** SciencePG**
Science Publishing Group

Research Article

Short-Term Traffic Flow Prediction Based on Wavelet Analysis and XGBoost

Xin Wang[*] , **Fang Fang**

School of Applied Science, Beijing Information Science and Technology University, Beijing, China

Abstract

Traffic flow prediction is of great significance for urban planning and alleviating traffic congestion. Due to the randomness and high volatility of urban road network short-term traffic flow, it is difficult for a single model to accurately estimate traffic flow and travel time. In order to obtain more ideal prediction accuracy, a combined prediction model based on wavelet decomposition and reconstruction (WDR) and the extreme gradient boosting (XGBoost) model is developed in this paper. Firstly, the Mallat algorithm is applied to perform multi-scale wavelet decomposition on the average travel time series of the original traffic data, and single branch reconstruction is performed on the components at each scale. Secondly, XGBoost is used to predict each reconstructed single-branch sequence, so as to obtain multiple sub-models, and the Bayesian algorithm is used to optimize the hyperparameters of the sub-models. Finally, the algebraic sum of the predicted values of all sub-models is used to obtain the overall traffic prediction result. To test the performance of the proposed model, actual traffic flow data has been collected from a certain link of the Brooklyn area in New York, USA. The performance of proposed WDR-XGBoost model has been compared with other existing machine learning models, e.g., support vector regression model (SVR) and single XGBoost model. Experimental findings demonstrated that the proposed WDR-XGBoost model performs better on multiple evaluation indicators and has significantly outperformed the other models in terms of accuracy and stability.

Keywords

Traffic Time Prediction, Wavelet Analysis, XGBoost, Bayesian Algorithm

1. Introduction

With the increasingly serious traffic congestion and traffic accidents in urban, the application of intelligent transportation system is more and more widely, involving information service, intelligent highway, operation management and so on. Traffic flow prediction is one of the key technologies to realize dynamic guidance system in intelligent transportation system, which is to use historical and real-time traffic data information to predict the traffic status of a specified road or area for a period of time in the future. The prediction content generally includes traffic flow, speed, density (or occupancy), travel time and other variables that reflect traffic status. It plays an important role in signal control, alleviating traffic congestion and improving the operation efficiency of transportation network. The parameters of traffic flow mainly include traffic volume, speed, density, etc. Short-term traffic flow predicting refers to the prediction of traffic volume in a short period (generally $< = 15$ minute) based on the current and past traffic flow data, so as to meet

the real-time requirements of traffic flow control and traffic guidance.

There are many methods are used for prediction in the field of transportation [1], mainly including: (1) models based on statistical analysis, such as autoregressive integrated moving average (ARIMA) model [2], Kalman filtering model etc. [3, 4]; (2) nonlinear theoretical models, such as chaos theory [5], catastrophe theory [6]; (3) simulation forecasting model [7]; (4) machine learning models, such as K-nearest neighbor, neural network, support vector regression, etc. [8-10]; (5) combined model.

In the prediction method based on statistical analysis, ARIMA is a commonly used time series model. Zhang [11] designed a research method for short-term passenger flow prediction of urban rail transit based on an improved ARIMA model. Aiming at the limitation of ARIMA in obtaining the nonlinear characteristics of time series, Wang [12] proposed a model combined the linear ARIMA algorithm and nonlinear generalized autoregressive conditional heteroscedasticity in mean (GARCH-M) algorithm considering the heteroscedasticity of the traffic flow time series. Li [13] used the chaos theory in the nonlinear theoretical model to determine the optimal delay time and embedding dimension of the original traffic flow time series, and obtained a more reasonable model data set with the same dynamic characteristics as the original data through phase space reconstruction. Based on long short-term memory neural network (LSTM), Weng [14] proposed a short-term traffic flow prediction method considering the characteristics of bus proportion. Wang (Wang et al., 2021) proposed a long short term memory (LSTM) model based on the encoder decoder (ED) framework, which can achieve multi-step prediction of traffic flow sequences. Subsequently, Wang [15] proposed a short-term passenger flow prediction model for road networks based on a multi-layer convolutional long short term memory neural network (ConvLSTM). The spatial characteristics of passenger flow are obtained through convolution operations in multi-layer ConvLSTM, and the temporal characteristics of passenger flow are obtained through the long short term memory part.

As the performance improvement of a single model gradually slows down, scholars are increasingly committed to the application of combined models in traffic flow forecasting [16-19]. Mallick [20] introduces transfer learning technology into traffic flow prediction. Song [21] uses the empirical mode decomposition (EMD) algorithm to decompose the traffic flow.

When the time interval is shortened to 5 ~ 15 minutes, the nonlinearity, randomness and time-varying characteristics of short-term traffic flow will be very obvious, which will reduce the accuracy and stability of the forecasting models [22, 23]. The wavelet analysis theory can project the original traffic flow data to different scales, then establish models and integrate them separately, so as to improve the forecasting accuracy [24, 25]. At the same time, the traditional models based

on wavelet analysis only model the data directly on individual scale after decomposition. Due to the different length of each scale sequence, it brings inconvenience and low accuracy to the calculation in the later forecasting and reconstruction. In recent years, the XGBoost algorithm has been introduced into the field of transportation research. It has shown significant advantages in solving problems such as slow convergence speed, overfitting, and susceptibility to local optima faced by models such as neural networks [26, 27]. Ye [28] proposed a short-term traffic flow prediction model based on convolutional neural network (CNN)-XGBoost, and the results showed that compared to SVR, LSTM and other models, the prediction error of traffic flow using XGBoost model was significantly reduced. Antypas [29] uses gradient boosting method to estimate the arrival time of autonomous vehicle. In a recent study in 2024, Chen et al. [30] utilizes a large dataset of taxi trajectories and uses advanced machine learning techniques, including XGBoost, to analyze the causes of imbalanced traffic networks.

Based on the above, a WDR-XGBoost short-term traffic flow predicting model is proposed. Firstly, the discrete wavelet decomposition of short-term traffic flow time series is carried out by using multi-scale wavelet analysis theory, and the approximate components and detail components are single-branch reconstructed to the original scale. Then the XGBoost model is used to forecast the reconstructed single-branch time series, and the Bayesian algorithm is used to obtain the optimal parameter combination of the model to accelerate the convergence speed. Finally, the predicting results of the traffic flow are obtained by summing the predicting values of all single-branch reconstruction sequences. The experimental results show that the WDR-XGBoost model can ultimately improve the short-term traffic flow predicting performance.

2. Problem Statement

Short-term traffic predicting at a certain time t refers to the real-time predicting of traffic characteristics in the next time $t + \Delta t$ even later, and the predictive time-span is not more than 15 minutes or less than 5 minutes. Generally, predicting is based on three basic parameters of macroscopic traffic flow: average traffic volume, average speed and average occupancy, average travel time et al. Those are the traffic characteristics from the perspective of the whole transportation system. In contrast, there are few studies regarding the average travel time of the link as the predicting result. Thus, the average travel time of the link refers to the average travel time of all vehicles through the observation link during a certain time interval, as shown in formula (1):

$$T(h) = \frac{1}{m} \sum_{i=1}^{m} t_i \qquad (1)$$

where $T(h)$ is the average link travel time in the hth time interval, m is the total number of vehicles passing through the link during the observation interval, t_i is the travel time of the ith vehicle through this link. Considering that the current road condition is related to the road condition in the past intervals, the average travel time in the past several intervals is used to forecast the average travel time in the future interval. Specifically, the average travel time $T(t)$ of the tth time interval on the target link can be described as a time series as follow:

$$T(t) = \{ T(t-1), T(t-2), \cdots, T(t-p) \} \tag{2}$$

where p is the time lag.

3. Research Methodology

3.1. Research Design Wavelet Transform and Multi-scale Analysis

The average travel time of a link can be regarded as a kind of time-varying information, which is nonlinear and random, and there will be noise interference. If the modeling and forecasting is directly based on the original data, it is obviously impossible to ensure the forecasting accuracy. Therefore, it is necessary to adopt some methods to extract noise and minimize its impact on the forecasting results.

Wavelet transform is a multi-scale time-frequency analysis tool for time series scaling in frequency domain and translation in time domain. In any space $L^2(\mathbf{R})$, the function $f(t)$ is expanded by a wavelet function, which is called continuous wavelet transform (CWT), and the expression is

$$W_f(a,b) = |a|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \overline{\psi \left(\frac{t-b}{a} \right)} dt = \left\langle f(t), \psi_{a,b}(t) \right\rangle \tag{3}$$

In equation (3), a is the scaling factor and b is the translation factor; $a,b \in \mathbf{R}$ and $a \neq 0$. $\psi(t)$ is basic wavelet; $\overline{\psi(t)}$ is the conjugate function; $\psi_{a,b}(t)$ is a family of functions generated by scaling and translation of basic wavelet, which is called wavelet function.

However, continuous wavelet transform will calculate the wavelet coefficients on all scales, which takes time and generates redundant data. Therefore, discrete wavelet transform (DWT) is usually used in the practical application. Discrete wavelet transform is the discretization of scaling factor a and translation factor b, namely

$$a = a_0^{-j}, \quad b = ka_0^{-j}b_0 \tag{4}$$

In equation (4), $a_0 > 1$, $b_0 > 0$, $\forall j,k \in Z$. Then the wavelet function is

$$\psi_{j,k}(t) = a_0^{\frac{j}{2}} \psi \left(a_0^{\,j} t - kb_0 \right) \tag{5}$$

Discrete wavelet transform is

$$W_f(j,k) = a_0^{\frac{j}{2}} \int_{-\infty}^{+\infty} f(t) \overline{\psi \left(a_0^{\,j} t - kb_0 \right)} dt = \left\langle f(t), \psi_{j,k}(t) \right\rangle \tag{6}$$

Then any function $f(t) \in L^2(\mathbf{R})$ can be expressed as a wavelet series

$$f(t) = \sum_{j,k} c_{j,k} \psi_{j,k}(t) \tag{7}$$

Where $c_{j,k} = \left\langle f(t), \psi_{j,k}(t) \right\rangle$ is called wavelet coefficients.

At the same time, according to multi-scale analysis, Mallat algorithm can be used to select appropriate scale function $\varphi_{j,k}(t)$ and wavelet function $\psi_{j,k}(t)$, and the corresponding decomposition coefficient sequence $\{a_n\}$, $\{b_n\}$ to reconstruct the coefficient sequence $\{p_n\}$, $\{q_n\}$. The original traffic flow data can be decomposed into low-frequency approximate component A_j and high-frequency detail component D_j at a certain scale j by low-pass filter and high-pass filter, namely

$$\begin{cases} A_j = \sum_k c_{j,k} \varphi \left(2^j t - k \right) \\ D_j = \sum_k d_{j,k} \psi \left(2^j t - k \right) \end{cases} \tag{8}$$

The noise part is usually contained in the high-frequency detail component. The low-frequency approximate component A_j can be further decomposed as

$$A_j = A_{j+1} + D_{j+1} \tag{9}$$

This decomposition process can be iterative many times until it reaches the maximum number set in advance. If j-scale decomposition is performed, the formula is as follows:

$$\begin{cases} c_{j,k} = \sum_l a_{l-2k} c_{j-1,l} \\ d_{j,k} = \sum_l b_{l-2k} c_{j-1,l} \end{cases} \tag{10}$$

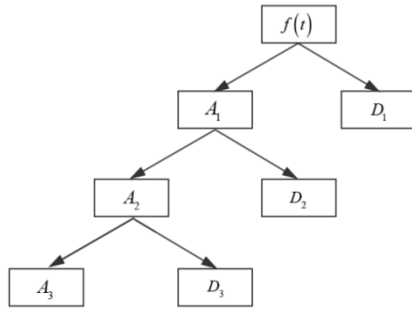The wavelet decomposition of 3-scale is shown in Figure 1:

Figure 1. *Wavelet decomposition of Mallat 3-scale.*

Then, the approximate component and the detail component are single-branch reconstructed to the original scale separately. The formula is as follow:

$$\tilde{c}_{j,k} = \sum_l \left(p_{k-2l} c_{j+1,l} + q_{k-2l} d_{j+1,l} \right) \tag{11}$$

Thus, the approximate component A_J^r and the detail components $D_1^r, D_2^r, \cdots, D_J^r$ can be obtained. Finally, approximate component and all detail components are algebraically added, and the reconstructed series is

$$f'(t) = A_J^r + \sum_{i=1}^{J} D_i^r \tag{12}$$

3.2. Principle of XGBoost Model

XGBoost is an integrated learning model, which can improve machine learning effect and forecasting accuracy by constructing multiple base learners. For data set $\{(X_i, y_i), i=1, 2, \cdots, N\}$, $X_i \in \mathbf{R}^m$ is the eigenvector of i th instance and y_i is the attribute value of X_i. XGBoost model is defined as the following addition model, and its base classifiers are classification and regression tree (CART):

$$L = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{13}$$

$$\hat{y}_i = \sum_{k=1}^{K} f_k(X_i), f_k \in \Gamma \tag{14}$$

$$\Gamma = \left\{ f_k(X) = \mathbf{w}_{q(X)}, k=1, 2, \cdots, K \right\}$$
$$\left(q : \mathbf{R}^m \to \{1, 2, \cdots, T\}, \mathbf{w} = \{w_1, \cdots, w_j\} \right) \tag{15}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 \tag{16}$$

In the above equation (13)-(15): L is the optimization function of XGBoost model and $l(y_i, \hat{y}_i)$ is the error function between the forecasting value $\hat{y}_i$ and the actual value y_i corresponding to X_i; $\sum_{k=1}^{K} \Omega(f_k)$ is the sum of the complexity of all K trees, which is added to the objective function as a regularization term to effectively prevent over fitting; γ and λ are regularization parameters. Γ is a set of CART. $f_k \in \Gamma$ is a tree model which means mapping a eigenvector of instance to the corresponding a leaf node; T is the number of leaf nodes of tree; $\mathbf{w} = \{w_1, \cdots, w_j\}$ represents the weight vector of all leaf nodes of a tree.

In this study, the forward step-by-step algorithm is used. In the tth step, the second-order Taylor expansion of the objective function is carried out, and the expression is as follow:

$$L^{(i)} = \sum_{i=1}^{N} \left[l\left(y_i, \hat{y}^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{17}$$

$$\begin{cases} g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right) \\ h_i = \partial_{\hat{y}^{(t-1)}}^2 l\left(y_i, \hat{y}^{(t-1)}\right) \end{cases} \tag{18}$$

After omitting the constant term of the first $t-1$ steps in equation (17), let $I_j = \{i | q(X_i) = j\}$ be the set of all samples belonging to the jth leaf node. By minimizing this formula, the optimal weight w_j^* of the jth leaf node of the tth sub-model and the corresponding optimal objective function value can be obtained by:

$$\begin{cases} w_j^* = -\dfrac{\sum\limits_{i \in I_j} g_i}{\sum\limits_{i \in I_j} h_i + \lambda} \\ obj = -\dfrac{1}{2} \sum\limits_{j=1}^{T} \dfrac{\left(\sum\limits_{i \in I_j} g_i\right)^2}{\sum\limits_{i \in I_j} h_i + \lambda} + \gamma T \end{cases} \tag{19}$$

Furthermore, the structure of the tree is determined by the information gain before and after segmentation:

$$Gain = \frac{1}{2} \left[\frac{\left(\sum\limits_{i \in I_L} g_i\right)^2}{\sum\limits_{i \in I_L} h_i + \lambda} + \frac{\left(\sum\limits_{i \in I_R} g_i\right)^2}{\sum\limits_{i \in I_R} h_i + \lambda} - \frac{\left(\sum\limits_{i \in I} g_i\right)^2}{\sum\limits_{i \in I} h_i + \lambda} \right] - \gamma \tag{20}$$

3.3. Bayesian Optimization Algorithm

Most widely adopted ways of parameter optimization are the so-called random parameters approach. However, grid search is slow, and random search is easy to miss some important information. Using the above methods to optimize parameter simply has many drawbacks, so the Bayesian optimization algorithm with fewer iterations, faster speed and better generalization ability is adopted. The idea of Bayesian optimization approach is an optimization algorithm based on probability distribution, and it finds the minimum value of objective function based on past evaluation results of objective function. The core process is surrogate function and acquisition function. Suppose a set of parameter combinations is $\mathbf{x} = (x_1, x_2, \cdots, x_n) \in \chi$, and the objective function $f : \chi \to \mathbf{R}^d$ is a black box with no analytics and higher evaluation cost. The optimal value satisfying the following equation (21) needs to be found:

$$\mathbf{x}^* = \arg\min_{\mathbf{x} \in \chi} f(\mathbf{x}) \tag{21}$$

For an optimization problem, there are given objective function f, parameter search space χ, surrogate function M, acquisition function S, number of iterations T. Then the algorithm steps can be expressed as:

Step1. Initialize several groups of parameters randomly and obtain the data set $D = \{(x_1, y_1), \cdots, (x_{t-1}, y_{t-1})\}$, where $y_i = f(\mathbf{x}_i)$;

Step2. Fit data set with surrogate function and selecting the optimum $\mathbf{x}^*$ with acquisition function;

$$\arg\max_{\mathbf{x} \in \chi} S(\mathbf{x}, M_{t-1}) \to \mathbf{x}^* \tag{22}$$

$$f(\mathbf{x}^*) \to y^* \tag{23}$$

Step3: Update the data set D by

$$D \cup (\mathbf{x}^*, y^*) \to D \tag{24}$$

Go back to step2 until the maximum number of iterations.

The Bayesian optimization process is shown in Figure 2. Among them, the surrogate function uses the tree-structured Parzen estimator (TPE), which is a Gaussian mixture model. The Bayesian rule is applied to construct the surrogate function $p(y|\mathbf{x})$, where $p(\mathbf{x}|y)$ is defined as:

$$p(\mathbf{x}|y) = \begin{cases} l(\mathbf{x}) & \text{if } y < y^* \\ g(\mathbf{x}) & \text{if } y \geq y^* \end{cases} \tag{25}$$

$l(\mathbf{x})$ and $g(\mathbf{x})$ are the generation models of all domain variables, and y^* is the specific quantile. The Eq. (25) means that two different distributions are made for the parameters.
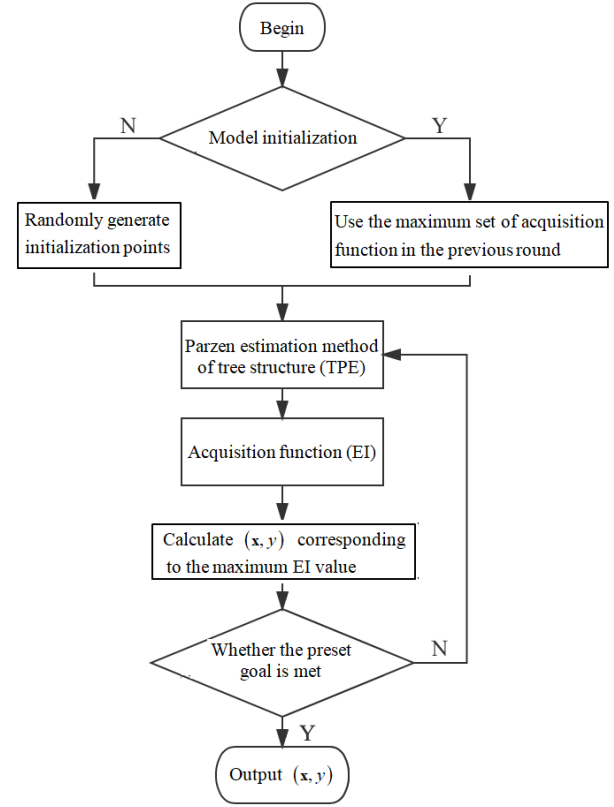

Figure 2. Calculation diagram of Bayesian optimization.

The acquisition function uses expected improvement method (EI) because EI is intuitive and has been proved to work well in various environments. EI is defined as:

$$\text{EI}_{y^*}(\mathbf{x}) = \int_{-\infty}^{y^*} (y^* - y) p(y|\mathbf{x}) dy \tag{26}$$

where y^* is the threshold of the objective function; y is the actual value of the objective function corresponding the parameter combination $\mathbf{x}$; $p(y|\mathbf{x})$ is an surrogate function expressed in probability. If the Eq. (26) is positive, it means that the parameter combination $\mathbf{x}$ is expected to produce better results than the threshold.

4. Numerical Example

The Materials and Methods section should provide comprehensive details to enable other researchers to replicate the study and further expand upon the published results. If you have multiple methods, consider using subsections with appropriate headings to enhance clarity and organization.

4.1. Data Collection and Preprocessing

The dataset in this paper is from the real-time traffic data of New York City (https://data.beta.nyc/dataset/nyc-real-time-traffic-speed-data-feed-archived), and the data source contains the real-time traffic information collected by the Ministry of transport in five administrative regions, which are mainly distributed on main roads and expressways. The total number of recorded from the 1st May to the 30th June, 2020 was 17480 with an interval of 5 minutes for the link-4616271 in Brooklyn administrative region. The missing values are completed by time series mean interpolation:

$$T(t) = \frac{T(t-1) + T(t+1)}{2} \qquad (27)$$

The descriptive statistics of average travel time series of link are shown in Table 1.

Table 1. Descriptive statistics of average travel time series of link.

Statistic	Average travel time(s)
Sample size	17480
Mean	503.22
Standard deviation	358.06
Maximum	3782
Minimum	249

Split the original data set, as shown in Figure 3. Take the data from the 1st May to the 20th June as the training set and the data from the 21th June to the 30th June as the test set.
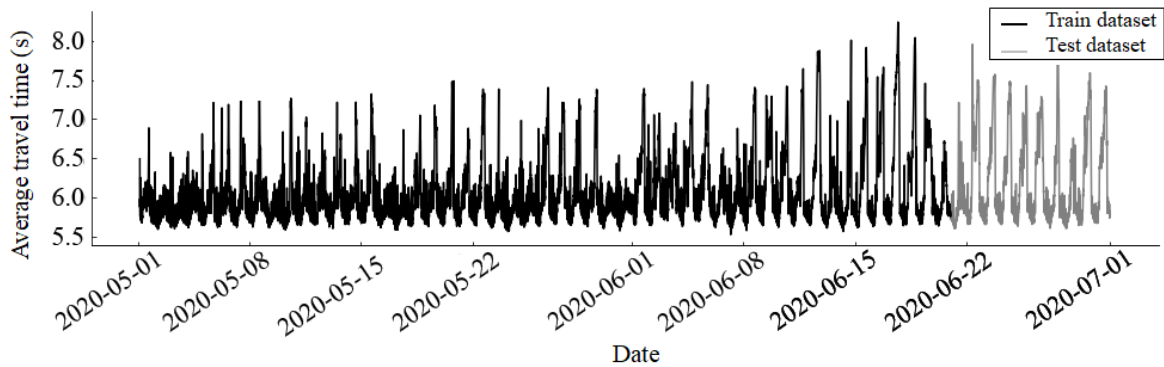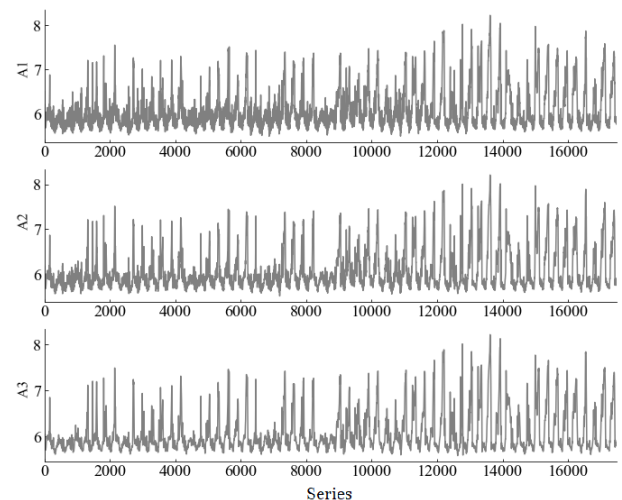

Figure 3. Schematic diagram of data set division.

When using discrete wavelet transform, there are two points worth attention. The first is to determine the type of basic wavelet. The commonly used basic wavelets include Haar wavelet, db wavelet, sym wavelet, coif wavelet and Morlet wavelet, which all belong to a wavelet family. Each wavelet family contains many specific wavelets. The second is the wavelet decomposition scale J, because with the increase of decomposition scale, the loss of information is more. These two points do not have specific selection instructions, and are mostly based on experience and multiple experimental results. In this paper, db5 is selected as the basic wavelet, which is one of the commonly used wavelets in db wavelet family. It indicates that the maximum decomposition scale is 5 in the process of wavelet transform. As shown as Figure 4, when the decomposition scale is 3, the approximate component A_3 can better show the trend of the original time series.

At the same time, the noise reflected in the detail component is removed, so the decomposition scale is finally selected as 3.
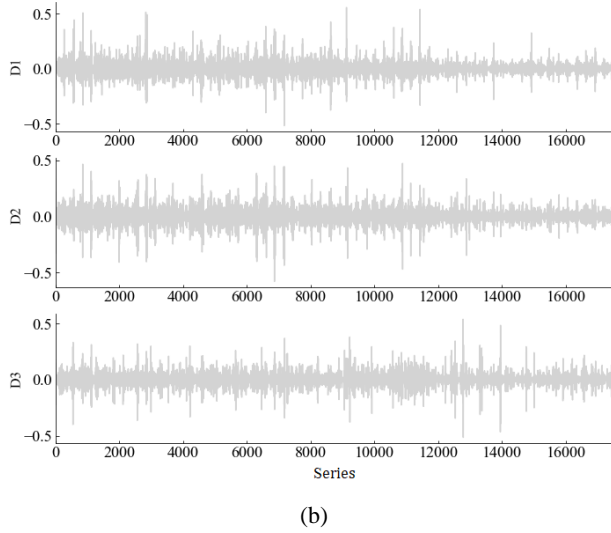

(a)

(b)

Figure 4. *Average travel time by 3-level wavelet decomposition.*

Since the length of the series after wavelet decomposition is different from that of the original series, the single-branch reconstruction of the approximate component A_3 and detail component D_1, D_2, D_3 after decomposition are shown in Figure 5.
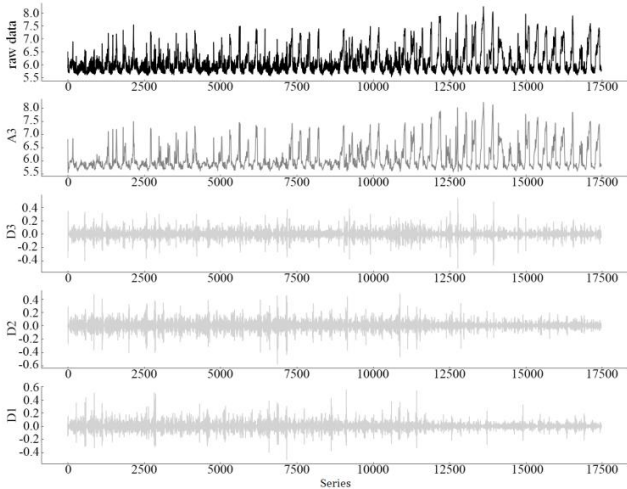

Figure 5. *The original average travel time and the single reconstruction data by three-level wavelet decomposition.*

4.2. Analysis and Extraction of Features

Since the short-term traffic flow forecasting is mainly based on the average travel time of the previous moments to forecast the average travel time of the next moment, thus the data of some moments in a certain day or a week will also affect the forecasted results.
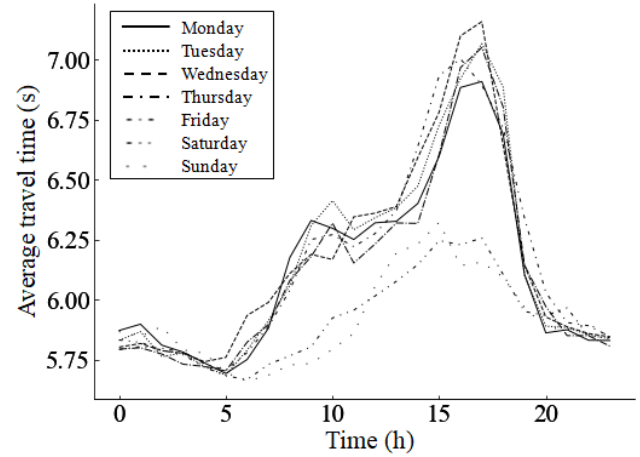

Figure 6. *Basic characteristics of average travel time.*

As can be seen from Figure 6, the change trends of average travel time on Saturday and Sunday are similar because they are both weekends. Similarly, from Monday to Friday, the change trends of average travel time are similar because they are all workdays. In the morning and evening peak period, the average travel time has an obvious upward trend. Therefore, the expression of the average travel time series can be described as

$$T(t) = \left\{ T(t-1), T(t-2), \cdots, T(t-p), H, D \right\} \qquad (28)$$

where H is the hourly feature and D is the weekly feature, and the time lag is $p = 5$.

4.3. Evaluation Index

Four indexes that aim to evaluate forecasting performance are used, i.e. root mean squared error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), and determination coefficient R^2. Both MAE and RMSE can measure absolute errors between the actual and forecasting values while the MAPE is employed to evaluate the relative errors of them. The closer R^2 to 1, the higher the forecasting performance is. These indexes are mathematically represented as Eqs. (29) to (32):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_i - \hat{y}_i \right)^2} \qquad (29)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \qquad (30)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} \left| y_i - \hat{y}_i \right| \qquad (31)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2} \qquad (32)$$

N is the total number of testing data used in prediction, $\hat{y}_i$ is the ith forecasting value, y_i is the ith actual value, and $\overline{y}$ is the mean of all actual value. The forecasting results of four WDR-XGBoost sub-models are shown in Table 2.
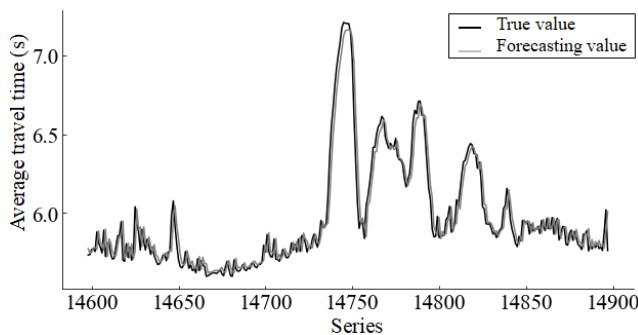
Table 2. *Performance of four WDR-XGBoost sub-models.*

Evaluation indexes	A₃	D₃	D₂	D₁
RMSE	0.0173	0.01	0.01	0.0141
MAPE	0.1738	78.8512	180.7309	199.0863
MAE	0.0113	0.0065	0.0077	0.0086
R^2	0.9990	0.9526	0.8844	0.8140

At the same time, the forecasting effects of SVR, XGBoost and WDR-XGBoost are compared, and evaluation indexes and forecasting results are as shown in Table 3 and Figure 7.

Table 3. *Comparison of forecasting performance of each modes.*

Evaluation indexes	SVR	XGBoost	WDR-XGBoost
RMSE	0.07791	0.06565	0.02627
MAPE	0.90762	0.78653	0.30937
MAE	0.05648	0.04837	0.01941
R^2	0.98073	0.98633	0.99781

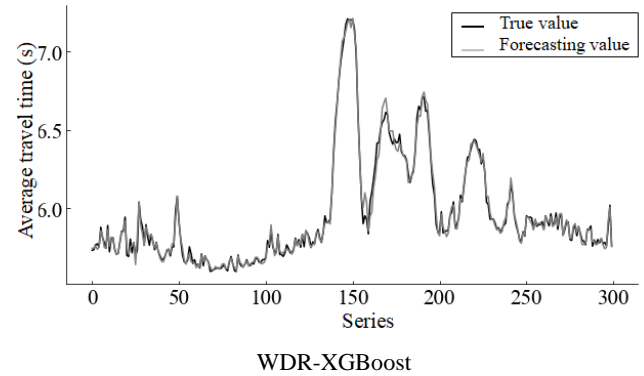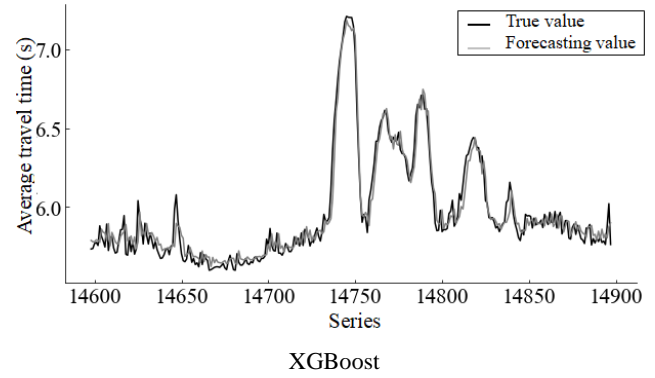SVR (C=9.818, gamma=0.65, kernel=linear)

XGBoost

WDR-XGBoost

Figure 7. *Comparison of forecasting results of three models.*

According to Table 3 and Figure 7, comparing the traffic flow forecasting performance of WDR-XGBoost model proposed in this paper with SVR model and XGBoost model, it can be intuitively found that the forecasting value of WDR-XGBoost model is closer to the real value.

5. Conclusions

It is difficult for a single model to obtain ideal accuracy and stability when predicting short-term traffic characteristics with complex nonlinearity and randomness. To solve this problem, a short-term travel time predicting model based on multi-scale WDR-XGBoost is proposed and the experimental results show that:

1) The db5 basic wavelet is used to decompose and reconstruct the original traffic data, and the predicting performance is improved. It shows that the discrete wavelet transform has the ability to completely reproduce the original data, and realizes the accurate predicting of time series data. By reconstructing the decomposed sequence to the original scale, each scale sequence has the same length as the original sequence, which simplifies the predicting process and improves the accuracy.

2) Compared with linear model and single XGBoost model, WDR-XGBoost algorithm has higher predicting accuracy, smaller error and faster training speed, which meets the timeliness requirements of short-term traffic

predicting.

3) However, the traffic flow state is affected by many factors in real life, such as weather conditions, road conditions, unexpected accidents, holidays and the spatial-temporal correlation of upstream and downstream links. This paper has not considered the above factors. At the same time, the experimental data are only for a single link, without considering other regions. In future research, more sufficient data collection and model validation are needed.

Abbreviations

WDR-XGBoost	Wavelet Decomposition and Reconstruction and the Extreme Gradient Boosting
SVR	Support Vector Regression Model
ARIMA	Autoregressive Integrated Moving Average model
GARCH-M	Generalized Autoregressive Conditional Heteroscedasticity in Mean Algorithm
LSTM	Long Short-Term Memory Neural Network
ED	Encoder Decoder
ConvLSTM	Convolutional LONG Short Term Memory Neural Network
EMD	Empirical Mode Decomposition
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
CART	Classification and Regression Tree
TPE	Tree-structured Parzen Estimator
EI	Expected Improvement Method
RMSE	Root mean squared error
MAPE	Mean absolute percentage error
MAE	Mean absolute error

Author Contributions

Xin Wang: Supervision, Validation, Writing – original draft, Writing – review & editing

Fang Fang: Methodology, Software, Visualization

Data Availability Statement

The data that support the findings of this study are openly available at: https://data.beta.nyc/dataset/nyc-real-time-traffic-speed-data-feed-archived.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Zhao, H., Zhai, D. M., Shi, Z. H. Review of short-term traffic flow forecasting models. Urban Rapid Rail Transit. 2019, 32(4), 50-54. https://doi.org/10.3969/j.issn.1672-6073.2019.04.009

[2] Li, W., L, J. Z., Wang, T. Improved ARIMA model traffic flow prediction method based on box-cox exponential transformation. Journal of Wuhan University of Technology (Transportation Science & Engineering). 2020, 44(6), 974-977. https://doi.org/10.3963/j.issn.2095-3844.2020.06.006

[3] Zhou, T. Jiang, D. Lin, Z. et al. Hybrid dual Kalman filtering model for short-term traffic flow forecasting. IET Intelligent Transport Systems. 2019, 13(6), 1023-1032. https://doi.org/10.1049/iet-its.2018.5385

[4] Cai, L. Zhang, Z. Yang, J. et al. A noise-immune Kalman filter for short-term traffic flow forecasting. Physica A: Statistical Mechanics and its Applications. 2019, 536, 122601. https://doi.org/10.1016/j.physa.2019.122601

[5] Liao, R. H., Lan, S., Liu, Z. X. Short-term traffic flow forecasting based on local prediction method in chaotic time series. Computer Technology and Development. 2015, 25 (1), 1-5. https://doi.org/10.3969/j.issn.1673-629X.2015.01.001

[6] Hu, J. R., He, L. Freeway Traffic flow condition criterion method based on cusp catastrophe theory. China Journal of Highway and Transport. 2017, 30(10), 137-144. https://doi.org/10.19721/j.cnki.1001-7372.2017.10.017

[7] Ma, Q. BP neural network short-term traffic flow prediction based on improved particle swarm optimization. Computer Simulation. 2019, 36(4), 94-98+323. https://doi.org/10.3969/j.issn.1006-9348.2019.04.020

[8] Liu, Z., Du, W., Yan, D. M., et al. Short-term traffic flow forecast based on combination of k-nearest neighbor algorithm and support vector regression. Journal of Highway and Transportation Research and Development. 2017, 34 (5), 122-128+158. https://doi.org/10.3969/j.issn.1002-0268.2017.05.017

[9] Yuan, H., Chen, Z. H. Short-term traffic flow prediction based on temporal convolutional networks. Journal of South China University of Technology (Natural Science Edition). 2020, 48(11), 107-113+122. https://doi.org/10.12141/j.issn.1000-565X.190276

[10] Fu, C. H., Yang, S. M., Zhang, Y. Promoted short-term traffic flow prediction model based on deep learning and support vector regression. Journal of Transportation Systems Engineering and Information Technology. 2019, 19(4), 130-134+148. https://doi.org/10.16097/j.cnki.1009-6744.2019.04.019

[11] Zhang, G. Y., Jin, H. Research on the prediction of short-term passenger flow of urban rail transit based on improved ARIMA model. Computer Applications and Software. 2022, 39(1), 339-344. https://doi.org/10.3969/j.issn.1000-386x.2022.01.052

[12] Wang, X. Q., Shao, C., F., Yin, C., Y., et al. Short-term traffic flow forecasting method based on ARIMA-GARCH-M model. Journal of Beijing Jiaotong University. 2018, 42(4), 83-88. https://doi.org/10.11860/j.issn.1673-0291.2018.04.011

[13] Li, Q. R., Chi, W. Y., Chen, L., et al. Short-term traffic flow forecast based on phase space reconstruction and PSO-GPR. Journal of Transport Information and Safety. 2019, 37(2), 70-76. https://doi.org/10.3963/j.issn.1674-4861.2019.02.010

[14] Weng, X. X. Hao, Y. Short-term traffic flow prediction based on LSTM algorithm with the characteristics of passenger car proportion. Journal of Chongqing Jiaotong University (Natural Science). 2020, 39(11), 20-25, 50. https://doi.org/10.3969/j.issn.1674-0696.2020.11.03

[15] Wang, B. W., Wang, J. S., Wang, T. Y., Zhang, Z. Q., Liu Y., Yu, H. An encoder-decoder multi-step traffic flow prediction model based on long short-time memory network. Journal of Chongqing University. 2021, 44(11), 71-80. https://doi.org/10.11835/j.issn.1000-582X.2021.160

[16] Wang, Q., Li, Y., Zhang, S. T., Zhang, L. Y. Research on short-term passenger flow prediction of urban rail transit based on multilayer convolution long and short-term memory neural network. Modern Urban Transit. 2023, 9, 95-99.

[17] Bui, K. N., Cho, J. Yi, H. Spatial temporal graph neural network for traffic forecasting: an overview and open research issues. Applied Intelligence. 2022, 52(3), 3763-3774. https://doi.org/10.1007/s10489-021-02587-w

[18] Li, L., Bi, J., Yang, K., et al. MGC-GAN: multi-graph convolutional generative adversarial networks for accurate citywide traffic flow prediction. International Conference on Systems, Man, and Cybernetics. IEEE, Czech, 2022; pp. 2557-2562.

[19] Aljuaydi, F., Wiwatanapataphee, B., Wu, Y. H. Multivariate machine learning-based prediction models of freeway traffic flow under non-recurrent events. Alexandria Engineering Journal. 2023, 65, 151-162. https://doi.org/10.1016/j.aej.2022.10.015

[20] Mallick, T., Balaprakash, P., Rask, E., et al. Transfer learning with graph neural networks for short-term highway traffic forecasting. The 25th International Conference on Pattern Recogintion. IEEE, Italy, 2020; pp. 10367-10374. https://doi.org/10.48550/arXiv.2004.08038

[21] Song, X. D., Ren, M. X. The short-term traffic flow prediction based on combination model. Computer Simulation. 2022, 39(7), 156-160. https://doi.org/10.3969/j.issn.1006-9348.2022.07.029

[22] Teresa, P. Impact of data loss for prediction of traffic flow on an urban road using neural networks. IEEE Transactions on Intelligent Transportation Systems. 2019, 20(3), 1000-1009. https://doi.org/10.1109/TITS.2018.2836141

[23] Zhan, H. Y., Gomes, G. Li, X. S., et al. Consensus ensemble system for traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems. 2018, 19(12), 3903-3914. https://doi.org/10.1109/TITS.2018.2791505

[24] Zheng, Z. B., Yang, Y. T., Liu, J. H., et al. Deep and embedded learning approach for traffic flow prediction in urban informatics. IEEE Transactions on Intelligent Transportation Systems. 2019, 20(10), 3927- 3939. https://doi.org/10.1109/TITS.2019.2909904

[25] Mou, Z. H., Li, K. P., Shen, D. F. Short-term traffic flow prediction based on wavelet denoising and Bayesian neural network model. Science Technology and Engineering. 2020, 20(33), 13881-13886. https://doi.org/10.3969/j.issn.1671-1815.2020.33.052

[26] Liu, B., Wu, Z. D., Yang, J. Y. Research on bio-intelligence algorithm optimized wavelet neural network and its plication in traffic flow prediction. Journal of Beijing Jiaotong University. 2020, 44(5), 17-26. https://doi.org/10.11860/j.issn.1673-0291.20200015

[27] Zhong, Y., Shao, Y. M., Wu, W. W. et al. Short-term traffic flow prediction model based on XGBoost. Science Technology and Engineering. 2019, 19(30), 337-342. https://doi.org/10.3969/j.issn.1671-1815.2019.30.050

[28] Ye, J., Li, L. J., Tang, Z X. Short-term traffic flow forecasting based on CNN-XGBoost. Computer Engineering and Design. 2020, 41(4), 1080-1086. https://doi.org/10.16208/j.issn1000-7024.2020.04.030

[29] Antypas, E., Spanos, G., Lalas, A., Votis, K., Tzovaras, D. A time-series approach for estimated time of arrival prediction in autonomous vehicles. Transportation Research Procedia. 2024, 78, 166-173. https://doi.org/10.1016/j.trpro.2024.02.022

[30] Chen, B. Y., Chen, X. Y., Chen, H. P., Huang, Y. B., Jia, T., Lam. W. H. Understanding user equilibrium states of road networks: Evidence from two Chinese mega-cities using taxi trajectory mining. Transportation Research Part A: Policy and Practice. 2024, 180(1), 103976. https://doi.org/10.1016/j.tra.2024.103976