



OPEN ACCESS

EDITED BY

Weimin Huang,
Memorial University of Newfoundland,
Canada

REVIEWED BY

Zhiding Yang,
Memorial University of Newfoundland,
Canada
Fan Meng,
Nanjing University of Information Science and
Technology, China

*CORRESPONDENCE

De'an Wu
✉ 20080003@hhu.edu.cn

RECEIVED 24 January 2024

ACCEPTED 20 February 2024

PUBLISHED 11 March 2024

CITATION

Ding T, Wu D, Li Y, Shen L
and Zhang X (2024) A hybrid CEEMDAN-
VMD-TimesNet model for significant wave
height prediction in the South Sea of China.
Front. Mar. Sci. 11:1375631.
doi: 10.3389/fmars.2024.1375631

COPYRIGHT

© 2024 Ding, Wu, Li, Shen and Zhang. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

A hybrid CEEMDAN-VMD-TimesNet model for significant wave height prediction in the South Sea of China

Tong Ding¹, De'an Wu^{1*}, Yuming Li², Liangshuai Shen²
and Xiaogang Zhang³

¹College of Harbor, Coastal and Offshore Engineering, Hohai University, Nanjing, China,

²Municipal Engineering Corporation Limited, Zhuhai, Guangdong, China, ³Power China Kunming Engineering Corporation Limited, Kunming, China

Accurate prediction of significant wave height is of great reference value for wave energy generation. However, due to the non-linearity and non-stationarity of significant wave height, traditional algorithms face difficulties in achieving satisfactory prediction results. In this study, a hybrid CEEMDAN-VMD-TimesNet model is proposed for non-stationary significant wave height prediction. Based on the significant wave height in the South Sea of China, the performance of the SVM model, the GRU model, the LSTM model, the TimesNet model, the CEEMDAN-TimesNet model and the CEEMDAN-VMD-TimesNet model are compared in terms of multi-step prediction. It is found that the prediction accuracy of the TimesNet model is higher than that of the SVM model, the GRU model and the LSTM model. The non-stationarity of significant wave height is reduced by CEEMDAN decomposition. Thus, the CEEMDAN-TimesNet model performs better than the TimesNet model in predicting significant wave height. The prediction accuracy of the CEEMDAN-VMD-TimesNet model is further improved by employing VMD for the secondary decomposition of components with high and moderate complexity. Additionally, the CEEMDAN-VMD-TimesNet model can accurately predict trends and extreme values of significant wave height with minimal phase shifts even during typhoon periods. The results demonstrate that the CEEMDAN-VMD-TimesNet model exhibits superiority in predicting significant wave height.

KEYWORDS

wave height prediction, CEEMDAN, VMD, TimesNet, CEEMDAN-VMD-TimesNet

1 Introduction

More than 70% of the world is composed of oceans. Wave energy, as a clean and renewable energy source (Reikard et al., 2015), is crucial for addressing the energy crisis. For wave energy generation, the significant wave height is a crucial metric (Ali and Prasad, 2019). As a result, there has always been a great deal of focus on accurately predicting significant

wave height (Deo et al., 2001). However, predicting significant wave height is challenging. In particular, during extreme weather conditions such as typhoons, the significant wave height undergoes rapid changes, posing a significant challenge to accurate prediction.

So far, experts and researchers from various countries have attempted to establish numerical models to simulate significant wave height (Reikard et al., 2017). Reikard and Rogers (2011) tried to predict significant wave height in the Pacific region and the Gulf of Mexico using the SWAN model. It was found that when the forecast time exceeds 6 hours, the SWAN model's prediction accuracy surpasses that of mathematical models. Chen et al. (2018) created a wave-current system that combined FVCOM with SWAN. This system allowed for on-line, two-way coupling and was also more computationally efficient and user-friendly for simulating wave-current interactions near the coast. Amunugama et al. (2020) employed the COAWST model to simulate multiple intense typhoons in Japan and found that the COAWST model effectively reproduced typhoon phenomena, storm surges, and wave height. Vijayan et al. (2023) utilized the dynamic coupling of SWAN and ADCIRC to simulate Hurricane Michael. They observed that the accuracy of the simulation was significantly improved by the dynamic coupling model of SWAN and ADCIRC.

However, traditional physics models often require high accuracy in terrain and boundary conditions and also require a significant amount of computing time (Wang et al., 2018). Consequently, as computer technology advances, an increasing number of researchers have shown a strong interest in machine learning. Soares and Cunha (2000) utilized a bivariate AR model incorporating significant wave height and average wave period to predict significant wave height at two adjacent stations in Portugal. Agrawal and Deo (2002) employed ARMA and ARIMA models to predict significant wave height at a nearshore location in India at intervals of 3, 6, 12, and 24 hours. By using Support Vector Machine (SVM), Mahjoobi and Mosabbebi (2009) successfully predicted significant wave height in Lake Michigan with acceptable accuracy.

Neural networks, known for the superiority in handling non-linear problems, have gained popularity among researchers. A growing number of experts have adopted the usage of neural networks for the purpose of predicting significant wave height. Chang et al. (2011) developed an Artificial Neural Network (ANN) model for typhoon waves, using local wind and simulated waves as the two critical factors for the proposed ANN model's input parameters. With the development of deep learning, long short-term memory (LSTM) network (Hochreiter and Schmidhuber, 1997) has gained attention. LSTM network can effectively capture the long-term dependencies in time series sequences and utilize historical data for prediction. Fan et al. (2020) used the wind speed data from the last 4 hours, together with the wave height and wind direction data from the previous 1 hour, as input variables to predict significant wave height at 10 distinct stations with varying oceanic environmental circumstances. The findings demonstrated that the LSTM model was capable of attaining precise prediction outcomes. Bethel et al. (2022) utilized LSTM to predict significant wave height during hurricanes Dorian, Sandy, and Igor. They found that the LSTM model is highly suitable for swiftly forecasting hurricane-forced wave height, with significantly lower computational expenses

when compared to numerical wave models. Song et al. (2022) developed a high spatial and temporal resolution regional significant wave height prediction model based on the ConvLSTM model. Additionally, the Mask method and Replace mechanism were applied to improve the model's long-term forecasting capabilities. Gate recurrent unit (GRU), as a variant of LSTM, has also been put into use. Meng et al. (2021) proposed a BiGRU network for wave height prediction during TC. The prediction by BiGRU of 10 buoys during a new typhoon was compared with machine learning models. The findings indicated that the BiGRU's predictive ability is stable, particularly for predicting 24 hours in advance. Li et al. (2022) predicted significant wave height at six stations along the coast of China. The experimental results showed that for a 3-hour forecast, the GRU network exhibited stronger robustness compared to the LSTM network. The prediction accuracy of significant wave height has been improved by the application of these deep learning techniques.

Unlike traditional convolutional neural networks and recurrent neural networks, the transformer utilized the attention mechanism to capture data features and has achieved good results (Vaswani et al., 2017). Shortly after, the Informer was proposed, which greatly reduced the computation time through the ProbSparse self-attention mechanism and the self-attention distilling (Zhou et al., 2021). Based on the Auto-Correlation mechanism, the Autoformer was successfully proposed. The accurate prediction was achieved using the Autoformer (Wu et al., 2021). Recently, the one-dimensional time series was transformed into a two-dimensional tensor to capture intra-period and inter-period changes. Following this approach, the TimesNet model was constructed, which achieved significant progress in time series prediction (Wu et al., 2022).

Due to the non-stationarity of significant wave height data, simply altering the network structure often fails to solve the problem effectively. Therefore, scholars have considered decomposing the significant wave height data to reduce the non-stationarity. Empirical mode decomposition (EMD) has demonstrated excellent performance in handling non-linear and non-stationary data (Huang et al., 1998). Duan et al. (2016) utilized EMD-AR to predict significant wave height. The results proved the effectiveness of EMD in handling non-linear and non-stationary significant wave height. Hao et al. (2022) decomposed significant wave height using EMD and then used the LSTM to predict the decomposed components. They found that the EMD-LSTM model significantly improves prediction accuracy. Compared to EMD, ensemble empirical mode decomposition (EEMD) exhibits greater robustness in terms of sampling and noise (Wu and Huang, 2009). EEMD and LSTM were used to propose a real-time TC wave height forecasting system (ATDNNS). The experimental findings revealed that ATDNNS performed much better compared to the baseline model and had the potential to address the issue of poor initial forecast performance in numerical models (Meng et al., 2022). The significant wave height in the Indian Ocean region was predicted using EEMD-LSTM. It was found that the prediction results by EEMD-LSTM were more accurate compared to the prediction results by EMD-LSTM (Song et al., 2023). Variational mode decomposition (VMD) is another improved version of

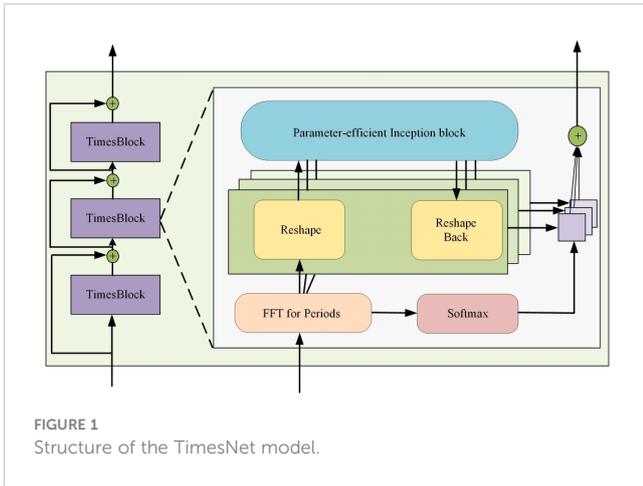


FIGURE 1 Structure of the TimesNet model.

where \mathbf{X}_{2D}^{li} is the transformed 2D tensor, $\hat{\mathbf{X}}_{2D}^{li}$ is the representation processed by the parameter-efficient inception block, and $\hat{\mathbf{X}}_{1D}^{li}$ is transformed from $\hat{\mathbf{X}}_{2D}^{li}$ for aggregation.

Given that amplitude is able to indicate the significance of a certain frequency and period, aggregation can be performed based on amplitude for one-dimensional representation. The process is formalized as follows:

$$\hat{\mathbf{A}}_{f_1}^{l-1}, \dots, \hat{\mathbf{A}}_{f_k}^{l-1} = \text{Softmax} \left(\mathbf{A}_{f_1}^{l-1}, \dots, \mathbf{A}_{f_k}^{l-1} \right) \quad (6)$$

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

It should be noted that the parameters representing the input and the output to the TimesNet model are \mathbf{X}_{1D} and \mathbf{X}_{1D}^l respectively.

2.2 CEEMDAN algorithm

By including IMF components with auxiliary noise after EMD decomposition, rather than simply introducing Gaussian white noise to the initial signal, the CEEMDAN algorithm overcomes the mode mixing problem and the transfer of white noise from high to low frequencies during signal decomposition (Torres et al., 2011). As a result, the CEEMDAN algorithm can better handle non-linear and non-stationary data, reduce the complexity of the original data, and improve the accuracy of prediction compared to EMD. Significant wave height is non-linear and non-stationary data. Therefore, we adopt CEEMDAN for the primary decomposition of significant wave height. However, there are a few drawbacks to using CEEMDAN. For example, in the beginning of the decomposition process, CEEMDAN could generate some “pseudo” modes, and it is difficult to fully extract the information from components with high and moderate complexity.

$f(t)$ is the input of CEEMDAN algorithm, which represents significant wave height. The steps of the algorithm are as follows:

1. By adding Gaussian white noise to the original signal $f(t)$ under decomposition, a new signal $y_i(t)$ is obtained:

$$y_i(t) = f(t) + \varepsilon_0 \delta_i(t), i = 1, 2, \dots, K \quad (7)$$

where K is the number of white Gaussian noises, ε_0 is a coefficient of intensity.

2. By taking the average of all the first decomposition components of EMD, the first IMF, denoted as, $\overline{\text{Imf}}_1(t)$ is obtained

$$\overline{\text{Imf}}_1(t) = \frac{1}{K} \sum_{i=1}^K \text{EMD}_1(y_i(t)) \quad (8)$$

$$r_1(t) = f(t) - \overline{\text{Imf}}_1(t)$$

where $\text{EMD}_1(\cdot)$ denotes the first IMF component produced by EMD method, $r_1(t)$ indicates the residual for the first stage.

3. To get the second IMF component, the signal $r_1(t) + \varepsilon_1 \text{EMD}_1(\delta_i(t))$ can be further decomposed and combined:

$$\overline{\text{Imf}}_2(t) = \frac{1}{K} \sum_{i=1}^K \text{EMD}_1(r_1(t) + \varepsilon_1 \text{EMD}_1(\delta_i(t))) \quad (9)$$

$$r_2(t) = r_1(t) - \overline{\text{Imf}}_2(t)$$

where $r_2(t)$ is the residual of the second stage.

4. The j th IMF component and j th residual can be calculated as:

$$\overline{\text{Imf}}_j(t) = \frac{1}{K} \sum_{i=1}^K \text{EMD}_1(r_{j-1}(t) + \varepsilon_{j-1} \text{EMD}_{j-1}(\delta_i(t))) \quad (10)$$

$$r_j(t) = r_{j-1}(t) - \overline{\text{Imf}}_j(t)$$

where $r_j(t)$ is the residual after the j th decomposition.

5. The algorithm terminates when the residual signal obtained cannot be further decomposed. The number of intrinsic mode components obtained at this time is N . The original signal $f(t)$ is decomposed into:

$$f(t) = \sum_{i=1}^N \overline{\text{Imf}}_i(t) + r_N(t) \quad (11)$$

2.3 Sample entropy algorithm

Sample entropy is a metric that assesses the complexity of a time series by assessing the probability of producing novel patterns in the signal (Richman and Moorman, 2000). A lower value of sample entropy indicates higher self-similarity in the sequence, while a higher value indicates greater complexity in the sample sequence. Both $\overline{\text{Imf}}_i(t) (i = 1, 2, \dots, N)$ and $r_N(t)$ generated in 2.2 need to calculate the sample entropy separately. $\overline{\text{Imf}}_1(t)$ is used as an example to illustrate the sample entropy algorithm. n represents the total number of data points in $\overline{\text{Imf}}_1(t)$. For $\overline{\text{Imf}}_1(t) = [x(1), x(2), \dots, x(n)]$, the steps for calculating sample entropy are as follows:

1. The vectors $x_m(i)$ can be calculated as follows:

$$x_m(i) = [x(i), x(i+1), \dots, x(i+m-1)], i = 1, 2, \dots, n-m+1 \quad (12)$$

where m denotes the length of sequences to be compared.

2. The distance between $x_m(i)$ and $x_m(j)$ is defined to be d_m :

$$d_m[x_m(i), x_m(j)] = \max [x_m(i+k) - x_m(j+k)], 0 \leq k \leq m-1 \quad (13)$$

3. Let r be the tolerance for accepting matrices. For a given $x_m(i)$, count the number of distances between $x_m(i)$ and $x_m(j)$ that are less than or equal to r and denote them as B_i . $B^m(r)$ is the probability of two sequences matching m points, which is defined as:

$$B^m(r) = \frac{1}{n-m} \sum_{i=1}^{n-m} \frac{1}{n-m+1} B_i \tag{14}$$

4. Increase the dimension to $m + 1$, count the number of distances between $x_{m+1}(i)$ and $x_{m+1}(j)$ that are less than or equal to r and denote them as A_i . $A^m(r)$ is the probability of two sequences matching $m + 1$ points, which is defined as:

$$A^m(r) = \frac{1}{n-m} \sum_{i=1}^{n-m} \frac{1}{n-m+1} A_i \tag{15}$$

5. Finally, the Sample Entropy is defined as $SampEn(m, r)$:

$$SampEn(m, r) = \lim_{n \rightarrow \infty} \left(-\ln \frac{A^m(r)}{B^m(r)} \right) \tag{16}$$

Sample Entropy can be defined using the following equation when n is a finite value:

$$SampEn(m, r, n) = -\ln \frac{A^m(r)}{B^m(r)} \tag{17}$$

The sample entropy of $\overline{Imf}_i(t) (i = 2, \dots, N)$ and $r_N(t)$ is computed just like $\overline{Imf}_1(t)$. All the calculated sample entropy is combined to form the sequence $s(N)$.

2.4 K-means clustering algorithm

The K-means clustering technique is a common unsupervised learning algorithm which uses the degree of similarity between samples to place them in appropriate groupings. A typical metric for comparing two objects' similarities is the Euclidean distance, which can be expressed as:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{18}$$

The sequence $s(N)$ formed in 2.3 is the input of K-means clustering algorithm, the steps for K-means clustering algorithm are as follows:

1. K samples are chosen as the first cluster centers $a_j(I)$. Set $I = 1$.
2. Compute the distance between each sample in the dataset $s(N)$ and the K cluster centers $a_j(I)$, then assign it to the class associated with the cluster center that has the shortest distance.
3. Determine the new cluster center using the following formula:

$$a_j(I + 1) = \frac{1}{|c_i|} \sum_{s(i) \in c_i} s(i) \tag{19}$$

where c_i is the cluster, $|c_i|$ is the number of data points in the cluster, and $s(i)$ is the data points in the cluster.

4. The clustering evaluation metrics J are computed based on the formula provided below:

$$J(I) = \sum_{i=1}^n \sum_{j=1}^k (\|s_i - a_j(I)\|)^2 \tag{20}$$

5. Exit the algorithm if $|J(I + 1) - J(I)| \leq \xi$ is met; otherwise, set $I = I + 1$ and return to step 2.

After completing the clustering, $\overline{Imf}_i(t) (i = 1, 2, \dots, N)$ and $r_N(t)$ that are in the same category are added together to form COIMF, which is denoted as the parameter $H(t)$.

2.5 VMD algorithm

VMD is a non-recursive signal processing method that decomposes time series data into a series of IMFs with finite bandwidth. It iteratively searches for the optimal solution of the variational mode.

By addressing a variational issue, the VMD method efficiently transforms decomposition into optimization. Hence, we think that VMD can extract the information in components with high and moderate complexity that CEEMDAN fails to extract completely, and further reduce the complexity of the data. Therefore, we use VMD for the secondary decomposition of significant wave height.

The procedure may be partitioned into two stages the formulation of the variational issue and its subsequent resolution.

VMD has redefined the intrinsic mode functions:

$$u_k(t) = A_k(t) \cos(\phi_k(t)) \tag{21}$$

where k is the mode number, $A_k(t)$ is the amplitude of the k mode, $\phi_k(t)$ is the phase of the k mode, $u_k(t)$ is the k mode function.

$H(t)$ is the data generated in 2.4, which is the input of VMD algorithm. The formulated constrained variational problem is as follows:

$$\begin{cases} \min_{\{u_k\}, \{\overline{\omega}_k\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\overline{\omega}_k t} \right\|_2^2 \right\} \\ \text{s. t. } \sum_{k=1}^K u_k(t) = H(t) \end{cases} \tag{22}$$

where u_k represents the respective mode functions, while $\overline{\omega}_k$ represents the respective mode center frequencies.

It is necessary to convert the constrained variational problem into an unconstrained variational problem in order to solve the described constrained optimization issue. By leveraging the advantages of quadratic penalty terms and the Lagrange multiplier method, an augmented Lagrangian function is introduced (Bertsekas, 1976).

$$\begin{aligned} L(\{u_k\}, \{\overline{\omega}_k\}, \lambda) = & \alpha \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\overline{\omega}_k t} \right\|_2^2 \\ & + \left\| H(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), H(t) - \sum_{k=1}^K u_k(t) \right\rangle \end{aligned} \tag{23}$$

where α represents the variance regularization parameter, while λ represents the Lagrangian multiplier.

To address the aforementioned variational problem, the ADMM is employed (Hestenes, 1969).

The procedure is outlined as follows:

1. Initialize $u_k^1, \bar{\omega}_k^1, \lambda_k^1$ and n .
2. The variable n is incremented by 1, and the program proceeds to enter the loop.
3. According to the Equation (24), the variables u_k and $\bar{\omega}_k$ are updated. The updating process stops when the number of iterations reaches k .

$$\begin{cases} \hat{u}_k^{n+1}(\omega) = \frac{\hat{H}(\omega) - \sum_{i < k} \hat{u}_i^{n+1}(\omega) + \sum_{i > k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \bar{\omega}_k^n)^2} \\ \bar{\omega}_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \end{cases} \quad (24)$$

4. According to the Equation (25), the variable λ is updated.

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left(\hat{H}(\omega) - \sum_{k=1}^K \hat{u}_k^{n+1}(\omega) \right) \quad (25)$$

5. If the user-defined variable ϵ satisfies the stopping condition (Equation 26), the loop is terminated, otherwise it proceeds to step 2 and continues the iteration.

$$\sum_{k=1}^K \frac{\|\hat{u}_k^{n+1}(\omega) - \hat{u}_k^n(\omega)\|_2^2}{\|\hat{u}_k^n(\omega)\|_2^2} < \epsilon \quad (26)$$

In fact, $H(t)$ after decomposition still has the residual. The residual is denoted as $r_k(t)$.

$$r_k(t) = H(t) - \sum_{k=1}^K u_k(t) \quad (27)$$

$u_k(t)$ and $r_k(t)$ are the output of VMD algorithm.

2.6 CEEMDAN-TimesNet

The CEEMDAN-TimesNet model for predicting significant wave height involves four distinct phases. The entire process is illustrated in Figure 2.

1. The initial action involves decomposing significant wave height $f(t)$ using the CEEMDAN algorithm. This process produces IMFs $\bar{Imf}_i(t)$ as well as a residual component $r_N(t)$.
2. For each IMF and residual component, the sample entropy is calculated using the sample entropy algorithm. All the calculated sample entropy is combined to form the sequence $s(N)$. Then, the K-means clustering technique is used to categorize $\bar{Imf}_i(t)$ and $r_N(t)$ into three distinct groups COIMF1, COIMF2 and COIMF3 based on $s(N)$.
3. The TimesNet model is utilized to independently predict COIMF1, COIMF2 and COIMF3.
4. Ultimately, the final result is achieved by the combination of the prediction outcomes of COIMF1, COIMF2 and COIMF3.

2.7 CEEMDAN-VMD-TimesNet

The process of predicting significant wave height via the CEEMDAN-VMD-TimesNet model has five distinct stages. The complete procedure is depicted in Figure 3.

1. The significant wave height $f(t)$ is decomposed through the utilization of the CEEMDAN algorithm, resulting in the generation of IMFs $\bar{Imf}_i(t)$ as well as a residual component $r_N(t)$.
2. The sample entropy algorithm is used to calculate the entropy values of each IMF and residual component. By combining all the computed sample entropy, the sequence $s(N)$ is formed. Based on $s(N)$, the K-means clustering algorithm is then applied to cluster $\bar{Imf}_i(t)$ and $r_N(t)$ into three categories COIMF1, COIMF2 and COIMF3.
3. The COIMF1 and COIMF2 are decomposed again using VMD to generate new IMFs $u_k(t)$ and a residual component $r_k(t)$.
4. The TimesNet model is employed to predict COIMF3 and new subsequences generated from COIMF1 and COIMF2 individually.

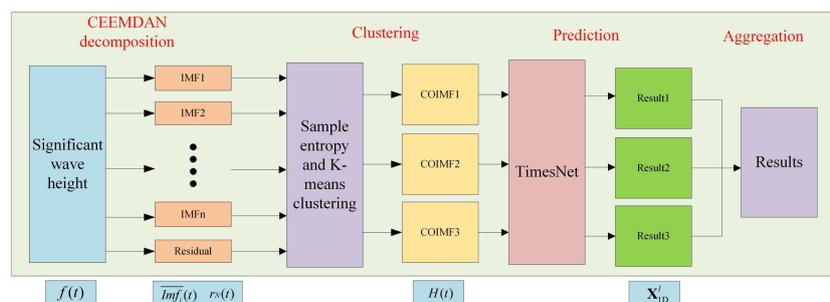


FIGURE 2 Flow chart of the hybrid CEEMDAN-TimesNet significant wave height prediction model.

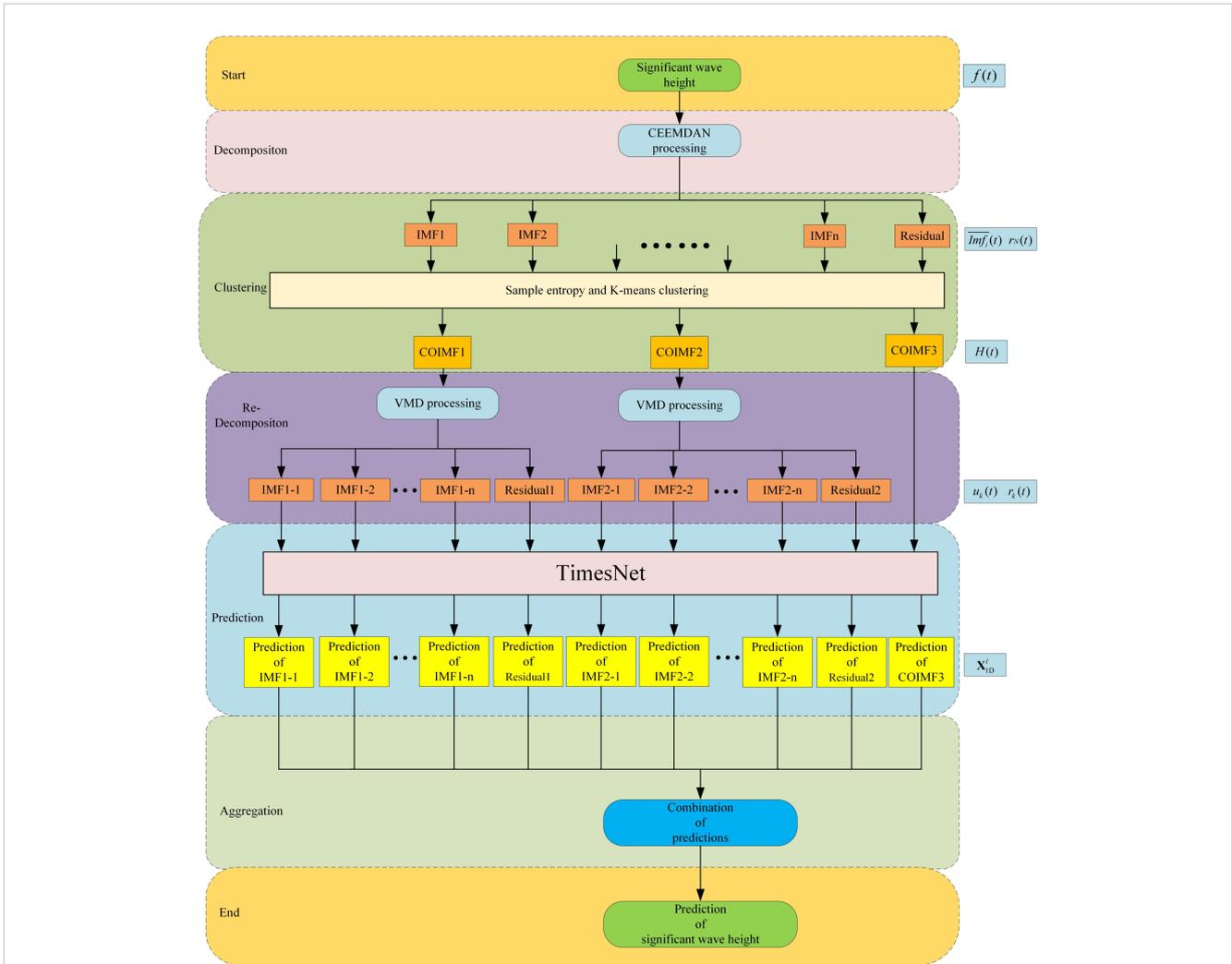


FIGURE 3 Flow chart of the hybrid CEEMDAN-VMD-TimesNet significant wave height prediction model.

5. In the final stage, the prediction results of COIMF3 and new subsequences generated from COIMF1 and COIMF2 are added together to form the ultimate result.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{30}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{31}$$

2.8 Evaluation metrics

R^2 , MSE , $RMSE$, and MAE are the four metrics used to quantify the errors between the predicted values and the actual values, which are used to evaluate the model's performance in this research. R^2 is used to evaluate the degree of agreement. MSE is used to quantify the average squared difference. $RMSE$ is used to measure the average difference. MAE is the mean of the absolute errors. The formulas are as follows:

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

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{29}$$

where \hat{y}_i represents the predicted value, y_i represents the actual value, \bar{y}_i represents the mean of the actual values, and n represents the number of data samples.

The larger R^2 indicates better agreement between the predicted values and the actual values, demonstrating stronger predictive capacity of the model. In contrast, the smaller MSE , $RMSE$, and MAE reflect the lower error and the better predictive performance.

3 Wave datasets preprocessing

3.1 Research area

The South Sea of China was chosen for this study. The original data for significant wave height was sourced from ERA5 of the

European Centre for Medium-Range Weather Forecasts (ECMWF) and covered the period from January 1, 2010, 00:00:00 UTC to December 31, 2018, 23:00:00 UTC. The time interval for significant wave height was one hour. There were a total of 78,888 valid data points. The dataset1 was selected at position 1, and the dataset2 was selected at position 2. The coordinates of position 1 were (21.50° N, 113.00° E), and the coordinates of position 2 were (23.00° N, 117.50° E). The location of the selected position in this paper is presented in Figure 4. The dataset1 and the dataset2 are illustrated in Figures 5, 6, respectively. The occurrence time of Typhoon Hato was from August 20, 2017, 14:00:00 UTC to August 24, 2017, 17:00:00 UTC. During the occurrence of Typhoon Hato, the maximum wind speed at position 1 is 25.08 m/s, and at position 2 is 20.84 m/s.

3.2 Data processing

3.2.1 Significant wave height decomposed by CEEMDAN

This study used the pyEMD package to implement the CEEMDAN. Figures 5, 6 depict the decomposed outcomes. As shown in Figures 5, 6, the dataset1 and the dataset2 are both decomposed into 18 subsequences. It can be observed that the dataset becomes stationary through CEEMDAN decomposition.

3.2.2 Clustering decomposed sub-sequences by Sample Entropy and K-means clustering

Sample Entropy was used to measure the complexity of each IMF by the Python sample module. During the computational process, mm and r were set to 1 and 0.1, respectively. Then the Kmeans clustering algorithm was then applied to cluster subsequences into three categories. We think this classification method effectively captures the information of various component types and saves computational time. Figures 7, 8 display the calculated sample entropy values and the newly clustered IMFs, where IMF18 is the residual. As shown in Figures 7, 8, for dataset1, these IMFs can be integrated into three new COIMFs: COIMF0 (IMF1-5), COIMF1

(IMF6-8), and COIMF2 (IMF9-18). For dataset2, these IMFs can also be integrated into three new COIMFs: COIMF0 (IMF1-4), COIMF1 (IMF5-8), and COIMF2 (IMF9-18). It is clear that COIMF1 is composed of highly complex IMF components, COIMF2 is composed of moderately complex IMF components, and COIMF3 is composed of lowly complex IMF components.

3.2.3 COIMF1 and COIMF2 decomposed by VMD

Following the first decomposition and reconstruction, the VMD was used again to decompose COIMF1 and COIMF2, aiming to enhance the stationarity and decrease the complexity of the data. Based on previous research and experience (Zhao et al., 2023), both COIMF1 and COIMF2 were decomposed into 10 IMFs and 1 residual. COIMF3 and new subsequences generated from COIMF1 and COIMF2 were subjected to maximum-minimum normalization. Then the data after normalization were predicted by TimesNet.

4 Results and discussion

4.1 Setup of model

First, the dataset partitioning should be emphasized. The dataset was partitioned into three subsets a training set including the initial 70% of the data consisting of 55221 samples, a validation set including the 10% of the data consisting of 7889 samples, and a testing set comprising the last 20% consisting of 15778 samples.

The SVM model, the GRU model, and the LSTM model are all commonly used for significant wave height prediction and are regularly employed in various studies for conducting comparative analyses (Meng et al., 2021; Song et al., 2022). Thus, in this research, we selected these models to assess the performance of our proposed model.

The experiment was conducted in a Python 3.7 environment. Based on previous research and experimental results (Meng et al., 2021), the kernel of the SVM model was Radial Basis Function (RBF). The tol, C and gamma of the SVM model were 0.001, 1 and

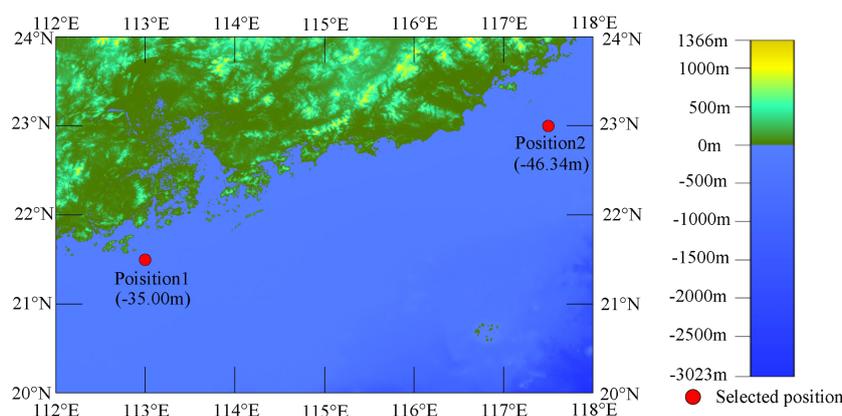


FIGURE 4
Distribution of the selected position in the South Sea of China.

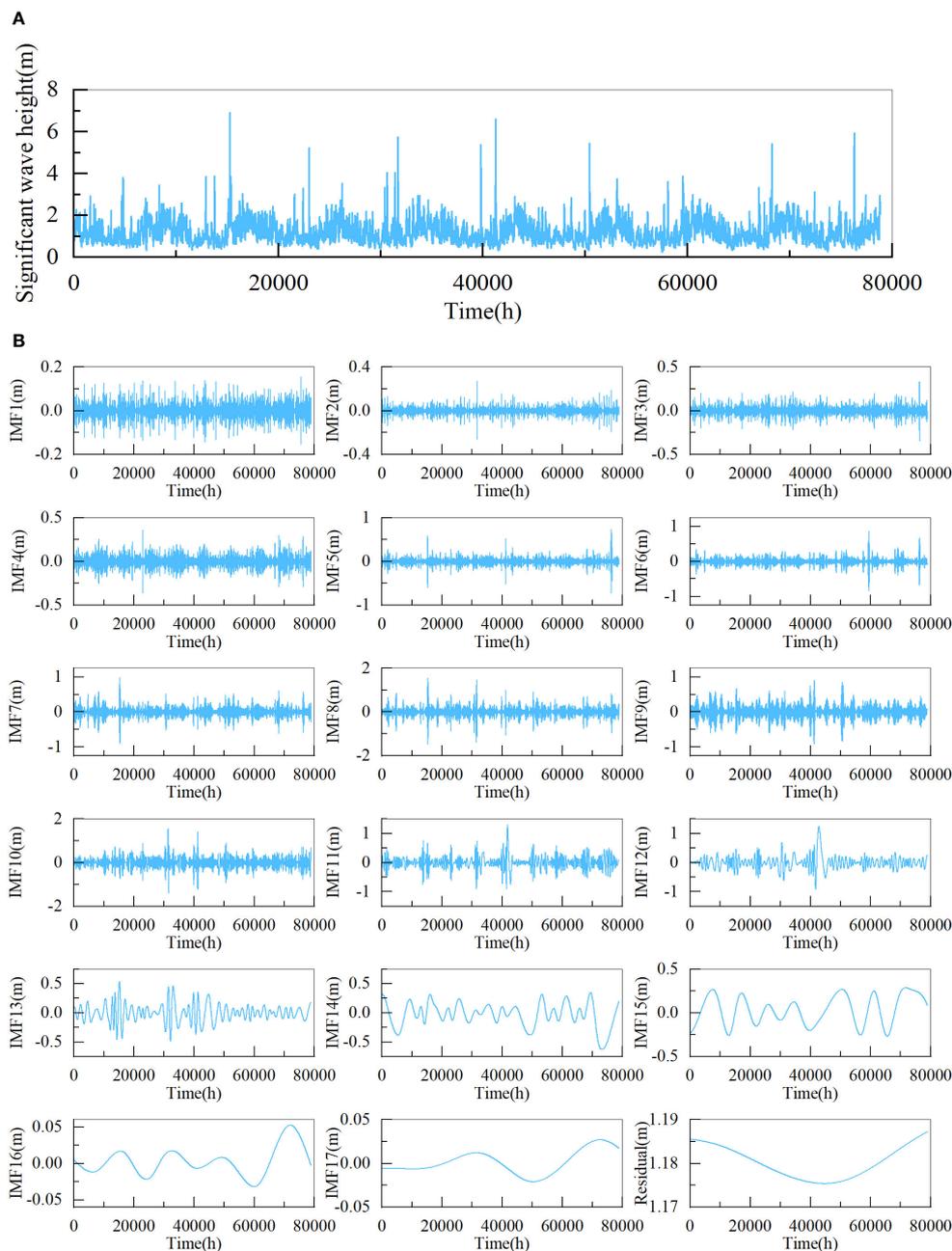


FIGURE 5

Significant wave height time series and decomposition results of dataset1 (A) significant wave height time series, (B) decomposition results.

0.1, respectively. The remaining parameters were set by default in the scikit-learn.

Based on previous experience and multiple experiments, the GRU model and the LSTM model were both divided into three layers input layer, hidden layer, and output layer. The hidden layer consisted of 32 neurons. The MSE was chosen as the loss function, and the Adam optimizer was used. Dropout was set to 0.2. The batch size was set to 32, and the number of epochs was set to 100. To prevent overfitting, early stopping was implemented with a patience of 10. The timestep was set to 7, meaning that the significant wave height of the current hour was considered to be related to the previous seven hours.

For the TimesNet model, the sequence length and label length were both set to 12 based on previous experience and multiple experiments. The loss function remained as MSE, and the optimizer was still Adam. The batch size and number of epochs were set to 32 and 7, respectively. Dropout was set to 0.2. To prevent overfitting, early stopping was implemented with a patience of 3. The number of heads was set to 8, the dimension of models was set to 16, and the attention factor was set to 5.

In order to facilitate the analysis, this study selected 400 data points to create comparative figures.

The time span covered was from August 18, 2017, 00:00:00 UTC to September 3, 2017, 15:00:00 UTC. It is worth noting that Typhoon Hato occurred during this time period.

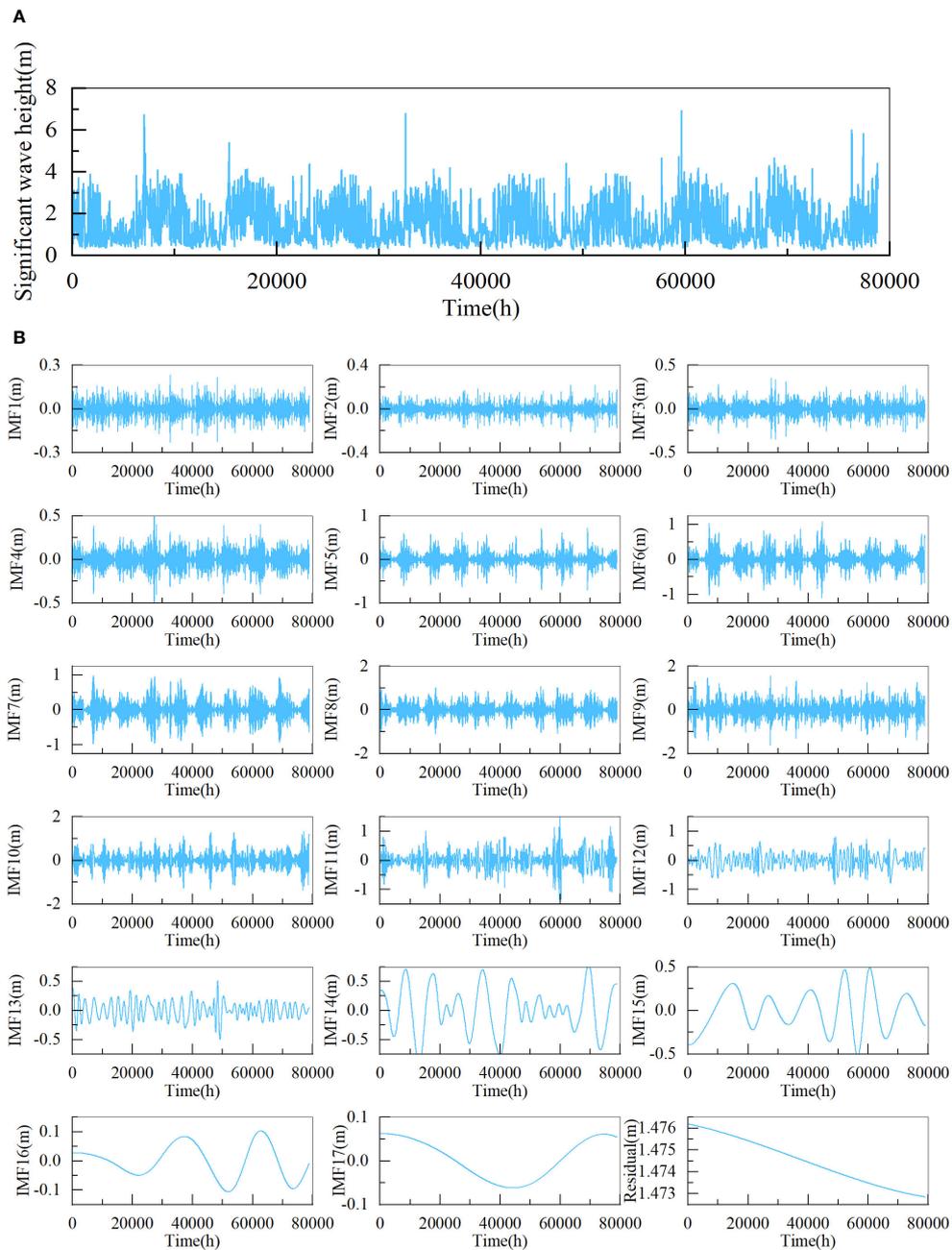


FIGURE 6 Significant wave height time series and decomposition results of dataset2 (A) significant wave height time series, (B) decomposition results.

4.2 Multi-step prediction results by different models

To evaluate the accuracy of the predictions, Tables 1, 2 record the error statistics of the results of the dataset by each model. Tables 3, 4 record the error statistics of the results of COIMF1, COIMF2 and COIMF3 by different models. Figures 9, 10 present the comparison between the predicted and actual values for both models. “1h” in Figures 9, 10 represents August 18, 2017, at 00:00:00 UTC. The occurrence time of Typhoon Hato in Figures 9, 10 is from 63h to 162h, which is between the dotted lines. Figure 11 shows

comparison of RMSE during the typhoon period in different prediction lead time.

4.3 Discussion of the prediction results

From Tables 1, 2, it can be observed that all models have shown good performance when the number of prediction steps is small, and the accuracy of the prediction results by the TimesNet model is higher compared to that of other models. Furthermore, as the number of steps increases, the difference becomes more

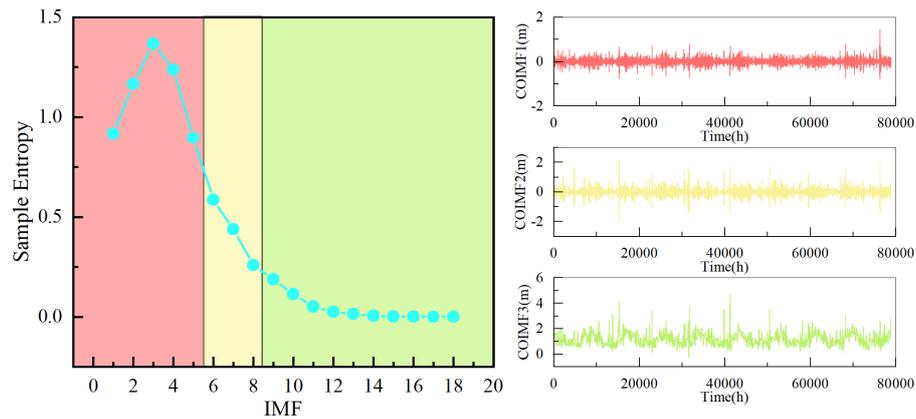


FIGURE 7
Sample entropy and clustering results of dataset1.

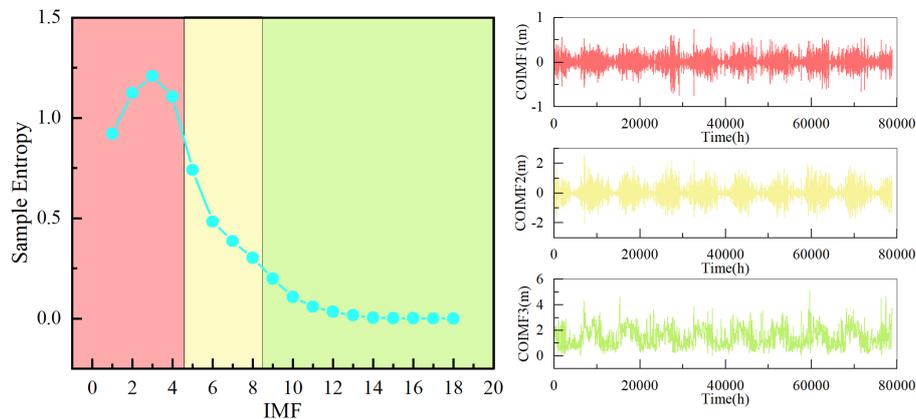


FIGURE 8
Sample entropy and clustering results of dataset2.

pronounced. For dataset1, when the number of prediction steps becomes twelve, R^2 of prediction results by the SVM model, the GRU model and the LSTM model are only 0.5118, 0.5363 and 0.5781, respectively. While R^2 of prediction results by the TimesNet model can reach 0.7414. This phenomenon occurs because the SVM model, the GRU model and the LSTM model all need to calculate the value of the current time before using it to predict the next value. This process introduces cumulative errors, which accumulate and become larger as the number of steps increases. On the other hand, the TimesNet model avoids cumulative errors by connecting each point in the time series with other points. Additionally, the TimesNet model can effectively capture intraperiod- and interperiod-variations, resulting in a minimal decrease in the accuracy of the predictions with an increasing number of steps. Wu et al. (2022) conducted research on both short-term and long-term forecasting for time series data and found that the TimesNet model outperformed the LSTM model in terms of prediction accuracy, which aligns with the findings of this study. Considering the advantages of the TimesNet model, this study chooses to conduct further research using the TimesNet model.

Observing Tables 1, 2, it is found that there has been a substantial improvement in the accuracy of the predictions after decomposing the original significant wave height with CEEMDAN and clustering IMFs. The decrease in accuracy of the CEEMDAN-TimesNet model is lower compared to the TimesNet model. For the 12-step ahead prediction results of dataset1, R^2 of the prediction results by the CEEMDAN-TimesNet model is improved by 21.85% compared to the prediction results by the TimesNet model. MSE , $RMSE$ and MAE are decreased by 62.66%, 38.86% and 34.57%, respectively. For the 12-step ahead prediction results of dataset2, R^2 of the prediction results by the CEEMDAN-TimesNet model is improved by 16.13% compared to the prediction results by the TimesNet model. MSE , $RMSE$ and MAE are decreased by 53.62%, 31.92% and 27.19%, respectively. The wave time series has non-stationary properties (Londhe and Panchang, 2006), which affects the prediction accuracy. Applying CEEMDAN and clustering improves prediction accuracy by reducing the non-stationarity of the original significant wave height.

A more detailed analysis is required for the prediction results of COIMFs for further enhancements to the CEEMDAN-TimesNet

TABLE 1 Error measures of multi-step predictions of dataset1 by different models.

Size of step	Index	SVM	GRU	LSTM	TimesNet	CEEMDAN	CEEMDAN-VMD
						-TimesNet	-TimesNet
2	R^2	0.9831	0.9789	0.9809	0.9862	0.9899	0.9993
	MSE	0.0047	0.0059	0.0053	0.0038	0.0028	0.0002
	RMSE	0.0687	0.0767	0.0730	0.0620	0.0530	0.0136
	MAE	0.0477	0.0542	0.0518	0.0355	0.0306	0.0092
4	R^2	0.9271	0.9210	0.9301	0.9544	0.9726	0.9984
	MSE	0.0204	0.0220	0.0195	0.0127	0.0077	0.0004
	RMSE	0.1427	0.1485	0.1398	0.1128	0.0875	0.0212
	MAE	0.1130	0.1077	0.1035	0.0671	0.0539	0.0144
6	R^2	0.8318	0.8361	0.8548	0.9095	0.9553	0.9965
	MSE	0.0470	0.0458	0.0406	0.0253	0.0125	0.0010
	RMSE	0.2167	0.2139	0.2014	0.1590	0.1118	0.0313
	MAE	0.1782	0.1551	0.1512	0.0965	0.0703	0.0215
8	R^2	0.7262	0.7362	0.7630	0.8577	0.9400	0.9940
	MSE	0.0765	0.0737	0.0662	0.0397	0.0168	0.0017
	RMSE	0.2766	0.2715	0.2573	0.1994	0.1295	0.0410
	MAE	0.2304	0.1967	0.1951	0.1218	0.0832	0.0281
10	R^2	0.6169	0.6330	0.6670	0.7999	0.9227	0.9912
	MSE	0.1071	0.1025	0.0930	0.0559	0.0216	0.0025
	RMSE	0.3272	0.3202	0.3050	0.2365	0.1470	0.0496
	MAE	0.2748	0.2320	0.2322	0.1451	0.0954	0.0336
12	R^2	0.5118	0.5363	0.5781	0.7414	0.9034	0.9882
	MSE	0.1365	0.1296	0.1179	0.0723	0.0270	0.0033
	RMSE	0.3694	0.3600	0.3434	0.2689	0.1644	0.0575
	MAE	0.3166	0.2610	0.2618	0.1649	0.1079	0.0386

TABLE 2 Error measures of multi-step predictions of dataset2 by different models.

Size of step	Index	SVM	GRU	LSTM	TimesNet	CEEMDAN	CEEMDAN-VMD
						-TimesNet	-TimesNet
2	R^2	0.9876	0.9859	0.9881	0.9916	0.9952	0.9997
	MSE	0.0095	0.0108	0.0091	0.0064	0.0037	0.0003
	RMSE	0.0972	0.1039	0.0956	0.0802	0.0609	0.0160
	MAE	0.0743	0.0750	0.0671	0.0456	0.3714	0.0113
4	R^2	0.9473	0.9463	0.9515	0.9687	0.9880	0.9992
	MSE	0.0403	0.0411	0.0371	0.0239	0.0092	0.0006
	RMSE	0.2007	0.2027	0.1926	0.1548	0.0957	0.0254
	MAE	0.1620	0.1471	0.1388	0.0905	0.0634	0.0181

(Continued)

TABLE 2 Continued

Size of step	Index	SVM	GRU	LSTM	TimesNet	CEEMDAN	CEEMDAN-VMD
						-TimesNet	-TimesNet
6	R^2	0.8748	0.8874	0.8917	0.9317	0.9779	0.9984
	MSE	0.0958	0.0861	0.0829	0.0523	0.0169	0.0012
	RMSE	0.3095	0.2935	0.2879	0.2286	0.1300	0.0352
	MAE	0.2548	0.2113	0.2079	0.1342	0.0877	0.0251
8	R^2	0.7925	0.8168	0.8149	0.8836	0.9604	0.9974
	MSE	0.1588	0.1402	0.1416	0.0891	0.0303	0.0020
	RMSE	0.3985	0.3744	0.3763	0.2985	0.1741	0.0449
	MAE	0.3302	0.2681	0.2718	0.1764	0.1154	0.0323
10	R^2	0.7059	0.7407	0.7287	0.8282	0.9318	0.9959
	MSE	0.2251	0.1984	0.2076	0.1315	0.0522	0.0031
	RMSE	0.4744	0.4455	0.4556	0.3626	0.2285	0.0557
	MAE	0.3957	0.3182	0.3289	0.2167	0.1495	0.0400
12	R^2	0.6045	0.6644	0.6396	0.7688	0.8928	0.9937
	MSE	0.3027	0.2569	0.2759	0.1770	0.0821	0.0049
	RMSE	0.5502	0.5068	0.5253	0.4207	0.2864	0.0697
	MAE	0.4636	0.3620	0.3791	0.2552	0.1858	0.0497

TABLE 3 Error measures of multi-step predictions of COIMFs for dataset1 by different models.

Size of step	Index	COIMF1 CEEMDAN -TimesNet	COIMF1 CEEMDAN -VMD -TimesNet	COIMF2 CEEMDAN -TimesNet	COIMF2 CEEMDAN -VMD -TimesNet	COIMF3 CEEMDAN -TimesNet
2	R^2	0.7579	0.9859	0.9984	0.9996	1.0000
	MSE	0.0027	0.0002	0.0001	0.0000	0.0000
	RMSE	0.0520	0.0126	0.0010	0.0049	0.0019
	MAE	0.0299	0.0081	0.0067	0.0034	0.0014
4	R^2	0.3450	0.9679	0.9917	0.9988	1.0000
	MSE	0.0073	0.0004	0.0005	0.0001	0.0000
	RMSE	0.0856	0.0189	0.0227	0.0087	0.0032
	MAE	0.0523	0.0123	0.0155	0.0058	0.0023
6	R^2	-0.0322	0.9326	0.9743	0.9971	0.9999
	MSE	0.0115	0.0008	0.0016	0.0002	0.0000
	RMSE	0.1074	0.0274	0.0398	0.0134	0.0050
	MAE	0.0658	0.0179	0.0274	0.0087	0.0035
8	R^2	-0.2721	0.8889	0.9384	0.9946	0.9997
	MSE	0.0142	0.0012	0.0038	0.0003	0.0001
	RMSE	0.1192	0.0352	0.0617	0.0184	0.0077
	MAE	0.0723	0.0228	0.0423	0.0119	0.0053
10	R^2	-0.3982	0.8434	0.8730	0.9915	0.9994

(Continued)

TABLE 3 Continued

Size of step	Index	COIMF1 CEEMDAN -TimesNet	COIMF1 CEEMDAN -VMD -TimesNet	COIMF2 CEEMDAN -TimesNet	COIMF2 CEEMDAN -VMD -TimesNet	COIMF3 CEEMDAN -TimesNet
	MSE	0.0156	0.0018	0.0079	0.0005	0.0001
	RMSE	0.1250	0.0418	0.0886	0.0230	0.0115
	MAE	0.0734	0.0264	0.0606	0.0151	0.0077
12	R ²	-0.4200	0.8028	0.7802	0.9871	0.9988
	MSE	0.0159	0.0022	0.0136	0.0008	0.0003
	RMSE	0.1260	0.0469	0.1166	0.0283	0.0163
	MAE	0.0717	0.0287	0.0795	0.0189	0.0107

TABLE 4 Error measures of multi-step predictions of COIMFs for dataset2 by different models.

Size of step	Index	COIMF1 CEEMDAN -TimesNet	COIMF1 CEEMDAN -VMD -TimesNet	COIMF2 CEEMDAN -TimesNet	COIMF2 CEEMDAN -VMD -TimesNet	COIMF3 CEEMDAN -TimesNet
2	R ²	0.5037	0.9782	0.9973	0.9995	1.0000
	MSE	0.0034	0.0001	0.0005	0.0001	0.0000
	RMSE	0.0581	0.0122	0.0216	0.0092	0.0042
	MAE	0.0341	0.0078	0.0143	0.0065	0.0029
4	R ²	-0.0572	0.9539	0.9848	0.9986	0.9999
	MSE	0.0072	0.0003	0.0027	0.0003	0.0001
	RMSE	0.0847	0.0177	0.0518	0.0159	0.0075
	MAE	0.0540	0.0119	0.0342	0.0110	0.0052
6	R ²	-0.3627	0.9214	0.9474	0.9971	0.9998
	MSE	0.0093	0.0005	0.0092	0.0005	0.0001
	RMSE	0.0962	0.0231	0.0961	0.0226	0.0120
	MAE	0.0621	0.0158	0.0633	0.0157	0.0082
8	R ²	-0.4127	0.8857	0.8698	0.9955	0.9994
	MSE	0.0096	0.0008	0.0229	0.0008	0.0004
	RMSE	0.0979	0.0279	0.1512	0.0280	0.0188
	MAE	0.0630	0.0191	0.0992	0.0197	0.0128
10	R ²	-0.3179	0.8498	0.7432	0.9936	0.9988
	MSE	0.0089	0.0010	0.0451	0.0011	0.0008
	RMSE	0.0946	0.0319	0.2123	0.0335	0.0276
	MAE	0.0608	0.0216	0.1403	0.0240	0.0186
12	R ²	-0.1811	0.8103	0.5745	0.9898	0.9975
	MSE	0.0080	0.0013	0.0746	0.0018	0.0015
	RMSE	0.0895	0.0359	0.2732	0.0423	0.0393
	MAE	0.0576	0.0241	0.1811	0.0303	0.0261

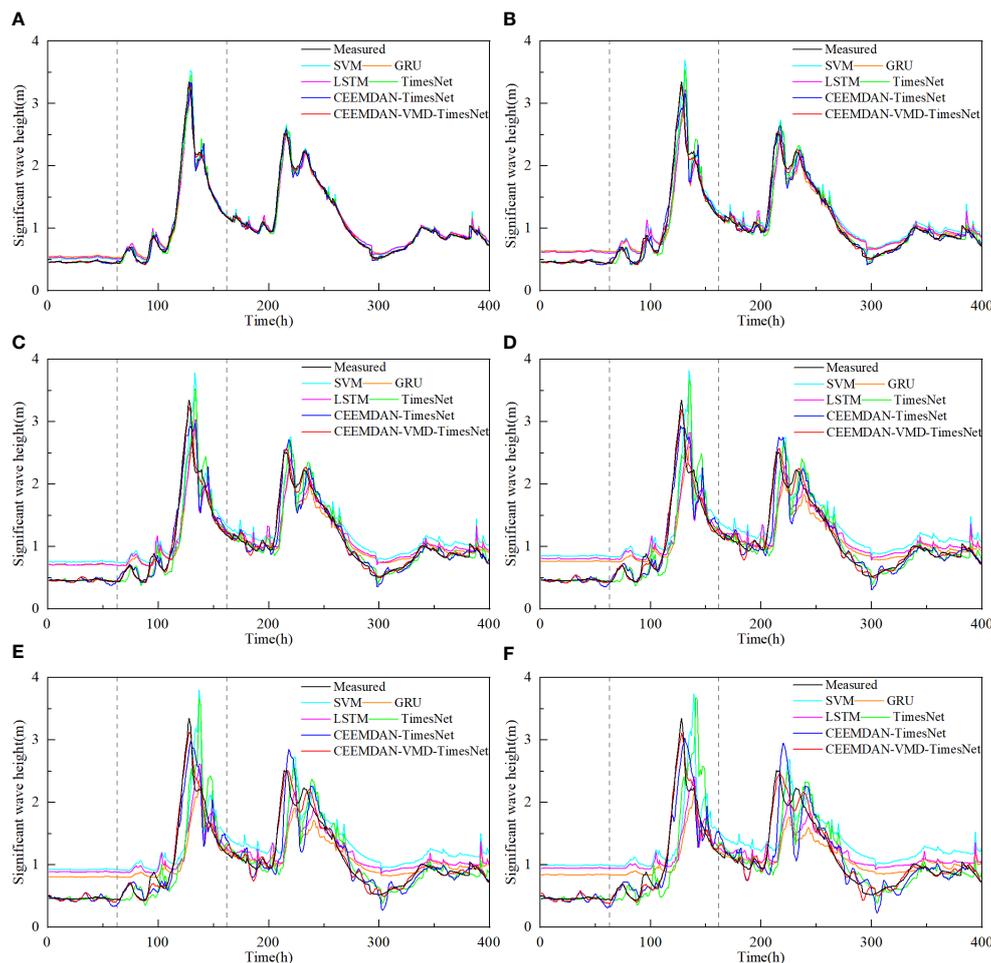


FIGURE 9 Predictions of significant wave height of dataset1 by different models for several future hours (A) 2 hours, (B) 4 hours, (C) 6 hours, (D) 8 hours, (E) 10 hours and (F) 12 hours.

model. From Tables 3, 4, it is evident that the main source of error is COIMF1 and COIMF2. The prediction results of COIMF1 are not satisfactory even for the small step-ahead prediction. As the prediction duration increases, the accuracy becomes worse. R^2 of the prediction results is even negative, which is unacceptable. COIMF2 performs better than COIMF1 in terms of prediction accuracy when the prediction steps are small. However, as the prediction steps increase, the prediction accuracy of COIMF2 also decreases noticeably. Surprisingly, COIMF3 maintains a relatively stable level, showing consistently outstanding performance in prediction. The reason for this phenomenon is that COIMF1 is formed by combining IMFs with high sample entropy, resulting in the highest complexity and non-stationarity, which introduces significant errors in the prediction results. COIMF3, on the other hand, is composed of IMF components with low complexity, making it stationary and highly accurate in prediction. COIMF2 consists of IMFs with moderate complexity. Therefore, COIMF2 falls between COIMF1 and COIMF3 in terms of prediction accuracy.

Given the poor prediction results of COIMF1 and COIMF2, COIMF1 and COIMF2 are decomposed by VMD again, while COIMF3 remains unchanged. From Tables 3, 4, it can be observed

that prediction results of COIMF1 and COIMF2 exhibit a significant improvement after decomposition, regardless of the number of prediction steps. The enhancement in the prediction of COIMF1 and COIMF2 has resulted in an overall improvement in prediction results of significant wave height. For the 12-step ahead prediction results of dataset1, R^2 of the prediction results by the CEEMDAN-VMD-TimesNet model is improved by 33.29% compared to the prediction results by the TimesNet model. MSE , $RMSE$ and MAE are decreased by 95.44%, 78.62% and 76.59%, respectively. For the 12-step ahead prediction results of dataset2, R^2 of the prediction results by the CEEMDAN-VMD-TimesNet model is improved by 29.25% compared to the prediction results by the TimesNet model. MSE , $RMSE$ and MAE are decreased by 97.23%, 83.43% and 80.53%, respectively. The accuracy of the CEEMDAN-VMD-TimesNet model far surpasses all the models mentioned earlier. The analysis leads to the conclusion that the CEEMDAN-VMD-TimesNet model has a clear advantage in prediction, especially when predicting multiple steps, significantly improving the prediction results.

Figures 9, 10 illustrate the performance of all models in predicting typhoons. When the number of prediction steps is small, all models perform well by capturing the overall

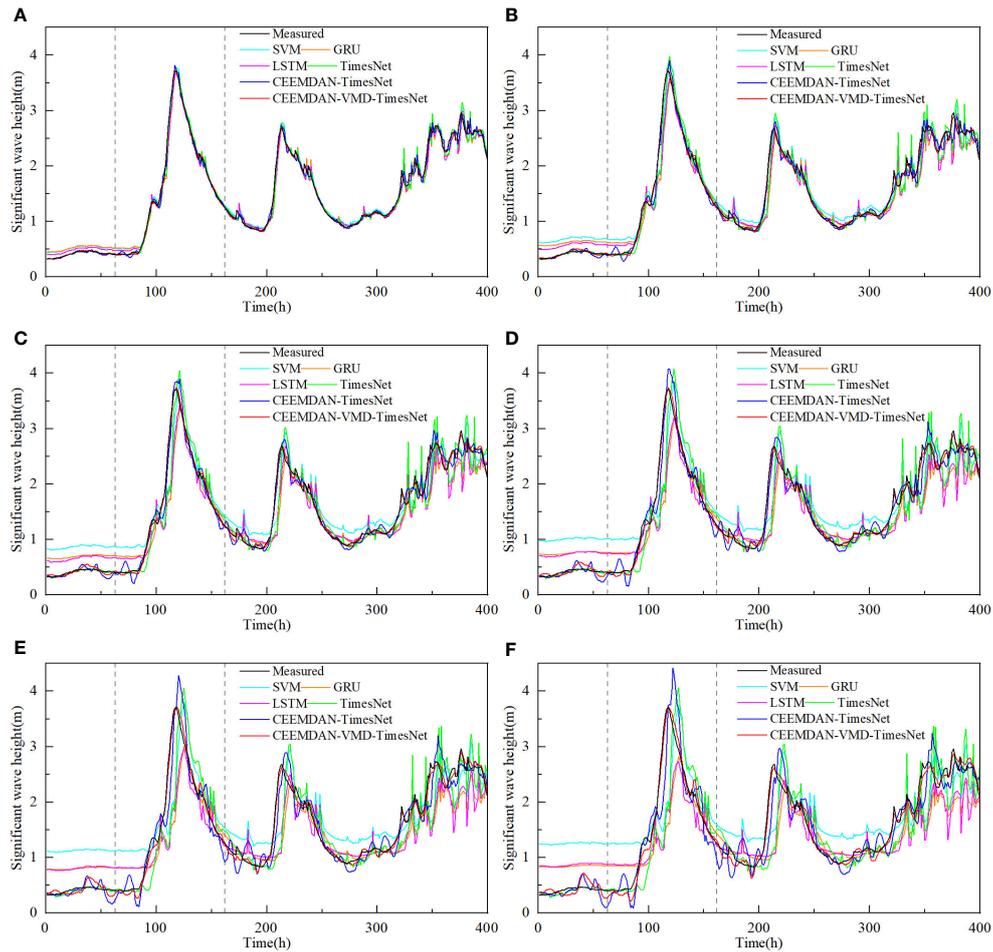


FIGURE 10 Predictions of significant wave height of dataset2 by different models for several future hours (A) 2 hours, (B) 4 hours, (C) 6 hours, (D) 8 hours, (E) 10 hours and (F) 12 hours.

characteristics of the original significant wave height and accurately predicting wave peaks and troughs without massive phase shifts. However, as the prediction duration increases, the SVM model, the GRU model, the LSTM model and the TimesNet model show noticeable phase shifts in the predictions of wave peaks and troughs due to the effects of non-stationarity. On the other hand,

the CEEMDAN-TimesNet model provides more accurate predictions compared to these models, greatly alleviating the occurrence of phase shifts. The CEEMDAN-VMD-TimesNet model stands out among all the models. The twelve-step-ahead prediction results by the CEEMDAN-VMD-TimesNet model remain highly consistent with the original significant wave height

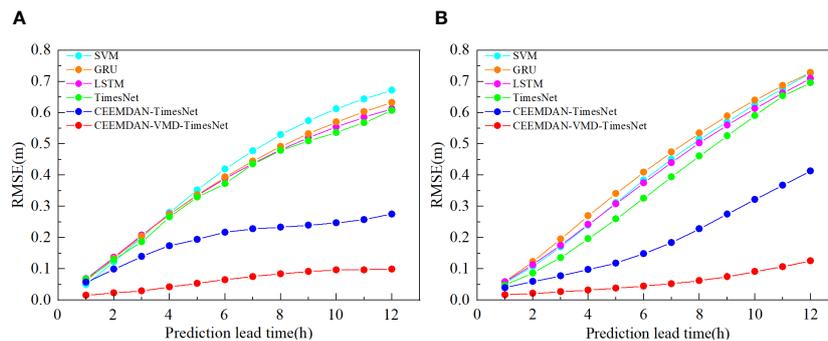


FIGURE 11 Comparison of RMSE during the typhoon period in different prediction lead time for dataset1 and dataset2 (A) dataset1, (B) dataset2.

in terms of both trend and extreme values, with minimal phase shifts. **Figure 11** shows the comparison of *RMSE* during the typhoon period in different prediction lead time between the CEEMDAN-VMD-TimesNet model and other models. Regardless of the prediction lead time, the CEEMDAN-VMD-TimesNet model consistently maintains the lowest *RMSE*. As the prediction lead time increases, *RMSE* grows at the slowest rate. When the prediction lead time is twelve, *RMSE* is much lower than other models, resulting in quite satisfactory accuracy. Thus, it can be inferred that the application of the CEEMDAN-VMD-TimesNet model for typhoon prediction is feasible.

5 Conclusion

Wave energy is a significant and environmentally friendly source of energy that has a wide range of applications. Accurate prediction of significant wave height plays a crucial role in wave energy generation. However, due to the complexity of significant wave height data, accurate prediction of significant wave height remains a challenge. In this study, a hybrid model called CEEMADN-VMD-TimesNet was proposed for predicting significant wave height. The significant wave height used was sourced from ERA5 of the European Centre for Medium-Range Weather Forecasts (ECMWF), and two positions in the South Sea of China were selected. The SVM model, the GRU model, the LSTM model, the TimesNet model, the CEEMADN-TimesNet model and the CEEMDAN-VMD-TimesNet model were built to predict significant wave height. By comparing the performance of these models, the following findings were obtained:

The SVM model, the GRU model and the LSTM model exhibit high accuracy for the small step-ahead prediction, but as the prediction steps increase, the prediction accuracy rapidly declines. Regardless of the number of prediction steps, the TimesNet model outperforms the SVM model, the GRU model and the LSTM model in terms of prediction accuracy, and this advantage becomes more pronounced for the large step-ahead prediction.

The non-stationarity of the original significant wave height is reduced by decomposing the original significant wave height with CEEMDAN and clustering IMFs. The CEEMADN-TimesNet model demonstrates superior accuracy in terms of prediction compared to the TimesNet model.

Components with high and moderate complexity are the main source of prediction errors. The accuracy of prediction results of significant wave height can be greatly improved by decomposing the components with high and moderate complexity using VMD. The CEEMADN-VMD-TimesNet model has the strongest prediction capability among the models used in this study.

The CEEMADN-VMD-TimesNet model can accurately predict trends and extreme values of significant wave height with minimal phase shifts during the Typhoon Hato period even for the large step-ahead prediction.

Although the proposed hybrid model achieves good accuracy, it still leaves questions for us to ponder. We did not compare the computation time of the secondary decomposition model with other models. The effect of the secondary decomposition on the

computation time needs to be further investigated. Besides, we did not compare the results of the CEEMADN-VMD-TimesNet model with the NWP model, which is a limitation of our study. In future, this will be the focus of our research.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

TD: Conceptualization, Methodology, Software, Writing – original draft. DW: Data curation, Funding acquisition, Supervision, Writing – review & editing. YL: Data curation, Software, Validation, Writing – review & editing. LS: Data curation, Software, Supervision, Writing – review & editing. XZ: Software, Validation, Writing – review & editing, Investigation.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The present work is sponsored by the National Natural Science Foundation of China (Grant No.42176166; No.41776024) and Major Project of Power Construction Corporation of China, Ltd (DJ-ZDXM-2022-28; KJXMWW-2023-01).

Acknowledgments

The significant wave height used was sourced from ERA5 of the European Centre for Medium-Range Weather Forecasts (ECMWF). We would like to acknowledge the ECMWF.

Conflict of interest

Authors YL and LS were employed by the company Municipal Engineering Corporation Limited. Author XZ was employed by the company Power China Kunming Engineering Corporation Limited.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Agrawal, J., and Deo, M. (2002). On-line wave prediction. *Mar. struct.* 15, 57–74. doi: 10.1016/S0951-8339(01)00014-4
- Ali, M., and Prasad, R. (2019). Significant wave height forecasting via an extreme learning machine model integrated with improved complete ensemble empirical mode decomposition. *Renewable and Sustain. Energy Rev.* 104, 281–295. doi: 10.1016/j.rser.2019.01.014
- Amunugama, M., Suzuyama, K., Manawasekara, C., Tanaka, Y., and Xia, Y. (2020). Typhoon-induced storm surge analysis with coast on different modelled forcing. *J. Japan Soc. Civil Eng. Ser. B3 (Ocean Engineering)* 76, I_210–I_215. doi: 10.2208/jsejoe.76.2_I_210
- Bertsekas, D. P. (1976). Multiplier methods A survey. *Automatica* 12, 133–145. doi: 10.1016/S1474-6670(17)67759-0
- Bethel, B. J., Sun, W., Dong, C., and Wang, D. (2022). Forecasting hurricane-forced significant wave heights using a long short-term memory network in the Caribbean sea. *Ocean Sci.* 18, 419–436. doi: 10.5194/os-18-419-2022
- Chang, H.-K., Liou, J.-C., Liu, S.-J., and Liaw, S.-R. (2011). Simulated wave-driven ANN model for typhoon waves. *Adv. Eng. Softw.* 42, 25–34. doi: 10.1016/j.advengsoft.2010.10.014
- Chen, T., Zhang, Q., Wu, Y., Ji, C., Yang, J., and Liu, G. (2018). Development of a wave-current model through coupling of fvcom and swan. *Ocean Eng.* 164, 443–454. doi: 10.1016/j.oceaneng.2018.06.062
- Deo, M. C., Jha, A., Chaphekar, A., and Ravikant, K. (2001). Neural networks for wave forecasting. *Ocean Eng.* 28, 889–898. doi: 10.1016/S0029-8018(00)00027-5
- Dragomiretskiy, K., and Zosso, D. (2013). Variational mode decomposition. *IEEE Trans. Signal Process.* 62, 531–544. doi: 10.1109/TSP.2013.2288675
- Duan, W. Y., Huang, L. M., Han, Y., and Huang, D. T. (2016). A hybrid emd-ar model for nonlinear and non-stationary wave forecasting. *J. Zhejiang University-SCIENCE A* 17, 115–129. doi: 10.1631/jzus.A1500164
- Fan, S., Xiao, N., and Dong, S. (2020). A novel model to predict significant wave height based on long short-term memory network. *Ocean Eng.* 205, 107298. doi: 10.1016/j.oceaneng.2020.107298
- Hao, W., Sun, X., Wang, C., Chen, H., and Huang, L. (2022). A hybrid emd-lstm model for non-stationary wave prediction in offshore China. *Ocean Eng.* 246, 110566. doi: 10.1016/j.oceaneng.2022.110566
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Las Vegas, NV, USA: IEEE) 770–778. doi: 10.1109/CVPR.2016.90
- Hestenes, M. R. (1969). Multiplier and gradient methods. *J. optim. Theory Appl.* 4, 303–320. doi: 10.1007/BF00927673
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London. Ser. A mathematic. Phys. Eng. Sci.* 454, 903–995. doi: 10.1098/rspa.1998.0193
- Li, X., Cao, J., Guo, J., Liu, C., Wang, W., Jia, Z., et al. (2022). Multi-step forecasting of ocean wave height using gate recurrent unit networks with multivariate time series. *Ocean Eng.* 248, 110689. doi: 10.1016/j.oceaneng.2022.110689
- Londhe, S., and Panchang, V. (2006). One-day wave forecasts based on artificial neural networks. *J. Atmos. ocean. Technol.* 23, 1593–1603. doi: 10.1175/JTECH1932.1
- Mahjoobi, J., and Mosabbe, E. A. (2009). Prediction of significant wave height using regressive support vector machines. *Ocean Eng.* 36, 339–347. doi: 10.1016/j.oceaneng.2009.01.001
- Meng, F., Song, T., Xu, D., Xie, P., and Li, Y. (2021). Forecasting tropical cyclones wave height using bidirectional gated recurrent unit. *Ocean Eng.* 234, 108795. doi: 10.1016/j.oceaneng.2021.108795
- Meng, F., Xu, D., and Song, T. (2022). Atdnns: An adaptive time–frequency decomposition neural network-based system for tropical cyclone wave height real-time forecasting. *Future Gen. Comput. Syst.* 133, 297–306. doi: 10.1016/j.future.2022.03.029
- Reikard, G., Robertson, B., and Bidlot, J.-R. (2015). Combining wave energy with wind and solar: Short-term forecasting. *Renewable Energy* 81, 442–456. doi: 10.1016/j.renene.2015.03.032
- Reikard, G., Robertson, B., and Bidlot, J.-R. (2017). Wave energy worldwide Simulating wave farms, forecasting, and calculating reserves. *Int. J. Mar. Energy* 17, 156–185. doi: 10.1016/j.ijome.2017.01.004
- Reikard, G., and Rogers, W. E. (2011). Forecasting ocean waves Comparing a physics-based model with statistical models. *Coast. Eng.* 58, 409–416. doi: 10.1016/j.coastaleng.2010.12.001
- Richman, J. S., and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. physiology-heart Circulatory Physiol.* 278, H2039–H2049. doi: 10.1152/ajpheart.2000.278.6.H2039
- Sharma, T., Bedi, J., Anand, A., and Aggarwal, A. (2023). Wave height prediction in maritime transportation using decomposition based learning. *IEEE Trans. Intell. Transport. Syst.* 1–10. doi: 10.1109/TITS.2023.3322192
- Soares, C. G., and Cunha, C. (2000). Bivariate autoregressive models for the time series of significant wave height and mean period. *Coast. Eng.* 40, 297–311. doi: 10.1016/S0378-3839(00)00015-6
- Song, T., Han, R., Meng, F., Wang, J., Wei, W., and Peng, S. (2022). A significant wave height prediction method based on deep learning combining the correlation between wind and wind waves. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.983007
- Song, T., Wang, J., Huo, J., Wei, W., Han, R., Xu, D., et al. (2023). Prediction of significant wave height based on emd and deep learning. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1089357
- Torres, M. E., Colominas, M. A., Schlotthauer, G., and Flandrin, P. (2011). “A complete ensemble empirical mode decomposition with adaptive noise,” in *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. (Prague, Czech Republic: IEEE) 4144–4147. doi: 10.1109/ICASSP.2011.5947265
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). “Attention is all you need,” in *Advances in neural information processing systems*. (Long Beach, California, USA: Neural Information Processing Systems Foundation, Inc. (NeurIPS)), 30. doi: 10.48550/arXiv.1706.03762
- Vijayan, L., Huang, W., Ma, M., Ozguven, E., Ghorbanzadeh, M., Yang, J., et al. (2023). Improving the accuracy of hurricane wave modeling in Gulf of Mexico with dynamically-coupled swan and adirc. *Ocean Eng.* 274, 114044. doi: 10.1016/j.oceaneng.2023.114044
- Wang, W., Tang, R., Li, C., Liu, P., and Luo, L. (2018). A bp neural network model optimized by mind evolutionary algorithm for predicting the ocean wave heights. *Ocean Eng.* 162, 98–107. doi: 10.1016/j.oceaneng.2018.04.039
- Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., and Long, M. (2022). Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv*. doi: 10.48550/arXiv.2210.02186
- Wu, Z., and Huang, N. E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv. adapt. Data Anal.* 1, 1–41. doi: 10.1142/S1793536909000047
- Wu, H., Xu, J., Wang, J., and Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* 34, 22419–22430. doi: 10.48550/arXiv.2106.13008
- Yang, Z., and Huang, W. (2022). Wave height estimation from x-band radar data using variational mode decomposition. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2022.3195675
- Zhao, L., Li, Z., Qu, L., Zhang, J., and Teng, B. (2023). A hybrid vmd-lstm/gru model to predict non-stationary and irregular waves on the east coast of China. *Ocean Eng.* 276, 114136. doi: 10.1016/j.oceaneng.2023.114136
- Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., et al. (2021). “Informer Beyond efficient transformer for long sequence time-series forecasting,” in *Proceedings of the AAAI conference on artificial intelligence*. (Palo Alto, California USA: AAAI), Vol. 35. 11106–11115. doi: 10.1609/aaai.v35i12.17325