remote sensing

MDPI

Article

NAMD-Net: A Dual-Branch Model for Significant Wave Height Prediction Based on Noise-Assisted Multivariate Empirical Mode Decomposition

Han Liu [1], Biao Meng [1], Xinqiong Liu [1,*], Qinghe Zhang [1] and Weimin Huang [2]

[1] Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China; liuh@ctgu.edu.cn (H.L.); mengbiao@ctgu.edu.cn (B.M.); zqh@ctgu.edu.cn (Q.Z.)
[2] Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada; weimin@mun.ca
* Correspondence: liuxinqiong@ctgu.edu.cn

Highlights

What are the main findings?

- NA-MEMD jointly decomposes SWH, WSPD, GST, and APD into scale-aligned intrinsic mode functions, alleviating mode mixing and avoiding the component-matching problem caused by separate decomposition of multiple input variables.
- Permutation entropy-guided grouping assigns fast-varying and slow-varying components to WaveNet and LSTM, respectively, achieving improved 1–24 h SWH prediction performance compared with benchmark models.

What are the implications of the main findings?

- Multivariate adaptive decomposition and dual-path deep learning can effectively model nonlinear and non-stationary SWH time series.
- NAMD-Net exhibits robustness under complex wave conditions, as supported by error accumulation, sea-state dependence, and seasonal variation analyses.

Abstract

Accurate prediction of significant wave height (SWH) is essential for ocean engineering safety, maritime transportation, and wave energy utilization. To address the nonlinear and non-stationary characteristics of SWH time series, this paper proposes a hybrid forecasting model that integrates signal decomposition, mode grouping and reconstruction, and deep learning. First, the model employs noise-assisted multivariate empirical mode decomposition (NA-MEMD) to adaptively decompose the multivariate time series into intrinsic mode functions (IMFs) to alleviate the problem of mode mixing. Subsequently, permutation entropy is employed to assess the complexity of the mode components and guide their grouping into two branches. Finally, WaveNet and long short-term memory (LSTM) networks are used to model and forecast the relatively high-complexity (fast-varying) and low-complexity (slow-varying) components, respectively, and the final forecast is generated by superimposing the predictions of all components. The proposed model was validated using measured data from two buoy stations, and evaluated by four metrics: root mean square error (RMSE), mean absolute error (MAE), symmetric mean absolute percentage error (sMAPE), and correlation coefficient (R). Experimental results show that the proposed model achieves lower RMSE, MAE, and sMAPE, and higher R compared with the baseline models across all evaluated forecast horizons and buoy sites.

Check for updates

Academic Editor: Sergei Badulin

Received: 11 August 2026
Revised: 15 October 2026
Accepted: 22 October 2026
Published: 27 October 2026

Copyright: © 2026 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.

Keywords: significant wave height; multistep prediction; deep learning; hybrid model

1. Introduction

Significant wave height (SWH), as the most fundamental parameter for describing ocean wave fields [1], serves as a key indicator for assessing the safety of marine engineering projects, optimizing maritime navigation, evaluating the risks of marine disasters, and quantifying the potential of wave energy resources [2–5]. Accurate prediction of SWH is not only an urgent requirement for enhancing marine environmental safety but also holds significant theoretical value for advancing marine science [6].

Currently, SWH prediction methods are primarily divided into two major categories: numerical methods and data-driven methods [7]. The former have a solid theoretical foundation with fluid mechanics and energy balance equations as their core, but the high computational cost and complex modeling process limit their flexibility in practical applications [8]. To overcome this limitation, data-driven methods have emerged which rely only on historical observations and make predictions by mining the intrinsic laws of the data. Within the framework of data-driven methods, these can be further subdivided into traditional statistical models and machine learning models. These traditional statistical models (such as autoregressive moving average (ARMA)) perform well when handling SWH time series that satisfy the linearity and stationarity assumptions [9]. However, constrained by the linearity assumptions in the models themselves and the strict requirements for data stationarity, such methods struggle to effectively capture the nonlinear and non-stationary characteristics in ocean wave processes [10–12]. In recent years, machine learning models such as support vector machine (SVM) [13], random forest (RF) [14], artificial neural network (ANN) [15], and long short-term memory (LSTM) have been widely applied in SWH forecasting due to their nonlinear fitting capabilities, robust adaptive learning abilities, and reasonable generalization performance [16].

Fan et al. [16] employed an LSTM network for SWH prediction and compared it with other models such as residual network (ResNet), back propagation neural network (BPNN), and extreme learning machine (ELM). Similarly, Lou et al. [17] also utilized an LSTM network for SWH prediction. In the context of SWH forecasting, Hao et al. [18] conducted a comprehensive comparison of three network models: recurrent neural network (RNN), gated recurrent unit (GRU), and LSTM. While these standalone machine learning models outperform traditional methods in handling the nonlinear characteristics of SWH time series and achieve relatively high forecasting accuracy, they face challenges in modeling non-stationary data [19,20].

Recently, a novel method that combines signal processing techniques with machine learning has been proposed for SWH forecasting to mitigate the impact of non-stationary features. For example, Zhou et al. [21] employed empirical mode decomposition (EMD) to decompose SWH time series into several intrinsic mode functions (IMFs), followed by forecasting using LSTM network. Feng et al. [22] introduced wind speed as an auxiliary variable and combined EMD with LSTM for multivariate prediction. Lv et al. [23] also introduced wind speed as an auxiliary variable, decomposed the time series using EMD, and then employed WaveNet network for prediction. Song et al. [24] employed ensemble empirical mode decomposition (EEMD) along with LSTM network to mitigate the impact of mode mixing in EMD. Huang and Dong [25] utilized improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) combined with LSTM network to further address the residual noise issue in the decomposed IMFs. Wang and Ying [26] introduced variational mode decomposition (VMD) to decompose SWH time

series, reconstructed the decomposed sub-sequences using refined composite multiscale fuzzy entropy (RCMFE), and then performed predictions using GRU network. Pang and Dong [12] incorporated wind speed and wind direction as auxiliary variables, combining ICEEMDAN with recurrence quantification analysis (RQA), and similarly employed LSTM for the final prediction.

Although the aforementioned hybrid methods have improved the prediction accuracy of SWH, several critical issues remain to be addressed:

1. Most studies decompose only a single SWH time series, neglecting the impact of multiple ocean variables highly correlated with SWH on its prediction and losing the dependencies among these variables [27].
2. In scenarios with multiple input variables, if each variable is decomposed separately, the number of IMFs generated for each variable often varies, making it impossible to directly map oscillations at the same scale. Traditional multivariate hybrid methods rely on methods such as stochastic components decomposition (DSD) [12], dynamic time warping (DTW) [28], and K-means to resolve this issue through manual theoretical clustering [29]. This not only significantly increases computational complexity but may also introduce subjective bias, reducing cross-variable scale consistency.
3. Most traditional hybrid methods employ a single prediction model (e.g., LSTM) to uniformly process all IMFs, failing to account for the dynamically different characteristics among IMF components. Relatively fast-varying components contain more complex local fluctuations, whereas slow-varying components preserve longer-term temporal dependencies. Since fast-varying components require capturing local nonlinear patterns, while slow-varying components are better suited for fitting using long-range dependencies, a single model struggles to optimally adapt to both simultaneously [28].

To overcome the above limitations, this paper proposes a noise-assisted multivariate decomposition network (NAMD-Net) for SWH buoy forecasting. Its core idea is as follows. First, noise-assisted multivariate empirical mode decomposition (NA-MEMD) is employed to jointly decompose the multivariate time series [30]. By introducing common auxiliary noise channels, NA-MEMD performs multivariate sifting within a shared projection space, which facilitates the generation of scale-aligned IMFs among different variables. This characteristic reduces the need for additional component matching or regrouping procedures required after multivariate decomposition in traditional hybrid methods, preserving the temporal-scale consistency among multivariate components. NA-MEMD has already been applied in the fields of photovoltaic power forecasting and carbon price forecasting [31,32]. Second, permutation entropy is introduced to characterize the complexity of the decomposed components and guide their grouping. Permutation entropy offers high computational efficiency and noise robustness, and it effectively captures the local dynamic characteristics of and complexity changes in nonlinear time series. It has been widely used in fault detection [33,34]. Finally, a dual-path parallel forecasting strategy is adopted. Complexity-guided fast-varying IMFs that contain more local fluctuations and abrupt changes are fed into WaveNet, a dilated convolutional network that excels at capturing local details and transient patterns [35]. Complexity-guided slow-varying IMFs that exhibit long-term dependencies are fed into LSTM network, which performs exceptionally well in capturing temporal dependencies in time series data [36]. The prediction results from the two paths are then summed to obtain the final SWH forecast [21].

2. Data Description

This study utilized standard meteorological data from the national data buoy center (NDBC) [37]. To minimize the random influence of individual station samples on the

prediction results, two buoy stations in the Atlantic, namely buoy 41,013 and buoy 41,046, were selected for SWH prediction and their geographical locations are shown in Figure 1.

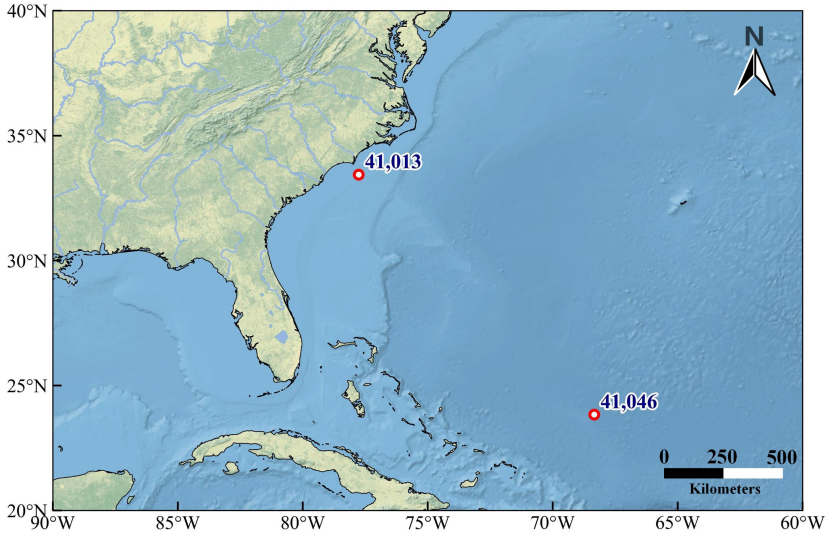


Figure 1. Specific locations of two buoy stations in the Atlantic.

The buoy records contain ten meteorological and wave-related variables. The first four are related to wind and waves: SWH, wind direction (WDIR), wind speed (WSPD), and wind gust (GST). The next three describe wave period and direction: dominant wave period (DPD), average wave period (APD), and dominant wave direction (MWD). The remaining three are atmospheric and water parameters: sea level pressure (PRES), air temperature (ATMP), and water temperature (WTMP). For missing observations, linear interpolation was applied to isolated one-hour gaps. To satisfy the continuity requirement of NA-MEMD, longer gaps were also interpolated strictly for the purpose of decomposition. However, during the subsequent sliding-window construction, any window containing temporal gaps (including those over the interpolated long gaps) was discarded, so that artificially interpolated values did not enter model training or testing. For feature screening, the ten variables from the training period (2017–2019) were individually quality-controlled and temporally aligned, and the Pearson correlation analysis was performed on this training period to examine their relationships with SWH [38]. The calculation process for the pearson correlation coefficient of data sequences X and Y is expressed as

$$\rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (1)$$

where σ_X and σ_Y represent the standard deviations of X and Y, respectively; μ_X and μ_Y, their mean values; and E, the expectation.

The specific analysis results are shown in Figure 2. According to the principles of pearson correlation analysis, a larger absolute value of the pearson correlation coefficient indicates a stronger linear relationship between variables, and typically implies that the feature has more direct predictive value for the target variable. Here, historical SWH, WSPD, GST, and APD were selected as model inputs since their pearson correlation coefficients with SWH are above 0.4 at both stations. Weakly correlated variables were excluded.

Table 1 presents the basic data information for the two buoy stations and the descriptive statistics of the main input variables, including the data study period, water depth, and the minimum, maximum, mean, and variance of SWH, WSPD, GST, and APD. Figure 3 further illustrates the temporal variation characteristics of SWH at the two buoy stations during the study period. The SWH time series at both stations exhibit distinct fluctuations and

intermittent spikes, with fluctuations predominantly at low values accompanied by a few high-value peaks.

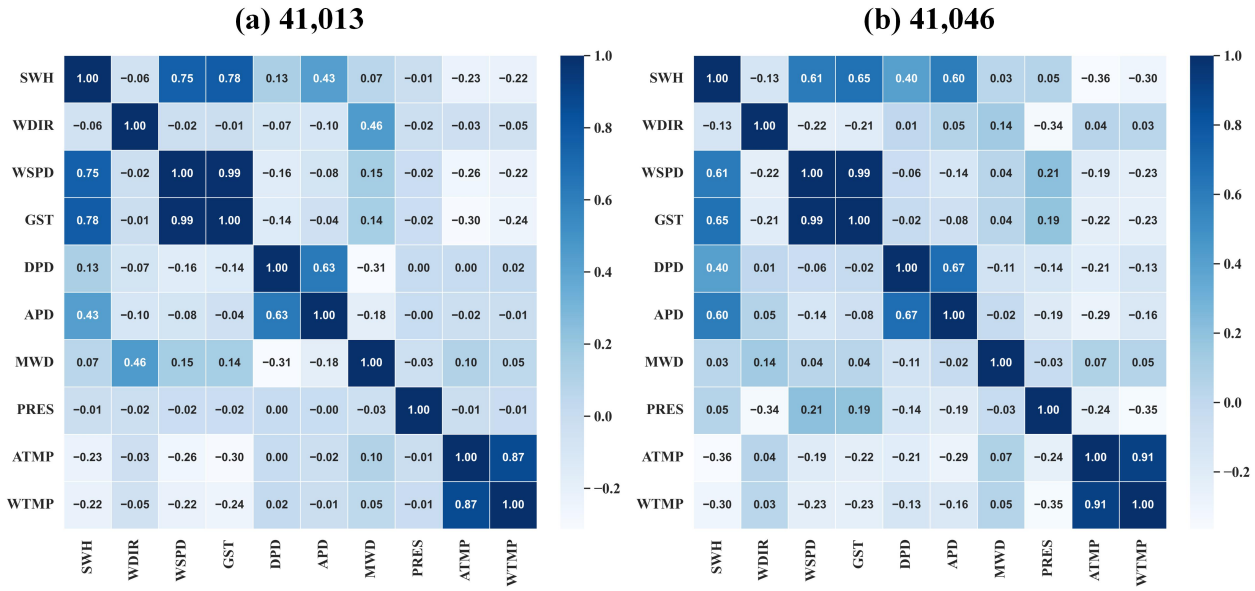


Figure 2. Pearson correlation coefficients between SWH and other variables. (**a**) At station 41,013. (**b**) At station 41,046.

Table 1. Specifications of the two buoy stations.

Station	Period	Depth	Variable	Min	Max	Mean	Variance
41,013	2017–2021	33 m	SWH (m)	0.28	6.98	1.34	0.43
			WSPD (m/s)	0.33	26.40	6.59	10.21
			GST (m/s)	0.10	39.20	8.42	15.63
			APD (s)	3.02	11.43	5.11	0.69
41,046	2017–2021	5553 m	SWH (m)	0.16	9.37	1.68	0.46
			WSPD (m/s)	0.10	24.80	5.82	5.49
			GST (m/s)	0.21	35.50	7.49	7.77
			APD (s)	3.68	13.41	6.15	1.06

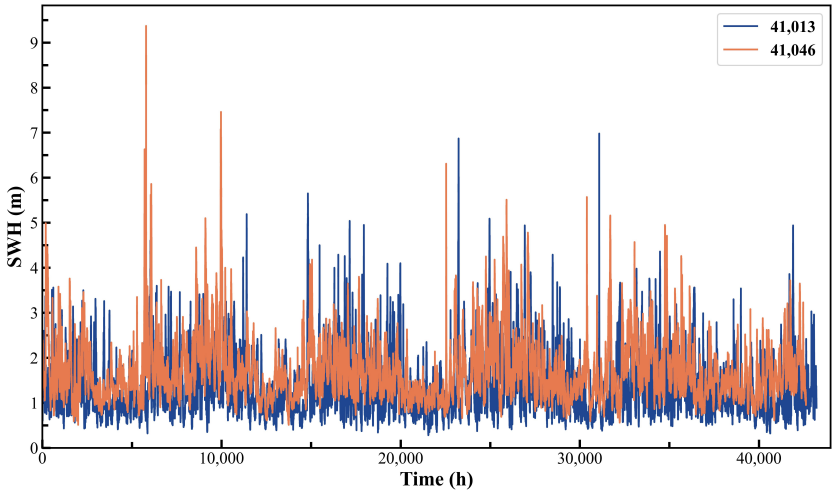


Figure 3. Time series of SWH at two buoy stations.

3. Methodology

This section respectively introduces the NA-MEMD algorithm, the permutation entropy algorithm, and the WaveNet and LSTM models. Finally, it presents the overall framework of NAMD-Net.

3.1. NA-MEMD Algorithm

EMD and its multivariate extension MEMD are widely used for adaptive signal decomposition, but both suffer from mode mixing [39,40]. To overcome this, Rehman et al. [30] proposed the NA-MEMD algorithm. By adding auxiliary Gaussian white noise channels to the multivariate time series and leveraging the inherent binary filter bank property of MEMD, NA-MEMD can alleviate mode mixing by improving scale separation among multivariate oscillatory components. After decomposition, the noise-channel IMFs are discarded, and the signal-channel IMFs are retained, yielding improved scale separation among modes. This method shows adaptability and robustness, providing stable decomposition performance under noisy conditions.

Assume that the original multivariate time series is denoted as $x(t) \in \mathbb{R}^{n\times T}$, where t is the time index, n is the number of original input variables, and T is the length of each time series. An auxiliary Gaussian white noise sequence is generated as $s(t) \in \mathbb{R}^{q\times T}$, where q denotes the number of auxiliary noise channels. The original signal and the auxiliary noise sequence are combined to form an augmented multivariate signal $z(t) \in \mathbb{R}^{(n+q)\times T}$. Here, $z(t)$ represents the composite signal used for MEMD decomposition, and $n + q$ is the total number of channels after adding the auxiliary noise.

For the augmented multivariate time series $z(t)$, the Hammersley sequence sampling method is used to generate a set of direction vectors in the multivariate space. The projection of $z(t)$ onto the k-th direction vector X_k is denoted as $\mathrm{Proj}_k(t)$, where $k = 1, \ldots, l$, and l represents the total number of direction vectors. The time locations t_i^k corresponding to the extrema of $\mathrm{Proj}_k(t)$ are then identified, where i denotes the extrema index.

Based on the extrema points $[t_i^k, z(t_i^k)]$, multivariate spline interpolation is performed to construct the envelope curve $E^k(t)$ corresponding to the k-th direction vector. To reduce boundary effects during envelope construction, two extrema were symmetrically extended at each end of the signal before cubic spline interpolation. After the envelope curves are obtained for all projection directions, their local mean function $L_m(t)$ is calculated as

$$L_m(t) = \frac{1}{l}\sum_{k=1}^{l} E^k(t), \tag{2}$$

A candidate IMF can be extracted after the $L_m(t)$ is removed from the current signal. A threshold-based stopping criterion with a threshold vector of $(0.075, 0.75, 0.075)$ was adopted; if the candidate satisfies the stopping criterion of a multivariate IMF, it is extracted as one mode; otherwise, the sifting operation is repeated, with a maximum of 1000 iterations imposed to avoid non-convergence. After each IMF is obtained, the corresponding residual is used for the next decomposition round until no further meaningful oscillatory component remains.

After NA-MEMD decomposition, the auxiliary noise-channel IMFs are discarded, and only the IMFs of the original variables are retained. In this way, different time series can produce an equal number of IMFs, and IMFs at comparable temporal scales remain aligned across different time series. This characteristic facilitates that subsequent machine learning models can learn consistently on IMFs with similar frequency features, thereby helping improve NAMD-Net's ability to predict SWH and its accuracy when processing complex marine environmental data.

3.2. Permutation Entropy Algorithm

Permutation entropy is used to characterize the irregularity and dynamic variations in time series [33]. It quantitatively assesses the complexity contained within a signal sequence and offers robust noise resistance and high computational efficiency [34]. The main calculation process is summarized below.

Set $x = \{x_1, x_2, ..., x_J\}$ be a time series, define the embedding dimension a, the delay time U_d, and $K = J - (a-1)U_d$ to construct the signal space matrix Z_p as

$$Z_p = \begin{pmatrix} x_1 & x_{1+U_d} & \cdots & x_{1+(a-1)U_d} \\ x_2 & x_{2+U_d} & \cdots & x_{2+(a-1)U_d} \\ x_c & x_{c+U_d} & \cdots & x_{c+(a-1)U_d} \\ \vdots & \vdots & \ddots & \vdots \\ x_K & x_{K+U_d} & \cdots & x_{K+(a-1)U_d} \end{pmatrix}$$

Each row of the signal space matrix Z_p corresponds to a reconstructed vector, forming the set of time series $v = \{v_1, v_2, ..., v_K\}$. To reflect the relative ordering of elements, each component must be sorted according to its temporal relationship. Here, $(c_1, c_2, \ldots, c_a)$ denotes the index sequence after the elements are ranked in ascending order. The corresponding ordinal relation is expressed as

$$x_i + (c_1 - 1)U_d \leq x_i + (c_2 - 1)U_d \leq ... \leq x_i + (c_a - 1)U_d$$

Let P_c denote the occurrence probability of the c-th ordinal pattern, where at most $a!$ different ordinal patterns exist. Next, the probability corresponding to the set of time series v is $P_1, P_2, ..., P_{a!}$. The permutation entropy of the original time series is then defined as

$$H_{pe} = -\sum_{c=1}^{a!} P_c \ln P_c \tag{3}$$

For convenience, permutation entropy typically requires normalization. The normalized permutation entropy $H(a)$ is

$$H(a) = \frac{H_{pe}}{\ln(a!)} \tag{4}$$

The embedding dimension a and delay time U_d are set to 3 and 1, respectively. For equal values within a reconstructed vector, their original temporal order is retained when determining the ordinal pattern [34].

Furthermore, permutation entropy is used as a complexity-based criterion for grouping the decomposed components. First, the time series of the four variables are decomposed into multiple IMF components using NA-MEMD decomposition. For each IMF component, the corresponding permutation entropy $H(a)$ is calculated. Subsequently, the $H(a)$ values of the IMF for all variable channels are averaged. This averaging operation reduces the influence of random fluctuations in individual IMF entropy values and provides a more stable complexity measure. Finally, by averaging the mean $H(a)$ of all variable channels, the result is used as a benchmark for determining complexity-guided fast-varying and slow-varying components: components with $H(a)$ above and below the average are assigned to the two prediction groups, respectively. Their spectral characteristics are further examined in Section 4.2.

3.3. WaveNet

WaveNet is a convolutional sequence modeling architecture built on stacked dilated causal convolution layers [35]. The causal convolution mechanism restricts that the output at each time step depends only on current and historical inputs, thereby preserving temporal

causality in sequence prediction. By introducing dilation factors between adjacent kernel elements, dilated convolution enlarges the receptive field without increasing the kernel size or substantially increasing the number of trainable parameters [41].

As shown in Figure 4, the dilation factor increases progressively from 1 to 8 across the stacked convolutional layers. This hierarchical receptive-field structure enables WaveNet to capture multi-scale temporal dependencies and local dynamic variations in time series, making it suitable for modeling short-term oscillations and abrupt changes in SWH prediction [23]. The corresponding mathematical formulation is provided in Van Den Oord et al. [35].

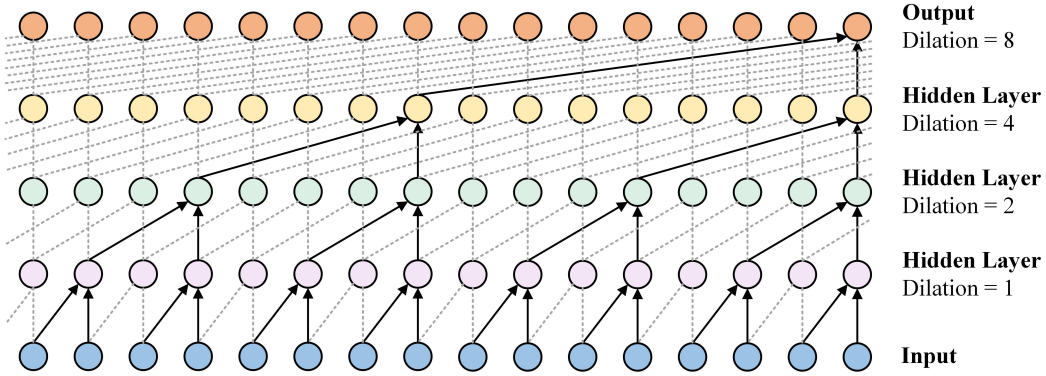


Figure 4. Schematic illustration of dilated causal convolutional layers.

3.4. LSTM

RNN can adjust its current hidden state using information from previous states and present input, making it applicable across various fields [42]. However, when dealing with large time intervals between input and target output, RNN encounters gradient vanishing or exploding. To mitigate this problem, Hochreiter et al. [36] developed LSTM by introducing a memory cell capable of retaining relevant information while discarding redundant information. As shown in Figure 5, the LSTM cell uses three gate units to control the flow of information across time steps. The LSTM gate activations and state updates are formulated in (5)–(10), following Hochreiter et al. [36].

$$f_t = \text{sigmoid}(W_{hf} h_{t-1} + W_{Xf} X_t + b_f) \tag{5}$$

$$i_t = \text{sigmoid}(W_{hi} h_{t-1} + W_{Xi} X_t + b_i) \tag{6}$$

$$\tilde{C}_t = \tanh(W_{hC} h_{t-1} + W_{XC} X_t + b_C) \tag{7}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{8}$$

$$o_t = \text{sigmoid}(W_{ho} h_{t-1} + W_{Xo} X_t + b_o) \tag{9}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{10}$$

where f_t, i_t, and o_t represent the forget gate, input gate, and output gate, respectively, X_t denotes the input vector, $\widetilde{C}_t$ denotes the candidate vector, C_t denotes the current vector, and h_t denotes the hidden state at time step t. The activation functions sigmoid and tanh correspond to the logistic sigmoid and hyperbolic tangent, mapping their inputs to $(0, 1)$ and $(-1, 1)$, respectively. Here, W_{h*} and W_{X*} denote the recurrent and input weight matrices for the corresponding candidate-state update, respectively, and b_* denotes the associated bias vector.

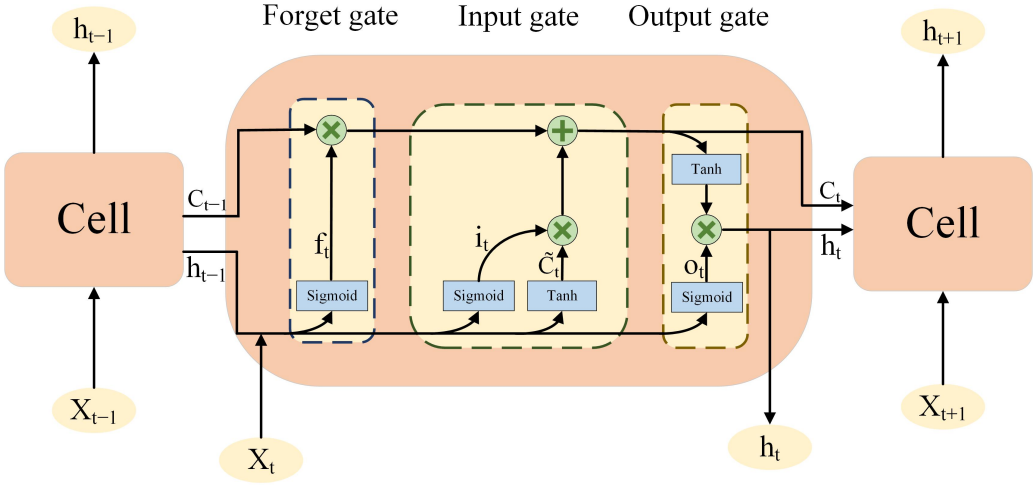


Figure 5. Diagram of LSTM memory cell structure.

Through the gated state-transition mechanism, LSTM can selectively preserve useful historical information and update temporal representations. This makes it suitable for modeling slowly varying components and long-term dependencies in time series, providing an appropriate prediction module for slow-varying components in the proposed framework.

3.5. NAMD-Net Framework

The NAMD-Net model prediction process for SWH is divided into five distinct stages. The complete process is shown in Figure 6.

1. The four input variable time series (SWH, WSPD, GST, and APD) selected via pearson correlation analysis are decomposed together using the NA-MEMD algorithm, yielding four corresponding subsets of time series

$$S = \{IMF_{S1}, IMF_{S2}, \ldots, IMF_{SN}, R_S\}$$
$$W = \{IMF_{W1}, IMF_{W2}, \ldots, IMF_{WN}, R_W\}$$
$$T = \{IMF_{T1}, IMF_{T2}, \ldots, IMF_{TN}, R_T\}$$
$$A = \{IMF_{A1}, IMF_{A2}, \ldots, IMF_{AN}, R_A\}$$

   where N denotes the number of IMFs decomposed for each variable, and R denotes the residual component.
2. The permutation entropy algorithm is used as a complexity-based criterion to group the corresponding components of the four subsets S, W, T, and A into two groups, denoted as G_H and G_L:

$$G_H = \{G_1^D, G_2^D, \ldots, G_M^D\}$$
$$G_L = \{G_{M+1}^D, G_{M+2}^D, \ldots, G_{N+1}^D\}$$

   where G_i^D denotes the i-th group, with each group containing the scale-aligned IMF components of the four variables after decomposition. Specifically, G_{N+1}^H denotes the residual component group, comprising the residual components of the four variables R_S, R_W, R_T, and R_A.
3. Each time series in the fast-varying group G_H and slow-varying group G_L is split into corresponding training, validation, and testing datasets.
4. For each component in the fast-varying group G_H, train WaveNet model using the training dataset, and use the validation set to tune model parameters and select the optimal

model. Finally, input the test set into the optimal model to obtain the corresponding fast-varying prediction results

$$Y_H = \{Y_1^H, Y_2^H, \ldots, Y_M^H\}$$

For each component in the slow-varying group G_L, train LSTM model using the training dataset, and use the validation set to tune model parameters and select the optimal model. Finally, input the test set into the optimal model to obtain the corresponding slow-varying prediction results

$$Y_L = \{Y_{M+1}^L, Y_{M+2}^L, \ldots, Y_{N+1}^L\}$$

5. The final SWH prediction is obtained by summing the corresponding components of the fast-varying predictions Y_H and the slow-varying predictions Y_L. This summation is essentially a mathematical reconstruction of the SWH time series and does not involve a physical decomposition of wave energy or a separation between wind-sea and swell.

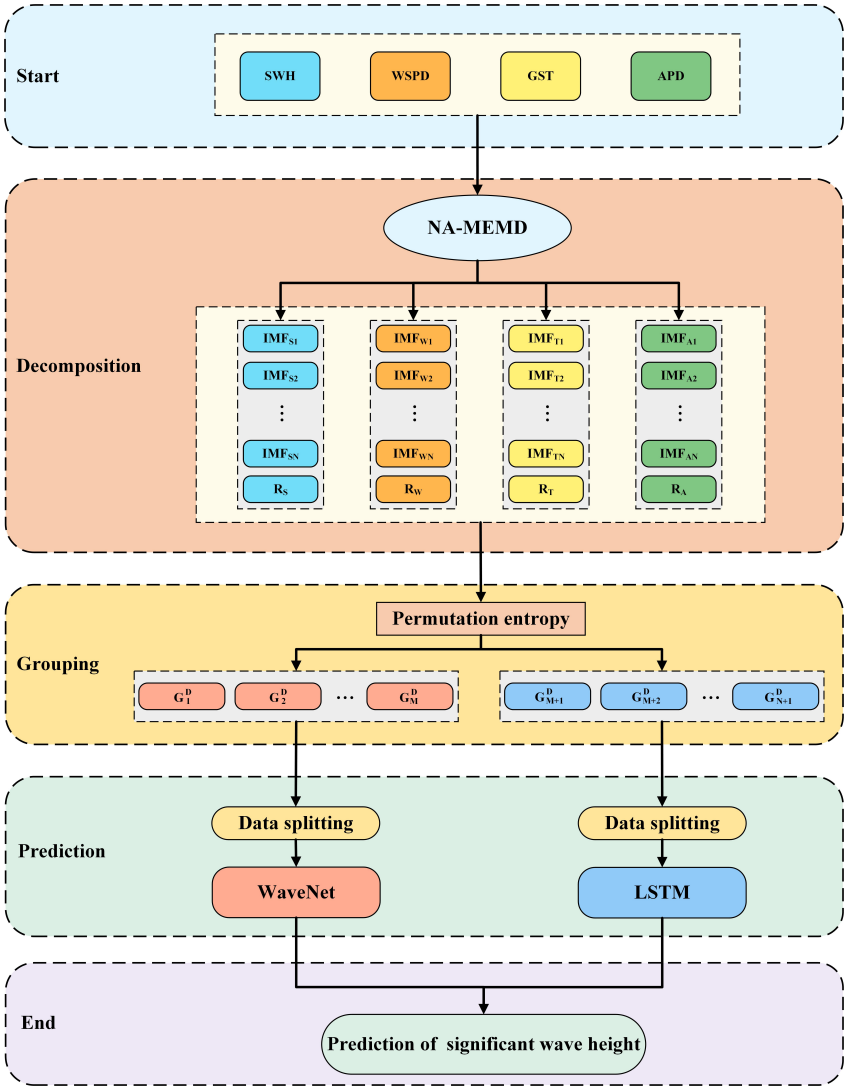


Figure 6. Flowchart of the NAMD-Net model.

In this study, NA-MEMD is performed offline on the available multivariate record before model training. Therefore, the current framework should be regarded as an offline decomposition-based forecasting framework rather than a strictly causal online decomposition scheme.

4. Experimental Background and Setup

4.1. NA-MEMD Decomposition Results

Since the selected variables have different physical units and magnitude ranges, all variables were normalized before NA-MEMD decomposition to reduce scale imbalance during multivariate processing. Normalization was performed using the min–max normalization method, calculated independently for each feature, as

$$o' = \frac{o - o_{\min}}{o_{\max} - o_{\min}} \tag{11}$$

where o denotes the original value; $o_{\min}$ and $o_{\max}$, the minimum and maximum values used for feature scaling, respectively; and o', the normalized value, ranging from 0 to 1.

In the NA-MEMD decomposition, white Gaussian noise with an intensity of 0.1 was added. The number of auxiliary noise channels was set to 4, and the number of direction vectors was set to 50. The added noise channels were only used to assist the multivariate decomposition process and were not interpreted as physical components.

Figures 7 and 8 respectively show the decomposition results obtained using the NA-MEMD algorithm after normalization for SWH from buoys 41,013 and 41,046. As shown in the figures, the SWH time series from buoys 41,013 and 41,046 were decomposed into seventeen IMFs and one residual component (Res). The other three variables were decomposed into the same number of IMF components due to the multivariate decomposition property of NA-MEMD. Since NA-MEMD introduces auxiliary Gaussian noise channels, the robustness of the decomposition against different noise realizations was further examined. Five random seeds were selected while keeping all other decomposition parameters unchanged. As summarized in Table 2, the number of extracted IMFs remained identical for both buoy stations under all noise realizations, indicating that the decomposition structure is stable with respect to the stochastic perturbations introduced by the auxiliary noise.

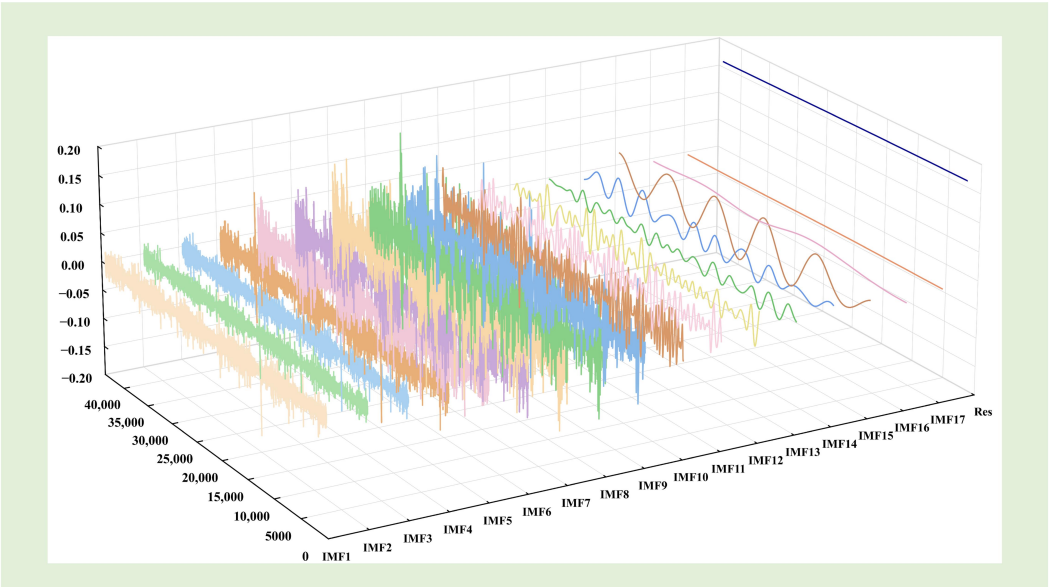


Figure 7. NA-MEMD decomposition of the normalized SWH time series at buoy 41,013, including 17 IMFs and Res.

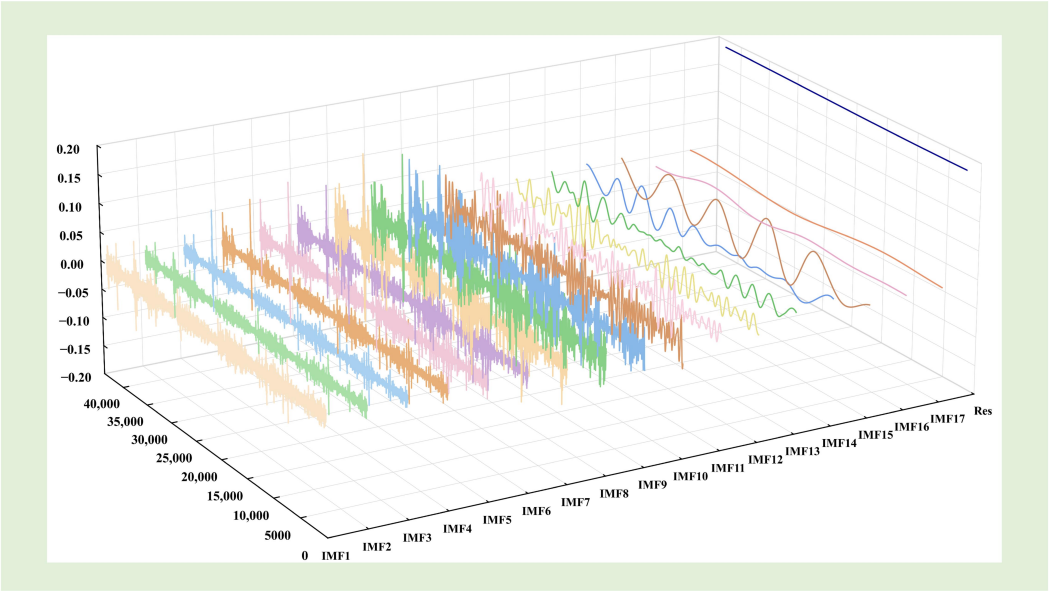


Figure 8. NA-MEMD decomposition of the normalized SWH time series at buoy 41,046, including 17 IMFs and Res.

Table 2. Stability analysis of the number of extracted IMFs from NA-MEMD under different random seeds.

Buoy	Seed 9	Seed 43	Seed 44	Seed 64	Seed 75
41,013	17	17	17	17	17
41,046	17	17	17	17	17

To further examine the mode-separation characteristics of NA-MEMD, the normalized power spectral densities (PSDs) of the first five SWH IMFs at buoy 41,013 were compared with those obtained using conventional EMD. Following the frequency-domain comparison approach in [43], the PSD of each IMF was estimated using Welch's method with a 512-sample Hann window, 50% overlap, and a 4096-point fast Fourier transform (FFT) at an hourly sampling frequency of $f_s = 1\ \mathrm{h}^{-1}$. Each PSD was normalized by its corresponding maximum value to facilitate comparison among IMFs. As shown in Figure 9, the EMD components exhibit broader spectral overlap among adjacent IMFs, whereas the NA-MEMD components show a clearer sequential separation of their dominant frequency ranges.

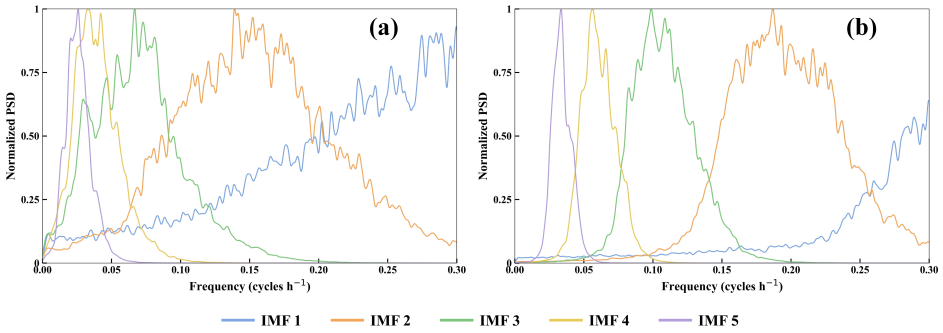


Figure 9. Normalized Welch PSDs of the first five SWH IMFs at buoy 41,013, where each PSD was normalized by its corresponding maximum value. (**a**) EMD. (**b**) NA-MEMD.

4.2. Permutation Entropy and Spectral Analysis

Permutation entropy was employed as a complexity-based criterion for grouping the decomposed components. For each component index, the normalized permutation entropy values of the four variables were averaged, and the mean value of all retained components was used as the grouping threshold. The Res was included because it forms part of the complete NA-MEMD representation and is subsequently modeled and incorporated into the final reconstruction. As shown in Figure 10, under the adopted parameters $a = 3$ and $U_d = 1$, IMF1 to IMF6 were above the corresponding threshold at both stations, whereas IMF7 to Res were below it. Accordingly, the components were assigned to the two prediction branches. Here, permutation entropy measures ordinal-pattern complexity rather than spectral frequency itself. Therefore, the permutation entropy results are treated as a complexity-guided grouping criterion.

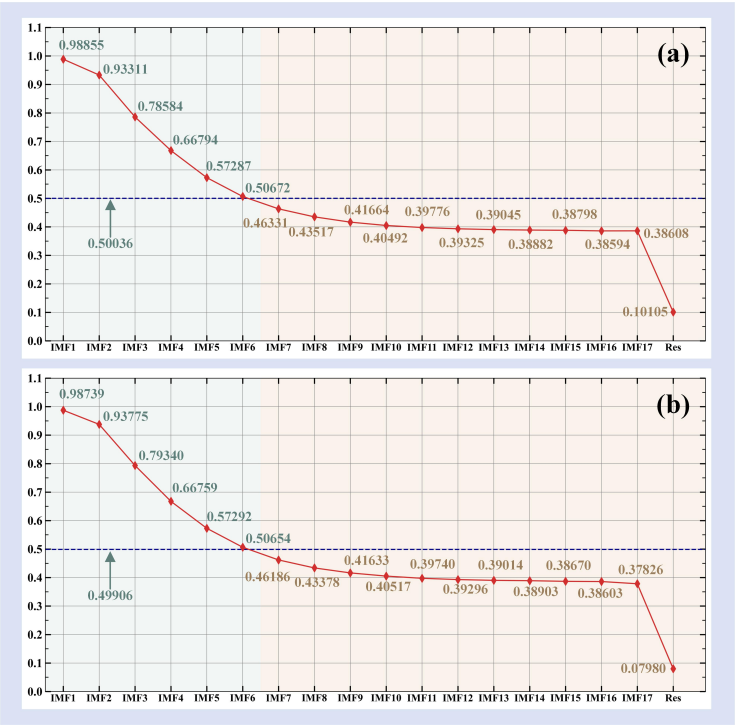


Figure 10. Permutation entropy analysis results for the IMFs and Res of each variable. (**a**) Buoy 41,013. (**b**) Buoy 41,046.

To evaluate the sensitivity of the grouping criterion, the embedding dimension was varied from 3 to 7 and the delay time from 1 to 3, resulting in 15 parameter combinations. At both buoy stations, when Res was included in the threshold calculation, the grouping boundary occurred between IMF6 and IMF7 in 10 cases, between IMF5 and IMF6 in four cases, and between IMF7 and IMF8 in one case. When Res was excluded, the boundary occurred between IMF6 and IMF7 in nine cases and between IMF5 and IMF6 in six cases. Excluding Res changed the boundary in only three of the 15 settings at each station, with a shift of one adjacent IMF. Under the baseline setting, IMF6 changed branch assignment when Res was excluded. These results indicate that the exact assignment of the transition component is sensitive to the threshold definition, whereas the overall complexity ordering remains confined to a narrow IMF5-IMF7 transition region.

For an independent quantitative characterization of the temporal scales of all decomposed components, the spectral centroid of each of the 17 oscillatory IMFs was calculated from its Welch PSD. A longer segment length of 8192 samples with 50% overlap was

adopted to provide finer frequency resolution. Res was excluded because it represents the non-oscillatory residual trend. The spectral centroid was calculated as

$$f_c = \frac{\sum_i f_i Q(f_i)}{\sum_i Q(f_i)} \tag{12}$$

where $Q(f_i)$ denotes the Welch PSD at frequency f_i. The spectral centroid represents the power-weighted mean frequency and provides a compact measure of the temporal scale of each IMF [44]. As shown in Figure 11, the spectral centroids decrease systematically with increasing IMF index at both stations, while the corresponding components of SWH, WSPD, GST, and APD exhibit closely matched frequency scales. The IMF5-IMF7 transition region therefore represents a continuous transition from relatively faster to slower temporal variations rather than indicating a sharp physical frequency cutoff. These spectral results provide independent temporal-scale support for the permutation entropy-guided branch grouping.

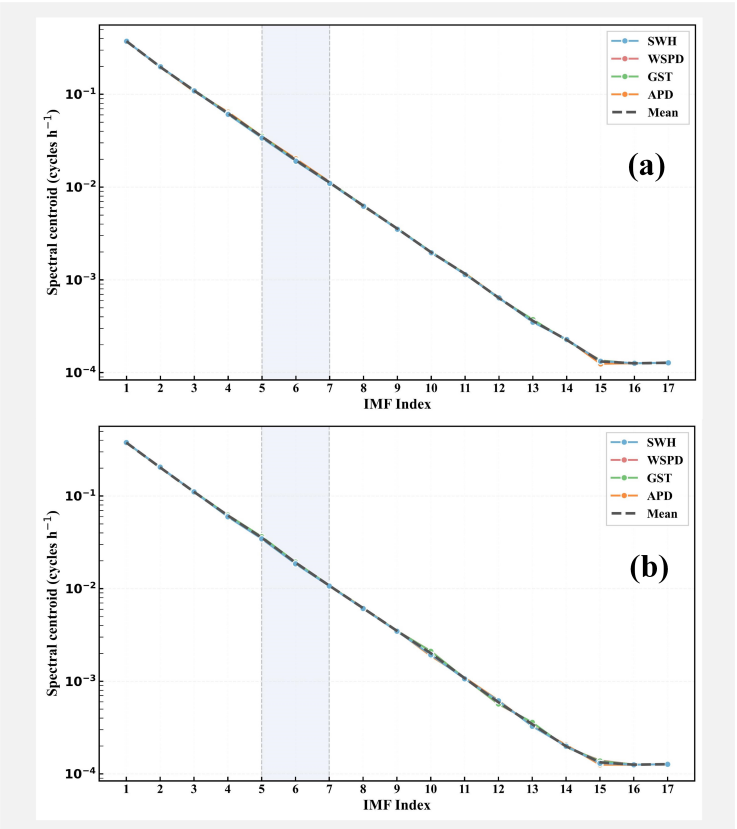


Figure 11. Spectral diagnostics of NA-MEMD components. The spectral centroid of each IMF was calculated using Welch power spectral density analysis. The shaded region represents the transition region (IMF5-IMF7) identified by the permutation entropy algorithm sensitivity analysis. (**a**) Buoy 41,013. (**b**) Buoy 41,046.

4.3. Model Training Settings

After offline NA-MEMD decomposition, the components were grouped by permutation entropy and then chronologically divided into training, validation, and testing subsets at a ratio of 3:1:1. The first 60% was used for training, the following 20% for validation, and the final 20% for testing. For buoy 41,013, the training, validation, and testing subsets contained 26,010, 8670, and 8670 temporal records, covering 1 January 2017–26 December 2019, 26 December 2019–28 December 2020, and 28 December 2020–31 December 2021, respectively. For buoy 41,046, the corresponding subsets contained 25,749, 8583, and 8583 records, covering

1 January 2017–2 January 2020, 2 January 2020–1 January 2021, and 1 January–31 December 2021, respectively. Sliding-window samples were then constructed independently within each subset using a 48-h historical input and a 24-h forecast horizon. Any window containing a temporal gap was discarded, and no window was allowed to cross subset boundaries. For model training, min–max scaling parameters were estimated only from the training subset and subsequently applied unchanged to the validation and testing subsets [45].

In the WaveNet branch, six gated dilated causal convolution blocks with residual and skip connections are used to construct a multi-scale receptive field, while moderate dropout is adopted to suppress overfitting to local noise. In the LSTM branch, two stacked LSTM layers are used for temporal modeling, and dropout is employed to improve generalization. To support training stability and reproducibility, both branches share consistent training strategy settings (such as batch size, patience, early stopping strategy, loss function, optimizer, and learning rate). The specific values for each parameter are detailed in Table 3. For reproducibility, all neural-network experiments were conducted with a fixed random seed of 42.

Table 3. Key Parameter Settings for WaveNet and LSTM.

Parameter	WaveNet	LSTM
Hidden layers	6	2
Hidden size	–	128
Kernel size	2	–
Dilation rate	1, 2, 4, 8, 16, 32	–
Residual channels	48	–
Skip channels	32	–
Dropout rate	0.15	0.5
Batch size	64	64
Epochs	150	150
Patience	30	30
Loss function	MSE	MSE
Optimizer	Adam	Adam
Learning rate	0.002	0.002

4.4. Evaluation Metrics

Four metrics, including root mean square error (RMSE), mean absolute error (MAE), symmetric mean absolute percentage error (sMAPE), and the correlation coefficient (R), are adopted for model performance assessment. These metrics quantify the discrepancy between predicted and observed values, as defined in (13)–(16) [46].

$$\mathrm{RMSE} = \sqrt{\frac{1}{n_s}\sum_{d=1}^{n_s}(ob_d - pr_d)^2} \tag{13}$$

$$\mathrm{MAE} = \frac{1}{n_s}\sum_{d=1}^{n_s}|ob_d - pr_d| \tag{14}$$

$$\mathrm{sMAPE} = \frac{1}{n_s}\sum_{d=1}^{n_s}\frac{2|ob_d - pr_d|}{|ob_d| + |pr_d|} \times 100\% \tag{15}$$

$$\mathrm{R} = \frac{\sum_{d=1}^{n_s}[(ob_d - \bar{ob})(pr_d - \bar{pr})]}{\sqrt{\sum_{d=1}^{n_s}(ob_d - \bar{ob})^2 \sum_{d=1}^{n_s}(pr_d - \bar{pr})^2}} \tag{16}$$

where ob_d and pr_d denote the d-th observed and predicted values, respectively; $\bar{ob}$ and $\bar{pr}$ denote the means; and n_s denotes the number of samples.

5. Results and Discussion

5.1. Results

Figure 12 shows the scatter plot results of NAMD-Net for buoys 41,013 and 41,046 at different lead times (1 h, 6 h, 12 h, and 24 h). The model adopts a direct multi-output strategy, using a 48-hour historical input sequence to directly generate a 24-dimensional forecast vector in a single run. No autoregressive iterative prediction is used during inference. Statistical analysis was performed on the 1st, 6th, 12th, and 24th output dimensions, corresponding to lead times of 1 h, 6 h, 12 h, and 24 h, respectively. At both stations, for all four lead times, most data points cluster near the 1:1 reference line, indicating close agreement between forecast and observed values. As lead time increases from 1 h to 24 h, scatterplot dispersion grows, especially for larger SWH values where more samples deviate from the reference line. This lead-time-dependent degradation reflects the increasing difficulty of longer-horizon forecasting [47]. Nevertheless, an excellent fit is maintained across the four lead times at both stations, indicating consistent forecasting performance at the two Atlantic stations.

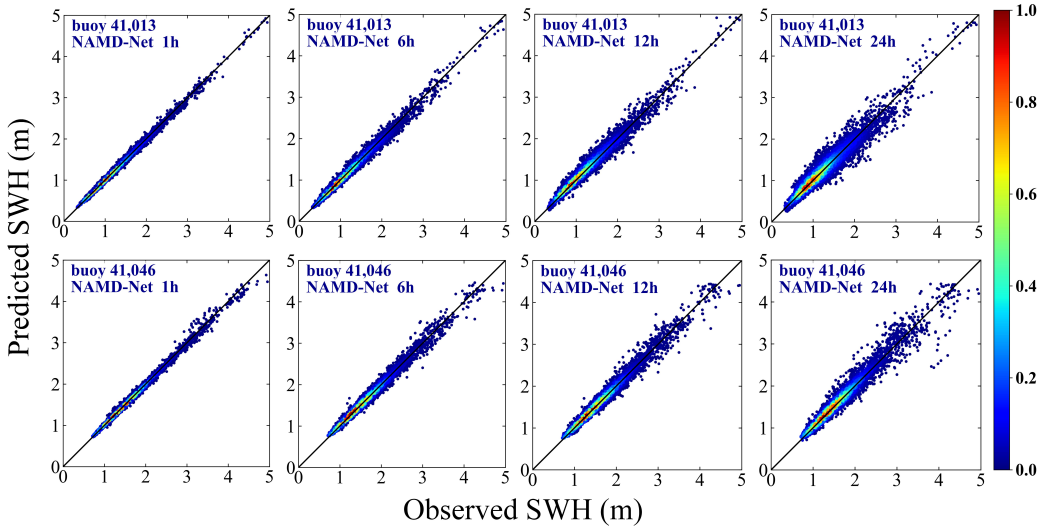


Figure 12. Scatter plot density distributions of observed SWH versus predicted SWH for buoys 41,013 and 41,046 under different lead times.

To comprehensively evaluate the proposed model, ten methods were compared: BP, RF, Transformer, Autoregressive with exogenous inputs (ARX), WaveNet, LSTM, MEMD+LSTM (ML), NA-MEMD+LSTM (NML), NA-MEMD+WaveNet (NMW), and NAMD-Net. These methods cover conventional machine-learning models, standalone deep-learning forecasting models, and decomposition-based forecasting models.

For a consistent comparison, all comparison models used the same chronological data partitions and forecasting setup wherever applicable. The BP model employed two hidden layers with 128 and 64 neurons. RF used 200 trees with the square-root feature-selection rule. The Transformer contained two encoder layers with a model dimension of 64, four attention heads, a feed-forward dimension of 128, and a dropout rate of 0.1. For ARX, the lag order was selected from 12, 24, 36, and 48 h according to the validation mean squared error. WaveNet and LSTM used the corresponding configurations listed in Table 3. ML and NML employed the same LSTM configuration after MEMD and NA-MEMD decomposition, respectively, whereas NMW applied the same WaveNet configuration to all NA-MEMD components.

At buoy stations 41,013 and 41,046, four forecast lead times (1 h, 6 h, 12 h, and 24 h) were set to compare the ten models using RMSE, MAE, sMAPE, and R. Tables 4 and 5 and Figures 13 and 14 present the results.

Table 4. Evaluation Metrics for Buoy 41,013 Across Various Lead Times.

Hour	Index	BP	RF	Transformer	ARX	WaveNet	LSTM	ML	NML	NMW	NAMD-Net
1	RMSE (m)	0.123	0.120	0.097	0.091	0.146	0.153	0.046	0.045	0.046	0.042
	MAE (m)	0.093	0.086	0.068	0.065	0.110	0.109	0.035	0.034	0.035	0.031
	sMAPE (%)	8.26	7.13	5.45	5.27	9.19	8.90	3.10	3.10	3.08	2.61
	R	0.977	0.977	0.985	0.987	0.972	0.962	0.997	0.997	0.997	0.997
6	RMSE (m)	0.284	0.281	0.272	0.262	0.288	0.384	0.082	0.078	0.077	0.072
	MAE (m)	0.204	0.201	0.187	0.186	0.204	0.270	0.061	0.057	0.059	0.052
	sMAPE (%)	16.88	16.35	14.99	14.95	16.55	20.87	5.19	4.88	4.82	4.32
	R	0.859	0.862	0.875	0.885	0.856	0.725	0.989	0.990	0.991	0.992
12	RMSE (m)	0.437	0.414	0.401	0.396	0.438	0.532	0.117	0.110	0.109	0.102
	MAE (m)	0.324	0.305	0.289	0.292	0.314	0.385	0.085	0.080	0.078	0.075
	sMAPE (%)	26.17	24.41	23.27	23.25	24.96	29.88	7.28	6.70	6.67	6.26
	R	0.631	0.663	0.686	0.710	0.624	0.444	0.977	0.980	0.981	0.984
24	RMSE (m)	0.556	0.520	0.516	0.507	0.563	0.643	0.186	0.171	0.167	0.155
	MAE (m)	0.423	0.397	0.381	0.386	0.419	0.483	0.141	0.125	0.122	0.111
	sMAPE (%)	33.12	31.49	30.60	30.64	32.34	37.54	11.70	10.40	10.03	9.15
	R	0.293	0.364	0.377	0.425	0.304	0.196	0.946	0.949	0.951	0.958

Table 5. Evaluation Metrics for Buoy 41,046 Across Various Lead Times.

Hour	Index	BP	RF	Transformer	ARX	WaveNet	LSTM	ML	NML	NMW	NAMD-Net
1	RMSE (m)	0.121	0.126	0.115	0.106	0.136	0.126	0.078	0.047	0.047	0.045
	MAE (m)	0.086	0.084	0.084	0.076	0.098	0.095	0.063	0.035	0.034	0.031
	sMAPE (%)	5.26	5.04	5.17	4.60	6.20	5.94	4.15	2.33	2.23	1.92
	R	0.978	0.975	0.980	0.982	0.973	0.980	0.995	0.997	0.997	0.997
6	RMSE (m)	0.223	0.218	0.212	0.208	0.224	0.221	0.095	0.094	0.090	0.077
	MAE (m)	0.145	0.144	0.140	0.137	0.149	0.158	0.070	0.071	0.067	0.056
	sMAPE (%)	8.59	8.70	8.42	8.31	9.14	9.61	4.36	4.71	4.46	3.44
	R	0.922	0.922	0.927	0.930	0.919	0.929	0.987	0.987	0.988	0.991
12	RMSE (m)	0.310	0.305	0.303	0.296	0.329	0.307	0.110	0.102	0.101	0.093
	MAE (m)	0.208	0.205	0.207	0.197	0.223	0.218	0.078	0.073	0.074	0.066
	sMAPE (%)	12.37	12.43	12.44	11.99	13.30	13.11	4.77	4.54	4.61	4.00
	R	0.843	0.844	0.849	0.854	0.821	0.854	0.981	0.985	0.983	0.987
24	RMSE (m)	0.438	0.437	0.430	0.419	0.481	0.424	0.160	0.147	0.144	0.135
	MAE (m)	0.295	0.303	0.301	0.294	0.324	0.306	0.108	0.100	0.099	0.089
	sMAPE (%)	17.42	18.15	17.86	17.71	18.93	18.30	6.47	6.31	6.08	5.28
	R	0.665	0.661	0.686	0.694	0.592	0.693	0.962	0.969	0.970	0.974

As shown in these tables and figures, the proposed model yields the lowest RMSE, MAE, and sMAPE, and the highest R at all lead times at both stations. At buoy 41,013, its RMSE values for lead times of 1, 6, 12, and 24 h are 0.042, 0.072, 0.102, and 0.155, with R values of 0.997, 0.992, 0.984, and 0.958. At buoy 41,046, the proposed model yields RMSE values of 0.045, 0.077, 0.093, and 0.135, with R values of 0.997, 0.991, 0.987, and 0.974. The visual comparison in Figures 13 and 14 corroborates these quantitative results.

With increasing lead time from 1 h to 24 h, all models show increasing RMSE, MAE, sMAPE, and declining R. This degradation is especially severe for BP, RF, and LSTM at 12 h and 24 h. In contrast, the metrics of ML, NML, NMW, and the proposed model change more gradually.

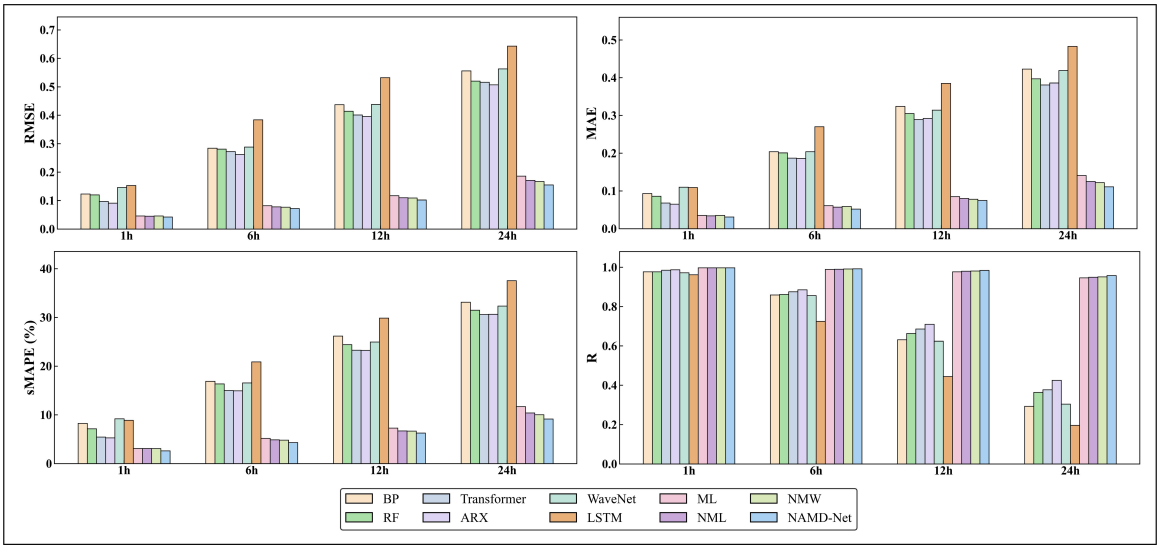


Figure 13. Comparison of the performance of multiple models using four evaluation metrics for buoy 41,013.

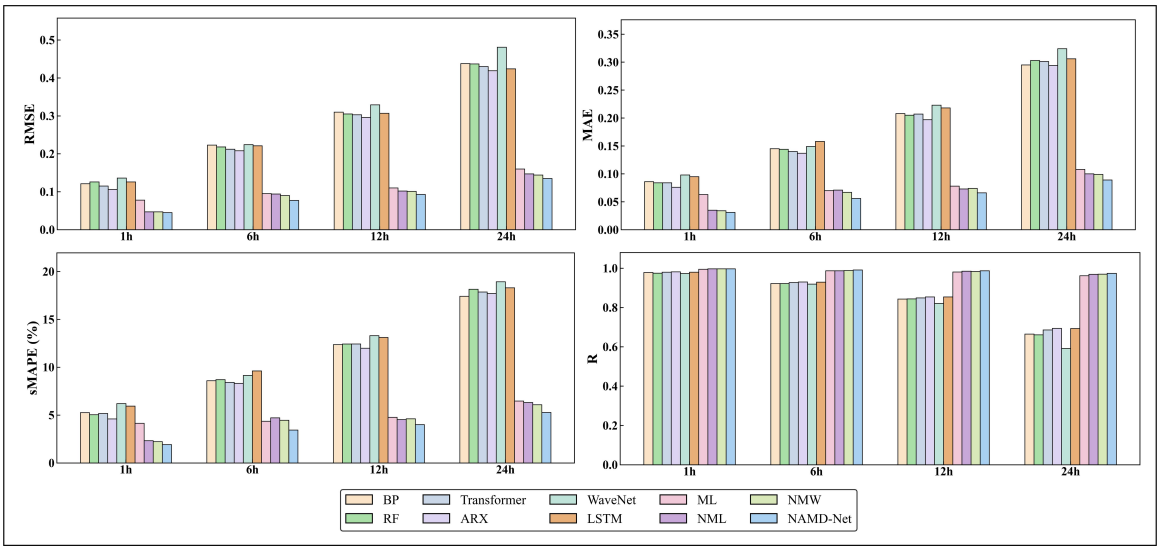


Figure 14. Comparison of the performance of multiple models using four evaluation metrics for buoy 41,046.

As shown in Tables 4 and 5, the forecasting results at buoys 41,013 and 41,046 are in line with the design motivation of NAMD-Net for nonlinear and nonstationary SWH prediction. The original SWH-related sequence contains both rapid local fluctuations and slower temporal evolution, which are difficult to represent using a single direct forecasting structure. NAMD-Net alleviates this difficulty by decomposing the multivariate signal into scale-aligned components and applying complexity-guided branch modeling. The comparison between ML and NML reflects the benefit of NA-MEMD over MEMD under the same LSTM forecasting structure. The WaveNet branch captures rapidly varying components, whereas LSTM models slowly varying components. Furthermore, the NML and NMW variants, which uniformly apply LSTM and WaveNet to all NA-MEMD components, respectively, perform consistently worse than NAMD-Net at both buoys, supporting the use of component-specific dual-branch modeling. Therefore, the long-lead forecasting advantage of NAMD-Net is associated with both multivariate scale-aligned decomposition and the differentiated modeling of components with different temporal characteristics.

Furthermore, forecasting performance at buoy 41,046 is slightly better than at buoy 41,013. For the 24 h forecast, the proposed model reduces RMSE, MAE, and sMAPE at buoy 41,046 by 12.9%, 19.8%, and 42.3% relative to buoy 41,013, while R increases from 0.958 to 0.974. The results also indicate modest site-dependent differences in forecasting performance [48,49].

To further assess the robustness of the reported improvements, multi-seed experiments and statistical significance analyses were conducted for the key model comparisons. Table 6 reports the RMSE averaged over five independent runs (random seeds 42, 43, 44, 64, and 75) together with the sample standard deviation (reflecting multi-seed training variability) [50]; while the paired RMSE differences are evaluated over the test set sequence (to account for temporal autocorrelation) with 95% confidence intervals [51] and Diebold–Mariano (DM) test [52] results.

Table 6. Multi-seed RMSE statistics and statistical significance analysis.

Buoy	Hour	Transformer RMSE (±SD)	NML RMSE (±SD)	NAMD-Net RMSE (±SD)	NAMD-Transformer ΔRMSE [95% CI], p	NAMD-NML ΔRMSE [95% CI], p
41,013	1 h	0.0957 ± 0.0016	0.0520 ± 0.0123	0.0479 ± 0.0070	−0.0478 [−0.0525, −0.0432] $p < 0.001$	−0.0041 [−0.0052, −0.0030] $p < 0.001$
41,013	6 h	0.2692 ± 0.0023	0.0790 ± 0.0062	0.0761 ± 0.0056	−0.1931 [−0.2072, −0.1743] $p < 0.001$	−0.0029 [−0.0038, −0.0020] $p < 0.001$
41,013	12 h	0.4018 ± 0.0033	0.1106 ± 0.0043	0.1050 ± 0.0029	−0.2968 [−0.3238, −0.2701] $p < 0.001$	−0.0056 [−0.0071, −0.0041] $p < 0.001$
41,013	24 h	0.5157 ± 0.0067	0.1639 ± 0.0066	0.1572 ± 0.0044	−0.3585 [−0.3999, −0.3157] $p < 0.001$	−0.0067 [−0.0071, −0.0062] $p < 0.001$
41,046	1 h	0.1127 ± 0.0018	0.0496 ± 0.0079	0.0451 ± 0.0034	−0.0676 [−0.0746, −0.0611] $p < 0.001$	−0.0045 [−0.0057, −0.0033] $p < 0.001$
41,046	6 h	0.2123 ± 0.0016	0.0867 ± 0.0069	0.0797 ± 0.0023	−0.1326 [−0.1549, −0.1106] $p < 0.001$	−0.0070 [−0.0083, −0.0058] $p < 0.001$
41,046	12 h	0.3030 ± 0.0036	0.1051 ± 0.0062	0.0966 ± 0.0036	−0.2064 [−0.2422, −0.1722] $p < 0.001$	−0.0085 [−0.0104, −0.0064] $p < 0.001$
41,046	24 h	0.4306 ± 0.0036	0.1445 ± 0.0040	0.1371 ± 0.0024	−0.2935 [−0.3463, −0.2451] $p < 0.001$	−0.0074 [−0.0098, −0.0051] $p < 0.001$

As shown in Table 6, SD denotes the sample standard deviation of RMSE over the five independent runs. In the last two columns, the three values in each cell denote the ΔRMSE, the 95% confidence interval, and the p-value, respectively. Negative ΔRMSE indicates a lower RMSE for NAMD-Net than for the baseline. NAMD-Net significantly outperforms Transformer at both buoys and all forecast horizons. Compared with NML, NAMD-Net yields lower RMSE across all evaluated horizons and both stations, with the DM test indicating statistically significant differences ($p < 0.001$).

Further analysis using time series is required to examine local peaks, abrupt changes, and dynamic details in the ocean wave evolution process. Taking buoy 41,013 as an example, a representative continuous segment from the test set was selected to depict comparison curves of observed and predicted time series for different models, as shown in Figure 15.

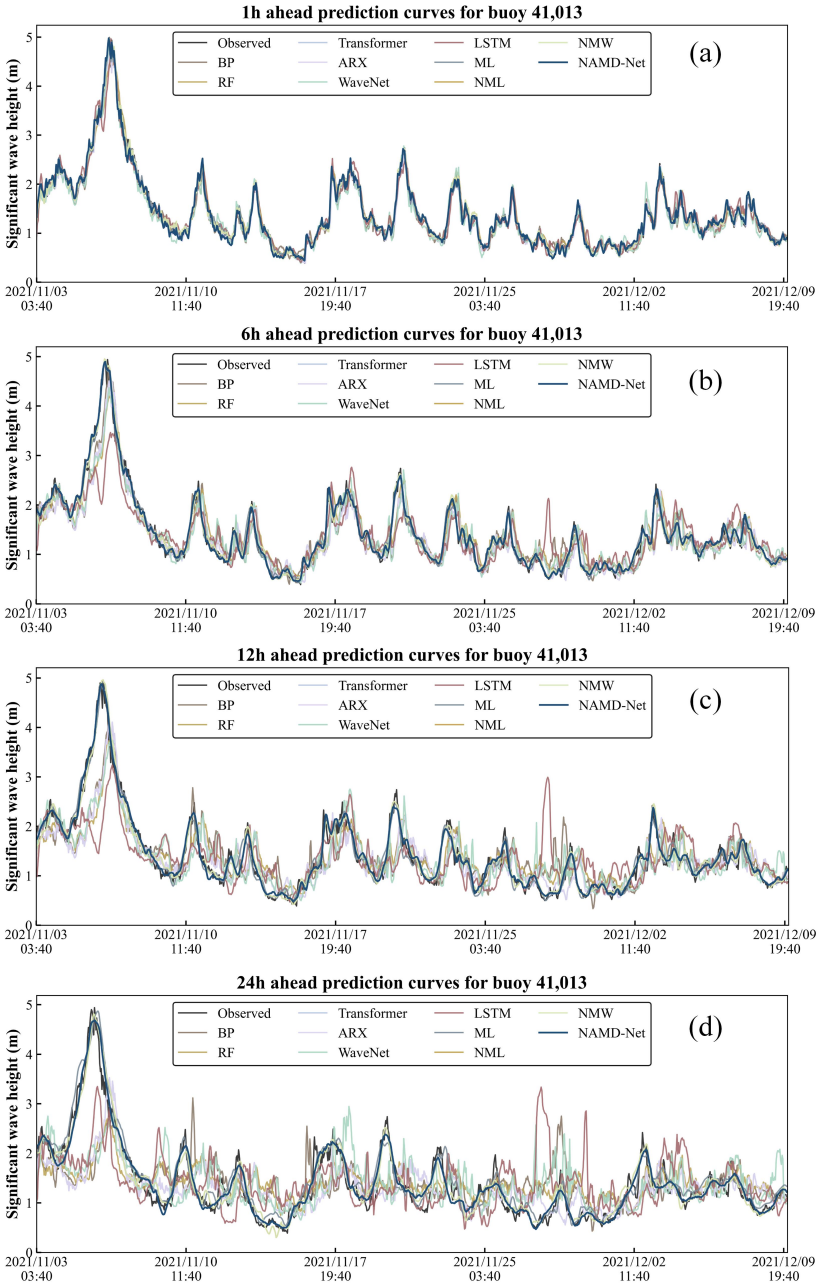


Figure 15. Comparison of time series data for buoy 41,013 between model forecasts and observed values at different lead times. (**a**) 1 h. (**b**) 6 h. (**c**) 12 h. (**d**) 24 h.

It can be observed that the non-decomposed model tends to underestimate the SWH at local abrupt changes and high-value peaks, resulting in a smooth curve. Furthermore, a certain time shift occurs as the forecast lead time increases. In contrast, NAMD-Net is able to stably track the observed sequence, maintaining a robust response particularly near high-value peaks, indicating that it not only captures the overall trend but also reconstructs local abrupt changes.

To further evaluate forecasting performance under extreme wave conditions, the 95th-percentile (P95) of the observed SWH during the test period at buoy 41,013 (2.24 m) was used as the threshold, yielding 358 extreme-wave samples at each forecast horizon [53,54]. NAMD-Net was compared with NML and Transformer at 1 h, 6 h, 12 h, and 24 h horizons, and the results are summarized in Table 7.

Table 7. Quantitative evaluation of predictive performance under extreme sea states for buoy 41,013 across different lead times.

Model	Index	1 h	6 h	12 h	24 h
Transformer	RMSE (m)	0.229	0.672	1.074	1.443
	MAE (m)	0.184	0.541	0.933	1.313
	sMAPE (%)	6.88	22.68	42.77	64.07
	R	0.942	0.664	0.510	0.254
NML	RMSE (m)	0.085	0.155	0.218	0.355
	MAE (m)	0.066	0.127	0.169	0.283
	sMAPE (%)	2.40	4.54	6.28	10.97
	R	0.990	0.967	0.938	0.864
NAMD-Net	RMSE (m)	0.081	0.150	0.189	0.304
	MAE (m)	0.063	0.122	0.146	0.243
	sMAPE (%)	2.25	4.13	5.49	9.46
	R	0.991	0.970	0.948	0.898

The results show that NAMD-Net achieved the lowest RMSE, MAE, and sMAPE among the three models under extreme-wave conditions, particularly as the forecast horizon increased.

In addition, extreme-wave samples exceeding the P95 threshold were grouped into independent events based on the autocorrelation-derived independence interval, and only events lasting more than 3 h were retained [53], resulting in 22 independent extreme-wave events.

These events were further examined at the representative 6 h forecast horizon. The mean absolute peak-SWH errors of NAMD-Net, NML, and Transformer were 0.1478, 0.1878, and 0.3132 m, respectively, while the corresponding mean absolute peak-timing errors were 1.23, 1.50, and 5.05 h [55].

To further evaluate the practical forecasting capability of the proposed model, wave products from the European Centre for Medium-Range Weather Forecasts (ECMWF) were introduced as an additional physics-based benchmark [56]. For buoys 41,013 and 41,046, the ECMWF SWH fields in the test period (2021) were bilinearly interpolated to the buoy coordinates and temporally aligned with the hourly observations, with only matched timestamps retained. Using the buoy observations as ground truth, the 1 h SWH predictions of NAMD-Net were then compared with the corresponding ECMWF estimates. The ECMWF results yielded an RMSE of 0.197 m and an R of 0.937 at buoy 41,013, and an RMSE of 0.214 m and an R of 0.932 at buoy 41,046. The comparison results are shown in Figure 16.

To evaluate the computational efficiency of NAMD-Net and examine the trade-off between forecasting accuracy and computational cost, the trainable parameter count, training time, and average inference time were compared with those of representative baseline models on buoy 41,013. As shown in Table 8, NAMD-Net contains 2.97 million trainable parameters, fewer than the 3.67 million parameters of NML, indicating that its improved forecasting performance is not simply attributable to increased model capacity. Its training time is comparable to that of NML (2218.1 s versus 2116.3 s), while its average inference time is lower (0.254 versus 0.877 ms/sample). Compared with NMW, NAMD-Net requires more parameters but substantially less training time. Although Transformer has the lowest computational cost, its forecasting accuracy is lower than that of the decomposition-based models. These results indicate that NAMD-Net achieves improved forecasting accuracy with a moderate additional computational cost.

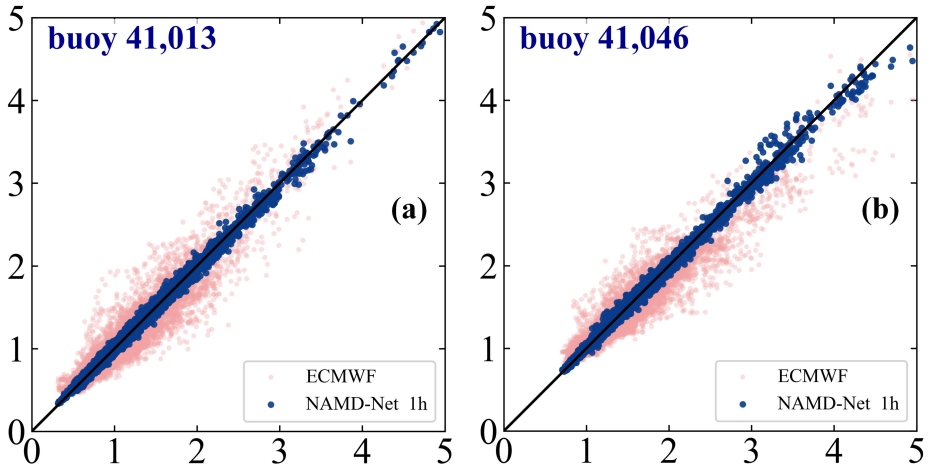


Figure 16. Scatter plots of observed versus predicted 1 h SWH for NAMD-Net and ECMWF. (**a**) Buoy 41,013. (**b**) Buoy 41,046.

Table 8. Computational cost and 24 h forecasting accuracy of different models at buoy 41,013.

Model	Trainable Parameters	Training Time (s)	Average Inference Time (ms/Sample)	24 h RMSE (m)
Transformer	87,136	55.334	0.0077	0.516
NML	3,668,400	2116.317	0.8767	0.171
NMW	1,579,536	3734.352	0.1303	0.167
NAMD-Net	2,972,112	2218.093	0.2540	0.155

5.2. Analysis

To examine the contribution of multivariate information, an auxiliary comparison was conducted at buoy 41,046, with the quantitative results shown in Table 9. The physical reason can be explained by the wave energy definition.

SWH H_s represents the integrated energy level of the wave spectrum and is related to its zeroth spectral moment m_0 by

$$H_s = 4\sqrt{m_0}.$$

Wave-spectrum evolution is governed by propagation together with physical source terms including wind input, nonlinear wave–wave interactions, and dissipation [57]. WSPD and GST provide information on the intensity and short-term variability of atmospheric forcing, whereas APD characterizes the temporal scale of the wave field. Therefore, the selected variables provide complementary information on wind forcing and the corresponding wave-state response.

Table 9. RMSE comparison for different input configurations at buoy 41,046.

Model	Input	1 h	6 h	12 h	24 h
LSTM	SWH	0.164	0.287	0.380	0.559
LSTM	SWH + WSPD + GST + APD	0.126	0.221	0.307	0.424
NAMD-Net	SWH + WSPD + GST + APD	0.045	0.077	0.093	0.135

As shown in Table 9, the multivariate LSTM clearly outperforms the SWH-only LSTM, indicating the value of additional wind-wave information. NAMD-Net further improves the

results, suggesting that joint multiscale decomposition can exploit these coupled variables more effectively than direct multivariate input.

To further examine the geographical applicability of the proposed framework, an additional validation was conducted at Pacific buoy 46,028 (35.763°N, 121.900°W; water depth: 1136 m) using data from 2020–2024 under the same preprocessing procedure, input variables, forecasting setup, and model configurations as the preceding experiments. To keep this additional validation concise, Transformer and NML were selected as representative non-decomposition and decomposition-based baselines, respectively, for comparison with the proposed model. The corresponding results are presented in Table 10. Similar to the visualization for the Atlantic buoys, the corresponding scatter density plots of observed versus predicted SWH at buoy 46,028 are shown in Figure 17.

Table 10. Forecasting performance at Pacific buoy 46,028.

Model	Index	1 h	6 h	12 h	24 h
Transformer	RMSE (m)	0.192	0.372	0.539	0.742
	MAE (m)	0.136	0.261	0.379	0.550
	sMAPE (%)	5.97	11.57	16.82	24.32
	R	0.978	0.913	0.807	0.605
NML	RMSE (m)	0.103	0.154	0.202	0.277
	MAE (m)	0.076	0.111	0.150	0.198
	sMAPE (%)	3.52	5.13	7.31	9.88
	R	0.994	0.986	0.979	0.958
NAMD-Net	RMSE (m)	0.082	0.143	0.167	0.229
	MAE (m)	0.053	0.101	0.120	0.162
	sMAPE (%)	2.39	4.53	5.37	7.43
	R	0.996	0.988	0.983	0.969

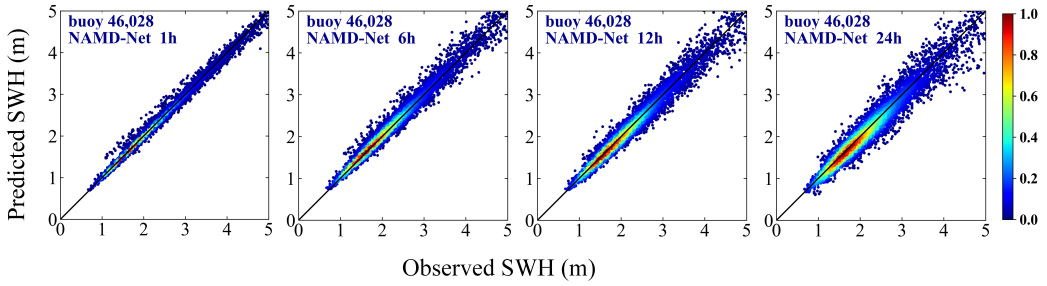


Figure 17. Scatter density distributions of observed versus predicted SWH for buoy 46,028 under different lead times.

The variation in forecasting performance with increasing lead time at buoy 46,028 is consistent with that observed at the two Atlantic buoys, with the proposed model maintaining the same overall advantage over the comparison models. This consistent pattern provides additional support for its applicability under different geographical conditions.

To further analyze the predictive performance of the proposed NAMD-Net model under different sea state conditions, this paper groups the test samples based on SWH, with SWH < 1.25 m, defined as low sea state, $1.25 \leq$ SWH < 2.5 m as moderate sea state, and SWH ≥ 2.5 m as high sea state [58]. The RMSE under different sea state conditions is shown in Figure 18. It can be seen that the RMSE values at both buoy stations are below 0.32 m, with overall smooth variations and no abnormal fluctuations. Meanwhile, RMSE exhibits a clear increasing trend: it increases with longer forecast lead times and progressively rises as sea state varies from low to high. This indicates that prediction errors are influenced not only by the length of the time horizon but also by increasing sea state complexity. Particularly

under high sea state conditions, ocean wave fluctuations are stronger and local peaks are more frequent, which will significantly increase the difficulty of long-term forecasting.

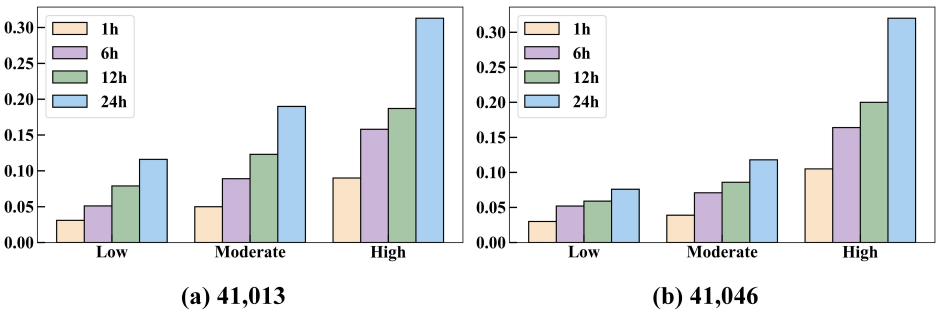


Figure 18. RMSE results for buoys 41,013 and 41,046 under different sea conditions and forecast lead times.

In addition to differences in sea conditions, seasonal background also influences the distribution of prediction errors. A further analysis of the RMSE results of the NAMD-Net model across seasons and forecast lead times is shown in Figure 19. It can be observed that the errors for the two buoys vary with season and forecast lead time, with the overall error levels being highest in winter, which is consistent with previously reported seasonal variability in SWH and wave climate [59,60]. For NAMD-Net specifically, the RMSE at both buoys exhibits a clear horizon-dependent increase across all evaluated horizons.

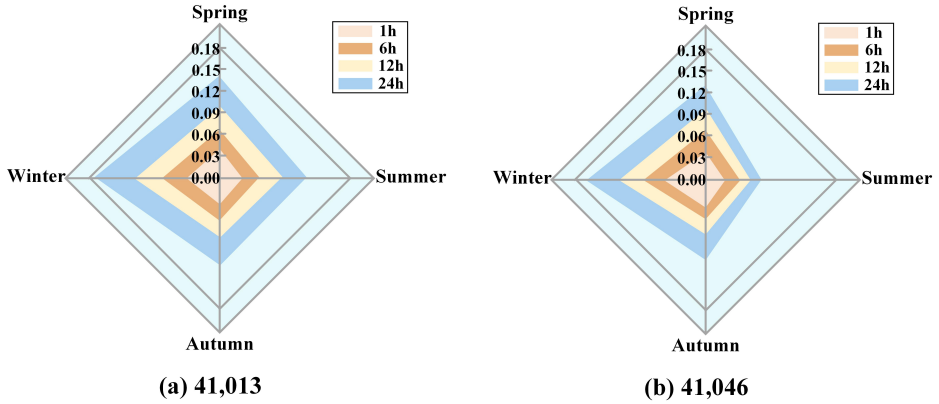


Figure 19. RMSE results for buoys 41,013 and 41,046 under different seasonal conditions and forecast lead times.

To assess the effect of future-information leakage associated with NA-MEMD, an additional causal re-decomposition experiment was conducted following the methods in [23,61,62]. Since NA-MEMD must be recalculated independently for each forecast origin, forecast origins were set at fixed 24 h intervals throughout the 2021 test period, yielding 365 candidate origins over the full year. At each origin, NA-MEMD was performed using only the observations available up to that time. The decomposed components from the preceding 48 h were then used as model inputs to directly generate the subsequent 24 h SWH forecast. Forecast origins containing temporal gaps longer than 3 h within the 48 h input window or the subsequent 24 h forecast period were excluded, yielding 346 valid forecast origins. This leakage-control experiment is illustrated using buoy 41,046, where the causal and offline results were evaluated on these same retained origins, as shown in Table 11.

Table 11. Comparison between offline and causal NA-MEMD implementations for NAMD-Net.

Hour	Metric	LSTM	Transformer	NAMD-Net (Offline)	NAMD-Net (Causal)
1 h	RMSE	0.137	0.111	0.052	0.064
	R	0.975	0.981	0.996	0.994
6 h	RMSE	0.224	0.203	0.084	0.112
	R	0.929	0.939	0.991	0.984
12 h	RMSE	0.293	0.275	0.103	0.151
	R	0.868	0.880	0.985	0.968
24 h	RMSE	0.404	0.400	0.141	0.214
	R	0.720	0.728	0.970	0.931

The results show that causal NA-MEMD yields higher RMSE and lower R than the offline implementation, suggesting that leakage may lead to optimistic accuracy estimates. Nevertheless, the causal NAMD-Net still achieves lower RMSE than the baselines at all horizons, indicating that its relative advantage is largely preserved despite the leakage effect.

6. Conclusions

This paper proposed the NAMD-Net model, grounded in multivariate adaptive decomposition and a dual-path deep neural network, for accurate prediction of SWH. The principal findings of this study are as follows:

1. SWH, WSPD, GST, and APD from buoy observations are jointly exploited so that cross-variable dependence in ocean-wave processes can be retained during decomposition and prediction.
2. NA-MEMD is employed to decompose the selected variables into scale-aligned IMFs, which alleviates mode mixing and avoids the component-matching problem caused by separate decomposition of multiple input variables.
3. A dual-branch architecture based on complexity-guided grouping and spectral validation is designed to assign fast-varying components to WaveNet and slow-varying components to LSTM, enabling the model to separately learn local abrupt variations and long-term trend information.
4. Extensive experiments on two Atlantic buoy stations indicate the effectiveness of the proposed model for 1 h to 24 h forecasts in the offline decomposition setting. In addition to overall accuracy comparisons, the analysis of lead-time-dependent error growth, the dependence of RMSE on sea state, and the seasonal variation in RMSE further indicate its robustness under complex wave conditions. Notably, the strictly causal re-decomposition experiment confirms that the relative advantage of NAMD-Net remains largely preserved.

A limitation of this study is the inherent optimistic bias of the offline decomposition framework. Although the strictly causal re-decomposition indicates that the relative advantage of NAMD-Net persists, a fully online decomposition framework remains an open research direction. Future research will explore integrating NAMD-Net with operational ocean reanalysis products and extending the current offline NA-MEMD framework toward real-time causal deployment in marine environmental monitoring networks.

Author Contributions: Conceptualization, Q.Z.; methodology, B.M.; software, B.M.; validation, H.L. and B.M.; formal analysis, X.L.; investigation, B.M.; data curation, B.M.; writing—original draft preparation, B.M.; writing—review and editing, W.H. and X.L.; visualization, B.M.; supervision, H.L., Q.Z. and W.H.; project administration, H.L.; funding acquisition, H.L. 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 under Grants 42406173 and 62371271; the Open Fund of Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering under Grant 2025SDSJ04; and the Cultivation Project of the Ezhou Innovation Development Joint Fund of Hubei Provincial Natural Science Foundation under Grant 2026AFC0653.

Data Availability Statement: All data used in this study are publicly available. The buoy data used in this study were obtained from the National Data Buoy Center https://www.ndbc.noaa.gov/ (accessed on 20 November 2025). The key implementation code of NAMD-Net is publicly available at GitHub v1.0 https://github.com/journeylife1115/NAMD-Net (accessed on 17 October 2026).

Acknowledgments: The authors would like to thank the National Oceanic and Atmospheric Administration for providing the buoy data used in this study.

Conflicts of Interest: The authors declare no conflicts of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Altunkaynak, A. Significant wave height prediction by using a spatial model. *Ocean Eng*. **2005**, *32*, 924–936. [CrossRef]
2. Huang, W.; Liu, X.; Gill, E.W. Ocean wind and wave measurements using X-band marine radar: A comprehensive review. *Remote Sens*. **2017**, *9*, 1261. [CrossRef]
3. Yan, Q.; Fan, C.; Meng, J.; Song, T.; Zhang, J.; Sun, W. Comparison of BPNN and dual-branch CNN for significant wave height estimation from polarimetric Gaofen-3 SAR wave mode data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens*. **2024**, *17*, 9582–9594. [CrossRef]
4. Lv, Z.; Wang, N.; Lou, R.; Tian, Y.; Guizani, M. Towards carbon Neutrality: Prediction of wave energy based on improved GRU in Maritime transportation. *Appl. Energy* **2023**, *331*, 120394. [CrossRef]
5. Barstow, S.; Mørk, G.; Mollison, D.; Cruz, J. The wave energy resource. In *Ocean Wave Energy: Current Status and Future Prespectives*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 93–132.
6. Ali, M.; Prasad, R. Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renew. Sustain. Energy Rev.* **2019**, *104*, 281–295. [CrossRef]
7. Dixit, P.; Londhe, S. Prediction of extreme wave heights using neuro wavelet technique. *Appl. Ocean. Res*. **2016**, *58*, 241–252. [CrossRef]
8. Güner, H.A.A.; Yüksel, Y.; Çevik, E.O. Estimation of wave parameters based on nearshore wind–wave correlations. *Ocean Eng*. **2013**, *63*, 52–62. [CrossRef]
9. Chen, Y.; Liu, B.; Wang, T. Analysing and forecasting China containerized freight index with a hybrid decomposition–ensemble method based on EMD, grey wave and ARMA. *Grey Syst. Theory Appl*. **2021**, *11*, 358–371. [CrossRef]
10. Zhang, J.; Zhao, X.; Jin, S.; Greaves, D. Phase-resolved real-time ocean wave prediction with quantified uncertainty based on variational Bayesian machine learning. *Appl. Energy* **2022**, *324*, 119711. [CrossRef]
11. Li, M.; Liu, K. Probabilistic prediction of significant wave height using dynamic Bayesian network and information flow. *Water* **2020**, *12*, 2075. [CrossRef]
12. 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]
13. Mahjoobi, J.; Mosabbeb, E.A. Prediction of significant wave height using regressive support vector machines. *Ocean Eng*. **2009**, *36*, 339–347. [CrossRef]
14. Fan, G.F.; Peng, L.L.; Hong, W.C. Short-term load forecasting based on empirical wavelet transform and random forest. *Electr. Eng*. **2022**, *104*, 4433–4449. [CrossRef]
15. Deo, M.; Naidu, C.S. Real time wave forecasting using neural networks. *Ocean Eng*. **1998**, *26*, 191–203. [CrossRef]
16. 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]
17. Lou, R.; Wang, W.; Li, X.; Zheng, Y.; Lv, Z. Prediction of ocean wave height suitable for ship autopilot. *IEEE Trans. Intell. Transp. Syst*. **2021**, *23*, 25557–25566. [CrossRef]
18. Hao, P.; Li, S.; Gao, Y. Significant wave height prediction based on deep learning in the South China Sea. *Front. Mar. Sci*. **2023**, *9*, 1113788. [CrossRef]
19. Cannas, B.; Fanni, A.; See, L.; Sias, G. Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Phys. Chem. Earth Parts A/B/C* **2006**, *31*, 1164–1171. [CrossRef]

20. Deka, P.C.; Prahlada, R. Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time. *Ocean Eng*. **2012**, *43*, 32–42. [CrossRef]
21. Zhou, S.; Bethel, B.J.; Sun, W.; Zhao, Y.; Xie, W.; Dong, C. Improving significant wave height forecasts using a joint empirical mode decomposition–long short-term memory network. *J. Mar. Sci. Eng*. **2021**, *9*, 744. [CrossRef]
22. 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]
23. Lv, T.; Tao, A.; Zhang, Z.; Qin, S.; Wang, G. Significant wave height prediction based on the local-EMD-WaveNet model. *Ocean Eng*. **2023**, *287*, 115900. [CrossRef]
24. Song, T.; Wang, J.; Huo, J.; Wei, W.; Han, R.; Xu, D.; Meng, F. Prediction of significant wave height based on EEMD and deep learning. *Front. Mar. Sci*. **2023**, *10*, 1089357. [CrossRef]
25. Huang, W.; Dong, S. Improved short-term prediction of significant wave height by decomposing deterministic and stochastic components. *Renew. Energy* **2021**, *177*, 743–758. [CrossRef]
26. Wang, M.; Ying, F. A hybrid model for multistep-ahead significant wave height prediction using an innovative decomposition–reconstruction framework and E-GRU. *Appl. Ocean Res*. **2023**, *140*, 103752. [CrossRef]
27. Asma, S.; Sezer, A.; Ozdemir, O. MLR and ANN models of significant wave height on the west coast of India. *Comput. Geosci*. **2012**, *49*, 231–237. [CrossRef]
28. Han, Y.; Zhao, R.; Wu, F.; Yan, J.; Dong, C. A two channel optimized SWH deep learning forecast model coupled with dimensionality reduction scheme and attention mechanism. *Ocean Eng*. **2025**, *330*, 121217. [CrossRef]
29. Ding, T.; Wu, D.; Li, Y.; Shen, L.; Zhang, X. A hybrid CEEMDAN-VMD-TimesNet model for significant wave height prediction in the South Sea of China. *Front. Mar. Sci*. **2024**, *11*, 1375631. [CrossRef]
30. Ur Rehman, N.; Mandic, D.P. Filter bank property of multivariate empirical mode decomposition. *IEEE Trans. Signal Process*. **2011**, *59*, 2421–2426. [CrossRef]
31. Rathore, A.; Gupta, P.; Sharma, R.; Singh, R. Day ahead solar forecast using long short term memory network augmented with Fast Fourier transform-assisted decomposition technique. *Renew. Energy* **2025**, *247*, 123021. [CrossRef]
32. Zhang, W.; Wu, Z.; Zeng, X.; Zhu, C. An ensemble dynamic self-learning model for multiscale carbon price forecasting. *Energy* **2023**, *263*, 125820. [CrossRef]
33. Ma, C.; Li, Y.; Wang, X.; Cai, Z. Early fault diagnosis of rotating machinery based on composite zoom permutation entropy. *Reliab. Eng. Syst. Saf*. **2023**, *230*, 108967. [CrossRef]
34. Xiao, M.; Wang, Z.; Zhao, Y.; Geng, G.; Dustdar, S.; Donta, P.K.; Ji, G. A new fault feature extraction method of rolling bearings based on the improved self-selection ICEEMDAN-permutation entropy. *ISA Trans*. **2023**, *143*, 536–547. [CrossRef]
35. Van Den Oord, A.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. Wavenet: A generative model for raw audio. *arXiv* **2016**, arXiv:1609.03499.
36. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput*. **1997**, *9*, 1735–1780. [CrossRef]
37. Teng, C.C. Wave measurements from NDBC buoys and C-MAN stations. In *Proceedings of the OCEANS'02 MTS/IEEE*; IEEE: Piscataway, NJ, USA, 2002; Volume 1, pp. 517–524. [CrossRef]
38. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4.
39. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci*. **1998**, *454*, 903–995. [CrossRef]
40. Rehman, N.; Mandic, D.P. Multivariate empirical mode decomposition. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci*. **2010**, *466*, 1291–1302. [CrossRef]
41. Han, Y.; Tang, J.; Jia, H.; Dong, C.; Zhao, R. A significant wave height prediction method based on improved temporal convolutional network and attention mechanism. *Electronics* **2024**, *13*, 4879. [CrossRef]
42. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput*. **2019**, *31*, 1235–1270. [CrossRef]
43. Lozano, M.; Fiz, J.A.; Jané, R. Performance evaluation of the Hilbert–Huang transform for respiratory sound analysis and its application to continuous adventitious sound characterization. *Signal Process*. **2016**, *120*, 99–116. [CrossRef]
44. Zhang, H.; Zhou, Q.Q.; Chen, H.; Hu, X.Q.; Li, W.G.; Bai, Y.; Han, J.X.; Wang, Y.; Liang, Z.H.; Chen, D.; et al. The applied principles of EEG analysis methods in neuroscience and clinical neurology. *Mil. Med. Res*. **2023**, *10*, 67. [CrossRef]
45. Wang, J.; Sun, X.; Cheng, Q.; Cui, Q. An innovative random forest-based nonlinear ensemble paradigm of improved feature extraction and deep learning for carbon price forecasting. *Sci. Total Environ*. **2021**, *762*, 143099. [CrossRef]
46. Obara, Y.; Nakamura, R. Transfer learning of long short-term memory analysis in significant wave height prediction off the coast of western Tohoku, Japan. *Ocean Eng*. **2022**, *266*, 113048. [CrossRef]

47. Fu, Y.; Ying, F.; Huang, L.; Liu, Y. Multi-step-ahead significant wave height prediction using a hybrid model based on an innovative two-layer decomposition framework and LSTM. *Renew. Energy* **2023**, *203*, 455–472. [CrossRef]
48. Altunkaynak, A.; Çelik, A.; Mandev, M.B. Hourly significant wave height prediction via singular spectrum analysis and wavelet transform based models. *Ocean Eng*. **2023**, *281*, 114771. [CrossRef]
49. Guo, Y.; Si, J.; Wang, Y.; Hanif, F.; Li, S.; Wu, M.; Xu, M.; Mi, J. Ensemble-Empirical-Mode-Decomposition (EEMD) on SWH prediction: The effect of decomposed IMFs, continuous prediction duration, and data-driven models. *Ocean Eng*. **2025**, *324*, 120755. [CrossRef]
50. Mehrer, J.; Spoerer, C.J.; Kriegeskorte, N.; Kietzmann, T.C. Individual differences among deep neural network models. *Nat. Commun*. **2020**, *11*, 5725. [CrossRef]
51. Jentsch, C.; Lunsford, K.G. Asymptotically valid bootstrap inference for proxy SVARs. *J. Bus. Econ. Stat*. **2022**, *40*, 1876–1891. [CrossRef]
52. Costantini, M.; Kunst, R.M. On using predictive-ability tests in the selection of time-series prediction models: A Monte Carlo evaluation. *Int. J. Forecast*. **2021**, *37*, 445–460. [CrossRef]
53. Jamous, M.; Marsooli, R. A multidecadal assessment of mean and extreme wave climate observed at buoys off the US East, Gulf, and West Coasts. *J. Mar. Sci. Eng*. **2023**, *11*, 916. [CrossRef]
54. Dhoop, T.; Mason, T. Spatial characteristics and duration of extreme wave events around the English coastline. *J. Mar. Sci. Eng*. **2018**, *6*, 14. [CrossRef]
55. Chao, Y.Y.; Tolman, H.L. Performance of NCEP regional wave models in predicting peak sea states during the 2005 North Atlantic hurricane season. *Weather Forecast*. **2010**, *25*, 1543–1567. [CrossRef]
56. Bekiryazıcı, Ş.; Amarouche, K.; Ozcan, N.; Akpınar, A. An innovative deep learning-based approach for significant wave height forecasting. *Ocean Eng*. **2025**, *323*, 120623. [CrossRef]
57. Janssen, P.A. Progress in ocean wave forecasting. *J. Comput. Phys*. **2008**, *227*, 3572–3594. [CrossRef]
58. Chen, Y.; Zhang, D.; Li, X.; Peng, Y.; Wu, C.; Pu, H.; Zhou, D.; Cao, Y.; Zhang, J. Significant wave height prediction through artificial intelligent mode decomposition for wave energy management. *Energy AI* **2023**, *14*, 100257. [CrossRef]
59. Colosi, L.V.; Villas Boas, A.B.; Gille, S.T. The seasonal cycle of significant wave height in the ocean: Local versus remote forcing. *J. Geophys. Res. Oceans* **2021**, *126*, e2021JC017198. [CrossRef]
60. Echevarria, E.; Hemer, M.; Holbrook, N. Seasonal variability of the global spectral wind wave climate. *J. Geophys. Res. Oceans* **2019**, *124*, 2924–2939. [CrossRef]
61. Yang, X.; Li, J.; Jiang, X. Research on information leakage in time series prediction based on empirical mode decomposition. *Sci. Rep*. **2024**, *14*, 28362. [CrossRef]
62. Lin, B.; Meng, H. Short-Term Prediction and Temporal Causality Analysis of Total Nitrogen in Wastewater Treatment Plant Effluent Based on LT-PR-LSTM. *Water* **2026**, *18*, 1607. [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.