



## Article

# Research on Long-Term Tidal-Height-Prediction-Based Decomposition Algorithms and Machine Learning Models

Wenchao Ban <sup>1</sup>, Liangduo Shen <sup>1,\*</sup>, Fan Lu <sup>2</sup>, Xuanru Liu <sup>3</sup> and Yun Pan <sup>2</sup>

<sup>1</sup> School of Marine Engineering Equipment, Zhejiang Ocean University, Zhoushan 316022, China; banwenchao@zjou.edu.cn

<sup>2</sup> School of Naval Architecture and Maritime, Zhejiang Ocean University, Zhoushan 316022, China; panyunhk@zjou.edu.cn (Y.P.)

<sup>3</sup> Zhoushan City Land Reserve Center, Zhoushan 316022, China; s20081500012@zjou.edu.cn

\* Correspondence: slduo@zjou.edu.cn

**Abstract:** Tidal-level prediction is crucial for ensuring the safety and efficiency of offshore marine activities, port and channel management, water transportation resource development, and life-saving operations. Although tidal harmonic analysis is among the most prevalent methods for predicting tidal water level fluctuations, it relies on extensive data, and its long-term prediction accuracy can be limited. To enhance prediction performance, this paper proposes a model that combines the variational mode decomposition (VMD) algorithm with the long short-term memory (LSTM) neural network. The initial step involves decomposing the original data using the VMD algorithm, followed by applying the LSTM to each decomposition component. Finally, all prediction results are superimposed and summed. The model is tested using the 2018 tidal time series data from the Lvsi station in Zhoushan City and the 2020 tidal time series data from the Ganpu station. The results are compared with those from the classical harmonic analysis model, the traditional machine learning model, and the decomposition-based machine learning method. The experimental outcomes demonstrate the superior predictive capabilities of the proposed model.

**Keywords:** tide level prediction; harmonic analysis; VMD algorithm; LSTM neural network; combined prediction model



**Citation:** Ban, W.; Shen, L.; Lu, F.; Liu, X.; Pan, Y. Research on Long-Term Tidal-Height-Prediction-Based Decomposition Algorithms and Machine Learning Models. *Remote Sens.* **2023**, *15*, 3045. <https://doi.org/10.3390/rs15123045>

Academic Editor: Chung-yen Kuo

Received: 21 April 2023

Revised: 8 June 2023

Accepted: 8 June 2023

Published: 10 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Tides [1,2] are among the most crucial physical elements of the ocean. This periodic rising and falling of seawater is due to the gravitational forces exerted on the ocean's surface by the movement of celestial bodies. Analyzing and understanding ocean tides allows us to comprehend the laws of change and internal dynamic mechanisms, which in turn enable us to project future tidal variations through tidal predictions. Tidal prediction is an essential technical means for identifying, researching, and effectively utilizing marine resources.

Given China's vast oceanic area and complex tidal variations, it is imperative to establish a detailed, accurate, convenient, and flexible marine tidal protection system. Analyzing and mastering tidal movements and improving the accuracy of tidal predictions [3,4] are crucial for ensuring the safety of marine activities, preventing and mitigating disasters, and harnessing tidal power generation. Therefore, making accurate ocean tide predictions is of utmost importance [5]. By thoroughly researching tides, we can utilize them effectively and mitigate potential hazards. Tidal harmonic analysis [6,7] involves calculating future tidal changes for a specific period by using actual tidal data from the predicted sea area and then superimposing this data to obtain future tidal variations, ultimately enabling tidal predictions. The classical method of tidal harmonic analysis was initiated by Kelvin in England, who proposed the technique in 1868 and invented the tidal prediction machine

based on this principle. Subsequently, Darwin further investigated tidal frequency division [8,9], followed by Dodson in the 1820s, who incorporated Brown's theory to increase the number of divisions and improve the accuracy of tidal summation analysis.

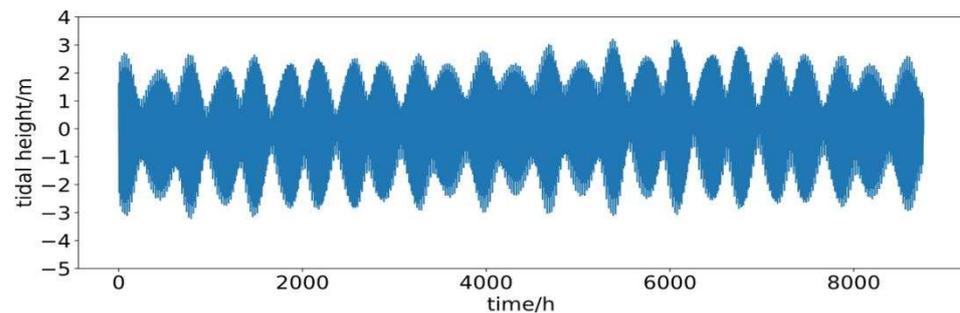
As generations of scholars have built upon their predecessors' work, the tidal harmonic analysis method has been continually refined [10,11], making significant contributions to tidal research in China. While the method can generally predict tidal height changes, it cannot guarantee the accuracy of tidal predictions. Unavoidable errors arise from factors such as the representation of astronomical image angles and the selection of tidal constituents, which often hinder our ability to achieve the desired level of accuracy in tidal forecast errors. In recent years, with the flourishing of deep learning technologies [12–17], novel methods have emerged for tide level prediction research. The primary distinction from traditional approaches is that machine learning time series forecasting methods do not necessitate large amounts of prior data or theoretical conjectures to determine model parameters. By learning from the input set's sample data, these methods can directly identify patterns among model arguments and data, enabling predictions. Their powerful data learning and information extraction capabilities facilitate high-precision prediction results. Consequently, machine learning techniques have effectively surpassed traditional forecasting methods in many areas involving modeling nonlinearity. For instance, Jung S used the LSTM [18] neural network to predict water level in a tidal river, Okwuashi O used SVM [19] to develop a tide level prediction model, and J. -C. Yin combined a discrete wavelet transform (DWT) with variable neural networks [20] to achieve real-time predictions for some tidal stations in the United States. The focus of this research is on time series machine learning models combined with decomposition algorithms [21–25]. It is acknowledged that improving the computational power of machine learning is challenging, but if the volatility and complexity in the original data can be reduced through decomposition algorithms, prediction difficulty in machine learning models can be diminished, ultimately enhancing the final prediction results. Currently, the more reliable decomposition algorithms include the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) algorithm [26,27] and the variational mode decomposition (VMD) [28,29] algorithm, which are seldom applied to the field of tidal prediction. The CEEMDAN decomposition algorithm is a step-by-step improvement based on the ensemble empirical mode decomposition (EEMD) [30–32] algorithm, which more effectively addresses the modal mixing phenomenon generated by the EEMD, reduces computational speed limitations, and can completely decompose the time series data from high to low frequencies into several intrinsic mode function (IMF) components [33], thereby decreasing data complexity. The VMD decomposition algorithm employs a non-recursive method for signal processing, decomposing the original signal by solving the constrained variable problem. This approach effectively avoids issues such as boundary effects and modal confusion, resulting in high-precision decomposition outcomes for complex data.

We recognize that the predictive capacity of a single model is ultimately limited [34–37], so we shift our focus to developing a combined model. In this paper, we choose to combine the VMD algorithm with the LSTM neural network to form a VMD-LSTM model. Based on two different sets of tide level data, we establish the classical tidal harmonic analysis, LSTM, SVM, CEEMDANN-LSTM, and VMD-LSTM models for comparative analysis in five groups, with the expectation [38–41] of providing new ideas and methods for high-accuracy tide level data prediction.

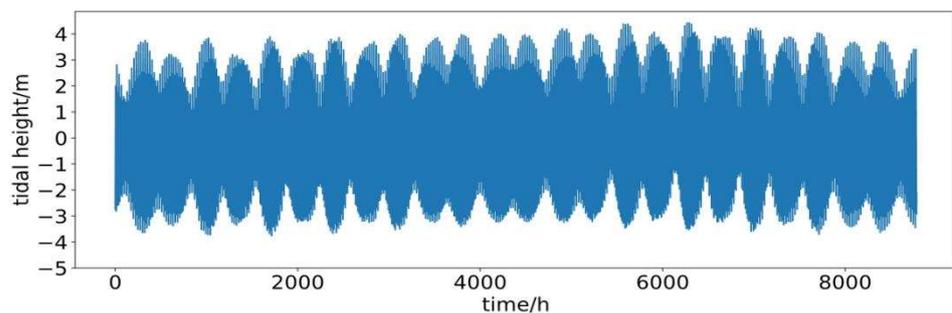
## 2. Materials and Methods

The data employed in this study were sourced from the National Marine Information Center (<https://www.cnss.com.cn/tide/> accessed on 12 December 2022) and include annual tidal information from the Lvsi station in 2018 and the Ganpu station in 2020. The sample sizes for these stations are 8760 and 8783, respectively, featuring an hourly sampling interval and tidal height measurements in meters. The observed tidal patterns are characterized as conventional semi-diurnal shallow tides with reciprocating flow properties.

To effectively compare the predictive performance of various methods, the first two months of the original data were designated as the training set, while the remaining ten months constituted the prediction set. The two-month data were employed to project the tidal trends throughout the subsequent ten months. Given the significant amount of data, it is impractical to visualize all of it, as this would result in a cluttered and unclear graph. Therefore, we have chosen to visualize tidal height data from periods with lesser tidal fluctuations (8100–8600 h) and greater tidal fluctuations (5500–6000 h), each spanning 500 h, for a more meaningful comparison. For the complete prediction set data, the error results will be directly analyzed in a tabular format. The unusual tidal height variations at both stations transpired in mid-August when the sun, moon, and earth were almost aligned, leading to the peak gravitational force on the seawater. This occurrence, in conjunction with the dominant southeast winds and tidal direction along the coast during summer, resulted in the abnormal tidal height shifts and an overall elevation in tidal height. The tidal time series data for the Lvsi and Ganpu stations in Zhoushan City are depicted in Figures 1 and 2.



**Figure 1.** Timing diagram of tidal height data at Lvsi Station.



**Figure 2.** Timing diagram of tidal height data at Ganpu Station.

2.1. Classical Tidal Harmonic Analysis Model

The classical tidal harmonic analysis model employed in this research utilized the T\_TIDE program, a widely recognized and effective tool for tidal analysis and prediction. In this study, a total of 36 subtidal tides were chosen for the analysis, with an emphasis on 11 major subtidal tides: K1, K2, M2, M4, M6, MS4, N2, O1, P1, Q1, and S2. These subtidal tides were deemed to be of critical importance due to their significant impact on tidal dynamics.

The harmonic analysis was conducted using observed real-time tidal height data, which allowed for the extraction of harmonic constants for the main sub-tides. These constants were then utilized for tidal prediction purposes. In the classical tidal harmonic analysis model, as described by Equation (1), each tidal constituent can be expressed as a trigonometric function. A linear superposition of these trigonometric functions is then used to represent the water level at any given time.

$$Z_{(t)} = S_0 + \sum_{j=1}^J [H_j \cos(\sigma_j t - g_j)] \tag{1}$$

$Z_{(t)}$  is the observed water level at time  $t$ .  $\sigma_j$ ,  $H_j$ , and  $g_j$  are the angular velocity, amplitude, and delay angle ( $^\circ$ ), respectively, corresponding to the  $j$ th subtide.  $j$  is the number of resolved subtides.  $J$  means the number of sub-tides,  $S_0$  is the mean sea level, and  $H_j$  in Equation (2) can be linearized using the variables  $a_j$  and  $b_j$ . In the equation,  $H_j = \sqrt{a_j^2 + b_j^2}$ ,  $g_j = \text{Arctan}(b_j/a_j)$ .

When doing linear regression, Equation (2) is required:

$$Z = AX \tag{2}$$

where  $Z$  is the observed water level matrix,  $A$  is the known coefficient matrix, and  $X$  is the unknown parameter matrix to be solved, as shown in Equations (3) and (4):

$$Z = \begin{bmatrix} Z(t_1) \\ Z(t_2) \\ \dots \\ Z(t_N) \end{bmatrix}, X = \begin{bmatrix} S_0 \\ a_1 \\ \dots \\ a_j \\ b_1 \\ \dots \\ b_j \end{bmatrix} \tag{3}$$

$$A = \begin{bmatrix} 1 & \cos\sigma_1 t_1 & \dots & \cos\sigma_j t_1 & \sin\sigma_1 t_1 & \dots & \sin\sigma_j t_1 \\ 1 & \cos\sigma_1 t_2 & \dots & \cos\sigma_j t_2 & \sin\sigma_1 t_2 & \dots & \sin\sigma_j t_2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & \cos\sigma_1 t_N & \dots & \cos\sigma_j t_N & \sin\sigma_1 t_N & \dots & \sin\sigma_j t_N \end{bmatrix} \tag{4}$$

According to the least squares method, the solution of  $X$  is:

$$X = (A^T A)^{-1} A^T Z \tag{5}$$

By employing the classical tidal harmonic analysis model, the research aimed to achieve a comprehensive understanding of tidal dynamics at the study site. This approach effectively accounted for the influences of various tidal constituents on water level changes over time. Consequently, the methodology provided a robust foundation for reliable tidal predictions, which are crucial for efficient coastal zone management, navigation, and infrastructure development.

## 2.2. CEEMDAN Model

The CEEMDAN decomposition algorithm can effectively decompose the highly fluctuating raw time series data from high to low frequencies into several intrinsic mode function (IMF) components. This helps improve the prediction accuracy by reducing the data fluctuations. The most significant advantage of the CEEMDAN decomposition algorithm is its completeness; the decomposed data can be aggregated back into the raw data without loss, and the efficiency of the CEEMDAN decomposition algorithm is very high, providing rapid computation. When decomposing the tidal data with CEEMDAN, the white noise of the standard normal distribution is added to the observation data sequence  $x(t) = (x_1, x_2, \dots, x_t)$  in  $\lambda (\lambda = 1, 2, \dots, I)$  times  $v^\lambda(t)$ , where  $T$  is the length of the tidal data,  $x_t$  is the value of tidal at time  $t$ ,  $I$  is the number of noise additions, and the reconstructed observation data sequence  $x^\lambda(t)$  is expressed in Equation (6).

$$x^\lambda(t) = x(t) + v^\lambda(t) \quad (6)$$

The first EMD solution of  $x^\lambda(t)$  yields the components  $IMF_1$  and its residuals  $m_1(t)$ , which are expressed in Equations (7) and (8).

$$IMF_1 = \frac{1}{I} \times \sum_{\lambda=1}^I IMF_1^\lambda \quad (7)$$

$$m_1(t) = x(t) - IMF_1 \quad (8)$$

Add  $v^\lambda(t)$  to  $m_1(t)$  and perform EMD decomposition to obtain  $IMF_2$  and  $m_2(t)$ ; add  $v^\lambda(t)$  to  $m_{n-1}(t)$  after decomposition  $n - 1$  time and perform EMD decomposition for  $m_{n-1}^\lambda(t)$  for the  $n$ th time to obtain  $IMF_n$  and  $m_n(t)$  as Equations (9) and (10), respectively.

$$IMF_n = \frac{1}{I} \times \sum_{\lambda=n-1}^I IMF_n^\lambda \quad (9)$$

$$m_n(t) = m_{n-1}(t) - IMF_n \quad (10)$$

After  $J$  decomposition,  $m_n(t)$  cannot be decomposed by EMD continuously, and the tidal observation data series  $x(t)$  is expressed in Equation (11).

$$x(t) = \sum_{n=1}^J IMF_n + m_J(t) \quad (11)$$

Compared to EMD, CEEMDAN combines positive and negative white noise added to the original data. The white noise property is utilized to amplify the uncorrelated degree of the difficult-to-separate modes, enabling the extraction of these modes and reducing the reconstruction error through the cancellation of positive and negative white noise.

Figures 3 and 4 display the decomposition results of the CEEMDAN algorithm, which was implemented using the PyEMD package in Python 3.8. The original tidal height time series data from the records of the Lvsi station and the Ganpu station in Zhoushan were fully decomposed according to high and low frequencies. The results of each component were predicted using the LSTM model and then summed and reconstructed. The flow chart is shown in Figure 5.

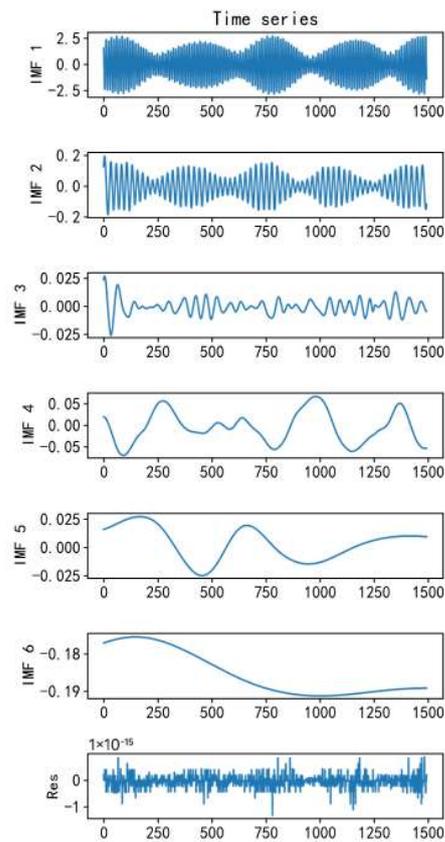


Figure 3. Decomposition results (Lvsi station).

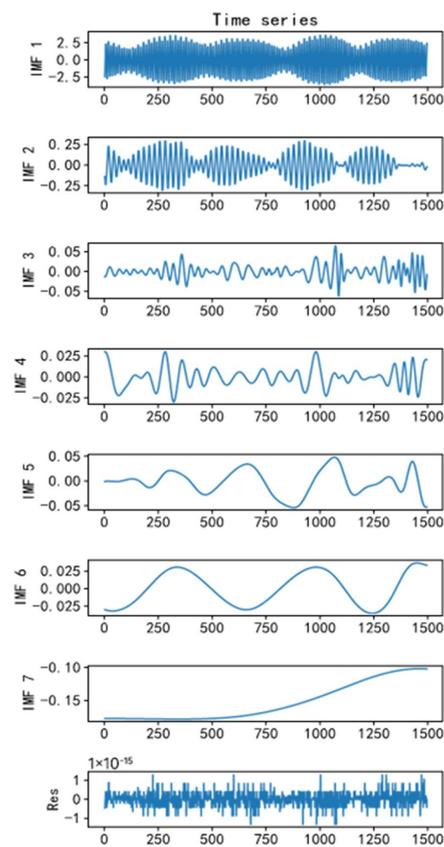


Figure 4. Decomposition results (Ganpu station).

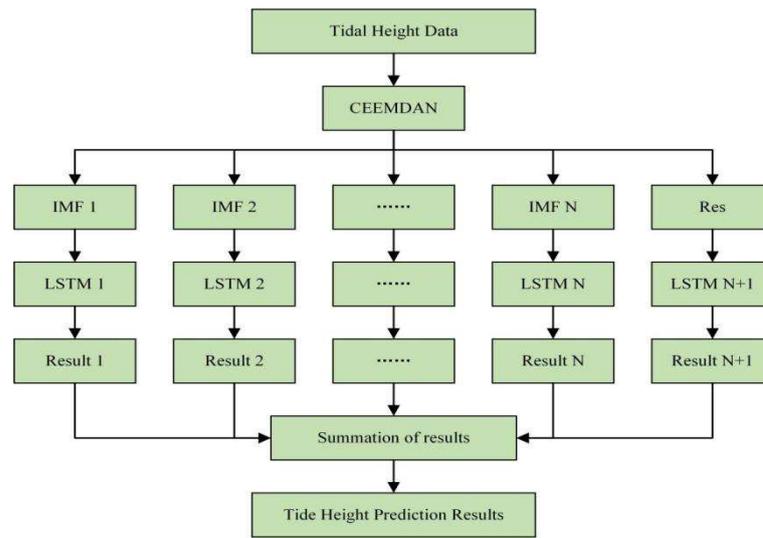


Figure 5. CEEMDAN-LSTM flow chart.

2.3. LSTM Model

LSTM is better equipped to handle time series data, as it processes a long span of time to capture the internal relationships within the data, thereby achieving long-term memory. Each LSTM unit consists of three gates: the forget gate, the input gate, and the output gate. The structure of a single neuron is shown in Figure 6.

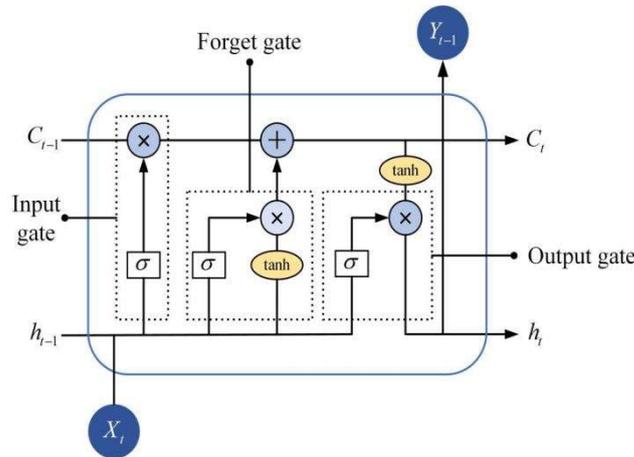


Figure 6. Neuronal structure.

$$i_t = \sigma_i(W_i \times [h_{t-1}, x_t] + b_i) \tag{12}$$

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \tag{13}$$

$$C_t = f \times C_{t-1} + i_t \times \tilde{C}_t \tag{14}$$

$$\tilde{C}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c) \tag{15}$$

$$O_t = \sigma(w_0 \times [h_{t-1}, x_t] + b_0) \tag{16}$$

$$h_t = O_t * \tanh(C_t) \quad (17)$$

In the above equation,  $i_t$ ,  $f_t$ ,  $c_t$ , and  $O_t$  represent the oblivion gate, the input gate, the cell state and the output gate, respectively;  $w_i$ ,  $w_f$ ,  $w_c$ , and  $w_o$  all imply different weight matrices; and  $b_f$ ,  $b_i$ ,  $b_c$ , and  $b_o$  represent the corresponding bias vectors. The activation function is  $\tanh$ ,  $h_t$  is the output at moment  $t$ ,  $x_t$  is the input at moment  $t$ , and  $\tilde{C}_t$  is the new value of the candidate vector update.

The LSTM models are implemented using Python 3.8. Our LSTM network consists of three hidden layers, with each hidden layer followed by a dropout layer to prevent overfitting. The network has a total of 300 dimensions and uses “relu” as the activation function and “Adam” as the optimizer and takes six steps to test the value of one step, for a total of 100 epochs. These crucial parameters are determined through continuous experimentation, testing numerous combinations of these parameters with different values until a stable network and optimal training options are obtained. To ensure better comparative experiments, we maintain these optimal parameter values without change.

#### 2.4. SVM Model

Support vector machines (SVM) is a widely used machine learning technique for regression analysis and statistical classification. The primary goal of SVM is to find the optimal hyperplane that maximizes the margin between different classes or minimizes the regression error. The SVM algorithm maps input vectors to a high-dimensional space using kernel functions, where it seeks to identify the optimal partitioning hyperplane and two parallel hyperplanes that best separate the data.

The total error of the SVM is minimized when the margin between the parallel hyperplanes is maximized. In this experimental comparison, stochastic gradient optimization is employed to optimize the model parameters, ensuring a more efficient convergence towards the optimal solution. MATLAB is used for the modeling process, with the following settings: KernelFunction is set to “linear”, KernelScale is set to 1, and the BoxConstraint range is defined as  $[1 \times 10^{-3}, 1 \times 10^3]$ .

The chosen settings for the experiment (linear KernelFunction, KernelScale of 1, and BoxConstraint range of  $[1 \times 10^{-3}, 1 \times 10^3]$ ) ensure a balance between model complexity and generalization performance. By using stochastic gradient optimization and these specific settings, the experimental comparison aims to showcase the effectiveness of SVM in solving regression and classification problems, highlighting its potential applications in various domains.

#### 2.5. VMD Algorithm

VMD is a novel time-frequency decomposition algorithm that aims to decompose the multi-component signal in the data into multiple single-component amplitude-modulated frequency signals. This approach helps to avoid issues such as pseudo-component problems and endpoint effects that may be encountered, resulting in more effective processing of nonlinear signals. The specific decomposition process is as follows: Initialize  $u_k^1$ ,  $w_k^1$ , and  $\lambda_k^1$  and  $n = 0$ .

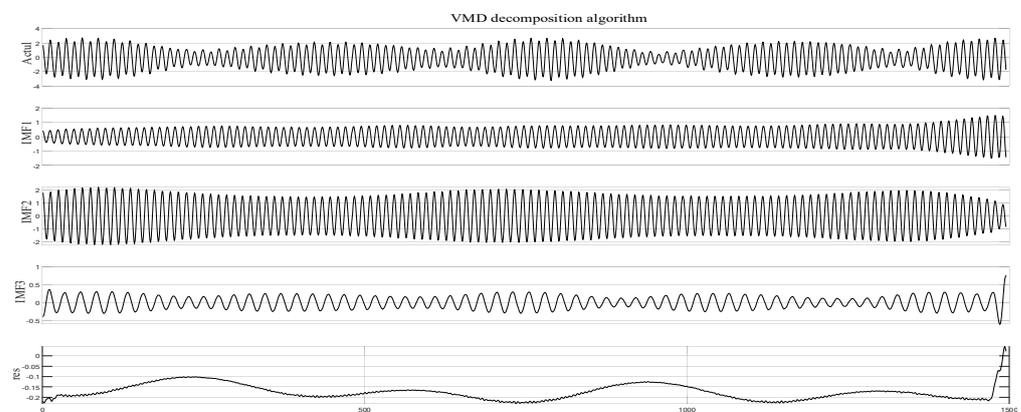
- (1)  $n = n + 1$ ; enter the loop.
- (2) Update according to the update formula of  $u_k$  and  $w_k$  until the number of decompositions is  $k$ .
- (3) Update  $\lambda$  according to the update formula of  $\lambda$ .
- (4) Given the accuracy  $\varepsilon$ , if the stopping condition is satisfied,

$$\sum_k \frac{\|u_k^{n+1} - u_k^n\|_2^2}{\|u_k^n\|_2^2} < \varepsilon \quad (18)$$

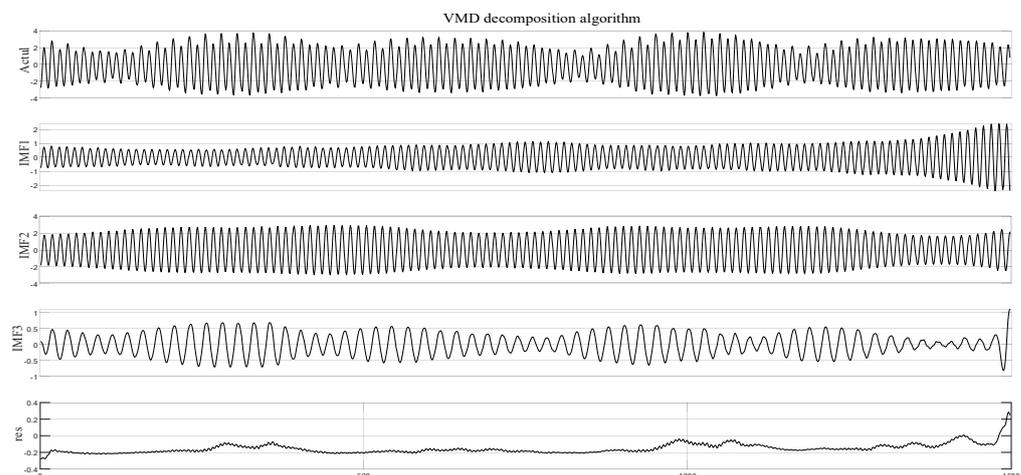
Terminate the cycle; otherwise, proceed to step (2) to continue the cycle.

In the above description,  $u_k$  represents the decomposed single-component AMF signal,  $w_k$  denotes the center frequency of each single-component AM/FM signal,  $\lambda$  is the Lagrangian multiplier, and  $n$  stands for the number of iterations.

When utilizing VMD to decompose the data, it is essential to set the parameters of the VMD algorithm (penalty factor  $\alpha$  and modal component  $K$ ) in advance. We reconstruct the corresponding time series by summing the sub-time series for each value of  $K$  and  $\alpha$ . When the reconstructed time series exhibits a perfect correlation with the original time series, the values of  $K$  and  $\alpha$  that yield the best efficiency for the VMD-based computational model are chosen as the optimal values. In this experiment, the penalty factor of the VMD decomposition algorithm is set to 2000, the specified number of decomposition modes is set to 4, and the articulation process of the LSTM neural network is identical to that of the CEEMDAN decomposition algorithm. The specific decomposition diagrams for the Lvsi and Ganpu stations are shown in Figures 7 and 8.



**Figure 7.** Splitting result based on variational modal decomposition (VMD) algorithm (Lvsi station).



**Figure 8.** Splitting result based on variational modal decomposition (VMD) algorithm (Ganpu station).

## 2.6. Model Evaluation Indexes

To emphasize the performance differences among the compared models, we opted to use three error metrics, namely root mean square error (RMSE), mean absolute error (MAE), and goodness of fit ( $R^2$ ), to prevent chance errors caused by individual metrics. The units of RMSE and MAE are meters, and smaller values indicate higher prediction

accuracy. Moreover, the closer the value of  $R^2$  is to 1, the better the fit. The expressions are shown in Equations (19)–(21):

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - z_i)^2} \quad (19)$$

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - z_i| \quad (20)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - z_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (21)$$

where  $m$  is the number of test samples, and  $y_i$  and  $z_i$  are the original and predicted values of the data, respectively.

### 3. Results

We initially chose to compare the predicted results of conventional tidal wave height. The conventional tidal wave height prediction results of the classical tidal harmonic analysis, SVM, LSTM, CEEMDAN-LSTM, and VMD-LSTM models for the Lvsi and Ganpu stations are displayed in Figures 9 and 10, respectively, while the error analysis results are shown in Tables 1 and 2. The raw data in the figures represent the 500 h of conventional tidal wave height data in the sample data range 8100 h–8600 h. It is evident that all five models perform well in conventional tidal wave height prediction, with high graphical fit and no significant errors, which means that all models can achieve the desired results in conventional tidal wave height prediction.

After comparing the error analysis metrics in Tables 1 and 2, it is concluded that the VMD-LSTM model exhibits smaller errors and better predictions. At both sites, the RMSE and MAE metrics of VMD-LSTM are improved by more than 50% compared to the comparison models, and the  $R^2$  is closer to 1. We believe that the conventional tidal wave height prediction is satisfactory for each model, and even if we use the more advanced VMD decomposition algorithm to enhance the prediction accuracy, the significant numerical improvement in the accuracy index is also a valuable addition.

**Table 1.** Error analysis results of conventional tidal height prediction at Lvsi station.

	Harmonic Analysis	SVM	LSTM	CEEMDAN-LSTM	VMD-LSTM
RMSE (m)	0.284	0.143	0.144	0.134	0.068
MAE (m)	0.228	0.119	0.119	0.114	0.058
$R^2$	0.956	0.989	0.979	0.980	0.992

**Table 2.** Error analysis results of conventional tidal height prediction at Ganpu station.

	Harmonic Analysis	SVM	LSTM	CEEMDAN-LSTM	VMD-LSTM
RMSE (m)	0.304	0.144	0.143	0.147	0.058
MAE (m)	0.245	0.119	0.112	0.124	0.048
$R^2$	0.977	0.995	0.995	0.995	0.998

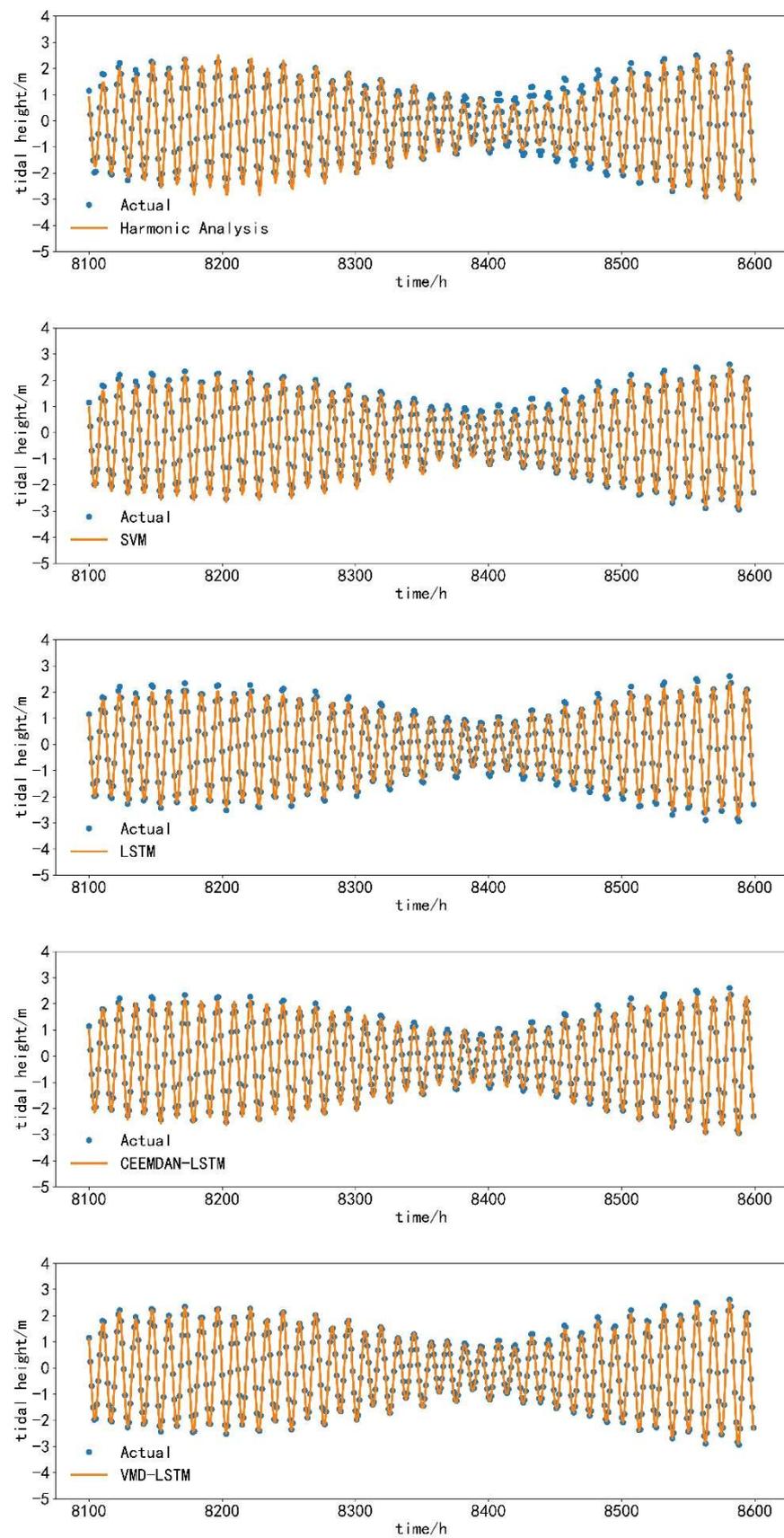


Figure 9. Comparison of prediction results of conventional tide level height (Lvsi station).

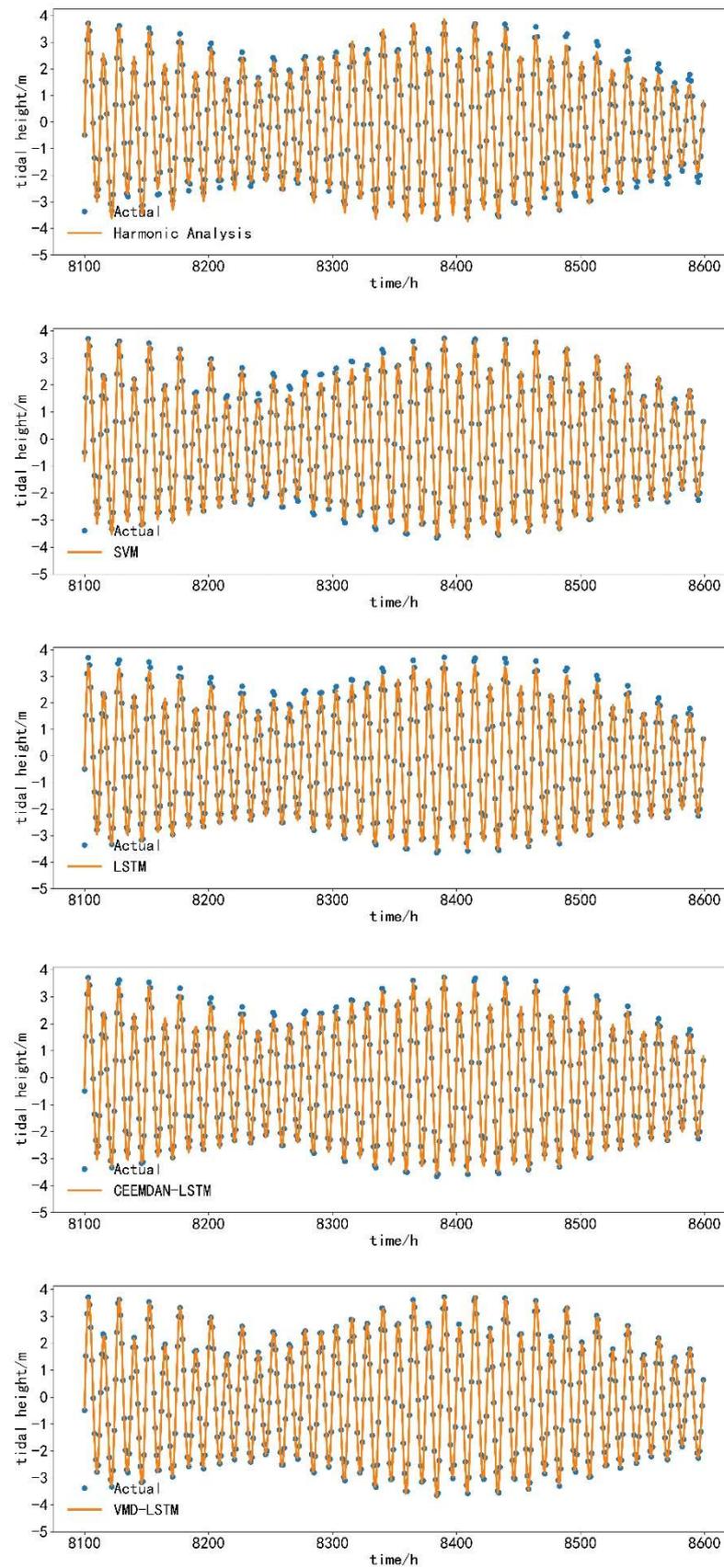


Figure 10. Comparison of prediction results of conventional tide level height (Ganpu station).

To better compare the performance of each model, we chose to continue analyzing the prediction results of the five models in abnormal tidal wave height conditions. We selected the abnormal tidal wave height data of 500 h from 5500 h to 6000 h in the original data for comparison. The prediction results of the five models in the abnormal tidal wave height are presented in Figures 11 and 12, and the error analysis results are shown in Tables 3 and 4, while the total tidal height prediction error results are presented in Tables 5 and 6.

In the comparison of anomalous tidal height prediction results, we can observe that the VMD-LSTM model's performance is superior. The four compared models have significant errors at the crests and troughs, with varying degrees of errors, but the VMD-LSTM model overcomes this problem. Its graphical fit is extremely high, and there are no obvious errors. After analyzing the errors of each model, the VMD-LSTM model shows a 36.5% improvement in both the lowest RMSE error metric and the lowest MAE error metric, compared to the four comparison models. The  $R^2$  is above 0.99 for both stations.

When we compare the error analysis indicators of the entire ten-month forecast set, we can also conclude that the comprehensive prediction performance of the VMD-LSTM model remains superior.

**Table 3.** Error analysis results of abnormal tidal height prediction at Lvsi station.

	Harmonic Analysis	SVM	LSTM	CEEMDAN-LSTM	VMD-LSTM
RMSE (m)	0.428	0.509	0.361	0.189	0.120
MAE (m)	0.380	0.495	0.335	0.156	0.102
$R^2$	0.892	0.854	0.924	0.980	0.992

**Table 4.** Error analysis results of abnormal tidal height prediction at Ganpu station.

	Harmonic Analysis	SVM	LSTM	CEEMDAN-LSTM	VMD-LSTM
RMSE (m)	0.594	0.465	0.288	0.178	0.072
MAE (m)	0.5	0.428	0.224	0.141	0.060
$R^2$	0.932	0.959	0.983	0.994	0.998

**Table 5.** Error analysis results of total tidal height prediction at Lvsi station.

	Harmonic Analysis	SVM	LSTM	CEEMDAN-LSTM	VMD-LSTM
RMSE (m)	0.372	0.339	0.260	0.171	0.091
MAE (m)	0.299	0.288	0.215	0.138	0.075
$R^2$	0.925	0.945	0.967	0.987	0.996

**Table 6.** Error analysis results of total tidal height prediction at Ganpu station.

	Harmonic Analysis	SVM	LSTM	CEEMDAN-LSTM	VMD-LSTM
RMSE (m)	0.504	0.323	0.224	0.147	0.067
MAE (m)	0.402	0.266	0.164	0.117	0.055
$R^2$	0.935	0.976	0.989	0.995	0.998

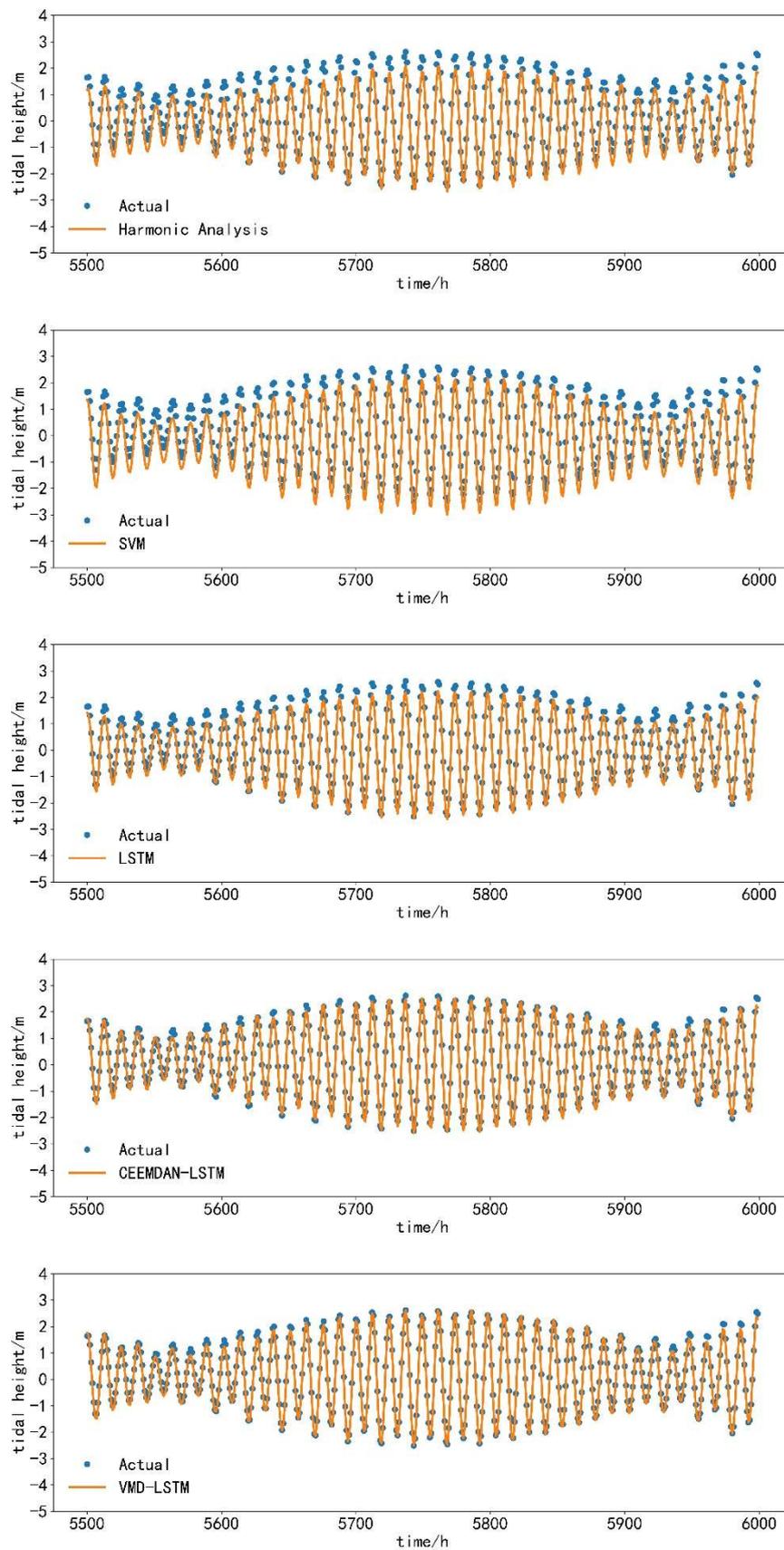


Figure 11. Comparison of abnormal tidal height prediction results (Lvsi station).

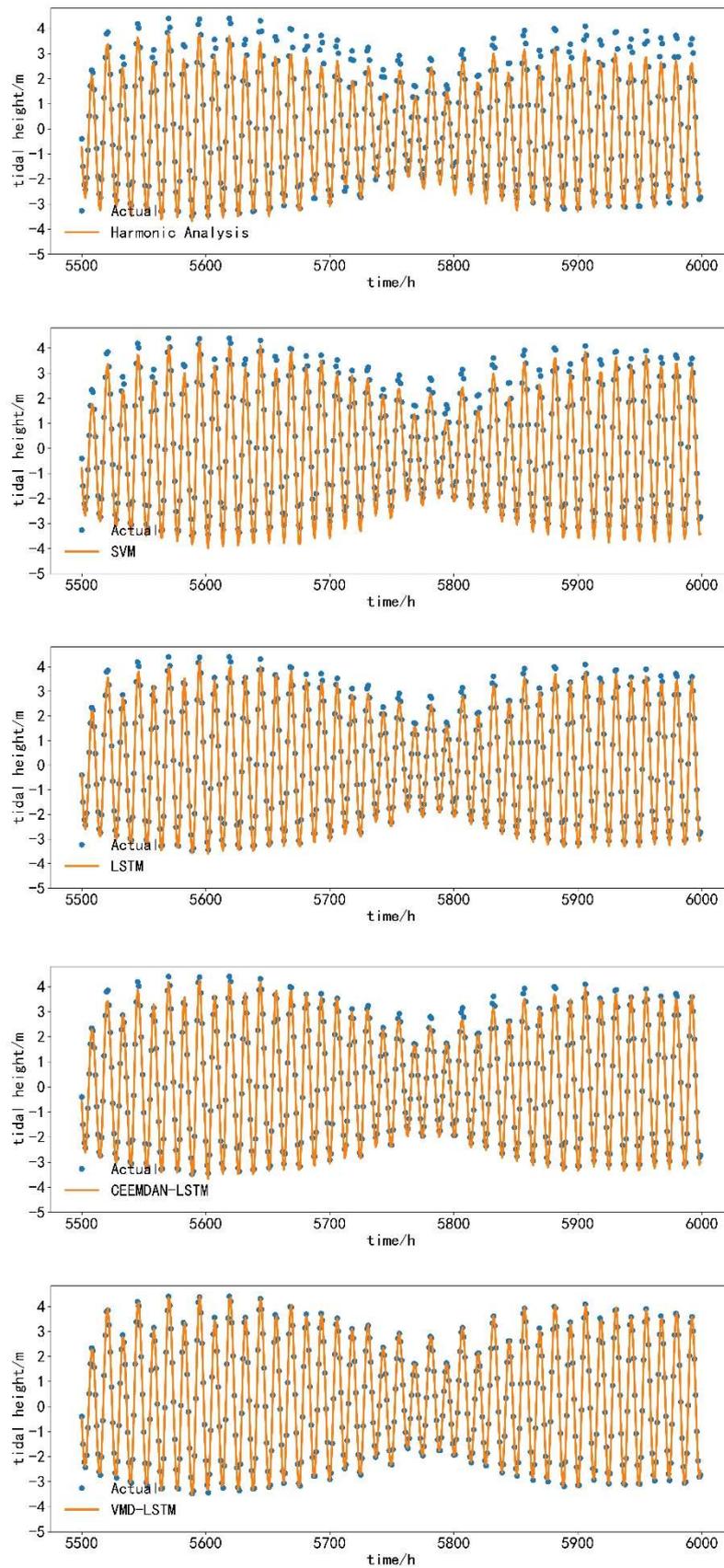


Figure 12. Comparison of prediction results of abnormal tide level height (Ganpu station).

#### 4. Discussion

Taking into account all the experimental results, we believe that the VMD decomposition algorithm is indeed a significant improvement and an ideal method for tidal height prediction accuracy. Comparing various methods, the classical tidal harmonic analysis method has high requirements for each data parameter. As Pan [1] and others have pointed out, the prediction accuracy of this method can be affected by numerous factors, such as the choice of subtidal zone, the number of subtidal zones, topographic effects, and more.

The CEEMDAN algorithm has been used successfully by Zhang [26] and others; however, although the CEEMDAN algorithm has high decomposition integrity, the cumulative error increases relatively due to the high number of low-frequency components generated by this decomposition method. Despite the good decomposition effect, it is still lacking in processing anti-noise and outlier fluctuations, and the data value is reduced after decomposition, but the handling of fluctuation frequency is lacking, making it inferior to the VMD algorithm. The experimental results show that the CEEMDAN-LSTM model's prediction performance is better than the classical harmonic analysis model and simple machine learning model but inferior to the VMD-LSTM model.

The VMD decomposition algorithm can simplify the data's physical structure, decomposing the raw signal of complexity and non-smoothness into feature mode components and capturing multi-scale information at different resolution levels. As a result, the decomposed training data can be trained more efficiently, and prediction performance can be better when applied to individual models. Using training data with a simplified structure in model training can learn data patterns more efficiently and reduce the effect of noise contained in the original data [28].

In contrast to Yin [20], who used a combination of discrete wavelet transform (DWT) and variable neural network to predict tidal heights, and Wang [32], who combined a hybrid model of EMD/EEMD and ARIMA to predict long-term streamflow, the EEMD- and DWT-based models for high-frequency sub-time sequences are less efficient and accurate than the low-frequency model, while the VMD-based model yields good efficiency and accuracy for both high-frequency and low-frequency sub-time sequences.

DWT is a reliable method for non-stationary signal analysis, but it can only choose a fixed wavelet basis function when used, which is weakly adaptive. The VMD non-recursive decomposition approach effectively overcomes its shortcomings. Compared with the EMD, EEMD, and CEEMD methods, it allows for a more effective separation of tidal signals and noise. The highly adaptive decomposition characteristics of VMD, the rigorous theoretical basis, and the advantages of high-frequency noise suppression, combined with LSTM's long-term dependence, achieve high-accuracy predictions of tidal forecasts. Moreover, the CEEMDAN decomposition algorithm in this study decomposes the original time series into more sub-time series than VMD, so the error accumulation of the CEEMDAN-based model may be greater than that of the VMD-based model. Therefore, VMD is a more effective model than CEEMDAN for developing tidal forecasts.

All models can obtain better prediction results in conventional tidal wave height prediction. However, when the tidal height increases, the prediction ability of the classical tidal harmonic analysis method and a single machine learning model is limited. The CEEMDAN-LSTM model can maintain good prediction accuracy, while the VMD-LSTM model can still maintain high-accuracy long-term prediction results, overcoming the problems that these models cannot solve. This indicates that the VMD algorithm is an effective way to improve the accuracy of tidal height prediction.

#### 5. Conclusions

- (1) The combination of the VMD decomposition algorithm and the LSTM neural network effectively increases the precision for long-term conventional tidal forecasting and addresses the problem of inaccurate prediction results of anomalous tidal data in long-term tide prediction. Achieving high-precision long-term tide prediction is essential for realistic marine activities. However, prediction results based on reconcil-

iation analysis often cannot meet the current refinement requirements. The pairing of the VMD decomposition algorithm and the LSTM model may provide a new method for this purpose, thus meeting the practical application needs of the growing marine economic activities.

- (2) The introduction of the VMD decomposition algorithm enables better processing of the original tide level data, effectively decomposing more stable modal components and improving the model's prediction accuracy. This approach provides new research ideas and methods for tide level prediction. This method can also be expanded, considering the combination of the VMD decomposition algorithm with more machine learning time prediction models to obtain a new combined model, or applying this model to more data types, such as waves.

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

**Funding:** This study is financially supported by General Projects of Zhoushan Science and Technology Bureau and the National Natural Science Foundation of China (No. 52101330).

**Data Availability Statement:** The data used to support the findings of this study are included within the article.

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

## References

1. Pan, H.; Lv, X.; Wang, Y.; Matte, P.; Chen, H.; Jin, G. Exploration of tidal-fluvial interaction in the Columbia river estuary using S\_TIDE. *J. Geophys. Res. Oceans* **2018**, *123*, 6598–6619. [[CrossRef](#)]
2. Zhu, X.H.; Ma, Y.L.; Guo, X.; Fan, X.; Long, Y. Tidal currents in the Qiongzhou Strait estimated from shipboard ADCP data during spring 2013. In Proceedings of the EGU General Assembly Conference Abstracts, Vienna, Austria, 27 April–2 May 2014; p. 2504.
3. Lee, T.L. Back-propagation neural network for long-term tidal predictions. *Ocean. Eng.* **2004**, *31*, 225–238. [[CrossRef](#)]
4. Chang, H.K.; Lin, L.C. Multi-point tidal prediction using artificial neural network with tide-generating forces. *Coast. Eng.* **2006**, *53*, 857–864. [[CrossRef](#)]
5. Li, M.; Liu, M.; Liu, X.; Peng, T.; Wang, S. Decomposition of long time-series fraction of absorbed photosynthetically active radiation signal for distinguishing heavy metal stress in rice. *Comput. Electron. Agric.* **2022**, *198*, 107111. [[CrossRef](#)]
6. Pawlowicz, R.; Beardsley, B.; Lentz, S. Classical tidal harmonic analysis including error estimates in MATLAB using T\_TIDE. *Comput. Geosci.* **2002**, *28*, 929–937. [[CrossRef](#)]
7. Foreman, M.; Henry, R. The harmonic analysis of tidal model time series. *Adv. Water Resour.* **1989**, *12*, 109–120. [[CrossRef](#)]
8. Sassi, H.G.; Hoitink, A.J.F.; De Brye, B.; Vermeulen, B.; Deleersnijder, E. Tidal impact on the division of river discharge over distributary channels in the Mahakam Delta. *Ocean. Dyn.* **2011**, *61*, 2211–2228. [[CrossRef](#)]
9. Jay, D.A.; Flinchem, E.P. Interaction of fluctuating river flow with a barotropic tide: A demonstration of wavelet tidal analysis methods. *J. Geophys. Res. Ocean.* **1997**, *102*, 5705–5720. [[CrossRef](#)]
10. Cacabelos, E.; Olabarria, C.; Incera, M.; Troncoso, J.S. Effects of habitat structure and tidal height on epifaunal assemblages associated with macroalgae. *Estuar. Coast. Shelf Sci.* **2010**, *89*, 43–52. [[CrossRef](#)]
11. Penna, N.T.; Stewart, M.P. Aliased tidal signatures in continuous GPS height time series. *Geophys. Res. Lett.* **2003**, *30*. [[CrossRef](#)]
12. Ji, L.; Xiao, Z.; Lei, Z. Research on the algorithm of education data mining based on big data. In Proceedings of the IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI), Xinxiang, China, 12–14 June 2020.
13. Guo, S.; Wang, Q. Application of knowledge distillation based on transfer learning of ERNIE model in intelligent dialogue intention recognition. *Sensors* **2022**, *22*, 1270. [[CrossRef](#)]
14. Ma, P.; Gao, Q. EEG signal and feature interaction modeling-based eye behavior prediction research. *Comput. Math. Methods Med.* **2020**, *2020*, 2801015. [[CrossRef](#)]
15. Xie, Y.; Zhao, J.; Qiang, B.; Mi, L.; Tang, C.; Li, L. Attention mechanism-based CNN-LSTM model for wind turbine fault prediction using SSN ontology annotation. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 6627588. [[CrossRef](#)]
16. Zhou, Y.; Li, T.; Shi, J.; Qian, Z. A CEEMDAN and XGBOOST-based approach to forecast crude oil prices. *Complexity* **2019**, *2019*, 4392785. [[CrossRef](#)]

17. Sakinah, N.; Tahir, M.; Badriyah, T.; Syarif, I. LSTM with adam optimization-powered high accuracy preeclampsia classification. In Proceedings of the International Electronics Symposium (IES), Surabaya, Indonesia, 27–28 September 2019; pp. 314–319. [[CrossRef](#)]
18. Jung, S.; Cho, H.; Kim, J.; Lee, G. Prediction of water level in a tidal river using a deep-learning based LSTM model. *J. Korea Water Resour. Assoc.* **2018**, *51*, 1207–1216.
19. Okwuashi, O.; Ndehedehe, C. Tide modelling using support vector machine regression. *J. Spat. Sci.* **2017**, *62*, 29–46. [[CrossRef](#)]
20. Yin, J.-C.; Perakis, A.N.; Wang, N. Ensemble Real-Time Tidal Level Prediction Mechanism Using Multiresolution Wavelet Decomposition Method. *IEEE Trans. Geosci. Remote. Sens.* **2018**, *56*, 4856–4865. [[CrossRef](#)]
21. Zhang, C.-Y.; Wang, Z.; Fei, C.-W.; Yuan, Z.-S.; Wei, J.-S.; Tang, W.-Z. Fuzzy multi-SVR learning model for reliability-based design optimization of turbine blades. *Materials* **2019**, *12*, 2341. [[CrossRef](#)]
22. Qu, B.; Xing, Z.; Liu, Y.; Chen, L. Research on short-term output power forecast model of wind farm based on neural network combination algorithm. *Wind. Energy* **2022**, *25*, 1710–1734. [[CrossRef](#)]
23. Cao, J.; Li, Z.; Li, J. Financial time series forecasting model based on CEEMDAN and LSTM. *Phys. A Stat. Mech. Its Appl.* **2019**, *519*, 127–139. [[CrossRef](#)]
24. Wang, W.; Chau, K.; Xu, D.; Chen, X.Y. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resour. Manag.* **2015**, *29*, 2655–2675. [[CrossRef](#)]
25. Zhang, Y.; Zhao, Y.; Kong, C.; Chen, B. A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic. *Energy Convers. Manag.* **2020**, *203*, 112254. [[CrossRef](#)]
26. Zhang, X.; Li, S.; Zhang, C. Short-Term Prediction of SDN Network Traffic Based on CEEMDAN and Mixed Kernel Least Squares Support Vector Machine. In Proceedings of the 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), IEEE, Chengdu, China, 19–21 August 2022; p. 1201.
27. Rongbin, C.; Sanming, L. Research on wind power prediction method based on CEEMDAN-SSA-GRU. In Proceedings of the 2021 IEEE International Conference on Electrical Engineering and Mechatronics Technology (ICEEMT), Qingdao, China, 2–4 July 2021; pp. 597–601. [[CrossRef](#)]
28. Wang, W.; Tong, M.; Yu, M. Blood glucose prediction with VMD and LSTM optimized by improved particle swarm optimization. *IEEE Access* **2020**, *8*, 217908–217916. [[CrossRef](#)]
29. Li, H.; Liu, T.; Wu, X.; Chen, Q. An optimized VMD method and its applications in bearing fault diagnosis. *Measurement* **2020**, *166*, 108185. [[CrossRef](#)]
30. Jin, X.-B.; Yang, N.-X.; Wang, X.-Y.; Bai, Y.-T.; Su, T.-L.; Kong, J.-L. Deep hybrid model based on EMD with classification by frequency characteristics for long-term air quality prediction. *Mathematics* **2020**, *8*, 214. [[CrossRef](#)]
31. Wang, Z.-Y.; Qiu, J.; Li, F.-F. Hybrid models combining EMD/EEMD and ARIMA for long-term streamflow forecasting. *Water* **2018**, *10*, 853. [[CrossRef](#)]
32. Lee, H.S. Estimation of extreme sea levels along the Bangladesh coast due to storm surge and sea level rise using EEMD and EVA. *J. Geophys. Res. Ocean.* **2013**, *118*, 4273–4285. [[CrossRef](#)]
33. Yan, K.; Hua, J. Deep learning technology for chiller faults diagnosis. In Proceedings of the IEEE International Conference on Dependable, Autonomic and Secure Computing, (DASC/PiCom/CBDCom/CyberSciTech), Fukuoka, Japan, 5–8 August 2019; pp. 72–79. [[CrossRef](#)]
34. Rosati, M.; Kelly, T.; Ringwood, J.V. Nonlinear data-based hydrodynamic modeling of a fixed oscillating water column wave energy device. *IEEE Access* **2021**, *9*, 149756–149765. [[CrossRef](#)]
35. Kumar, N.K.; Savitha, R.; Al Mamun, A. Regional ocean wave height prediction using sequential learning neural networks. *Ocean Eng.* **2017**, *129*, 605–612. [[CrossRef](#)]
36. Kavousi-Fard, A.; Su, W. A combined prognostic model based on machine learning for tidal current prediction. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3108–3114. [[CrossRef](#)]
37. Liu, J.; Shi, G.; Zhu, K. High-precision combined tidal forecasting model. *Algorithms* **2019**, *12*, 65. [[CrossRef](#)]
38. Wei, J.; Jiang, G.-Q.; Liu, X. Parameterization of typhoon-induced ocean cooling using temperature equation and machine learning algorithms: An example of typhoon Soulik (2013). *Ocean Dyn.* **2017**, *67*, 1179–1193. [[CrossRef](#)]
39. Yen, P.H.; Jan, C.D.; Lee, Y.P.; Lee, H.F. Application of Kalman filter to short-term tide level prediction. *J. Waterw. Port Coast. Ocean. Eng.* **1996**, *122*, 226–231. [[CrossRef](#)]
40. Granata, F.; Di Nunno, F. Artificial Intelligence models for prediction of the tide level in Venice. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 2537–2548. [[CrossRef](#)]
41. Wu, W.; Li, L.; Yin, J.; Zhang, W. A modular tide level prediction method based on a NARX neural network. *IEEE Access* **2021**, *9*, 147416–147429. [[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.