

## Article

# Neural-Network-Based Prediction of Wave Heights and Periods Southeast of Fujian Province

Ting Zhang \*, Min Liu and Zhenwei Su

School of Computing and Data Science, Fujian University of Technology, Fuzhou 350118, China; minliuedword@163.com (M.L.); zhenweisufjut@163.com (Z.S.)

\* Correspondence: zhangting2021@fjut.edu.com; Tel.: +86-19959198792

## Abstract

The prediction of ocean wave heights and periods is crucial for reducing maritime accidents. The application of neural networks to predictive modeling eliminates the need for complex parameter schemes that describe specific physical processes, thereby reducing computational costs. This study presents two single and three fusion models based on neural-network architectures: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), LSTM with an attention mechanism (LSTM+Attention), LSTM fused with a CNN (CNN+LSTM), and LSTM with Ensemble Empirical Mode Decomposition (EEMD+LSTM). The models were compared on multiple prediction lengths based on multivariate datasets from two oceanographic stations southeast of Fujian Province. All models performed well for prediction lengths of less than 24 h, with the coefficients of determination for wave height and period prediction higher than 0.7 and 0.6, respectively. For prediction lengths longer than 24 h, the predictive capability of EEMD+LSTM was superior to that of the other models. For a 48 h prediction length, the predictive results at each time step were compared with the true values across four typhoons. At the 24th and 48th time steps, phase lags were observed in the results from the CNN and LSTM. In contrast, EEMD+LSTM better captured the trends of the true values.

**Keywords:** wave height; wave period; CNN; LSTM; LSTM+Attention; CNN+LSTM; EEMD+LSTM

## 1. Introduction

With the development of deep learning in data mining, neural-network models have become widely used in the Earth sciences [1,2]. At the same time, increasing amounts of meteorological and hydrological data are driving the continuous improvement of deep learning techniques [3]. Compared to numerical models, deep learning offers high computational efficiency and accurate predictions, making it popular among researchers [4,5].

The long-term prediction of ocean wave heights and periods along China's coast is crucial for disaster prevention and coastal risk management. It is also important for coastal infrastructure construction, navigation safety, and marine engineering [6,7]. Currently, methods such as numerical wave models, wave spectral inversion, and deep learning have led to significant progress in short-term predictions of wave height and period (less than 24 h) [8,9]. Wave motions are affected by many different sea conditions and are described by nonlinear equations. Previous studies have used different parameter schemes to characterize the sea conditions, influencing the accuracy of the predictive equations. Additionally, solving nonlinear equations requires differential techniques, which produce



Academic Editor: A. Miguel P. Santos

Received: 12 May 2026

Revised: 16 June 2026

Accepted: 10 July 2026

Published: 14 July 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

errors during the computing process. For long-term predictions, these errors generally accumulate and limit the equations' predictive ability. The complex physical processes of ocean waves and their nonlinearity make accurate long-term prediction (beyond 48 h) challenging, particularly during tropical cyclones.

Traditional numerical models, such as WAVEWATCH III, FVCOM, and SWAN, were constructed based on physical equations [10–12]. Their predictive performance mainly depends on the reasonableness of approximate assumptions of the physical mechanism. Both the equations and the assumptions fail to capture the strong nonlinearity and randomness of ocean waves, limiting the models' predictive ability in extreme weather conditions. Thus, researchers have devoted themselves to optimizing parameter schemes to improve model predictive accuracy [13–16]. In addition, more mathematical equations increase the program's complexity and consume more computational resources. Wave spectral inversion predicts wave heights and periods using available wave data, thereby avoiding the need to model multiscale physical processes [17]. Prediction accuracy is primarily determined by data quality and the spectral analysis method. The spectral inversion also fails to overcome the nonlinearity of ocean wave propagation. In contrast to numerical wave models, this method performs poorly under extreme conditions because it cannot model the dynamics of ocean waves. Combining this method with other prediction models has been shown to improve predictive performance [18].

To compensate for shortcomings in handling the nonlinearity of ocean waves, machine learning techniques have been used to predict ocean waves over the last few decades. The mainstream methods include support vector machines (SVM), decision trees (DT), random forests (RF), and artificial neural networks (ANN) [19–22]. Among these, ANN-based methods have developed rapidly and have been modified into various versions suitable for different scenarios, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Instead of searching for parameterization schemes and solving nonlinear equations, the neural networks make complete predictions using historical data of ocean waves alone. This process avoids solving many nonlinear equations to shorten computational resources. The information contained in the data can be effectively extracted through training the neural networks. Well-trained models can effectively recognize the features and regularities of similar datasets, generally achieving highly accurate predictions. Recent studies have shown that deep neural networks perform well at predicting time series. In particular, the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) neural networks are effective at identifying long-term dependencies in time series, and are widely used for predicting significant wave height [23,24]. In a previous study, univariate time series were used to train a convolutional LSTM to predict regional wave height, and the results were compared with a SWAN model [25]. The predicted wave heights were essentially the same, and the predictive accuracy decreased with the increase in prediction lengths. The computational time of the LSTM was about 1% of that of SWAN. A previous study reported that among six machine learning methods, the physics-fusion neural network SWAN-LSTM performed the best for 1–3-day single-point prediction of significant wave heights [26]. Reanalysis data from the ERA5 dataset (European Centre for Medium-Range Weather Forecasts Reanalysis version 5) were used to train single neural models (i.e., RNN, LSTM, and GRU) to predict significant wave heights in the South China Sea [27]. The models' prediction accuracy was found to be sensitive to the input window and forecast horizon lengths of the time series, as well as the model's complexity. The proposed CLTS-Net fusion model outperformed single models on three-day predictions of significant wave heights [28]. The error in each time step was not discussed. A novel method was proposed to integrate wavelet decomposition and particle swarm optimization into machine learning [29]. This study revealed that their proposed hybrid model performs

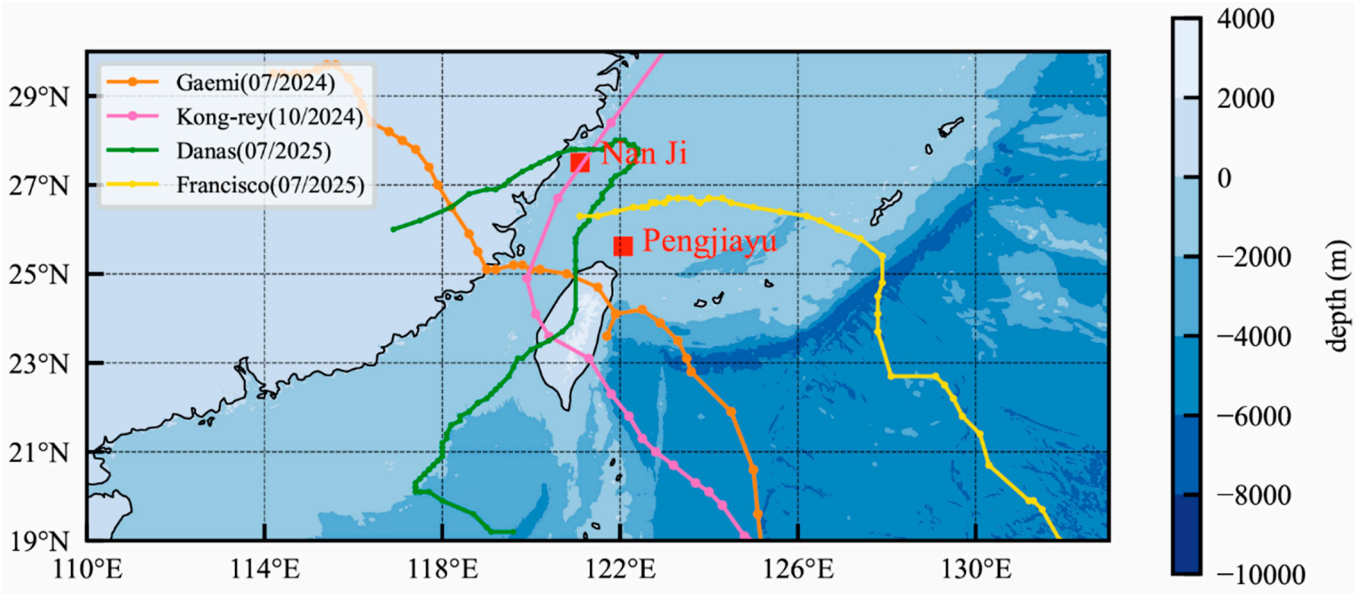
best among other models at predicting wave heights when the prediction time is 72 h. However, the study did not analyze model performance for longer prediction times. In another study, the Hodrick–Prescott filter was applied to univariate wave height data to achieve lag-free prediction [30], although multivariate wave data were not used. Another study used the XGBoost (i.e., eXtreme Gradient Boosting) algorithm to optimize the LSTM’s prediction of wave heights and periods [31], using wave data from Lake Erie to train a fusion model. The variation in significant wave heights and peak wave periods with time was not dramatic, and so this model may not be applicable for extreme ocean conditions. There is also a lack of discussion concerning the use of EEMD+LSTM on multivariate datasets and its predictive accuracy at each time step [8].

Although considerable progress has been made in adopting deep learning for wave prediction, some issues still require further investigation. For example, the predictive models are highly sensitive to the quality of the input data. Wave data with multiple variables and dramatic variations can pose difficulties for accurate model prediction. Both the model architectures and the parameter configurations also influence the predictive accuracy, and so it is important to compare the performances of various models across datasets. The long-term accurate prediction of typhoon waves is the ultimate objective for model optimization. Therefore, this study used two single and three fusion neural networks to perform multi-step predictions of wave heights and periods from two in situ multivariable datasets from ocean observation stations southeast of Fujian Province. The impact of output length on the predictive abilities of different models was analyzed. To assess typhoon wave prediction, errors at each time step were used to estimate the models’ performances in short-, medium-, and long-term predictions. The rest of this paper is organized as follows. The data and neural networks are described in Section 2. The training schemes are described, and the results are analyzed in Section 3. The Discussion and Conclusions are provided in Section 4.

## 2. Materials and Methods

### 2.1. Data and Study Area

In this study, we used data from Nan Ji and Pengjiayu in China, two observational stations that provide meteorological and oceanographic measurements (Figure 1). The Nan Ji station is located at (27.5° N, 121.1° E) and is run by the National Marine Data Center. It has obtained hourly near-real-time observations since 1945. The seven observational variables from Nan Ji station are shown in Table 1. To avoid the large amounts of missing data from before 31 December 2022 in Nan Ji station, this study used the data from 1 December 2022 to 31 August 2025 to train, validate, and test the neural-network models. There are only a few missing records in these 24096 continuous data points, which better reflect the characteristics of the time series. Pengjiayu station is located at (25.62° N, 122.069° E), and is run by the Central Weather Bureau. To avoid missing data, we selected 36,251 continuous data points from 30 May 2020 to 31 December 2024. Ten variables are measured at Pengjiayu station (Table 1). Wave height and wave period are the target variables. The mean values and standard deviations of wave heights and periods in Nan Ji station are 1.075 m, 0.528 m, 6.280 s, and 1.179 s, respectively. For Pengjiayu station, they are 1.367 m, 0.736 m, 5.452 s, and 0.8041 s. The fact that the standard deviations are higher for wave periods than for wave heights indicates that accurately predicting wave periods is more challenging.



**Figure 1.** The red squares denote the Nan Ji and Pengjiayu stations. The colored curves denote the best track of the four typhoons.

**Table 1.** Observational variables from Nan Ji and Pengjiayu stations.

| Station   | Observational Variables                                                                                                                                    |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Nan Ji    | Air temperature, air pressure, sea surface temperature, wind speed, wind direction, wave height, and wave period.                                          |
| Pengjiayu | Air temperature, air pressure, sea temperature, mean-wind speed, mean wind direction, gust wind speed, wave height, spectral peak period, and wave period. |

Nan Ji station recorded the typhoon waves caused by Kong-rey in October 2024, and Danas and Francisco in July 2025. Pengjiayu station captured the typhoon waves caused by Gaemi and Kong-rey during July and October 2024. Danas and Francisco also passed over Pengjiayu station in 2025; however, the station failed to record any wave data during these two typhoons. The four typhoon tracks are shown in Figure 1. A partial time series of wave heights and periods containing the typhoon waves is shown in Figure 2. The maximum wave heights observed by Nan Ji station exceeded 3 m during typhoon Francisco (Figure 2a). The wave heights observed by Pengjiayu station were higher, with the maximum being greater than 6 m (Figure 2c). During typhoon periods, the wave periods were also high, with maximum values generally exceeding 8 s (Figure 2b,d).

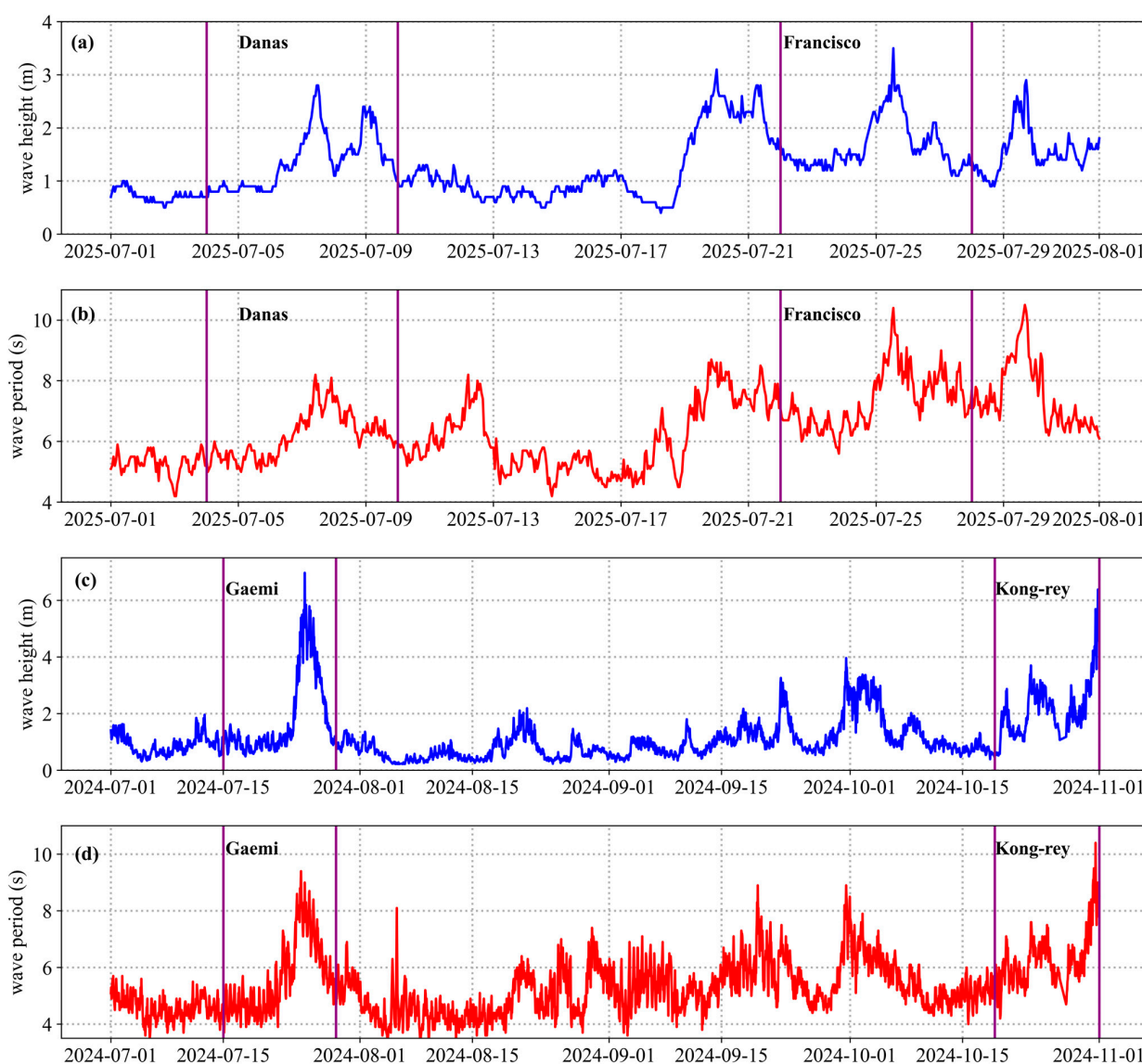
2.2. Data Correlation Analysis

There are no outliers in the two original datasets, as the observations were conducted under official quality control verification. Specifically, the data collection agency applied rigorous control and verification procedures before release. The two datasets thus contain no erroneous outliers caused by sensor malfunctions or data entry errors. The sampling interval for both datasets is 1 h. We only needed to handle missing values in the datasets, which was carried out using linear interpolation. Data from 17 February 2022 to 2 August 2022 from Pengjiayu station, which contained continuous missing values, was removed.

In statistical theory, the Pearson correlation coefficient (PCC) is widely used to quantitatively evaluate the linear correlation between two variables [32]. It is defined as follows:

$$\text{PCC} = \left[ \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y}) \right] / \left[ \sqrt{\sum_{k=1}^n (x_k - \bar{x})^2} \cdot \sqrt{\sum_{k=1}^n (y_k - \bar{y})^2} \right], \quad (1)$$

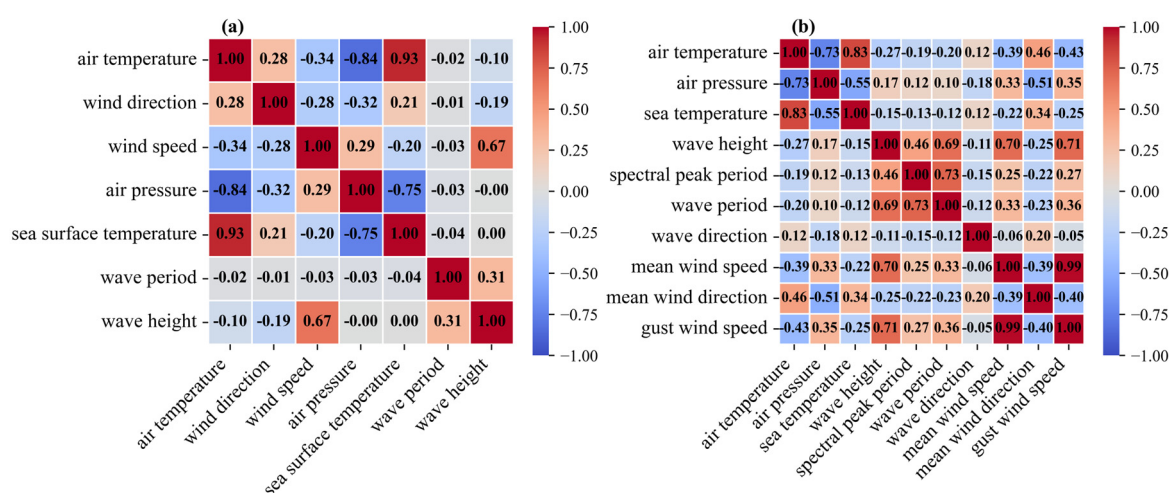
where  $n$  is the number of an observational dataset,  $x_k$  and  $y_k$  are the indicator values of the  $k$ th observation, and  $\bar{x}$  and  $\bar{y}$  are the mean values of variables  $x$  and  $y$ . The absolute value and sign of the PCC determine the linear correlation of  $x$  and  $y$  through two aspects. The positive PCC indicates that an increase in  $x$  results in an increase in  $y$ , or a decrease in  $x$  results in a decrease in  $y$ . The negative PCC indicates that an increase in  $x$  results in a decrease in  $y$ , or a decrease in  $x$  results in an increase in  $y$ . For two variables, the absolute value of PCC (i.e.,  $|\text{PCC}|$ ) greater than 0.7 (i.e.,  $|\text{PCC}| > 0.7$ ) implies a strong linear correlation,  $0.5 < |\text{PCC}| < 0.7$  implies a moderate linear correlation, and  $|\text{PCC}| < 0.5$  implies a weak linear correlation.



**Figure 2.** Partial time series of wave heights and periods observed from Nan Ji (a,b) and Pengjiaiyu (c,d). The purple vertical lines denote the time periods of the four typhoons.

This study focuses on predicting wave heights and periods, and so it is necessary to identify variables with strong or moderate linear correlation with these two target

variables. The Pearson correlation matrix was calculated for the two datasets (Figure 3). For the 7 variables at Nan Ji station (Figure 3a), wave heights show a moderate linear correlation with wind speed ( $PCC = 0.69$ ), and weak or no linear correlations with other variables ( $PCC < 0.5$ ). No variables present a strong or moderate linear correlation to wave periods ( $PCC < 0.5$ ). For the 10 variables from Pengjiayu station (Figure 3b), wave heights show strong or moderate linear correlations with three variables, i.e., gust wind speed ( $PCC = 0.71$ ), mean wind speed ( $PCC = 0.70$ ), and wave period ( $PCC = 0.69$ ). Wave periods show a strong correlation with the wave spectral peak period ( $PCC = 0.73$ ), and a moderate correlation to the wave height ( $PCC = 0.69$ ). Wave heights were moderately correlated with wave periods at Pengjiayu, but not at Nan Ji. This can be explained by the following. Commonly, in the offshore stations dominated by local wind and sea conditions, the wave height and wave period exhibit a strong linear correlation. In the nearshore stations, this correlation prominently weakens due to depth-induced wave breaking, bottom friction, and the mixing of multi-directional waves.



**Figure 3.** PCC matrices of the 7 variables from Nan Ji station (a) and the 10 parameters from Pengjiayu station (b).

### 2.3. Neural-Network Architectures

This study used five neural-network models to predict wave heights and periods. The first two models are single models: a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) neural network. The last three are fusion models: LSTM with an attention mechanism (LSTM+Attention), a convolutional LSTM network (CNN+LSTM), and an Ensemble Empirical Mode Decomposition LSTM (EEMD+LSTM). The detailed descriptions of the construction of the five models are presented below.

#### 2.3.1. CNN

CNNs are designed to handle data with grid structures, such as two-dimensional image data with a pixel grid and a one-dimensional speech signal. Their function is to automatically extract local features of the data. Other than the input layer and output layers, a CNN has three layers: convolution layers (Conv), pooling layers (Pool), and fully connected layers (Linear). Conv uses a convolution kernel that slides over the data to extract local features and increase the dimensions. Pool continues to extract the main features of the data by reducing the dimensionality. Linear layers integrate the extracted features and output the prediction results [33]. Figure 4a shows the design of the CNN model in our study. The “Conv1D” indicates that the convolution kernel is shifted along one direction (i.e., the time axis). The “input length” and “prediction length” denote the length of past time series data and that of future time series data, respectively (applicable

to Figure 4b–d). The “input dim” denotes the number of variables put into the model (applicable to Figure 4b–d). Notably, only one pooling layer was added after the last Conv1D to avoid losing continuous information during dimension reduction. A similar design was adopted in which the pool layer was removed from their model [28].



**Figure 4.** Structure of CNN (a), LSTM (b), LSTM+Attention (c), and CNN+LSTM (d) for one sample. “...” denotes the repetition of the structure of previous modules. The arrows denote the program flow.

To depict the process of the “Conv1D” module in Figure 4a in detail, the mathematical description for the first arrow in Figure 4a is represented as follows. Here, the following assumptions are made: the “input\_length” =  $T$ , “input\_dim” =  $n$ , “hidden\_dim\_1” =  $m$ , “kernel\_size\_1” =  $K$ . For a single sample, the input matrix  $X$  is given as

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1T} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2T} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nT} \end{pmatrix}, \quad X \in \mathbb{R}^{n \times T}. \quad (2)$$

The operation of convolution in the “Conv1D” module results in a  $Z$  matrix, i.e.,

$$Z = \begin{pmatrix} z_{11} & z_{12} & z_{13} & \dots & z_{1(T-k+1)} \\ z_{21} & z_{22} & z_{23} & \dots & z_{2(T-k+1)} \\ \dots & \dots & \dots & \dots & \dots \\ z_{m1} & z_{m2} & z_{m3} & \dots & z_{m(T-k+1)} \end{pmatrix}, \quad Z \in \mathbb{R}^{m \times (T-k+1)}, \quad (3)$$

where

$$z_{ij} = \sum_{c=1}^n \sum_{k=0}^{K-1} x_{c, j+k-P} \omega_{i,c,k} + b_i, \quad i = 1, \dots, m; \quad j = 1, \dots, T - k + 1; \quad (4)$$

in which  $P = [(K - 1)/2]$ ,  $\omega_{i,c,k}$  is the kernel weights, and  $b_i$  is the bias.

$Z$  is then standardized, scaled, and translated to obtain the normalized  $\tilde{Z}$ . The elements in  $\tilde{Z}$  are obtained as follows:

$$\hat{z}_{ij} = \frac{z_{ij} - \mu_i}{\sqrt{\sigma_i^2 + \varepsilon}}, \tilde{z}_{ij} = \gamma_i \cdot \hat{z}_{ij} + \beta_i, \quad i = 1, \dots, m; \quad j = 1, \dots, T - k + 1; \quad (5)$$

where  $\mu_i$  and  $\sigma_i$  are the mean and standard deviation for the  $i$ th dimension of  $Z$ ; and  $\gamma_i$  and  $\beta_i$  are the scaling and translation coefficients.

Finally, the ReLU activation function was applied to the normalized  $\tilde{Z}$  to generate output matrix  $A$ . The elements of  $A$  are given as

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1(T-k+1)} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2(T-k+1)} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{m(T-k+1)} \end{pmatrix}, \quad A \in \mathbb{R}^{m \times (T-k+1)}, \quad (6)$$

where

$$a_{ij} = \max(0, \tilde{z}_{ij}), \quad i = 1, \dots, m; \quad j = 1, \dots, T - k + 1. \quad (7)$$

### 2.3.2. LSTM, LSTM+Attention, and CNN+LSTM

LSTM can effectively capture long-term dependencies in sequences and is widely adopted in wave prediction. An LSTM layer has four key components: the input gate, the forget gate, the output gate, and the cell state. The three gates respectively determine how much new information the cell state should receive, how much old information it should forget, and how much information it should output [34]. Figure 4b shows the design of the LSTM model in this study. The red arrows indicate that an LSTM layer handles the input information at the current moment and the output at the previous moment (applicable to Figure 4c,d).

To depict the process of the “LSTM” module in Figure 4b in detail, the mathematical description for the first arrow in Figure 4b is represented as follows. Here, the following assumptions are made: the “input\_length” =  $T$ , “input\_dim” =  $n$ , “hidden\_dim\_1” =  $m$ . For a single sample, the input matrix  $X$  is given by Equation (2). The forget gate  $f_t$ , input gate  $i_t$ , and candidate cell state  $\tilde{c}_t$  are given as

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ \text{and } \tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \quad (8)$$

where  $x_t$  is the input at time  $t$ ;  $h_{t-1}$  is the hidden state at time  $t - 1$ ;  $W_f$ ,  $W_i$ , and  $W_c$  are weight matrices;  $b_f$ ,  $b_i$ , and  $b_c$  are the biases; and  $\sigma$  is the sigmoid function.

Then, the cell state is updated to  $c_t$ , the output gate  $o_t$  is computed, and the hidden state  $h_t$  is updated, i.e.,

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \quad o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad \text{and } h_t = o_t \odot \tanh(c_t), \quad (9)$$

where  $\odot$  is the Hadamard product,  $W_o$  is the weight matrix, and  $b_o$  is the bias. The matrix  $H$  consists of all hidden states  $h_1, h_2, \dots, h_T$  is fed to the next LSTM layer, i.e.,

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1T} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2T} \\ \dots & \dots & \dots & \dots & \dots \\ h_{m1} & h_{m2} & h_{m3} & \dots & h_{mT} \end{pmatrix}, \quad H \in \mathbb{R}^{m \times T}, \quad (10)$$

where  $h_t = (h_{1t}, h_{2t}, \dots, h_{mt}) \in \mathbb{R}^m$ .

Unlike RNN, LSTM is designed to resolve the vanishing and exploding gradient problems in long-term time series. For RNN, the hidden state is updated by the following formula:

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + b_t), \quad (11)$$

where  $W_{hh}$  and  $W_{xh}x_t$  are the weight matrices. The partial derivative of  $h_t$  to  $h_{t-1}$  is

$$\frac{\partial h_t}{\partial h_{t-1}} = \text{diag}(\tanh'(W_{hh}h_{t-1} + W_{xh}x_t + b_t)) \cdot W_{hh}^T, \quad (12)$$

Assuming the loss function is  $L$ , the gradient of  $L$  to  $h_t$  is

$$\frac{\partial L}{\partial h_T}, \frac{\partial L}{\partial h_{t-1}} = \frac{\partial L}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_{t-1}}, t = T, T-1, T-2, \dots, 1. \quad (13)$$

Using Equation (12),  $\frac{\partial L}{\partial h_1}$  could be written as

$$\frac{\partial L}{\partial h_1} = \frac{\partial L}{\partial h_T} \cdot \prod_{t=2}^T \frac{\partial h_t}{\partial h_{t-1}} = \frac{\partial L}{\partial h_T} \cdot \prod_{t=2}^T [\text{diag}(\tanh'(W_{hh}h_{t-1} + W_{xh}x_t + b_t)) \cdot W_{hh}^T]. \quad (14)$$

If  $\|W_{hh}^T\| < 1$ , the factor  $\prod_{t=2}^T [\text{diag}(\tanh'(W_{hh}h_{t-1} + W_{xh}x_t + b_t)) \cdot W_{hh}^T]$  in Equation (14) decreases with time  $t$ . This causes the gradient  $\frac{\partial L}{\partial h_1}$  to approach zero, leading to a vanishing gradient. In contrast, if  $\|W_{hh}^T\| > 1$ , the factor  $\prod_{t=2}^T [\text{diag}(\tanh'(W_{hh}h_{t-1} + W_{xh}x_t + b_t)) \cdot W_{hh}^T]$  in Equation (14) increases with time  $t$ . This causes the gradient  $\frac{\partial L}{\partial h_1}$  to keep increasing, leading to an exploding gradient.

LSTM computes the gradient of the loss function  $L$  with respect to the cell state  $c_t$  instead of a hidden state  $h_t$ , i.e.,

$$\frac{\partial L}{\partial c_T}, \frac{\partial L}{\partial c_{t-1}} = \frac{\partial L}{\partial c_t} \cdot \frac{\partial c_t}{\partial c_{t-1}}, t = T, T-1, T-2, \dots, 1. \quad (15)$$

Based on Equation (9), the partial derivative of  $c_t$  to  $c_{t-1}$  is given as

$$\frac{\partial c_t}{\partial c_{t-1}} = \text{diag}(f_t), \quad (16)$$

which does not include a weight matrix. Using Equation (16),  $\frac{\partial L}{\partial c_1}$  is written as

$$\frac{\partial L}{\partial c_1} = \frac{\partial L}{\partial c_T} \cdot \prod_{t=2}^T \frac{\partial c_t}{\partial c_{t-1}} = \frac{\partial L}{\partial c_T} \cdot \prod_{t=2}^T [\text{diag}(f_t)], \quad (17)$$

which solves the vanishing and exploding gradient problems in RNNs.

The inclusion of an attention mechanism in the LSTM layer strengthens the model's memory capacity. The attention mechanism helps each LSTM layer review all historical states of the input time sequence, providing a weight vector to the LSTM for dynamic selection of important information [35]. Figure 4c shows the design of the LSTM+Attention model in this study. An "Attention mechanism" was added after the last LSTM layer to integrate the entire history through a weight vector. The red arrows pointing from "Attention" to "LSTM" indicate that the last LSTM layer receives all historical states of the input sequence. The mathematical description of the "Attention" module in Figure 4c is given here. For a single sample, the attention score  $e$  is given as

$$e = (e_1, e_2, e_3, \dots, e_T), e_t = W_{att}h_t, t = 1, 2, \dots, T, \quad (18)$$

where  $W_{att}$  is the attention weight matrix and  $h_t$  is the input hidden state. Then, the Softmax function is used to normalize the  $e$  to obtain  $\alpha$ , i.e.,

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_T) = \left( \frac{e_1}{\sum_{t=1}^T e_t}, \frac{e_2}{\sum_{t=1}^T e_t}, \frac{e_3}{\sum_{t=1}^T e_t}, \dots, \frac{e_T}{\sum_{t=1}^T e_t} \right). \quad (19)$$

Finally, the weights in Equation (19) and hidden states are fused in a “Fusion Layer” module as follows:

$$\sum_{t=1}^T \alpha_t \cdot h_t. \quad (20)$$

The CNN+LSTM model was constructed to compare the predictive abilities of the single models CNN and LSTM, and the fusion model CNN+LSTM. Figure 4d shows the design of CNN+LSTM, where CNN captures local features first, and LSTM then learns long-term dependencies in the multivariable time series. Specifically, the output of the CNN in Equation (6) is fed to Equation (8) of the LSTM.

### 2.3.3. EEMD+LSTM

Along with the models mentioned in Section 2.3.2, we constructed a fusion model based on EEMD and LSTM to complete the decomposition and prediction of the time series, i.e., EEMD+LSTM (Figure 5). EEMD (Ensemble Empirical Mode Decomposition) is an improved algorithm for addressing the mode mixing issue of EMD (Empirical Mode Decomposition). It repeatedly adds white noise, leveraging its uniform distribution to facilitate automatic signal separation across different scales [8]. The steps of EEMD decomposition are as follows:

- Set the number of times to add white noise,  $M$ ;
- Complete  $M$  cycles. In the  $i$ th cycle, a white noise sequence  $n_i$  is generated and added to the original signal  $x(t)$ , obtaining a new signal  $x'_i(t)$ , i.e.,

$$x'_i(t) = x(t) + n_i(t); \quad (21)$$

Then, apply the standard EMD to  $x'_i(t)$  to obtain a set of Intrinsic Mode Function (IMF) components, i.e.,

$$x'_i(t) = \sum_{j=1}^J c_{ij}(t) + r_i(t), \quad (22)$$

where  $c_{ij}(t)$  is the  $j$ th IMF component,  $J$  is the number of IMF components, and  $r_i(t)$  is the residual component.

- Average the IMF and residual components to remove the noise, i.e.,

$$c_j(t) = \frac{1}{M} \sum_{i=1}^M c_{ij}(t), \quad r(t) = \frac{1}{M} \sum_{i=1}^M r_i(t), \quad (23)$$

where  $c_j(t)$  and  $r(t)$  are noise-reduced IMF components and the residual.

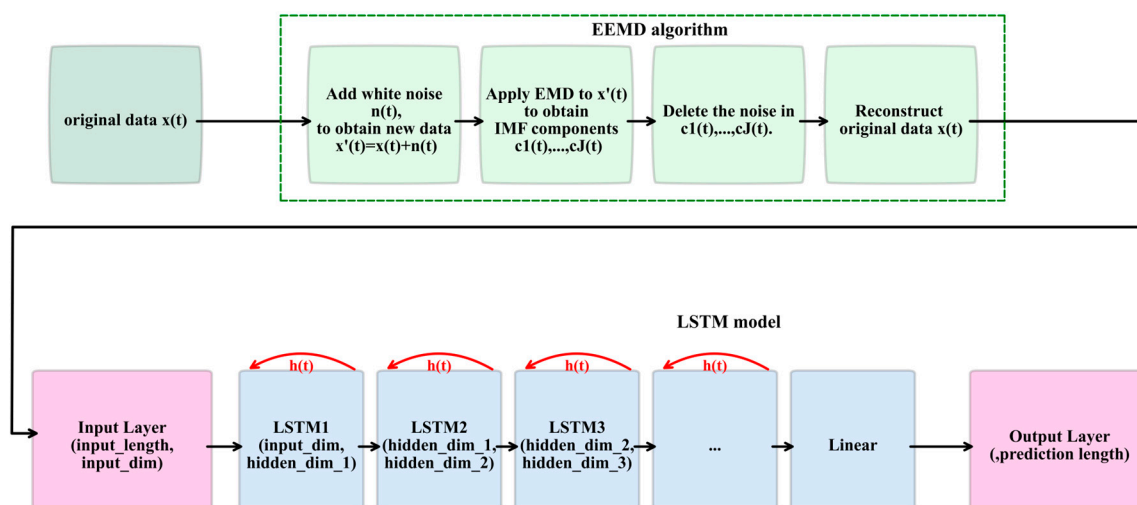
- Reconstruct the original signals, i.e.,

$$x(t) = \sum_{j=1}^J c_j(t) + r(t). \quad (24)$$

In this study, only the target variables (i.e., wave heights and wave periods) were analyzed using EEMD. As shown in Figure 5, the reconstructed signals and other variables were fed into the LSTM to complete the prediction.

To determine the appropriate parameters, an experimental approach was adopted. The white noise amplitude was varied from 0.1 to 0.5 in steps of 0.05, while the number of IMFs was varied from 2 to 20 in steps of 1. Based on model performance, the optimal values were determined to be 0.2 for the white noise amplitude and 15 for the number of IMFs. The ensemble size of the EEMD algorithm was set to 30 since the reconstruction error exhibited

diminishing returns, and the computation time continued to increase approximately linearly. The hyperparameters for LSTM in EEMD+LSMT were the same as those for the LSTM in Section 2.3.2, including hidden layers, neurons, initial learning rate, batch size, and epochs.



**Figure 5.** Structure of EEMD+LSTM. The arrows denote the program flow. The three dots denote the repetition of the previous module.

### 3. Training Design and Results Analysis

#### 3.1. Training Design

The training design consisted of two components. One was the design of input variables and target variables for the model. The corresponding five experiments (abbreviated as EX) are shown in Table 2. The first three, EX1, EX2, and EX3, were based on the observational data from Nan Ji station, and the remaining two, EX4 and EX5, were based on observations from Pengjiayu station. Notably, in EX2, EX4, and EX5, the input variables have a moderate or strong linear correlation with the output variables (Figure 3). The other aspect of the design was the architecture of the five models described in Sections 2.3.1–2.3.3. According to the experimental results of a previous study [27], an input length of 24 h, where  $h$  denotes hours below, performed better for multi-time-step predictions. In this study, we set the input lengths as 12 h, 24 h, 36 h, 48 h, 60 h, 72 h, 84 h, 96 h, 108 h, and 120 h to train the model. For an input length lower than 24 h, the predictive errors were dramatically greater when the output length exceeded 12 h. For the input lengths longer than 24 h, the results show that the longer the input length, the lower the predictive accuracy. Therefore, the optimal input length for all models was set to 24 h. Since the study aimed to investigate the models' performance on the short-term (1–24 h), the medium-term (24–48 h), and the long-term (48–120 h) predictions, nine output lengths (or prediction lengths) were set to train the models, i.e., 1 h, 6 h, 12 h, 24 h, 36 h, 48 h, 72 h, 96 h, and 120 h. The detailed information about the model architectures is shown in Table 3. Each dataset was chronologically split into training, validation, and test sets at 0.7:0.15:0.15. The data observed during typhoons Danas, Francisco, Gaemi, and Kong-rey were assigned to the test sets. The splitting process featured strict chronological order preservation and subset boundary isolation. Typhoon periods were at the end of the test set, where no information leaked into the training set through window sliding or normalization.

All models were trained on a computer equipped with an Intel Core Ultra 7 265 K CPU (20 cores, 20 threads, 3.9 GHz base, 5.5 GHz boost) and 64 GB DDR5 RAM. The software environment consisted of Python 3.11 and PyTorch 2025.3.3. The proposed model was trained for a maximum of 100 epochs with a batch size of 32, using early stopping with

patience = 15. Each epoch approximately required an average of 20 s to process; the total training time for one model was approximately 0.3 h (100 epochs), and the total inference time for the test set was 5 s.

The models were trained using the Adam optimizer, which combines adaptive learning rates with momentum to smooth the noisy gradients of the time series data. This enables fast and stable convergence in CNN and LSTM models without extensive tuning. The initial learning rate in the Adam optimizer was set to 0.001. If the validation loss did not improve for five consecutive epochs, the learning rate was multiplied by a factor of 0.2 to facilitate finer convergence. The validation set was used to select the optimal model, adjust the learning rate, and incorporate early stopping to avoid overfitting.

**Table 2.** Designs of input and output variables for the models.

| Name | Input Variables                                                               | Output Variables |
|------|-------------------------------------------------------------------------------|------------------|
| EX1  | All seven variables from Nan Ji                                               | Wave height      |
| EX2  | Wave height and wind speed from Nan Ji                                        | Wave height      |
| EX3  | All seven variables from Nan Ji                                               | Wave period      |
| EX4  | Wave height, wave period, mean-wind speed, and gust-wind speed from Pengjiayu | Wave height      |
| EX5  | Wave height, wave period, and spectral-peak period from Pengjiayu             | Wave period      |

**Table 3.** Model architectures and corresponding input and output lengths.

| Model                         | CNN        | LSTM       | LSTM+Attention | CNN+LSTM   |
|-------------------------------|------------|------------|----------------|------------|
| Input length (unit: hours)    | 24         | 24         | 24             | 24         |
| Kernel size                   | 3          | –          | –              | 3          |
| Layers                        | 2          | 3          | 3              | 2 and 3    |
| The hidden size of each layer | 64<br>128  | 128        | 128            | 64         |
|                               |            | 128        | 128            | 128        |
|                               |            | 128        | 128            | 128        |
|                               |            |            |                | 128        |
| Stride                        | 1          | 1          | 1              | 1          |
| Output length (unit: hour)    | 1/6/12     | 1/6/12     | 1/6/12         | 1/6/12     |
|                               | /24/36/48  | /24/36/48  | /24/36/48      | /24/36/48  |
|                               | /72/96/120 | /72/96/120 | /72/96/120     | /72/96/120 |
| Batch size                    | 32         | 32         | 32             | 32         |
| Epoch                         | 100        | 100        | 100            | 100        |

### 3.2. Metrics

Here, the model's performance on test sets was evaluated using four statistical metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of determination ( $R^2$ ). The four metrics were calculated through the following formulas:

$$\text{RMSE} = \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2 / n}, \quad (25)$$

$$\text{MAE} = \sum_{i=1}^n |\hat{y}_i - y_i| / n, \quad (26)$$

$$\text{MAPE} = \left( \sum_{i=1}^n |(\hat{y}_i - y_i) / y_i| \right) / n \times 100\%, \quad (27)$$

$$R^2 = 1 - \sum_{i=1}^n (y_i - \hat{y}_i)^2 / \sum_{i=1}^n (y_i - \bar{y})^2, \quad (28)$$

where  $n$  is the number of samples in the test set, and  $\hat{y}_i$ ,  $y_i$ , and  $\bar{y}$  are the predicted value, true value, and mean value of indicator  $y$ . Lower RMSE, MAE, and MAPE values indicate less error between the predicted and actual values. In contrast, a higher  $R^2$  indicates a more accurate prediction.

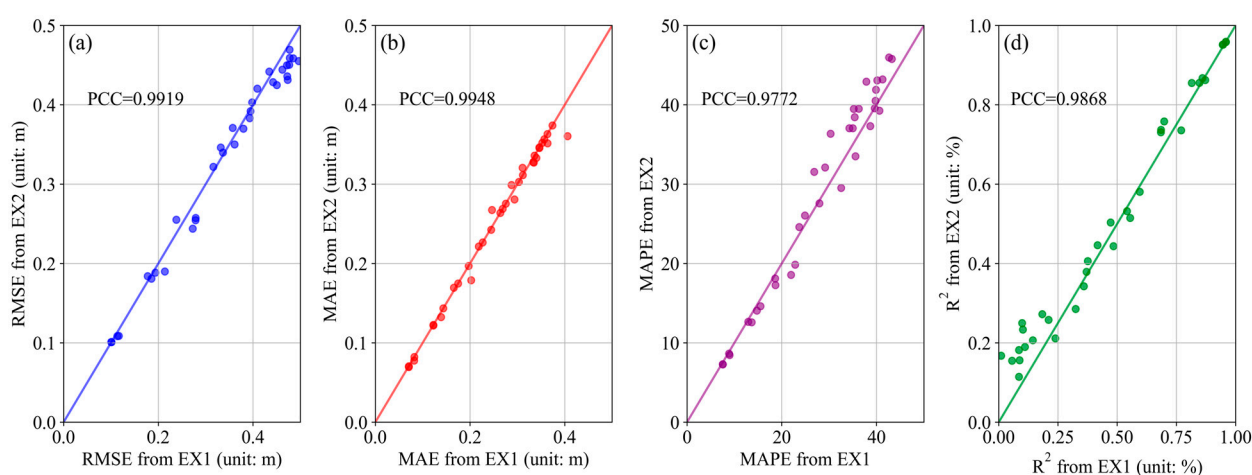
### 3.3. Results Analysis

#### 3.3.1. Performance of CNN, LSTM, LSTM+Attention, and CNN+LSTM at Nan Ji Station

EX1 and EX2 were designed to evaluate the impact of input variables and output length on wave height predictions. At Nan Ji station, only the wind speed shows a linear correlation with the wave height. To study whether the input of variables with weak or nonlinear correlations with wave height could improve the models' prediction accuracy, EX2 was designed to compare the results of EX1.

The metrics for the test sets of EX1 and EX2 are shown in Table 4, with the optimal values highlighted using colored backgrounds. Here, the optimal results for wave height predictions satisfy the criteria  $\text{RMSE} < 0.4$ ,  $\text{MAE} < 0.4$ ,  $\text{MAPE} < 30\%$ , and  $R^2 > 0.7$ .

The comparison of the first and second rows in each cell of Table 4 indicates little difference between the metrics of EX1 and EX2. These small differences were plotted in scatter plots, as shown in Figure 6. The calculated PCC of RMSE, MAE, MAPE, and  $R^2$  between EX1 and EX2 are higher than 0.95 (Figure 6), signifying that variables that have weak linear correlations with wave heights in the models failed to improve their accuracy when predicting wave heights. The four models performed well for 1, 6, and 12 h predictions (Table 3). Models performed poorly for prediction lengths greater than 12 h, with values of RMSE, MAE, and MAPE exceeding 0.3 m, 0.2 m, 20%, respectively, and with an  $R^2$  lower than 0.7. In particular, for a prediction length of 72 h, the  $R^2$  for all models is below 0.3, indicating that the models lost predictive ability. In fact, the models performed even worse on 96 h and 120 h predictions. The three LSTM-based models performed consistently on wave height predictions, with neither LSTM+Attention nor CNN+LSTM exhibiting a clearly superior predictive ability to LSTM.



**Figure 6.** Scatters of EX1 and EX2 at four metrics: (a) RMSE, (b) MAE, (c) MAPE, and (d)  $R^2$ .

EX3 was designed to analyze the impact of output length on the accuracy of wave period forecasts across various models. The test set metrics are shown in Table 5, with the optimal metrics highlighted. Since the standard deviation of wave periods is 1.179 s, higher than that of wave height (0.528 m), the optimal results for wave period predictions adhere to the rule

that  $RMSE < 0.7$ ,  $MAE < 0.5$ ,  $MAPE < 10\%$ , and  $R^2 > 0.6$ . Table 5 shows that the CNN's predictive ability rapidly declines when the prediction length exceeds 24 h. For prediction lengths of 1 h, 6 h, and 12 h, the LSTM and LSTM+Attention models performed similarly and better than the CNN. The attention mechanism did not improve LSTM's predictive ability, possibly because the attention mechanism led the model to capture irrelevant information from the multivariable time series. Table 5 also shows that CNN+LSTM performs worse than LSTM, particularly for prediction lengths higher than 24 h. The contents of Table 5 are also illustrated in Figure A1 of Appendix A.

**Table 4.** Prediction results of EX1 (first row) and EX2 (second row) on the test set. The optimal results are highlighted in the colorful background.

| Model              | Metrics        | Prediction Length (Unit: Hours) |          |          |          |          |          |
|--------------------|----------------|---------------------------------|----------|----------|----------|----------|----------|
|                    |                | 1                               | 6        | 12       | 24       | 48       | 72       |
| CNN                | RMSE           | 0.1131 m                        | 0.1938 m | 0.2786 m | 0.3364 m | 0.4422 m | 0.4720 m |
|                    |                | 0.1086 m                        | 0.1886 m | 0.2546 m | 0.3398 m | 0.4286 m | 0.4360 m |
|                    | MAE            | 0.0818 m                        | 0.1386 m | 0.2022 m | 0.2444 m | 0.3106 m | 0.3346 m |
|                    |                | 0.0776 m                        | 0.1325 m | 0.1789 m | 0.2424 m | 0.3206 m | 0.3272 m |
|                    | MAPE           | 8.97%                           | 15.51%   | 22.82%   | 27.95%   | 35.17%   | 36.25%   |
|                    |                | 8.46%                           | 14.61%   | 19.86%   | 27.58%   | 39.49%   | 39.51%   |
|                    | R <sup>2</sup> | 0.9479                          | 0.8473   | 0.6849   | 0.5414   | 0.2100   | 0.1017   |
|                    |                | 0.9520                          | 0.8554   | 0.7368   | 0.5321   | 0.2578   | 0.2335   |
| LSTM               | RMSE           | 0.1002 m                        | 0.1775 m | 0.2378 m | 0.3160 m | 0.3947 m | 0.4500 m |
|                    |                | 0.1010 m                        | 0.1840 m | 0.2552 m | 0.3218 m | 0.3920 m | 0.4249 m |
|                    | MAE            | 0.0705 m                        | 0.1226 m | 0.1658 m | 0.2182 m | 0.2875 m | 0.3333 m |
|                    |                | 0.0695 m                        | 0.1219 m | 0.1695 m | 0.2214 m | 0.2989 m | 0.3278 m |
|                    | MAPE           | 7.55%                           | 13.63%   | 18.59%   | 23.67%   | 34.25%   | 39.79%   |
|                    |                | 7.29%                           | 12.59%   | 18.11%   | 24.59%   | 37.06%   | 40.51%   |
|                    | R <sup>2</sup> | 0.9592                          | 0.8720   | 0.7704   | 0.5954   | 0.3704   | 0.1835   |
|                    |                | 0.9585                          | 0.8624   | 0.7356   | 0.5805   | 0.3792   | 0.2721   |
| LSTM+<br>Attention | RMSE           | 0.1012 m                        | 0.1855 m | 0.2726 m | 0.3316 m | 0.4089 m | 0.4729 m |
|                    |                | 0.1009 m                        | 0.1807 m | 0.2440 m | 0.3461 m | 0.4205 m | 0.4314 m |
|                    | MAE            | 0.0706 m                        | 0.1226 m | 0.1748 m | 0.2265 m | 0.3029 m | 0.3360 m |
|                    |                | 0.0697 m                        | 0.1216 m | 0.1631 m | 0.2356 m | 0.3144 m | 0.3184 m |
|                    | MAPE           | 7.51%                           | 12.90%   | 18.67%   | 24.84%   | 35.37%   | 38.69%   |
|                    |                | 7.29%                           | 12.66%   | 17.28%   | 26.04%   | 38.44%   | 37.13%   |
| CNN+<br>LSTM       | RMSE           | 0.1166 m                        | 0.2135 m | 0.2787 m | 0.3609 m | 0.4340 m | 0.4764 m |
|                    |                | 0.1088 m                        | 0.1889 m | 0.2575 m | 0.3501 m | 0.4418 m | 0.4504 m |
|                    | MAE            | 0.0822 m                        | 0.1435 m | 0.1968 m | 0.2688 m | 0.3115 m | 0.3565 m |
|                    |                | 0.0781 m                        | 0.1297 m | 0.1730 m | 0.2530 m | 0.3170 m | 0.3352 m |
|                    | MAPE           | 8.86%                           | 14.74%   | 21.95%   | 32.54%   | 35.00%   | 40.59%   |
|                    |                | 8.65%                           | 14.06%   | 18.56%   | 29.52%   | 37.05%   | 39.26%   |
|                    | R <sup>2</sup> | 0.9447                          | 0.8147   | 0.6846   | 0.4722   | 0.2388   | 0.0849   |
|                    |                | 0.9518                          | 0.8550   | 0.7308   | 0.5034   | 0.2112   | 0.1821   |

**Table 5.** Prediction results of EX3 on the test set. The optimal results are highlighted by the colorful background.

| Model              | Metrics        | Prediction Length (Unit: Hours) |          |          |          |          |          |
|--------------------|----------------|---------------------------------|----------|----------|----------|----------|----------|
|                    |                | 1                               | 6        | 12       | 24       | 48       | 72       |
| CNN                | RMSE           | 0.3149 s                        | 0.5105 s | 0.6312 s | 0.7407 s | 0.9805 s | 1.1089 s |
|                    | MAE            | 0.2366 s                        | 0.3673 s | 0.4602 s | 0.5353 s | 0.7367 s | 0.8798 s |
|                    | MAPE           | 4.13%                           | 6.04%    | 8.15%    | 9.51%    | 13.37%   | 16.35%   |
|                    | R <sup>2</sup> | 0.9141                          | 0.7743   | 0.6545   | 0.5227   | 0.1647   | −0.0669  |
| LSTM               | RMSE           | 0.2894 s                        | 0.4668 s | 0.5691 s | 0.7131 s | 0.8458   | 1.0392 s |
|                    | MAE            | 0.2143 s                        | 0.3384 s | 0.4025 s | 0.5060 s | 0.5990 s | 0.7460 s |
|                    | MAPE           | 3.75%                           | 6.05%    | 7.03%    | 9.01%    | 10.67%   | 13.50%   |
|                    | R <sup>2</sup> | 0.9274                          | 0.8113   | 0.7191   | 0.5576   | 0.3784   | 0.0631   |
| LSTM+<br>Attention | RMSE           | 0.2900 s                        | 0.4415 s | 0.5828 s | 0.7702 s | 0.9994 s | 1.0130 s |
|                    | MAE            | 0.2160 s                        | 0.3171 s | 0.4134 s | 0.5439 s | 0.7113 s | 0.7543 s |
|                    | MAPE           | 3.77%                           | 5.59%    | 7.22%    | 9.51%    | 12.88%   | 13.83%   |
|                    | R <sup>2</sup> | 0.9271                          | 0.8312   | 0.7055   | 0.4840   | 0.1322   | 0.1097   |
| CNN+<br>LSTM       | RMSE           | 0.3187 s                        | 0.5489 s | 0.6398 s | 0.8511 s | 0.8816 s | 1.0794 s |
|                    | MAE            | 0.2417 s                        | 0.4015 s | 0.4675 s | 0.6253 s | 0.6356 s | 0.8387 s |
|                    | MAPE           | 4.23%                           | 7.12%    | 8.22%    | 11.22%   | 11.25%   | 15.51%   |
|                    | R <sup>2</sup> | 0.9120                          | 0.7391   | 0.6450   | 0.3699   | 0.3246   | −0.0109  |

### 3.3.2. Performance of CNN, LSTM, LSTM+Attention, and CNN+LSTM at Pengjiayu Station

The results of EX1 and EX2 in Section 3.3.1 demonstrate that inputting variables with moderate or strong linear correlations with the output indicator into the models can effectively present the models' predictive abilities. In EX4 and EX5, only the variables with moderate and strong linear correlations with wave heights and periods were included in the models. The metrics evaluating the predictive ability at Pengjiayu station are shown in Tables 6 and 7. The optimal metrics for wave height predictions at Pengjiayu station are  $RASE < 0.4$ ,  $MAE < 0.4$ ,  $MAPE < 30\%$ , and  $R^2 > 0.7$ . The optimal results are highlighted by the colorful background. In EX4, the four models performed well for wave height predictions at 1 h, 6 h, and 12 h (Table 6). Although LSTM and LSTM+Attention performed better than both CNN and CNN+LSTM at 24 h prediction, their forecast accuracy continued to weaken as the prediction length increased.

The model performances of EX5 are shown in Table 7, with the optimal results highlighted. In this case, the optimal results for wave period predictions are  $RMSE < 0.7$ ,  $MAE < 0.5$ ,  $MAPE < 10\%$ , and  $R^2 > 0.6$ . The four models performed well for prediction lengths lower than 12 h. Specifically, both LSTM and LSTM+Attention present better predictive abilities than CNN and CNN+LSTM models. With increasing prediction lengths, the predictive ability of all modes decreased, particularly for CNN and CNN+LSTM.

The forecast performance of CNN, LSTM, LSTM+Attention, and CNN+LSTM on the Pengjiayu station dataset was nearly identical to that of the Nan Ji station dataset. The four models performed well for prediction lengths lower than 24 h. The MASE, MAE, and MAPE for wave height predictions were lower than 0.4 m, 0.4 m, and 30%, and  $R^2$  was higher than 0.7. The four metrics for wave period predictions were lower than 0.7 s, 0.5 s, and 10%, and  $R^2$  was higher than 0.6. Their predictive capabilities deteriorated when the prediction length exceeded 24 h. With the increase in prediction length, LSTM and LSTM+Attention performed better than CNN and CNN+LSTM. It indicates that, in our model architectures, CNNs did not exhibit an advantage over LSTMs for predicting single-point time series.

**Table 6.** Prediction results of EX4 on the test set. The optimal results are highlighted by the colorful background.

| Model              | Metrics        | Prediction Length (Unit: Hours) |          |          |          |          |          |
|--------------------|----------------|---------------------------------|----------|----------|----------|----------|----------|
|                    |                | 1                               | 6        | 12       | 24       | 48       | 72       |
| CNN                | RMSE           | 0.2028 m                        | 0.2829 m | 0.3414 m | 0.4397 m | 0.5378 m | 0.5875 m |
|                    | MAE            | 0.1368 m                        | 0.1912 m | 0.2305 m | 0.3123 m | 0.3765 m | 0.4096 m |
|                    | MAPE           | 12.09%                          | 16.99%   | 20.13%   | 31.30%   | 38.09%   | 40.30%   |
|                    | R <sup>2</sup> | 0.9285                          | 0.8610   | 0.7976   | 0.6650   | 0.5007   | 0.4065   |
| LSTM               | RMSE           | 0.1900 m                        | 0.2673 m | 0.3190 m | 0.4130 m | 0.5263 m | 0.5846 m |
|                    | MAE            | 0.1262 m                        | 0.1784 m | 0.2260 m | 0.2819 m | 0.3688 m | 0.3949 m |
|                    | MAPE           | 10.42%                          | 15.76%   | 22.20%   | 26.87%   | 36.87%   | 38.30%   |
|                    | R <sup>2</sup> | 0.9373                          | 0.8758   | 0.8233   | 0.7045   | 0.5219   | 0.4123   |
| LSTM+<br>Attention | RMSE           | 0.1906 m                        | 0.2746 m | 0.3211 m | 0.4023 m | 0.5465 m | 0.5903 m |
|                    | MAE            | 0.1264 m                        | 0.1874 m | 0.2215 m | 0.2713 m | 0.3706 m | 0.4073 m |
|                    | MAPE           | 10.47%                          | 17.20%   | 20.23%   | 24.42%   | 35.36%   | 40.59%   |
|                    | R <sup>2</sup> | 0.9369                          | 0.8689   | 0.8210   | 0.7197   | 0.4844   | 0.4007   |
| CNN+<br>LSTM       | RMSE           | 0.2044 m                        | 0.2926 m | 0.3501 m | 0.4306 m | 0.5489 m | 0.6016 m |
|                    | MAE            | 0.1356 m                        | 0.1954 m | 0.2330 m | 0.3059 m | 0.3780 m | 0.4222 m |
|                    | MAPE           | 11.53%                          | 17.10%   | 20.35%   | 31.19%   | 35.96%   | 43.24%   |
|                    | R <sup>2</sup> | 0.9273                          | 0.8513   | 0.7872   | 0.6788   | 0.4799   | 0.3777   |

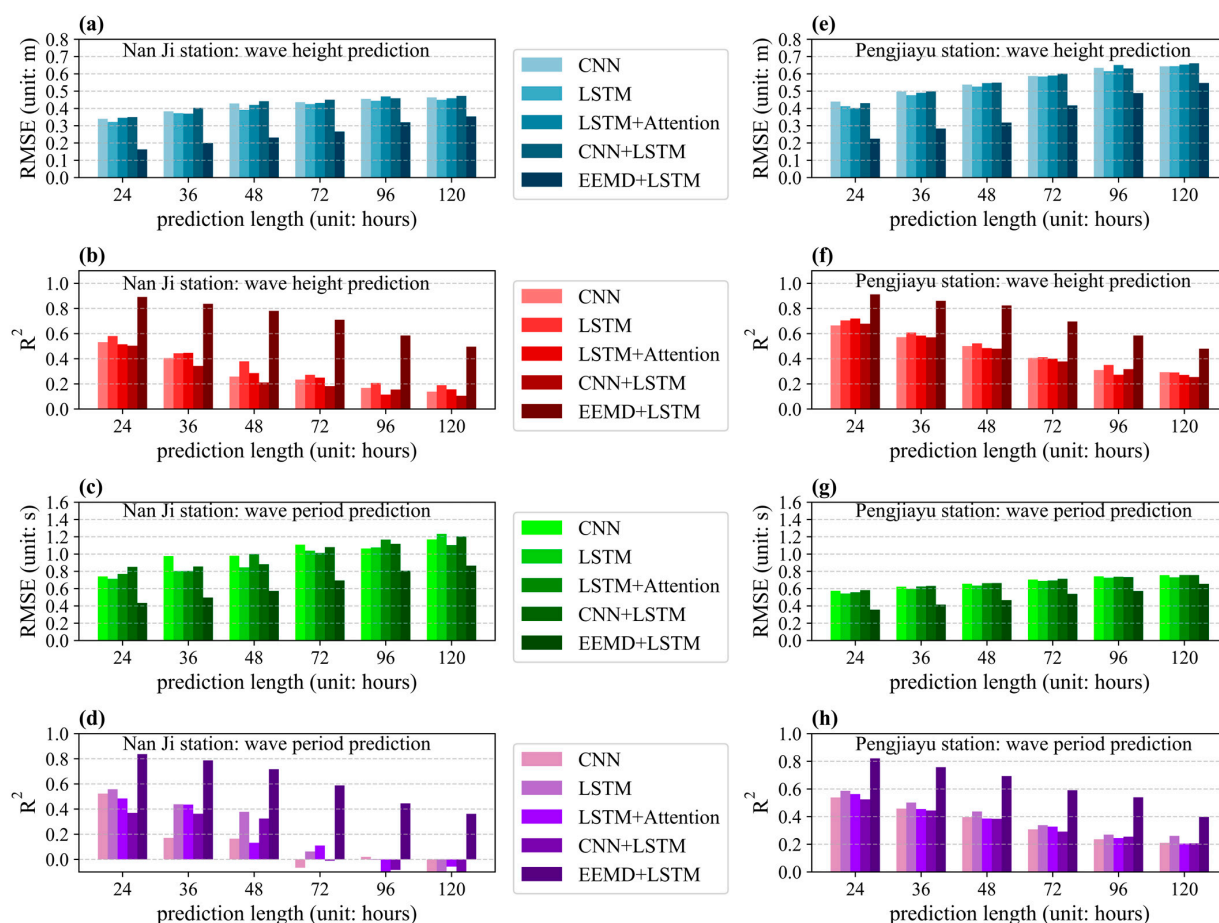
**Table 7.** Prediction results of EX5 on the test set. The optimal results are highlighted by the colorful background.

| Model              | Metrics        | Prediction Length (Unit: Hours) |          |          |          |          |          |
|--------------------|----------------|---------------------------------|----------|----------|----------|----------|----------|
|                    |                | 1                               | 6        | 12       | 24       | 48       | 72       |
| CNN                | RMSE           | 0.3409 s                        | 0.4646 s | 0.5029 s | 0.5740 s | 0.6572 s | 0.7048 s |
|                    | MAE            | 0.2554 s                        | 0.3479 s | 0.3735 s | 0.4258 s | 0.4887 s | 0.5217 s |
|                    | MAPE           | 4.75%                           | 6.51%    | 6.96%    | 7.9%     | 9.13%    | 9.66%    |
|                    | R <sup>2</sup> | 0.8366                          | 0.6968   | 0.6451   | 0.5384   | 0.3965   | 0.3079   |
| LSTM               | RMSE           | 0.3252 s                        | 0.4229 s | 0.4685 s | 0.5432 s | 0.6347 s | 0.6893 s |
|                    | MAE            | 0.2444 s                        | 0.3169 s | 0.3498 s | 0.4037 s | 0.4745 s | 0.5143 s |
|                    | MAPE           | 4.57%                           | 5.95%    | 6.54%    | 7.58%    | 8.93%    | 9.65%    |
|                    | R <sup>2</sup> | 0.8514                          | 0.7488   | 0.6920   | 0.5865   | 0.4372   | 0.3380   |
| LSTM+<br>Attention | RMSE           | 0.3271 s                        | 0.4244 s | 0.4776 s | 0.5582 s | 0.6629 s | 0.6947 s |
|                    | MAE            | 0.2446 s                        | 0.3186 s | 0.3598 s | 0.4076 s | 0.4855 s | 0.5264 s |
|                    | MAPE           | 4.56%                           | 5.98%    | 6.77%    | 7.56%    | 9.08%    | 9.90%    |
|                    | R <sup>2</sup> | 0.8496                          | 0.7470   | 0.6798   | 0.5635   | 0.3860   | 0.3275   |
| CNN+LSTM           | RMSE           | 0.3398 s                        | 0.4660 s | 0.5168 s | 0.5824 s | 0.6640 s | 0.7131 s |
|                    | MAE            | 0.2547 s                        | 0.3487 s | 0.3900 s | 0.4338 s | 0.5000 s | 0.5373 s |
|                    | MAPE           | 4.75%                           | 6.50%    | 7.36%    | 8.11%    | 9.36%    | 10.09%   |
|                    | R <sup>2</sup> | 0.8377                          | 0.6950   | 0.6252   | 0.5247   | 0.3840   | 0.2914   |

### 3.3.3. Performance of EEMD+LSTM at Nan Ji and Pengjiayu Stations

EX1, EX3, EX4, and EX5 were conducted with the EEMD+LSTM model. EX1 and EX4 were designed to implement the wave height predictions based on data from the Nan Ji and Pengjiayu stations. The metrics on test sets of EX1 and EX4 are shown in Table 8. Optimal metrics are defined as RMSE < 0.4 m, MAE < 0.4 m, MAPE < 30%, and R<sup>2</sup> > 0.7; these metrics are highlighted in Table 8. For prediction lengths of 24 h, 36 h, 48 h, and 72 h, the EEMD+LSTM performed well, with an R<sup>2</sup> greater than 0.7 (Table 8). It can also be seen that EEMD+LSTM performed better on the dataset from Nan Ji station than that

from Pengjiayu station for prediction lengths of 72 h, 96 h, and 120 h. A comparison of RMSE and  $R^2$  for CNN, LSTM, LSTM+Attention, and EEMD+LSTM in EX1 and EX4 is shown in Figure 7a,b,e,f. It shows that, for medium- and long-term wave height predictions, EEMD+LSTM significantly outperformed the other four models.



**Figure 7.** Comparison of the prediction accuracy of five models. Values of RMSE and  $R^2$  for Nan Ji station (a–d) and Pengjiayu station (e–h).

For both LSTM and EEMD+LSTM, experiments were repeated using 20 different random seeds. The results of paired  $t$ -tests indicated that EEMD achieved superior metrics compared to LSTM ( $p < 0.05$ ). The box plots of the four metrics for a prediction length of 24 h are shown in Figure A3.

The EX3 and EX5 were designed to implement wave period predictions based on data from the Nan Ji and Pengjiayu stations. The metrics on test sets of EX3 and EX5 are shown in Table 9. The optimal definition for wave prediction is that  $RMSE < 0.7$  s,  $MAE < 0.5$  s,  $MAPE < 10\%$ , and  $R^2 > 0.6$ . The optimal metrics are highlighted in Table 9. For prediction lengths lower than 96 h, the EEMD+LSTM performed well on both datasets. Figure 7c,d,g,h shows the RMSE and  $R^2$  of five models in EX3 and EX5. The EEMD+LSTM performs the best among the five models in wave period prediction. The RMSE from EEMD+LSTM is significantly lower than that of the other four models (Figure 7c,g). For prediction lengths of 72 h, 96 h, and 120 h at Nan Ji station, the  $R^2$  from EEMD+LSTM is higher than 0.36, whereas that from CNN, LSTM, LSTM+Attention, and CNN+LSTM is much lower, even reaching negative values. At Pengjiayu station, the EEMD+LSTM for wave period prediction produced substantially better wave periods than other models (Figure 7h). The learning and validation curves for EEMD+LSTM at 12 h, 24 h, 48 h, and 72 h are shown in Figure A2.

**Table 8.** EEMD+LSTM performances in EX1 and EX4 on test sets. The optimal results are highlighted by the colorful background.

| Station   | Prediction Length (Unit: Hours) | Metrics  |          |        |                |
|-----------|---------------------------------|----------|----------|--------|----------------|
|           |                                 | RMSE     | MAE      | MAPE   | R <sup>2</sup> |
| Nan Ji    | 24                              | 0.1628 m | 0.1101 m | 11.39% | 0.8915         |
|           | 36                              | 0.1997 m | 0.1343 m | 14.43% | 0.8366         |
|           | 48                              | 0.2313 m | 0.1578 m | 17.34% | 0.7807         |
|           | 72                              | 0.2664 m | 0.1871 m | 21.12% | 0.7100         |
|           | 96                              | 0.3198 m | 0.2394 m | 29.15% | 0.5843         |
|           | 120                             | 0.3532 m | 0.2483 m | 28.15% | 0.4958         |
| Pengjiayu | 24                              | 0.2247 m | 0.1529 m | 12.89% | 0.9120         |
|           | 36                              | 0.2832 m | 0.1916 m | 16.61% | 0.8602         |
|           | 48                              | 0.3183 m | 0.2143 m | 17.95% | 0.8235         |
|           | 72                              | 0.4176 m | 0.2855 m | 24.92% | 0.6963         |
|           | 96                              | 0.4887 m | 0.3325 m | 29.86% | 0.5843         |
|           | 120                             | 0.5468 m | 0.3691 m | 32.34% | 0.4798         |

**Table 9.** EEMD+LSTM performances in EX3 and EX5 on test sets. The optimal results are highlighted by the colorful background.

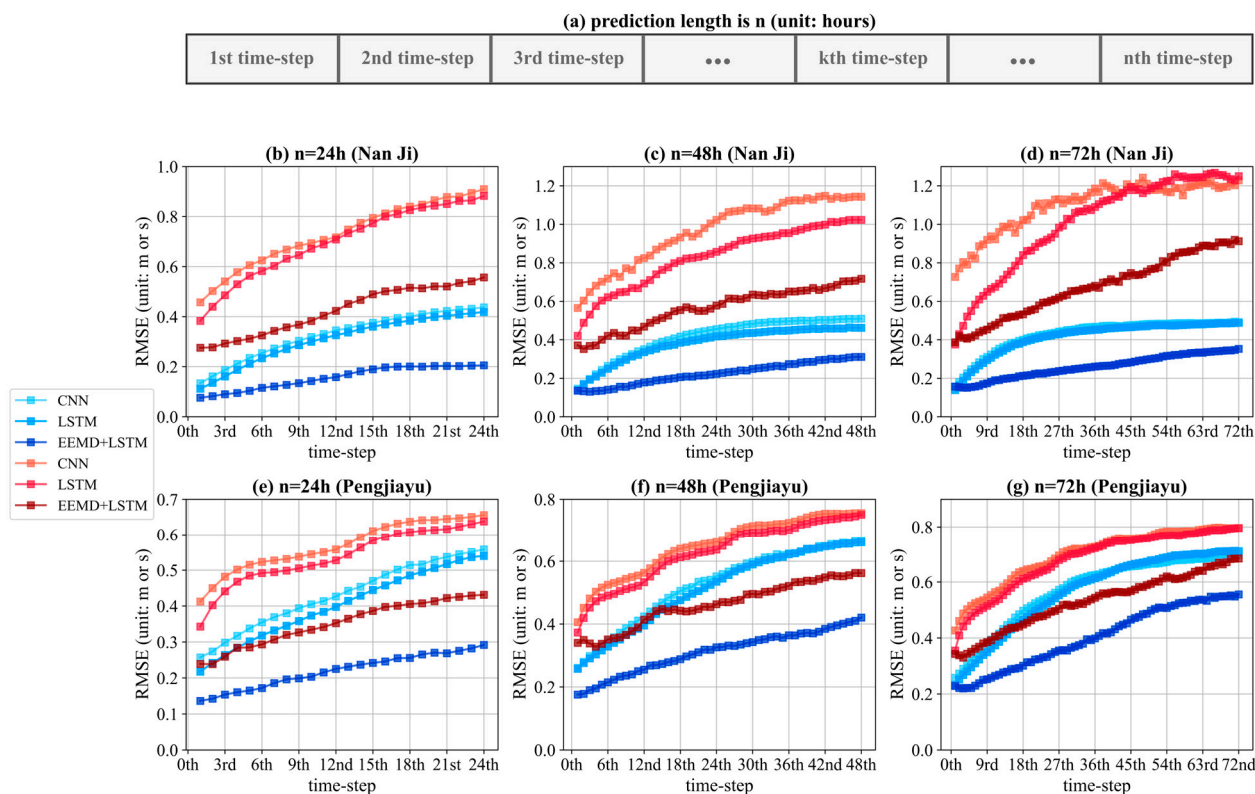
| Station   | Prediction Length (Unit: Hours) | Metrics  |          |        |                |
|-----------|---------------------------------|----------|----------|--------|----------------|
|           |                                 | RMSE     | MAE      | MAPE   | R <sup>2</sup> |
| Nan Ji    | 24                              | 0.4346 s | 0.3129 s | 5.50%  | 0.8370         |
|           | 36                              | 0.4971 s | 0.3534 s | 6.21%  | 0.7873         |
|           | 48                              | 0.5737 s | 0.4098 s | 7.2%   | 0.7172         |
|           | 72                              | 0.6938 s | 0.4872 s | 8.24%  | 0.5884         |
|           | 96                              | 0.8070 s | 0.5719 s | 9.83%  | 0.4451         |
|           | 120                             | 0.8656 s | 0.6304 s | 10.92% | 0.3625         |
| Pengjiayu | 24                              | 0.3569 s | 0.2678 s | 5.03%  | 0.8206         |
|           | 36                              | 0.4154 s | 0.3113 s | 5.84%  | 0.7571         |
|           | 48                              | 0.4668 s | 0.3553 s | 6.71%  | 0.6932         |
|           | 72                              | 0.5388 s | 0.4105 s | 7.72%  | 0.5912         |
|           | 96                              | 0.5720 s | 0.4415 s | 8.35%  | 0.5393         |
|           | 120                             | 0.6544 s | 0.5035 s | 9.56%  | 0.3975         |

### 3.3.4. Performance of CNN, LSTM, EEMD+LSTM for Typhoon Wave Prediction

For a given prediction length, it is important to understand the errors at each time step. Figure 8a shows a diagram of a multi-step prediction for an output length of  $n$ . Correspondingly, there is an error in each time step, which is denoted as 1st time step, 2nd time step, ...,  $k$ th time step, ..., and  $n$ th time step. The error is estimated by RMSE. The results in Sections 3.3.1–3.3.3 indicate that the performance of LSTM+Attention and CNN+LSTM on test sets is essentially consistent with that of LSTM. Here, the CNN, LSTM, and EEMD+LSTM models were selected to study the variation in RMSE with time step on test sets.

The relationship between RMSE and time step at Nan Ji station for prediction lengths  $n$  of 24 h, 48 h, and 72 h is shown in Figure 8b–d. The fact that the red curves are greater than the blue ones signifies that, in each time step, the values of RMSE for wave height predictions are higher than those for wave period predictions. For wave height predictions at Nan Ji station, the RMSE of CNN and LSTM rapidly increased within the first 24 time steps, gradually leveling off after the 24th time step (Figure 8b–d). In contrast, the RMSE from EEMD+LSTM maintained lower values than those of CNN and LSTM at each time

step, and the maximum did not surpass 0.4 m. The comparison of the three blue-toned curves in Figure 8b–d reveals that EEMD+LSTM performed better in each time step for medium- and long-term wave height predictions. Similar conclusions were also obtained by comparing the three red-toned curves in Figure 8b–d. The maximum RMSE from EEMD+LSTM did not exceed 1.0 s for all time steps, whereas the RMSE values of CNN and LSTM are higher than 1.0 s after the 24th time step for prediction lengths  $n$  of 48 h and 72 h, respectively (Figure 8c,d).

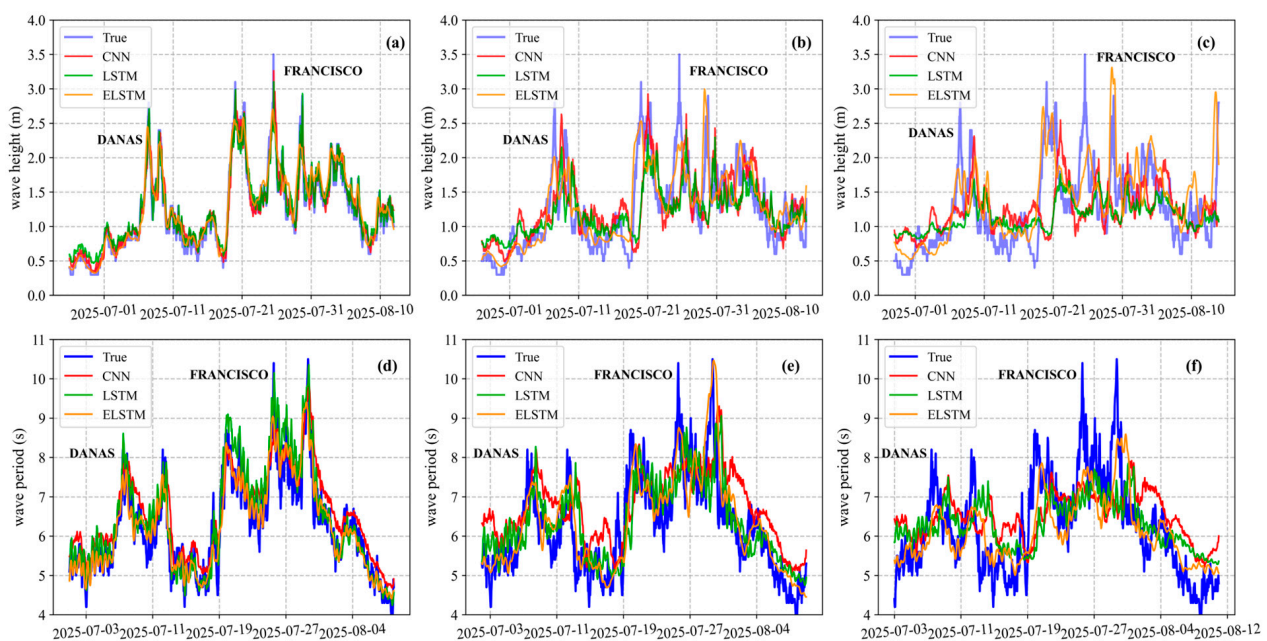


**Figure 8.** (a) Diagram for a prediction length of  $n$ . Variations in RMSE with time step for  $n = 24$  h,  $48$  h, and  $72$  h at Nan Ji (b–d) and Pengjiayu stations (e–g). Blue and red curves represent the wave height and period predictions, respectively. The three dots denote the omitted time description.

The variations in RMSE with time step at Pengjiayu station are shown in Figure 8e–g. For wave height predictions, the RMSE generally increased with time step, and the values of RMSE from EEMD+LSTM are much lower than those of CNN and LSTM for three prediction lengths. The maximum RMSE from EEMD+LSTM is lower than  $0.6$  m, which again reflects the benefits of EEMD+LSTM in multi-step wave height predictions. For wave period predictions, the RMSE from EEMD+LSTM is also much lower than that of other models in each time step. In conclusion, the variations in RMSE versus time step at Pengjiayu and Nan Ji stations, based on CNN, LSTM, and EEMD+LSTM, are consistent. For all models, the values of RMSE increased with increasing time step, with that from EEMD+LSTM exhibiting the slowest increase rate. The models performed better at predicting wave height than wave periods.

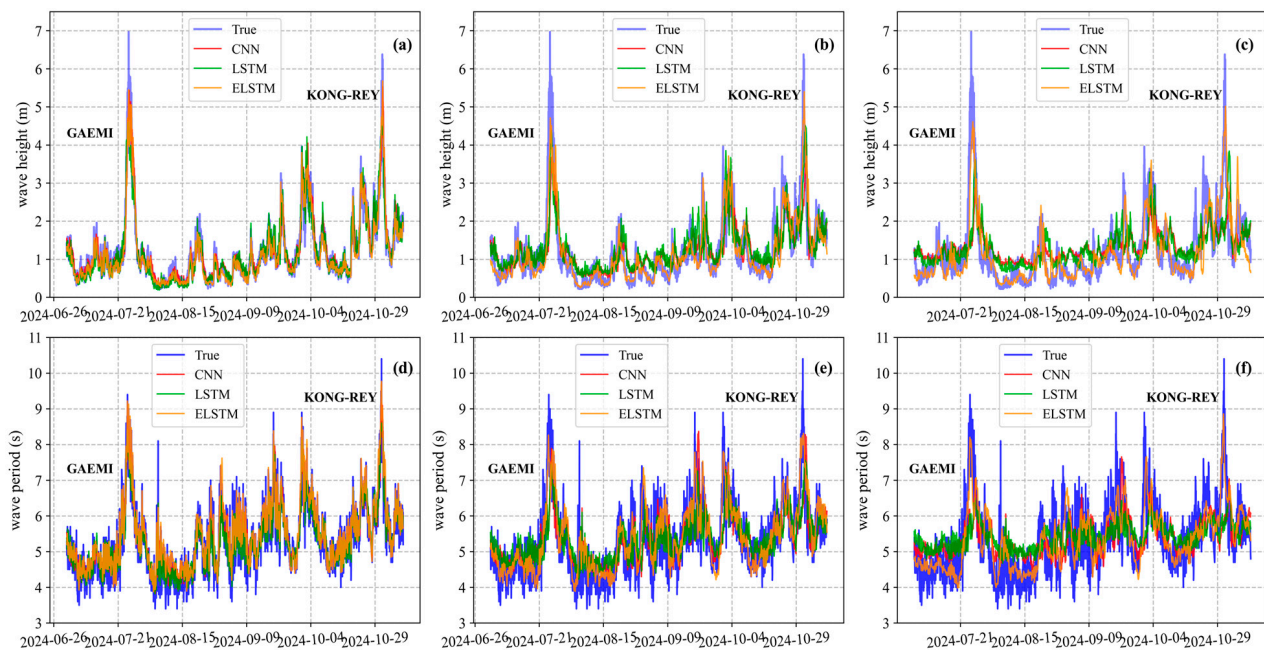
Figures 9 and 10 show the predicted wave heights and periods during typhoons Danas, Francisco, Gaemi, and Kong-rey at three time steps, 1st, 24th, and 48th, when the prediction time is  $48$  h. The EEMD+LSTM is replaced by ELSTM for short in the two figures. The typhoon waves caused by Danas and Francisco were captured by Nan Ji station. The predicted wave heights based on CNN, LSTM, and EEMD+LSTM agree well with the true values at the 1st time step (Figure 9a). At the 24th time step, the wave heights predicted using CNN and LSTM show a phase lag relative to the true values (Figure 9b). Specifically,

the wave height peaks predicted by CNN and LSTM emerged several hours later than the true wave height peaks. The wave heights predicted using EEMD+LSTM overcome this issue and better depict the trends of wave height sequences at the 24th time step. At the 48th time step, the phase lag problem in CNN and LSTM becomes more serious (Figure 9c). Conversely, the EEMD+LSTM outperformed CNN and LSTM in the 48th time step, fitting well with several wave height peaks (Figure 9c). At Nan Ji station, the three models performed worse on predictions of wave periods than wave heights at the 1st, 24th, and 48th time steps (Figure 9d–f). The higher wave period standard deviation contributes to this result. The predicted wave periods agree better with the truth at the 1st time step than at the 24th and 48th time steps (Figure 9d–f). However, the predicted wave periods from the three models fail to capture the local trend, which shows a rapidly changing amplitude. At the 24th and 48th time steps, there was also a phase lag between the predicted wave periods from CNN and LSTM and the true values, with it being more pronounced for the CNN predictions. In this aspect, the EEMD+LSTM performs better than the others.



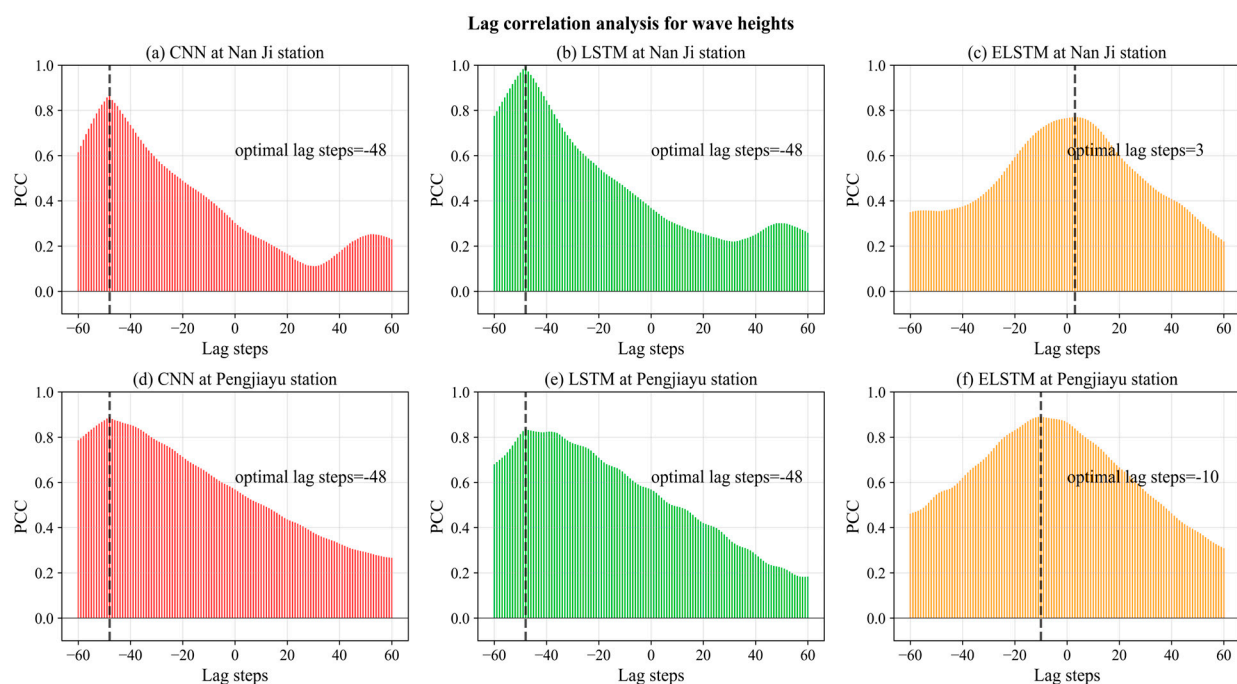
**Figure 9.** Predictions of typhoon waves at Nan Ji station for prediction length  $n = 48$  h. Predictions at the 1st time step (a,d), 24th time step (b,e), and 48th time step (c,f).

The typhoon waves caused by Gaemi and Kong-rey were captured by Pengjiayu station. At the 1st time step, the predicted wave heights based on CNN, LSTM, and EEMD+LSTM fit well with the true values (Figure 10a). At the 24th and 48th time steps, the phase lag also appeared in the predicted wave heights based on CNN and LSTM, which did not exist in EEMD+LSTM predictions (Figure 10b,c). Notably, only the EEMD+LSTM predicted the wave height peaks exceeding 5 m at the 24th and 48th time steps. For the wave period predictions at Pengjiayu station, the predictions based on three models agree well with the truth at the 1st time step, except for some local minima and maxima (Figure 10d). The ability of the three models to capture the high wave periods during typhoons deteriorates at the 24th and 48th time steps (Figure 10e,f). In particular, at the 48th time step, the CNN and LSTM's predicted results show phase lags, whereas the EEMD+LSTM predicted wave periods closely followed the trend of the true values and captured peak values around 21 July 2024 and 29 October 2024, caused by Gaemi and Kong-rey.

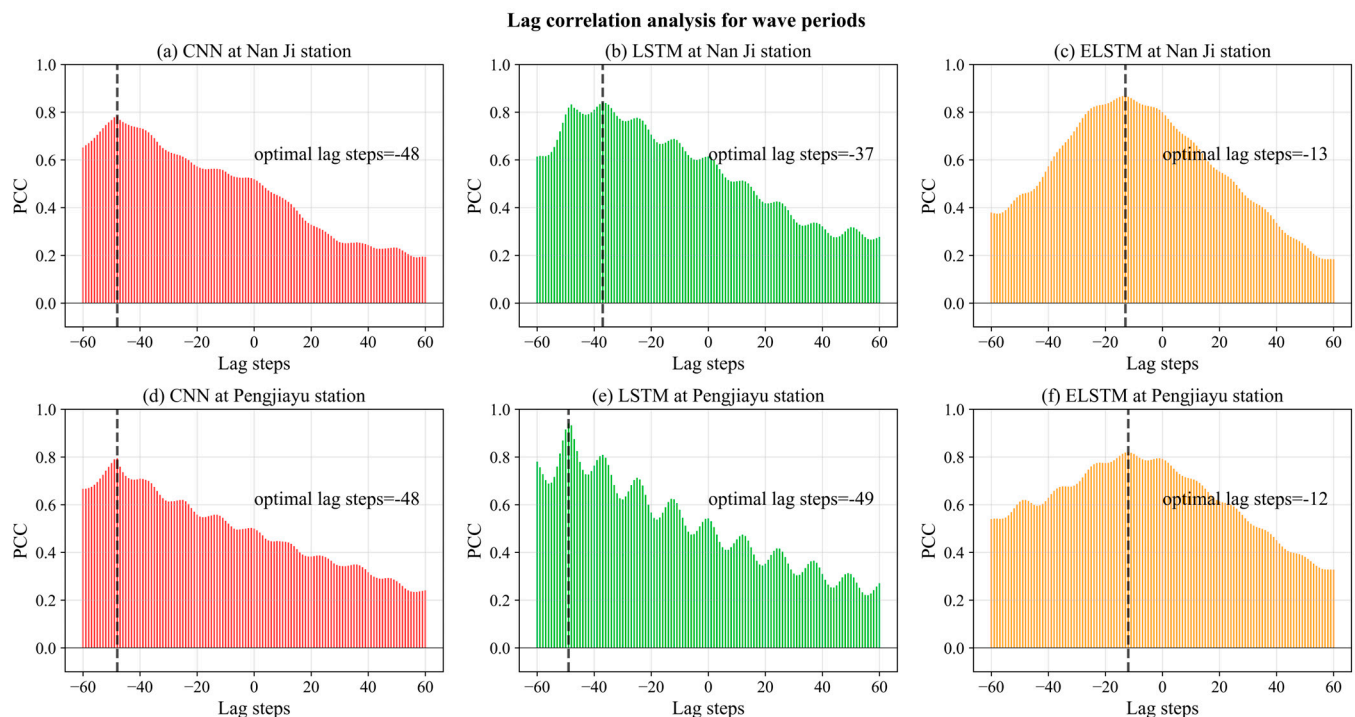


**Figure 10.** Predictions of typhoon waves at Pengjiayu station for prediction length  $n = 48$  h. Predictions at the 1st time step (a,d), 24th time step (b,e), and 48th time step (c,f).

Figures 11 and 12 show the lag correlation analyses between the predictions and true values at the 48th time step for a prediction length of 48 h. For wave height predictions at both stations (Figure 11), the optimal lag step of CNN and LSTM is  $-48$ , which is significantly greater than that of EEMD+LSTM. This quantitatively reveals the prominent phase lag in the prediction results from CNN and LSTM. This phenomenon also exists in wave period predictions. As shown in Figure 12, the optimal lag steps from CNN and LSTM are  $-48$  and  $-37$  at Nan Ji station, and  $-48$  and  $-49$  at Pengjiayu station. By contrast, the phase lag issue is not pronounced in the predictions from EEMD+LSTM.



**Figure 11.** Lag correlation analysis for wave heights at the 48th time step for a prediction length  $n = 48$  h. The vertical dashed lines denote the optimal lag steps.



**Figure 12.** Lag correlation analysis for wave periods at the 48th step for a prediction length  $n = 48$  h. The vertical dashed lines denote the optimal lag steps.

#### 4. Discussion and Conclusions

The multi-step prediction of wave heights and periods is important for reducing marine accidents. In this study, based on the datasets from Nan Ji and Pengjiayu, two observational stations located in the ocean southeast of Fujian Province, the performances of CNN, LSTM, LSTM+Attention, CNN+LSTM, and EEMD+LSTM were analyzed and discussed with regard to short-, medium-, and long-term predictions of wave heights and periods. For all models, the input length was set to 24 h, and the prediction lengths were set to 1 h, 6 h, 12 h, 24 h, 36 h, 48 h, 72 h, 96 h, and 120 h. The ratio of training, validation, and test sets was 0.70:0.15:0.15. The wave data from typhoons Danas and Francisco in 2025 were split into the test set at Nan Ji station. The wave data from the typhoons Gaemi and Kong-rey in 2024 were split into the test set at Pengjiayu station. The main research methods and principal findings of this study are as follows.

Linear correlation analysis was applied to both datasets. It was found that at Nan Ji station, wave height only showed a moderate linear correlation with wind speed, and wave period showed a weak linear correlation with all variables. At Pengjiayu station, wave heights showed moderate or strong linear correlations with wave period, mean wind speed, and gust wind speed, and wave period showed strong linear correlations with wave height, wave period, and spectral peak period. It was found that inputting variables with a weak correlation to the target variables failed to improve the predictive accuracy. CNN, LSTM, LSTM+Attention, and CNN+LSTM performed well for prediction lengths of 1 h, 6 h, and 12 h. The metrics in test sets satisfied  $RMSE < 0.4$  m,  $MAE < 0.4$  m,  $MAPE < 30\%$ , and  $R^2 > 0.7$  for wave height predictions, and  $RMSE < 0.7$  s,  $MAE < 0.5$  s,  $MAPE < 10\%$ , and  $R^2 > 0.6$  for wave period predictions. For prediction lengths longer than 24 h, the prediction capabilities of CNN, LSTM, LSTM+Attention, and CNN+LSTM weakened. Generally, the performances of LSTM and LSTM+Attention were better than those of CNN and CNN+LSTM. This indicates that our CNN-based model architectures did not show an advantage in single-variable predictions. It was also found that incorporating an attention mechanism into the LSTM did not effectively boost its performance when predicting a single indicator.

The application of EEMD to LSTM significantly improved LSTM's predictive capability for prediction lengths longer than 24 h. Across the test sets, for prediction lengths of 24 h, 36 h, 48 h, and 72 h, the EEMD+LSTM performed much better than the other four models. Specifically, the coefficient of determination  $R^2$  for wave height predictions exceeds or approaches 0.7, and that for wave period predictions exceeds or approaches 0.6. In addition, for a given prediction length, the errors at each time step for the predictive results of CNN, LSTM, and EEMD+LSTM were analyzed. Using the RMSE to estimate the errors reveals that the RMSE increases with the time step for both wave height and period predictions. At each time step, the RMSE values from EEMD+LSTM are much lower than those from CNN and LSTM. For a prediction length of 48 h, at the 1st, 24th, and 48th time steps, the predictions of time series of wave heights and periods during four typhoon cases were illustrated. The CNN, LSTM, and EEMD+LSTM agree well with the truth at the 1st time step, particularly for wave height predictions. The predicted results from CNN and LSTM show a phase lag at the 24th and 48th time steps, which did not appear in the EEMD+LSTM. It also found that, at the 24th and 48th time steps, EEMD+LSTM successfully captured several typhoon-induced wave heights and periods, whereas the CNN and LSTM did not.

In conclusion, the EEMD+LSTM architecture developed in this study performed better on medium- and long-term predictions than CNN, LSTM, LSTM+Attention, and CNN+LSTM. In fact, for prediction lengths of 96 h and 120 h, the performance of EEMD+LSTM was unsatisfactory. In the future, we will apply this fusion model to other datasets featuring different wave climates and meteorological conditions to evaluate its performance.

There is a contradiction in this study: we applied EEMD to the data before splitting it. While this approach is reasonable from the standpoint of physics, from a machine learning perspective, it results in possible information leakage. Our further work will focus on improving the prediction performance of EEMD+LSTM, considering both the physical meaning and information protection.

Further work will also focus on addressing the above issues to enhance the model performance, such as fusing other models or adopting a transformer framework.

**Author Contributions:** Conceptualization, T.Z. and M.L.; methodology, T.Z.; software, T.Z., M.L. and Z.S.; validation, T.Z., M.L. and Z.S.; formal analysis, T.Z.; investigation, M.L.; resources, M.L. and Z.S.; data curation, T.Z.; writing—original draft preparation, T.Z.; writing—review and editing, T.Z. and M.L.; visualization, T.Z. and M.L.; supervision, T.Z.; project administration, T.Z.; funding acquisition, T.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Research Setup Funding of Fujian University of Technology, grant number GY-Z220208.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Acknowledgments:** During the preparation of this manuscript, the author used PyCharm (Version 2025.2.3) for model architecture and result visualization. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

## Abbreviations

The following abbreviations are used in this manuscript:

|      |                                       |
|------|---------------------------------------|
| CNN  | Convolutional Neural Network          |
| LSTM | Long Short-Term Memory neural network |
| EEMD | Ensemble Empirical Mode Decomposition |

PCC  
RMSE  
MAE  
MAPE  
 $R^2$

Pearson correlation coefficient  
Root Mean Square Error  
Mean Absolute Error  
Mean Absolute Percentage Error  
Coefficient of Determination

## Appendix A

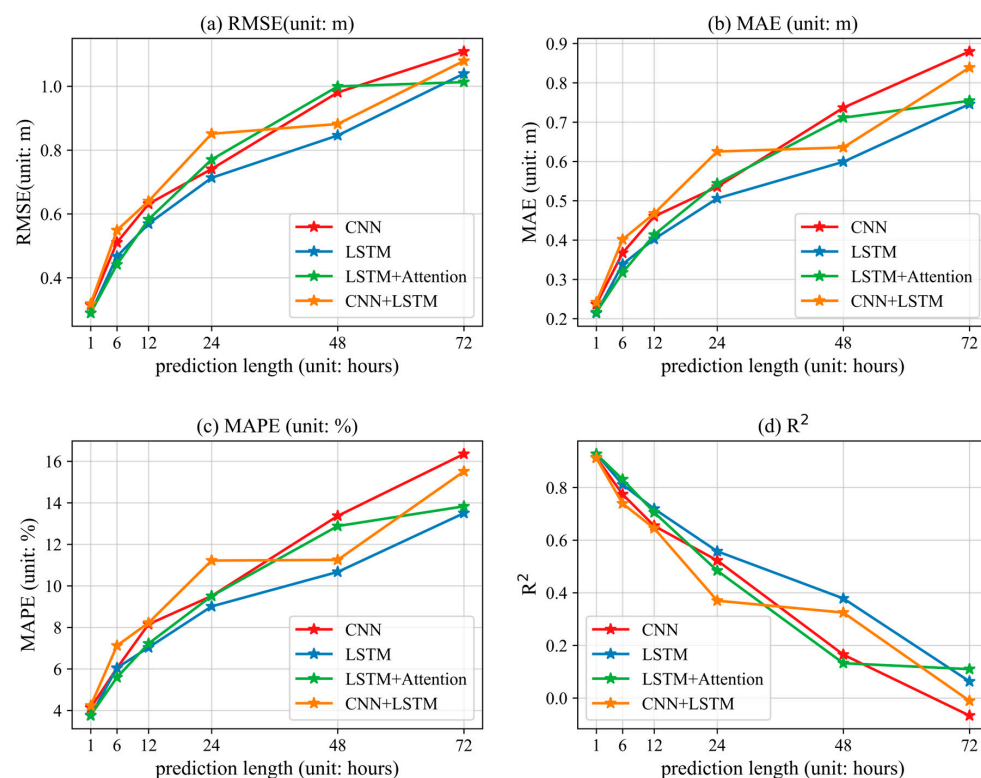


Figure A1. Graphical form of Table 5.

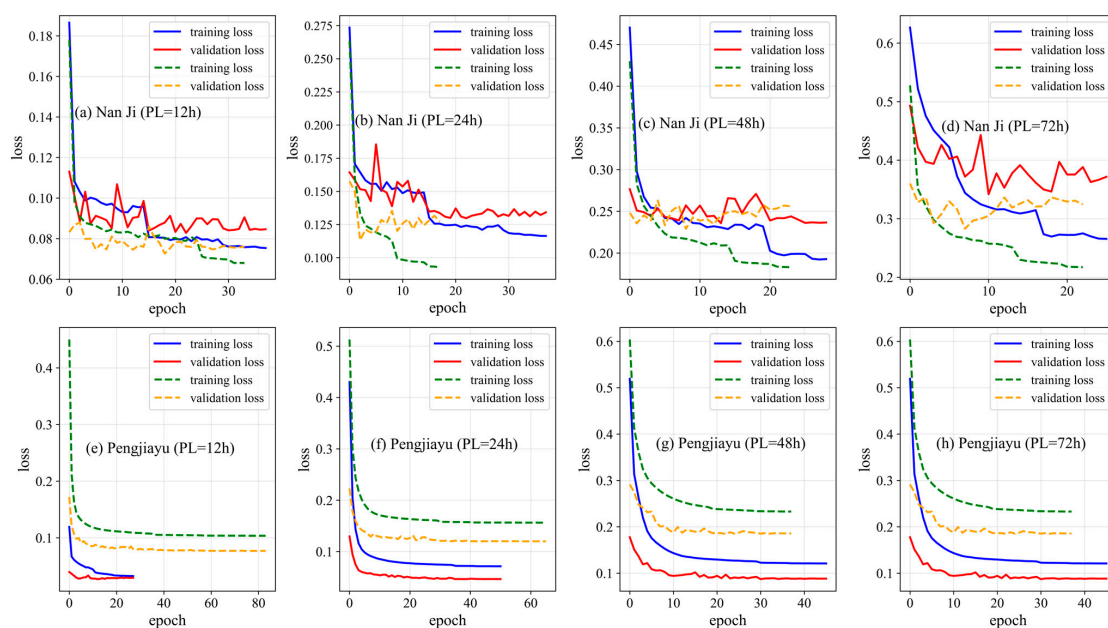
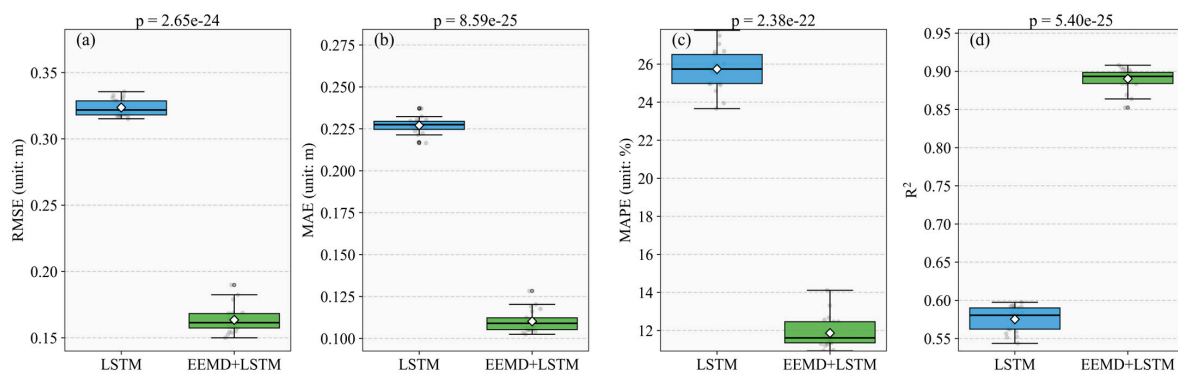


Figure A2. Training and validation losses for different prediction lengths (abbreviated as PL) at Nan Ji and Pengjiayu stations.



**Figure A3.** Statistical significance tests for the prediction length of 24 h. Note that  $2.65 \times 10^{-24}$ ,  $8.59 \times 10^{-25}$ ,  $2.38 \times 10^{-22}$ , and  $5.40 \times 10^{-25}$ . The box plots of (a) RMSE, (b) MAE, (c) MAPE, and (d)  $R^2$  for LSTM and EEMD+LSTM.

## References

1. Ayyad, M.; Hajj, M.R.; Marsooli, R. Artificial intelligence for hurricane storm surge hazard assessment. *Ocean Eng.* **2022**, *245*, 110435. [\[CrossRef\]](#)
2. Ellenson, A.; Pei, Y.; Wilson, G.; Özkan-Haller, H.T.; Fern, X. An application of a machine learning algorithm to determine and describe error patterns within wave model output. *Coast. Eng.* **2020**, *157*, 103595. [\[CrossRef\]](#)
3. Qian, C.C.; Huang, B.X.; Yang, X.Q.; Ge, C.G. Data science for oceanography: From small data to big data. *Big Earth Data* **2022**, *6*, 236–250. [\[CrossRef\]](#)
4. Luo, Q.; Hu, H.; Bai, L. Prediction of significant wave height in hurricane area of the Atlantic Ocean using the Bi-LSTM with attention model. *Ocean Eng.* **2022**, *266*, 112747. [\[CrossRef\]](#)
5. Xie, W.H.; Xu, G.J.; Zhang, H.C.; Dong, C.M. Developing a deep learning-based storm surge forecasting mode. *Ocean Model.* **2023**, *182*, 102179. [\[CrossRef\]](#)
6. Young, I.R.; Ribal, A. Multiplatform evaluation of global trends in wind speed and wave height. *Science* **2019**, *364*, 548–552. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Chen, Z.W.; Jiao, J.L.; Chen, Y.M.; Jiang, C.; Chen, S. Influence of wave directional spreading of short-crested irregular waves on ship motions and wave loads. *Mar. Struct.* **2025**, *103*, 103828. [\[CrossRef\]](#)
8. 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. [\[CrossRef\]](#)
9. Ali, A.; Fathalla, A.; Salah, A.; Bekhit, M.; Eldesouky, E. Marine data prediction: An evaluation of machine learning, deep learning, and statistical predictive models. *Comput. Intell. Neurosci.* **2021**, *13*, 8551167. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Lin, Y.; Bianucci, L. Seasonal variability of the ocean circulation in Queen Charlotte Strait, British Columbia. *Atmosphere-Ocean* **2023**, *62*, 35–57. [\[CrossRef\]](#)
11. Papandreou, A.; Hara, T.; Ginis, I. Impact of wave-current interaction in coupled storm surge-wave model. *J. Geophys. Res. Ocean.* **2025**, *19*, 7–89. [\[CrossRef\]](#)
12. Cheng, L.; Wang, C.; Guo, B.; Liang, Q.; Xie, Z.; Yuan, Z.; Chen, X.; Hu, H.; Du, P. Numerical investigation on the interaction between large-scale continuously stratified internal solitary wave and moving submersible. *Appl. Ocean Res.* **2024**, *145*, 103938. [\[CrossRef\]](#)
13. Kazeminezhad, M.H.; Siadatmousavi, S.M. Performance evaluation of WAVEWATCH III model in the Persian Gulf using different wind resources. *Ocean Dyn.* **2017**, *67*, 839–855. [\[CrossRef\]](#)
14. Chen, Y.; Yu, X. Enhancement of wind stress evaluation method under storm conditions. *Clim. Dyn.* **2016**, *47*, 3833–3843. [\[CrossRef\]](#)
15. Liu, Q.; Rogers, W.E.; Babanin, A.V.; Young, I.R.; Romero, L.; Zieger, S.; Qiao, F.; Guan, C. Observation-based source terms in the third-generation wave model WAVEWATCH III: Updates and verification. *J. Phys. Oceanogr.* **2019**, *49*, 489–517. [\[CrossRef\]](#)
16. Zhang, T. Impact of Sea Spray on the Simulation of Air–Sea Momentum Fluxes and Wave Heights during Severe Typhoon Soulik (2018). *J. Phys. Oceanogr.* **2025**, *55*, 1121–1137. [\[CrossRef\]](#)
17. Ling, W.; Torgeir, M.; Zhen, G.; Wei, S. A review on the technical development of combined wind and wave energy conversion systems. *Energy* **2024**, *294*, 130885. [\[CrossRef\]](#)
18. Ning, C.; Li, H.; Wang, Z.; Li, C.; Zeng, L.; Shao, W.; Nie, S. Significant Wave Height Prediction Using LSTM Augmented by Singular Spectrum Analysis and Residual Correction. *J. Mar. Sci. Eng.* **2025**, *13*, 1635. [\[CrossRef\]](#)

19. Ding, T.; Wu, D.; Shen, L.; Liu, Q.; Zhang, X.; Li, Y. Prediction of significant wave height using a VMD-LSTM-rolling model in the South Sea of China. *Front. Mar. Sci.* **2024**, *11*, 1382248. [[CrossRef](#)]
20. James, S.C.; Zhang, Y.; O'Donncha, F. A machine learning framework to forecast wave conditions. *Coast. Eng.* **2018**, *137*, 1–10. [[CrossRef](#)]
21. Chen, C.; Lin, H.; Guan, D.; Cai, F.; Wang, Q.; Liu, Q. Enhancing typhoon wave hindcasting with random forests and BP neural networks in the SWAN model. *Front. Mar. Sci.* **2024**, *11*, 1472047. [[CrossRef](#)]
22. Luo, Y.; Shi, H.; Zhang, Z.; Zhang, C.; Zhou, W.; Pan, G.; Wang, W. Wave field predictions using a multi-layer perceptron and decision tree model based on physical principles: A case study at the Pearl River Estuary. *Ocean Eng.* **2023**, *277*, 114246. [[CrossRef](#)]
23. Sadeghifar, T.; Motlagh, M.N.; Azad, M.T.; Mahdizadeh, M.M. Coastal Wave Height Prediction using Recurrent Neural Networks (RNNs) in the South Caspian Sea. *Mar. Geod.* **2017**, *40*, 454–465. [[CrossRef](#)]
24. Lian, J.; Dong, P.P.; Zhang, Y.P.; Pan, J.G. A Novel Deep Learning Approach for Tropical Cyclone Track Prediction Based on Auto-Encoder and Gated Recurrent Unit Networks. *Appl. Sci.* **2020**, *10*, 3965. [[CrossRef](#)]
25. Gao, L. Wave Forecasting of Taiwan Strait and Its Surrounding Waters Based on Deep Learning. Master's Thesis, Fujian Agriculture and Forestry University, Fuzhou, China, 2019.
26. 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](#)]
27. 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](#)]
28. Li, S.; Hao, P.; Yu, C.; Wu, G. CLTS-Net: A More Accurate and Universal Method for the Long-Term Prediction of Significant Wave Height. *J. Mar. Sci. Eng.* **2021**, *9*, 1464. [[CrossRef](#)]
29. Kaloop, M.R.; Kumar, D.; Zarzoura, F.; Roy, B.; Hu, J.W. A wavelet-Particle swarm optimization-Extreme learning machine hybrid modeling for significant wave height prediction. *Ocean Eng.* **2020**, *213*, 107777. [[CrossRef](#)]
30. Domala, V.; Kim, T.W. Application of Empirical Mode Decomposition and Hodrick Prescott filter for the prediction single step and multistep significant wave height with LSTM. *Ocean Eng.* **2023**, *285*, 115229. [[CrossRef](#)]
31. Hu, H.G.; André, J.; Westhuysen, V.D.; Chu, P.; Ayumi, F.M. Predicting Lake Erie wave heights and periods using XGBoost and LSTM. *Ocean Model.* **2021**, *164*, 101832. [[CrossRef](#)]
32. Xu, H.H.; Deng, Y. Dependent evidence combination based on shearman coefficient and pearson coefficient. *IEEE Access* **2018**, *6*, 11634–11640. [[CrossRef](#)]
33. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [[CrossRef](#)]
34. Benjamin, L.; Timo, M.; Hannes, V.; Nasser, J.; Michael, W. Survey on long short-term memory networks for time series prediction. *Procedia CIRP* **2021**, *99*, 650–655. [[CrossRef](#)]
35. Guo, M.-H.; Xu, T.-X.; Liu, J.-J.; Liu, Z.-N.; Jiang, P.-T.; Mu, T.-J. Attention Mechanisms in Computer Vision: A Survey. *Comput. Vis. Media* **2022**, *8*, 331–368. [[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.