Journal of Advanced Transportation

Synchronization and Resource Sharing in Intelligent Transport Systems

Lead Guest Editor: Yong Wang
Guest Editors: Jinjun Tang and Xiangyang Guan

Synchronization and Resource Sharing in Intelligent Transport Systems

Journal of Advanced Transportation

Synchronization and Resource Sharing in Intelligent Transport Systems

Lead Guest Editor: Yong Wang
Guest Editors: Jinjun Tang and Xiangyang Guan

Contents

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 1422381, 12 pages
https://doi.org/10.1155/2021/1422381

Research Article

Modeling Tourists' Departure Time considering the Influence of Multisource Traffic Information

Shijun Yu,[1] Siyuan Zhang,[2] Shejun Deng,[1] Tao Ji,[1] Peng Zhou,[1] and Lang Peng[1]

[1]*Yangzhou University, University South Road No. 88, Yangzhou, Jiangsu, China*
[2]*Jiangsu College of Tourism, Yuxiu Road No. 88, Yangzhou, Jiangsu, China*

Correspondence should be addressed to Siyuan Zhang; zhangsiyuan2019@126.com and Shejun Deng; dsj@yzu.edu.cn

Received 22 July 2021; Accepted 5 October 2021; Published 25 October 2021

Academic Editor: Jinjun Tang

The development of tourism brings economic benefits as well as additional pressure on the urban traffic system. For example, the travel time of tourists coincides with the rush hour of urban residents' daily commuting. Limited urban traffic resources cannot meet the travel needs of tourists and urban residents at the same time, resulting in traffic congestion and low travel efficiency. Now, with the development of intelligent technology, tourists can obtain real-time information about transportation systems through various channels and adjust their travel behavior accordingly. This study shows tourists' travel behavior based on a survey conducted to the tourists in Yangzhou city. 1500-interview data are analyzed, and a Multinomial Logit Model (MNL) was employed to establish the probability prediction model of tourists' departure time choice. The results presented that sync traffic information and some other tourism-related factors determine the choice of tourists' departure time. These factors distinguish the travel behavior of tourists from the daily travel behavior of urban residents. This study can provide suggestions for the urban tourism management department to formulate more targeted and efficient policies while creating a more comfortable tourism environment for tourists.

1. Introduction

With the rapid growth of economy, tourism has experienced a dramatic development. It has changed from a traditional hospitality business to a new growth point in the national social economy [1]. The development of tourism brings spiritual joy to the people and huge economic benefits to society, along with a series of accompanying effects [2–4]. The tourism traffic demand causes serious traffic congestions in many cities with tourism as a pillar industry (such as Yangzhou and Suzhou). Many traffic management policies and strategies have been taken to solve this problem, but the result is not as good as expected [5–7]. One of the most important reasons is that decision makers cannot know exactly the tourists' travel demand. Travel demand analysis studies have traditionally modeled many travel components (e.g., travel mode, departure time, and trip destination) using residents' travel data [8–11].

Many researchers have intensely focused on departure time choice, especially the workday departure time choice of urban commuters. Traditionally departure time has been modeled using discrete choice models. Bhat [12] believed that travel mode and departure time choice of urban trips are important determinants of urban travel demand and adopted a multinomial logit (MNL) model to examine the joint nature of mode and departure time choice for urban shopping trips. Lemp et al. [13] viewed departure time choice as a continuous response variable and formulated a continuous cross-nested logit (CCNL) model using Bayesian estimation techniques and San Francisco Bay Area data. Shabanpour et al. [14] adopted the discrete choice approach to model activity timing decisions and a hybrid utility maximization and developed a regret minimization model to account for the heterogeneity of decision rules across choice variables, while some recent studies have argued that because of the continuous nature of time, it is more accurate to consider it as a continuous variable. Furthermore, modeling departure time as a continuous variable frees researchers from constraints of discrete choice models. Gadda et al. [15]

offered Bayesian estimates of continuous departure time models using accelerated failure time specifications for various trip purposes with several distributional specifications. Habib et al. [16] used a discrete-continuous econometric model to investigate the joint decisions of trip timing and mode choice for commuting trips in the Greater Toronto Area (GTA). Amirgholy et al. [17] proposed a continuum approximation model to optimize the line spacing, stop spacing, headway, and fare of the transit system. Moreover, since the mode choice decision is closely intertwined with the decision on trip departure time, several studies considered them as a joint decision to account for the interrelation or causal effects between them. Paleti et al. [18] developed an integrated model of mode and trip departure TOD choices by using both HP and stated preference (SP) data from the large-scale household travel survey undertaken in Jerusalem in 2010. Shabanpour et al. [19] presented a cluster-based joint modeling approach to investigate heterogeneous travelers' behavior toward trip mode and departure time choices by considering those choices as a joint decision.

Recently, information technologies have been rapidly advanced. They have the greatest potential for use in the research of travelers' travel behavior. In terms of data collection, different from the traditional questionnaire survey, advanced information technology makes it possible to obtain data more conveniently and accurately. Mobile communication systems such as GPS (global positioning systems), cellular phone, and RF-ID (Radio Frequency Identification) system can be used as the core instruments of tracking survey of a traveller's behavior [20]. At the same time, information technology is also widely used in the analysis of the impact on travelers' travel behavior. Hadi et al. analyzed multimodal journeys for information on transit planning using the smart card system in London by identifying the multimodal transfer combinations of bus-to-underground, underground-to-bus, and bus-to-bus. The Florida Department of Transport (FDOT) is already preparing for the expected rise of advanced traffic management strategies summarized as Intelligent Transportation Systems (ITS) [21]. Meanwhile, many studies have shown that real-time information can adapt to changing traffic conditions and assist travelers to make better travel decisions in uncertain networks [22–24].

Although all of these studies have been conducted to research urban commuters' travel behavior, the travel behavior of tourists has received relatively little attention in previous studies. Tourists share many preferences with commuters, on whom most of our knowledge of traveler behavior is based [25–27]. However, due to the restrictions of travel purpose and traffic conditions, the transportation choice behavior of tourists is significantly different from that of urban commuters. Tourists mainly travel for leisure and sightseeing, so they may choose a more comfortable mode of transportation and departure time with less urban traffic pressure. Moreover, there are fewer transportation mode options for tourists than urban commuters, and public transportation and private cars are the main ones. More importantly, tourists' travel behavior is the main factor that affects their departure time choices, such as the acquisition of real-time information around scenic spots and tour schedules. However, existing studies on departure time choice are mostly based on the characteristics of urban residents or commuters. There are only a few studies focused on modeling the travel choices of tourists during the tour based on tourist data and considering the influence of the known destination traffic information on the travel behavior of tourists.

To contribute to this gap, this study takes tourists' choice of departure time as the research object and obtains tourist data through a survey. Based on the tourist data obtained from the survey, a tourists' departure time choice prediction model is established. The model is proposed to improve the understanding of tourists' travel behavior and predict accurate tourists' travel demand. Compared with the traditional model based on urban commuters, the model proposed in this paper can more accurately describe the traffic behavior of tourists.

2. Data Preparation

2.1. The Influence Mechanism of Multisource Traffic Information. The multisource traffic information system provides travel information to tourists through multiple channels. The purpose is to provide multiple information to help tourists make the best travel decisions and play a role in guiding tourists' travel behavior. After obtaining multisource traffic information, tourists will recognize and judge the diversified travel information based on their understanding and then make corresponding decision-making behaviors. The basic mode of travel behavior decision-making of tourists under the influence of multisource traffic information is shown in Figure 1.

2.2. Factors Affecting the Departure Time of Tourists

2.2.1. Tourist Personal Attributes Factors. The attributes of tourists have an important influence on the choice of travel time. Due to differences in individual attributes, tourists will have different choices of travel time. Specific personal attributes include gender, age, income, and occupation.

(1) Age. The age of tourists is also one of the factors that affect the choice of departure time. Young people generally prefer a later departure time, while most elderly people generally choose an earlier departure time. At the same time, different age groups may choose different departure times due to their income level, tourism purpose, and travel distance.

(2) Income Level. Tourists with higher income pay more attention to the pursuit of comfort during the travel process, and the cost of travel has less influence on their choice of departure time. High-income groups are more inclined to individual modes of transportation and generally use private cars or taxis when traveling, so it is easier to choose a later departure time. The middle- and lower-income groups take into account the travel cost, which is an important factor considered when making travel decisions. Therefore, this

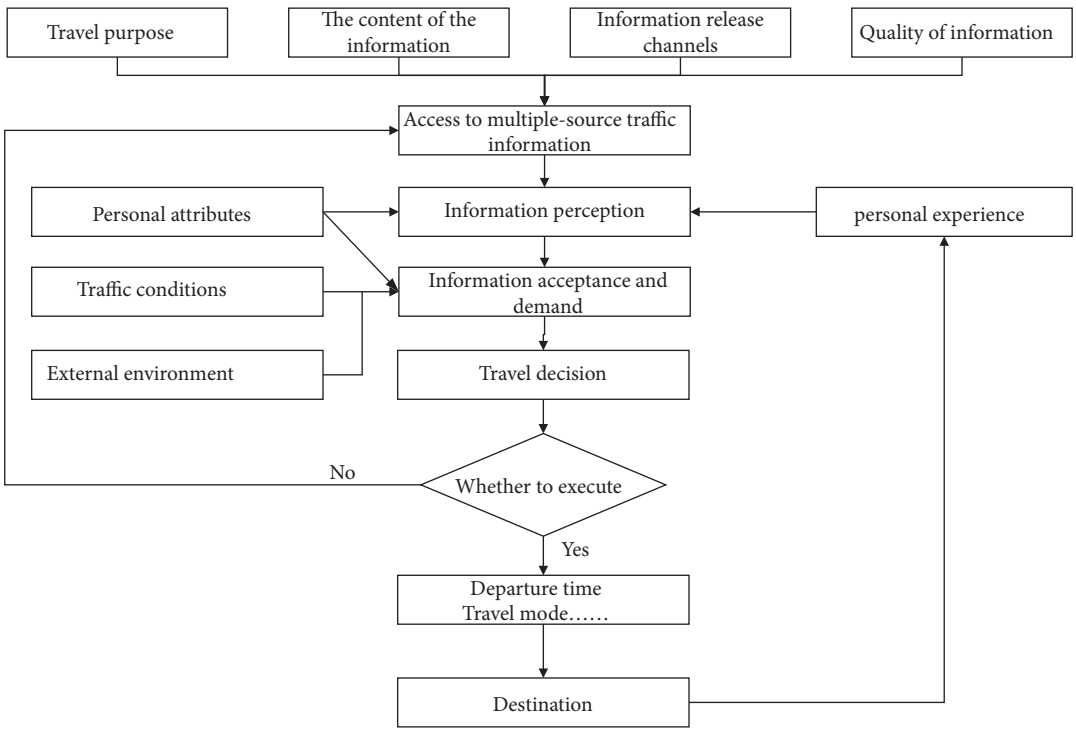


Figure 1: The basic function mode of the multisource traffic system.

group of tourists is highly dependent on public transportation travel modes and is more likely to choose an earlier departure time.

(3) Occupation. There is a corresponding relationship between the occupation of tourists and their income level. In general, the income level of tourists with stable occupations is relatively high. Therefore, tourists with stable careers are more likely to choose an earlier departure time.

2.2.2. Tour-Related Factors. Tour-related factors refer to factors that are closely related to the tourist activities, such as the frequency of trips to the city, the tour schedule, composition of people traveling together, and the desired tour experience.

(1) The Frequency of Trips to the City. This factor mainly reflects whether it is the first time for tourists to travel to the city. Generally, tourists who travel for the first time may arrange more excursions, so they will choose an earlier departure time. Conversely, tourists who have visited several times may be more likely to pursue detailed tours of some specific attractions, and it is easier to choose a later departure time.

(2) Tour Schedule. This factor directly reflects the number of days tourists will spend in this city. For example, tourists who have only one day to visit are more urgent than those who can stay for multiple days, and they may choose an earlier departure time.

(3) Composition of People Traveling Together. If the people traveling together include the elderly and children, for the elderly and children to have a better travel experience, such tourists often regard the comfort and safety of the travel process as the primary factors in travel decision-making, so they may choose early departure time. On the contrary, if there is no participation of the elderly or children, the tourists may travel more freely and may choose an earlier departure time.

2.2.3. Multisource Traffic Information-Related Factors. (1) The Ways to Obtain Travel Information. In recent years, with the rapid development of technology such as big data, information technology, and smart terminals, tourists have more and more ways to obtain traffic information, such as smartphones, navigation equipment, and broadcasting. However, tourists will rely on their habitual information acquisition ways due to differences in the effectiveness of the information update on different routes. Therefore, this will also lead to differences in the choice of departure time for tourists.

(2) The Types of Information. The information obtained by tourists is often composite rather than single, so it is necessary to know which information is in line with the actual needs of tourists and which information has a significant impact on the travel behavior of tourists. The information that tourists can obtain through the information system can be divided into three categories: traffic information, scenic area information, and external environment information. Traffic information includes real-time road conditions, public transport information, and parking information. Scenic area information mainly reflects the real-time status of the scenic area, including the number of tourists in the scenic spot, the number of queued tourists, and the ticket purchase situation. In addition, external

information such as weather conditions is also very important for tourists' travel decisions.

2.3. Survey Design and Implementation. To minimize the possible errors in the questionnaire survey, this article considers the use of Revealed Preference (RP) investigation combined with Stated Preference (SP) investigation. In terms of survey content, the RP survey method mainly focuses on the family property, personal property, and some travel characteristics surveys of tourists, and the SP survey mainly focuses on the survey of the first choice of scenarios under hypothetical scenarios.

The questionnaire was designed to reveal three catalogs of information.

2.3.1. Basic Tourist Personal Attributes. The basic information of tourists includes gender, age, occupation, educational background, and family car ownership.

2.3.2. Tour-Related Information. It includes the tour schedule, the purpose of this tour, the composition of people traveling together, travel mode during the tour, and departure time.

2.3.3. The Use of Multisource Traffic Information. It includes the ways to obtain travel information, the types of information, and the degree of influence of travel information on tourists' travel time choices.

The survey was designed and conducted to the tourists over three days (1 to 3 May 2019), in Yangzhou city, a famous tourist city of China. We collect data on tourists visiting Slender West Lake through face-to-face interviews. They were requested to report their information by answering a questionnaire. The dataset contains detailed information of 1500 tourists' fundamental socioeconomics and demographical information, tour information, and transport information. After cleaning the dataset and removing invalid records, 1463 valid tourist questionnaires were selected for this study.

2.4. Data Collation and Analysis. By analyzing the survey data of tourists' travel time, as shown in Figure 2, it was found that the peak of tourists reaching the attractions appears between 9 am and 10 am. Considering the time spent by tourists on their way to the scenic spot, it can be inferred that the departure time of tourists may be between 8 am and 9 am. The morning peak in Yangzhou appears to be from 7 am to 9 am, and the evening peak appears to be from 5 pm to 7 pm. The general business hours of the attractions are from 8 am to 4 pm. Taking the above conditions into consideration, we divide the departure time of tourists into three categories: before 8 am, between 8 am and 9 : 30 am, and between 9 am and 4 pm. These three options constitute a selection set of tourists' departure times.

The data show that the main channels for tourists to obtain traffic information are smartphones, navigation equipment, and broadcasting, which account for more than 10% of the total sample. As shown in Figure 3, the use of smartphones to obtain traffic information accounted for the highest proportion of 48%. Only 8% of tourists obtained travel information through roadside smart devices, indicating that tourists are relatively not reliant on it due to their unfamiliarity with the city. With the improvement of information technology, the ways for urban tourists to obtain travel information will become more and more diversified.

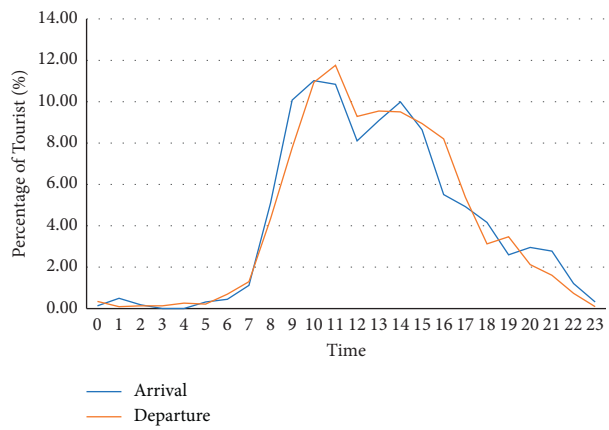


Figure 2: Tourists' travel time statistics.

As shown in Figure 4, the information tourists most want to obtain includes weather information, real-time traffic information, public transportation-related information, parking information, and real-time information on the number of tourists in the target scenic spot. Public transportation information mainly includes rail or conventional bus arrival information and transfer and connection methods. Parking information mainly includes the remaining status of parking spaces and charging standards.

This article uses 1–4 to represent the degree of influence of various types of information on the departure time of tourists; 1 represents the least impact, and 4 represents the largest. As shown in Figure 5, all kinds of information will have an impact on the departure time of tourists, of which real-time traffic information and public transportation information have the greatest impact.

3. Model Specification

3.1. Model Selection. To predict and analyze the rules of tourists' choices of departure time, it is necessary to employ mathematical models and methods to simulate tourists' travel behavior based on collected survey data. Discrete choice models are econometric models that describe the behavior of decision makers when making choices between discrete alternatives [28]. Various types of discrete choice models have been developed and applied to a wide variety of choice situations in many diverse fields, especially in transport [12, 29, 30].

Multiple selection problems can be divided into three categories: general multiple selection problems, ranking

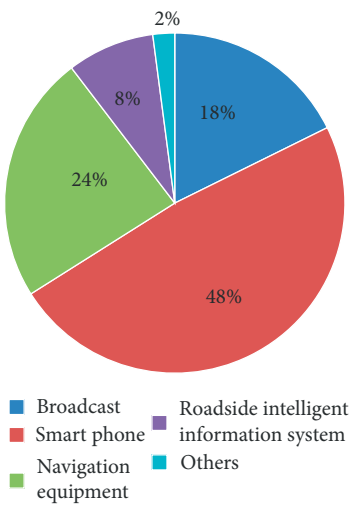


Figure 3: Distribution of the ways tourists obtain information.

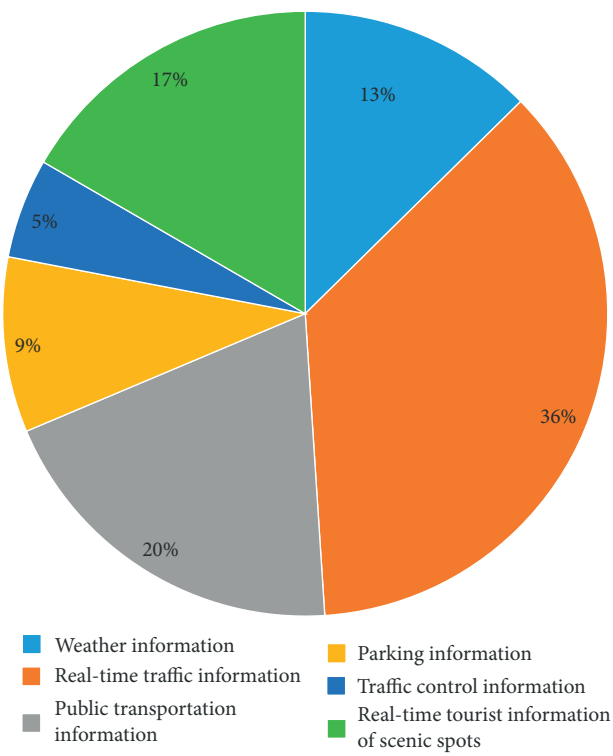


Figure 4: Distribution of the types of information tourists want to obtain.

selection problems, and nested selection problems. This research mainly studies tourists' departure time choice behavior during the tour, and it should be modeled by the general multiple discrete choice model. The Multinomial Logit (ML) formulation is employed in this research to describe the tourists' departure time choice behavior during the tour.

$$P_{in} = \frac{e^{V_{1n}}}{\sum_{j \in A_n} e^{V_{jn}}}. \quad (1)$$

In (1), P_{in} is the probability of the tourist n choosing the choice i ($i = 1, \cdots, n$); V_{in} is the utility function of the tourist n choosing the choice i ($i = 1, \cdots, n$); A_n is the choice collection of tourists n.

Further, the utility function of the nth tourist choosing i is

$$V_{in} = f(\theta, X_{in}) = \sum_{k=1}^{K} \theta_k X_{ink}. \quad (2)$$

In (2), X_{in} is the influencing variable of the nth tourist choosing the ith choice; X_{ink} is the kth influencing variable of the nth tourist choosing the ith choice. θ_k is the corresponding parameter of the kth influence variable.

3.2. Tourists' Departure Time Choice Model

3.2.1. Variable Correlation Analysis. To further explore the selection behavior of tourists' departure times, this paper selects factors from three perspectives: the personal attributes of tourists, tour-related information, and the acquisition of multiple traffic information based on the data obtained from the questionnaire survey.

Firstly, it is necessary to test the correlation between independent variables and dependent variables and eliminate independent variables with small correlations. The x2 independence test is used to analyze the correlation between the independent variable and the dependent variable. In the results of the x2 independence test, for the independent variables such as age and income level, there are more than 20% of the cells having an expected frequency of less than 5, which makes the test results unreliable. Then, to ensure the reliability of the test results, the Fisher exact test in SPSS is used. The test results are shown in Table 1.

The x^2 independence test and Fisher's exact test both take the 95% confidence level; that is, when the significance level P value is less than 0.05, the null hypothesis is rejected and the independent variable is considered to be correlated with the dependent variable. From the results, it can be obtained that the tourist's tour schedule, whether to drive for self-guided tours, and the real-time information acquisition are related. It can be considered that the above variables affect the choice of the tourist's departure time.

3.2.2. Tourists' Departure Time Prediction Model. (1) Variable Conversion and Screening. To use the most concise variables to build the model, it is necessary to screen the independent variables. In the disordered multicategory logistic model, four independent variable screening methods are provided, which are forward entry, backward removal, forward stepping, and backward stepping. The forward stepping method is used in this study to select independent variables.

Independent variables and dependent variables need to be transformed when they are included in the analysis model. The independent variables and dependent variables are transformed into numerical variables, and then the model parameters are calibrated, as shown in Tables 2 and 3.

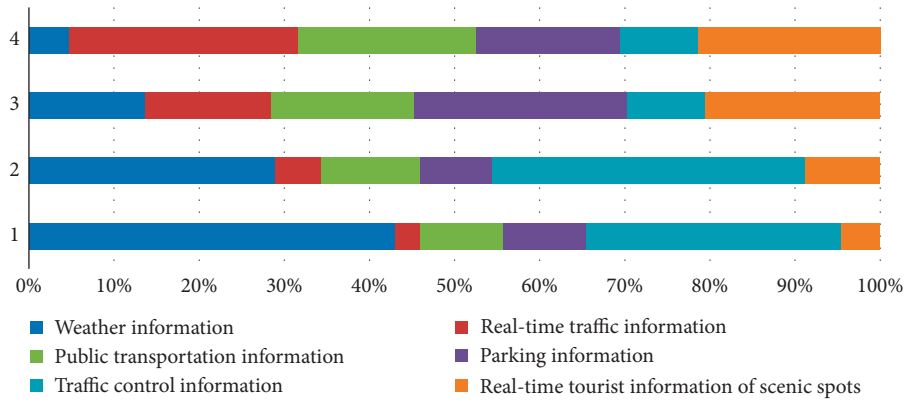


Figure 5: Influence degree characteristics.

Table 1: Correlation analysis.

Independent variable	Dependent variable	P value of Pearson chi-square (bidirectional)	Fisher's exact test P value (bidirectional)	Cells with an expected number of less than 5 (%)
Gender	DT	0.629	0.655	0.00
Age	DT	0.142	0.183	38.10
Occupation	DT	0.85	0.847	0.00
Income level	DT	0.862	0.908	33.30
Family car ownership	DT	0.917	0.890	0.00
Frequency of trips to Yangzhou	DT	0.301	0.303	0.00
Purpose of travel	DT	0.218	0.222	0.00
Tour schedule	DT	0.013	0.014	0.00
Self-driving tour or group tour	DT	0.001	0.001	0.00
Composition of people traveling together	DT	0.282	0.270	16.70
The number of attractions planned to visit	DT	0.242	0.246	0.00
Real-time information acquisition	DT	0.007	0.006	0.00

*DT: departure time.

Table 2: Conversion value of independent variables.

Independent variables	Category	Conversion value
Gender (x_1)	Male	1
	Female	2
Age (x_2)	<10	1
	10–19	2
	20–29	3
	30–39	4
	40–49	5
	50–59	6
	≥60	7
Occupation (x_3)	Have a regular occupation	1
	No fixed occupation	2
Income level (x_4)	<10 thousand yuan	1
	10 thousand yuan—20 thousand yuan	2
	20 thousand yuan—50 thousand yuan	3
	>50 thousand yuan	4
Family car ownership (x_5)	Yes	1
	No	2

Table 2: Continued.

Independent variables	Category	Conversion value
Frequency of trips to Yangzhou (x_6)	First time	1
	Second time	2
	> Second time	3
Purpose of travel (x_7)	Leisure travel	1
	Nonleisure travel	2
Tour schedule (x_8)	One day	1
	Multidays	2
Self-driving tour or group tour (x_9)	Self-driving	1
	Group tour	2
Composition of people traveling together (x_{10})	Have elderly or children	1
	No elderly or children	2
The number of attractions planned to visit (x_{11})	≤3	1
	>3	2
Real-time information acquisition (x_{12})	Yes	1
	No	2

(2) Model Construction and Analysis. The model is constructed based on the processed sample data, and the model fitting information is shown in Table 4.

In Table 4, the −2 times log-likelihood value is a statistic used to measure whether the logistic model fits the classified data. For the intercept-only −2 log-likelihood value and the final model −2 times, the likelihood ratio test of the difference of the log-likelihood value of 76.466 results in a P value of $0.006 < 0.05$. The independent variables included in the model have significant explanatory power over the dependent variable, indicating that the model is meaningful. The likelihood ratio test is performed on the difference of 76.466 between the −2 times log-likelihood value of the intercept only and the −2 times log-likelihood value of the final model, and the P value is $0.006 < 0.05$. This shows that the independent variables included in the model have significant explanatory power for the dependent variables. The model is valid and meaningful.

The parameter estimation results of the tourists departure time selection model are shown in Table 5.

According to statistical principles, under the condition of 95% confidence level, if the absolute value of the t-test (Wald value) is greater than 1.96, it is determined that this variable has a significant impact on the choice of travel time of tourists. The parameter estimated value and Wald value of each variable are shown in Figures 6 and 7.

According to the parameter estimation results calibrated by the above model, the tourists' departure time prediction model can be obtained as follows:

$$\ln \frac{P_1}{P_3} = 1.239 + 0.284X_2 + 0.606X_3 - 0.382X_6 - 0.737X_8 - 1.444X_9 + 1.029X_{10} - 0.823X_{12}, \tag{3}$$

$$\ln \frac{P_2}{P_3} = 0.081 - 0.301X_6 - 1.347X_9 + 1.140X_{10} - 0.554X_{12}. \tag{4}$$

In (3) and (4), P_1 is the probability that the tourist chooses to depart before 8 : 00; P_2 is the probability that the tourist chooses to depart between 8 am and 9.30 am; P_3 is the probability that the tourist chooses to depart between 9.30 am and 4 pm; X_i is the ith independent variable.

According to the obtained tourist departure time prediction model, the probability that tourists choose to depart at each period is

$$\frac{P_1}{P_3} = \mathrm{EXP}\left(1.239 + 0.284X_2 + 0.606X_3 - 0.382X_6 - 0.737X_8 - 1.444X_9 + 1.029X_{10} - 0.823X_{12}\right),$$

$$\frac{P_2}{P_3} = \mathrm{EXP}\left(0.081 - 0.301X_6 - 1.347X_9 + 1.140X_{10} - 0.554X_{12}\right),$$

$$P_1 + P_2 + P_3 = 1,$$

Table 3: Conversion value of dependent variables.

Dependent variable	Category	Conversion value
Departure time	Before 8 am	1
	Between 8 am and 9.30 am	2
	Between 9.30 am and 4 pm	3

Table 4: Model fitting information.

Model	Model fitting conditions		Likelihood ratio test	
	−2 log-likelihood value	Bangla	Degree of freedom	Statistical significance
Intercept	495.01	—	—	—
	418.544	76.466	48	0.006

$$
\begin{aligned}
P_3 &= \frac{1}{M+N+1},\\
P_1 &= P_3 \cdot M,\\
P_2 &= P_3 \cdot N,\\
M &= \mathrm{EXP}\left(1.239 + 0.284X_2 + 0.606X_3 - 0.382X_6 - 0.737X_8 - 1.444X_9 + 1.029X_{10} - 0.823X_{12}\right),\\
N &= \mathrm{EXP}\left(0.081 - 0.301X_6 - 1.347X_9 + 1.140X_{10} - 0.554X_{12}\right).
\end{aligned} \tag{5}
$$

P_1 is the probability that the tourist chooses to depart before 8 : 00; P_2 is the probability that the tourist chooses to depart between 8 am and 9.30 am; P_3 is the probability that the tourist chooses to depart between 9.30 am and 4 pm; X_i is the ith independent variable.

3.2.3. Model Verification. To verify whether the established tourists' departure time prediction model can accurately reflect the actual departure time selection behavior of tourists, this study uses survey data as the basis to compare the travel time distribution probabilities predicted by the model with the actual probability distribution of the survey. The results are shown in Table 6.

The comparative analysis of the model prediction results and the survey results shows that the model established in this study can reflect the choice behavior of tourists during the travel period, and the model has good accuracy.

3.3. Parameter Estimation Analysis. By analyzing the model obtained, it can be found that the factors that affect tourists' departure time choice are not only the personal attributes of tourists but also factors related to tourism aspects and real-time information. The combined effect of these factors together determines the departure time decision of tourists.

From Table 7, factors affecting tourists' choice of departure before or after the morning peak include age, gender, frequency of trips to Yangzhou, tour schedule, self-driving tour or group tour, the composition of people traveling together, and real-time information acquisition. The estimated values of the parameters of age and occupation in the personal attributes of tourists are positive, which indicates that young female tourists are more inclined to choose the departure time before the morning peak. Among tour-related factors, the parameter estimated values of trips to Yangzhou, tour schedule, self-driving tour, or group tour are negative, indicating that tourists who travel to Yangzhou by self-driving for the first time for a day trip are more likely to depart before the morning peak. The parameter estimated value of the factor of people traveling together is positive, indicating that tourists are more likely to choose to start before the morning peak without the company of the elderly or children. The acquisition of real-time information also has a significant impact on the choice of departure time for tourists. The estimated value of this parameter is negative, which indicates that tourists are more likely to travel after the morning peak. This may be due to the decision made by tourists to avoid the congested traffic environment based on their judgment after obtaining real-time traffic information.

In addition, some common variables appear in both formulas (3) and (4), including frequency of trips to Yangzhou, self-driving tour or group tour, composition of people traveling together, and real-time information acquisition. But except that the parameter estimates of composition of people traveling together are all positive, for the parameter estimates of other variables in the two formulas, one is positive and the other one is negative. This result shows that even the same variable has different effects on tourists' travel behavior in different time periods.

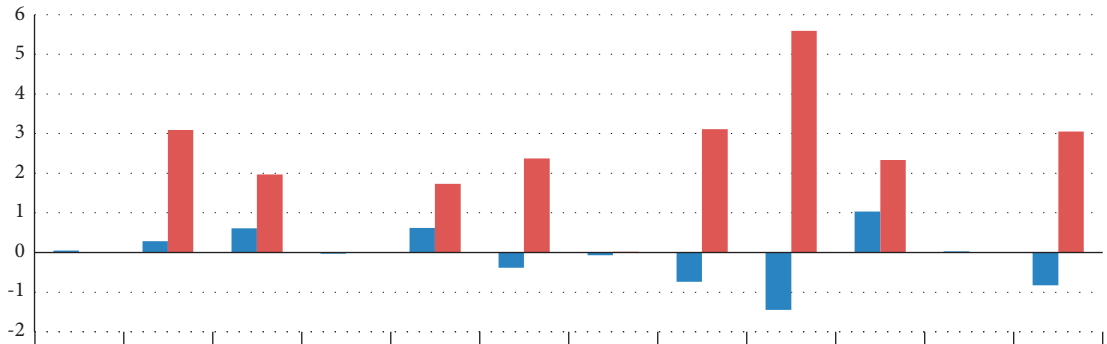
Further analysis of the data in the table, we can find that it is easier for tourists to travel to a city for the first time without the accompaniment of the elderly or children to choose an earlier departure time. At the same time,

Table 5: Parameter estimation results.

	Departure time[a]	B	Standard error of mean	Wald	Degree of freedom	Statistical significance	Exp (B)
1.0	Intercept	1.239	2.252	0.302	1	0.582	
	Gender	0.044	0.411	0.012	1	0.914	1.045
	Age	0.284	0.161	3.090*	1	0.079	1.328
	Occupation	0.606	0.456	1.969*	1	0.183	1.834
	Income level	−0.034	0.307	0.012	1	0.913	0.967
	Family car ownership	0.621	0.472	1.732	1	0.188	1.861
	Frequency of trips to Yangzhou	−0.382	0.248	2.376*	1	0.123	0.682
	Purpose of travel	−0.073	0.494	0.022	1	0.883	0.93
	Tour schedule	−0.737	0.418	3.109*	1	0.078	0.478
	Self-driving tour or group tour	−1.444	0.611	5.589*	1	0.018	0.236
	Composition of people traveling together	1.029	0.674	2.336*	1	0.126	2.799
	The number of attractions planned to visit	0.027	0.421	0.004	1	0.948	1.028
	Real-time information acquisition	−0.823	0.471	3.049*	1	0.081	0.439
2.0	Intercept	0.081	1.822	0.002	1	0.965	
	Gender	0.2	0.32	0.39	1	0.532	1.221
	Age	0.034	0.139	0.059	1	0.807	1.034
	Occupation	0.243	0.356	0.465	1	0.495	1.274
	Income level	0.039	0.247	0.024	1	0.876	1.039
	Family car ownership	0.314	0.374	0.706	1	0.401	1.369
	Frequency of trips to Yangzhou	−0.301	0.195	2.383*	1	0.123	0.74
	Purpose of travel	−0.343	0.39	0.773	1	0.379	0.709
	Tour schedule	0.135	0.341	0.158	1	0.691	1.145
	Self-driving tour or group tour	−1.347	0.453	8.838*	1	0.003	0.26
	Composition of people traveling together	1.14	0.56	4.140*	1	0.042	3.126
	The number of attractions planned to visit	0.308	0.333	0.851	1	0.356	1.36
	Real-time information acquisition	−0.554	0.342	2.625*	1	0.105	0.575

*B is the parameter estimated value; Wald is a chi-square value; Exp (B) is the odds ratio; the reference category is 3.0; *the variable is significant.

Figure 6: Value distribution of B and Wald (1).



	Gender	Age	Occupation	Income Level	Family Car Ownership	Frequency of trips to Yangzhou	Purpose of Travel	Tour Schedule	Self-driving Tour or Group Tour	Composition of People Traveling Together	The Number of Attractions planned to visit	Real-time information acquisition
B	0.044	0.284	0.606	-0.034	0.621	-0.382	-0.073	0.737	-1.444	1.029	0.027	-0823
Wald	0.012	3.09	1.969	0.012	1.732	2.376	0.022	3.109	5.589	2.336	0.004	3.049

Figure 7: Value distribution of B and Wald (2).

Table 6: Model validation.

Departure time	Model prediction probability (%)	Survey probability (%)	Relative difference (%)
Before 8 am	23.47	21.63	8.51
Between 8 am and 9.30 am	42.83	45.72	−6.32
Between 9.30 am and 4 pm	33.70	32.65	3.22

Table 7: Model parameter estimation results.

Variable	1		2	
	B	Wald	B	Wald
Age	0.284	3.09	—	—
Occupation	0.606	1.969	—	—
Frequency of trips to Yangzhou	−0.382	2.376	−0.301	2.383
Tour schedule	−0.737	3.109	—	—
Self-driving tour or group tour	−1.444	5.589	−1.347	8.838
Composition of people traveling together	1.029	2.336	1.14	4.14
Real-time information acquisition	−0.823	3.049	−0.554	2.625

*B is the parameter estimated value; Wald is a chi-square value.

according to our survey results, tourists are more inclined to choose convenient devices such as smartphones and navigation equipment to receive traffic-related information such as real-time traffic information and public traffic information. Therefore, the research results of this study have important guiding significance for the city management department to explore the travel behavior of tourists and examine how to release information that guides tourists' travel behavior through which way.

4. Conclusion

With the rapid development of information technology, the traffic information provided by the intelligent information platform tends to be diversified from singleness. The multisource traffic information can help tourists make reasonable travel decisions during travel, which can achieve the purpose of alleviating traffic congestion and improving the quality of travel.

When increasing cities pay attention to taking management measures to eliminate traffic pressures caused by tourism, it is important to get deep learning about tourists' travel behavior from the angle of tourists instead of traditional urban residents and guide tourists' travel behavior through information sharing technology. The choice of departure time is an important aspect that affects the characteristics of traffic demand, as well as the factor that reflects the difference in travel behavior between tourists and urban residents.

In this study, we obtained the data of tourists traveling in Yangzhou through a survey and based on which a probability model of tourists' departure was established. Different from the previous study on departure time based on urban residents' data, the model established in this article can reveal tourists' departure time selection behavior more specifically from the perspective of tourism and information synchronization sharing. The results obtained in this research can provide a reference for obtaining more accurate tourist traffic demand and at the same time provide technical support for urban management departments to formulate more reasonable tourism traffic management measures. City management departments can guide tourists' travel behavior by publishing real-time transportation information and at the same time formulate more targeted and efficient tourism traffic management policies to alleviate the pressure of tourism traffic on urban traffic while creating a more comfortable tourism environment for tourists and promoting better tourism development.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the Research Planning Fund for Humanities and Social Sciences of the Ministry of Education (19YJAZH011), Support for the Open Project of Key Laboratory of Intelligent Traffication Technology and Traffication Industry (F262019016), and the Natural Science Foundation of Jiangsu Province BK20210833.

References

[1] Y. Y. Yu Cheng and D. X. Zhen Liu, "Review and prospect of China's tourism development," *East China Economic Management*, vol. 34, no. 3, pp. 1–9, 2020.

[2] D. A. Gillmor, "Evolving air-charter tourism patterns: change in outbound traffic from the Republic of Ireland," *Tourism Management*, vol. 17, no. 1, pp. 9–16, 1996.

[3] J. Lungren, "The development of the tourist travel systems: a metropolitan economic hegemony par excellence?" *Tourist Review*, vol. 23, no. 1, pp. 2–14, 1973.

[4] O. Saenz-de-Miera and J. Rosselló, "The responsibility of tourism in traffic congestion and hyper-congestion: a case study from Mallorca, Spain," *Tourism Management*, vol. 33, no. 2, pp. 466–479, 2012.

[5] X. Zhao and G. Wen, "Optimizing method for evacuation route of dual-object traffic jam during peak tourist period," *Journal of Yanshan University*, vol. 42, no. 3, pp. 278–282, 2018.

[6] L. Wang, Y. Zhang, Q. Ji, and G. Chai, "Research on the traffic status and countermeasures in wutai mountain scenic area," *Journal of Shanxi Normal University (Philosophy and Social Sciences edition)*, vol. 31, no. 2, pp. 121–124, 2017.

[7] M.-W. Jo, H. Kim, and H.-J. Shin, "Understanding traffic congestion to improve tourist satisfaction in local tourism," *International Journal of Tourism and Hospitality Research*, vol. 30, no. 4, pp. 85–95, 2016.

[8] G. Zhong, T. Yin, J. Zhang, S. He, and B. Ran, "Characteristics analysis for travel behavior of transportation hub passengers using mobile phone data," *Transportation*, vol. 46, no. 5, pp. 1713–1736, 2018.

[9] S. Bamberg, I. Ajzen, and P. Schmidt, "Choice of travel mode in the theory of planned behavior: the roles of past behavior, habit, and reasoned action," *Basic and Applied Social Psychology*, vol. 25, no. 3, pp. 175–187, 2003.

[10] M. J. Sirgy and C. Su, "Destination image, self-congruity, and travel behavior: toward an integrative model," *Journal of Travel Research*, vol. 38, no. 4, pp. 340–352, 2000.

[11] W. R. Zhao and Y. Zhao, "Research on passenger travel model choice in shenyang-dalian transport corridor," *Applied Mechanics and Materials*, vol. 744–746, pp. 2090–2093, 2015.

[12] C. R. Bhat, "Analysis of travel mode and departure time choice for urban shopping trips," *Transportation Research Part B: Methodological*, vol. 32, no. 6, pp. 361–371, 1998.

[13] J. D. Lemp, K. M. Kockelman, and P. Damien, "The continuous cross-nested logit model: formulation and application for departure time choice," *Transportation Research Part B: Methodological*, vol. 44, no. 5, pp. 646–661, 2010.

[14] R. Shabanpour, N. Golshani, J. Auld, and A. Mohammadian, "Dynamics of activity time-of-day choice," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2665, no. 1, pp. 51–59, 2017.

[15] S. Gadda, K. M. Kockelman, and P. Damien, "Continuous departure time models," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2132, no. 1, pp. 13–24, 2009.

[16] K. M. N. Habib and J. A. Carrasco, "Investigating the role of social networks in start time and duration of activities: trivariate simultaneous econometric model," *Transportation Research Record*, vol. 2230, no. 1, pp. 1–8, 2018.

[17] M. Amirgholy, M. Shahabi, and H. O. Gao, "Optimal design of sustainable transit systems in congested urban networks: a macroscopic approach," *Transportation Research Part E: Logistics and Transportation Review*, vol. 103, pp. 261–285, 2017.

[18] R. Paleti, P. S. Vovsha, D. Givon, and Y. Birotker, "Joint modeling of trip mode and departure time choices using revealed and stated preference data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2429, no. 1, pp. 67–78, 2014.

[19] R. Shabanpour, N. Golshani, S. Derrible, A. Mohammadian, and M. Miralinaghi, "Joint discrete-continuous model of travel mode and departure time choices," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2669, no. 1, pp. 41–51, 2017.

[20] Y. Asakura and E. Hato, "Tracking survey for individual travel behaviour using mobile communication instruments," *Transportation Research Part C: Emerging Technologies*, vol. 12, no. 3–4, pp. 273–291, 2004.

[21] M. Hadi, X. Yan, I. Shahadat, K. Samaneh, and S. Purser II, *Lehman Center of Transportation Research, Florida International University; Southwest Research Institute (SwRI). USA*, Florida Department of Transportation, Tallahassee, FL, USA, 2017.

[22] G. Song and H. He, "Real-time traveler information for optimal adaptive routing in stochastic time-dependent networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 196–213, 2012.

[23] M. Gendreau, G. Ghiani, and E. Guerriero, "Time-dependent routing problems: a review," *Computers & Operations Research*, vol. 64, pp. 189–197, 2015.

[24] J. Haworth, J. Shawe-Taylor, and T. Cheng, "Local online kernel ridge regression for forecasting of urban travel times," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 151–178, 2014.

[25] L. U. Song, J. I. Hui, and C. Yunfeng, "A study on the spatial travel behavior of self-driving tourists into Huangshan City," *Geographical Research*, vol. 32, no. 1, pp. 179–190, 2013.

[26] H. Zhu, H. Guan, Y. Han, and W. Li, "A study of tourists' holiday rush-hour avoidance travel behavior considering psychographic segmentation," *Sustainability*, vol. 11, 2019.

[27] C. Mok and T. Lam, "Travel-related behavior of Japanese leisure tourists: a review and discussion," *Journal of Travel & Tourism Marketing*, vol. 9, no. 1-2, pp. 171–184, 2000.

[28] M. Bierlaire, "A theoretical analysis of the cross-nested logit model," *Annals of Operations Research*, vol. 144, no. 1, pp. 287–300, 2006.

[29] G. D. Jong, A. Daly, M. Pieters, C. Vellay, M. Bradley, and F. Hofman, "A model for time of day and mode choice using error components logit," *Transportation Research Part E: Logistics and Transportation Review*, vol. 39, no. 3, pp. 245–268, 2003.

[30] D.-S. Yun, "Analysis of travel mode and home departure time choice for university students' commuting trips," *Journal of Korea Planning Association*, vol. 141, no. 4, 2001.

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 2739129, 17 pages
https://doi.org/10.1155/2021/2739129

Research Article

An Extended Car-Following Model in Connected and Autonomous Vehicle Environment: Perspective from the Cooperation between Drivers

Shenzhen Ding,[1] **Xumei Chen**,[1,2] **Zexin Fu**,[1] **and Fei Peng**[1]

[1]*Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing 100044, China*
[2]*Xuchang University, Henan Province, Xuchang 461000, China*

Correspondence should be addressed to Xumei Chen; xmchen@bjtu.edu.cn

Received 22 July 2021; Accepted 17 September 2021; Published 13 October 2021

Academic Editor: Francesco Galante

The development of connected and autonomous vehicle (CAV) technology has received increasing attention in recent years. Although car-following behavior in mixed traffic with CAVs and human-driven vehicles (HDVs) is a core component of microscopic traffic simulation, intelligent traffic systems, etc., the current study of car-following behavior in mixed traffic has some limitations. Furthermore, actual data do not support its applicability to the Chinese traffic environment. To address this gap, this paper designs and organizes a car-following experiment in mixed traffic in Beijing, extracts the trajectory data of CAVs and HDVs based on video recognition, and reconstructs the extracted trajectory data using the Lagrangian theory and Kalman filter theory to ensure the accuracy of the data. Based on this data set, this paper develops an extended car-following model. The model considers the cooperation between drivers by reformulating the prospect theory (PT). The root mean square percentage error (RMSPE) is selected to calibrate and validate the parameters of the proposed model, and the results show that there is significant heterogeneity between CAVs and HDVs in mixed traffic, and the proposed model captures this heterogeneity well. The model presented in this paper provides theoretical support for microscopic traffic simulation in mixed traffic.

1. Introduction

Traffic congestion has been a hot issue for decades and will remain so in the future. Generally, traffic congestion causes oscillations and stop-and-go waves, which are a nuisance to drivers and result in more accidents and lower traffic efficiency. Many fascinating traffic flow phenomena, such as hysteresis, traffic breakdown, and capacity drop, result from the complexity and randomness of driving behavior [1]. To better alleviate traffic congestion and control traffic flow, it is necessary to understand the mechanism of these phenomena. However, our understanding of traffic flow is still not comprehensive and is even somewhat ambiguous [2].

In addition, understanding and exploring these human behaviors are just as important in the field of CAVs. At present, despite the enormous investments in CAV technologies that the automotive industry has made to compete in this potential market, it will still require decades for CAVs to be fully deployed [3]. It is a fact that mixed traffic, including CAVs and HDVs, will appear and last for a long time. In the second and third level automatic driving stages, although the driver does not need to perform driving operations, they need to monitor the road environment and the autonomous driving system and respond to takeover requests from the system in time [4, 5]. How will these two types of vehicles, CAVs and HDVs, operate together in mixed traffic? How can we better understand the driving behavior in mixed traffic and thus improve traffic conditions? A model of the collaborative interaction between two types of vehicles is one of the many essential tools.

As one of the microscopic traffic flow models, the car-following model has been developed for many decades

because of its ability to control the longitudinal part of the traffic flow. During the following process, novice drivers feel very nervous due to a lack of confidence or inaccurate judgment of the relative speed and distance between the cars in front and behind, which causes drivers to have very high stress and thus affects their following behavior. Skilled drivers may feel bored during car-following, but their rich driving experience allows them to drive steadily behind the preceding vehicle [6]. For different types of drivers, driving behaviors will vary during the following process, and the study of these behavioral characteristics is essential for microscopic traffic flow. The first car-following model, the Pipes model, was proposed in 1953 [7]. Since then, numerous car-following models have been proposed, reflecting the evolution of the study of traffic flow from different perspectives. These models mainly include the Gazis–Herman–Rothery model [8], the intelligent driver model [9], the optimal velocity model [10], the Gipps model [11], and the Wiedemann model [12]. These models describe the driving behavior during the car-following from different perspectives, such as stimulus-response, safety-distance, optimal velocity, and psycho-physical. Among them, many methods and conclusions have been proposed to enhance traffic flow stability, improve safety, and alleviate traffic congestion [13]. However, the analysis of vehicle interactions cannot be separated from human factors, which are often neglected in previous studies. The introduction of human factors complicates the interaction between vehicles [14]. An accurate description of the human-like collaborative relationship between cars during the following process is conducive to understanding the operation rules of heterogeneous traffic flow and then to implement traffic management and control measures more effectively.

Recently, emerging automated vehicle (AV) technologies push the study of driver behavior to a new stage, especially car-following behavior. As the precursor of AV technology, the adaptive cruise control (ACC) has been very common. It has been more than 20 years since the advent of the first vehicle equipped with ACC technology [15]. Moreover, the ACC technology is likely to be the primary component for longitudinal control of higher-level AVs in the future [16]. Therefore, it is critical to understand the changes brought about by vehicles equipped with ACC systems compared with traditional HDVs. Although vehicles equipped with ACC systems have become common, only a few studies have been conducted on the driving behavior of ACC vehicles based on experimental field data. Several studies have been conducted in the literature on the stability of ACC systems [17–19]. Notably, in recent years, pioneering work has been done to understand the behavior of ACC, and more sophisticated control experiments have been conducted. Specifically, Gunter et al. have undertaken a set of car-following experiments to collect empirical data. They calibrated an optimal velocity car-following model in their research and analyzed the string stability for both calibrated models based on the observed data [20]. Moreover, the Joint Research Centre of the European Commission has conducted a set of ACC experiments, and some basic features of ACC have already been revealed [21]. However, ACC systems make car-following decisions that depend only on the preceding vehicle's information, weakening the connectivity of cars equipped with the ACC system. Compared with the ACC system, the cooperative adaptive cruise control (CACC) system makes car-following decisions with more information from either a single vehicle or multiple vehicles using connectivity technologies [22]. The CACC system allows vehicles to follow in a platoon by extra layers of communication and automation [23]. As one of the most popular CAV technology applications, scholars have paid a significant amount of attention to the CACC system [24–27]. However, the research on CACC systems has focused on what benefits CAVs may bring to traffic networks, with little consideration of the impact of CACC systems on vehicles not equipped with the CAV technology (HDVs). So, it is particularly important to comprehensively consider the cooperative relationship between CAVs and HDVs in mixed traffic to conduct interaction analysis and achieve the optimal system.

A common and perhaps the most critical gap identified in the above studies is the insufficient consideration of human factors during the following process. The cooperation between human drivers (or between a human driver and a CAV with a higher autopilot system) significantly affects the car-following operation. In addition, a few studies directly used existing car-following models to simulate the behavior of vehicles in mixed traffic without examining the applicability of the models, either for CAVs or for HDVs. Further, most previous studies have used the vehicle trajectory data from driving simulators and numerical simulations, and they employed trajectory data of HDVs to calibrate the model. Still, these data are not representative of real-world vehicle trajectory data in connected and autonomous vehicle environment. The car-following behavior of both CAVs and HDVs has changed in connected and autonomous vehicle environment. The traditional car-following model is no longer sufficient to describe their driving behavior. Therefore, how can we obtain accurate and credible vehicle trajectory data in connected and autonomous vehicle environment? How can we take human factors into account when modeling car-following behavior in mixed traffic? These questions are still to be explored.

Thus, the paper focuses on modeling car-following behavior in connected and autonomous vehicle environment by incorporating human factors, i.e., cooperation between drivers. The well-known PT describes the cooperation between drivers in this paper. Three variables—speed difference, acceleration, and distance headway—are utilized to model the car-following behavior of vehicles. Also, we designed and conducted a car-following experiment in connected and autonomous vehicle environment to obtain the vehicle trajectory data required for the calibration and validation of the model. In the experiment, three Tesla vehicles, equipped with Level 2 autonomous driving functions, completed the designed driving operations on the test road section. Using the Lagrangian theory and Kalman filter theory, we developed a vehicle trajectory reconstruction method to reconstruct the extracted data from the collected videos.

The remainder of the paper is structured as follows. The following section (Section 2) is about data collection, which mainly explains the experimental and data reconstruction method. Section 3 presents the analysis of cooperation for car-following behavior in mixed traffic. Section 4 describes the methodologies to model car-following behavior in connected and autonomous vehicle environment based on PT. The calibration and verification of the model are provided in Section 5. The last section concludes the paper.

2. Data Collection and Processing

High-precision vehicle trajectory data are necessary to support the research of micro-driving behavior in connected and autonomous vehicle environment. However, the vehicle trajectory data acquired based on in-vehicle devices or videos have non-negligible measurement errors that adversely affect the establishment and calibration of microscopic driving models. Thus, a data acquisition experiment was designed and implemented.

In this paper, the experiment was carried out using three Tesla vehicles. Some well-known experiments on mixed traffic were also done with the help of a few test vehicles. For example, the CACC experiments of PATH Laboratory at the University of California, Berkeley. In the study of modeling cooperative and autonomous adaptive cruise control dynamic responses, their experiments were carried out using four production Infiniti M56s (the vehicle equipped with a commercial ACC system) provided by Nissan [28]. Then, we extracted the collected videos to obtain the original vehicle trajectory data. To obtain richer data, a total of 90 minutes of video data were collected for trajectory extraction (greater than 45 minutes in NGSIM), with a granularity of 0.1 seconds. Finally, a two-step reconstruction method was presented by combining the Lagrange algorithm and Kalman filter theory to achieve interpolation replacement and noise reduction of the abnormal data in the data set.

2.1. Experimental Implementation of the Cooperative Adaptive Cruise Control. The video data used for vehicle trajectory extraction in this study were collected on December 27, 2019. Ninety minutes of video data, divided into six segments, were collected every 15 minutes. The collection site was located at Majiabao West Road, Fengtai District, Beijing, with a total length of about 230 m, and the collection method was 4K high-definition video. Figure 1 shows the survey location and camera position.

To obtain actual data for mixed traffic, three Tesla vehicles, which can achieve Level 2 autonomous driving, were driven back and forth on Majiabao West Road in Beijing for 1.5 hours. The videos are recorded with a high-resolution camera located on a high building near Majiabao West Road. The autonomous driving functions, including adaptive cruise control, lane keep assistance, and lane departure warning, are active when the drivers enter the test area. To simulate a connected environment, the drivers of the three Tesla vehicles interact via real-time video communication. For example, when a Tesla driver performs a lane-changing operation, the other two Tesla drivers in adjacent lanes are informed through real-time video communication and then cooperate with the driver to complete the operation. In addition, the three Tesla vehicles share speed information in real time. In this way, the experimental vehicles can simulate a CACC scenario. The vehicles could receive speed advice from other vehicles via real-time video communication and then decide whether to perform coordinated acceleration or deceleration. The three Teslas' roofs were marked in purple, green, and yellow to facilitate identifying the test cars in the video, as shown in Figure 2. During trajectory extraction, we combine two methods to improve the accuracy of secondary tracking of Tesla vehicles. On the one hand, we use support vector machines (SVMs) as a classifier to classify the Tesla vehicles and then detect the Tesla vehicle data. The realization of this process uses color filters to generate image features. On the other hand, convolutional neural networks (CNNs) are used to learn the features for the Tesla vehicles, and the accuracy of the Tesla vehicle tracking is improved by controlling the Euclidean distance between different classes of feature vectors.

2.2. Vehicle Trajectory Reconstruction. Professional collaborators have performed video-based extraction of vehicle trajectory data. Vehicle trajectory information extraction includes vital steps, such as coordinate system establishment and conversion, vehicle detection and positioning, vehicle classification, target vehicle tracking, and trajectory information extraction. The vehicle trajectory data obtained by video extraction include a vehicle identification number, recording time, vehicle type, lane number, horizontal coordinate X, vertical coordinate Y, speed, and acceleration.

The error of vehicle trajectory data extraction based on video is directly related to the sampling frequency of video frames [29]. This error can be divided into two categories. One is the system error based on different detection methods; the other is the random error, i.e., the noisy data generated during the detection process.

For the speed data calculated by the vehicle longitudinal position information, there are two kinds of abnormal values: negative values and extreme values. In addition, the analysis of vehicle acceleration data finds that there is a specific moment when the vehicle speed data are correct. However, the acceleration data are too large, beyond the limit of the physical properties of the vehicle and the bearing capacity of the human body, so this kind of abnormal value also needs to be corrected. The possible causes of system errors include light, occlusion, and other uncontrollable factors. In the process of vehicle trajectory extraction, the unsuccessful recognition of the vehicle position in the image and the abnormal disturbance of the vehicle detection edge can cause system errors. The random error is mainly generated because the original vehicle trajectory data obtained from video extraction will be perturbed up and down around the vehicle's actual position, resulting in obvious noise sequences in the data. This random error is unavoidable, and it will be transferred to the speed and acceleration data. There are many reasons for random errors, such as the external

(a) (b)

Figure 1: (a) Survey location and (b) camera position.

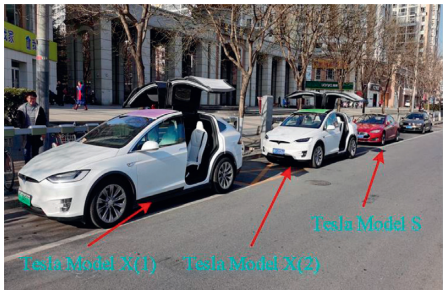


Figure 2: Tesla vehicles marked in different colors.

environment, the construction of the instrument itself, and the temporal granularity of the vehicle trajectory extraction. Random errors cannot be eliminated, but they can be filtered by signal filtering techniques, such as Kalman filter techniques [30].

We design a two-step reconstruction method of vehicle trajectory based on the error analysis, as shown in Figure 3. First, according to the physical performance of the vehicle and the human body's tolerance limit, combined with the recorded actual speed data of the experimental vehicles, the speed extremes, negative speed values, and abnormal values of acceleration are identified. The Lagrange interpolation method is used to re-estimate these outliers. Second, due to the random error in video trajectory extraction, the Kalman filter method is used for noise reduction.

Step 1. Vehicle trajectory data correction based on the Lagrangian interpolation method.

Interpolation is the process of solving for the values corresponding to the missing or incorrect discrete points when the series of discrete points and their corresponding values are known [31]. This study uses the Lagrange interpolation method to correct the abnormal values of speed and acceleration. The thresholds of detecting outliers are mainly determined according to the physical performance of the vehicle and the human body's tolerance limit, combined with the actual speed data recorded by the experimental vehicles. In this paper, we adopted detection thresholds of $5\,\mathrm{m/s^2}$ for acceleration and $-8\,\mathrm{m/s^2}$ for deceleration. These thresholds were chosen after large-scale testing [32]. In fact, lower absolute thresholds (close to the physical acceleration of the vehicle) will result in the detection and subsequent removal of a large number of points, and loss of information

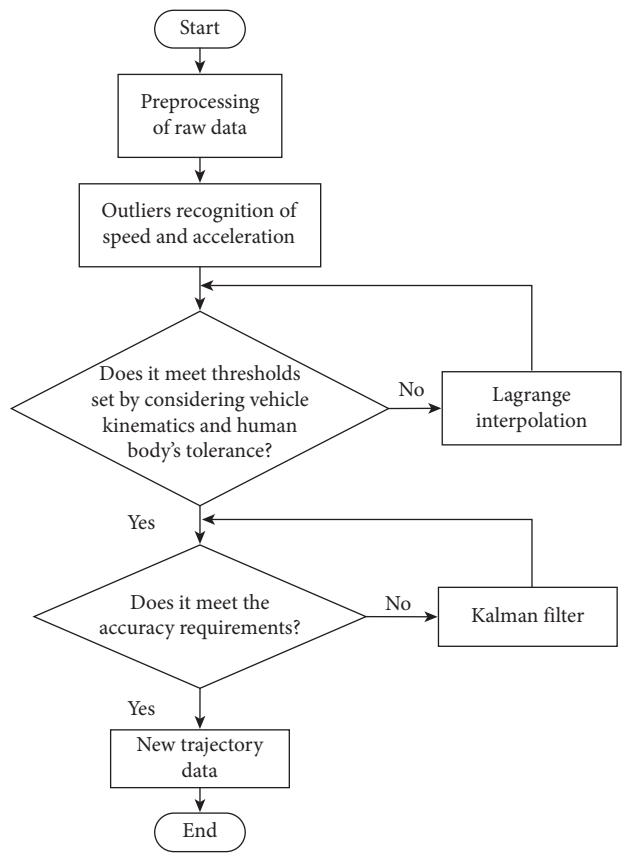


Figure 3: Flow chart of vehicle trajectory reconstruction.

about the trajectory trend. Taking the speed correction of the experimental vehicle No. 1 as an example, as shown in Figure 4, by identifying the outliers, the negative speed values, speed extremes, and the vehicle position are effectively corrected.

In this step, the threshold boundary set for identifying abnormal values of speed and acceleration should not be too strict to reduce the impact of the human correction on the original structural characteristics of the vehicle trajectory data.

Step 2. Noise reduction based on Kalman filter method.

After Lagrangian interpolation, the Kalman filter method is used to reduce noise for speed and acceleration.

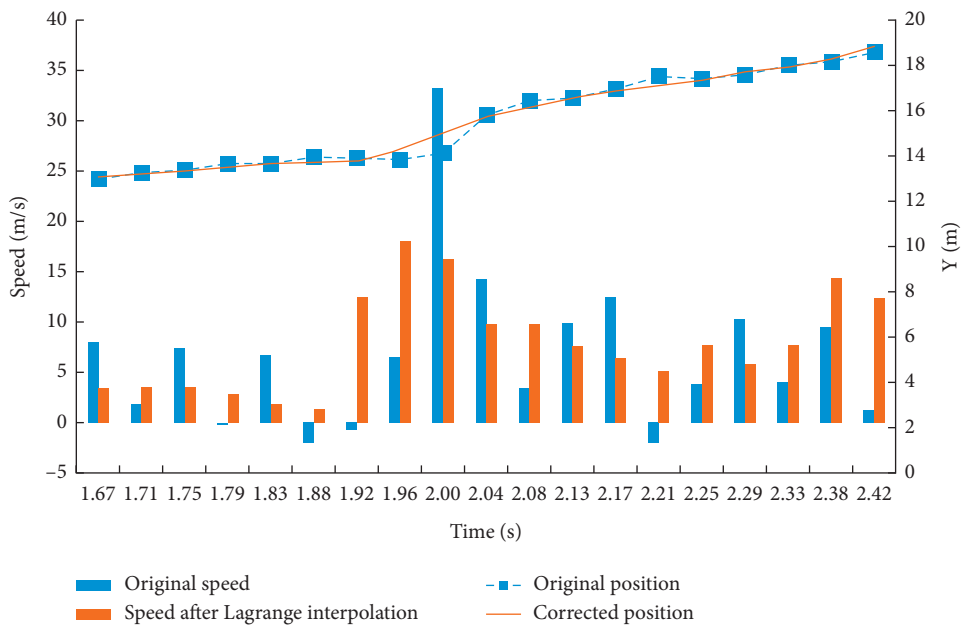


Figure 4: Schematic diagram of data correction based on Lagrangian interpolation.

The Kalman filtering is affected by the Kalman gain, which depends on the ratio between the system noise and the observed noise. The two noises are expressed using their respective covariances. The system noise covariance is expressed as Q, and the observed noise covariance is expressed as R. Since the calculation of the Kalman gain depends on the ratio between Q and R, four groups are designed in this study with R as the variable, that is, $Q = 0.0001$, $R = 0.001$, $Q = 0.0001$, $R = 0.005$, $Q = 0.0001$, $R = 0.01$, and $Q = 0.0001$, $R = 0.1$, to compare and analyze the efficiency of the Kalman filtering, as shown in Figure 5.

Figure 5 shows that the filtered curves of different groups display different trends. Group 1 is closest to the original curve, and Group 4 has the smoothest curve after noise reduction. Still, some speed measurements of Group 1 are greater than 70 km/h after noise reduction, and the maximum speed value of Group 4 is less than 50 km/h after noise reduction, neither of which matches the actual value recorded in the experimental field. The effect of noise reduction for Groups 2 and 3 could not be judged by Figure 5 alone, so this study uses the recorded actual speed data of the experimental vehicles to calculate the percentage error of different speed intervals, as shown in Figure 6. As Figure 6 shows, Group 2 has the best noise reduction effect and the smallest cumulative percentage error, 9.4%. In addition, the percentage error of Groups 4 and 1 after noise reduction is more considerable, presumably due to the under-reduction and overfitting phenomena in the Kalman filter process. So, the ratio between the system noise and the observed noise should fall reasonably within the accuracy requirements and the original data characteristics.

Analysis of the reconstructed trajectory data of Group 2 reveals that the speed values are all within 70 km/h, which is consistent with the actual value recorded in the experimental field. In addition, the acceleration values of Group 2 also meet the constraints of vehicle dynamics limits and the human body's tolerance, all concentrated between $-3\ \mathrm{m/s^2}$ and $3\ \mathrm{m/s^2}$, as shown in Figure 7.

3. Analysis of Cooperation for Car-Following Behavior in Mixed Traffic

The experiment collects six 15-minute videos. After trajectory extraction and reconstruction, about 821,045 frames of vehicle trajectory data of 9730 vehicles are obtained (1 frame is equal to 0.1 s). It is necessary to define the types of car-following behavior to model the car-following behavior of vehicles in mixed traffic. The distributions, such as the speed, acceleration, and distance headway, are analyzed and then used to understand the cooperation between drivers during the following process.

3.1. Different Types of Car-Following Behavior in Mixed Traffic. There are two vehicle types in mixed traffic, CAV and HDV, so a total of four car-following types are derived, as shown in Figure 8. Since CAVs are equipped with automatic driving and communication functions, they can obtain information about the speed and distance of the vehicle in front of them. However, HDVs do not have these functions, so we speculate that their car-following behavior differs from each other.

Figure 8 shows that there are four car-following types in mixed traffic, namely CAVs followed by CAVs, CAVs followed by HDVs, HDVs followed by CAVs, and HDVs followed by HDVs. Twenty pairs of car-following groups (5 pairs for each car-following type) are selected from the vehicle trajectory set to study the distribution of the vehicle's speed, acceleration, and distance headway during the following process. Figure 9 shows the vehicle's speed

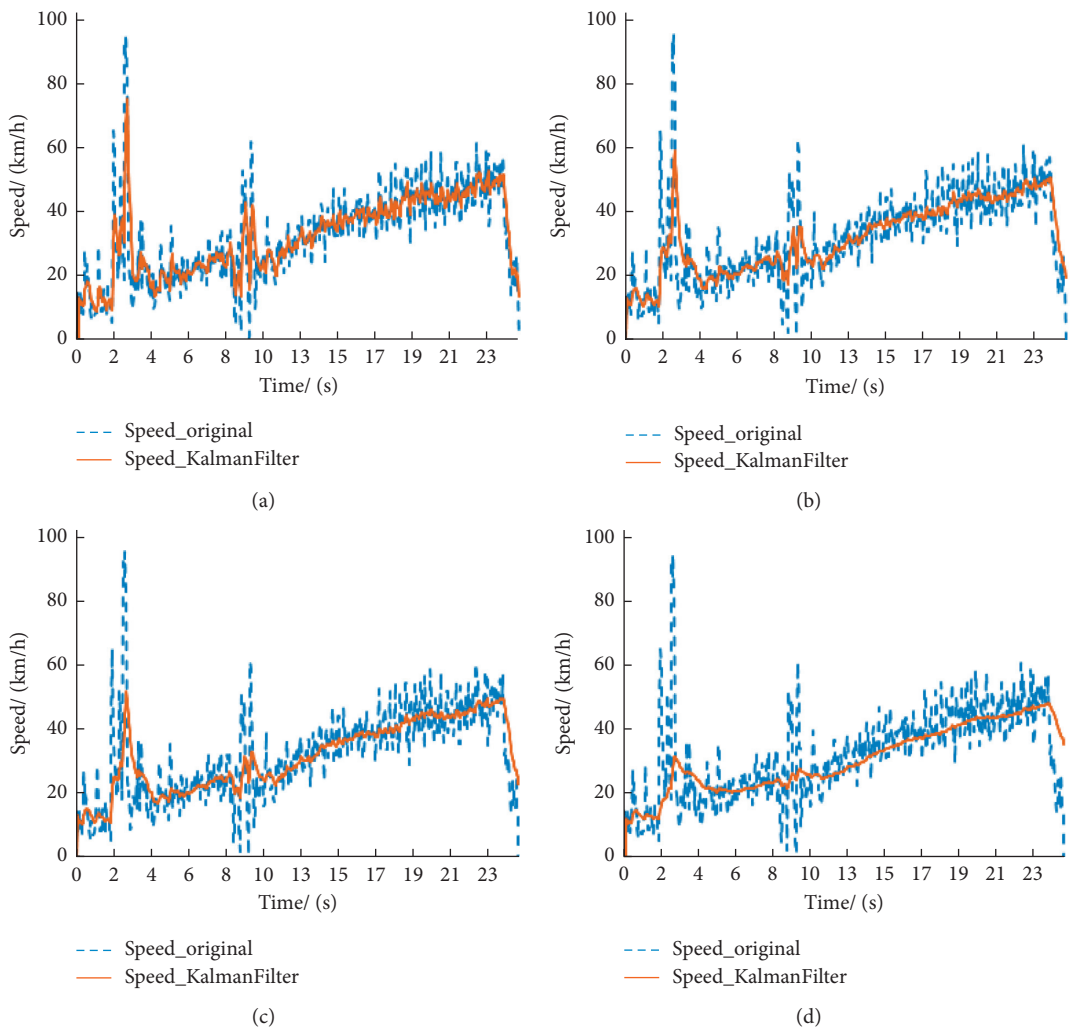


Figure 5: Comparison of different Kalman filter parameter settings. (a) Group 1; (b) Group 2; (c) Group 3; and (d) Group 4.

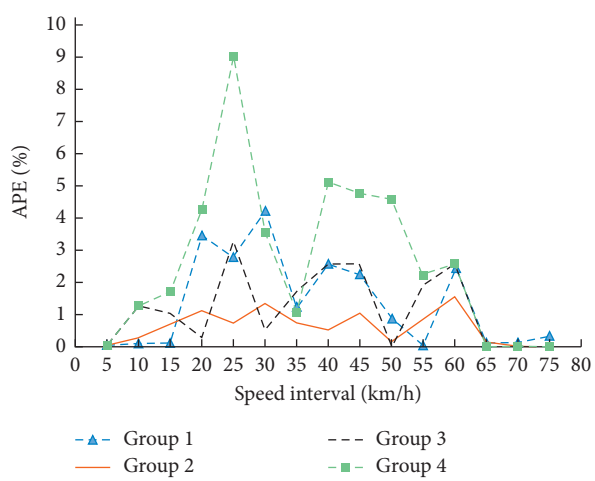


Figure 6: Absolute percentage error (APE) for different speed intervals.

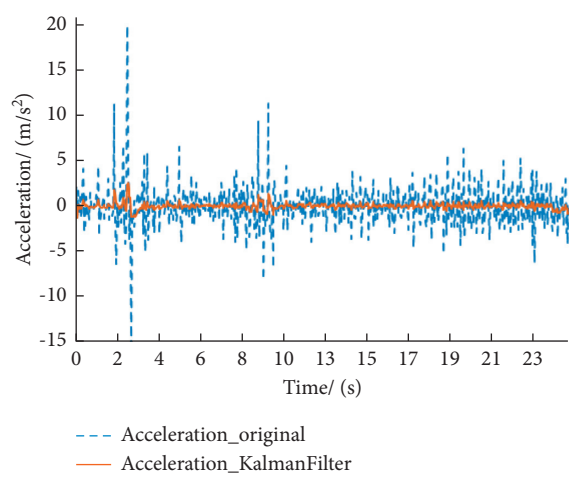


Figure 7: Comparison of acceleration before and after Kalman filter for Group 2.

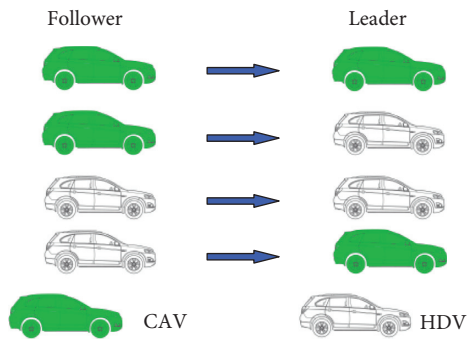


Figure 8: Four car-following types in mixed traffic.

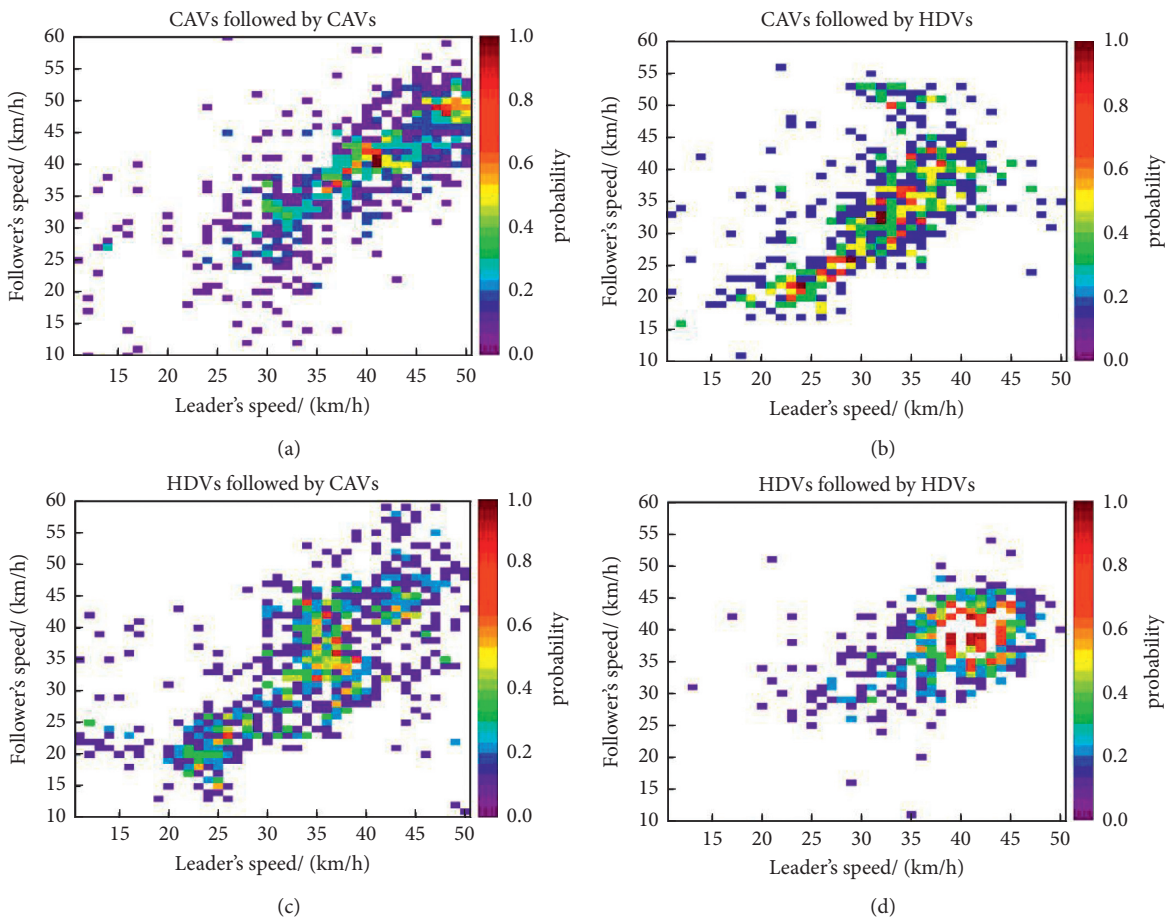


Figure 9: Vehicle's speed distributions of four car-following types. (a) Type 1; (b) Type 2; (c) Type 3; and (d) Type 4.

distributions of the leader and follower for the different car-following types.

As shown in Figure 9, for the first of the four types of car-following, when CAVs follow CAVs, the vehicle speed is distributed in the high-speed region. The possible reason is that both vehicles use CAV technologies, and thus the willingness to cooperate between drivers is stronger. The speed distributions of the other three types are similar. There is no communication between the front and rear vehicles for these three car-following types, which affects the car-following behavior of the driver. The distribution relationship between the speed difference and the follower's acceleration for different car-following types appears in Figure 10.

Although the speed distributions are concentrated in the high-speed region when the vehicles in the car-following pairs are both CAVs, the distributions of the speed difference and the follower's acceleration are more dispersed, as shown in Figure 10. Figure 10 also shows that the distributions are more concentrated for the car-following pairs of Types 2 and 4, which indicates that when the follower is a CAV, the vehicle is more sensitive to the speed change of the front vehicle. Therefore, the acceleration change of the follower is

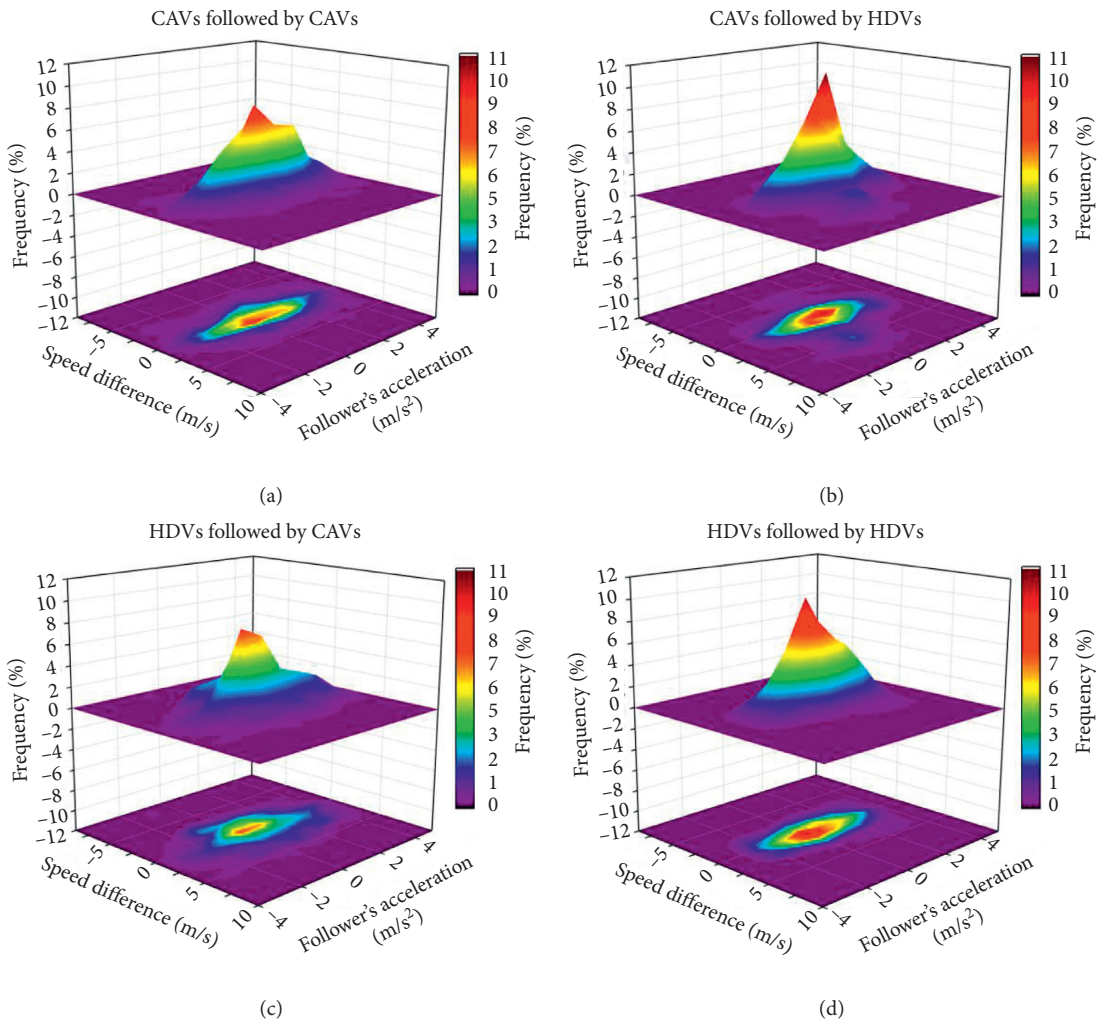


Figure 10: The distribution relationship between the speed difference and the follower's acceleration for different car-following types. (a) Type 1; (b) Type 2; (c) Type 3; and (d) Type 4.

also more sensitive. The situation reverses when the follower is an HDV.

We further analyzed the distance headway distribution when the follower is a CAV or an HDV, respectively, as shown in Figure 11. From Figure 11(a), we can see that, in general, the distance headway distribution is more discrete when the follower is a CAV, but the median value is smaller than that when the follower is an HDV. Further analysis of the distance headway distributions for the four car-following types shows that the second car-following type causes this phenomenon. When both vehicles in the car-following pair are the CAV, the distance headway distribution is more concentrated, and the median value is smaller, which can be seen in Figure 11(b).

3.2. Analysis of Cooperation during the Following Process. In this part, taking the car-following pairs in the data set as an example, the cooperation between vehicles is explained from the variation of speed, headway, and the follower's acceleration. Figure 12 shows the speed variation of the vehicles, where Figures 12(a) and 12(b) shows the case that the follower is a CAV and an HDV, respectively.

Figure 12 shows that the speed variation of the leader and follower displays the same trend during the following process, regardless of whether the follower is a CAV or an HDV, which indicates that the speed cooperation between vehicles does exist during the following process. However, a comparison of Figures 12(a) and 12(b) reveals that the speed change of the follower is smoother when the vehicle is an HDV. Conversely, when the follower is a CAV, the speed variation of the vehicle fluctuates more. To further verify the cooperative relationship between vehicles, Figure 13 compares the variation trends of acceleration and distance headway when a CAV and an HDV are the followers.

Analysis of the variation of the follower's acceleration with the distance headway shows that there is also a cooperative relationship between vehicles during the following process. This relationship is more noticeable when the follower is a CAV because the follower's acceleration is more sensitive to the change of distance headway, as shown in Figure 13(a).

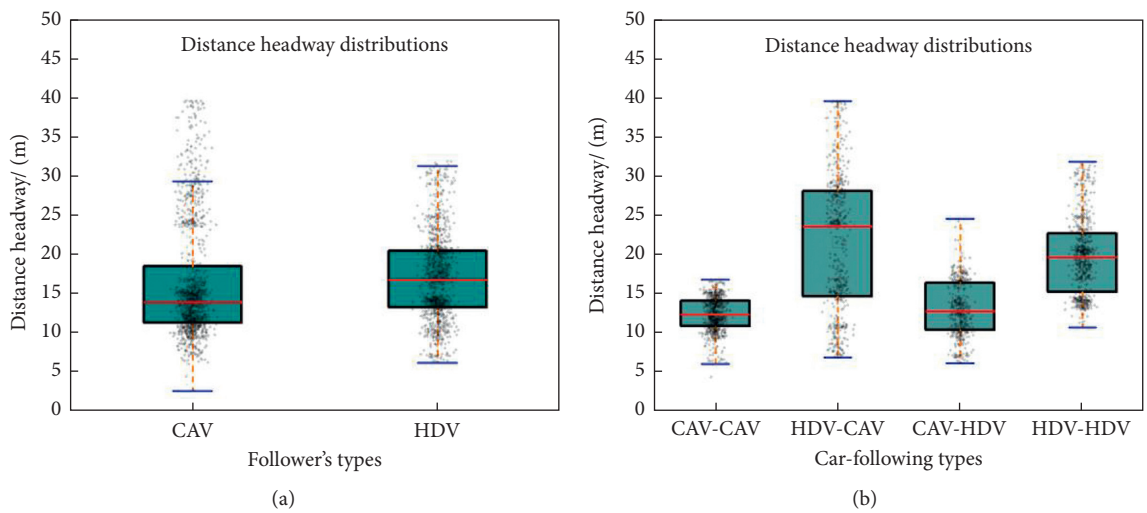


Figure 11: The distance headway distributions. (a) The distributions for different follower's types. (b) The distributions for different car-following types.

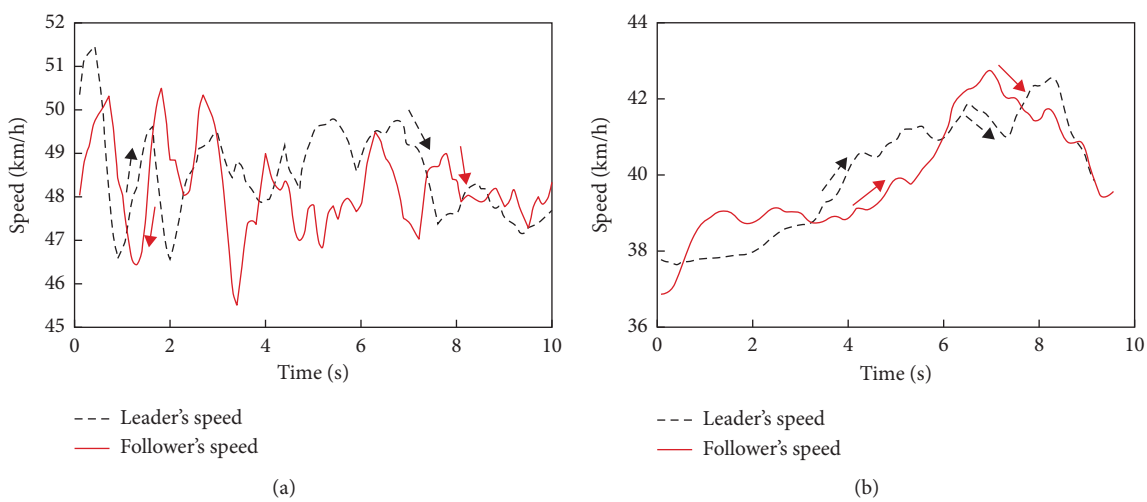


Figure 12: Speed variation during the following process. (a) Speed variation of CAV follower. (b) Speed variation of HDV follower.

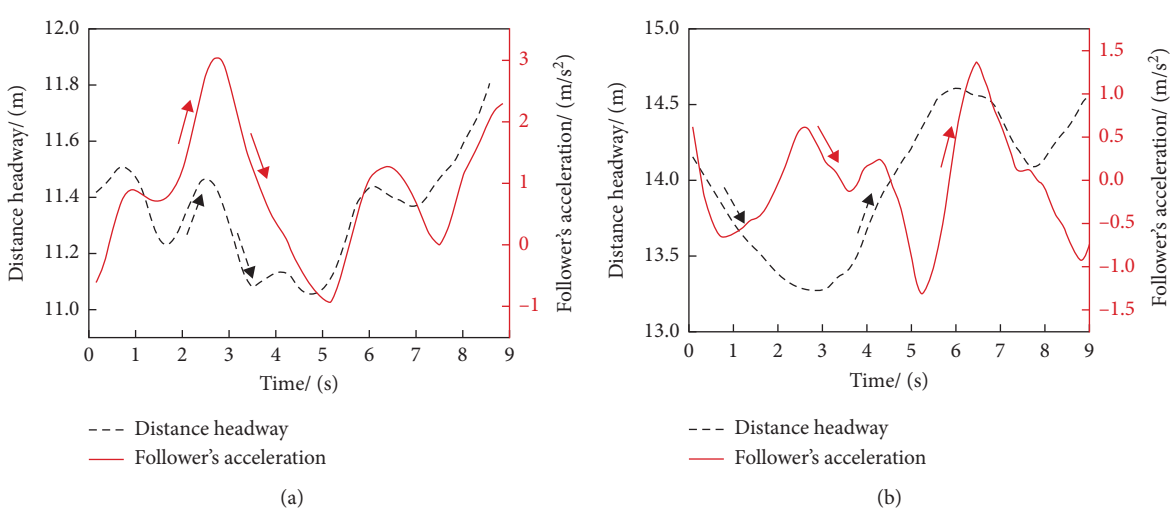


Figure 13: Variation of acceleration and distance headway during the following process. (a) CAV follower. (b) HDV follower.

4. Modeling Car-Following Behavior Based on Prospect Theory

In this section, the cooperation between drivers is measured based on the driver's response to the motion state of the front vehicle, and the cooperation is divided into two levels. A driver's cooperation decision for the state change of the front vehicle (e.g., "leader slowing down") is a classic case of a decision under risk. Decision-making under risk is often modeled by prospect theory (PT) or expected utility theory (EUT). In this paper, PT is used to model the cooperation between drivers because it can describe rational and irrational driving behaviors realistically and consistently, while EUT is best suited for modeling rational decision-makers [14].

4.1. Prospect Theory. Prospect theory, proposed by Daniel Kahneman and Amos Tversky [33], applies psychological research to economics and has made outstanding contributions to the study of human judgment and decision-making under uncertainty. In PT, the decision-maker relates the perceived utility of each available choice and selects the most significant perceived utility. In transportation engineering, PT is mainly used to model route choice behavior and is rarely used to model car-following behavior, except for the work by Anshuman et al. [14]. This study uses PT to model driver's compliance in a connected environment. To our knowledge, PT has not been applied to model the cooperation behavior between drivers in a connected and autonomous driving environment.

The options given to the decision-maker are called prospects in PT. The prospects are first formulated to simulate the decision-maker's choices. Then, the utility value of each prospect is calculated. The prospect with the highest utility describes the decision-maker's final choice. A simple prospect includes two outcomes: Gain or Loss expressed by x and Neutral expressed by 0. Gain means that there are some positive or beneficial supplements to the status of the decision-maker. Loss means that there are some unfavorable supplements to the current status of the decision-maker, and Neutral means that the status of the decision-maker has not changed. The utility related to a simple prospect represented by $U(x, p)$ is the product of the value related to x and p:

$$U(x, p) = v(x) \cdot w(p). \tag{1}$$

The value function $v(x)$ and weight function $w(p)$ can be expressed by equations (2) and (3) [14]:

$$v(x) = \begin{cases} x^{\alpha}, & x > 0, \\ -\lambda(-x)^{\beta}, & x \leq 0, \end{cases} \tag{2}$$

$$w(p) = \frac{p^{\gamma}}{\left(p^{\gamma} + (1-p)^{\gamma}\right)^{1/\gamma}}, \tag{3}$$

where the functions $v(x)$ and $w(p)$ are the values and weights related to the outcome x and probability p. The parameters α, λ, and γ control the shape of the PT curve. When the value of α is less than 1, the value function $v(x)$ will have a concave shape in the gain part and a convex shape in the loss part. When $\lambda > 1$, the loss part of the curve is steeper than the gain part, implying that decision-makers are more sensitive to losses. The value of γ determines how decision-makers perceive the probability. More specially, when γ is equal to 1, the weight function $w(p)$ becomes linear.

As mentioned by Anshuman et al. in their study, PT has four basic properties, including diminishing sensitivity, reference dependence, loss aversion, and probability weighting. The first three properties are related to the value function, and the last one is related to the weight function. Usually, the sensitivity of decision-makers to gains and losses decreases as the value increases, especially for some large values where the sensitivity value is low. Reference dependence refers to the fact that most people's judgments of gains and losses are often determined by a reference point rather than an absolute value. For example, in a choice between "someone else earns $50,000 a year and you earn $60,000 a year" and "someone else earns $80,000 a year and you earn $70,000 a year," most people will choose the former. Loss aversion shows that decision-makers are more sensitive to loss, which is characterized by making the loss part of the value function steeper than the gain part. For example, the pleasure associated with gaining $100 cannot offset the pain associated with losing $100. Finally, probability weighting shows the characteristics of decision-makers' perceptions in response to probabilities and is quantified by equation (3). For example, in the face of small probability events, human beings have an ambivalent attitude toward risk. One can be a risk-lover or a risk-averse person, and events with low probability tend to be given high weight by decision-makers.

4.2. Cooperation Modeling Using Prospect Theory. We assumed that all changes in driving behavior due to CAVs technologies are attributable to the cooperative urgency of the driver's current state. Zero cooperation does not lead to a change in the following behavior. In addition, the driver's choice of the level of cooperation is usually dependent on the distance headway when the information is obtained from the vehicle in front. The distance headway reflects the driver's intuitive experience better than the time headway. As the distance headway decreases, the more urgent and easier it is for the driver to cooperate with the vehicle in front, and vice versa. Thus, the cooperation willingness of a driver in response to the change of leader's moving state can be categorized as two levels: low cooperation level and high cooperation level.

The two cooperation levels can be regarded as two prospects, and we use PT to simulate the prospects chosen by the driver. Both prospects are currently treated as simple prospects. In the car-following process, any level of cooperation in response to perceptually obtained information about changes in the driving state of the front vehicle indicates a gain for the driver. In contrast, a complete lack of cooperation suggests a loss. Therefore, we can consider that both prospects consist of gains. Based on the utility formula

of simple prospect, as shown in equation (1), we define the cooperation level utility as the product of urgency and weight, as shown in equation (4). Urgency indicates the willingness of drivers to cooperate when the state of the front vehicle changes, while weight indicates how drivers perceive different levels of cooperation. The urgency value function is used to calculate the urgency value, and the weighting function is used to calculate the weight.

$$\text{Cooperation utility} = \text{urgency} \times \text{weight}. \tag{4}$$

Since the cooperation utility is proportional to the urgency, the greater the urgency, the higher the driver's willingness to cooperate. We assume that the urgency and the observed distance headway have the same inverse relationship as the cooperation. When the driver perceives a change in the motion state of the front vehicle, the smaller the observed distance headway at that time, the greater the driver's willingness to operate collaboratively, and vice versa. In addition, the logistic function is selected for modeling. Considering the nature of PT's urgency function, the urgency value function is expressed by

$$V\left(dh_{\text{obs}}\right) = \frac{1}{1 + e^{\lambda\left(\alpha \cdot dh_{\text{obs}} - 1\right)}}, \tag{5}$$

where dh_{obs} is the observed distance headway, and α and λ are the parameters to be calibrated.

The inversely proportional relationship between the urgency value and the observed distance headway is shown in Figure 14. For small and large distance headways, the sensitivity of urgency values decreases (flatter at both ends of the curve), which is consistent with the diminishing sensitivity of PT. In addition, the urgency value function can be used to estimate the distance headway when the urgency value is close to zero and one. When the urgency value is one, it is very urgent, and when the urgency value is zero, it is not necessary to cooperate with the front vehicle. The estimated minimum and maximum distance headway will be used to calculate the weight function.

We assume that each driver has a distance headway range corresponding to each cooperation level. The value of the weight at a particular cooperation level is determined by the probability that the observed distance headway falls within that level distance headway range. Therefore, the weight is higher for this distance headway at high cooperation levels and lower at low cooperation levels. The weighting functions for low and high levels of driver cooperation are also formulated in this way. In addition, the functions reflect the PT properties of probability weighting, i.e., higher weighting for small probability events and lower weighting for significant probability events. Referring to Anshuman et al.'s study [14], the weighting functions are shown as follows.

The low cooperation weighting function is given by equations (6) and (7):

$$W_{LC}\left(P_{LC}\right) = \frac{P_{LC}^{\gamma}}{\left(P_{LC}^{\gamma} + \left(1 - P_{LC}\right)^{\gamma}\right)^{1-\gamma}}, \tag{6}$$

$$P_{LC} = \min\left(\frac{dh_{\text{obs}}}{dh_{\max}}, 1\right). \tag{7}$$

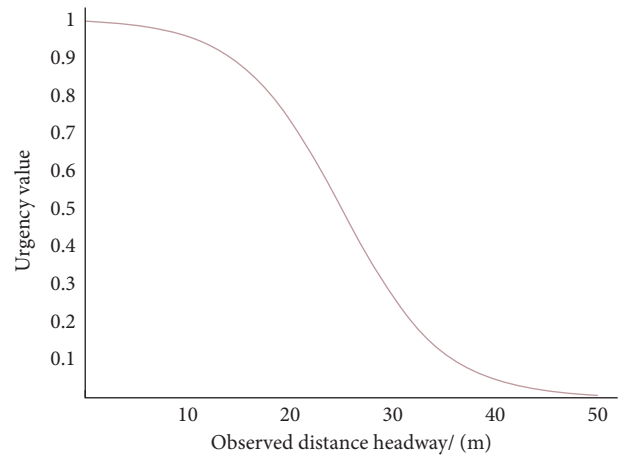


Figure 14: Schematic diagram of the urgency value function.

The high cooperation weighting function is given by equations (8) and (9):

$$W_{HC}\left(P_{HC}\right) = \frac{P_{HC}^{\gamma}}{\left(P_{HC}^{\gamma} + \left(1 - P_{HC}\right)^{\gamma}\right)^{1-\gamma}}, \tag{8}$$

$$P_{HC} = \min\left(\frac{dh_{\min}}{dh_{\text{obs}}}, 1\right), \tag{9}$$

where $W_{LC}\left(P_{LC}\right)$ and $W_{HC}\left(P_{HC}\right)$ are the low cooperation weighting function and high cooperation weighting function, respectively. P_{LC} and $P_{HC}$are the probabilities that an observed distance headway falls in a driver's low and high cooperation ranges. $dh_{\max}$ and $dh_{\min}$ are the distance headways when the values of $V\left(dh_{\text{obs}}\right)$ are close to zero and one, respectively. γ is the shape parameter.

Thus, the cooperation utility functions for low and high cooperation levels can be defined by equations (10) and (11), and the final cooperation utility value U is the larger of the low and high utility values, as shown in equation (12):

$$U_{LC}\left(db_{\text{obs}}, P_{LC}\right) = V\left(db_{\text{obs}}\right) \cdot W_{LC}\left(P_{LC}\right), \tag{10}$$

$$U_{HC}\left(db_{\text{obs}}, P_{HC}\right) = V\left(db_{\text{obs}}\right) \cdot W_{HC}\left(P_{HC}\right), \tag{11}$$

$$U = \max\left(U_{LC}, U_{HC}\right). \tag{12}$$

4.3. Modeling Car-Following Behavior Considering Cooperation between Drivers. A car-following model is a mathematical description of the influence of the driving environment on driving behavior. The classical car-following model originates from driving dynamics and uses vehicle acceleration to reflect the car-following decision. Based on the classic car-following model of PATH Laboratory, this section models car-following behavior in mixed traffic with HDVs and CAVs from the perspective of cooperation between drivers.

Scholars from the PATH Laboratory at the University of California, Berkeley, have conducted a long-term study on the ACC/CACC car-following model [34]. In the proposed

car-following model, the acceleration of the following vehicle depends on three components: the acceleration of the front vehicle, the speed difference between the front and rear vehicles, and the error term between the actual distance headway and the desired distance headway. The structure of the car-following model by PATH Laboratory can be expressed as follows:

$$a_i = k_0 a_{i-1} + k_1 e + k_2 (v_{i-1} - v_i), \tag{13}$$

$$e = x_{i-1} - x_i - T v_i, \tag{14}$$

where a_i and a_{i-1} represent the acceleration of the rear and front vehicles. x_i and x_{i-1} represent the position of the rear and front vehicles. v_i and v_{i-1} represent the speed of the rear and front vehicles, respectively. Moreover, e is the difference between the actual distance headway and the desired distance headway. T is the desired time headway. k_0, k_1, and k_2 are the weights to be calibrated.

Based on the cooperation analysis for car-following behavior in mixed traffic, the car-following behavior of CAVs and HDVs are different. First, CAVs and HDVs have different acceleration, speed, and distance headway distributions. The distributions of acceleration and distance headway for CAVs fluctuate more, and the speed distribution of CAVs also concentrates in the high-speed region. In addition, the cooperative phenomenon between vehicles is more evident when the follower is a CAV than an HDV. Still, the speed and acceleration are more turbulent when the follower is a CAV, probably, because CAVs are more responsive to the front vehicle's state changes than HDVs. This different cooperative nature exhibited by different vehicle types can be characterized by PT, see Section 4.2. In this paper, the extended car-following model in mixed traffic is developed considering the acceleration of the front vehicle, the speed difference between the front and rear vehicles, and the distance headway between the front and rear vehicles. Also, the cooperation willingness of different vehicle types in response to the state changes of the leader is described using PT. Equations (15)–(17) present the mathematical formulations of the proposed model:

$$\begin{aligned} a_n(t) = {} & \delta \cdot a_{n-1}(t) + \eta \cdot \left[S_n(t) - S_n^*(t)\right] \\ & + \theta \cdot \left[V_{n-1}(t) - V_n(t)\right], \end{aligned} \tag{15}$$

$$S_n(t) = Y_{n-1}(t) - Y_n(t), \tag{16}$$

$$S_n^*(t) = S_0 + \left[1 + U\left(dh_{\text{obs}}\right)\right] \cdot T V_n(t) + \frac{V_n(t) \cdot \Delta V_n(t)}{2\sqrt{ab}}, \tag{17}$$

where $a_n(t)$ and $a_{n-1}(t)$ are the accelerations of the nth and $(n-1)^{\text{th}}$ vehicles at moment t. $S_n(t)$ is the distance headway of the nth vehicle at moment t. $S_n^*(t)$ is the desired distance headway of the nth vehicle at moment t. $V_n(t)$ and $V_{n-1}(t)$ are the speeds of the nth and $(n-1)^{\text{th}}$ vehicles at moment t. $Y_n(t)$ and $Y_{n-1}(t)$ are the positions of the nth and $(n-1)^{\text{th}}$ vehicles at moment t. Moreover, S_0, T, a, b, $V_n(t)$, and $U(dh_{\text{obs}})$ are the standstill distance, desired time headway, maximum acceleration, desired deceleration, speed difference between vehicles, and cooperation utility value, respectively. δ, η, and θ are the parameters to be calibrated.

5. Calibration and Verification Methodology

In this paper, RMSPE of distance headway, mentioned by many studies [13, 35], is used as the objective function for parameter calibration:

$$\text{RMSPE} = \sqrt{\frac{1}{n} \cdot \frac{\sum_{i=1}^{n} \left(s_i^{\text{sim}} - s_i^{\text{act}}\right)^2}{\sum_{i=1}^{n} \left(s_i^{\text{act}}\right)^2}}, \tag{18}$$

where RMSPE of distance headway as the measure of model performance denotes the objective function, n is the number of observations, and s_i^{sim} and s_i^{act} are the ith simulated distance headway and actual distance headway, respectively.

The genetic algorithm (GA) is used to solve the objective function, and each parameter of GA is set as follows: the population size, the maximum number of generations, the number of stall generations, and the function tolerance are set to 100, 300, 100, and $1e-6$. If the change is less than the function tolerance, the algorithm will stop, and the algorithm finds a different solution in each optimization run. The optimization is repeated ten times for each driver to obtain a solution closer to the global optimum, and the set of parameters with the smallest RMSPE is selected. In addition, to improve the computational tractability of the genetic algorithm optimization process, the paper sets the upper and lower bounds for the parameters of the extended car-following model, as shown in Table 1.

In the parameter calibration, the performance of the proposed model calibration method is tested using the so-called "synthetic data" generated from each driver's trajectory data, as suggested by Punzo et al. [36]. For example, the parameters in the extended car-following model are set to $\alpha = 0.3$, $\lambda = 5.5$, $\gamma = 0.55$, $\delta = 2.5$, $\eta = 2.5$, $\theta = 2.5$, $S_0 = 0.55$, $T = 2.5$, $a = 2.5$, and $b = 2.5$, and synthetic car-following data are generated. Then, a calibration process is implemented to derive the best model parameters based on the set calibration accuracy.

As mentioned earlier, there are four car-following types in mixed traffic. In particular, the car-following behavior changes significantly whether the rear vehicle is a CAV or an HDV. Therefore, the paper performs model calibration and verification for the rear vehicle as a CAV or as an HDV. We select twenty leader-follower pairs from the data set, including ten pairs of CAV followers and ten pairs of HDV followers. Nine out of ten pairs in each group are used for calibration, and the remaining one is used for calibration. In addition, in order to retain generality, twenty groups of initial parameter data are selected for each leader-follower pair.

6. Results and Discussions

Table 2 shows the results of the parameter calibration for a leader-follower pair, and it only offers the calibration

Table 1: Parameter bounds in optimization.

Parameters	Definition	Bounds
α	Parameter of the urgency value function	[0.1, 0.5]
λ	Parameter of the urgency value function	[1, 10]
γ	Parameter of the weighting function	[0.1, 1]
δ	Weight of acceleration	[0.1, 5]
η	Weight of distance headway	[0.1, 5]
θ	Weight of speed difference	[0.1, 5]
S_0	Standstill distance	[0.1, 10]
T	Desired time gap	[0.1, 5]
a	Maximum acceleration	[0.1, 5]
b	Desired deceleration	[0.1, 5]

Table 2: Parameter calibration results for a leader-follower pair.

α	λ	γ	δ	η	θ	S_0	T	a	b	RMSPE (%)
0.33	9.18	0.84	3.83	3.80	4.57	0.70	1.38	3.96	4.90	2.17
0.20	3.73	0.67	2.79	1.59	4.96	9.15	3.85	1.15	1.71	2.28
0.49	2.10	0.30	2.40	2.60	3.66	6.72	1.43	4.05	4.92	2.36
0.36	8.01	0.36	3.34	2.28	4.98	2.16	4.18	4.90	3.04	2.10
0.45	5.35	0.45	1.59	0.70	3.11	7.50	2.43	2.53	0.70	2.11
0.18	**7.72**	**0.90**	**1.41**	**1.96**	**4.00**	**5.13**	**2.09**	**1.83**	**2.11**	**1.92**
0.21	5.92	0.91	2.95	0.71	2.12	6.54	4.20	3.93	2.46	2.02
0.45	1.76	0.69	2.00	2.17	4.99	0.38	0.54	3.79	0.99	2.44
0.48	4.03	0.36	2.13	1.40	3.66	1.86	2.05	2.83	1.74	2.50
0.16	9.63	0.53	2.45	1.54	4.81	4.46	4.54	3.95	1.89	2.38

Bold values represent the optimal value of each parameter and the minimum RMSPE value.

results corresponding to the minimum error in each set of initial values. The minimum calibration error of 1.92% is obtained for the 6th set of initial values, which is also the final calibration error for this leader-follower pair. Figure 15 shows the distribution of the parameter calibration errors.

Figure 16 shows the distributions of the calibrated parameter values in the cooperation utility. The figure also shows that, for CAVs, the value of γ is larger than that for HDVs. γ is the parameter in the weighting function, and the larger value means that the willingness of the follower to cooperate is stronger when the driving state of a leader changes.

The distributions of the calibrated parameter values in the desired distance headway appear in Figure 17. The figure shows that the distributions of the maximum acceleration and desired time gap are significantly different. Compared with HDVs, the maximum acceleration value of CAVs is larger. At the same time, the desired time gap is smaller, which leads to a shorter desired distance headway of the CAV, indicating that the CAV is more willing to maintain a smaller distance headway with the vehicle in front.

Figure 18 presents the distributions of the calibrated parameter values of δ, η, θ, which respectively represent the weight of the speed difference, the distance headway, and the leader's acceleration. The figure reveals that the value of η is larger for CAVs compared with HDVs. A larger value of η means that the CAV's acceleration is more influenced by the distance headway compared with the HDV. In particular, the difference in parameters is

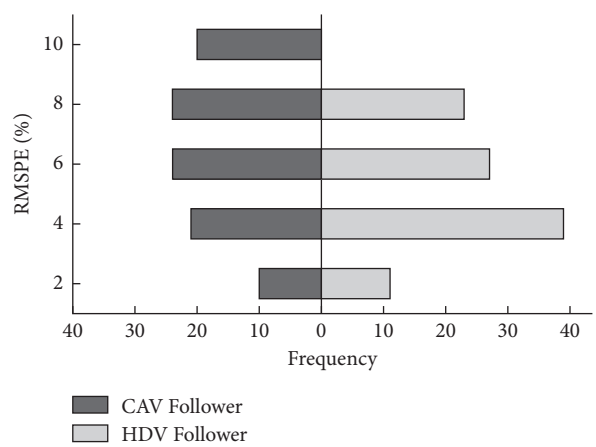


Figure 15: Distribution of calibration errors.

attributed to the heterogeneity of CAVs and HDVs. The traffic behavior of CAVs and HDVs in mixed traffic is very different due to the different levels of CAV technology and the involvement of human factors. After determining the optimal values of each parameter, the proposed car-following model is used to calculate the traffic state parameters of the follower. Figure 19 presents the variation of the follower's speed and the variation of the difference between the actual and simulated follower's speed. As shown in Figure 19, the proposed model can well characterize the car-following behavior of the follower, whether it is a CAV or an HDV as a rear vehicle.

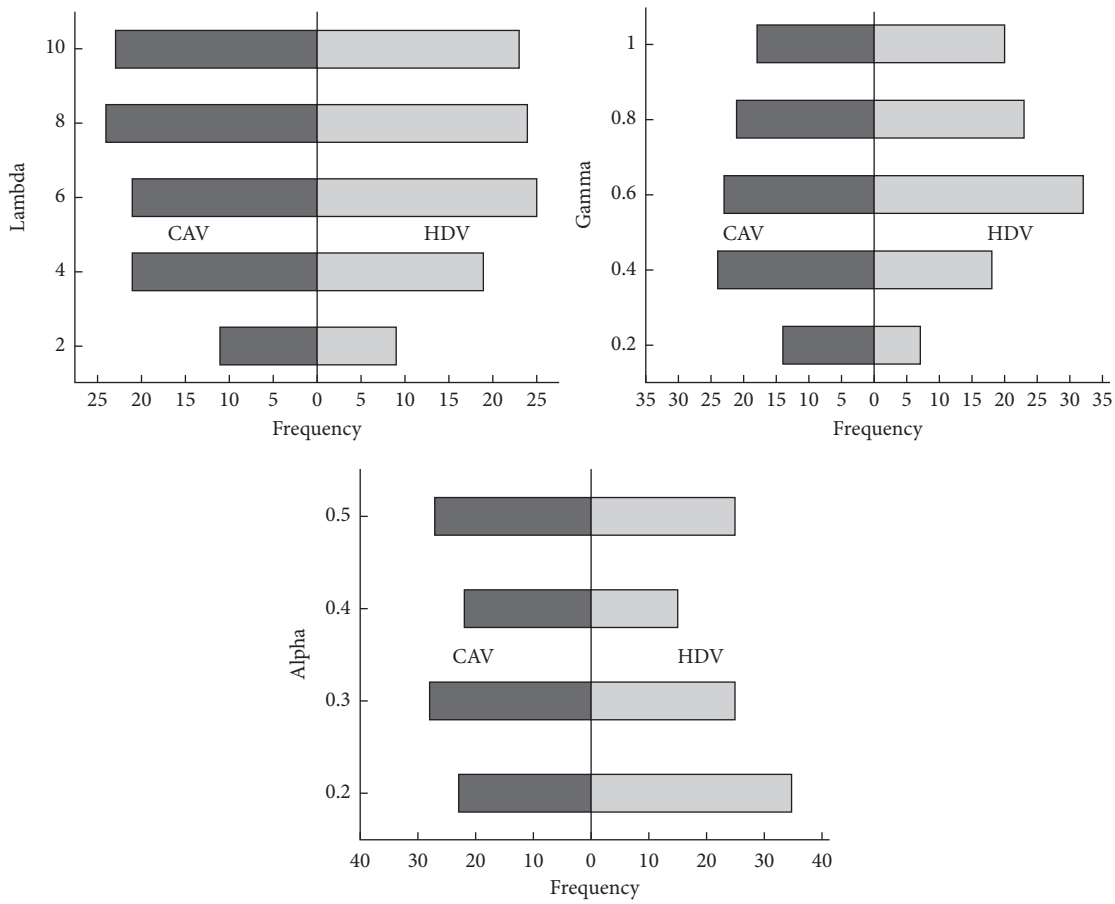


Figure 16: Distributions of the calibrated parameter values in the cooperation utility.

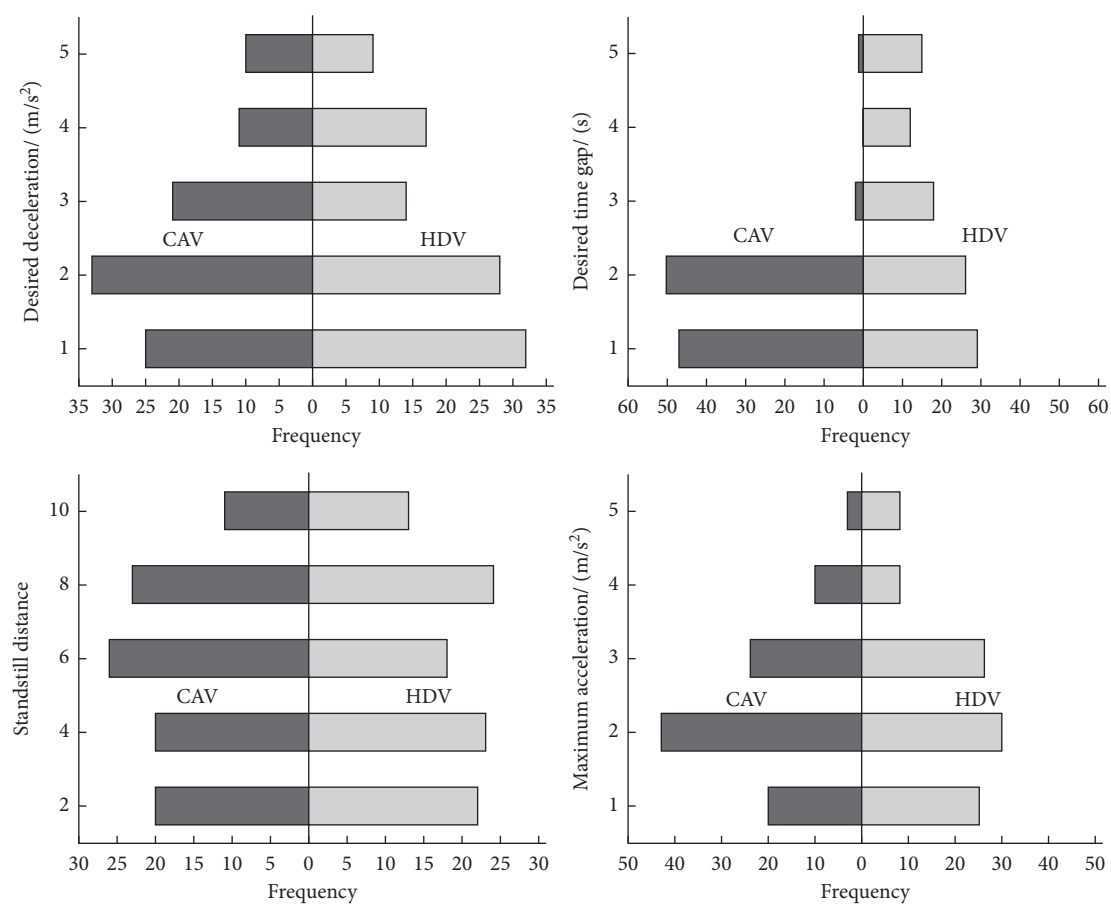


Figure 17: Distributions of the calibrated parameter values in the desired distance headway.

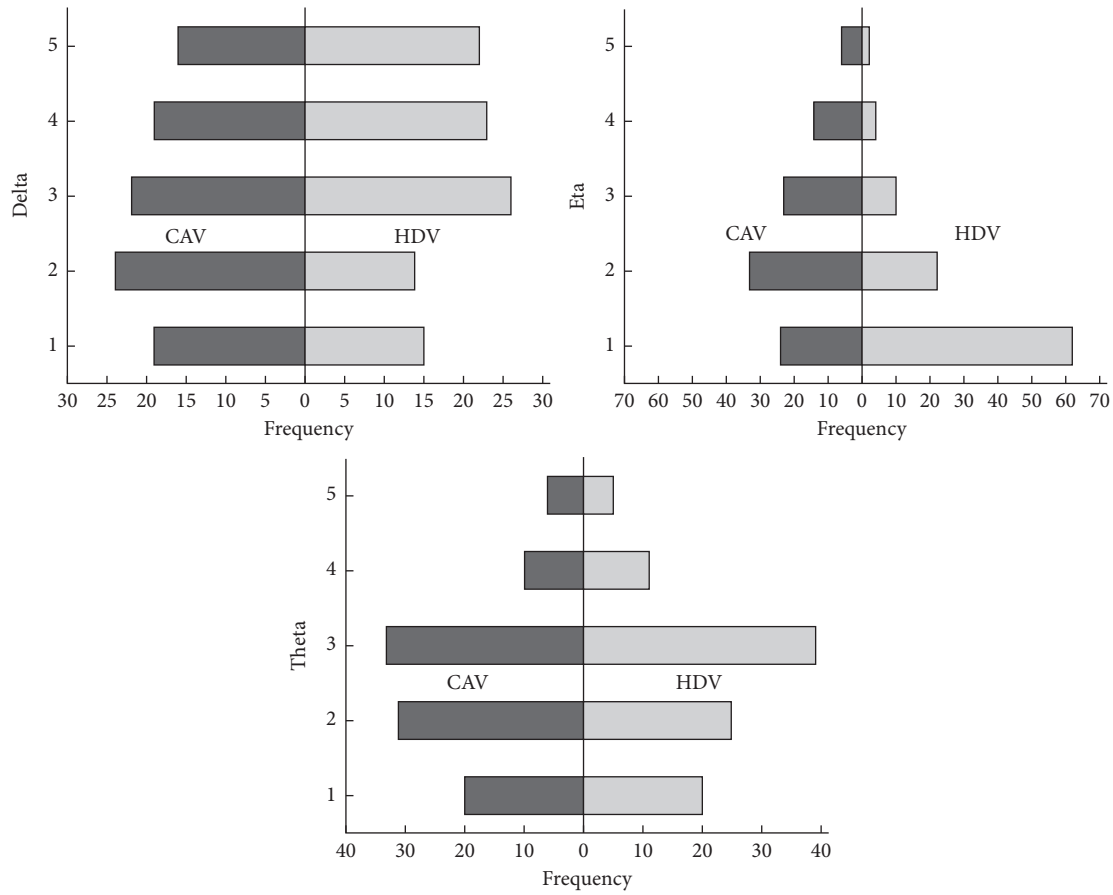


Figure 18: Distributions of the calibrated parameter values of δ, η, θ.

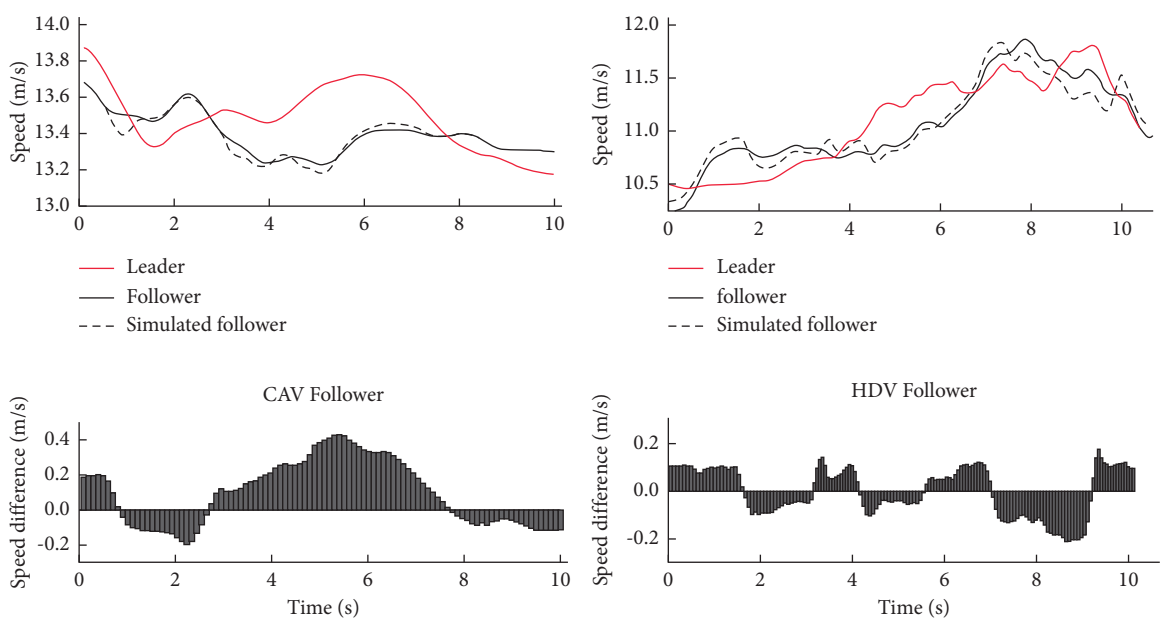


Figure 19: Verification for the actual and simulated follower's speed.

7. Conclusions and Future Research

This paper proposes an extended car-following model for mixed traffic in a connected and autonomous driving environment, which can characterize the car-following behavior of CAVs and HDVs in the heterogeneous traffic flow. Using PT, the cooperation between drivers, a human factor, is well integrated into the model. The weight and value functions of PT are modified to simulate the willingness of the follower to cooperate with the front vehicle. To specifically overcome the unavailability of vehicle trajectory data in connected and autonomous vehicle environment, we design and implement a car-following experiment in the CACC scenario and extract vehicle trajectory data based on video recognition. Furthermore, the Lagrangian theory and Kalman filter theory are used to reconstruct the extracted trajectory data to ensure the accuracy of the data.

The model calibration and verification results, combined with the analysis of cooperation for car-following behavior, reveal that the heterogeneity of CAVs and HDVs in mixed traffic is manifested in various ways. Compared with HDVs, CAVs tend to maintain a smaller distance headway with the vehicle in front and have a greater willingness to cooperate with the leader. When the distance headway becomes larger, CAVs are more responsive than HDVs, which is one of the reasons for the more significant fluctuations in the speed and the acceleration of CAVs. It is worth noting that the results reported in this paper are obtained from a relative rather than an absolute perspective. Different results may be obtained depending on the experimental design and the development of CAV technologies.

This paper has several limitations that need further study. One issue is that the experiment is designed without considering intersection signals. The experiment can be improved in the future to study the effect of vehicle-intersection interaction on car-following behavior and some critical traffic flow phenomena, e.g., capacity drop. Another issue is that, when developing the car-following model, only the influence of one vehicle in front is considered. In the future, multiple vehicles in front can be included.

Data Availability

The data used to support the findings of this study are available from the first author upon request (Shenzhen Ding, 17114228@bjtu.edu.cn).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge that this paper is prepared based on the National Natural Science Foundation of China (NSFC) under grant no. 71871013.

References

[1] J. Tian, C. Zhu, D. Chen, R. Jiang, G. Wang, and Z. Gao, "Car following behavioral stochasticity analysis and modeling: perspective from wave travel time," *Transportation Research Part B: Methodological*, vol. 143, no. 2021, pp. 160–176, 2021.

[2] M. Treiber, A. Kesting, and D. Helbing, "Three-phase traffic theory and two-phase models with a fundamental diagram in the light of empirical stylized facts," *Transportation Research Part B: Methodological*, vol. 44, no. 2010, pp. 983–1000, 2010.

[3] H. S. Mahmassani, "50th anniversary invited article-autonomous vehicles and connected vehicle systems: flow and operations considerations," *Transportation Science*, vol. 50, no. 4, pp. 1140–1162, 2016.

[4] SAE International in United States, *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, 2018, https://saemobilus.sae.org/content/J3016_201806/.

[5] J. Chen, D. Sun, Y. Li, M. Zhao, W. Liu, and S. Jin, "Human-machine cooperative scheme for car-following control of the connected and automated vehicles," *Physica A: Statistical Mechanics and its Applications*, vol. 573, Article ID 125949, 2021.

[6] F. M. Favarò, S. O. Eurich, and S. S. Rizvi, ""Human" problems in semi-autonomous vehicles: understanding drivers' reactions to off-nominal scenarios," *International Journal of Human-Computer Interaction*, vol. 35, no. 11, pp. 956–971, 2019.

[7] L. A. Pipes, "An operational analysis of traffic dynamics," *Journal of Applied Physics*, vol. 24, no. 3, pp. 274–281, 1953.

[8] D. C. Gazis, R. Herman, and R. W. Rothery, "Nonlinear follow-the-leader models of traffic flow," *Operations Research*, vol. 9, no. 4, pp. 545–567, 1961.

[9] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E*, vol. 62, no. 2, pp. 1805–1824, 2000.

[10] M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, "Dynamical model of traffic congestion and numerical simulation," *Physical Review E*, vol. 51, no. 2, pp. 1035–1042, 1995.

[11] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation Research Part B: Methodological*, vol. 15, no. 2, pp. 105–111, 1981.

[12] R. Wiedemann, "Simulation des strassenverkehrsflusses,," in *Proceedings of the Schriftenreihe des instituts fir Verkehrswesen der Universitiit Karlsruhe*, Germany, 1974.

[13] M. Zhu, X. Wang, A. Tarko, and S. e. Fang, "Modeling car-following behavior on urban expressways in Shanghai: a naturalistic driving study," *Transportation Research Part C: Emerging Technologies*, vol. 93, pp. 425–445, 2018.

[14] A. Sharma, Z. Zheng, A. Bhaskar, and M. M. Haque, "Modelling car-following behaviour of connected vehicles with a focus on driver compliance," *Transportation Research Part B: Methodological*, vol. 126, pp. 256–279, 2019.

[15] L. Xiao and F. Gao, "A comprehensive review of the development of adaptive cruise control systems," *Vehicle System Dynamics*, vol. 48, no. 10, pp. 1167–1192, 2010.

[16] T. Li, D. Chen, H. Zhou, J. Laval, and Y. Xie, "Car-following behavior characteristics of adaptive cruise control vehicles based on empirical experiments," *Transportation Research Part B: Methodological*, vol. 147, pp. 67–91, 2021.

[17] G. J. L. Naus, R. P. A. Vugts, J. Ploeg, M. J. G. van de Molengraft, and M. Steinbuch, "String-stable CACC design and experimental validation: a frequency-domain approach," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 9, pp. 4268–4279, 2010.

[18] V. Milanés, S. E. Shladover, J. Spring, C. Nowakowski, H. Kawazoe, and M. Nakamura, "Cooperative adaptive cruise

control in real traffic situations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 296–305, 2013.

[19] Z. Yang, Z. Yang, A. Soyoung, W. Meng, and H. Serge, "Stabilizing mixed vehicular platoons with connected automated vehicles: an H-infinity approach," *Transportation Research Part B: Methodological*, vol. 132, pp. 152–170, 2020.

[20] G. Gunter, C. Janssen, W. Barbour, R. E. Stern, and D. B. Work, "Model-based string stability of adaptive cruise control systems using field data," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 1, pp. 90–99, 2020.

[21] M. Makridis, K. Mattas, A. Anesiadou, and B. Ciuffo, "OpenACC: an open database of car-following experiments to study the properties of commercial ACC systems," *Transportation Research Part C: Emerging Technologies*, vol. 125, Article ID 103047, 2021.

[22] J. Wang, S. Gong, S. Peeta, and L. Lu, "A real-time deployable model predictive control-based cooperative platooning approach for connected and autonomous vehicles," *Transportation Research Part B: Methodological*, vol. 128, pp. 271–301, 2019.

[23] Z. Zhong, E. E. Lee, M. Nejad, and J. Lee, "Influence of CAV clustering strategies on mixed traffic flow characteristics: an analysis of vehicle trajectory data," *Transportation Research Part C: Emerging Technologies*, vol. 115, Article ID 102611, 2020.

[24] W. Zhao, D. Ngoduy, S. Shepherd, R. Liu, and M. Papageorgiou, "A platoon based cooperative eco-driving model for mixed automated and human-driven vehicles at a signalised intersection," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 802–821, 2018.

[25] H. Wang, Y. Qin, W. Wang, and J. Chen, "Stability of CACC-manual heterogeneous vehicular flow with partial CACC performance degrading," *Transportmetrica B: Transport Dynamics*, vol. 7, no. 1, pp. 788–813, 2019.

[26] A. Ghiasi, X. Li, and J. Ma, "A mixed traffic speed harmonization model with connected autonomous vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 104, pp. 210–233, 2019.

[27] M. Amirgholy, M. Shahabi, and H. Oliver Gao, "Traffic automation and lane management for communicant, autonomous, and human-driven vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 477–495, 2020.

[28] V. Milanés and S. E. Shladover, "Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 285–300, 2014.

[29] T. Zhang and P. J. Jin, "A longitudinal scanline based vehicle trajectory reconstruction method for high-angle traffic video," *Transportation Research Part C: Emerging Technologies*, vol. 103, pp. 104–128, 2019.

[30] G. Welch and G. Bishop, "An introduction to the Kalman filter," in *Proceedings of the SIGGRAPH, Course*, vol. 8, p. 41, Los Angeles, CA, USA, August 2001.

[31] H. Liu, J. Huang, and W. Zhang, "Numerical algorithm based on extended barycentric Lagrange interpolant for two dimensional integro-differential equations," *Applied Mathematics and Computation*, vol. 396, Article ID 125931, 2021.

[32] M. Montanino and V. Punzo, "Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns," *Transportation Research Part B: Methodological*, vol. 80, pp. 82–106, 2015.

[33] K. A. Tversky, "Prospect theory: an analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–291, 1979.

[34] S. E. Shladover, D. Su, and X.-Y. Lu, "Impacts of cooperative adaptive cruise control on freeway traffic flow," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2324, no. 1, pp. 63–70, 2012.

[35] J. Sangster, H. Rakha, and J. Du, "Application of naturalistic driving data to modeling of driver car-following behavior," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2390, no. 1, pp. 20–33, 2013.

[36] V. Punzo, B. Ciuffo, and M. Montanino, "Can results of car-following model calibration based on trajectory data be trusted?" *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2315, no. 1, pp. 11–24, 2012.

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 9219176, 22 pages
https://doi.org/10.1155/2021/9219176

Research Article

Modeling Intercity Travel Mode Choice with Data Balance Changes: A Comparative Analysis of Bayesian Logit Model and Artificial Neural Networks

Xiaowei Li,[1] **Yuting Wang**,[1] **Yao Wu**,[2] **Jun Chen**,[1] **and Jibiao Zhou**[3]

[1]*School of Civil Engineering, Xi'an University of Architecture & Technology, Xi'an 710055, China*
[2]*School of Modern Posts and Institute of Modern Posts, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*
[3]*College of Transportation Engineering, Tongji University, Shanghai 201804, China*

Correspondence should be addressed to Yao Wu; wuyao@njupt.edu.cn

Received 1 July 2021; Revised 26 August 2021; Accepted 1 September 2021; Published 14 September 2021

Academic Editor: Xiangyang Guan

This study conducts a comprehensive comparative analysis of regression-based multinomial models and artificial neural network models in intercity travel mode choices. The four intercity travel modes of airplane, high-speed rail (HSR), train, and express bus were used for analysis. Passengers' activity data over the process of intercity travel were collected to develop the models. The standard multinomial logit (MNL) regression and Bayesian multinomial logit (BMNL) regression were compared with the radial basis function (RBF) and multilayer perceptron (MLP). The results show that MLP performs best in terms of predictive accuracy, followed by BMNL and MNL, and RBF is the least accurate. The performances of all models were examined against changes in data balance, and it was found that rebalancing can improve fitting performance while slightly reducing the predictive performance. This comparative study and its parameter estimation shed new light on the comparison of traditional and emerging models in travel behavior studies, and the findings can be used as heuristic guidance for all stakeholders.

1. Introduction

To model passengers' travel behaviors is of value to better understand mobility modes in the complex travel environment [1]. Policies and managerial strategies rely on the accurate estimation of travel mode choices of passengers. In 2020, COVID-19 has profoundly influenced passengers' travel behaviors, causing a dramatic shift in intracity and intercity mobility modes, inevitably affecting society, production, and the global economy. Scholars have investigated contextual factors that influence travel modes, aiming to better understand passengers' choices and develop suitable models.

Previous studies have shown that travel mode choice can be affected by social and demographic factors, including gender [2–4], age [4–6], occupation [3], income [2, 4, 7, 8], and car ownership [4]. Miskeen et al. [4] found that males were more likely to use public transportation than cars, while females were less likely to shift to public transportation. Cheng et al. [6] indicated that age was the most significant individual-related attribute. Tourists were more likely to choose a plane or train than a coach [3]. Forinash et al. [7] found that high- and low-income groups preferred air travel and bus, respectively. It was similarly reported that an increase in passengers' incomes decreased their use of buses [4]. Lower-income individuals were found to be more sensitive to cost and less sensitive to out-of-vehicle time than middle- and high-income individuals [8]. Related attributes, such as travel demand, service quality of transport modes, and accessibility of transportation hubs, have been found to influence travel mode choices [1, 3, 9, 10].

The most widely used modeling techniques in travel mode choice are discrete choice models, such as the binomial logit (BL) [11], multinomial logit (MNL) [4, 12],

multinomial probit (MNP) [3], nested logit (NL) [13, 14], and mixed logit (ML) models [15–17], which have high interpretability of estimation results on input variables, as well as high transferability and validity. Regression-based models form maximum likelihood estimates of parameters [4, 5, 11, 12, 17, 18]. Apart from the popular logit model, Bayesian parameter estimation methods have shown good accuracy and performance [19–22]. For example, Wong and Farooq [23] developed an algorithm based on the restricted Boltzmann machine, which has multiple discrete-continuous layers and can be expressed as a variational Bayesian inference optimization problem.

Emerging machine learning techniques have been studied for travel mode choice [24–34]. Lindner et al. [33] found an artificial neural network (ANN) and classification tree (CT) to outperform binary logit regression in motorized travel mode choice. Cheng et al. [6] found the random forest (RF) to have significantly better prediction accuracy than support vector machine (SVM), adaptive boosting (AdaBoost), and MNL in modeling travel mode choice. Zhao et al. [24] compared the model development, evaluation, and behavior interpretation of MNL and ML with that of the naive Bayes, CT, AdaBoost, bag fruit tree, RF, SVM, and ANN machine learning classifiers. Among machine learning approaches, the multilayer perceptron (MLP) and radial basis function (RBF) have been widely applied due to their better classification accuracy compared to naive Bayes, K-nearest neighbors, and backpropagation neural networks [35–37]. Hence, they have potential use in the study of travel mode choice.

The influence of data balance on the accuracy of multimode choice models has not been widely reported. Imbalanced sample data can influence the accuracy of estimation in multiclass discrete choice prediction [38, 39], and methods such as oversampling and undersampling have been proposed to address this issue [40, 41]. However, there is no commonly agreed best method to resolve this issue in multiclass classification. This is a well-known issue in travel mode choice, and the effectiveness of rebalancing methods when using different regression-based and neural network models in empirical studies of modeling travel mode choice requires study.

This study has three objectives: (1) to investigate the predictive performance of modeling techniques including Bayesian multinomial logit (BMNL), MNL, MLP, and RBF for intercity travel mode choice; (2) to assess the predictive performance of the above techniques after data balancing; and (3) to evaluate the factors affecting intercity travel mode choice and their relative importance using a comprehensive dataset. Passengers' activity data over the whole process of intercity travel were collected. The travel modes of airplane, HSR, train, and express bus were investigated. The BMNL, MNL, MLP, and RBF models were developed and validated. A receiver operating characteristic (ROC) curve and confusion matrix were employed to evaluate the models' predictive performance.

The remainder of this paper is organized as follows. Section 2 introduces the methodological background of the selected models, followed by a description of the dataset in Section 3. Section 4 presents the results and findings. We summarize our conclusions and propose future work in Section 5.

2. Methodology

2.1. Bayesian Multinomial Logit Model. MNL regression generalizes logistic regression into multiclass problems that consist of more than two possible discrete groups [1, 19]. It can be expressed as [19]

$$P\left(Z_j = i\right) = \frac{\exp\left(\beta_0^i + \beta_1^i x_{j1} + \beta_2^i x_{j2} + \cdots + \beta_k^i x_{jk}\right)}{\sum_{i=1}^{I} \exp\left(\beta_0^i + \beta_1^i x_{j1} + \beta_2^i x_{j2} + \cdots + \beta_k^i x_{jk}\right)}, \tag{1}$$

where $\mathbf{X} = [x_{j1}, x_{j2}, \ldots, x_{jk}]$ is a vector of independent variables x_{jk}, $\beta = [\beta_0^i, \beta_1^i, \beta_2^i, \ldots, \beta_k^i]^T$ is the corresponding coefficient vector, and $Z_j = i$ is the choice of travel mode i for the jth observation.

The likelihood function can be expressed as

$$\begin{aligned} f(\mathbf{Z} \mid \boldsymbol{\beta}) &= \prod_{j=1}^{N} \prod_{i=1}^{I} \left[\varepsilon_{ji} \times P\left(Z_j = i\right)\right] \\ &= \prod_{j=1}^{N} \prod_{i=1}^{I} \left[\varepsilon_{ji} \times \frac{\exp\left(\beta_0^i + \beta_1^i x_{j1} + \beta_2^i x_{j2} + \cdots + \beta_k^i x_{jk}\right)}{\sum_{m=1}^{I} \exp\left(\beta_0^i + \beta_1^i x_{j1} + \beta_2^i x_{j2} + \cdots + \beta_k^i x_{jk}\right)}\right], \end{aligned} \tag{2}$$

where N is the number of samples, I is the number of outcomes, and ε_{ij} equals 1 when the discrete outcome of sample j is i and is 0 otherwise.

The Bayesian approach using Markov chain Monte Carlo (MCMC) was utilized for model estimation. Based on Bayesian inference, the posterior joint distribution of parameters β conditional on dataset $\mathbf{Z}$ can be estimated as [19]

$$f(\boldsymbol{\beta} \mid \mathbf{Z}) = \frac{f(\mathbf{Z}, \boldsymbol{\beta})}{f(\mathbf{Z})} = \frac{f(\mathbf{Z} \mid \boldsymbol{\beta}) \pi(\boldsymbol{\beta})}{f(\mathbf{Z} \mid \boldsymbol{\beta}) \pi(\boldsymbol{\beta}) d(\boldsymbol{\beta})} \propto f(\mathbf{Z} \mid \boldsymbol{\beta}) \pi(\boldsymbol{\beta}), \tag{3}$$

where $f(\mathbf{Z}, \beta)$ is the joint probability distribution of $\mathbf{Z}$ and β, $f(\mathbf{Z} \mid \beta)$ is the likelihood of the conditional function based on β, and $\pi(\beta)$ is the prior distribution of β. Due to lack of information on the random parameters, we used the non-informative prior distributions [1]:

$$\boldsymbol{\beta} \sim N\left(0_k, 10^6 M_k\right), \tag{4}$$

where 0_k is a vector of zeros and M_k is the $k \times k$ identity matrix.

The posterior joint distribution can be derived as [42]

$$f(\boldsymbol{\beta} \mid \mathbf{Z}) \propto f(\mathbf{Z} \mid \boldsymbol{\beta})\pi(\boldsymbol{\beta}) = \prod_{j=1}^{N}\prod_{i=1}^{I}\left[\varepsilon_{ji} \times \frac{\exp\left(\beta_0^i + \beta_1^i x_{j1} + \beta_2^i x_{j2} + \cdots + \beta_k^i x_{jk}\right)}{\sum_{m=1}^{I}\exp\left(\beta_0^i + \beta_1^i x_{j1} + \beta_2^i x_{j2} + \cdots + \beta_k^i x_{jk}\right)}\right]$$
$$\times \prod_{j=1}^{N}\prod_{i=1}^{I}\left[\frac{1}{\sqrt{2\pi}\,10^3}\exp\left(-\frac{1}{2}\frac{\left(\beta_k^i\right)^2}{10^6}\right)\right] \propto \exp\left\{\sum_{j=1}^{N}\sum_{i=1}^{I}\left[\varepsilon_{ij} \times \frac{\exp\left(\beta_0^i + \beta_1^i x_{j1} + \beta_2^i x_{j2} + \cdots + \beta_k^i x_{jk}\right)}{\sum_{m=1}^{I}\exp\left(\beta_0^i + \beta_1^i x_{j1} + \beta_2^i x_{j2} + \cdots + \beta_k^i x_{jk}\right)}\right] - \sum_{j=1}^{N}\sum_{i=1}^{I}\left[\frac{1}{2}\frac{\left(\beta_k^i\right)^2}{10^6}\right]\right\}. \tag{5}$$

2.2. Radial Basis Function (RBF) Neural Network. An RBF neural network is a typical three-layer neural network model with input, hidden, and output layers, as shown in Figure 1, where k is the number of input variables, H is the number of hidden neurons, I is the number of output neurons (travel modes), $\mathbf{X} = [x_1, x_2, \ldots, x_k]^T$ is the input, $\mathbf{Y} = [y_1, y_2, \ldots, y_I]^T$ is the output, and w_{hi} is the connection weight of the hth hidden layer neuron to the ith output layer neuron.

A Gaussian function is generally used as the hidden layer excitation function. The output of the hth hidden layer neuron is

$$G_h(x) = e^{\left(-X - c_h^2/2\sigma_h^2\right)}, \quad h = 1, 2, \ldots, H, \tag{6}$$

and the linear mapping relationship between $G_h(x)$ and the ith output layer neuron is

$$G_h(x) = y_i = \sum_{h=1}^{H} w_{hi} G_h(x), \quad i = 1, 2, \ldots, I, \tag{7}$$

where c_h and σ_h are, respectively, the center vector of the Gaussian function and the base width of the hth hidden neuron.

RBF has been criticized as a "black box" that lacks interpretability [43]. Various tools have been developed to address this issue, the most common being variable importance analysis [44–46], which measures the relative importance of each independent variable in predicting dependent variables.

2.3. Multilayer Perceptron (MLP) Neural Network. MLP is a commonly used supervised ANN model that can be used for both pattern recognition and function approximation. Compared to RBF, MLP can have multiple hidden layers (shown in Figure 2) [47]. The hyperbolic tangent function is selected as the activation function of MLP hidden neurons. The output from a hidden neuron is

$$y = \frac{e^u - e^{-u}}{e^u + e^{-u}}, \tag{8}$$

and the connection weight is the output of the net function,

$$u = b + \sum_{p=1}^{k} w_p x_p, \tag{9}$$

where k is the number of inputs, x_p is the input, w_p is the weight of the corresponding input (w_p; $1 \le p \le k$), b is the bias weight, and the Levenberg–Marquardt training algorithm is selected [28].

2.4. Model Comparison and Validation. The multi-classification confusion matrix (see Table 1) is used to calculate the accuracy of each model [48], where s_{im} is the number of samples in which mode i is predicted as mode m. The recall and precision of mode i are

$$\text{Recall}_i = \frac{S_{ii}}{\sum_{m=1}^{I} s_{im}}, \qquad \text{Precision}_i = \frac{S_{ii}}{\sum_{m=1}^{I} S_{mi}}, \tag{10}$$

and the accuracy of the model can be calculated as [1]

$$\text{accuracy} = \frac{\sum_{i=1}^{I} S_{ii}}{N}. \tag{11}$$

The ROC curve and area under the curve (AUC) were also used to measure the predictive ability. A higher AUC value indicates better predictive accuracy [42, 49].

3. Data Collection

Data from Li et al.'s work [42] were used in this study. A total of 985 random samples collected in Xi'an from March 1–10, 2018, were used for analysis, where 161 samples reported the choice of airplane, accounting for 16.3% of intercity travel records, and 369 (37.5%) were reported as HSR, 299 (30.4%) as train, and 156 (15.8%) as express bus. Among them, 80% were randomly selected for training, and the remaining 20% were used for prediction. In addition to the original information included in the database, the travel distance was calculated by Baidu Maps using the real route between the cities of origin and destination. The intercity travel time was obtained according to the identification number of the carrier, transportation schedule, and origin and destination cities.

Undersampling and oversampling are the most frequently used techniques to balance data for machine learning and pattern recognition [38–41]. Undersampling

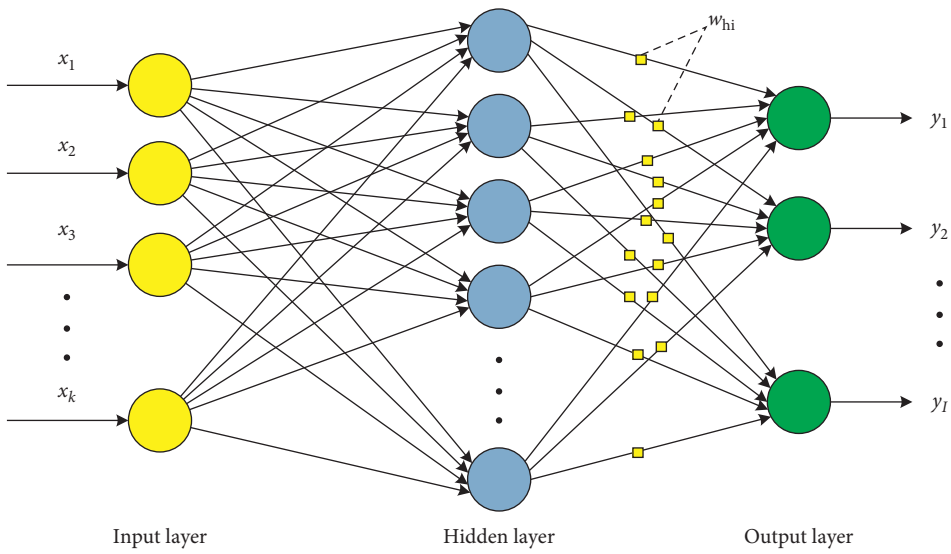


Figure 1: RBF network.

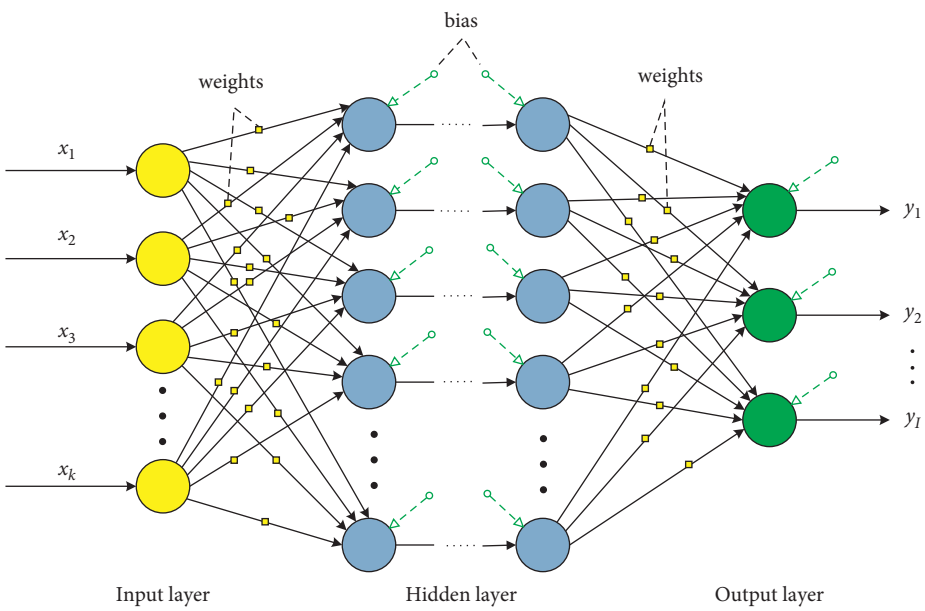


Figure 2: General topology of MLP.

Table 1: Multiclassification confusion matrix.

		Predictive class					
	Mode	1	2	3	...	I	Recall
Actual class	1	s_{11}	s_{12}	s_{13}	...	s_{1I}	Recall_1
	2	s_{21}	s_{22}	s_{23}	...	s_{2I}	Recall_2
	3	s_{31}	s_{32}	s_{33}	...	s_{3I}	Recall_3
	...	...	...	...	...	...	
	I	s_{I1}	s_{I2}	s_{I3}		s_{II}	Recall_I
	Precision	Precision_1	Precision_2	Precision_3		Precision_I	Accuracy

achieves relative equilibrium among classes by reducing the number of samples of classes with more samples. Using this method, the number of samples of each travel mode was 156, with 80% randomly selected for training, and the remaining 20% selected for prediction. Oversampling is to add samples of classes with fewer samples to equal the number of samples

in a class with more samples. Through oversampling, the sample size of each transportation mode became 369; 80% of samples were randomly selected for training, and the remaining 20% were selected for prediction.

Tables 2 and 3 describe the categories and continuous variables for imbalanced and rebalanced data, respectively.

4. Results

Stata 15.0 software was used for parameter estimation of the BMNL and MNL models, confusion matrix, and ROCs. SPSS 25.0 was used for relative importance analysis of variables by the RBF and MLP models.

4.1. Model Results. Table 4 presents the estimated means of variables from the BMNL model, and Tables 5–7 show their parameter estimates. The frequently used train was considered as the reference in the model. The typical variables including gender, age, occupation, travel purpose, monthly income, intercity travel distance, intercity travel cost, intercity travel time, safety, comfort, punctuality, access time, and departure time were selected for modeling after collinearity testing. The MCMC simulation-based full Bayesian approach was employed to estimate the posterior distributions of parameters. Variables with confidence intervals not including zero were regarded as significant [19]. As shown in Table 4, we found that the parameter estimates of certain variables differed slightly between the imbalanced and balanced data. For example, the intercity travel distance is significantly related to the choice of express bus when using balanced data, but not when using imbalanced data. The signs of variables were found to be consistent between balanced and imbalanced data.

Table 8 shows the estimated coefficients of variables from the MNL model using the same variables. Parameter estimates of variables are shown in Tables 9–11. Similar to the BMNL model, the parameter estimates differ to some extent between imbalanced and balanced data, and the signs of significant variables are consistent. The symbols of significant variables in the MNL model were consistent with those in the BMNL model. However, the significant variables in MNL were not completely consistent with those in the BMNL model. For example, the travel purpose is significant in the BMNL model but not in the MNL model. Gender was significantly related to the choice of HSR in the BMNL model, but not in the MNL model.

Figures 3 and 4 show the relative importance of the factors obtained by RBF and MLP, respectively. There is a slight difference in the order of relative importance of factors. For example, using imbalanced data, intercity travel cost is most important in the RBF model, but second in importance in the MLP model, after intercity travel time. Slight differences exist in the relative importance of factors between imbalanced and balanced data in the same model. For example, in the MLP model, gender is the least important using imbalanced data, and travel purpose is the least important using balanced data. Overall, intercity travel cost, intercity travel time, intercity travel distance, comfort, safety, and punctuality are the most important factors in the intercity travel mode choice, followed by the monthly income, age, and occupation. Access and departure times, which reflect the accessibility of a transport hub, show moderate importance. Travel purpose and gender are the least important.

4.2. Model Comparison and Validation

4.2.1. Model Performance for Imbalanced Data. AUCs and confusion matrices were employed to compare the fitting and predictive performance of the MNL, BMNL, MLP, and RBF models. The confusion matrix of the four models using imbalanced data is shown in Table 12. Through the analysis of the accuracy, it can be found that MLP has the best fitting performance (80.70%), and RBF is the worst (67.30%). BMNL (76.36%) and MNL (76.10%) have similar fitting performance. For the predictive set, the predictive performance of MLP (78.70%) is the best, followed by MNL (75.76%), BMNL (75.25%), and RBF (65.50%).

The ROC curves of the four models are shown in Figures 5 and 6. For the training set, the AUC of the MLP for the airplane is 0.9857, which indicates that its fitting performance is better than that of BMNL (0.9732), MNL (0.9731), and RBF (0.9443). The MLP model is almost perfect, as its ROC curve rises rapidly toward the upper-left corner of the graph. Similarly, the AUCs of MLP for HSR and train are the largest, followed by BMNL, MNL, and RBF. These findings confirm that the MLP model outperforms BMNL and MNL, followed by RBF.

For the predictive set, the AUC of MLP for airplane is 0.9905, which is better than RBF (0.9823), BMNL (0.9784), and MNL (0.9767). Similarly, MLP is almost perfect, as its ROC curve rises rapidly toward the upper-left corner of the graph. The AUC of MLP for HSR is also the largest, followed by BMNL, MNL, and RBF. The AUC of MLP for train is 0.9054, which indicates that its predictive performance is significantly better than that of MNL (0.8637), BMNL (0.8624), and RBF (0.8280). However, the AUC of BMNL for express bus is the largest, followed by MNL, MLP, and RBF.

4.2.2. Model Performance for Rebalanced Data. MNL, BMNL, MLP, and RBF were used to train and verify the balanced data with the same variables used for the imbalanced data. The confusion matrices of each model for undersampled and oversampled balanced data are shown in Tables 13 and 14. The ROC curves for the undersampled training and predictive set are presented in Figures 7 and 8. The ROC curves for the oversampled training and predictive sets are presented in Figures 9 and 10.

The results show that MLP provides the best fitting for both oversampled and undersampled data, followed by BMNL and MNL, and RBF has the poorest fitting performance. The results are consistent with those of the four models using imbalanced data. Hence, whether the data are balanced will not affect the relative fitting performance of the models.

Table 2: Description of categorical variables.

Variable	Description	Value	Imbalanced data				Oversampling balanced data				Undersampling balanced data			
			Training set		Predictive set		Training set		Predictive set		Training set		Predictive set	
			Frequency	Proportion (%)	Frequency	Proportion (%)	Frequency	Proportion (%)	Frequency	Proportion (%)	Frequency	Proportion (%)	Frequency	Proportion (%)
Dependent														
Travel modes	Airplane	1	129	16.39	32	16.16	295	25	74	25	125	25	31	25
	HSR	2	295	37.48	74	37.37	295	25	74	25	125	25	31	25
	Train	3	239	30.37	60	30.31	295	25	74	25	125	25	31	25
	Express bus	4	124	15.76	32	16.16	295	25	74	25	125	25	31	25
Independent														
Gender	Female	0	335	42.57	81	40.91	475	40.25	125	42.23	213	42.6	58	46.77
	Male	1	452	57.43	117	59.09	705	59.75	171	57.77	287	57.4	66	53.23
Age	<19	1	18	2.29	5	2.53	26	2.2	4	1.35	9	1.8	1	0.81
	20–29	2	324	41.17	83	41.92	501	42.46	131	44.26	211	42.2	54	43.55
	30–39	3	261	33.16	58	29.29	377	31.95	83	28.04	157	31.4	37	29.84
	40–49	4	116	14.74	38	19.19	191	16.19	43	14.53	83	16.6	23	18.55
	50–59	5	54	6.86	11	5.56	63	5.34	28	9.46	30	6	7	5.65
	60 and above	6	14	1.78	3	1.51	22	1.86	7	2.36	10	2	2	1.60
Occupation	Enterprise unit	1	168	21.35	45	22.73	245	20.76	66	22.3	102	20.4	29	23.39
	Personnel of institutions	2	134	17.03	40	20.20	197	16.69	48	16.22	94	18.8	19	15.32
	Student	3	225	28.59	53	26.77	366	31.02	92	31.08	152	30.4	33	26.61
	Farmer	4	50	6.35	7	3.54	65	5.51	18	6.08	28	5.6	10	8.06
	Self-employed	5	113	14.36	30	15.15	171	14.49	37	12.5	67	13.4	21	16.94
	Other	6	97	12.33	23	11.61	136	11.53	35	11.82	57	11.4	12	9.68
Monthly income	<3K yuan	1	249	31.64	59	29.80	393	33.31	99	33.45	166	33.2	37	29.84
	3-4K yuan	2	147	18.68	41	20.71	198	16.78	53	17.91	98	19.6	20	16.13
	4-5K yuan	3	200	25.41	54	27.27	301	25.51	64	21.62	123	24.6	37	29.84
	5-6K yuan	4	112	14.23	29	14.65	177	15	50	16.89	68	13.6	21	16.94
	6-7K yuan	5	29	3.68	4	2.02	38	3.22	6	2.03	11	2.2	2	1.61
	>7K yuan	6	50	6.35	11	5.55	73	6.19	24	8.11	34	6.8	7	5.65
Travel purpose	Mandatory	1	380	48.28	94	47.47	584	49.49	159	53.72	251	50.2	66	53.23
	Leisure	0	407	51.72	104	52.53	596	50.51	137	46.28	249	49.8	58	46.77
Access mode	Public transit	1	548	69.63	129	65.15	808	68.47	200	67.57	358	71.6	74	59.68
	Private car or taxi	0	239	30.37	69	34.85	372	31.53	96	32.43	142	28.4	50	40.32
Access time	0–30 min	1	252	32.02	66	33.33	383	32.46	104	35.14	166	33.2	44	35.48
	30–60 min	2	283	35.96	80	40.41	414	35.08	97	32.77	165	33	42	33.87
	60–90 min	3	252	32.02	52	26.26	383	32.46	95	32.09	169	33.8	38	30.65
Safety	Very unsafe	1	0	0	0	0	0	0	0	0	0	0	0	0
	Unsafe	2	8	1.02	2	1.01	17	1.44	3	1.01	6	1.2	2	1.61
	General	3	145	18.42	30	15.15	216	18.31	67	22.64	92	18.4	31	25
	Safe	4	390	49.56	105	53.03	594	50.34	144	48.65	251	50.2	55	44.35
	Very safe	5	244	31	61	30.81	353	29.91	82	27.7	151	30.2	36	29.03

Table 2: Continued.

Variable	Description	Value	Imbalanced data				Oversampling balanced data				Undersampling balanced data			
			Training set		Predictive set		Training set		Predictive set		Training set		Predictive set	
			Frequency	Proportion (%)	Frequency	Proportion (%)	Frequency	Proportion (%)	Frequency	Proportion (%)	Frequency	Proportion (%)	Frequency	Proportion (%)
Comfort	Very uncomfortable	1	5	0.64	3	1.52	15	1.27	2	0.68	6	1.2	2	1.61
	Uncomfortable	2	40	5.08	12	6.06	70	5.93	27	9.12	32	6.4	10	8.06
	General	3	198	25.16	47	23.74	301	25.51	74	25	119	23.8	32	25.81
	Comfortable	4	411	52.22	105	53.03	609	51.61	140	47.3	262	52.4	59	47.58
	Very comfortable	5	133	16.9	31	15.65	185	15.68	53	17.91	81	16.2	21	16.94
Punctuality	Very unpunctual	1	0	0	0	0	0	0	0	0	0	0	0	0
	Unpunctual	2	30	3.81	9	4.54	46	3.9	22	7.43	24	4.8	5	4.03
	General	3	204	25.92	47	23.74	340	28.81	76	25.68	141	28.2	37	29.84
	Punctual	4	414	52.6	111	56.06	624	52.88	160	54.05	248	49.6	70	56.45
	Very punctual	5	139	17.66	31	15.66	170	14.41	38	12.84	87	17.4	12	9.68
Departure mode	Public transit	1	571	72.55	142	71.72	816	69.15	220	74.32	356	71.2	90	72.58
	Private car or taxi	0	216	27.45	56	28.28	364	30.85	76	25.68	144	28.8	34	27.42
Departure time	0–30 min	1	387	49.24	102	51.52	594	50.42	134	45.27	260	52.1	45	36.29
	30–60 min	2	254	32.31	68	34.34	369	31.32	100	33.78	151	30.26	51	41.13
	60–90 min	3	145	18.45	28	14.14	215	18.25	62	20.95	88	17.64	28	22.58

Table 3: Description of continuous variables.

Data	Variable	Unit	Training set				Predictive set			
			Min	Max	Mean	SD	Min	Max	Mean	SD
Imbalanced data	Intercity travel distance	km	16.00	2831.00	797.54	579.86	16.00	2540.00	792.66	632.00
	Intercity travel cost	Yuan	7.00	2600.00	310.46	319.44	7.00	1400.00	283.07	288.70
	Intercity travel time	Hour	0.22	52.00	6.48	6.49	0.25	33.00	6.00	5.87
	Access cost	Yuan	1.00	300.00	11.59	22.89	1.00	100.00	10.19	16.36
	Departure cost	Yuan	1.00	150.00	13.52	20.72	1.00	150.00	13.23	21.89
Oversampling balanced data	Intercity travel distance	Km	16.00	2831.00	807.54	621.35	16.00	2801.00	808.45	617.20
	Intercity travel cost	Yuan	7.00	2600.00	321.23	343.35	7.00	2600.00	324.57	344.48
	Intercity travel time	Hour	0.22	52.00	5.92	6.33	0.25	37.00	5.85	6.50
	Access cost	Yuan	1.00	300.00	13.03	25.51	1.00	150.00	11.57	20.38
	Departure cost	Yuan	1.00	150.00	15.19	23.36	1.00	150.00	14.24	22.69
Undersampling balanced data	Intercity travel distance	Km	17.00	2831.00	823.55	616.21	27.00	2500.00	731.52	597.32
	Intercity travel cost	Yuan	7.00	2600.00	330.79	360.54	13.50	1200.00	277.20	275.22
	Intercity travel time	Hour	0.22	37.00	5.92	6.14	0.25	28.00	5.22	5.90
	Access cost	Yuan	1.00	300.00	12.77	25.43	1.00	120.00	12.49	19.88
	Departure cost	Yuan	1.00	150.00	14.68	23.11	1.00	120.00	14.51	22.79

Table 4: Parameter estimation in BMNL.

Variable	Imbalanced data			Oversampling balanced data			Undersampling balanced data		
	Airplane Mean	HSR Mean	Express bus Mean	Airplane Mean	HSR Mean	Express bus Mean	Airplane Mean	HSR Mean	Express bus Mean
Gender									
Male vs. female	0.555	0.186	0.368	0.683	0.197	0.713	1.231	0.384	0.733
Age	0.282	0.350	0.247	0.416	0.357	0.346	0.396	—	0.380
Occupation									
Personnel of institutions vs. enterprise unit	0.710	0.408	0.651	0.344	0.143	0.313	—	—	—
Student vs. enterprise unit	—	−0.300	1.348	0.316	−0.349	1.930	—	—	1.454
Farmer vs. enterprise unit	−0.473	—	—	−0.898	−0.561	−0.333	−2.097	−0.584	−0.574
Self-employed vs. enterprise unit	—	—	0.686	−1.157	−0.384	0.558	−2.540	−1.114	—
Others vs. enterprise unit	−0.844	−0.193	0.679	−1.480	−0.475	0.616	−2.160	−0.704	0.473
Monthly income	—	−0.273	—	—	−0.221	—	−0.163	−0.222	—
Travel purpose									
Mandatory travel vs. leisure travel	−0.477	−0.284	−0.425	−0.341	−0.388	—	—	−0.230	0.138
Intercity travel distance	0.004	0.001	—	0.002	—	−0.002	0.003	—	−0.002
Intercity travel cost	0.018	0.015	—	0.027	0.022	0.001	0.026	0.023	—
Intercity travel time	−1.239	−0.397	—	−1.265	−0.417	0.002	−1.255	−0.482	0.036
Safety	0.566	0.609	—	0.818	0.685	—	−0.098	0.621	—
Comfort	—	0.314	−0.433	0.195	0.369	−0.463	0.639	0.312	−0.480
Punctuality	−0.335	0.319	−0.574	−0.377	0.317	−0.525	−0.496	0.323	−0.817
Access time	0.390	—	−0.268	0.455	—	−0.284	0.351	−0.220	—
Departure time	0.694	—	—	1.026	0.264	—	0.796	—	−0.427
Constant	−7.703	−5.803	3.089	−9.503	−6.798	2.328	−5.985	−5.589	3.884

Note: parameters that were significant at the 95% confidence level are shown in the table.

For the predictive performance of the models, we found that MLP performs best regardless of oversampling or undersampling balanced data. More importantly, BMNL and MNL show the same predictive performance when using oversampling balanced data, and RBF models have the worst predictive performance. Similarly, BMNL and MNL have the same predictive performance using undersampled data, but their performance is lower than that of RBF.

The fitting performance of models based on balanced data is a slight improvement over using imbalanced data. For example, the fitting performance of MLP model is 80.70% using imbalanced data, and 81.80% and 83.10%, respectively, with undersampled and oversampled data. However, except for the RBF model, the predictive performance of these models based on balanced data is slightly lower than that when using imbalanced data.

The ROC curve was also used to intuitively judge the predictive performance of each model, and AUCs were used to quantitatively compare their predictive accuracy under different modeling techniques. We found that the results

Table 5: Parameter estimation in BMNL using imbalanced data.

Variable	Airplane				HSR				Express bus			
	Mean	SD	Credit interval		Mean	SD	Credit interval		Mean	SD	Credit interval	
			2.50%	97.50%			2.50%	97.50%			2.50%	97.50%
Gender												
Male vs. female	0.555	0.143	0.275	0.833	0.186	0.057	0.067	0.295	0.368	0.167	0.040	0.677
Age	0.282	0.089	0.114	0.454	0.350	0.091	0.180	0.538	0.247	0.084	0.072	0.407
Occupation												
Personnel of institution vs. enterprise unit	0.710	0.173	0.377	1.051	0.408	0.059	0.295	0.527	0.651	0.129	0.403	0.913
Student vs. enterprise unit	0.114	0.171	−0.217	0.435	−0.300	0.068	−0.440	−0.167	1.348	0.181	1.002	1.727
Farmer vs. enterprise unit	−0.473	0.226	−0.930	−0.027	−0.078	0.063	−0.205	0.423	−0.031	0.158	−0.321	0.288
Self-employed vs. enterprise unit	−0.147	0.187	−0.522	0.228	0.074	0.082	−0.083	0.235	0.686	0.139	0.409	0.967
Others vs. enterprise unit	−0.844	0.161	−1.173	−0.526	−0.193	0.066	−0.316	−0.059	0.679	0.144	0.396	0.953
Monthly income	−0.098	0.090	−0.270	0.089	−0.273	0.062	−0.394	−0.154	−0.121	0.070	−0.251	0.016
Travel purpose												
Mandatory travel vs. leisure travel	−0.477	0.196	−0.871	−0.096	−0.284	0.053	−0.384	−0.185	−0.425	0.130	−0.699	−0.192
Intercity travel distance	0.004	0.001	0.003	0.005	0.001	0.001	0.001	0.002	−0.001	0.001	−0.002	0.001
Intercity travel cost	0.018	0.002	0.014	0.022	0.015	0.002	0.011	0.186	0.001	0.002	−0.005	0.004
Intercity travel time	−1.239	0.096	−1.433	−1.059	−0.397	0.044	−0.485	−0.311	−0.009	0.029	−0.064	0.048
Safety	0.566	0.149	0.268	0.847	0.609	0.086	0.440	0.782	0.039	0.065	−0.079	0.173
Comfort	0.245	0.124	−0.017	0.474	0.314	0.053	0.211	0.427	−0.433	0.099	−0.634	−0.254
Punctuality	−0.335	0.13	−0.586	−0.089	0.319	0.066	0.189	0.440	−0.574	0.04	−0.655	−0.498
Access time	0.390	0.151	0.074	0.677	−0.08	0.056	−0.189	0.034	−0.268	0.106	−0.472	−0.066
Departure time	0.694	0.142	0.428	0.981	0.029	0.060	−0.080	0.151	−0.161	0.132	−0.419	0.094
Constant	−7.703	0.081	−7.863	−7.537	−5.803	0.098	−5.997	−5.62	3.089	0.124	2.857	3.342

Table 6: Parameter estimation in BMNL using oversampling of balanced data.

Variable	Airplane				HSR				Express bus			
	Mean	SD	Credit interval		Mean	SD	Credit interval		Mean	SD	Credit interval	
			2.50%	97.50%			2.50%	97.50%			2.50%	97.50%
Gender												
Male vs. female	0.683	0.118	0.474	0.947	0.197	0.075	0.052	0.339	0.713	0.085	0.556	0.872
Age	0.416	0.106	0.190	0.615	0.357	0.028	0.298	0.410	0.346	0.054	0.241	0.452
Occupation												
Personnel of institution vs. enterprise unit	0.344	0.076	0.218	0.510	0.143	0.054	0.041	0.256	0.313	0.107	0.050	0.491
Student vs. enterprise unit	0.316	0.078	0.126	0.436	−0.349	0.050	−0.444	−0.256	1.930	0.060	1.817	2.041
Farmer vs. enterprise unit	−0.898	0.058	−1.014	−0.787	−0.561	0.077	−0.712	−0.417	−0.333	0.054	−0.438	−0.229
Self-employed vs. enterprise unit	−1.157	0.120	−1.398	−0.928	−0.384	0.075	−0.531	−0.232	0.558	0.142	0.269	0.818
Others vs. enterprise unit	−1.480	0.072	−1.643	−1.348	−0.475	0.116	−0.695	−0.249	0.616	0.101	0.423	0.794
Monthly income	−0.044	0.061	−0.192	0.058	−0.221	0.047	−0.315	−0.129	0.015	0.052	−0.079	0.128
Travel purpose												
Mandatory travel vs. leisure travel	−0.341	0.033	−0.396	−0.264	−0.388	0.086	−0.556	−0.216	−0.127	0.064	−0.243	0.004
Intercity travel distance	0.002	0.001	0.001	0.004	−0.001	0.001	−0.002	0.001	−0.002	0.001	−0.003	−0.001
Intercity travel cost	0.027	0.002	0.023	0.031	0.022	0.002	0.018	0.026	0.001	0.002	−0.002	0.005
Intercity travel time	−1.265	0.075	−1.426	−1.128	−0.417	0.022	−0.459	−0.373	0.002	0.024	−0.047	0.045
Safety	0.818	0.069	0.671	0.947	0.685	0.047	0.594	0.780	0.155	0.104	−0.042	0.359
Comfort	0.195	0.120	0.007	0.490	0.369	0.033	0.303	0.434	−0.463	0.094	−0.644	−0.284
Punctuality	−0.377	0.116	−0.602	−0.164	0.317	0.041	0.242	0.399	−0.525	0.080	−0.680	−0.370
Access time	0.455	0.081	0.255	0.587	−0.136	0.110	−0.357	0.066	−0.284	0.058	−0.409	−0.175
Departure time	1.026	0.171	0.717	1.356	0.264	0.033	0.201	0.329	−0.202	0.111	−0.418	0.012
Constant	−9.503	0.080	−9.678	−9.362	−6.798	0.092	−6.976	−6.605	2.328	0.067	2.164	2.433

Table 7: Parameter estimation in BMNL using undersampling of balanced data.

Variable	Airplane				HSR				Express bus			
	Mean	SD	Credit interval		Mean	SD	Credit interval		Mean	SD	Credit interval	
			2.50%	97.50%			2.50%	97.50%			2.50%	97.50%
Gender												
Male vs. female	1.231	0.266	0.719	1.764	0.384	0.114	0.167	0.620	0.733	0.165	0.435	1.075
Age	0.396	0.102	0.206	0.595	0.232	0.143	−0.080	0.506	0.380	0.083	0.220	0.541
Occupation												
Personnel of institution vs. enterprise unit	−0.162	0.130	−0.403	0.081	−0.176	0.106	−0.393	0.035	0.139	0.171	−0.235	0.444
Student vs. enterprise unit	0.041	0.170	−0.274	0.366	−0.121	0.210	−0.529	0.288	1.454	0.131	1.198	1.718
Farmer vs. enterprise unit	−2.097	0.191	−2.450	−1.722	−0.584	0.061	−0.709	−0.458	−0.574	0.223	−1.017	−0.128
Self-employed vs. enterprise unit	−2.540	0.232	−3.011	−2.105	−1.114	0.081	−1.280	−0.954	0.021	0.231	−0.449	0.485
Others vs. enterprise unit	−2.160	0.211	−2.563	−1.735	−0.704	0.085	−0.906	−0.574	0.473	0.134	0.217	0.740
Monthly income	−0.163	0.067	−0.305	−0.031	−0.222	0.101	−0.418	−0.027	−0.128	0.075	−0.284	0.020
Travel purpose												
Mandatory travel vs. leisure travel	−0.038	0.215	−0.470	0.352	−0.230	0.113	−0.471	−0.013	0.138	0.205	−0.253	0.554
Intercity travel distance	0.003	0.001	0.001	0.005	0.000	0.001	−0.002	0.002	−0.002	0.001	−0.004	−0.001
Intercity travel cost	0.026	0.004	0.020	0.035	0.023	0.003	0.016	0.031	0.000	0.003	−0.006	0.006
Intercity travel time	−1.255	0.062	−1.378	−1.146	−0.482	0.056	−0.598	−0.388	0.036	0.043	−0.049	0.126
Safety	−0.098	0.133	−0.357	0.151	0.621	0.099	0.432	0.824	0.077	0.096	−0.113	0.274
Comfort	0.639	0.035	0.574	0.706	0.312	0.120	0.056	0.526	−0.480	0.134	−0.753	−0.223
Punctuality	−0.496	0.158	−0.800	−0.190	0.323	0.018	0.290	0.363	−0.817	0.118	−1.065	−0.582
Access time	0.351	0.073	0.216	0.490	−0.220	0.030	−0.278	−0.159	0.042	0.070	−0.098	0.178
Departure time	0.796	0.109	0.584	1.029	−0.050	0.063	−0.178	0.064	−0.427	0.115	−0.641	−0.200
Constant	−5.985	0.135	−6.242	−5.731	−5.589	0.207	−6.001	−5.200	3.884	0.173	3.540	4.203

Table 8: Parameter estimation in MNL.

Variable	Imbalanced data			Oversampling balanced data			Undersampling balanced data		
	Airplane	HSR	Express bus	Airplane	HSR	Express bus	Airplane	HSR	Express bus
	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Gender									
Male vs. female	—	—	—	—	—	0.681	—	—	0.723
Age	—	0.349	—	—	0.366	0.360	—	—	—
Occupation									
Personnel of institutions vs. enterprise unit	—	—	—	—	—	—	—	—	—
Student vs. enterprise unit	—	—	1.298	—	−0.352	1.856	—	—	1.642
Farmer vs. enterprise unit	—	—	—	—	—	—	—	—	—
Self-employed vs. enterprise unit	—	—	—	—	—	—	—	—	—
Others vs. enterprise unit	—	—	—	—	—	—	—	—	—
Monthly income	−0.128	−0.289	—	—	−0.226	—	—	—	—
Travel purpose									
Mandatory travel vs. leisure travel	—	—	—	—	—	—	—	—	—
Intercity travel distance	0.004	0.001	—	0.002	—	−0.002	0.003	—	—
Intercity travel cost	0.017	0.014	—	0.026	0.022	—	0.031	0.027	—
Intercity travel time	−1.211	−0.380	—	−1.242	−0.421	—	−1.395	−0.638	—
Safety	—	0.585	—	0.801	0.678	—	—	—	—
Comfort	—	—	−0.458	—	—	−0.461	—	—	−0.418
Punctuality	—	0.299	−0.575	—	—	−0.513	—	—	−0.617
Access time	—	—	—	—	—	—	—	—	—
Departure time	0.731	—	—	1.030	—	−0.185	—	—	—
Constant	−7.673	−5.821	3.058	−9.522	−6.798	2.303	—	−5.297	3.420

Note: parameters that were significant at the 95% confidence level are shown in the table.

Table 9: Parameter estimation in MNL using imbalanced data.

Variable	Airplane				HSR				Express bus			
	Coefficient	$P > z$	Credit interval		Coefficient	$P > z$	Credit interval		Coefficient	$P > z$	Credit interval	
			2.50%	97.50%			2.50%	97.50%			2.50%	97.50%
Gender												
Male vs. female	0.474	0.321	−0.462	1.410	0.160	0.543	−0.355	0.674	0.409	0.121	−0.107	0.926
Age	0.326	0.269	−0.252	0.904	0.349	0.035	0.024	0.675	0.253	0.128	−0.073	0.579
Occupation												
Personnel of institution vs. enterprise unit	0.725	0.290	−0.618	2.069	0.410	0.316	−0.391	1.210	0.642	0.191	−0.320	1.604
Student vs. enterprise unit	0.116	0.885	−1.453	1.685	−0.362	0.390	−1.189	0.464	1.298	0.002	0.464	2.132
Farmer vs. enterprise unit	−0.425	0.658	−2.307	1.458	−0.080	0.892	−1.234	1.074	−0.059	0.929	−1.359	1.241
Self-employed vs. enterprise unit	−0.207	0.789	−1.720	1.306	0.081	0.849	−0.754	0.916	0.611	0.195	−0.314	1.537
Others vs. enterprise unit	−0.824	0.366	−2.610	0.962	−0.173	0.732	−1.161	0.816	0.705	0.178	−0.32	1.730
Monthly income	−0.128	0.565	−0.563	0.307	−0.289	0.013	−0.516	−0.062	−0.106	0.329	−0.318	0.106
Travel purpose												
Mandatory travel vs. leisure travel	−0.443	0.340	−1.352	0.466	−0.302	0.256	−0.823	0.219	−0.412	0.112	−0.921	0.096
Intercity travel distance	0.004	0.001	0.002	0.005	0.001	0.278	−0.001	0.002	−0.001	0.009	−0.003	0.001
Intercity travel cost	0.017	0.001	0.013	0.021	0.014	0.001	0.010	0.018	0.001	0.828	−0.005	0.004
Intercity travel time	−1.211	0.001	−1.444	−0.978	−0.380	0.001	−0.483	−0.276	−0.003	0.917	−0.061	0.055
Safety	0.491	0.137	−0.156	1.138	0.585	0.004	0.188	0.982	0.059	0.760	−0.319	0.437
Comfort	0.251	0.410	−0.345	0.846	0.358	0.056	−0.01	0.725	−0.458	0.011	−0.811	−0.105
Punctuality	−0.278	0.356	−0.866	0.311	0.299	0.104	−0.061	0.660	−0.575	0.002	−0.936	−0.213
Access time	0.404	0.180	−0.187	0.994	−0.035	0.841	−0.374	0.304	−0.221	0.172	−0.539	0.096
Departure time	0.731	0.017	0.131	1.331	0.047	0.791	−0.304	0.398	−0.210	0.269	−0.584	0.163
Constant	−7.673	0.002	−12.518	−2.828	−5.821	0.001	−8.592	−3.051	3.058	0.012	0.667	5.450

Table 10: Parameter estimation in MNL using oversampling of balanced data.

Variable	Airplane				HSR				Express bus			
	Coefficient	$P > z$	Credit interval		Coefficient	$P > z$	Credit interval		Coefficient	$P > z$	Credit interval	
			2.50%	97.50%			2.50%	97.50%			2.50%	97.50%
Gender												
Male vs. female	0.661	0.118	−0.167	1.490	0.196	0.443	−0.305	0.697	0.681	0.001	0.275	1.087
Age	0.386	0.126	−0.108	0.880	0.366	0.020	0.058	0.673	0.360	0.007	0.097	0.624
Occupation												
Personnel of institution vs. enterprise unit	0.343	0.564	−0.824	1.510	0.138	0.723	−0.627	0.903	0.318	0.401	−0.424	1.061
Student vs. enterprise unit	0.298	0.672	−1.082	1.679	−0.352	0.399	−1.171	0.466	1.856	0.001	1.193	2.519
Farmer vs. enterprise unit	−0.898	0.309	−2.627	0.831	−0.546	0.376	−1.757	0.664	−0.338	0.488	−1.291	0.616
Self-employed vs. enterprise unit	−1.107	0.126	−2.525	0.312	−0.387	0.376	−1.243	0.469	0.554	0.119	−0.142	1.250
Others vs. enterprise unit	−1.518	0.055	−3.067	0.030	−0.479	0.305	−1.394	0.437	0.545	0.179	−0.250	1.340
Monthly income	−0.040	0.836	−0.416	0.336	−0.226	0.043	−0.445	−0.007	−0.017	0.840	−0.182	0.148
Travel purpose												
Mandatory travel vs. leisure travel	−0.343	0.413	−1.166	0.479	−0.384	0.140	−0.895	0.126	−0.134	0.506	−0.529	0.261
Intercity travel distance	0.002	0.001	0.001	0.004	−0.001	0.264	−0.002	0.001	−0.002	0.001	−0.003	−0.001
Intercity travel cost	0.026	0.001	0.021	0.031	0.022	0.001	0.017	0.026	0.001	0.580	−0.003	0.005

Table 10: Continued.

Variable	Airplane				HSR				Express bus			
	Coefficient	$P > z$	Credit interval		Coefficient	$P > z$	Credit interval		Coefficient	$P > z$	Credit interval	
			2.50%	97.50%			2.50%	97.50%			2.50%	97.50%
Intercity travel time	−1.242	0.001	−1.432	−1.053	−0.421	0.001	−0.524	−0.318	0.001	0.964	−0.050	0.052
Safety	0.801	0.006	0.225	1.378	0.678	0.001	0.279	1.077	0.146	0.333	−0.149	0.440
Comfort	0.236	0.404	−0.318	0.791	0.371	0.051	−0.001	0.743	−0.461	0.001	−0.737	−0.185
Punctuality	−0.353	0.194	−0.886	0.179	0.321	0.082	−0.040	0.682	−0.513	0.001	−0.808	−0.219
Access time	0.456	0.076	−0.048	0.959	−0.122	0.462	−0.447	0.203	−0.244	0.060	−0.499	0.010
Departure time	1.030	0.001	0.506	1.554	0.271	0.130	−0.080	0.621	−0.185	0.232	−0.487	0.118
Constant	−9.522	0.001	−13.873	−5.170	−6.798	0.001	−9.530	−4.066	2.303	0.020	0.356	4.250

Table 11: Parameter estimation in MNL using undersampling of balanced data.

Variable	Airplane				HSR				Express bus			
	Coefficient	$P > z$	Credit interval		Coefficient	$P > z$	Credit interval		Coefficient	$P > z$	Credit interval	
			2.50%	97.50%			2.50%	97.50%			2.50%	97.50%
Gender												
Male vs. female	0.847	0.137	−0.268	1.962	0.175	0.637	−0.552	0.901	0.723	0.010	0.176	1.271
Age	0.439	0.213	−0.252	1.130	0.305	0.183	−0.144	0.754	0.259	0.135	−0.080	0.598
Occupation												
Personnel of institution vs. enterprise unit	0.252	0.748	−1.284	1.787	−0.057	0.916	−1.113	0.999	0.220	0.657	−0.751	1.191
Student vs. enterprise unit	0.317	0.742	−1.565	2.199	−0.128	0.831	−1.304	1.048	1.642	0.001	0.774	2.510
Farmer vs. enterprise unit	−1.600	0.179	−3.931	0.732	−0.704	0.399	−2.339	0.931	−0.518	0.420	−1.777	0.741
Self-employed vs. enterprise unit	−1.939	0.059	−3.953	0.076	−0.828	0.200	−2.096	0.439	0.646	0.176	−0.291	1.584
Others vs. enterprise unit	−1.860	0.093	−4.028	0.307	−0.782	0.289	−2.229	0.664	0.632	0.256	−0.458	1.721
Monthly income	−0.139	0.612	−0.676	0.398	−0.223	0.180	−0.550	0.103	−0.048	0.671	−0.271	0.175
Travel purpose												
Mandatory travel vs. leisure travel	−0.337	0.549	−1.439	0.764	−0.313	0.408	−1.055	0.428	−0.283	0.300	−0.819	0.252
Intercity travel distance	0.003	0.006	0.001	0.005	0.001	0.960	−0.002	0.002	−0.002	0.008	−0.003	0.001
Intercity travel cost	0.031	0.001	0.023	0.040	0.027	0.001	0.019	0.036	0.001	0.875	−0.006	0.005
Intercity travel time	−1.395	0.001	−1.686	−1.105	−0.638	0.001	−0.837	−0.439	−0.013	0.763	−0.095	0.070
Safety	0.084	0.838	−0.727	0.896	0.514	0.074	−0.050	1.077	0.014	0.945	−0.392	0.421
Comfort	0.301	0.424	−0.437	1.039	0.466	0.074	−0.044	0.976	−0.418	0.023	−0.780	−0.057
Punctuality	−0.389	0.256	−1.061	0.283	0.189	0.448	−0.298	0.675	−0.617	0.002	−1.003	−0.230
Access time	0.422	0.237	−0.277	1.120	−0.194	0.430	−0.676	0.288	−0.135	0.434	−0.475	0.204
Departure time	0.655	0.073	−0.061	1.371	−0.034	0.892	−0.530	0.461	−0.226	0.259	−0.617	0.166
Constant	−5.696	0.050	−11.396	0.004	−5.297	0.006	−9.039	−1.555	3.420	0.008	0.912	5.928

from AUCs are consistent with those from the confusion matrices for each model.

4.3. Model Interpretation. We use the results of the BMNL with better statistical performance and MLP models with better predictive performance to explain the effects of factors on intercity travel mode choice.

From Table 4 and Figure 4, it is found that gender was positively correlated with the choice of HSR and express bus, indicating that men were prone to traveling by HSR or express bus, and women by train. This finding is consistent with a previous study [2], which revealed that women preferred using train more than men. The models show that personnel of government-sponsored institutions were more likely than enterprise personnel to choose an airplane. Farmers and the self-employed were less likely than enterprise personnel to travel by airplane. Similarly, students and farmers were not prone to choosing HSR, and farmers were prone to using an express bus [2, 8]. These results are supported by a previous study [1, 3] that found that passengers working in the state sector are likely to choose airplane over coach. Monthly income was found to be positively associated with airplane choice, and negatively

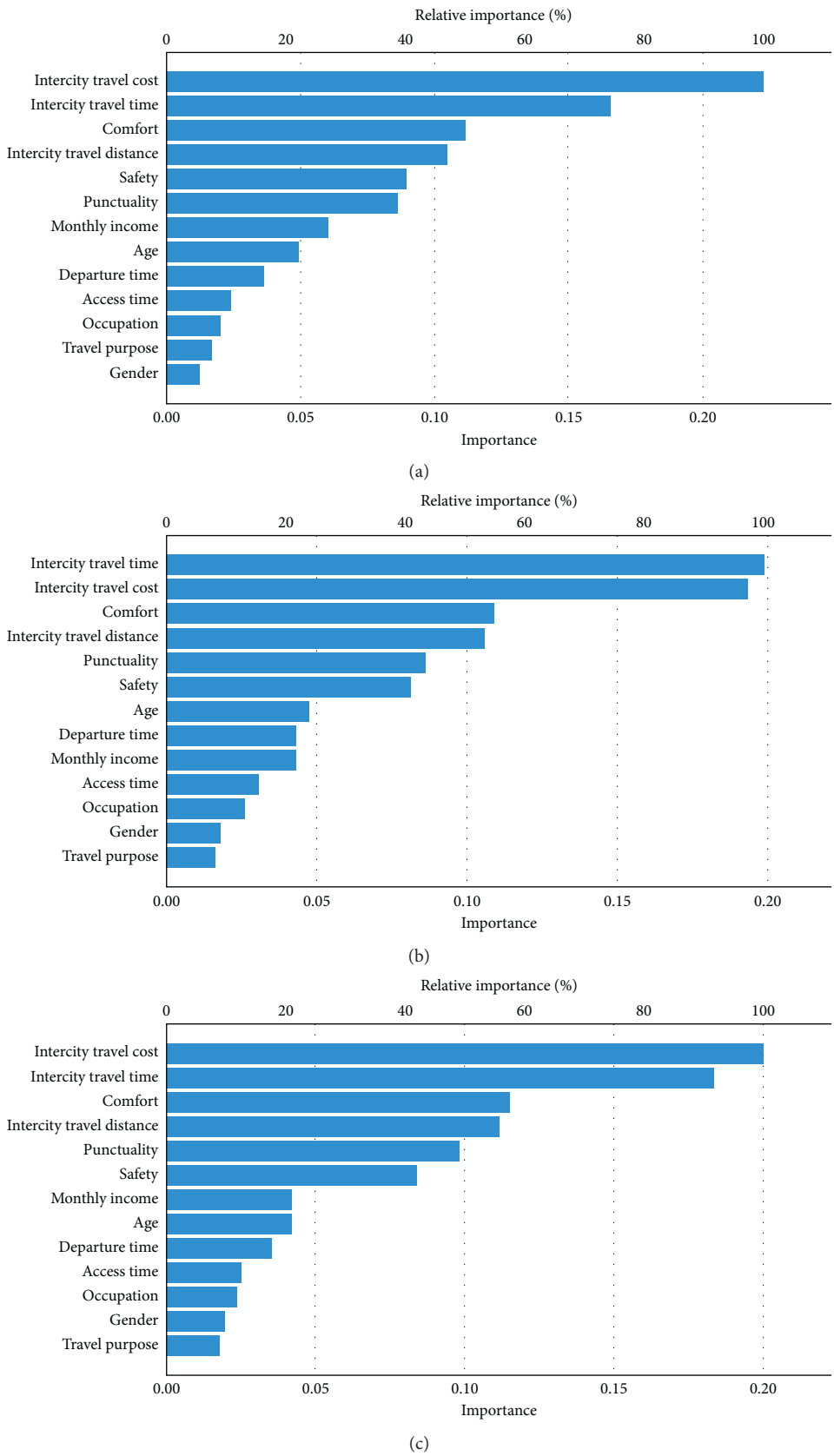


Figure 3: Relative importance of each variable using RBF. (a) Imbalanced data. (b) Oversampling balanced data. (c) Undersampling balanced data.

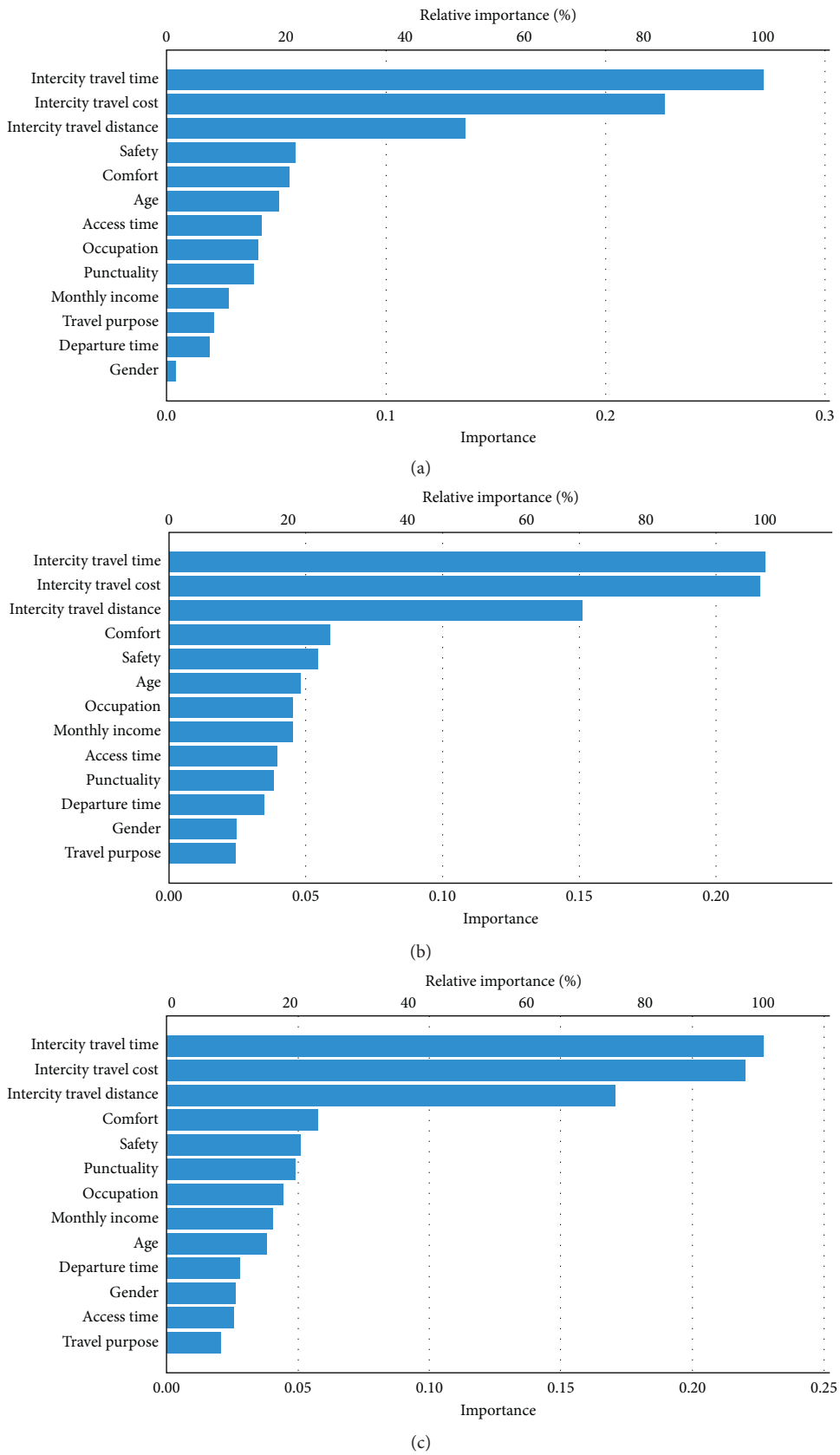


Figure 4: Relative importance of each variable using MLP. (a) Imbalanced data. (b) Oversampling balanced data. (c) Undersampling balanced data.

Table 12: Confusion matrix and recall, precision, and accuracy of each model for imbalanced data.

Model			Training set					Predictive set				
			Predictive class					Predictive class				
		Mode	Airplane	HSR	Train	Express bus	Recall	Airplane	HSR	Train	Express bus	Recall
MNL	Actual class	Airplane	109	18	2	0	84.50%	30	2	0	0	93.80%
		HSR	16	246	27	6	83.40%	3	57	11	3	77.00%
		Train	3	25	181	28	76.40%	3	5	44	8	73.30%
		Express bus	0	8	54	59	48.80%	0	1	12	19	59.40%
		Precision	85.16%	82.83%	68.56%	63.44%	**76.10%**	83.33%	87.69%	65.67%	63.33%	**75.76%**
BMNL	Actual class	Airplane	113	14	2	0	87.60%	30	2	0	0	93.80%
		HSR	18	243	28	6	82.40%	3	55	13	3	74.30%
		Train	5	24	184	26	77.00%	3	5	45	7	75.00%
		Express bus	3	7	53	61	49.20%	0	0	13	19	59.40%
		Precision	81.29%	84.38%	68.91%	65.59%	**76.36%**	83.33%	88.71%	63.38%	65.52%	**75.25%**
MLP	Actual class	Airplane	118	9	2	0	91.50%	29	3	0	0	90.60%
		HSR	12	268	13	2	90.80%	3	64	6	1	86.50%
		Train	1	30	183	23	77.20%	0	7	45	7	76.30%
		Express bus	0	5	54	62	51.20%	0	2	13	17	53.10%
		Precision	90.08%	85.90%	72.62%	71.26%	**80.70%**	90.63%	84.21%	70.31%	68.00%	**78.70%**
RBF	Actual class	Airplane	80	41	6	2	62.00%	21	11	0	0	65.60%
		HSR	17	216	45	17	73.20%	4	53	12	5	71.60%
		Train	0	39	181	17	76.40%	1	9	38	11	64.40%
		Express bus	0	18	54	49	40.50%	0	1	14	17	53.10%
		Precision	82.47%	68.79%	63.29%	57.65%	**67.30%**	80.77%	71.62%	59.38%	51.52%	**65.50%**

The bold values represent the accuracy of models.

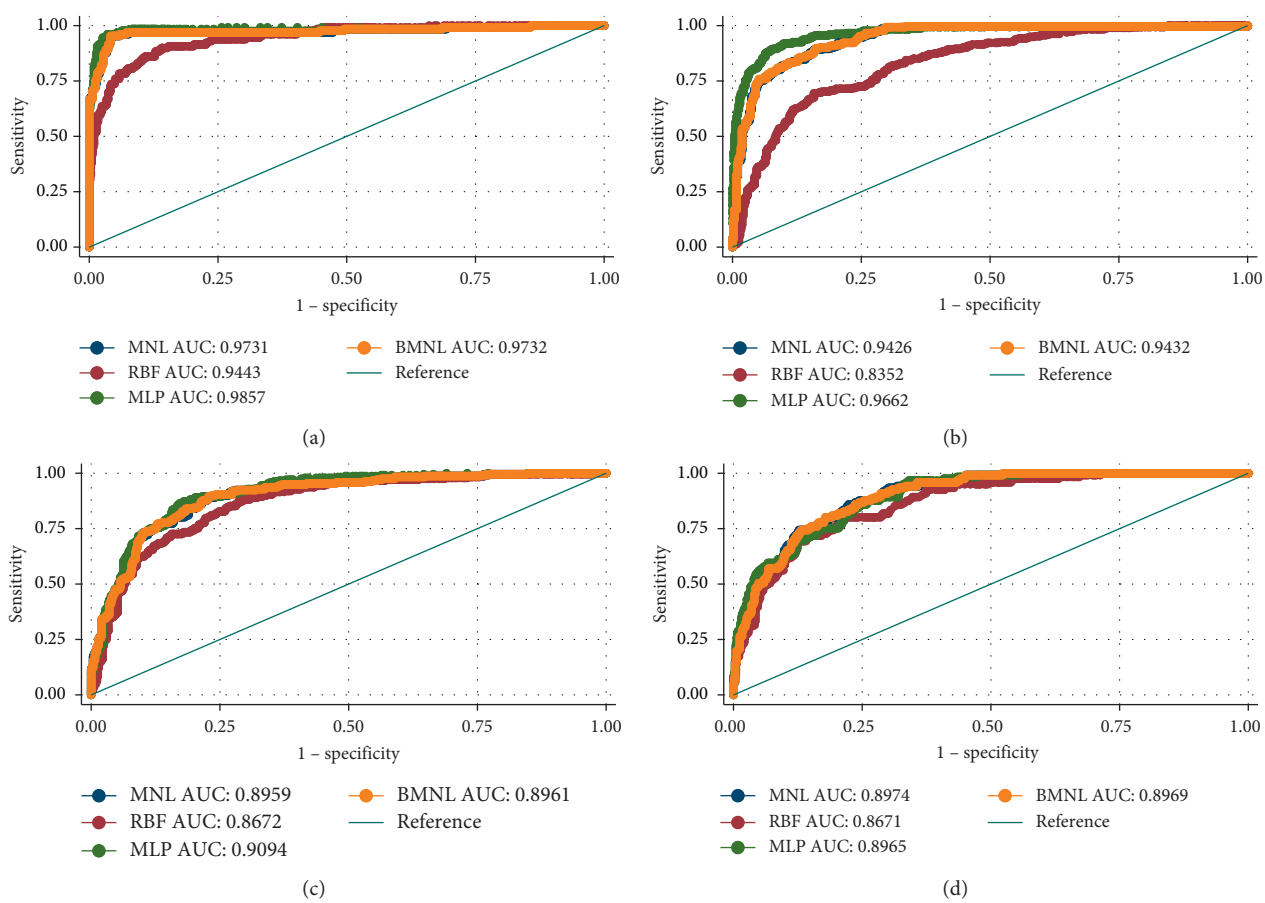


Figure 5: ROC curves of models for imbalanced data training set. (a) Airplane. (b) HSR. (c) Train. (d) Express bus.

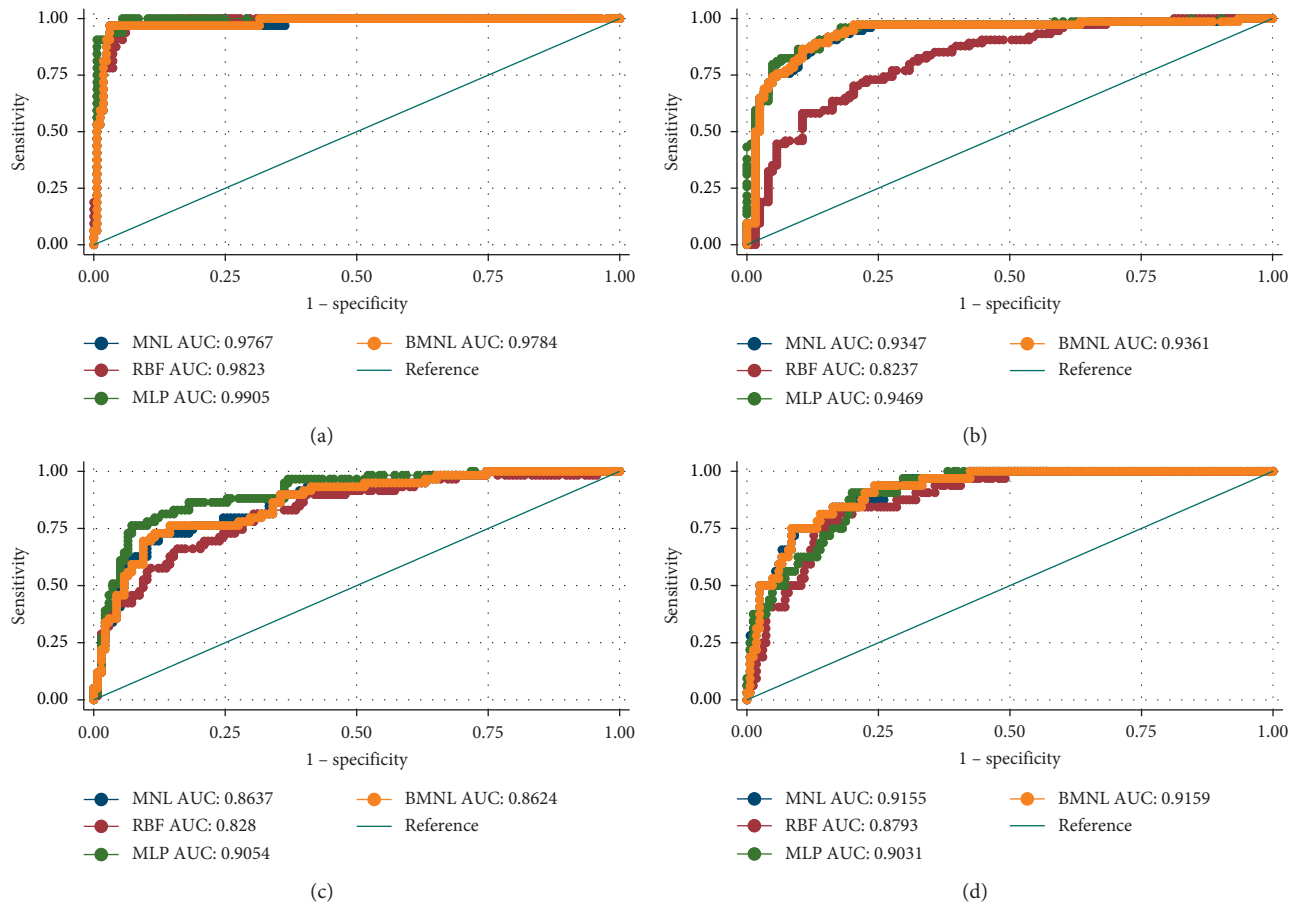


Figure 6: ROC curves of models for imbalanced data predictive set. (a) Airplane. (b) HSR. (c) Train. (d) Express bus.

Table 13: Confusion matrix and recall, precision, and accuracy of each model for undersampling of balanced data.

Model			Training set					Predictive set				
			Predictive class					Predictive class				
		Mode	Airplane	HSR	Train	Express bus	Recall	Airplane	HSR	Train	Express bus	Recall
MNL	Actual class	Airplane	115	8	1	1	92.00%	27	4	0	0	87.10%
		HSR	8	102	7	8	81.60%	5	21	1	4	67.74%
		Train	3	15	79	28	63.20%	0	4	15	12	48.39%
		Express bus	3	3	25	94	75.20%	0	1	8	22	70.97%
		Precision	89.15%	79.69%	70.54%	71.76%	**78.00%**	84.38%	70.00%	62.50%	57.89%	**68.55%**
BMNL	Actual class	Airplane	116	7	1	1	92.80%	28	3	0	0	90.32%
		HSR	12	88	14	11	70.40%	5	18	3	5	58.06%
		Train	3	6	81	35	64.80%	0	2	17	12	54.84%
		Express bus	3	0	28	94	75.20%	0	1	8	22	70.97%
		Precision	86.57%	87.13%	65.32%	66.67%	**75.80%**	84.85%	75.00%	60.71%	56.41%	**68.55%**
MLP	Actual class	Airplane	121	2	1	1	96.80%	28	3	0	0	90.30%
		HSR	6	107	7	5	85.60%	2	25	1	3	80.60%
		Train	1	13	92	17	74.80%	0	4	19	8	61.30%
		Express bus	0	3	30	89	73.00%	0	2	10	19	61.30%
		Precision	94.53%	85.60%	70.77%	79.46%	**81.80%**	93.33%	73.53%	63.33%	63.33%	**73.40%**
RBF	Actual class	Airplane	105	12	3	5	84.00%	25	5	0	1	80.60%
		HSR	20	79	15	11	63.20%	3	20	3	5	64.50%
		Train	1	12	83	27	67.50%	0	2	22	7	71.00%
		Express bus	1	9	28	84	68.90%	0	2	8	21	67.70%
		Precision	82.68%	70.54%	64.34%	66.14%	**70.20%**	89.29%	68.97%	66.67%	61.76%	**71.00%**

The bold values represent the accuracy of models.

Table 14: Confusion matrix of each model for oversampling of balanced data.

Model			Training set					Predictive set				
			Predictive class					Predictive class				
		Mode	Airplane	HSR	Train	Express bus	Recall	Airplane	HSR	Train	Express bus	Recall
MNL	Actual class	Airplane	275	18	1	1	93.22%	71	1	1	1	95.95%
		HSR	18	235	23	19	79.66%	9	55	5	5	74.32%
		Train	6	23	190	76	64.41%	1	10	40	23	54.05%
		Express bus	7	8	63	217	73.56%	0	1	27	46	62.16%
		Precision	89.87%	82.75%	68.59%	69.33%	**77.71%**	87.65%	82.09%	54.79%	61.33%	**71.62%**
BMNL	Actual class	Airplane	276	17	1	1	93.56%	71	1	1	1	95.95%
		HSR	18	235	22	20	79.66%	9	54	6	5	72.97%
		Train	6	23	194	72	65.76%	1	10	41	22	55.41%
		Express bus	7	5	73	210	71.19%	0	0	28	46	62.16%
		Precision	89.90%	83.93%	66.90%	69.31%	**77.54%**	87.65%	83.08%	53.95%	62.16%	**71.62%**
MLP	Actual class	Airplane	277	16	1	1	93.90%	72	0	1	1	97.30%
		HSR	13	241	21	20	81.70%	6	55	5	8	74.30%
		Train	3	21	209	58	71.80%	1	13	42	18	56.80%
		Express bus	0	13	31	244	84.70%	0	2	15	57	77.00%
		Precision	94.54%	82.82%	79.77%	75.54%	**83.10%**	91.14%	78.57%	66.67%	67.86%	**76.40%**
RBF	Actual class	Airplane	244	33	6	12	82.70%	58	10	3	3	78.40%
		HSR	51	176	37	31	59.70%	14	43	8	9	58.10%
		Train	2	37	188	64	64.60%	2	15	41	16	55.40%
		Express bus	2	18	72	196	68.10%	0	3	18	53	71.60%
		Precision	81.61%	66.67%	62.05%	64.69%	**68.14%**	78.38%	60.56%	58.57%	65.43%	**65.90%**

The bold values represent the accuracy of models.

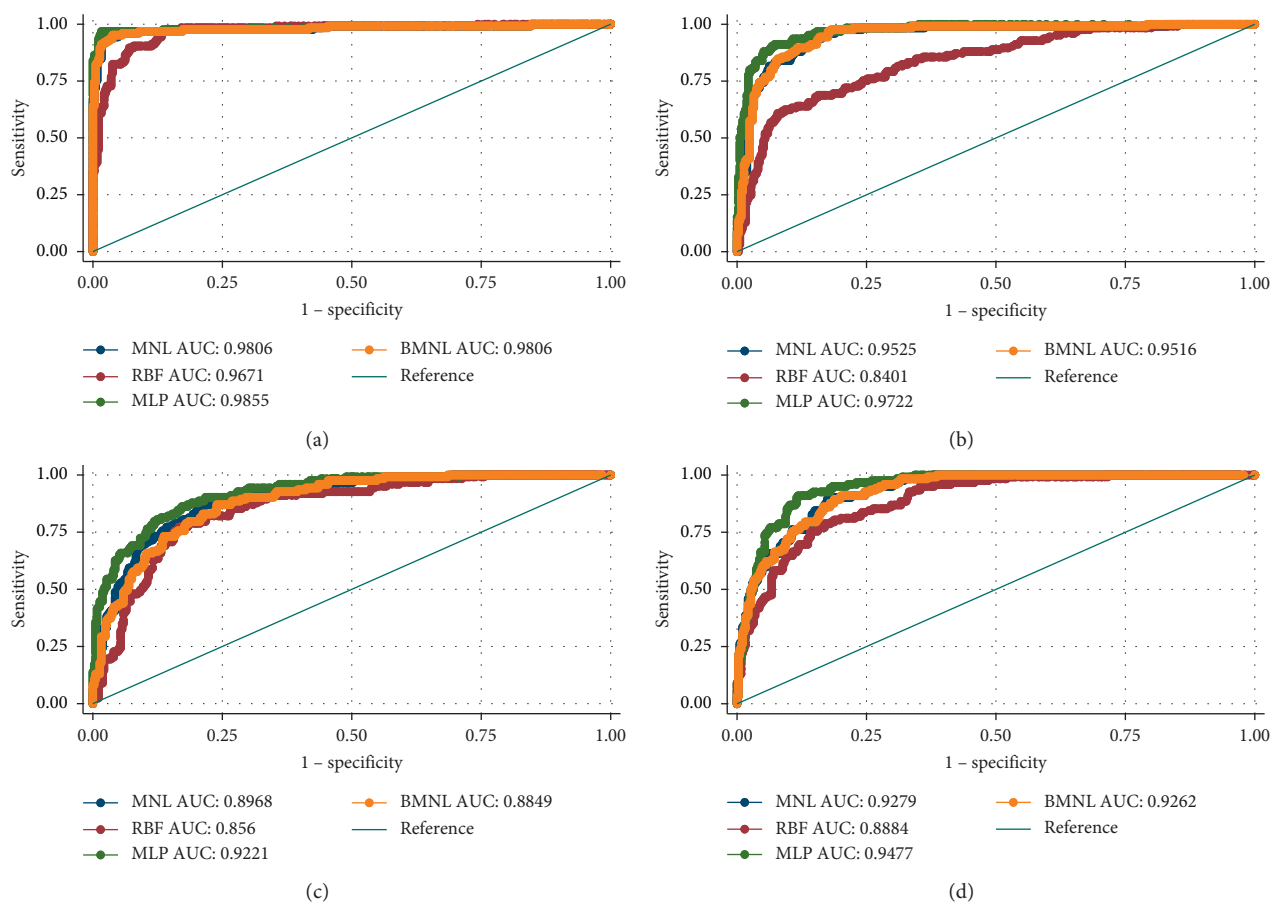


Figure 7: ROC curves for undersampled balanced data training set. (a) Airplane. (b) HSR. (c) Train. (d) Express bus.

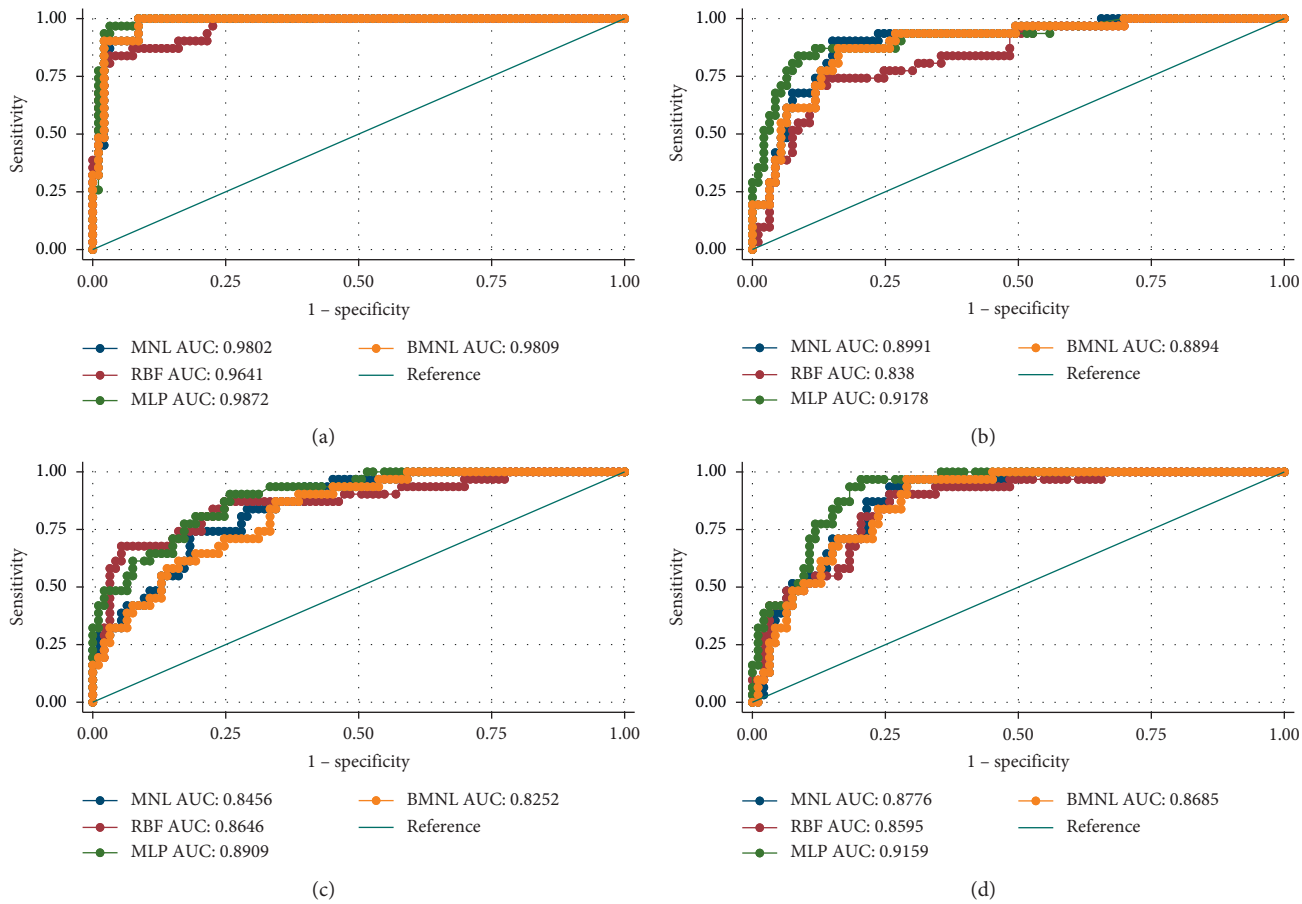


Figure 8: ROC curves for undersampled balanced data predictive set. (a) Airplane. (b) HSR. (c) Train. (d) Express bus.

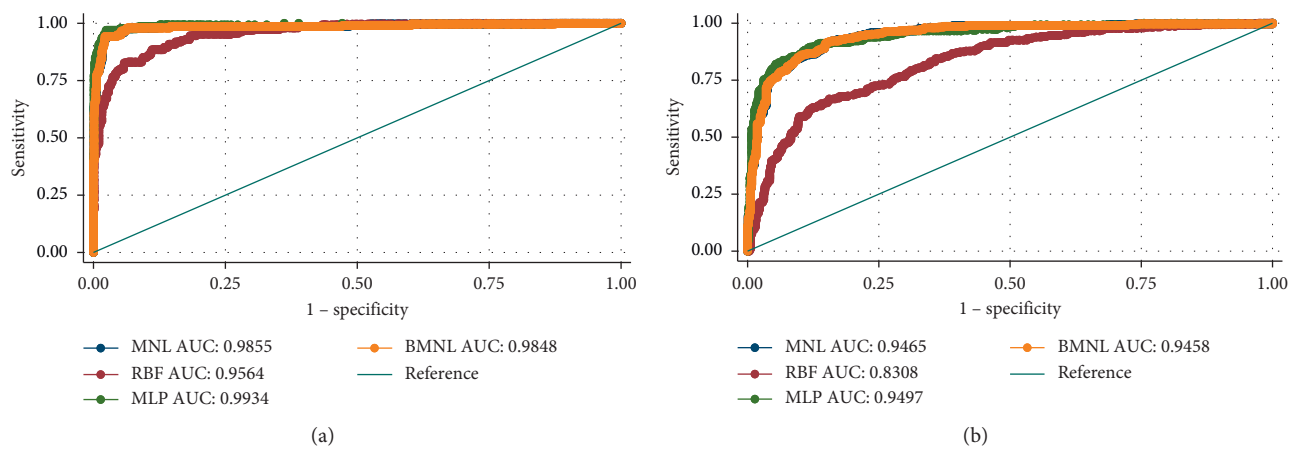


Figure 9: Continued.

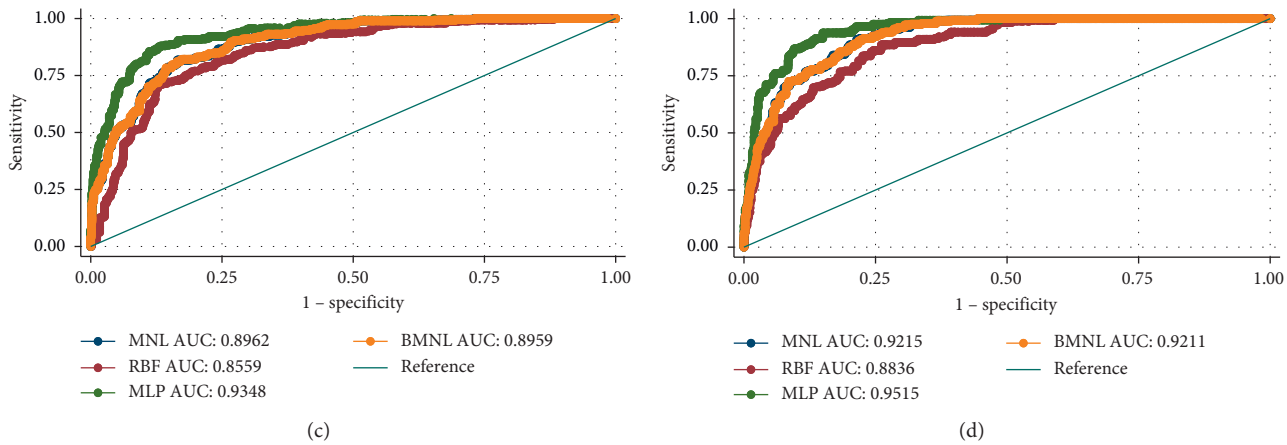


Figure 9: ROC curves for oversampled balanced data training set. (a) Airplane. (b) HSR. (c) Train. (d) Express bus.

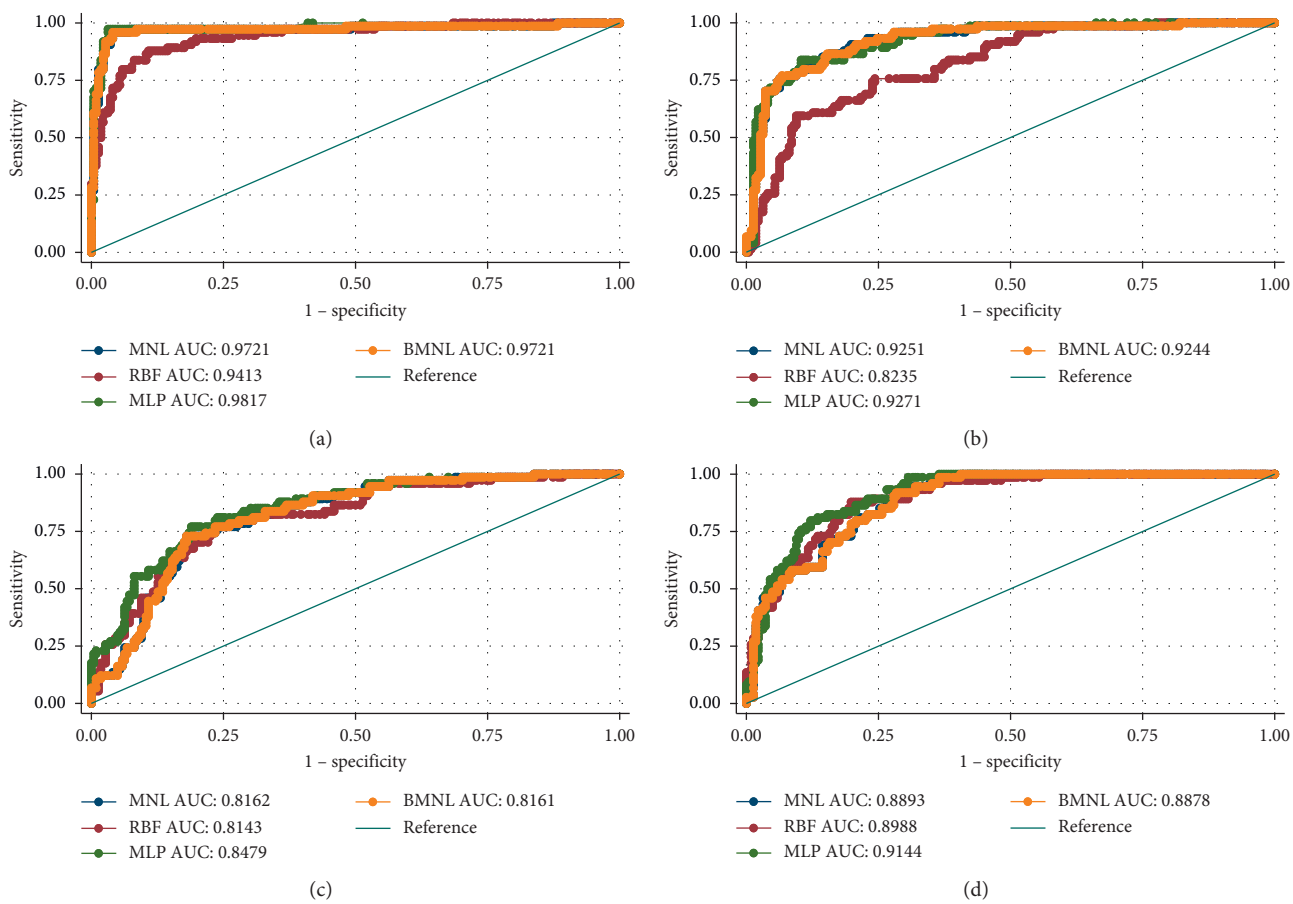


Figure 10: ROC curves for oversampled balanced data training set. (a) Airplane. (b) HSR. (c) Train. (d) Express bus.

with express bus choice. This result agrees with the result of a study [4, 7] that found that higher- and lower-income individuals favor air and bus travel, respectively.

The variable of travel purpose had a significant positive effect on HSR choice and was ranked 11th in the relative importance of all variables. This finding is similar to that of a past study [1] and reveals that, compared to the train, leisure passengers prefer HSRs or airplanes more than passengers for mandatory travel. It is possible that leisure passengers can afford the higher travel cost and are more willing to travel in a comfortable mode. The modeling results also show the significant impact of travel distance on

intercity travel mode choice, which is third most important. This result implies that, compared to the train, the longer travel distance favors airplane and HSR, and the shorter distance favors express bus, which is consistent with previous studies [3, 10]. Intercity travel cost is the most important variable and is a positive sign for the airplane or HSR choice, indicating that passengers incurring higher travel costs are more likely to travel by airplane or HSR. Intercity travel time is the second most important factor, showing a negative association with the choice of airplane and HSR, indicating that passengers spending less travel time are more likely to select airplane or HSR. This finding is intuitive because airplanes and HSRs are faster than trains.

Safety ranked fourth, and this variable affects the choice of airplane and HSR, showing that passengers with a higher safety demand are more likely to travel by airplane or HSR. Comfort is the fifth most important factor; it positively influences the choice of airplane and HSR and is negatively associated with express bus. This result is expected, as airplanes and HSRs have better service facilities and environments than trains [1]. Punctuality ranked ninth and is positively related to HSR choice and negatively associated with airplane and express bus. This shows that a higher punctuality demand favors HSR and does not favor airplane and express bus. This result is expected, as external conditions such as bad weather can easily affect the operation of airplanes and express buses, but its impact on HSRs and trains is relatively small [1].

Access time ranked seventh in relative importance and is found to have a positive effect on airplane choice and a negative effect on express bus compared to train, indicating that passengers spending longer access time prefer traveling by airplane and are less likely to travel by express bus. The finding is straightforward because the airport is generally farther than the railway station from the city center, and the highway passenger station is closer [10]. A similar result was found for the effect of departure time.

5. Discussion and Conclusions

We investigated modeling techniques BMNL, MNL, MLP, and RBF for passengers' intercity travel mode choices. Data from a large individual-level survey in the city of Xi'an were used to develop the model. More comprehensive factors such as socioeconomics, travel demand, service quality, and accessibility of transport hub were incorporated in the models.

The comparison results show that MLP has the best predictive performance, BMNL and MNL have approximately equal predictive accuracy, and RBF has the poorest performance using imbalanced data. It was found that the fitting performance of the four models with balanced data was slightly higher than those with imbalanced data. However, it was surprising that the predictive performance of these models with balanced data was slightly lower than those with imbalanced data. A potential reason could be that the degree of imbalance for the original data is very small. These findings suggest that the MLP and BMNL modeling approaches are recommended for the analysis of passengers' intercity travel mode choice. Significant variables in the BMNL model include gender, age, occupation, travel purpose, intercity travel distance, intercity travel cost, intercity travel time, safety, punctuality, access time, and departure time, which is not completely consistent with those in the MNL model. However, the signs of significant variables in the BMNL model were in line with those in the MNL model. Regarding the MLP modeling results, the travel cost was found to be the most important factor in intercity mode choice, followed by travel time and travel distance. Comfort, safety, and punctuality were relatively important factors for passenger travel mode choices. The influence of individual characteristics on intercity travel mode choices was relatively low, and monthly income was the most important factor among individual characteristics.

These findings can provide a reference for traffic management departments to formulate traffic demand management strategies and provide technical support for data analysts and high-tech enterprises to develop intelligent decision-making systems for the choice of passenger intercity travel modes. Through our research conclusion, we can find that intercity travel time, intercity travel cost, intercity travel distance, and the service quality of a transportation mode are important factors affecting intercity travel mode choices. Traffic transportation management departments can accordingly develop a green transportation development strategy by optimizing ticket prices, increasing vehicle speeds, and improving the quality of service, so as to push travelers from transportation with high energy consumption to that with low energy consumption. Our findings show that the predictive performance of models does not significantly improve when using balanced data instead of imbalanced data. This can provide a basis for data analysts to fully understand the impact of data structures on the predictive performance of models.

There are some limitations to this study. The results may only apply to the selected dataset and therefore must be verified using datasets from more cities. The degree of data imbalance and proportion between the training set and the prediction set may also affect the fitting and predictive performance of the models, and it is necessary to explore the fitting and predictive performance of models using extremely unbalanced data and other proportions in the future. In addition, although no significant multicollinearity was found in the independent variables for the models, intercity travel time and intercity travel cost varied with travel distance. It is necessary to generate the fare rate and intercity travel time per kilomile by standardizing the intercity travel time and intercity travel cost and incorporate the transformed variables into the models to eliminate the potential impact of travel distance. Moreover, more variables that might be associated with intercity travel mode choices, such as the characteristics of the destination city, weather, and coronavirus disease, should be investigated. Advanced modeling techniques, such as the Bayesian random parameter model capturing more unobserved heterogeneity, the Probit model with endogenous variables, and the XGBoost model, should be applied in future studies.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 52002282), Shaanxi Natural Science Foundation Youth Project (Grant no. 2017JQ5086), Shaanxi Education Department Special Science and Technology Project Science (Grant no. 19JK0477), and Philosophy and Social Science Foundation of Zhejiang Province (Grant no. 21NDJC163YB).

References

[1] X. Li, J. Tang, X. Hu, and W. Wang, "Assessing intercity multimodal choice behavior in a touristy city: a factor analysis," *Journal of Transport Geography*, vol. 86, Article ID 102776, 2020.

[2] C. R. Bhat, "Accommodating variations in responsiveness to level-of-service measures in travel mode choice modeling," *Transportation Research Part A: Policy and Practice*, vol. 32, no. 7, pp. 495–507, 1998.

[3] V. V. Can, "Estimation of travel mode choice for domestic tourists to Nha Trang using the multinomial probit model," *Transportation Research Part A: Policy and Practice*, vol. 49, pp. 149–159, 2013.

[4] M. A. A. B. Miskeen, A. M. Alhodairi, and R. A. A. B. O. Rahmat, "Modeling a multinomial logit model of intercity travel mode choice behavior for all trips in Libya," *International Journal of Civil & Environmental Engineering*, vol. 7, no. 9, pp. 636–645, 2013.

[5] S. Hess, G. Spitz, M. Bradley, and M. Coogan, "Analysis of mode choice for intercity travel: application of a hybrid choice model to two distinct US corridors," *Transportation Research Part A: Policy and Practice*, vol. 116, pp. 547–567, 2018.

[6] L. Cheng, X. Chen, J. De Vos, X. Lai, and F. Witlox, "Applying a random forest method approach to model travel mode choice behavior," *Travel Behaviour and Society*, vol. 14, pp. 1–10, 2019.

[7] C. V. Forinash and F. S. Koppelman, "Application and interpretation of nested logit models of intercity mode choice," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1413, no. 1413, pp. 98–106, 1993.

[8] C. R. Bhat, "An endogenous segmentation mode choice model with an application to intercity travel," *Transportation Science*, vol. 31, no. 1, pp. 34–48, 1997.

[9] Z.-C. Li and D. Sheng, "Forecasting passenger travel demand for air and high-speed rail integration service: a case study of Beijing-Guangzhou corridor, China," *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 397–410, 2016.

[10] C. Román, J. C. Martín, R. Espino et al., "Valuation of travel time savings for intercity travel: the Madrid-Barcelona corridor," *Transport Policy*, vol. 36, pp. 105–117, 2014.

[11] Y. Guo, Z. Li, Y. Wu, and C. Xu, "Evaluating factors affecting electric bike users' registration of license plate in China using Bayesian approach," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 59, pp. 212–221, 2018.

[12] C. R. Bhat, "A heteroscedastic extreme value model of intercity travel mode choice," *Transportation Research Part B: Methodological*, vol. 29, no. 6, pp. 471–483, 1995.

[13] Y. Wang, L. Li, L. Wang, A. Moore, S. Staley, and Z. Li, "Modeling traveler mode choice behavior of a new high-speed rail corridor in China," *Transportation Planning and Technology*, vol. 37, no. 5, pp. 466–483, 2014.

[14] S. Ashiabor, H. Baik, and A. Trani, "Logit models for forecasting nationwide intercity travel demand in the United States," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2007, no. 1, pp. 1–12, 2007.

[15] J.-H. Lee, K.-S. Chon, and C. Park, "Accommodating heterogeneity and heteroscedasticity in intercity travel mode choice model: formulation and application to HoNam, South Korea, high-speed rail demand analysis," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1898, no. 1, pp. 69–78, 2004.

[16] Y. Guo, P. Liu, Y. Wu, and J. Chen, "Evaluating how right-turn treatments affect right-turn-on-red conflicts at signalized intersections," *Journal of Transportation Safety & Security*, vol. 12, no. 3, pp. 419–440, 2020.

[17] S. Srinivasan, C. R. Bhat, and J. Holguin-Veras, "Empirical analysis of the impact of security perception on intercity mode choice," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1942, no. 1, pp. 9–15, 2006.

[18] X. Li, W. Wang, C. Xu, Z. Li, and B. Wang, "Multi-objective optimization of urban bus network using cumulative prospect theory," *Journal of Systems Science and Complexity*, vol. 28, no. 3, pp. 661–678, 2015.

[19] Y. Guo, Z. Li, Y. Wu, and C. Xu, "Exploring unobserved heterogeneity in bicyclists' red-light running behaviors at different crossing facilities," *Accident Analysis & Prevention*, vol. 115, pp. 118–127, 2018.

[20] C.-X. Zhang, S. Xu, and J.-S. Zhang, "A novel variational Bayesian method for variable selection in logistic regression models," *Computational Statistics & Data Analysis*, vol. 133, pp. 1–19, 2019.

[21] A. P. Afghari, M. M. Haque, S. Washington, and T. Smyth, "Effects of globally obtained informative priors on Bayesian safety performance functions developed for Australian crash data," *Accident Analysis & Prevention*, vol. 129, pp. 55–65, 2019.

[22] X. Zhou, M. Wang, and D. Li, "Bike-sharing or taxi? Modeling the choices of travel mode in Chicago using machine learning," *Journal of Transport Geography*, vol. 79, 2019.

[23] M. Wong and B. Farooq, "A bi-partite generative model framework for analyzing and simulating large scale multiple discrete-continuous travel behaviour data," *Transportation Research Part C: Emerging Technologies*, vol. 110, pp. 247–268, 2020.

[24] X. Zhao, X. Yan, A. Yu, and P. Van Hentenryck, "Prediction and behavioral analysis of travel mode choice: a comparison of machine learning and logit models," *Travel Behaviour and Society*, vol. 20, pp. 22–35, 2020.

[25] C. Xie, J. Lu, and E. Parkany, "Work travel mode choice modeling with data mining: decision trees and neural networks," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1854, no. 1, pp. 50–61, 2003.

[26] Y. Zhang and Y. Xie, "Travel mode choice modeling with support vector machines," *Transportation Research Record:*

Journal of the Transportation Research Board, vol. 2076, no. 1, pp. 141–150, 2008.

[27] Y. Wang, S. Peng, and M. Xu, "Emergency logistics network design based on space-time resource configuration," *Knowledge-Based Systems*, vol. 223, Article ID 107041, 2021.

[28] J. Hagenauer and M. Helbich, "A comparative study of machine learning classifiers for modeling travel mode choice," *Expert Systems with Applications*, vol. 78, pp. 273–282, 2017.

[29] Y. Wang, Y. Yuan, X. Guan et al., "Collaborative two-echelon multicenter vehicle routing optimization based on state-space-time network representation," *Journal of Cleaner Production*, vol. 258, Article ID 120590, 2020.

[30] F. Wang and C. L. Ross, "Machine learning travel mode choices: comparing the performance of an extreme gradient boosting model with a multinomial logit model," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 47, 2018.

[31] M. Wong, B. Farooq, and G.-A. Bilodeau, "Discriminative conditional restricted Boltzmann machine for discrete choice and latent variable modelling," *Journal of Choice Modelling*, vol. 29, pp. 152–168, 2018.

[32] Y. Wang, S. Peng, X. Zhou, M. Mahmoudi, and L. Zhen, "Green logistics location-routing problem with eco-packages," *Transportation Research Part E: Logistics and Transportation Review*, vol. 143, Article ID 102118, 2020.

[33] A. Lindner, C. S. Pitombo, and A. L. Cunha, "Estimating motorized travel mode choice using classifiers: an application for high-dimensional multicollinear data," *Travel Behaviour and Society*, vol. 6, pp. 100–109, 2017.

[34] S. Rasouli and H. J. P. Timmermans, "Using ensembles of decision trees to predict transport mode choice decisions: effects on predictive success and uncertainty estimates," *European Journal of Transport and Infrastructure Research*, vol. 14, pp. 412–424, 2014.

[35] H. B. Celikoglu, "Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modelling," *Mathematical and Computer Modelling*, vol. 44, no. 7-8, pp. 640–658, 2006.

[36] L. Wang and Z. Zuo, "Travel mode recognition using RBF neural network," *CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems*, pp. 711–721, 2014.

[37] J. Sun, J. Sun, and P. Chen, "Use of support vector machine models for real-time prediction of crash risk on urban expressways," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2432, no. 2432, pp. 91–98, 2014.

[38] J. Xie and Z. Qiu, "The effect of imbalanced data sets on LDA: a theoretical and empirical analysis," *Pattern Recognition*, vol. 40, no. 2, pp. 557–562, 2007.

[39] A. D'Addabbo and R. Maglietta, "Parallel selective sampling method for imbalance and large data classification," *Pattern Recognition Letters*, vol. 62, pp. 61–67, 2015.

[40] W. A. Rivera and P. Xanthopoulos, "A priori synthetic oversampling methods for increasing classification sensitivity in imbalanced data sets," *Expert Systems with Applications*, vol. 66, pp. 124–135, 2016.

[41] V. García, J. S. Sánchez, A. I. Marqués, R. Florencia, and G. Rivera, "Understanding the apparent superiority of oversampling through an analysis of local information for class-imbalanced data," *Expert Systems with Applications*, vol. 158, Article ID 113026, 2020.

[42] X. Li, R. Ma, Y. Guo, W. Wang, B. Yan, and J. Chen, "Investigation of factors and their dynamic effects on intercity travel modes competition," *Travel Behaviour and Society*, vol. 23, pp. 166–176, 2021.

[43] H. A. Klaiber and R. H. Von Haefen, "Do random coefficients and alternative specific constants improve policy analysis? An empirical investigation of model fit and prediction," *Environmental and Resource Economics*, vol. 73, pp. 75–91, 2011.

[44] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[45] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, "Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation," *Journal of Computational & Graphical Statistics*, vol. 24, no. 1, pp. 44–65, 2015.

[46] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, Lulu, Morrisville, NC, USA, 2018.

[47] R. Hecht-Nielsen, "Kolmogorov's mapping neural network existence theorem," *Proceedings of the International Conference on Neural Networks*, vol. 3, pp. 11–13, 1987.

[48] S. V. Stehman, "Estimating standard errors of accuracy assessment statistics under cluster sampling," *Remote Sensing of Environment*, vol. 60, no. 3, pp. 258–269, 1997.

[49] C. Xu, W. Wang, P. Liu, and F. Zhang, "Development of a real-time crash risk prediction model incorporating the various crash mechanisms across different traffic states," *Traffic Injury Prevention*, vol. 16, no. 1, pp. 28–35, 2015.

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 6937228, 15 pages
https://doi.org/10.1155/2021/6937228

Research Article

Analysis of Spatial-Temporal Characteristics of Operations in Public Transport Networks Based on Multisource Data

Hui Zhang,[1] Yanjun Liu,[1] Baiying Shi,[1] Jianmin Jia,[1] Wei Wang,[2] and Xiang Zhao[3]

[1]*School of Transportation Engineering, Shandong Jianzhu University, Jinan 250101, China*
[2]*School of Economics, Ocean University of China, Qingdao 266100, China*
[3]*Institute of Computing Technology, China Academy of Railway Sciences, Beijing 100044, China*

Correspondence should be addressed to Baiying Shi; shibaiying@sdjzu.edu.cn

Received 26 June 2021; Accepted 30 August 2021; Published 9 September 2021

Academic Editor: Yong Wang

Operational efficiency and stability are two critical aspects to measure bus systems. Influenced by many stochastic factors, buses always suffer from delay and bunching. Traditional studies focus on a single route and lack research on the systematic evaluation of bus network. In this paper, we propose a data-driven framework to analyze the efficiency and stability based on small granularity GPS trajectory data from the perspective of entire bus network. The IC card data and route data are used to extract the boarding passenger number and topological structure, respectively. The results show that the average headway of stations follows a lognormal distribution. Moreover, the distribution of arrival efficiency of stations is inhomogeneous and a small number of stations have large values. In addition, the relationships among average headway of stations, boarding passenger number, bus number, and complex network indicators are revealed. It is found that the average headway of station is negatively correlated with other indicators, which implies that complex network connections and more passenger flows could weaken the efficiency of bus operations. This paper provides a way to evaluate the operational performance of bus networks and could give help for monitoring and optimizing the daily operation of bus systems.

1. Introduction

Nowadays, public transport plays a growing role in alleviating traffic congestion and reducing greenhouse gas emission. Building effective, convenient, and stable public transport has become a crucial step to solve urban traffic problems in many countries. The efficiency and stability of public transport are two core problems that are concerned by both travelers and operators. Affected by many stochastic factors such as weather, congestion, passenger flow, and drivers' behavior, bus delay and bunching occur in many routes during the operation time [1–4]. Bus bunching is that the adjacent buses belonging to the same route are too close to each other. Bus bunching could impact the uniformity of buses and lead to big interval from the other buses, which enhance passengers' waiting time and result in low efficiency. In some bus systems with dense stations and large passenger flow, a small disturbance of buses in a route may spread all over the network. So, it is imperative to construct a robust public transport system to provide better service.

The development of public transport information technology provides a powerful tool to monitor and manage the transportation system. The Automatic Vehicle Location (AVL) system and Automatic Passenger Count (APC) system could record the travel trajectory and passenger information. In recent years, many studies focused on travel time, delay, and reliability of buses based on AVL data [5–7]. To address these problems, numerous control strategies have been proposed, like speed control [8], holding strategies [9, 10], and skip-stop strategies [11]. The prediction methods give great help to improve the control effects, which is better than traditional models without predictions [12]. High-resolution bus GPS data could also be used to identify congestion hotspots in the urban street [13]. The APC data are commonly used to extract OD information, estimate waiting time, and find missing transfers [14–16]. Recently,

some researchers use APC data to identify public transit corridors and transit network flow characteristics [17, 18]. Typically, the bus operational status can be achieved by AVL data incorporating APC data [19].

Essentially, the delay and bunching of bus are caused by small disturbances such as more boarding passengers and bad traffic condition of road. These small disturbances may spread, superpose, and amplify, which results in heavy disorder of the entire bus system. The public transit system is a complex spatiotemporal network embedded in complicated urban surroundings [5]. Understanding the mechanism of bus operation is conducive to enhance the efficiency of transit service. Most researches have focused on a single route [8, 20], and few studies pay attention to the whole transit network's operational stability. The structures and dynamics of bus networks are so complicated. In the past years, researchers studied the bus networks from complex network perspective using line information and IC card data [21–23]. They pay main attention to the structure characteristics of bus such as community structure and "small world."

Recently, there are many researches focusing on extract bus network characteristics by merging AVL data and APC data. Chen et al. used IC card data and GPS data to estimate passenger boarding and alighting station of the entire bus and found that passenger flow is mainly distributed in an east-west belt-shaped downtown area [19]. Sui et al. constructed a layered network model to depict public transport network, OD flows, and transfer flows [24]. The data-driven methods appear to analyze the whole bus network performance in recent years. Zhang et al. studied the average headway and headway deviation of the entire bus network in Jinan, China, and found that the two indicators follow lognormal distributions [25]. Iliopoulou et al. used AVL data to identify spatiotemporal patterns of bus bunching by clustering method. However, they did not study the passenger flow factor due to the lack of APC data [26]. Nowadays, network-based studies using big data have been successfully applied in many transportation systems [27–31].

Most aforementioned studies focus on the operational performance of a single route using AVL data. However, there is limited research for the whole network. Actually, the operational performance of bus network is more significant in bus network design and optimization of operation. For example, planners consider passenger flow, network structure, accessibility, and transfer when designing or adding new routes. They rarely consider the operation performance systematically. Therefore, there exist many stations having very bad operation performance. Passengers should wait a long time or suffer from bus bunching at these stations. This is a big problem for bus systems. At present, the network-based studies of bus network concentrated on the topological structure rather than operational performance. How to evaluate the bus operation performance and find the hub stations is the key to solve the problem. To fill this gap, this paper designs proper indicators based on GPS trajectory data to evaluate the operational performance of the entire bus network.

In this study, we conduct a data-driven framework to measure the operational status of bus network-based incorporating GPS trajectory data, IC card data, and route data. First, we propose average headway of station and average arrival rate of station to assess the stability and efficiency of bus network. Second, the relationships among the two proposed indicators, boarding passengers, bus supply, and topological structure are studied. This paper aims to construct a framework based on multisource data to evaluate the operation of bus systems. The rest paper is organized as follows. Section 2 gives the literature review. Section 3 proposes evaluation indicators of bus network performance. Section 4 introduces the data used in this paper. Section 5 shows the results. Section 6 concludes the paper.

2. Literature Review

2.1. Application of Locator Data in Transportation Area. The GPS trajectory data are space-time continuum, which is the foundation to study bus operation status and residents' activity. In transportation area, it has been widely applied in travelers' behavior, estimation of traffic demand, traffic status, and traffic model optimization. The GPS trajectory data are playing a more and more important role in traffic planning and management. In some areas, it has become an alternative to traditional traffic survey.

At present, the studies of travelers' behavior and estimation of traffic demand concentrated on taxi trajectory data and mobile phone data. Tang et al. divided the studied area into small cells and estimated the distribution of OD [32]. Zhang et al. utilized complex network theory to reveal the urban traffic demand based on taxi trajectory data [33]. Mobile data are a good source to study travelers' behavior, which is a research hotspot [34, 35]. Dockless bike sharing is an emerging traffic mode, which plays a significant role in connecting with other traffic modes. The trip data can be used to analyze travelers' behavior, traffic demand estimation, and bicycle rebalance [36–38]. In the transit area, there are many researches that focus on travel time prediction and delay [5, 39]. The GPS trajectory data could reflect the operational status of bus, which is crucial for operation management. Besides that, the trajectory data can be used to detect the travelers' behavior incorporating IC card. Tu et al. studied the dynamic characteristics of multimode travel by trajectory data, IC card data, and taxi trajectory data [40]. Tang et al. used GPS trajectory data and smart card data to optimize the timetable of the bus line [41]. The GPS trajectory data could also be used to identify transportation mode by GIS information and machine learning methods [42, 43].

2.2. Operational Stability of Public Transport. The operational stability of public transport has drawn much attention since the AVL and APC devices were applied. Chepuri et al. pointed out that the travel times during peak hours followed normal distributions [44]. Fan and Machemehl verified the relationship between the waiting

time of passengers and headway deviation and they found a positive relationship between them [45]. The travel time of bus contains running time and stopping time. Studies show that the running time is affected by traffic conditions of roads and traffic signals [46, 47]. Kieu et al. analyzed the distribution of transit travel time used transit signal priority data and found that they followed lognormal distributions [48]. The stopping time comprises door opening, door closing, and the time of passengers boarding and alighting. Research shows that the time of stopping could account for 26% of total travel time [49]. Generally, the main factor that impacts the stopping time is the number of boarding and alighting passengers [50]. Ji et al. claimed that enlarging the platform areas and installed guide guardrails could reduce the variation of dwell time, but not the time [51]. Passengers' arrival times are related to the headway. When the headway is smaller than 12 minutes, passengers arrive randomly [45]. Chepuri et al. proposed new reliability indicators to measure the bus route by trajectory data from both route level and segment level [52]. Paudel revealed that high volumes of bus ridership could cause a significant increase in the variance of bus service reliability [53].

2.3. Evaluation of Transit System. The service quality of the transit system is influenced by many components such as transit network structure and management level. The transit systems are fragile when they meet periodic passenger flow fluctuations and other stochastic factors. It is necessary to grasp the operational status of the transit system. Traditional studies focused on a single route rather than the entire network. From the evaluation of transit network perspective, there have been various researches in the past years. Zhang et al. constructed the evaluation framework which includes convenience, comfort, security, reliability, and facility level according to survey data [54]. Lots of studies analyzed transit network topology structure using complex network theory [55–57]. Sun et al. estimated the transit ridership by points of interest [58]. Chen et al. used GPS trajectory data and IC card data to estimate the number of boarding and alighting passengers and the distributions of passenger flow [19]. Illiopoulou et al. used the clustering method to distinguish different types of bus bunching [26]. Zuo et al. proposed a holistic accessibility measurement considering land use, timetable, and individual factors to evaluate the bus system [59]. Zhang et al. identified the bus station patterns based on network structure, operational status, and POI data [60]. Wei et al. highlighted the bus lines value and proposed a "line-line" network to examine the spatial characteristics of cross-administration bus lines [61]. Bree et al. studied the relationship between transit ridership and local accessibility and found that it was more closely to predict public transit ridership when including gravity-based accessibility in the model [62]. All in all, there were many studies on the evaluation of transit system, but there is little study focus on the operational status of the entire transit network.

3. Evaluation Indicators of Bus Network Performance

The performance of a bus network is significant in daily operations. Reliability and efficiency are two key indicators to measure the performance of bus systems. Headway is an important indicator to study the reliability. The headway between adjacent buses varies, affected by many stochastic factors such as traffic congestion and bad weather. The GPS trajectory data are a feasible source to detect the operational status of buses. Figure 1 shows the GPS trajectory data of bus routes. Traditional studies mainly focus on the operational status of a single route. In this paper, we use the average headway of stations and the average arrival rate of stations to evaluate the operational performance of bus networks. We intend to extract the macroscopic operation status of bus networks by mining the GPS trajectory data.

3.1. Average Headway of Stations. The bus run is considered as a series of events that comprises section running, arrivals, and departures. Headway is the time difference between two successive vehicles belonging to the same routes. It is defined as [25]

$$\Delta H_k^{i,j} = \begin{cases} 0, & j = 1, \\ \left| h_k^{i,j+1} - h_k^{i,j} \right|, & \text{otherwise}, \end{cases} \tag{1}$$

where $\Delta H_k^{i,j}$ is the headway between vehicle j and the former vehicle of line i at station k; set the headway for the first vehicle as 0; $h_k^{i,j}$ is the arrival time of vehicle j of line i at station k.

Generally, the headways of buses are even according to their schedule timetable when they depart at original stations. The headways fluctuate due to the complicated external environment and drivers' factors. Sometimes, the headway between two consecutive buses is smaller than the plan headway. When the headway is too small that two consecutive buses arrive at the same time, the bus bunching appears. On the contrary, there is larger headway. Figure 2 shows the four kinds of bus operation status. Bus bunching and large headway disturb the balance of operation, which could result in long waiting time for passengers.

In this paper, the average headway of stations during time period t is defined as

$$\Delta \overline{H}_k^t = \sum_{i=1}^{m_t} \sum_{j=1}^{n_t^i} \frac{\Delta H_k^{i,j}}{s_k^t}, \tag{2}$$

where m_t is the number of routes that stop at station k, n_t^i is the number of buses of route i during time period t, and s_k^t is the total number of buses that stop at station k during time period t.

3.2. Average Arrival Rate of Stations. Bus operational efficiency is crucial for passengers, which could reflect the quality of service. In this paper, we introduce the average arrival rate of stations to measure the operational efficiency of the entire bus network. It is defined as the mean value of

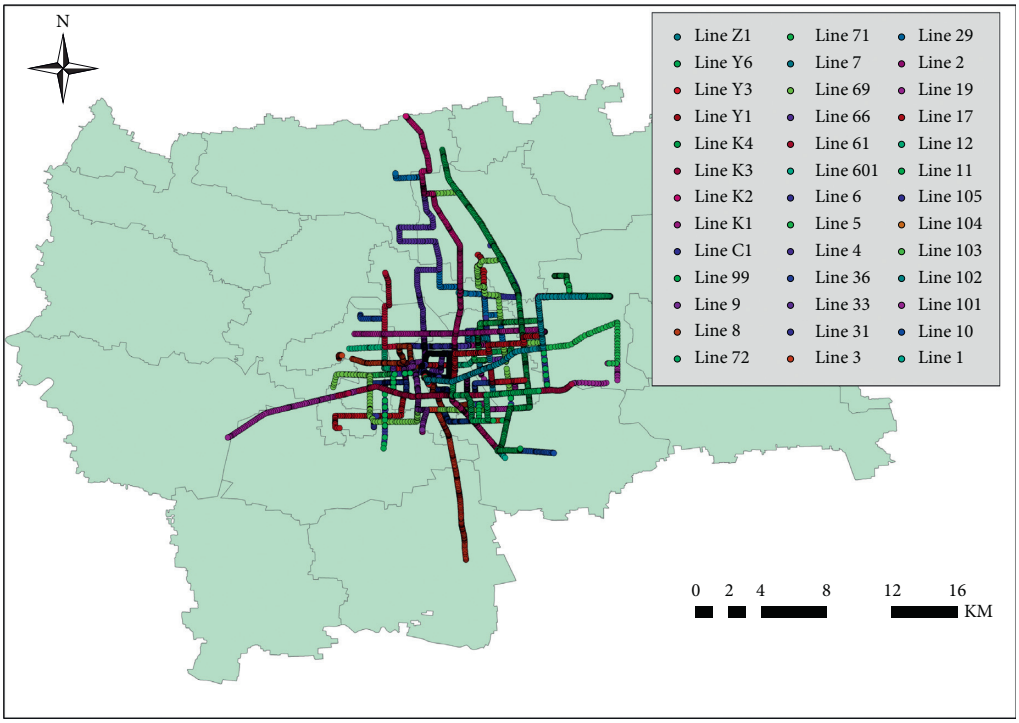


Figure 1: GPS trajectory points of bus lines.

time of the nearest bus that runs to a station at a certain time. This index is correlated with the position of the buses. When a bus delays, the operational efficiency decreases. It is known that the time of sending GPS message is out of synchronization. To solve the problem, we introduce a small time period and choose the last GPS message in the time period.

The average arrival rate of stations is defined as

$$\lambda_k^t = \begin{cases} \dfrac{1}{\left(1/s_{\Delta t}^k\right)\sum_{i=1}^{L}\sum_{j=1}^{O_i}\left(t_{ij}^k - t_{ij}^c + \Delta t\right)\gamma_{ijk}^{\Delta t}}, & t_{ij}^k - t_{ij}^c \geq 0, \\ 0, & t_{ij}^k - t_{ij}^c < 0, \end{cases} \tag{3}$$

where L is the number of bus routes, O_i is the number of buses of i-th route that running to k station, Δt is the given time period, $\Delta t \geq \max(t_{\text{send}})$, t_{send} is the time interval of sending message, t_{ij}^k is the arrival time of i-th route of j-th vehicle at station k, t_{ij}^c is the last message time of i-th route of j-th vehicle in the time period $t + \Delta t$, and $s_{\Delta t}^k$ is the number of buses that runs to station k. $\gamma_{ijk}^{\Delta t} = 1$ if i-th route of j-th vehicle is running to station k; otherwise, $\gamma_{ijk}^{\Delta t} = 0$. Figure 3 shows the sketch map of arrival efficiency of bus station.

3.3. Structural Measures. In many complex networks, the structure of network has an important impact on system functions. Some nodes with a large degree or betweenness play a critical role in system dynamics. During the past two decades, there are many studies focusing on the topological structure of public transport networks such as bus and metro. In this part, we introduce four main network-based indicators to study the relationship between bus network structure and operational performance.

3.3.1. Node Degree. Node degree is defined as the number of nodes that are connected with the node. Typically, nodes with a very large degree account for a very small proportion. It is defined as

$$k_i = \sum_{j \neq i}^{N} e_{ij}, \tag{4}$$

where k_i denotes the degree of node i, e_{ij} is the connection status between node i and node j, and $e_{ij} = 1$ means that there exist connections between node i and node j; otherwise $e_{ij} = 0$.

3.3.2. Betweenness. In transportation networks, transport efficiency is very crucial. Betweenness is another index to evaluate the importance of nodes in propagation. It is calculated as follows:

$$C_B(v) = \sum_{s \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \tag{5}$$

where $C_B(v)$ is the betweenness value of node v, σ_{st} is the number of shortest paths going from s to t, and $\sigma_{st}(v)$ is the number of the shortest paths going from s to t through the node v.

3.3.3. Clustering Coefficient. Clustering coefficient reflects the connections among the neighbors of a node. It is defined as

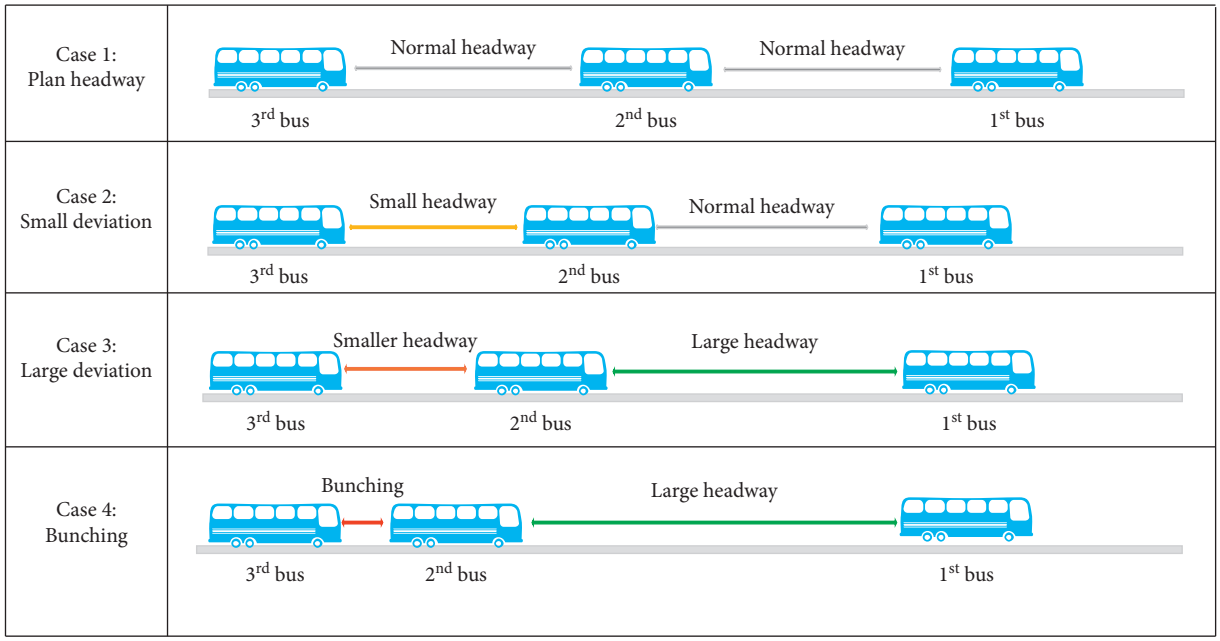


Figure 2: Four kinds of bus operation status.

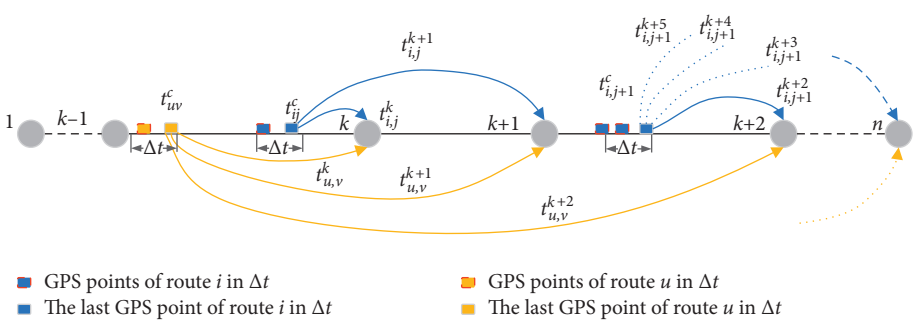


Figure 3: Sketch map of arrival efficiency of bus station.

$$C(v_i) = \frac{2e_i}{m_i(m_i - 1)}, \tag{6}$$

where $C(v_i)$ is the clustering coefficient of node v_i, e_i is the number of edges among local neighbors of node v_i, and m_i is the connection degree of neighbors of node v_i.

3.3.4. PageRank. PageRank is widely used to measure node importance in many networks. PageRank could grasp the global topological information. Thus, we introduce the PageRank algorithm to achieve the importance of nodes, which is defined as

$$p_i = \frac{1}{N}(1 - p) + p * \sum_{j=1}^{N} \frac{a_{ji}}{k_j^{\text{out}}} * p_j, \tag{7}$$

where p_i denotes the influence score of ith node, p is the damping coefficient, k_j^{out} means the out-degree of jth node, and a_{ji} is the adjacency matrix.

4. Data Description

The data were collected on October 25, 2018, in Xuchang, China. Xuchang is a famous historical city that is located in the central Henan Province. The studied bus network of Xuchang comprises 39 routes and 629 stations. Figure 4 exhibits the bus network of Xuchang, China. There are three kinds of data that contain GPS trajectory data, IC card data, and bus routes data.

Table 1 shows the basic information of bus route data, which includes line number, direction, station index, station name, longitude, and latitude. Direction "0" represents the upstream route and "1" represents the downstream route. The bus network can be constructed by the bus route information. It is noted that the station index in the table is given for a route. In this paper, each station will be given a unique index when generating the bus network. Table 2 gives the main information of GPS trajectory data, which contains line number, bus number, date, time, longitude, latitude, station index, and direction. The GPS devices of bus send a message every 10–15 seconds. The station index "0" means

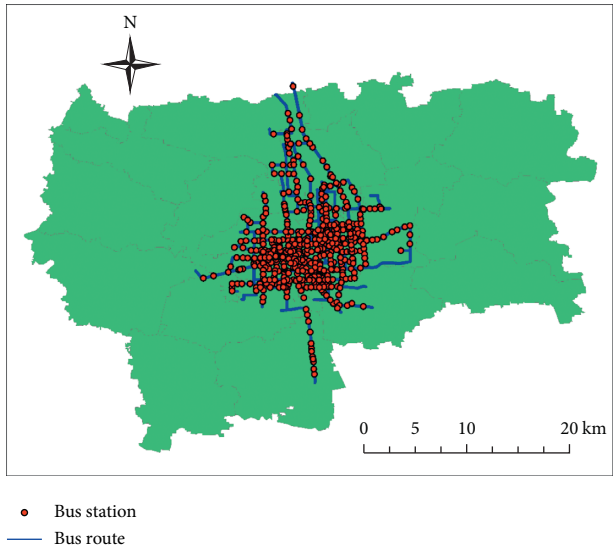


Figure 4: Bus network of Xuchang, China.

Table 1: Illustration of bus routes data of Xuchang.

Line no.	Direction	Station index	Station name	Longitude	Latitude
1	0	1	Huilong Guoji	113.7723	34.0141
1	0	2	Shiba Zhongxue	113.7761	34.0150
.....	.....	.....	.....	.....	.....
1	0	36	Shangmaocheng	113.8607	33.9769
2	1	1	Gongjiao Gongsi	113.8808	34.0675
2	1	2	Dazhang Shequ	113.8777	34.0648
.....	.....	.....	.....	.....	.....

Direction: "0" represents upstream route and "1" represents downstream route.

Table 2: Main information of GPS trajectory data.

Line no.	Bus no.	Date	Time	Longitude	Latitude	Station index	Direction
1	1511	2018/10/25	6:38:17	113.7723	34.0141	0	0
1	1511	2018/10/25	6:38:31	113.7725	34.0142	0	0
.....	.....	.....	.....	.....	.....	.....	.....
1	1511	2018/10/25	6:40:27	113.7761	34.0150	2	0
.....	.....	.....	.....	.....	.....	.....	.....
1				113.7812	34.0161	0	0
.....	.....	.....	.....	.....	.....	.....	.....

Station index: "0" represents nonstation.

that the buses are running in the sections between stations. The running status of each bus can be achieved by the GPS trajectory data. In addition, we can get the departure and arrival information of stations for each bus. Table 3 introduces the main information of IC card, which comprises card number, line number, bus number, date, time, and direction. We can obtain the boarding number of passengers by joining with GPS trajectory data.

The data were preprocessed to eliminate the outliers. The duplicated and missing values are removed. Moreover, the longitudes and latitudes out of the studied area are removed. The bus network will be constructed by the bus routes data.

Table 3: Information of IC card data.

Card no.	Line no.	Bus no.	Date	Time	Direction
EB3F10	1	1511	2018/10/25	7:27:16	0
000F2A	1	1511	2018/10/25	7:36:28	0
EB278D	1	1511	2018/10/25	8:28:06	0
.....	.....	.....	.....	.....	.....

The performance of the bus network could be calculated by the proposed indicators with given data. Then the statistical analysis and visualization will be done. In this paper, we use ArcGIS 10.3 to visualize the data.

5. Results

5.1. Traffic Demand and Bus Vehicle. Passenger flow and bus supply of bus play critical roles in daily management. On one hand, bus will delay at stations if there are a large number of boarding and alighting passengers. On the other hand, bus operational efficiency will enhance if the operational company provides more buses. Figure 5 shows the number of boarding passengers and bus vehicles during different hours in a day. We can see that the distributions of boarding passengers exhibit two obvious peaks in morning and afternoon. During the morning peak (7 : 00–9 : 00) and evening peak (16 : 00–18 : 00), there are many commuters. The peaks are not sharp; that is because some elder people prefer to travel to avoid peak hours [63]. The number of bus distributions is smooth and the large value appears in the morning peak hours.

Figure 6 shows the distributions of bus number and boarding passenger number of stations in a day. As we can see that the bus number of stations follows a lognormal distribution and the boarding passenger number of stations follows an exponential distribution. The results indicate that the stations that have a large value of bus number and boarding passenger number account for small proportions. In bus networks, the lognormal and exponential distributions are ubiquitous. It is reported the average dwell time and headway deviation of stations follow lognormal distributions [25]. Bus network structure could have some impacts on the results. A small proportion of stations in the bus network have more connections and most stations have a few connections. The degree of stations follows power-law distributions or exponential distributions [55].

Figure 7 shows the spatial distributions of bus number and boarding passenger numbers. We can see that the stations with large values are located in the central city. The stations with a large number of boarding passengers are more concentrated. Operators should pay more attention to the balance of operational status of the bus network.

5.2. Average Headway of Stations. Bus headway is an important index to measure the operational efficiency and reliability. The headway of a route is changing according to many stochastic factors such as the number of boarding and alighting passengers, traffic signal, and weather conditions. For passengers, a stable and small headway is expected. Unstable headway will disturb passengers' travel plan and large headway will enhance the waiting time. For operators, stable headway could enhance the efficiency of bus routes. There are a large number of studies focusing on strategies to keep the headway stable by holding bus, skipping stations, and controlling signals [64–66]. Bus bunching is a serious problem about headway stability in daily operation, which means two or more buses of one route arrive at the same station simultaneously. Figure 8 exhibits the running maps of route 9 and route 16. As can be seen, there are many buses running with small headways in route 16 during the morning and evening peak hours.

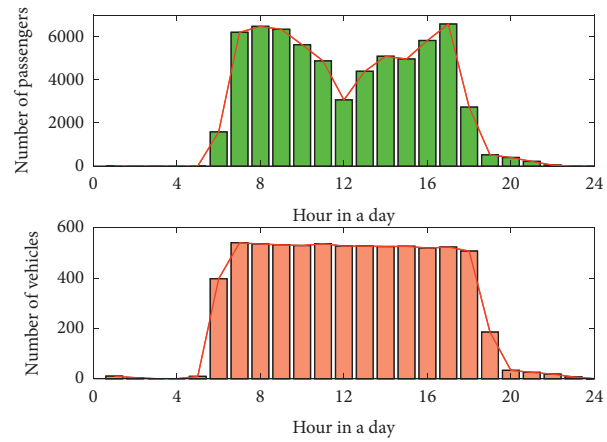


Figure 5: Number of passengers and vehicles of bus system per hour in a day.

Traditional headway studies mostly focus on single routes. The situation of the entire network headway is not very clear. Understanding the headway condition of bus networks could provide a macroscopic view for planner and manager when making schedules and strategies. In this paper, we use the proposed average headway of stations to evaluate the whole network headway conditions. Figure 9 shows the distributions of average headway of stations in different time periods: (a) 8 : 00–9 : 00, (b) 10 : 00–11 : 00, (c) 12 : 00–13 : 00, (d) 14 : 00–15 : 00, (e) 16 : 00–17 : 00 , and (f) 18 : 00–19 : 00. The red curve is the fitted curves of lognormal. We can see that the average station headways follow lognormal distributions with obvious right tails. Most values are concentrated between 100 seconds and 1000 seconds. Stations with large values more than 1000 seconds account for a small proportion.

Figure 10 shows the average headway of station values of all stations in a day. We can see that there are some stations that have very large values. The top five stations are Ruixianglu-Gongnonglu, Xuchang Shiyan School, Jianglijijie, Yangguangdadao Dongkou, and Nongji Wuliuyuan. In addition, there is a large difference in the deviation of the values in different hours in a day among the stations. The standard deviations range from 6.32 to 2780.

5.3. Average Arrival Rate of Station. The average arrival rate of station is an index to measure the instantaneous operational efficiency of stations, which can be achieved by the high dense GPS trajectory data. In this part, we mainly consider the buses that run to the nearest station. Figure 11 shows the average arrival rate of station that there exit buses running to them at 8 : 00, 10 : 00, 12 : 00, 14 : 00, 16 : 00, and 18 : 00. As we can see, the distributions of the values are inhomogeneous. A small number of stations have large values, while most stations have small values. In the morning and evening peak hours, there are many stations that have a large value. The reason is that there are more buses running on the road in those time periods. For one station, the average arrival rate is changing as time goes. The indicator could measure the station service quality instantaneously.

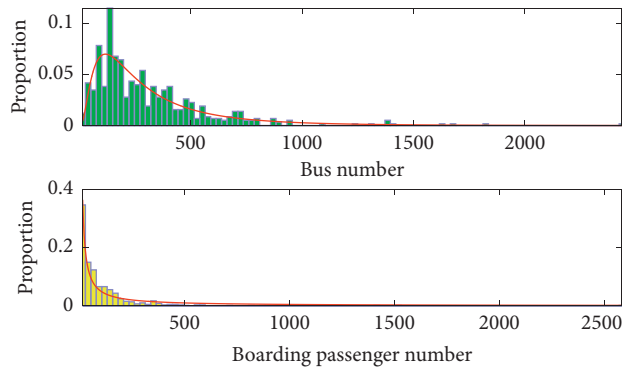


Figure 6: Distributions of bus number and boarding passenger number.

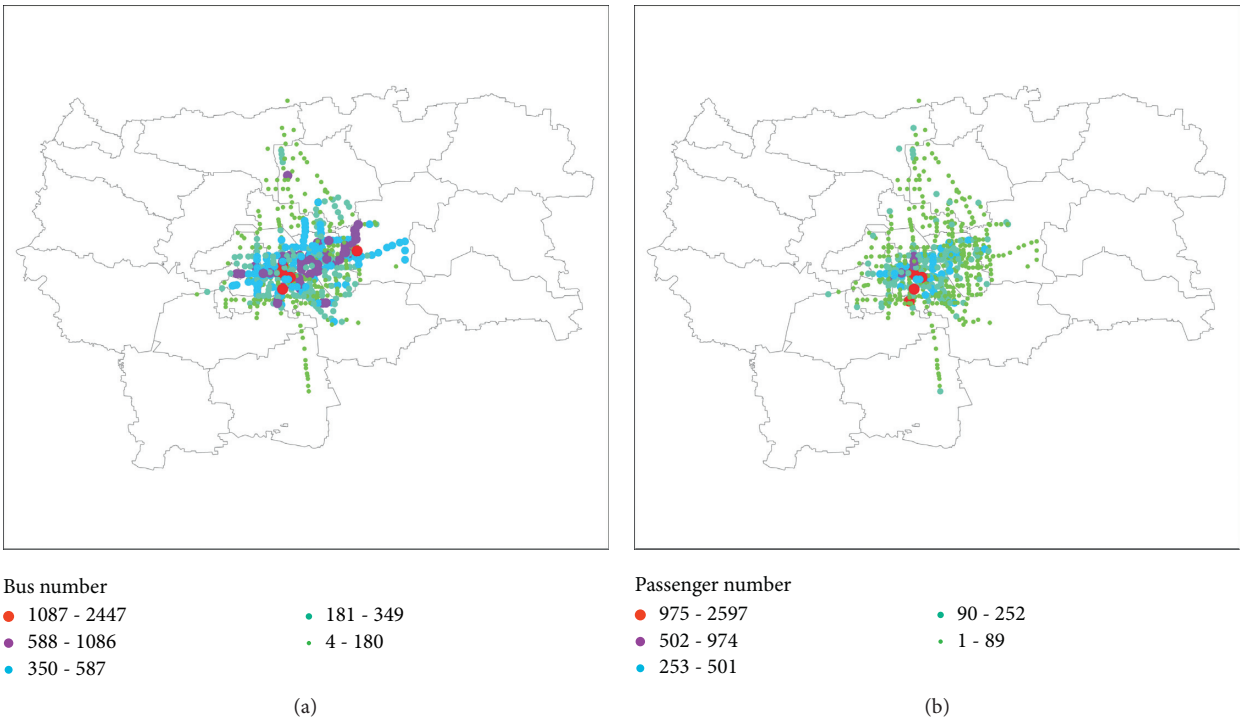


Figure 7: Maps of spatial distributions of (a) bus number and (b) passenger number.

Operators could make the targeted policy to enhance the arrival efficiency of stations by providing more buses.

Figure 12 shows the mean value of the average arrival rate of station at different hours of a day. We can see that the values are large in the peak hours and small in the off-peak hours. The value is very small after 20 : 00, because most routes stop running on the road. It is noted that there are some time periods that have larger values. Due to some stochastic factors, the headways are affected and lose homogeneity. Operators should provide a more robust timetable to keep the values stable. As we know, the stability of headway plays an important role in residents' travel mode choice. High operation efficiency and stable headway could attract more people to use the public transport. The headway control could not pay attention to a single route but the entire bus network.

5.4. Topological Structural Characteristics. In many complex networks, the topology structure is of significance in system function. We introduce node degree, betweenness, clustering coefficient, and PageRank index to evaluate the structural characteristics of the studied bus network. Figure 13 shows the spatial distributions of the four indicators. There is a small proportion of stations that have very large values of the four indicators, which are concentrated on the center of the city. Take node degree, for example; there is

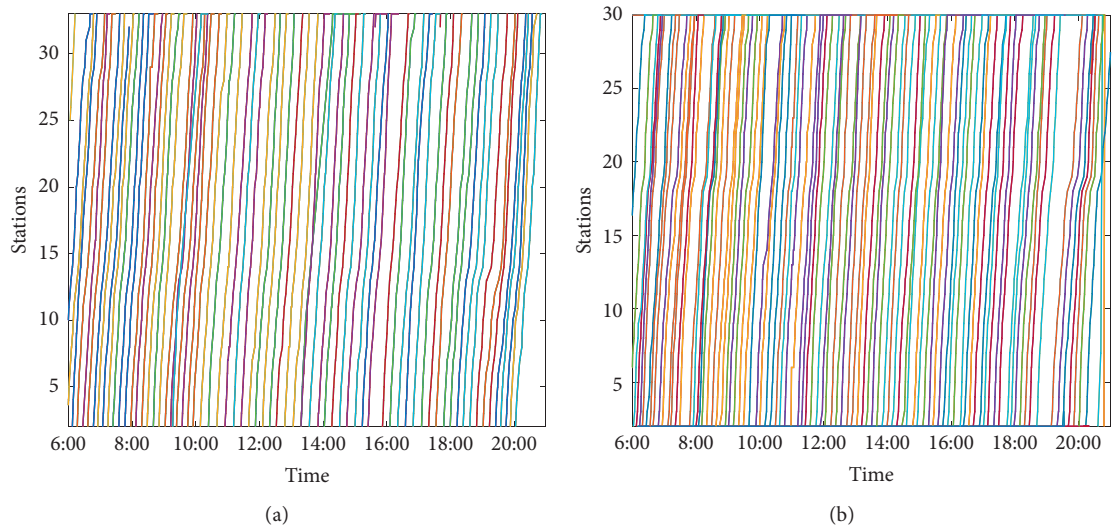


Figure 8: Running maps of (a) route 9 and (b) route 16.

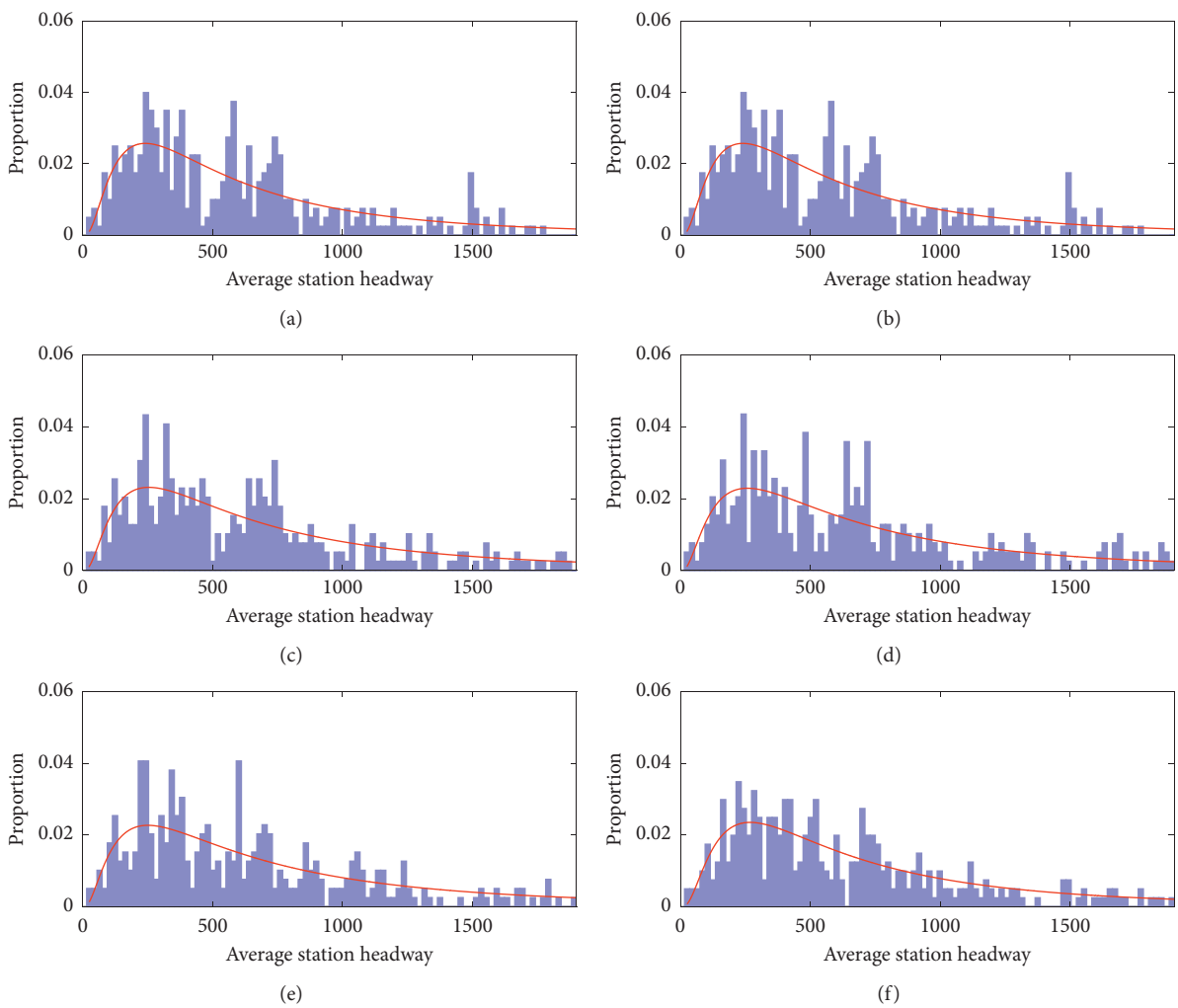


Figure 9: Average station headway distributions in different time periods: (a) 8 : 00–9 : 00, (b) 10 : 00–11 : 00, (c) 12 : 00–13 : 00, (d) 14 : 00–15 : 00, (e) 16 : 00–17 : 00, and (f) 18 : 00–19 : 00.

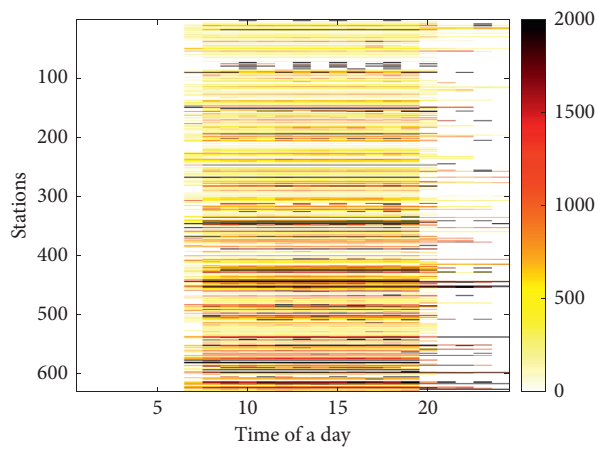


Figure 10: Heatmap of average headway of stations in a day.

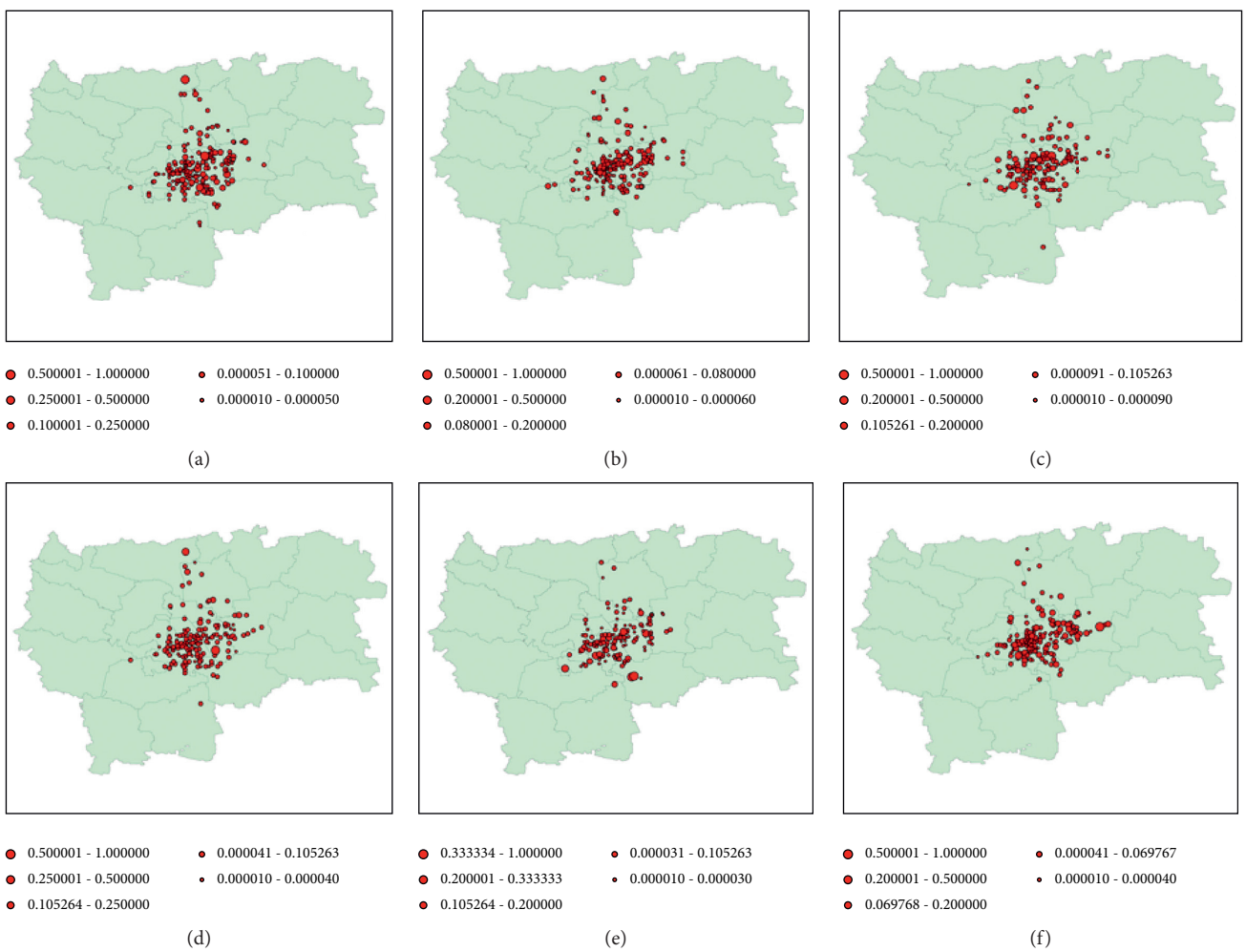


Figure 11: Average arrival rate of stations: (a) 8 : 00, (b) 10 : 00, (c) 12 : 00, (d) 14 : 00, (e) 16 : 00, and (f) 18 : 00.

only 5.2% of stations that the degree is larger than 8. The top five stations are Gaotiedong Zhan, Keyunbei Zhan, Beihai Park, Railway station, and Jianan Hospital. Those stations are transport hub which attract numerous passengers.

5.5. Relationship between Indicators. To explore the relationship among passengers, operational performance, and the connections of stations, Figure 14 shows the heatmap of the Pearson correlation coefficient among these indicators.

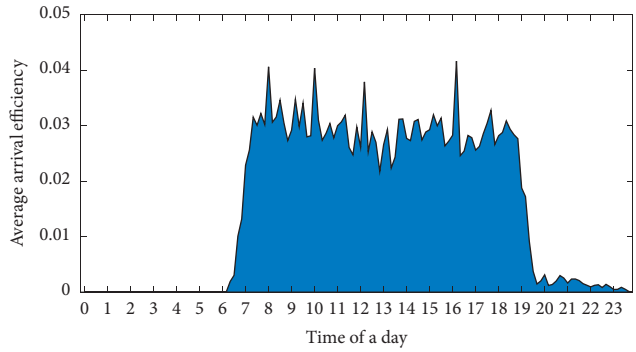


Figure 12: Mean value of average arrival rate of stations in a day.

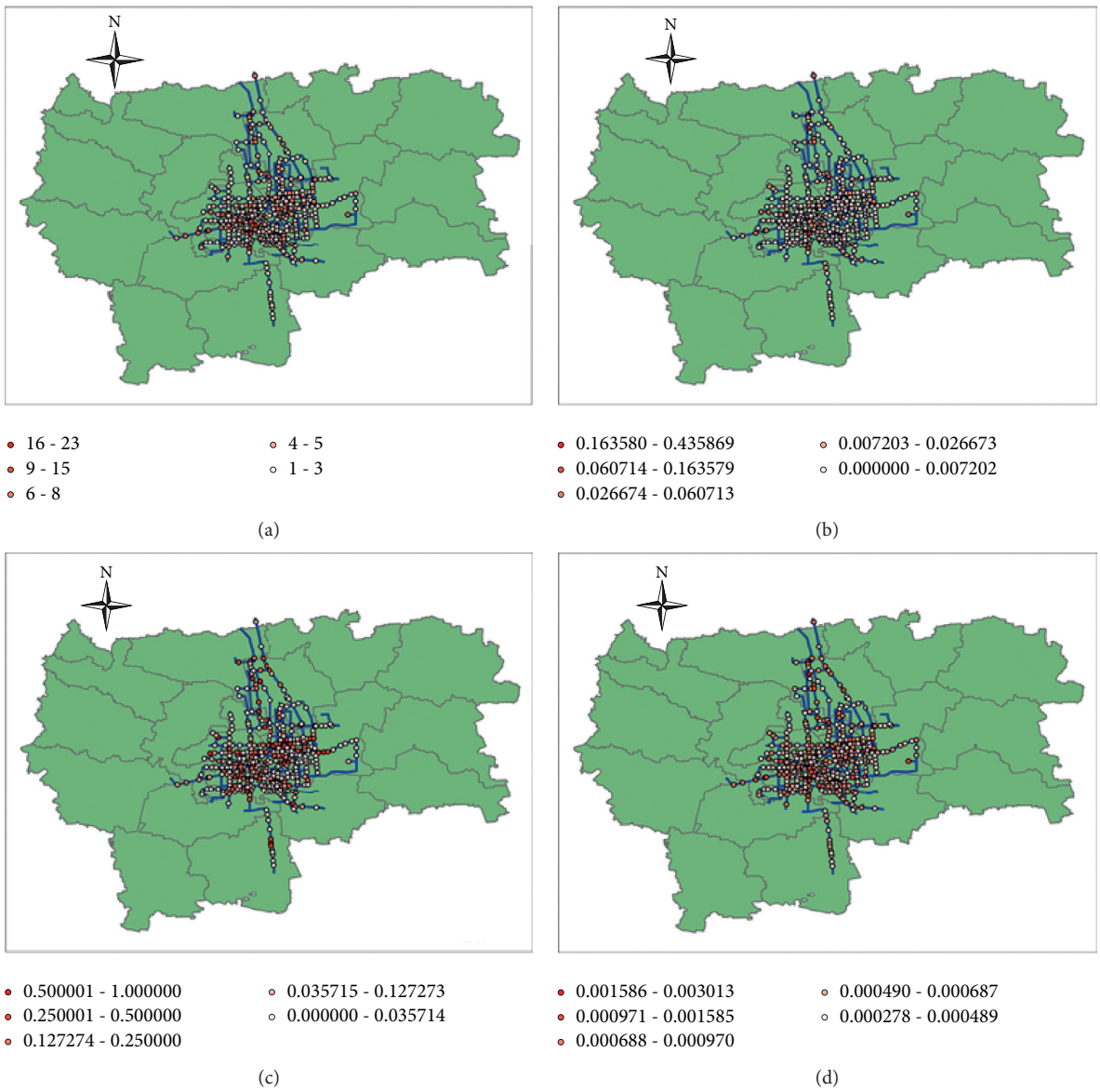


Figure 13: Distributions of bus topological structure properties: (a) degree, (b) betweenness, (c) clustering coefficient, and (d) PageRank.

As can be seen, the boarding number is strongly correlated with the bus number. Moreover, the boarding number is positively correlated with topology structure indicators, which means that a better network connection could attract more passenger flows. It is remarkable that the average headway of station is negative with boarding passenger

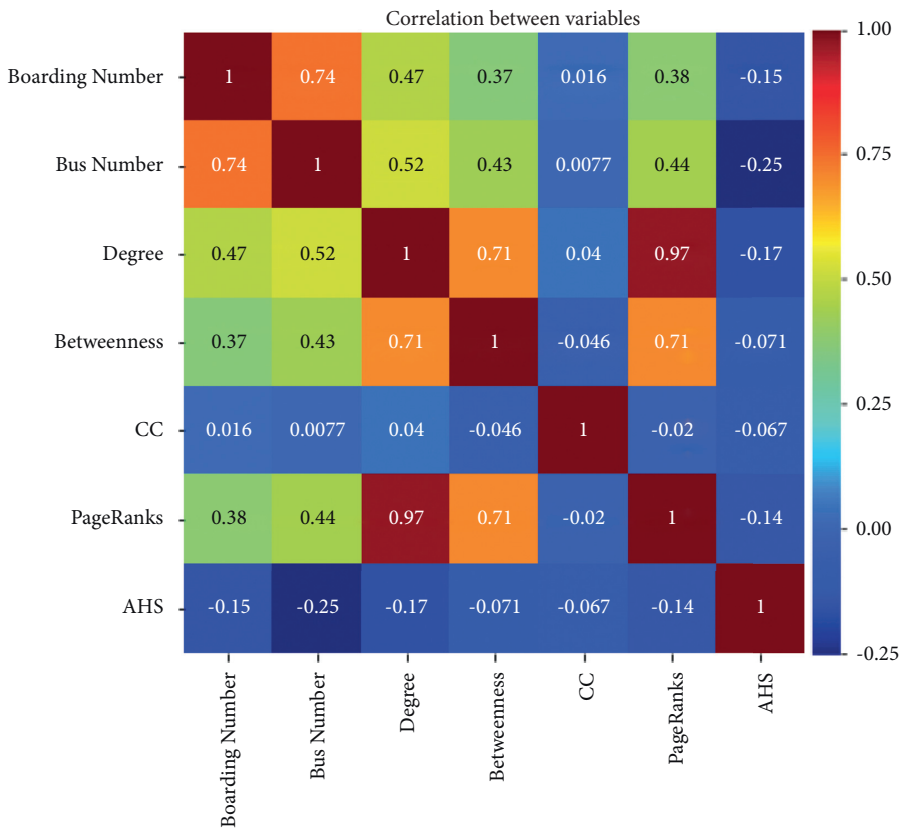


Figure 14: Correlations of main indicators.

number, which demonstrates that the passenger number is a critical factor to disturb the stability of operation. For the network structure, the PageRank index has a strong positive correlation with node degree. Clustering coefficient could reflect the local connections. The figure shows clustering coefficient nearly has no correlations with other indicators, which indicates that local connections hardly impact the operation. Operators should pay more attention to the entire network. We can see that the average headway of station is negatively correlated with other indicators, which implies that complex network connections and more passenger flows could weaken the efficiency of bus operation.

6. Conclusions

Bus system is a complicated spatiotemporal network, which plays a key role in alleviating traffic congestions. Understanding the operational status of the whole network is of significance for improving the service quality. The bus system often suffers from delay and bunching due to many factors such as traffic condition of road, weather, traffic light, and number of boarding and alighting passengers. Any disturbance of bus operation could cause the cascade reaction from stops to a route, even to the entire network. The efficiency and stability are two key indicators to measure the bus operation.

To evaluate the operational performance of the bus network, this paper proposes a spatiotemporal analysis of the bus network based on GPS trajectory data, IC card data, and route data. We build average headway of station and arrival efficiency of station to evaluate the bus network operation. The results show that the bus number and boarding passenger number of bus network follow lognormal and exponential distributions. Moreover, the average headway of stations follows a lognormal distribution. There exist some stations, where the average headway of the station is very large during the operation time. Managers should arrange more buses for the routes that serve these stations. The distributions of the average arrival rate of stations are inhomogeneous. A small number of stations have large values, while most stations have small values. We test the relationships among passengers, operational performance, and the connections of stations. It is found that the average headway of station is negatively correlated with other indicators, which implies that complex network connections and large passenger flows could weaken the efficiency of bus operation. In addition, the boarding number is strongly correlated with bus number. The boarding number is positively correlated with topology structure indicators, which means that a better network connection could attract more passengers to use public transport.

This paper will promote the study of stability and efficiency of bus system from a single route to the entire network, which has important theoretical and practical meaning for bus systematic management and control. The limitations of this study are as follows. This paper did not involve the number of alighting passengers due to data limitation. The external factors, such as traffic condition and

weather, have not been considered in the paper. In the future study, more factors will be considered and prediction model will be studied.

Data Availability

The data are available by contacting the corresponding author.

Conflicts of Interest

The authors declare there are no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (42001396 and 41901396) and the Beijing Jingwei Information Technology Co., Ltd. Scientific Research Project (DZYF20-02).

References

[1] C. R. Sekhar, T. Iryo, and Y. Asakura, "Analysis of travel-time variation over multiple sections of Hanshin expressway in Japan," *Current Science*, vol. 102, no. 11, pp. 1527–1533, 2012.

[2] N. Julio, R. Giesen, and P. Lizana, "Real-time prediction of bus travel speeds using traffic shockwaves and machine learning algorithms," *Research in Transportation Economics*, vol. 59, pp. 250–257, 2016.

[3] A. H. F. Chow, S. Li, and R. Zhong, "Multi-objective optimal control formulations for bus service reliability with traffic signals," *Transportation Research Part B: Methodological*, vol. 103, pp. 248–268, 2017.

[4] S. M. H. Moosavi, A. Ismail, and A. Balali, "Evaluating bus dwell time at key stops using automatic data collection systems," *ITE Journal-Institute of Transportation Engineers*, vol. 87, no. 11, pp. 45–49, 2017.

[5] Y. Park, J. Mount, L. Liu, N. Xiao, and H. J. Miller, "Assessing public transit performance using real-time data: spatiotemporal patterns of bus operation delays in Columbus, Ohio, USA," *International Journal of Geographical Information Science*, vol. 34, no. 2, pp. 367–392, 2020.

[6] Y. J. Deng, X. Luo, X. B. Hu, Y. F. Ma, and R. Ma, "Modeling and prediction of bus operation states for bunching analysis," *Journal of Transportation Engineering Part A*, vol. 146, no. 9, Article ID 04020106, 2020.

[7] A. Khadhir, B. A. Kumar, and L. D. Vanajakshi, "Analysis of global positioning system based bus travel time data and its use for advanced public transportation system applications," *Journal of Intelligent Transportation Systems*, vol. 25, no. 1, pp. 58–76, 2021.

[8] Y. J. Deng, X. H. Liu, X. B. Hu, and M. Zhang, "Reduce bus bunching with a real-time speed control algorithm considering heterogeneous roadway conditions and intersection delays," *Journal of Transportation Engineering Part A*, vol. 146, no. 7, Article ID 04020048, 2020.

[9] J. Argote-Cabanero, C. F. Daganzo, and J. W. Lynn, "Dynamic control of complex transit systems," *Transportation Research Part B: Methodological*, vol. 81, pp. 146–160, 2015.

[10] J.-D. Schmöcker, W. Sun, A. Fonzone, and R. Liu, "Bus bunching along a corridor served by two lines," *Transportation Research Part B: Methodological*, vol. 93, pp. 300–317, 2016.

[11] Q. Huang, B. Jia, R. Jiang, and S. Qiang, "Simulation-based optimization in a bidirectional A/B skip-stop bus service," *IEEE Access*, vol. 5, pp. 15478–15489, 2017.

[12] S. J. Berrebi, E. Hans, N. Chiabaut, J. A. Laval, L. Leclercq, and K. E. Watkins, "Comparing bus holding methods with and without real-time predictions," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 197–211, 2018.

[13] N. B. Stoll, T. Glick, and M. A. Figliozzi, "Using high-resolution bus GPS data to visualize and identify congestion hot spots in urban arterials," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2539, no. 1, pp. 20–29, 2016.

[14] J. Seo, S. H. Cho, D. K. Kim, and P. Y. J. Park, "Analysis of overlapping origin-destination pairs between bus stations to enhance the efficiency of bus operations," *IET Intelligent Transport Systems*, vol. 14, no. 6, pp. 545–553, 2020.

[15] A. Webb, P. Kumar, and A. Khani, "Estimation of passenger waiting time using automatically collected transit data," *Public Transport*, vol. 12, no. 2, pp. 299–311, 2020.

[16] L. Liu and H. J. Miller, "Measuring risk of missing transfers in public transit systems using high-resolution schedule and real-time bus location data," *Urban Studies*, 2020.

[17] T. Zhang, Y. Li, H. Yang, C. Cui, J. Li, and Q. Qiao, "Identifying primary public transit corridors using multi-source big transit data," *International Journal of Geographical Information Science*, vol. 34, no. 6, pp. 1137–1161, 2020.

[18] J. Li, P. Zheng, and W. Zhang, "Identifying the spatial distribution of public transportation trips by node and community characteristics," *Transportation Planning and Technology*, vol. 43, no. 3, pp. 325–340, 2020.

[19] F. Chen, J. Zhang, Z. Wang, S. Shi, and H. Liu, "Passenger travel characteristics and bus operational states: a study based on IC card and GPS data in Yinchuan, China," *Transportation Planning and Technology*, vol. 42, no. 8, pp. 825–847, 2019.

[20] H. Zhang, B. Y. Shi, S. G. Song, Q. M. Zhao, X. M. Yao, and W. Wang, "Statistical analysis of the stability of bus vehicles based on GPS trajectory data," *Modern Physics Letters B*, vol. 33, no. 3, Article ID 1950015, 2019.

[21] H. Zhang, C. Zhuge, and X. Yu, "Identifying hub stations and important lines of bus networks: a case study in Xiamen, China," *Physica A: Statistical Mechanics and Its Applications*, vol. 502, pp. 394–402, 2018.

[22] Z. Neal, "Is the urban world small? The evidence for small world structure in urban networks," *Networks and Spatial Economics*, vol. 18, no. 3, pp. 615–631, 2018.

[23] T. N. Maeda, J. Mori, I. Hayashi, T. Sakimoto, and I. Sakata, "Comparative examination of network clustering methods for extracting community structures of a city from public transportation Smart Card Data," *IEEE Access*, vol. 7, pp. 53377–53391, 2019.

[24] Y. Sui, F. Shao, X. Yu, R. Sun, and S. Li, "Public transport network model based on layer operations," *Physica A: Statistical Mechanics and Its Applications*, vol. 523, pp. 984–995, 2019.

[25] H. Zhang, H. Cui, and B. Shi, "A data-driven analysis for operational vehicle performance of public transport network," *IEEE Access*, vol. 7, pp. 96404–96413, 2019.

[26] C. A. Iliopoulou, C. P. Milioti, E. I. Vlahogianni, and K. L. Kepaptsoglou, "Identifying spatio-temporal patterns of bus bunching in urban networks," *Journal of Intelligent Transportation Systems*, vol. 24, no. 4, pp. 365–382, 2020.

[27] J. J. Tang, Z. T. Li, F. Gao, and F. Zong, "Identifying critical metro stations in multiplex network based on D-S evidence theory,," *Physica A*, vol. 574, Article ID 126018, 2021.

[28] A. A. De Bona, M. D. Rosa, K. V. O. Fonseca, and R. Luders, "A reduced model for complex network analysis of public transportation systems," *Physica A*, vol. 567, Article ID 125715, 2021.

[29] Y. Wang, Y. Y. Yuan, X. Y. Guan, M. Z. Xu, L. Wang, and H. Z. Wang, "Collaborative two-echelon multicenter vehicle routing optimization based on state-space-time network representation," *Journal of Cleaner Production*, vol. 258, Article ID 120590, 2020.

[30] Y. Wang, S. G. Peng, X. S. Zhou, M. Mahmoudi, and L. Zhen, "Green logistics location-routing problem with eco-packages," *Transportation Research Part E*, vol. 143, Article ID 102118, 2020.

[31] T. Li and L. Rong, "Impacts of service feature on vulnerability analysis of high-speed rail network," *Transport Policy*, vol. 110, pp. 238–253, 2021.

[32] J. Tang, S. Zhang, X. Chen, F. Liu, and Y. Zou, "Taxi trips distribution modeling based on entropy-maximizing theory: a case study in Harbin city-China," *Physica A: Statistical Mechanics and Its Applications*, vol. 493, pp. 430–443, 2018.

[33] H. Zhang, L. Zhang, F. Che, J. Jia, and B. Shi, "Revealing urban traffic demand by constructing dynamic networks with taxi trajectory data," *IEEE Access*, vol. 8, pp. 147673–147681, 2020.

[34] L. Ni, X. Wang, and X. Chen, "A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 510–526, 2018.

[35] M. G. Demissie, S. Phithakkitnukoon, and L. Kattan, "Trip distribution modeling using mobile phone data: emphasis on intra-zonal trips," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 7, pp. 2605–2617, 2019.

[36] H. Zhang, C. X. Zhuge, J. M. Jia, B. Y. Shi, and W. Wang, "Green travel mobility of dockless bike-sharing based on trip data in big cities: a spatial network analysis," *Journal of Cleaner Production*, vol. 313, Article ID 127930, 2021.

[37] Z. Tian, J. Zhou, and M. Wang, "Dynamic evolution of demand fluctuation in bike-sharing systems for green travel," *Journal of Cleaner Production*, vol. 231, pp. 1364–1374, 2019.

[38] Y. L. Lu, U. Benlic, and Q. H. Wu, "An effective memetic algorithm for the generalized bike-sharing rebalancing problem," *Engineering Applications of Artificial Intelligence*, vol. 95, Article ID 103890, 2020.

[39] Z. Gurmu and W. Fan, "Artificial neural network travel time prediction model for buses using only GPS data," *Journal of Public Transportation*, vol. 17, no. 2, pp. 45–65, 2014.

[40] W. Tu, R. Cao, Y. Yue, B. Zhou, Q. Li, and Q. Li, "Spatial variations in urban public ridership derived from GPS trajectories and smart card data," *Journal of Transport Geography*, vol. 69, pp. 45–57, 2018.

[41] J. Tang, Y. Yang, W. Hao, F. Liu, and Y. Wang, "A data-driven timetable optimization of urban bus line based on multi-objective genetic algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2417–2429, 2021.

[42] R. Zhang, P. Xie, C. Wang, G. Y. Liu, and S. H. Wang, "Classifying transportation mode and speed from trajectory data via deep multi-scale learning," *Computer Networks*, vol. 162, Article ID 106861, 2019.

[43] J. Li, X. Pei, X. Wang, D. Yao, Y. Zhang, and Y. Yue, "Transportation mode identification with GPS trajectory data and GIS information," *Tsinghua Science and Technology*, vol. 26, no. 4, pp. 403–416, 2021.

[44] A. Chepuri, J. Ramakrishnan, S. Arkatkar, G. Joshi, and S. S. Pulugurtha, "Examining travel time reliability-based performance indicators for bus routes using GPS-based bus trajectory data in India," *Journal of Transportation Engineering Part A*, vol. 144, no. 5, Article ID 04018012, 2018.

[45] W. Fan and R. B. Machemehl, "Do transit users just wait for buses or wait with strategies?" *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2111, no. 1, pp. 169–176, 2009.

[46] S. S. Moghaddam, R. Noroozi, J. M. Casello, and B. Hellinga, "Predicting the mean and variance of transit segment and route travel times," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2217, no. 1, pp. 30–37, 2011.

[47] A. Gal, A. Mandelbaum, F. Schnitzler, A. Senderovich, and M. Weidlich, "Traveling time prediction in scheduled transportation with journey segments," *Information Systems*, vol. 64, pp. 266–280, 2017.

[48] L. M. Kieu, A. Bhaskar, and E. Chung, "Public transport travel-time variability definitions and monitoring," *Journal of Transportation Engineering*, vol. 141, no. 1, Article ID 04014068, 2015.

[49] R. Rajbhandari, S. I. Chien, and J. R. Daniel, "Estimation of bus dwell times with automatic passenger counter information," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1841, no. 1, pp. 120–127, 2003.

[50] A. Tirachini, "Bus dwell time: the effect of different fare collection systems, bus floor level and age of passengers," *Transportmetrica: Transportation Science*, vol. 9, no. 1, pp. 28–49, 2013.

[51] Y. Ji, L. Gao, D. Chen, X. Ma, and R. Zhang, "How does a static measure influence passengers' boarding behaviors and bus dwell time? Simulated evidence from Nanjing bus stations," *Transportation Research Part A: Policy and Practice*, vol. 110, pp. 13–25, 2018.

[52] A. Chepuri, S. Joshi, S. Arkatkar, G. Joshi, and A. Bhaskar, "Development of new reliability measure for bus routes using trajectory data," *Transportation Letters*, vol. 12, no. 6, pp. 363–374, 2020.

[53] J. Paudel, "Bus ridership and service reliability: the case of public transportation in Western Massachusetts," *Transport Policy*, vol. 100, pp. 98–107, 2021.

[54] C. Zhang, Z. Juan, Q. Luo, and G. Xiao, "Performance evaluation of public transit systems using a combined evaluation method," *Transport Policy*, vol. 45, pp. 156–167, 2016.

[55] X. P. Yang, S. W. Lu, W. F. Zhao, and Z. Y. Zhao, "Exploring the characteristics of an intra-urban bus service network: a case study of Shenzhen, China," *ISPRS International Journal of Geo-Information*, vol. 8, no. 11, Article ID 486, 2019.

[56] G. L. Jia, R. G. Ma, and Z. H. Hu, "Urban transit network properties evaluation and optimization based on complex network theory," *Sustainability*, vol. 11, p. 2007, 2019.

[57] Y. J. Wang, Y. Deng, F. Ren et al., "Analysis the spatial configuration of urban bus networks based on the geospatial network analysis method," *Cities*, vol. 96, Article ID 102406, 2020.

[58] L.-S. Sun, S.-W. Wang, L.-Y. Yao, J. Rong, and J.-M. Ma, "Estimate of transit ridership based on spatial analysis and precise land use data," *Transportation Letters*, vol. 8, no. 3, pp. 140–147, 2016.

[59] Y. Zuo, Z. Liu, and X. Fu, "Measuring accessibility of bus system based on multi-source traffic data," *Geo-Spatial Information Science*, vol. 23, no. 3, pp. 248–257, 2020.

[60] H. Zhang, X. Li, L. Zhang, W. Wang, J. Jia, and B. Shi, "Discovering station patterns of urban transit network with multisource data: empirical evidence in Jinan, China," *KSCE Journal of Civil Engineering*, vol. 25, no. 2, pp. 680–691, 2021.

[61] S. Wei, W. Zheng, and L. Wang, "Understanding the configuration of bus networks in urban China from the perspective of network types and administrative division effect," *Transport Policy*, vol. 104, pp. 1–17, 2021.

[62] S. Bree, D. Fuller, and E. Diab, "Access to transit? Validating local transit accessibility measures using transit ridership," *Transportation Research Part A: Policy and Practice*, vol. 141, pp. 430–442, 2020.

[63] W. Y. Szeto, L. Yang, R. C. P. Wong, Y. C. Li, and S. C. Wong, "Spatio-temporal travel characteristics of the elderly in an ageing society," *Travel Behaviour and Society*, vol. 9, pp. 10–20, 2017.

[64] K. Gkiotsalitis and E. C. Van Berkum, "An analytic solution for real-time bus holding subject to vehicle capacity limits," *Transportation Research Part C*, vol. 121, Article ID 102815, 2020.

[65] X. M. Chen, X. M. Han, L. Yu, and C. H. Wei, "Does operation scheduling make a difference: tapping the potential of optimized design for skipping-stop strategy in reducing bus emissions," *Sustainability*, vol. 9, no. 10, p. 1737, 2017.

[66] Y. Bie, X. Xiong, Y. Yan, and X. Qu, "Dynamic headway control for high-frequency bus line based on speed guidance and intersection signal adjustment," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 1, pp. 4–25, 2020.

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 8380247, 15 pages
https://doi.org/10.1155/2021/8380247

Research Article

Research on the Evolution Mechanism of Congestion in the Entrances and Exits of Parking Facilities Based on the Improved Spatial Autoregressive Model

Hongru Yu,[1] **Shejun Deng**,[1] **Caoye Lu**,[1] **Yucheng Tang**,[1,2] **Shijun Yu**,[1] **Lu Liu**,[1] **and Tao Ji**[1]

[1]*College of Civil Science and Engineering, Yangzhou University, Yangzhou 225002, China*
[2]*China Iconic Technology Company Limited, Hefei 230088, China*

Correspondence should be addressed to Shejun Deng; yzrx6@163.com

Received 18 June 2021; Revised 1 August 2021; Accepted 13 August 2021; Published 30 August 2021

Academic Editor: Jinjun Tang

The entrance and exit area of parking facilities has the characteristics of high concentration of urban traffic and prominent traffic intertwining phenomenon, which easily induces rapid congestion of mixed heterogeneous traffic at specific times and local locations and quickly spreads to the entire road section or even a larger area. In order to better understand the congestion distribution characteristics and propagation effects of access section of the parking entrance and exit from the mid and microperspective, a 5 m $*$ lane width pixel grid is used to divide the frontage road research. It also proposes a spatially robust autoregressive model and complex network tools suitable for analysis of local traffic flow to analyze it. The results show that as spatial scale increases, the congestion propagation decreases sharply and spatial adjacency within the fourth order can account for more than 90% of the propagation; the frontage road to the entrance and exit is the place where the congestion first happens, and the congestion gradually attenuates as it propagates to the inner lane and the upstream of the road segments; the lateral congestion propagation attenuates faster, so the area affected by congestion is mainly distributed in the outermost lane. This paper can provide theoretical guidance for alleviating traffic congestion in the entrance and exit areas of parking facilities and has theoretical and empirical significance.

1. Introduction

Increasing car park [1], which exacerbates the contradiction between the supply and demand of urban parking, along with the lag of parking facilities planning [2–4] and other factors, causes more severe traffic congestion problems in the entrances and exits of parking facilities. Since the entrances and exits of parking facilities are generally set up on secondary arterial roads or branch roads with low capacity, the interweaving of incoming and outgoing vehicles with external traffic will reduce the traffic capacity [5, 6], increase the delay time of main road traffic [7, 8], and even cause a bottleneck and congestion on the upstream road.

At present, research focused on the traffic conditions in entrances and exits of the off-street parking is still very rare. Previous research mainly focuses on the complex impact of the bus stop [9–14] and on-street parking [15–18] on the traffic flow. Compared with other road facilities, the settings of entrances and exits of parking facilities can bring more complicated impact: on the one hand, due to the uncertainty of the occurrence and attraction of parking facilities, the temporal and spatial distribution of parked vehicles are more disorderly; the intermittently concentrated entering and departing of vehicles enhance the occasional disturbance to the road; on the other hand, the interwoven flow of incoming and outgoing vehicles, the traffic flow of the main road, and the flow of motor vehicles and nonmotorized vehicles [19] increase the heterogeneity of the traffic flow and the complexity of the occurrence of congestion in this area to a certain extent.

In the analysis of traffic congestion, scholars have built various models to reveal the essence of congestion. At the macrolevel [20–24], traffic congestion is generally regarded as a process of compression, blockage, and spread of traffic

flow. At the microlevel [25–29], more attention is paid to conflicts, queuing, and easing of vehicles. Spatial autocorrelation is a geographic concept, which refers to the potential interdependence between observed data of some variables in the same distribution area [30, 31]. The development of this theory in the field of transportation has resulted in its wide use in traffic feature recognition [32] at the macrolevel and mid level and accident analysis [33–36]. In addition, complex network theory, as a new scientific theory for revealing network complexity phenomena, has a wide range of applications for the organization and optimization of transportation logistics [37–40] and the macrointerpretation of transportation systems [41–46], but it is rarely used in the study of traffic flow characteristics at the meso-microlevels.

To sum up, although there are rich research outcomes on congestion propagation, research on the frontage roads of parking facilities is still relatively rare; besides, although the spatial autocorrelation theory has developed in the field of transportation, due to its higher requirements on spatial stability, it is not suitable for the research of transportation theory at the mid and microlevel. In response to the abovementioned problems, improving the spatial autocorrelation model, building a model of the spatial congestion propagation in the entrances and exits of the parking facilities from the mid and microperspective under the guidance of the complex network theory and other theories, and exploring the attenuation effect and key nodes in the process of congestion propagation based on the observed data are of great significance for alleviating traffic congestion in the entrances and exits of parking facilities.

2. Data Description

In order to determine the objective of the study, the frontage road of this article is defined as follows: the free road segments from about 140 m upstream to about 20 m downstream from the entrance and exit access points. This paper selects the entrance and exit of the parking lot of the West Affiliated Hospital of Yangzhou University as a typical off-street parking area. The frontage road is the one-way three-lane road (a part of the two-way six-lane road with a separation zone in the middle), and the upstream area is far away from the intersection. The traffic volume and entry rate vary significantly in different periods, and the road segment is free of vegetation, billboards, and other structures, which are conducive for data collection. The data set in this paper is obtained by aerial filming, with a wind-proof drone MAVIC_AIR2 recording the traffic conditions in the experimental area at an altitude of 100–150 m (each video lasts 10–15 minutes, 100 videos in total). Then, the open-source vehicle-tracking algorithm based on YOLOV5 and DeepSort [47] is used to monitor the location parameters of the moving vehicle [48, 49] in real time (as shown in Figure 1).

In order to describe the location of the lanes in an easy way, the lanes in the experimental area are numbered, where the outermost lane is Lane 3, the middle lane is Lane 2, and the innermost lane is Lane 1. In addition, in order to study the relationship of the traffic state of each spatial unit in the entrance and exit and the spatial congestion propagation, this paper divides the road into three lanes and then further divides these three lanes into 5-meter wide spatial units, which can ensure the spatial accuracy of the model and meet the requirements of velocity measurement, as shown in Figure 2.

The investigation finds that the time for vehicles to enter the parking lot is generally within the range of 3 to 13 seconds. In order to extract the traffic information of each spatial unit to the greatest extent, this paper uses a time interval of 3 seconds to calculate the average speed of vehicles in each spatial unit. The sample data are shown in Table 1.

3. Method

3.1. Improved Spatial Autoregressive Model. The traditional spatial autoregressive model requires the data to have spatial stability. However, due to the influence of external factors such as the actual environment and traffic conditions, the spatial stability of the measured data is often poor. In order to reduce the influence of external factors on the modeling results, this paper improves the traditional spatial autoregressive model and builds a robust spatial autoregressive model based on repeated observation data that consider spatial instability and multiple correlations.

The traditional spatial autoregressive model adds a spatial lag term to the general regression model, that is, the dependent variable of a spatial object is related to the independent variable on the same object, and it is also related to the independent variable and dependent variable of adjacent objects. Its general form is shown in

$$\begin{cases} z = \rho W_1 z + X\beta + \mu, \\ \mu = \lambda W_2 \mu + \varepsilon, \\ \varepsilon \sim N\left(0, \sigma^2 I\right). \end{cases} \tag{1}$$

Here, z represents the spatial attributes of the research object, or the dependent variable; X represents the spatial attribute of the adjacent objects, or the independent variable; ρ is the spatial autoregressive coefficient; β is the regression coefficient; μ represents the residual of the model; λ represents the residual spatial regression coefficient; ε represents the random error; and W_1 and W_2 represent the spatial adjacency matrix of the spatial units of frontage road.

Observing the influencing factors of the attribute value of the spatial unit in the study area, it is not difficult to find that the change of the attribute value of the spatial unit is mainly affected by the adjacent spatial unit, while the explanatory variable X has almost no contribution. Therefore, X and W_2 in formula (1) can be assigned to 0, and the first-order spatial autoregressive model can be obtained, as shown in

$$\begin{cases} z = \rho W_1 z + \varepsilon, \\ \varepsilon \sim N\left(0, \sigma^2 I\right). \end{cases} \tag{2}$$

In order to study the propagation mechanism of congestion at different spatial scales, this paper divides adjacent spatial units according to "distance" and takes the influence

Figure 1: Vehicle tracking.

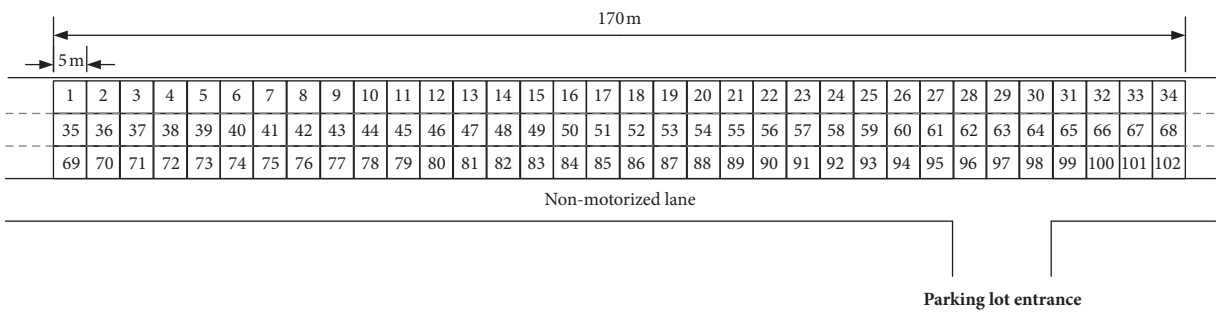


Figure 2: Grid division of road sections.

Table 1: Samples of average speeds of spatial units (unit: km/h).

Spatial unit	Time					
	1 0 s–3 s	2 3 s–6 s	3 6 s–9 s	4 9 s–12 s	... ...	110 327 s–330 s
1	41.46	41.21	41.03	43.82	...	35.78
2	41.11	40.99	40.84	44.02	...	35.56
3	40.80	40.85	40.98	44.15	...	35.39
4	40.46	40.83	41.12	44.27	...	35.23
...	...	...	...	...	...	...
102	45.12	44.93	44.46	44.46	...	40.45

of spatial units under each adjacent number on the research object as an independent variable. Taking into account the correlation between the independent variables, the direct use of traditional parameter estimation methods based on the least squares method will cause larger errors. Therefore, this paper proposes an improved first-order spatial autoregressive model that takes into account spatial instability and multiple correlations.

In the improved spatial autoregressive model, the disturbance of the research object is the dependent variable and the disturbances of other spatial units are the independent variables to describe the relationship between the spatial units in the research attributes, as shown in formula

$$y' = \sum_{k=1}^{K} \rho_k \overline{W}^k y' + \gamma l + \varepsilon. \tag{3}$$

Here, y' represents the velocity disturbance sequence of the research object; $\overline{W}^k$ represents the k-order spatial adjacency matrix considering spatial instability; ε represents the error term of the improved spatial autoregressive model; l is the constant term; ρ_k represents the spatial lag term; the regression coefficient of $\overline{W}^k$; γ represents the regression coefficient of the constant term; k represents the number of spatial adjacencies; and K is the maximum number of spatial adjacencies considered.

For the space unit under study, the historical sequence of the previous c cycle formula (4) can be constructed to obtain the velocity sequence of each space unit:

$$Y = \begin{bmatrix} y_{1,1} & y_{2,1} & \cdots & y_{n-1,1} & y_{n,1} \\ y_{1,2} & y_{2,2} & & y_{n-1,2} & y_{n,2} \\ \vdots & & \ddots & \vdots & \\ y_{1,c} & y_{2,c} & \cdots & y_{n-1,c} & y_{n,c} \end{bmatrix}. \tag{4}$$

Here, y_i represents the data of space unit i in c cycles and $y_i = [y_{i,1}, y_{i,2}, y_{i,3}, \ldots, y_{i,c}]^T$ is the speed history data sequence of space unit i.

3.1.1. Speed Disturbance Sequence. "Urban Traffic Management Evaluation Index System" (2002, China) stipulates that when the average speed of motor vehicles on the main road is not less than 30 km/h, the road is unblocked. Therefore, the speed perturbation sequence of the unit of the space is calculated according to the speed history data sequence. As shown in formulas (5) and (6), y_i is the historical velocity sequence of space unit i and $\widehat{y}$ is the historical velocity perturbation sequence:

$$y' = \widehat{y} = [\widehat{y}_1, \widehat{y}_2, \widehat{y}_3, \ldots, \widehat{y}_n]^T, \tag{5}$$

$$\widehat{y}_i = -(y_i - 30). \tag{6}$$

3.1.2. Considering the Spatial Adjacency Matrix That Is Not Stationary. The traditional spatial adjacency matrix based on the boundary adjacency method or the center-of-gravity distance method assumes that all adjacent objects have the same weight value, which is inconsistent with the actual situation of congestion propagation. On the one hand, the number of adjacencies in different spatial units is different. On the other hand, the traffic flow is not homogenous and the effects of adjacent units in different directions are not the same.

The spatial adjacency matrix determined based on the boundary adjacency method is shown in formulas (5) and (6). In the formula, $w_{i,j}^k$ indicates that there is a k-order adjacency relationship between space unit i and space unit j. Determining the spatial adjacency matrix based on the boundary adjacency method is a commonly used method to determine the adjacency relationship between area units; if two spatial units have the same boundary, they have a first-order spatial adjacency relationship and the k-order adjacency relationship can be determined according to the transitivity of the spatial adjacency relationship.

$$W^k = \begin{bmatrix} w_{1,1}^k & w_{1,2}^k & \cdots & w_{1,j}^k & \cdots & w_{1,n}^k \\ w_{2,1}^k & w_{2,2}^k & \cdots & w_{2,j}^k & \cdots & w_{2,n}^k \\ w_{i,1}^k & w_{i,2}^k & \cdots & w_{i,j}^k & \cdots & w_{i,n}^k \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ w_{n,1}^k & w_{n,2}^k & \cdots & w_{n,j}^k & \cdots & w_{n,n}^k \end{bmatrix}, \tag{7}$$

$$w_{i,j}^k = \begin{cases} 1, & i \text{ and } j \text{ have } k-\text{order adjacency}, \\ 0, & i \text{ and } j \text{ do not have } k-\text{order adjacency}. \end{cases}$$

In order to accurately describe the relationship between the spatial units, first, according to the spatial adjacency matrix W^k, determine the set of spatial units that have a k-order adjacency relationship with each spatial unit, denoted as $f_k(i, j), j = 1, 2, 3, \ldots, r, \ldots, R$; then, the rth space unit that has an adjacency relationship of order k with the space unit i is denoted as $f_k(i, r)$.

Let vector $\widehat{y}_i = [\widehat{y}_{i,1}, \widehat{y}_{i,2}, \widehat{y}_{i,3}, \ldots, \widehat{y}_{i,r}, \ldots, \widehat{y}_{i,R}]^T$ be the velocity history perturbation sequence of space unit i, and H_i^k is the matrix of velocity history perturbation sequences of all space units that have a k-order adjacency relationship with space unit i, as shown in

$$H_i^k = \left[\widehat{y}_{f_k(i,1)}, \widehat{y}_{f_k(i,2)}, \widehat{y}_{f_k(i,3)}, \ldots, \widehat{y}_{f_k(i,r)}, \ldots, \widehat{y}_{f_k(i,R)}\right]. \tag{8}$$

In order to preserve the features in H_i^k to the greatest extent and reduce their dimensionality, the principal component analysis method is used to extract the first principal component that is most similar to the velocity history disturbance sequence $\widehat{y}_i$ of space unit i from H_i^k.

Let the vector $\widehat{u}_i^k = [u_{f_k(i,1)}^k, u_{f_k(i,2)}^k, u_{f_k(i,3)}^k, \ldots, u_{f_k(i,r)}^k, \ldots, u_{f_k(i,R)}^k]^T$ be the weight value of each spatial unit that has a kth order adjacency with the spatial unit i, in order to ensure each weight value can accurately describe the degree of association between spatial units, and the covariance between $H_i^k \widehat{u}_i^k$ and $\widehat{y}_i$ should be the largest, as shown in

$$\max\left(H_i^k \widehat{u}_i^k\right)^T \widehat{y}_i. \tag{9}$$

Use the Lagrangian method to solve the model, as in formula (9):

$$F = \left(H_i^k \widehat{u}_i^k\right)^T \widehat{y}_i - \lambda\left(\widehat{u}_i^{kT} \widehat{u}_i^k - 1\right). \tag{10}$$

Calculate the partial derivative of F with $\widehat{u}_i^k$ and λ, as shown in

$$\frac{\partial F}{\partial \widehat{u}_i^k} = H_i^{kT} \widehat{y}_i - 2\lambda \widehat{u}_i^k, \tag{11}$$

$$\frac{\partial F}{\partial \lambda} = -\left(\widehat{u}_i^{kT} \widehat{u}_i^k - 1\right). \tag{12}$$

According to formula (12),

$$\widehat{u}_i^k = \frac{H_i^{kT} \widehat{y}_i}{\left\| H_i^{kT} \widehat{y}_i \right\|}. \tag{13}$$

The value $u_{f_k(i,r)}^k$ in $\widehat{u}_i^k$ represents the spatial association relationship of the rth spatial unit that has a k-order adjacency with the spatial unit i; the sign of $u_{f_k(i,r)}^k$ indicates the positive or negative correlation between the space unit i and the rth space unit that has a k-order adjacency relationship, and the absolute value indicates the relationship between the space units. The greater the correlation, the greater the absolute value, and vice versa. In actual road conditions, the propagation of congestion is the same direction, so only the positive numbers in the vector are retained to construct the smoothed spatial adjacency matrix $\overline{W}^k$, as shown in

$$\overline{W}^k = \begin{bmatrix} \overline{w}_{1,1}^k & \overline{w}_{1,2}^k & \cdots & \overline{w}_{1,j}^k & \cdots & \overline{w}_{1,n}^k \\ \overline{w}_{2,1}^k & \overline{w}_{2,2}^k & \cdots & \overline{w}_{2,j}^k & \cdots & \overline{w}_{2,n}^k \\ \overline{w}_{i,1}^k & \overline{w}_{i,2}^k & \cdots & \overline{w}_{i,j}^k & \cdots & \overline{w}_{i,n}^k \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \overline{w}_{n,1}^k & \overline{w}_{n,2}^k & \cdots & \overline{w}_{n,j}^k & \cdots & \overline{w}_{n,n}^k \end{bmatrix}, \tag{14}$$

$$\overline{w}_{i,j}^k = \begin{cases} 0, & u_{f_k(i,r)}^k \le 0, \\ \dfrac{u_{f_k(i,r)}^k}{\sum_{r=1}^{R} u_{f_k(i,r)}^k}, & u_{f_k(i,r)}^k > 0. \end{cases} \tag{15}$$

3.2. Parameter Calculation Method. In order to study the spatial propagation process of traffic congestion in the entrance and exit area of parking facilities, in the improved spatial autoregressive model, the disturbance of the research unit is used as the dependent variable and the spatial lag term $\overline{W}^k y'$ of the disturbance under different spatial adjacency k is used as an independent variable; there is a linear correlation between the spatial lag terms $\overline{W}^k y'$ under different spatial adjacency numbers k.

According to the space lag term $\overline{W}^1 y'$ of the disturbance under different number of space delay periods, the independent variable matrix is constructed, as shown in

$$Z(K) = \left[\overline{W}^1 y', \overline{W}^2 y', \overline{W}^3 y', \ldots, \overline{W}^k y', \ldots, \overline{W}^K y'\right]. \tag{16}$$

3.2.1. Improved Spatial Autoregressive Model Parameter Estimation Process. For the convenience of calculation, the dependent variable matrix y' of the improved spatial autoregressive model is marked as F_0, the independent variable matrix is marked as E_0, and the parameter estimation is performed.

Extract the first principal component t_1 from the independent variable matrix, where t_1 is the linear combination of E_0 of the independent variables as shown in formulas (17) and (18). Since the dependent variable contains only one variable, the first principal component of the dependent variable is F_0.

$$t_1 = E_0 u_1, \tag{17}$$

$$u_1 = \left[u_{1,1}, u_{1,2}, \ldots, u_{1,k}\right]^T. \tag{18}$$

For the needs of regression analysis, the extracted principal components must simultaneously meet the requirements of the first principal components t_1 and F_0 of the variables to reflect the characteristics of the variable group to the greatest extent and the correlation between the two first principal components to reach the maximum. Considering the requirements of modeling comprehensively, covariance is selected to characterize the principal components, as shown in

$$\begin{cases} < t_1, F_0 \ge u_1^T E_0^T F_0 \Longrightarrow \max, \\ u_1^T u_1 = 1. \end{cases} \tag{19}$$

According to the Lagrangian multiplier method, the following formula can be obtained:

$$L = u_1^T E_0^T F_0 - \lambda\left(u_1^T u_1 - 1\right). \tag{20}$$

Respectively, find the partial derivatives of u_1 and λ to get the following formulas:

$$\frac{\partial L}{\partial u_1} = E_0^T F_0 - 2\lambda u_1 = 0, \tag{21}$$

$$\frac{\partial L}{\partial \lambda} = -\left(u_1^T u_1 - 1\right) = 0. \tag{22}$$

According to formula (22), we can get

$$\widehat{u}_1 = \frac{E_0^T F_0}{\left\|E_0^T F_0\right\|}. \tag{23}$$

According to the principle of principal component analysis, the sample is expressed as a linear combination of components, that is, regression on t_1 is performed on E_0 and F_0, respectively, as shown in formulas (24) and (25), where E_1 and F_1 are residual matrixes.

$$E_0 = t_1 p_1 + E_1, \tag{24}$$

$$F_0 = t_1 r_1 + F_1. \tag{25}$$

According to (25), the following formulas can be obtained from the least square estimation:

$$\widehat{p}_1 = \frac{E_0^T e_1}{\left\|e_1\right\|}, \tag{26}$$

$$\widehat{r}_1 = \frac{F_0^T e_1}{\left\|e_1\right\|}. \tag{27}$$

Replace E_0 and F_0 with the residual terms E_1 and F_1. Repeat the above steps to obtain the weight vector $\widehat{u}_2$ (formula (28)) of the second principal component t_2 of E_0 and the regression coefficients $\widehat{p}_2$ (formula (29)) and r of E_1 and F_1 with respect to $\widehat{r}_2$ (formula (30)).

$$\widehat{u}_2 = \frac{E_1^T F_1}{\left\|E_1^T F_1\right\|}, \tag{28}$$

$$\widehat{p}_2 = \frac{E_1^T e_2}{\left\|e_2\right\|}, \tag{29}$$

$$\widehat{r}_2 = \frac{F_1^T e_2}{\left\|e_2\right\|}. \tag{30}$$

Assuming that a total of h principal components are extracted, the regression model between F_0 and E_0 can be obtained, as shown in

$$F_0 = e_1\hat{r}_1 + e_2\hat{r}_2 + e_3\hat{r}_3 + \cdots + e_H\hat{r}_H + F_H, \tag{31}$$

$$F_0 = E_0B_0 + F_H, \tag{32}$$

$$B_0 = \hat{r}_1\hat{u}_1^* + \hat{r}_2\hat{u}_2^* + \hat{r}_3\hat{u}_3^* + \ldots \hat{r}_H\hat{u}_H^*, \tag{33}$$

where B_0 is the perturbation sequence F_0 of the dependent variable y' of the improved spatial autoregressive model and the regression coefficient vector between the perturbation matrix E_0 of the independent variable. Determine the value of the regression coefficient according to the position of the respective variable in $Z(K)$. For example, the kth element in B_0 represents the estimated value of ρ_k.

3.2.2. Determination of the Optimal Number of Principal Components. The partial least squares method is a method of establishing a regression model by extracting the principal components of the independent variables. If the number of principal components is too small, the relationship between the variables cannot be reflected; otherwise, overfitting will occur. In order to ensure the validity of the model, the cross-validity criterion is selected to determine the number of principal components to be extracted.

The cross-validity criterion is mainly to test whether the extracted hth principal component can significantly improve the prediction performance of the model compared to the $(h-1)$ th principal component. For the improved spatial autoregressive model, the cross-validity criterion is selected to determine the number of principal components, which is mainly calculated by using all the spatial unit disturbance value sequences to extract the difference between the predicted value and the true value of the spatial unit disturbance when the number of principal components is $(h-1)$. The sum of squared deviations SS_{h-1} is shown in formula (34). And, extract the hth principal component based on the sequence after removing the perturbation value of its own spatial unit and calculate the sum of squared deviations of the predicted value and the true value based on the new sequence PRESS_h, as shown in

$$\mathrm{SS}_{h-1} = \sum_{i=1}^{n} \left(y_t'(i) - \hat{y}_t'(i)\right)^2, \tag{34}$$

$$\mathrm{PRESS}_h = \sum_{i=1}^{n} \left(y_t'(i) - \hat{y}_{t,-i}'(i)\right)^2. \tag{35}$$

According to SS_h and PRESS_h, calculate the effectiveness of extracting the hth principal component Q_h^2 (formula (36)). When $Q_h^2 \geq 1 - 0.95^2$, it indicates that the hth principal component has significantly improved the prediction performance of the model, and it is necessary to continue to extract the principal components; when $Q_h^2 < 1 - 0.95^2$, indicating the introduction after the hth principal component, the model has no obvious improvement, and there is no need to continue to extract principal components.

$$Q_h^2 = 1 - \frac{\mathrm{PRESS}_h}{\mathrm{SS}_{h-1}}. \tag{36}$$

3.3. Improving the Validation of the Spatial Autoregressive Model. For the speed disturbance sequence studied, the improved spatial autoregressive model is used to fit and the model's fitting effect is tested from two aspects: the model's goodness-of-fit (R^2) index and visual observation. When R^2 is closer to 1, it means that the fitting effect is good. On the contrary, when it is close to 0, it means that the fitting effect is poor. When the change trend of the measured value curve and the predicted curve is close, it indicates that the model can reasonably describe the space average speed at each position. On the contrary, it means that the model cannot reasonably reflect the real situation.

3.4. Analysis Method of Congestion Evolution in the Entrance and Exit Area. The improved spatial autoregressive model takes the velocity disturbance of each spatial unit as the dependent variable and independent variable and uses the partial least square method to estimate the parameters of the improved spatial autoregressive model and obtain the parameter estimated value ρ_k under the number of adjacent spaces in each space. The estimated value ρ_k represents the degree of correlation between the degree of velocity disturbance of the k-order adjacent unit of each spatial unit and the degree of correlation between the spatial unit.

Therefore, the parameters in the improved spatial autoregressive model are defined as the propagation structure of each spatial unit velocity disturbance in space. The influence of the disturbance of the space unit on its adjacent space unit of order k is denoted as $V(\overline{W}^k)$, as shown in

$$V\left(\overline{W}^k\right) = \rho_k\overline{W}^k \, (k = 1, 2, \ldots, k, \ldots, K). \tag{37}$$

The spatial propagation process of congestion can be summarized as a process in which the influence of local spatial units gradually spreads outwards through neighboring units. Congestion first spreads from spontaneously congested spatial units to neighboring spatial units and then from neighboring spatial units to second-order units. The continuous spread to the surroundings eventually leads to the proliferation and transformation of congestion. Therefore, the spatial propagation effects of local congestion can be calculated by formula (37), and the distribution of congestion spatial propagation effects under different spatial adjacency numbers can be investigated and then the spatial process of congestion propagation can be studied.

3.4.1. Propagation Effect of Congestion. Suppose that the space unit is congested and the speed of its first-order adjacent unit is disturbed through the adjacency relationship. The first-order propagation effect generated by the space

unit, that is, the influence on each adjacent unit can be represented by $V(\overline{W}^1)_{i,j}$ (j is the first-order adjacent space unit of i). Similarly, after congestion is propagated k times, the influence of spatial unit i on its k-order adjacent unit is expressed as $V(\overline{W}^k)_{i,j}$. Then, the ratio of the kth propagation effect to the total propagation effect is shown in

$$\mu_k = \frac{V\left(\overline{W}^k\right)_{i,j}}{\sum_{k=1}^{K} V\left(\overline{W}^k\right)_{i,j}}\#. \tag{38}$$

3.4.2. Interaction Degree Matrix between Spatial Units. The change of velocity perturbation of each space unit in the study area is the result of the combined effect of multiple space units. In order to quantify the degree of mutual influence between spatial units, the disturbance formed by the influence of spatial unit j by i is recorded as $v_{i,j}$ and then the degree of influence of spatial unit i on spatial unit j is calculated:

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,j} & \cdots & w_{1,n} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,j} & \cdots & w_{2,n} \\ w_{i,1} & w_{i,2} & \cdots & w_{i,j} & \cdots & w_{i,n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{n,1} & w_{n,2} & \cdots & w_{n,j} & \cdots & w_{n,n} \end{bmatrix}, \tag{39}$$

$$w_{i,j} = \frac{v_{i,j}}{\sum_{i=1}^{102} v_{i,j} \sum}, \tag{40}$$

$$v_{i,j} = \sum_{k=1}^{K} \rho_k \left(\overline{w}^k\right)_{i,j} y_i \quad . \tag{41}$$

3.4.3. Identification of Key Nodes. In the discipline of complex networks, nodes that are easily damaged and have a greater impact on other nodes are called key nodes. With reference to the definition of key nodes in a complex network, this paper proposes a method for identifying key nodes in the congestion evolution process of the entrance and exit areas: the space where the speed of the space unit is susceptible to influence and has a greater impact on the speed of adjacent space units. The unit is defined as the key node in the congestion evolution process of the entrance and exit area.

Based on the abovementioned definition of traffic congestion and related standards, "Urban Traffic Management Evaluation Index System" (2002, China), the spatial unit under the transmission effect is divided into 5 states: spontaneously congested units, heavily congested units, moderately congested units, lightly congested units, and immune units. The specific classification criteria are shown in Table 2.

The first of the above five traffic states is the congestion formed under the influence of its own factors, while the other four are under the comprehensive influence of the congestion propagation effect of neighboring units. When the disturbance rises sharply and reaches a certain threshold, the traffic state gradually changes from unblocked to congested, indicating that the propagation effect is the main reason for the decline in service level and even congestion in the entrance and exit areas.

Based on the method of judging the traffic state, this paper defines the space unit with a speed greater than 10 km/h and an impact on adjacent space units greater than 30% as a key node in the congested traffic state/process.

4. Results

4.1. Goodness-of-Fit Test. The goodness-of-fit test is conducted on the velocity disturbance sequence in the experimental area under the improved spatial autoregressive model. According to the parameter fitting results, the goodness-of-fit (R^2) value of the improved spatial autoregressive model is 0.7423, indicating that the model has a good fitting effect.

According to the test shown in Figure 3, divergence between the actual observed value and predicted value in each spatial unit is small and the actual observed value curve and the predicted value curve are very close, indicating the improved spatial autoregressive model can reasonably describe the trend of speed at various locations. In a nutshell, the model performs well in predicting the speed disturbance of each spatial unit and can truly and reasonably reflect the spatial correlation of the traffic of various units in the area.

4.2. Analysis on the Evolution Law of Congestion in the Entrance and Exit. Congestion evolution analysis, the method proposed in this paper, is used to study the evolution of congestion in the entrance and exit. Twenty sets of observed data whose traffic flow is 2000 pcu/h and entry rate is 20% are collected for research. By calculating the congestion propagation in the entrance and exit and the influence matrix of different spatial units, the key nodes of the evolution of congestion in the area can be perceived and its formation mechanism can be analyzed.

4.2.1. Analysis of Congestion Propagation in the Entrance and Exit. Based on the criteria in Table 2, the traffic of each spatial unit in the entrance and exit can be determined and the spatial distribution is shown in Figure 4. It can be seen from the distribution of various congestion units that the spread of local congestion is not isotropic; the frontage road to the entrance and exit of the parking lot is the culprit of local congestion, and the units where they are located are mostly the spontaneous congestion units; the congestion propagates to the inner lane and the downstream of the road segments from the spontaneous congestion units. The outer lane is affected most, and the affected area is also the largest; the middle lane and the inner lane are the least affected; and

Table 2: Definition of the traffic status of the space unit (unit: km/h).

Status	Disturbance range	Speed range	Description
Spontaneously congested unit			Congestion occurs first
Heavily congested unit	>20	<10	Vehicles are in a queue
Moderately congested unit	10~20	10~20	Low speed with a standstill
Lightly congested unit	0~10	20~30	Moderate speed with no standstill
Immune unit	≤0	>30	Faster speed in a free-flow state

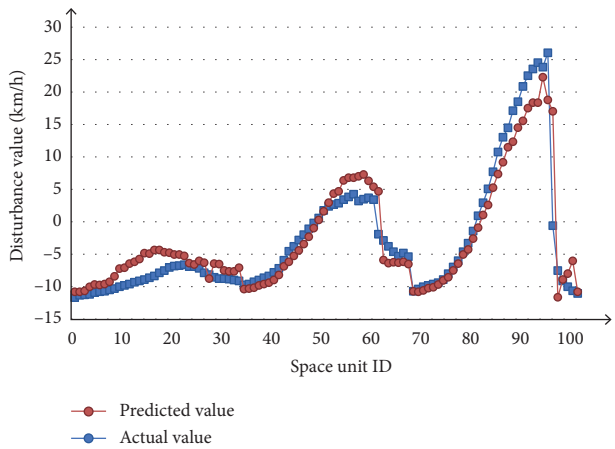


Figure 3: Comparison between the predicted value curve and actual observed value curve.

the downstream area of the entrance and exit and the upstream of the affected area are almost no longer affected.

In order to further explore the interaction between spatial units and understand the characteristics of congestion propagation, five spatial units with different traffic are selected and the distributions of congestion of spatial units of different orders are compared. The result is shown in Figure 5. It can be seen from the figure that the propagation whose spatial adjacency is within the fourth order accounts for more than 90% of the total propagation. Propagation gradually decreases as the spatial adjacency increases. This shows that when congestion happens in a spatial unit, its impact on the most adjacent spatial unit (first-order adjacent unit) is the greatest, and as the spatial scale increases, the impact decreases sharply.

4.2.2. Spatial Evolution Mechanism of Congestion in the Entrance and Exit. In order to further analyze the evolution mechanism of congestion in the entrance and exit, this paper calculates the influence matrix of spatial units based on the improved spatial autoregressive model and quantifies the degree of mutual influence between spatial units under the influence of congestion propagation. The spatial unit is regarded as the node, the mutual influence between the spatial units as the edge, and the degree of mutual influence as the weight, and a directed weighted network of the spatial units can be constructed, as shown in Figure 6.

It can be seen from Figure 6 that the directed weighted network of spatial units in the experimental area takes the shape of a "band," which is consistent with the above-mentioned attenuation characteristics of congestion propagation. To better analyze and study the evolution mechanism of congestion in the area, this paper divides the congestion evolution process into entrance and exit section area, transition area, and upstream area based on different spatial locations where the congestion happens. Figure 7 draws a directed weighted network for the three areas and uses the abstract network to analyze the evolution process of congestion.

(1) Section Area of the Entrance and Exit. The directed weighted network near the entrance and exit shows that affected by the parking behavior, there are significant differences in the weight values of the spatial units near the section. Based on the directed weighted network near the section (Figure 8), Figure 9 is drawn. All units in Figure 9 are spatial units in the section area of the entrance and exit; A, B, and C indicate Lane 1, Lane 2, and Lane 3, respectively, 1, 2, 3, 4, and 5 indicate adjacent spatial units from upstream to downstream of the road; C3, C4, and C5 are spontaneous congestion units, and other units are general units. It can be seen from the figure that the lateral interference on the spatial unit accounts for about 85% of the total in Lane 1 (outermost lane), and the lateral interference on the spatial unit accounts for 60% of the total in Lane 2 (middle lane). The main reason for the spontaneous congestion units in Lane 3 (innermost lane) is that the service capacity of the entrance and exit of the parking lot cannot meet the needs of parking and the entry time is prolonged due to the conflict between motor vehicles and nonmotorized vehicles. These ultimately make it impossible to ease the vehicles and cause congestion to propagate upstream.

(2) Transition Area. From the directed weighted network in the transition area (Figure 10) and the diagram of spatial units (Figure 11) (all the units in the figure are units in the transition area and all are general units, and the meanings of all the other symbols are the same as the preceding text), it can be seen that under the influence of the vehicle stagnation, the ratio of longitudinal interference to the total interference the spatial units receive in Lane 3 varies within the range of [0.5, 0.65], and the ratio of longitudinal interference to the total interference and the distance between the spatial units and the entrance and exit is inversely correlated. The ratio of the longitudinal interference to the total interference the spatial units receive on Lane 2 varies within the range of [0.35, 0.6], and the ratio of the lateral interference to the total interference the spatial units receive on Lane 2 varies within the range of [0.1, 0.45]. Moreover, the ratio of longitudinal interference to the total interference is positively correlated with the distance between the spatial units and the entrance and exit, and the ratio of lateral interference to the total interference is negatively correlated

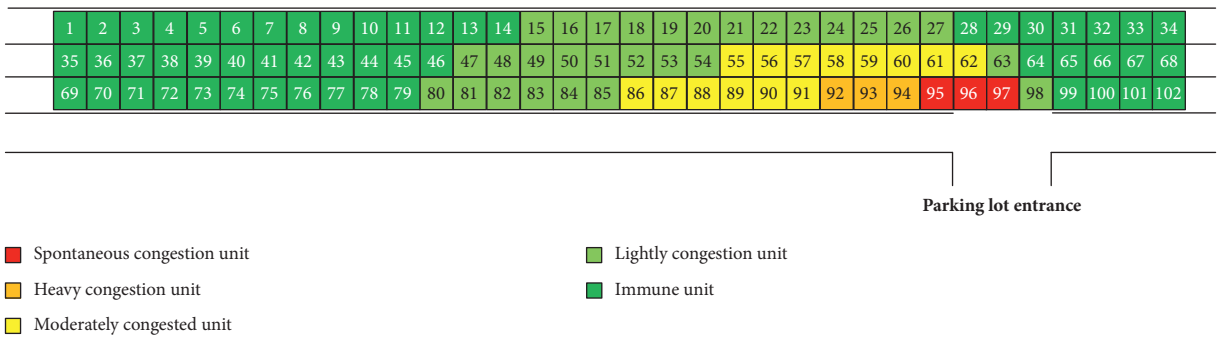


Figure 4: Distribution of the spatial unit.

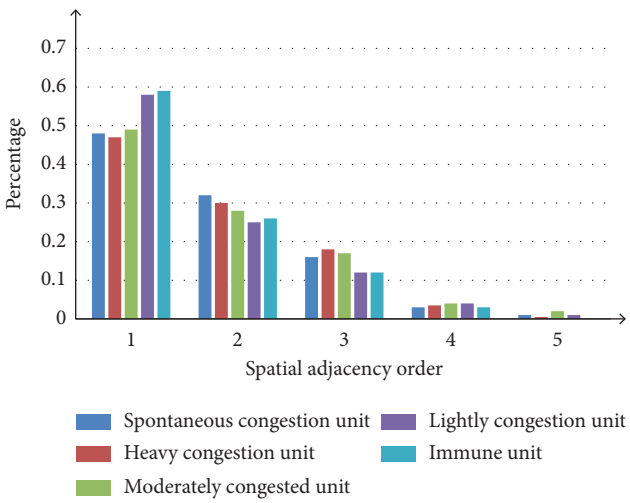


Figure 5: The attenuation process of spatial congestion propagation.

with the distance between the spatial units and the entrance and exit. This is because the spatial units in Lane 2 is affected by the combination of stagnation of vehicles in the same lane and lane changing of driving vehicles in Lane 3, and the speed is disturbed. As the distance increases, the impact of lane changing on Lane 2 gradually decreases, and the impact of stagnation on the upstream units gradually becomes the main factor, which is manifested by the gradual increase in the weight value of the units in Lane 2. The longitudinal interference the spatial units receive on Lane 1 accounts for about 65% of the total, which is mainly caused by the vehicles changing lanes.

(3) Upstream Area. The directed weighted network of the spatial units in the upstream area of the entrance and exit is shown in Figure 12. The weight value between the spatial units in the upstream area tends to be stable, which means that the congestion is propagated upstream in the same pattern. The relationship between spatial units is illustrated by Figure 13 (all the units in the figure are units in the upstream area and all are general units, and the meanings of all the other symbols are the same as the preceding text). It can be seen from the figure that the lateral interference the spatial units receive in Lane 1 accounts for more than 90% of the total. The main reason for this phenomenon is that the vehicles in Lane 1 rarely stagnate, and the spatial units are mainly affected by "friction effect"; in Lane 2, the lateral interference the spatial units receive in Lane 2 is about 60%. This is because the intermittent queuing of vehicles parked near the entrance and exit causes the vehicles in the upstream area to stagnate. Therefore, the mutual influence between the spatial units in the same lane is relatively high. Lane 3 is the lane directly connected to the entrance and exit. Due to the combination of frequent queuing, following the traffic flow and lane changing of parked vehicles, the longitudinal and lateral interferences the spatial units receive in Lane 3 are both about 50%.

To sum up, from the perspective of the impact value, the congestion propagation is the strongest in the frontage road to entrance and exit of the parking lot, where the congestion first happens; there is moderate propagation in the transition area, the main area where congested vehicles stagnate; congestion propagation is weak in the downstream area, stagnation of vehicles is barely seen, and there is a relatively fast-flowing traffic. From the perspective of the proportion, from the frontage road to the transition area and then to the upstream area, the proportion of longitudinal interference gradually increases, while the proportion of lateral interference gradually decreases. This shows that the lateral propagation of congestion attenuates faster than longitudinal propagation.

4.2.3. Key Congestion Nodes. In order to further analyze the relationship between the location and number of key nodes in the process of congestion evolution and the traffic volume of the road segments and the entry rate, this paper calculates the location of key nodes of congestion propagation when the traffic volume is 1500 pcu/h, 2000 pcu/h, and 2500 pcu/h, respectively, and when the entry rate of parked vehicles stands at 15%, 20%, and 25%, respectively (as shown in Table 3). Obviously, the traffic volume and the entry rate are positively correlated with the number of key nodes. As the traffic volume and the entry rate increase, key nodes spread from the outer lane to the inner lane and from the frontage road to the entrance and exit of the parking lot to the upstream area. Further comparative analysis shows that when the traffic volume is 1500 pcu/h, the change of the entry rate does not have a significant impact on the number of key nodes; when the traffic volume increases to 2000 pcu/h (especially when it increases to 2500 pcu/h), the number of

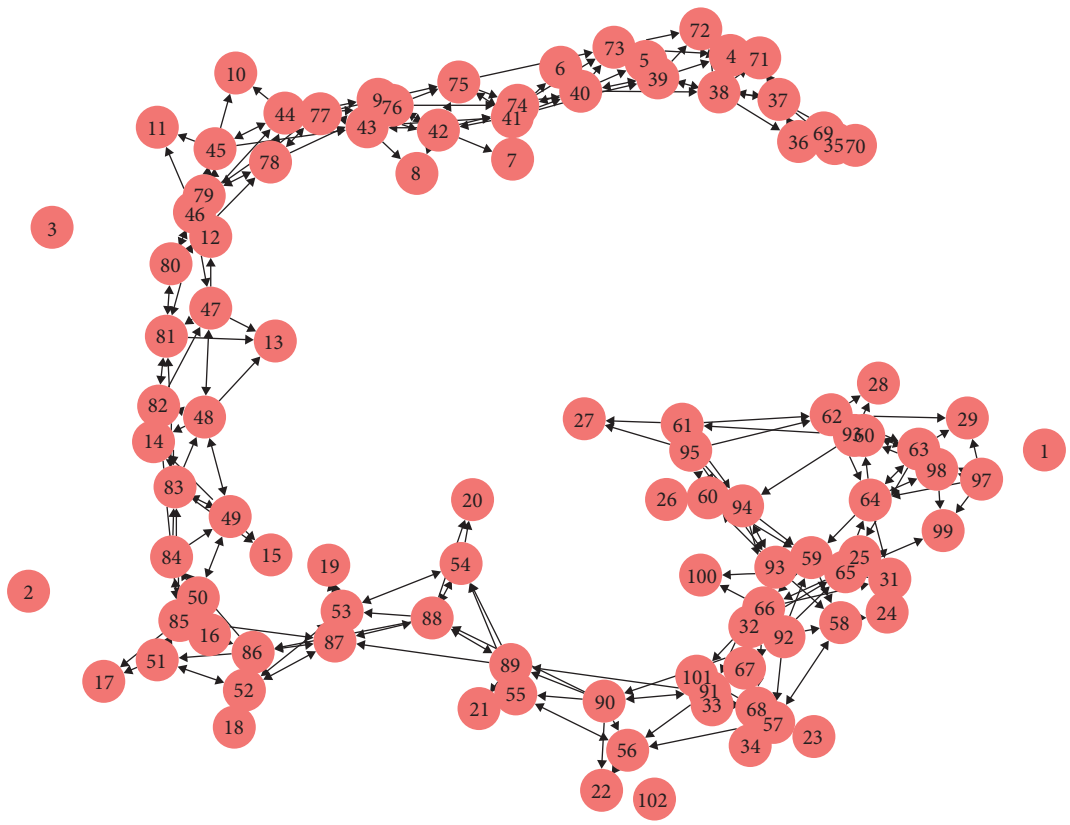


Figure 6: Directed weighted network of the entrance and exit.

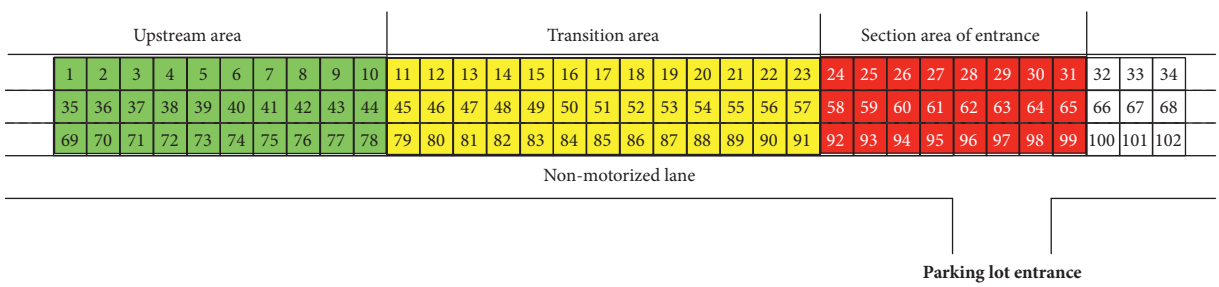


Figure 7: Diagram of the division of the entrance and exit.

key nodes increases significantly along with the increase in the entry rate, which shows that the impact of the entry rate on the key nodes is enhanced as the traffic volume of the road segments increases.

The analysis suggests that the degree, distribution, and number of congestion of spatial units in the entrance and exit are constantly changing. When the traffic volume of the road segments and entry rate are high, the degree and number of congestion of the spatial units of the entrance and exit will increase significantly, and vice versa. However, for each specific unit, whether it will become a congestion unit is not determined solely by the traffic volume or the entry rate, but by the combination of the traffic volume, the entry rate, and the location of the spatial unit. In the directed weighted network in the entrance and exit, some spatial units (key nodes under the effect of congestion propagation) quickly evolve into congestion units because of greater turbulence caused by the increase in traffic volume or entry rate. However, some units are less or not susceptible to the increase in traffic volume or entry rate. Therefore, in daily traffic management, the key nodes in the evolution process should be eased; for spontaneous congestion units, by improving the service capacity of entrances and exits and reducing the entry time of parked vehicles, the impact of spontaneous congestion units on adjacent units can be reduced; in terms of the key nodes under the influence of propagation, reducing stagnation of vehicles can alleviate the longitudinal propagation of congestion; in addition, guiding vehicles to enter the parking lot from other entrances with facilities such as parking grading guidance screens to reduce entry rate improve the frontage roads to the entrances and exits and enhance the quality of travel for citizens.

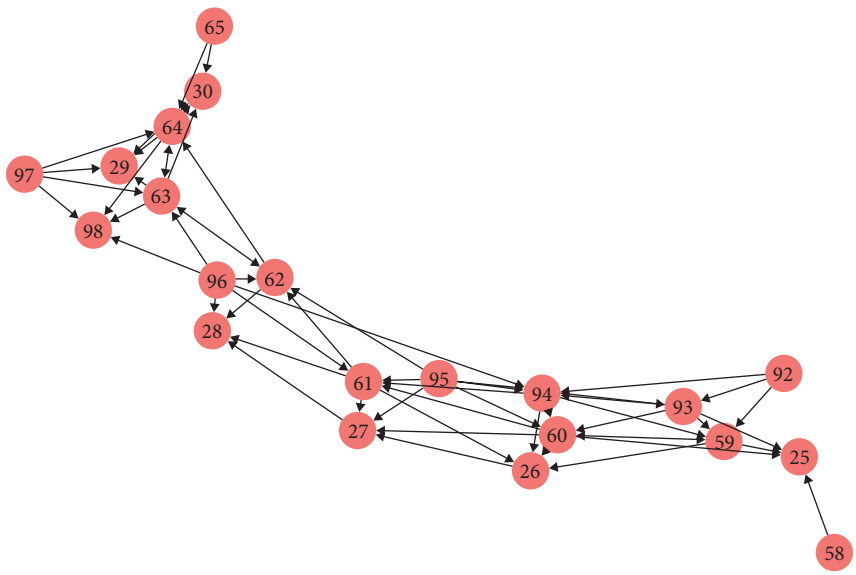


Figure 8: Diagram of a directed weighted network in the section area of the entrance and exit.

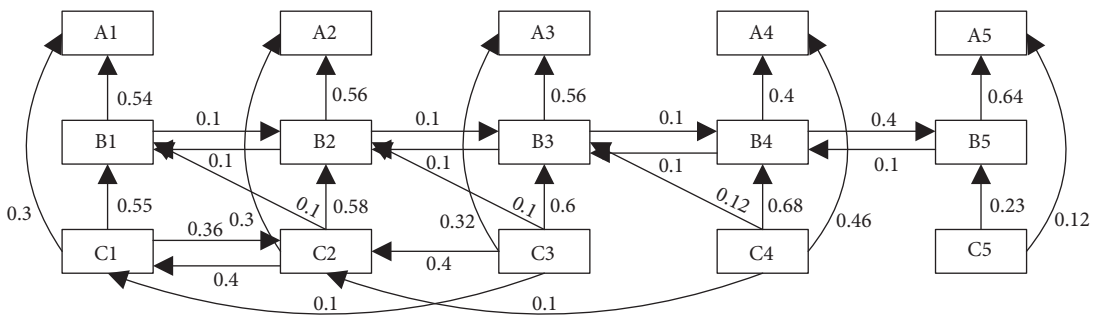


Figure 9: The diagram of spatial units in the section area of the entrance and exit.

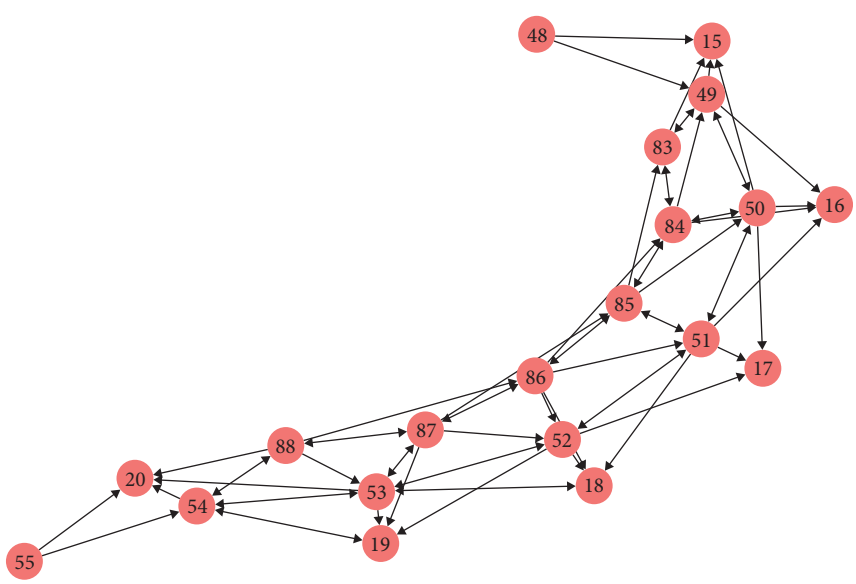


Figure 10: Diagram of a directed weighted network in the transition area.

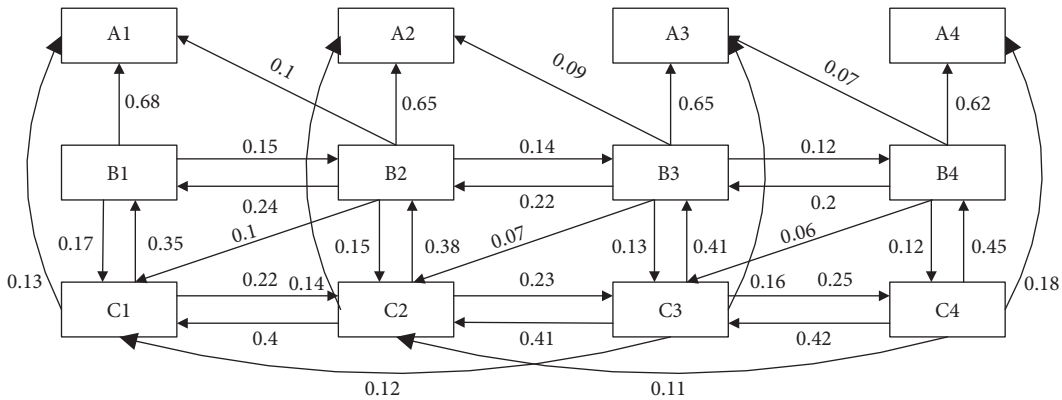


Figure 11: The diagram of spatial units in the transition area.

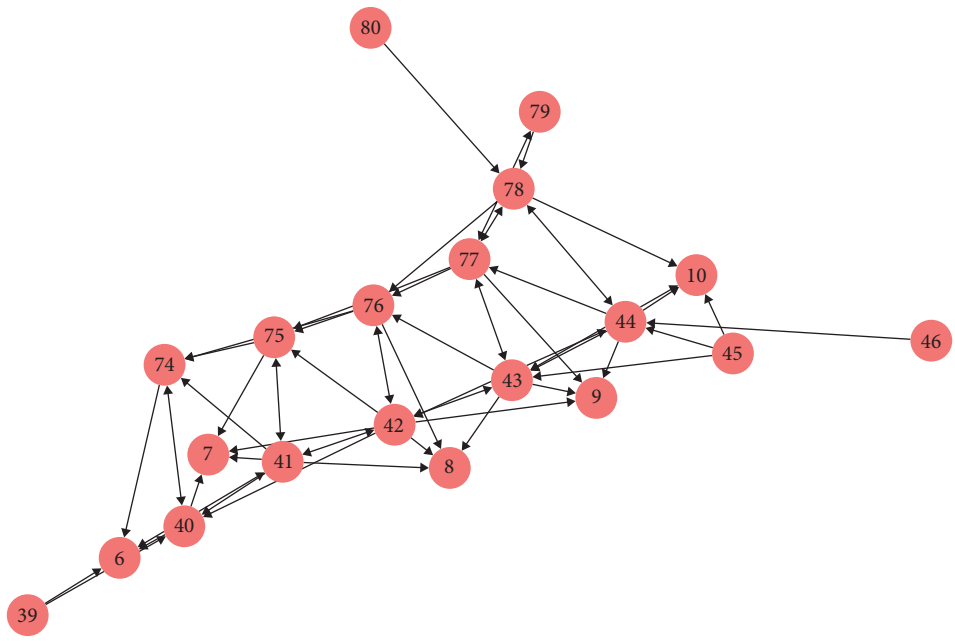


Figure 12: Diagram of a directed weighted network in the upstream area.

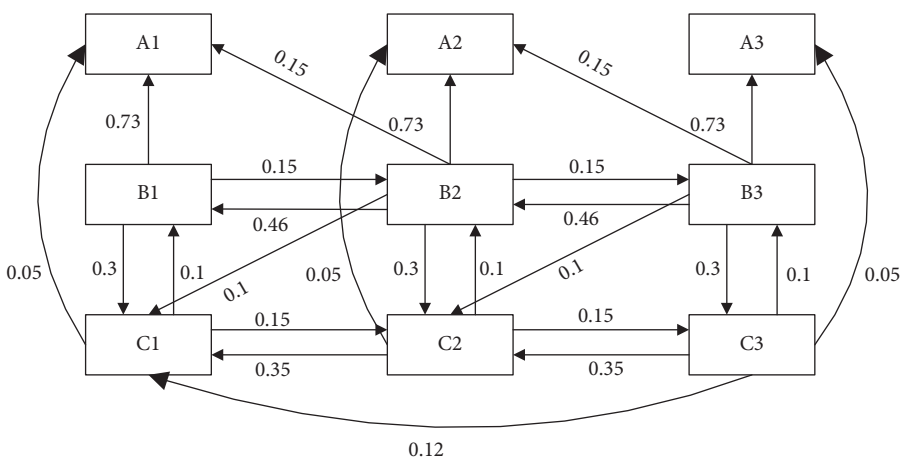


Figure 13: The diagram of spatial units in the upstream area.

Table 3: Key nodes of congestion propagation.

Traffic flow (±100 pcu/h)	Entry rate (±5 (%))	Key nodes
1500	15	95, 96, 97
	20	95, 96, 97
	25	93, 94, 95, 96, 97, 61, 62
2000	15	92, 93, 94, 95, 96, 97, 58, 59, 60, 61, 62
	20	89, 90, 91, 92, 93, 94, 95, 96, 97, 57, 58, 59, 60, 61, 62
	25	86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 54, 55, 56, 57, 58, 59, 60, 61, 62
2500	15	88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 54, 55, 56, 57, 59, 60, 61, 62
	20	83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62
	25	76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57

5. Conclusions and Discussion

This paper starts with studying the spatial evolution pattern of congestion propagation in the entrance and exit of a parking lot. First, it discusses the shortcomings of traditional spatial autoregressive model and proposes an improved robust spatial autoregressive model based on the actual characteristics of the traffic flow in the entrance and exit. Second, it constructs a smooth spatial adjacency matrix based on the principal component analysis and designs a parameter estimation method based on partial least squares. Third, it selects the frontage road to entrance and exit of a typical parking lot for research, and the fitting result (R^2 is 0.7423) shows that the model has a good fitting performance for the velocity disturbance sequence of the spatial units. Fourth, by identifying and observing the distribution of different types of congestion units in the research section, the mechanism of congestion formation and the main propagation direction are analyzed. Finally, it explores and analyzes the congestion propagation and the spatial evolution mechanism of the congestion in the experimental area. The results show that as the spatial scale increases, the congestion propagation decreases sharply, and spatial adjacency within the fourth order can account for more than 90% of the propagation; the frontage road to the entrance and exit is the place where the congestion first happens; the congestion gradually attenuates as it propagates to the inner lane and the upstream of the road segments; the lateral congestion propagation attenuates faster, so the area affected by congestion is mainly distributed in the outermost lane. In summary, the proposed method for analyzing the spatial propagation characteristics of traffic congestion can be better applied to the frontage roads of parking facilities. The results of the case analysis show that for the local traffic congestion at the entrance and exit, the relevant departments should proceed from a holistic perspective to accurately monitor and orderly channel the key space nodes. The conclusions are of great theoretical and practical significance for improving the frontage roads to the entrances and exits of parking facilities and the quality of travel for citizens.

Data Availability

The measured data used to support the findings of this study are included within the article and the supplementary files.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by Research Planning Fund for Humanities and Social Sciences of the Ministry of Education (19YJAZH011), Support for the Open Project of Key Laboratory of Intelligent Traffication Technology and Traffication Industry (F262019016), Science and Technology Project of Traffication Department of Jiangsu Provincial Department of Communications (KY2018049), Jiangsu Provincial Department of Science and Technology Industry-University-Research Cooperation Project "Traffic Safety-Oriented Intelligent Supervision and Decision Support System Platform Research and Development" (BY2019263), and the Key Project of Transportation Science and Technology Achievement & Transformation of Jiangsu Province (2019Z01).

Supplementary Materials

The following is a concise description to the supplementary file. The data include our observation records of the frontage road section of the parking lot of the West Affiliated Hospital of Yangzhou University, when the entry rate is about 20% and the traffic flow is around 2000 pcu/h. The columns of data represent the labels of 13 groups of observations; the indexes represent the IDs of the spatial units, and values represent the average values of the instantaneous speed of 100 observations, recorded every 3 s for 5 min, for each group of each unit. (*Supplementary Materials*)

References

[1] China Statistical Yearbook, 2021, https://www.stats.gov.cn/tjsj/ndsj/.

[2] J. Chen and G. Yang, "Off-street parking," in *International Encyclopedia of Transportation*, R. Vickerman and R. Vickerman, Eds., Elsevier, Oxford, UK, 2021.

[3] T. Shen, Y. Hong, M. M. Thompson, J. Liu, X. Huo, and L. Wu, "How does parking availability interplay with the land use and affect traffic congestion in urban areas? The case study of Xi'an, China," *Sustainable Cities and Society*, vol. 57, Article ID 102126, 2020.

[4] P. Zhao, H. Guan, H. Wei, and S. Liu, "Mathematical modeling and heuristic approaches to optimize shared parking resources: a case study of Beijing, China," *Transportation Research Interdisciplinary Perspectives*, vol. 9, Article ID 100317, 2021.

[5] X. Cai, L. Lu, J. J. Lu, and W. Lin, "Impacts of access density on traffic capacity of arterial roads," in *Proceedings of the 4th International Conference on Transportation Engineering*, pp. 359–364, Chengdu, China, October 2013.

[6] J. Zhao, P. Li, and X. Zhou, "Capacity estimation model for signalized intersections under the impact of access point," *PLoS One*, vol. 11, no. 1, Article ID e0145989, 2016.

[7] C. Jiang and L. Lu, "Influences of access density of minor roads on arterial speed variance by TSIS," in *Proceedings of the 4th International Conference on Transportation Engineering*, Chengdu, China, October 2013.

[8] J. A. Bonneson, "Delay to major-street through vehicles due to right-turn activity," *Transportation Research Part A: Policy and Practice*, vol. 32, no. 2, pp. 139–148, 1998.

[9] X. Hu, T. Liu, X. Hao, Z. Su, and Z. Yang, "Research on the influence of bus bay on traffic flow in adjacent lane: simulations in the framework of Kerner's three-phase traffic theory," *Physica A: Statistical Mechanics and its Applications*, vol. 563, Article ID 125495, 2021.

[10] M. Johari, M. Keyvan-Ekbatani, and D. Ngoduy, "Impacts of bus stop location and berth number on urban network traffic performance," *IET Intelligent Transport Systems*, vol. 14, no. 12, pp. 1546–1554, 2020.

[11] S. Liang, M. Ma, S. He, H. Zhang, and Z. Tang, "Influence of bus stop location on traffic flow," *Proceedings of the Institution of Civil Engineers-Municipal Engineer*, vol. 174, no. 1, pp. 24–31, 2021.

[12] F. Han, Y. Han, M. Ma, and D. Zhao, "Research on traffic wave characteristics of bus in and out of stop on urban expressway," *Procedia Engineering*, vol. 137, pp. 309–314, 2016.

[13] P. Bansal, R. Agrawal, and G. Tiwari, "Impacts of bus-stops on the speed of motorized vehicles under heterogeneous traffic conditions: a case-study of Delhi, India," *International Journal of Transportation Science and Technology*, vol. 3, no. 2, pp. 167–178, 2014.

[14] P. Raj, G. Asaithambi, and A. U. Ravi Shankar, "Effect of curbside bus stops on passenger car units and capacity in disordered traffic using simulation model," *Transportation Letters*, vol. 2, 2020.

[15] Y. Cao, Z. Z. Yang, and Z. Y. Zuo, "The effect of curb parking on road capacity and traffic safety," *European Transport Research Review*, vol. 9, no. 1, 2017.

[16] Z. Mei and J. Chen, "Modified motor vehicles travel speed models on the basis of curb parking setting under mixed traffic flow," *Mathematical Problems in Engineering*, vol. 2012, Article ID 351901, 14 pages, 2012.

[17] S. Salini and R. Ashalatha, "Analysis of traffic characteristics of urban roads under the influence of roadside frictions," *Case Studies on Transport Policy*, vol. 8, no. 1, pp. 94–100, 2020.

[18] M. Patkar and A. Dhamaniya, "Effect of on-street parking on effective carriageway width and capacity of urban arterial roads in India," *European Transport-Trasporti Europei*, no. 73, 2019.

[19] K. Seggerman and K. Williams, "Managing the indirect impacts of bypasses on small and medium-sized communities in Florida," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2453, no. 1, pp. 46–53, 2014.

[20] J. Kim and H. S. Mahmassani, "Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories," *Transportation Research Part C: Emerging Technologies*, vol. 59, pp. 375–390, 2015.

[21] B. Bae, Y. Liu, L. D. Han, and H. Bozdogan, "Spatio-temporal traffic queue detection for uninterrupted flows," *Transportation Research Part B: Methodological*, vol. 129, pp. 20–34, 2019.

[22] D. Ni and J. D. Leonard, "A simplified kinematic wave model at a merge bottleneck," *Applied Mathematical Modelling*, vol. 29, no. 11, pp. 1054–1072, 2005.

[23] G. F. Newell, "A moving bottleneck," *Transportation Research Part B: Methodological*, vol. 32, no. 8, pp. 531–537, 1998.

[24] G. Gentile, "Using the general link transmission model in a dynamic traffic assignment to simulate congestion on urban networks," *Transportation Research Procedia*, vol. 5, pp. 66–81, 2015.

[25] S. Kurata and T. Nagatani, "Spatio-temporal dynamics of jams in two-lane traffic flow with a blockage," *Physica A*, vol. 318, no. 3, pp. 537–550, 2003.

[26] C. Wright and P. Roberg, "The conceptual structure of traffic jams," *Transport Policy*, vol. 5, no. 1, pp. 23–35, 1998.

[27] M. Saeedmanesh and N. Geroliminis, "Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks," *Transportation Research Part B: Methodological*, vol. 105, pp. 193–211, 2017.

[28] 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," *Computers, Environment and Urban Systems*, vol. 77, Article ID 101364, 2019.

[29] B. Alkouz and Z. Al Aghbari, "SNSJam: road traffic analysis and prediction by fusing data from multiple social networks," *Information Processing & Management*, vol. 57, no. 1, Article ID 102139, 2020.

[30] Y. Song, X. Liang, Y. Zhu, and L. Lin, "Robust variable selection with exponential squared loss for the spatial autoregressive model," *Computational Statistics & Data Analysis*, vol. 155, Article ID 107094, 2021.

[31] S. P. Zhou and J. J. Zhang, "Quasi-maximum likelihood estimations and applications for spatial dynamic autoregressive panel model with fixed effects," *Systems Engineering-Theroy & Practice*, vol. 41, no. 1, pp. 45–57, 2021.

[32] F. Han and F. Sui, "Effect of global moran's I and space-time permutation scanning method in shanghai metro traffic based on ecological transportation system," *Ekoloji*, vol. 28, pp. 4741–4749, 2019.

[33] H. Rong, A. P. Teixeira, and C. Guedes Soares, "Spatial correlation analysis of near ship collision hotspots with local maritime traffic characteristics," *Reliability Engineering & System Safety*, vol. 209, Article ID 107463, 2021.

[34] A. Soltani and S. Askari, "Exploring spatial autocorrelation of traffic crashes based on severity," *Injury*, vol. 48, no. 3, pp. 637–647, 2017.

[35] T. Ban, T. Usui, and T. Yamamoto, "Spatial autoregressive model for estimation of visitors' dynamic agglomeration patterns near event location," *Sensors*, vol. 21, no. 13, p. 4577, 2021.

[36] X. Zhang, G. Chen, J. Wang, M. Li, and L. Cheng, "A GIS-based spatial-temporal autoregressive model for forecasting marine traffic volume of a shipping network," *Scientific Programming*, vol. 2019, Article ID 2345450, 14 pages, 2019.

[37] Y. Wang, Y. Yuan, X. Guan et al., "Collaborative two-echelon multicenter vehicle routing optimization based on state-

space-time network representation," *Journal of Cleaner Production*, vol. 258, Article ID 120590, 2020.

[38] Y. Wang, S. Peng, and M. Xu, "Emergency logistics network design based on space-time resource configuration," *Knowledge-Based Systems*, vol. 223, Article ID 107041, 2021.

[39] Y. Wang, X. Ma, Z. Li, Y. Liu, M. Xu, and Y. Wang, "Profit distribution in collaborative multiple centers vehicle routing problem," *Journal of Cleaner Production*, vol. 144, pp. 203–219, 2017.

[40] Y. Wang, S. Peng, X. Zhou, M. Mahmoudi, and L. Zhen, "Green logistics location-routing problem with eco-packages," *Transportation Research Part E: Logistics and Transportation Review*, vol. 143, Article ID 102118, 2020.

[41] N. G. Álvarez, B. Adenso-Díaz, and L. Calzada-Infante, "Maritime traffic as a complex network: a systematic review," *Networks and Spatial Economics*, pp. 1–31, 2021.

[42] L. Siozos-Rousoulis, D. Robert, and W. Verbeke, "A study of the U.S. domestic air transportation network: temporal evolution of network topology and robustness from 2001 to 2016," *Journal of Transportation Security*, 2021.

[43] Y. Sui, F. Shao, X. Yu, R. Sun, and S. Li, "Public transport network model based on layer operations," *Physica A: Statistical Mechanics and its Applications*, vol. 523, pp. 984–995, 2019.

[44] H.-Z. Lin and J. Wei, "Optimal transport network design for both traffic safety and risk equity considerations," *Journal of Cleaner Production*, vol. 218, pp. 738–745, 2019.

[45] Y. Sui, F.-j. Shao, R.-c. Sun, and S.-j. Li, "Space evolution model and empirical analysis of an urban public transport network," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 14, pp. 3708–3717, 2012.

[46] R. Khalid, M. K. M. Nawawi, L. A. Kawsar, N. A. Ghani, A. A. Kamil, and A. Mustafa, "Optimal routing of pedestrian flow in a complex topological network with multiple entrances and exits," *International Journal of Systems Science*, vol. 51, no. 8, pp. 1325–1352, 2020.

[47] YOLOV5-deepsort-pytorch, 2021, https://github.com/topics/yolov5-deepsort-pytorch.

[48] X. Chen, L. Qi, Y. Yang et al., "Video-based detection infrastructure enhancement for automated ship recognition and behavior analysis," *Journal of Advanced Transportation*, vol. 2020, Article ID 7194342, 12 pages, 2020.

[49] X. Chen, Z. Li, Y. Yang, L. Qi, and R. Ke, "High-resolution vehicle trajectory extraction and denoising from aerial videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 3190–3202, 2021.

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 9950417, 13 pages
https://doi.org/10.1155/2021/9950417

Research Article

Optimization Model and Method of Variable Speed Limit for Urban Expressway

Shubin Li,[1] **Tao Wang**,[2,] **Hualing Ren**,[3] **Baiying Shi**,[4] **and Xiangke Kong**[4]

[1]*Department of Traffic Management Engineering, Shandong Police College, Jinan 250014, China*
[2]*Department of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China*
[3]*MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China*
[4]*Department of Traffic Engineering, Shandong Jianzhu University, Jinan 250101, China*

Correspondence should be addressed to Shubin Li; li_shu_bin@163.com, Tao Wang; wangtao@qust.edu.cn, and Baiying Shi; shibaiying@sdjzu.edu.cn

Received 26 March 2021; Revised 20 April 2021; Accepted 5 May 2021; Published 19 May 2021

Academic Editor: Yong Wang

The urban expressway network is the main part of the urban traffic network carrying most of the city's traffic pressure for its continuity and rapidity, but the control method of the traffic flow was too simple to other control methods in application in addition to the ramp control and the fixed speed control. In this paper, the theory of variable speed limit (VSL) was used to develop an optimal control model based on the improved traffic flow simulation model according to the characteristics of urban expressway traffic flow. The objective of the proposed model is to minimize the delay and maximize the traffic flow. It can adjust the traffic flow on the network in space time so that the whole network is in a state of equilibrium which not only is conducive to the control of the local traffic congestion and avoids the spread of congestion but also improves the traffic safety. The SPSA-based solution algorithm was proposed by taking into account the needs of real-time online applications. It can not only ensure the accuracy of the solution but also meet the requirements of the simulation time. The simulation results show that the variable speed limit can be optimized in moderate demand, and the proposed model and algorithm are effective and feasible in this paper. The conclusions are useful to help the traffic management department to formulate reasonable traffic control strategies.

1. Introduction

Thanks to the rapid development of economy and mechanization, traffic congestion has become one of the most prominent city problems in China. It gradually spreads from big cities to medium-sized and small ones. Even metropolises like Beijing, Shanghai, Guangzhou, and Shenzhen have not conquered the traffic congestion and therefore they are nicknamed "congestion city." The Annual Report of Beijing Traffic Development of 2015 shows that the traffic congestion becomes worse per hour, and the traffic flow of the central part of Beijing increased by 2.7% compared to that of the previous year. The time quantum of congestion is one hour and fifty-five minutes and the distance of severe traffic congestion is 10.4%. Apparently, this worldwide problem has caused Chinese top leaders concern. President Xi Jinping listened to relevant reports while he was inspecting the traffic in Beijing. He demanded that the transport sector analyzes the traffic problems and performs effective measures, which promotes the efforts and research to relieve the noncapital function of Beijing and the scheme of coordinated development of the Beijing-Tianjin-Hebei region. Developing the intelligent transportation system is one of the strategies to approach traffic congestion, namely, balancing the conflict between the traffic demand and the traffic supply by virtue of modernized science and technology, together with macroregulation of government. In this way, it is likely to achieve smooth transportation. Taking the traffic of Beijing as a case in point, each car serves 1.99 people daily before the appearance of an intelligent travel

platform which can serve 2.24 people each day. The availability of cars increases by 12.1%. In October 2014, Chinese Premier Li Keqiang pointed out that the so-called intelligent sustainable transportation should be given a complete definition. It refers to addressing the traffic problems through the combination of traffic factors (people, car, and road) and the achievement of intelligent traffic through education, law enforcement, and technology. Recently, Huabei Area Intelligent Travel Big Data Report released by Uber indicates that the first five months of 2014 have witnessed 3.5 billion people in Huabei province who used the intelligent travel application when going out. Beijing, Taiyuan, and Tianjin take the lead in the employment of this app among cities in Huabei area. The focal point in the study of intelligent traffic technology lies in the evaluation and prediction of the current traffic network and the implementation of online control.

As the main part of the urban traffic network, the urban expressway bears the heaviest urban traffic pressure and is characterized by high speed, large flow, and continuous traffic flow. However, more effective methods are in urgent demand for the management of urban expressway because it is not compatible with those for the management of highway and for the tackling of conventional traffic problems. Hence, it is a significant try to introduce the intelligent control of variable speed limit (VSL) to the management of urban expressway.

The management of variable speed limit is subject to current traffic conditions, which requires sense and judges the condition of the traffic network by checking the current and previous data. The major concern about the evaluation of the traffic network lies in the use of the traffic flow simulation model which can be roughly divided into macrotraffic flow simulation model [1], mesotraffic flow simulation model [2, 3], microtraffic flow simulation model [4], and all kinds of the optimized simulation model [5–8]. Among them, the mesotraffic flow model can simulate different levels of traffic details with its unique modeling ideas and simulation methods. Considering the computational accuracy and speed, the mesotraffic flow model is of great significance in the study of the intelligent transportation system (ITS) and hybrid simulation model. Felipe proposed an agent-based mesoscopic traffic flow simulation model, which could satisfy both the macrosimulation efficiency and the microsimulation accuracy [9]. Anahita proposed a midrange flow-emission model considering the aggregation behavior of different vehicle groups, which verified that the model was suitable for real-time simulation [10].

Making positive traffic control is based on the online simulation of the real-time traffic condition through simulated dynamic characteristics. Currently, the function of maintaining traffic safety is analyzed in most studies of variable speed limit [11–14]. As a part of positive traffic management, the variable speed limit can optimize and control the speed of the traffic flow in the artery movement, thus maintaining traffic safety and order. However, little emphasis is put on the study of alleviating traffic congestion. Zhang established a cellular automaton model in highway variable speed limit zone to study the effect of speed limit value and speed limit zone on traffic congestion [15]. Wang W regarded the variable speed limit control of expressway as the discrete-time Markov decision process, proposed the variable speed limit control model based on reinforcement learning and Markov decision, and used the Paramics simulation software to verify that the model can improve the traffic flow [16]. Zhou proposes a variable speed limit control strategy based on reinforcement learning and verifies that the strategy has obvious advantages in reducing congestion and travel time [17].

In order to relieve the traffic pressure and realize the change from passive reactive traffic management to active precautionary traffic management, this paper focuses on how to achieve traffic safety and smooth traffic flow. The main innovations of this paper include the following: (1) In view of the traffic jam of Jinan viaduct expressway, an improved medium traffic flow model more suitable for local traffic conditions is adopted to simulate the movement of vehicles. (2) In order to achieve the effect of system optimization, an optimization model of variable speed limit is proposed to adjust the speed, so as to achieve the balanced distribution of traffic density and promote the improvement of traffic efficiency. (3) *e* solution algorithm based on SPSA is designed to help the traffic management department formulate reasonable traffic control strategies.

2. The Improved Mesoscopic Traffic Flow Simulation Model

The mesoscopic traffic flow simulation model quickly arouses scholars' interest in traffic flow field because it can not only ensure the accuracy of simulation but also satisfy the demand for calculating speed of real-time online simulation. The anisotropic mesoscopic traffic flow simulation model in literature [3] is adopted in this paper and the model is improved according to speed limit control strategy. The construction of the mesoscopic traffic emulator usually includes four categories of models: the road network model, the static queuing model, the speed model, and the vehicle movement model.

2.1. The Road Network Model. The analysis method based on the network model is widely used in the fields of social, transportation, and logistics, such as logistics distribution network [18] and state space-time network [19, 20]. By constructing the network topology of the system, the evolution rule and essence of the system can often be reflected. The urban expressway network is composed of ramp and artery. Owing to the similarity with the freeway, the node, the connecting link, the lane group, the lane, detector, the turning prohibition line, and the traffic zone can be used to describe the topological structure of the urban expressway. As Figure 1, a road consists of the following elements: node, connecting line, lane, lane group, and segment. And traffic network consists of node and link.

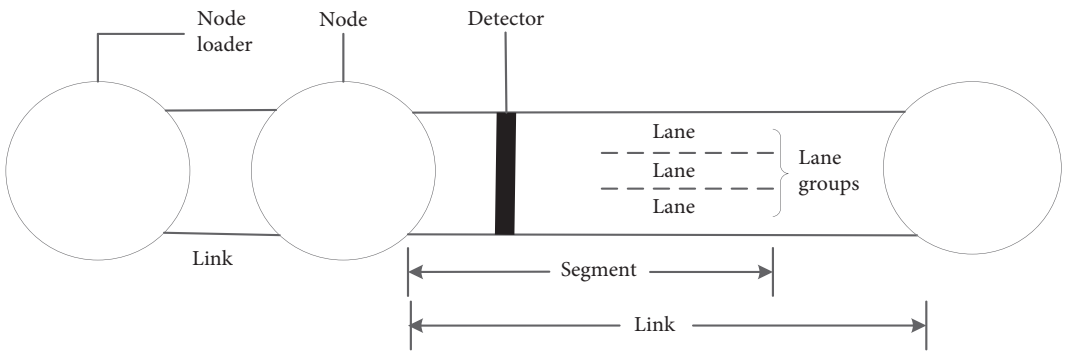


Figure 1: Road network topology.

The mentioned factors have a relationship as in Figure 2. VMS is the variable message signs, while VSL is the variable speed limit system.

2.2. The Static Queuing Model. When the expressway exit ramp or bottleneck section is congested, the application of this model to the upstream section of the congested section can effectively alleviate the traffic congestion. The vehicle behavior is similar to the queue in signal control intersection. It is necessary to calculate the process when a vehicle joins a queue, the movement in the queue, and the process when the vehicle leaves the queue. Suppose a vehicle i is in a queue; then the queue will definitely incur delay which can be signified by the following formula:

$$T = \frac{i}{c}. \tag{1}$$

According to what is mentioned in Figure 2, to define sections of a road with identical physical property as a segment, the lane group can be defined on the basis of segment. Generally, a lane group consists of one or more lanes. Then, c represents the output capacity. On this segment, the vehicles are divided into moving ones and queuing ones and the above model is used to describe queuing behavior. t refers to time, and then ct refers to the number of vehicles that leave the lane group. Suppose a vehicle reaches the end of a queue at t; then its position can be expressed by the following formula:

$$q(t) = q(0) + l_{\text{car}}(ct - m), \tag{2}$$

where $q(0)$ is the end position of queue at $t = (0)$. $l_{\text{car}} = (1/k_{\text{jam}})$ represents the standard vehicle length including headway. k_{jam} refers to the jam density. m refers to the total number of vehicles between the one at the end of the queue and the considered one. That is to say, it refers to the number of vehicles that join the queue before the considered one.

It is worthy of note that L refers to the length of segment and the model is worked with $0 < q(t) < L$. When $q(t) \geq L$, the queue dissipated. This means that the considered vehicle failed to catch up with the queue and it is still in movement. Because there cannot be a negative number, $q(t) < 0$ never occurs.

2.3. The Speed Model. Vehicles move at a certain speed on segment. The speed model can describe a vehicle with a certain speed in the network. According to the definition of the anisotropy mesoscopic traffic flow, the speed of the vehicle is determined by the traffic density of some areas downstream. In other words, the vehicles in movement are influenced by the ones in front of them. This influence fades with the increase of distance and has nothing to do with the vehicles left behind. Therefore, the anisotropy in mesoscopic traffic flow can be achieved through the definition of the length of the downstream region, which is called Speed Influencing Region (SIR). It is shown in Figure 3.

The traffic density of each vehicle in movement can be calculated according to the amount of vehicle in SIR. The first case is shown in Figure 3(a). The density of vehicle i in SIR is as follows:

$$k_i^{t-1} = \frac{N_i^{t-1}}{nl}, \tag{3}$$

where N refers to the amount of vehicle i in SIR when the time is $t - 1$, n refers to the amount of lane in SIR, and l refers to the length of SIR. From 3(b) and 3(c), the formula of the region density is as follows:

$$k_i^{t-1} = \min\left[k_{\text{jam}} \frac{N_i^{t-1}}{mx + n(1-x)}\right]. \tag{4}$$

According to the density of SIR, the current speed of vehicle i can be calculated by means of the following exponential function:

$$V_i^t = \begin{cases} V_f^i, & k \leq k_{\max}, \\ \max\left\{V_{\min}, V_{\max}\left\{1 - \left(\dfrac{V_i^{t-1} - k_{\max}}{k_{jam}}\right)^{\beta}\right\}^{\alpha}\right\}, & k > k_{\max}, \end{cases} \tag{5}$$

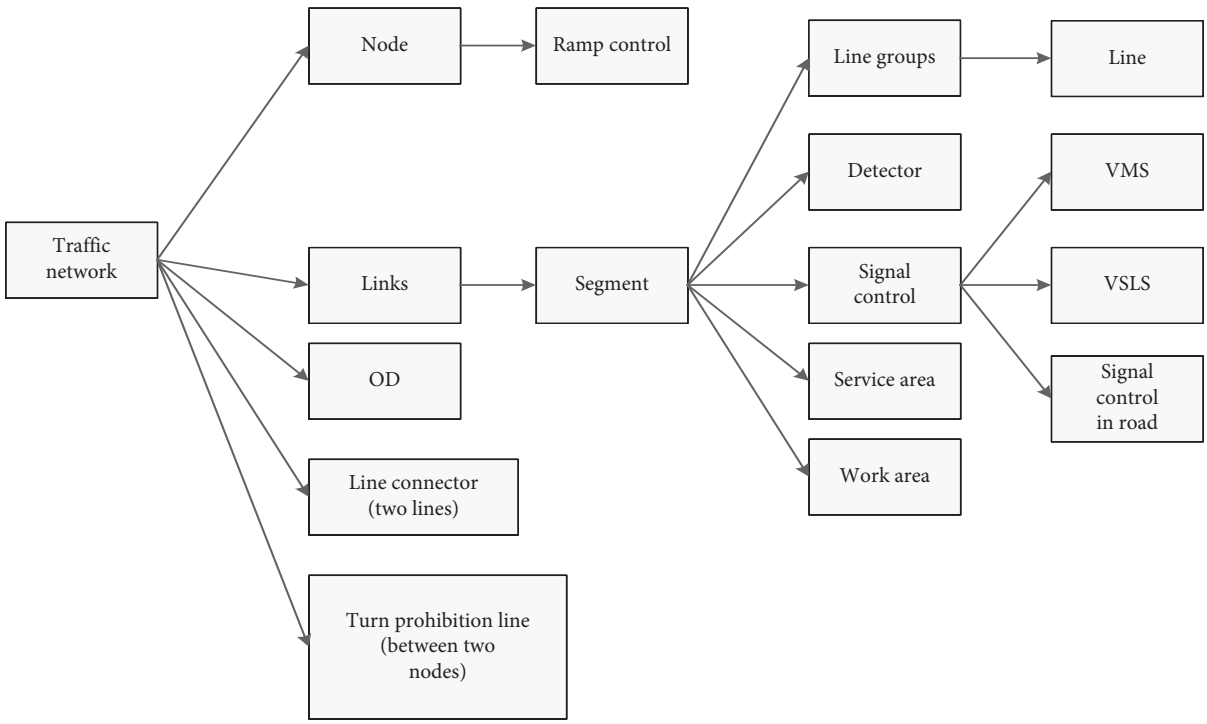


Figure 2: The relationship between network elements.

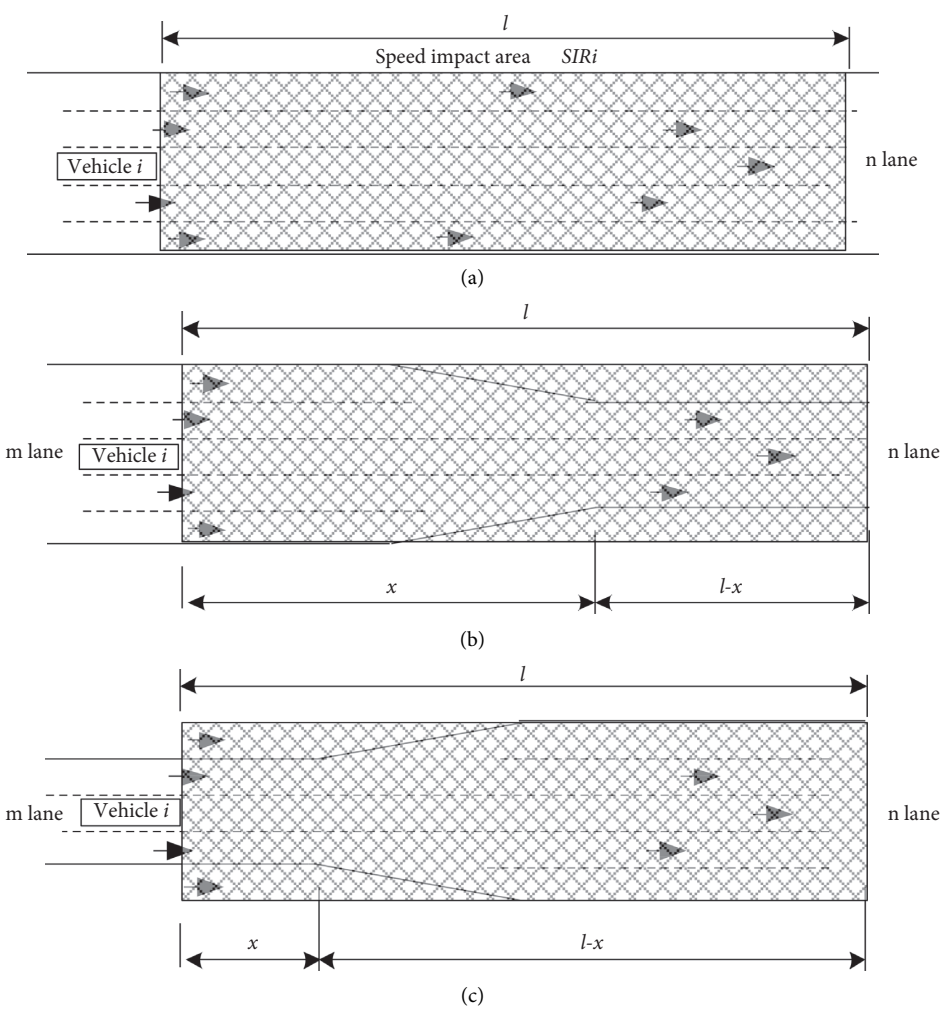


Figure 3: SIR of the Model. (a) Equal number of lane road sections. (b) Reduced number of lanes roads. (c) Increased number of lanes roads.

where V_i^t refers to the speed of vehicle i at the time of t; V_f^i refers to the free flow speed on this segment; $V_{\min}$ refers to the minimum speed that can enable free flow to move; α and β refer to model parameter. The relation between speed and density is shown in the following Figure 4.

The above formula demonstrates the state of the traffic flow without any control. Once there is a VSL control, the vehicles have to obey the speed limit. The model about the relation between speed and density changes.

$$v_i^t = \begin{cases} \min\left(v_f^i, v_{\lim}^i\right), & k \le k_{\max}, \\ \max\left\{ v_{\min}, \min\left(v_{\max}, v_{\lim}^i\right)\left\{1 - \left(\dfrac{\min\left(v_i^{t-1}, v_{\lim}^i\right) - k_{\max}}{k_{\text{jam}}}\right)^{\beta}\right\}^{\alpha}\right\}, & k > k_{\max}. \end{cases} \tag{6}$$

2.4. The Vehicle Movement Model. The speed of vehicle is determined by the density of SIR in front of him when it moves on the segment. If there is no queuing, the vehicle is positioned at z_0 when $t = 0$. Then the vehicle will arrive at z when the time is $t(z)$. Their relation is shown as follows:

$$t(z) = \begin{cases} \dfrac{1}{\lambda}\log\dfrac{\lambda_{z+v_u}}{\lambda_{z_0+v_u}}, & if\ v_u \ne v_d, \\ \dfrac{z - z_0}{V_u}, & if\ v_u = v_d, \end{cases} \tag{7}$$

where λ is determined by the following formula, $\lambda = (v_d - v_u)/L_s$; V_d refers to the downstream speed of the segment; V_u refers to the upstream speed of the segment; L_s refers to the length of segment s. Suppose the time $t = 0$, the vehicle is positioned at z_0, then the position of vehicle is signified by the following formula when the time is t:

$$z(t) = \begin{cases} e^{\lambda t}\left(z_0 + \dfrac{v_u}{\lambda}\right), & if\ v_u \ne v_d, \\ v_u t + z_0, & if\ v_u = v_d. \end{cases} \tag{8}$$

If there is a queue on segment when the vehicle is positioned at $z(t)$, then the position of the vehicle is signified by the following formula when t refers to any time:

$$\begin{aligned} z(t) &= e^{\lambda(t)t}\left(z_0 + \frac{V_u}{\lambda(t)}\right) - \frac{V_u}{\lambda(t)}, \\ \lambda(t) &= \frac{-V_u}{q_0 + l(ct - m)}. \end{aligned} \tag{9}$$

The meaning of q_0, l, c, m is identical with that of the above formula.

If there is a VSL, the above formula is signified as follows:

$$\begin{aligned} z(t) &= \begin{cases} e^{\lambda(t)t}\left(z_0 + \dfrac{\min\left(V_u, V_{\lim}\right)}{\lambda(t)}\right) - \dfrac{\min\left(V_u, V_{\lim}\right)}{\lambda(t)}, & if\ t < t^* \\ q_0 + l(ct - m) & if\ t > t^*, \end{cases} \\ z(t) &= e^{\lambda(t)t}\left[z_0 + \frac{\min\left(V_u V_{\lim}\right)}{\lambda(t)}\right] - \frac{\min\left(V_u, V_{\lim}\right)}{\lambda(t)}, \\ \lambda(t) &= \frac{-\min\left(V_u, V_{\lim}\right)}{q(0) + l(ct - m)}. \end{aligned} \tag{10}$$

3. The Improved VSL Simulation Model

3.1. The Optimization Model. Every connected link of the urban expressway network is defined as several segments. The segment set defined as I, i is one of them. The simulation time K is divided into equally several intervals, and each interval is signified by k. Apparently, $i \in I, k \in K$. Each segment has installed a monitor which includes a variable speed limit sign to obtain the speed, density, flow, and occupancy. The speed, density, and flow of the segment i in the time interval k are signified by $V_i(k)$, $\rho_i(k)$, and $f_i(k)$, respectively. $\vartheta_i(k)$ means occupancy; the direction where i decreases is the vehicle moving one. The speed limit of the time interval k in segment i is $V_{\lim}(i, k)$. The target function studies the largest traffic flow of the whole network during the time interval. Then the improved variable speed limit simulation model is signified by the following formula:

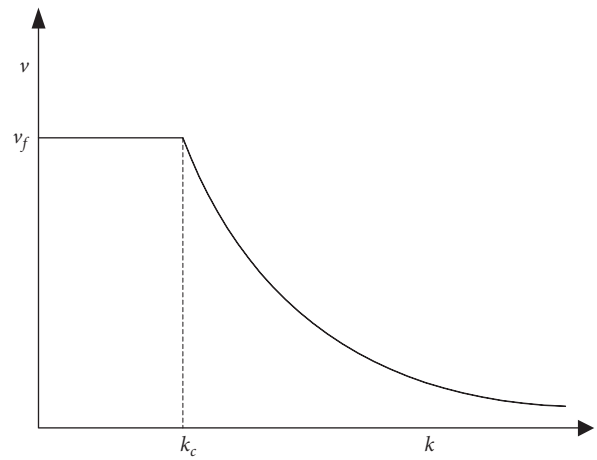


Figure 4: Speed-density curve in SIR.

$$\begin{aligned} &\max \sum_k \sum_i f_i(k), \\ &V_{\lim}(i,k) = \{20, 30, 40, 50, 60, 70, 80, 90\}, \\ &V_{\lim}(i,k) \le V_{\text{uplim}}(i,k), \\ &\quad i = 1, 2, \ldots N; k = 1, 2, \ldots K. \end{aligned} \tag{11}$$

$V_{\text{uplim}}(i,k)$ refers to the maximal value of the speed limit calculated according to the traffic of simulation time interval.

$$\begin{aligned} \vartheta(k) &= f(\vartheta_i(k,l), \vartheta_i(k,2), \ldots, \vartheta_i(k,m_i)), \quad i = 1, 2, \ldots, N; k = 1, 2, \ldots K, \\ V_i(k) &= f(\rho_i(k)), \quad i = 1, 2, \ldots, N; k = 1, 2, \ldots K. \end{aligned} \tag{12}$$

m_i refers to the amount of the detector group included in segment i

$$\begin{aligned} \rho_i(k) &= f(\vartheta_i(k)), \quad i = 1, 2, \ldots, N; k = 1, 2, \ldots K, \\ V_{\lim}(i,k) - V_{\lim}(i-1,k) &\le p(i), \quad i = 1, 2, \ldots, N; k = 1, 2, \ldots K. \end{aligned} \tag{13}$$

$p(i)$ refers to the speed difference between adjacent sections of expressway. Hence, the VSL value is determined by the following formula:

$$V_{\lim}(i,k) = \min\left\{V_{\text{uplim}}(i,k), V_{\lim}(i-1,k) + p(i), V_i(k)\right\}. \tag{14}$$

3.2. The SPSA-Based Solution Algorithm. The conventional analytic algorithm can be used to solve the proposed model, such as the Newton algorithm and the projection algorithm. However, for real-time application, the mesoscopic traffic emulator is adopted to achieve online optimization of VSL control, so it is necessary to meet the real-time requirement of simulation as well as the accuracy of the calculation. Fortunately, the Simultaneous Perturbation Stochastic Approximation algorithm (SPSA) algorithm can cover all the characteristics listed above and it becomes the first choice for the real-time online simulation. It disturbs one time the vector components to be optimized at the same time and has nothing to do with the number of optimized parameters. Compared with the finite-difference stochastic approximation algorithm, its computing capability improves dramatically, which plays an essential role in optimization of large-scale and numerous parameters. The SPSA should reevaluate the gradient vector in each iterative process while the algorithm only needs to perform calculation of objective function twice, which has nothing to do with the amount of parameter n that needs estimating. This feature implies that the algorithm is applied to large-scale optimization.

The SPSA produces a series of estimated parameter values that make the gradient of objective function tend to be 0. Then the updated formula of iteration for the i time is as follows:

$$\theta^{i+1} = \theta^i - a_i \hat{g}\left(\theta^i\right), \tag{15}$$

where θ^i refers to the parameter vector when the algorithm starts to iterate for the i time; $\hat{g}(\theta^i)$ refers to the estimated value of gradient vector; a_i refers to the step length. The approximation of gradient can be obtained through two function evaluations by SPSA. The formula for that is as follows:

$$\hat{g}\left(\theta^i\right) = \frac{z\left(\theta^i + c^i \Delta_i\right)}{2c^i} \begin{bmatrix} \Delta_{i1}^{-1} \\ \Delta_{i2}^{-1} \\ \vdots \\ \Delta ik \end{bmatrix}. \tag{16}$$

The most striking advantage of the above formula lies in the fact that the calculational value of each iterative step is identical, which indicates that it has nothing to do with latitude k. The reason is that Δi is the random perturbation of latitude k, and its numerator keeps identical to k no matter if it is 1, 2, 3, or more. Of course, if an algorithm wants to have a better application value, then it needs a reasonable number of iterations to achieve convergence. The detailed steps of the SPSA algorithm designed in this paper are as follows:

(i) *Step 1.* Initialization: set $i = 0$, $\theta^i = \theta^0$ is the initial value. In actual practice, set a list of nonnegative parameters [21, 22] as the variable speed limit value. At the same time, initialize network state, define traffic network, node, and division of segment, and load on traffic demand and the initiative value of variable speed limit.

(ii) *Step 2.* Set the times of grad-rep for the calculation of gradient vector in each iteration; namely, choose the average value as the gradient estimated value in each step of iteration.

(iii) *Step 3.* Let $i = i + 1$, and calculate formula (16) again.

(iv) *Step 4.* Produce Δ_j, the disturbance random vector of latitude k. Each element Δ_{jk}, $k = 1, 2, \ldots, k$, is obtained through random selection of Bernoulli distribution. The density function of this probability distribution is symmetrical about the vertical axis. And there is an upper bound for $|\Delta_{jk}|$ and $E|\Delta_{jk}^{-1}|$. This upper bound is in accordance with the characteristic of variable speed limit studied in this paper.

(v) *Step 5.* Evaluate the target function of variable speed limit model by the mesoscopic traffic model proposed in this paper; obtain value from $\theta^{i+} = \theta^i + c^i \Delta_j$ and $\theta^{i-} = \theta^i + c^i \Delta_j$. Each point can satisfy the upper and lower bound of optimization. The following mesoscopic traffic simulation is as follows:

(a) *Step 5.1.* Define the update time interval and advance time interval. The relation is as follows:

$$\begin{aligned} T &= k_u \Delta t_{\text{update}}, \\ \Delta t_{\text{update}} &= k_A \Delta_{\text{advance}}. \end{aligned} \tag{17}$$

(b) k_A and k_u are positive integer, and T is the simulation time. The specific parameter of the traffic flow model is mainly updated during update time and the vehicles are moved to a new place in advance interval.

(c) *Step 5.2.* Simulation circulation starts, and the update time interval is utilized. Counter c plus 1, $c = k_u$, happens to discontinue the algorithm.

(d) *Step 5.3.* Optimization circulation starts to output the information about the traffic state, such as the average flow, speed and density of the segments, and the length of the queue. Return to step 5.2.

(vi) *Step 6.* Calculate the gradient vector of latitude k random approximation according to formula (15). The identical numerator of k component serves as the essential content that distinguishes it from finite difference.

(vii) *Step 7.* Repeat step 4 to step 6 for grad-rep times, and each time draw sample independently for Δ_j. Finally, calculate the average value at θ^i.

(viii) *Step 8.* Obtain a new solution point at θ^{i+1}. Then make adjustment according to the question.

(ix) *Step 9.* Return to step 3, and iterate to convergence. The condition for convergence is that the iterative times reach the upper bound, or the function value corresponding to θ^i keeps stable in several continuous iterative processes. Then iteration stops.

Literature [21] indicates that the most satisfactory convergence speed of the SPSA algorithm is around $k^{1/3}$. As direction of search is randomly chosen, hence it cannot be ensured that every step of iteration is downward. But it can be ensured that expectation is unbiased estimate:

$$E\left[\hat{g}\left(\hat{\theta}_k\right)\right] = g\left(\hat{\theta}_k\right) + b^k. \tag{18}$$

$b^k = uc^{k^2}$, u is a constant. As opposed to the comparatively large k, the deviation b^k will gradually disappear with c^k approaching 0. Therefore, it can be concluded that the outcome is closed to the optimal one.

4. Results and Discussion

4.1. Subheadings. The traffic network composed of three elevated expressways of Jinan in Shandong province is adopted as a case study. The three elevated expressways are Beiyuan elevated expressway, Shunhe elevated expressway, and east Erhuan elevated expressway, respectively. Especially, Beiyuan elevated expressway is 13.6km, including 8 pairs of ramp entrances and exits and 6 two-way lanes; east Erhuan elevated expressway is 9.5 km including 4 entrances and exits and 4 two-way lanes; Shunhe elevated expressway is 7.4 km and its width is 18.5 m, including 4 lanes. As one of

Jinan first-class arteries, its total area is 12.25 million square meters, including 1 toll station, 7 entrances and exits, and 3 interchange overpasses. The simplified diagram of traffic network is shown in Figure 5.

In this network, the diamond marked with number represents node, and it also represents point OD for the traffic demand, namely, the place where vehicles appear and disappear. For the convenience of simulation, we define the lanes between nodes which have unified physical characteristics; namely, they can be defined as a segment. The focus of this paper is how the VSL control method diffuses congested vehicles and seeks the law of traffic flow diffusion. In order to simulate the real traffic situation and reveal its essence, the parameters of the traffic flow model are set according to the specific characteristics of the road section (length, width, level, capacity, and traffic flow), where the free flow speed is 90 km/h, the defined ρ_{jam} is 0.1150, and α is 3.500, and β is 0.75004. The capacity for input and output of each segment is 0.6667 per second per car. The time for step length predicted and estimated by the dynamic OD is 15 minutes. The whole simulation lasts for 3 h. This means that there are 12 simulation time intervals. All areas of SIR influenced by heterogeneous directions of the mesoscopic traffic flow simulation model are defined as 150 m. The time for advance step length is 10 s and for update step length is 20 s.

It deserves to explain that emphasis in this paper is exerted on the test of effect and feasibility of the proposed VSL control strategy. Hence set the ideal demand level of dynamic OD, and adopt a periodic loading method that is characterized by a closed cycle. At the beginning of each simulation cycle, the dynamic OD demand is loaded on the network. The normal distribution is adopted to load vehicles to the corresponding point in the network when vehicles are loaded to node or segment in loading point. The VSL cannot perform a function on traffic flow with high density or low speed, nor can it exert influence on fluent traffic network. Thus it only works for moderate demand. Due to the shortage of the real OD demand data, different demand levels are tested to figure out the amount of OD adopted in the simulation. When all demands are set as 1 per interval, all the connected lines in the network appear green, which shows that the traffic runs smoothly. Then the demand is in order turned to 2, 3, 4, and 5, and the results show that the traffic network is relatively smooth, smooth, slight congestion, and congestion. Therefore, the demand of 4 is selected in this paper as the demand level studied concerning variable speed limit control. Additionally, 30% random disturbance is complemented in each time interval of the traffic demand. That is to say, there is 30% random change of time demand by OD in each time interval.

During simulation, all the segments of expressway network are installed with detection, on which the average density, speed, and flow of the traffic flow can be perceived and computed. The variable speed limit value can be shown on VMS installed on roadside with comparatively long physical length to limit the speed of vehicles. Also, it is wise to select the integer among all the variable speed limit values displayed on the VMS; for example, the computed speed of the optimized model is 37 km/h. In this case, displaying 40 km/h is a better choice. In terms of the network of Jinan elevated expressway, it is highly admirable to set 6 VMS, to limit speed from 40 to 60 km/h , and to select integral 10 digits. The simulation process is conducted by one laptop 4G internal storage. First, execute non-variable speed limit network. Then execute the variable speed limit strategy process. Finally, execute the network of optimal variable speed limit strategy. It is shown in Figure 6.

There are two aspects of practical significance in this study. On the one hand, the variable speed limit control proposed in this paper can achieve the purpose of stabilizing traffic flow. According to the traffic condition of the elevated road network, the dynamic adjustment of the speed limit value can achieve the purpose of balanced distribution of traffic flow and improve traffic safety. On the other hand, according to the traffic condition of the urban main road, the traffic efficiency of the urban main road can be improved by adjusting the traffic flow near the ramp exit of the viaduct through variable speed restriction.

4.2. The Simulation Results. Firstly, build the standard state of the traffic network without any control. Then observe its macro- and microcharacteristics. Secondly, set control strategy of VSL and observe the macro- and microcharacteristics of the traffic network with the same demand. Lastly, operate the optimal model of the VSL and the algorithm, also observe and discuss the macro- and microcharacteristics of the traffic network after optimization. From the macroscopic view, the major concern is carried on the observation of the total flow, the total density, and the total speed on the network during the simulation time. Observation of the distribution and variety of the average density, the average speed, and the average flow on each road during each simulation time interval become the priority from the microscopic aspect. At the same time, the traffic impedance of the three cases is compared.

As previously mentioned, the traffic demand in each time interval is defined as 4. Then perform 30% random disturbance to build the standard state of the traffic network. In this way, the traffic demand in each interval is subtlety different. Under this standard, the time of the whole simulation witnesses the following aspects: the total travelling time is 2759.7 h, the total travelling distance is 156169.3 km, the total average speed is 42.6 km/h , the total traffic flow is 188399 vehicles, the total delay is 879.1 h, the largest delay on node is 153.4 h, the total density is 4.5412, and the average flow is 96.0930. In terms of the microscopic aspect, the distribution of the average density, the average speed, and the average flow in each segment is shown in Figures 7–9.

Under the above standard state, the initial VSL control scheme is set. The Beiyuan elevated expressway, the Shunhe elevated expressway, and the Erhuan elevated expressway are installed with eight VMS, respectively, including four in each direction. These VMS are placed on the middle part of the road with comparatively long physical length with the aim of controlling the speed of the traffic flow. Considering the traffic safety and the traffic congestion of the elevated

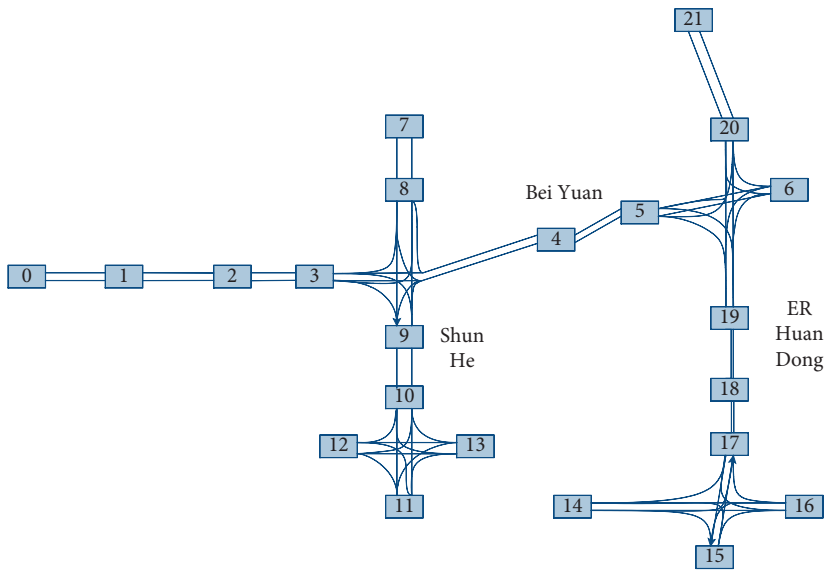


Figure 5: The elevated expressway network of Jinan.

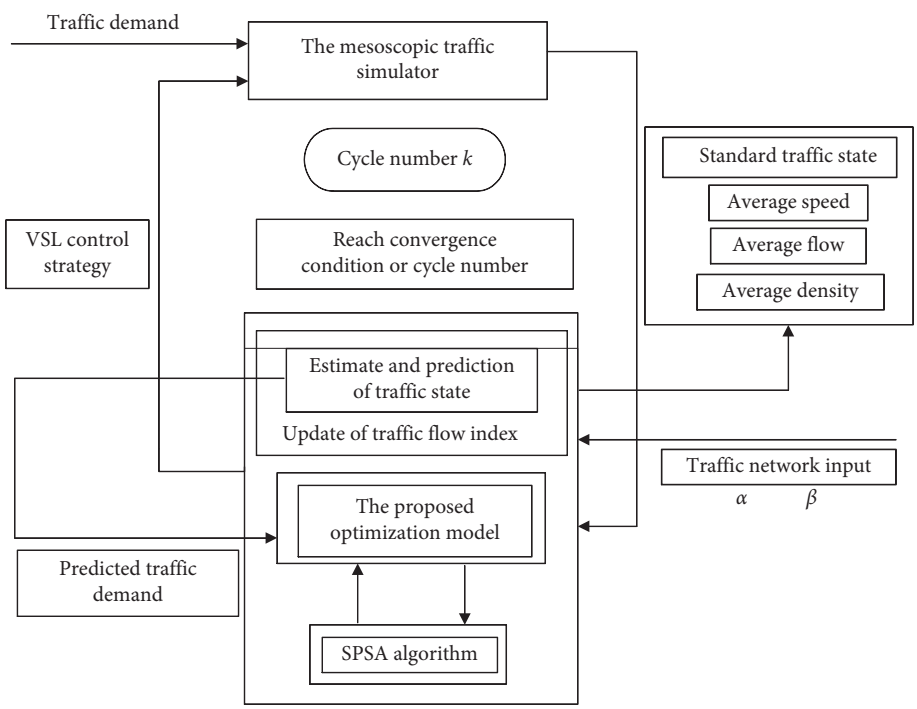


Figure 6: The simulation structure.

expressway connecting the network of the urban arteries, the speed of the traffic flow on the expressway should be decreased during the rush hour. Therefore, the initial value of variable speed limit is set on 7 : 30, from which the speed is limited to 60 km/h. From 8 : 00, the speed is limited to 50 km/h and the speed is limited to 40 km/h at 9 : 00. Then the above information variables have changed. The travelling time is 5092.6 h; the total travelling distance is 157856.5; the total average speed is 31 km/h; the total traffic flow is 187008 per car unit; the total delay is 3194.3 h; the total density is 6.3272 per car; the average traffic flow is 94.4721 cars. There is no traffic congestion under standard state. But when there is a control, the traffic congestion area reaches 19% of the whole network. According to the data, the traffic safety gets further ensured when there is a control over variable speed limit. However, the efficiency of the network operation decreases sharply. In particular, the traffic congestion occurs in some areas under network state. The whole network state improves greatly. From the perspective of microscopic aspect, the distribution of the average density, the average speed, and the average flow of each segment during each simulation time interval is shown in Figures 10–12, which

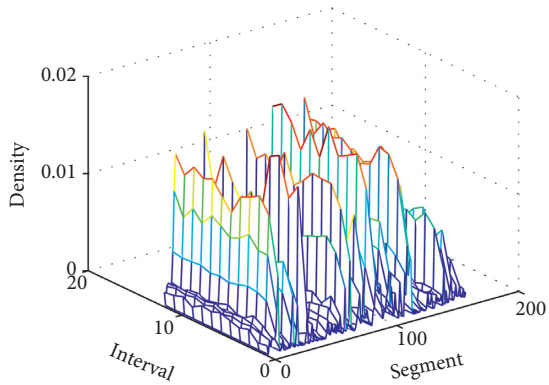


Figure 7: The averaged density without control.

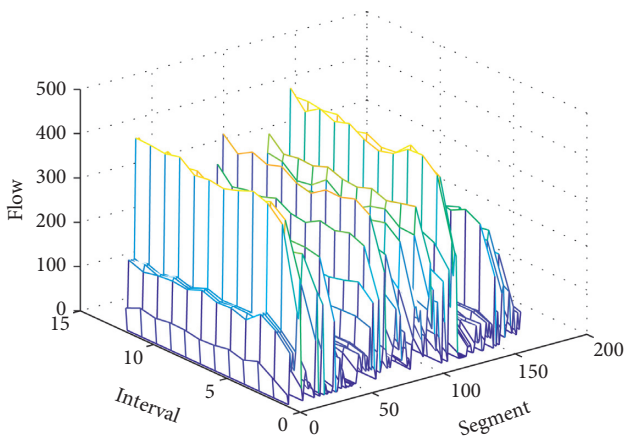


Figure 8: The averaged flow without control.

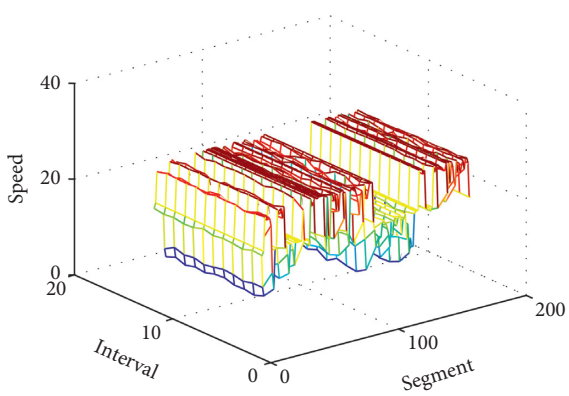


Figure 9: The averaged speed without control.

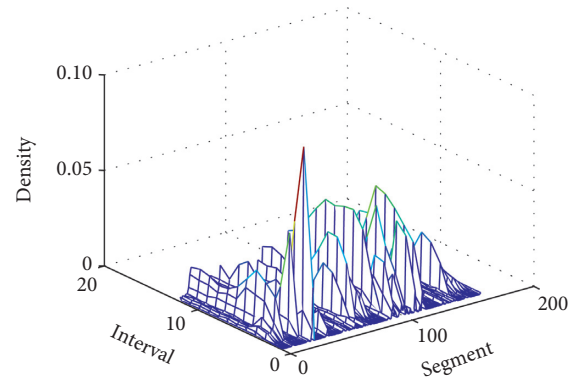


Figure 10: The averaged density with control.

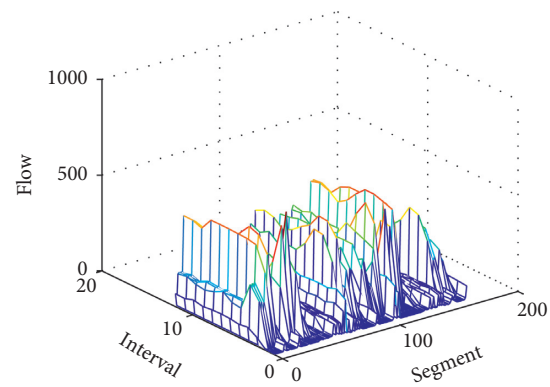


Figure 11: The averaged flow with control.

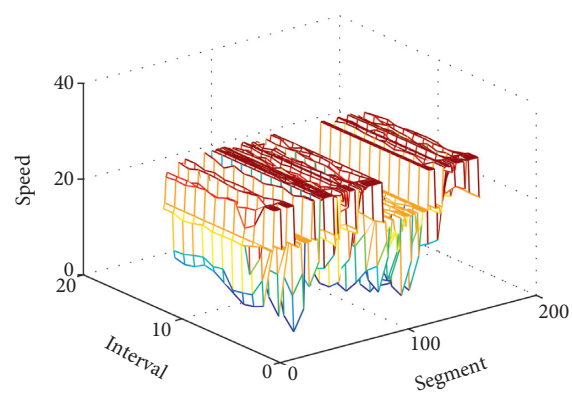


Figure 12: The averaged speed with control.

shows that the distribution of the network density is in imbalance and so do the distribution of speed and flow. Particularly, the average speed decreases dramatically. Thus, utilizing variable speed limit without control may lead to dramatic exacerbation of the traffic network. Obviously, the operation of the traffic flow can be adjusted through variable speed limit once there is a traffic accident. Therefore, the most satisfying scheme for setting control strategy of variable speed limit should be selected in reality.

Under the above state, the implementation of the improved model of variable speed limit and the SPSA algorithm can figure out the most satisfying scheme of variable speed limit and then make it executed, thus testing all indexes of the traffic network. Again the above information has changed. The total travelling time is 3511.7 h; the total travelling distance is 136520.4 km; the total average speed is 36.8 km/h; the total traffic flow is 161809 cars; the total delay is 1870 h; the largest delay at node is 337.2 h; the total density

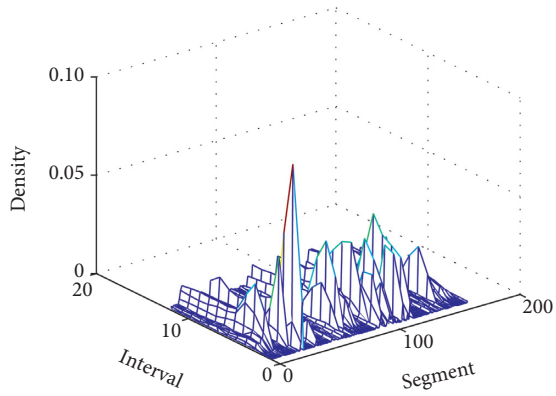


Figure 13: The optimal averaged density.

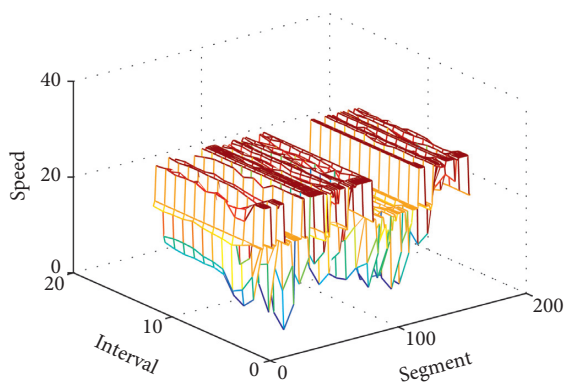


Figure 14: The optimal averaged speed.

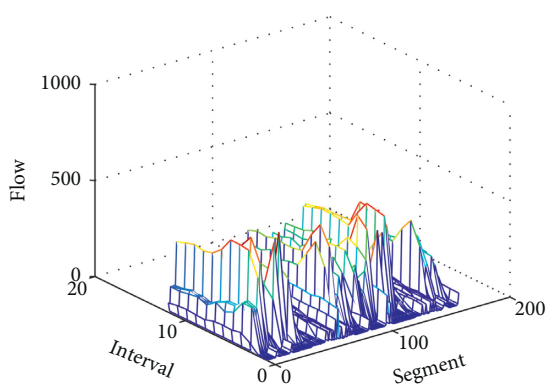


Figure 15: The optimal averaged flow.

is 4.7195; the average flow is 82.3237 cars. The congestion area is 1% of the total network under the state of optimized control. The light congestion road is 12%. It can be seen that the density appears well-distributed in network compared with that of the control. The efficiency of vehicles' movement improves a lot. From the perspective of microscopic aspect, the distribution of the average density, the average speed, the average flow of each segment during each simulation time interval is shown in Figures 13–15. Apparently, the density, speed, and flow appear well-distributed and the average speed improves greatly. This saves passengers' time for travelling. Meanwhile, the VSL ensures the traffic safety by optimize the traffic network indexes and enables the traffic flow to move orderly.

The simulation results show that the VSL exerts considerable influence on the traffic flow distribution and the dynamic characteristics. The existence of the speed limitation in different time will inevitably lead to the decrease of the average speed. And with the increase of the total density, the total mileage of vehicles decreased, but it can be seen that the delay on the largest node does not increase significantly. If the proposed optimization control strategy is employed, it will decrease mostly the negative effect resulting from the speed limit and optimize the whole network of this area. According to the relevant study, the proper reduction of speed is beneficial for ensuring traffic safety and reducing the traffic accident.

It is worth noting that the movement of vehicle is in accordance with some regulations whether the model and algorithm are used in simulation test. This is similar to the fact that vehicles obey traffic rules and move orderly at intersections where there is no signal control. In reality, intersections with heavy traffic flow but no signal control, however, are usually the places where severe traffic congestion occurs. This often leads to vicious circle which results from violation against traffic rules and competition between drivers. Therefore, it must implement the control over traffic flow. However, the effect exerted by obeying traffic rules is much more admirable than that by control. Therefore, under the most satisfying state presented in this paper, the network state is much better than that of the time when there is no optimization but control implementation. Moreover, the efficiency of travelling improves rapidly. Compared to state without control, the network is much worse.

Although the time interval divided is a little longer, it does not affect the analysis of questions. The time interval can be shortened according to reality. This means that there is one more variable value domain in the SPSA algorithm, in which the advantage of this algorithm lies. The more the variables are, the more advantageous the algorithm is. That is to say, the speed and astringency of the algorithm have nothing to do with a variable number. The refinement of time interval is supportive to obtaining a better control strategy, therefore, making the control strategy over variable speed limit in each time interval more effective.

One of the objectives to employ VSL control is to maintain the traffic safety. The adjustment of speed according to time interval can contribute to not only traffic safety but also the travelling efficiency of road network. Secondly, the VSL can manage effectively the distribution of the traffic flow that always changes and guide the traffic flow for the sake of traffic structure. Thirdly, in some congestion areas, especially the places in which the travelling capacity of lanes decrease sharply (construction area; accident area), the VSL can be used to create a free flow area in the upstream, thus making traffic capacity of it equivalent to or slightly less than that of the largest lane group of this area. This can minimize the traffic congestion and the resulting chain reaction.

5. Research Conclusions and Prospects

As an effective way to control the continuous traffic flow of freeway or expressway, the VSL control strategy can maintain traffic safety and adjust the distribution of the traffic flow. Considering the real time and accuracy of the VSL control strategy, this paper selects the mesoscopic traffic flow model as the study means, embeds VSL control strategy into the model, builds the improved mesoscopic traffic flow simulation model, and then designs its simulation flow. Next, the improved model of VSL control and the SPSA-based algorithm are proposed. The advantage of this algorithm lies in the fact that it can not only guarantee the accuracy of simulation but also meet online demands for real-time application. The simulation results show that the improved model and algorithm can enhance dramatically the efficiency of the network, the average speed, and the traffic capacity and enable the traffic density to appear well-distributed. Meanwhile, it can also ensure the traffic safety and be applied to such traffic circumstances as construction area or accident area. Compared to all vehicles that are intelligently equipped and obey unified traffic rules, the traffic network without any implementation of control strategy works better, which makes no sense in reality because areas without control strategy tend to be places in which the traffic congestion frequently occurs. The emphasis of the latter study is on the combination of VSL control strategy of urban expressway and signal control of urban arteries and initiation of aggregative model and algorithm. This is supportive for traffic sector to make decision.

Data Availability

The main data of this study are obtained by mesoscopic simulation software.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 71871130, 71771019, and 71971125); the University Science and Technology Program Funding Projects of Shandong Province (Grant no. J17KA211); the Social Science Foundation of Shandong Province (Grant no. 12BJJJ13); and the Science and Technology Program Funding Projects of Shandong Police College (Grant no. YSKYB2005).

References

[1] J. Mccrea and S. Moutari, "A hybrid macroscopic-based model for traffic flow in road networks," *European Journal of Operational Research*, vol. 207, no. 2, pp. 676–684, 2010.

[2] Y. Diao, Z. C. Juan, and A. N. Ni, "Review of mesoscopic traffic flow modeling and system simulation," *Application Research of Computers*, vol. 26, no. 7, pp. 2411–2415, 2009.

[3] S. B. Li, *Research on the Run State Evaluation of Urban Transportation System and Control Strategies*, Beijing Jiao Tong University, Beijing, China, 2012.

[4] D. F. Xie, *Analyzing of Typical Problems of Urban Road Traffic Flow Based on Microscopic Methods*, Beijing Jiao Tong University, Beijing, China, 2010.

[5] T. Wang, *Research on Traffic Flow Modeling and Simulating Based on the Lattice Hydrodynamic Model*, Beijing Jiao Tong University, Beijing, China, 2015.

[6] J. F. Tian, Z. Z. Yuan, J. Bin et al., "Phase transitions and the Korteweg-devries equation in the density difference lattice hydrodynamic model of traffic flow," *International Journal of Modern Physics C*, vol. 24, no. 3, Article ID 1350016, 2013.

[7] T. Wang, Z. Gao, W. Zhang, J. Zhang, and S. Li, "Phase transitions in the two-lane density difference lattice hydrodynamic model of traffic flow," *Nonlinear Dynamics*, vol. 77, no. 3, pp. 635–642, 2014.

[8] T. Wang, Z. Y. Gao, J. Zhang et al., "A new lattice hydrodynamic model for two-lane traffic with the consideration of density difference effect," *Nonlinear Dynamics*, vol. 75, no. 1-2, pp. 27–34, 2014.

[9] D. S. Felipe, V. Omer, and J. Auld, "Mesoscopic traffic flow model for agent-based simulation," *Procedia Computer Science*, vol. 151, pp. 858–863, 2019.

[10] J. Anahita, P. Ioannis, P. Markos et al., "A mesoscopic integrated urban traffic flow-emission model," *Transportation Research Part C*, vol. 75, pp. 45–83, 2017.

[11] A. A. Mohamed, R. J. Cunningham, V. V. Gayah et al., "Dynamic variable speed limit strategies for real-time crash risk reduction on freeways," *Transportation Research Record*, vol. 2078, no. 1, pp. 108–116, 2008.

[12] Y. Pu, L. Hu, Y. S. Jiang et al., "Variable speed-limit control before expressway mainline toll station," *Journal of Traffic and Transportation Engineering*, vol. 12, no. 5, pp. 119–126, 2012.

[13] J. Soobin, P. Chongmyung, and S. Dongmahn, "The multi-station based variable speed limit model for realization on urban highway," *Electronics*, vol. 9, no. 5, 2020.

[14] Y. Zhou, B. Lin, X. Yang et al., "Application of the bayesian model averaging in analyzing freeway traffic incident clearance time for emergency management," *Journal of Advanced Transportation*, vol. 2021, no. 4, 9 pages, Article ID 6671983, 2021.

[15] J. J. Zhang, M. B. Pang, and S. S. Ren, "Characteristic analysis of traffic flow in variable speed limit section of freeway based on cellular automaton model," *Acta Physica Sinica*, vol. 61, no. 24, pp. 340–347, 2012.

[16] W. Wang, Z. S. Yang, and D. X. Zhao, "Control model of variable speed limit based on finite horizon Markov Decision-making," *Journal of Traffic and Transportation Engineering*, vol. 11, no. 5, pp. 109–114, 2011.

[17] W. Y. Zhou, M. F. Yang, M. H. Lee et al., "Q-Learning-Based coordinated variable speed limit and hard shoulder running control strategy to reduce travel time at freeway corridor," *Transportation Research Record*, vol. 2674, no. 11, 2020.

[18] Y. Wang, X. Ma, Z. Li, Y. Liu, M. Xu, and Y. Wang, "Profit distribution in collaborative multiple centers vehicle routing problem," *Journal of Cleaner Production*, vol. 144, pp. 203–219, 2017.

[19] Y. Wang, S. G. Peng, X. S. Zhou et al., "Green logistics location-routing problem with eco-packages," *Transportation Research Part E: Logistics and Transportation Review*, vol. 143, Article ID 102118, 2020.

[20] Y. Wang, Y. Y. Yuan, X. Guan et al., "Collaborative two-echelon multicenter vehicle routing optimization based on

state–space–time network representation," *Journal of Cleaner Production*, vol. 258, Article ID 120590, 2020.

[21] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 332–341, 2002.

[22] J. C. Spall, "Implementation of the simultaneous perturbation algorithm for stochastic optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 3, pp. 817–823, 1998.