IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 7, 2021, accepted May 15, 2021, date of publication May 20, 2021, date of current version May 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3082188

Research on Traffic Situation Analysis for Urban Road Network Through Spatiotemporal Data Mining: A Case Study of Xi'an, China

RUIYU ZHOU[1], HONG CHEN[1], HENGRUI CHEN[1], ENZE LIU[1], AND SHANGJING JIANG[2]
[1]College of Transportation Engineering, Chang'an University, Xi'an 710000, China
[2]School of Geography, Nanjing Normal University, Nanjing 210023, China

Corresponding author: Hong Chen (glch@chd.edu.cn)

This work was supported by the 111 Project of Sustainable Transportation for Urban Agglomeration in Western China under Grant B20035.

ABSTRACT Severe traffic congestion has promoted the development of the Intelligent Transportation System (ITS). Accurately analyzing and predicting the traffic states of the urban road networks has important theoretical significance and practical value for improving traffic efficiency and formulating ITS scheme according to local conditions. This study aims to identify and predict the traffic operation status in the road network within the Third Ring Road in Xi'an and explore spatiotemporal patterns of traffic congestion. In this paper, firstly, we discriminated the traffic status of the urban road network used the GPS data of floating vehicles (e.g., taxis and buses) in Xi'an by the Travel Time Index (TTI). Secondly, we used the emerging hot spot analysis method to locate different hot spot patterns. The time series clustering method was used to divide the whole road network's locations into distinct clusters with similar spatiotemporal characteristics. Thirdly, we applied three different time series forecasting models, including Curve Fit Forecast (CFF), Exponential Smoothing Forecast (ESF), Forest-based Forecast (FBF), to predict the traffic operation status. Finally, we summarized the spatiotemporal characteristics of the whole-network congestion. The results of this study can contribute some helpful insights for alleviating traffic congestion. For instance, it is essential to speed up the construction of urban traffic microcirculation and increase the road network density. Moreover, it is crucial to adhere to the urban public transport priority development strategy and increase public transportation travel sharing.

INDEX TERMS Urban traffic congestion, spatiotemporal pattern, short-term prediction, taxi trajectory, road traffic performance index.

I. INTRODUCTION

With the rapid growth of private vehicle ownership, the corresponding transportation infrastructure is insufficiently supplied, traffic congestion has become more serious, affecting people's travel and limiting the city's economic stable development [1], [2]. Due to traffic congestion, the most influential cities in China suffer economic losses of $1 billion every year [3]. The European Commission calculates that the annual cost related to traffic congestion is about 100 billion euros (1% of GDP) [4]. The traffic congestion raises gas consumption, aggravates environmental pollution, and raises residents' travel time [5]. Understanding the spatiotemporal patterns of traffic congestion and predicting the short-term traffic state are significant urban management challenges [6]. Despite much research on route optimization and congestion prediction, there is still an inadequate understanding of the whole road network's spatiotemporal congestion patterns. Accurately assessing the spatiotemporal characteristics of urban traffic congestion and deeply mining its complex operation regularity has important theoretical significance and practical value for improving traffic operation efficiency and intelligent traffic management technology [7].

Most of the existing studies focus on improving models and algorithms of traffic data from a single source and focusing on analyzing and predicting traffic conditions in congested sections or regions. There is a lack of research on multi-source traffic big data for the whole road network.

The associate editor coordinating the review of this manuscript and approving it for publication was Nabil Benamar.

Traditional methods are mostly microscopic methods based on simulation, which are difficult to be applied to real life [8]. Moreover, it is challenging to explore the spatiotemporal distribution and evolution trend of recurrent congestion at the macro level. Therefore, we divided the road network of Xi'an into a small grid with a geographic position, which has significant application value and practical significance. Simultaneously, the calculation and analysis of GPS data at the grid level can eliminate the process of map matching, improving the calculation efficiency [9]. Besides, due to the complexity of roads and taxi drivers' driving characteristics, obtaining current or future road traffic operation status is a huge challenge. Because of its operating model, no-load taxis will travel slowly to look for passengers, which reduces the road's capacity and, in severe cases, may cause traffic congestion. Compared with the running speed of taxis, the driving trajectory of buses is relatively stable and regular [3]. To make the spatiotemporal patterns more informative, we integrated the GPS data from taxis and buses.

Although various algorithms have been used to mine the traffic congestion evolution regularity, there are little researches on the formation mechanism and internal reason of the whole road network traffic condition. There are two main types of traffic congestion in urban roads: recurrent congestion and occasional congestion. Occasional congestion is usually generated by emergencies, such as traffic accidents and temporary traffic control, while periodic traffic flows usually cause recurrent congestion [10]. Unlike the randomness of occasional congestion, recurrent congestion has specific rules in forming, spreading, and dissipating. On different days, the generation and dissipation time of congestion is always the same in the temporal dimension, while the location, propagation direction, and influence range of congestion are highly similar in spatial dimension [11]. This study focused on the spatiotemporal patterns of recurrent congestion on working days. The spatiotemporal pattern of recurrent congestion should be minded continuously, adopting management and control measures at the macroscopic road network level to settle the congestion problem.

The remainder of this article is structured as follows: Section 2 reviews the research results in associated areas; Section 3 presents the study area and dataset. Section 4 details the used method; Section 5 discusses model results; Section 6 summarizes the conclusions.

II. LITERATURE REVIEW

The first step in traffic control and management is to quantify urban roads' operating status. The road traffic performance index is a comprehensive indicator reflecting the whole road network's traffic operation status. It is also a comprehensive and straightforward description of traffic congestion characteristics in space, time, and intensity. Previous studies have proposed many road traffic performance indices, such as Road Congestion Index (RCI [12]), Lane Kilometer Duration Index (LKDI [13]), Travel Time Index (TTI [14]), and Congestion Severity Index (CSI [15]). With wireless communication and location technology development, GPS data has been widely used in traffic congestion identification algorithms. Chen *et al.* [16] established a novel speed performance index (SPI) to assess traffic conditions taking the traffic flow speed and road speed limit into account based on the floating car data. Zhao *et al.* [3] calculated the actual and standard bus travel time by processing the GPS trajectory of the bus and bus station data. They then built the congestion index and classification (CI-C) model to explore the traffic status and predict the future bus congestion time. Su et al [17] used the fuzzy c-means (FCM) algorithm and fuzzy entropy weight method to describe traffic status and conducted a detailed analysis in Beijing. Speed is one of the basic parameters in most traffic performance indices.

By accurately identifying the traffic operation status of the whole road network and timely release of relevant information to reasonably plan the vehicles' driving path, it can effectively alleviate traffic congestion. Based on massive trajectory data, more and more scholars dedicate to identifying traffic congestion in recent years. Yu *et al.* [18] used the Markov model and Back Propagation Neural Network (BPNN) model to identify the campus traffic congestion connected with the pedestrian, vehicle, motorbike, and bike speed. Hao *et al.* [19] proposed the multiple criteria decision-making method to evaluate the intersection traffic congestion. Taking traffic congestion at intersections as an example, five evaluation indicators were used to verify the hybrid method's effectiveness. Yang *et al.* [20] propose a traffic congestion prediction method based on real-time and predictive traffic data. They established two evaluation frameworks to verify the accuracy and computational effectiveness of the proposed method. Shen *et al.* [21] employed the semi-supervised extreme learning machine (SSELM) to connect congestion conditions and road information. However, most of the existing studies focus on improving models and algorithms of traffic data from a single source and focus on analyzing and predicting traffic conditions in congested sections or regions. There is a lack of research on big data with multi-source traffic for the whole road network.

The clustering method based on time series is widely used in the field of spatiotemporal data mining. Classifying the spatiotemporal congestion patterns makes it possible to mine the hidden rules and features behind the massive traffic data. Furthermore, it reveals the heterogeneity of the distribution of traffic flow status in space and time dimensions and extracts homogeneous traffic status change patterns. Jiang *et al.* [22] used the fuzzy clustering method to classify a single road section's traffic status, which can dynamically identify its traffic status. Weijermars *et al.* [23] extracted traffic patterns from traffic flow data without considering the road network's spatial distribution pattern and used hierarchical clustering technology to analyze the time-varying characteristics of daily heterogeneous traffic states of a single detector. Wen *et al.* [24] clustered the time series of traffic performance index and identified typical urban congestion patterns by

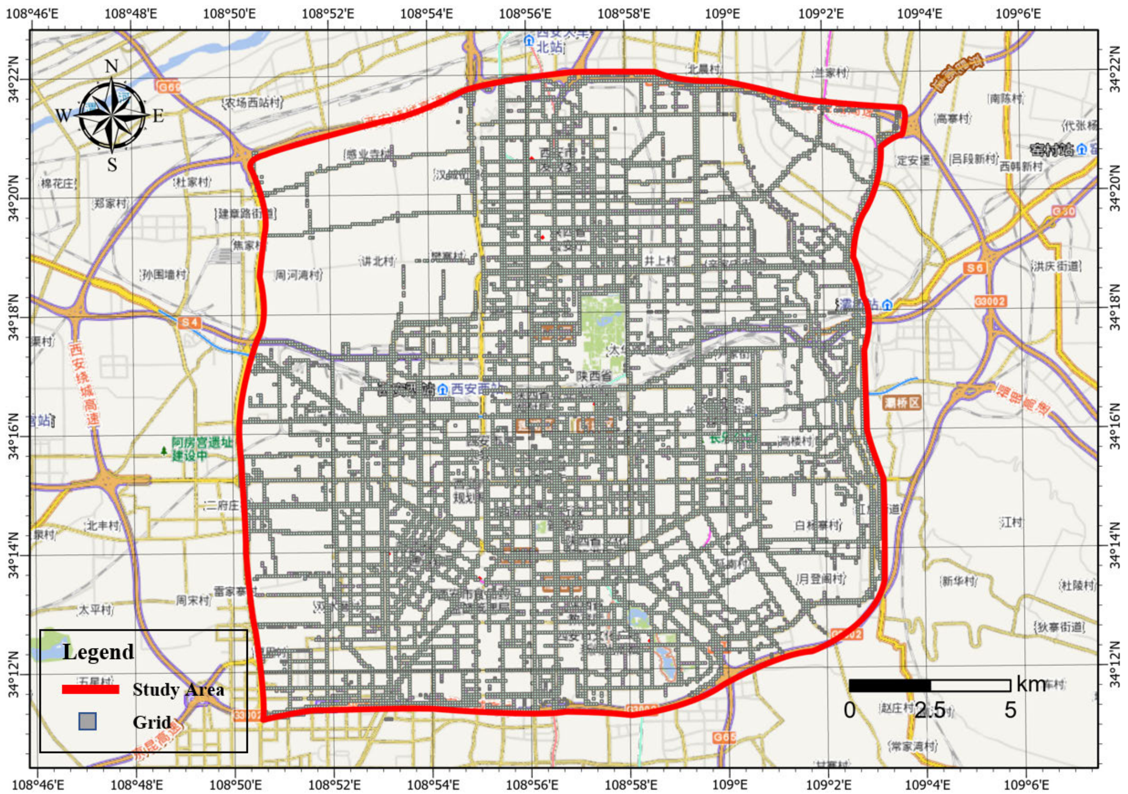


FIGURE 1. The study area.

combining the factors such as dates, transportation demand management policies, holidays, and weather conditions.

Short-term traffic flow prediction is the most dynamic and classic research in intelligent transportation systems. It is based on historical data to estimate the traffic state in the adjacent time, which is the research foundation for realizing adaptive control. It is critical for traffic management and route guidance at the whole road network level, and it is also the hot spot of intelligent transportation application research. With intelligent computing development, traffic state prediction models have gradually changed from simple linear methods to complex non-parametric methods. Traditional short-term traffic flow prediction models are based on statistical methods, such as ARIMA and Kalman filtering methods, make full use of the significant time dependence between historical univariate time series data. Such methods usually set the model structure in advance and estimate the model parameters from the historical data, which has strong interpretability, but the prediction accuracy is easily influenced by the unstable traffic state [25]. Non-parametric methods based on machine learning are extensively applied in short-term traffic flow forecasting. They can mine the nonlinear characteristics between variables and have high prediction accuracy, including BPNN [26], Support Vector Machine Algorithm (SVM) [27], K Nearest Neighbor Algorithm (KNN) [28]. These methods can effectively extract potential information from massive data and achieve better prediction effects but lack interpretability. Many previous studies have focused on predicting traffic status variables at designated points, only considering historical traffic information's influence in the time dimension, ignoring the spatial dimension characteristics. A specific location's traffic status is primarily affected by the adjacent location's traffic status in space. Considering the influence of the combined information of time and space traffic is a primary development direction in traffic flow prediction.

This study has two contributions to the existing literature. Firstly, a hybrid data model integrating GPS data and road grid model is built, adopting the TTI to evaluate the urban road traffic state based on processing the GPS track data and road network data of taxis and buses in Xi'an. Secondly, we mine the spatiotemporal pattern of recurrent congestion in the whole road network in Xi'an and provide valuable insights to alleviate the traffic congestion.

III. STUDY AREA AND DATA DESCRIPTION

This study took the road network within the Third Ring Road of Xi'an as its study area (108°49′57″E∼108°49′57″E, 34°11′20″N∼34°22′10″N), including 87 arterial roads and 110 secondary roads. The study area was divided into nearly 14,000 grids according to the size of 100∗100m, as shown in FIGURE 1. Xi'an, the capital city of Shaanxi Province in west China, has more than 3.6 million vehicles. The contradiction

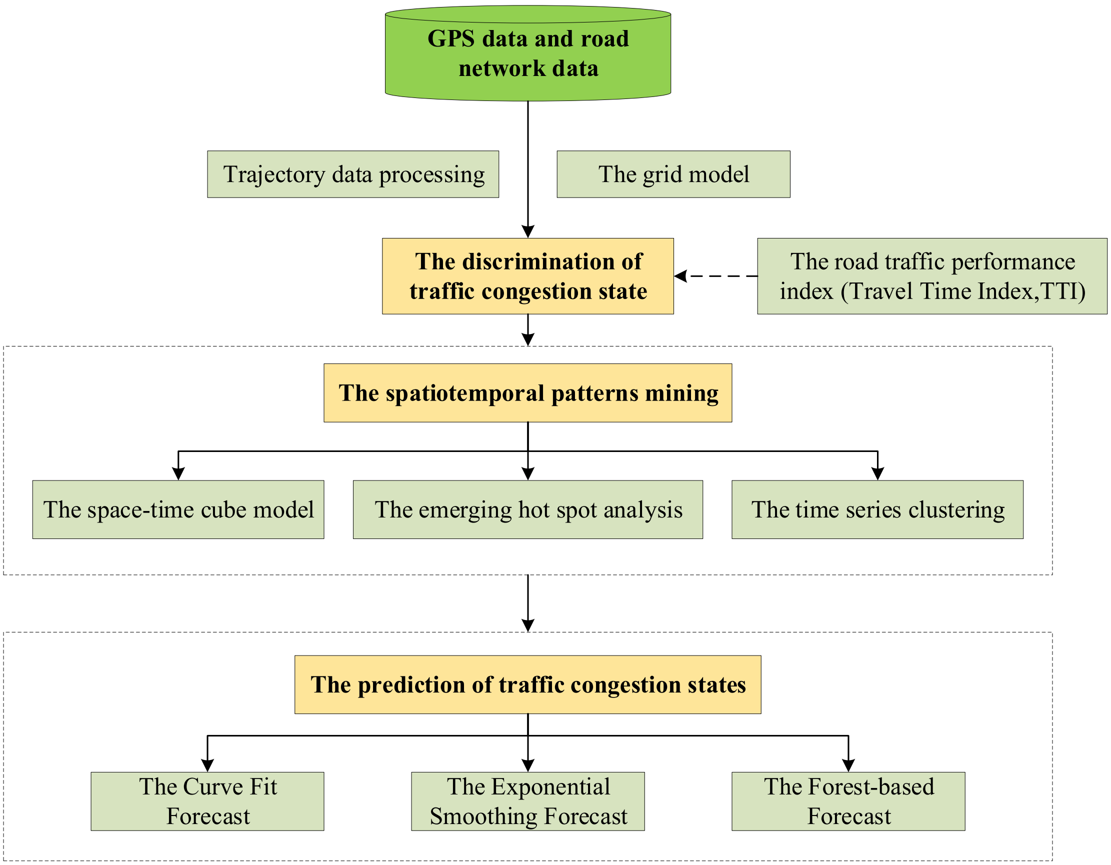


FIGURE 2. Overview of the framework for the study.

between vehicle supply and demand in Xi'an has further intensified, resulting in increasingly severe traffic congestion.

This study is based on the GPS trajectory data of floating vehicles (e.g., taxis and buses) in Xi'an provided by the Xi'an Transportation Administration of China. The data include location information, vehicle state information, time information, and vehicle speed information. The GPS data were recorded every 5s for 30 days in June 2019. This study focuses on the discrimination and prediction of traffic congestion states on weekdays, so the data of 21 working days are selected for analysis, totaling more than 600 million data, with a storage scale of about 65GB. An example of taxi and bus GPS data is shown in TABLE 1.

TABLE 1. An instance of taxi and bus GPS data.

Field	Type	Sample	Comment
GpsId	String	8191105186	GPS_ID
timestamp	String	1592150400109	Unix timestamp, in seconds
longitude	String	108.917202	WGS84 Coordinate System
latitude	String	34.288117	WGS84 Coordinate System
vehicle state	String	1	0: Without passenger 1: With passenger
speed	String	27.3	Vehicle Speed

IV. METHODOLOGY

This section includes three main parts: discrimination and prediction of traffic status of the urban road network and mining the spatiotemporal patterns of traffic congestion using multi-source data. The discrimination of traffic congestion status includes the processing of trajectory data, the construction of the grid model, and the use of road traffic performance indicators. The spatiotemporal patterns mining includes emerging hot spot analysis and time series clustering. We carry out the prediction of traffic congestion states of urban roads network based on space-time cube and the time series method. The research framework is shown in FIGURE 2.

A. THE DISCRIMINATION OF TRAFFIC CONGESTION STATE

A comprehensive data model combining GPS trajectory and the urban grid was constructed by analyzing the taxi and bus GPS data and urban road network data, providing data support for the following study. This study determined the square grid dividing method by comparing several grid dividing methods [9]. This study used Travel Time Index (TTI) to evaluate road traffic performance. The TTI is the most commonly used evaluation index of urban congestion in the industry, which is the ratio of the actual travel time and the free flow time. The current study used Python and ArcGIS Pro to process and analyze the GPS data.

1) THE TRAJECTORY DATA PROCESSING

The critical step of trajectory data processing is to clear invalid data and extract the vehicle's free-flow speed and actual speed. The trajectory data processing process is shown in FIGURE 3. The primary processing steps are as follows:

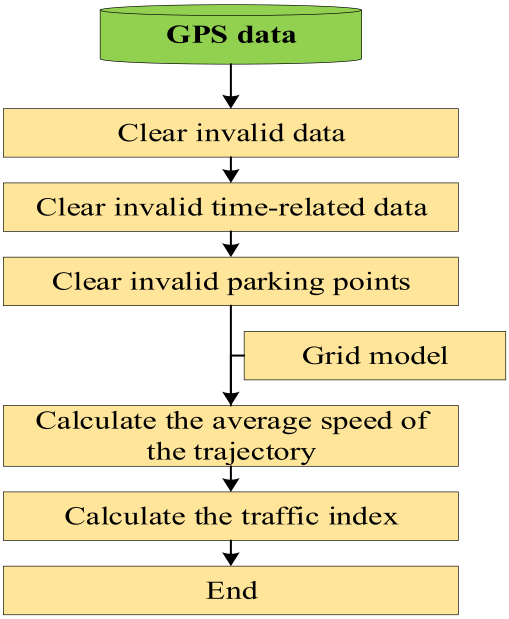


FIGURE 3. **The trajectory data processing process.**

Step 1: Clear invalid data. We cleared the trajectory data of the vehicle under abnormal states (such as alarm and shutdown state); We cleared the trajectory points whose trajectory range is not in the study area; We cleared abnormal trajectories with vehicle speed greater than 100km/h.

Step 2: Clear invalid time-related data. We sorted the same vehicle's trajectory points in chronological order and deleted the trajectory points with repeated time stamps. If the time interval between a trajectory points and two adjacent trajectory points exceeds 5 minutes, it will be regarded as an isolated point and deleted.

Step 3: Clear invalid parking points. Considering the invalid data generated by vehicles parked at the roadside waiting for passengers, we cleared the trajectory points whose speed was always 0 within 3 minutes.

Step 4: Calculate the average speed of the trajectory. Firstly, the entire road network was divided into uniform grids according to the spatial dimensions. Then each grid was rasterized by latitude and longitude, and the grids were numbered. Finally, calculated the average speed of the trajectory in each grid.

Step 5: Calculate the road traffic performance index. We calculated the TTI of each grid using the free-flow speed and actual speed of the vehicle in each grid.

2) THE CALCULATION METHOD OF TTI

The TTI is the most commonly used evaluation index of urban congestion in the industry, which is the ratio of the actual travel time and the free flow time [29]. The higher the TTI value is, the more congested the road is. Some unusual weather situations (e.g., rain, snow, fog) or unusual road statuses may similarly influence the value of TTI. [30].

The fundamental concept of speed: If a link has two time slices, t1 and t2, and the link's length is S, then between the period from t1 to t2, the average speed v of the link is: $v = 2 \cdot S/(t1 + t2)$. The speed is calculated as (2):

$$S = \{Link_1, Link_2, Link_3 \ldots\ldots Link_N, \} \tag{1}$$

$$speed = \frac{\sum_{i=1}^{N} L_i * W_i}{\sum_{i=1}^{N} \frac{L_i}{V_i} * W_i} \tag{2}$$

The fundamental concept of TTI: On the same link in a time slice, TTI = free flow speed / actual speed. The time mean speed of 3-5 a.m. on the link is used as the free-flow speed of the vehicle. The TTI is calculated as (3):

$$TTI = \frac{\sum_{i=1}^{N} \frac{L_i}{V_i} * W_i}{\sum_{i=1}^{N} \frac{L_i}{V_{free\_i}} * W_i} \tag{3}$$

where the length of link is L_i, the weight of link is W_i, the free flow of link is V_{free_i}, the real time speed of link is V_i. According to the Specification for Urban Traffic Performance Evaluation (GB/T 33171-2016) [31], the traffic performance index is defined as five categories based on TTI, as shown in TABLE 2.

TABLE 2. **The evaluation criteria of TTI.**

TTI	[0, 2)	[2, 4)	[4, 6)	[6, 8)	[8, 10]
traffic status	fast	smooth	light congestion	medium congestion	severe congestion

B. THE SPATIOTEMPORAL PATTERNS MINING

This study used ArcGIS Pro for spatiotemporal patterns mining. We analyzed spatiotemporal data by using time series analysis and realized visualization by creating space-time cubes. The emerging hot spot analysis methods take multi-dimensional data sets as input and distinguish statistically meaningful hot and cold spot trends over time. We used the emerging hot spot analysis method to analyze the GPS data to locate different hot spot patterns at one-hour intervals. The time series clustering method divides the positions in a space-time cube into different clusters where each cluster has the same time-series features.

1) THE SPACE TIME CUBE MODEL

We created a space-time cube from the multidimensional raster layer to construct a multidimensional cube data structure and input the analysis data. The output space-time cube will have the same spatial and temporal resolution as the

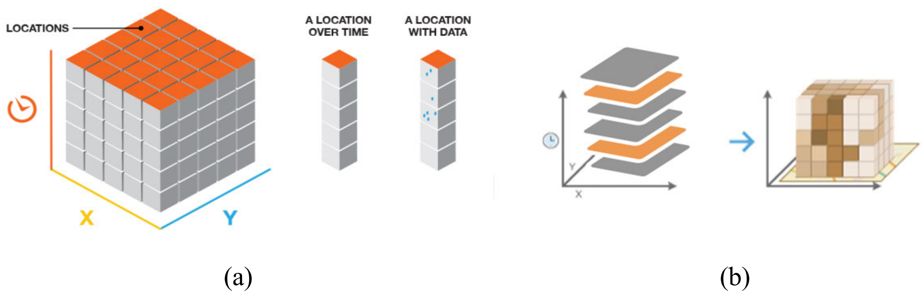


FIGURE 4. The Space-Time Cube: (a) the grid cube; (b) Create a space-time cube from a multidimensional raster layer.

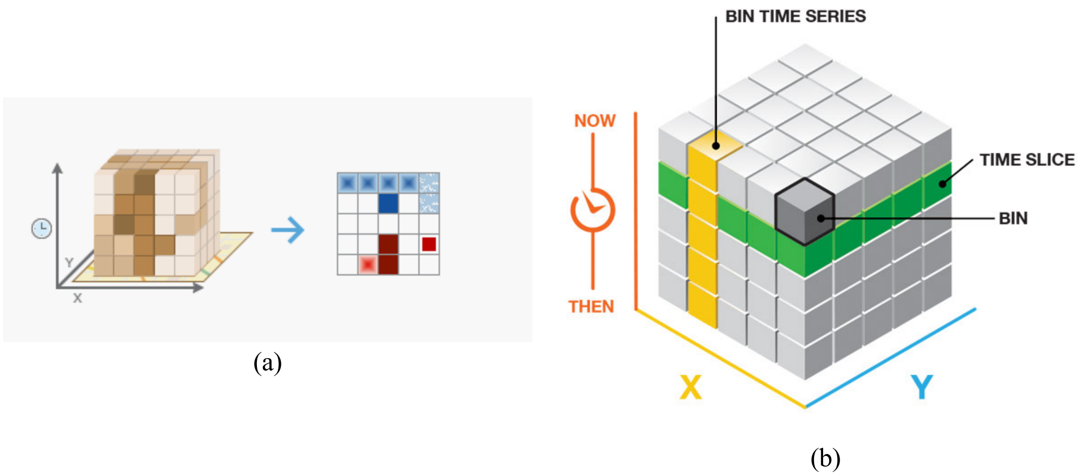


FIGURE 5. The structure of emerging hot spot analysis: (a) Hot spot trends; (b) Each bin has a location ID, a time-step ID, and a count.

multidimensional raster, in which each raster cell of each dimension is converted to a single space-time bin. Interested readers can refer to the study [32], [33] for further information. The space-time cube consists of rows, columns, and time steps. The number of bins in the space-time cube is equal to the number of rows multiplied by the number of columns multiplied by the time step. The rows and columns decide the cube's spatial range, while the time steps decide the temporal range. The space-time cube is shown in FIGURE 4.

2) THE EMERGING HOT SPOT ANALYSIS

Recognize the changing trend of each bin in the space-time cube, including persistent, sporadic, and oscillating hot and cold spots. We used the Getis-Ord Gi* statistic to estimate clustering's intensity, which regards each bin's value within neighborhood bins. To decide which bins will be contained in a specific investigation neighborhood, we first found neighborhood bins that drop within the specified conceptual of spatial relationships. Next, each bin must include bins in the same position from N former time steps, Where N is the value of the neighborhood time step we defined. Then two analyses were conducted: (1) Each bin was analyzed in the neighboring bins to estimate both high and low values' clustering intensity. This analysis results come from the z-score, p-value, and bin's category of each bin in the space-time cube. (2) Then the Mann-Kendall statistic was used to evaluate the time series of these z-scores at the analyzed location. This analysis result was from a clustering trend z-score, p-value, and bin's category for each location. The structure of emerging hot spot analysis is shown in FIGURE 5.

3) THE TIME SERIES CLUSTERING

The time series clustering method is used to identify the types of congestion in the whole network, where the members of each cluster have the same time-series features. Congestion types can be clustered because they show similar evolution patterns over time [34]. The illustration of the time series clustering is shown in FIGURE 6. Clustering aims to divide the space-time cube into multiple groups, where the time series of locations within each group and the locations outside the group are more different from each other. The character's correlation option is used to cluster time series that tend to stay in proportion to each other and increase and decrease value simultaneously. For example, this method helps users distinguish the stores whose sales increase during the Christmas shopping season and decrease after Christmas from the

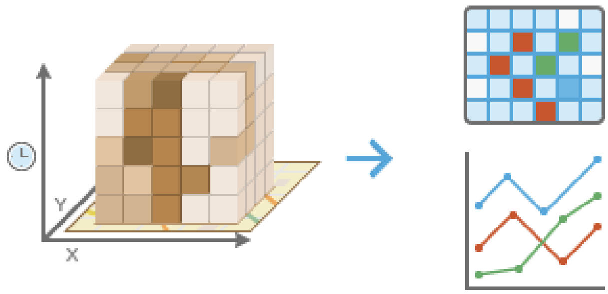


FIGURE 6. The illustration of the time series clustering.

stores whose sales do not have this pattern. The stores with different sales patterns are worth further investigation. This information can also help the retailer predict demand and ensure that stores have sufficient inventory. This option measures the similarity in time series based on their statistical correlation across time. The method calculates the difference between two time series. Based on the definition of similarity between time series, we used the k-medoids algorithm, also called the Partitioning Around Medoids (PAM) algorithm, to cluster the locations of the space-time cube [35], [36].

Determining the number of clusters is one of the most challenging aspects of clustering workflows. This method will evaluate the optimal number of clusters using a pseudo-F statistic [37]. The pseudo-F statistic's larger values indicate that the time series is more similar to the cluster than the entire dataset, indicating effective clustering.

C. THE PREDICTION OF TRAFFIC STATES

We used the time series forecasting methods (e.g., Curve Fit Forecast (CFF), Exponential Smoothing Forecast (ESF), Forest-based Forecast (FBF)) to forecast the traffic states. The description of each model is shown in TABLE 3.

TABLE 3. The description of each model.

Model	Description
Curve Fit Forecast (CFF) [38]	Forecasts the values of each location of a space-time cube using curve fitting
Exponential Smoothing Forecast (ESF) [39]	Forecasts the values of each location of a space-time cube using the Holt-Winters exponential smoothing method by decomposing the time series at each location cube into seasonal and trend components
Forest-based Forecast (FBF) [27], [40]	Forecasts the values of each location of a space-time cube using an adaptation of Leo Breiman's random forest algorithm. The forest regression model is trained using time windows on each location of the space-time cube.

1) FORECASTING AND VALIDATION

These methods build two models when predicting time series. One is the forecast model, which is adapted to predict the future time step value. The other is the validation model, which validates the predicted values.

The forecast model is created by fitting the selected curve type to the time series values at each space-time cube position. Then extrapolate this curve to the future to forecast the value of the future time slice. The curve's fit to each time series is estimated by the Forecast Root Mean Squared Error (F_{RMSE}), which equals the square root of the average squared difference between the curve and the time series' values.

$$F_{RMSE} = \sqrt{\frac{\sum_{t=1}^{T}(c_t - r_t)^2}{T}} \tag{4}$$

where T is the number of time steps, c_t is the value of the curve, and r_t is the raw value of the time series at time t.

FIGURE 7 illustrates how the forecast model and validation model work. The F_{RMSE} estimates the difference between the model's fit value and the original time series value and its fit degree to the original time series value. It does not estimate how well the forecast model forecasts future values. It often closely fit time series but fail to provide accurate predictions when extrapolating. Fortunately, the validation model solves this problem.

The validation model is applied to determine the predictive ability of the prediction model for each time series' future values. It is built by excluding the last part of each time series and fitting the forecast model to the previous data. The accuracy of the forecasts is calculated by the Validation Root Mean Squared Error (V_{RMSE}), which equals the square root of the average squared difference between the forecasted and raw values of the excluded time steps. The validation model is essential because it can directly compare predicted value with the original value to measure the accuracy of the prediction model. Although it is not applied for forecasting, it is applied to justify forecasting models.

$$V_{RMSE} = \sqrt{\frac{\sum_{t=T-M+1}^{T}(c_t - r_t)^2}{m}} \tag{5}$$

where T is the number of time steps, m is the number of time steps withheld for validation, c_t is the value forecasted from the first T-m time steps, and r_t is the raw value of the time series withheld for validation at time t.

V. RESULTS AND DISCUSSION

A. THE DISCRIMINATION OF TRAFFIC STATES

As mentioned before, this study took the road network within the Third Ring Road of Xi'an as a case study. We divided the road network of Xi'an into a small grid with a geographic position. The study area was divided into nearly 14,000 grids according to the size of 100*100m and recorded each grid's number and the road number and road length contained in the grid. We believe that it is essential to decrease the impact of accidental conditions such as traffic accidents and severe weather on TTI value. Therefore, we calculated the average hourly TTI value from 6 a.m. to 10 p.m. on all working days at the one-hour interval to represent the average traffic status. Here we only show the traffic status of the whole-network

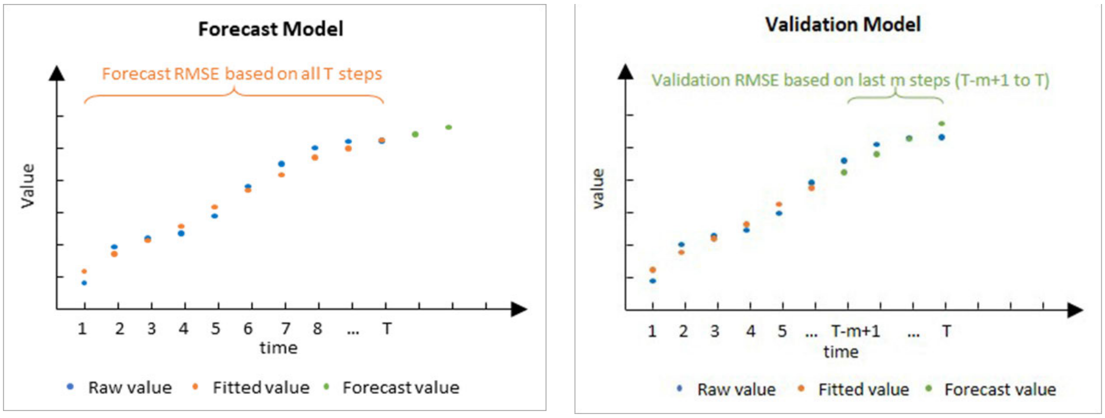
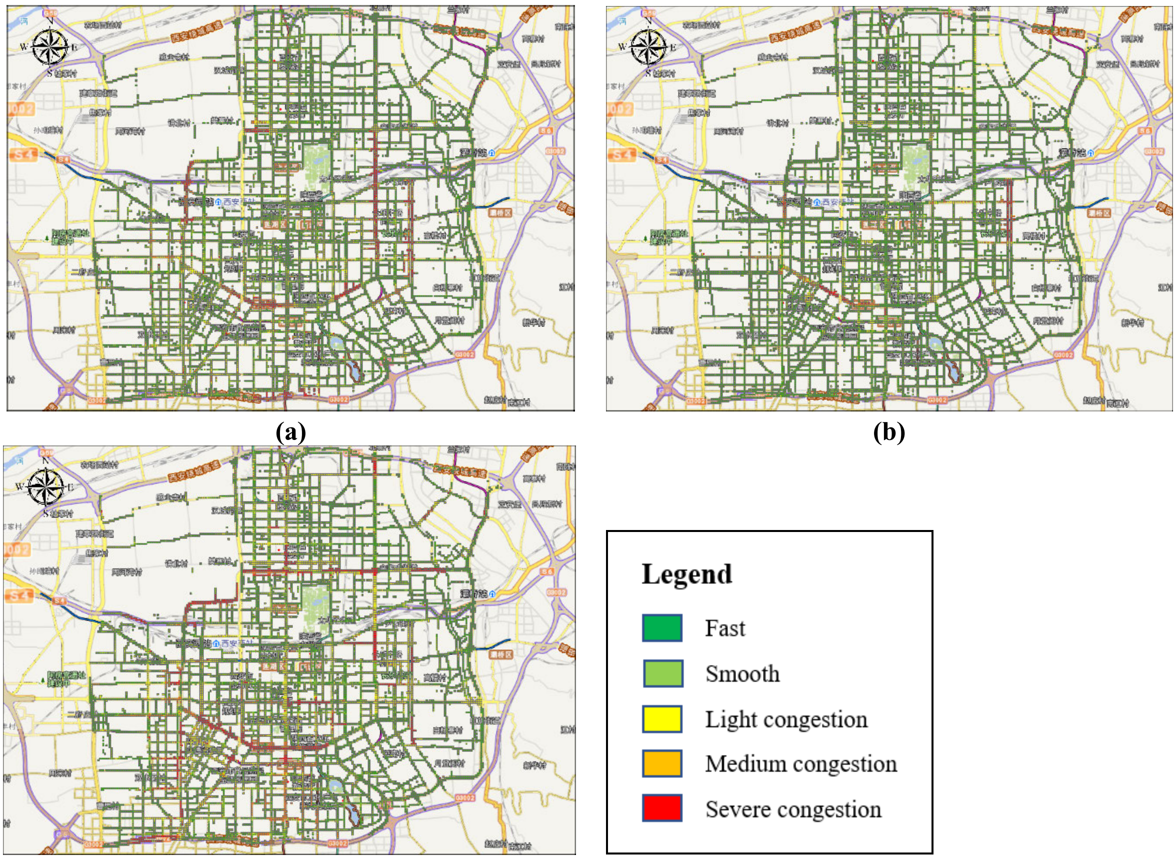


FIGURE 7. **Forecasting and validation: (a) The forecast model; (b) The validation model.**



(c)

FIGURE 8. **The distribution of traffic congestion: (a) The traffic status at 8 a.m.; (b) The traffic status at 1 p.m.; (c) The traffic status at 6 p.m.**

during the morning, noon, and evening rush hours of the working days, as shown in FIGURE 8.

From the above figure, we can intuitively observe the traffic status in each grid in the study area. The different color represents the different traffic statuses in the grid. We can observe that recurrent congestion mainly occurs on expressways and arterial roads and fewer on the urban secondary trunk road. During the morning and evening rush hour, the roads with the most severe recurrent congestion are the same, such as the South Second Ring Road, the East Second Ring Road, and the West Section of North Second Ring Road. The traffic congestion in the evening rush hour is more severe than in the morning and is mainly concentrated in the southwest region of the city because it is the city's political center with a strong economy and a large population. Recurrent congestion is not easy to occur in the north of the city's economic and technological development zone because the population density in this area is minimal.

B. THE SPATIOTEMPORAL PATTERNS MINING

To better describe the spatiotemporal distribution of traffic status, we also analyzed the spatiotemporal patterns of congestion based on the space-time cube model, as shown in FIGURE 9.

According to FIGURE 9, the hot spots of the whole-network congestion are relatively concentrated, including urban arterial roads, commercial centers, and road sections for construction and reconstruction. The persistent hot spots are mainly concentrated on the West Section of South Second Ring Road and the North Section of East Second Ring Road. These locations remained statistically significant hotspots over 90% of the time step interval, with no significant trend indicating an increase or decrease in clustering intensity over time. The sporadic hot spots are mainly concentrated on the Longshou North Road and Fengcheng First Road. These locations are an on-again, off-again hot spot with no time interval being a statistically significant cold spot. The oscillating hot spots are mainly concentrated on the Kunming Road, Xiaozhai West Road, and Changle Middle Road. It indicates that these locations are congested most of the time.

Moreover, the cold spots of the whole-network congestion are widely distributed, mainly including persistent cold spots and oscillating cold spots. The persistent cold spots are mainly concentrated on the north of the East Third Ring Road, the south of the West Third Ring Road, and the northeast region of the city. These locations remained statistically significant cold spots over 90% time step intervals, with no significant trend indicating an increase or decrease in clustering intensity over time. The oscillating cold spots are mainly concentrated in the center and southwest region of the city. It indicates that these locations are congested during part of the time but are less congested most of the time.

Under the interaction of residents' activities and urban planning, urban traffic is affected by time and space variables and has significant aggregation on both time and space scales. Therefore, urban traffic congestion hot spots belong to spatiotemporal events. The cold and hot spots of traffic congestion are closely related to the land use type. Different regions have different functions, resulting in a different spatiotemporal pattern of congestion during weekday periods. It can be seen from FIGURE 9 that the persistent hot spots are clustered at intersections and urban arterial roads, such as the intersection of Yanta North Road and the west section of South Second Ring Road. It indicates that recurrent congestion mainly occurs in the interaction area between the arterial roads and the express road and rarely occurs in the secondary and branch roads. Moreover, tourist areas (such as Dayan Pagoda Park and Xi'an Circumvallation) usually destroy road connectivity, reducing the density and accessibility of the road network, leading to traffic congestion. This result is consistent with previous findings that a sparse road network decreases accessibility and connectivity [41] because it has fewer alternative paths and long detour distances. As a result, vehicles have to use the same road to pass through the area, leading to traffic congestion. The oscillating hot spots were located in a mixed land-use area with commerce, public service departments including the railway stations and hospitals. Traffic congestion near public service facilities and high work-density commercial areas can exacerbate traffic congestion and duration during off-peak hours. The results show that large public departments such as the railway station serve the people of Xi'an and serve other people who come to Xi'an, which may cause severe traffic congestion when it is located in dense residential, tourism, or employment centers. Such a large public facility receives many tourists every day, which means that specific traffic management strategies should be developed to consider the area's actual needs and detailed street design, balancing traffic demand with network capacity. Commuting between residential and work areas intensifies traffic congestion on the road in the morning and evening rush hours. For instance, according to FIGURE 8 and FIGURE 9, the oscillating cold spots are mainly concentrated in the center and southwest region of the city, and these places show severe traffic congestion during the morning and evening rush hours, while the congestion is not apparent during other times. The results show medium congestion in the central urban area, while the southern and western districts were severe congested during the evening rush hours. In general, congestion is more widespread and more severe in the evening rush hours than in the morning. Because in China, most people commute to work between 07:30-09:00. This concentration of travel time has led to a rapid increase in traffic flow and caused severe congestion. The pattern is different at night. People may not choose to go home immediately after work but remain at the office or participate in entertainment activities. Thus, return trips become more distributed over time, reducing rush-hour traffic and reducing nighttime congestion, but extending its duration. These conclusions are consistent with previous studies' results that commuting between work and home generates congestion on arterial roads [5].

C. THE TIME SERIES CLUSTERING

To identify the most similar locations of congestion types in the whole-network, we used time series clustering to divide each grid into different clusters, where each cluster has similar spatiotemporal characteristics. The clustering result ((FIGURE 10 and FIGURE 11) shows three types of congestion within the Third Ring Road of Xi'an on working days. According to three clusters, over time, the congestion of the whole-network has the following spatiotemporal characteristics:

The distribution of cluster 1(FIGURE 10) is consistent with the persistent cold spots in FIGURE 9, verifying the method's

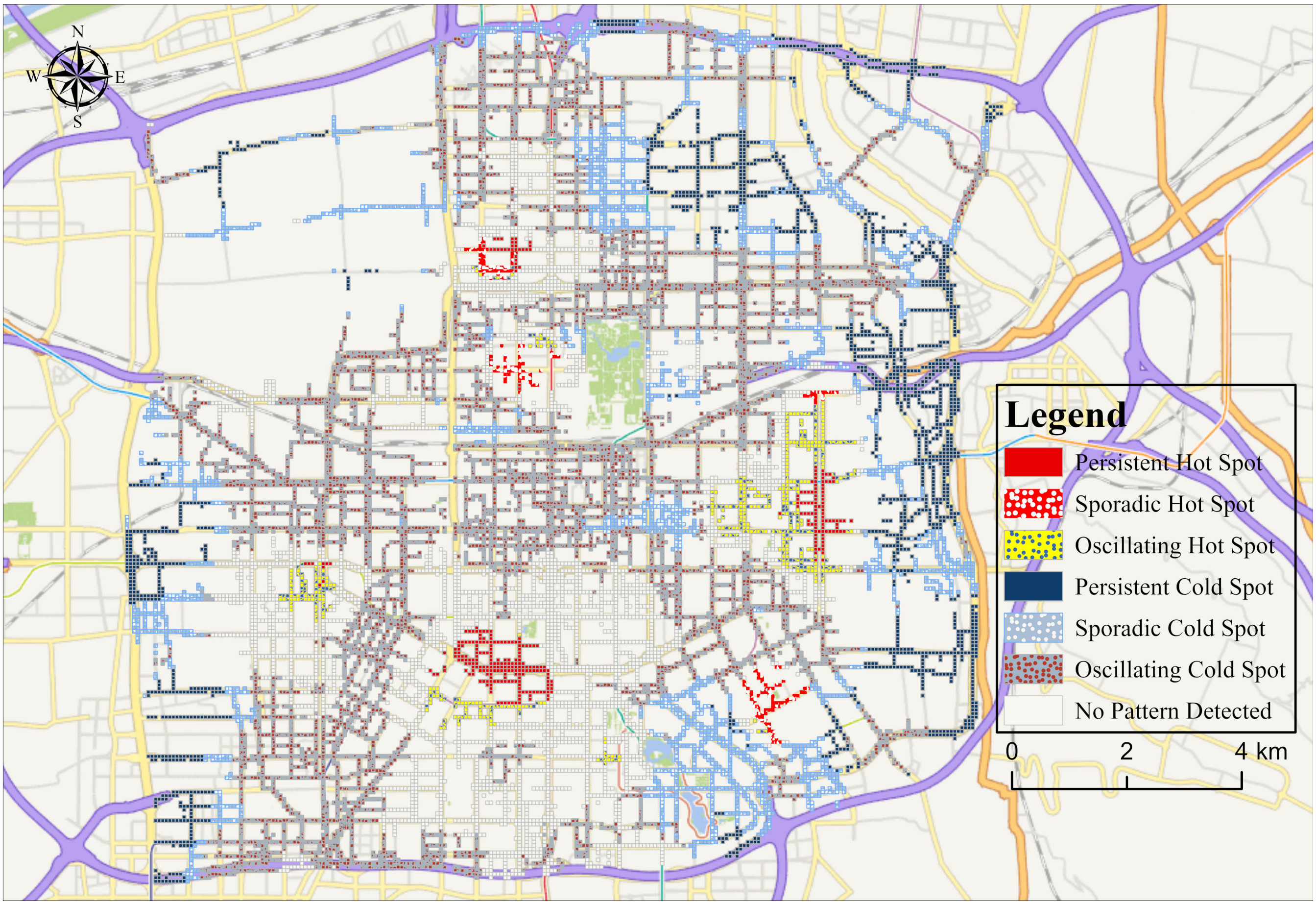


FIGURE 9. The spatiotemporal patterns of traffic status.

accuracy from the side. Cluster 1 is mainly concentrated in the northern, eastern and westernmost economic and technological development zones because of the low population density and the less traffic demand. The TTI value of cluster 1 is always maintained between 2-4, indicating that the traffic status is smooth (TABLE 2).

Cluster 2 has an apparent tendency to be congested in the evening rush hour. Combined with POI distribution in Xi'an, cluster 2 is mainly concentrated in arterial roads, business districts, and entertainment districts. Unlike the feature that the morning rush hour congestion quickly forms and rapidly dissipates, the evening rush hour congestion lasts for a long time and is more widely distributed [16]. The morning rush hour is mainly for commuting and school trips. The time is concentrated, and the purpose is single, which leads to the high peak of the flow of people and vehicles in the morning rush hour, causing more severe traffic congestion. On the one hand, in the evening rush hours, the time after work and school is more scattered, so people can avoid the rush hours to travel. On the other hand, many people have the habit of going shopping, dining, and entertainment after work, which postpones returning time and increases the scope and duration of congestion. However, the evening rush hour appears to be more "congested" mainly due to poor lighting conditions in the evening rush hour, prone to traffic accidents. The driving speed may be slower when coming home from work than when going to work. The parking space in shopping, catering, and entertainment places are insufficient. The side parking phenomenon is widespread, seriously affecting the level of road service [42].

Cluster 3 is the commuter roads with obvious morning and evening rush hour congestion. It is mainly distributed in the city center (where most people work) and the Yanta District, which has the highest population of the five core districts (1.30 million) [43]. Thus, when commuters return home, numerous vehicles flood into the area for a short time, causing severe traffic congestion. The most congested time in the morning and evening of the whole road network is 8 a.m. and 6 p.m. respectively. It shows the most typical working hours for residents: many firms and schools start between 7.30 and 8:30 a.m., and commuters need to leave home an hour or two earlier, so traffic volume increases rapidly from 6 a.m. and dissipating after 8 a.m. At night, most companies and schools close between 5 p.m. and 6.30 p.m. Traffic flow usually lag behind the closing time, so the evening rush hour occurs from 5 p.m. to 7 p.m. There is more severe traffic congestion near

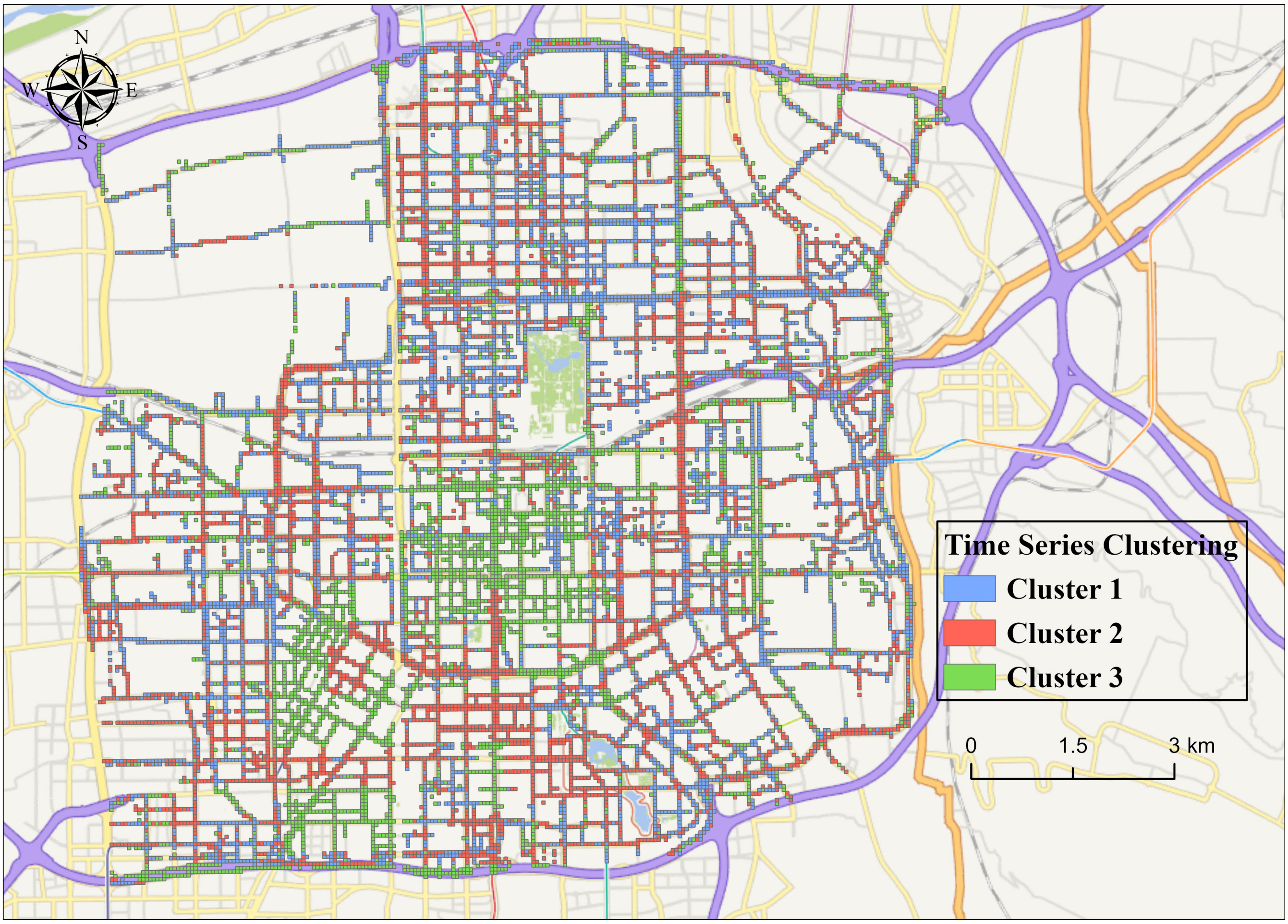


FIGURE 10. The distribution of time series clustering.

large hospitals, shopping malls, and schools. For instance, the severe traffic congestion near a primary school, which means the most significant vehicle source is the parents picking up their children—especially in the rush hours, increasing traffic congestion on the roads. It is consistent with previous research found in the morning rush-hour driving to school accounts for 15% of all trips [44].

D. THE PREDICTION OF TRAFFIC STATES

More than 600 million trajectory data from 21 working days was used as the dataset in the current study. Approximately 80% of the data (17 days) was used as the training set, and the remaining 20% (4 days) was the test set. As mentioned above, we chose F_{RMSE} and V_{RMSE} indicators to measure three different models' performance. To better illustrate the experimental results, we calculated each model's indicators values as shown in TABLE 4.

The F_{RMSE} measures the curve's fit to the raw time series values and the V_{RMSE} measures the curve's ability to predict future values. According to TABLE 4, it can be found that in the Curve Fit Forecast (CFF) model, the average value of the two indicators is the minimum, indicating that the model's prediction accuracy is higher than the other two models. Furthermore, the standard deviation of the two indicators in the CFF model is also the minimum, indicating that the model is more stable than the other two models. In summary, the CFF model can accurately and stably predict the road network's traffic operation status within the Third Ring Road in Xi'an, thereby providing road condition information in advance and providing convenience for residents' travel. At the same time,

TABLE 4. The result of three different models.

Indicators		Models		
		CFF	ESF	FBF
F_{RMSE}	minimum value	0.00	0.01	0.00
	maximum value	3.07	3.14	3.15
	average value	0.61	0.87	0.68
	standard deviation	0.87	1.34	1.25
V_{RMSE}	minimum value	0.00	0.00	0.00
	maximum value	1.89	2.42	2.59
	average value	0.57	0.70	0.64
	standard deviation	0.71	1.08	0.94

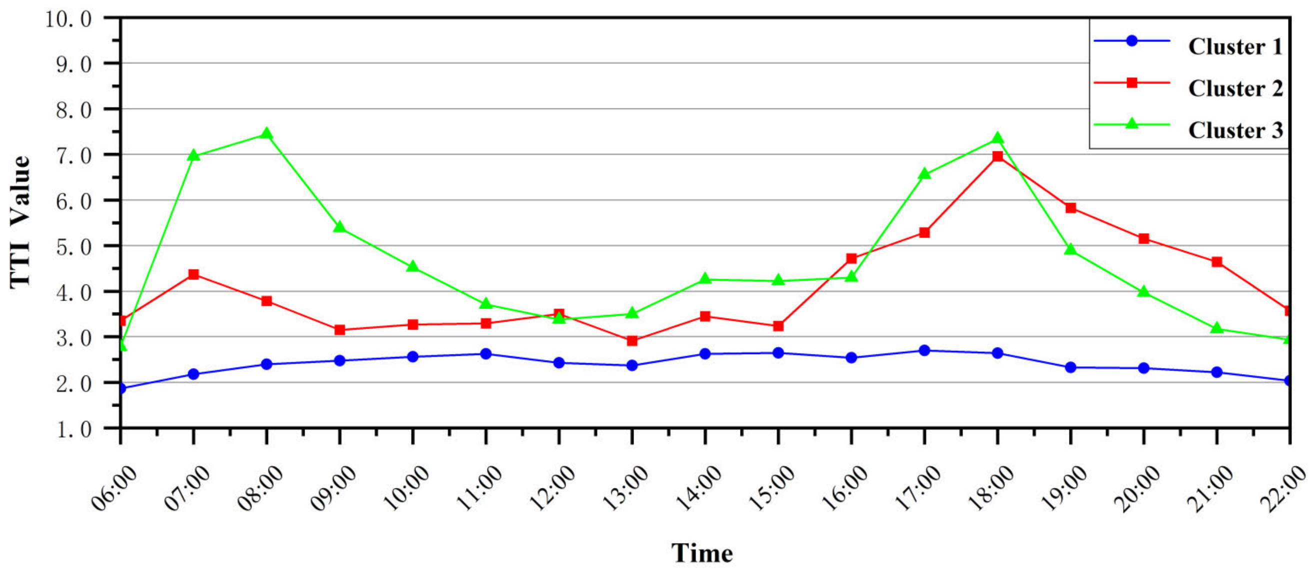


FIGURE 11. The result of time series clustering.

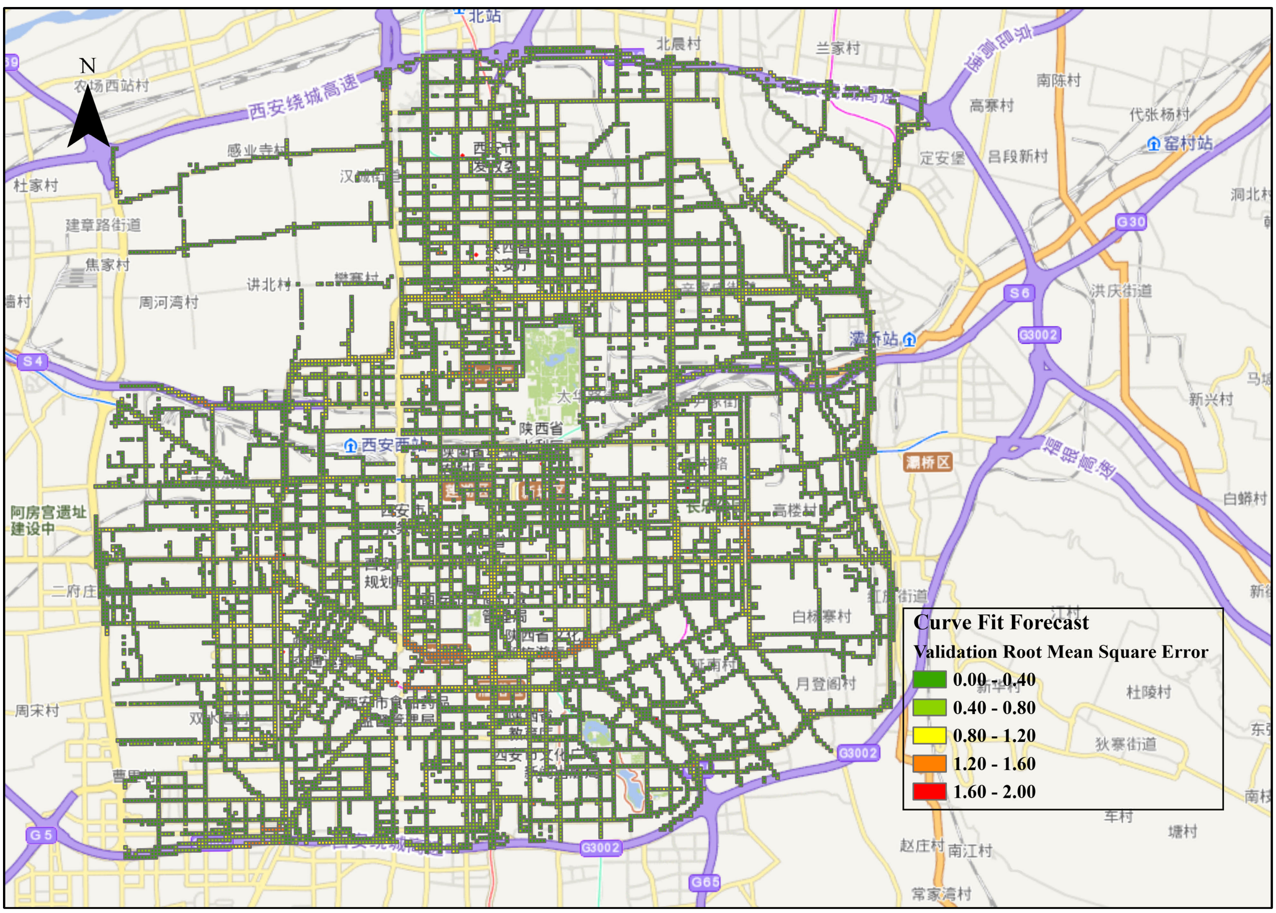


FIGURE 12. The prediction result of Curve Fit Forecast model.

it lays the foundation for realizing the source diagnosis of traffic congestion and avoiding recurrent congestion.

To better illustrate the CFF model's result, we show the distribution of V_{RMSE} within the Third Ring Road in Xi'an in FIGURE 12. The predicted value of most areas is very close to the actual value, which shows that the CFF model can genuinely predict the traffic operation status of the road network.

Based on the spatiotemporal characteristics of traffic congestion in Xi'an demonstrated in this study, we contribute some helpful insights for alleviating traffic congestion:

(1) Based on the congestion evaluations of the whole network, it is essential to alleviate the congestion on arterial roads. It is crucial to speed up the construction of urban traffic microcirculation, increase the road network's density, and improve its capacity. It is worth highlighting the need to continue to make specific improvements to transport facilities, such as traffic signs and traffic line markings, redrawing zebra crossings, adding safety islands, and building pedestrian overpasses to ensure smooth roads.

(2) Our results proved that the most-congested area in the evening rush hour was clustered in shopping, catering, and entertainment places. More severe is that there are insufficient parking spaces near these locations, leading to widespread side parking phenomenon, seriously affecting road service levels. It is essential to make full use of idle land and increase the construction of parking spaces. Given the different characteristics of vehicle parking requirements in different areas, the shared parking policy is comprehensively promoted to alleviate parking problems actively.

(3) It is essential to adhere to the urban public transport priority development strategy and increase the public transportation travel sharing rate. Furthermore, we will continue to optimize the public transportation network to attract citizens to travel public transportation and improve service quality. We will strengthen our capacity for an emergency response to traffic congestion and promote the intelligent transformation and networked application of all signal lights in the city.

VI. CONCLUSION

This study aims to identify and predict the traffic operation status in the road network within the Third Ring Road of Xi'an and explore spatiotemporal patterns of traffic congestion. For the modeling purpose, we used the urban road network data and GPS trajectory data of floating vehicles (e.g., taxis and buses) in Xi'an provided by the Xi'an Transportation Administration of China. This study focuses on the traffic congestion states on weekdays, so 21 working days were selected for analysis, totaling more than 600 million data, with a storage scale of about 65GB. A comprehensive data model combining GPS trajectory and the urban grid was constructed by analyzing the GPS data of taxis and buses and urban road network data. This study uses Travel Time Index (TTI) to evaluate road traffic performance. We used the emerging hot spot analysis method to identify new, intensifying, persistent, or sporadic hot spot patterns at one-hour intervals. We use the time-series clustering method to cluster the traffic status of the whole network into three types so that each cluster has similar spatiotemporal characteristics. In order to compare the prediction performance of different models, three different time series prediction models are used in this study: Curve Fit Forecast (CFF), Exponential Smoothing Forecast (ESF), Forest-based Forecast (FBF). We selected Validation Root Mean Square Error (V_{RMSE}) and Forecast Root Mean Square Error (F_{RMSE}) to measure the forecast performance.

According to the result of this study, over time, the congestion of the whole-network has the following spatiotemporal characteristics: (1) The recurrent congestion mainly occurs on the intersections and urban arterial roads, such as the Yanta North Road and the west section of South Second Ring Road. (2) The result shows three traffic status types within the Third Ring Road of Xi'an on working days, including the traffic status are always smooth, apparent congestion during the evening rush hour, and the commuter roads with obvious morning and evening rush hour congestion. (3) The average value of the F_{RMSE} and V_{RMSE} in the CFF model is the minimum indicating that it can accurately and stably predict the road network's traffic operation status.

The results of this study can contribute some helpful insights for alleviating traffic congestion. For instance, it is essential to speed up the construction of urban traffic microcirculation, increase the road network's density, and improve the road network's capacity. Furthermore, it is necessary to make full use of idle land and increase parking spaces' construction. Given the different characteristics of vehicle parking requirements in different areas, the shared parking policy is comprehensively promoted to alleviate parking problems actively. Moreover, it is crucial to adhere to the urban public transport priority development strategy and increase the public transportation travel sharing rate.

Meanwhile, the study has some limitations. Firstly, the data applied in this study could be enlarged to promote the reliability of model results. Future studies should consider integrating the GPS data of private vehicles to overcome the instability of taxi GPS. Secondly, although this study used three methods to predict traffic operation status, other machine learning approaches are also worth exploring. Future studies should try to integrate different methods to predict traffic conditions. Thirdly, this study uses only TTI for different periods to predict. Future research should rely on existing multi-source data and deep learning technologies to dynamically modify model parameters and algorithms to improve the accuracy and scientific road traffic performance index.

REFERENCES

[1] Y. Zheng, Y. Li, C.-M. Own, Z. Meng, and M. Gao, "Real-time predication and navigation on traffic congestion model with equilibrium Markov chain," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 4, Apr. 2018, Art. no. 155014771876978, doi: 10.1177/1550147718769784.

[2] F. Wen, G. Zhang, L. Sun, X. Wang, and X. Xu, "A hybrid temporal association rules mining method for traffic congestion prediction," *Comput. Ind. Eng.*, vol. 130, pp. 779–787, Apr. 2019, doi: 10.1016/j.cie.2019.03.020.

[3] Z. Huang, J. Xia, F. Li, Z. Li, and Q. Li, "A peak traffic congestion prediction method based on bus driving time," *Entropy*, vol. 21, no. 7, p. 709, Jul. 2019, doi: 10.3390/e21070709.

[4] J. Mena-Oreja and J. Gozalvez, "A comprehensive evaluation of deep learning-based techniques for traffic prediction," *IEEE Access*, vol. 8, pp. 91188–91212, 2020, doi: 10.1109/ACCESS.2020.2994415.

[5] Y. Long and J.-C. Thill, "Combining smart card data and household travel survey to analyze jobs–housing relationships in beijing," *Comput., Environ. Urban Syst.*, vol. 53, pp. 19–35, Sep. 2015, doi: 10.1016/j.compenvurbsys.2015.02.005.

[6] S. Lauf, D. Haase, and B. Kleinschmit, ''The effects of growth, shrinkage, population aging and preference shifts on urban development—A spatial scenario analysis of Berlin, Germany,'' *Land Use Policy*, vol. 52, pp. 240–254, Mar. 2016, doi: 10.1016/j.landusepol.2015.12.017.
[7] S. El Hamdani and N. Benamar, ''A comprehensive study of intelligent transportation system architectures for road congestion avoidance,'' in *Ubiquitous Networking* (Lecture Notes in Computer Science), vol. 10542, E. Sabir, A. G. Armada, M. Ghogho, and M. Debbah, Eds. Cham, Switzerland: Springer, 2017, doi: 10.1007/978-3-319-68179-5_9.
[8] V.-T. Ta and A. Dvir, ''A secure road traffic congestion detection and notification concept based on V2I communications,'' *Veh. Commun.*, vol. 25, Oct. 2020, Art. no. 100283, doi: 10.1016/j.vehcom.2020.100283.
[9] S. An, H. Yang, J. Wang, N. Cui, and J. Cui, ''Mining urban recurrent congestion evolution patterns from GPS-equipped vehicle mobility data,'' *Inf. Sci.*, vol. 373, pp. 515–526, Dec. 2016, doi: 10.1016/j.ins.2016.06.033.
[10] T. Afrin and N. Yodo, ''A survey of road traffic congestion measures towards a sustainable and resilient transportation system,'' *Sustainability*, vol. 12, no. 11, pp. 1–23, 2020, doi: 10.3390/su12114660.
[11] C. Ma, J. Zhou, X. Xu, and J. Xu, ''Evolution regularity mining and gating control method of urban recurrent traffic congestion: A literature review,'' *J. Adv. Transp.*, vol. 2020, pp. 1–13, Jan. 2020, doi: 10.1155/2020/5261580.
[12] J. C. Falcocchio and H. S. Levinson, ''Measuring traffic congestion,'' in *Road Traffic Congestion: A Concise Guide* (Springer Tracts on Transportation and Traffic), vol. 7. Cham, Switzerland: Springer, 2015, doi: 10.1007/978-3-319-15165-6_8.
[13] M. Vaziri, ''Development of highway congestion index with fuzzy set models,'' *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1802, no. 1, pp. 16–22, Jan. 2002, doi: 10.3141/1802-03.
[14] J. D. Hall, ''Pareto improvements from lexus lanes: The effects of pricing a portion of the lanes on congested highways,'' *J. Public Econ.*, vol. 158, pp. 113–125, Feb. 2018, doi: 10.1016/j.jpubeco.2018.01.003.
[15] M. Aftabuzzaman, ''Measuring traffic congestion—A critical review,'' in *Proc. 30th Australas. Transp. Res. Forum*, Sep. 2007, pp. 1–6.
[16] Y. Chen, C. Chen, Q. Wu, J. Ma, G. Zhang, and J. Milton, ''Spatial-temporal traffic congestion identification and correlation extraction using floating car data,'' *J. Intell. Transp. Syst. Technol. Planning, Oper.*, vol. 25, no. 3, pp. 1–18, 2020, doi: 10.1080/15472450.2020.1790364.
[17] F. Su, H. Dong, L. Jia, and X. Sun, ''On urban road traffic state evaluation index system and method,'' *Modern Phys. Lett. B*, vol. 31, no. 1, Jan. 2017, Art. no. 1650428, doi: 10.1142/S0217984916504285.
[18] X. Yu, S. Xiong, Y. He, W. E. Wong, and Y. Zhao, ''Research on campus traffic congestion detection using BP neural network and Markov model,'' *J. Inf. Secur. Appl.*, vol. 31, pp. 54–60, Dec. 2016, doi: 10.1016/j.jisa.2016.08.003.
[19] N. Hao, Y. Feng, K. Zhang, G. Tian, L. Zhang, and H. Jia, ''Evaluation of traffic congestion degree: An integrated approach,'' *Int. J. Distrib. Sensor Netw.*, vol. 13, no. 7, Jul. 2017, Art. no. 155014771772316, doi: 10.1177/1550147717723163.
[20] X. Yang, S. Luo, K. Gao, T. Qiao, and X. Chen, ''Application of data science technologies in intelligent prediction of traffic congestion,'' *J. Adv. Transp.*, vol. 2019, pp. 1–14, Apr. 2019, doi: 10.1155/2019/2915369.
[21] Q. Shen, X. Ban, and C. Guo, ''Urban traffic congestion evaluation based on kernel the semi-supervised extreme learning machine,'' *Symmetry*, vol. 9, no. 5, p. 70, May 2017, doi: 10.3390/sym9050070.
[22] G.-y. Jiang, J.-f. Wang, X.-d. Zhang, and L.-h. Gang, ''The study on the application of fuzzy clustering analysis in the dynamic identification of road traffic state,'' in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Oct. 2003, pp. 1149–1152, doi: 10.1109/ITSC.2003.1252723.
[23] W. Weijermars and E. van Berkum, ''Analyzing highway flow patterns using cluster analysis,'' in *Proc. IEEE Intell. Transp. Syst.*, Sep. 2005, pp. 308–313, doi: 10.1109/ITSC.2005.1520157.
[24] H. Wen, J. Sun, and X. Zhang, ''Study on traffic congestion patterns of large city in China taking Beijing as an example,'' *Procedia Social Behav. Sci.*, vol. 138, pp. 482–491, Jul. 2014, doi: 10.1016/j.sbspro.2014.07.227.
[25] Y. Kamarianakis, W. Shen, and L. Wynter, ''Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO,'' *Appl. Stochastic Models Bus. Ind.*, vol. 28, no. 4, pp. 297–315, Jul. 2012, doi: 10.1002/asmb.1937.
[26] J. W. C. van Lint, S. P. Hoogendoorn, and H. J. van Zuylen, ''Accurate freeway travel time prediction with state-space neural networks under missing data,'' *Transp. Res. C, Emerg. Technol.*, vol. 13, nos. 5–6, pp. 347–369, Oct. 2005, doi: 10.1016/j.trc.2005.03.001.
[27] H. Chen, H. Chen, R. Zhou, Z. Liu, and X. Sun, ''Exploring the mechanism of crashes with autonomous vehicles using machine learning,'' *Math. Problems Eng.*, vol. 2021, pp. 1–10, Feb. 2021.
[28] 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 Behav. Sci.*, vol. 96, pp. 653–662, Nov. 2013, doi: 10.1016/j.sbspro.2013.08.076.
[29] S. An, H. Yang, J. Wang, N. Cui, and J. Cui, ''Estimating freeway route travel time reliability from data on component links and associated cost implications,'' *Inf. Sci.*, vol. 373, no. 3, pp. 515–526, 2016, doi: 10.1016/j.ins.2016.06.033.
[30] Y. Zou, T. Zhu, Y. Xie, L. Li, and Y. Chen, ''Examining the impact of adverse weather on travel time reliability of urban corridors in shanghai,'' *J. Adv. Transp.*, vol. 2020, pp. 1–11, Dec. 2020, doi: 10.1155/2020/8860277.
[31] *Specification for Urban Traffic Performance Evaluation*, document GB/T 33171-2016, 2016.
[32] H. Shimazaki and S. Shinomoto, ''A method for selecting the bin size of a time histogram,'' *Neural Comput.*, vol. 19, no. 6, pp. 1503–1527, Jun. 2007, doi: 10.1162/neco.2007.19.6.1503.
[33] G. R. Terrell, D. W. Scott, G. R. Terrell, and D. W. Scott, ''Oversmoothed non para metric density estimates,'' *J. Amer. Stat. Assoc.*, vol. 80, no. 389, pp. 209–214, 2014, doi: 10.2307/2288074.
[34] P. Montero and J. A. Vilar, ''TSclust: AnRPackage for time series clustering,'' *J. Stat. Softw.*, vol. 62, no. 1, pp. 1–43, 2014, doi: 10.18637/jss.v062.i01.
[35] D. Arthur and S. Vassilvitskii, ''K-means++: The advantages of careful seeding,'' in *Proc. Annu. ACM-SIAM Symp. Discret. Algorithms*, vols. 7–9, 2007, pp. 1027–1035.
[36] S. Lloyd, ''Least squares quantization in PCM,'' *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982, doi: 10.1109/TIT.1982.1056489.
[37] A. K. Jain, ''Data clustering: 50 years beyond K-means,'' *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
[38] R. Klosterman, K. Brooks, J. Drucker, E. Feser, and H. Resnski, *Planning Support Methods: Urban and Regional Analysis*. Lanham, MD, USA: Rowman. 2018.
[39] R. Hyndman, A. Koehler, K. Ord, and R. Snyder, ''Forecasting with exponential smoothing,'' in *The State Space Approach*, 2008, doi: 10.1007/978-3-540-71918-2.
[40] H. Chen, H. Chen, Z. Liu, X. Sun, and R. Zhou, ''Analysis of factors affecting the severity of automated vehicle crashes using XGBoost model combining POI data,'' *J. Adv. Transp.*, vol. 2020, pp. 1–12, Nov. 2020, doi: 10.1155/2020/8881545.
[41] J. Song, C. Zhao, S. Zhong, T. A. S. Nielsen, and A. V. Prishchepov, ''Mapping spatio-temporal patterns and detecting the factors of traffic congestion with multi-source data fusion and mining techniques,'' *Comput., Environ. Urban Syst.*, vol. 77, Sep. 2019, Art. no. 101364, doi: 10.1016/j.compenvurbsys.2019.101364.
[42] Z. Chen, Y. Jiang, and D. Sun, ''Discrimination and prediction of traffic congestion states of urban road network based on spatio-temporal correlation,'' *IEEE Access*, vol. 8, pp. 3330–3342, 2020, doi: 10.1109/ACCESS.2019.2959125.
[43] C. Yan, X. Wei, X. Liu, Z. Liu, J. Guo, Z. Li, Y. Lu, and X. He, ''A new method for real-time evaluation of urban traffic congestion: A case study in Xi'an, China,'' *Geocarto Int.*, vol. 35, no. 10, pp. 1033–1048, Jul. 2020, doi: 10.1080/10106049.2018.1552325.
[44] M. Lu, C. Sun, and S. Zheng, ''Congestion and pollution consequences of driving-to-school trips: A case study in Beijing,'' *Transp. Res. D, Transp. Environ.*, vol. 50, pp. 280–291, Jan. 2017, doi: 10.1016/j.trd.2016.10.023.

RUIYU ZHOU received the B.E. degree in accounting from the Taiyuan University of Technology, Taiyuan, China, in 2017. She is currently pursuing the Ph.D. degree in transportation engineering with Chang'an University, Xi'an, China. Her research interests include transportation economy and traffic control.

HONG CHEN received the M.S. and Ph.D. degrees in transportation engineering from Chang'an University, China, in 1999 and 2006, respectively. She is currently a Professor with the College of Transportation Engineering, Chang'an University. Her research interests include traffic simulation, transportation planning, and traffic safety.

HENGRUI CHEN received the B.E. degree in transportation engineering from Chang'an University, Xi'an, China, in 2018, where he is currently pursuing the Ph.D. degree in transportation engineering. His research interests include traffic safety analysis and traffic simulation.

ENZE LIU received the B.E. degree in transportation engineering from Chang'an University, Xi'an, China, in 2018, where he is currently pursuing the Ph.D. degree in transportation engineering. His research interest includes intelligent transportation systems.

SHANGJING JIANG was born in 1995. She is currently pursuing the master's degree. Her main research interests include spatial-temporal analysis and field model.