RESEARCH ARTICLE | MARCH 17 2025

# A Swin-Transformer-based deep-learning model for rolled-out predictions of regional wind waves

Weikai Tan (谈伟恺) ; Caihao Yuan (袁才昊); Sudong Xu (徐宿东) ; Yuan Xu (徐元) ; Alessandro Stocchino



*Physics of Fluids* 37, 036625 (2025)
https://doi.org/10.1063/5.0256654

# A Swin-Transformer-based deep-learning model for rolled-out predictions of regional wind waves

Cite as: Phys. Fluids **37**, 036625 (2025); doi: 10.1063/5.0256654
Submitted: 6 January 2025 · Accepted: 25 February 2025 ·
Published Online: 17 March 2025



Weikai Tan (谈伟恺),[1] Caihao Yuan (袁才昊),[1] Sudong Xu (徐宿东),[1,a)] Yuan Xu (徐元),[2] and Alessandro Stocchino[3]

## AFFILIATIONS

[1]Department of Port, Waterway and Coastal Engineering, School of Transportation, Southeast University, Nanjing 210096, China
[2]State Key Laboratory of Estuarine and Coastal Research, East China Normal University, Shanghai 200241, China
[3]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China

[a)]Author to whom correspondence should be addressed: sudongxu@seu.edu.cn

## ABSTRACT

Short-term predictions of regional wind waves are crucial for coastal and ocean engineering. In this study, we introduce a novel Swin-Transformer-based model, named ST-RWP (Swin Transformer for Regional Wave Prediction), designed to leverage the spatiotemporal relationships of wind velocities and significant wave heights. The model considers inductive bias to capture both local and global dependencies via Convolution and Swin Transformer layers, enabling accurate short-term wave field predictions on unseen data. A rolled-out prediction scheme is employed to extend the forecast horizon efficiently. Trained on the reanalysis dataset offered by European Center for Medium-Range Weather Forecasts, ST-RWP demonstrates excellent performance in predicting wave fields with lead times of 6 and 12 h. However, the model's accuracy degrades when the lead time exceeds 24 h, primarily due to the limited spatial information available at boundary nodes and the low autocorrelation value for such large time span. The dataset exhibits strong spatial and temporal correlations, which are key to the model's success. Our findings indicate that ST-RWP offers an efficient tool for real-time wave field nowcasting, representing a significant advancement in the application of Transformer-based deep neural networks to wave prediction.





## I. INTRODUCTION

Prediction of short-term wave fields has garnered significant attention within both coastal and offshore research communities. Accurate forecasts of nearshore wave fields are indispensable in early-warning system for storm surge (Naeini and Snaiki, 2024a; Luo *et al.*, 2022), tsunami (Wang *et al.*, 2023), wave runup, and coastal inundation (Naeini and Snaiki, 2024b). They also play a vital role in coastal wetland restoration (Liu *et al.*, 2021a), harbor (Zanuttigh *et al.*, 2013), and port (Zheng *et al.*, 2020) structure health management. In offshore engineering, reliable wave field prediction favors ship autopilot (Lou *et al.*, 2021), wave energy forecasting (Bento *et al.*, 2021; Yang *et al.*, 2021), remotely operated-vehicle (Law *et al.*, 2020), among other applications. Notably, accurate short-term wave field predictions, such as those with a 12-h lead time, are crucial for simulating marine pollution processes (Daliri *et al.*, 2025). Thus, there is a need for wave prediction models that are both accurate and computationally efficient.

Over the past several decades, extensive research has been devoted to simulating wave fields through numerical models. These models can generally be categorized into two types, i.e., phase-resolving and phase-averaged models. A phase-resolving model (e.g., Peregrine, 1967; Liu, 1995; and Liu *et al.*, 2018) solves for intra-wave details using two-dimension horizontal frameworks. However, these models are constrained by high computational costs when simulating large-scale wave fields. On the other hand, phase-averaged models, solving the spectral action balance equation, are widely utilized for simulating wave fields at oceanic scales, e.g., the WAM (WAve Model) developed by Hasselmann *et al.* (1988). Follow them, the SWAN (Simulating Waves Nearshore, Booij *et al.*, 1999) model, by incorporating implicit schemes and additional physical processes relevant to shallow water, is particularly suited for calculating nearshore wave fields. Similarly, the WW3 (WAVEWATCH III) model (e.g., Mentaschi *et al.*, 2015) is designed with a broader applicability, enabling its use across a wider coastal or oceanic range.

Despite the success in numerical models for wave field simulation, challenges remain in achieving real-time forecasting (e.g., James *et al.*, 2018). For example, both pre- and post-processing often require professional intervention, and their high computational costs constrain their applicability in hindcasting. To this end, Machine Learning

(called ML hereafter) models gain popularity in wave field forecasting (e.g., Roome *et al.*, 2024).

In most data-driven studies, the data used to train ML-based wave models may be classified into three types: (i) measured wave conditions from buoys, (ii) simulated wave data using the aforementioned numerical models, and (iii) regional-scale reanalysis datasets (e.g., ERA5). Various strategies have been proposed to fully exploit these datasets.

To forecast wave conditions at buoy stations, Ellenson *et al.* (2020), followed by Wang *et al.* (2022), integrated bagged regression trees with physics-based wave models (WW3 and SWAN, respectively) to predict waves conditions. James *et al.* (2018) successfully applied a multi-layer perceptron for estimating significant wave height ($H_s$), and a support vector machine to classify characteristic wave periods ($T$). Yet, these approaches do not account for temporal dynamics, which are typically modeled using Recurrent Neural Networks (RNNs). For instance, Minuzzi and Farina (2023) and Fan *et al.* (2020) adopted a Long Short-Term Memory (LSTM) model to predict $H_s$ across various lead times with impressive accuracy, while Luo *et al.* (2022) successfully implemented a bidirectional LSTM with attention mechanism to predict $H_s$ in hurricane area of the Atlantic Ocean. Li *et al.* (2022) applied a Gate Recurrent Unit (GRU) model to simulate $H_s$ and incorporated multivariate time series to enhance predictive accuracy. Hao *et al.* (2022) integrated the empirical mode decomposition method with LSTM to favor the prediction of non-stationary time series of waves.

ML-based models excel at predicting wave conditions at single or a limited number of buoy stations, whereas their efficiency diminishes when applied to multiple locations. To solve this issue, more advanced deep-learning models have been studied, leveraging spatiotemporal information to enhance predictive performance. Naeini and Snaiki (2024a) proposed a hybrid neural network to predict storm surge and waves by combining a deep autoencoder, designed to handle high-dimensional spatial systems, and a deep neural network for encoding storm parameters, wherein a good model performance was achieved. As the number of spatial points further increases, Convolutional Neural Networks (CNNs), using a few trainable kernels to quantify relationship among adjacent nodes, become more efficient. Consequently, Guan (2020), Huang *et al.* (2022), and Bai *et al.* (2022) successfully integrated CNN for capturing spatial relationships with LSTM for modeling temporal dynamics. Noticeably, the applicability of CNNs is limited to structured data. Therefore, Feng and Xu (2024) also combined graph convolution neural network and GRU to simulate $H_s$ at unevenly distributed stations.

Beyond the CNN-RNN framework, Transformer-based models (e.g., Vaswani *et al.*, 2017) have gained unprecedented attention due to their remarkable success in nature language processing. Recently, Bi *et al.* (2023) successfully applied a three-dimensional transformer for medium-range global weather forecasting, highlighting its potential in metocean area. Liu *et al.* (2023) pioneered the application of a Vision Transformer (ViT) model in wave forecasting, which extracts spatiotemporal information from wind fields over preceding 6 h to reconstruct the wave field for the subsequent hour. ViT shows its superiority over CNN-based Regional Wave Prediction (RWP) models, which is not surprising as the multi-head self-attention (MSA) mechanism in transformer models efficiently captures global features across the entire input field. Moreover, the MSA addresses temporal dynamics in a parallel mode, which is more efficient than RNN. It is noteworthy that the prediction horizon in Liu *et al.* (2023) is relatively short, being limited to 1-h lead time. Moreover, their wave field prediction exhibits mesh-like error, which is a common issue encountered by vision transformers (due to the up-sampling layer) in image-to-image learning tasks. These relatively high errors may be exacerbated in long-term wave field forecasting.

To address the above-mentioned issues and advance the transformer-based RWP further, we implement a novel ST-RWP model, which combines Swin Transformer and CNN to fully extract spatiotemporal information of wind wave fields at different scales. It is shown to be a promising tool for real-time wave field nowcasting. The rest of the paper is arranged as follows: Sec. II provides an overview of the dataset and study area. Details on data preparation for training, validating, and testing of the ST-RWP model are also given. Section III presents the details of the ST-RWP model and its training process. The model's performance is then evaluated in Sec. IV. Section V offers some discussions on the inductive bias and hyperparameters, and finally, Sec. VI draws the conclusions.

## II. DATA DESCRIPTION

The ST-RWP model requires high-quality, sufficient data for training. Therefore, we select the fifth generation European Center for Medium-Range Weather Forecasts (ECMWFs) reanalysis (abbreviated as ERA5) (Hersbach *et al.*, 2020) dataset for training, validating, and testing. The following section provides details on the dataset and data preparation.

### A. ERA5 dataset and study area

As the data-driven ST-RWP model presented in this study requires a large dataset to learn potential spatiotemporal relationship within wave fields, the ERA5 reanalysis dataset is adopted herein. It integrates diverse observations with numerical simulations through data assimilation to develop a high-quality, long-term metocean dataset.

To validate the performance of the ST-RWP model, we utilize the wave field data at the North Atlantic Ocean as a pilot site, which is extracted from the ERA5 dataset. Figure 1 delineates the study area,

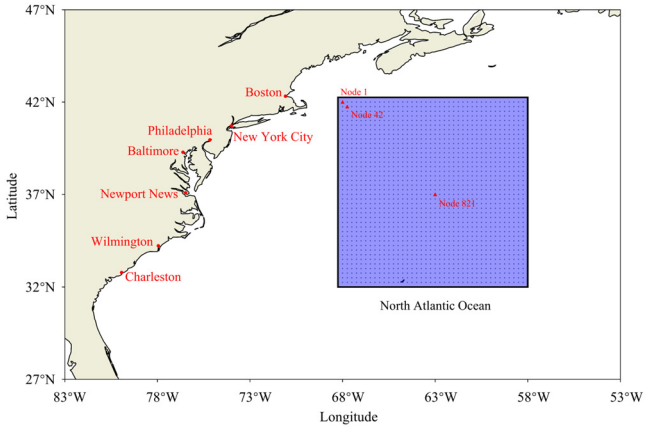


**FIG. 1.** A sketch of the study area at the North Atlantic Ocean. The spatial data nodes (40 × 40) are marked, and the node ID is assigned row-by-row.

with the shaded blue region representing the $40 \times 40$ spatial nodes of interest. It ranges from 68° W to 58° W longitude and from 32° N to 42° N latitude, with a spatial resolution of 0.25°. The data nodes are numbered for identification, with Node 1 located at the top-left corner. At each data node, we consider the significant wave height ($H_s$) and two wind velocity components ($U_w$, $V_w$) at 10 $m$ above the sea surface. For the present analysis, we extracted the hourly data from 2017 to 2021.

### B. Data preparation

The original 5-year dataset consists of sequential data of wave and wind velocity fields, with each time step representing the two-dimensional spatial distributions of variables across the target region.

When forecasting wave fields with neural network models, the goal is to design an efficient architecture that robustly captures spatiotemporal dynamics of the key variables for reliable predictions. To fully utilize the dataset, these features should be seamlessly integrated and fused into the model inputs. To this end, we posit that the last $K$ time steps of the two-dimensional $H_s$, $U_w$, and $V_w$ fields are crucial for forecasting their future $M$ time steps. The values of $K$ and $M$ should be carefully selected based on autocorrelation analysis and model performance.

Consequently, a sliding window sampling strategy is implemented on the original data, see the data preparation module in Fig. 2. Starting from the first time step, the window extracts $K$ consecutive steps as input and $M$ subsequent steps as output and then moves forward with a stride of 1. Subsequently, the sampled data are partitioned into training, validating, and testing datasets, wherein the initial four years are used for training (80%) and validating (20%) the model, while the final year is reserved for testing. This division ensures a robust evaluation of the model's performance on unseen data. For instance, with $K = 6$ and $M = 1$, similar to Liu *et al.* (2023), we obtain 28 046 samples for training, 7012 samples for validating, and 8754 samples for testing.

## III. METHODOLOGY

The proposed ST-RWP model exploits the strengths of convolutional and Swin Transformer layers to predict short-term wave fields. To better introduce the ST-RWP model, this section first highlights the advantages of transformer-based neural networks, followed by a detailed description of the ST-RWP model.

### A. Extract spatiotemporal relationships of wave fields using Transformer-based model

The Transformer architecture was initially introduced to capture temporal dependencies in sequential data through a parallel-processing mode (Vaswani *et al.*, 2017). This is achieved via Multi-head Self-Attention (MSA) mechanisms, where each head learns a distinct set of attention weights, allowing the model to flexibly encode various aspects of the sequence, e.g., local and long-range dependencies or recurring patterns. Consequently, the Transformer has demonstrated superior performance in handling sequential data, enabling efficient and effective extraction of temporal dynamics in wave fields.

On the other hand, the spatial relationships within wave fields can also be effectively extracted using the Vision Transformer (ViT) architecture (Dosovitskiy, 2020). ViT segments input two-dimensional fields into patches, encoding each as a vector. MSA then learns spatial relationships among these patches according to those encoded vectors. Yet, ViT is less effective for high-resolution two-dimensional fields. To address this, the Swin Transformer (Liu *et al.*, 2021b) introduced a hierarchical shifted-window mechanism, capturing both local and global features, making it particularly suited for high-resolution image processing.

In this study, we integrate Swin Transformer with CNN modules to effectively exploit spatiotemporal information from wind waves for forecasting future wave fields. Details of the model will be provided in the following section. Notably, a similar model architecture was successfully applied to image restoration (see Liang *et al.*, 2021), and our model is inspired by it.

### B. ST-RWP model for wave field prediction

The structure of the ST-RWP model is illustrated in Fig. 2. The model takes $K$ steps of wind and wave fields as inputs, which first pass through a convolution layer designed to extract low-frequency information, such as general contours and large-scale structures of wind and wave fields,

$$I_0 = F_{Conv_1}[U_w(t_K, x, y), V_w(t_K, x, y), H_s(t_K, x, y)], \tag{1}$$

where $I_0$ denotes low-frequency information and $F_{Conv_1}$ indicates the early convolution layer (see Fig. 2). $U_w$ and $V_w$, and $H_s$ are wind velocities and significant wave height, respectively. $t_K$ denotes $K$-step inputs. Such design follows the findings by Zeiler (2014) and Olah *et al.* (2018), who demonstrated that early layers in convolution neural networks primarily capture low-frequency features. Subsequently, the low-frequency features are directly passed to the last layer of the ST-RWP model via residual connection (e.g., He *et al.*, 2016), ensuring that the large-scale structure in wind and wave fields is preserved.

Then, the processed data are passed through $k_2$ layers of Residual Swin Transformer (RST, shaded in blue) block, wherein $k_2$ is selected to be 4 through testing, i.e.,

$$I_i = F_{RST_i}(I_{i-1}), \quad i = 1, \ldots, k_2, \tag{2}$$

where $I_i$ represents the output from the $i$ RST block. At the end of the RST blocks, a convolution layer is added, i.e., $I_D = F_{Conv_3}(I_4)$, with $I_D$ being the features extracted by the RST blocks.

More specifically, each RST block consists of $2k_1$ ST blocks (shaded in yellow in Fig. 2), which reads

$$I_{i-1,j} = F_{ST_{i,j}}(I_{i-1,j-1}), \quad j = 1, \ldots, 2k_1, \tag{3}$$

where $I_{i-1,j}$ represents the extracted information by the $(i-1)$ RST block and $j$ ST block. In this study, we choose $k_1 = 1$ for simplicity. Notably, a Swin Transformer block must be performed twice for window partition and shifted-window partition, with the later for resolving cross-window relationships. Hence, the number of ST blocks is even, i.e., $2k_1$. Next, a convolutional layer ($Conv_2$) is added, and a residual connection is utilized to generate results from the $i$ RST block, i.e.,

$$I_i = F_{Conv_2}(I_{i-1,2}) + I_{i-1,0}. \tag{4}$$

It should be noted that $I_{i-1,0}$ is equivalent to $I_{i-1}$, which represents the output from the $(i-1)$ RST block. As clarified by Liang *et al.* (2021), a stack of ST blocks excels at extracting both global relationship and high-frequency information, whereas the convolution layer can effectively capture local relationship (i.e., taking advantage of the inductive bias). This combination harnesses the strengths of both neural network structures, which is believed to be beneficial for wave field forecasting.

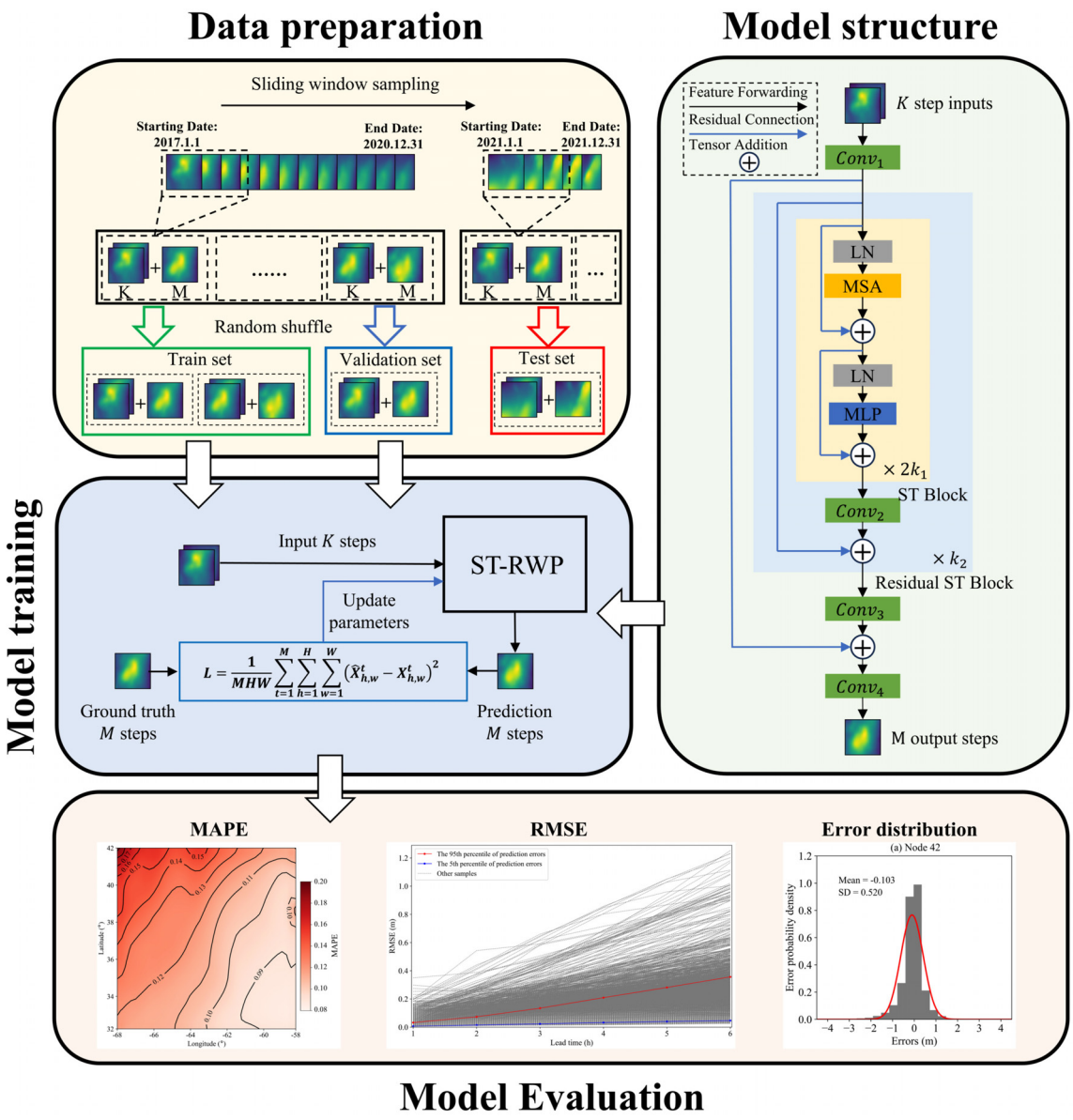


**FIG. 2.** An overview of data preparation, model structure, model training, and evaluation.

Additionally, the residual connections facilitate the aggregation of features extracted across different layers.

Within each ST block, a Layer Normalization (LN) is applied to enhance training stability. Afterward, MSA mechanisms are used to extract spatiotemporal information from the input fields.

Finally, a residual connection and a convolution layer ($Conv_4$) are adopted to output $M$-step wave fields, viz.,

$$\left[\hat{U}_w(t_M,x,y),\hat{V}_w(t_M,x,y),\hat{H}_s(t_M,x,y)\right]=F_{Conv_4}(I_0+I_D), \quad (5)$$

where $t_M$ denotes $M$-step wave fields, and $\hat{U}_w$ and $\hat{V}_w$, and $\hat{H}_s$ indicate the predicted wind velocity and significant wave height by the ST-RWP model, respectively. The key target is $\hat{H}_s$, while $\hat{U}_w$ and $\hat{V}_w$ are indispensable for a "rolled-out" prediction, i.e., feeding the predicted results as inputs to further forecast the future wave field, following Lam *et al.* (2023).

### C. Model training and evaluation

Given the above-mentioned model setup, the ST-RWP contains 416, 103 trainable parameters in total. We adopt a mean square-error (MSE) loss function to train the model, i.e.,

$$L=\frac{1}{MHW}\sum_{t=1}^{M}\sum_{h=1}^{H}\sum_{w=1}^{W}\left[\hat{X}(t,h,w)-X(t,h,w)\right]^2, \quad (6)$$

where $L$ denotes the loss function, $H$ and $W$ are the number of nodes in $x$ and $y$ directions, respectively, which are both 40 in this work. $\hat{X} = [\hat{U}_w, \hat{V}_w, \hat{H}_s]$ and $X$ are integrated tensors.

The training dataset is iterated for 500 times (i.e., epoch = 500), which is sufficient for model convergence. The learning rate is chosen to be 0.001, and the batch size is 128. A workstation with NVIDIA GeForce RTX 4080 SUPER (16 GB of video memory) is employed to conduct the training. For instance, given $K = 6$ and $M = 1$, it takes around 10.8 h to finish the training process.

Afterward, the performance of the trained ST-RWP model is rigorously validated from multiple perspectives. For instance, the evaluation encompasses key metrics such as the Mean-Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE), which are computed across all testing data samples. Furthermore, a careful analysis of the error distributions at selected nodes is conducted to provide insights into localized model behavior. Through these comprehensive assessments, the overall performance and reliability of the ST-RWP model are discussed.

Once trained, the model requires a limited amount of computational resources for inference and can be efficiently deployed on a standard laptop. The currently trained architecture occupies just 2.28 MB of disk space. On a laptop equipped with an NVIDIA RTX 4060 GPU, the model achieves a per-time step computation time of 0.09 s, while a laptop without a GPU requires 0.63 s per time step. Consequently, a full 12-h forecast can be completed in under 10 s using consumer-grade hardware, demonstrating the model's efficiency.

## IV. RESULTS

Currently, the ST-RWP is a data-driven model that requires thorough validation. Therefore, this section presents a detailed comparison between model predictions and ground truth data, evaluating the model's performance across different input steps ($K$) and output steps ($M$).

### A. Rolled-out prediction of wave fields with 6-h leading time

To flexibly utilize the ST-RWP model, a 1-hour rolled-out prediction scheme is implemented. Initially, the model receives the past 6 h of wind wave field data ($K = 6$) to forecast the subsequent 1-h wind wave conditions ($M = 1$). The predicted 1-h results are then concatenated with the preceding 5 h of the data to perform the next 1-h forecast. This iterative process continues until a 6-h lead time is reached.

Notably, the key advantage of the rolled-out prediction approach is that it requires only a single training procedure (with $K = 6$, $M = 1$), which significantly reduces computational costs and enhances flexibility in forecasting for any lead time (Lam *et al.*, 2023). Figure 3 shows the convergence of the training and validation loss functions as the number of epoch increases. Before the 100th epoch, the MSE loss function for both the training and validation datasets decreases rapidly and then gradually diminishes to below 0.1. Although the loss function of the validation dataset exhibits some oscillations up to the 300th epoch, the overall trends of the training and validation losses converge closely. After the 320th epoch, the ST-RWP model is well-trained, as evidenced by the stable and converged loss functions for both train and validation dataset. Finally, we selected the model trained at the 500th epoch.

To better visualize the model's performance, Fig. 4 shows the root mean square-error (RMSE) of significant wave height ($H_s$) for each data sample in the test dataset across 1- to 6-h lead times (Wang *et al.*, 2012). As expected, the RMSE for most test samples increases with lead time, indicating the typical accumulation of error in rolled-out predictions. Such accumulated errors occur both in numerical and deep-learning wave models, which may be mitigated through data assimilation. In Fig. 4, the samples at the 95th and 5th percentiles of RMSE errors are highlighted red and blue, respectively. The associated wave field predictions are shown in Figs. 5 and 6.

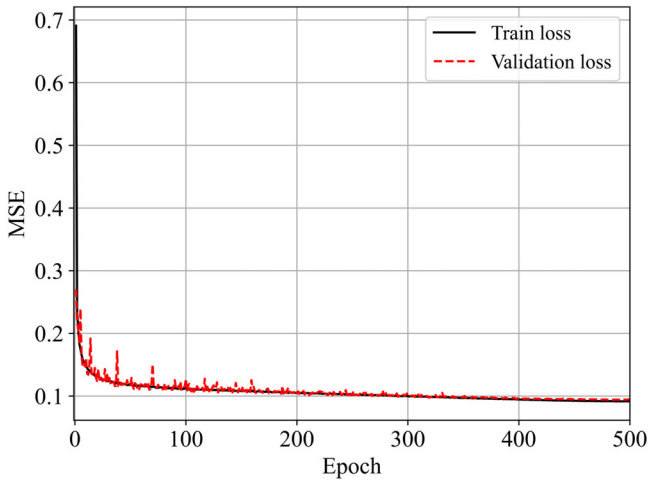


**FIG. 3.** Loss function for the training and validation datasets with varying epoch for $K = 6$ and $M = 1$.

In Fig. 5, the three columns depict the ground truth, ST-RWP predictions, and percentage error defined as $(\hat{H}_s - H_s)/H_s$, respectively. The significant wave height ($H_s$) peaks at approximately 11 $m$ in the bottom-left region, possibly due to a storm event. The ST-RWP model accurately captures the local maxima of $H_s$ across all six time steps. For lead times (abbreviated as $LT$ hereafter) up to 4 h, the percentage error remains consistently below 10% across the entire domain, suggesting an excellent model performance in predicting wave distribution over the study area. For $LT \geq 4h$, 10% error of $H_s$ appears along the right boundary, which is expected given the absence of Dirichlet boundary conditions in the current model. Future work

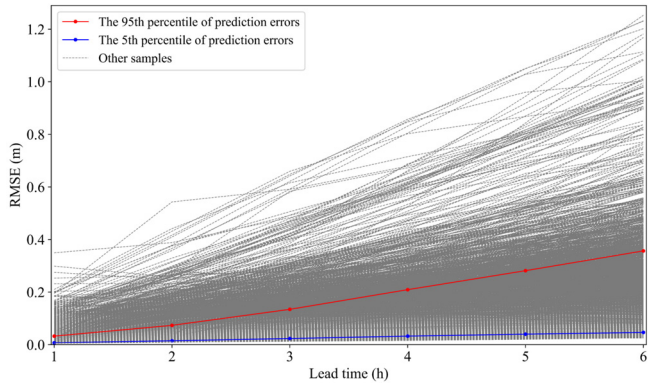


**FIG. 4.** RMSE of $H_s$ predictions for samples in the testing dataset with 1–6 h lead times (rolled-out prediction with $K = 6$ and $M = 1$). The red line and blue line represent the samples at the 95th and 5th percentiles of prediction errors, respectively.

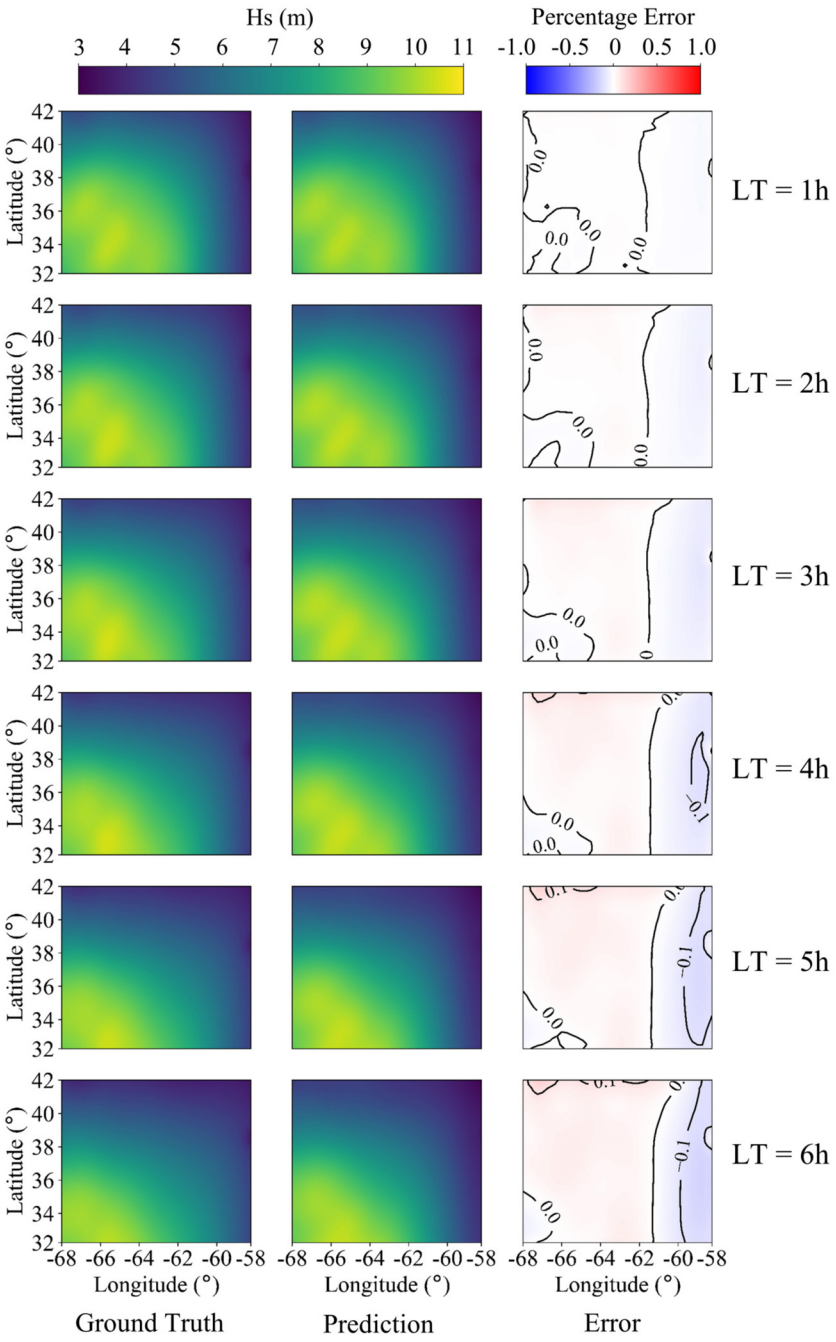


**FIG. 5.** Ground truth, predictions, and percentage errors of $H_s$ for sample 1462 (from 3:00 to 8:00 on March 3, 2021) in the test set, referring to the red line in Fig. 4, with lead times of 1–6 h.

may incorporate boundary conditions into the ST-RWP model to improve its predictive accuracy. In the present study, a 10% error in $H_s$ is considered acceptable. Meanwhile, it should be noted that this particular sample exhibits a relatively high RMSE, as indicated in Fig. 4. To further illustrate the model's performance, we also present a sample with a relatively low RMSE in Fig. 6.

In contrast to the high $H_s$ in Fig. 5, Fig. 6 shows mild $H_s$, with a maximum value of approximately 2.3 $m$, which is typical for normal daily conditions at the North Atlantic Ocean (Woolf *et al.*, 2002). In this case, the ST-RWP model exhibits high accuracy, with the percentage error consistently below 10% throughout the computational domain and across all the six lead times. The predicted wave fields closely align with the ground truth, particularly in capturing local extremes of $H_s$ at the left-bottom corner of the study area. Moreover, the ST-RWP model accurately captures the evolution of wave fields over 6 h, demonstrating its ability to learn temporal dynamics effectively.

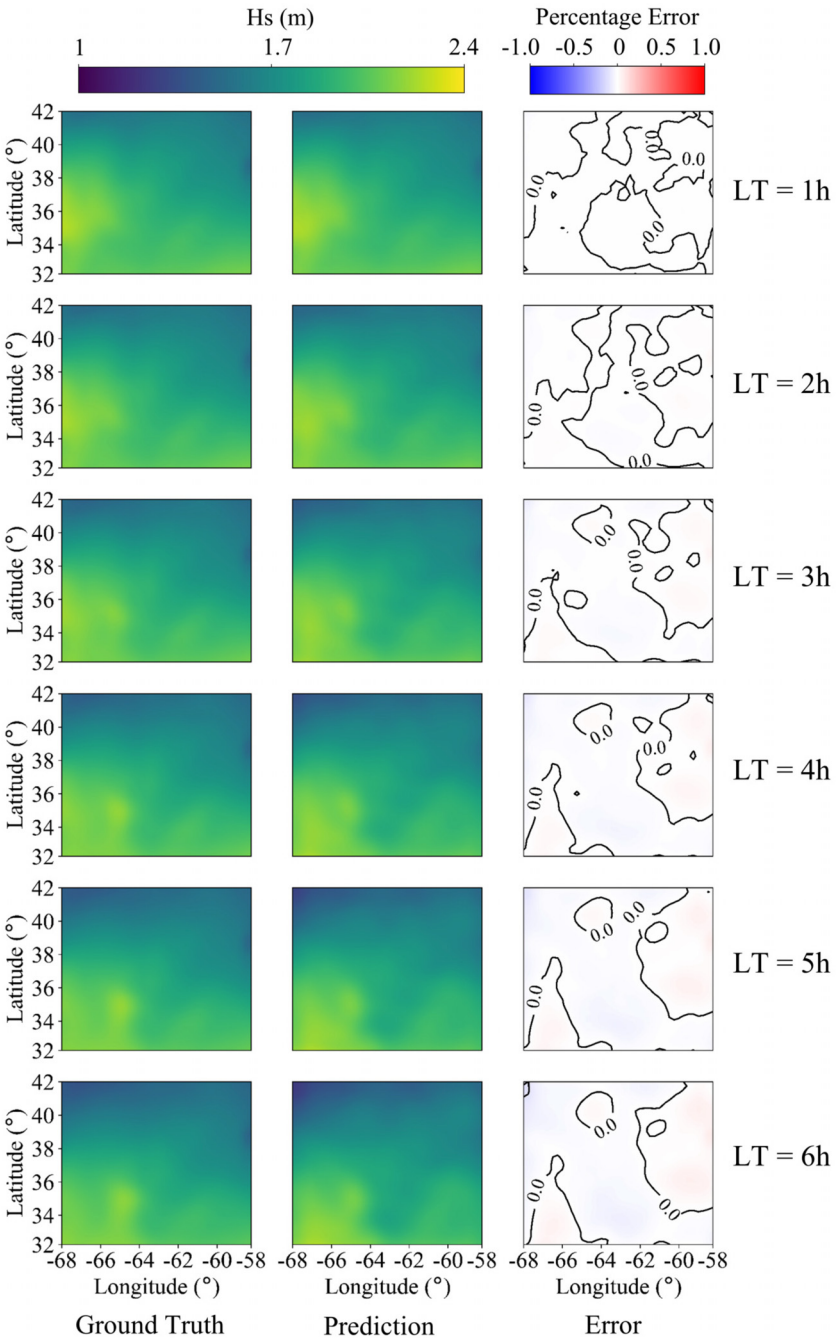


**FIG. 6.** Ground truth, predictions, and percentage errors of $H_s$ for sample 7321 (from 6:00 to 11:00 on November 2, 2021) in the test dataset, referring to the blue line in Fig. 4, with lead times of 1–6 h.

Consequently, the ST-RWP model is robust in predicting short-term wave fields (up to a 6-h lead time) under both strong-wave and mild-wave conditions.

### B. Rolled-out prediction of wave fields for 12-h lead times

In this section, we extend the rolled-out prediction to 12-h lead times. Notably, no re-training of the ST-RWP model is required for such extension.

As an overview of the model performance at the 12-h lead time, Fig. 7 presents the Mean-Absolute-Percentage Error (MAPE) for all samples in the test dataset, which reads

$$MAPE(h,w) = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{H}_i(h,w) - H_i(h,w)}{H_i(h,w)}\right|, \tag{7}$$

where $N$ denotes 8749 samples in the test dataset. As shown in this figure, the right half of the computational domain exhibits a MAPE of

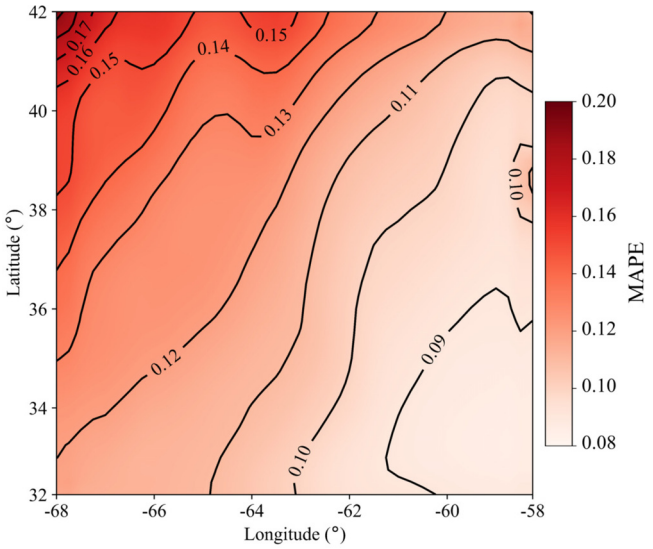


**FIG. 7.** MAPE of $H_s$ predictions for samples in the testing dataset with 12-h lead time.

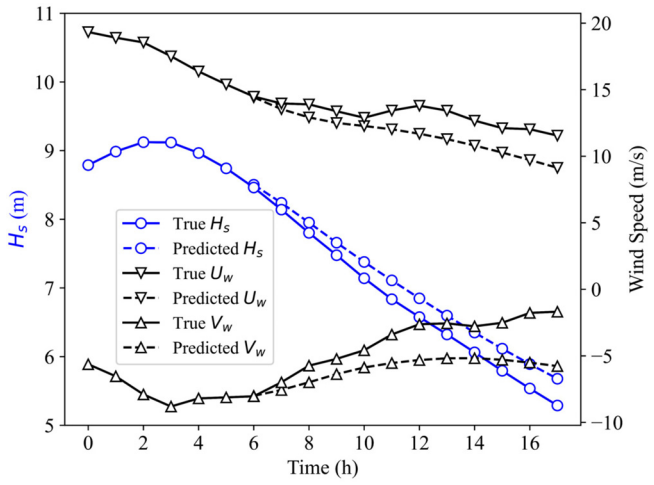


**FIG. 8.** True and predicted $H_s$, $U_w$, and $V_w$ values for sample 1462 (from 21:00 on March 2, 2021, to 14:00 on March 3, 2021) at Node 821 (37° N, 63° W). The first 6 h of data are used as model input for the subsequent 12 h through rolled-out prediction.

approximately 0.10, while the left section has a relatively higher MAPE ($< 0.2$). Indeed, for the 12-h lead time prediction, the preceding 6 h inputs are all calculated values by the ST-RWP model. Such MAPE in our 12-h lead time prediction is comparable to the MAPE reported in Liu *et al.* (2023) for their 1-h lead time wave field prediction (cf. their Fig. 10), suggesting the superiority of the ST-RWP model. Noticeably, the percentage error in our 1-h lead time wave field prediction is around 0.01, which is much smaller than that in Liu *et al.* (2023).

More interestingly, Fig. 7 does not exhibit significant mesh-like errors, which were observed in Liu *et al.* (2023) due to the up-sampling of latent feature maps in their deconvolution layer. In contrast, our ST-RWP model avoids compressing the feature fields into a latent space, maintaining constant spatial dimensions ($H$ and $W$) throughout the CNN and Swin Transformer blocks. This approach is crucial for the rolled-out wave field predictions, as the mesh-like errors may be drastically developed and transported within the computational domain, leading to contaminated forecasting results.

Although the wind speeds ($U_w$ and $V_w$) are not the primary focus of this study, their predictive accuracy still influences the rolled-out prediction strategy. Therefore, Fig. 8 illustrates the predicted $U_w$, $V_w$, and $H_s$ over a 12-h lead time for sample 1462 at node 821. Notably, a relatively high $H_s$ is observed during the initial 6 h. Both the errors in the significant wave height and wind velocities increase gradually with lead time. The relative errors in wind velocities are similar to those in significant wave height (slightly higher), suggesting that the model performs reasonably well in predicting both wind velocities and wave height.

The model has demonstrated superior performance in predicting short-term wave fields, specifically for 6-h and 12-h lead times. Several recent studies using deep-learning models have also achieved significant success in wave field prediction. This prompts an important question: what makes such data-driven model effective? In the next section, we will delve into the underlying mechanisms and discuss potential reasons for its success.

### C. The primary error sources in rolled-out prediction of long-term wave fields

The long-term performance of the ST-RWP model is also of interest. Predicting long-term wave fields presents substantial challenges even for numerical wave models, where data assimilation techniques are usually adopted to reduce error accumulation (e.g., Long and Thacker, 1989; Yoon *et al.*, 2015; and Wang and Pan, 2021). A thorough understanding of the ST-RWP model's limitations not only facilitates its practical application but also offers valuable guidance for future enhancements.

For instance, Fig. 9 presents a model-data comparison of significant wave height ($H_s$) at four selected lead times for each sample in the testing dataset at Node 821. The Correlation Coefficients (CC) between the predicted and observed $H_s$ values, along with the associated Root Mean Square Error (RMSE), are provided. The model demonstrates excellent performance for the lead times of 6 and 12 h, with the majority of the samples clustered around the line of perfect match, i.e., the solid line in Figs. 10(a) and 10(b). For mild waves ($H_s \leq 5\,m$), the model remains reasonably accurate at 18-h lead times, though it performs less well for higher $H_s$ values. This is expected, given that extreme wave events (e.g., during storms) represent only a small fraction of the training data. Afterward, the associated model inaccuracy further increases for $LT = 24\,h$ as shown in Fig. 10(d), suggesting that the model becomes invalid for such long-term prediction. These results highlight the need for event recognition methods in future works. For instance, historical data for extreme events can be extracted from long time-series datasets and resampled multiple times to ensure a balanced representation between normal and extreme $H_s$ conditions. Such strategy may enhance the model's ability to improve its accuracy in predicting large $H_s$ (e.g., Naeini and Snaiki, 2024a).

To further visualize the spatial distribution of errors across the entire computational domain, Fig. 10 presents a typical example (sample 1411 from the testing dataset) of the percentage error in the predicted significant wave height ($H_s$). Some marginal errors in $H_s$

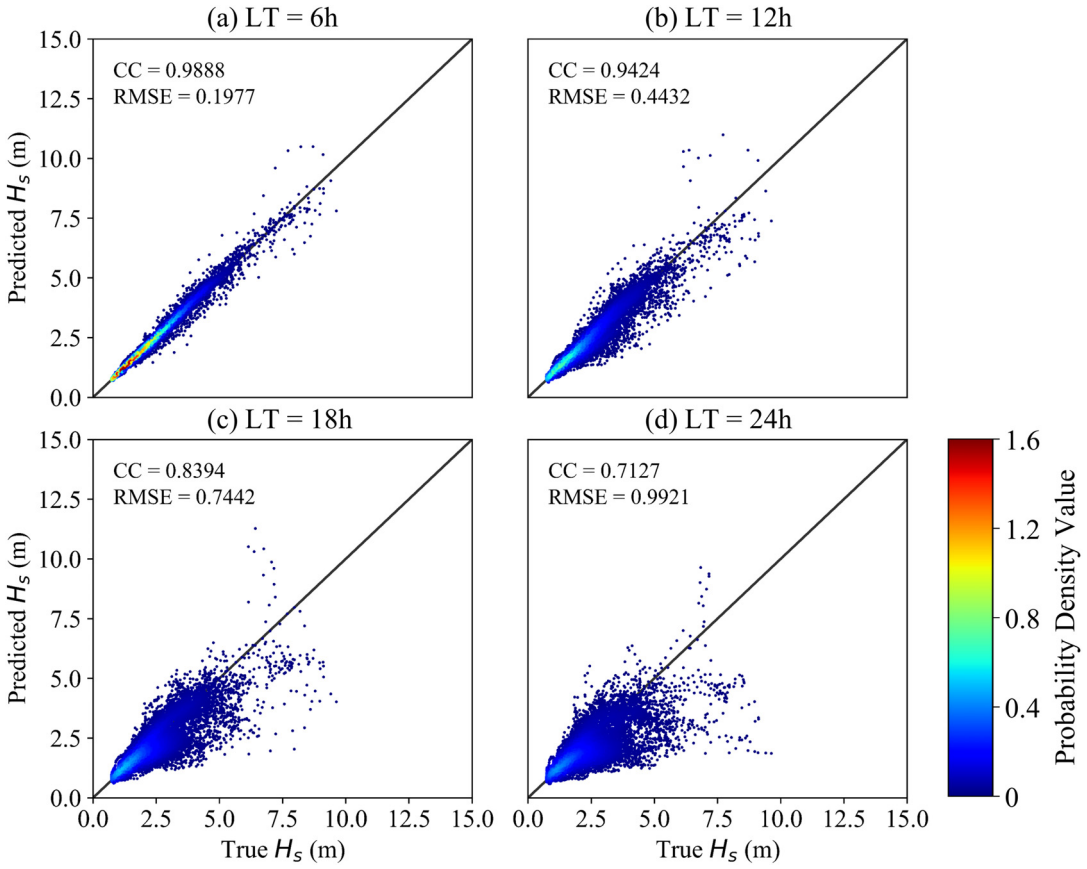


**FIG. 9.** Comparison of the predicted and true $H_s$ for the samples in the testing dataset at Node 821 (37° N, 63° W) with four different lead times: (a) $LT = 6\,h$, (b) $LT = 12\,h$, (c) $LT = 18\,h$, and (d) $LT = 24\,h$. Probability density values are marked.

(around 10%) originate from the top and right boundaries of the domain at a lead time of 6 h, which are then enlarged to be 20% ∼ 30% at $LT = 12\,h$. As the lead time further increases to 18 and 24 h, the prediction errors exceed 50% and propagate into the central region of the domain. Therefore, it is plausible that the major source of error is from the boundaries of the computational domain. To confirm it, we select two nodes: one situated at the top boundary (Node 42) and the other located at the center of the domain (Node 821, cf. Fig. 1). The error distributions for the two nodes, based on all testing samples at a 12-h lead time, are depicted in Figs. 11(a) and 11(b), respectively. To quantify the errors, we fit a Gaussian distribution to the data. As shown in the figure, the Node 42 has an absolute mean error 0.103, more than twice that of Node 821 (0.047). Similarly, the standard deviation (SD) of the errors is notably higher for Node 42. Hence, the errors in the boundary Node 42 are, indeed, higher than in the central Node 821.

An overview of the prediction error distribution at each node, for all testing samples at a 1-h lead time, is presented in Fig. 12, wherein both error percentiles and Mean-Absolute Error (MAE) are provided. Given that node IDs are assigned row-by-row from the upper boundary to the lower one (as shown in Fig. 1), a periodic pattern in the prediction errors becomes rather evident. A zoom-in view of Nodes 761–840 reveals this periodicity more clearly. Node 761 is located at the left boundary, while node 800 is at the right boundary. The prediction error decreases from Node 761 to Node 790 but then increases sharply as it approaches Node 800. This pattern repeats approximately every 40 nodes, indicating a consistent cyclical behavior in the error distribution.

In effect, the boundary nodes inherently have fewer neighboring nodes compared to central nodes, which restricts the ability of CNNs and Swin Transformer blocks to effectively leverage spatial information when predicting $H_s$ at these locations. This limitation underscores the importance of incorporating boundary conditions in the design of deep-learning-based RWP models. Future work should focus on addressing this challenge to improve prediction accuracy at boundary nodes.

## V. DISCUSSION

### A. Why does ST-RWP perform well in wave field prediction: Inductive bias

A deep-learning model should not be viewed merely as a tool for data fitting. A well-designed neural network structure takes advantage of the so-called inductive bias to guide the learning process, enabling

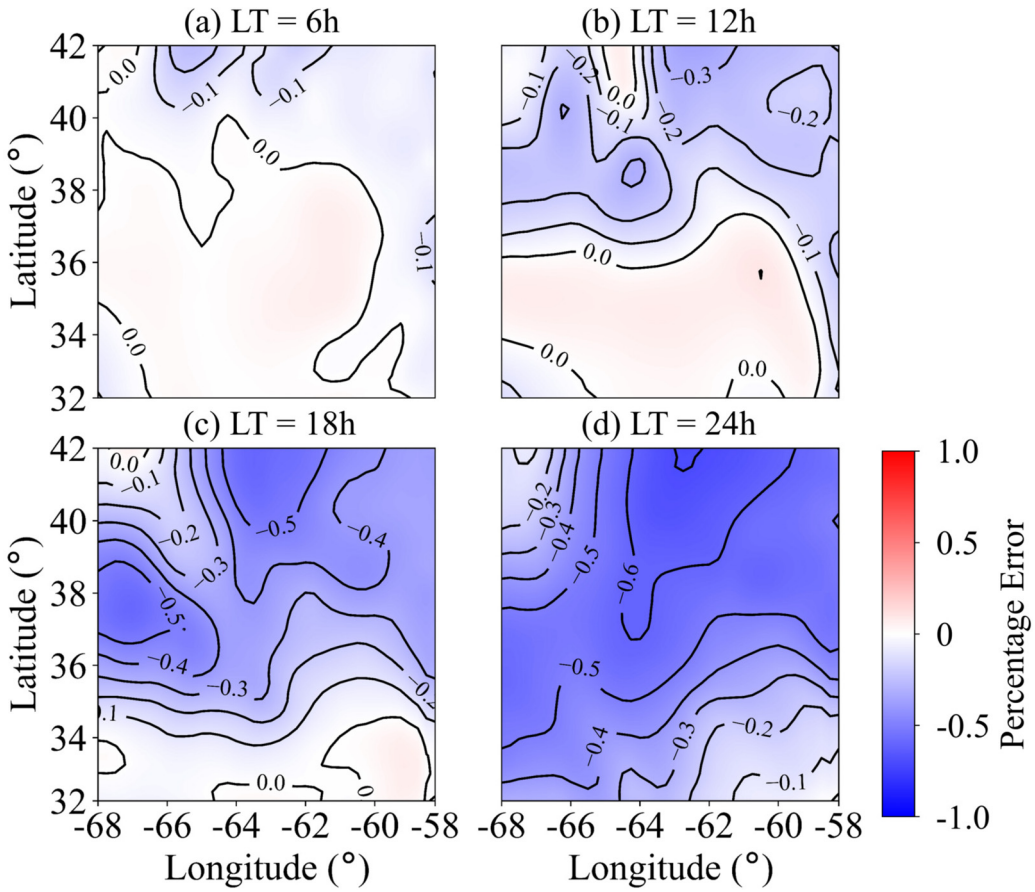


**FIG. 10.** Percentage errors of the predicted $H_s$ for sample 1411 (March 1, 2021, at 05:00, 11:00, 17:00, and 23:00) in the testing dataset at 4 selected lead times.

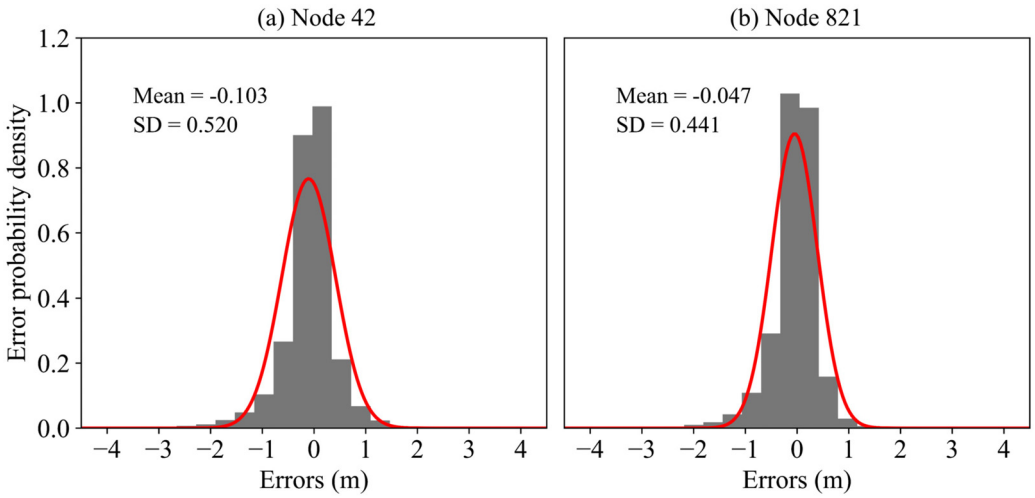


**FIG. 11.** Prediction errors for $H_s$ at lead time of 12 h in the test set samples at Node 42 (41.75° N, 67.75° W) and Node 821 (37° N, 63° W). The red lines represent the fitted Gaussian distributions.

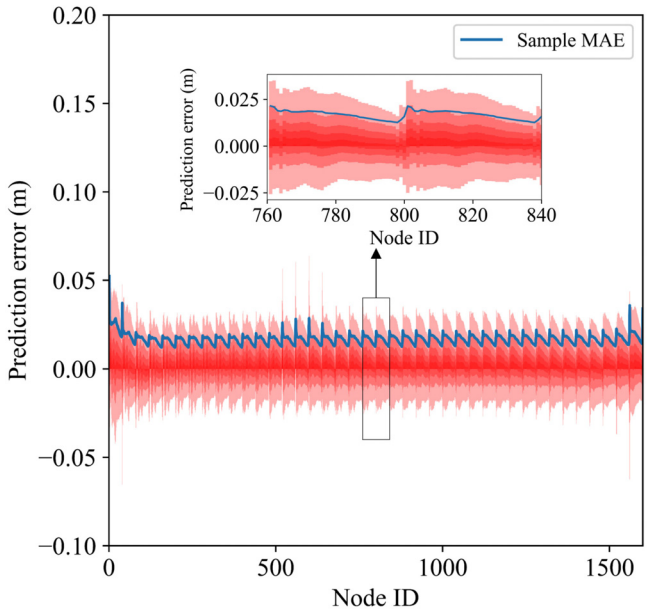


**FIG. 12.** Distribution of prediction errors for $H_s$ at a lead time of 1 h (with $K = 6$ and $M = 1$) in the testing dataset samples across all nodes. The error percentiles are presented from top to bottom, corresponding to the 90th, 80th, and down to the 10th percentile. The blue line represents the Mean-Absolute Error (MAE) for each node.

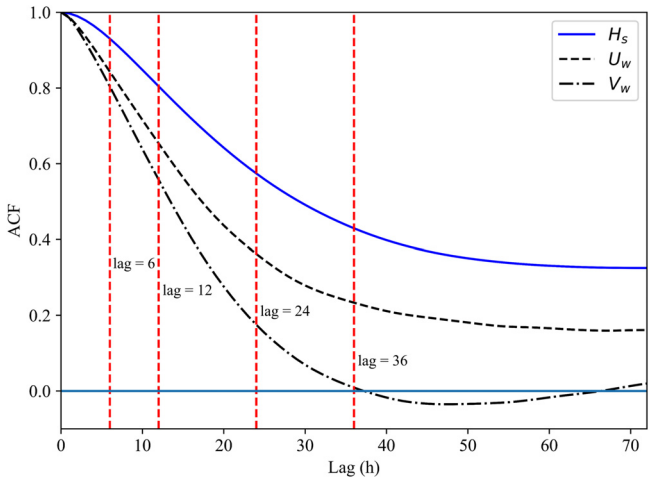


**FIG. 13.** Autocorrelation functions of $H_s$, $U_w$, and $V_w$ at node 821 (37° N, 63° W) from 2017 to 2021.

the model to generalize effectively from training data to unseen data (e.g., Wang and Wu, 2023).

As a black-box model, the ST-RWP integrates both CNNs and Swin Transformer modules to learn the historical data. The effectiveness of this architecture can be attributed to its carefully designed inductive biases, which align with the spatiotemporal characteristics of wind wave data.

When designing a deep neural network, inductive bias assumes that the data contain hidden relationships, which should be extracted through a well-structured model. For example, a CNN layer employs small, localized receptive fields (e.g., $3 \times 3$ filters used in this study) to capture local connectivity among neighboring nodes. Meanwhile, a Swin Transformer block excels at modeling long-range dependencies, making it particularly effective for long sequential wind wave fields. Correspondingly, the dataset should encompass both local and global relationships in spatial and temporal dimensions, respectively.

In this context, we calculate the autocorrelation function (*ACF*) at a selected central node (Node 821) to measure the temporal correlation of data at different time lags. The *ACF* quantifies the correlation between pairs of data points separated by a specified time interval. Following Tan *et al.* (2024), we employ a sliding window approach, where the window width is equal to the given time lag, to extract paired samples. The correlation coefficient is then computed for these sampled pairs over the period from 2017 to 2021. The time lags considered herein range from 1 h to 72 h. The resulting *ACF* values for $U_w$, $V_w$, and $H_s$ are presented in Fig. 13.

As shown in Fig. 13, the *ACF* values are very high ($>0.9$) for the three variables at 1-h time lag, indicating strong temporal correlation between consecutive wind wave fields. At a 6-h lag, the *ACF* remains high, with values of approximately 0.82 for $U_w$ and $V_w$, and around 0.9 for $H_s$. Such high correlation may explain the good prediction accuracy observed at a 6-h lead time, e.g., Fig. 9(a). Notably, the *ACF* values for $U_w$ and $V_w$ drop more rapidly than that for $H_s$, which may account for the superior performance of $H_s$ predictions, as illustrated in Fig. 8. For a 12-h time lag, the *ACF* for $H_s$ remains relatively high at 0.81, whereas it decreases to around 0.6 for both $U_w$ and $V_w$. At longer lags of 24 and 36 h, the *ACF* values for all three variables drop significantly below 0.6, which may be the reason to the deteriorated model performance at $LT \geq 24\,h$ observed in Fig. 10. After a 36-h time lag, the *ACF* value for $V_w$ turns negative, suggesting that its temporal correlation almost vanishes at such large time lag. Notably, while $U_w$ seems to have a slightly higher *ACF* (around 0.2) compared to $V_w$ at 70-h time lag, it may not be meaningful since we only consider one location (Node 821). The *ACF* for $H_s$ exhibits a more gradual decline, only approaching 0 (marked as the horizontal line in Fig. 13) until the time lag reaches 700 *h* (not shown in this figure).

Therefore, it is plausible that using an autoregressive strategy for predicting long-term wave fields (e.g., $LT \geq 24\,h$) is not feasible due to the rapid decline in temporal correlations. For short-term wave predictions, the ST-RWP model, by incorporating Swin Transformer blocks, effectively captures temporal dependencies and favors the predictions.

In addition to the temporal dynamics, the spatial correlation is also investigated among the 1600 nodes. The correlation coefficients for each pair of nodes are calculated and visualized in Fig. 14. Figure 14(a) clearly demonstrates high spatial correlation along the diagonal regions, indicating strong relationships between neighboring nodes. Notably, a patch-type pattern emerges, with each patch spanning a $40 \times 40$ area (bearing in mind that the node IDs are assigned row-by-row, and each row has 40 nodes). To better visualize this, a zoomed-in view of the correlation between Nodes 801–840 is provided in Fig. 14(b). This figure reveals that the spatial correlation decreases gradually from the diagonal areas toward the corners, confirming the high correlation between nearby nodes. Additionally, the maximum correlation values within these patches diminish as they move away from the diagonal regions, as seen in Fig. 14(a). Consequently, the high

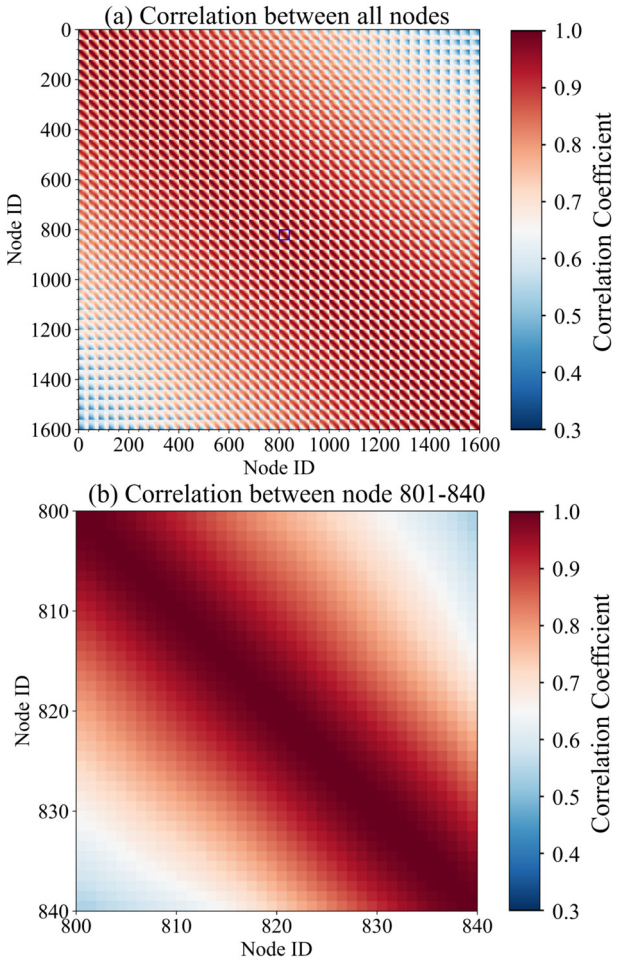


**FIG. 14.** Correlation coefficients of $H_s$ from 2017 to 2021 between different nodes. (a) The correlation coefficients between all points. (b) The correlation coefficients between all Node 801 to Node 840, as highlighted by the blue box in (a).

spatial correlation among neighboring nodes motivates the incorporation of CNN blocks in the ST-RWP model.

## B. Optimization of ST-RWP model structure

The CNN blocks in the ST-RWP model are designed to extract local connectivity among neighboring, while the Swin Transformer block captures both local and global relationships. To investigate the importance of the CNN blocks, we conducted ablation studies by removing the $Conv_2$ and $Conv_3$ layers from the ST-RWP model. The same training dataset and setups are utilized to ensure a fair comparison. Model performance is evaluated using the RMSE. The results are summarized in Table I.

As shown in the table, the model achieves the best performance across all lead times when both $Conv_2$ and $Conv_3$ are included. Removing $Conv_3$ results in a modest increase in RMSE, while the removal $Conv_2$ leads to a higher rise in RMSE. We conjecture that this difference arises from the distinct roles of these layers: $Conv_3$, placed at the end of the RST blocks, primarily refines local relationships through

**TABLE I.** The RMSE of $H_s$ predictions averaged across all testing samples with different lead times. ST-RWP is the base model, while other models have omitted $Conv_2$ and/or $Conv_3$.

| Lead time | ST-RWP | No $Conv_2$ | No $Conv_3$ | No $Conv_2$ and $Conv_3$ |
|---|---|---|---|---|
| 1 h | **0.0231** | 0.0246 | 0.0234 | 0.0247 |
| 2 h | **0.0469** | 0.0503 | 0.0472 | 0.0502 |
| 3 h | **0.0723** | 0.0779 | 0.0736 | 0.0784 |
| 4 h | **0.0985** | 0.1061 | 0.1017 | 0.1079 |
| 5 h | **0.1257** | 0.1347 | 0.1320 | 0.1387 |
| 6 h | **0.1543** | 0.1641 | 0.1648 | 0.1713 |

residual connections, while $Conv_2$ is integrated multiple times within each RST block, playing a more critical role in capturing fine-grained local features. Notably, removing both $Conv_2$ and $Conv_3$ simultaneously does not further deteriorate the model too much, indicating that the Swin Transformer blocks can still capture some local relationships, though with reduced effectiveness. This suggests that the Swin Transformer blocks compensate for the absence of CNN layers to some extent, highlighting their robustness in modeling both local and global relationships.

Then, we investigate several hyperparameters to optimize the model performance. Specifically, we focus on the number of RST blocks, the number of heads in the multi-head self-attention (MSA) mechanism, and the embedded feature dimensions due to $Conv_1$. To evaluate the impact of these hyperparameters, we use the RMSE as the performance metric. We re-train the model 11 times, each with different combinations of hyperparameters, to identify the locally optimal values. The optimal settings are found to be 4 RST blocks, 10 attention heads, and 60 embedding dimensions, which are used in this study (see Fig. 15).

## C. Comparison with existing models

Recent advances in deep learning have inspired diverse models for wave field prediction. To rigorously evaluate ST-RWP's performance, we benchmark it against two recent approaches, i.e., the Vision Transformer (ViT, Liu *et al.*, 2023) and the CNN-LSTM model (Zhang *et al.*, 2024). We faithfully replicate the ViT architecture following the implementation by Liu *et al.* (2023), while adopting comparable hyperparameters and training protocols for the CNN-LSTM reported by Zhang *et al.* (2024), to ensure fair comparison. We utilize the RMSE of $H_s$ predictions averaged across all testing samples for comparison. The results (see Table II) for ViT and CNN-LSTM are consistent with those reported in their respective papers.

As evidenced by Table II, ST-RWP demonstrates superior accuracy across all six lead times, achieving more than 30% and 54% reductions in RMSE compared to ViT and CNN-LSTM, respectively. This performance gain underscores the efficacy of our hybrid CNN-Swin Transformer architecture in capturing spatiotemporal features for short-term wave forecasting.

## VI. CONCLUSIONS

The present study introduces a novel Swin Transformer-based model for regional wave predictions (ST-RWPs), which demonstrates exceptional performance in short-term wave field forecasting while

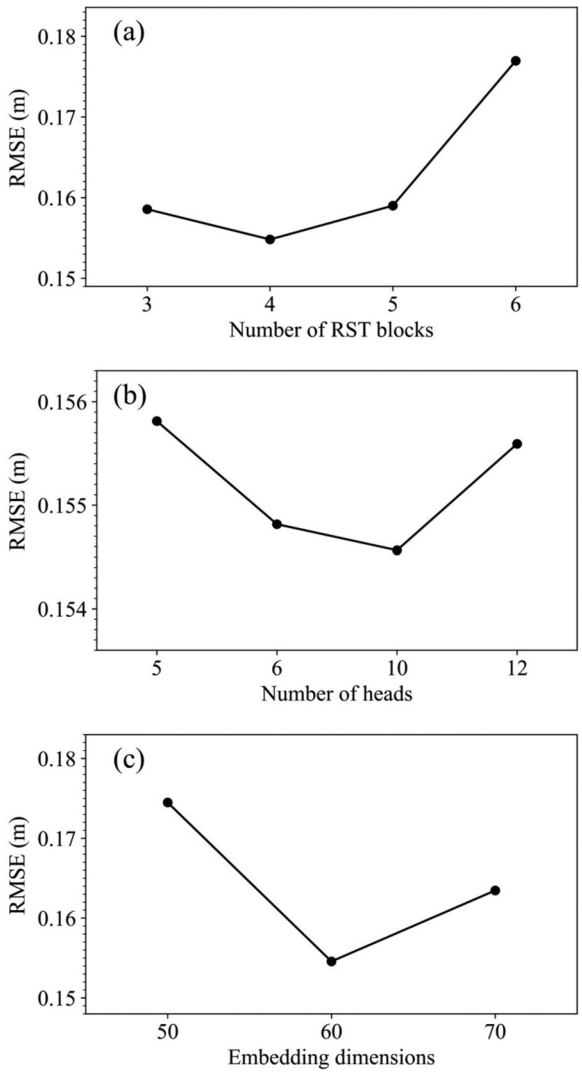


**FIG. 15.** The mean spatial RMSE of $H_s$ predictions at a 6-h lead time for all samples in the testing dataset with different hyperparameters.

proving less effective for long-term predictions. A systematic evaluation of the ST-RWP model for wave field prediction is conducted.

The model architecture is specifically designed for spatiotemporal wind-wave data, incorporating convolutional layers to capture local spatial relationships and Swin Transformer layers to extract global spatial and temporal features. Residual connections are extensively employed to integrate features extracted across different layers. To minimize the mesh-like predictive errors, the size of the latent feature maps remains invariant throughout the model, eliminating the need for up-sampling at the final layer of the model.

To train the model, we adopt a 5-year ERA5 reanalysis dataset of significant wave heights and wind velocities over the North Atlantic Ocean. A sliding window approach is applied to sample the dataset, using the first 4 years for training and validation, and the final year for testing. Model training employs mean square-error as the loss function. The training process achieves stabilization and convergence after 500 iterations. It should be noted that the currently selected data serve to demonstrate the ST-RWP model's performance in wave field forecasting. Future studies may consider larger regions and longer durations to fully take advantage of the ST-RWP model.

**TABLE II.** The RMSE of wave field predictions averaged across all testing samples for ST-RWP, ViT, and CNN-LSTM, with 6 different lead times.

| Lead time | ST-RWP | ViT | CNN-LSTM |
|---|---|---|---|
| 1 h | 0.0231 | 0.0483 | 0.1558 |
| 2 h | 0.0469 | 0.0746 | 0.1871 |
| 3 h | 0.0723 | 0.1067 | 0.2208 |
| 4 h | 0.0985 | 0.1425 | 0.2579 |
| 5 h | 0.1257 | 0.1816 | 0.2969 |
| 6 h | 0.1543 | 0.2240 | 0.3371 |

The well-trained ST-RWP model is rigorously tested using a rolled-out prediction scheme for ease of practical application. The model takes the preceding 6-h wind velocities and significant wave heights as inputs to forecast the subsequent hour. While the model excels in short-term predictions, its accuracy diminishes for long-term forecasts.

An analysis of prediction errors reveals that extreme wave heights are not well predicted at long lead times. This is attributed to the scarcity of storm surge events in the training dataset, limiting the model's ability to learn such patterns. Additionally, prediction errors are more pronounced at boundary nodes due to their limited spatial information compared to central nodes.

The model's inductive bias, a key factor for its success, is explored through the spatiotemporal correlation of the data. Autocorrelation analysis shows high temporal correlation for short lags in both wave height and wind velocities, which decreases significantly over longer lags (e.g., 24 h). Strong spatial correlations among neighboring nodes further enhance the model's performance, enabling efficient predictions for unseen data.

A systematic ablation study highlights the role of convolutional layers in improving the model performance, while the Swin Transformer block is able to capture both local and global relationships. Hyperparameters are carefully tuned for optimal results.

This study provides valuable insights into the application of Transformer-based deep learning approaches for wave field prediction, offering a foundation for future research in this area.

## ACKNOWLEDGMENTS

The work described in this paper was supported by a grant from the National Natural Science Foundation of China (Grant No. 52301316) and the National Key Research and Development Program of China (Grant No. 2023YFB2603803). It was also partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Reference Number: AoE/P-601/23-N).

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

## Author Contributions

**Weikai Tan:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Supervision (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Caihao Yuan:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – review & editing (equal). **Sudong Xu:** Funding acquisition (equal); Investigation (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal). **Yuan Xu:** Writing – review & editing (equal). **Alessandro Stocchino:** Conceptualization (equal); Writing – review & editing (equal).

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## APPENDIX: MODEL PERFORMANCE WITH VARYING INTRODUCTION PERIODS

The impact of introduction periods (i.e., the number of time steps input to the model) on model performance is investigated. To assess this effect, numerical experiments are conducted using eight different introduction periods (3, 6, 9, 12, 15, 18, 21, and 24 h) for predicting three lead times (1, 6, and 12 h). Model performance is evaluated using spatially averaged RMSE. As shown in Fig. 16, extending the introduction period has only marginal effects on performance for 1-h and 6-h lead times. For the 12-h lead time, increasing the introduction period does not consistently improve performance. While the model performs best with an 18-h introduction period (RMSE = 0.33), this improvement is limited compared to a 6-h introduction period (RMSE = 0.36). Notably, incorporating longer introduction periods beyond 6 h can even degrade performance in some cases.

This result is unsurprising, as the *ACF* values indicate that temporal correlation decreases significantly after 12 h. Including excessively long introduction periods may introduce irrelevant inputs, which could hinder rather than improve model performance. Additionally, incorporating more time steps increases the requirements of memory and computational costs during both model training and inference, potentially reducing model efficiency. Therefore, we opt for a 6-h introduction period in this study.

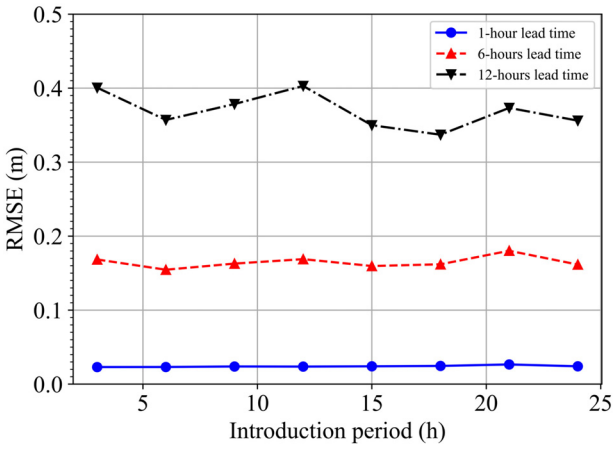


**FIG. 16.** The RMSE of $H_s$ predictions at the 6-h lead time averaged across all testing samples with 8 different introduction periods.

## REFERENCES


Bai, G., Wang, Z., Zhu, X., and Feng, Y., "Development of a 2-d deep learning regional wave field forecast model based on convolutional neural network and the application in south China sea," Appl. Ocean Res. **118**, 103012 (2022).

Bento, P., Pombo, J., Calado, M., and Mariano, S., "Ocean wave power forecasting using convolutional neural networks," IET Renewable Power Gener. **15**, 3341–3353 (2021).

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q., "Accurate medium-range global weather forecasting with 3d neural networks," Nature **619**, 533–538 (2023).

Booij, N., Ris, R. C., and Holthuijsen, L. H., "A third-generation wave model for coastal regions: 1. model description and validation," J. Geophys. Res. Oceans **104**, 7649–7666, https://doi.org/10.1029/98JC02622 (1999).

Daliri, M., De Leo, F., Loarca, A. M. L., Scovenna, M., Stocchino, A., Capello, M., Cutroneo, L., and Besio, G., "From hindcast to forecast: A statistical framework for real-time coastal circulation bulletins in the Gulf of Genoa," Appl. Ocean Res. **154**, 104337 (2025).

Dosovitskiy, A., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv:2010.11929 (2020).

Ellenson, A., Pei, Y., Wilson, G., Özkan-Haller, H. T., and Fern, X., "An application of a machine learning algorithm to determine and describe error patterns within wave model output," Coastal Eng. **157**, 103595 (2020).

Fan, S., Xiao, N., and Dong, S., "A novel model to predict significant wave height based on long short-term memory network," Ocean Eng. **205**, 107298 (2020).

Feng, X. C. and Xu, H., "Multi-station collaborative wave height prediction based on multi-feature identification and interpretable analysis," Phys. Fluids **36**, 076617 (2024).

Guan, X., "Wave height prediction based on CNN-LSTM," in *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)* (IEEE, 2020), pp. 10–17.

Hao, W., Sun, X., Wang, C., Chen, H., and Huang, L., "A hybrid EMD-LSTM model for non-stationary wave prediction in offshore China," Ocean Eng. **246**, 110566 (2022).

Hasselmann, K., Hasselmann, S., Bauer, E., Janssen, P., Komen, G., Bertotti, L., Lionello, P., Guillaume, A., Cardone, V., Greenwood, J *et al.*, "The WAM model-a third generation ocean wave prediction model," J. Phys. Oceanogr. **18**, 1775–1810 (1988).

He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 770–778.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D *et al.*, "The era5 global reanalysis," Q J. R. Meteorol. Soc. **146**, 1999–2049 (2020).

Huang, L., Jing, Y., Chen, H., Zhang, L., and Liu, Y., "A regional wind wave prediction surrogate model based on CNN deep learning network," Appl. Ocean Res. **126**, 103287 (2022).

James, S. C., Zhang, Y., and O'Donncha, F., "A machine learning framework to forecast wave conditions," Coastal Eng. **137**, 1–10 (2018).

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., Ewalds, T., Eaton-Rosen, Z., Hu, W *et al.*, "Learning skillful medium-range global weather forecasting," Science **382**, 1416–1421 (2023).

Law, Y., Santo, H., Lim, K., and Chan, E., "Deterministic wave prediction for unidirectional sea-states in real-time using artificial neural network," Ocean Eng. **195**, 106722 (2020).

Li, X., Cao, J., Guo, J., Liu, C., Wang, W., Jia, Z., and Su, T., "Multi-step forecasting of ocean wave height using gate recurrent unit networks with multivariate time series," Ocean Eng. **248**, 110689 (2022).

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R., "Swinir: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (IEEE, 2021), pp. 1833–1844.
Liu, P. L. F., "Model equations for wave propagations from deep to shallow water," Adv. Coastal Ocean Eng. **1**, 125–157 (1995).
Liu, Y., Huang, L., Ma, X., Zhang, L., Fan, J., and Jing, Y., "A fast, high-precision deep learning model for regional wave prediction," Ocean Eng. **288**, 115949 (2023).
Liu, Z., Fagherazzi, S., and Cui, B., "Success of coastal wetlands restoration is driven by sediment availability," Commun. Earth Environ. **2**, 44 (2021a).
Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (IEEE, 2021b), pp. 10012–10022.
Liu, Z. B., Fang, K. Z., and Cheng, Y., "A new multi-layer irrotational boussinesq-type model for highly nonlinear and dispersive surface waves over a mildly sloping seabed," J. Fluid Mech. **842**, 323–353 (2018).
Long, R. B. and Thacker, W. C., "Data assimilation into a numerical equatorial ocean model. i. the model and the assimilation algorithm," Dyn. Atmos. Oceans **13**, 379–412 (1989).
Lou, R., Wang, W., Li, X., Zheng, Y., and Lv, Z., "Prediction of ocean wave height suitable for ship autopilot," IEEE Trans. Intell. Transp. Syst. **23**, 25557–25566 (2022).
Luo, Q. R., Xu, H., and Bai, L. H., "Prediction of significant wave height in hurricane area of the Atlantic Ocean using the Bi-LSTM with attention model," Ocean Eng. **266**, 112747 (2022).
Mentaschi, L., Besio, G., Cassola, F., and Mazzino, A., "Performance evaluation of Wavewatch III in the Mediterranean Sea," Ocean Modell. **90**, 82–94 (2015).
Minuzzi, F. C. and Farina, L., "A deep learning approach to predict significant wave height using long short-term memory," Ocean Modell. **181**, 102151 (2023).
Naeini, S. S. and Snaiki, R., "A novel hybrid machine learning model for rapid assessment of wave and storm surge responses over an extended coastal region," Coastal Eng. **190**, 104503 (2024a).
Naeini, S. S. and Snaiki, R., "A physics-informed machine learning model for time-dependent wave runup prediction," Ocean Eng. **295**, 116986 (2024b).
Olah, C., Mordvintsev, A., and Schubert, L., "Feature visualization: How neural networks build up their understanding of images," *Distill* (2018).
Peregrine, D. H., "Long waves on a beach," J. Fluid Mech. **27**, 815–827 (1967).
Roome, E., Christie, D., and Neill, S., "Predicting coastal wave conditions: A simple machine learning approach," Appl. Ocean Res. **153**, 104282 (2024).
Tan, W., Stocchino, A., and Cai, Z., "Subspace time series clustering of meteocean data to support ocean and coastal hydrodynamic modeling," Ocean Eng. **313**, 119417 (2024).
Vaswani, A., Shazeer, N., Parmar, N., Uszkorei, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I., "Attention is all you need," in Advances in *Neural Information Processing Systems* (2017).
Wang, G. and Pan, Y., "Phase-resolved ocean wave forecast with ensemble-based data assimilation," J. Fluid Mech. **918**, A19 (2021).
Wang, N., Chen, Q., Zhu, L., and Sun, H., "Integration of data-driven and physics-based modeling of wind waves in a shallow estuary," Ocean Modell. **172**, 101978 (2022).
Wang, X. L., Feng, Y., and Swail, V., "North Atlantic wave height trends as reconstructed from the 20th century reanalysis," Geophys. Res. Lett. **39**, L18705, https://doi.org/10.1029/2012GL053381 (2012).
Wang, Y., Imai, K., Miyashita, T., Ariyoshi, K., Takahashi, N., and Satake, K., "Coastal tsunami prediction in Tohoku region, Japan, based on s-net observations using artificial neural network," Earth. Planets Space **75**, 154 (2023).
Wang, Z. and Wu, L., "Theoretical analysis of the inductive biases in deep convolutional networks," Adv. Neural Inf. Process. Syst. **36**, 74289–74338 (2023).
Woolf, D. K., Challenor, P., and Cotton, P., "Variability and predictability of the north Atlantic wave climate," J. Geophys. Res. Oceans **107**, 9-1–9-14, https://doi.org/10.1029/2001JC001124 (2002).
Yang, S., Deng, Z., Li, X., Zheng, C., Xi, L., Zhuang, J., Zhang, Z., and Zhang, Z., "A novel hybrid model based on STL decomposition and one-dimensional convolutional neural networks with positional encoding for significant wave height forecast," Renewable Energy **173**, 531–543 (2021).
Yoon, S., Kim, J., and Choi, W., "An explicit data assimilation scheme for a nonlinear wave prediction model based on a pseudo-spectral method," IEEE J. Ocean. Eng. **41**, 112–122 (2015).
Zanuttigh, B., Formentin, S. M., and Briganti, R., "A neural network for the prediction of wave reflection from coastal and harbor structures," Coastal Eng. **80**, 49–67 (2013).
Zhang, J., Luo, F., Quan, X., Wang, Y., Shi, J., Shen, C., and Zhang, C., "Improving wave height prediction accuracy with deep learning," Ocean Modell. **188**, 102312 (2024).
Zeiler, M., "Visualizing and understanding convolutional networks," in Computer Vision–ECCV 2014: 13th European Conference (2014).
Zheng, Z., Ma, X., Ma, Y., and Dong, G., "Wave estimation within a port using a fully nonlinear boussinesq wave model and artificial neural networks," Ocean Eng. **216**, 108073 (2020).