


## Article

# A Hybrid Model for Vessel Traffic Flow Prediction Based on Wavelet and Prophet

Dangli Wang<sup>1,2,3</sup>, Yangnan Meng<sup>1</sup> , Shuzhe Chen<sup>1,2,3,\*</sup>, Cheng Xie<sup>1,2</sup> and Zhao Liu<sup>1,2</sup><sup>1</sup> School of Navigation, Wuhan University of Technology, Wuhan 430063, China; w48236917@gmail.com (D.W.); mengyr@whut.edu.cn (Y.M.); cxie@whut.edu.cn (C.X.); zhaoliu@whut.edu.cn (Z.L.)<sup>2</sup> Hubei Key Laboratory of Inland Shipping Technology, Wuhan 430063, China<sup>3</sup> National Engineer Research Center Water Transport Safety, Wuhan 430063, China

\* Correspondence: cszcsz79@whut.edu.cn; Tel.: +86-130-2630-7744

**Abstract:** Accurate vessel traffic flow prediction is significant for maritime traffic guidance and control. According to the characteristics of vessel traffic flow data, a new hybrid model, named DWT-Prophet, is proposed based on the discrete wavelet decomposition and Prophet framework for the prediction of vessel traffic flow. First, vessel traffic flow was decomposed into a low-frequency component and several high-frequency components by wavelet decomposition. Second, Prophet was trained to predict the components, respectively. Finally, the prediction results of the components were reconstructed to complete the prediction. The experimental results demonstrate that the hybrid DWT-Prophet outperformed the single Prophet, long short-term memory, random forest, and support vector regression (SVR). Moreover, the practicability of the new forecasting method was improved effectively.

**Keywords:** vessel traffic flow; prediction; wavelet decomposition; Prophet



**Citation:** Wang, D.; Meng, Y.; Chen, S.; Xie, C.; Liu, Z. A Hybrid Model for Vessel Traffic Flow Prediction Based on Wavelet and Prophet. *J. Mar. Sci. Eng.* **2021**, *9*, 1231. <https://doi.org/10.3390/jmse9111231>

Academic Editor: Claudio Ferrari

Received: 23 October 2021

Accepted: 2 November 2021

Published: 7 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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

Reliable vessel traffic flow prediction not only can provide the basis for port resource management and allocation to achieve the purpose of improving channel port operation efficiency, but it also is conducive to the implementation of maritime traffic safety guarantee measures [1]. In recent years, with the rapid development of the shipping economy, vessel traffic has become increasingly busy, and water traffic safety is facing tremendous pressure. At the same time, with the rise of artificial intelligence technology, the construction of an intelligent water transportation system is of great help for reducing maritime accidents. Accurate vessel traffic flow prediction also provides a basis for the effective operation of artificial intelligence systems [2].

There are plenty of methods for the prediction of vessel traffic flow. The traditional prediction models are mainly linear models based on mathematical statistics. Among the linear models, the grey model, based on small samples and poor information, was established to predict the unknown part through the relevance of the part of the known information [3,4]; time series prediction models can make full use of data and have the characteristics of fast calculation speed and high accuracy; the Kalman filter has the advantages of dynamically adjusting model parameters and processing observations in real time [5]. Linear models have strong interpretability of prediction results obtained through the analysis of historical data when the data are not complex. Because of the advantages of the linear models, they have been introduced to many fields including energy prediction [6], traffic flow prediction [7], passenger flow prediction [8], and air quality prediction [9,10]. However, when the data are nonlinear and complex, these linear models are challenging for the exploration of the potential laws of the data, resulting in poor prediction performance [11]. To overcome the defects of the linear models, nonlinear prediction models in artificial intelligence fields, such as artificial neural networks (ANNs)

and support vector regression (SVR), have been proposed and widely used [12–14]. These models have nonlinear solid mapping and adaptive learning capabilities to mine complex relationships in data and improve prediction accuracy. For example, Liu et al. [15] built a dynamic model based on a back propagation (BP) neural network to predict the operating parameters of a nuclear power plant (NPP). Ahmad et al. [16] employed the support vector regression (SVR) to forecast the splice strength in reinforced concrete, and the experimental results showed that the SVR had better prediction performance. With the rapid development of research methods, researchers have discovered that deep learning can further explore the underlying laws of data, overcome the shortcomings of artificial neural networks, and have better prediction results. For example, Jorges et al. [17] introduced long short-term memory (LSTM) neural networks to predict the wave height. Single linear and nonlinear models have their own advantages; in certain specific studies, they can show good predictive ability. However, when the data are more complex or the data change drastically, the limitations of these models reduce the predictive performance of the models. Therefore, researchers began to try to use two or more hybrid models for prediction, and the recent research results show that hybrid models can make full use of the advantages of every single model and overcome the limitations of the single model, greatly improving the model's prediction performance [18–20]. The hybrid method is widely used in many fields because of its superiority in predictive performance. Ma et al. [21] adopted a hybrid forecast model based on the DWT-TCN-PSO-SVR, which can effectively forecast traction load. Tian et al. [22] combined the empirical mode decomposition and LSTM which had parameters that were optimized by the sparrow search algorithm to make predictions for wind speed data. Han et al. [23] proposed ensemble learning model based on a recurrent neural network (RNN), extreme learning machine (ELM), support vector regression (SVR), and least squares support vector machine (LSSVM) to forecast the exhaust emissions, and the method made full use of all four models in this way. He et al. [24] built a SARIMA-CNN-LSTM model to predict daily travel demand, used the seasonal autoregressive integrated moving average (SARIMA) model to capture linear features in the data, and used a convolutional neural network (CNN) and LSTM to mine nonlinear features; the results showed that the prediction accuracy of the hybrid model was significantly higher than those of a single SARIMA and LSTM. Du et al. [25] proposed an improved whale optimization algorithm to optimize wavelet neural networks and predict short-term traffic flow in cities, and the experimental results showed that this method could effectively compensate for the low prediction accuracy of the wavelet neural network. Chen et al. [26] built an EEMD-LSTM model to predict road traffic flow. Vessel traffic flow is easily affected by factors such as environment and weather, so its data sequence had the characteristics of nonlinearity, complexity, randomness, etc. A single predictive model cannot characterize the potential law of vessel traffic flow either, while a combined model can overcome it. For example, Zhao et al. [27] proposed the SARIMA-BP neural network to predict vessel traffic flow in the Shenzhen Port and achieved good prediction results. Li et al. [1] used the chaotic cloud-simulated annealing genetic algorithm to optimize the parameters of the robust support vector regression model (RSVR), and then used the kernel principal component analysis method to determine the input variables of the model, and the prediction effect was effectively improved after this hybrid method was used to predict the vessel traffic flow in Tianjin Port. Zhang et al. [28] established a regression-Kalman filter model to predict vessel traffic flow in the channel's regulation area.

In order to further improve the prediction accuracy of vessel traffic flow, this paper proposes a vessel traffic flow prediction method based on the combination of discrete wavelet decomposition and the Prophet framework. Wavelet decomposition was introduced to decompose the data into an approximate component and several high-frequency components, and then the Prophet framework was trained to predict every component. The final prediction result was obtained by reconstructing every component prediction result.

The main contributions of the DWT-Prophet model are described as follows:

- (1) Aiming at the nonlinear and non-stationary characteristics of vessel traffic flow data, the discrete wavelet decomposition method was used to decompose the data into sub-sequences of different frequencies that made it easier for the prediction model to characterize its internal characteristics;
- (2) A Bayesian curve fitting method was adopted to smooth and predict time series data when using the Prophet, and the data with periodic and trend changes and significant outliers could effectively be processed to obtain a better fitting effect;
- (3) To verify the prediction performance of the DWT–Prophet combined model, the prediction results were compared with other models, such as Prophet, LSTM, random forest, and SVR, using vessel traffic flow data from the Wuhan Port Yangtze River Bridge section. The experimental results showed that the DWT–Prophet combination model had a better prediction effect.

## 2. Theory and Methodologies

### 2.1. Wavelet Theory

#### 2.1.1. Principle of Decomposition and Reconstruction

The core idea of wavelet analysis is to decompose and reconstruct signals and overcome the limitations of short-time Fourier transform by providing time–frequency windows that vary with frequency. Wavelet transform, which is often used as an effective tool to extract time–frequency information on non-stationary series, is not constrained by a stationary hypothesis and can automatically adapt to the requirements of time–frequency signal analysis and focus on the details of signals. Therefore, wavelet transform has become a major breakthrough since Fourier transform. Wavelet transform is divided into continuous wavelet transform (CWT) and discrete wavelet transform (DWT). As the data of vessel traffic flow are a discrete structure, this paper chose DWT to decompose and reconstruct the data of vessel traffic flow.

The original time series was decomposed into a low-frequency series and several high-frequency series by scale function,  $\phi(t)$ , and wavelet basis function,  $\psi(t)$ , when wavelet transform was used. The approximate coefficient,  $a_0(k)$ , and detail coefficient,  $d_j(k)$ , of the discrete wavelet transform are expressed by the following formulae:

$$a_0(k) = \frac{1}{\sqrt{N}} \sum_{n=1}^N y(n) \phi_{0,k}(n) \quad (1)$$

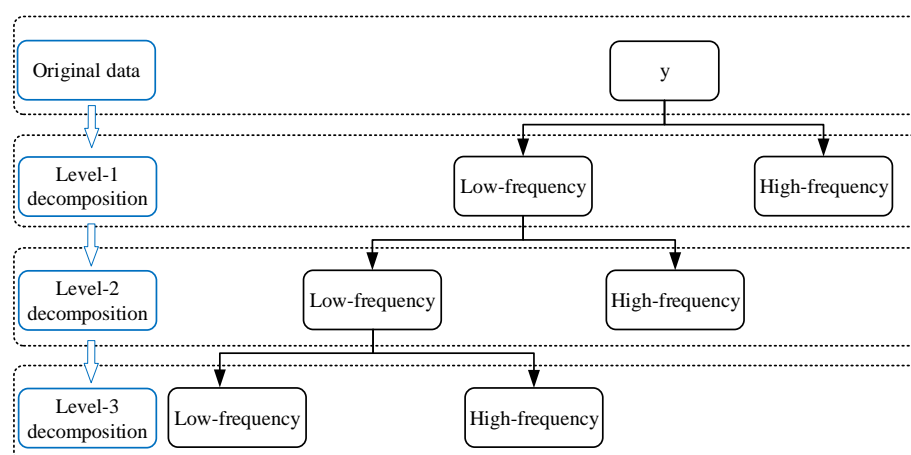
$$d_j(k) = \frac{1}{\sqrt{N}} \sum_{n=1}^N y(n) \psi_{j,k}(n) \quad (2)$$

where  $y(n)$  is the original signal;  $j$  and  $k$  are the expansion and contraction of the sub-signals in the frequency domain and the translation in the time domain.

The approximate coefficients obtained from the decomposition of the layer  $j$  and the detail coefficients obtained from the decomposition of each layer were reconstructed to obtain the approximate components,  $A_j$ , and detailed components,  $D_1, D_2, \dots, D_j$ , of each layer. Then, the multi-layer decomposition of the original signal was completed. The relationship between the original signal and each component can be expressed as:

$$y(t) = D_1 + D_2 + \dots + D_j + A_j \quad (3)$$

Figure 1 is the schematic diagram of three-layer wavelet decomposition in which the low-frequency parts of each layer correspond to the approximate components of the original signal, such as  $A_1, A_2$ , and  $A_3$ , while the high-frequency parts of each layer correspond to the detailed components of the original signal such as  $D_1, D_2$ , and  $D_3$ . The approximate component of each layer can be subdivided into the approximate component and detail component of the next layer. Compared with the original signal, the decomposed components can more effectively and accurately mine the potential information of the data.



**Figure 1.** Schematic diagram of three-layer wavelet decomposition.

### 2.1.2. Selection of Wavelet Basis

Wavelet basis has five characteristics, which are orthogonality, symmetry, vanishing moment, regularity, and compact support. Generally speaking, the orthogonality of norms can make the wavelet analysis simpler and facilitate signal reconstruction. The symmetry of the wavelet basis function will not lead to signal distortion, which is beneficial to the efficient operation of the algorithm. The higher the vanishing moment, the faster the attenuation at high frequency, which is beneficial to the energy concentration of the decomposed signal, but the support length will be too long if the vanishing moment is too high. Regularity determines the smoothing degree after reconstruction. The better the regularity of wavelet, the better the smoothing effect after signal reconstruction, but good regularity will lead to long support length. The shorter the compact support, the more beneficial to the efficient implementation of the algorithm. The features of the wavelet basis function are shown in Table 1.

**Table 1.** Features of several different wavelet basis functions.

Functions	Orthogonality	Bi-Orthogonality	Symmetry	Disappearing Moment	Support Length	Tight Supportability
haar	✓	✓	✓	1	1	✓
dbN	✓	✓	✓	N	2N-1	✓
coifN	✓	✓	Approximate	2N	6N-1	✓
symN	✓	✓	Approximate	N	2N-1	✓
morl	×	×	✓	×	∞	×

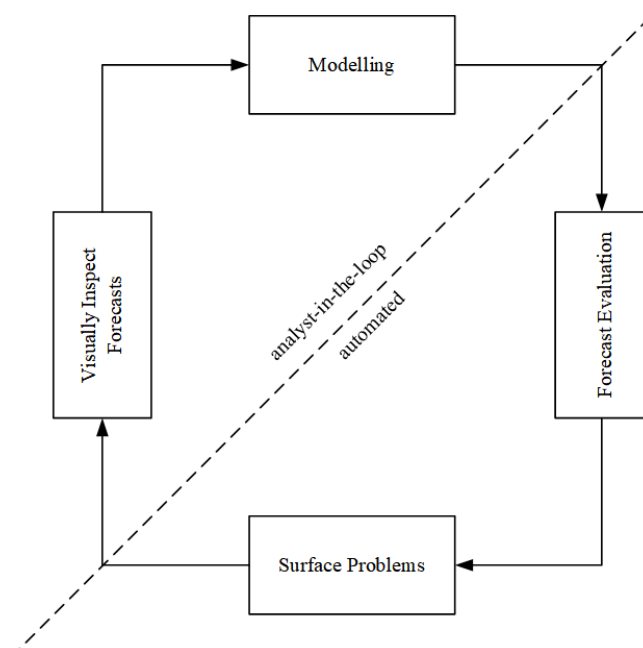
According to the characteristics of vessel traffic flow, the selection of wavelet basis function should meet the requirements of good regularity, appropriate vanishing moment, and orthogonality of specifications, etc. Therefore, the wavelet basis functions of the db family are more suitable for vessel traffic flow prediction.

### 2.2. Prophet Framework

Prophet is an open-source model that was developed by the Facebook team for time series prediction in 2017 [29]. The model has three basic characteristics named trend term, seasonal period term, and holidays. Different from other forecast methodologies, the Prophet framework is simple and easy to explain. Except strong prediction ability, Prophet can also effectively deal with data with periodic and trend changes plus large abnormal values. Affected by various factors, the data of vessel traffic flow have periodic and trend changes, while there are also abnormal values.

Using the Prophet framework to predict vessel traffic flow can overcome the limitations of traditional forecasting models and obtain ideal forecasting results. As shown in

Figure 2, the modeling process of vessel traffic flow using the Prophet framework was divided into the following four stages:



**Figure 2.** Modeling process of Prophet.

- (1) In the stage of modeling, the analyst builds an appropriate model according to the characteristics and laws of vessel traffic flow;
- (2) In the stage of forecast evaluation, through simulating the historical data of vessel traffic flow, the forecast effect is evaluated and the parameters are constantly adjusted;
- (3) In the stage of surface problem, if the prediction effect is difficult to meet the requirements, the model presents the potential reasons to the analyst;
- (4) In the stage of visually inspecting forecasts, the analyst adjusts the parameters of the model or reconstructs the model according to the visual forecast results and problems.

In fact, the process of predicting vessel traffic flow with the Prophet framework is a process of dynamic adjustment and circulation between the analyst part and the automation part according to the prediction results, which, to a certain extent, improves the accuracy of vessel traffic flow prediction. Compared with other models, the Prophet framework has the characteristics of extensibility, data flexibility, high speed, and easy interpretation of model variables.

The Prophet framework can be expressed by the following formula:

$$Y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (4)$$

where,  $g(t)$  is the trend item, modeling the aperiodic changes in time series;  $s(t)$  is a seasonal periodic item, reflecting the nature of periodic changes in the day, week, or year;  $h(t)$  is a holiday item, indicating the influence of special events on time series;  $\varepsilon_t$  is an error term, meaning that it does not react abnormally in the model and obeys normal distribution.

#### 2.2.1. Trend Items

There are two growth models in the Prophet framework,  $g(t)$ , one is the logistic model and the other one is the linear model. The difference between which is whether the carrying capacity is limited. Trend items are affected by change points.

The expression of the growth trend of logistic regression nonlinear saturation is:

$$g(t) = \frac{C(t)}{1 + e^{-k(t)(t-m(t))}} \quad (5)$$

In the formula,  $C(t)$  is the bearing capacity;  $k(t)$  is growth rate;  $m(t)$  is the offset. All three parameters are functions of time ( $t$ ).

The expression of piecewise linear growth trend is:

$$g(t) = (k(t) + a(t)^T \delta)t + (m(t) + a(t)^T \gamma) \quad (6)$$

In the formula  $k(t)$  is the slope variation;  $a(t)$  is the parameter adjustment vector;  $m(t)$  indicates the offset.

The trend item growth model should be selected according to the characteristics of the predicted objects. In this paper, the piecewise linear growth model was selected as trend items.

### 2.2.2. Seasonal Period Items

Vessel traffic flow time series often contain periodic changes such as day, week, and year. Only by grasping these periodic changes and effectively combining these trends can we obtain accurate prediction results. The function  $s(t)$  uses Fourier series to approximate the seasonal periodicity of time series, with its expression as follows:

$$s(t) = \sum_{n=1}^N (a_n \cos \frac{2n\pi t}{P} + b_n \sin \frac{2n\pi t}{P}) \quad (7)$$

where,  $a_n$  and  $b_n$  are learning parameters, respectively;  $P$  is the period,  $P = 365$  means that the time series is estimated in years, and  $P = 7$  means that the time series is estimated in weeks;  $N$  is the order of fitting, and the bigger the value  $N$ , the better the fitting effect on the periodic changes of the model, but it is easier to cause over-fitting.

It can be assumed as:

$$\beta = (a_1, b_1, \dots, a_N, b_N)^T \quad (8)$$

$$X(t) = [\cos(\frac{2\pi(1)t}{P}), \sin(\frac{2\pi(1)t}{P}), \dots, \cos(\frac{2\pi(N)t}{P}), \sin(\frac{2\pi(N)t}{P})] \quad (9)$$

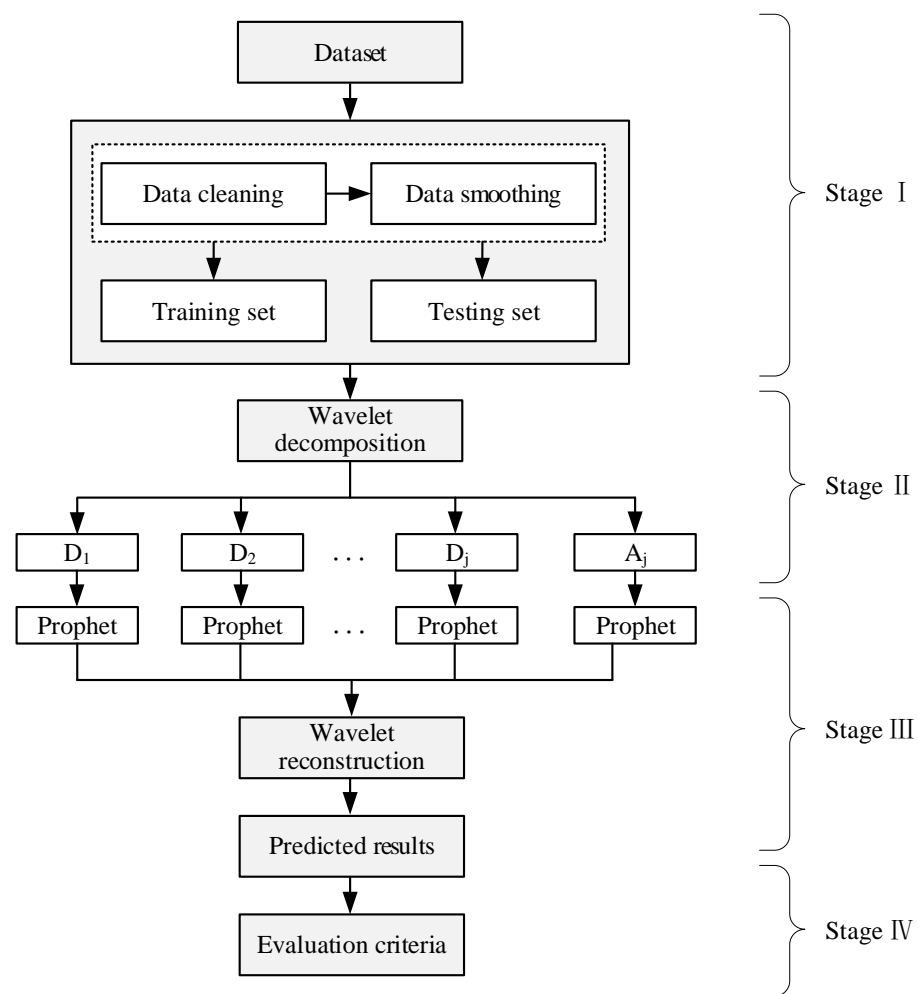
The seasonal periodic function  $s(t)$  can also be expressed as the product of two vectors:

$$s(t) = \beta X(t) \quad (10)$$

In the formula,  $\beta$  obeys  $Normal(0, \sigma)$  distribution.

### 2.3. DWT-Prophet Combination Model

Researchers often build decomposing–predicting–integrating models for time series prediction to improve prediction accuracy. The flow chart of the combined prediction model of wavelet decomposition and Prophet constructed in this paper can be seen in Figure 3.



**Figure 3.** Flow chart of the combination model prediction.

When the combination model is used to predict the vessel traffic flow, the time series of the vessel traffic flow can be decomposed into one low-frequency approximate part  $A_j$  by wavelet decomposition, and several high-frequency detail parts  $D_1, D_2, \dots, D_j$ . In this way, the time–frequency information of the signal can effectively be captured. Secondly, the Prophet framework is used to predict each component, and the final prediction result is obtained by superimposing the predicted values of each component.

#### 2.4. Model Evaluation Indicators

In order to verify the effectiveness of the DWT–Prophet model, the following four indexes were used to measure the prediction effect of vessel traffic flow, and the expressions are as follows:

- (1) Mean absolute percentage error:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (11)$$

- (2) Mean absolute error:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (12)$$

- (3) Root mean square error:



$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|^2} \quad (13)$$

(4) Coefficient of determination:

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (14)$$

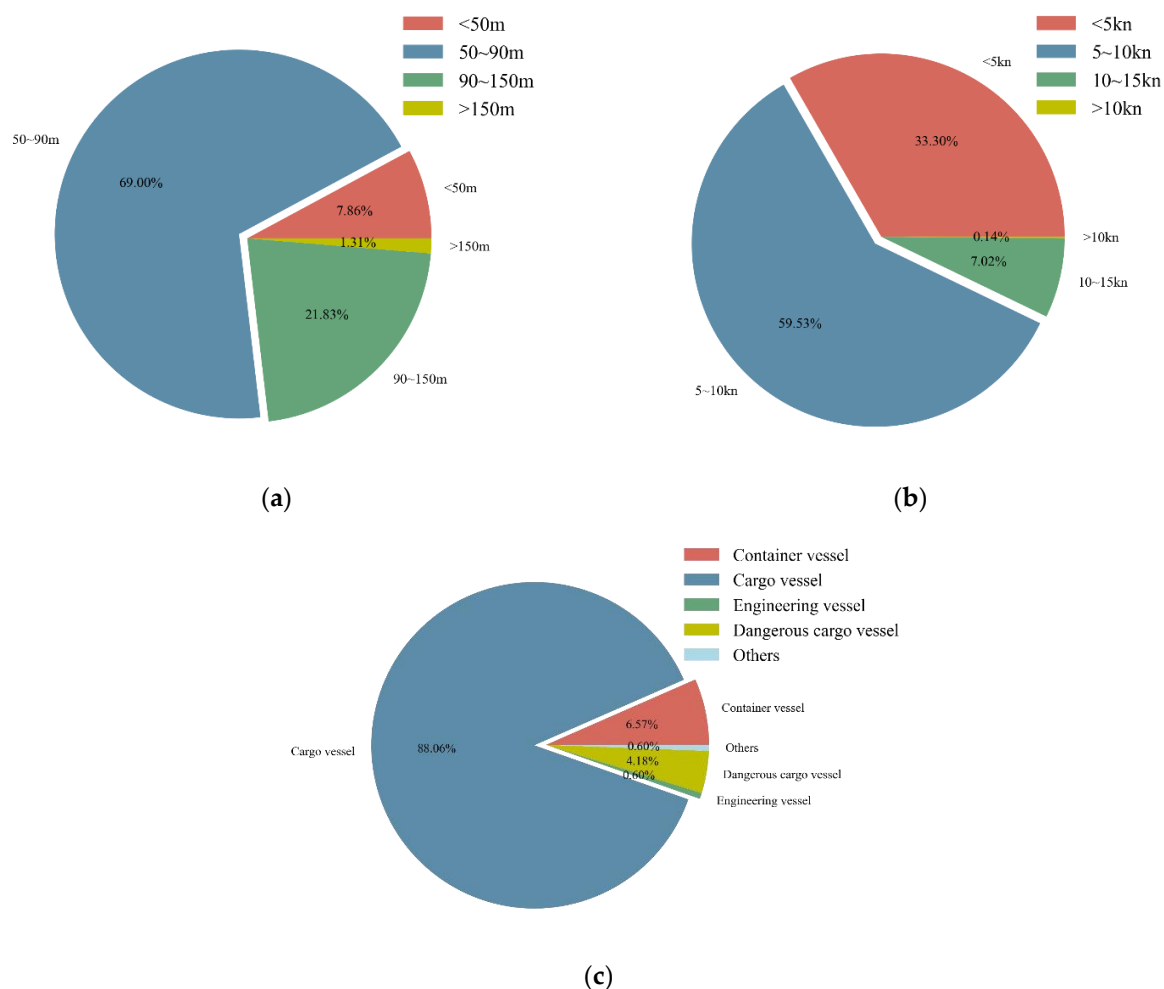
where  $y_t$  is the observed value at time  $t$ ;  $\hat{y}_t$  is the predicted value at time  $t$ .

The smaller the MAPE, MAE, and RMSE, and the closer the coefficient of determination is to 1, the better the prediction effect of the model.

### 3. Experiment and Result Analysis

#### 3.1. Data Preprocessing

In this paper, the hourly vessel traffic flow of the Wuhan Port Yangtze River Bridge in January 2018 was selected as the research object. Figure 4 shows the distribution of vessel traffic flow characteristics in this section. As shown in Figure 4, the main types of vessels were cargo vessels, accounting for 88.06%, followed by container vessels, accounting for 6.57%; vessels between 50 and 90 m in length accounted for the most significant percentage, reaching 69.00%, followed by 90 to 150 m in length, accounting for 21.83%; the vessels' speeds were less than 10 kn, most of which were between 5 and 10 kn.



**Figure 4.** Distribution of vessel traffic flow characteristics: (a) distribution of vessel length; (b) distribution of vessel speed; (c) distribution of vessel style.

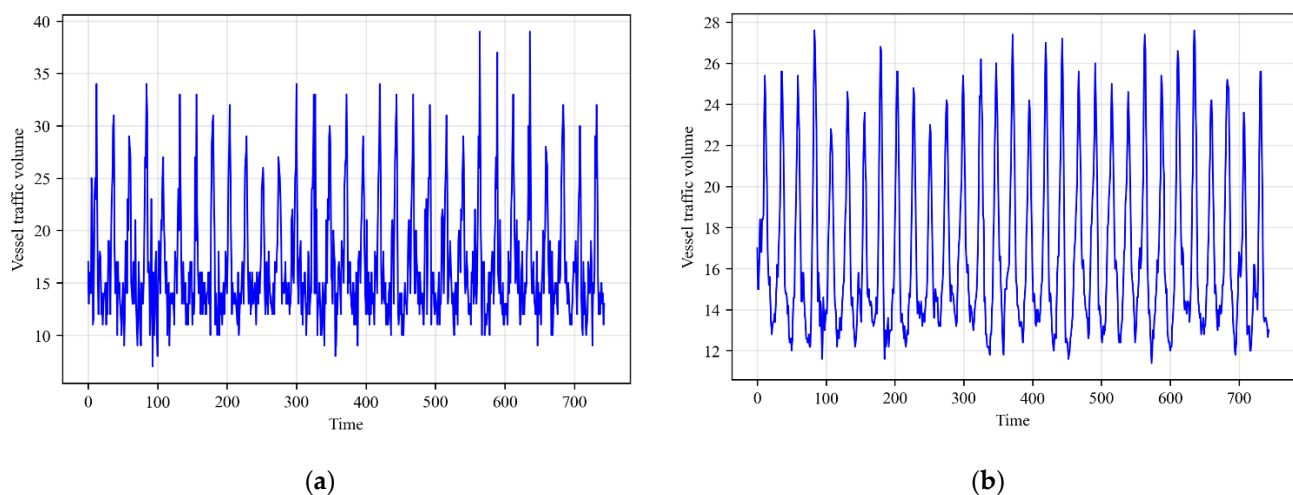


There were 744 sets of data in total including 672 sets of training data from January 1st to 28 January 2018, and 72 sets of test data from 29 January to 31 January. Due to the presence of various factors, the data on vessel traffic flow were noisy and often showed great volatility, and forecasting will directly lead to great errors. In order to avoid this situation, it was necessary to preprocess the original data on vessel traffic flow. Mean filtering is a simple and efficient filtering method that can suppress signal noise and random error. Mean filtering refers to smoothing the data by calculating the arithmetic mean of the data at a certain time and before and after and taking it as the observed value at that time. Using Formula (15), we obtained the mean filtered value:

$$y_t = \frac{1}{2N+1} \sum_{i=t-N}^{t+N} x_i \quad (15)$$

where  $x_i$  is the actual observed value of vessel traffic flow at time  $i$ ;  $2N$  is the number of neighbors of  $x_t$  including  $N$  neighbors, such as  $x_{t-N}, x_{t-N+1} \dots x_{t-1}$ , and  $N$  neighbors such as  $x_{t+1}, x_{t+2} \dots x_{t+N}$ ;  $y_t$  is the arithmetic mean of the  $2N+1$  samples.

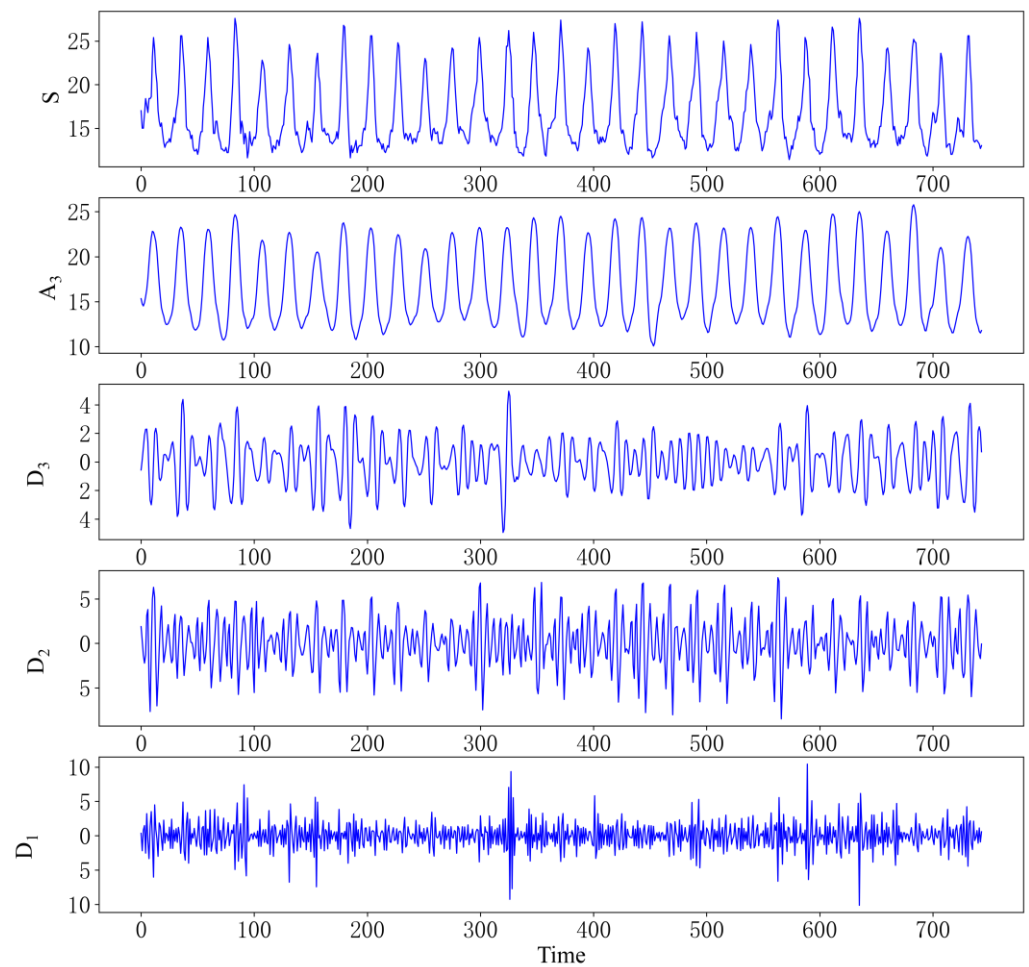
Formula (15) was adopted to smooth the vessel traffic flow data in this paper, where is 1. Figure 5 shows the comparison between the processed data and the original data. It can be seen that after smoothing, the vessel traffic flow data not only maintained the changing trend of the original data but also reduced the fluctuation between the data.



**Figure 5.** Comparison of the data before and after processing: (a) original data; (b) processed data.

### 3.2. Construction of the DWT–Prophet Model

When using discrete wavelet to decompose vessel traffic flow at multi-scale, it is necessary to not only ensure that the approximate and detailed components after decomposition are smooth but also to minimize signal loss. On this basis, the lower the complexity of the model, the better it will be. According to previous research experience, it is better to choose 1–3 layers for decomposition, and the scale should not be too large, otherwise it will lead to too much calculation and increase the complexity of the model, which will lead to too many errors. After repeated experiments, it is optimal to use db3 wavelet basis to decompose traffic flow signals into three layers, which not only keeps the trend of vessel traffic flow but also separates high-frequency information. Figure 6 is a three-layer decomposition effect diagram of db3.



**Figure 6.** Three-layer wavelet decomposition.

It can be seen from Figure 6 that  $S$  is the actual vessel traffic flow,  $A_3$  is the low-frequency component after wavelet decomposition, which keeps the same trend as  $S$ , while  $D_1$ ,  $D_2$  and  $D_3$  are the high-frequency components, which fluctuate violently.

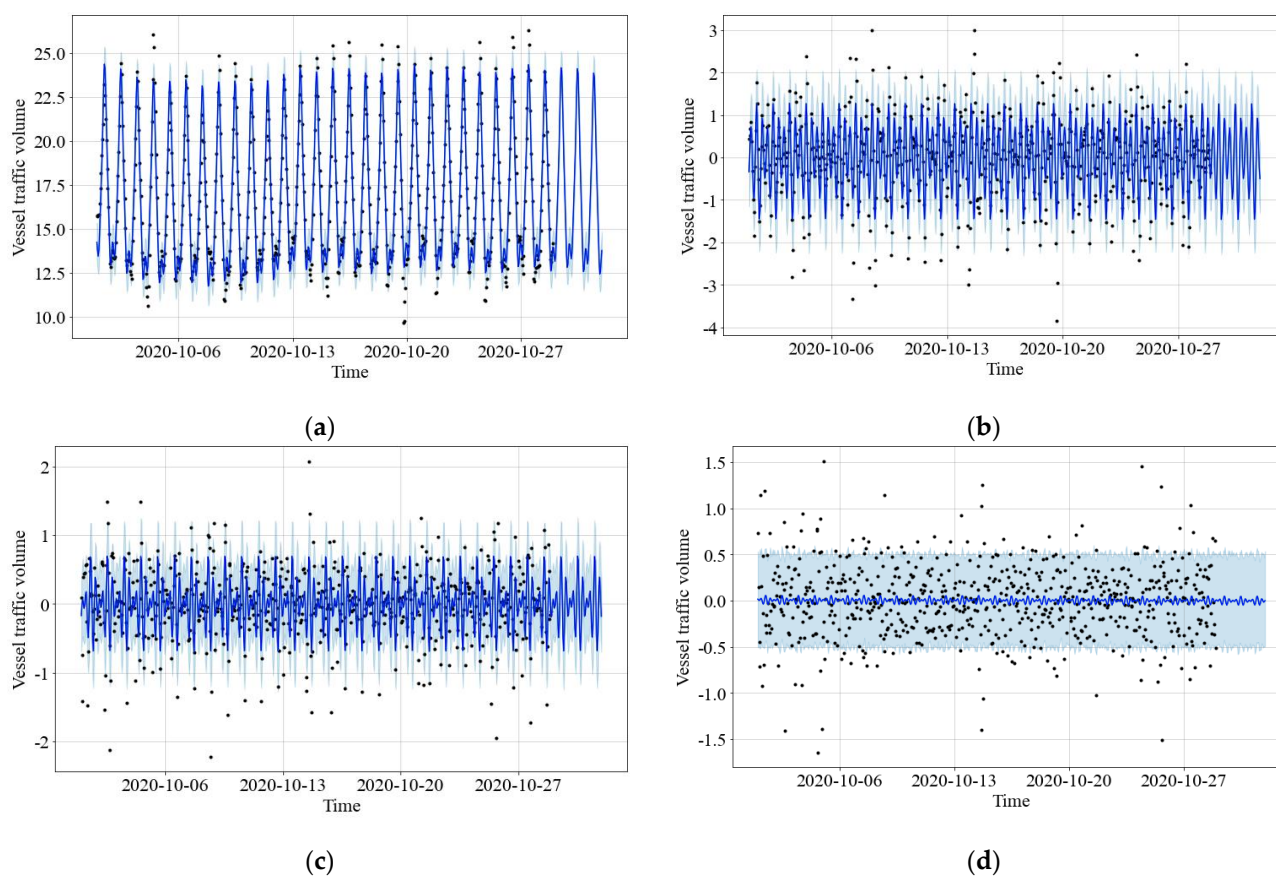
The approximate component  $A_3$  and detailed components  $D_1$ ,  $D_2$ , and  $D_3$  were predicted by the Prophet framework. The Prophet framework can decompose time series into trend items, seasonal periodic items, etc. When the Prophet framework is used to predict time series, it is necessary to set and adjust the parameters of the model in order to achieve the best prediction effect of the model. Table 2 shows the settings of the Prophet's parameters.

**Table 2.** Prophet parameter settings.

Parameter Name	Value
Growth	Linear
n_changepoints	Auto
changepoint_prior_scale	0.5
changepoint_range	0.95
yearly_seasonality	False
Weekly_seasonality	True
daily_seasonality	True

### 3.3. Analysis of Prediction Results

As shown in Figure 7, the training data and the parameters were inputted into the Prophet framework, and the prediction result of vessel traffic flow was obtained.



**Figure 7.** Prediction results of four components: (a)  $A_3$ ; (b)  $D_3$ ; (c)  $D_2$ ; (d)  $D_1$ .

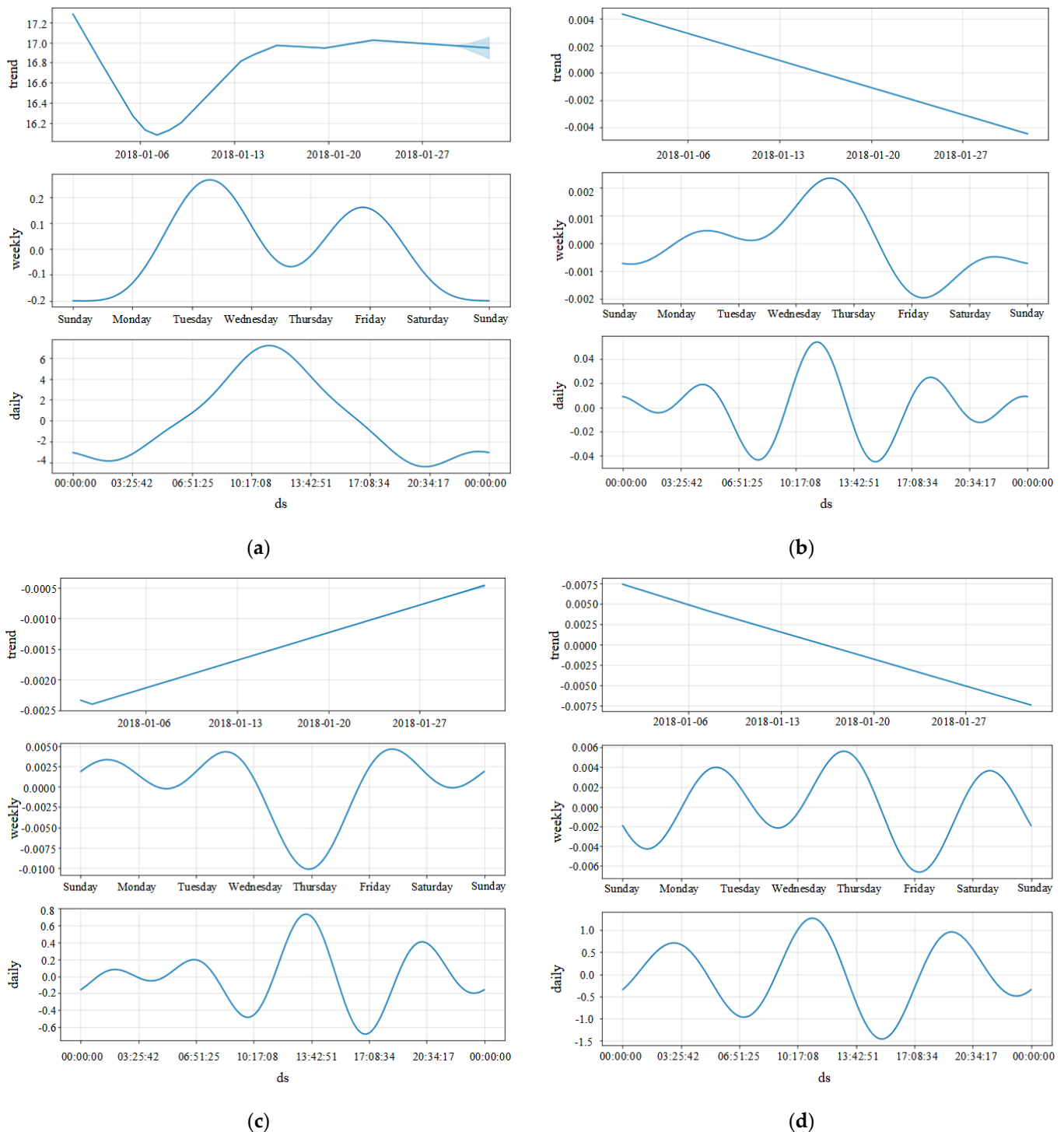
In Figure 7, the part of the shadow indicates the confidence interval of prediction. From this figure, we know that the prediction framework fit  $A_3$ ,  $D_2$ , and  $D_3$  well, and the prediction confidence interval contained most of the real data, which indicates that the prediction model can grasp well the trend of  $A_3$ ,  $D_2$  and  $D_3$ . However, the model did not fit to  $D_1$  well, that is because  $D_1$  contained too much noise. Figure 8 shows the trend term change and periodic change in the four components, and the final prediction results were the sum of the trend term and periodic term of these components.

The Prophet framework was used to predict the high- and low-frequency components of the 72 sets of test data, and the prediction results are shown in Figure 9. Comparing the predicted and actual values of each component, we know that the change trends in predicted and actual values were basically consistent, and the Prophet framework could better predict the subsequence of actual values after wavelet decomposition.

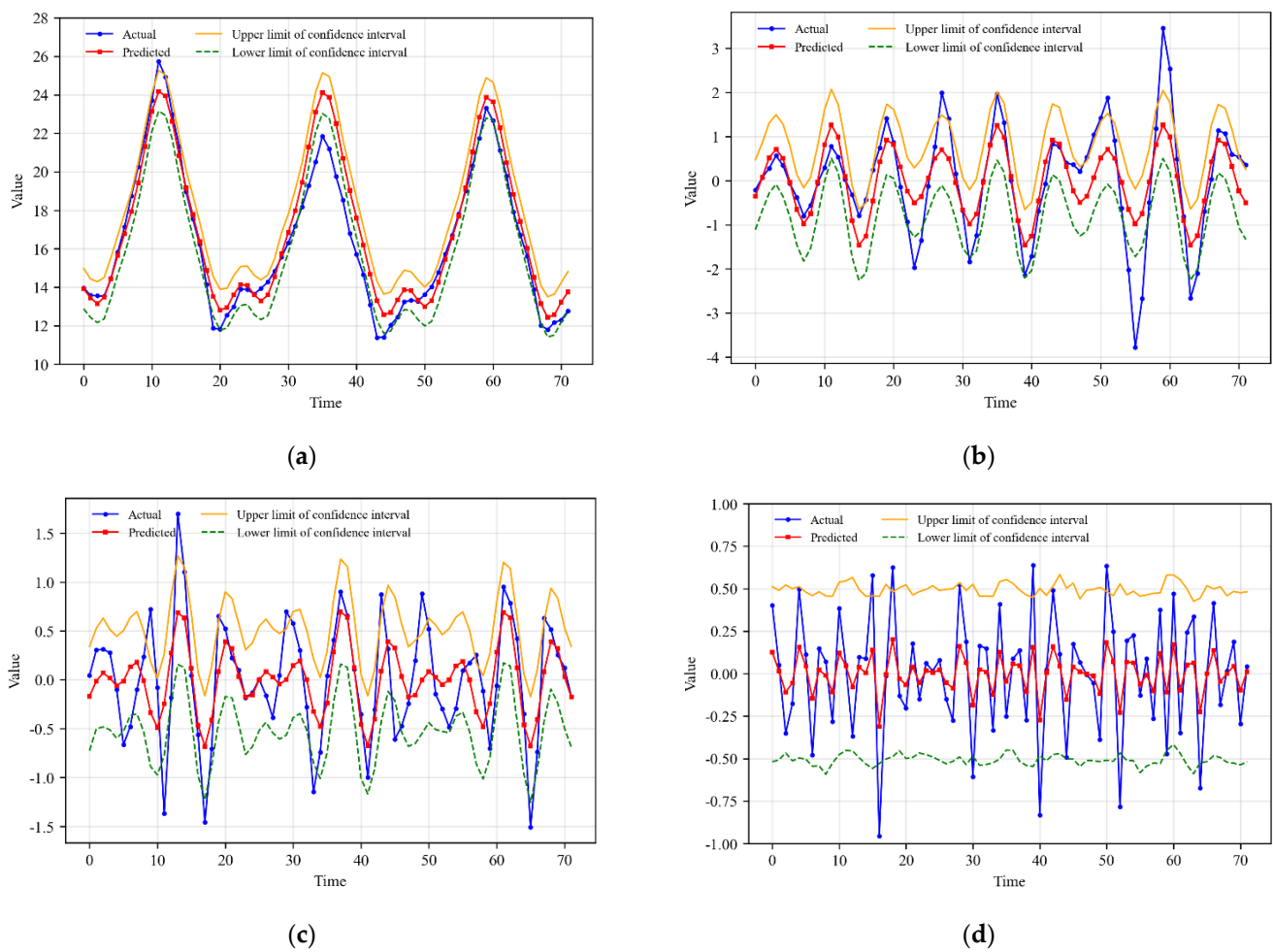
In order to verify the performance of the hybrid model based on DWT-Prophet proposed in this paper, the Prophet framework, the LSTM, the RF, and the SVR model were selected as the comparison models to predict the vessel traffic flow in the testing data set. The comparison between the actual values and the predicted values of the five models are shown in Figure 10.

It can be seen from Figure 10 that all five models were able to predict vessel traffic flow. The combination model proposed in this paper had the best prediction performance, followed by the Prophet, the random forest, and the SVR model, and the LSTM had the worst prediction performance. The prediction curve of the DWT-Prophet model was closer to the original vessel traffic flow curve, and the prediction accuracy of the hybrid model was much higher in the time period with more noise influence when compared with the comparison models, which indicates that the combination model can grasp the changing trends in vessel traffic flow and mine the potential information on vessel traffic flow data more effectively. Figure 11 reflects the relative errors of these models. As

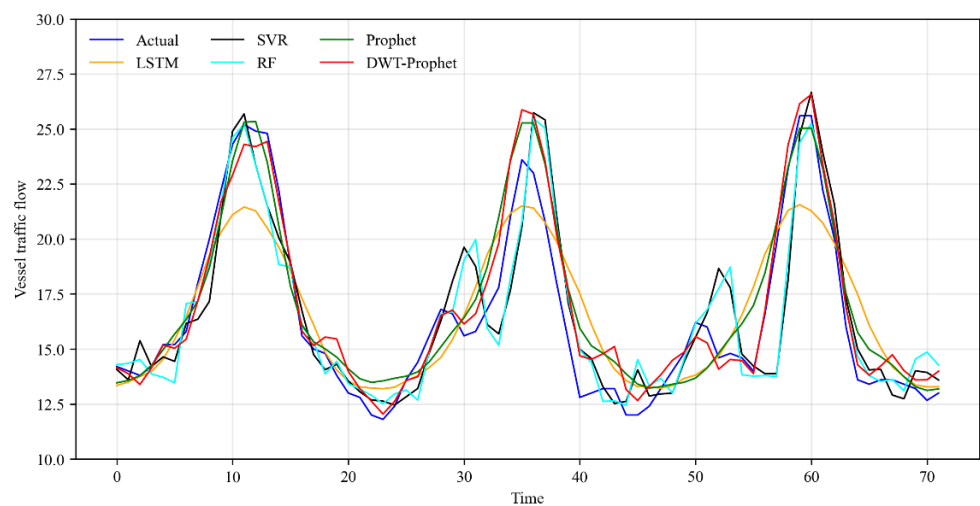
shown in Figure 11, the relative error of the hybrid model proposed in this paper was smallest on the whole, which means that DWT-Prophet outperformed the comparison models significantly.



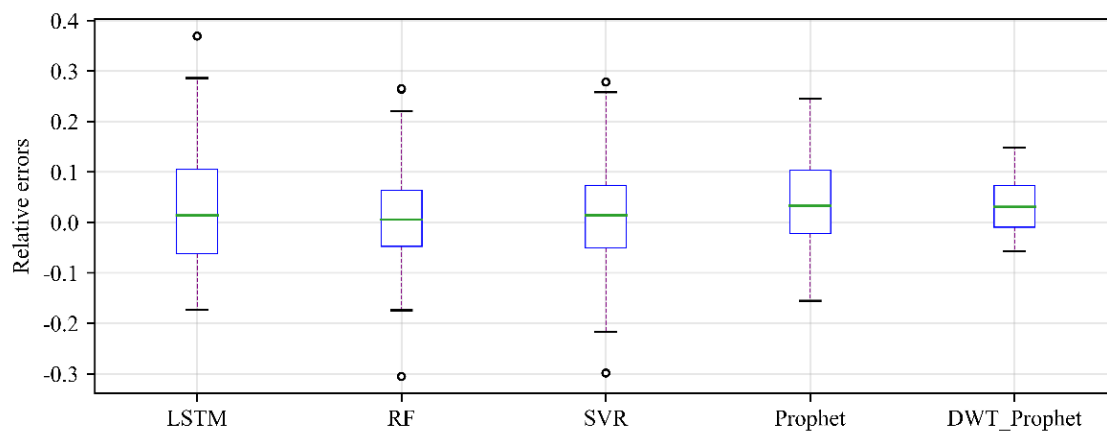
**Figure 8.** The trend term and periodic change in four components: (a)  $A_3$ ; (b)  $D_3$ ; (c)  $D_2$ ; (d)  $D_1$ .



**Figure 9.** Comparison of the actual and predicted results of the four components of the test set: (a)  $A_3$ ; (b)  $D_3$ ; (c)  $D_2$ ; (d)  $D_1$ .

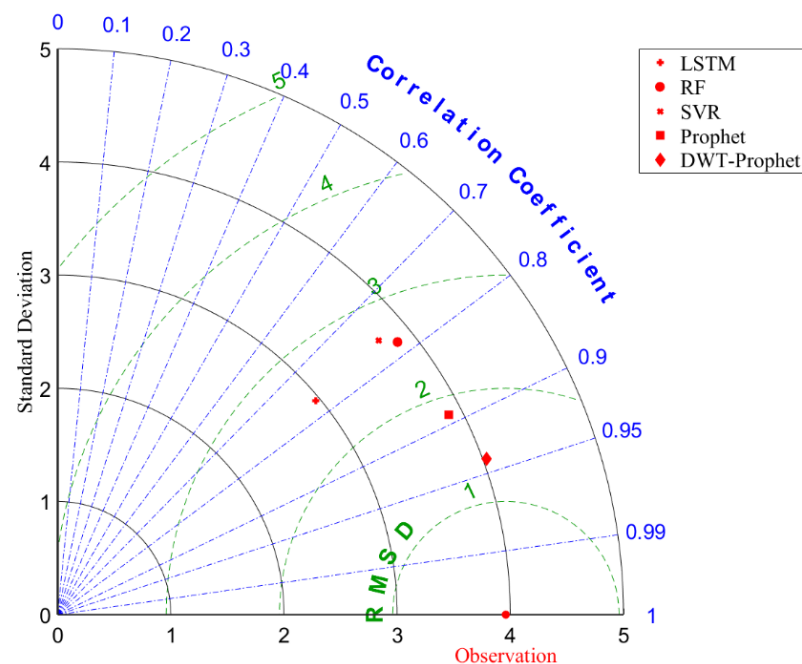


**Figure 10.** Comparison results of the actual and predicted values of the five models.



**Figure 11.** Comparison of the relative errors of the five models.

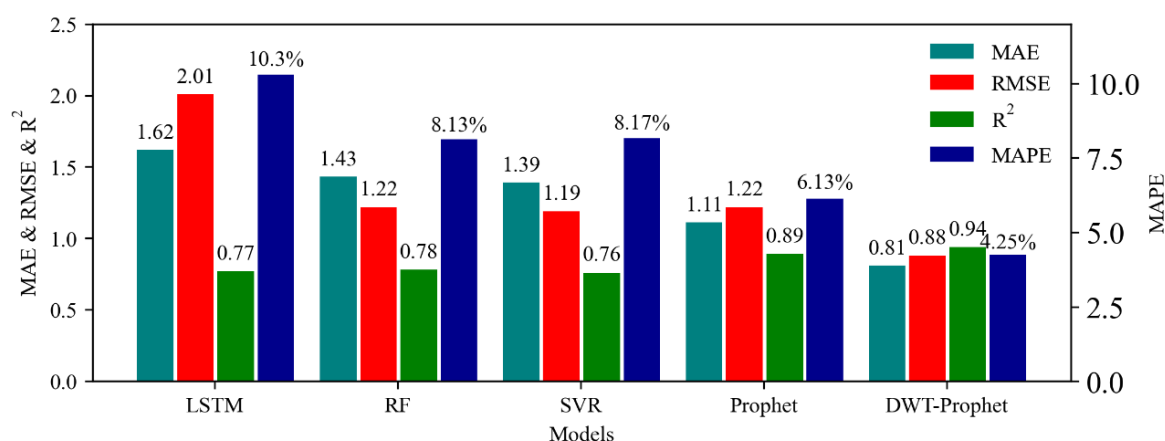
A Taylor diagram is an effective method to evaluate the prediction performance of prediction models intuitively. As shown in Figure 12, the DWT–Prophet had the smallest RMSE and the largest correlation coefficient; moreover, it can be easily seen that the DWT–Prophet prediction value was closest to the observation value, which indicates that the DWT–Prophet had the best prediction performance.



**Figure 12.** The Taylor diagram of the prediction errors results.

The MAPE, MAE, RMSE, and the  $R^2$  are usually used to evaluate the prediction performance of prediction models. The results of these four evaluation indexes are shown in Figure 13. The MAPE, MAE, and RMSE of the DWT–Prophet combination model were 4.25%, 0.81%, and 0.88% lower than those of the comparison models and the  $R^2$  of the DWT–Prophet was 0.94, higher than that of the comparison models, which means that the prediction performance of the proposed hybrid model was better than the comparison models. Therefore, it can be concluded that the proposed hybrid model can fit the nonlinear trend of vessel traffic flow effectively and had a good forecast performance when employed to predict vessel traffic flow in the future.





**Figure 13.** Comparison of the prediction errors for the different methods.

#### 4. Conclusions

Reliable vessel traffic flow prediction is significant for the improvement of vessel navigation efficiency and maritime traffic safety. Aimed at the nonlinearity and uncertainty in vessel traffic flow, this paper proposed a hybrid forecast model for vessel traffic flow based on the discrete wavelet decomposition and Prophet framework fusion. Discrete wavelet decomposition was employed to decompose the vessel traffic flow data into one low-frequency component and three high-frequency components, and the Prophet framework was introduced to the prediction for every component separately; in this way, the final accurate prediction result was obtained by reconstructing the prediction result of every component. In order to test the prediction performance of the proposed hybrid model, we selected the Prophet framework, the LSTM, the random forest, and the SVR models as the comparison models. The prediction results demonstrated that the hybrid model of DWT-Prophet outperformed the comparison models and could effectively overcome the limitations of a single model.

Although the DWT-Prophet had strong predictive capabilities, the model only digs out the underlying laws from the data itself and cannot take the factors that affect vessel traffic flow, such as weather, hydrology, and policies, as input variables to further improve the accuracy of the prediction. In future research, models that can take influence factors as input variables will be considered for combined use with wavelet transform and Prophet to predict vessel traffic flow. At the same time, the amount of data will be increased to reduce accidental results.

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

**Funding:** This research was supported by the National Nature Science Foundation (No.: 51809207).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data in this study are not available publicly, but are available on request from the corresponding author.

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



## References

1. Yu, Q.; Liu, K.Z.; Chang, Z.L.; Yang, Z.L. Realising advanced risk assessment of vessel traffic flows near offshore wind farms, America. *Reliab. Eng. Syst. Saf.* **2020**, *203*, 107086. [\[CrossRef\]](#)
2. Han, C.; Song, S.; Wang, C. A real-time short-term traffic flow adaptive forecasting method based on ARIMA model. *Syst. Simul.* **2004**, *7*, 1530–1534.
3. Wang, Z.X.; Li, D.D.; Zheng, H.H. Model comparison of GM(1,1) and DGM(1,1) based on montecarlo simulation. *Phys. A Stat. Mech. Its Appl.* **2019**, *542*, 123341. [\[CrossRef\]](#)
4. Zhan, W.; Qian, S.; Yin, J. An improved GM(1,1) model for monthly electric power consumption forecasting. *South. Power Syst. Technol.* **2012**, *6*, 92–96.
5. Assimakis, N.; Adam, M.; Ktena, A.; Manasis, C. Steady state Kalman filter design for cases and deaths prediction of covid-19 in greece. *Results Phys.* **2021**, *26*, 104391. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Li, Q.; Liu, M.; Qian, F. Analysis and forecast of residential building energy consumption in Chongqing on carbon emissions. *J. Cent. South. Univ.* **2009**, *16*, 214–218.
7. Li, M.W.; Han, D.F.; Wang, W.L. Vessel traffic flow forecasting by RSVR with chaotic cloud simulated annealing genetic algorithm and KPCA. *Neurocomputing* **2015**, *157*, 243–255. [\[CrossRef\]](#)
8. Wei, Z.M.; Song, H.C. Prediction scheme of railway passenger flow based on multiplicative holt-winters model. *Appl. Mech. Mater.* **2013**, *416–417*, 1949–1953.
9. Guo, J.; Tu, L.; Qiao, Z.; Wu, L. Forecasting the air quality in 18 cities of Henan province by the compound accumulative grey model. *J. Clean. Prod.* **2021**, *310*, 127582. [\[CrossRef\]](#)
10. Wu, L.; Zhao, H. Using FGM(1,1) model to predict the number of the lightly polluted day in jing-jin-ji region of China. *Atmos. Pollut. Res.* **2019**, *10*, 552–555. [\[CrossRef\]](#)
11. Li, X.; Peng, L.; Yao, X.; Cui, S.; Hu, Y.; You, C.; Chi, T. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollut.-Lond. Barking* **2017**, *231*, 997–1004. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Mohamed, O.A.; Ati, M.; Kewalramani, M.; Hawat, W.A. Application of ANN for prediction of chloride penetration resistance and concrete compressive strength. *SSRN Electron. J.* **2021**, *17*, 101123. [\[CrossRef\]](#)
13. Tian, J.; Liu, Y.; Bi, H.; Li, F.; Bao, L.; Han, K.; Zhou, W.; Ni, Z.; Lin, Q. Experimental study on the spray characteristics of octanol diesel and prediction of spray tip penetration by ANN model. *Energy* **2021**, *239*, 121920. [\[CrossRef\]](#)
14. Hong, W.C. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy* **2011**, *36*, 5568–5578. [\[CrossRef\]](#)
15. Liu, Y.K.; Xie, F.; Xie, C.L.; Peng, M.J.; Wu, G.H.; Xia, H. Prediction of time series of npp operating parameters using dynamic model based on bp neural network. *Ann. Nucl. Energy* **2015**, *85*, 566–575. [\[CrossRef\]](#)
16. Ahmad, M.S.; Adnan, S.M.; Zaidi, S.; Bhargava, P. A novel support vector regression (SVR) model for the prediction of splice strength of the unconfined beam specimens. *Constr. Build. Mater.* **2020**, *248*, 118475. [\[CrossRef\]](#)
17. Joges, C.; Berkenbrink, C.; Stumpe, B. Prediction and reconstruction of ocean wave heights based on bathymetric data using LSTM neural networks. *Ocean Eng.* **2021**, *232*, 109046. [\[CrossRef\]](#)
18. Dong, X.J.; Shen, J.N.; He, G.X.; Ma, Z.F.; He, Y.J. A general radial basis function neural network assisted hybrid modelling method for photovoltaic cell operating temperature prediction. *Energy* **2021**, *234*, 121212. [\[CrossRef\]](#)
19. Hu, Y.L.; Liang, C. A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic elm and differential evolution algorithm. *Energy Convers. Manag.* **2018**, *173*, 123–142. [\[CrossRef\]](#)
20. Mohammed, E.D.; Salim, A.H.; Spiros, P. Hybrid harmonic analysis and wavelet network model for sea water level prediction. *Appl. Ocean Res.* **2018**, *70*, 14–21.
21. Ma, Q.; Wang, H.; Peng, Y.S.; Li, Q. Ultra-short-term Railway traction load prediction based on DWT-TCN-PSO\_SVR combined model. *Int. J. Electr. Power Energy Syst.* **2021**, *135*, 107595. [\[CrossRef\]](#)
22. Tian, Z.D.; Chen, H. A novel decomposition-ensemble prediction model for ultra-short-term wind speed. *Energy Convers. Manag.* **2021**, *248*, 114775. [\[CrossRef\]](#)
23. Han, Z.Z.; Li, J.A.; Hossain, M.M.; Qi, Q.; Zhang, B.; Xu, C. An ensemble deep learning model for exhaust emissions prediction of heavy oil-fired boiler combustion. *Fuel* **2021**, *308*, 121975.
24. He, K.J.; Ji, L.; Wu DW, C.; Tso GF, K. Using SARIMA–CNN–LSTM approach to forecast daily tourism demand. *J. Hosp. Tour. Manag.* **2021**, *49*, 25–33. [\[CrossRef\]](#)
25. Du, W.; Zhang, Q.; Chen, Y.; Ye, Z. San urban short-term traffic flow prediction model based on wavelet neural network with improved whale optimization algorithm. *Sustain. Cities Soc.* **2021**, *69*, 102858. [\[CrossRef\]](#)
26. Chen, X.; Chen, H.; Yang, Y.; Wu, H.; Xiong, Y. Traffic flow prediction by an ensemble framework with data denoising and deep learning model. *Phys. A Stat. Mech. Its Appl.* **2021**, *565*, 125574. [\[CrossRef\]](#)
27. Zhao, L.W.; Chang, D.F.; Zhu, Z.L.; Gao, Y.P. Port traffic volume forecast with SARIMA-BP model. *Navig. China* **2020**, *43*, 94.
28. Zhang, S.Y.; Yang, Y.H.; Chen, C.; Yang, Y.C.; Li, F.L. Spatio-temporal prediction of ship traffic flow in waterway regulation area based on regression Kalman filter combination mode. *J. Dalian Marit. Univ.* **2021**, *47*, 37–44.
29. Taylor, S.J.; Letham, B. Forecasting at scale. *PeerJ Prepr.* **2017**, *5*, e3190v2. [\[CrossRef\]](#)