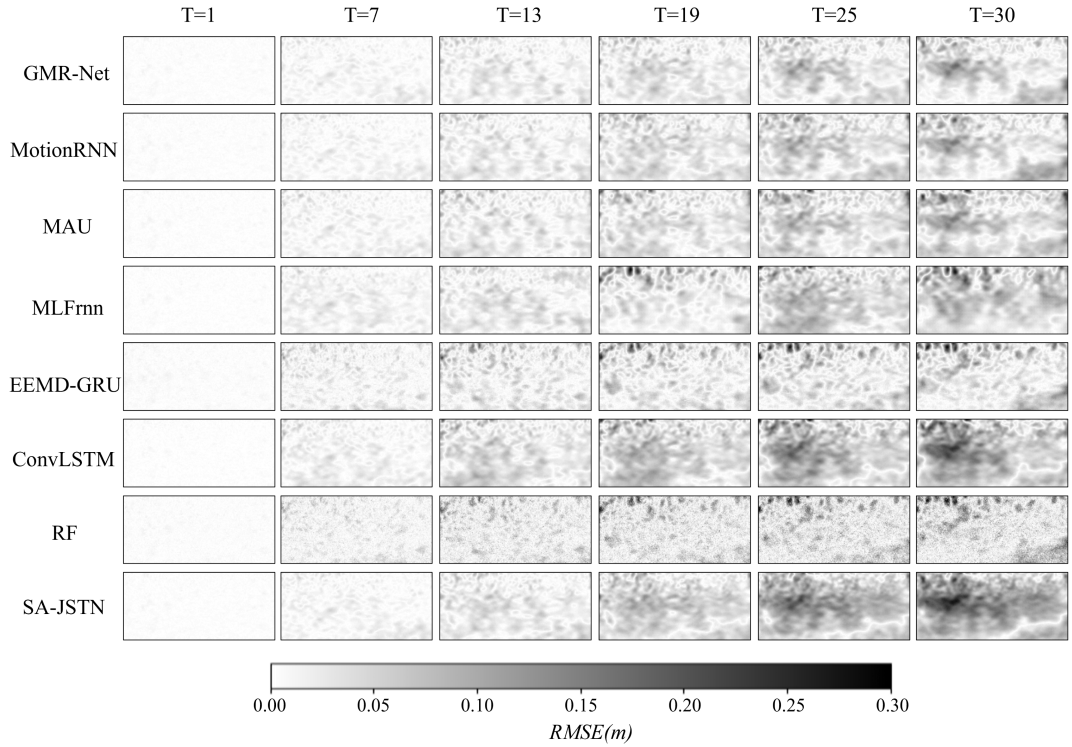


Fig. 7. *Spatial distributions of RMSE for different methods during the 30-d forecast period on TWP.* The spatial distributions of RMSE for different methods were calculated on a grid-by-grid basis using the forecasting errors of all samples during the testing period. The lighter the color, the smaller the prediction error and the better the prediction accuracy. Compared with other methods, the GMR-Net model yields prediction results with the best visual qualities at each frame.

information during the prediction process. In short, the proposed GMR-Net model generated clearer predictions and is able to memorize detailed visual appearances.

4) Comparison of Forecast Results in Different Months: SSH is influenced by seasonal variations. To further validate the prediction performance of the GMR-Net model, we tested the prediction results of the model in different months. Fig. 8 shows the comparison of RMSE and MAE results predicted by different models in different months on TWP. It can be seen that the prediction results of other comparative models, particularly the SA-JSTN model, exhibit significant fluctuations with seasonal changes. In contrast, our proposed GMR-Net model exhibits minimal sensitivity to seasonal changes, providing evidence of its robustness and reliability in predicting SSHA for various months. In addition, it is noteworthy that our model also exhibits the best prediction performance for each month, with the smallest RMSE.

F. Experiment Results on SCS

1) Comparison of Daily and Average RMSE and MAE: We further evaluated the performance of the GMR-Net model on SCS. Table III and IV, respectively, shows RMSE and MAE

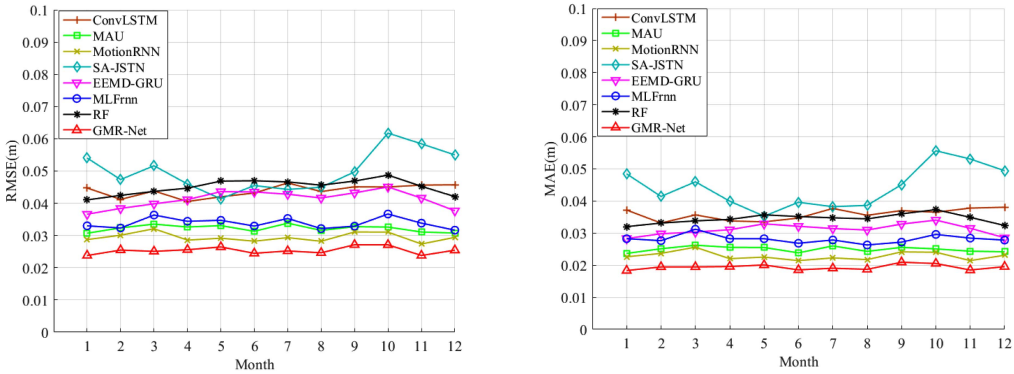


Fig. 8. *Comparison of RMSE and MAE results predicted by different models in different months on TWP.* (a) RMSE comparison in different months. (b) MAE comparison in different months. The RMSE and MAE of GMR-Net remains stable despite seasonal variations. Compared to other models, GMR-Net achieves the smallest RMSE and the best prediction accuracy in every month.

TABLE III
DAILY AND AVERAGE RMSE COMPARISON RESULTS OF DIFFERENT METHODS DURING THE 30-D FORECAST PERIOD ON SCS

MODEL	1	7	13	19	25	30	Average
ConvLSTM	0.0036±0.00046	0.019±0.0029	0.037±0.0062	0.052±0.0095	0.065±0.0122	0.075±0.0138	0.041±0.0066
MAU (2021)	0.0035±0.00046	0.018±0.0028	0.035±0.0054	0.048±0.0078	0.059±0.0100	0.067±0.0119	0.038±0.0054
MotionRNN (2021)	0.0035±0.00051	0.018±0.0027	0.033±0.0055	0.046±0.0083	0.057±0.0108	0.065±0.0128	0.037±0.0057
SA-JSTN (2021)	0.0035±0.00046	0.019±0.0027	0.036±0.0056	0.050±0.0090	0.062±0.0128	0.072±0.0162	0.040±0.0064
EEMD-GRU (2022)	0.0045±0.00052	0.022±0.0034	0.040±0.0065	0.051±0.0086	0.058±0.0099	0.062±0.0106	0.040±0.0061
RF (2023)	0.0048±0.00046	0.024±0.0031	0.043±0.0058	0.057±0.0076	0.064±0.0085	0.069±0.0089	0.044±0.0053
GMR-Net (proposed)	**0.0033±0.00044**	**0.017±0.0023**	**0.033±0.0048**	**0.045±0.0071**	**0.054±0.0086**	**0.060±0.0097**	**0.035±0.0048**

The bold entities represent the best prediction results.

TABLE IV
DAILY AND AVERAGE MAE COMPARISON RESULTS OF DIFFERENT METHODS DURING THE 30-D FORECAST PERIOD ON SCS

MODEL	1	7	13	19	25	30	Average
ConvLSTM	0.0027±0.00028	0.015±0.0022	0.028±0.0051	0.040±0.0082	0.051±0.0106	0.059±0.0120	0.032±0.0056
MAU (2021)	0.0026±0.00027	0.014±0.0021	0.027±0.0042	0.037±0.0061	0.044±0.0076	0.050±0.0088	0.029±0.0041
MotionRNN (2021)	0.0026±0.00029	0.013±0.0020	0.025±0.0042	0.035±0.0065	0.044±0.0087	0.051±0.0107	0.028±0.0044
SA-JSTN (2021)	0.0026±0.00027	0.014±0.0020	0.027±0.0046	0.038±0.0076	0.04±0.0108	0.056±0.0135	0.031±0.0054
EEMD-GRU (2022)	0.0034±0.00032	0.017±0.0025	0.030±0.0048	0.039±0.0065	0.044±0.0076	0.047±0.0082	0.030±0.0046
RF (2023)	0.0036±0.00026	0.018±0.0022	0.033±0.0043	0.043±0.0057	0.049±0.0065	0.052±0.0069	0.034±0.0040
GMR-Net (proposed)	**0.0025±0.00026**	**0.013±0.0018**	**0.025±0.0039**	**0.034±0.0057**	**0.040±0.0068**	**0.045±0.0076**	**0.026±0.0038**

The bold entities represent the best prediction results.

comparison results with different methods on SCS. The prediction performance of the MAU and MotionRNN models is not as excellent as their performance on TWP. This observation may be attributed to the intricate marine environment in the SCS area, which poses a significant challenge for models to accurately capture and predict complex spatiotemporal dynamics. The mid-term prediction performance of the SA-JSTN model on SCS is acceptable, indicating that the spatiotemporal information captured by the SA-JSTN model in the short term is helpful for mid-term forecasting. This may be because in the SCS area, it is crucial for the model to pay more attention to the capture of spatial dynamics. It is noteworthy that the GMR-Net model also exhibits superior prediction performance on SCS compared to other models. It can be concluded that the GMR-Net model can capture mid- and long-term spatiotemporal dynamic changes in complex marine environments, such as the SCS, and achieve high-precision mid-term SSHA forecasting.

2) Comparison of Frame-by-Frame RMSE and MAE: Fig. 9(a) and (b), respectively, show the frame-by-frame RMSE and MAE comparisons of different models on SCS. It can be seen that all models achieved good predictive results in the first five days. However, as the prediction lead increases, the disparities in prediction results between different models begin to widen. Specifically, as the prediction lead increases from 5 to 20 d, the prediction errors of RF and EEMD-GRU are significantly higher than other models, probably because they made predictions for each site separately. As the prediction lead increases from 25 to 30 d, ConvLSTM and SA-JSTN exhibit poor results, possibly due to the lack of relevant modules for mid- and long-term forecasting, resulting in severe error accumulation. As expected, our GMR-Net model has performed well in the prediction results of each day and has achieved the best prediction results. This indicates that our model has a good ability to capture mid- and long-term temporal dependencies, and can achieve accurate mid-term prediction performance for SSHA field.

3) Comparison of Spatial Distributions of RMSE: Fig. 10 shows the RMSE spatial distribution of different methods on SCS. Similarly, our model achieves the lowest error and the best visual quality in each frame. The EEMD-GRU and RF models yield nonuniformly distributed prediction errors and

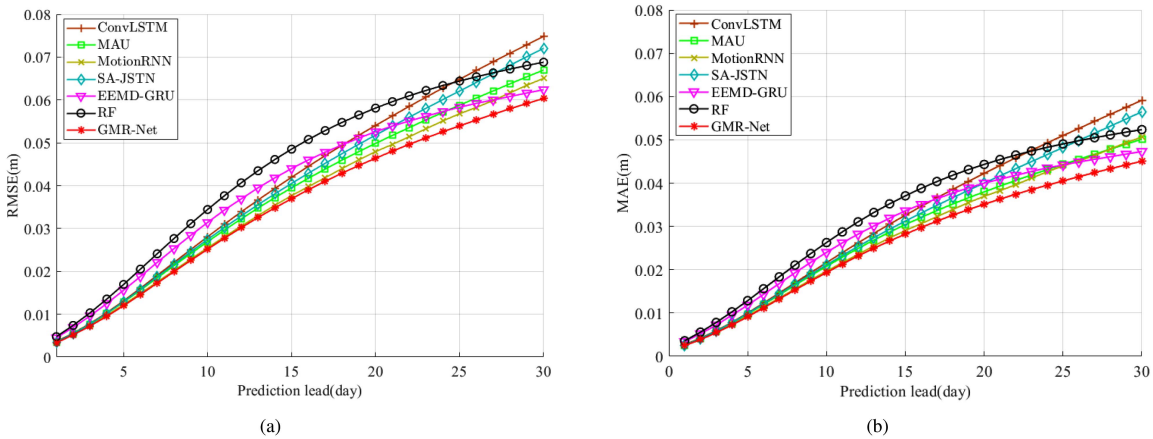


Fig. 9. *Framewise RMSE and MAE comparisons of different models during the 30-d forecast period on SCS.* (a) Framewise RMSE comparisons of different models. (b) Framewise MAE comparisons of different models. The lower the value of RMSE and MAE, the better the prediction accuracy. Compared with other models, GMR-Net has the lowest RMSE and MAE on every prediction lead.

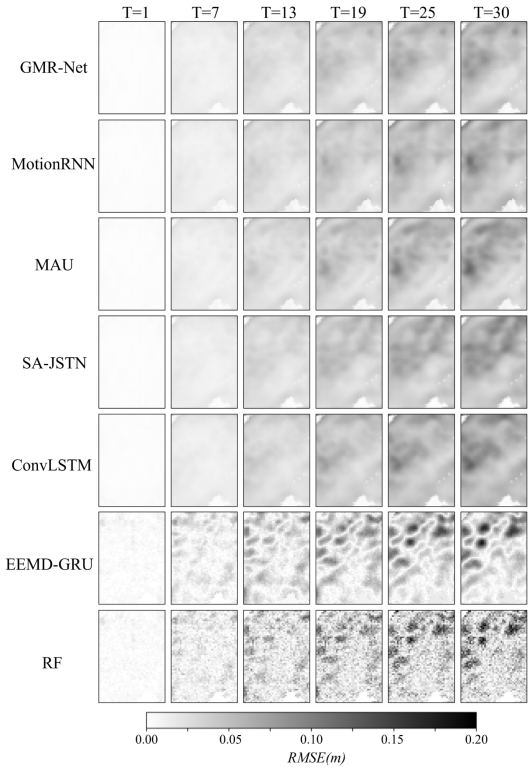


Fig. 10. *Spatial distributions of RMSE for different methods during the 30-d forecast period on SCS.* The spatial distributions of RMSE for different methods were calculated on a grid-by-grid basis using the forecasting errors of all samples during the testing period. The lighter the color, the smaller the prediction error and the better the prediction accuracy. Compared with other methods, the GMR-Net model yields prediction results with the best visual qualities at each frame.

numerous noise points. This is because they perform SSHA predictions separately for individual sites, lacking the ability to comprehensively capture spatial information. Spatiotemporal prediction models, such as ConvLSTM, MAU, MotionRNN, and SA-JSTN overcome the shortcomings of models that predict individual sites separately, but there will still be large errors in locations where sea-level fluctuations are large. Thanks to the exceptional capability of our model in capturing mid- and long-term spatiotemporal dynamics, it exhibits smaller prediction errors compared to other methods.

G. Ablation Study

1) Effectiveness of MSR Module and Fast Track Mechanism: In order to investigate the contributions of the proposed MSR module and the effect of adding the Fast Track mechanism in the network, an ablation study was conducted in this part of the experiments.

Table V shows the daily and average RMSE of the 30-d prediction results for the different models on TWP. The results presented in rows 1 and 2 show the contribution of the two essential components in the proposed GMR-Net: 1) MSR module, and 2) Fast Track mechanism. Specifically, the first baseline method was used to evaluate the effectiveness of the MSR module in capturing mid- and long-term spatiotemporal patterns, and the results indicate that this component can significantly learn mid- and long-term dependencies and improve the mid-term SSHA forecasting performance. Similarly, the contribution of the Fast Track mechanism is isolated in the second baseline method. The results indicate that Fast Track maintains the motion context information and compensates existing models for the additional useful information. The proposed GMR-Net exhibits the best prediction performance, illustrating that both components of the GMR-Net are essential and crucial for effectively modeling spatiotemporal data. Table VI shows the comparison of daily and average MAE predicted by different models 1–30 d ahead, which are consistent with the RMSEs in Table V.

2) Effect of the Number of Stacked LMR Cell Layers: Our proposed GMR-Net model employs a stacking scheme of four-layer LMR cells. To investigate the impact of different stacking configurations on the SSHA prediction performance, we vary the number of stacked LMR cell layers L from 2 to 5 while keeping other settings unchanged. Fig. 11 shows the prediction results of the GMR-Net model under different numbers of LMR cell layers on TWP. Specifically, when $L = 2$, the model shows a steep increase in prediction error as the forecast horizon exceeds 15 d, which can be attributed to the model's inadequate capacity for capturing SSHA feature information and information transmission. In contrast, when $L = 3$, the model exhibits a slower accumulation of prediction error as the prediction lead increases. This is possibly due to the higher order interactions

TABLE V
RMSE COMPARISON OF SSHA FORECASTING 1–30 D AHEAD WITH DIFFERENT MODELS ON TWP

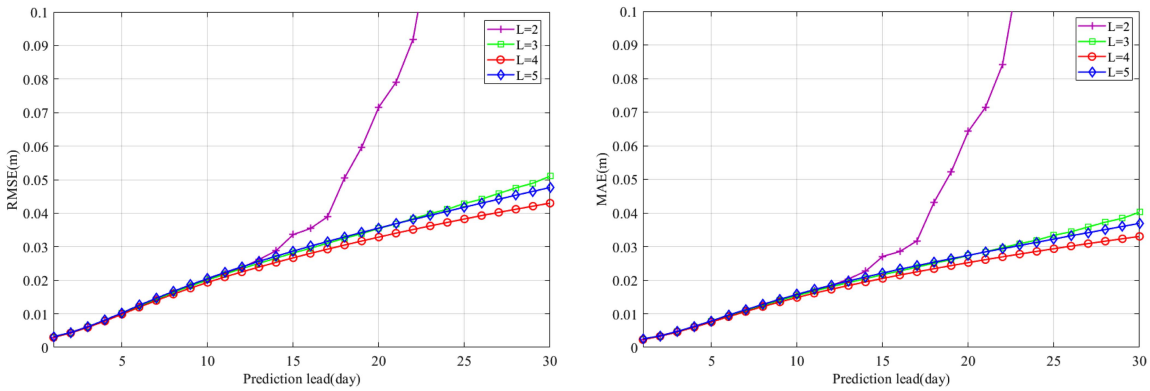
Prediction lead(day)	1	7	13	19	25	30	Average
GMR-Net (W/O MSR module)	0.0032	0.016	0.028	0.037	0.046	0.053	0.030
GMR-Net (W/O Fast Track)	0.0031	0.014	0.026	0.036	0.045	0.052	0.029
GMR-Net	**0.0030**	**0.014**	**0.024**	**0.032**	**0.038**	**0.043**	**0.026**

The bold entities represent the best prediction results.

TABLE VI
MAE COMPARISON OF SSHA FORECASTING 1–30 D AHEAD WITH DIFFERENT MODELS ON TWP

Prediction lead(day)	1	7	13	19	25	30	Average
GMR-Net (W/O MSR module)	0.0025	0.012	0.022	0.029	0.036	0.041	0.024
GMR-Net (W/O Fast Track)	0.0025	0.011	0.020	0.028	0.036	0.041	0.023
GMR-Net	**0.0023**	**0.011**	**0.018**	**0.024**	**0.029**	**0.033**	**0.020**

The bold entities represent the best prediction results.



(a) (b)

Fig. 11. *Prediction results of the GMR-Net model with different numbers of LMR cell layers on TWP.* (a) Framewise RMSE comparisons of GMR-Net model stacked with different numbers of LMR cell layers. (b) Framewise MAE comparisons of GMR-Net model stacked with different numbers of LMR cell layers. When $L = 4$, the GMR-Net model achieves the smallest RMSE and MAE.

captured by the increased number of layers, enabling the model to extract more comprehensive features from the surrounding oceanic environments. Notably, the GMR-Net model with $L = 4$ demonstrates no significant advantage in predicting the SSHA for the first 10 d, yet the slowest accumulation of prediction error and the best overall performance as the prediction lead increases. Conversely, the predictive performance of the model deteriorates as the number of LMR cell layers increases to 5, which can be attributed to the loss or confusion of the captured SSHA feature information over time as the number of LMR cell layers of the model increased. Overall, we observe that the GMR-Net model achieves the best performance when $L = 4$, as evidenced by the smallest RMSE and MAE values. Based on this observation, we select the stacking configuration of four-layer LMR cells in the final model to obtain the most accurate SSHA prediction results.

H. Case Study: Observations of Rossby Waves

Rossby et al. [55], [56] discussed the classical free Rossby waves solution in the Cartesian coordinate system. The linearized large-scale and low-frequency motion equations can be employed to obtain Rossby waves with infinitely many modes when static fluid serves as the background field [57]. The dispersion relation of Rossby waves propagating along the zonal direction is as follows:

$$c_n = -\frac{\beta}{k^2 + l^2 + \lambda_n{}^{-2}} \tag{19}$$

where c_n represents the speed of Rossby waves in the nth mode, k is the latitudinal wavenumber, l is the meridional wavenumber, λ_n is the deformation radius of Rossby waves in the nth mode, and β is the change rate of the Coriolis parameter with latitude, which is a fundamental parameter that affects the formation of Rossby waves. The lowest order mode $(n = 0)$ is the barotropic mode, and the other modes are all baroclinic Rossby wave modes. The most important one is the first-order baroclinic mode when $n = 1$. By applying the long-wave approximation $(k^2, l^2 \ll \lambda^{-2})$ to (19), the first-order baroclinic Rossby waves can be further simplified as

$$c = -\beta\lambda^2. \tag{20}$$

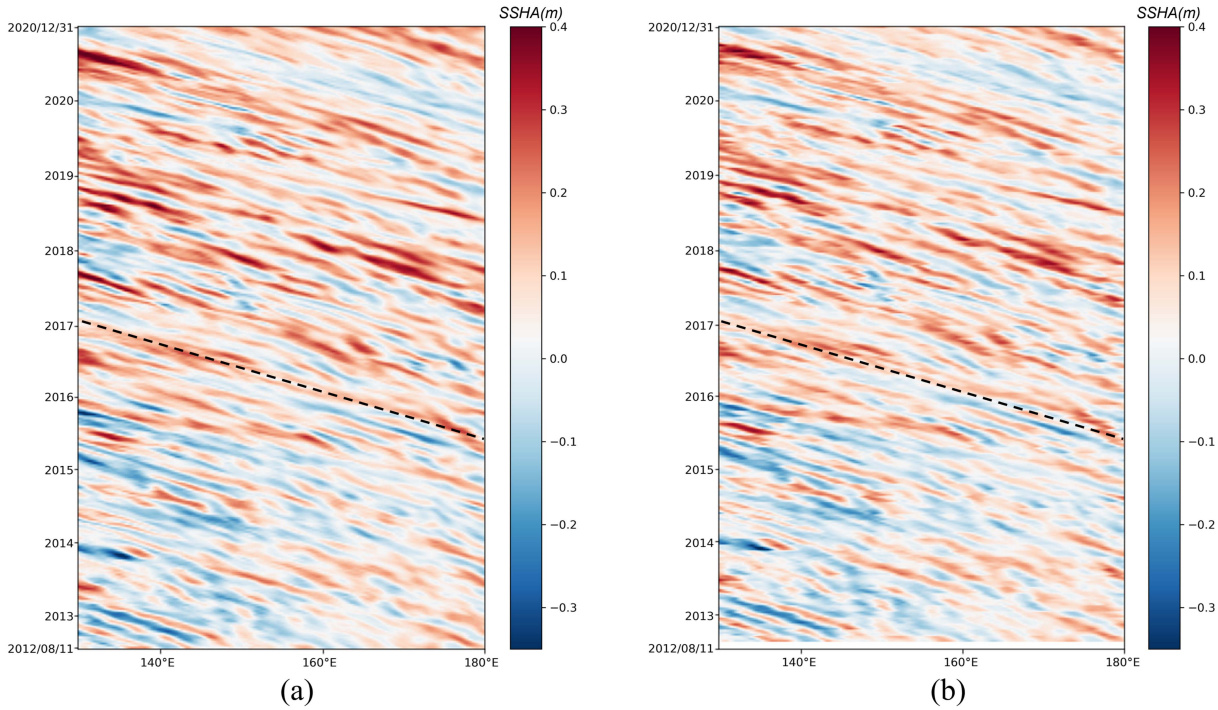


Fig. 12. *Longitude-time plot of the satellite-observed and model-forecasted SSHA field at 19.875°N from August 11, 2012, to December 31, 2020 on TWP.* (a) Longitude-time plot of the satellite-observed SSHA field at 19.875°N from August 11, 2012, to December 31, 2020. (b) Longitude-time plot of the model-forecasted SSHA field at 19.875°N from August 11, 2012, to December 31, 2020. The movement of the model-forecasted SSHA field closely approximates that of the satellite-observed data.

To further demonstrate the effectiveness of our proposed forecast model, we conducted a case study in which we investigated the observations of Rossby waves from the forecasted SSHA field. Since Rossby waves propagate primarily along the zonal (east to west) direction [27], we can observe them using the longitude-time plots of the SSHA field at given latitudes, a technique typically used to detect Rossby waves [27], [58]. Fig. 12 shows the longitude-time distribution of the satellite-observed and model-forecasted SSHA field along 19.875°N during the testing period, which clearly reflects the westward-propagating signal of the SSHA and the changes of sea level on interannual scale. From Fig. 12, it can be seen that the movement of the model-forecasted SSHA field closely approximates that of the satellite-observed data, which can further prove the accuracy of our GMR-Net model for mid-term SSHA forecasting.

We can observe the following phenomena from Fig. 12.

1) SSHA propagates from east to west (the black dotted line highlights the propagation path of SSHA), which is consistent with the propagation characteristics that Rossby waves mainly propagate along the zonal (east to west) direction.
2) It can be directly inferred from the slope of the black dotted line that the westward-propagating speed of the SSHA field is about 0.11 m/s, which is also consistent with the propagation speed of Rossby waves.
3) To detect Rossby waves, we also utilized the dispersion relationship of Rossby waves along the zonal direction. According to the data on the deformation radius of Rossby waves provided by Chelton et al. [59], the deformation radius of Rossby waves is approximately 70 km in the range of 15°N-20°N in the TWP. When the latitude is 19.875°N, according to the β calculation formula provided by Chelton et al. [59], it can be calculated that β is about $2.15 \times 10^{-11} m^{-1} s^{-1}$. By substituting the Rossby wave deformation radius, β, and our predicted westward-propagating speed of SSHA into (20), we found that the Rossby waves dispersion relationship is satisfied, confirming that we observed Rossby waves from the forecasted SSHA field.

These observations further illustrate the validity of our prediction method.

V. Conclusion

The SSHA field is intricately linked with a range of oceanic, air–sea, biophysical, and climate change processes, making it a critical factor in the study of various ocean phenomena, including Rossby waves. In this study, we develop a novel end-to-end GMR-Net to capture the mid- and long-term temporal dependencies of the SSHA field, thereby enhancing the mid-term prediction performance of SSHA. The core of our proposed model is a new LMR cell that extracts mid- and long-term temporal dependencies of the SSHA field using the proposed graph representation mechanism. Besides, the Fast Track mechanism is proposed in the network to solve the problem of predicted SSHA field motion vanishing. The model was, respectively, trained and tested using the daily SSHA spatiotemporal sequence data distributed by CMEMS in the TWP and SCS. Comparisons with advanced prediction models, including ConvLSTM, MAU, MotionRNN, SA-JSTN, EEMD-GRU, MLFrnn, and RF, demonstrated that

our proposed GMR-Net model has better mid-term prediction performance.

An assessment of the effect of the proposed LMR cell and Fast Track mechanism on the model was carried out, with RMSE and MAE utilized as statistical indicators for comparison. Experimental results demonstrate that our model can reliably predict the SSHA by capturing the mid- and long-term temporal dependencies and spatial motion correlations of SSHA sequence change.

Moreover, we successfully observed Rossby waves from our forecast data, further underscoring the validity of our prediction method. The present study represents a significant extension of the utilization of ocean satellite data and serves to enrich the application of DL technology within the field of oceanography. Furthermore, our findings have the potential to stimulate further multidisciplinary research within this nascent domain.

REFERENCES

[1] R. J. Nicholls and A. Cazenave, "Sea-level rise and its impact on coastal zones," *Science*, vol. 328, pp. 1517–1520, 2010.

[2] J. A. Church et al., "Understanding global sea levels: Past, present and future," *Sustainability Sci.*, vol. 3, pp. 9–22, 2008.

[3] A. B. A. Slangen, C. A. Katsman, R. S. W. van de Wal, L. Vermeersen, and R. E. M. Riva, "Towards regional projections of twenty-first century sea-level change based on IPCC SRES scenarios," *Climate Dyn.*, vol. 38, pp. 1191–1209, 2012.

[4] J. A. Church, J. R. Hunter, K. L. McInnes, C. Marine, and N. White, "Sea-level rise around the australian coastline and the changing frequency of extreme sea-level events," *Australian Meteorological Mag.*, vol. 55, pp. 253–260, 2006.

[5] K. L. McInnes, I. Macadam, G. D. Hubbert, and J. O'Grady, "A modelling approach for estimating the frequency of sea level extremes and the impact of climate change in Southeast Australia," *Natural Hazards*, vol. 51, pp. 115–137, 2009.

[6] Y. Cheng, Q. Xu, and X. Li, "Spatio-temporal variability of annual sea level cycle in the Baltic sea," *Remote Sens.*, vol. 10, 2018, Art. no. 528.

[7] J. L. Chen, C. R. Wilson, D. P. Chambers, R. S. Nerem, and B. D. Tapley, "Seasonal global water mass budget and mean sea level variations," *Geophysical Res. Lett.*, vol. 25, no. 19, pp. 3555–3558, 1998.

[8] J. M. Gregory and J. A. Lowe, "Predictions of global and regional sea-level rise using AOGCMs with and without flux adjustment," *Geophysical Res. Lett.*, vol. 27, no. 19, pp. 3069–3072, 2000.

[9] E. R. Miles, C. M. Spillman, J. A. Church, and P. C. McIntosh, "Seasonal prediction of global sea level anomalies using an ocean catmosphere dynamical model," *Climate Dyn.*, vol. 43, pp. 2131–2145, 2014.

[10] T. Niedzielski and W. Kosek, "Forecasting sea level anomalies from TOPEX/Poseidon and Jason-1 satellite altimetry," *J. Geodesy*, vol. 83, pp. 469–476, 2009.

[11] T. Ezer, L. Atkinson, W. B. Corlett, and J. L. Blanco, "Gulf stream's induced sea level rise and variability along the U.S. mid-atlantic coast," *J. Geophys. Res.*, vol. 118, pp. 685–697, 2013.

[12] Y. Fu, X. Zhou, W. Sun, and Q. Tang, "Hybrid model combining empirical mode decomposition, singular spectrum analysis, and least squares for satellite-derived sea-level anomaly prediction," *Int. J. Remote Sens.*, vol. 40, pp. 7817–7829, 2019.

[13] M. Imani et al., "Spatiotemporal prediction of satellite altimetry sea level anomalies in the tropical pacific ocean," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1126–1130, Jul. 2017.

[14] H. Cui, D. Tang, H. Liu, Y. Sui, and X. Gu, "Composite analysis-based machine learning for prediction of tropical cyclone-induced sea surface height anomaly," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2644–2653, 2023.

[15] X. Yao, G. Zheng, J. Yu, J. Shao, and J. Zuo, "Robust prediction of sea surface temperature based on SSPGAN," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 9396–9405, 2023.

[16] Y. Ren and X. Li, "Predicting the daily sea ice concentration on a subseasonal scale of the pan-arctic during the melting season by a deep learning model," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4301315.

[17] A. Braakmann-Folgmann, R. Roscher, S. Wenzel, B. Uebbing, and J. Kusche, "Sea level anomaly prediction using recurrent neural networks," in *Proc. Conf. Big Data Space*, 2017, pp. 297–300.

[18] Q. Sun, J. Wan, and S. Liu, "Estimation of sea level variability in the China sea and its vicinity using the SARIMA and LSTM models," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3317–3326, 2020.

[19] J. Liu, B. Jin, L. Wang, and L. Xu, "Sea surface height prediction with deep learning based on attention mechanism," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2020, Art. no. 1501605.

[20] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.

[21] T. Song, J. Jiang, W. Li, and D. Xu, "A deep learning method with merged LSTM neural networks for SSHA prediction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2853–2860, 2020.

[22] Y. Zhou, C. Lu, K. Chen, and X. Li, "Multilayer fusion recurrent neural network for sea surface height anomaly field prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4205111.

[23] G. Zheng, X. Li, R. Zhang, and B. Liu, "Purely satellite data–driven deep learning forecast of complicated tropical instability waves," *Sci. Adv.*, vol. 6, no. 29, 2020, Art. no. eaba1482.

[24] J. M. Gregory et al., "Twentieth-century global-mean sea-level rise: Is the whole greater than the sum of the parts?," *J. Climate*, vol. 26, pp. 4476–4499, 2013.

[25] R. A. Madden, "Large-scale, free Rossby waves in the atmosphere an update," *Tellus A: Dyn. Meteorol. Oceanogr.*, vol. 59, pp. 571–590, 2007.

[26] R. A. Madden, "Observations of large-scale traveling Rossby waves," *Rev. Geophys.*, vol. 17, pp. 1935–1949, 1979.

[27] P. Challenor, P. Cipollini, and D. Cromwell, "Use of the 3D radon transform to examine the properties of oceanic Rossby waves," *J. Atmospheric Ocean. Technol.*, vol. 18, pp. 1558–1566, 2001.

[28] J. Potemra, "Contribution of equatorial pacific winds to southern tropical indian ocean Rossby waves," *J. Geophys. Res.*, vol. 106, pp. 2407–2422, 2001.

[29] J. L. Pinault, "A review of the role of the oceanic Rossby waves in climate variability," *J. Mar. Sci. Eng.*, vol. 10, no. 4, 2022, Art. no. 493.

[30] D. Chelton and M. G. Schlax, "Global observations of oceanic Rossby waves," *Science*, vol. 272, pp. 234–238, 1996.

[31] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?," in *Proc. Int. Conf. Learn. Representations*, 2019, pp. 1–17.

[32] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," *IEEE Data Eng. Bull.*, vol. 40, pp. 52–74, 2017.

[33] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W.-K. Wong, and W. Chun Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[34] D. Gong et al., "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1705–1714.

[35] G. Sun, Y. Hua, G. Hu, and N. M. Robertson, "Mamba: Multi-level aggregation via memory bank for video object detection," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 2620–2627.

[36] A. G. Nidheesh, M. Lengaigne, J. Vialard, A. S. Unnikrishnan, and H. Dayan, "Decadal and long-term sea level variability in the tropical indo-pacific ocean," *Climate Dyn.*, vol. 41, pp. 381–402, 2013.

[37] S. Hu et al., "Observed strong subsurface marine heatwaves in the Tropical Western Pacific ocean," *Environmen. Res. Lett.*, vol. 16, no. 10, 2021, Art. no. 104024.

[38] K. S. Whan et al., "Trends and variability of temperature extremes in the Tropical Western Pacific," *Int. J. Climatol.*, vol. 34, no. 8, 2014, pp. 2585–2603.

[39] R. C. Y. Li and W. Zhou, "Changes in western pacific tropical cyclones associated with the El Niño–Southern oscillation cycle," *J. Climate*, vol. 25, pp. 5864–5878, 2012.

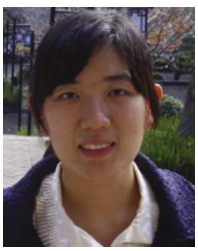
[40] G. Meyers, "On the annual Rossby wave in the tropical north pacific ocean," *J. Phys. Oceanogr.*, vol. 9, pp. 663–674, 1979.

[41] D. Chelton, M. G. Schlax, J. M. Lyman, and G. C. Johnson, "Equatorially trapped Rossby waves in the presence of meridionally sheared baroclinic flow in the pacific ocean," *Prog. Oceanogr.*, vol. 56, pp. 323–380, 2003.

[42] W. S. Kessler, "Observations of long Rossby waves in the northern tropical pacific," *J. Geophysical Res.*, vol. 95, pp. 5183–5217, 1990.

[43] A. Bonaduce, N. Pinardi, P. Oddo, G. Spada, and G. Larnicol, "Sea-level variability in the mediterranean sea from altimetry and tide gauges," *Climate Dyn.*, vol. 47, pp. 2851–2866, 2016.

[44] J. A. Carton, B. S. Giese, and S. A. Grodsky, "Sea level rise and the warming of the oceans in the simple ocean data assimilation (SODA) ocean reanalysis," *J. Geophysical Res.*, vol. 110, no. C9, 2005, Art. no. C09006.
[45] R. D. Norris, S. K. Turner, P. M. Hull, and A. Ridgwell, "Marine ecosystem responses to Cenozoic global change," *Science*, vol. 341, pp. 492–498, 2013.
[46] F. Nan, Z. He, H. Zhou, and D. Wang, "Three longlived anticyclonic eddies in the northern South China sea," *J. Geophysical Res.*, vol. 116, no. C5, 2011, Art. no. C05002.
[47] Q. Zheng et al., "Standing wave modes observed in the South China sea deep basin," *J. Geophysical Res.*, vol. 119, pp. 4185–4199, 2014.
[48] Z. Zhao, B. Liu, and X. Li, "Internal solitary waves in the China seas observed using satellite remote-sensing techniques: A review and perspectives," *Int. J. Remote Sens.*, vol. 35, pp. 3926–3946, 2014.
[49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.
[50] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Representations*, 2017, pp. 1–19.
[51] Z. Chang, "Mau: A. motion-aware unit for video prediction and beyond," in *Proc. Neural Inf. Process. Syst.*, 2021, pp. 26950–26962.
[52] H. Wu, Z. Yao, M. Long, and J. Wan, "Motionrnn: A. flexible model for video prediction with spacetime-varying motions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 15430–15439.
[53] L. Shi, N. Liang, X. Xu, T. Li, and Z. Zhang, "SA-JSTN: Self-attention joint spatiotemporal network for temperature forecasting," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9475–9485, 2021.
[54] X. Liu et al., "Multistep-ahead prediction of ocean SSTA based on hybrid empirical mode decomposition and gated recurrent unit model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7525–7538, 2022.
[55] C.-G. Rossby, "Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action," *J. Mar. Res.*, vol. 2, pp. 38–55, 1939.
[56] C.-G. Rossby, "On the propagation of frequencies and energy in certain types of oceanic and atmospheric waves," *J. Meteorol.*, vol. 2, pp. 187–204, 1945.
[57] A. E. Gill, *Atmosphere-Ocean Dynamics*, vol. 30, San Diego, CA USA: Academic Press, 1982.
[58] P. Cipollini, D. Cromwell, P. Challenor, and S. Raffaglio, "Rossby waves detected in global ocean colour data," *Geophysical Res. Lett.*, vol. 28, no. 2, pp. 323–326, 2001.
[59] D. Chelton, R. A. Deszoeke, M. G. Schlax, K. Naggar, and N. Siwertz, "Geographical variability of the first baroclinic Rossby radius of deformation," *J. Phys. Oceanogr.*, vol. 28, pp. 433–460, 1998.



Yuan Zhou (Senior Member, IEEE) received the B.Eng., M.Eng., and Ph.D. degrees from Tianjin University, Tianjin, China, in 2006, 2008, and 2011, respectively, all in electronic engineering and communication engineering.

Since 2011, she has been a Faculty Member with the School of Electronic Information Engineering, Tianjin University, where she is currently a Professor. From 2013 to 2014, she was a Visiting Scholar with the School of Mechanical and Electrical Engineering, University of Southern Queensland, Toowoomba, QLD, Australia. From 2016 to 2017, she was a Visiting Scholar with the Department of Electrical Engineering, Princeton University, Princeton, NJ, USA. Her research interests include computer vision, machine learning, image processing, and artificial intelligence oceanography.

Tian Ren received the B.Eng. degree in communication engineering from Taiyuan University of Technology, Taiyuan, China, in 2021. She is currently working toward the M.Eng. degree in information and communication engineering with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China.

Her research interests include computer vision and artificial intelligence oceanography.

Keran Chen received the B.Eng. degree in new energy science and engineering from the Hebei University of Technology, Tianjin, China, in 2018, and the M.Sc. degree in control science and engineering, in 2021, from Tianjin University, Tianjin, where he is currently working toward the Ph.D. degree in information and communication engineering.

His research interests include artificial intelligence oceanography and multimodal understanding.

Le Gao (Member, IEEE) received the B.S. degree in geodesy and surveying engineering from the Shandong University of Science and Technology, Qingdao, China, in 2008, the joint M.S. degree in cartography and geography information system from the Shandong University of Science and Technology, Qingdao, China, and the Chinese Academy of Surveying and Mapping, Beijing, China, in 2011, and the joint Ph.D. degree in cartography and geography information system from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, and the Space Research Institute, Austrian Academy of Sciences, Graz, Austria, in 2014.

He is currently an Associate Professor with the Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China. His research interests include artificial intelligence oceanography, big data, target detection and semantic segmentation from ocean remote sensing data, and ocean prediction based on deep learning.

Xiaofeng Li (Fellow, IEEE) received the bachelor's degree in optical engineering from Zhejiang University, Hangzhou, China, in 1985, the master's degree in physical oceanography from the First Institute of Oceanography, Qingdao, China, in 1992, and the Ph.D. degree in physical oceanography from North Carolina State University, Raleigh, NC, USA, in 1997.

From 1997 to 2019, he was with the National Environmental Satellite, Data, and Information Service, National Ocean and Atmospheric Administration, Washington, DC, USA, where he was involved in developing many operational satellite ocean remote sensing products. He is currently with the Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China. His research interests include synthetic aperture radar applications in oceanography and marine meteorology, the application of artificial intelligence in oceanographic studies, big data analytics, and satellite image processing.

Dr. Li holds editorial positions as an Associate Editor for both IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the *International Journal of Remote Sensing*, as an Editor for *JGR: Machine Learning and Computation*, and as the Executive Editor-in-Chief for the *Journal of Remote Sensing*.