

## Research Article

# Urban Traffic Travel Time Short-Term Prediction Model Based on Spatio-Temporal Feature Extraction

Leilei Kang,<sup>1</sup> Guojing Hu,<sup>2</sup> Hao Huang,<sup>1</sup> Weike Lu,<sup>3</sup> and Lan Liu<sup>1,4</sup> 

<sup>1</sup>School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, Sichuan 610031, China

<sup>2</sup>Department of Mathematics and Statistical Sciences, Jackson State University, Jackson 39217, MS, USA

<sup>3</sup>Alabama Transportation Institute, The University of Alabama, Tuscaloosa 35487, AL, USA

<sup>4</sup>National United Engineering Laboratory of Integrated and Intelligent Transportation, Chengdu, Sichuan 610031, China

Correspondence should be addressed to Lan Liu; [jianan\\_l@swjtu.edu.cn](mailto:jianan_l@swjtu.edu.cn)

Received 1 February 2020; Revised 4 July 2020; Accepted 30 July 2020; Published 14 August 2020

Academic Editor: Ludovic Leclercq

Copyright © 2020 Leilei Kang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the accuracy of short-term travel time prediction in an urban road network, a hybrid model for spatio-temporal feature extraction and prediction of urban road network travel time is proposed in this research, which combines empirical dynamic modeling (EDM) and complex networks (CN) with an XGBoost prediction model. Due to the highly nonlinear and dynamic nature of travel time series, it is necessary to consider time dependence and the spatial reliance of travel time series for predicting the travel time of road networks. The dynamic feature of the travel time series can be revealed by the EDM method, a nonlinear approach based on Chaos theory. Further, the spatial characteristic of urban traffic topology can be reflected from the perspective of complex networks. To fully guarantee the reasonability and validity of spatio-temporal features, which are dug by empirical dynamic modeling and complex networks (EDMCN), for urban traffic travel time prediction, an XGBoost prediction model is established for those characteristics. Through the in-depth exploration of the travel time and topology of a particular road network in Guiyang, the EDMCN-XGBoost prediction model's performance is verified. The results show that, compared with the single XGBoost, autoregressive moving average, artificial neural network, support vector machine, and other models, the proposed EDMCN-XGBoost prediction model presents a better performance in forecasting.

## 1. Introduction

Intelligent transportation system (ITS) is currently the most effective technical solution to improve public transportation service and management [1, 2]. The successful application of ITS is inseparable from the accurate identification and prediction of urban traffic status, usually measured by travel time. The advantage of travel time lies in that it can be easily understood by road administrators and users [3]. Also, managers can take precautions and actively intervene in the state of urban traffic from the overall situation, and users can choose their driving routes according to the released information. Previous studies have proved that using travel time as an indicator can better guide users and make the traffic flow in the peak period reach a state of balanced distribution in the road network [4, 5]. However, the prediction of urban road network travel time is a complicated

and challenging work, mainly reflected in the acquisition and analysis of the spatio-temporal characteristics of the road network. Therefore, the prediction of urban traffic conditions has attracted much attention from scholars committed to fully exploring public transportation information, grasping the law of urban traffic situation and development and providing real-time and accurate traffic information for decision-makers [6]. Initially, traditional statistical models, including autoregressive integrated moving average (ARIMA) [7] and Kalman filter [8], were leveraged to forecast the traffic state. However, the statistical methods are based on linear assumptions about data, while the actual traffic data are usually nonlinear. Thus, state prediction has gradually shifted from the previous statistical models to the current data-driven models [9], which mainly includes machine learning and deep learning methods. Both machine learning algorithms such as artificial neural

network [10], Bayesian network [11], the  $k$ -nearest neighbor [12], and support vector machine [13], as well as deep learning models such as deep confidence network [14], and long short-term memory models [15], have achieved excellent results in the field of traffic prediction. Furthermore, some prediction approaches have evolved in recent years based on multiinput multioutput supervised regression (M-SVR and M-KNN, for instance), and these approaches strive to capture the spatio-temporal relationship within the regression process [16, 17]. Some studies also intend to capture traffic states' spatial dependencies through the graph convolutional network and learn the dynamic changes of traffic data by the recursive neural network to capture the time dependence of traffic states [18]. Although these models perform well in describing the influence of spatio-temporal correlation on traffic state evolution, detailed explanations of temporal and spatial feature information are still lacking [19–21].

Because various factors affect the travel time of the road network, only by thoroughly understanding these factors can we better plan and manage urban traffic. To better complement the point, this study conducts an exploratory analysis of traffic factors to understand the physical mechanisms of the system by introducing a spatio-temporal approach that only uses the nonlinear method, which includes empirical dynamic modeling (EDM) [22, 23] and complex networks (CN) [24]. The method provides a perspective of highlighting in an understanding way the spatio-temporal features without binding assumptions. EDM has been verified promising in nonlinear testing of time series, external driving factors for the evolution of mining systems [25]. Therefore, EDM can be adopted to analyze travel time's influencing factors and quantify the causal relationship between these factors. Simultaneously, the urban road network's topological characteristics and spatial structure play an irreplaceable role in the analysis of urban traffic characteristics. To reveal it better, the study uses the relevant knowledge of complex networks to deeply explore the spatial structure characteristics of urban road networks and their impact on urban traffic conditions.

Recently, extreme gradient boosting (XGBoost) [26, 27], as an efficient implementation of gradient enhancement tree algorithm, has become promising candidates for short-term forecasting not only because it can flexibly represent complex nonlinear traffic systems but also because of its ability to map from input variables to output variables directly. Simultaneously, it also has significant advantages to explain the features imported into a predictive model. For example, Zhang and Haghani [28] obtained quite good experimental results when applying the gradient enhancement tree algorithm to predict the travel time of road sections, and the model was well interpreted. Nevertheless, the shortcomings in his paper are only a few road sections which were concerned, the prediction of road network level was not considered, and the selection of features was not explained in detail. To overcome these shortages and maximize the XGBoost forecast method's advantages, a hybrid prediction model named EDMCN-XGBoost for spatio-temporal feature extraction

and prediction of urban road network travel time is proposed in this research.

## 2. Methodology

*2.1. Research Thinking and Architecture.* Figure 1 shows the complete pipeline of our proposed approach, consisting of four interacting steps. First, data preprocessing is conducted to transform raw data into understandable format. Second, the complete mining process of EDM on road segment data is carried out to explore the temporal dependence and influencing factors of travel time. Third, the complex network theory is used to extract the spatial statistical features of traffic networks. Finally, the linkages between spatio-temporal features and corresponding travel time of the road network are built using XGBoost, and key features are identified according to importance ranking and recursive elimination. Afterwards, we establish the XGBoost predictive model.

### 2.2. EDMCN

*2.2.1. Empirical Dynamic Modeling.* Due to the stochastic and nonlinear nature of actual traffic conditions, linear-based methods are no longer suitable for the study of urban traffic data, while the application of nonparametric methods in traffic data mining has gained more attention. Empirical dynamic modeling (EDM) is a data-driven nonparametric modeling framework for nonlinear dynamic systems, including simplex projection, S-map, multivariate embedding, convergent cross-mapping (CCM), and multiview embedding [22, 29–33]. In contrast to fitting parametric equations used by many pieces of research, EDM instead relies on data to infer the operating mechanism of the dynamic system and reveal the relationship among variables, because of the unknown mechanism equation. The main idea of EDM is to regard time series as projections of complex dynamic system behaviors. The state of the system is described as a point in the high-dimensional space whose axis can be regarded as a fundamental state variable, also known as an attractor. An analogy would be with the traffic system. The travel time variable is the projection point of the high-dimensional space of the complex traffic system. The coordinate axis is the primary state variable such as the lagging variable, the speed variable, and the indicator variable. In general, many state variables in complex systems are unobservable, but some systems can recover their evolution using only time series. As Takens' theorem [33] tells us, although the behavior of a system is nominally determined by the high-dimensional state space, it can replace some unknown variables with its lag variable, that is, construct the attractor using the lag variable.

Therefore, this paper studies the dimension and nonlinearity of the time series using EDM quantitative analysis. Based on the nonlinear system's dimension, the causal relationship between other external primary state variables and the target variable is determined by the optimal lag variable of the time series and the EDM convergent cross-mapping test.

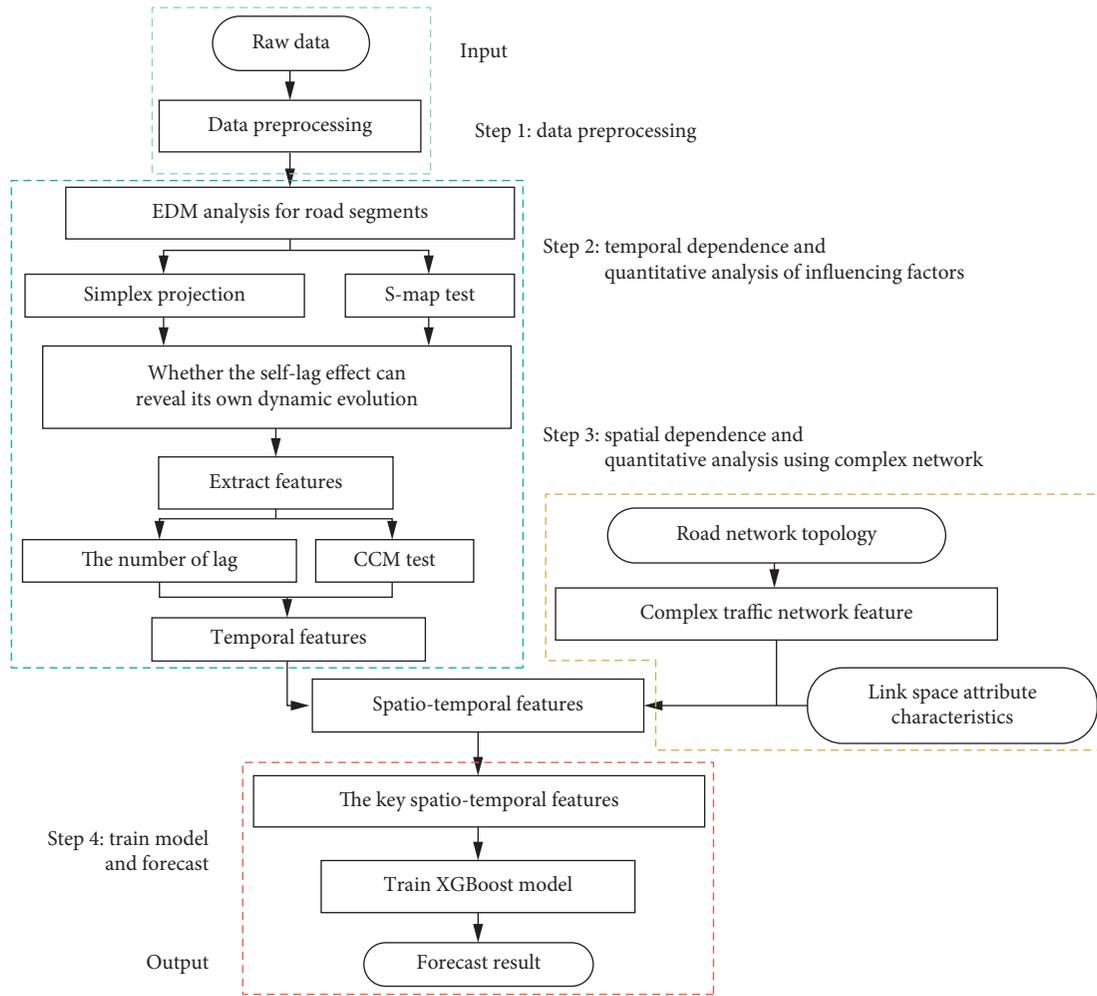


FIGURE 1: Framework of the proposed method.

(1) *System Dimension and Nonlinearity Testing.* The univariate time series of the target variable (travel-time) is analyzed and tested to find its optimal dimension and the nonlinear degree. The optimal dimension ( $E$ ) of the target variable, determined by simple projection, is defined as the lag number that can maximize the prediction ability. When using the simplex projection method, the time series is divided into two parts, one of which is used as a sample to predict the remaining part. It is noteworthy that the simplex projection method does not divide the training set and the prediction set, and the prediction is performed outside the sample. Simplex projection is a nonparametric analysis method in the state space of dynamic systems that obtains the prediction set by mapping the neighbors of the predicted points. The nonlinearity of the dynamic system can be quantified by the S-map analysis method (S-map represents sequence-local weighted global linear mapping) to obtain the nonlinear parameter ( $\theta$ ), which is the state dependence of nonlinear systems.

(2) *Causality Analysis.* EDM can be used to reveal the relationship between different variables. According to the Takens' theorem, the univariate reconstructed attractor  $X_Z$  is

the shadow version of the original multivariate attractor  $X$ , i.e., topological invariance. Therefore,  $X$ 's reconstructed shadow version gives a one-to-one mapping with  $X$ . Furthermore, when there is a causal relationship between two variables, it will lead to a one-to-one mapping of their corresponding reconstructed attractors (e.g.,  $X_Z'$ ,  $X_p'$ ) [23]. Based on the principle and the simplex projection method, Sugihara et al. [30] proposed a convergent cross-mapping (CCM) method for testing the causal relationship between a pair of variables in a dynamic system. If two variables belong to the same variable system with causal links, it must be convergent. The convergence means that the cross-mapping technique ( $p$ ) increases as the size of the library increases, because the more data in the library, the greater the density of reconstructed polylines, and then the higher resolution attractors can improve the prediction accuracy based on neighbors (i.e., simplex projection).

2.2.2. *Complex Networks.* Originated from graph theory, complex network is a large-scale complex system theory tool applied to the real world. In fact, the urban traffic state has spatial autocorrelation characteristics, which are affected by

spatial structure and topology features of road networks. Previous studies have identified the prominent impact of network spatial features and hierarchical features [34, 35] on the network traffic flow. Therefore, this paper mainly applies the theory of complex networks to mine the topological structure characteristics and network structure characteristics of urban traffic networks. For details, please refer to Section 4.3.

**2.3. Extreme Gradient Boosting Prediction Model.** XGBoost [27], an integrated learning parallel processing algorithm based on tree structure, is nonparameterized and can deal with the complex nonlinear relationships between features. Compared to other algorithms, XGBoost has higher interpretability, predictive accuracy, and computational speed. The idea of this algorithm is to continuously add trees, continuously perform feature splitting to grow a tree and each time to add a tree, needing to learn a new function to fit the last predicted residual. When we intend to forecast a sample score, we should get  $K$  trees through a training sample dataset. This process corresponds to the characteristics of the sample, each characteristic will fall into a corresponding leaf node, and each leaf node will correspond to a score. Finally, we need to add each tree's corresponding score to obtain the predicted value of the sample. XGBoost's tree integration model is essentially a set of classification regression trees (CARTs). Each tree (CART) means a decision model  $f(\cdot)$ . The XGBoost algorithm uses model integration; to be precise, it uses  $k$  decision tree models to output results based on input  $x_i$  and sums the  $K$  output results to obtain  $\hat{y}$ . The set of decision tree models is called  $F$ .

$$\hat{y} = \varnothing(x_i) = \sum_{k=1}^K f_k(x_i), \quad (1)$$

$$F = \{f(x) = w_{q(x)}\}, \quad (q: \mathbb{R}^m \longrightarrow T, w \in \mathbb{R}^T),$$

where  $F$  is a set that incorporates  $k$  decision tree models, among them  $F = f_1(\cdot), f_2(\cdot), \dots, f_k(\cdot)$ ;  $f_k(\cdot)$  represents the input/output function relationship of the  $k^{\text{th}}$  regression tree; and  $f_k(\cdot)$  corresponds to the structure  $q$  and leaf node weight  $w$  of the  $k^{\text{th}}$  regression tree.  $w_i$  represents the score of the  $i^{\text{th}}$  leaf node. The first few steps of actually fitting the predicted value include (1)  $\hat{y}^{(0)} = 0$ ; (2)  $\hat{y}^{(1)} = f_1(x_i)$ ; (3)  $\hat{y}^{(2)} = f_1(x_i) + f_2(x_i) = \hat{y}^{(1)} + f_2(x_i)$ . Therefore, the  $t^{\text{th}}$  iteration can be displayed as

$$\hat{y}^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}^{(t-1)} + f_t(x_i). \quad (2)$$

Further optimization of the objective function of the prediction model can be obtained by fd3

$$\text{obj}^{(t)} = \sum_i^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t), \quad (3)$$

where  $\hat{y}_i^{(t-1)}$  represents the prediction of the actual value  $i$  at iteration  $t-1$ ;  $l(y_i, \hat{y}_i^{(t-1)})$  corresponds to the training loss

function; and  $\Omega$  is the regularization term. The regularization term is shown as

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2. \quad (4)$$

Equation (4) represents the regularization term for the decision tree function  $f$ , where  $\gamma, \lambda$  are regularization coefficients, corresponding to gamma and lambda in XGBoost's parameters;  $T$  represents the number of leaf nodes of the decision tree function  $f$ ; and  $\|w\|^2$  represents the sum of the output squares of all leaf nodes of the decision tree function  $f$ .

Equation (4) is used to control the variance of the fitting in order to enhance the flexibility of learning task and to obtain a better prediction model for unknown data. Therefore, the over-fitting of training data is avoided and the complexity of prediction model is penalized.

As can be seen from the above, the newly generated tree is the new mapping relationship to fit the residual of the last prediction, so when after generating  $t$  trees, the predicted value can be written as  $\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i)$ , so the target function can be rewritten as equation (3).

In order to get the minimum value of the objective function, XGBoost approximates it with Taylor's second-order expansion, so that the objective function can be approximated as

$$\text{obj}^{(t)} \approx \sum_{i=1}^n \left[ l\left(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right) \right] + \Omega(f_t), \quad (5)$$

where  $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$  is the first derivative and  $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$  is the second derivative.

Since the prediction score of the former  $t-1$  tree does not affect the optimization of the objective function, it can be neglected, so that the objective function can be simplified as

$$\overline{\text{obj}}^{(t)} \approx \sum_{i=1}^n \left[ l\left(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right) \right] + \Omega(f_t). \quad (6)$$

Equation (6) sums the loss function values of each sample. According to the abovementioned analysis, each sample will eventually fall to a leaf node, so all the samples of the same leaf node can be reorganized as follows:

$$\begin{aligned} \text{obj}^{(t)} &\approx \sum_{i=1}^n \left[ g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \\ &= \sum_{i=1}^n \left[ g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2 \right] + \gamma T + \lambda \frac{1}{2} \sum_{j=1}^T w_j^2 \end{aligned} \quad (7)$$

$$= \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T.$$

It can be seen from the abovementioned analysis that the objective function can be transformed into a unary quadratic function about the leaf node score  $w$ , and the optimal  $w$  and

the objective function  $\text{obj}$  can be directly solved by using the vertex formula of the unary quadratic function.

$$\begin{aligned} w_j^* &= -\frac{G_j}{H_j + \lambda}, \\ \text{obj}^* &= -\frac{1}{2} \sum_{j=1}^T \frac{G_j}{H_j + \lambda} + \gamma T, \end{aligned} \quad (8)$$

where  $G_j = \sum_{i \in I_j} g_i$ ;  $H_j = \sum_{i \in I_j} h_i$ .

In the previous lifting tree model, the classification regression trees are generally arranged in a sequence, but the trees generated by the XGBoost training are arranged in a parallel manner. When the XGBoost training data fits a model, all the cores of the computer CPU are called to construct the tree itself in a parallel way, which in turn increases the speed of calculation. For more details on the XGBoost model, see literature [27].

### 3. Data Description

The dataset used in this study comes from the Alibaba Cloud Tianchi dataset platform, the actual desensitization data of a specific location in Guiyang, providing the attributes of urban road segments and travel time of each segment during the historical period (April 2017). The research dataset is from 132 roads of a particular place in Guiyang, including the average travel time of motor vehicles passing the road sections. The period is 6:00–9:00, 14:00–16:00, and 17:00–19:00 in April 2017, and the time interval is 2 minutes. It should be noted that the original data is not aggregated. Some attributes of the road network are shown in Figure 2.

**3.1. Data Preprocessing.** The target variable shows a long tail distribution through statistical analysis of the dataset, indicating that the overall quality of the dataset is good, but there is still a small number of unreasonable extreme values. To eliminate the influence of abnormal data, the paper first carried out noise reduction processing on data. Next, the target variable is transformed by log; after the log transformation, the target variable presents a normal distribution, which is more suitable for input into the prediction model to fit.

Another problem with the dataset is the existence of missing values. In April 2017, there were 1069200 sample data, but the actual data were only 973978, so there were 95222 missing data. Although XGBoost has a good handle on missing values, due to the nature of time-series data, the values at first few moments have a significant impact on the values at the next moments. Therefore, to minimize the impact of missing values, this study uses the trend of the data to complete the missing data.

**3.2. Road Network Topology.** The topological characteristics and spatial structure of the urban road network play an irreplaceable role in the analysis of urban traffic characteristics. In this study, the relevant knowledge of complex network is employed to deeply explore the spatial structure

characteristics of urban road networks and their impact on urban traffic conditions. The road network studied in this paper includes 132 road sections and their upstream and downstream relationships. The actual dataset does not contain the geographic coordinates of the real road sections and only provides the adjacency relations of 132 road sections. The topology structure of the road network is expressed according to the adjacency relation of road sections [36, 37], as shown in Figure 3.

## 4. Spatio-Temporal Features Extraction and Analysis

**4.1. Basic Spatio-Temporal Characteristics.** The attributes of the road network dataset include the link-ID of each road section, travel time of the two-minute slice, and the relationship between the upstream and downstream sections. A large number of basic spatio-temporal features can be extracted from the raw data, as shown in Table 1. Nos. 1–3 are the lagging values of the target variable; Nos. 4–6 represent the timing variables; Nos. 7–10 are the time and space shift variable of the target variable; Nos. 11–16 are the indicator variable; Nos. 17–19 are some basic spatial properties of the road segment.

Features are derived values from raw data and used as input to a machine learning algorithm. High-quality features (e.g., being informative, relevant, interpretable, and non-redundant) are the basis for modeling and problem-solving and generating reliable and convincing results. Due to the spatio-temporal characteristics of urban road network travel time, this study focuses on two kinds of features: time-related and spatial correlation features. First, the EDM method is used to analyze the basic time-series features quantitatively. Then, the theory of complex network is leveraged to deeply dig the urban road network to obtain its spatial characteristics.

**4.2. Quantitative Analysis by Empirical Dynamic Modeling.** The results of the simplex projection of the dataset show that the optimal embedding dimension ( $E$ ) of the univariate analysis of the travel time sequence of the road segment is six (see Figure 4(a)). The nonlinear test of the travel time series by  $S$ -map shows that the dataset exhibits a relatively high nonlinearity in  $\theta = 2$  (see Figure 4(b)). Forecast skill ( $\rho$ ) represents the Pearson correlation coefficient ( $\rho$ ) between the observed value and the predicted value by the EDM method, which is used to maximize the prediction performance. The above indicates that it is reasonable to use the EDM method to process and analyze the data. It should be reminded that before using the empirical dynamic modeling, the dataset needs to be normalized to eliminate the influence of different data units.

**4.2.1. EDM Multivariate Analysis.** Firstly, a prediction model is established by using EDM. The prediction of travel time is to use the maximum available time lag determined by the simplex projection. The target variable (travel time at time  $t$ ) is fixed time coordinates, and new coordinates  $(1, 2, \dots, \text{or } (E-1))$  are

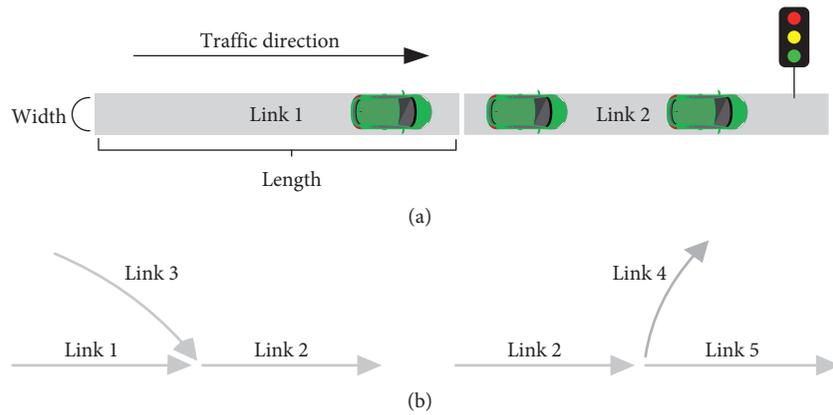


FIGURE 2: Main spatial information of roads. (a) The attribute information of links. (b) The upstream and downstream relationship of road segments.

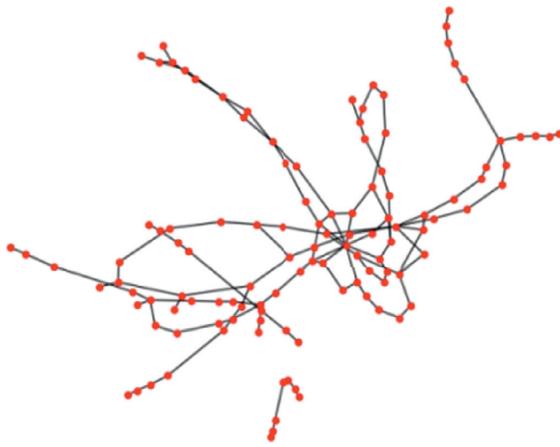


FIGURE 3: Road network topology map (desensitized data).

considered as the construction model predictor coordinates. Correctly, the delay effect of the target variable itself is used as a predictor, and the “block-lnlp” function in EDM can be used to accurately show whether the lagging effect of the target variable itself can fully explain the complex system dynamics of the target variable. The result in Figure 5 shows that the lag coordinates of travel time itself cannot reveal its complex dynamic evolution. After adding the external driving factor of the indication, the model prediction skill is not significantly improved, which indicates that the evolution of travel time series is a highly nonlinear complex phenomenon, and the lag variable of the target variable is a very crucial driving factor that is of greater importance than other external driving factors such as a vacation.

The driving factors affecting travel time evolution have been proved to be diverse and complex, so if we want to restore the complex dynamic behavior of travel time series as much as possible, it is necessary to find more external driving factors.

**4.2.2. Nonlinear Causality Test.** It can be seen from the abovementioned analysis that the evolution of urban road network travel time is a highly nonlinear complex

phenomenon. According to the results of EDM multivariate analysis, the five important driving factors of travel time evolution can be preliminarily determined as  $TT_{t-1}$ ,  $TT_{t-2}$ ,  $TT_{t-3}$ ,  $TT_{t-4}$ , and  $TT_{t-5}$ . On this basis, the nonlinear causality relationship (CCM) test is applied to identify the travel time-driven time series variables, time-space time-shifting features, time-indicating features, and state-indicating features of the target road segments. Sugihara et al. [29, 30] developed a cross-mapping algorithm to test the causation between a pair of variables in dynamic systems. If there is a causal relationship between two variables, the cross-mapping between them shall be “convergent.” Convergence means that the cross-mapping skill ( $\rho$ ) improves with increasing library size [23]. A typical road segment determines the causal association between the following characteristics and the target variable, specifically Minute, Hour, Dayofweek, Dayof-year, Vacation, Average\_speed, out\_link\_lagging,  $T_1$ , etc. It should be noted that the black-marked numbers in Table 2 are identified as causal variables. The detailed quantitative analysis results are shown in Table 2 and Figure 6.

Note that, the direction of cross-mapping is opposite to the direction of cause-effect in Figure 6. According to the cross-convergence test of some external driving factors, the magnitude of the cross-mapping rise of external factors is not high. Therefore, based on the quantitative analysis results, it can be determined that there is no strong causal relationship between these features and target variables, only a medium causal or weak relationship.

The interesting point is the first picture and the last picture of Figure 6. The CCM test does not treat time-space time-shifting features (in\_link\_lagging) in some road segments as a powerful causal, even weak causal association features. By contrast, another time-space time-shifting feature (out\_link\_lagging) is regarded as a causal connection feature. Maybe the time step is quite long regarding the urban traffic dynamic. It can explain that no relationship between the upstream and downstream sections is statistically highlighted. From the perspective of urban traffic congestion, the downstream section’s causal

TABLE 1: The basic spatio-temporal features.

Symbol	Description	No.
$TT_{t-1}$	The first lagging values of target variable	1
$TT_{t-2}$	The second lagging values of target variable	2
$TT_{t-3}$	The third lagging values of target variable	3
$T_1$	The first values before the start of the forecast period	4
$H_1$	The true values of the hour before the forecast period	5
Average_speed	The first speed before the prediction point	6
in_link1_lagging1	The true values of the $i^{\text{th}}$ upstream road of the target	7
in_link2_lagging1	Road before the predicted point	8
out_link1_lagging1	The true values of the $i^{\text{th}}$ downstream road of the target	9
out_link2_lagging1	Road before the predicted point	10
Minute	The minute of the forecast point	11
Hour	The hour of the forecast point	12
Dayofweek	The forecast point represents the day of the week	13
Dayofyear	The date of the forecast point	14
Vacation	Holiday or not	15
Link-ID	The ID of each road segment	16
Length	The length of each road segment	17
Width	The width of each road segment	18
Area	The area of each road segment	19

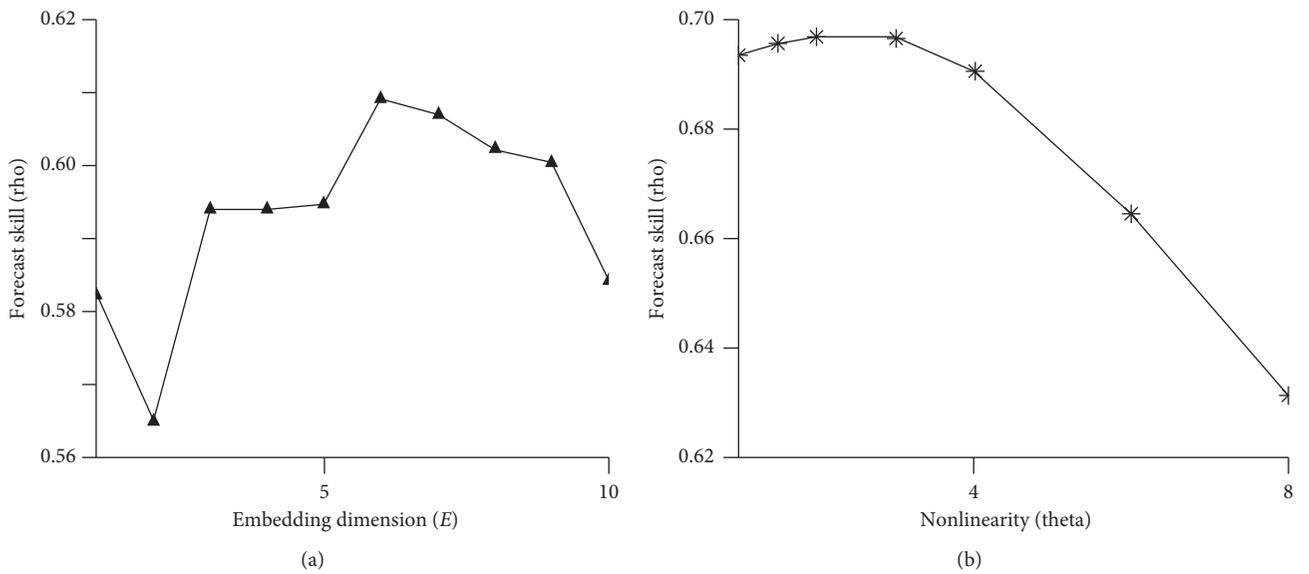


FIGURE 4: The embedding dimension and nonlinear analysis of travel time series. (a) Simplex projection. (b) S-map nonlinear analysis.

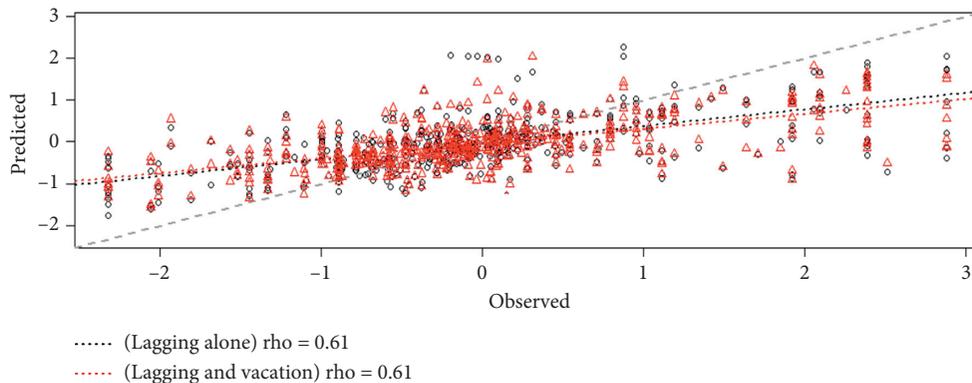


FIGURE 5: The influence of lagging variables on the evolution of target variable sequences.

TABLE 2: CCM test of external driving factor and its correlation with target variables.

Candidate variable ( $x_i$ )	Cross-map skill (travel_time~ $x_i$ )	Linear correlation (travel_time~ $x_i$ )
Minute	<b>0.1030</b>	-0.0143
Hour	<b>0.0824</b>	0.0072
Dayofweek	<b>0.0718</b>	-0.1250
Dayofyear	-0.0364	0.0125
Vacation	<b>0.0303</b>	-0.0484
$T_1$	<b>0.0681</b>	0.6939
$H_1$	0.0082	0.6701
Average_speed	<b>0.0540</b>	-0.2800
in_link_lagging	-0.0240	0.0140
out_link_lagging	<b>0.1171</b>	0.1644

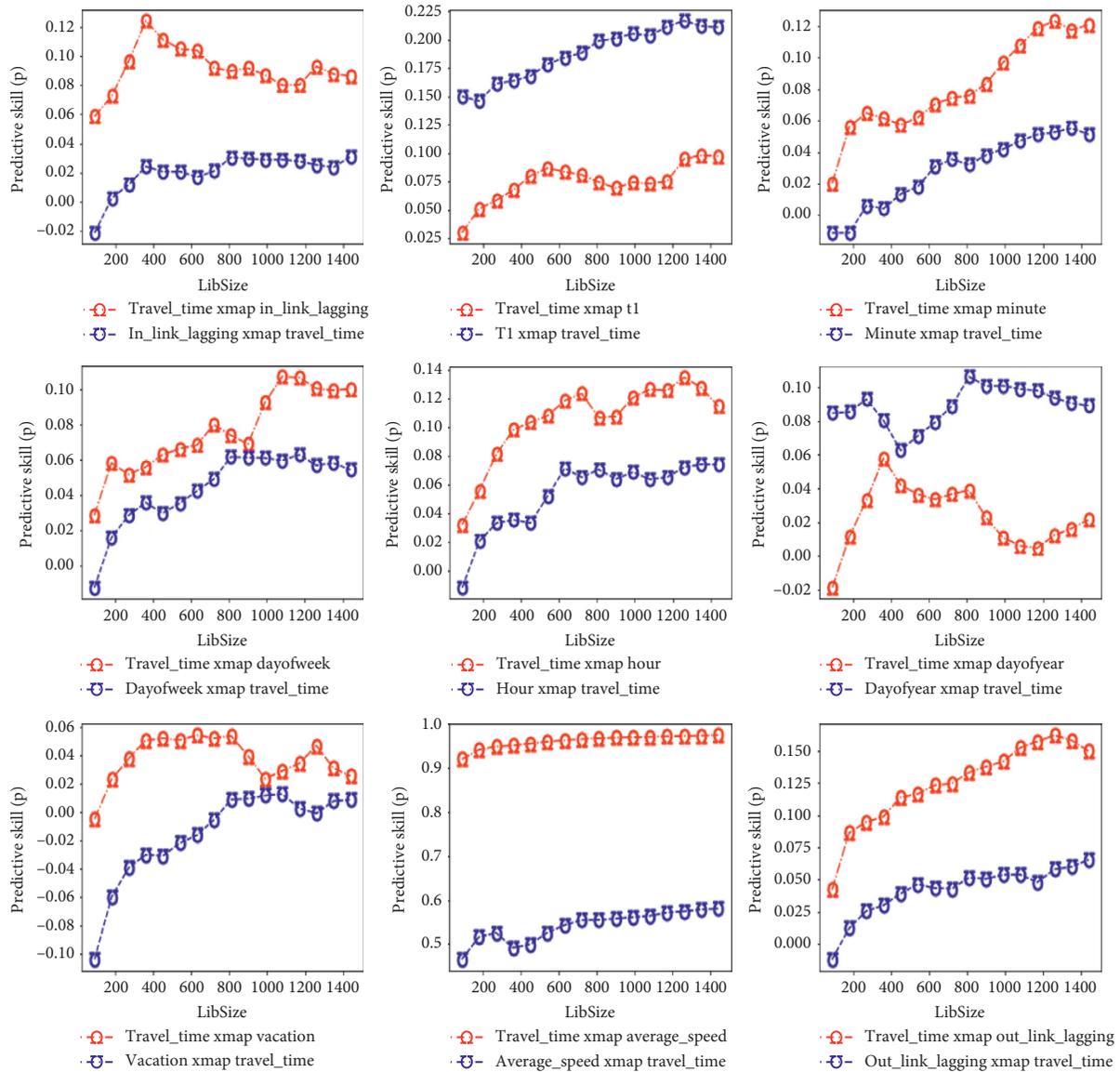


FIGURE 6: External driver factor CCM test for the travel time series.

impact on the target section is following the traffic jam propagation law. It can be seen from Table 2 that the linear relationship cannot be related as the basis for judging the causal connection.

4.3. *Complex Traffic Network Spatial Characteristics.* To further reveal the spatial dependence of the road network, it is necessary to dig out more valuable information from the perspective of spatial characteristics. Based on the raw

dataset, we can extract some essential spatial characteristics, that is, length, width, area, and link-ID of different road segments, but the above four features are only the primary attributes of a road segment and do not involve the spatial correlation characteristics between road segments. Therefore, this study tries to understand the traffic network from the perspective of the complex network to mine useful features and obtain the quantitative characteristics of mathematical statistics of a large number of complex traffic networks such as degree, in-degree, out-degree, closeness centrality, the importance of nodes, betweenness-centrality, degree-centrality, and PageRank. To explore the spatial aggregation behavior of different road segments, the community structure is introduced into the complex traffic network. The following is a detailed explanation of some spatial features.

#### 4.3.1. Complex Traffic Network Characteristic Parameters.

Complex networks have been widely concerned by traffic scholars. A large number of studies have found that transportation networks have the same complex network structural characteristics as social networks and computer networks. Therefore, a road network of a certain place in Guiyang city modeled by the dual-method is taken as the research object [37], and the complex network is adopted to study the interaction between the structural characteristics and the topology of the traffic network. The research is shown below.

- (1) *Degree of node.* The degree of a node represents the number of nodes directly connected to the node. In a directed graph, the degree of a node is divided into the degree of out (out\_degree) and the progress (in\_degree), indicating the intensity of the road segment in the road network to some extent. Out-degree is expressed as the number of road segments connected downstream of the road segment, and in-degree means the number of road segments connected upstream. Thus, the higher the degree (in or out), the node has a more prominent connection effect in the road network.
- (2) *Closeness centrality.* The closeness centrality of a node in the network measures the proximity of the node to all other nodes. The larger the value of closeness centrality, the more central the node is in the road network, and the faster it can reach other nodes. Therefore, it is used to emphasize the values of the different nodes in a traffic network; the specific calculation method is as follows:

$$C_c(i) = \frac{N-1}{\sum_{j=1}^N d_{ij}}, \quad (9)$$

where  $d_{ij}$  means the shortest distance from node  $i$  to node  $j$  and  $N$  represents the number of road segments.

- (3) *Node importance ordering.* The typical node importance ranking methods are PageRank, LeaderRank,

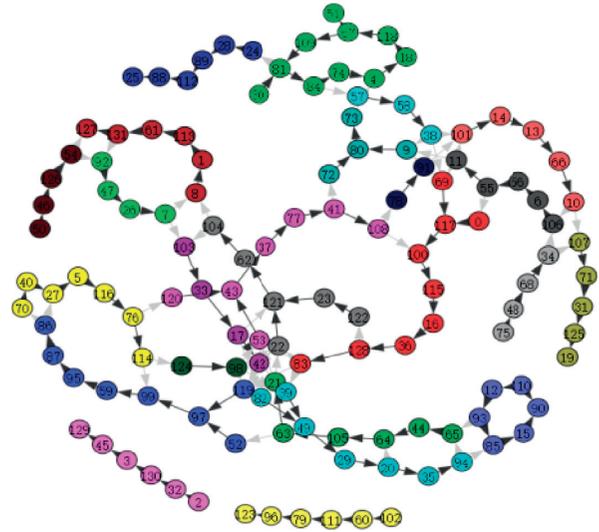


FIGURE 7: Community structure of the traffic network.

and HITS algorithms, among which PageRank [38] is the core sorting algorithm of Google search. The main idea of the PageRank algorithm is that if the quality of webpage  $A$  is high, and page  $A$  points to page  $B$ , then the quality of page  $B$  is also high. Because the actual page links are much more complex, iterations are required to get the final result of the page sort. It is mapped to a complex traffic network with specific quantized PageRank values (referred to as  $PR$ ) after repeated iterations. The specific calculation steps are mainly referenced in [34].

#### 4.3.2. Complex Networks Community Detection.

Community detection, also known as community discovery, is a way to reveal network aggregation behavior. Commonly used community discovery algorithms include Louvain, label propagation, and infomap. After comprehensive comparisons, this study adopts the infomap algorithm to identify community clusters of road network topology, and the specific calculation steps have mainly referenced the literature [31].

Taking the complex traffic network topology (dual-map structure) of a certain place in Guiyang city as input, the community division result of the traffic network is obtained in Figure 7, where nodes of the same color represent the same community. Each node represents a specific road segment, and each arrow indicates the upstream and downstream relationship of adjacent nodes.

**4.4. Feature Extraction Results.** Quantitative analysis characteristics of each road segment are obtained by the EDM method; it has obtained a large number of spatial statistical features of urban traffic networks through complex network theory. The abovementioned steps make deep excavation and quantitative description of the spatio-temporal features, which greatly enhance the interpretability and richness of the features. However, through the EDM method, the

TABLE 3: The key spatio-temporal features of the traffic network.

Predictors (symbol)					No.
$TT_{t-1}$	$TT_{t-2}$	$TT_{t-3}$	$TT_{t-4}$	$TT_{t-5}$	1~5
$T_1$	Average_speed	Minute	Hour	Dayofweek	6~10
Vacation	Length	Width	Area	Link_ID	11~15
in_degree	out_degree	closeness_centrality	PageRank	infomap	16~20
out_link1_lag1	out_link1_lag2				21~22

quantitative analysis results of each road segment are not precisely the same. If the analysis result of each road segment is taken as the feature input of the road network, it will cause a lot of information redundancy and lead to conflicting results. Therefore, it is necessary to extract the road network characteristics that conform to most road segments. The characteristics of the road network extracted through the complex network are plentiful and specific, but at the same time, there is also redundancy. It is necessary to streamline the complex network special features to seek the causality between those features with the target variable.

The relationship between spatio-temporal features and travel time of road segments is established by XGBoost, and redundant variables are removed by its feature importance ordering and recursive elimination methods. Table 3 shows the road network spatio-temporal features finally determined for this study.

## 5. Analysis of Experimental Results

The travel time prediction of the traffic network is a critical step in constructing an urban intelligent transportation system. According to the above empirical dynamic modeling and quantification results, the evolution of travel time in an urban road network is proved to be highly nonlinear, which is difficult to predict by purely mechanical equations. Thus, to accurately forecast the evolution of complex time series, it is necessary to extract the key driving factors of system evolution from the data itself. Combining the broad prospects of data-driven methods in traffic prediction applications in recent years and the considerable advantage of tree models based on parallel integration processing in dealing with the relationship between a large number of feature data and processing features, this study decides to adopt XGBoost as the predictive model.

**5.1. Dataset Description and Division.** In the experiment, 80% of the data is used as the training set and 20% as the test set. In other words, the data from April 1 to April 24, 2017, is taken as the training set to predict the travel time in the traffic network from April 25 to 30, 2017. The main target period in this study is the morning peak (8:00–9:00), the noon stable period (15:00–16:00), and the evening peak (18:00–19:00). The topology of the road network is detailed in Section 3.2.

Table 4 shows the model input and output, the first 21 lines ( $f_0 \sim f_{21}$ ) in the table are the eigenvalue data as input, and the last line is the output value of the XGBoost model. It should be pointed out that some of the data in the model are

log-transformed, and the output data also experience the log conversion. So, when we verify the prediction result, the transformed data should be restored to the original scale.

**5.2. Model Optimization.** In the machine learning model, data quality is the root cause of the prediction accuracy, but different parameter combinations will also have a particular impact on the prediction model. Therefore, the model is compared several times through the experiment, and the evaluation index used in the comparison process is the root mean square error (RMSE). The derived optimal parameter combination of the research dataset is determined, as shown in Table 5.

**5.3. Model Interpretation.** It is well known that the prediction accuracy of the XGBoost model is very demanding on the predictor variables (i.e., feature vectors). In order to further analyze the influence of the characteristic variables of EDMCN on the model output (response variables), the feature importance is understood by the trained XGBoost model (see Figure 8).

It can be seen from Figure 8 that  $TT_{t-1}$  is the most critical characteristic variable, indicating that the value at the previous moment of the prediction point has the most considerable influence on the next point. It is consistent with our common sense, and the performance of other lagging variables demonstrates that the lagging variables selected before the model input are also reasonable. It is proved that the evolution of travel time series is highly time-dependent, and the time dependence is highly valued in urban traffic management. Average\_speed is also a vital feature. In urban traffic management, attention should be paid to the driving speed of road sections. Surprisingly, the  $T_1$  feature performed well. The inspiration to us in this article is to pay attention to before making predictions the first values before the start of the forecast period, which may be significant.

Simultaneously, the space-time time-shifted variables, out\_link1\_lag1 and out\_link1\_lag2, we need to the critical focus on spatio-temporal variable, are outstanding, which is also in line with the law of congestion propagation of roads. Among the time indication features, Minute performs the best, which indicates that in the management and control of urban traffic, attention should be paid to the excellent time granularity. Among the spatial indication features, link\_ID and length are the most prominent, which means that each road needs careful management and detailed analysis of its characteristics in urban traffic management. Among the sophisticated traffic network features, closeness\_centrality, PageRank, and infomap perform better. It means that the

TABLE 4: XGBoost model input and output.

Symbol	The values of features									No.
Minute	0	2	4	6	8	10	12	14	...	$f_0$
Hour	8	8	8	8	8	8	8	8	...	$f_1$
Dayofweek	2	2	2	2	2	2	2	2	...	$f_2$
Vacation	0	0	0	0	0	0	0	0	...	$f_3$
in_degree	1	1	1	1	1	1	1	1	...	$f_4$
out_degree	1	1	1	1	1	1	1	1	...	$f_5$
Length	48	48	48	48	48	48	48	48	...	$f_6$
Width	3	3	3	3	3	3	3	3	...	$f_7$
Area	144	144	144	144	144	144	144	144	...	$f_8$
Link_ID	43	43	43	43	43	43	43	43	...	$f_9$
$T_1$	1.931	1.931	1.931	1.931	1.931	1.931	1.931	1.931	...	$f_{10}$
closeness_centrality	0.053	0.053	0.053	0.053	0.053	0.053	0.053	0.053	...	$f_{11}$
PageRank	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	...	$f_{12}$
infomap	5	5	5	5	5	5	5	5	...	$f_{13}$
$TT_{t-1}$	2.197	1.903	1.911	1.918	1.921	1.918	1.906	1.899	...	$f_{14}$
$TT_{t-2}$	1.995	2.197	1.903	1.911	1.918	1.921	1.918	1.906	...	$f_{15}$
$TT_{t-3}$	2.197	1.995	2.197	1.903	1.911	1.918	1.921	1.918	...	$f_{16}$
$TT_{t-4}$	2.197	2.197	1.995	2.197	1.903	1.911	1.918	1.921	...	$f_{17}$
$TT_{t-5}$	2.140	2.197	2.197	1.995	2.197	1.903	1.911	1.918	...	$f_{18}$
Average_speed	6.000	8.403	8.329	8.262	8.231	8.264	8.378	8.447	...	$f_{19}$
out_link1_lag1	2.360	2.379	2.360	2.370	2.476	2.595	2.803	2.687	...	$f_{20}$
out_link2_lag1	-999	-999	-999	-999	-999	-999	-999	-999	...	$f_{21}$
$TTt$	2.042	1.932	1.949	1.948	1.948	1.950	1.945	1.937	...	...

TABLE 5: The optimal parameter combination of the XGBoost prediction model.

$n\_estimators$	eta	min_child_weight	max_depth	colsample_bytree	Lambda	Gamma
200	0.26	66	7	0.7	1	0.2

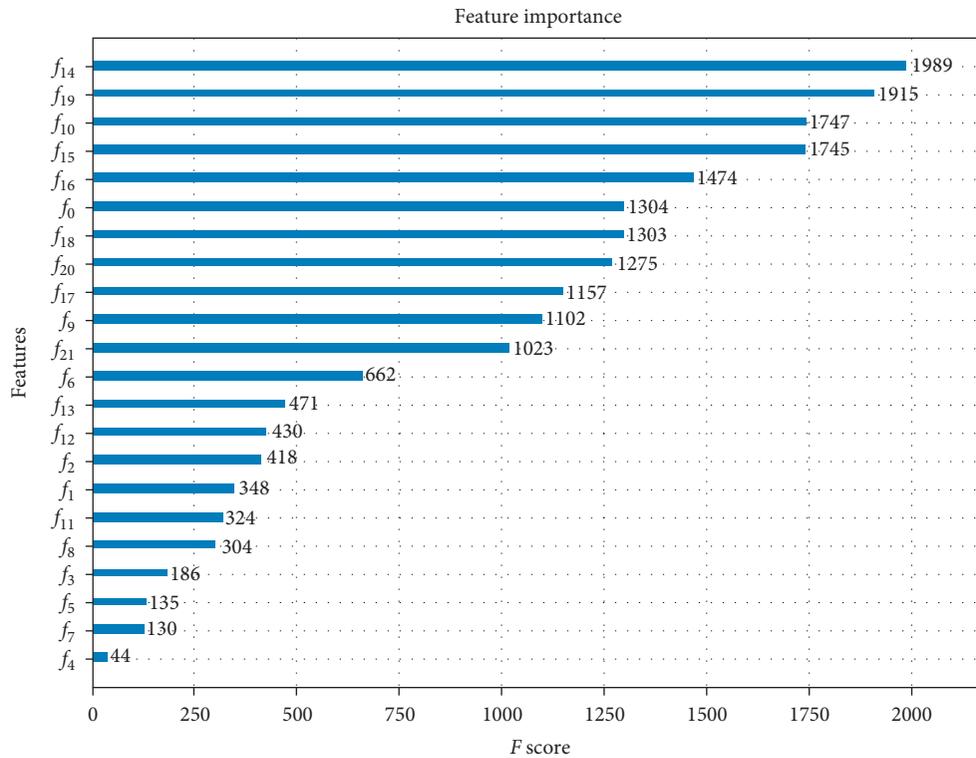
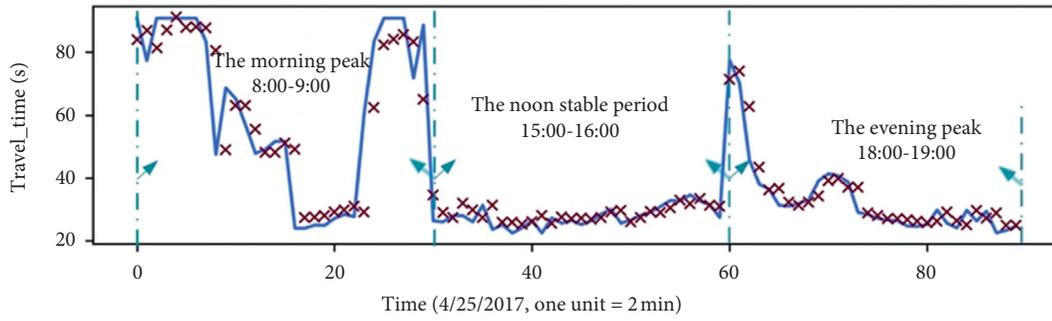
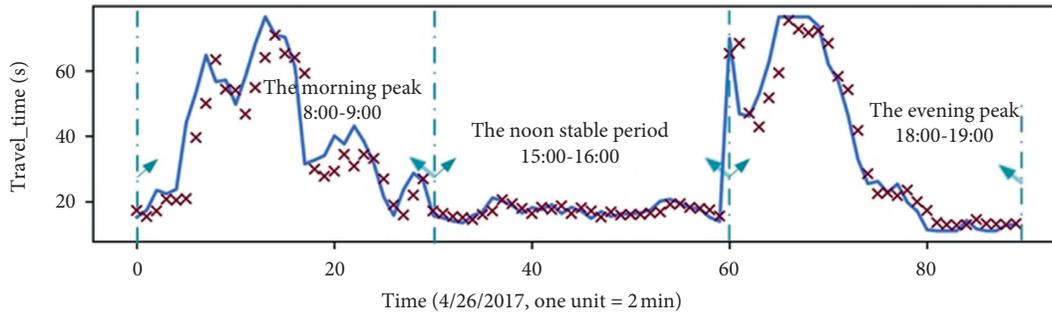


FIGURE 8: F score of feature vector in the EDMCN-XGBoost model.



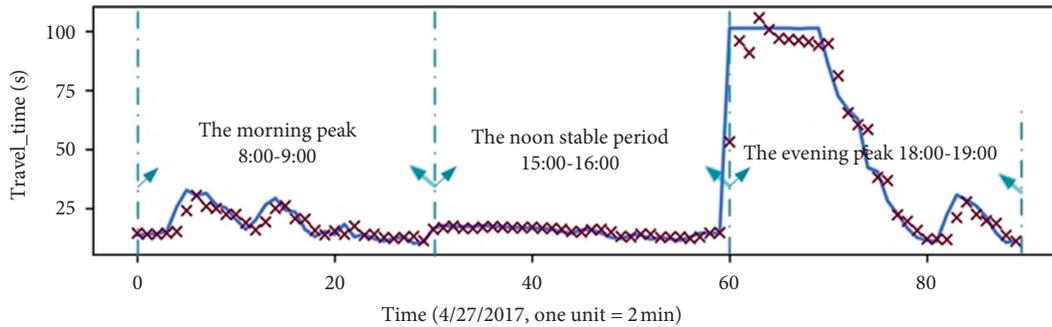
— Original value  
x Predicted value

(a)



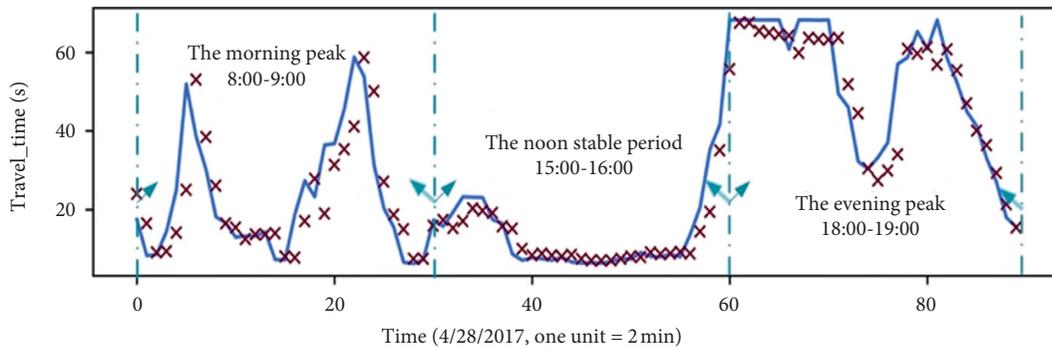
— Original value  
x Predicted value

(b)



— Original value  
x Predicted value

(c)



— Original value  
x Predicted value

(d)

FIGURE 9: Continued.

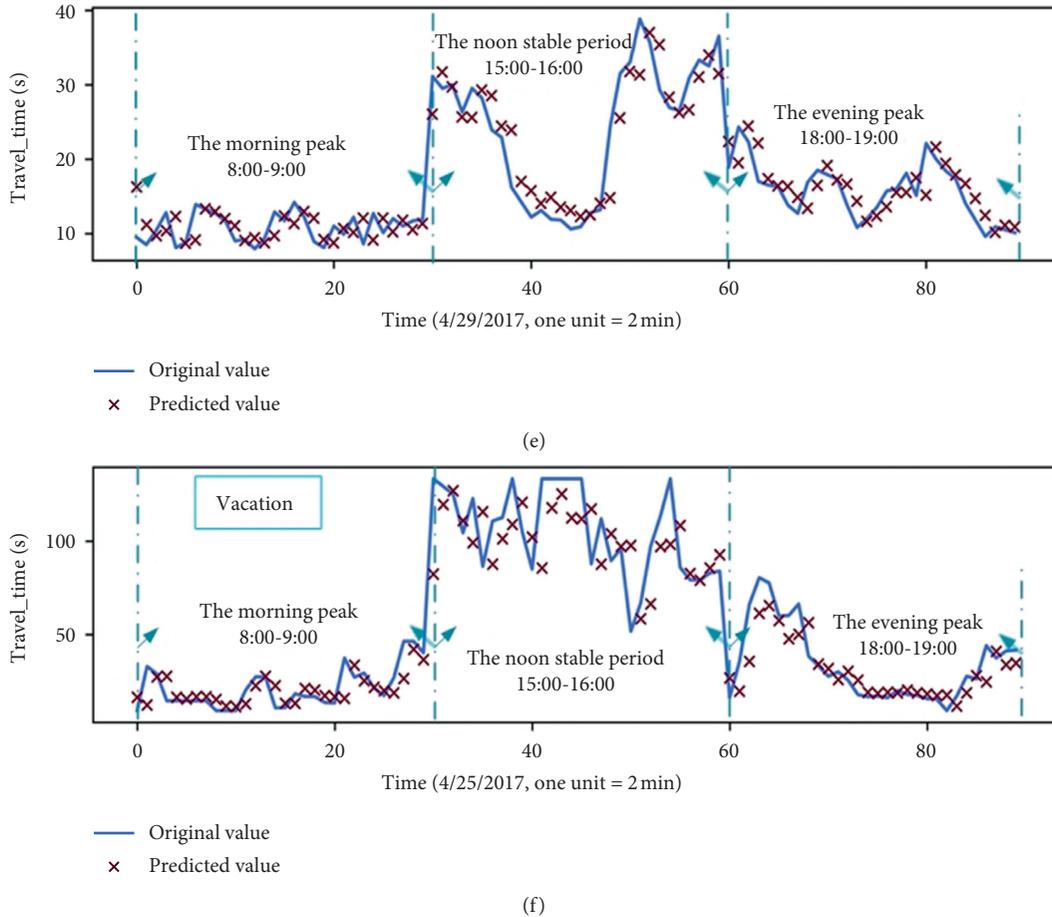


FIGURE 9: The prediction effect of the EDMCN-XGBoost model.

evolution of travel time in the urban road network is spatially dependent. Therefore, urban traffic management and control need to consider the specific characteristics of each road and the spatial connection between roads. The unexpected point is that *in\_degree* has the worst performance in this study.

Figure 9 shows the comparison between the travel times of the 132 roads in Guiyang and the travel times predicted by the EDMCN-XGBoost model from April 25, 2017, to April 30, 2017. The blue line represents the actual travel time, and the cross represents the travel time predicted by the forecast model. The marked morning peak, noon stable period, and evening peak in the figure represent the time concept of urban traffic commuting and does not mean the road's actual travel status. The horizontal axis 0–29 represents the morning peak 8:00–9:00, 30–59 represents the noon stable period 15:00–16:00, and 60–89 stands for the evening peak 18:00–19:00. Each unit in the horizontal axis denotes two minutes.

The time scale of the vertical axis in Figure 9 is measured in seconds to reflect the real-time variation law of the traffic state. It can be seen that the overall performance of XGBoost is quite well, and the prediction effect of different roads in different periods is useful. It is worth noting that some roads of the road network are not strictly following the morning

TABLE 6: Comparison between the forecasting model and baseline model evaluation index.

Forecast model	MAPE (%)	MAE	RMSE	$R^2$
ARIMA	20.627	3.576	7.688	0.879
ANN	20.653	3.740	8.080	0.866
SVR	20.265	3.381	7.477	0.885
XGBoost	19.262	3.319	6.858	0.903
CN-XGBoost	19.111	3.305	6.840	0.904
EDM-XGBoost	18.744	3.256	6.728	0.908
EDMCN-XGBoost	<b>18.695</b>	<b>3.230</b>	<b>6.710</b>	<b>0.909</b>

and evening peak travel rules, and the holiday travel rules are also different from the usual. Although the travel time rules of different road segments at different periods are different, the spatio-temporal dynamics of this travel time can still be captured by the model and accurately predicted.

**5.4. Model Comparison.** In order to test the validity of the EDMCN-XGBoost model, the currently accepted traffic prediction baseline models ARIMA [8], ANN [10], and SVR [13] and some hybrid models are comprehensively evaluated, and four classical evaluation index functions MAPE, MAE, RMSE, and  $R^2$  are used to evaluate the prediction models.

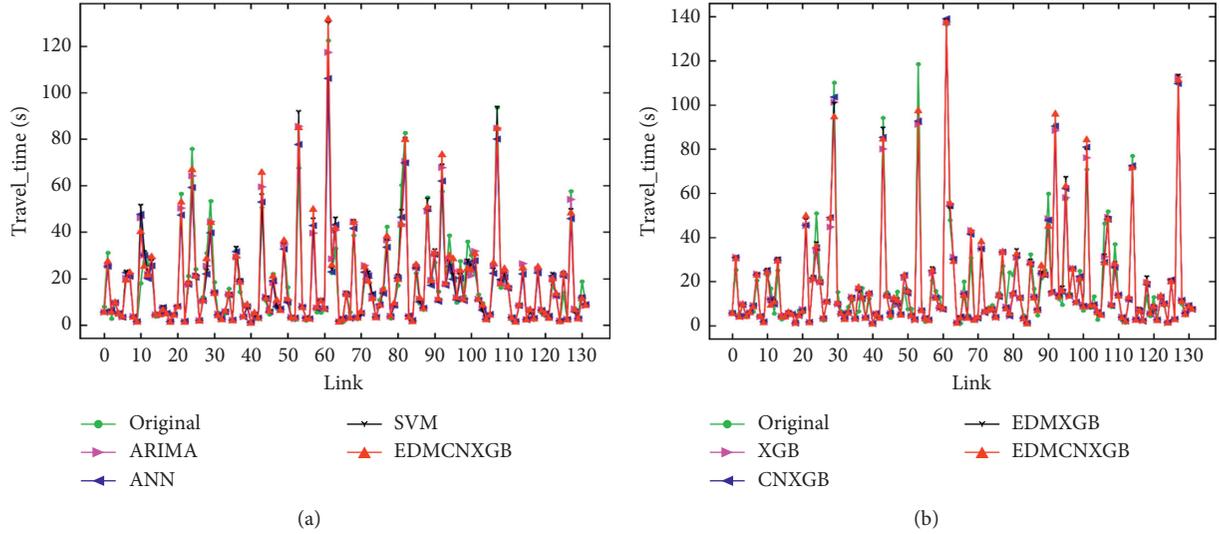


FIGURE 10: The actual operation status of the road network at a specific moment and the predicted values of different models. (a) Longitudinal comparison of predictive models. (b) Horizontal comparison of predictive models.

$$\text{The mean absolute percentage error : MAPE} = \frac{1}{Q} \sum_{i=1}^Q \frac{|t_i - \hat{t}_i|}{t_i},$$

$$\text{The mean absolute error : MAE} = \frac{1}{Q} \sum_{i=1}^Q |t_i - \hat{t}_i|,$$

$$\text{The root mean square error : RMSE} = \sqrt{\frac{1}{Q} \sum_{i=1}^Q |t_i - \hat{t}_i|^2},$$

$$\text{The coefficient of determination : } R^2 = 1 - \frac{\sum_i |t_i - \hat{t}_i|^2}{\sum_i |\bar{t}_i - \hat{t}_i|^2}, \quad (10)$$

where  $t_i$  represents the actual value;  $\hat{t}_i$  is the predicted value;  $\bar{t}_i$  represents the mean of the test set; and  $Q$  represents the number of test sets.

In this study, data from the first 24 days of April (see Section 5.1) were selected as the training set, and data from the last six days of April were used as the verification set. To thoroughly verify the validity of the EDMCN-XGBoost model, it was compared with other prediction models from vertical and horizontal directions. For ensuring the fairness of the input of different predicted models. In the longitudinal comparison, the  $(p, d, q)$  parameters obtained by the ARIMA model based on the dataset optimization in the longitudinal comparison are  $(4, 0, 2)$ ; the inputs to the ANN and SVM models are based on the features of the EDMCN screening. In the horizontal comparison, when the study relies solely on XGBoost, the input of the lagging variables is determined by ARIMA's 4; the spatial features of road segments are length, width, area, and link-ID; XGBoost's feature importance ranking selects features variables. The input of CN-XGBoost adds complex network features to the XGBoost input, increasing the spatial connectivity between

roads. The input of EDM-XGBoost aims to select and filter the feature vector of the roads by EDM quantitative analysis before the input. It is to notice that the dataset is a bit short of producing an efficient neural network, but it remains useful for comparison [39].

It can be seen from the experimental results that the effect of EDMCN-XGBoost is superior to other models from a vertical and horizontal direction. From comparison, result further demonstrates the scientific validity of spatio-temporal features from EDMCN and more explanatory. One thing to add is that it is useful to compare the prediction performance according to the considered peak period instead of the full day-period, especially when aggregated indicators are considered. The off-peak period is not of paramount importance for the prediction as it is a stable state. The specific experimental results are shown in Table 6.

Figure 10 compares the actual operating conditions of the road network at a specific time interval (2 min) and the forecast performance of different models. The  $x$ -axis indicates the link\_ID of each road segment of the road network. It can be seen from the comparison that the EDMCN-XGBoost model performs best in prediction accuracy.

From the experimental results in Figures 8–10, and Table 6, the following conclusions can be drawn: (a) the EDMCN-XGBoost forecast model is useful for forecasting the travel time of the urban road network and superior to the baseline models. (b) The evolution of travel time series is dynamic and highly nonlinear, and the influencing factors are relatively diverse. (c) Travel time prediction at the network level must consider both spatial and temporal dependencies. (d) During the peak period, the change of travel time in the road network is more apparent, and the travel rules of the traffic network on working days, weekends, and holidays are not exactly the same. (e) It is necessary to construct high-quality feature engineering as much as possible, making it rich, interpretable, and nonredundant. Only in this way can the evolution of time series be restored to the greatest extent.

## 6. Conclusion

This study designed a framework for analyzing, mining, quantifying, and predicting spatio-temporal travel time of the urban road network. The framework is mainly divided into the following steps: (a) firstly, a large amount of primary feature data is obtained from the original data. (b) The EDM is used to quantify the dynamic and highly nonlinear evolution of the travel time series. Meanwhile, the time series lagging variable ( $TT_{t-i}$ ) is determined as the crucial variable of travel time evolution for the strong causal relationship between  $TT_{t-i}$  and  $TT_t$ . (c) The complex network theory is adopted to explore the topological structure of urban traffic networks deeply, and a large number of spatial statistical features of the road network is obtained. (d) The feature importance ordering and recursive elimination principle of XGBoost are used to remove the redundancy of acquired spatio-temporal features and finally establish the prediction model. The example shows that the framework proposed in this study performs well in obtaining high-quality features that are rich, interpretable, and nonredundant to establish a predictive model. The established XGBoost prediction model is superior to other comparable models in both interpretability and prediction accuracy.

Overall, the road network travel rules are complex and changeable, challenging to capture. By digging out the information behind raw data as much as possible, we can better seize the evolution rules of urban traffic travel time. To further improve the prediction accuracy of road network travel time, based on the framework of this paper, we will continue to refine the feature engineering of road network travel time and try to combine more forecasting methods.

## Data Availability

The terms of use of the data used in this study do not allow the authors to distribute or publish the data directly. However, these data can be obtained directly from Tianchi dataset platform via the following webpage: <https://tianchi.aliyun.com/dataset/dataDetail?dataId=1079>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors would like to thank the Alibaba Cloud Tianchi platform which provided the traffic dataset for them and the fund from the National Natural Science Foundation of China (61873216).

## References

- [1] Y. Gao, S. Sun, and D. Shi, "Network-scale traffic modeling and forecasting with graphical lasso," in *Advances in Neural Networks—ISNN 2011*, pp. 151–158, Springer, 2011. Springer, Berlin, Germany, vol. 6676 of Lecture Notes in Computer Science.
- [2] K. N. Qureshi and A. H. Abdullah, "A survey on intelligent transportation systems," *Middle East Journal of Scientific Research*, vol. 15, pp. 629–642, 2013.
- [3] J. Yeon, L. Elefteriadou, and S. Lawphongpanich, "Travel time estimation on a freeway using discrete time Markov chains," *Transportation Research Part B: Methodological*, vol. 42, no. 4, pp. 325–338, 2008.
- [4] T. Choe, A. Skabardonis, and P. Varaiya, "Freeway performance measurement system: operational analysis tool," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1811, no. 1, pp. 67–75, 2002.
- [5] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining—KDD'11*, pp. 316–324, San Diego, CA, USA, August 2011.
- [6] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we're going," *Transportation Research Part C Emerging Technologies*, vol. 43, no. 1, pp. 3–19, 2014.
- [7] Z. Zhang, M. Li, X. Lin, Y. Wang, and F. He, "Multistep speed prediction on traffic networks: a deep learning approach considering spatio-temporal dependencies," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 297–322, 2019.
- [8] G. Box, E. P. Pierce, and A. David, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American Statistical Association*, vol. 65, no. 332, pp. 1509–1526, 1970.
- [9] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research Part B: Methodological*, vol. 18, no. 1, pp. 1–11, 1984.
- [10] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211–234, 2005.
- [11] G. Fusco, C. Colombaroni, and N. Isaenko, "Short-term speed predictions exploiting big data on large urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 73, pp. 183–201, 2016.
- [12] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong, "An improved K-nearest neighbor model for short-term traffic flow prediction," *Procedia—Social and Behavioral Sciences*, vol. 96, pp. 653–662, 2013.
- [13] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [14] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.
- [15] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.
- [16] X. Ma, H. Yu, Y. Wang, and Y. H. Wang, "Large-scale transportation network congestion evolution prediction using deep learning theory," *PLoS One*, vol. 10, no. 3, Article ID e0119044, 2015.
- [17] M. Fouladgar, M. Parchami, R. Elmasri, and A. Ghaderi, "Scalable deep traffic flow neural networks for urban traffic

- congestion prediction,” in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN) IEEE*, pp. 2251–2258, Anchorage, AK, USA, May 2017.
- [18] L. Zhao, Y. Song, C. Zhang et al., “T-GCN: a temporal graph convolutional network for traffic prediction,” *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2018.
- [19] M. G. Karlaftis and E. I. Vlahogianni, “Statistical methods versus neural networks in transportation research: differences, similarities and some insights,” *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 3, pp. 387–399, 2011.
- [20] Y. Zhang, Y. Zhang, and A. Haghani, “A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model,” *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 65–78, 2014.
- [21] L. N. N. Do, N. Taherifar, and H. L. Vu, “Survey of neural network-based models for short-term traffic state prediction,” *WIREs Data Mining and Knowledge Discovery*, vol. 9, no. 1, p. pe1285, 2019.
- [22] H. Ye, R. J. Beamish, S. M. Glaser et al., “Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 13, pp. E1569–E1576, 2015.
- [23] C. W. Chang, M. Ushio, and C. H. Hsieh, “Empirical dynamic modeling for beginners,” *Ecological Research*, vol. 32, no. 6, pp. 785–796, 2017.
- [24] J. Liu, Z. Ren, Q. Guo, and B. Wang, “Research progress on order importance ranking of nodes in complex networks,” *Acta Physica Sinica*, vol. 62, no. 17, p. 178901, 2013.
- [25] M. Wang, C. Yoshimura, A. Allam, F. Kimura, and T. Honma, “Causality analysis and prediction of 2-methylisoborneol production in a reservoir using empirical dynamic modeling,” *Water Research*, vol. 163, p. 114864, 2019.
- [26] T. Chen and C. Guestrin, “XGBoost: a scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, San Francisco, CA, USA, August 2016.
- [27] S. Xiupeng, D. W. Yiik, Z.-F. L. Michael, P. Chandrasekar, and C. Chai, “A feature learning approach based on XGBoost for driving assessment and risk prediction,” *Accident Analysis & Prevention*, vol. 129, pp. 170–179, 2019.
- [28] Y. Zhang and A. Haghani, “A gradient boosting method to improve travel time prediction,” *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 308–324, 2015.
- [29] G. Sugihara and R. M. May, “Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series,” *Nature*, vol. 344, no. 6268, pp. 734–741, 1990.
- [30] G. Sugihara, R. May, H. Ye et al., “Detecting causality in complex ecosystems,” *Science*, vol. 338, no. 6106, pp. 496–500, 2012.
- [31] E. R. Deyle, R. M. May, S. B. Munch, and G. Sugihara, “Tracking and forecasting ecosystem interactions in real time,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 283, no. 1822, Article ID 20152258, 2016.
- [32] J. A. McGowan, E. R. Deyle, H. Ye et al., “Predicting coastal algal blooms in Southern California,” *Ecology*, vol. 98, no. 5, pp. 1419–1433, 2017.
- [33] F. Takens, R. Rand, Young, Detecting strange attractors in turbulence,” in *Dynamical Systems and Turbulence, Warwick 1980*, pp. 366–381, Springer, 1981Springer, Berlin, Germany, vol. 898 of Lecture Notes in Mathematics.
- [34] B. Jiang, J. Yin, and S. Zhao, “Characterizing the human mobility pattern in a large street network,” *Physical Review E*, vol. 80, no. 2, p. 21136, 2009.
- [35] X. P. Han, Q. Hao, B. H. Wang, and T. Zhou, “Origin of the scaling law in human mobility: hierarchy of traffic systems,” *Physical Review E*, vol. 83, no. 3, p. 21136, 2011.
- [36] S. Porta, P. Crucitti, and V. Latora, “The network analysis of urban streets: a primal approach,” *Environment and Planning B: Planning and Design*, vol. 33, no. 5, pp. 705–725, 2006.
- [37] S. Porta, P. Crucitti, and V. Latora, “The network analysis of urban streets: a dual approach,” *Physica A: Statistical Mechanics and Its Applications*, vol. 369, no. 2, pp. 853–866, 2006.
- [38] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” *Computer Networks and ISDN Systems*, vol. 30, no. 1–7, pp. 107–117, 1998.
- [39] C. Makoto, G. Prateek, V. Varun et al., “On the possibility of short-term traffic prediction during disaster with machine learning approaches: an exploratory analysis,” *Transport Policy*, 2020, in press.