Journal of
Marine Science and Engineering

MDPI

Article

Significant Wave Height Forecasting Based on EMD-TimesNet Networks

Zhuxin Ouyang [1], Yaoting Gao [2], Xuefeng Zhang [1], Xiangyu Wu [3,*] and Dianjun Zhang [1,4,*]

[1] School of Marine Science and Technology, Tianjin University, Tianjin 300072, China; zhuxin98@foxmail.com (Z.O.); xuefeng.zhang@tju.edu.cn (X.Z.)
[2] Army 31016, PLA, Beijing 100094, China; jiaweiqi97@163.com
[3] Key Laboratory of Research on Marine Hazards Forecasting, National Marine Environmental Forecasting Center, Beijing 100081, China
[4] Key Laboratory of Ocean Observation Technology, Ministry of Natural Resources, National Ocean Technology Center, Tianjin 300112, China
* Correspondence: wxy@nmefc.cn (X.W.); zhangdianjun@tju.edu.cn (D.Z.)

Abstract: Significant Wave Height (SWH) is a crucial parameter in ocean wave dynamics, impacting coastal safety, maritime transportation, and meteorological research. Building upon the TimesNet neural network, a recent advancement in the realm of time series prediction in deep learning, this study proposes an integrated approach combining Empirical Mode Decomposition (EMD) with TimesNet, introducing the EMD-TimesNet model for SWH forecasting. The TimesNet model's multidimensional spatial mapping guarantees effective historical information extraction, while the EMD approach makes it easier to decompose subsequence characteristics inside the original SWH data. The predicted Root Mean Square Error (RMSE) and Correlation Coefficient (CC) values of the EMD-TimesNet model are 0.0494 m and 0.9936; 0.0982 m and 0.9747; and 0.1573 m and 0.9352 at 1 h, 3 h, and 6 h, respectively. The results indicate that the EMD-TimesNet model outperforms existing models, including the TimesNet, Autoformer, Transformer, and CNN-BiLSTM-Attention models, both in terms of overall evaluation metrics and prediction performance for diverse sea states. This integrated model represents a promising advancement in enhancing the accuracy of SWH predictions.

Keywords: significant wave height forecasting; TimesNet; empirical mode decomposition; deep learning

check for updates

Citation: Ouyang, Z.; Gao, Y.; Zhang, X.; Wu, X.; Zhang, D. Significant Wave Height Forecasting Based on EMD-TimesNet Networks. *J. Mar. Sci. Eng.* **2024**, *12*, 536. https://doi.org/10.3390/jmse12040536

Academic Editor: Merrick C. Haller

Received: 3 March 2024
Revised: 21 March 2024
Accepted: 22 March 2024
Published: 24 March 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Extreme marine phenomena impact human life, impede coastal city development, and threaten human well-being [1]. Hurricanes, in particular, are extremely dangerous marine weather phenomena that endanger the security and welfare of people living in coastal cities [2]. The passage of hurricanes entails intense winds and rainfall. Simultaneously, the enormous waves generated by hurricanes can cause ships to capsize, damage sea platforms, and result in significant disasters for sea transportation, construction, and fisheries. Accurate wave prediction can mitigate specific disasters and decrease economic losses in marine transportation, offshore oil and gas exploration, marine scientific research, water operations, and military activities [3]. Hence, precise analysis and forecasting of ocean waves can furnish essential marine meteorological reference parameters for the normal progression of maritime activities and prevent the detrimental impacts of extreme marine meteorological events on human production and life [4].

The marine meteorological environment is highly intricate, characterized by a combination of numerous waves originating from various directions, each with different amplitudes and periods [5]. A single wave is often not indicative. As a result, significant wave height (SWH) is used as the main ocean wave measurement metric [6]. Hence, the analysis of waves can be simplified to focus on SWH. Recognizing the significance and practical value of SWH prediction, methods for predicting SWH have evolved over the past few decades. The third-generation wave model is currently the most widely used numerical model for wave prediction. In 1988, the Wave Model (WAM) was first proposed by the cited

group for operational use [7]. Subsequently, in 1999, Booij and Holthuijsen developed the Simulating Waves Nearshore (SWAN) model for near-coastal applications [8]. Tolman later introduced the WAVEWATCH III model in 2002, building upon the WAM model [9]. While the third-generation wave model offers a more precise depiction of waves compared to its predecessors, its use still involves numerous empirical parameters, leading to limitations in the accuracy of wave simulation [10].

The recent surge in machine learning has significantly contributed to the rapid development of various fields [11,12]. With their benefits of strong nonlinear learning capabilities, low computing cost, and quick calculation speed, SWH prediction methods based on machine learning have garnered significant attention from researchers in recent years [13,14]. Deep learning methods require only the identification of factors related to the desired physical quantities. Mahjoobi employed a classical machine learning algorithm, decision trees, to predict SWH. Five years of historical data were used to train and test the model [15]. Etemad-Shahidi utilized M5′ model trees for predicting SWH in Lake Superior [16]. Due to its comprehensive mathematical foundation, Support Vector Regression (SVR) was employed by Mahjoobi [17], Cornejo-Bueno [18], Berbić [19], and others to acquire short- and medium-term predictions of SWH. SVR demonstrates faster speed and superior accuracy in low-feature-dimensional scenarios. However, in cases of large input dimensions, the training time of SVR significantly increases.

In recent years, deep learning has experienced rapid growth as a crucial branch of machine learning [20]. By combining the best features of each technique, the hybrid approach improves a model's resilience and predictive ability [21]. In particular, Long Short-Term Memory (LSTM) networks, enhanced versions of RNNs, exhibit robust capabilities in regard to processing time series data. Minuzzi and Farina [22] employed LSTM networks for forecasting SWH at seven distinct locations along the Brazilian coast. These networks were trained using buoy data and the ERA5 dataset, and SWH was predicted with an accuracy of nearly 87% when compared to the actual buoy data. Zhang [23] employed a spatio-temporal deep learning method to refine the SWH grid-point forecasts generated by the European Center for Medium-Range Weather Forecasts Integrated Forecast System global model (ECMWF-IFS). For real-time, rolling revisions of the 0 to 240 h SWH forecasts from ECMWF-IFS, this technique involves the use of a deep neural network with trajectory-gated recurrent cells. In comparison to the initial ECMWF SWH projections, the spring correction proved to be the most successful, with a mean absolute error reduction of 12.972–46.237%. Ikram and Cao [24] investigated the performance of a unique hybrid neuro-fuzzy model for the short-term (one hour to one day) prediction of SWH. This model combines the Adaptive-Network-based Fuzzy Inference System (ANFIS) method with the Marine Predator Algorithm (MPA), improving the accuracy of ANFIS-PSO and ANFIS-GA by 8.30% and 11.20%, respectively, in terms of root mean square errors during the prediction of a 1 h lead time in the test period. Peng and Li [25] proposed CBA-Net, a deep neural network model, for the prediction of regional SWH. This model integrates the attention mechanism, and training involved the use of wind and wave height data from the ERA5 dataset covering the South China Sea waters during 2011–2018 as input features. The results showed that SWH cannot be accurately predicted by using a convolutional neural network alone; the addition of the Bi-LSTM layer and the attention mechanism significantly improve the prediction of SWH. Luo and Xu [26] proposed a Bi-LSTM-with-attention-mechanism model. They also looked into the influencing factors, such as the input–output ratio and the combination of input features. Pang and Dong [27] incorporated recursive quantitative analysis (RQA) and the improved complete integrated empirical modal decomposition of adaptive noise (ICEEMDAN) into a deterministic and stochastic component decomposition (DSD) method. The results showed that wind data had a favorable effect on long-term forecasts and that the DSD methodology improved forecast accuracy. This hybrid model combines all three machine learning models with the DSD method for forecasting SWH. The unique nested artificial neural networks were proposed by Amin Mahdavi-Meymand, and Wojciech Sulisz [28] and used to predict SWH at 20 North Sea locations. According to

the findings, the nonlinear machine learning model that was developed outperformed the linear regression technique in terms of prediction accuracy by about 18.39%. The nested artificial neural network can increase the accuracy of the traditional model by up to 34%, according to statistical measures. Shi and Su [29] put forth a Transformer model based on the attention process. In order to achieve continuous time series prediction, the Transformer model is designed to be able to capture contextual data and sequence relationships.

Due to the potent nonlinear modeling capabilities of deep learning models, researchers have proposed numerous strategies for capturing intricate temporal variations in real-world time series for predicting SWH [30]. One approach essentially involves the utilization of RNNs to model continuous time points based on Markov assumptions [31]. Nevertheless, these methods frequently encounter challenges in capturing long-term correlations, and their effectiveness is impeded by the sequential computing paradigm. An alternative category of methods utilizes Temporal Convolutional Networks (TCNs) focused on the time dimension to extract change information [32]. However, owing to the localized nature of the one-dimensional convolutional kernel, they are constrained to modeling changes between adjacent time points, resulting in limitations in addressing long-term dependence. Transformers with attention processes have been employed extensively in sequence modeling recently [33]. To capture pairwise temporal connections between time points, many Transformer-based models used in time series analysis use the attention mechanism or one of its variations. However, because temporal dependencies can be extensively obscured by complicated temporal patterns, it is challenging for the attention mechanism to discern reliable connections directly from scattered time points [34].

In order to overcome the limitations of the existing models, in April 2023, a new model named TimesNet [35] was introduced, showcasing cutting-edge performance across multiple tasks in time series analysis and emerging as the leader in five key areas: long-term forecasting, short-term forecasting, missing-value filling, anomaly detection, and classification. Researchers have investigated complex temporal shifts in variations within periods (intra-period) and variations between periods (inter-period). With the use of Fast Fourier Transform (FFT)-based periodicity extraction, they used multi-period alignment to convert one-dimensional time series into two-dimensional tensors. The drawbacks of representing one-dimensional time series can be overcome using this trans-formative method.

Although the TimesNet model has demonstrated excellent prediction accuracy for numerous datasets, it is challenging to optimally fit the TimesNet model alone since the sources of waves are tied to a variety of factors and the SWH data fluctuate greatly. This study proposes an empirical mode decomposition (EMD) method that decomposes the complete set of empirical modes, combining the adaptive noise and TimesNet and eventually yielding good prediction accuracy. The EMD method is based on the Hilbert–Huang transform [36], which is able to reduce the limitations of nonlinear problems in SWH forecasting and is highly adaptable. The EMD method decomposes any input data into a series of intrinsic modal functions and their residual terms, which improves the processing efficiency of time series data. Based on this, this paper adopts an EMD-TimesNet neural network model based on the empirical modal decomposition method, i.e., EMD decomposition is used as a preprocessing tool for nonlinear data, and the decomposed data are forecasted. In previous studies, many researchers used a single buoy for prediction without considering the influence of terrain and water depth factors on the accuracy of prediction models. In this study, the latest research results from the deep learning community were used to improve prediction accuracy, and the SWH prediction performance of EMD-TimesNet under different terrain and water depth conditions was evaluated, yielding good results.

The rest of this paper is organized as follows: Section 2 describes the data and methodologies used in this investigation. Section 3 discusses the experimental data on each model's performance. Finally, Section 4 presents the conclusions.

2. Study Area and Data

2.1. Study Area and Data Collection

The National Oceanic and Atmospheric Administration (NOAA) of the United States provides data for monitoring and researching ocean and atmospheric conditions. The data used in this study were gathered from the Nation Data Buoy Center (NDBC, https://www.ndbc.noaa.gov/ (accessed on 21 March 2024)). It operates a network of buoys and gauging stations across the world's oceans and seas, designed to collect information on ocean weather, waves, ocean temperature, salinity, and other relevant data. The NDBC's buoys and gauging stations transmit real-time data through automated systems, and this information is publicly accessible on its website. The NDBC website offers access to both real-time and historical ocean and meteorological data, supporting various applications, including weather forecasting, marine research, navigation, and other areas related to ocean and atmospheric conditions. Twelve sets of wave-related data are available from this center, including WSPD, WVHT, WDIR, and MWD. These data are mainstream inputs in the SWH forecast [26]. Therefore, these four data features were selected for this study. Table 1 presents a thorough explanation of these properties.

Table 1. Description of variables and units.

Variables	Descriptions	Units	Sampling Interval
WVHT	Significant wave height	m	Every hour
WSPD	Wind speed	m/s	Every 10 min
WDIR	Wind direction	deg	Every 10 min
MWD	The direction of wave propagation during the dominant period (DPD).	deg	Every hour

To enhance the model's applicability in assessing various marine phenomena, five buoys were selected in the Atlantic hurricane zone. Figure 1 shows the locations of the buoy stations, while Table 2 provides corresponding geographic information. To ensure the model's capability to evaluate different ocean bathymetry scenarios, we opted for buoys with both nearshore and offshore attributes. Of all the stations, two (41008, 41025) are located in the nearshore area, with a depth of less than 100 m, while station 41010 is located in water that is 890 m deep. The other two stations (41046 and 41044) are situated far from land, with depths exceeding 5000 m.

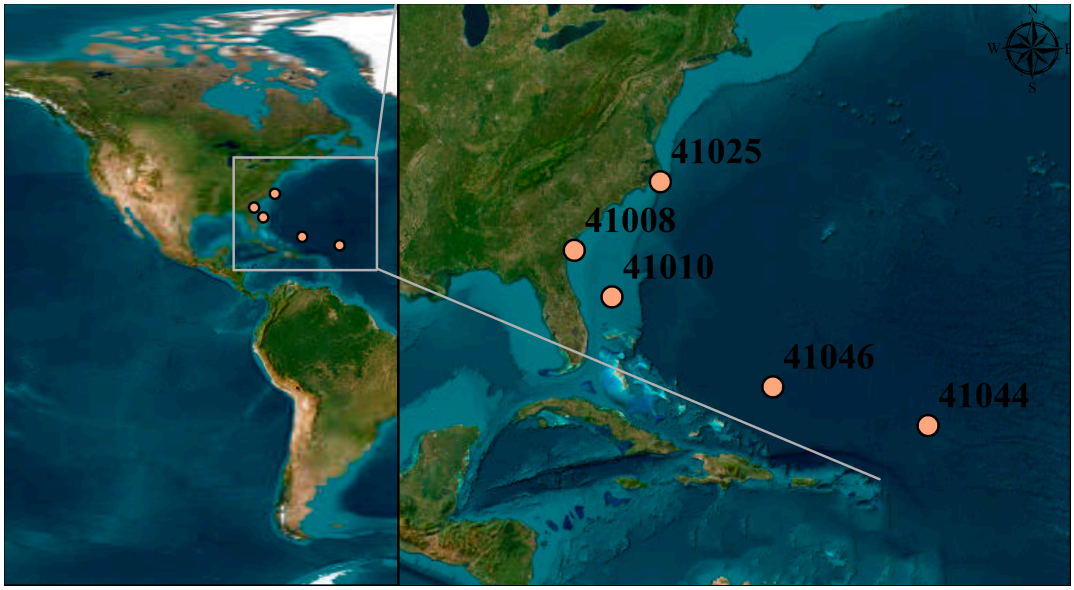


Figure 1. The locations of five stations.

Table 2. Variable descriptions and units.

ID	Latitude (N)	Longitude (W)	Depth (m)
41008	31.400	80.866	16
41025	35.010	75.454	48.8
41010	28.878	78.485	890
41044	21.582	58.630	5419
41046	23.822	68.393	5490

2.2. Data Preprocessing

Two buoy data, 41008 (at a water depth of 16 m) and 41046 (at a water depth with 5490 m), with different water depths and corresponding to the period of January 2019 to December 2020 were selected as the model's training set in this study to improve the prediction ability of the model at different water depths. The test set consisted of the remaining three stations. These stations, which are dispersed at various water depths, are 41010 (at a water depth of 890 m), 41025 (at a water depth of 48.8 m), and 41044 (at a water depth of 5419 m). The test set was completely independent from the training set, and the time span is from January to December 2021.

As depicted in Table 1, the sampling intervals for the four wave features, namely, WVHT, WSPD, WDIR, and MWD, exhibit inconsistency. Notably, WVHT and MWD have the lowest sampling frequency, set as one hour. Consequently, the data in this current study were collected at hourly intervals. For this purpose, the data were first processed via resampling to generate hour-by-hour datasets for the five stations. Furthermore, there were a few instances of missing values in the data, and these gaps were filled using a forward-filling method.

3. Methodology

3.1. EMD

EMD is a signal processing method designed for nonlinear and non-stationary time series. Following EMD processing, the original signal is decomposed into the sum of various Intrinsic Mode Functions (IMFs) determined by the local characteristic time scale of the signal. At the same time, the decomposed EMD signal is gradually smoothed. The basic EMD process is as follows:

(1) Let the original sequence be $x(t)$. Calculate the local maxima and minima in $x(t)$, and use the cubic spline difference method to obtain the upper and lower envelopes, $e_+(t)$ and $e_-(t)$, of the original sequence. The mean value is taken as the mean envelope $m_1(t)$ of the initial signal, corresponding to the expression

$$m_1(t) = \frac{e_+(t) + e_-(t)}{2} \tag{1}$$

where $e_+(t)$ represents the upper envelope of the original sequence, $e_-(t)$ represents the lower envelope of the original sequence, and $m_1(t)$ represents the average envelope.

(2) The initial signal is subtracted from the mean envelope to obtain a new signal that eliminates the low-frequency component of the new signal $h_1^1(t)$:

$$h_1^1(t) = x(t) - m_1(t) \tag{2}$$

(3) $h_1^1(t)$ is usually a non-smooth signal that does not satisfy the definitional conditions of the IMF, i.e., the difference between the number of extreme points and the number of points past zero for the entire initial signal is not greater than one, and the mean value of the upper and lower envelopes at any moment of the initial signal is 0. Repeating

the above process continuously until the kth repetition yields $h_1^k(t)$, which satisfies the above definition. The first-order IMF component of the initial signal is

$$c_1(t) = imf_1(t) = h_1^k(t) \tag{3}$$

(4) The initial signal is subtracted from $c_1(t)$ to obtain a new signal $r_1(t)$ with the high-frequency component removed:

$$r_1(t) = x(t) - c_1(t) \tag{4}$$

(5) Repeat the acquisition process of $c_1(t)$ for $r_1(t)$ to obtain the corresponding second IMF component $c_2(t)$, and continue the above process until the nth-order IMF component $c_n(t)$ or the residual component $r_n(t)$ of the signal is less than the set termination value, or the residual component $r_n(t)$ is a monotonic function or a constant; then, EMD signal separation is complete, and $x(t)$ decomposes into

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{5}$$

where $r_n(t)$ is the trend term, reflecting the average trend or mean of the signal.

3.2. TimesNet

3.2.1. Converting 1D Variants into 2D Variants

As shown in Figure 2, each time point involves two types of time series changes: intra-periodic changes and inter-periodic changes. However, the original one-dimensional time series structure can only represent changes between neighboring time points. To address this problem, TimesNet extends the time series variation to a two-dimensional structure, which can explicitly present intra-periodic and inter-periodic variations, thus providing additional advantages in terms of representation capability and facilitating subsequent representation learning.

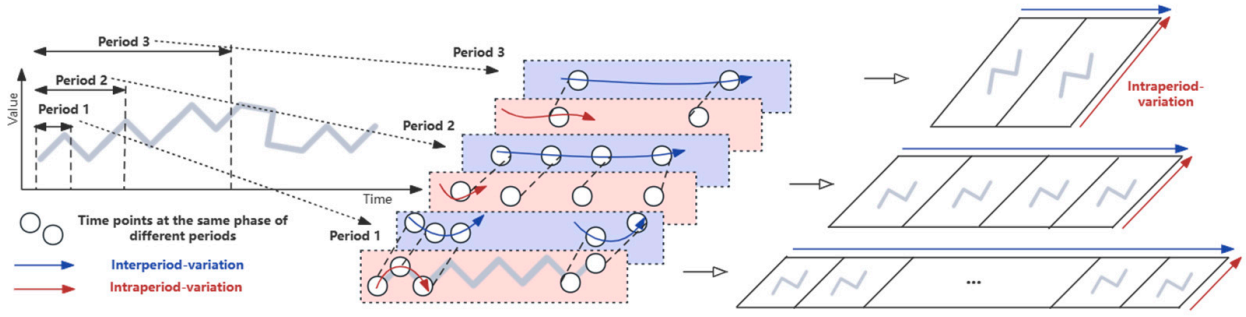


Figure 2. Multi-periodicity and temporal 2D-variation of time series.

Finding the time series' periodicity is a prerequisite for evenly representing the temporal variations within and between cycles. The FFT of the time dimension can be used to calculate the periodicity of a 1D time series $\mathbf{x}_{1D} \in \mathbb{R}^{T \times C}$ with length of time T and recorded variates C:

$$\mathbf{A} = Avg(Amp(FFT(\mathbf{X}_{1D}))) \tag{6}$$

$$\begin{gathered} f_1, \ldots, f_k = argTopk(A) \\ f_* \in \left\{1, \ldots, \left[\frac{T}{2}\right]\right\} \end{gathered} \tag{7}$$

$$p_i = \left[\frac{T}{f_i}\right], \; i \in \{1, \ldots, k\} \tag{8}$$

where $\mathbf{A} \in \mathbb{R}^T$ represents the intensity of each frequency component of $\mathbf{x}_{1D}$; the most significant k cycle lengths $\{p_1, \cdots, p_k\}$ are correlated with the highest-intensity k frequencies $\{f_1, \cdots, f_k\}$. The method described above can be shortened as follows:

$$\mathbf{A}, \{f_1, \cdots, f_k\}, \{p_1, \cdots, p_k\} = \text{Period}(\mathbf{X}_{1D}) \tag{9}$$

The initial one-dimensional time series can then be folded based on chosen periods, as shown in Figure 3, a process that is as simple as this:

$$\mathbf{X}^i_{2D} = \text{Reshape}_{p_i, f_i}(\text{Padding}(\mathbf{X}_{1D})), i \in \{1, \cdots, k\} \tag{10}$$

where Padding$(\cdot)$ is the sequence's complement of 0 at the conclusion, making the sequence's length divisible by p_i. After performing the aforementioned processes, a collection of two-dimensional tensors $\left\{\mathbf{X}^1_{2D}, \mathbf{X}^2_{2D}, \ldots, \mathbf{X}^k_{2D}\right\}$ is generated, with $\mathbf{X}^i_{2D}$ denoting temporal fluctuations in two dimensions that are dominated by cycle p_i.

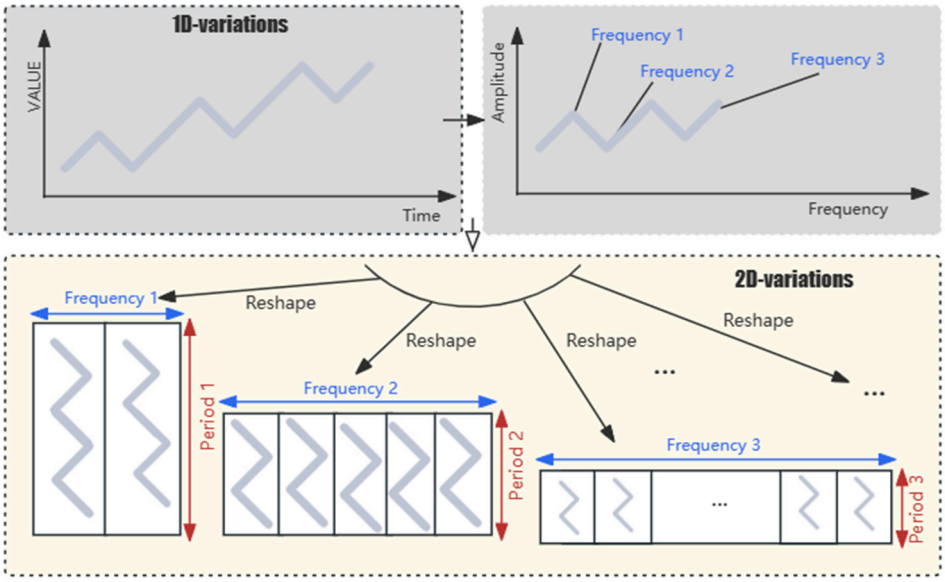


Figure 3. Conversion of a 1D time series into 2D.

It is important to note the previously described 2D vectors since every column and row correlates with a surrounding moment and a neighboring period, and similar temporal fluctuations are frequently indicated by the neighboring moments and periods. Consequently, the aforementioned 2D tensor will display 2D locality, making it simple to use 2D convolution to collect the information.

3.2.2. TimesBlock

Figure 4 depicts the TimesNet structure; here, it can be seen that TimesNet is a stack of multiple TimesBlocks with residual connections. In each TimesBlock, the sequence first finds the different cycles in the data via FTT [37]. It is then reshaped into a 2D vector and sent to the Inception block, where it learns and predicts the 2D representation of the sequence. Finally, this depth representation must be reshaped back into a 1D vector using adaptive aggregation.

To obtain the depth features, the input sequence $\mathbf{X}^0_{1D} \in \mathbb{R}^{T \times d_{\text{model}}}$ is initially passed through the embedding layer. $\mathbf{X}^{l-1}_{1D} \in \mathbb{R}^{T \times d_{\text{model}}}$ is the input for the TimesBlock's lth layer. Next, by using 2D convolution, the 2D temporal variations are extracted:

$$\mathbf{X}^l_{1D} = \text{TimesBlock}\left(\mathbf{X}^{l-1}_{1D}\right) + \mathbf{X}^{l-1}_{1D} \tag{11}$$

where $\mathbf{X}_{1D}^{l}$ denotes the lth layer TimesBlock and the input is $\mathbf{X}_{1D}^{l-1} \in \mathbb{R}^{T \times d_{\text{model}}}$.

Specifically, as shown in Figure 5, TimesBlock contains the following sub-processes:

1. The conversion of 1D features to 2D: to represent the two-dimensional timing changes, the cycles are first extracted for the input one-dimensional timing feature $\mathbf{X}_{1D}^{l-1}$ and then converted into a two-dimensional tensor, a procedure that is outlined in Equations (5) and (6).
2. Extracting 2D time-varying representations: for a 2D tensor $\left\{\mathbf{X}_{2D}^{l,1}, \mathbf{X}_{2D}^{l,2}, \ldots, \mathbf{X}_{2D}^{l,k}\right\}$, because of its 2D localization, 2D convolution can be used to extract information. Here, the classical Inception model is selected [38]:

$$\hat{\mathbf{X}}_{2D}^{l,i} = \text{Inception}\left(\mathbf{X}_{2D}^{l,i}\right) \tag{12}$$

where $\hat{\mathbf{X}}_{2D}^{l,i}$ denotes the 2D tensor $\left\{\mathbf{X}_{2D}^{l,1}, \mathbf{X}_{2D}^{l,2}, \ldots, \mathbf{X}_{2D}^{l,k}\right\}$.

3. The transformation from 2D to 1D: timesblock returns the extracted time-series features to a one-dimensional space in order to aggregate the data:

$$\hat{\mathbf{X}}_{1D}^{l,i} = \text{Trunc}(\text{Reshape}_{1,(p_i \times f_i)}(\hat{\mathbf{X}}_{2D}^{l,i})), i \in \{1, \cdots, k\} \tag{13}$$

where $\hat{\mathbf{X}}_{1D}^{l,i} \in \mathbb{R}^{T \times d_{\text{model}}}$, $Trunc(\cdot)$ denote the removal of the 0 supplemented by the operation in the $Padding(\cdot)$ of step 1.

4. Adaptive fusion: comparable to the Autoformer's design [39], the resulting one-dimensional representation $\{\hat{\mathbf{X}}^{l,1}, \cdots, \hat{\mathbf{X}}^{l,k}\}$ is weighted and then summed with the intensity of its matching frequency:

$$\hat{\mathbf{A}}_{f_1}^{l-1}, \cdots, \hat{\mathbf{A}}_{f_k}^{l-1} = \text{Softmax}\left(\mathbf{A}_{f_1}^{l-1}, \cdots, \mathbf{A}_{f_k}^{l-1}\right) \tag{14}$$

$$\mathbf{X}_{1D}^{l} = \sum_{i=1}^{k} \hat{\mathbf{A}}_{f_i}^{l-1} \times \hat{\mathbf{X}}_{1D}^{l,i}. \tag{15}$$

where $\hat{\mathbf{A}}_{f_k}^{l-1}$ denotes the intensity of the one-dimensional representation $\left\{\hat{\mathbf{X}}^{l,1}, \cdots, \hat{\mathbf{X}}^{l,k}\right\}$ corresponding to the frequency, and $\mathbf{X}_{1D}^{l}$ denotes the final output.

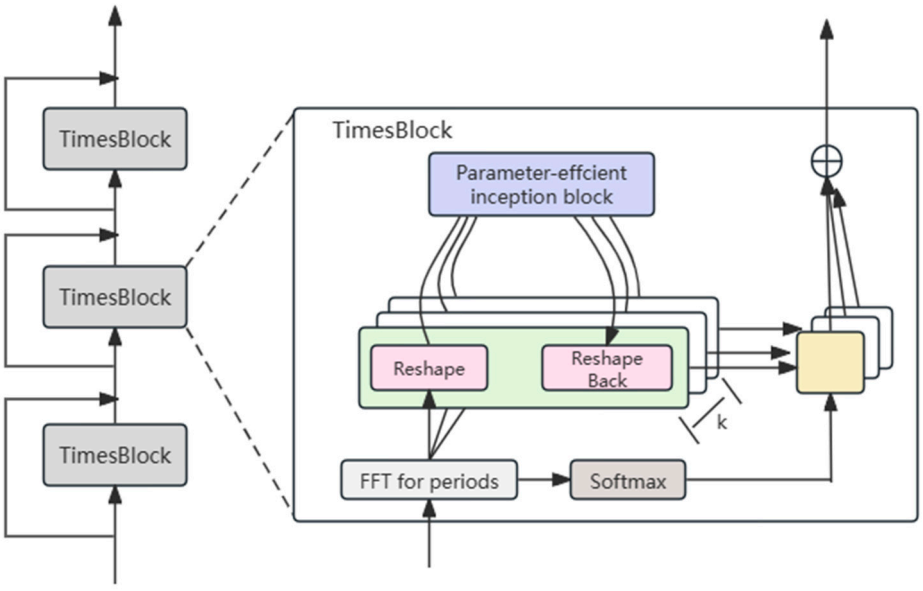


Figure 4. The structure of TimesNet.

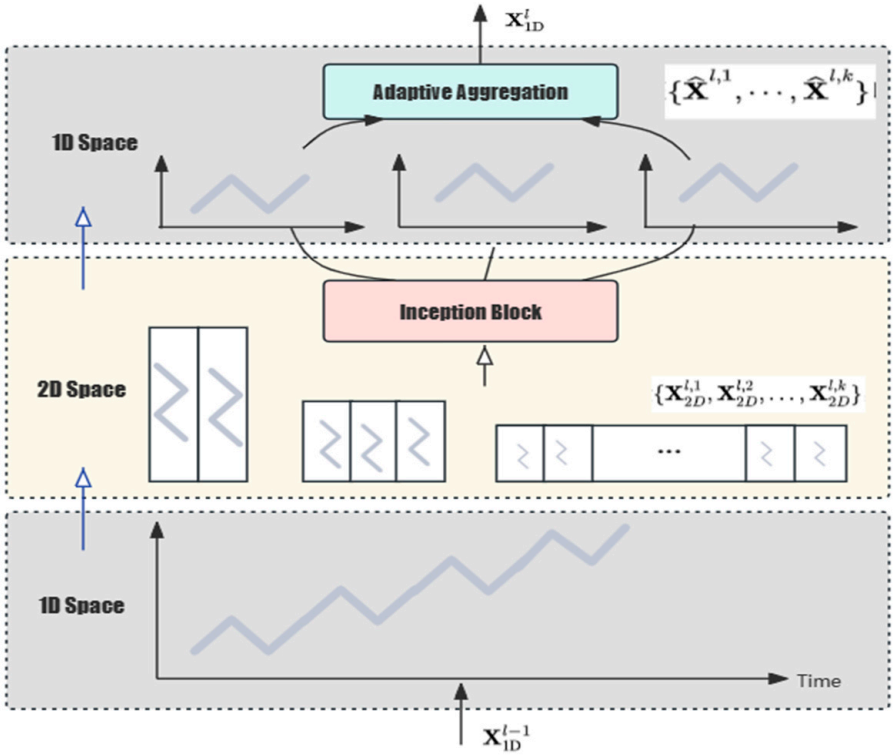


Figure 5. Subprocesses of TimesBlock.

Through the above design, TimesNet completes the time-varying modeling process of extracting 2D time-varying features from multiple cycles and then fusing them adaptively [35]. It is worth noting that since TimesNet converts 1D temporal features into 2D tensors for analysis, it can directly adopt advanced visual backbone networks for feature extraction, such as Swin Transformer, ResNeXt, ConvNeXt, etc. This design also allows the tasks of temporal and time-series analysis to directly benefit from the advanced visual backbone networks.

3.3. Parameter Settings

Before training, a few crucial parameters for deep learning neural networks must be manually established [40]. Each of the five models applied in this study was established with the appropriate parameters, and in order to minimize the impact of different parameter settings on model performance, the same parameters were set for EMD-TimesNet and TimesNet, which have similar model structures, as well as for Autoformer and Transformer. In addition, the CNN-BiLSTM-attention model has a different structure, so different parameters were chosen. Furthermore, in this study, three prediction ranges of 1 h, 3 h, and 6 h were set. The input/output ratios were configured as 9:1, 9:3, and 9:6, meaning that we utilized data spanning 9 h to forecast the subsequent 1 h, 3 h, and 6 h of SWH, respectively. The details of the parameter settings are shown in Table 3.

Table 3. Model parameter settings.

Model	Memory Cell	Batch Size	Attention Cell	Optimizer	Dropout	Activation
EMD-TimesNet	d_model-64/fcn-2048	32	multi_head-8	adam	0.1	gelu
TimesNet	d_model-64/fcn-2048	32	multi_head-8	adam	0.1	gelu
Autoformer	d_model-512/fcn-2048	32	multi_head-8	adam	0.05	gelu
Transformer	d_model-512/fcn-2048	32	multi_head-8	adam	0.05	gelu
CNN-BiLSTM-Attention	80	256	160	adam	0.01	tanh

3.4. Evaluation Metrics

Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Absolute Percentage Error (MAPE), and Correlation Coefficient (CC) were used to evaluate the effectiveness of the model. The formulas of the four metrics are provided as follows:

$$CC = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(\hat{y}_i - \overline{y'})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2(\hat{y}_i - \overline{y'})^2}} \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{17}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{18}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{19}$$

where y_i is the original measurement value, $\hat{y}_i$ is the forecasted value, $\overline{y}$ is the mean of the original measurement values, $\overline{y'}$ is the mean of the forecasted values, and n is the number of samples.

4. Results and Discussion

In this study, three prediction ranges, 1 h, 3 h, and 6 h, were set. The input/output ratios were configured as 9:1, 9:3, and 9:6, meaning that we utilized data spanning 9 h to forecast the subsequent 1 h, 3 h, and 6 h of SWH data, respectively. The evaluation metrics MAE, MAPE, RMSE, and CC are used in this section to evaluate various elements of the prediction outcomes. In order to represent the overall prediction level of each model, the forecasts were averaged after being repeated five times in every delivery cycle at each site, taking into account the stochastic nature of deep learning predictions.

4.1. Overall Forecast Assessment

4.1.1. Overall Predictive Assessment of Buoy 41010

The specific values of MAE, MAPE, RMSE, and CC for buoy 41010 are shown in Table 4. Additionally, as shown in Figure 6, each of the distinct prediction metrics is displayed as a line graph to better illustrate the comparison of the various models' prediction levels.

As illustrated in Figure 6, the ranked order of performance for each model is as follows: EMD-TimesNet > TimesNet > Autoformer > Transformer ≈ CNN-BiLSTM-Attention. Based on the specific values, when the prediction time is 1 h, the accuracy of each model is very good; the accuracy of CNN-BiLSTM-attention is the worst, but the CC value can still reach 0.9408, showing a good correlation. In terms of accuracy, EMD-TimesNet performs the best: all of its evaluation index values are the best among the five models, with an MAE of 0.0348 m, an RMSE of 0.0494 m, an MAPE of 2.51%, and a CC value of 0.9936, showing very strong correlation and excellent prediction performance. For 1 h prediction, the MAE of EMD-TimeNet is 0.0348 m, which is 21.26%, 50.36%, 52.72%, and 65.68% lower than that of TimesNet, Autoformer, Transformer, and CNN-BiLSTM-Attention, respectively; EMD-TimeNet's RMSE is 0.0494 m, which is 27.46%, 48.54%, 58.10%, and 67.15% lower than that of TimesNet, Autoformer, Transformer, and CNN-BiLSTM-Attention, respectively; and the CC value of EMD-TimeNet is 0.9936, which is lower than that of TimesNet, Autoformer, Transformer, and CNN-BiLSTM-Attention by 0.59%, 1.81%, 3.11%, and 5.61%, respectively. When the prediction time is 3 h, EMD-TimesNet's performance is still the best: all of its evaluation index values are the best among the five models; its MAE is 0.0653 m, its RMSE is 0.0982 m, its MAPE reached 4.58%, and its CC value is 0.9747. The accuracies of

Transformer and CNN-BiLSTM-Attention are similar; that is, the accuracy of the prediction of SWH by these two models is roughly comparable. When the prediction time is 6 h, EMD-TimesNet still provides the best performance: all of its e evaluation indices are better than those of the other four models, with an MAE of 0.1026 m, an RMSE of 0.1573 m, an MAPE reaching 7.12%, and a CC value of 0.9352. The accuracies of Transformer and CNN-BiLSTM-Attention are similar; specifically, their prediction accuracies for SWH are about the same.

Table 4. Comparison of prediction errors for 1 h, 3 h, and 6 h between EMD-TimesNet, TimesNet, Autoformer, Transformer, and CNN-BiLSTM-Attention at 41010. The bold text in the table indicates the best prediction.

Lead Time	Model	MAE	RMSE (m)	CC	MAPE
1 h	EMD-TimesNet	**0.0348**	**0.0494**	**0.9936**	**2.51%**
	TimesNet	0.0442	0.0681	0.9878	3.05%
	Autoformer	0.0701	0.0960	0.9759	5.12%
	Transformer	0.0736	0.1179	0.9636	4.93%
	CNN-BiLSTM-Attention	0.1014	0.1504	0.9408	6.97%
3 h	EMD-TimesNet	**0.0653**	**0.0982**	**0.9747**	**4.58%**
	TimesNet	0.0796	0.1246	0.9593	5.41%
	Autoformer	0.1093	0.1609	0.9322	7.76%
	Transformer	0.1352	0.2040	0.8910	9.49%
	CNN-BiLSTM-Attention	0.1348	0.2007	0.8945	9.19%
6 h	EMD-TimesNet	**0.1026**	**0.1573**	**0.9352**	**7.12%**
	TimesNet	0.1427	0.2233	0.8693	9.66%
	Autoformer	0.1613	0.2358	0.8544	11.40%
	Transformer	0.1810	0.2664	0.8140	12.56%
	CNN-BiLSTM-Attention	0.1817	0.2675	0.8125	12.39%

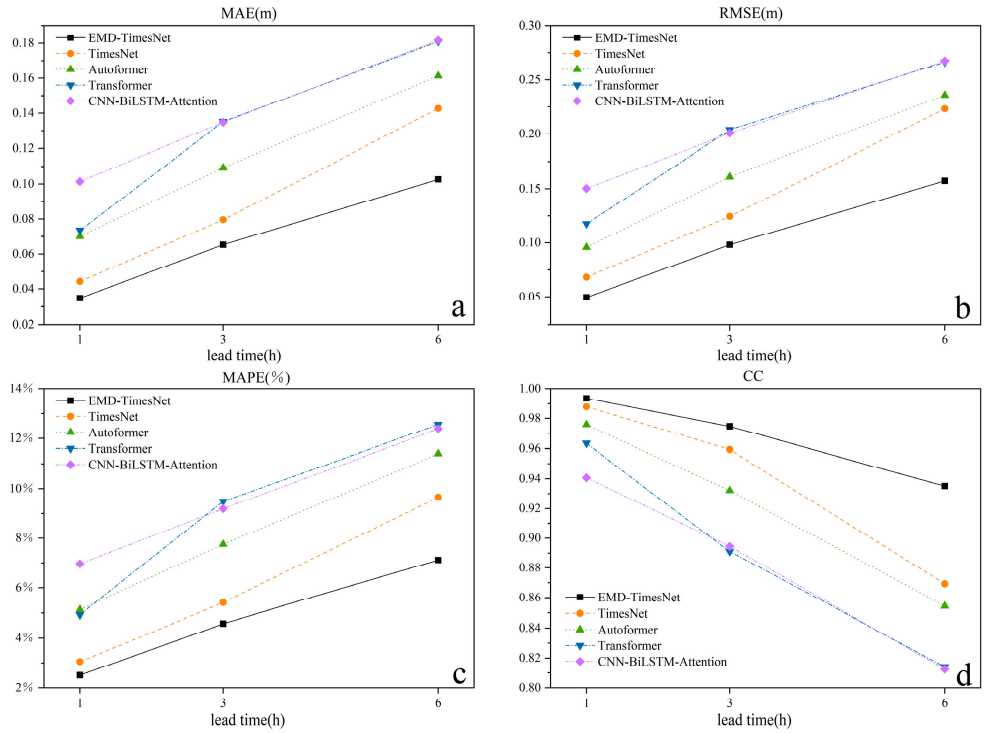


Figure 6. Comparative plots of the results obtained for EMD-TimesNet (black), TimesNet (orange), Autoformer (green), Transformer (blue), and CNN-BiLSTM-Attention (purple) models for predicting 1 h, 3 h, and 6 h results at 41010. (**a**) MAE; (**b**) RMSE; (**c**) MAPE; (**d**) CC.

4.1.2. Overall Predictive Assessment of Buoy 41025

The specific MAE, MAPE, RMSE, and CC values for buoy 41025 are shown in Table 5, and corresponding line graphs are shown in Figure 7.

Table 5. Comparison of prediction errors for 1 h, 3 h, and 6 h between EMD-TimesNet, TimesNet, Autoformer, Transformer, and CNN-BiLSTM-Attention at 41025. The bold text in the table indicates the best prediction.

Lead Time	Model	MAE(m)	RMSE (m)	CC	MAPE
1 h	EMD-TimesNet	0.0458	**0.0679**	**0.9886**	3.79%
	TimesNet	**0.0405**	0.0697	0.9864	**2.10%**
	Autoformer	0.0823	0.1131	0.9685	7.00%
	Transformer	0.1101	0.1916	0.9094	10.17%
	CNN-BiLSTM-Attention	0.1708	0.2300	0.8696	12.43%
3 h	EMD-TimesNet	0.0890	**0.1315**	**0.9573**	6.87%
	TimesNet	**0.0756**	0.1511	0.9360	**3.83%**
	Autoformer	0.1439	0.2180	0.8828	10.90%
	Transformer	0.1687	0.2591	0.8345	13.26%
	CNN-BiLSTM-Attention	0.2103	0.2913	0.7907	15.31%
6 h	EMD-TimesNet	0.1502	**0.2279**	**0.8719**	11.18%
	TimesNet	**0.1250**	0.2552	0.8173	**6.26%**
	Autoformer	0.2054	0.3057	0.7695	15.32%
	Transformer	0.2312	0.3544	0.6905	17.46%
	CNN-BiLSTM-Attention	0.2685	0.3729	0.6571	19.60%

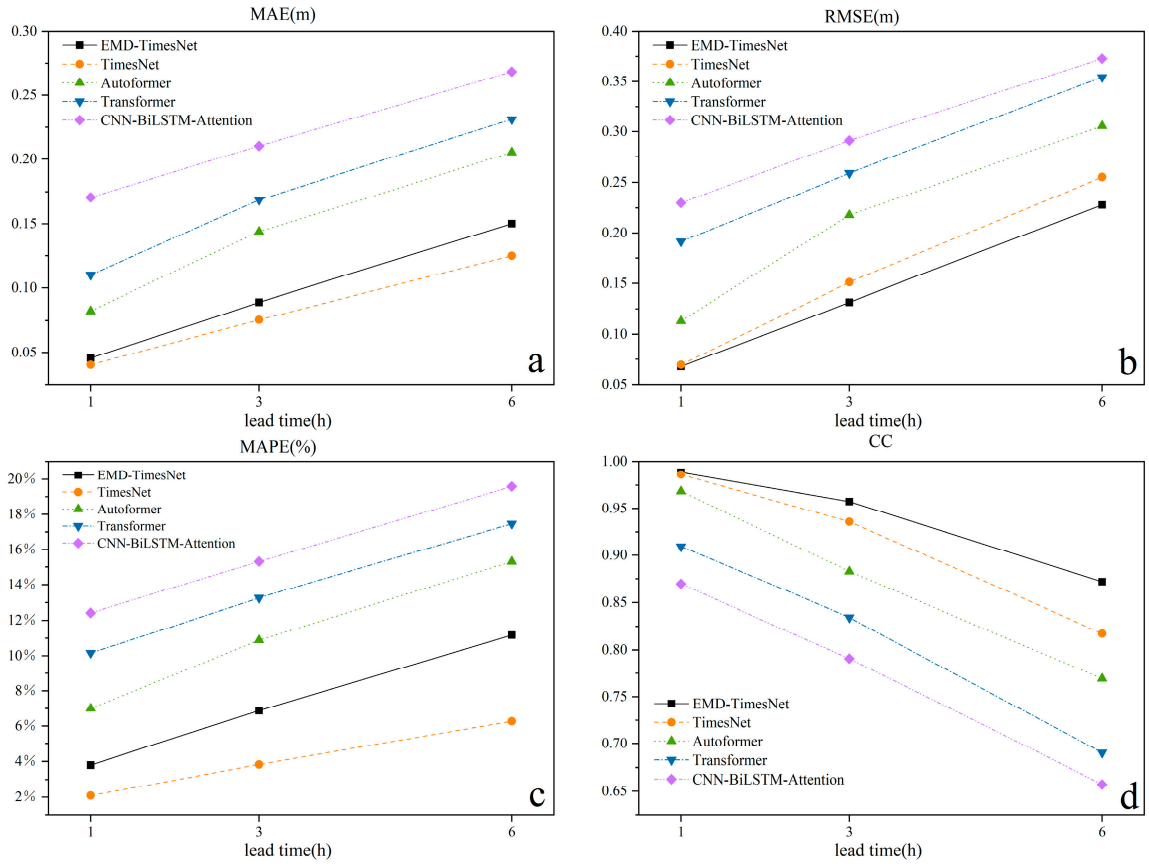


Figure 7. Comparative plots of the results obtained for EMD-TimesNet (black), TimesNet (orange), Autoformer (green), Transformer (blue), and CNN-BiLSTM-Attention (purple) models for predicting 1 h, 3 h, and 6 h results at 41025. (**a**) MAE; (**b**) RMSE; (**c**) MAPE; (**d**) CC.

Based on Figure 7, the ranked order of performance for each model is as follows: EMD-TimesNet ≈ TimesNet > Autoformer > Transformer > CNN-BiLSTM-Attention. When the prediction time is 1 h, the accuracy of EMD-TimesNet is superior to that of TimesNet in terms of RMSE and CC values, but based on the MAE and MAPE values, the accuracy of TimesNet is better than that of EMD-TimesNet; the difference in the accuracy of these two models is very small, so the prediction performance of EMD-TimesNet and TimesNet for SWH is roughly equivalent. They all perform better than the other three models. The performance of the Transformer algorithm is superior to that of CNN-BiLSTM-Attention. When the prediction time is 3 h, according to the values of RMSE and CC, the accuracy of EMD-TimesNet is also better than that of TimesNet, but based on the values of MAE and MAPE, the accuracy of TimesNet is better than that of EMD-TimesNet; however, the difference in the accuracy of these two models is very small, so the prediction performance of EMD-TimesNet and TimesNet for SWH is roughly equivalent. They perform better than the other three models. Notably, when the prediction time is set to 6 h, the accuracy of EMD-TimesNet surpasses that of TimesNet, as evident from the RMSE and CC values. However, when considering MAE and MAPE values, TimesNet exhibits superior accuracy compared to EMD-TimesNet. It is worth noting that the difference in accuracy between these two models is marginal. Consequently, the predictive performance of EMD-TimesNet and TimesNet for SWH is roughly equivalent. Moreover, both EMD-TimesNet and TimesNet outperform the remaining three models.

4.1.3. Overall Predictive Assessment of Buoy 41044

Table 6 displays the precise values of MAE, MAPE, RMSE, and CC for buoy 41044. The line graph is depicted in Figure 8.

Table 6. Comparison of prediction errors for 1 h, 3 h, and 6 h between EMD-TimesNet, TimesNet, Autoformer, Transformer, and CNN-BiLSTM-Attention at 41044. The bold text in the table indicates the best prediction.

Lead Time	Model	MAE (m)	RMSE (m)	CC	MAPE
1 h	EMD-TimesNet	**0.0394**	**0.0658**	**0.9897**	**2.05%**
	TimesNet	0.0406	0.0667	0.9886	2.08%
	Autoformer	0.0858	0.1372	0.9516	4.54%
	Transformer	0.1950	0.3140	0.7468	9.42%
	CNN-BiLSTM-Attention	0.2154	0.3249	0.7289	13.97%
3 h	EMD-TimesNet	**0.0772**	**0.1341**	**0.9538**	3.95%
	TimesNet	0.0785	0.1485	0.9434	**3.94%**
	Autoformer	0.1134	0.2132	0.8832	5.73%
	Transformer	0.1914	0.3277	0.7242	9.49%
	CNN-BiLSTM-Attention	0.2571	0.3943	0.6007	16.97%
6 h	EMD-TimesNet	**0.1195**	**0.2285**	**0.8659**	**5.96%**
	TimesNet	0.1311	0.2597	0.8268	6.49%
	Autoformer	0.1749	0.3157	0.7440	8.90%
	Transformer	0.3173	0.5075	0.3387	15.01%
	CNN-BiLSTM-Attention	0.3223	0.4880	0.3885	23.76%

Based on Figure 8, the ranked order of performance for each model is as follows: EMD-TimesNet > TimesNet > Autoformer > Transformer > CNN-BiLSTM-Attention. When the prediction time is 1 h, all the accuracy rating metrics of EMD-TimesNet and TimesNet are very close, the prediction performance of EMD-TimesNet and TimesNet for SWH is roughly comparable, and EMD-TimesNet performs a little bit better than TimesNet, and when compared with the other models, they all perform better. The following order applies for the other three models: Autoformer > Transformer > CNN-BiLSTM-Attention. When the prediction time is 3 h, EMD-TimesNet performs the best, yielding the best values among the five models for each evaluation metric. When the prediction time is 6 h, EMD-

TimesNet still performs the best, with all its evaluation metrics being the best among the five models. However, for buoy 41044, based on Figure 8b,d, upon scrutinizing the specific values at buoy 41044, it can be observed that when the prediction time was set to 6 h, the CC values for CNN-BiLSTM-Attention and Transformer diminished to 0.3387 and 0.3885, respectively. This decline in correlation signifies a substantial reduction, reaching a notably low level. Additionally, the RMSE values increased to 0.5075 m and 0.4880 m, indicating an increase in prediction errors to a relatively high level. In light of these observations, the prediction performance of both CNN-BiLSTM-Attention and the Transformer algorithm can be characterized as suboptimal during this specific prediction period.

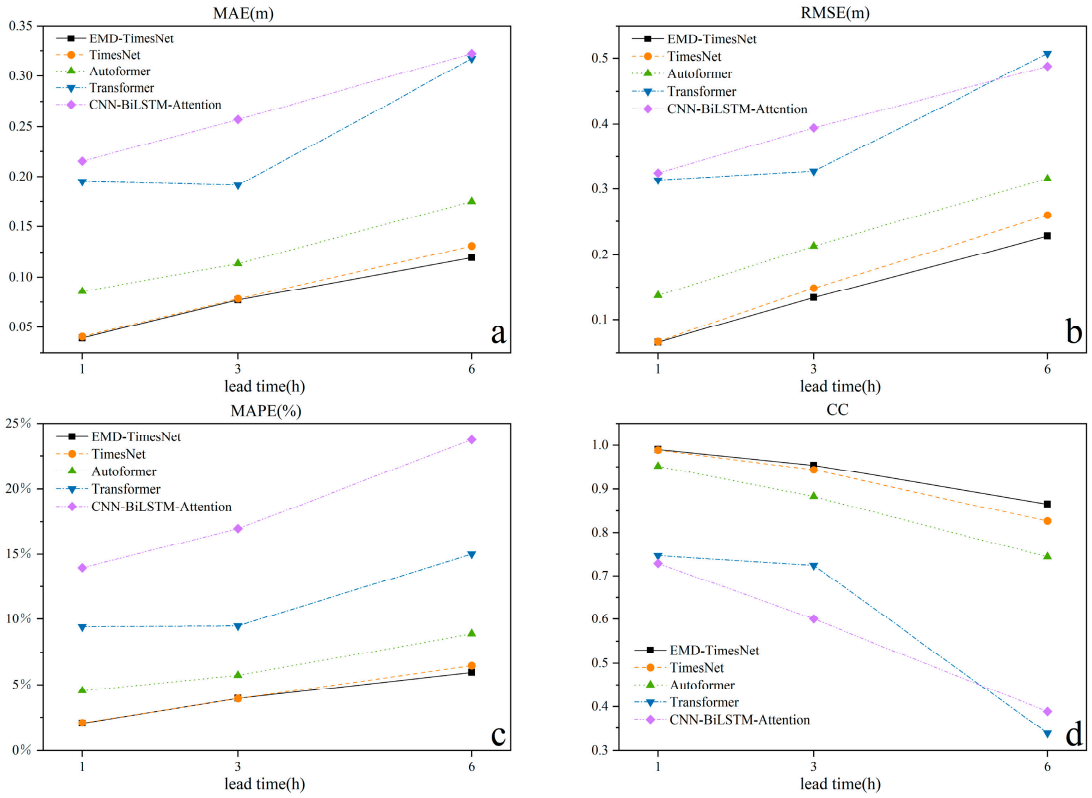


Figure 8. Comparative plots of the results obtained for EMD-TimesNet (black), TimesNet (orange), Autoformer (green), Transformer (blue), and CNN-BiLSTM-Attention (purple) models for predicting 1 h, 3 h, and 6 h results at 41044. (**a**) Represents MAE; (**b**) represents RMSE; (**c**) represents MAPE; (**d**) represents CC.

4.2. Predictive Performance of Models for Different Sea States

Waves are affected by a variety of factors, which are not only related to the input characteristics of the model but also bottom topography, water depth, etc. In this subsection, the effect of different monitoring stations on model accuracy is explored. Initial scrutiny involves a comprehensive analysis of the data derived from three buoys. The results are visually represented through box-and-line plots, as illustrated in Figure 9.

In the box plot representation in Figure 9, the two small horizontal lines at the top and bottom of each box signify the maximum and minimum values, respectively. The width of the central box reflects the degree of data dispersion, with the horizontal line inside indicating the median and the small square denoting the mean. Upon examining Figure 9a, it becomes apparent that the wind conditions across the stations are closely aligned, with 41025 exhibiting the most variable winds. Regarding wind direction, as illustrated in Figure 9b, while other stations show more frequent shifts in wave direction, station 41044 maintains a rather steady wave direction. While station 41010 experiences more significant changes in wave direction, stations 41025 and 41044 exhibit a generally

consistent wave direction, as shown in Figure 9c. Lastly, Figure 9d underscores that stations 41010 and 41025 exhibit lower and more stable wave heights, whereas station 41044 demonstrates higher and less stable wave heights.

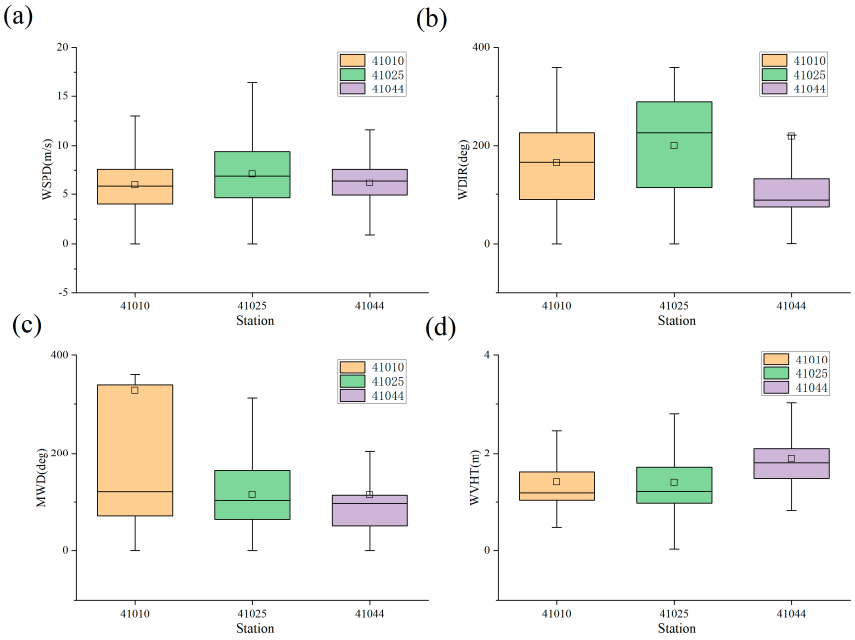


Figure 9. Data distribution for three buoys: (**a**) WSPD; (**b**) WDIR; (**c**) MWD; and (**d**) WVHT.

The data analysis indicates significant differences among these stations, primarily attributed to variations in wave height. Stations 41010 and 41025 display smoother wave conditions, whereas station 41044 exhibits steeper waves. Additionally, station 41010 experiences frequent changes in wave direction, contributing to the overall differentiation observed among these monitoring stations.

Next, each model's prediction performance for various sites at the same prediction time was examined. To gauge accuracy, CC and RMSE were chosen. Figure 10 provides a visual representation of the models' accuracy at a prediction time of 1 h.

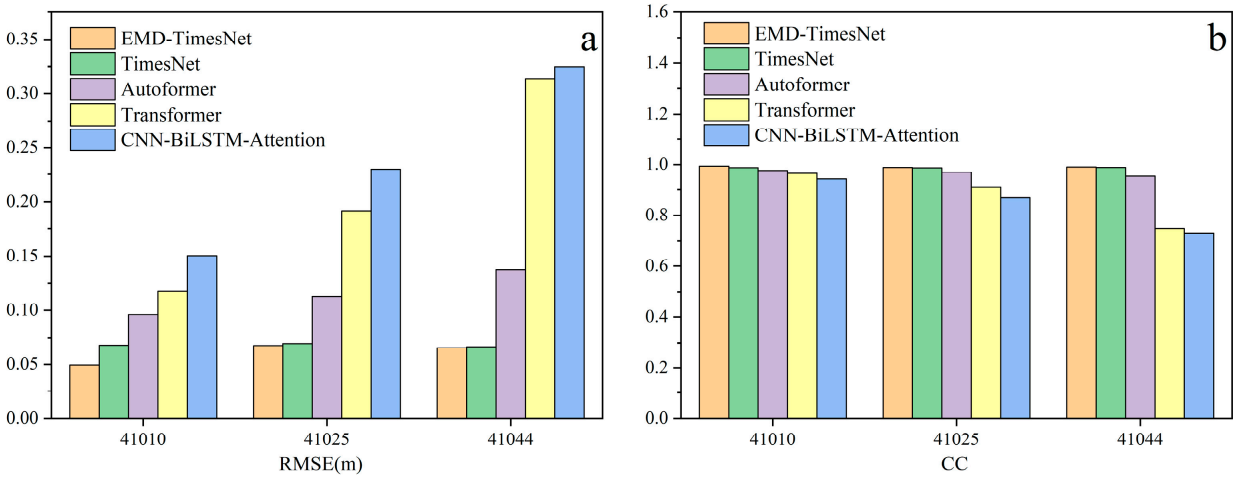


Figure 10. Predictive performance of each model at different stations for a prediction time of 1 h: (**a**) RMSE; (**b**) CC.

As depicted in Figure 10a, for different sites, the RMSE values of EMD-TimesNet basically remain in a relatively stable range when predicting 1 h SWH and are the lowest among the five models. TimesNet maintains a relatively stable range as well but with

higher RMSE values compared to EMD-TimesNet. In contrast, Autoformer, Transformer, and CNN-BiLSTM-Attention display notable disparities in prediction performance at different sites for the same prediction time, and they could not make accurate predictions for different sites.

As shown in Figure 10b, in the context of predicting SWH at 1 h intervals, the CC values for EMD-TimesNet consistently reside within a stable range, surpassing those of the other models. While TimesNet and Autoformer also maintain relatively stable ranges, their CC values are lower than the same value for EMD-TimesNet. Transformer and CNN-BiLSTM-Attention, on the other hand, exhibit significant variability in CC values when predicting SWH at different sites for the same time interval, indicating a limited capacity to adapt to site-specific changes.

Illustrated in Figure 11a, when predicting SWH at 3 h intervals across different sites, the RMSE for EMD-TimesNet increases compared to the 1 h prediction but remains within a relatively stable range, securing EMD-TimesNet's position as the model with the lowest error among the five. TimesNet also sustains a relatively stable range, albeit with higher RMSE values compared to those of EMD-TimesNet. Conversely, Autoformer, Transformer, and CNN-BiLSTM-Attention exhibit substantial variations in prediction performance at different sites for the same 3 h prediction time, precluding accurate predictions.

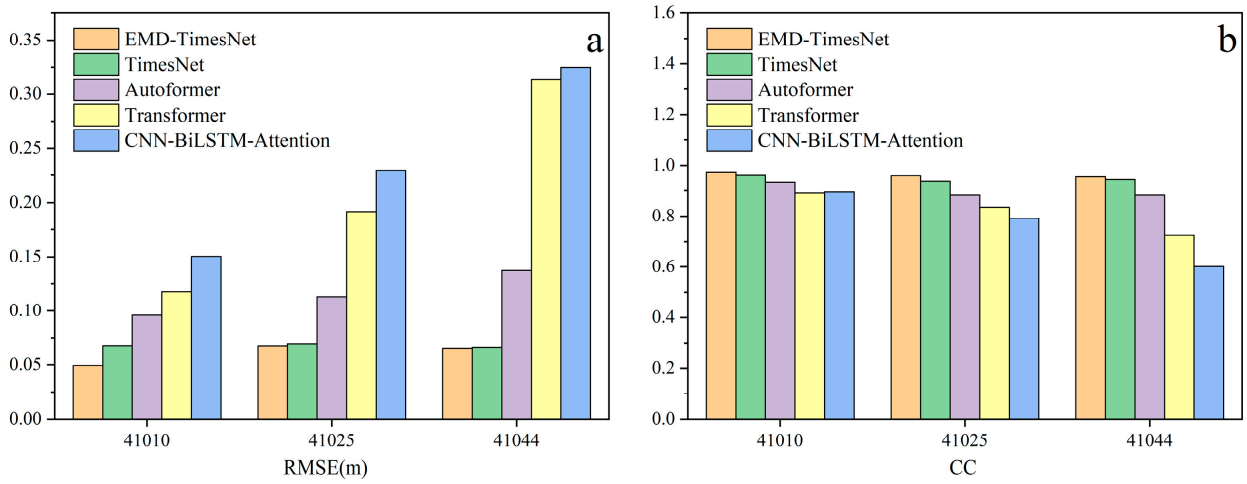


Figure 11. Predictive performance of each model at different stations for a prediction time of 3 h: (**a**) for RMSE; (**b**) for CC.

Based on Figure 11b, the CC values for EMD-TimesNet maintain a relatively stable range when predicting SWH at 3 h intervals, remaining the highest among the five models. While TimesNet and Autoformer also sustain relatively stable ranges, their CC values are lower than the same value for EMD-TimesNet. Transformer and CNN-BiLSTM-Attention display considerable discrepancies in CC values for the same 3 h prediction time across different sites, indicating a limited capacity to adapt to site-specific changes.

In Figure 12a, when predicting SWH at 6 h intervals across different sites, EMD-TimesNet continues to exhibit RMSE values within a relatively low range, maintaining the lowest error among the five models. TimesNet also maintains a relatively stable range yet with higher RMSE values compared to EMD-TimesNet. Conversely, Autoformer, Transformer, and CNN-BiLSTM-Attention display increased RMSE values compared to the 1 h and 3 h predictions. Notably, the RMSE values for Transformer and CNN-BiLSTM-Attention rose significantly, reaching a high level, indicative of a substantial decrease in prediction accuracy.

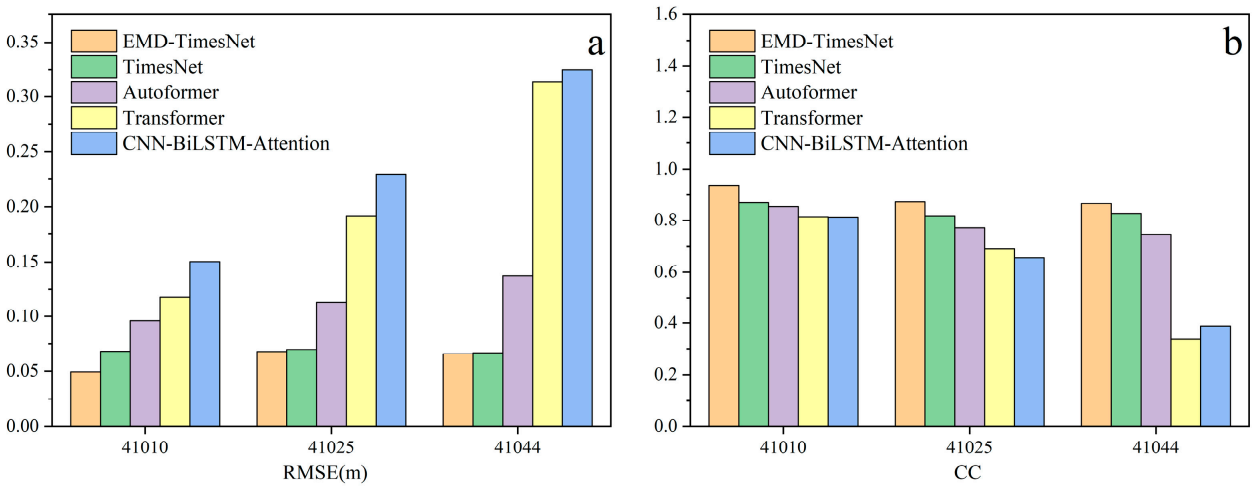


Figure 12. Predictive performance of each model at different stations for a prediction time of 6 h: (**a**) for RMSE; (**b**) for CC.

Based on Figure 12b, when forecasting SWH at 6 h intervals, EMD-TimesNet maintains a relatively high range of CC values, ranking as the highest among the five models. TimesNet and Autoformer also sustain a relatively stable range but with CC values lower than the CC of EMD-TimesNet. Conversely, Transformer and CNN-BiLSTM-Attention exhibit a pronounced drop in CC values, signaling poor model performance at this extended prediction time.

4.3. Discussion

Overall, EMD-TimesNet outperformed the other five models in terms of SWH prediction performance at all three buoys. This model's prediction performance decreased as prediction time increased, but even at 6 h, EMD-TimesNet continued to maintain a reasonably good level of prediction performance. At buoy 41010, EMD-TimesNet still maintained a strong correlation, with an RMSE value of 0.1573 m, an MAE value of 0.1026 m, and a CC value of 0.9352. At buoy 41025, EMD-TimesNet had an RMSE value of 0.2279 m, an MAE value of 0.1502 m, and a CC of 0.8719, also boasting good performance. At buoy 41044, EMD-TimesNet's RMSE value was 0.2285 m, its MAE value was 0.1195 m, and its CC reached 0.8659, showing that it still maintained high levels.

Upon comparing EMD-TimesNet and TimesNet, when the EMD module is added, the predictive performance of the models is similar at buoy 41025. At buoy 41010, the prediction performance of EMD-TimesNet compared to TimesNet was improved by 37.85% and 0.59% for RMSE and CC values, respectively, when the prediction time was 1 h; by 26.88% and 1.61% for RMSE and CC values, respectively, when the prediction time was 3 h; and by 41.96% and 1.61% for RMSE and CC values, respectively, when the prediction time was 6 h. were improved by 41.96%, 7.58%, respectively. At buoy 41044, regarding prediction performance, EMD-TimesNet also outperforms TimesNet across the board. Upon combining the results obtained for the three buoys, as expected, the overall performance of the models decreases with the increase in prediction time, but EMD-TimesNet still maintains high performance in long-term prediction. EMD-TimesNet maintains high accuracy and outperforms the other models. For the three buoys at different water depths and topographic conditions, EMD-TimesNet shows adaptability to various conditions; i.e., its performance is optimal at each buoy.

Based on these findings, it is of great practical importance to combine EMD with TimesNet. EMD contributes by offering additional information for the learning process, and, concurrently, the proven efficacy of TimesNet in time-series prediction tasks further underscores its utility. The experimental results affirm a consistent enhancement in overall assessment metrics and SWH prediction across various sea states for the EMD-TimesNet

model, outperforming alternative models. Therefore, combining these two mechanisms is a proven-successful and promising approach.

5. Conclusions

This paper introduces a novel approach utilizing EMD modal decomposition for the prediction of SWH in the Atlantic Ocean. Given the inherent instability of wave data time series, the EMD modal decomposition method demonstrates advantages in handling non-smooth and realistic signals. By using EMD, the proposed EMD-TimesNet neural network breaks down input data into discrete IMF signals. Data from January 2019 to December 2020 were used to train the model, and for the purpose of forecasting SWH at 1 h, 3 h, and 6 h intervals, the 2021 dataset was used as an independent experimental sample. The EMD-TimesNet model was juxtaposed with existing models, including TimesNet, Autoformer, Transformer, and CNN-BiLSTM-Attention. To assess prediction performance, four error assessment metrics were employed, providing a comprehensive evaluation of these five time-series prediction models.

The experimental findings reveal that EMD-TimesNet demonstrates accurate predictions for 1 h, 3 h, and 6 h SWH. Acknowledging the diverse factors influencing waves, including water depth and seabed topography, experiments simulating different sea states were conducted at onshore and offshore buoy sites, and comparative tests were performed. The TimesNet model uses convolution operations to extract temporal features from a one-dimensional time series and converts them into a two-dimensional representation. By using this method, the receptive field can be successfully expanded, and TimesNet is able to collect both intra- and inter-periodic information with accuracy. Because of this feature, it is especially suitable for SWH prediction, in which it improves predictive capabilities. The ability to capture time series features while maintaining signal non-stationary frequencies is greatly enhanced by the EMD decomposition method. In comparison with other exemplary forecasting models, the use of the EMD-TimesNet model for SWH forecasting surpasses TimesNet in terms of overall evaluation metrics and prediction under various sea states. Notably, it outperforms Autoformer, Transformer, and CNN-BiLSTM-Attention models as well.

While this paper makes a contribution, it is important to acknowledge certain limitations that require further refinement, prompting ongoing work. One notable limitation is the relatively large prediction error for SWH under extreme climate conditions. To improve prediction accuracy in extreme climate scenarios, future efforts will use other data models, like numerical predictions and satellite data. Further experiments are also needed to systematically explore the effects of different input combinations and different input/output ratios on model accuracy. These refinements are intended to bolster the robustness and effectiveness of our approach in addressing diverse environmental conditions.

Author Contributions: Conceptualization, Z.O. and X.W.; methodology, Z.O. and X.Z.; validation, Z.O. and Y.G.; writing—original draft preparation, Z.O., D.Z. and X.Z.; writing—review and editing, D.Z. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 42375143); the Key Research and Development Program, sponsored by the Ministry of Science and Technology (MOST) under grants 2023YFC3107701 and 2023YFC3107901; and the Key Laboratory of Smart Earth, NO. KF2023YB03-03. Funding was also provided by the Key Laboratory of Ocean Observation Technology, Ministry of Natural Resources, (klootB06).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Buoy data used in this study are available from the National Data Buoy Center at https://www.ndbc.noaa.gov/ (accessed on 21 March 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hsiao, S.-C.; Chen, H.; Wu, H.-L.; Chen, W.-B.; Chang, C.-H.; Guo, W.-D.; Chen, Y.-M.; Lin, L.-Y. Numerical Simulation of Large Wave Heights from Super Typhoon Nepartak (2016) in the Eastern Waters of Taiwan. *J. Mar. Sci. Eng.* **2020**, *8*, 217. [CrossRef]
2. Paerl, H.W.; Bales, J.D.; Ausley, L.W.; Buzzelli, C.P.; Crowder, L.B.; Eby, L.A.; Fear, J.M.; Go, M.; Peierls, B.L.; Richardson, T.L.; et al. Ecosystem Impacts of Three Sequential Hurricanes (Dennis, Floyd, and Irene) on the United States' Largest Lagoonal Estuary, Pamlico Sound, NC. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5655–5660. [CrossRef] [PubMed]
3. Wang, H.; Fu, D.; Liu, D.; Xiao, X.; He, X.; Liu, B. Analysis and Prediction of Significant Wave Height in the Beibu Gulf, South China Sea. *JGR Oceans* **2021**, *126*, e2020JC017144. [CrossRef]
4. Fan, S.; Xiao, N.; Dong, S. A Novel Model to Predict Significant Wave Height Based on Long Short-Term Memory Network. *Ocean. Eng.* **2020**, *205*, 107298. [CrossRef]
5. Trigo, R.M.; Valente, M.A.; Trigo, I.F.; Miranda, P.M.A.; Ramos, A.M.; Paredes, D.; García-Herrera, R. The Impact of North Atlantic Wind and Cyclone Trends on European Precipitation and Significant Wave Height in the Atlantic. *Ann. N. Y. Acad. Sci.* **2008**, *1146*, 212–234. [CrossRef]
6. Aarnes, O.J.; Reistad, M.; Breivik, Ø.; Bitner-Gregersen, E.; Ingolf Eide, L.; Gramstad, O.; Magnusson, A.K.; Natvig, B.; Vanem, E. Projected Changes in Significant Wave Height toward the End of the 21st Century: Northeast A Tlantic. *JGR Oceans* **2017**, *122*, 3394–3403. [CrossRef]
7. Group, T.W. The WAM Model—A Third Generation Ocean Wave Prediction Model. *J. Phys. Oceanogr.* **1988**, *18*, 1775–1810. [CrossRef]
8. Booij, N.; Ris, R.C.; Holthuijsen, L.H. A Third-generation Wave Model for Coastal Regions: 1. Model Description and Validation. *J. Geophys. Res.* **1999**, *104*, 7649–7666. [CrossRef]
9. Tolman, H.L. Distributed-Memory Concepts in the Wave Model WAVEWATCH III. *Parallel Comput.* **2002**, *28*, 35–52. [CrossRef]
10. Wang, W.; Tang, R.; Li, C.; Liu, P.; Luo, L. A BP Neural Network Model Optimized by Mind Evolutionary Algorithm for Predicting the Ocean Wave Heights. *Ocean. Eng.* **2018**, *162*, 98–107. [CrossRef]
11. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
12. Feng, Z.; Hu, P.; Li, S.; Mo, D. Prediction of Significant Wave Height in Offshore China Based on the Machine Learning Method. *J. Mar. Sci. Eng.* **2022**, *10*, 836. [CrossRef]
13. Wang, J.; Yu, T.; Deng, F.; Ruan, Z.; Jia, Y. Acquisition of the Wide Swath Significant Wave Height from HY-2C through Deep Learning. *Remote Sens.* **2021**, *13*, 4425. [CrossRef]
14. Wang, J.K.; Aouf, L.; Dalphinet, A.; Zhang, Y.G.; Xu, Y.; Hauser, D.; Liu, J.Q. The Wide Swath Significant Wave Height: An Innovative Reconstruction of Significant Wave Heights from CFOSAT's SWIM and Scatterometer Using Deep Learning. *Geophys. Res. Lett.* **2021**, *48*, e2020GL091276. [CrossRef]
15. Mahjoobi, J.; Etemad-Shahidi, A. An Alternative Approach for the Prediction of Significant Wave Heights Based on Classification and Regression Trees. *Appl. Ocean. Res.* **2008**, *30*, 172–177. [CrossRef]
16. Etemad-Shahidi, A.; Mahjoobi, J. Comparison between M5′ Model Tree and Neural Networks for Prediction of Significant Wave Height in Lake Superior. *Ocean. Eng.* **2009**, *36*, 1175–1181. [CrossRef]
17. Mahjoobi, J.; Adeli Mosabbeb, E. Prediction of Significant Wave Height Using Regressive Support Vector Machines. *Ocean. Eng.* **2009**, *36*, 339–347. [CrossRef]
18. Cornejo-Bueno, L.; Nieto Borge, J.C.; Alexandre, E.; Hessner, K.; Salcedo-Sanz, S. Accurate Estimation of Significant Wave Height with Support Vector Regression Algorithms and Marine Radar Images. *Coast. Eng.* **2016**, *114*, 233–243. [CrossRef]
19. Berbić, J.; Ocvirk, E.; Carević, D.; Lončar, G. Application of Neural Networks and Support Vector Machine for Significant Wave Height Prediction. *Oceanologia* **2017**, *59*, 331–349. [CrossRef]
20. Lim, B.; Zohren, S. Time-Series Forecasting with Deep Learning: A Survey. *Phil. Trans. R. Soc. A* **2021**, *379*, 20200209. [CrossRef] [PubMed]
21. Huang, X.; Li, Q.; Tai, Y.; Chen, Z.; Zhang, J.; Shi, J.; Gao, B.; Liu, W. Hybrid Deep Neural Model for Hourly Solar Irradiance Forecasting. *Renew. Energy* **2021**, *171*, 1041–1060. [CrossRef]
22. Minuzzi, F.C.; Farina, L. A Deep Learning Approach to Predict Significant Wave Height Using Long Short-Term Memory. *Ocean Model.* **2023**, *181*, 102151. [CrossRef]
23. Zhang, W.; Sun, Y.; Wu, Y.; Dong, J.; Song, X.; Gao, Z.; Pang, R.; Guoan, B. A Deep-Learning Real-Time Bias Correction Method for Significant Wave Height Forecasts in the Western North Pacific. *Ocean Model.* **2024**, *187*, 102289. [CrossRef]
24. Ikram, R.M.A.; Cao, X.; Sadeghifar, T.; Kuriqi, A.; Kisi, O.; Shahid, S. Improving Significant Wave Height Prediction Using a Neuro-Fuzzy Approach and Marine Predators Algorithm. *J. Mar. Sci. Eng.* **2023**, *11*, 1163. [CrossRef]
25. Hao, P.; Li, S.; Yu, C.; Wu, G. A Prediction Model of Significant Wave Height in the South China Sea Based on Attention Mechanism. *Front. Mar. Sci.* **2022**, *9*, 895212. [CrossRef]
26. Luo, Q.-R.; Xu, H.; Bai, L.-H. Prediction of Significant Wave Height in Hurricane Area of the Atlantic Ocean Using the Bi-LSTM with Attention Model. *Ocean Eng.* **2022**, *266*, 112747. [CrossRef]
27. Pang, J.; Dong, S. A Novel Multivariable Hybrid Model to Improve Short and Long-Term Significant Wave Height Prediction. *Appl. Energy* **2023**, *351*, 121813. [CrossRef]
28. Mahdavi-Meymand, A.; Sulisz, W. Application of Nested Artificial Neural Network for the Prediction of Significant Wave Height. *Renew. Energy* **2023**, *209*, 157–168. [CrossRef]

29. Shi, J.; Su, T.; Li, X.; Wang, F.; Cui, J.; Liu, Z.; Wang, J. A Machine-Learning Approach Based on Attention Mechanism for Significant Wave Height Forecasting. *J. Mar. Sci. Eng.* **2023**, *11*, 1821. [CrossRef]
30. Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are Transformers Effective for Time Series Forecasting? In Proceedings of the 37th AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 11121–11128. [CrossRef]
31. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**. [CrossRef]
32. An, J.; Yang, H.; Gui, X.; Zhang, W.; Gui, R.; Kang, J. TCNS: Node Selection with Privacy Protection in Crowdsensing Based on Twice Consensuses of Blockchain. *IEEE Trans. Netw. Serv. Manag.* **2019**, *16*, 1255–1267. [CrossRef]
33. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In Proceedings of the 35th AAAI Conference on Artificial Intelligence, Palo Alto, CA, USA, 2–9 February 2021; Volume 35, pp. 11106–11115. [CrossRef]
34. Xu, J.; Wu, H.; Wang, J.; Long, M. Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy. *arXiv* **2021**, arXiv:2110.02642. [CrossRef]
35. Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; Long, M. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In Proceedings of the Eleventh International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
36. Huang, N.E. Introduction to the Hilbert–Huang Transform and Its Related Mathematical Problems. In *Interdisciplinary Mathematical Sciences*; World Scientific: Singapore, 2014; Volume 16, pp. 1–26, ISBN 978-981-4508-23-0.
37. Almeida, L.; Pedreiras, P.; Fonseca, J.A.G. The FTT-CAN Protocol: Why and How. *IEEE Trans. Ind. Electron.* **2002**, *49*, 1189–1201. [CrossRef]
38. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv* **2015**, arXiv:1512.00567. [CrossRef]
39. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *arXiv* **2022**, arXiv:2106.13008. [CrossRef]
40. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.