Journal of Advanced Transportation

Advanced Models and Practice for Addressing Emerging Cross-Cutting Issues in Multimodal Transportation Research

Lead Guest Editor: Wei (David) Fan
Guest Editors: Xingju Wang, Lu Gao, and Zhigang Xu

Advanced Models and Practice for Addressing Emerging Cross-Cutting Issues in Multimodal Transportation Research

Journal of Advanced Transportation

Advanced Models and Practice for Addressing Emerging Cross-Cutting Issues in Multimodal Transportation Research

Lead Guest Editor: Wei (David) Fan
Guest Editors: Xingju Wang, Lu Gao, and Zhigang Xu

Contents

Forecasting Urban Rail Transit Vehicle Interior Noise and Its Applications in Railway Alignment Design
Yifeng Wang, Ping Wang, Zihan Li, Zhengxing Chen, and Qing He
Research Article (13 pages), Article ID 5896739, Volume 2020 (2020)

A Gradient Boosting Crash Prediction Approach for Highway-Rail Grade Crossing Crash Analysis
Pan Lu, Zijian Zheng, Yihao Ren, Xiaoyi Zhou, Amin Keramati, Denver Tolliver, and Ying Huang
Research Article (10 pages), Article ID 6751728, Volume 2020 (2020)

Compressive Strength Gain Behavior and Prediction of Cement-Stabilized Macadam at Low Temperature Curing
Yongli Xu, Guang Yang, and Hongyuan Zhao
Research Article (10 pages), Article ID 2469436, Volume 2020 (2020)

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 8863487, 12 pages
https://doi.org/10.1155/2021/8863487

Research Article

Variable Speed Limit Control Method of Freeway Mainline in Intelligent Connected Environment

Xingju Wang,[1,2] **Rongqun Zhang**,[2] **Yang Gou**,[3] **Jiayu Liu**,[2] **Lin Zhao**,[2] **and Yanting Li**[2]

[1]*State Key Laboratory of Mechanical Behavior and System Safety of Traffic Engineering Structures, Shijiazhuang Tiedao University, Shijiazhuang 050043, China*
[2]*Shijiazhuang Tiedao University, Shijiazhuang 050043, China*
[3]*China Railway Design Corporation, Tianjin 300000, China*

Correspondence should be addressed to Xingju Wang; wangxingju@stdu.edu.cn

Received 16 July 2020; Revised 26 June 2021; Accepted 14 July 2021; Published 23 July 2021

Academic Editor: Giuseppe Musolino

Freeway is an important component of transportation system. Bottleneck areas on freeway reduce driving safety and traffic efficiency. The development of intelligent connected technology provides a new idea for traffic management. In order to alleviate traffic congestion on the freeway bottleneck area, this paper proposes a variable speed limit (VSL) control method in intelligent connected environment. In this paper, the METANET model is improved by combining intelligent connected environment and VSL control theory. The total traffic capacity (TTC), total travel time (TTT), and total speed difference (TSD) are used to build multiobjective function. The microsimulation at SUMO by using the data from PeMS is employed as a case study to validate the proposed model. The results show that the VSL online control method in intelligent connected environment has better control effect. And the improvement is more obvious with increasing penetration rate of intelligent connected vehicle (ICV).

1. Introduction

Due to the rapid growth of freeway transport volume and the influence of holiday travel peak, large-scale congestion of freeway has become an increasingly serious problem. Congestion often occurs where the road segment is temporarily closed for construction. Bottleneck areas, such as the intersection of the main line and the ramp, are also likely to cause congestions. Traditionally static speed limit control technology of freeway cannot meet the existing complex traffic environment.

Many studies have proposed different approaches and models to reduce traffic congestion and delay [1]. Variable speed limit control is a method of traffic management commonly used in freeway systems [2]. Grumert and Tapani [3] summarized important characteristics that affect performance of VSL system, indicating that different characteristics and control algorithms have different impact on traffic performance. Cao et al. [4] analyzed the congestion distribution in various situations under VSL control and non-VSL control. The key of the VSL control is the analysis of the characteristics of traffic state. Therefore, whether the traffic flow model is reasonable directly affects the application effect of the VSL control strategy. Festa et al. [5] analyzed two traffic models in noncongested traffic conditions and discussed models potentialities and limits in large-scale applications. Yang et al. [6] proposed an active VSL control method by minimizing two control objectives, travel time minimization and velocity variance minimization, and found that the coincidence rate of drivers was an important factor for the consistency between reality and VSL system. Hegyi et al. [7] presented a model predictive control (MPC) approach to optimally coordinate variable speed limit for freeway traffic. Li and Ranjitkar [8, 9] designed a VSL control algorithm based on fuzzy logic in 2015 and a method based on particle swarm optimization in 2019. Yang et al. [10] adopted the Kalman filter VSL control system enhancement module based on the macroscopic traffic flow model to solve

the displacement capacity of the upstream and downstream of the freeway. Kotsialos et al. [11] used the second-order traffic flow METANET model as a core model, which had good expression characteristics for VSL control. These experiment results showed that advanced models can effectively improve the traffic capacity of road congested sections, including SCTM-METANET [12], proactive VSL based on Dyna TAM-VSL [13], a variety of vehicles model [14], and predictive control based on the cell transmission model [15].

ICV is equipped with advanced sensors, controllers, and modern communication technology, which not only provide more real-time information to infrastructure but also can be controlled and optimized [16]. Traffic flow on the road is generally mixed with ICVs and manual vehicles for a long time to come [17]. Optimal traffic control strategies contribute to reducing congestion, travel time, and fuel consumption [18]. The introduction of intelligent connected technology can optimize traffic flow, reduce total travel time on the road [19], optimize control distance and vehicle performance [20, 21], and decrease the vehicle delay and fuel consumption [22]. The construction method of intelligent transportation system based on ICV and big data technology had also been discussed [23]. Cellular automaton model containing the collaborative components of the network and autonomous vehicles (CAVs) [24] and MVDE model considering the interaction between front and rear vehicles [25] both confirmed the important role of intelligent connected technology to improve transport service level. Furthermore, Grumert and Tapani [26] confirmed VSL system can manage to decrease the difference in speeds between individual vehicles by proposed evaluation method. Guo et al. [27] proposed an effective real-time traffic information sharing mechanism which is VANET-assisted. Song et al. [28] proposed a dynamic vehicle path guidance model based on global adaptive optimization scheduling under the Internet of vehicles. Erdağı et al. [29] created two hypothetical test networks with different levels of complexity, in which total time spent and total emission were considered to find an optimum penetration rate of CACC in urban road. Yao et al. [30] established a method for the stability of mixed traffic flow and obtained the fundamental diagram model under different penetration rates of CAVs.

Many studies have applied VSL control to freeway management control system. Most of these works are based on historical traffic data, which cannot reflect the time-varying of traffic flow parameters. The control objective is single, and the method to solve the objective function is a trial algorithm that is difficult to be applied to complex control system. Few studies discuss the adaptive improvement of freeway control methods according to the characteristics of intelligent connected technology. In this paper, the METANET model is optimized and improved by considering characteristics of intelligent connected environment. The VSL control objective function is proposed by maximizing TTC and minimizing TTT and TSD. Finally, the microsimulation is used to verify the effectiveness of the proposed methodology.

2. Methodology

2.1. Enhanced METANET Model. As shown in Figure 1, the typical freeway section is divided into N segments with length of l. The METANET model calculates the state of traffic flow at the next time interval based on the recurrence formula of the model using the flow, speed, and density of the subsection.

For each segment i, the relationship between average traffic volume, average traffic density, and average traffic speed during time interval k can be expressed as

$$q_i(k) = \rho_i(k) v_i(k) \lambda_i, \tag{1}$$

where $q_i(k)$ is average volume of vehicles on segment i during time interval k; $\rho_i(k)$ is average density on segment i during time interval k; $v_i(k)$ is average speed on segment i during time interval k; λ_i is the number of lanes.

The density can be determined by on-ramp flow and off-ramp flow as

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{\lambda_i l_i}\left(q_{i-1}(k) - q_i(k) + r_i(k) - s_i(k)\right), \tag{2}$$

where T is discrete time step used in the METANET model; l_i is the length of segment i; $r_i(k)$ is the on-ramp traffic volume on the segment i during time interval k; $s_i(k)$ is the off-ramp traffic volume on the segment i during time interval k.

In METANET model, the convection of the speed dynamic function is more consistent with the actual situation in the free flow state. However, under the condition of congestion, this is relatively deviated from the actual operating state. This paper enhanced the third term to alleviate the impact of upstream speed value on downstream traffic flow. Improved speed dynamics formula [31] is as follows:

$$\begin{aligned} v_i(k+1) = {} & v_i(k) + \frac{T}{\tau}\left[V\left[\rho_i(k)\right] - v_i(k)\right] + \frac{T}{l_i} v_i(k)\left[\sqrt{\frac{v_{i-1}^2(k) + v_i^2(k)}{2}} - v_i(k)\right] \\ & - \frac{1}{\tau}\left[\frac{T\nu}{L_i}\frac{\rho_{i+1}(k) - \rho_i(k)}{\rho_i(k) + \kappa}\right], \end{aligned} \tag{3}$$

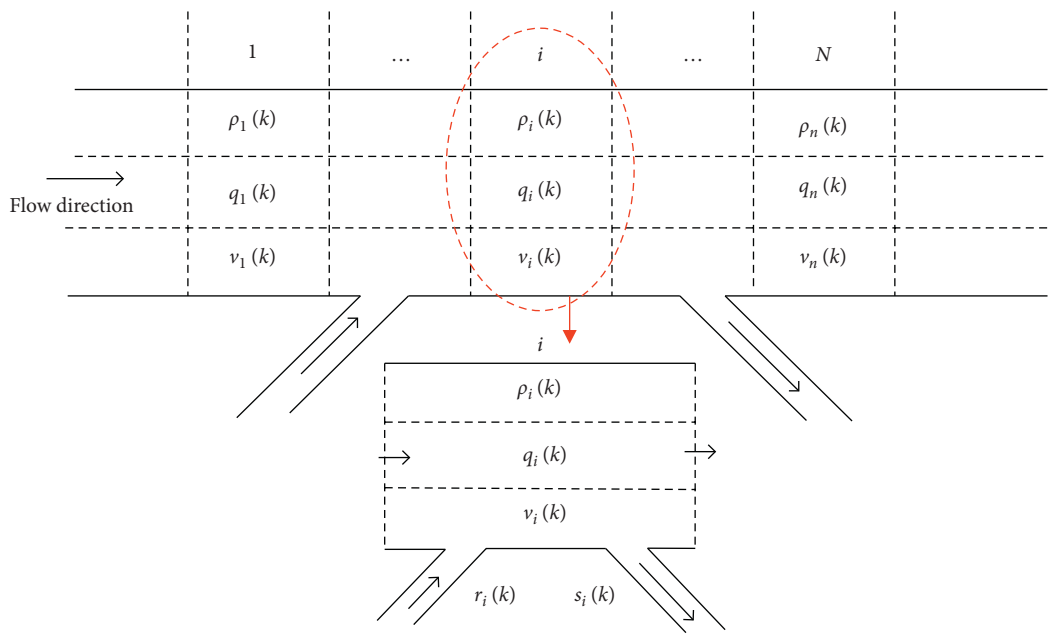


Figure 1: Typical freeway segment diagram.

where τ, ν, and κ are global parameters of METANET model, determined jointly by characteristics such as road conditions, driver behavior, and vehicle performance.

For most traffic flow models, under the condition of no speed limit control, the driver's desired speed is free flow speed. However, in the VSL control environment, the driver's desired speed is not free flow speed, but the modified speed limit of the road section; that is, each road section aims to control the driver's desired speed below the reasonable speed limit value. Improved driver's desired speed [32] can be expressed as follows:

$$V[\rho_i(k)] = u_i(k)\exp\left[-\frac{1}{o_i}\left(\frac{\rho_i(k)}{\rho_c}\right)^{o_i}\right], \tag{4}$$

where o_i is set to 2 in this paper; $u_i(k)$ is the speed limit value on segment i during time interval k.

Compared with manual vehicles, ICV has a higher compliance rate. The maximum speed of a vehicle can be adjusted according to vehicles with higher compliance rate [33]. Therefore, in the intelligent connected environment, the control center improves the speed limit value of each control section by collecting real-time data from the road traffic state monitoring device and the ICV. This can be expressed as follows [14]:

$$\begin{aligned} u_i^{\mathrm{acc}}(k) &= (1+\alpha_i(k)) \times u_i(k), \\ \alpha_i(k) &= \frac{v_i^{\mathrm{acc}}(k-1) - u_i^{\mathrm{acc}}(k-1)}{u_i^{\mathrm{acc}}(k-1)}, \end{aligned} \tag{5}$$

where $u_i^{\mathrm{acc}}(k)$ is the modified speed limit in the intelligent connected environment on segment i during time interval k; $\alpha_i(k)$ is the relation coefficient between the ICV speed and the speed limit on segment i during time interval k; $v_i^{\mathrm{acc}}(k-1)$ is average speed of ICV on segment i during time interval $k-1$; $u_i^{\mathrm{acc}}(k-1)$ is the speed limit on segment i during time interval $k-1$.

2.2. Control Objective Function. On the premise of ensuring the traffic capacity of the road, the control objective function is established by combining TTC, TTT, and TSD to improve the traffic efficiency of the road and the driving comfort. The weight relationship among the three terms is balanced to realize the coordinated control of VSL.

Increasing traffic volume on the road segment can improve the efficiency of the road traffic. The TTC as one of control objectives during the predicted time length Np can be expressed as

$$\mathrm{TTC} = T\sum_{k=1}^{N_p}\sum_{i=1}^{N}\lambda_i l_i \rho_i(k) v_i(k). \tag{6}$$

Reducing the TTT can effectively improve the efficiency of road traffic. The TTT as one of control objectives can be expressed as

$$\mathrm{TTT} = T\sum_{k=1}^{N_p}\sum_{i=1}^{N}\lambda_i l_i \rho_i(k). \tag{7}$$

When the difference of driver's speed changes is large, the vehicle will accelerate and decelerate to a large extent. It is easy to cause traffic collisions and to reduce stability of traffic flow. The TSD selected as the control target can be expressed as

$$\mathrm{TSD} = \sum_{k=1}^{N_p}\sum_{i=1}^{N}\left(u_i^{acc}(k) - v_i(k)\right)^2. \tag{8}$$

In this paper, control objective function includes TTC, TTT, and TSD. In the optimal state, TTC is the maximum

while TTT and TSD are the minimum. In this case, the speed limit value is the optimal solution of our optimization model. The multiobjective function normalizes TTC, TTT, and TSD to obtain S_1, S_2, and S_3 and combines with the corresponding weight. Then the optimal speed limit value is obtained by minimization of multiobjective function. It can be expressed as

$$\begin{aligned} S_1 &= \frac{\mathrm{TTC} - \mathrm{TTC}_{\min}}{\mathrm{TTC}_{\max} - \mathrm{TTC}_{\min}}, \\ S_2 &= \frac{\mathrm{TTT} - \mathrm{TTT}_{\min}}{\mathrm{TTT}_{\max} - \mathrm{TTT}_{\min}}, \\ S_3 &= \frac{\mathrm{TSD} - \mathrm{TSD}_{\min}}{\mathrm{TSD}_{\max} - \mathrm{TSD}_{\min}}, \\ &\mathrm{Min}\left\{-w_1\frac{\mathrm{TTC} - \mathrm{TTC}_{\min}}{\mathrm{TTC}_{\max} - \mathrm{TTC}_{\min}} + w_2\frac{\mathrm{TTT} - \mathrm{TTT}_{\min}}{\mathrm{TTT}_{\max} - \mathrm{TTT}_{\min}} + w_3\frac{\mathrm{TSD} - \mathrm{TSD}_{\min}}{\mathrm{TSD}_{\max} - \mathrm{TSD}_{\min}}\right\}, \end{aligned} \tag{9}$$

where w_1, w_2, and w_3 are weight parameter of TTC, TTT, and TSD, respectively; $\mathrm{TTC}_{\max}$ and $\mathrm{TTC}_{\min}$ are the maximum and minimum of TTC, respectively; $\mathrm{TTT}_{\max}$ and $\mathrm{TTT}_{\min}$ are the maximum and minimum of TTT, respectively; $\mathrm{TSD}_{\max}$ and $\mathrm{TSD}_{\min}$ are the maximum and minimum of TSD, respectively.

The weight was determined by the 1–9 scale method in the analytic hierarchy process to calibrate the importance of the control target. The quantitative rules of the 1–9 scale method are shown in Table 1.

By analyzing the importance of these three functions, the judgment matrix of the objective function is obtained, as shown in Table 2.

The weight coefficient of each objective function is calculated according to the square root method, which can be expressed as

$$w_i = \left(\prod_{j=1}^{n} a_{ij}\right)^{1/n}, \tag{10}$$

where a_{ij} is the element value of the i th row and j th column in the target judgment matrix.

Normalization according to the formula can be expressed as

$$w = \frac{w_i}{\sum_i^n w_i}. \tag{11}$$

The weight calculation results are shown in Table 3.

The consistency test is carried out on the results in Table 3 and the verification result CR = 0.0474 < 0.1, which is in line with inspection standards.

The speed limit value of the freeway should be controlled within a reasonable range to ensure the comfort and safety of driving behavior while improving the traffic efficiency of the freeway. Our multiobjective function is given by the following formula:

$$\mathrm{Min}\left\{-0.6491\frac{\mathrm{TTC} - \mathrm{TTC}_{\min}}{\mathrm{TTC}_{\max} - \mathrm{TTC}_{\min}} + 0.2790\frac{\mathrm{TTT} - \mathrm{TTT}_{\min}}{\mathrm{TTT}_{\max} - \mathrm{TTT}_{\min}} + 0.0719\frac{\mathrm{TSD} - \mathrm{TSD}_{\min}}{\mathrm{TSD}_{\max} - \mathrm{TSD}_{\min}}\right\}. \tag{12}$$

It is subject to

$$u_i^{\mathrm{acc}}(k) \le V_{\max}, \tag{13}$$

$$u_i^{\mathrm{acc}}(k) \ge V_{\min}, \tag{14}$$

$$\left|u_i^{acc}(k+1) - u_i^{\mathrm{acc}}(k)\right| \le 20\frac{\mathrm{km}}{\mathrm{h}}, \tag{15}$$

Table 1: 1–9 scale quantization.

Scale value	Description
1	Same importance
3	The former is slightly more important than the latter
5	The former is significantly more important than the latter
7	The former is more important than the latter
9	The former is extremely more important than the latter
2, 4, 6, 8	Intermediate value of importance

Table 2: Objective function judgment matrix.

Objective function	S_1	S_2	S_3
S_1	1	3	7
S_2	1/3	1	5
S_3	1/7	1/5	1

Table 3: Weight calculation results.

Objective function	S_1	S_2	S_3
Weight	2.7589	1.1856	0.3057
Normalized result (w)	0.6491	0.2790	0.0719

$$\left|u_{i+1}^{\mathrm{acc}}(k)-u_i^{\mathrm{acc}}(k)\right| \le 20\frac{\mathrm{km}}{\mathrm{h}}, \tag{16}$$

$$u_i^{\mathrm{acc}}(k) \in \left\{\begin{array}{c} 40\frac{\mathrm{km}}{\mathrm{h}}, 45\frac{\mathrm{km}}{\mathrm{h}}, 50\frac{\mathrm{km}}{\mathrm{h}}, 55\frac{\mathrm{km}}{\mathrm{h}} \\ \ldots 105\frac{\mathrm{km}}{\mathrm{h}}, 110\frac{\mathrm{km}}{\mathrm{h}}, 115\frac{\mathrm{km}}{\mathrm{h}}, 120\frac{\mathrm{km}}{\mathrm{h}} \end{array}\right\}, \tag{17}$$

$$\rho_i(k+1)=\rho_i(k)+\frac{T}{\lambda_i l_i}\left(q_{i-1}(k)-q_i(k)+r_i(k)-s_i(k)\right), \tag{18}$$

$$\begin{aligned} v_i(k+1)=v_i(k)&+\frac{T}{\tau}\left[V\left[\rho_i(k)\right]-v_i(k)\right]+\frac{T}{l_i}v_i(k)\left[\sqrt{\frac{v_{i-1}^2(k)+v_i^2(k)}{2}}-v_i(k)\right] \\ &-\frac{1}{\tau}\left[\frac{Tv}{L_i}\frac{\rho_{i+1}(k)-\rho_i(k)}{\rho_i(k)+\kappa}\right], \end{aligned} \tag{19}$$

$$V\left[\rho_i(k)\right]=\frac{u_i^{\mathrm{acc}}(k)}{\left(1+\left(\left(v_i^{\mathrm{acc}}(k-1)-u_i^{\mathrm{acc}}(k-1)\right)/u_i^{\mathrm{acc}}(k-1)\right)\right)}\exp\left[-\frac{1}{o_i}\left(\frac{\rho_i(k)}{\rho_c}\right)^{o_i}\right]. \tag{20}$$

Formula (13) takes into account the safety of the open freeway. The speed limit value should be less than the maximum speed limit of the freeway. Formula (14) considers the characteristics of high speed, high efficiency, and high traffic volume of freeway. The speed limit value should be higher than the minimum speed limit so that the freeway can realize its normal road transportation function. Formulas (15) and (16) are to consider the driver's driving comfort and driving safety and to prevent the driver from causing a traffic accident due to excessive acceleration and deceleration in the VSL area. The difference between the speed limits on the controlled road section during the two adjacent time intervals should not exceed the maximum value of driver's speed change. Formula (17) is to consider driver's maneuverability to driving speed, and the speed interval is set to an integer multiple of 5 km/h. Formulas (18) to (20) are dynamic constraints from enhanced METANET model.

2.3. Control Flow. The traditional method of formulating VSL control strategies are based on historical traffic flow data through technical experience. That is no longer applicable to the rapid development of information technology because of neglecting the time-varying nature of traffic data. As shown in Figure 2, this paper constructs an online control method and formulates control strategies based on the time-varying nature of traffic flow.

The principle of VSL system is as follows: the fundamental traffic flow data from monitors of freeway system is transferred to the internal controller after analysis and processing. Internal controller calculates traffic flow prediction by the enhanced METANET model from time interval k to the predicted time. Then the traffic flow data can be passed to the optimization model. The control objective function is solved to get the speed limit of target section. The optimal solution of the optimization problem is fed back to VSL control system. Due to real-time information interaction and acquisition of intelligent connected environment, the variable speed limit can be dynamically updated so as to continuously optimize and achieve global optimization.

Data collection: collect and update traffic flow data in real time using the detectors set in the freeway and intelligent connected vehicles.

METANET model: predict traffic flow data by the acquired data from freeway system.

Optimization model: achieve optimized control target according to different control objectives, such as reducing

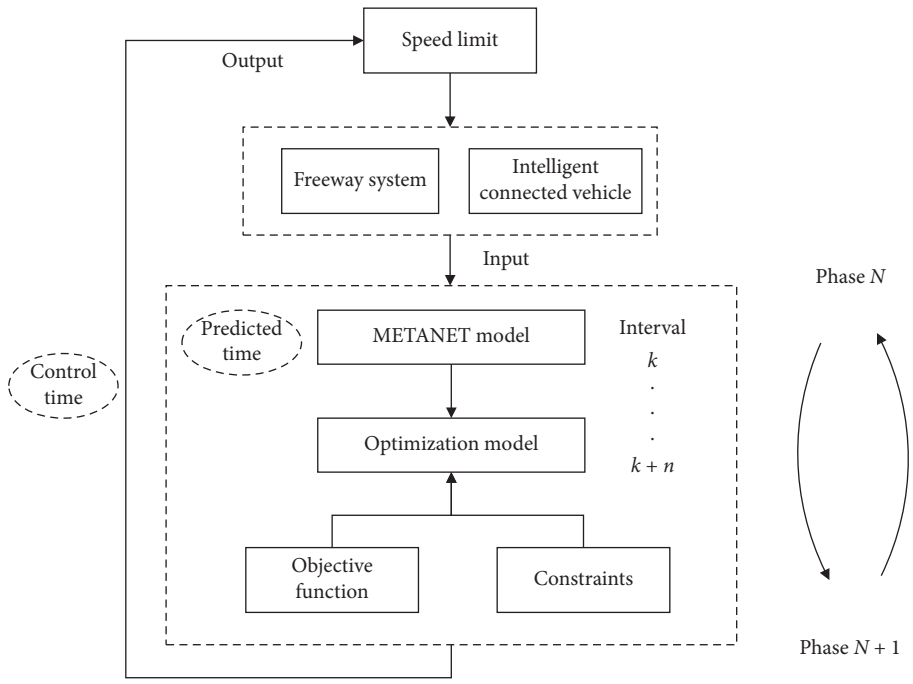


Figure 2: Control flow diagram.

travel time and increasing total traffic flow by the objective function and constraint equations.

Solving tool: solve the control objective function within the predicted time, get the speed limit value of each road segment, and feed back to the freeway to realize the control within the control time.

The control model contains two important parameters, prediction time length Np and control time length Nc, and the values are generally an integer multiple of discrete time step T. The combination of the prediction time length and the control time length will directly affect the control effect [34, 35]. The length of the prediction time should not be less than the travel time consumed by the vehicle in the control section. The excessively large prediction step will increase the calculation cost and is not conducive to simulation experiment. When the traffic flow is close to congestion, the traffic flow state is extremely unstable. The controller step should be updated in real time to ensure that measures can be taken to the traffic state changes. In this article, the value of Np is set to 5 min and Nc is set to 3 min.

3. Case Study

3.1. Simulation Scenario. Simulation of Urban Mobility (SUMO) is an open source, multimodal traffic simulation software that can modify vehicle state, operation mode, and other parameters to a greater extent based on the user's algorithm control requirements. This article selects SUMO as the basic simulation tool. The simulation scenario is as shown in Figure 3, and the road is divided into 7 sections. At the end of the 7th section, the road is closed due to factors such as construction occupation, and the second section to the 6th section are set to be in the variable speed limit control area. A traffic detection device is set at the front end of each section of road. The seventh section is the set variable speed limit dissipation area, and the length of the dissipation area is set to 500–700 m. The simulation parameter input is shown in Tables 4 and 5.

Intelligent connected technology makes vehicles no longer individual driving unit, but an information body that can interact with the driving environment in real time. More information can be obtained between the vehicle and the control center. The operation state of ICV can be adjusted in real time to make these vehicles safer and more reasonable. Therefore, traffic efficiency is effectively improved. The driving behavior of the vehicle is an important factor that affects the performance of the traffic flow. This is also the research basis of the traffic control strategy in the intelligent connected environment.

Treiber [36] proposed an intelligent driver model (IDM) in 2000. The IDM is widely applicable in intelligent connected environment. The ICV has better driving stability and average minimum time headway than manual vehicles. In free flow, the car accelerates to desired speed under the control of the model and keeps running at that speed. Under congested state, when the difference between the front and rear car speeds is not large, small changes in vehicle spacing will not cause the rear car to decelerate. Each parameter of the IDM represents a specific physical meaning. Different driving strategies can be implemented among vehicles by modifying the parameters. Therefore, this paper selects the IDM as the car-following model of intelligent connected vehicle.

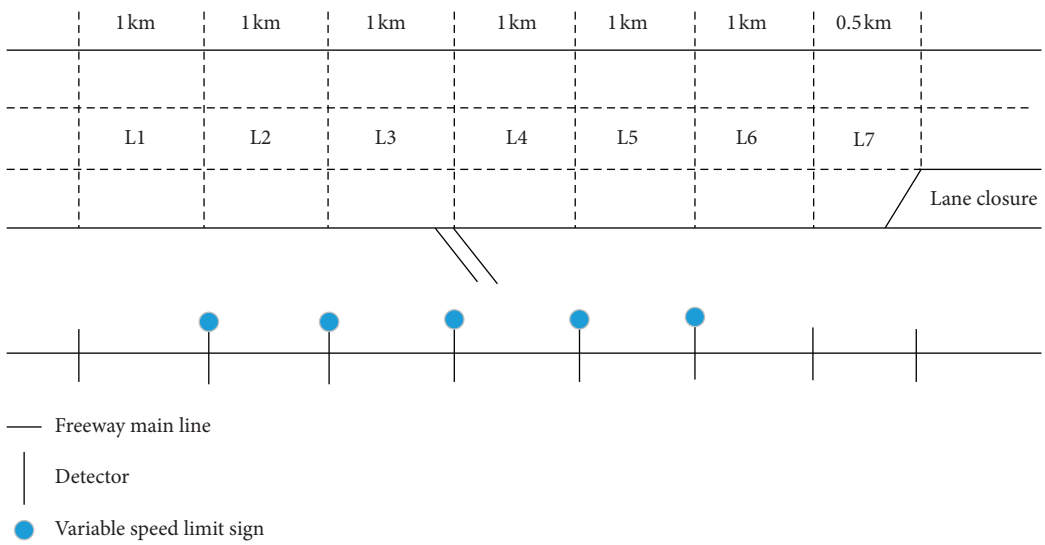


Figure 3: Simulation scenario.

Table 4: Simulation parameter input.

No.	Time interval (s)	Input flow (veh/h)
1	0–1000	3000
2	1000–5000	4500
3	5000–10000	2000

The IDM can be expressed as follows:

$$a = a_0\left[1 - \left(\frac{v_i}{v_0}\right)^{\delta} + \left(\frac{s_1}{s}\right)^2\right],$$
$$s_1 = s + Tv_i + \frac{v_i \Delta v_i}{2\sqrt{a_0 b}}, \tag{21}$$

where a is the acceleration of vehicle; v_i is the speed of vehicle; Δv_i is the speed difference between vehicle and the front vehicle; v_0 is the driver's desired speed in free flow; a_0 is the maximum acceleration of the vehicle; b is the braking deceleration of the vehicle; s is the minimum safety distance; T is the expected time headway; δ is the model parameter.

The driving parameters of the IDM of the intelligent connected vehicle in this paper are shown in Table 6 [37].

Table 5: Simulation scenarios.

No.	Description
Scenario 1	With no variable speed limit and 100% manual vehicle
Scenario 2	With no variable speed limit and 25% ICV
Scenario 3	With no improved variable speed limit and 25% ICV
Scenario 4	With improved variable speed limit and 25% ICV

Table 6: The parameter of IDM.

Parameter	a_0 (m·s^{-2})	b (m·s^{-2})	δ	s (m)	T (s)
ICV	1	2	4	0	1.1
Manual vehicle	1	2	4	2.5	1.6
Truck	1	2	4	2.5	1.6

3.2. Calibration of METANET Model. It can be seen from the METANET model that there are global parameters τ, ν, and κ. They are closely related to the characteristics of road segments, vehicles, and traffic flow. In order to narrow the difference between the METANET model and the actual road conditions, this paper uses the data obtained by the simulation of the SUMO to calibrate the improved METANET model parameters. To make the parameters more reasonable, the predicted speed and traffic volume of the model are compared with actual data on road sections. The optimization objective function is minimized to reduce the error.

The optimization function can be expressed as

$$\min f = \sum_{i=1}^{N}\sum_{k=1}^{K}\left(\left(\frac{v_{i,\text{actual}}(k) - v_{i,\text{predicted}}(k)}{v_{i,\text{actual}}(k) - v_{i,\text{predicted}}(k)/2}\right)^2 + \left(\frac{q_{i,\text{actual}}(k) - q_{i,\text{predicted}}(k)}{q_{i,\text{actual}}(k) - q_{i,\text{predicted}}(k)/2}\right)^2\right). \tag{22}$$

It is subject to

$$\begin{aligned} &X_{\min} \le X \le X_{\max}, \\ &X_{\min} = [0.05, 10, 10], \\ &X_{\max} = [0.2, 60, 60]. \end{aligned} \tag{23}$$

The simulation scenario is the condition with no road closed and the results are shown in Table 7.

4. Results

The simulation results are shown in Table 8. When the intelligent connected vehicle permeability is from 0 to 0.25, the simulation results show that average travel time, average delay, number of passing vehicles, average CO2 emissions, and average PMx emissions are all improved by 6.68%, 8.72%, 2.37%, 2.99%, and 2.39%, respectively. When intelligent connected vehicle permeability is 0.25 and the improved intelligent connected environment VSL control is used, average travel time, average delay, number of passing vehicles, average CO_2 emissions, and average PMx emissions are all improved by 27.43%, 46.13%, 10.88%, 16.55%, and 22.38%, respectively. The control center can obtain the driving information of the ICV at any time to correct the speed limit value. There is an improvement compared with the uncontrolled state. It can be seen that the VSL online control method can more effectively improve the driving state of the road environment and reduce exhaust emissions.

When in scenario 2 situation, the density changes of the fifth and sixth sections are shown in Figure 4. When the simulation time is 4000 s, the congestion spreads from the road closure to the sixth section of the road. At this time, the road section density is higher and the vehicles are in serious congestion state. When the simulation time is 6000 s, the traffic jam phenomenon spreads to the fifth section. When the simulation time is 9000 s, as the queuing gradually dissipates, the congestion state of the fifth road section can be reduced.

When in scenario 4, that is, with the improved VSL control in the intelligent connected environment, the density changes of the fifth and sixth segments are shown in Figures 5 and 6. It can be seen from Figures 4 and 5 that the improved VSL control can control the road density by controlling the driving speed of the vehicle on the road section. The density of the road section is significantly reduced. The average driving speed of the vehicle is greatly improved, which greatly eases the phenomenon of traffic congestion and improves the traffic efficiency and driving safety of vehicles.

When in scenario 4, the flow comparison curve of the sixth section is shown in Figure 7. Before the simulation time 4000 s, there is no congestion phenomenon. The traffic has no change even with the VSL control. When the simulation time is over 4000 s, it can be seen that the improved VSL control system can properly increase the number of vehicles passing and improve the throughput of congested roads.

From the above analysis, it can be seen that the use of improved VSL control in an intelligent connected environment can effectively alleviate traffic congestion in bottleneck areas. It also improves vehicle traffic efficiency and driving stability and effectively reduces CO_2 and PMx emissions, which is beneficial to environmental protection and green transportation.

In the application process of VSL control system, the improvement effect is important, and the operating efficiency of the system also can not be ignored. Long calculating time will lead to a significant reduction on the effectiveness of the system in practical applications. In this system, the calculation part is mainly integrated in the genetic algorithm solution process. The time consumed by the system is mainly from the time occupied by the genetic algorithm solution. In this paper, solution process iterations of the VSL control system are commonly between 60 and 90 times. The convergence speed is relatively fast. It can be seen from Figure 8 that the average running time is around 2 s, which confirms the high efficiency of improved VSL control system.

4.1. Sensitivity Analysis. The relationship between average travel time and the penetration rate of intelligent connected vehicles is shown in Figure 9. As can be seen from Figure 9, the efficiency of road sections is improved and average travel time continues to decrease with continuous increase of the penetration rate of ICV. When penetration rate of ICV is 25%, 50%, and 75%, the travel time will be reduced by 6.7%, 14.1%, and 21.2%, respectively, compared with no ICV. Furthermore, the improved VSL control system can have a better improvement effect on average travel time under the conditions of the above penetration rate of ICV. Compared with no VSL control, the proportion of improvement is 27.4%, 26.2%, and 26.7%, respectively.

The relationship between average delay and the penetration rate of intelligent connected vehicles is shown in Figure 10. It can be seen that as the penetration rate of ICV continues to increase, the average delay continues to decrease. When penetration rate of ICV is 25%, 50%, and 75%, the average delay is reduced by 8.7%, 18.3%, and 27.3% compared with no ICV. In addition, the improved VSL control system can significantly reduce average delay under the conditions of the above penetration rate of ICV. Compared with no VSL control, the proportion of improvement is 46.1%, 50.6%, and 21.2%, respectively.

The relationship between the number of vehicles and the penetration rate of intelligent connected vehicles is shown in Figure 11. The number of vehicles actually passing through a road segment is one of the most important parameters that

Table 7: Parameter calibration results.

ICV permeability (%)	τ	ν	κ	Free flow speed/(km/h)	Critical density/(veh/(km·lane))
0	0.05	13.9196	53.9389	110	13
25	0.05	13.9196	53.9389	110	15
50	0.05	13.9196	53.9389	110	17
75	0.05	13.9196	53.9389	110	19
100	0.05	13.9196	53.9389	110	22

Table 8: Simulation result.

No.	Average travel time (s)	Average delay (s)	Passing vehicles (veh)	Average CO_2 emissions (mg)	Average PMx emissions (mg)
Scenario 1	812.92	621.49	5789	2067.95	44.86
Scenario 2	758.65	567.27	5926	2006.04	43.79
Scenario 3	574.04	350.79	6550	1760.20	36.41
Scenario 4	550.53	305.57	6571	1674.00	33.99

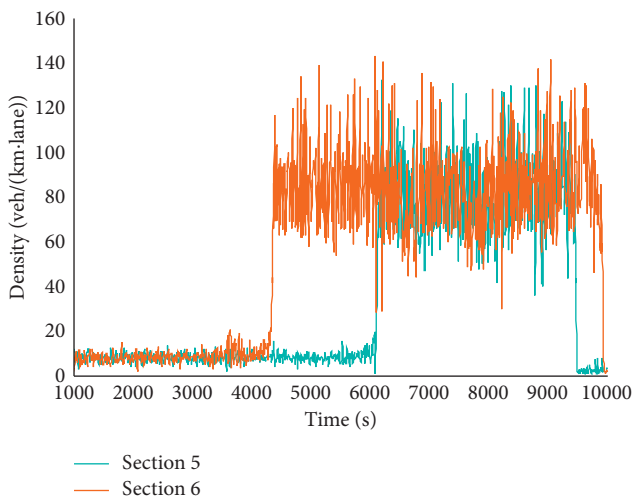


Figure 4: Scenario 2 simulation result diagram.

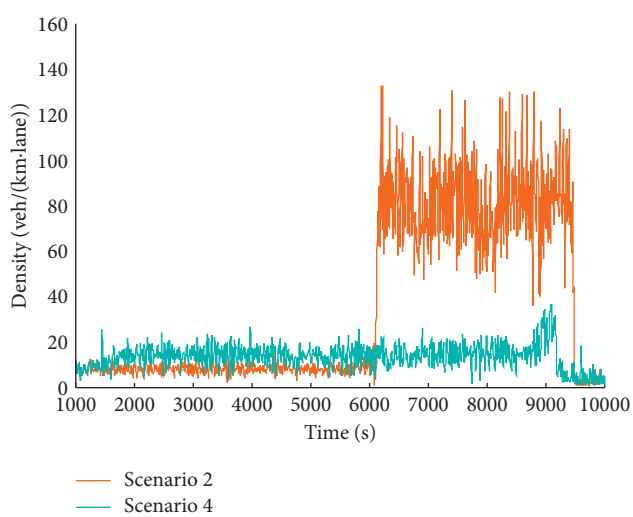


Figure 5: The fifth density comparison diagram between scenario 2 and scenario 4.

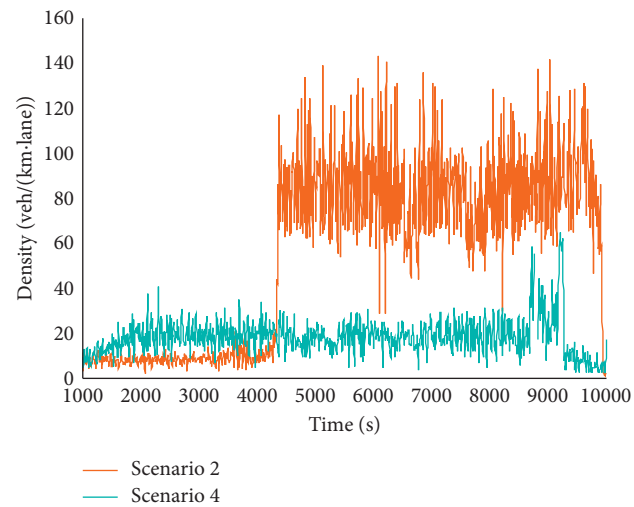


Figure 6: The sixth density comparison diagram between scenario 2 and scenario 4.

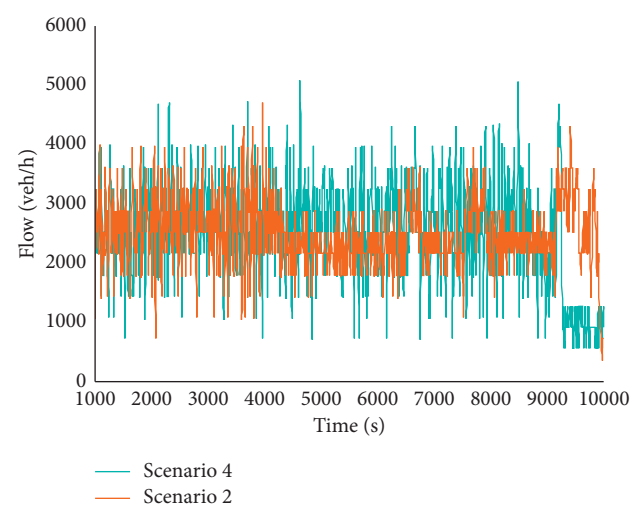


Figure 7: The sixth traffic flow comparison diagram between scenario 2 and scenario 4.

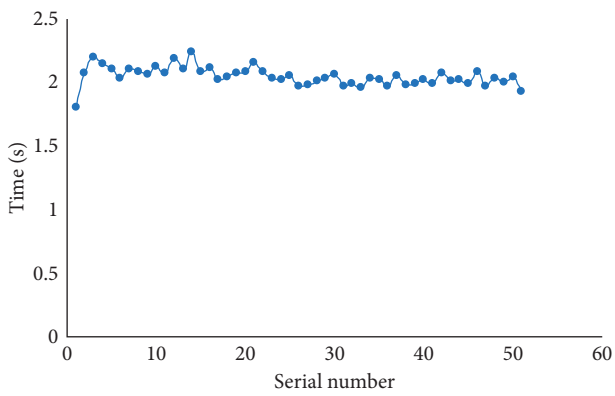


Figure 8: The calculating time.

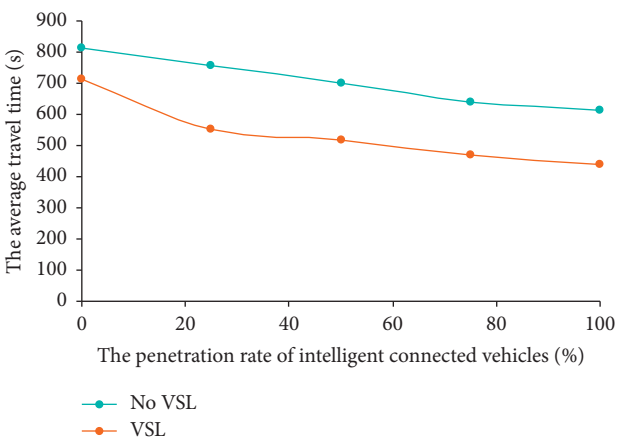


Figure 9: The average travel time result.

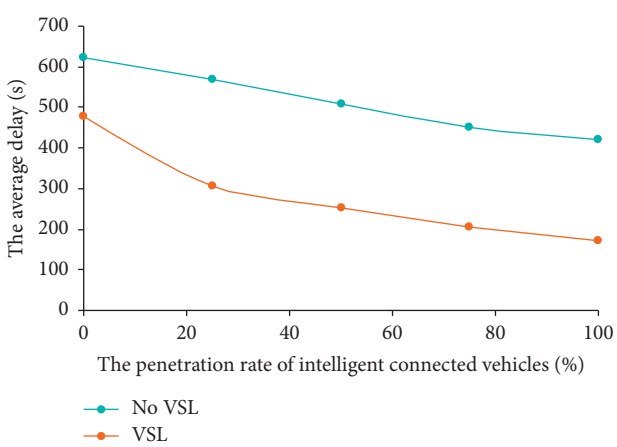


Figure 10: The average delay result.

can directly reflect the actual road capacity. It can be seen from Figure 11 that, with the increasing penetration rate of ICV, the number of actually passing vehicles continues to increase. For example, with ICV permeability by 25%, 50%, and 75%, the improvement efficiency is 2.7%, 14.7%, and 24.3% compared with no ICV. In addition, the improved

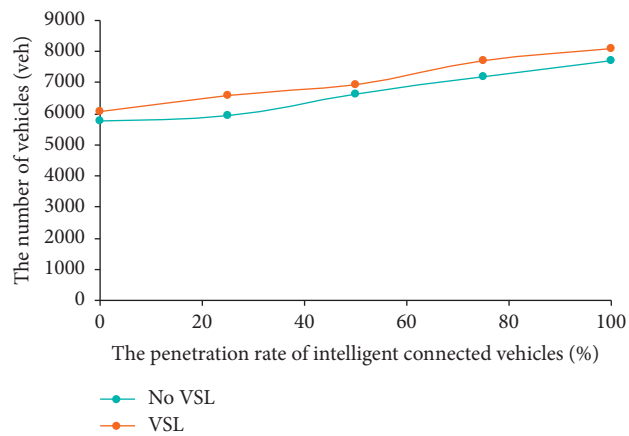


Figure 11: The number of vehicles' result.

VSL control system can increase the number of actually passing vehicles in the bottleneck area. Compared with no VSL control, the ratio of improvement is 10.9%, 4.6%, and 6.9%, respectively.

5. Conclusions

With the innovation and development of intelligent connected technology, ICV will be widely used in real life in the near future. In this paper, considering the characteristics of intelligent connected technology, the VSL online control method combined intelligent connected technology and VSL theory to optimize the VSL control system. The improved VSL control system in intelligent connected environment has a better improvement effect under different ICV penetration rates. The increase of ICV permeability can further alleviate the traffic congestion, which can effectively improve the traffic efficiency and ease the congestion in the bottleneck area.

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 there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was sponsored by the key project supported by the Scientific Research Foundation, Education Department of Hebei Province of China (ZD2021336). This work was further sponsored by the Social Science Development Research Project of Hebei Province of China (No. 20200302079).

References

[1] B. Alonso, Á. I. Pòrtilla, G. Musolino, C. Rindone, and A. Vitetta, "Network Fundamental Diagram (NFD) and traffic

signal control: first empirical evidences from the city of Santander," *Transportation Research Procedia*, vol. 27, pp. 27–34, 2017.

[2] M. Sadat and H. B. Celikoglu, "Simulation-based variable speed limit systems modelling: an overview and a case study on Istanbul freeways," *Transportation Research Procedia*, vol. 22, pp. 607–614, 2017.

[3] E. F. Grumert and A. Tapani, "Characteristics of variable speed limit systems," *European Transport Research Review*, vol. 10, no. 2, pp. 1–12, 2018.

[4] J. Cao, D. Hu, Y. Luo, T. Z. Qiu, and Z. Ma, "Exploring the impact of a coordinated variable speed limit control on congestion distribution in freeway," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 2, no. 3, pp. 167–178, 2015.

[5] D. C. Festa, G. Longo, G. Mazzulla et al., "Experimental analysis of different simulation models for motorway traffic flow," in *Proceedings of 2001 IEEE Intelligent Transportation Systems*, Oakland, CA, USA, August 2001.

[6] X. Yang, Y. C. Lu, G.-L. Chang et al., "Exploratory analysis of an optimal variable speed control system for a recurrently congested freeway bottleneck," *Journal of Advanced Transportation*, vol. 49, no. 2, pp. 195–209, 2015.

[7] A. Hegyi, B. DeSchutter, and J. Hellendoorn, "Optimal coordination of variable speed limits to suppress shock waves," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 102–112, 2005.

[8] D. Li and P. Ranjitkar, "A fuzzy logic-based variable speed limit controller," *Journal of Advanced Transportation*, vol. 49, no. 8, pp. 913–927, 2015.

[9] D. Li, P. Ranjitkar, and Y. Zhao, "Mitigating recurrent congestion via particle swarm optimization variable speed limit controllers," *KSCE Journal of Civil Engineering*, vol. 23, no. 7, pp. 3174–3179, 2019.

[10] X. Yang, Y. Lin, Y. Lu, and N. Zou, "Optimal variable speed limit control for real-time freeway congestions," *Procedia-Social and Behavioral Sciences*, vol. 96, pp. 2362–2372, 2013.

[11] A. Kotsialos, M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham, "Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool METANET," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 282–292, 2002.

[12] Y. Bie, M. Seraj, C. Zhang et al., "Improving traffic state prediction model for variable speed limit control by introducing stochastic supply and demand," *Journal of Advanced Transportation*, vol. 2018, Article ID 7959815, 12 pages, 2018.

[13] X. Wang, M. Seraj, Y. Bie et al., "Implementation of variable speed limits: preliminary test on whitemud drive, edmonton, Canada," *Journal of Transportation Engineering*, vol. 142, no. 12, Article ID 05016007, 2016.

[14] M. Yu and W. D. Fan, "Optimal variable speed limit control in connected autonomous vehicle environment for relieving freeway congestion," *Journal of Transportation Engineering, Part A: Systems*, vol. 145, no. 4, Article ID 04019007, 2019.

[15] M. Hadiuzzaman and T. Z. Qiu, "Cell transmission model based variable speed limit control for freeways," *Canadian Journal of Civil Engineering*, vol. 40, no. 1, pp. 46–56, 2013.

[16] Y. S. Jiang, B. Zhao, M. Liu et al., "A two-level model for traffic signal timing and trajectories planning of multiple CAVs in a random environment," *Journal of Advanced Transportation*, vol. 2021, Article ID 9945398, 13 pages, 2021.

[17] Z. H. Yao, T. Xu, Y. S. Jiang et al., "Linear stability analysis of heterogeneous traffic flow considering degradations of connected automated vehicles and reaction time," *Physica A: Statistical Mechanics and Its Applications*, vol. 561, Article ID 125218, 2021.

[18] B. Alonso, Á. Ibeas, G. Musolino, C. Rindone, and A. Vitetta, "Effects of traffic control regulation on Network Macroscopic Fundamental Diagram: a statistical analysis of real data," *Transportation Research Part A: Policy and Practice*, vol. 126, pp. 136–151, 2019.

[19] M. Tashiro, H. Motoyama, Y. Ichioka, T. Miwa, and T. Morikawa, "Simulation analysis on optimal merging control of connected vehicles for minimizing travel time," *International Journal of Intelligent Transportation Systems Research*, vol. 18, no. 1, pp. 65–76, 2020.

[20] M. Wang, W. Daamen, S. P. Hoogendoorn, and B. van Arem, "Connected variable speed limits control and car-following control with vehicle-infrastructure communication to resolve stop-and-go waves," *Journal of Intelligent Transportation Systems*, vol. 20, no. 6, pp. 559–572, 2016.

[21] L. Pu, X. Xu, H. He, H. Zhou, Z. Qiu, and Y. Hu, "A flexible control study of variable speed limit in connected vehicle systems," *International Journal of Embedded Systems*, vol. 7, no. 2, pp. 180–188, 2015.

[22] Z. H. Yao, H. R. Jiang, and Y. Cheng, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, 2020.

[23] H. Su, "Research about the construction of intelligent transportation system under the big data of vehicle networking," *IOP Conference Series: Materials Science and Engineering*, vol. 688, no. 2, Article ID 022011, 2019.

[24] Y. Zou and X. Qu, "On the impact of connected automated vehicles in freeway work zones: a cooperative cellular automaton model-based approach," *Journal of Intelligent and Connected Vehicles*, vol. 1, no. 6, 2018.

[25] H. T. Zhao, L. Lin, C. P. Xu et al., "Cellular automata model under Kerner's framework of three-phase traffic theory considering the effect of forward–backward vehicles in internet of vehicles," *Physica A: Statistical Mechanics and its Applications*, vol. 553, Article ID 124213, 2020.

[26] E. F. Grumert and A. Tapani, "Using connected vehicles in a variable speed limit system," *Transportation Research Procedia*, vol. 27, pp. 85–92, 2017.

[27] C. Guo, D. Li, G. Zhang, and M. Zhai, "Real-time path planning in urban area via vanet-assisted traffic information sharing," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 7, pp. 5635–5649, 2018.

[28] M. L. Song, R. R. Li, Y. N. Wei et al., "Dynamic vehicle route guidance model under internet of vehicles," *Advances in Transportation Studies*, vol. 1, pp. 39–48, 2019.

[29] İ. G. Erdağı, M. A. Silgu, and H. B. Çelikoğlu, "Emission effects of cooperative adaptive cruise control: a simulation case using SUMO," *EPiC Series in Computing*, vol. 62, pp. 92–100, 2019.

[30] Z. Yao, R. Hu, Y. Wang et al., "Stability analysis and the fundamental diagram for mixed connected automated and human-driven vehicles," *Physica A: Statistical Mechanics and Its Applications*, vol. 533, 2019.

[31] X. Y. Lu, T. Z. Qiu, R. Horowitz et al., "METANET model improvement for traffic control," in *Proceedings of 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 2148–2153, Washington, DC, USA, October 2011.

[32] M. Yu and W. Fan, "Optimal variable speed limit control at a lane drop bottleneck: genetic algorithm approach," *Journal of Computing in Civil Engineering*, vol. 32, no. 6, Article ID 04018049, 2018.

[33] B. Khondaker and L. Kattan, "Variable speed limit: a microscopic analysis in a connected vehicle environment," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 146–159, 2015.

[34] D. Wang, Y. Zhou, and X. He, "Radial basis function neural network-based model predictive control for freeway traffic systems," *International Journal of Intelligent Systems Technologies and Applications*, vol. 2, no. 4, pp. 370–388, 2007.

[35] C. Portilla, F. Valencia, J. Espinosa, A. Núñez, and B. De Schutter, "Non-linear model predictive control based on game theory for traffic control on highways," *IFAC Proceedings Volumes*, vol. 45, no. 17, pp. 436–441, 2012.

[36] 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.

[37] M. J. Cassidy and J. Rudjanakanoknad, "Increasing the capacity of an isolated merge by metering its on-ramp," *Transportation Research Part B: Methodological*, vol. 39, no. 10, pp. 896–913, 2005.

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 8872889, 12 pages
https://doi.org/10.1155/2021/8872889

Research Article

An Indoor Vehicle-in-the-Loop Simulation Platform Testing Method for Autonomous Emergency Braking

Ying Gao, Xiangmo Zhao, Zhigang Xu, Jingjun Cheng, and Wenwei Wang

School of Information and Engineering, Chang'an University, Xi'an, Shaanxi, China

Correspondence should be addressed to Xiangmo Zhao; xmzhao@chd.edu.cn and Zhigang Xu; xuzhigang@chd.edu.cn

Received 17 July 2020; Revised 6 February 2021; Accepted 12 April 2021; Published 20 April 2021

Academic Editor: Vittorio Astarita

Autonomous vehicle (AV) is expected to be the ultimate solution for traffic safety, while autonomous emergency braking (AEB), as a crucial and fundamental active safety function of AV, has excellent potential for reducing fatalities and improving road safety. Although AV has the ability to cope with harsh conditions, it is supposed to be tested fully, systematically, and rigorously before it is officially commercialized. This study developed a novel indoor AV-in-the-loop (AVIL) simulation platform based on Client-Server (C/S) architecture for real full-scale AV testing. The proposed AVIL simulation platform consists of three parts: physical hardware components, software components, and various electrical interfaces that ensure the bidirectional virtual reality (VR) interaction. To validate the functionality and performance of the platform, this paper then adopted the Udwadia–Kalaba (U-K) approach to build the AEB system based on a typical driving situation due to the explicitness and simplicity of the U-K approach. Further, a group of real road-based experiments and AVIL-based experiments were conducted. The experimental results showed that the testing data obtained from the proposed AVIL platform have a high similarity to those of the real road tests, which means that the proposed AVIL platform is capable of simulating the AV running condition when it performs linear emergency braking on the road, thus confirming the feasibility and effectiveness of the AVIL platform for AV AEB testing. Simultaneously, the testing time and repeatability of the latter performed better. The findings of this study provide a new safe, effective, and fast solution to AV testing, and the practicability of this method has been verified.

1. Introduction

According to the World Health Organization [1, 2], more than one million people died in traffic accidents each year. This figure is 1.2 million in 2015 [3], 1.24 million in 2013 [4], and 1.3 million in 2009 [5]. From these figures, road traffic injuries are still one of the top ten causes of death worldwide. Besides, road traffic accidents also cause a great deal of cost, which is estimated to account for approximately 3% of the entire world's GDP [3]. In China, 63000 people are killed, and 226000 are injured in traffic accidents every year [6].

Autonomous vehicles (AVs) are designed to free drivers [7] from driving tasks and are expected to improve traffic safety [8] and efficiency [9]. They also have great potential in dealing with traffic problems [10–12]. AVs with great adaptability that use advanced sensing, communication, and control technologies have the potential to increase road capacity and improve traffic operations [7, 13–15].

The fully automated vehicle would eventually eliminate driver errors, and ultimately, 90% of traffic accidents could be avoided [16]. Therefore, the development of AV has become an ideal way to reduce road traffic accidents and improve traffic safety [17, 18]. Although AVs have many advantages, several recent accidents have slowed down their process of commercialization. Tesla canceled the propaganda of self-driving feature, and Uber also terminated all AV-related tests. In fact, both the government and academia have reached an agreement that the AV needs to be fully, systematically, and rigorously tested before it is officially commercialized.

The AV testing methods are mainly classified into four categories: real-world-based, simulation-based, hardware-in-the-loop- (HIL-) based, and vehicle-in-the-loop- (VIL-) based. Actual road tests are more accurate, but they are time-consuming, costly, and susceptible to weather, illumination conditions, and changing traffic scenarios. In contrast, the

simulation test is always based on specific software that has the advantages of time-saving, cost-effective, and convenience. However, it is generally based on the ideal mathematical model and lacks several realistic considerations, which might lead to a big gap between the simulation results and the actual results.

HIL has been adopted in automotive industries [19] since the early 1990s [20]. Over the past 30 years, many researchers have proposed various HIL testing platforms. Deng et al. developed a HIL simulation platform, which is an integrated system of various autonomous driving programmers as an indoor environment to support the development, testing, and verification of multiple functions and algorithms related to sensor-guided AVs [21]. Galko et al. brought a real car into the HIL loop, building a vehicle-hardware-in-the-loop system to evaluate the new algorithm and new hardware for driver assistance systems directly on a complete vehicle [22]. Zulkefli et al. proposed a powertrain research platform under the laboratory setup with a microscopic traffic simulator (VISSIM), which could be used to test various intelligent transportation systems applications such as cooperative adaptive cruise control, eco-driving, and speed harmonization [23]. Chen et al. developed a simulation-based HIL system that can flexibly construct vehicle kinematic models, virtual sensors, and multiple simulation scenarios and developed hardware control interfaces to perform an effective closed-loop simulation in conjunction with software interfaces [24]. Xu et al. proposed a HIL-scaled platform for testing AV trajectory tracking with a mechatronic testbed, consisting of the scaled roadway, scaled vehicles, indoor positioning subsystem, and computer-aided graphical user interfaces. It is intended to verify the accuracy of the algorithms that would be transferred to the actual AV [25].

These HIL platforms more or less have the following one or several shortcomings: (1) only a virtual car or a part of the car's actuators is introduced to the test loop; (2) a scaled AV is introduced; (3) although significant results have been achieved, the effectiveness of the proposed and validated function on the real full-scale AV still lacks verification; (4) no bridging is found between software-defined AV and real AV.

As the name of VIL indicates, the method of directly testing a real vehicle can be considered as VIL test. Therefore, all actual road tests can be considered as VIL tests. To the best of our knowledge, there is a driving simulator-in-the-loop testbed under the laboratory environment [26], but there is almost no in-loop testing platform for real AVs. To this end, a real AV-in-the-loop (AVIL) simulation platform is proposed, which consists of a real AV, a set of mechanical equipment, sensor-related black boxes, other types of hardware, and software components. The presented AVIL simulation platform has the following advantages. First, a real AV is introduced into the testing loop to connect the software-based vehicle dynamics model supported by software with the real AV. The algorithm modules and parameters verified by software can be directly reflected on the real AV. Second, the real AV runs on the bench, which can be set with external factors such as road friction coefficient, road geometry models, and road gradient to simulate various real road conditions for the AV. Third, by deploying sensor-related black boxes, we have made it possible for real sensors to interact with virtual scenes, thereby maximizing the restoration of the real road testing process to ensure the homogeneity of the AVIL testing process and the real road testing process. The core contribution of this study is focused on the AVIL platform we established. With some customized components, it becomes possible to use this platform for full-scale AV testing in the laboratory environment, and the AVIL test results are of great similarity to those of real road test.

In this paper, the system structure of this AVIL platform has been described. Further, the autonomous emergency braking (AEB) control algorithm has been developed based on Udwadia–Kalaba (U-K) approach and tested on this platform. AVIL feasibility analysis is then performed. The remainder of this paper is organized as follows: A detailed introduction of the proposed AVIL simulation platform is illustrated in Section 2. Section 3 presents the AEB control algorithm and AVIL feasibility analysis and then is followed by discussion in Section 4 and conclusion with some future work in Section 5.

2. Overview of the Proposed AVIL Platform

The proposed AVIL simulation platform (see Figure 1) is an indoor closed-loop system based on Client-Server (C/S) architecture, which is inspired from [27]. Our proposed AVIL simulation platform consists of physical hardware components, software components, and various electrical interfaces that ensure real and virtual bidirectional interaction. Due to the benefits of C/S mode, (1) C/S mode separates clients and servers, and the system is stable and flexible. (2) The C/S mode is equipped with a local area network (LAN) appropriate point-to-point structure mode and has reliable security. Under laboratory conditions, the AVIL simulation platform we proposed is being carried out. Therefore, we chose the C/S mode to realize data transmission by way of LAN. In this way, every single component in this system can easily communicate with others through their specific protocols independently, even if the entire system is not fully started. Therefore, any part of the system can be individually debugged and tested. It is also convenient for the maintenance, replacement, and expansion of the submodules of the AVIL platform system.

2.1. Hardware in AVIL. The most important hardware equipment is that we have customized a three degrees-of-freedom test bench consisting of four independent components (shown in Figure 2) and can interact with the virtual scene through its specific communication protocol. The test bench in the AVIL platform can efficiently use the simulation environment to avoid the potential risks of real road tests. The rotating tires are the tires of the AV under test. The drum roller has two working modes, active and passive. Passive mode means that it rotates with the longitudinal rotation of the AV's tires, while active mode means that it

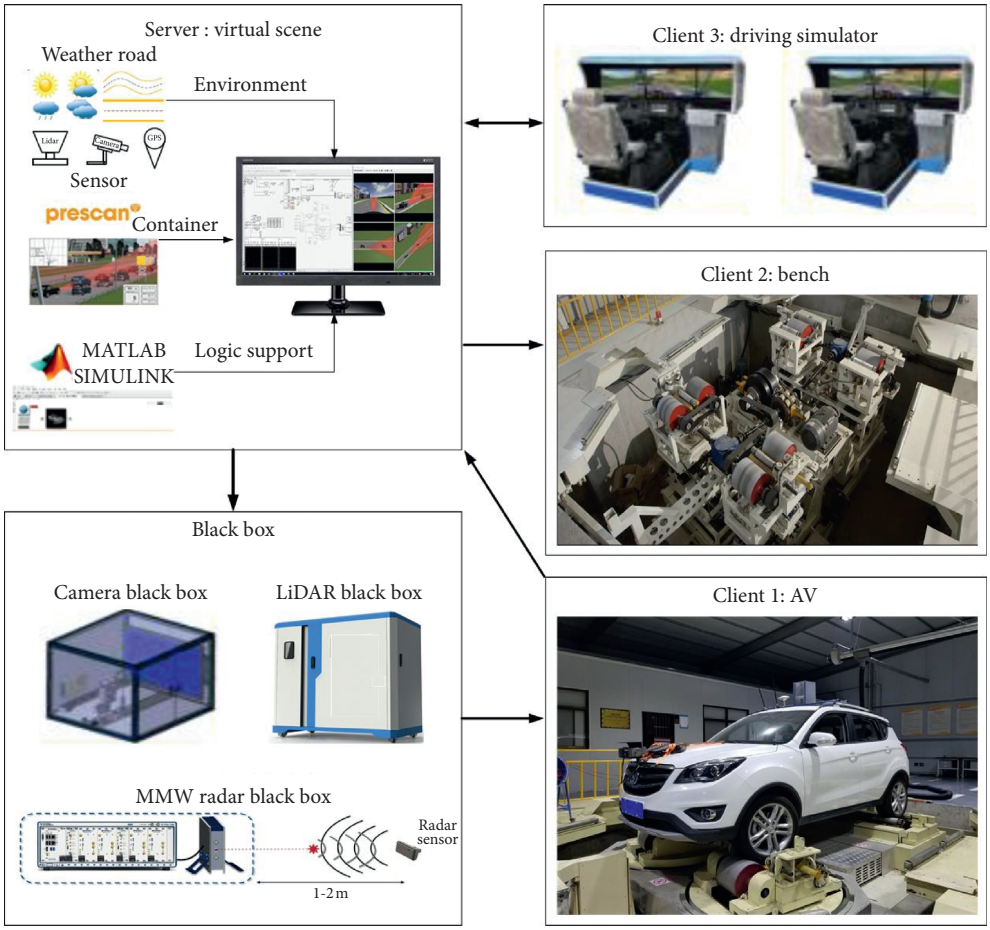


Figure 1: The structure of the AVIL simulation platform. Arrows indicate the direction of data transmission.

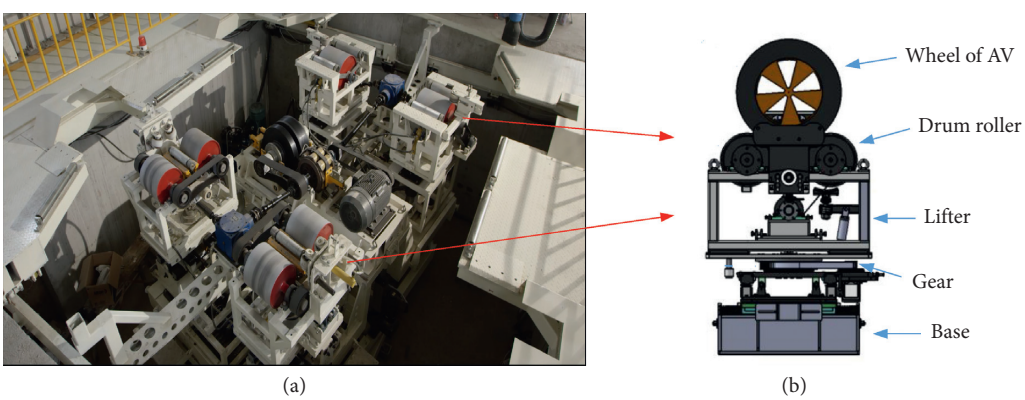


Figure 2: The structure of the test bench. (a) Top view. (b) Subcomponent.

can rotate spontaneously after power on. Through the combination of these two modes, the continuous road surface is simulated, as well as the ground adhesion coefficient. The role of the lifter is to simulate the geometric characteristics of the ground, to simulate the slope road surface through the height difference between the front and rear lifters, and to add a degree of freedom to AV under test. Regarding the gear, it adds another degree of freedom to AV under test. It only has the passive working mode; that is, it can only rotate with the lateral rotation of the AV's tires, providing the possibility for steering, as well as the safety guarantee for steering operation. Specifications of our test bench are given in Table 1.

We have deployed some driving simulators [28], as shown in Figure 1, which are also connected to our testbed through C/S mode as clients. The driving simulator makes the background vehicle in the virtual scenarios behave more realistically without following a predetermined route. In this way, while reducing the programming workload, uncertain factors such as driving intention and driving behavior can be

Table 1: Specifications of the test bench.

Parameter	Value
Wheelbase	2550 mm–2720 mm
Track width	1225 mm–1825 mm
Drum configuration	4-wheel independent drive
Drum diameter	318 mm in front; 452 mm in rear
Total peak power	80 KW
Traction force	3500 N in front
Response time	≤ 200 ms
Maximum velocity	60 km/h
Lifter range	0.6 m
Gear angle	−35°–35°
Slope angle	−15°–15°
Rated torque	600 Nm

quickly and easily introduced. The effectiveness of driving simulators has been verified via experiments, which does not show up in this research. Although it is interesting and worthy of discussion, it is beyond the scope of this research and will be discussed and applied in our follow-up research.

The other hardware devices in the simulation loop, such as sensors, actuators, processors, and communication buses, can significantly reduce the mathematical modeling effort and improve the simulation fidelity. The development efficiency can be dramatically increased in the early stages of designing the production-intent or production-ready hardware and software.

A real AV, test bench, radar simulator, video black box, several actuators, and sensors are connected to one or two of the USBCAN buses, net cable, or RS232. An actual AV is brought in to conduct the simulation and evaluation with greater fidelity under the lab settings, from which we can test whether the processors, sensors, actuators, and bus protocols normally work according to the design specifications of various aspects of hardware and software. This is particularly beneficial for testing, debugging, and verifying such a complex AV system in the laboratory environment.

2.2. Software in AVIL. The AVIL simulation consists of both mathematical models and physical hardware devices that interact in real time through specific interfaces and perform joint hardware testing and software simulation tasks. Under the AVIL simulation, math models receive various hardware signals, and the hardware devices receive calculated variables from the math models simultaneously. For software model-driven hardware systems, by ensuring computational efficiency and model fidelity, mathematical models with real-time capabilities are critical to the success of the AVIL simulation.

Our AVIL simulation platform is developed based on C/S mode. Various submodules are connected through LAN. That is, they can work together to complete joint tasks as well as run independently. Therefore, our virtual scene module can be supported by different simulation software [29], such as Prescan, Unity3D, Carla, and CarMake. Prescan is an automatic driving simulation platform for intelligent vehicle system based on physical models, which has a good built-in vehicle dynamics model and supports the development and application of camera, radar, LiDAR, GPS, V2X communication, and other applications, as well as model-in-the-loop, software-in-the-loop, HIL, and sensor-in-the-loop. Prescan fits our research needs very well. Therefore, in this research, Prescan is selected for the virtual scenario construction as shown in Figure 3, as well as the sensor-in-the-loop model. Unity3D is actually a game development software driven by unreal engine; the virtual scene built with Unity3D is more realistic than that of Prescan. Therefore, we also tried to build some virtual scenes with Unity3D, as shown in Figure 4.

No matter which software our virtual scene is built with, the way they connect to the developed AVIL platform is the same. The virtual scene always serves as the server, and the AV under test and driving simulators play the role of clients connected to the virtual scene through LAN to realize data interaction. The AV performs different driving strategies according to the driving environment or driving tasks in the virtual scene. The driving state of the AV and driving simulators will drive the vehicles in the virtual scene concurrently.

It can be seen that the Unity3D-built virtual scene is more realistic than Prescan, the model is more refined, and it has a variety of customizable features. Most notably, Prescan does not support virtual reality (VR), while Unity3D does, making it possible to perform an immersive driving experience with VR helmets. Our AVIL platform is very inclusive. According to different testing requirements, we could pick up the most appropriate simulation software as the server.

2.3. Sensor in AVIL. Sensor data simulation and injection realize the virtual sensor's sensing function by using sensor-in-the-loop (SIL) test method, such as camera simulator, millimeter-wave (MMW) radar, and LiDAR simulator. In our proposed AVIL simulation platform, sensors-related black boxes are customized. They are not one of-the-shelves components, nor just a concept. They are one of the independent parts of our AVIL simulation platform. Figure 5 shows a flowchart of the SIL simulation. In AVIL test, the black boxes are next to the cabinet on the right side of Figure 6.

In the actual road test process, the braking is triggered by the developed AEB system relying on radar in this research. Therefore, during the AVIL test, we still use the real radar sensor to activate the developed AEB system to ensure the homogeneity of the AVIL test process and the real road test process. Taking the MMW radar black box as an example (workflow shown in Figure 7), the virtual scene information is converted into a real millimeter wave (*SignalA*) through a certain algorithm and hardware and then transmitted through a transmitter. The real radar is detached from the AV and placed in the black box, and a special material covers the black box. This material can absorb millimeter wave without generating echoes, meaning that real radar cannot receive the echo wave emitted by itself when it hits this material. Therefore, in the black box, only *SignalA* exists and can be received by the real radar sensor, thus realizing the interaction between the real radar and virtual scene and eliminating the influence of the black box on the real radar at the same time.

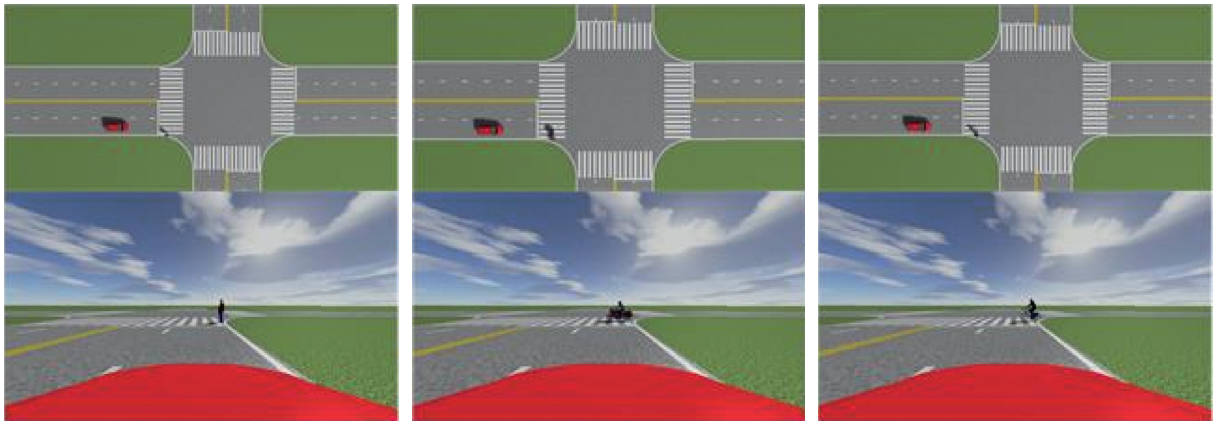


Figure 3: Virtual scenarios built with Prescan used in this study.

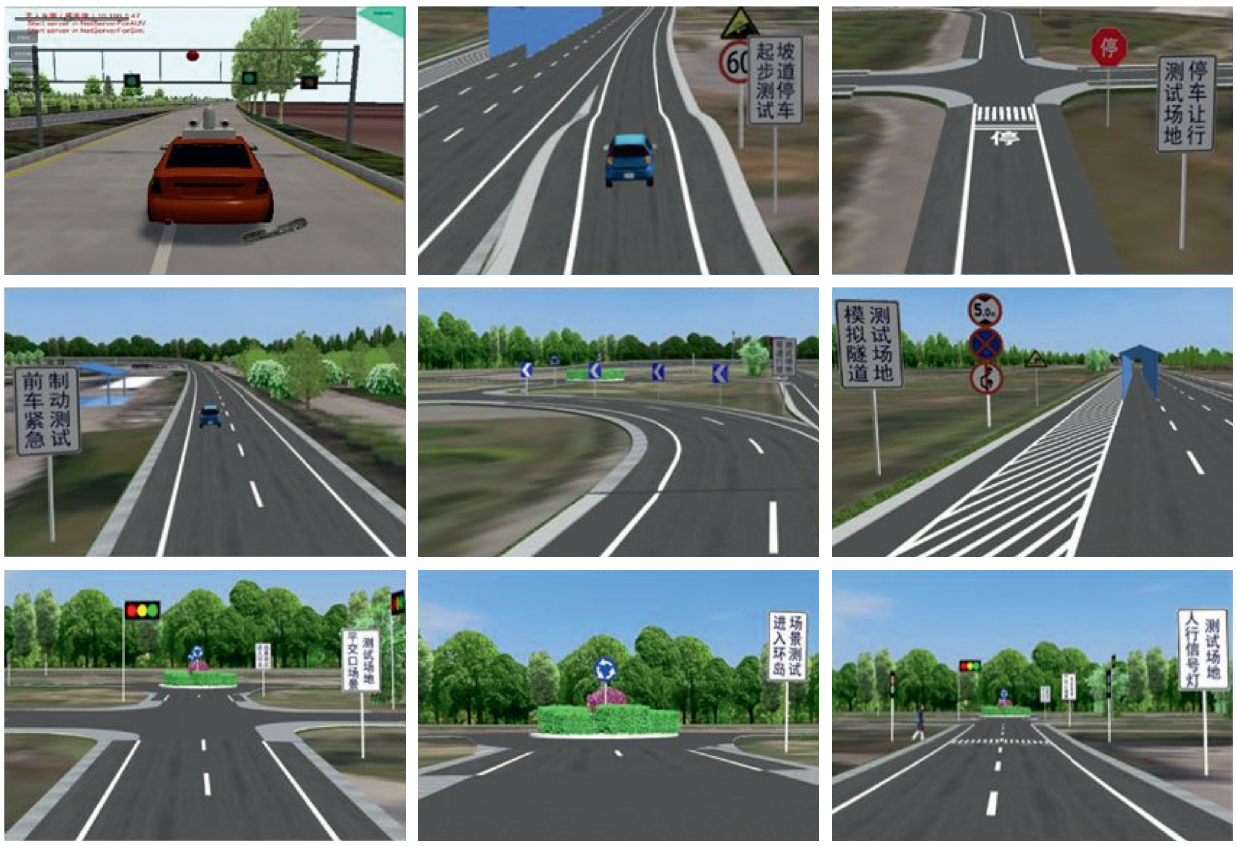


Figure 4: Virtual scenarios constructed with Unity3D.

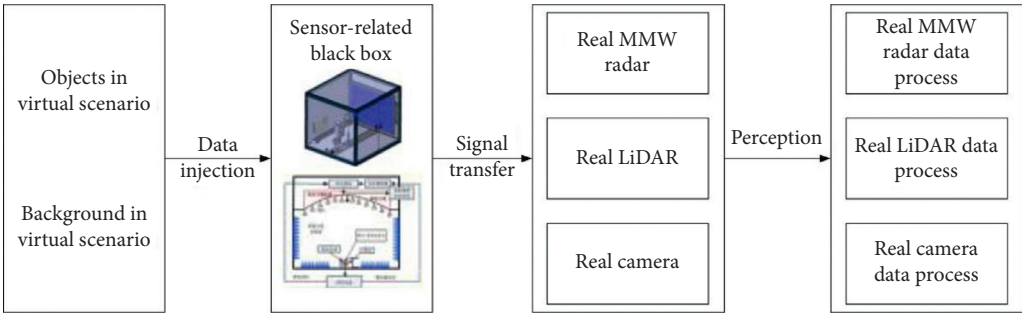


Figure 5: Flowchart of SIL simulation.

(a) (b)

Figure 6: A snapshot of AVIL test. (a) Inner viewpoint. (b) External viewpoint.

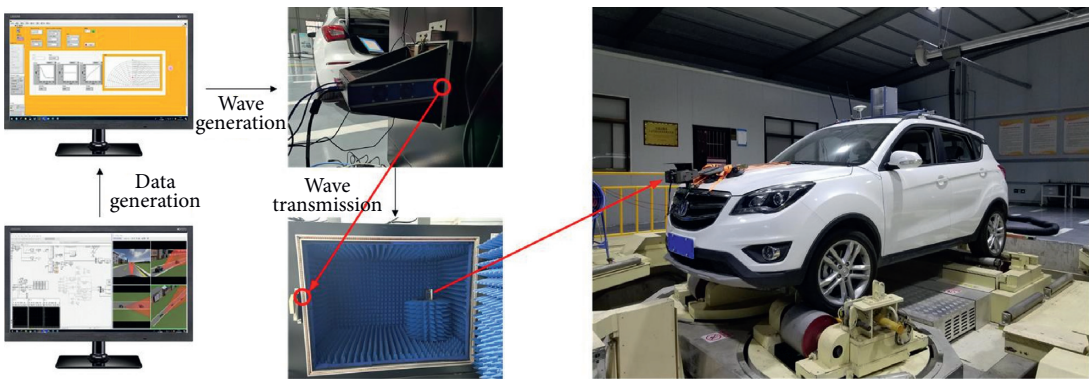


Figure 7: The workflow of MMW radar black box. There are two small red circles in this figure, referring to the position of the transmitter. The blue ones with bumps are special absorbing materials. In AVIL test, the real radar is detached from the AV and placed in the black box, as shown in the figure.

It has been verified but not shown in this paper that the black box does interfere with the real radar sensor, such as detection noise, if this special absorption material is not applied. It is also a research gap worthy of study and discussion, that is, in the case of interference, for the algorithm or system robustness research, since we can effectively model different scenarios by changing the material coverage and the attitude of the actual radar.

3. AVIL Feasibility Analysis

In this section, the fundamental equations of the excellent U-K theory are illustrated at first. AEB control algorithm based on U-K approach is then proposed. Further, AVIL effectiveness is demonstrated.

3.1. The U-K Approach. In nature, all motions follow Newton's second law, and the generalized form of the equation of motion can be described as

$$F = ma, \tag{1}$$

where F is the total force exerting on the system, m is the mass, and a is the acceleration.

The motion equation of U-K mechanics for an unconstrained system can be obtained by Newtonian mechanics, which can be written in the form of

$$M(q,t)\ddot{q} = Q(q,\dot{q},t), \tag{2}$$

where M is a positive definite inertia $n \times n$ matrix, q is the coordinate which is an n-vector, $\dot{q}$ is the velocity, $\ddot{q}$ is the acceleration, t is the independent variable and generally refers to time, and Q is the force imposed on the system [30].

From equation (2), the generalized acceleration denoted by $a(q,\dot{q},t)$ of the unconstrained system at time t can be obtained.

$$\ddot{q} = M^{-1}(q,t)Q(q,\dot{q},t) = a(q,\dot{q},t). \tag{3}$$

There are inevitably some constraints presented in the system that need to be considered. These constraints can be roughly divided into two types, such as m holonomic constraints which can perform differential twice in the form of

$$\varphi_i(q,t) = 0, \quad (i = 1, 2, \ldots, m), \tag{4}$$

and n nonholonomic constraints which can perform differential once in the form of

$$\varphi_j(q,\dot{q},t) = 0, \quad (j = 1, 2, \ldots, n). \tag{5}$$

It is assumed that equations (4) and (5) are sufficiently smooth and consistent; we can obtain the nominal generalized form of the constraints by differentiating $\varphi_i(q,t)$ twice and differentiating $\varphi_j(q,\dot{q},t)$ once with regard to time t, given by

$$A(q,\dot{q},t)\ddot{q} = b(q,\dot{q},t), \tag{6}$$

where A is a constraint matrix in the shape of $m \times n$, $\ddot{q}$ denotes the quadratic differential of generalized coordinates, and b is an m-dimensional vector.

The equation of motion with constraints can be obtained by combining equations (2) and (6). Additional "generalized forces of constraints" are applied to the system. Therefore, the actual motion equation of the constrained system can be written as

$$M(q,t)\ddot{q} = Q(q,\dot{q},t) + Q^c(q,\dot{q},t), \tag{7}$$

where Q^c is the constraint force causing the change of the acceleration, which can be regarded as a set of control forces acting on the unconstrained system.

In U-K equation, both ideal and nonideal constraints are considered in the system. Q^c can be written in the form of

$$Q^c(q,\dot{q},t) = Q^c_{\text{id}}(q,\dot{q},t) + Q^c_{\text{nid}}(q,\dot{q},t), \tag{8}$$

where Q^c_{id} and Q^c_{nid} indicate the ideal and nonideal constraint force vectors, respectively.

Udwadia and Kalaba have been developed the nominal explicit expression of Q^c_{id} and Q^c_{nid}. As there is no nonideal constraint involved in the vehicle system studied in this paper, only the ideal constraint force vector is taken into consideration, and equation (8) can be rewritten as

$$Q^c = Q^c_{id} = M^{1/2}B^+(b - Aa), \tag{9}$$

where $B = AM^{-(1/2)}$, $a = QM^{-1}$, and "+" represents Moore–Penrose generalized inverse [31].

Substituting equation (9) into (7), the general equation of motion in U-K mechanics can be yielded

$$M\ddot{q} = Q + M^{1/2}B^+(b - Aa). \tag{10}$$

3.2. A Case Study of AEB. In the present study, a common driving scenario is used to define the AEB maneuver. As shown in Figure 8, there are one vehicle and one pedestrian in the scenario. The vehicle is uniformly traveling along the straight lane, while a pedestrian is crossing the road at a constant speed. Therefore, the vehicle needs to slow down to avoid collision with the pedestrian. This vehicle is equipped with on-board sensors that can sense the surroundings during the driving process. After perceiving the pedestrian, the vehicle conducts a risk assessment and then takes emergency braking if necessary, to avoid the traffic accident.

Axis AA' depicts the trajectory of the pedestrian, and BB' is the centerline of the vehicle. L is the impact position for 50% of scenarios, which is also the hypothetical origin. Let x_r denote the lateral position of the vehicle head on the geodetic

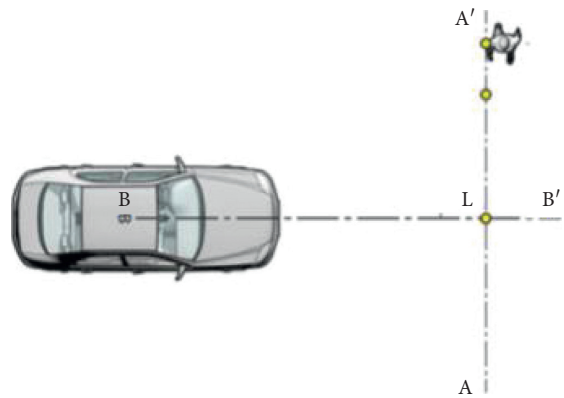


Figure 8: A common car-to-pedestrian scenario.

coordinates. Then, the nonlinear vehicle longitudinal dynamics model can be described as

$$\begin{aligned} \dot{x}_r &= v_r(t), \\ M_r\dot{v}_r(t) &= u_r(t) - F_a(t) - F_r(t), \\ d &= -x_r, \end{aligned} \tag{11}$$

where t is the time and v_r is the velocity. u_r is the vehicle driving force input, which is to be designed. M_r represents the mass of the vehicle. F_a denotes the aerodynamic resistance. F_r is the other external disturbances working on the vehicle. d is the distance between the vehicle and pedestrian.

Assuming that the final distance between the vehicle and the pedestrian is d_d, the difference between the stopping position and the current vehicle position can be expressed as

$$e_r(t) = d_d - d = d_d + x_r. \tag{12}$$

Based on equations (11) and (12), the U-K equation of this specific problem can be constructed as

$$M_r\ddot{e}_r(t) = u_r(t) + Q(e,\dot{e},t), \tag{13}$$

where $Q(e,\dot{e},t) = -F_a - F_r$.

According to our expectation, each time the vehicle is at a specific distance from the pedestrian d_s, its velocity v_s and deceleration a_s are consistent with our expectations. This constraint can be expressed as

$$\begin{aligned} \dot{d}_d = v_s \longrightarrow 0 \text{ when } d_s \longrightarrow d_d \\ \text{or } e_r \longrightarrow 0. \end{aligned} \tag{14}$$

Equation (14) can be differentiated again to be expressed in the form of (6), given by

$$\begin{aligned} \ddot{d}_d = a_s \longrightarrow 0 \text{ when } d_s \longrightarrow d_d \\ \text{or } e_r \longrightarrow 0, \end{aligned} \tag{15}$$

where $A(q,\dot{q},t) = 1$ and $b(q,\dot{q},t) = a_s$.

Substituting A and b into equation (9), we have

$$Q^c = Q^c_{\text{id}} = M^{1/2}B^+(a_s - a), \tag{16}$$

where the constraint force Q^c can be viewed as a control input. The vehicle will move in the way we expected under the control law of equation (16).

3.3. Testing on the AVIL Simulation Platform. A simulation experiment was carried out using MATLAB to validate the effectiveness of the proposed AEB control algorithm. Specifications of our test bench have been displayed in Table 1; as can be seen, due to the physical limitations of the machinery, only low-speed experiments can be feasibly performed at present. Therefore, only the simulation experiment under low-speed conditions will be verified here. We demonstrated medium-speed and high-speed verification in the "Discussion" section of this paper.

To verify the efficiency of the developed AEB algorithm, it is specified that when braking starts, there is only braking action acting on the vehicle, and it will eventually stop. Besides, the ability of the pedestrian to avoid the crash is not taken into consideration.

In the simulation, $F_a(t)$ can be estimated as $F_a(t) = c \cdot v(t) \cdot |v(t)|$. The parameters of the simulation experiment are set as follows: $M_r = 1300\,\text{kg}$, $c = 0.6$, and $F_r = 280\,\text{N}$. The simulation results are consistent with our expectations that, in the braking process, as the vehicle gets closer to the pedestrian, the deceleration of the vehicle will gradually increase. The deceleration and speed changes during braking are shown in Figures 9 and 10.

A group of 10 real road tests and AVIL tests were carried out, respectively. The braking time, braking distance, speed, and deceleration during braking are recorded for each test. Besides, the vehicle's initial braking velocity is maintained between [28–30] km/h to ensure the safety of the experiments fully, and the vehicle was traveling at a uniform speed before braking. The results of the comparative experiments are shown in Figures 11–14. Please note that the velocity data are obtained from the vehicle's OBD bus. The snapshot of AVIL test has been given in Figure 8.

Figures 11 and 12 demonstrate that, in the AVIL experiments, the braking time and braking distance are incredibly similar to those of the actual road tests. The AVIL tests and actual road tests reach an average braking distance of 10.26 m and 10.32 m, respectively. 2.03 s and 1.94 s are the average braking time of AVIL tests and real road tests. Overall, the braking time in AVIL tests is marginally longer than in the actual road tests by about 0.1 seconds. The braking time is close to 2.05 s of the simulation results, on both the real road tests and AVIL tests, and the experiment data in each group are relatively concentrated and less fluctuated. As shown in Figure 12, the braking distance of real road tests is more focused and shifts slightly. In contrast, the data in AVIL tests vary largely, and the range is significantly more extensive than that of the real road tests. However, the average braking distance of the AVIL tests is close to that of the real road tests.

Figures 13 and 14 show a set of velocity and its corresponding deceleration from comparative tests. Observing Figures 13 and 14 simultaneously, the change of speed has a tendency to be concave, which is the same as the original design of our expectation; that is, the braking force gradually increases, which is also in line with the driving behavior of most drivers. The reason why the deceleration is not a smooth curve is that the AV selected in this study is a wire-controlled modification, in which the throttle valve can be directly operated through the control command. The interval between two consecutive braking commands should be no less than 0.3 seconds. Therefore, deceleration will increase each time the command is sent; otherwise, the deceleration will decrease, causing the variation.

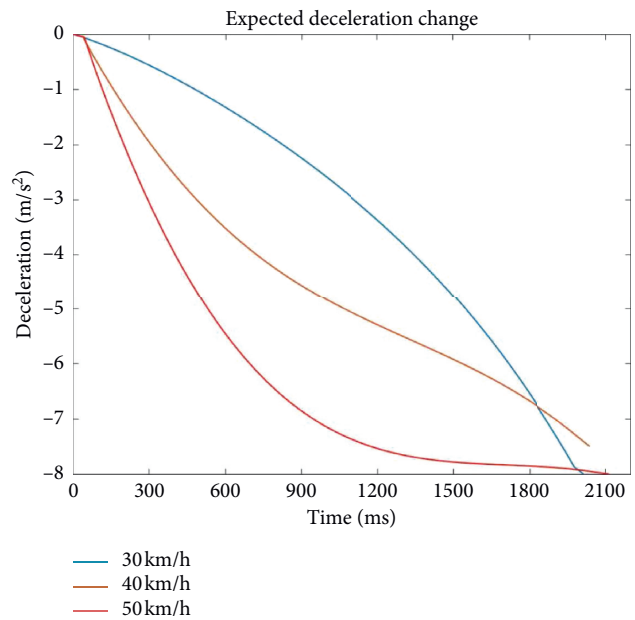


Figure 9: Expected deceleration.

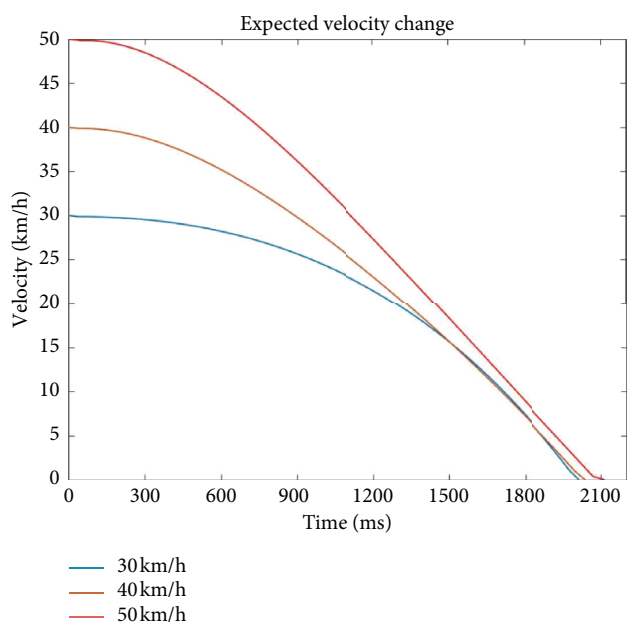


Figure 10: Expected velocity.

The big screen as shown in Figure 10 is not part of any modules of our AVIL simulation platform. The purpose of setting such a big screen is as follows: (1) In the early development process and debugging process, to ensure safety, generally, the driver operates the vehicle instead of the program. The driver can perceive the current driving environment surroundings through the big screen and then perform

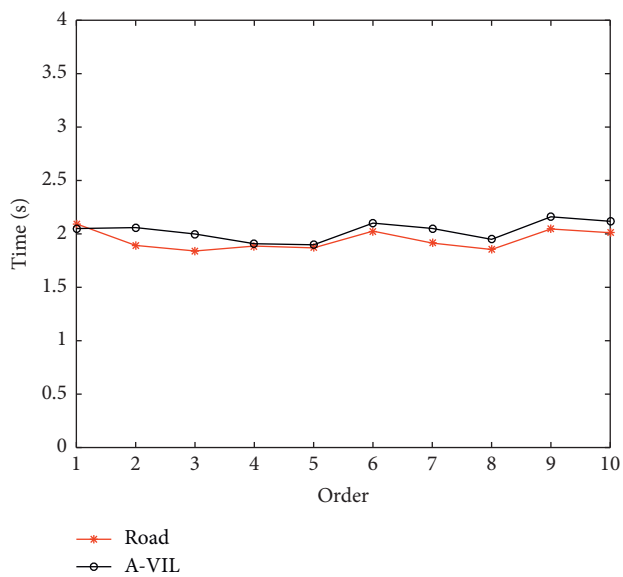


Figure 11: Time.

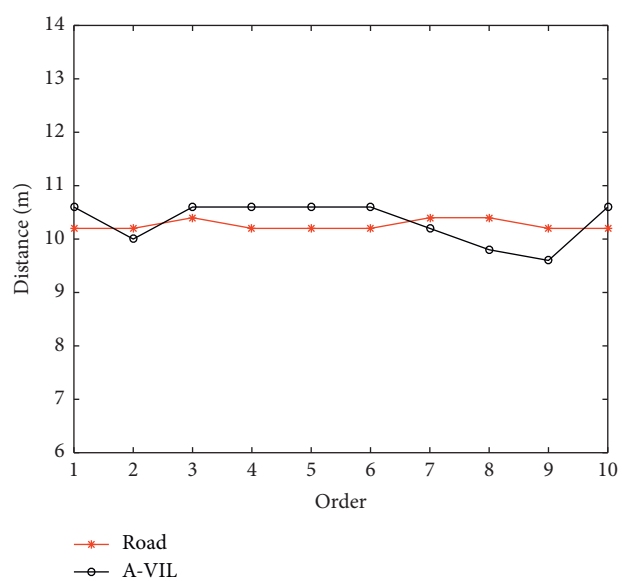


Figure 12: Distance.

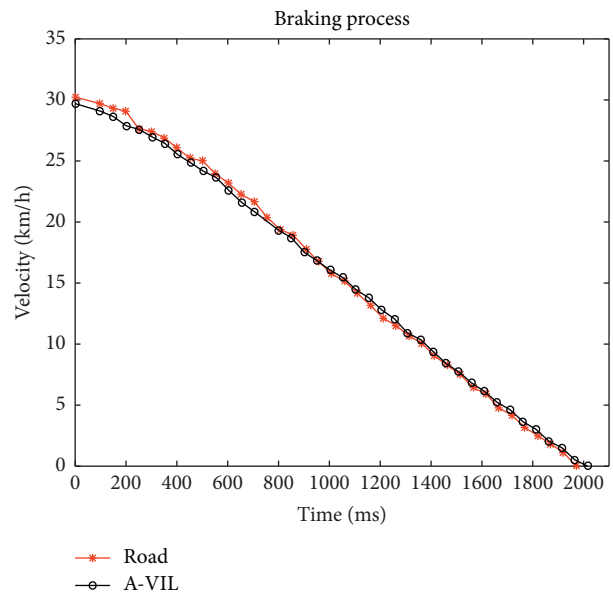


Figure 13: Speed.

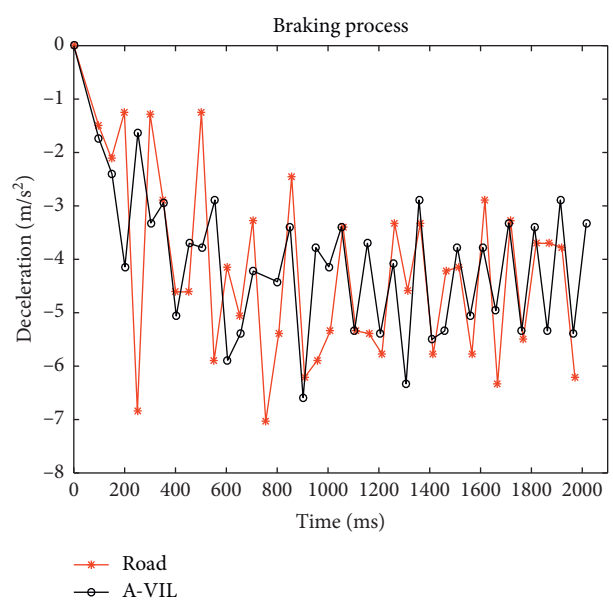


Figure 14: Deceleration.

operations. (2) To conduct research on driver or passenger driving experience in the future, test and evaluate different functions of AV. At this time, people in the vehicle need to know the current driving environment of the vehicle and the upcoming event, so as to not panic. (3) In the real test process, we tend to add a safety officer who does not usually take any driving behavior. When the AV is in an abnormal state or does not run as expected, the safety officer can terminate the test in time. For these above reasons, a big screen is needed.

4. Discussion

In the next few decades, road transport will undergo a significant revolution with the emergence of intelligent vehicle, such as AV, connected vehicle, and connected AV, which will certainly change the way we travel. In order to ensure the safe, reliable, and efficient operation of intelligent vehicles in different scenarios, a large number of tests must be carried out before deployment and implementation.

Compared with other VIL testbeds, such as [22], the VIL-based chassis-dynamometer testbed named SERBER is introduced to test and evaluate advanced driver assistance systems (ADAS) of vehicles. Similarly, [32] studied ADAS with a full-scale vehicle and a VIL-based chassis-dynamometer testbed. The difference is that [32] uses moving bases to represent other traffic participants instead of using simulation, improving the results' significance. The study [33] uses a VIL-based nonchassis-dynamometer simulator to study lane-keeping and cooperative cruise control

function. Although significant results have been achieved on the simulator, the effectiveness of the proposed functions on the real vehicles still lacks consideration and evaluation.

Although [22] achieves the interaction between the real vehicle and the virtual scene, it is not based on sensor-in-the-loop; therefore, the fidelity of the test remains to be discussed. Although [32] is based on the sensor-in-the-loop test, all test objects are real, which increases the cost, and there are problems such as the difficulty of completely reproducing the scene. Both of them have some common shortcomings, such as only one degree of freedom offered to the vehicle under test, the vehicle should be "fixed" on the chassis-dynamometer, and no sensor simulation.

The AVIL platform we proposed solves these problems completely. By using the customized bench, different subcomponents offer different degrees of freedom of the vehicle. They also guarantee the vehicle to perform various operations, and our vehicle does not need to be "fixed" on the bench. We also proposed a sensor-in-the-loop test method based on black boxes, which realizes the interaction between real sensors and virtual scenes. The most important feature is that our AVIL platform is easy to expand and maintain due to C/S mode. In this way, each submodule of the AVIL platform can be tested and debugged independently, and joint testing can also be performed on multiple submodules.

It can be seen from Figures 11–14 that the trends of all the two curves are almost the same, the data at each data collection moment is also similar, and the correlation of each group of data is more significant than 0.93. Therefore, the effectiveness of the AVIL simulation platform for testing the AV AEB system is confirmed. Since the initial braking speed fluctuates slightly each time, this leads to the fluctuation of braking distance and braking time. Although the fluctuation range of the braking distance of AVIL tests in Figure 12 is larger than that of actual road tests, this problem will be eventually solved by improving the platform and adjusting the parameters.

The experimental results show that when the initial braking speed is 30 km/h, the braking time is very close to the simulation experiment, and the similarity is more than 94%. The change in velocity has a small convex trend in the first half of the curve, similar to that in Figure 9. Due to the limitation of our AV, we set the maximum limit for the control of the throttle in the experiments. As a result, there is a trend of uniform speed change in the second half of the curve. The deceleration will tend to be approaching a constant value corresponding to this feature. The deceleration is not a smooth curve because the control of the throttle valve is similar to the piston. When the braking command is issued, the piston pushes in, the braking force increases, and the deceleration increases. During the empty window period of the two braking commands (0.3 s mentioned in Section 3.3), the piston will move outward. The braking force decreases, causing the decrease of deceleration, resulting in deceleration fluctuations.

Regarding the performance of the AEB control algorithm we proposed in medium-speed and high-speed conditions, we conducted simulation experiments, and the results are shown in Figures 15 and 16. As shown in these figures, the AEB control algorithm we proposed is still applicable to medium-speed and high-speed situations. Reasons accounting for only performing the low-speed experiment in the proposed AVIL platform are as follows:

(1) Due to the physical limitations of the test bench itself, the medium-speed and high-speed experiment cannot be carried out on our AVIL platform at present

(2) Based on reason (1) and also to fully ensure the safety of the experiment, the 1/2 of the maximum speed limit is determined, that is, 30 km/h

Carrying out actual road tests is not only time-consuming and laborious [34], but many practical factors such

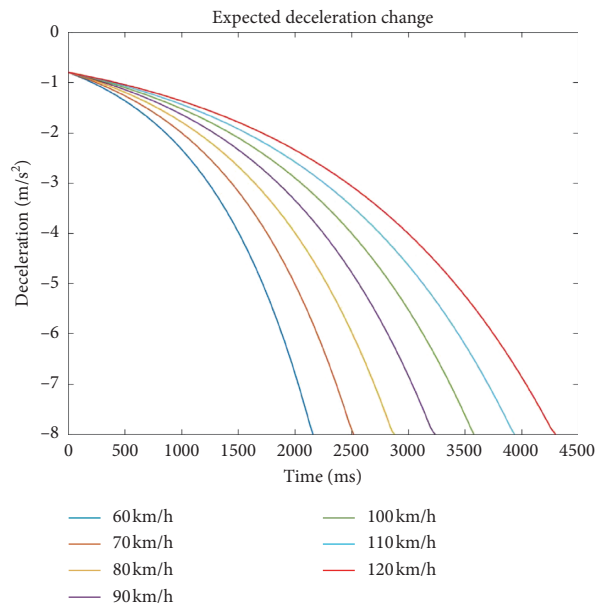


Figure 15: Expected deceleration.

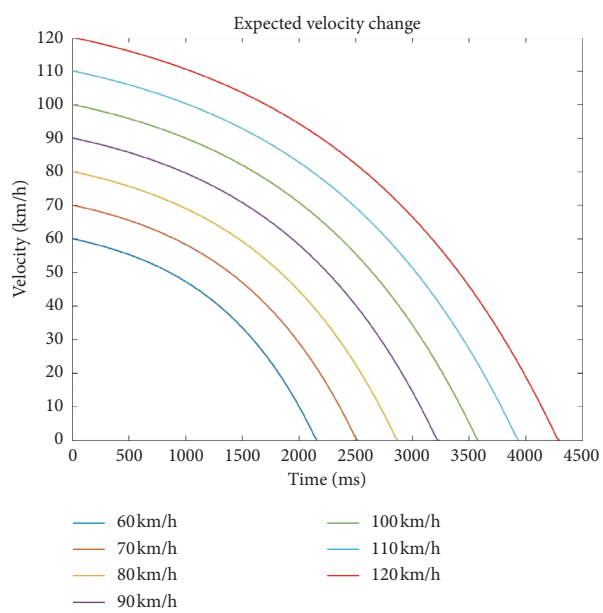


Figure 16: Expected velocity.

as weather conditions, illumination conditions, and ground geometry conditions are uncontrollable. Moreover, dangerous scenario testing is difficult, and its reproduction is complicated. Based on the developed indoor AVIL simulation platform, these become much easier. For example, we can add snow, rain, and fog models directly to the test scenario, change the behavior of traffic participants, or seamlessly switch the test scenarios. All of these will be truly reflected in the real AV based on our VR interaction method.

Compared with performing actual road tests or building a closed field, it is more efficient to construct an AVIL simulation platform in which uncertain factors can be controllable, and the test scenes are flexible, variable, and safe. As depicted in Figure 3, it is effortless to change the model, behavior, and other properties of the testing objects. By changing the model of the testing object, the properties of the object, such as length and height, can be changed; it can also easily change the operating behavior of the testing object, such as changing the trajectory or triggering interactive events. Some of the most basic metatest scenarios have been constructed as shown in Figure 4, such as curve lane, tunnel lane, and intersection lane, which can be switched at will and combined arbitrarily to build more complex test scenarios. Although it is still worth arguing whether the testing results in these scenarios still have high fidelity, it is undeniable that the AV testing method based on the AVIL simulation platform constructed in this paper has the potential to replace AV real road tests. Thus, AVIL-based AV testing might become an efficient and influential approach for real AV testing.

5. Conclusion

It is difficult to test real AV under harsh conditions. Using AVIL simulation platform under the laboratory environment to test real AV becomes an apparent alternative. In this paper, we developed an indoor full-scale AV test platform, which has the following advantages:

(1) This platform is designed as C/S mode so that each submodule can perform joint tasks and work independently, which is convenient for this platform expansion, debug, and maintenance.

(2) The test bench is mainly composed of four independent submodules as shown in Figure 2. Each submodule contains a drum roller, lifter, and gear components as shown in the figure, and each component has its corresponding role. In addition to offering freedom for AV, the drum roller can simulate continuous road surface and different road adhesion coefficient, gear can provide AV with the safety guarantee for steering operation, and lifter can simulate slop road surface.

(3) We designed a black box-based sensor data injection method to achieve sensor simulation, making it possible to interact between real sensors and virtual scenes. We designed and customized the MMW radar black box, and its feasibility has been verified.

In order to demonstrate the feasibility of the proposed AVIL simulation platform, we developed AEB system relying on MMW radar and then designed and conducted a comparison test between real road test and AVIL test. The results of these tests are encouraging as shown in Figures 11–14. The AVIL-based test results of AEB have a high similarity to those of the real road test. These results confirm the possibility that our proposed indoor AVIL simulation platform has a great potential to be used to substitute for the real road test, at least in the aspect of AEB test.

In the future, our team will focus on the following two aspects. First, restricted by the test bench, only the feasibility of the proposed AVIL platform for testing the AEB system at low speed under simple operating conditions has been demonstrated. Our next-generation test bench is currently under construction, and compared with the current test bench, its response time is only one-third, and the maximum speed of the vehicle can reach 120 km/h. Therefore, our team will discuss the feasibility of the AVIL platform with the new test bench for testing the performance of the AEB system under high-speed and extreme conditions. Second, our team will investigate the effectiveness of other applications with the same functions as AEB, such as V2X communication or steering avoidance, as it is therefore not possible to quantify the relative effectiveness of AEB compared to other competing technologies.

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 there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded mainly by the National Key Research and Development Program (no. 2019YFB1600100), partially by the National Natural Science Foundation of China (nos. 61973045 and 61903046), and partially by the Key Research and Development Program of Shaanxi (nos. 2018ZDCXL-GY-05-02 and S2018-YF-ZDGY-0300).

References

[1] World Health Organization, *Global Status Report on Road Safety*, World Health Organization, Geneva, Switzerland, 2018, https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/.

[2] World Health Organization, *Global Status Report on Road Safety 2018*, WHO, Geneva, Switzerland, 2019, https://apps.who.int/iris/bitstream/handle/10665/276462/9789241565684-eng.pdf?ua=1.

[3] WHO, *World Health Organization Report,* WHO, Geneva, Switzerland, 2015, https://scholar.google.com/scholar_lookup?title=World%20Health%20organization%20report&author=WHO&publication_year=2015.

[4] World Health Organization, *Global Status Report on Road Safety*, World Health Organization, Geneva, Switzerland,

2013, https://www.who.int/violence_injury_prevention/road_safety_status/2013/en/.

[5] World Health Organization, *Global Status Report on Road Safety*, World Health Organization, Geneva, Switzerland, 2009, https://www.who.int/violence_injury_prevention/road_safety_status/2009/en/.

[6] The Ministry of Public Security of the People's Republic of China, *Annual Report on Road Traffic Accidents of the People's Republic of China*, The Ministry of Public Security of the People's Republic of China, Beijing, China, 2017.

[7] A. F. Idriz, A. Abdul Rachman, and S. Baldi, "Integration of auto-steering with adaptive cruise control for improved cornering behaviour," *IET Intelligent Transport Systems*, vol. 11, no. 10, pp. 667–675, 2017.

[8] W. Cho, H. Heo, K. Yi et al., "Design and evaluation of an integrated vehicle safety system for longitudinal safety and lateral stability," in *Proceedings of the 22nd International Technical Conference on the Enhanced Safety of Vehicles*, pp. 1–9, Washington, DC, USA, June 2011.

[9] Z. Xu, Y. Wang, G. Wang et al., "Trajectory optimization for a connected automated traffic stream: comparison between an exact model and fast heuristics," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10. In press, 2020.

[10] S. E. Li, Y. Zheng, K. Li et al., "Dynamical modeling and distributed control of connected and automated vehicles: challenges and opportunities," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 3, pp. 46–58, 2017.

[11] E. Moradi-Pari, H. N. Mahjoub, H. Kazemi, Y. P. Fallah, and A. Tahmasbi-Sarvestani, "Utilizing model-based communication and control for cooperative automated vehicle applications," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 1, pp. 38–51, 2017.

[12] S. Moon, I. Moon, and K. Yi, "Design, tuning, and evaluation of a full-range adaptive cruise control system with collision avoidance," *Control Engineering Practice*, vol. 17, no. 4, pp. 442–455, 2009.

[13] B. V. Arem, C. J. G. V. Driel, and R. Visser, "The impact of cooperative adaptive cruise control on traffic-flow characteristics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 4, pp. 429–436, 2006.

[14] K. C. Dey, L. Yan, X. Wang et al., "A review of communication, driver characteristics, and controls aspects of cooperative adaptive cruise control (CACC)," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 491–509, 2016.

[15] M. Wang, W. Daamen, S. P. Hoogendoorn, and B. van Arem, "Cooperative car-following control: distributed algorithm and impact on moving jam features," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 5, pp. 1459–1471, 2016.

[16] I. Y. Noy, D. Shinar, and W. J. Horrey, "Automated driving: safety blind spots," *Safety Science*, vol. 102, pp. 68–78, 2018.

[17] X. Kuang, F. Zhao, H. Hao, and Z. Liu, "Intelligent connected vehicles: the industrial practices and impacts on automotive value-chains in China," *Asia Pacific Business Review*, vol. 24, no. 1, pp. 1–21, 2018.

[18] X. Kuang, F. Zhao, H. Hao et al., "Assessing the socioeconomic impacts of intelligent connected vehicles in China: a cost–benefit analysis," *Sustainability*, vol. 11, no. 12, Article ID 3273, 2019.

[19] E. Bringmann and A. Krämer, "Model-based testing of automotive systems," in *Proceedings of the 2008 1st International Conference on Software Testing, Verification, and Validation*, pp. 485–493, Lillehammer, Norway, April 2008.

[20] H. Hanselmann, "Hardware-in-the loop simulation as a standard approach for development, customization, and production test of ECU's," SAE Technical Paper 931953, SAE International, Troy, MI, USA, 1993.

[21] W. Deng, Y. H. Lee, and A. Zhao, "Hardware-in-the-loop simulation for autonomous driving," in *Proceedings of the 34th Annual Conference of the IEEE Industrial Electronics*, pp. 1742–1747, Orlando, FL, USA, November 2008.

[22] C. Galko, R. Rossi, and X. Savatier, "Vehicle-Hardware-In-The-Loop system for ADAS prototyping and validation," in *Proceedings of the 2014 International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS XIV)*, pp. 329–334, Agios Konstantinos, Greece, July 2014.

[23] M. A. M. Zulkefli, P. Mukherjee, Z. Sun et al., "Hardware-in-the-loop testbed for evaluating connected vehicle applications," *Transportation Research Part C: Emerging Technologies*, vol. 78, pp. 50–62, 2017.

[24] Y. Chen, S. T. Chen, T. Zhang, S. Y. Zhang, and N. N. Zheng, "Autonomous vehicle testing and validation platform: integrated simulation system with hardware in the loop," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, pp. 949–956, Changshu, China, June 2018.

[25] Z. Xu, M. Wang, F. Zhang, S. Jin, J. Zhang, and X. Zhao, "PATAVTT: a hardware-in-the-loop scaled platform for testing autonomous vehicle trajectory tracking," *Journal of Advanced Transportation*, vol. 2017, Article ID 9203251, 11 pages, 2017.

[26] J. Duan, R. Li, W. Wang et al., "Driver braking behavior analysis to improve autonomous emergency braking systems in typical Chinese vehicle-bicycle conflicts," *Accident Analysis & Prevention*, vol. 108, pp. 74–82, 2017.

[27] A. S. A. Rachman, A. F. Idriz, S. Li, and S. Baldi, "Real-time performance and safety validation of an integrated vehicle dynamic control strategy," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 13854–13859, 2017.

[28] B. A. Guvenc and E. Kural, "A low-cost, multiple-driver-in-the-loop adaptive cruise control simulator," *IEEE Control Systems*, vol. 26, no. 3, pp. 42–55, 2006.

[29] Andrew, "AutonoVi: Autonomous vehicle planning with dynamic maneuvers and traffic constraints," 2017, https://arxiv.org/abs/1703.08561.

[30] F. E. Udwadia and R. E. Kalaba, "On the foundations of analytical dynamics," *International Journal of Non-linear Mechanics*, vol. 37, no. 6, pp. 1079–1090, 2002.

[31] B. Noble and J. W. Daniel, *Applied Linear Algebra*, Prentice Hall, Hoboken, NJ, USA, 1988.

[32] O. Gietelink, J. Ploeg, B. De Schutter, and M. Verhaegen, "Development of advanced driver assistance systems with vehicle hardware-in-the-loop simulations," *Vehicle System Dynamics*, vol. 44, no. 7, pp. 569–590, 2006.

[33] S. Y. Gelbal, S. Tamilarasan, M. R. Cantas, L. Guvenc, and B. Aksun-Guvenc, "A connected and autonomous vehicle hardwareinthe-loop simulator for developing automated driving algorithms," in *Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3397–3402, Banff, Canada, October 2017.

[34] J. Xu, K. Yang, Y. Shao, and G. Lu, "An experimental study on lateral acceleration of cars in different environments in Sichuan, Southwest China," *Discrete Dynamics in Nature and Society*, vol. 2015, Article ID 494130, 16 pages, 2015.

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 8813461, 13 pages
https://doi.org/10.1155/2021/8813461

Research Article

Metro Train Operation Plan Analysis Based on Station Travel Time Reliability

Ruihua Xu,[1,2] Fangsheng Wang,[1,2] and Feng Zhou[1,2]

[1]*Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China*
[2]*Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, Tongji University, Shanghai 201804, China*

Correspondence should be addressed to Fangsheng Wang; wangfangsheng@tongji.edu.cn and Feng Zhou; zhoufeng24@tongji.edu.cn

Received 18 July 2020; Accepted 3 April 2021; Published 15 April 2021

Academic Editor: Zhixiang Fang

The train operation plan plays an essential role in metro systems and directly affects transportation organization efficiency and passenger service level. In metro systems, passengers have paid more attention to the travel time reliability (TTR), reflecting the reliability of metro operation management. This article proposes an analysis method of train operation plan based on TTR in the station dimension. First, an automated fare collection (AFC) data-driven framework is established to calculate the station travel time reliability (STTR) and analyze the train operation plan at different periods. The framework structure consists of four steps: AFC data preprocessing, STTR calculation and assignment, clustering algorithm design based on SOM neural network, and train operation plan analysis and optimization. Second, the proposed method is applied to the Beijing metro network as a case study. Several promising results are analyzed that allow the optimization of the existing train operation plan. Our research shows that STTR is a good supplement for the existing metro operation assignment studies, which can help analyze and optimize the train operation plan effectively. This study is also applicable to other metro networks with AFC systems.

1. Introduction

With the ongoing socioeconomic development, urban traffic congestion has become increasingly severe, especially in large cities like Beijing and Shanghai. Metro is playing an increasingly important role in urban public transportation, owing to the outstanding advantages of faster velocity, higher reliability, and larger capacity. With the continuous expansion of the scale of metro networks, passenger demand shows a high-speed growth, while the distribution of passenger demand presents the unbalance characteristics in the time-space dimension. There is an increasingly prominent contradiction between the transportation capacity supply and the passenger flow demand, and it puts forward higher requirements for train operation organizations under networked operating conditions.

As a significant part of the metro operation and management, the train operation plan directly affects transportation organization efficiency and passenger service level. In many large cities, train running intervals are continuously shrinking during the morning rush hours, while some stations are still highly congested. Passengers have particular travel characteristic, which generally concentrates in individual stations or periods. Restricted by objective conditions of metro network structure, metro transport capacity cannot meet the passenger flow demand at specific locations and periods, resulting in severe partial congestion in the network. The fundamental reason is that the configuration of network transport capacity does not match the distribution of passenger travel needs in the space-time dimension.

Recent research focused on extracting relevant indexes to reflect the train operation plan quality, such as train full-load rate [1–4] and platform congestion degree [5–7]. On one hand, the existing research methods screen the top/bottom ranking sections/stations according to the operation indicators, including section full-load rate and station passenger volume, and essentially sort operation indicators and get the concerned sections or stations, but cannot obtain the potential causes. On the other hand, passengers have

paid more attention to the travel time reliability (TTR) in public transportation, and TTR has been one of the most significant factors affecting transportation services level [8, 9]. In general, passengers will always ride the first train to arrive after they reach the platform unless the train is too crowded. When the transportation capacity cannot meet the passenger demand in some stations and sections, there will be a backlog of passengers waiting in station platforms and, thus, there will be a direct impact on TTR. As for an OD pair in the metro network, TTR typically has two definitions: (1) the probability that passengers can complete a trip within a specified time; (2) the fluctuation degree of the average travel time of passengers.

As the node for passengers to start and finish the journey in the train operation plan, stations are the core of transportation organizations in the metro system. Based on the characteristics of metro network structure and operation management, we propose the definition of *station travel time reliability* (STTR) as the fluctuation degree between the actual time and standard travel time of each OD from this station as the starting station to other stations. Based on the support of big data, STTR analyzes and evaluates the TTR of inbound passenger flow totally to reflect the passenger service level at different stations and periods. Combined with operation experience and travel investigations, the factors that affect the fluctuation of STTR value consist of the following three aspects:

(i) Passenger flow of the station is excessive.

(ii) Train running interval of the line is large; that is, the transportation capacity is insufficient.

(iii) Station location: trains are too crowded when arriving at the station because their capacity has been used in front of this station.

To sum up, the train operation plan analysis should not be limited to the ranking of indicators but also should pay attention to the analysis of potential causes. This study aims to develop a data-driven approach to analyze train operation plans based on the STTR of all stations in the network. The contributions of this article are as follows:

(1) Based on the AFC data, an STTR measurement model is built to calculate the value of passenger TTR from station dimension and principal component analysis (PCA) is used to process clustering elements.

(2) Combining the Self Organizing Maps (SOM) neural network, a station clustering framework is established with the STTR values and influence factors to analyze the train operation plan more objectively and comprehensively and explore the specific reasons for low STTR level.

(3) Apply the proposed approach to the Beijing metro as a case study, and several results are analyzed that inspire the optimization of the existing train operation plan.

2. Literature Review

Numerous studies in the literature related to the train operation plan analysis consist of the following aspects: operation organization [1–4] and station service level [5–7]. Li et al. [1, 2] constructed an interaction model of trains and passengers and obtained evaluation indexes such as full-load rate, number of passengers, and average waiting time and optimized the train operation plan based on the matching degree of capacity supply and passenger demand. Wang et al. [3] evaluated the adaptability of train operation schemes and passenger demand from three aspects: total adaptability, structural adaptability, and quality adaptability. Lu [4] divided transportation efficiency into three levels: capacity output efficiency, capacity utilization efficiency, and transport demand satisfaction efficiency. Tian [5] used the passenger flow aggregation and congestion as an indicator to measure the service level of the station and as one of the bases for the preparation of the train operation plan. Liu and Chen [6] used the minimization of factors such as the waiting time of passengers at the station as the objective function to establish a multiobjective nonlinear mixed-integer optimization model evaluates and optimizes the line operation plan. Shafahi and Khani [7] considered the minimum transfer waiting time as the goal and combined heuristic algorithms to optimize the transfer of the road network. We find these analysis methods focused on restoring the passenger travel process to extract relevant indexes, such as train full-load rate, platform waiting time. However, in the process of path restoration, the parameters such as the passenger walking time and the train maximum passenger capacity will have a few differences and fluctuations in the space-time dimension. The pattern of empirical values for these parameters will cause vast subjectivity and randomness in evaluation results.

The theory of TTR was first proposed on urban road traffic, and there are several types of research about the TTR analysis in public transportation. Considering travel behavior analysis in the road network, Asakura and Kashiwadani [10] gave a concept of TTR, the probability that passengers can complete the trip within the specified time, and measured TTR of an OD pair in a deteriorated road network [11, 12]. Lam and Xu [13] calculated TTR's value by establishing a traffic flow simulator model and access the reliability of metro systems organization management. Bell and Chirs [14] analyzed travel time change based on sensitivity analysis and described TTR by travel time variance. While some scholars [15, 16] used the buffer time index (BTI) to describe the TTR, BTI is the fluctuation degree between the actual and planned travel time at a specified period. Besides, Lomax et al. [16] defined the unit distance travel time and defined the BTI as the rate between the average travel time and the time of passengers having a 95% chance of arriving at the destination on time.

To our knowledge, little attention has been paid to introducing TTR to the train operation plan analysis in the station dimension, which is of great significance to the metro operation management. Zhang et al. [17] presented a new unit distance TTR evaluation index and method to assess the Beijing metro network. Li et al. [18] proposed a TTR calculation algorithm to analyze the reliability of transfer time quantitatively. Chen [19] proposed the definition and evaluation method of metro network operation reliability

and established a train operation delay propagation model. Based on the data-driven method [20, 21], this article focuses on calculating the STTR and analyzing and optimizing the train operation plan combined with the clustering algorithm.

3. Data Description

3.1. AFC Data. The study addressed in this article requires passenger travel time data extracted from the automated fare collection (AFC) data. The AFC system has become the primary method of collecting metro fares in many cities throughout the world. AFC system provides a large quantity of passenger flow information, recording passengers' activities with original station ID, destination station ID, tap-in time, and tap-out time. Necessary elements for the model formula are summarized (Table 1).

3.2. Train Diagram. The train diagram illustrates the relationship between space and time for train operation (Figure 1). Necessary elements for the model formula are summarized (Table 2). According to the train diagram data, we can extract each line's running interval at different periods for the model formula.

4. Methodology

As mentioned, the existing analysis methods emphasize screening the top/bottom ranking sections/stations according to the operation indicators, which are essentially the index ranking methods. However, an increasing number of researchers and professionals have identified shortcomings in traditional analysis methods. For example, these indicators may be subject to bias and error in evaluation results. Moreover, the manual methods usually only focus on getting the concerned sections or stations but cannot obtain the potential causes. For these reasons, alternative concepts and methods need to be developed. This article proposes a cluster-driven method for analyzing the train operation plan, consisting of four steps: AFC data preprocessing, STTR measurement model, cluster-based analysis method, and train operation plan optimization.

Step 1: AFC data processing

Input AFC data, calculate the lower and upper bound of each OD pair's travel time thresholds, and remove abnormal records that are not between the lower and upper bound of thresholds.

Step 2: STTR measurement model

Based on the Cumulative Chance Measurement Model (CCMM), calculate the values of STTR (NSTTR and LSTTR) by the actual and standard of travel time, and PCA is used to process clustering elements.

Step 3: Cluster-based analysis method

Introduce SOM neural network to clustering algorithm for station classification and explore the specific reasons for low STTR level.

Step 4: Train operation plan optimization

By combining the level of STTR and influencing factors, including passenger flow, train running intervals, and station location coefficient, analyze stations characteristics of different clusters, and design appropriate optimization measures in train operation plans for low-reliability stations and lines. For the convenience of model formulation, relevant sets and parameters are listed in Table 3.

4.1. AFC Data Processing. In general, passengers' travel time between the same OD will be within a reasonable section. Typically, the threshold of the route travel time is determined by the results of travel surveys. First, obtain the actual travel time set of each OD by extracting each passenger travel time from the network AFC ticket dataset. Passenger travel time is the difference between the passenger's tap-in time and tap-out time in the smart card. Secondly, sort the actual travel time data for each OD pair in ascending order. The lower and upper bound of the travel time threshold of the OD (Station i to Station j) are obtained from the following formulae:

$$\begin{aligned} t_{ij}^{\text{low}} &= t_{\text{ceil}\left(5\%*M_{ij}\right)}, \\ t_{ij}^{\text{up}} &= \min\left(t_{ij}^{\text{low}} * (1+a),\ t_{ij}^{\text{low}} + U\right), \end{aligned} \tag{1}$$

where t_{ij}^{low} is the lower bound of the travel time threshold; t_{ij}^{up} is the upper bound of the travel time threshold; $t_{\text{ceil}(5\%*M_{ij})}$ is the actual travel times value for the fifth percent [22]; M_{ij} is the number of passengers; a is the relative threshold coefficient; U is the absolute threshold.

The values of a and U are determined through travel surveys; normally, a is 0.6 and U is 20 minutes [22]. Then, the data with the actual travel time at $[t_{ij}^{\text{low}}, t_{ij}^{\text{up}}]$ are retained, and the noise data are removed for each OD travel time set.

4.2. STTR Measurement Model. The measure indicating TTR includes two types: probability and fluctuation. The former indicates the probability that the passenger could complete the trip within the specified time, and the latter reflects the fluctuation degree between the actual and planned travel time. The study in this article focuses on the quantitative relationship between passenger travel time and train operation plan so that we decide to use the fluctuation indicator as the basis of the model.

As distinct from manual methods, the proposed method integrates multiple indicators (STTR, passenger flow, train running intervals, geographic location, etc.) for cluster analysis and classification of stations. Thus, through the analysis of various categories, we can evaluate the operation effect of the train operation plan of stations and lines. Therefore, we propose a measurement model to calculate the STTR value and analyze the correlation between STTR values with these factors and provide the basis for cluster analysis in the next charter.

Firstly, the lower bound of the travel time threshold (t_{ij}^{low}) is used as the standard travel time (t_{ij}^{stand}) of the OD, and the passengers' TTR of one OD pair (station i to station

Table 1: Necessary elements of AFC data.

Date	Original station ID	Tap-in time	Destination station ID	Tap-out time	Smart card no.
20161018	000521	07 : 15 : 00	001313	09 : 19 : 15	200141995
20161018	000551	10 : 04 : 56	000429	10 : 21 : 20	200141996
20161018	000535	10 : 02 : 32	000559	10 : 45 : 23	200141997
. . .. . .					

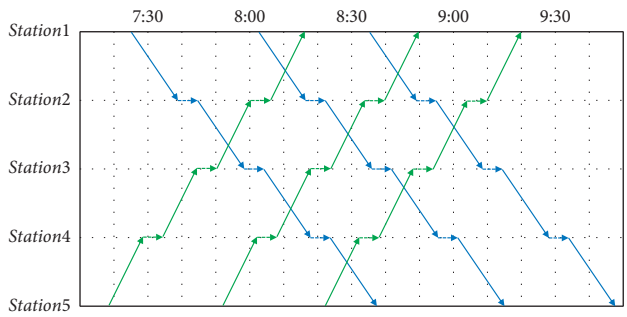


Figure 1: Example of actual train diagram.

j) is measured by the average and standard of travel time, as shown in equation (2).

$$\mathrm{TTR}_{ij} = \frac{\left(t_{ij}^{\mathrm{ave}} - t_{ij}^{\mathrm{stand}}\right)}{t_{ij}^{\mathrm{stand}}}, \tag{2}$$

where t_{ij}^{ave} is the average travel time between i and station j.

We divide STTR into Network STTR (NSTTR) and Line STTR (LSTTR). NSTTR is the relationship between this station and all other stations in the network, whereas LSTTR is the relationship between this station and all other stations in the same line. Based on the CCMM presented in TTR studies [23], we measure the STTR (NSTTR, LSTTR) of station i, as shown in equation (3).

$$\begin{aligned} \mathrm{NSTTR}_i &= \frac{\sum_{j\in S, j\neq i}\left(\mathrm{TTR}_{ij} * t_{ij}^{\mathrm{stand}} * M_{ij}\right)}{\sum_{j\in S} M_{ij}}, \\ \mathrm{LSTTR}_i &= \frac{\sum_{j\in S_l, j\neq i}\left(\mathrm{TTR}_{ij} * t_{ij}^{\mathrm{stand}} * M_{ij}\right)}{\sum_{j\in S_i} M_{ij}}, \end{aligned} \tag{3}$$

where S is the stations set of the metro network and S_i is the stations set of the line which station i belongs to.

Secondly, according to the previous analysis in this article, the influencing factors for the STTR level include passenger flow, train running intervals, and station location, so that we use the three influencing factors and the values of NSTTR and LSTTR as clustering elements:

(i) *Passenger flow*: the inbound passenger flow of the station, that is, the total OD passenger flow with the station as the departure station during this period;

(ii) *Train running intervals*: the train operation plan running intervals of the line where the station locates during this period;

(iii) *Station location coefficient*: analyze the geographic location of all stations in the network, and extract the central station, and set its station location coefficient as 0, and the station location coefficients of other stations are determined by the OD standard travel time from it to the central station;

(iv) NSTTR;

(v) LSTTR.

These five elements are the add-in values in this model. The first three are derived from the passenger flow statistics system, train operation plans, and geographical statistics data. Moreover, the calculations of NSTTR and LSTTR are directly related to AFC data. Traditional evaluation methods focus on the calculation and simple ranking of indicators, but the specific causes of the station or line's poor indicators are not enough.

Thirdly, we use PCA to reduce the clustering elements' dimensional reduction and analyze the correlation between STTR values with these factors. As a multivariate statistical method based on orthogonal transformation, PCA indexes multiple related variables of the research object into a few unrelated variables and retains feature vectors with significant contributions [24]. These unrelated comprehensive variables include most information provided by the original variables, thereby achieving dimensionality reduction. Specific steps are as follows:

Step 1: Normalization

Scale clustering elements to a normal distribution with a mean of 0 and a variance of 1.

Step 2: Correlation coefficient matrix

Compose the normalized clustering elements into a 5-dimensional random vector:

$$X = (x_1, x_2, x_3, x_4, x_5), \tag{4}$$

where the covariance of x_m and x_n is the correlation coefficient of them, namely,

$$p_{m,n} = Cov(x_m, x_n). \tag{5}$$

The correlation between x_m and x_n is

(i) positive correlation, when $p_{m,n} > 0$;
(ii) negative correlation, when $p_{m,n} < 0$;
(iii) irrelevant, when $p_{m,n} = 0$.

The larger the absolute value of $p_{m,n}$, the stronger the linear correlation of x_m and x_n. Finally, obtain the correlation coefficient matrix $D(X)$ of X.

Step 3: Principal components extraction

Extract the feature root of $D(X)$ and convert it to the corresponding standard feature vector μ_k, which is the

Table 2: Necessary elements of train diagram.

Date	Train ID	Destination station ID	Stop station ID	Stop order	Arrival time	Departure time
20161018	0429	000941	000933	5	07 : 12 : 15	07 : 12 : 45
20161018	0431	000941	000933	5	17 : 15 : 20	17 : 15 : 50
20161018	0103	000521	000545	13	10 : 45 : 23	10 : 45 : 53
. . . . .						

Table 3: Relevant sets and parameters in model formulation.

Sets/parameters	Definition
M_{ij}	Passengers number of OD (station i to station j)
a	The relative threshold coefficient
U	The absolute threshold
t_{ij}^{low}	Travel time threshold lower bound of OD (station i to station j)
t_{ij}^{up}	Travel time threshold up bound of OD (station i to station j)
t_{ij}^{stand}	The standard travel time value of OD (station i to station j)
$t_{ceil(5\%*M_{ij})}$	Actual travel times value for the fifth percent
t_{ij}^{ave}	Travel time average value of OD (station i to station j)
TTR_{ij}	Travel time reliability value of OD (station i to station j)
S	Stations set of the metro network
S_i	Stations set of the line which station i belongs to
$NSTTR_i$	The weighted average TTR value between station i and all other stations in the network
$LSTTR_i$	The weighted average TTR value between station i and all other stations in the same line

contribution rate of the main component Z_k, and sequentially extract Z_1, Z_2, ..., Z_r. Moreover, the cumulative contribution rate of these principal components reaches the specified threshold, which is 70% generally.

Therefore, previous approaches focus on index calculation and ranking, screens the top/bottom ranking sections/stations according to the operation indicators. Compared with the traditional measurement methods about TTR, the model proposed in this article has the following differences: (1) the evaluation index (STTR) in this model is directly calculated by AFC data and does not need path restoration; (2) because of the metro network complexity and OD quantity diversity, this model calculates TTR values from station dimension and divides them into two levels of network and line; (3) by integrating analysis with other factors, the model analyzes the correlation between STTR values with these factors and provides data support for cluster analysis method and train operation plan optimization in the next charters.

4.3. Cluster-Based Analysis Method. In this section, we will identify the stations with low reliability and provide some suggestions for improving the STTR level. Clustering analysis is commonly used to categorize large amounts of data. Considering different clusters tend to show distinct differences in the clustering analysis results, and the abnormal points can help distinguish the potential outliers in the data. In this article, by analyzing the different parts determined by cluster analysis, we can identify low-reliability metro stations and propose optimization suggestions in the train operation plan.

We use the SOM neural network to categorize and analyze stations. Based on the values of principal components, the stations with higher similarity are in the same group, and the attributes of these stations are considered to be the same. As an unsupervised learning neural network, SOM [25] has strong self-organization characteristics and only an input layer-competitive layer (Figure 2).

Compared with the K-means clustering algorithm, the advantages of the SOM neural network include the following:

(i) Not affected by the initialization of the cluster centroid

(ii) Improving the processing ability of nonlinear data

(iii) Reducing the influence of noise data

The SOM neural network cluster algorithm includes the following parts.

4.3.1. Determine the Number of Clusters. We use Silhouette Coefficient (SC) method to determine the number of cluster groups, that is, the number of station categories. The SC method combines the clustering degree of Cohesion and Separation. The Cohesion refers to the average distance between the sample point i and all other elements in the same cluster, denoted as $a(i)$. The Separation means the average distance between the sample point i and the points in the other cluster, traversing other clusters to obtain the minimum value, denoted as $b(i)$; the cluster is the neighbor cluster of i. The sample point i contour coefficient is

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}. \tag{6}$$

The larger the average of all stations' contour coefficient, the better the number of clusters.

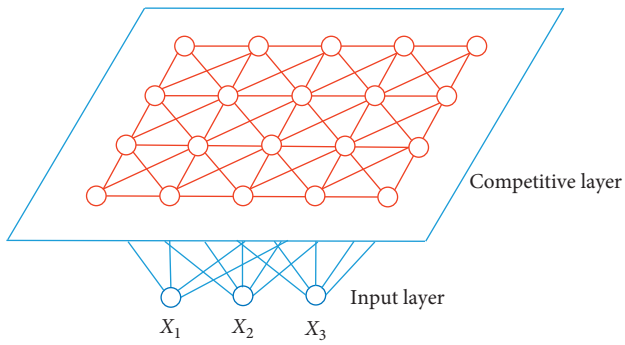


Figure 2: The structure of SOM neural network.

4.3.2. SOM Network Initialization. Import the principal component data of each station into the input layer of the SOM neural network. The data format is

$$U_k = \left[u_1^k, u_2^k, \ldots, u_N^k\right]^T, \tag{7}$$

where $k = 1, 2, .., r$, r is the number of principal components, and N is the number of stations.

We construct the initial neuron network of the competition layer, and the weight vector expression of the neuron node j and the input layer data is

$$W_j = \left[w_{j,1}, w_{j,2}, \ldots, w_{j,s}\right]^T, \tag{8}$$

where $j = 1, 2, .., M$, M is the number of neuron nodes in the competition layer.

4.3.3. Competitive Learning in SOM Network. SOM neural network adopts the method of competitive learning during the training process. Each input data point finds a node that matches it best in the competitive layer, called its activation neuron (WN). Then, use the stochastic gradient descent method to update the parameters of the active node and the data points it covers. The competitive learning process includes the following steps:

Step 1: The initialization parameters of the competition layer nodes have the same parameter dimensions as the input layer data dimensions.

Step 2: According to the Euclidean distance, match point i of the input layer to the nearest node(WN) in the competitive layer:

$$\left\|u_i - \mathrm{WN}\right\| = \min\left\{\left\|u_i - w_j\right\|\right\}, \quad j = 1, 2, .., M. \tag{9}$$

Step 3: Set WN as the center, the connection weights between other neurons in the neighborhood of the competition layer and the input layer neurons are modified:

$$w_j(t+1) = w_j(t) + h_{cj}(t)\left[X_j(t) - w_j(t)\right], \quad j = 1, 2, .., M, \quad j \neq c, \tag{10}$$

where t is the number of iterations, $w_j(t)$ is the connection weight of the node j and the input layer at the moment t, $X_j(t)$ is the input sample vector of the node j at the moment t, and $h_{cj}(t)$ is the neighborhood kernel function of WN at the moment t, namely:

$$h_{cj}(t) = \exp\left(-\frac{d_{cj}^2(t)}{2 * \delta^2(t)}\right), \tag{11}$$

where $d_{cj}^2(t)$ is the lateral distance between neuron j and WN and $\delta(t)$ is the amount of network width at the moment t, that is:

$$\delta(t) = \delta(0) * \exp\left(-\frac{n * \log \delta(0)}{1000}\right), \tag{12}$$

where $\delta(0)$ is set as the radius of the initial grid.

Step 4: Update the node parameters until the feature map gradually converges. The neurons in the SOM competition layer continually iterate and cluster simultaneously to divide the stations covered by each neuron into groups.

4.4. Train Operation Plan Optimization. After these steps, we obtain several station clusters. Combining the values of NSTTR and LSTTR, we can analyze the characteristics of all clusters and identify the stations with low reliability. By combining passenger flow analysis, train running interval, and station location, we could put forward several suggestions for train operation plans from these three aspects.

Take Line X as an example; as shown in Figure 3 there is only a long routing with the train running interval being 4 minutes in the train operation plan of Line X.

And the measures we could apply include the following:

(1) Minify train running interval. It is the most convenient method to improve the STTR level of all stations by increasing the number of trains per hour, as shown in Figure 4.

(2) Adopt the long-short routing operation mode. As shown in Figure 5, if the stations with low-reliability concentrate in a certain section (Station C to Station B), a short routing can be introduced.

(3) Fare incentives or congestion alerts.

When the train operation plan's transport capacity is close to saturation, we can adopt some other measures to encourage passengers to choose other routes to the destination station. In metro systems, fare incentives are emerging as a method to manage peak-hour congestion, including two strategies: a time-based fare incentive strategy (TBFIS) and a route-based fare incentive strategy (RBFIS). With the development of science and technology, passengers can be reminded of the congestion in some stations and sections in real-time through mobile apps or large screens in the stations to switch paths in time.

5. Case Study on Beijing Metro

5.1. The Network and Existing Analysis Method. In this section, the quality of methodology will be illustrated using a

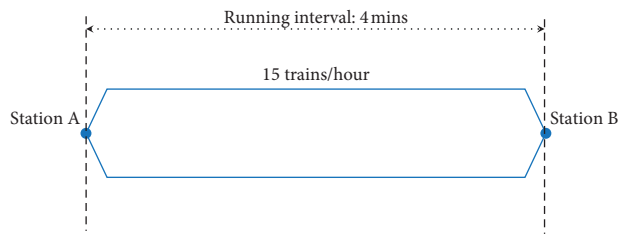


Figure 3: Original train operation plan of Line X.

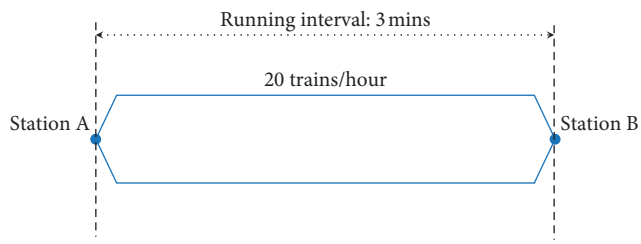


Figure 4: Minify train running interval.

real case of the Beijing metro system. In 2016, its network consisted of 18 lines and 326 stations (Figure 6), and there were more than 6,000,000 daily trips on average.

Over the past few years, the Beijing metro system has developed rapidly and is now one of the world's largest. According to the aggregation and dissipation of passenger flow, we divide the area inside the red frame into urban districts. Divide the area outside the red frame into suburban districts. In this article, we use a total of 39,453,138 AFC records on weekdays, calculate all the results by C#. Net and PL/SQL database programming. According to tap-in time in AFC data, we divide the study period into three parts: (1) morning peak periods: 6 : 00–10 : 00; (2) off-peak periods: 10 : 00–16 : 00; (3) evening peak periods: 16 : 00–20 : 00.

The train operation plan analysis method currently employed by the Beijing metro system is the index ranking method. This method screens the top/bottom ranking sections/stations according to the operation indicators, including section full-load rate and station passenger volume. The existing method is essentially to sort operation indicators and get the concerned sections or stations.

5.2. General Analysis. First, we calculate the values of NSTTR and LSTTR of each station at different periods based on the STTR measurement model. The values of NSTTR and LSTTR reflect the STTR level. Figure 7 shows the visualization of NSTTR and LSTTR values of all stations at different periods; the darker the color, the bigger the NSTTR/LSTTR value, that is, the lower the STTR level.

The values of NSTTR and LSTTR present a significant difference in the space-time dimension. In the morning peak, the urban stations' NSTTR and LSTTR values are small, while the suburban stations' values are larger, but there is the opposite in the evening peak. Furthermore, there is a relative balance in the off-peak periods.

Then, we select six lines with huge passenger demand, the NSTTR and LSTTR values are shown in Figure 8, and the stations at the red frame in the horizontal axis are urban stations. The results show that the LSTTR value of most stations is slightly lower than the NSTTR value, and their changing trend is nearly consistent in one line. In the morning peak, the values of NSTTR and LSTTR in urban stations are smaller; that is, the STTR level of urban stations is higher than suburban stations generally.

5.3. Clustering Analysis. Based on the general analysis, we choose the morning peak as the study period for clustering analysis. First, we obtain each station's passenger demand from the AFC data and calculate the running interval of each line (Table 4) from the train diagram data. According to the Beijing metro network's geographical location analysis, we regard Tiananmen West Station as the central station of the network, as the red pentagram in Figure 6. Then, we calculate the location coefficient of all stations.

5.3.1. Clustering Elements Processing. We use SPSS statistical software for the PCA on cluster elements, and the correlation coefficient matrix of cluster elements is shown in Table 5. There are positive correlations between the NSTTR, LSTTR and passenger flow, train running intervals, and station location coefficient.

The principal components with the top two rankings are extracted, and the cumulative variance is 76.8%. As shown in the principal component matrix (Table 6), the principal component PC1 represents mainly passenger flow, train running intervals, and location coefficient, while the principal component PC2 represents mainly NSTTR and LSTTR.

5.3.2. Clustering Results. First, we use the Contour Coefficient Method [26] to determine the optimal clustering number as 4. Then, we use MATLAB software for cluster analysis. The initial network of the SOM neural network is a $6 * 6$ neuron network. During the competitive learning process, each neuron updates its position and stations connected to it. Then, the categories of stations in the network are classified. As shown in Figure 9, the final positions of all neurons are in the red network, and the black points are the station points. The number of stations covered by each neuron is shown in Figure 10.

Finally, the station clustering results and the distribution in the network are shown in Figures 11 and 12, where the same color points are the stations of the same group, and the size of the station shape in Figure 11 is proportional to the passenger volume.

Based on the above analysis, we analyze each station's group characteristics.

Cluster 1. In Cluster 1, the values of PC1 and PC2 are low; that is, the levels of STTR and influence factors are high. These stations are distributed mainly in the urban districts, their passenger demand stress is weak, and transport capacity supply is high. These stations do not need to improve the train operation plan.

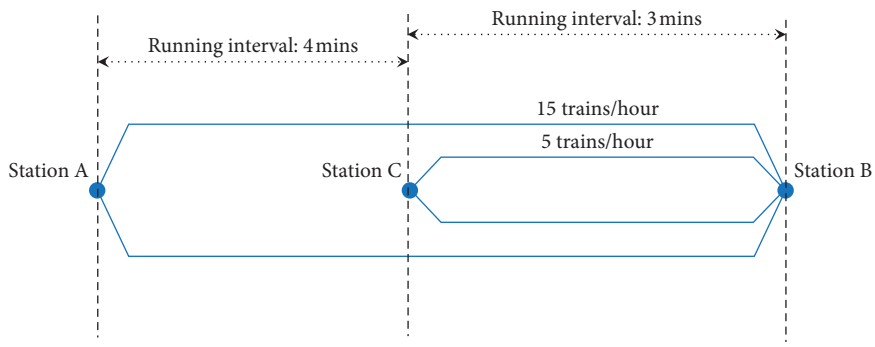


Figure 5: Long-short routing operation mode.

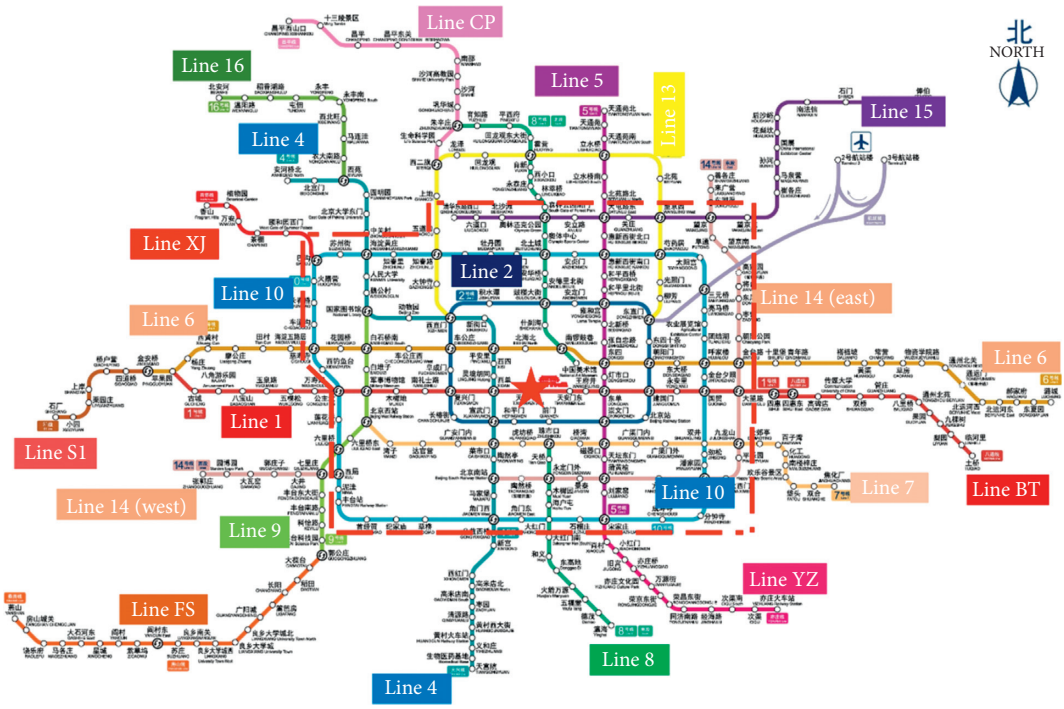


Figure 6: Beijing metro network.

Cluster 2. In Cluster 2, the values of PC1 are low, while PC2 is high; that is, the STTR levels are high, but influence factors are weak. There is a high matching of transport capacity and passenger demand, but a weak level of STTR in these stations. The representative lines and stations are Line YZ and Line FS (East). Regard these stations as potential stations that need attention.

Cluster 3. In Cluster 3, the values of PC1 are high, while PC2 is low; that is, the STTR levels are weak, but influence factors are high. The representative lines and stations are Line4 (South), Line9 (South), nad Line BT.

Cluster 4. In Cluster 4, the values of PC1 and PC2 are high; that is, the levels of STTR and influence factors are weak. The representative lines and stations are Line CP(North), Line8(North), Line5(North), Line15 (Northeast), and Line14 (West).

5.4. Train Operation Plan Analysis. Considering the distribution of each cluster station in the metro network, in this charter, we focus on the lines and stations in Cluster 3 and Cluster 4 and divide the potential causes of weak STTR into the following three aspects:

(1) Passenger flow stress is strong.

There is intense passenger flow stress in Line5 (North), Line8 (North), Line6 (East), and Line BT. These lines and stations may require minifngyi train running intervals. Furthermore, introduce the additional custom buses to divert commuter passenger flow.

(2) Train running interval is large.

There is a weak STTR level in Line4(North) because the train running intervals in these stations are 240 seconds, while those of other stations of Line4 are 120 seconds; thus, Line4(North) requires

Figure 7: The visualization of NSTTR and LSTTR.

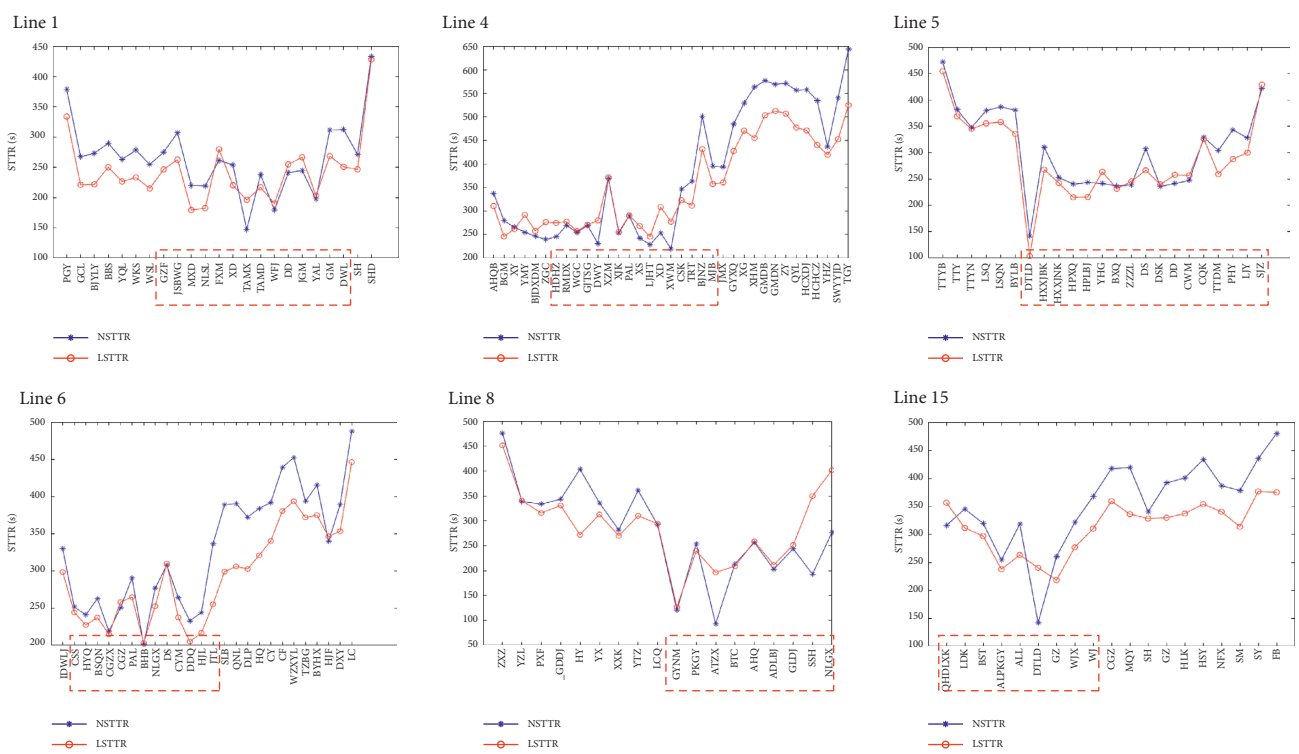


Figure 8: The values of NSTTR and LSTTR of six lines.

Table 4: Train running interval of lines in the morning peak.

Line name	Running interval(s)	Line name	Running interval(s)	Line name	Running interval(s)
Line1	120	Line6	257	Line14-East	327
Line2	129	Line7	200	Line15	327
Line3	129	Line8	212	Line CP	360
Line4-main	120	Line9	133	Line FS	360
Line4-DX	240	Line13	180	Line YZ	360
Line5	180	Line14-west	514	Line BT	180

Table 5: Correlation coefficient matrix.

	Passenger flow	Train running intervals	Location coefficient	NSTTR	LSTTR
Passenger flow	1	−0.21	−0.03	0.16	0.23
Train running intervals	−0.21	1	0.48	0.18	0.22
Location coefficient	−0.03	0.48	1	0.23	0.19
NSTTR	0.16	0.18	0.23	1	0.63
LSTTR	0.23	0.22	0.19	0.63	1

Table 6: Principal component matrix of clustering elements processing.

	Passenger flow	Train running interval	Location coefficient	NSTTR	LSTTR
PC1	0.6221	0.53	0.4002	0.2749	0.3104
PC2	0.1621	0.2746	0.216	0.62	0.6835

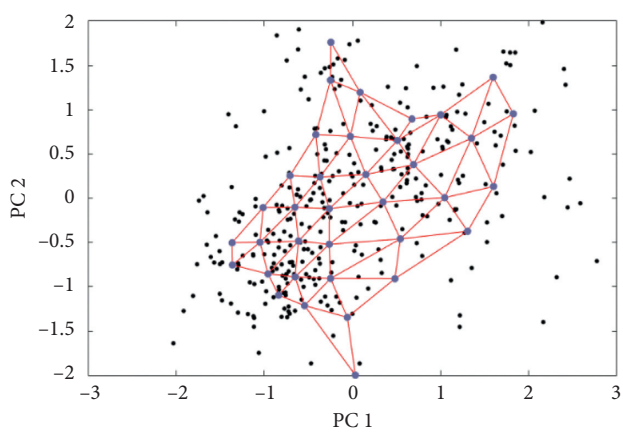


Figure 9: SOM neurons final locations.

minifying train running intervals. Similar lines and stations need to minify train running intervals including Line CP, Line14 (West), and Line15, and it is also appropriate to adopt the long and short routing operation mode for them.

(3) Station location problems

As shown in Figure 13, Line FS is located in the southwest of Beijing and has only a transfer station (GGZ), which connects with Line9. GGZ is a terminal station of the two lines, which means the passengers of Line FS who want to go to urban districts must pass through Line9.

According to the AFC data, in the morning peak, the proportion of transfer passengers in the Line FS's total passenger flow is 82.7%. Thus, the STTR level of Line9 is weak, while the passenger flow stress is weak, and the transport capacity supply is adequate in Line9.

Minify the running intervals of Line9 (South) appropriately to increase the transport capacity and reduce the impact from Line FS to Line 9. At the same time, Line9 is affected by the transfer passenger flow in the west section of Line 14 at QLZ and intensify the passenger flow pressure of Line9. Specific strategies, including fare incentives or congestion alerts, can be used to encourage more passengers from Line14 (West) to choose to transfer to Line 10 at Station XJ instead of Line9 at Station QLZ so that it can reduce the transportation pressure of Line9.

5.5. Comparison and Analysis. Take the existing analysis method using station passenger volume to compare it with the proposed approach in this article. Figure 14 shows the top 20 stations with inbound passenger volume. The X-axis represents the station name; Y-axis represents the inbound passenger volume; the color represents the station's cluster in Figure 11. It can be seen that most of these stations belong to Cluster 4 (red) and Cluster 2 (green).

In the proposed approach, the TTR and other influencing factors are integrated and analyzed by SOM neural network; these stations' passenger service level reflected by STTR values varies greatly and is divided into different clusters. Moreover, according to its influencing factors, analyze the optimization measures that need to be taken. In the existing method, all of the top 20 stations are considered in the train operation plan, so the comprehensive rating is insufficient and cannot explain why the station service level is low.

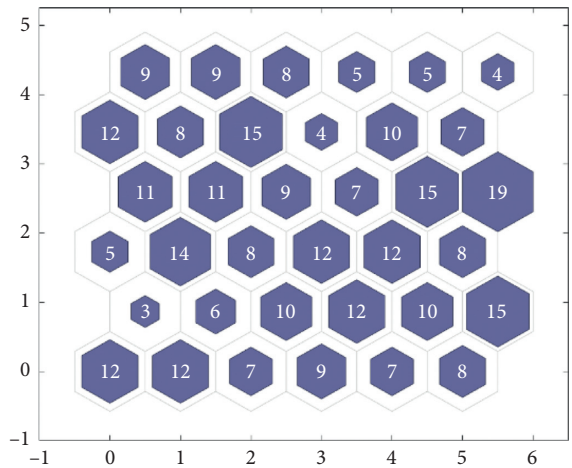


Figure 10: The number of stations covered by SOM neurons.

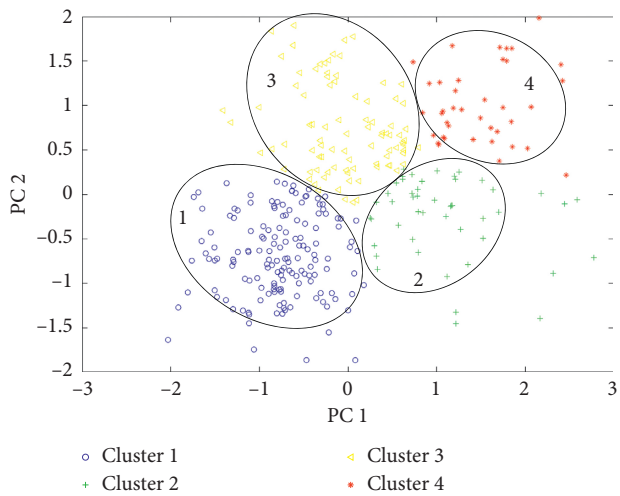


Figure 11: Stations clustering renderings.

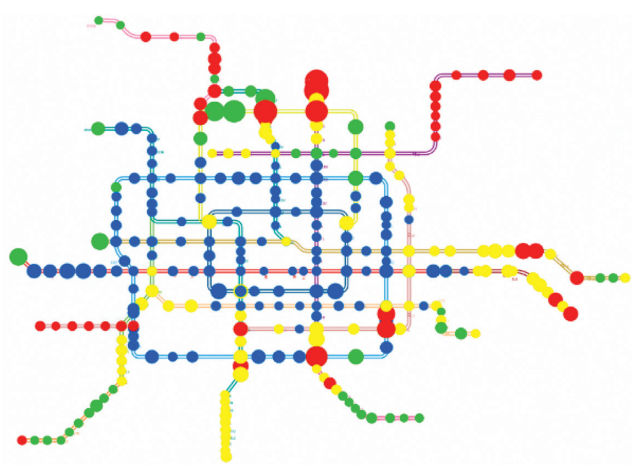


Figure 12: Distribution of various cluster stations.

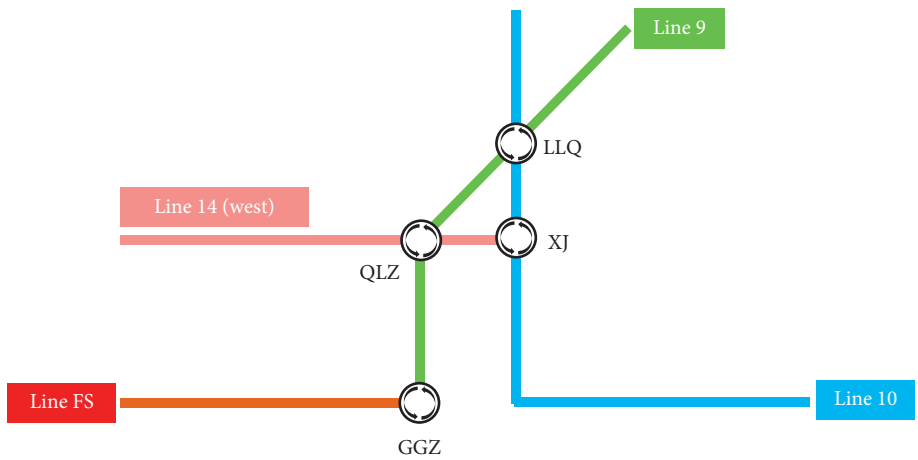


Figure 13: The lines connected with Line9.

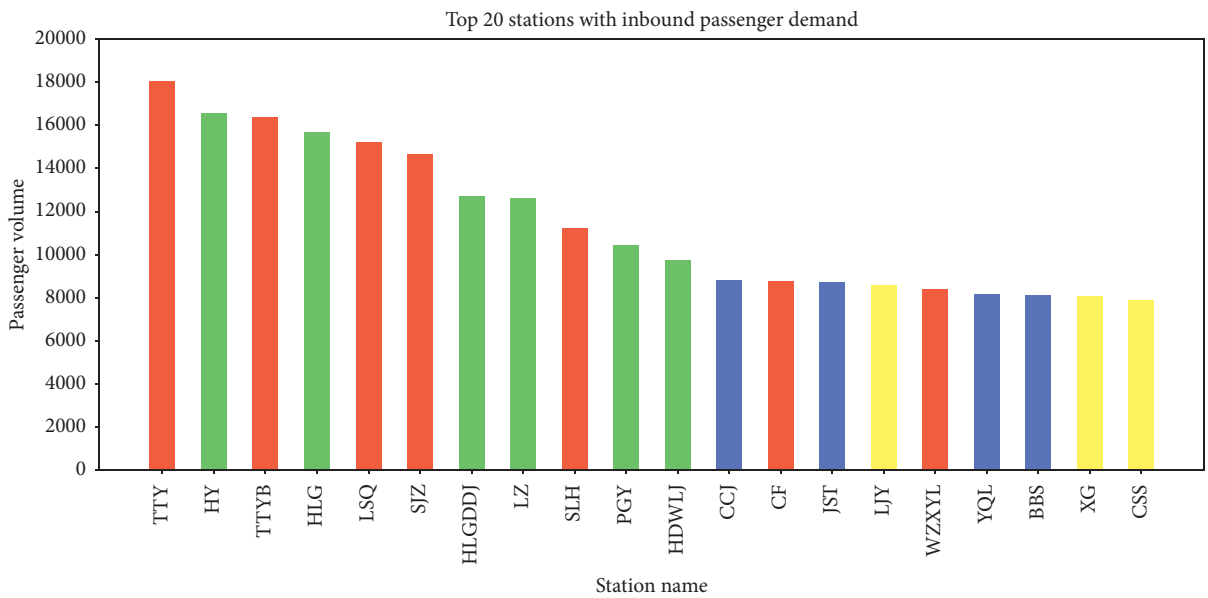


Figure 14: Top 20 stations with inbound passenger demand.

6. Conclusion

This article contributes a method for analyzing the train operation plan based on the STTR in the metro system:

(1) STTR, which is the fluctuation degree between the actual time and standard travel time of each OD from this station as the starting station to other stations, is calculated by AFC data.

(2) The clustering algorithm based on SOM neural network is efficient in classifying stations and identifying the potential causes of weak STTR level. SOM is an unsupervised learning neural network with strong self-organization and visualization characteristics.

(3) Taking the Beijing metro network as an example, the framework is applied, and the results are given and discussed in detail. Besides, several suggestions are put forward to optimize the train operation plan.

The application case of the Beijing metro network shows that the proposed method can be used to analyze the train operation plan effectively. It is also applicable to other metro networks with AFC systems. A possible future research direction is to expand the methodology framework to the reliability of transfer time in the time-space dimension. More efforts are also necessary to adopt diversified measures for optimizing operation management, such as asymmetric operation plans.

Data Availability

The AFC data used to support the findings of this study were supplied by Beijing Metro Co., Ltd., under license and so cannot be made freely available. Requests for access to these data should be sent to Mr. Wang, 15221628818@qq.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Key R & D Program of China (2018YFB1201402). The project was funded by the Ministry of Science and Technology. The authors are grateful for this support.

References

[1] S. Li, R. Xu, and Z. Jiang, "Evaluation method for matching degree between train diagram capacity and passenger demand for urban rail transit," *China Railway Science*, vol. 38, no. 3, pp. 137–144, 2017.

[2] S. Li, R. Xu, and K. Han, "Demand-oriented train services optimization for a congested urban rail line: integrating short turning and heterogeneous headways," *Transportmetrica A Transport Science*, vol. 15, no. 2, pp. 1459–1486, 2019.

[3] W. Wang, L. Cheng, T. Chen, and S. Ni, "A research on adaptability evaluation of train operation plan and passenger flow," *Railway Transport and Economy*, vol. 41, no. 7, pp. 65–71, 2019.

[4] F. Lu, *Theory of Transport Efficiency of Urban Rail Transit Network*, Beijing Jiaotong University, Beijing, China, 2016.

[5] S. Tian, *Study on Optimization of Train Routing Planning of Urban Rail Transit Considering Congestion Degree of Transfer Station*, Beijing Jiaotong University, Beijing, China, 2019.

[6] Y. Liu and D. Chen, "An optimization of long and short routing train plan of urban rail transit," *Urban Rail Transit*, vol. 41, no. 02, pp. 117–122, 2019.

[7] Y. Shafahi and A. Khani, "A practical model for transfer optimization in a transit network: model formulations and solutions," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 6, pp. 377–389, 2010.

[8] L. Bos, D. Ettema, and E. Molin, "Modeling effect of travel time uncertainty and traffic information on use of park-and-ride facilities," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1898, no. 1, pp. 37–44, 2004.

[9] A. J. Pel, N. H. Bel, and M. Pieters, "Including passengers' response to crowding in the Dutch national train passenger assignment model," *Transportation Research Part A: Policy and Practice*, vol. 66, pp. 111–126, 2014.

[10] Y. Asakura and M. Kashiwadani, "Road network reliability caused by daily fluctuation of traffic flow," in *Proceedings of the 19th PTRC Summer Annual Meeting, Proceeding Seminar G*, pp. 73–84, Brighton, England, 1991.

[11] Y. Asakura, "Reliability measure of an origin and destination pair in a deteriorated road network with variable flow," in *Proceeding of 4th Meeting of the EURO Working Group in Transportion*, pp. 75–77, Mayaguez, Puerto Rico, April 1996.

[12] Y. Asakura, M. Kashiwadani, and K. I. Fujiwara, "Functional hierarchy of a road network and its relations to time reliability," *Proceedings of JSCE*, vol. 583, no. 583, pp. 51–60, 2010.

[13] W. H. K. Lam and G. Xu, "A traffic flow simulator for network reliability assessment," *Journal of Advanced Transportation*, vol. 33, no. 2, pp. 159–182, 1999.

[14] E. Bell and C. Chirs, *Reliability of Transport Networks*, Research Studies Press Ltd., Biggleswade, England, 2000.

[15] S. Dai, C. Zhu, and Y. Chen, "Research on time reliability of urban public transit," *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, vol. 5, pp. 869–871, 2008.

[16] T. Lomax, D. Schrank, S. Turner et al., *Selecting Travel Reliability Measure*, Texas Transportation Institute, College Station, TX, USA, 2003.

[17] W. Zhang, P. Zhao, and X. Yao, "Reliability evaluation of Beijing metro rate travel time," *Shandong Science*, vol. 26, no. 6, pp. 77–81, 2013.

[18] W. Li, J. Zhou, F. Zhou, and R. Xu, "Calculation of travel time reliability of park-and-ride based on structural reliability algorithm," *Journal of Southeast University(Natural Science Edition)*, vol. 46, no. 1, pp. 226–230, 2016.

[19] J. Chen, *Research on Operation Reliability of Urban Rail Transit Network*, Tongji University, Shanghai, China, 2007.

[20] P. V. S. Rao, P. K. Sikdar, K. V. K. Rao, and S. L. Dhingra, "Another insight into artificial neural networks through behavioural analysis of access mode choice," *Computers Environment & Urban Systems*, vol. 22, no. 5, pp. 485–496, 1998.

[21] T.-H. Tsai, C.-K. Lee, and C.-H. Wei, *Neural Network Based Temporal Feature Models for Short-Term Railway Passenger Demand Forecasting*, Pergamon Press, Inc., Oxford, England, 2009.

[22] W. Zhu, W. L. Fan, A. M. Wahaballa, and J. Wei, "Calibrating travel time thresholds with cluster analysis and Afc data for passenger reasonable route generation on an urban rail transit network," *Transportation*, vol. 47, 2020.

[23] M. Wachs and T. G. Kumagai, "Physical accessibility as a social indicator," *Socio-Economic Planning Sciences*, vol. 7, no. 5, pp. 437–456, 1973.

[24] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 6, 1901.

[25] T. Kohonen, "Self-organizing maps, in Springer Series in Information Sciences," vol. 30, Springer-Verlag, Berlin, Germany, 3rd edition, 2001.

[26] R. J. Peter, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational & Applied Mathematics*, vol. 20, 1999.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 3976910, 9 pages
https://doi.org/10.1155/2020/3976910

Research Article

Simulation Study on Cascading Failure of Multimodal Transport Network

Jingni Guo,[1,2] Junxiang Xu,[1,2] Zhenggang He,[1,2] and Wei Liao[3]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China*
[2]*National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu 611756, China*
[3]*School of Logistics, Chengdu University of Information Technology, Chengdu 610225, China*

Correspondence should be addressed to Wei Liao; 3481531980@qq.com

Received 29 November 2019; Revised 3 November 2020; Accepted 16 November 2020; Published 1 December 2020

Academic Editor: Xingju Wang

Cascading failure in multimodal transport network may cause huge economic loss and social impact, which has gradually attracted public attention. In view of the coupling effect of nodes in multimodal transport network and the higher complexity of cascading failure process, the concepts of node correlation degree and node cooperation degree are proposed to characterize the characteristics of the network, and a logit model is introduced to calculate the initial load of nodes. In the case of ignoring network interruption, we propose two load redistribution methods: local allocation and global-local allocation. Taking the multimodal transport network in Sichuan–Tibet region of China as an example, the cascading failure effect of multimodal transport network in Sichuan–Tibet region is quantified by sensitivity analysis. The results show that when the load of the multimodal transport network in Sichuan–Tibet region exceeds the maximum capacity but does not exceed 150%~170% of the network capacity, the network can still operate normally. In addition, the nodes in the multimodal transport network should have 0.3~0.5 scalable space. In the cascading failure control method, load redistribution based on global-local allocation can minimize the impact of node overload.

1. Introduction

Cascading failure theory was first studied by Motter and Lai [1], and the classical ML model was proposed. Most of the existing researches are based on this. At present, the cascading failure theory is mainly used in power [2, 3], traffic [4, 5], and information network [6]. In the past few decades, scholars and researchers have done a lot of research on the cascading failure and its control methods in the network. Peng [5] built a cascade model on the urban road traffic network based on the disaster spreading dynamics. Considering the influence of self-healing ability of nodes and delay time factor on the disaster spreading mechanism, taking the number of collapse nodes and the repair rate of nodes as evaluation factors, the influence of different parameter values on the network was studied; Liu et al. [7] studied the cascading failure model in the traffic network under the emergency and determined the initial load by the incremental loading method under the multipath probability. Considering the influence of road resistance on the driver's decision, the redistribution strategy of the load based on the foreground theory can be determined. In the case of node failure, Zhong and Shuai [8] established a cascading failure model of transport network of dangerous goods under the continuous attack. Taking the node survival rate and the number of relative secondary joint failures as the evaluation factors, the node load was spread by means of average distribution, and the impact of different parameters on the network invulnerability was analyzed. Zhang et al. [9] analyzed the effects of different parameters on cascading failure process; Qian [10] studied the effects of network time delay and self-healing on cascading failure process. Hao et al. [11] and Peng et al. [12] studied the scale-free networks (BA networks), small world networks (NW networks), and

random networks (ER networks) and proposed to weaken the cascading failure in the network by controlling the node weighting and initial load. Hong et al. [13] studied the inhibition effect of multiple recovery strategies on the cascading failure in interdependent network. On the whole, the shortcomings of the existing studies are as follows: first, existing researches have simplified the idealization of the network, ignoring some of the attributes in the real network; second, the complex coupling relationship between nodes of multimodal transport network is ignored in the existing research.

It can be seen that the relevant researches focus on: first, the determination of initial load and update mechanism; second, the research of load redistribution after node failure; and third, the influence of different factors on cascading failure effect. In this paper, the relationships between nodes of multimodal transport network are redefined and applied to the cascading failure model. According to the characteristics of the network, the corresponding initial load distribution mode and load redistribution mode are proposed, which can provide a theoretical basis for effectively controlling the evolution process of cascading failure in multimodal transport network.

The structure of this paper is as follows: Section 2 introduces the definitions, including node relevance and node collaboration; in Section 3, the cascading failure model is established; in Section 4, the empirical analysis of the multimodal transport network in Sichuan–Tibet region of China is performed; and Section 5 summarizes the work of this paper.

2. Definitions

2.1. Node Relevance. In a multimodal transport network, there are multiple connection modes between nodes, and different node connections will affect the carrying capacity and load propagation speed of the network. Therefore, the calculation of the node relevance in the network is the premise of cascading failure modeling. According to the connection mode of organizational elements, the relationships between nodes of multimodal transport network can be divided into five types [14]. First, direct subordination: the nodes are directly connected, and the downstream nodes are subordinate to the upstream nodes. Second, indirect subordination: the nodes are connected indirectly, and the downstream nodes are subordinate to the upstream nodes. Third, direct control relationship: the nodes are directly connected, and the upstream node controls the downstream node. Fourth, indirect control relationship: the nodes are connected indirectly, and the upstream node controls the downstream node. Fifth, parallel relationship: the nodes are in parallel state, and there is no subordinate and control relationship.

Through the two steps of determining the node order of different connection modes and the calculation of relative importance contrast matrix, the relevance between nodes can be calculated: first, determine the node order under different connection modes. In a multimodal transport network, the number of nodes $X = \{x_1, x_2, \ldots, x_n\}$ is determined in a certain order of location. Second, determine a comparison matrix of the relative importance of nodes x_i and x_j. On the basis of determining the order relation, experts judge the importance degree between two adjacent nodes and obtain the importance degree matrix between nodes in the network. The numerical description of connection strength is shown in Table 1.

In the process of network failure and load redistribution, the load tends to flow in the direction of high node correlation.

2.2. Node Collaboration. The concept of node collaboration is often used in command and control network to measure the effect of collaborating between two or more nodes in order to complete one task [6]. In the multimodal transport network, the node collaboration K_{ij}^{w} is used to measure the effect of the two nodes collaborating with each other in order to complete the transport task using the w_{th} transport mode. In this paper, transport cost c_{ij}^{w}, transport time t_{ij}^{w}, and transport reliability r_{ij}^{w} are used to measure the node collaboration. The higher the node collaboration, the better the transport effect. When the two nodes are transfer nodes, the w_{th} transportation mode represents the way of goods outflow.

Let f_{ij}^{w} be the traffic volume for completing a transport task, l_{ij}^{w} be the transport operation between two nodes, o_{ij}^{w} be the unit operation cost, and u_{ij}^{w} be the unit operation time; then the transport cost c_{ij}^{w} and the transport time t_{ij}^{w} are shown as follows:

$$c_{ij}^{w} = f_{ij}^{w} \cdot l_{ij}^{w} \cdot o_{ij}^{w}, \tag{1}$$

$$t_{ij}^{w} = f_{ij}^{w} \cdot l_{ij}^{w} \cdot u_{ij}^{w}. \tag{2}$$

Among them, $c_{ij}^{w} > 0$, and $t_{ij}^{w} > 0$. Let $d(s)$ be a function of transport failure rate, which indicates the times of damage or failure of transport in unit time, and then the transport reliability r_{ij}^{w} is

$$r_{ij}^{w} = 1 - e^{-\int_{0}^{t_{ij}^{w}} d(s)\mathrm{d}s}, \tag{3}$$

where $r_{ij}^{w} \in (0, 1]$ and $\mathrm{d}(s) \geq 0$.

Due to the nonuniform dimension of cost, time, and reliability, the standard deviation method is adopted to standardize the cost and time, so that they are distributed between [0, 1]. The conversion formula is shown as follows:

$$x^{*} = \frac{x - \min}{\max - \min}, \tag{4}$$

where maxis the maximum value in the data set, min is the minimum value in the data set, andmax − min is the range. The normalized values of transit cost c_{ij}^{w} and transit time t_{ij}^{w} are recorded as c_{ij}^{w*} and t_{ij}^{w*}, and the subjective weighting method is used for weight distribution. The collaboration of transport nodes can be obtained as shown below, $K_{ij}^{w} \in (-(1/2), (1/2)]$:

Table 1: Numerical description of connection strength.

Significance	Same importance	Slightly important	More important	Obviously important	Very important	Extremely important
Numerical value	0.50	0.60	0.70	0.80	0.90	1.0

$$K_{ij}^{w} = -\frac{1}{4}c_{ij}^{w*} - \frac{1}{4}t_{ij}^{w*} + \frac{1}{2}r_{ij}^{w}. \tag{5}$$

By considering the influence of transport cost, transport time, and transport reliability, this paper establishes a node collaboration model.

3. Cascading Failure Model

In the cascading failure model, according to the different risk sources of cascading failure, it can be divided into node failure mode [15], edge failure mode [16], and node-edge mixed failure mode [17]. In the multimodal transport network, three failure modes coexist, but due to the limitation of space, this paper only discusses the cascading failure effect based on node failure [18].

3.1. Initial Load of Node. At present, the existing method of determining the initial load considers that the larger the capacity of the node is, the higher the initial load is. This paper considers that in the multimodal transport network, the initial load will be affected by the node collaboration, and the goods will tend to flow in the direction of high node collaboration, but with the continuous accumulation of goods, the node collaboration will also change. In this paper, a logit model is introduced to calculate the selection probability of the w_{th} transport mode between nodes i and j at time T, as shown below:

$$P_{ij}^{w}(T) = \frac{\exp\left(K_{ij}^{w}\right)}{\sum_{w}\exp\left(K_{ij}^{w}\right)}. \tag{6}$$

When the load of the network in a stable state is called the initial load in the network as $L(0)$, then the initial load $L_i(0)$ of node i is shown as follows, where ξ is a constant:

$$L_i(0) = \sum_{w}\sum_{j} P_{ij}^{w} L(0) + \xi. \tag{7}$$

The initial load $L_{ij}^{w}(0)$ of the w_{th} transport mode between nodes i and j is shown as follows, where δ is a constant:

$$L_{ij}^{w}(0) = P_{ij}^{w} L(0) + \delta. \tag{8}$$

3.2. Load Capacity. Suppose that the maximum load capacity of the connection edge of the w_{th} transport mode between nodes i and j in the network is C_{ij}^{w}, the maximum load capacity of node i is C_i; $\alpha > 0$ and $\varphi > 0$ are the tolerance factors of the edge and node, respectively, and then,

$$\begin{aligned} C_{ij}^{w} &= (1+\alpha)L_{ij}^{w}(0), \\ C_i &= (1+\varphi)L_i(0). \end{aligned} \tag{9}$$

Among them, the algebraic relationship between C_i and C_{ij}^{w} is $C_i \geq \sum_{w}\sum_{j} C_{ij}^{w}$.

3.3. Load Redistribution. In multimodal transport network $G = (V, E)$, the set of nodes is $V = \{v_1, v_2, \cdots, v_n\}$. After the state of node v_i changes from "normal" to "overload" due to the influence of internal and external factors, the load that exceeds the maximum load capacity on the node will propagate to other nodes in certain rules. This paper discusses two load redistribution methods: load redistribution method based on local distribution and load redistribution method based on global-local distribution.

3.3.1. Load Redistribution Method Based on Local Distribution. The basic idea of local distribution is as follows: considering the influence of the node collaboration and node relevance with the adjacent nodes, node v_i in overload state will give additional load to the adjacent "normal" nodes, and the load of the adjacent "normal" nodes is

$$L_j(t+1) = L_j(t) + \frac{\sum_{w}\beta_{ij}^{w}K_{ij}^{w}}{\sum_{j}\sum_{w}\beta_{ij}^{w}K_{ij}^{w}}\left(L_i(t) - C_i\right). \tag{10}$$

3.3.2. Load Redistribution Method Based on Global-Local Distribution. The basic idea of global-local distribution is as follows: on the basis of considering node state, node collaboration, and node relevance, supplemented by the remaining capacity factor of adjacent "normal" nodes, the load is redistributed with the goal of network balance. In this way, the load of adjacent "normal" nodes is as follows:

$$\begin{aligned} L_j(t+1) = L_j(t) &+ \frac{\left(C_j - L_j(t)\right)\cdot\left(L_i(t) - C_i\right)}{\lambda_1\sum_{j}\left(C_j - L_j(t)\right)} \\ &+ \frac{\left(L_i(t) - C_i\right)\sum_{w}\beta_{ij}^{w}K_{ij}^{w}}{\lambda_2\sum_{j}\sum_{w}\beta_{ij}^{w}K_{ij}^{w}}. \end{aligned} \tag{11}$$

Among them, $0 < \lambda_1, \lambda_2 < 1$, and $\lambda_1 + \lambda_2 = 1$.

When the load exceeds the maximum load C_j of node v_j, the excess part will continue to propagate to the adjacent nodes of v_j until the load of all nodes in the network is within the maximum load, and the cascading failure ends.

3.4. Measures of Cascading Failure Effect. In this paper, the proportion of abnormal nodes and average network efficiency are used to measure the cascading failure effect in multimodal transport network.

3.4.1. The Proportion of Abnormal Nodes. Suppose that the total number of nodes in the multimodal transport network is N, in which the total number of nodes in abnormal state is N_f and the proportion of nodes in abnormal state is R_f. The higher the proportion is, the stronger the cascading failure in the multimodal transport network is, and vice versa:

$$R_f = \frac{N_f}{N}. \tag{12}$$

3.4.2. Average Network Efficiency. The average network efficiency E is used to measure the connectivity in multimodal transport network. When nodes are abnormal, the higher the average network efficiency is, the weaker the cascading failure in the network is. Otherwise, the stronger the cascading failure is.

In this paper, overload function [19] $F_i(t)$ is used to express the difficulty of load passing through nodes:

$$F_i(t) = \begin{cases} 1, L_i(t) \le L_i(0), \\ 1 + \dfrac{L_i(t) - L_i(0)}{C_i - L_i(0)}(N-1), L_i(0) \le L_i(t) \le C_i, \\ N, L_i(t) > C_i, \\ N^2, \text{failure}. \end{cases} \tag{13}$$

Let d_{ij} represent the path with the highest load trafficability between node v_i and node v_j, and $P(i, j, w)$represents the set of all paths between node v_i and node v_j; then, the average network efficiency is as shown as follows:

$$d_{ij} = \min\left\{\sum_p F_p(t), p \in P(i, j)\right\}, \tag{14}$$

$$E = \frac{1}{N(N-1)}\sum_{i \ne j}\frac{1}{d_{ij}}. \tag{15}$$

4. Case Study

Taking the multimodal transport network in Sichuan–Tibet region as an example, this paper selects the hubs of Sichuan–Tibet railway, Sichuan–Tibet highway, and aviation as network nodes, and extracts 23 transport mode intersections in turn as 1–23 nodes in the multimodal transport network. The corresponding network structure is shown in Figure 1.

4.1. Node Relevance and Collaboration. According to the correlation between 23 nodes, the relevance between nodes is calculated under different connection modes. Among the adjacent nodes, according to the distribution of node order, the relationship of the direct control between the upstream node and the downstream node has the same relevance with the relationship of the direct subordinate between the downstream node and the upstream node. But the connection strength between the two nodes will change with the change of the upstream and downstream relationship. In general, the node relevance is 0.5. In this paper, only nongeneral node relationships are represented as shown in Table 2.

The node collaboration is affected by the transport cost, transport time, and transport reliability. At present, the multimodal transport network in Sichuan–Tibet region has not been built yet, and there is no adequate data. Therefore, the Delphi method is adopted in this paper to get the estimated data of the transport cost and transport time of the multimodal transport network by visiting the relevant staff in Sichuan–Tibet region. Then, through the analysis of the external environment of the node and the use of G1 method, the relevant data of transport reliability are obtained. According to the calculation formula in Section 2.1, the node collaboration under different transport modes can be obtained, as shown in Tables 3–5.

4.2. Numerical Analysis. In the multimodal transport network, the load will only propagate to the adjacent nodes, so only the nodes with direct and parallel relationships are calculated and analyzed. This paper sets Chengdu as the upstream direction and Lhasa as the downstream direction for the following simulation. Based on the local distribution and the global-local distribution methods, the cascading failure effect of multimodal transport network is analyzed by using MATLAB and Origin software. If the initial load in the network is $L(0)$, then the initial load $L_i(0)$ of each node can be obtained according to the calculation steps and order in Sections 2.1 and 3.1. Set the maximum tolerance factor of the node $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and the overload load proportion η as, respectively, 10%, 30%, 50%, 70%, and 90% of the initial load of the corresponding node:

(1) Considering the maximum tolerance factor of the node changes, the overload load proportion is 50%. Under the local distribution method, the effects on the proportion of abnormal nodes and average network efficiency of the network are shown in Tables 6 and 7.

(2) Considering the maximum tolerance factor of the node changes, the overload load proportion is 50%. Under the global-local distribution method, the effects on the proportion of abnormal nodes and average network efficiency of the network are shown in Tables 8 and 9.

(3) Considering the overload load proportion changes, the tolerance factor is 50%. Under the local distribution method, the effects on the proportion of abnormal nodes and average network efficiency of the network are shown in Tables 10 and 11.

(4) Considering the overload load proportion changes, the tolerance factor is 50%. Under the global-local distribution method, the effects on the proportion of

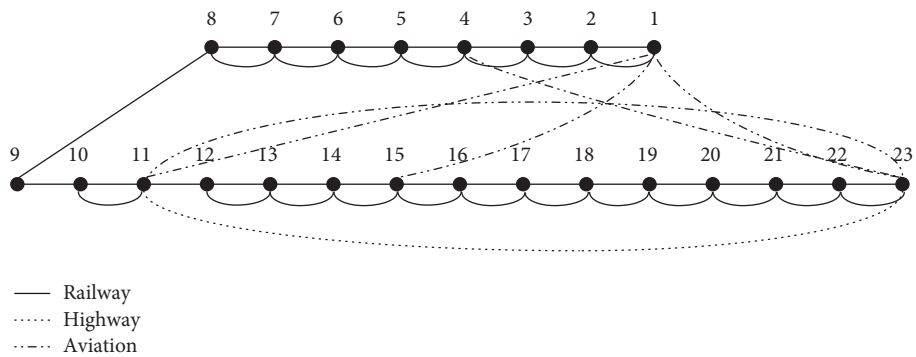


Figure 1: Multimodal transport network map in Sichuan–Tibet region.

Table 2: Relevance of nodes.

Node number	1	2	4	5	11	15	23
1	—	Direct control, 0.8	—	—	Direct control, 0.7	Direct control, 0.7	Direct control, 0.6
2	Direct subordinate, 0.8	—	—	—	Parallel	Parallel	Parallel
4	—	—	—	Direct control, 0.6	—	—	Direct control, 0.6
5	—	—	Direct subordinate, 0.6	—	—	—	Parallel
11	Direct subordinate, 0.7	Parallel	—	—	—	Parallel	Direct control, 0.6
15	Direct subordinate, 0.7	Parallel	—	—		—	Parallel
23	Direct subordinate, 0.6	Parallel	Direct subordinate, 0.6	Parallel	Direct subordinate, 0.6	Parallel	—

Table 3: Collaboration degree of railway transportation nodes.

Number	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9
Collaboration	0.885	0.836	0.793	0.72	0.625	0.639	0.606	0.619
Number	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
Collaboration	0.612	0.632	0.632	0.636	0.621	0.545	0.519	0.71
Number	17-18	18-19	19-20	20-21	21-22	22-23		
Collaboration	0.685	0.616	0.632	0.635	0.673	0.656		

Table 4: Collaboration degree of highway transportation nodes.

Number	1-2	2-3	3-4	4-5	5-6	6-7	7-8
Collaboration	0.842	0.875	0.525	0.477	0.414	0.424	0.402
Number	10-11	12-13	13-14	14-15	15-16	16-17	17-18
Collaboration	0.42	0.378	0.366	0.319	0.301	0.412	0.398
Number	18-19	19-20	20-21	21-22	22-23	11-23	
Collaboration	0.412	0.423	0.425	0.451	0.44	0.579	

Table 5: Collaboration degree of airline transportation nodes.

Number	1–11	1–15	1–23	4–23	11–23
Collaboration	0.614	0.592	0.522	0.612	0.572

Table 6: The proportion of abnormal nodes (a).

The proportion of overloaded nodes	The tolerance factor				
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
0.100	0.420	0.240	0.100	0.100	0.100
0.200	0.770	0.530	0.350	0.200	0.200
0.300	0.980	0.850	0.690	0.460	0.300
0.400	1.000	1.000	0.930	0.710	0.580
0.500	1.000	1.000	1.000	0.920	0.790
0.600	1.000	1.000	1.000	1.000	0.930
0.700	1.000	1.000	1.000	1.000	1.000
0.800	1.000	1.000	1.000	1.000	1.000
0.900	1.000	1.000	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000	1.000

Table 7: Average network efficiency (a).

The proportion of overloaded nodes	The tolerance factor				
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
0.100	0.045	0.067	0.098	0.098	0.098
0.200	0.021	0.038	0.063	0.098	0.098
0.300	0.004	0.018	0.036	0.070	0.098
0.400	0.000	0.000	0.019	0.034	0.069
0.500	0.000	0.000	0.000	0.016	0.036
0.600	0.000	0.000	0.000	0.000	0.013
0.700	0.000	0.000	0.000	0.000	0.000
0.800	0.000	0.000	0.000	0.000	0.000
0.900	0.000	0.000	0.000	0.000	0.000
1.000	0.000	0.000	0.000	0.000	0.000

Table 8: The proportion of abnormal nodes (b).

The proportion of overloaded nodes	The tolerance factor				
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
0.100	0.320	0.180	0.100	0.100	0.100
0.200	0.670	0.330	0.250	0.200	0.200
0.300	0.930	0.750	0.390	0.360	0.300
0.400	1.000	0.920	0.680	0.580	0.480
0.500	1.000	1.000	0.910	0.770	0.620
0.600	1.000	1.000	1.000	0.970	0.790
0.700	1.000	1.000	1.000	1.000	0.930
0.800	1.000	1.000	1.000	1.000	1.000
0.900	1.000	1.000	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000	1.000

abnormal nodes and average network efficiency of the network are shown in Tables 12 and 13.

The results of numerical analysis can be obtained as follows:

(1) When the overload load proportion gradually increases between 10% and 90%, the speed of risk propagation in the multimodal transport network will increase, but it will not affect the final state of the

Table 9: Average network efficiency (b).

The proportion of overloaded nodes	The tolerance factor				
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
0.100	0.065	0.087	0.098	0.098	0.098
0.200	0.041	0.058	0.081	0.098	0.098
0.300	0.014	0.038	0.057	0.079	0.098
0.400	0.000	0.013	0.039	0.054	0.080
0.500	0.000	0.000	0.016	0.036	0.057
0.600	0.000	0.000	0.000	0.017	0.033
0.700	0.000	0.000	0.000	0.000	0.019
0.800	0.000	0.000	0.000	0.000	0.000
0.900	0.000	0.000	0.000	0.000	0.000
1.000	0.000	0.000	0.000	0.000	0.000

Table 10: The proportion of abnormal nodes (c).

The proportion of overloaded nodes	The tolerance factor				
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
0.100	0.100	0.100	0.100	0.210	0.290
0.200	0.200	0.200	0.350	0.590	0.680
0.300	0.300	0.410	0.690	0.880	1.000
0.400	0.520	0.620	0.930	1.000	1.000
0.500	0.740	0.910	1.000	1.000	1.000
0.600	0.920	1.000	1.000	1.000	1.000
0.700	1.000	1.000	1.000	1.000	1.000
0.800	1.000	1.000	1.000	1.000	1.000
0.900	1.000	1.000	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000	1.000

Table 11: Average network efficiency (c).

The proportion of overloaded nodes	The tolerance factor				
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
0.100	0.098	0.098	0.098	0.079	0.063
0.200	0.098	0.098	0.063	0.047	0.024
0.300	0.098	0.065	0.036	0.200	0.000
0.400	0.071	0.032	0.019	0.000	0.000
0.500	0.034	0.018	0.000	0.000	0.000
0.600	0.015	0.000	0.000	0.000	0.000
0.700	0.000	0.000	0.000	0.000	0.000
0.800	0.000	0.000	0.000	0.000	0.000
0.900	0.000	0.000	0.000	0.000	0.000
1.000	0.000	0.000	0.000	0.000	0.000

network. Among them, when $\eta = 10\%$, the risk propagation speed is the slowest in multimodal transport networks. When $\eta = 90\%$, the risk propagation speed is the fastest in the network. As the proportion of overloaded nodes in the network increases, the network will eventually collapse. According to the results of the Delphi method, the proportion of overloaded nodes in the multimodal transport network of Sichuan–Tibet region is not expected to be 0.4, so the overload load proportion that can be carried is between 50% and 70%.

(2) When the tolerance factor α increases gradually from 0.1 to 0.9, the cascading failure speed in the multimodal transport network will also slow down, but with the increase of the proportion of overloaded nodes, the network will eventually collapse completely. Among them, when $\alpha = 0.1$, the cascading failure speed of multimodal transport network is the fastest, and when $\alpha = 0.9$, the speed is the slowest. Therefore, according to the results of the Delphi method, the tolerance factor in the network can be set between 0.3 and 0.5.

Table 12: The proportion of abnormal nodes (d).

The proportion of overloaded nodes	The tolerance factor				
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
0.100	0.100	0.100	0.100	0.260	0.350
0.200	0.200	0.200	0.250	0.480	0.670
0.300	0.300	0.390	0.430	0.760	0.890
0.400	0.470	0.580	0.680	0.920	1.000
0.500	0.620	0.730	0.910	1.000	1.000
0.600	0.790	0.900	1.000	1.000	1.000
0.700	0.960	1.000	1.000	1.000	1.000
0.800	1.000	1.000	1.000	1.000	1.000
0.900	1.000	1.000	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000	1.000

Table 13: Average network efficiency (d).

The proportion of overloaded nodes	The tolerance factor				
	$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$	$\alpha = 0.7$	$\alpha = 0.9$
0.100	0.098	0.098	0.098	0.082	0.073
0.200	0.098	0.098	0.081	0.059	0.045
0.300	0.098	0.079	0.057	0.036	0.022
0.400	0.079	0.058	0.039	0.017	0.000
0.500	0.053	0.031	0.016	0.000	0.000
0.600	0.036	0.020	0.000	0.000	0.000
0.700	0.017	0.000	0.000	0.000	0.000
0.800	0.000	0.000	0.000	0.000	0.000
0.900	0.000	0.000	0.000	0.000	0.000
1.000	0.000	0.000	0.000	0.000	0.000

(3) When the load in the network is redistributed by the local distribution method and the global-local distribution method, respectively, it can be seen that when the proportion of overloaded nodes is within 0.3, the effect of the two distribution methods is not much different. When the proportion of overloaded nodes is more than 0.3, the global-local distribution method can make the cascading failure speed in the network slower. Therefore, in multimodal transport network of Sichuan–Tibet region, the method of global-local distribution should be used for load redistribution.

5. Conclusion

According to the characteristics of node coupling in multimodal transport network, the concept of node degree is replaced by node relevance and node collaboration, and it is applied in the cascading failure model. The logit model is introduced into the model to calculate the initial load of nodes in the network. Based on the overload state of nodes, two load redistribution methods are proposed: local distribution and global-local distribution. Taking the proportion of abnormal nodes and average network efficiency as the measured index of the cascading failure results, the paper analyzes the cascading failure evolution process of the multimodal transport network in Sichuan and Tibet regions. The results show that when the proportion of overloaded nodes of the multimodal transport network in Sichuan and Tibet is less than 0.4, the overload load proportion that the network can carry is 50%~70%. When the tolerance factor in the network is set between 0.3 and 0.5, the global-local redistribution method can more effectively control the network cascading failure process. In other words, when the load of the multimodal transport network in Sichuan–Tibet region exceeds the maximum capacity but does not exceed 150%~170% of the network capacity, the network can still operate normally. In addition, the nodes in the multimodal transport network should have 0.3~0.5 scalable space. In the cascading failure control method, load redistribution based on global-local allocation can minimize the impact of node overload.

Data Availability

The corresponding data of multimodal transport network in Sichuan–Tibet region in the manuscript are obtained by interviewing related department and simulation of MATLAB software, and all data are available.

Conflicts of Interest

The authors declare there are no conflicts of interest.

Authors' Contributions

All authors have contributed to the creation of this manuscript for important intellectual content and read and approved the final manuscript.

Acknowledgments

The project was supported by National Key R&D Plan in China (2018YFB1601400).

References

[1] A. E. Motter and Y. Lai, "Cascade-based attacks on complex networks," *Physical Review E Statistical Nonlinear & Soft Matter Physics*, vol. 66, no. 6, p. 65102, 2002.

[2] Q. Lan, Y. Zou, and C. Feng, "Cascading failure of power grid under three side attack modes," *Computational Physics*, vol. 29, no. 6, pp. 943–948, 2012.

[3] J. Song, E. Cotilla-Sanchez, G. Ghanavati et al., "Dynamic modeling of cascading failure in power systems," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2085–2095, 2016.

[4] C. Liu, Y. Lv, B. Liu et al., "Study on cascading failure and invulnerability of urban rail transit network," *Transportation System Engineering and Information*, vol. 18, no. 5, pp. 82–87, 2018.

[5] S. Peng, *Study on the Dynamic Model of Disaster Spread for Cascading Failure of Urban Road Traffic Network*, Changsha University of Technology, Changsha, China, 2014.

[6] Q. Cui, J. Li, P. Wang et al., "Study on cascading failure of two layer coupling network model of command information system," *Journal of Harbin University of Technology*, vol. 49, no. 5, pp. 100–108, 2017.

[7] W. Liu, K. Chen, Z. Tian et al., "Partition model of road traffic accident influence area based on density entropy," *Journal of Traffic and Transportation Engineeing*, vol. 19, no. 6, pp. 163–170, 2019.

[8] P. Zhong and B. Shuai, "Cascaded failure modeling of dangerous goods transportation network under terrorist attacks," *Theory and Practice of System Engineering*, vol. 34, no. 4, pp. 1059–1106, 2014.

[9] Y. Zhang, Y. Lu, G. Lu, P. Chen, and C. Ding, "Analysis of road traffic network cascade failures with coupled map lattice method," *Mathematical Problems in Engineering*, vol. 2015, no. 4, 8 pages, Article ID 101059, 2015.

[10] Y. Qian, B. Wang, Y. Xue et al., "A simulation of the cascading failure of a complex network model by considering the characteristics of road traffic conditions," *Nonlinear Dynamics*, vol. 80, no. 1-2, pp. 413–420, 2015.

[11] Y. Hao, L. Jia, and Y. Wang, "Robustness of weighted networks with the harmonic closeness against cascading failures," *Physica A: Statistical Mechanics and Its Applications*, vol. 541, p. 123373, 2020.

[12] X. Peng, H. Yao, J. Du et al., "Invulnerability of scale-free network against critical node failures based on a renewed cascading failure model," *Physica A Statistical Mechanics & Its Applications*, vol. 421, 2015.

[13] S. Hong, C. Lv, T. Zhao, B. Wang, J. Wang, and J. Zhu, "Cascading failure analysis and restoration strategy in an interdependent network," *Journal of Physics A: Mathematical and Theoretical*, vol. 49, no. 19, p. 195101, 2016.

[14] M. Zhang and Y. Yuan, "Cascaded failure modeling of lifeline system based on disaster spread dynamics," *Systems Engineering*, vol. 32, no. 6, pp. 64–70, 2014.

[15] Y. Moreno, J. B. Gómez, and A. F. Pacheco, "Instability of scale-free networks under node-breaking avalanches," *Europhysics Letters (EPL)*, vol. 58, no. 4, pp. 630–636, 2002.

[16] Y. Moreno, R. Pastor-Satorras, A. Vázquez et al., "Critical load and congestion instabilities in scale-free networks," *Europhysics Letters (EPL)*, vol. 62, no. 2, pp. 292–298, 2003.

[17] C. Paolo, L. Vito, and M. Massimo, "Model for cascading failures in complex networks," *Physical Review E Statistical Nonlinear & Soft Matter Physics*, vol. 69, no. 4, p. 45104, 2004.

[18] T. Zhu, G. Chang, S. Zhang et al., "Research on cascaded failure model of command and control based on complex network," *Journal of System Simulation*, vol. 22, no. 8, pp. 1817–1820, 2010.

[19] G. Jingni, X. Junxiang, H. Zhenggang et al., "Cascading failure model establishment and empirical analysis of multimodal transport network," in *Proceedings of the 2020 IEEE 5th International Conference on Intelligent Transportation Engineering (ICITE)*, IEEE, Beijing, China, September 2020.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8878911, 13 pages
https://doi.org/10.1155/2020/8878911

Research Article

A Simultaneous Safety Analysis of Crash Frequency and Severity for Highway-Rail Grade Crossings: The Competing Risks Method

Amin Keramati,[1] Pan Lu,[2] Xiaoyi Zhou,[1] and Denver Tolliver[1]

[1]*Upper Great Plains Transportation Institute, North Dakota State University, NDSU Dept., 2880 P. O. Box 6050, Fargo, ND 58108-6050, USA*
[2]*Department of Transportation, Logistics and Finance, Upper Great Plains Transportation Institute, North Dakota State University, NDSU Dept., 2880 P. O. Box 6050, Fargo, ND 58108-6050, USA*

Correspondence should be addressed to Pan Lu; pan.lu@ndsu.edu

Received 19 March 2020; Revised 19 June 2020; Accepted 16 July 2020; Published 3 August 2020

Academic Editor: Lu Gao

This paper proposes a mathematical model, the competing risks method, to investigate highway-rail grade crossing (HRGC) crash frequency and crash severity simultaneously over a 30-year period. The proposed competing risks model is a special type of survival analysis to accommodate the competing nature of multiple outcomes from the same event of interest; in this case, the competing multiple outcomes are crash severities, while event of interest is crash occurrence. Knowledge-gain-based benefits to be discovered through the application of this model and 30-year dataset are as follows: (1) a straightforward and integrated one-step estimation process that considers both crash frequency and severity likelihood in the same model, so direct hazard ranking considering both crash frequency and severity likelihood is possible; (2) interpretative effects of identified covariates from both the direction and magnitude perspectives; and (3) the long-term cumulative effect of contributors with the cumulative incidence function.

1. Introduction

The highway-rail grade crossing (HRGC) is a unique spatial location where two transportation modes intersect with each other at grade level. Crashes at HRGCs are attributed to potential points of conflict between roadway traffic and train traffic. The crashes usually have relatively more severe results because of the huge mass difference between vehicle and train. Moreover, the economic consequences of crashes at HRGCs can be further extended by traffic delays on both the railway and the roadway. Crash modeling has been crucially important and intensively researched by decision-makers and transportation agencies for decades. Stakeholders and decision-makers rely on those models to provide direction to improve HRGCs safety performance.

A large volume of literature has been found to analyze and predict transportation accidents. Most of the previous research studies have focused on roadway intersection or roadway crashes [1–14]. Relatively little research effort has focused on HRGC accidents compared to roadway accidents [15–24]. Moreover, among all the previous HRGC accident analyses, the majority of them focus only either on crash frequency, often based on crossing inventory databases [19, 25–39], or on crash severity analysis, often based on historical crash police report databases [16, 40–52]. To understand and predict crash frequency and severity simultaneously and consistently is important for agencies seeking to improve safety so they can account for the common factors affecting both crash frequency and severity. Separate forecasting models help determine what factors affect the likelihoods of a crash occurrence or crash severity levels; however, there are several application obstacles: (1) Identified contributors are not consistent. Policy-reported surface conditions are often used in severity models but are

not available for crash occurrence models. (2) Estimated crash severity likelihoods are conditional probabilities given crash occurring based on identified unique set of contributors and not transferable for agencies to calculate absolute probability for a specific crash level. For example, separate forecasting models could provide 20% crash likelihoods with one set of contributors, say *A* to *E*, and 25% level-one crash severity, 30% level-two crash severity, and 45% level-three severity with another set of contributors, say *D* to *H*. Because of *F*, *G*, and *H* contributors, the probabilities are not transferable among different models; however, safety improvement agencies need consistent and commonly available information to assist in safety improvement decision-makings based on both crash occurrence and crash severity. The same forecasting model to account for both crash frequency and severities with commonly available contributors is needed so that unmeasurable variance can be accounted for in the same error term and the estimated likelihoods can be directly used by agencies. Little research has tried to incorporate crash frequency and crash severity in the same model.

This paper identified and proposed a modeling approach, the competing risks method, for HRGC accident analysis. This method has been verified and applied extensively in the medical field. However, it has not been used in safety analysis. The authors investigate the model's interpretive capabilities in frequency of crashes and crash severity simultaneously through the application of highway-rail crash and inventory data.

2. Literature on Crash Frequency and Severity

Previous HRGC safety studies focus either on crash frequency or on crash severity. Due to the random, discrete, and nonnegative nature of crash data, generalized liner models (GLMs) have been the most common technique for modeling crash frequency. Heydari and Fu [28] investigated a Poisson–Weibull model. Other statistical models [19, 33, 36, 53] such as zero inflated, hurdle, and generalized event count models were adopted to address data issues such as the excess number of zero accidents and overdispersions/underdispersions. Previous literatures [51, 54–56] also used data mining methods such as the hierarchical tree-based regression technique [39], neural network [23, 57], and random forests [58] to predict train-vehicle crash frequencies or severities at highway-rail grade crossings. Recently, Heydari et al. [27] presented a method to compare different geographic areas in terms of prespecified safety performance, which can measure crash frequency of a given type.

There is a vast body of previous researches focusing on modeling crash severity outcomes. Most of the modeling techniques are discrete choice approaches because of the discrete nature of crash severity levels. Hao and Daniel [45] and Abdel-Aty and Keller [57] applied ordered probit model in the United States considering crash severities that are naturally ordered. Hao and Daniel [44] followed up on the previous work and adopted an ordered probit model by considering various control measures at HRGCs. Eluru et al. [40] present a latent ordered response model considering ten years of crash data at HRGCs in the Unites States. Hu et al. [46] applied a generalized logit model to assess factors affecting crash severity at Taiwan's railroad grade crossings. Recently, Zhao et al. [51] used binary logit models and a generalized linear mixed model to study the association of potential factors with pedestrian injury severity levels using ten years of data at HRGCs in the United States.

All the abovementioned methods are very useful to model and understand crash frequency and severity outcomes separately. However, to predict crash frequency and the likelihood of crash severity outcomes simultaneously is important for agencies that need to account for the common factors affecting crash frequencies and severities based on the same set of available data. Unaccounted variables that affect crash occurrence and severity will be accounted for in error terms for each separate model. However, it is likely that these error terms are interrelated because they are describing the same concerns. Abdel-Aty and Nawathe [59] proposed a two-step approach to estimate crash frequency based on simulated geometric and traffic exposure information and then estimate crash severity with a neural network. Zalinger et al. [60] try to develop an integrated hazard regression model that considered both the number of crashes and their severity. However, in the final model development, the number of crashes and their severity are treated as accident history and only the number of accidents is selected as the hazard. Ye et al. [61] developed a simultaneous modeling approach for crash analysis, but the model only considers crash frequencies by collision types simultaneously using crash data rather than inventory data.

This paper attempts to further contribute to the HRGC safety by presenting a modeling methodology that decision-makers can use to estimate crash frequency and crash severity outcomes simultaneously. Moreover, the research will investigate the contributors' effects on the rate of crash and severity outcomes and their long-term cumulative effects over time.

3. Modeling Methodology

The competing risks model is a special type of survival analysis designed to correctly estimate the marginal probability of incidence outcomes when more than one cause of failure is possible. This method is intensively used in medical research [62–68] to study patient deaths attributable to competing events such as cardiovascular and non-cardiovascular causes. Survival analysis intends to estimate the probability of an occurrence of an event of interest. In transportation safety analysis, the event of interest is crashes. A unique feature of survival data is that not all targets experience the failure (e.g., crash) by the end of the analysis period. This feature is true for HRGCs. This phenomenon is referred to as censoring. Furthermore, transportation crash analysis is suitable for the competing risks model if each HRGC is viewed as a patient with a crash indicating a survival failure and where different severity outcomes at each HRGC are seen as different causes of survival failure or outcomes. In this study, crash severity outcomes can be

property damage only (PDO), injury, and fatality and they can be considered as competing outcomes for crashes. Figure 1 indicates the modeling structure with the three severity levels. The model starts with an initial state of "no crash" at year zero, which is 1989 for all records in this study. So all 3,310 crossings included in the study have zero crash record at the year 1989. As time passes, some crossings experience crashes. As shown in Figure 1, there are a total of 261 PDO triggered failures, 147 injury triggered failures, and 67 fatal triggered failures for a total of 3,310 crossings within the 30-year analysis period (1990 to 2018). Note, crossings with multiple crashes are excluded from the study database.

Consequently, each record in the dataset has three main variables of status (D), time (t), and crash severity (j), whereby competing risk algorithm is able to estimate the crash occurrence and severity likelihood. Status (D) is a binary variable that is equal to 1 if crash occurs and 0 otherwise. Considering t as the occurrence crash time, model is able to use variables D and t to estimate the crash occurrence likelihood. Crash severity variable can be equal to 1, 2, or 3 indicating crash severity levels of PDO, injury, and fatality, respectively. Therefore, if D related to one of the crash severity levels j is equal to 1 for a specific record at time t, the model will be able to estimate probability of crash occurrence ($D = 1$) with severity level j by time t. Note, in the situation where a crossing status associated with all severity levels is equal to 0 ($D = 0$) for whole study time span ($t \in [0, 30]$), crossing is counted as censored. Details about censored data and the way the competing model deals with them are explained in the Modeling Methodology section.

3.1. Competing Risks Model and Cause-Specific Hazard Function. The observable data in competing risks models can be represented by (1) the time of failure T or time of crash occurrence, (2) the cause of failure of D or crash severity levels (PDO, injury, and fatality), and (3) a covariate vector Z indicating contributors such as AADT and percentage of trucks. Equation (1) indicates the *cause-specific hazard function*, which is the fundamental concept in competing risks models. The cause-specific hazard function describes the instantaneous rate of Kth event failure in subjects. It is event-free currently:

$$\lambda_k(t) = \lim_{\Delta t \longrightarrow 0} \frac{\text{Prob}(t \le T < t + \Delta t,\ D = k \mid T \ge t)}{\Delta t}. \quad (1)$$

Competing risks model is a multivariate failure time model, where each individual patient is assumed to have a potential failure time for each type of failure [24]. The earliest of these failures is actually observed and the others are latent. If $\widetilde{T}_k$ is considered as the time to failure of cause k, only $T = \min\{\widetilde{T}_k\}$ and D can be observed, while D is an index variable specifying which event happened first. In competing risks models, if the endpoint of interest has not yet occurred at the end of the observation window, the event time is right censored. Therefore, in this study, the crossings that have no crash records during the 30-year study period are considered as right censored data. One of the advantages of the competing risks model is the ability to consolidate and utilize all the available data, even for the crossings that have no crash records. Previous studies only considered crossings with crash records [40, 48, 69]. Therefore, if some individuals (crossing records in HRGC safety analysis) are censored for all events by end of study, they have $D = 0$, and an extra censoring distribution $C \sim G$ is introduced, which is assumed to be independent of all the other events [70].

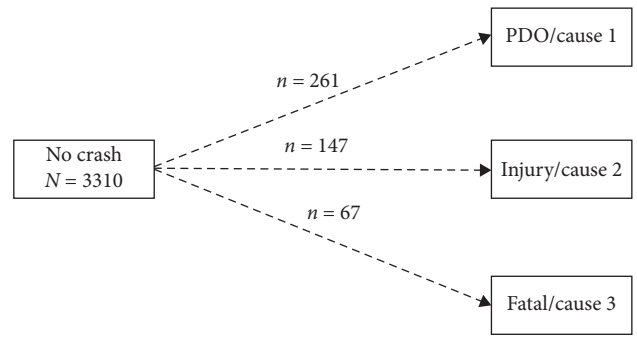


Figure 1: HRGC crash competing risks modeling structure.

3.2. Regression on Cause-Specific Hazards. The cause-specific hazard of cause K, $\lambda_k(t \mid Z)$, for a crossing record with covariate values $Z = (Z_1, \ldots, Z_p)$ is estimated in the following equation [70]:

$$\lambda_k(t \mid Z) = \lambda_{k,0}(t) \exp(\beta_k^{\top} Z), \quad (2)$$

where the hazard ratio of crash is the instantaneous risk of failure from a particular severity level k for subject with covariate vector Z which is defined as $\exp(\beta_k^{\top} Z)$ in equation (2). It is the conditional probability that a crossing with contributor vector Z has crash with severity level k at time t given it is crash-free just before time t. In equation (2), $\lambda_{k,0}(t)$ is the baseline cause-specific hazard of severity level k, and β_k indicates the contributors' effects on severity level k. The baseline hazard of severity level k can be estimated by the Breslow estimator, which is described by the following equation [71]:

$$\Delta \widetilde{A}_k(t, \widehat{\beta}) = \frac{\Delta N_k(t)}{S_k^{(0)}(\widehat{\beta}, t)}, \quad (3)$$

where $\widehat{\beta}$ is the maximum likelihood estimator of β and $\Delta N_k(t)$ denotes the number of records having crash with severity level k at t, and $S_k^{(0)}(\widehat{\beta}, t)$ indicates the number of records at risk. The term $S_k^{(0)}$ represents the weighted risk set and is estimated in

$$S_k^{(0)}(\beta, t) = \sum_{i=1}^{n} \exp(\beta^{\top} Z_{ki}(t)) Y_{ki}(t), \quad (4)$$

where $Y_{ki}(t)$ denotes the at-risk process; that is, $Y_{ki}(t) = 1$, and if the record i is at risk for severity level k at time t, the time point is just before time t. Note that, to estimate the effect of covariates on crash frequency, $\Delta N_k(t)$ in equation (3) is defined as the number of records that have crashes with any severity level.

The cause-specific function assumes independent censoring in estimating HR and coefficients. For instance, when a HRGC crash resulting in PDO is the event of interest, crossings with crashes resulting in other severities such as injury and fatality will be treated as censored observations. In other words, a crossing coded as PDO crash failure at time t is no longer at risk of injury crash or fatal crash failure at time t and will be treated as a censored observation. It is not easy to test whether a crossing that failed (crash occurred) in PDO would have failed from injury if the crossing did not fail from PDO, since the possible injury failure is unobservable for the crossings that actually are a PDO failure. But, in safety applications, dependence between competing risks should exist. In other words, a crossing having PDO crash can also experience a crash from other severity levels. Accordingly, the cumulative incidence function allows for estimation of the probability of the occurrence of an event while taking a competing risk into account, which does not assume the independence of competing events.

3.3. Cumulative Incidence Function. Cause-specific densities have the property of summing to the overall density as $\lambda(t) = \sum_k \lambda_k(t)$. Thus, the integral of the cause-specific density $I_k(t)$ is the cumulative incidence function (CIF) of crash severity *level k*. It describes the probability of crash occurrence with severity level k by time t, which is expressed in

$$I_k = \int_0^t \lambda_k(t) S(t)\, \mathrm{d}t = \mathrm{prob}\left(T \le t_j,\, D = k\right), \tag{5}$$

where the overall survival probability $S(t)$ estimates the overall probability of failure not happening at time t. Equation (6) denotes the estimation of $S(t)$ at t without considering the crash severities that are estimated by the Kaplan–Meier estimator [70]:

$$\widehat{S}(t) = \prod_{j:t_j \le t} \left(1 - \frac{d_j}{n_j}\right). \tag{6}$$

In equation (6), let $0 < t_1 < t_2 < \ldots < t_N$ be the ordered distinct crash time. If d_{kj} denotes the number of records having a crash with severity level k at t_j, then $d_j = \sum_{k=1}^{K} d_{kj}$ estimates the total number of crashes that occurred at t_j. The number of records at risk, n_j, indicates the number of records that are in follow-up status and have not had a crash until time t_j. A discretized version of cause-specific hazard of equation (1) can be revised as

$$\lambda_k\left(t_j\right) = \mathrm{Prob}\left(T = t_j,\, D = k \mid T > t_{j-1}\right). \tag{7}$$

$\lambda_k(t_j)$ can be estimated by $\widehat{\lambda_k}(t_j) = d_{kj}/n_j$ indicating the proportion of records at the risk of crash occurrence with severity level k. To simplify and to estimate the effect of covariates on cumulative probability, equation (6) can be expressed as equation (8) considering the crash severity level k:

$$\widehat{S(t)} = \prod_{j:t_j \le t} \left(1 - \sum_{k=1}^{k} \widehat{\lambda_k}\left(t_j\right)\right). \tag{8}$$

The estimator of the cumulative incidence function (CIF) is expressed in

$$\widehat{I_k}(t) = \prod_{j:t_j \le t} \widehat{S}\left(t_j\right)\left(\frac{d_{kj}}{n_j}\right). \tag{9}$$

4. Case Study

4.1. Data. Three main data resources are used for this research: (1) North Dakota (ND) roadway network, railway network, roadway intersections, and HRGCs from North Dakota Department of Transportation; (2) highway-rail grade crossing accident/incident data from the Federal Railway Administration (FRA); and (3) the highway-rail grade crossing inventory from FRA. The data include all reported crashes/incidents and their associated information, current and historical crossing inventory information for each crossing, and geometric features relative to the connecting highway and railways during the analysis period in North Dakota.

The final database contains information for 3,310 crossings, including 475 crash records and 2,835 no-crash records (total 3,310 records) for ND public HRGCs from 1989 to 2018 with three crash severity levels: PDO, injury, and fatality. Table 1 summarizes all the variables used in the study. Most records experienced no crash (86%) and the proportions of crash severities with PDO, injury, and fatality are 8%, 4%, and 2%, respectively. The key variables inputs are determined based on data availability and intuitive judgement, and the model will provide key variables with their significance based on model's statistical significance test. More detailed variables information associated with all crossings with or without crash records over 30-year period can be found in Table 1. In this study, not only is the crash information updated through 30-year analysis period for each crossing but also all contributors' inventory values for 30 years are kept to account for changes, so range, mode, and mean values of a variable are provided in Table 1.

4.2. Estimated Coefficient and Hazard Ratio Interpretation. Table 2 represents estimated contributors' coefficient (Coe) and hazard ratio (HR) for each severity level (PDO, injury, and fatality) and crash frequency. The estimated coefficient (β_k) and hazard ratio ($\exp(\beta_k^{\top} Z)$) are estimated based on equations (2) and (3). For a categorical variable, HR indicates the relative risk for a crossing with that contributor's-value level compared to the reference level. For a continuous contributor, HR is interpreted as the relative independent risk associated with a one-unit change in variable [73]. The regression coefficient from a cause-specific hazard model indicates the magnitude of the corresponding change in the cause-specific hazard function associated with a one-unit change in the contributor, while HR indicates the magnitude of the corresponding change in crash likelihood. As indicated in Table 2, positive Coe of 0.6 and HR of 1.82 indicate an 82% increase in PDO crash likelihood associated with crossings with passenger train service compared to freight train service. Negative Coe of −0.8 and HR of 0.45 indicate a

Table 1: Descriptive statistics of key variables.

Variable	Values	Frequency	Mean or percentage	Min	Max	Mode
Crash information	No crash	2835	86%			
	PDO	261	8%			
	Injury	147	4%			
	Fatality	67	2%			
Type of train service	Freight	2855	86%			
	Intercity passenger	455	14%			
Train detection	None	2450	74%			
	Constant warning time	438	13%			
	Motion detection	42	1%			
	DC	379	11%			
Is commercial power available?	Yes	2202	67%			
	No	1108	33%			
Is roadway/pathway paved?	Yes	625	19%			
	No	2685	81%			
Total daylight through trains			7.26	0	35	5
Total night time through trains			6.42	0	33	4.75
Total switching trains			0.77	0	12	0.55
Maximum train speed			41.02	5	79	37
Annual average daily traffic (AADT)			360.16	5	23438	45
Percent trucks			14.06	1	22.67	15.44
Distance to nearby roadway intersection			199.28	0.775	2502	27.04
Smallest crossing angle			68.37	7.869	90	90
Number of traffic lanes			1.812	1	4	2

Note. One important influential variable, warning device type, is not included in the model mainly due to unreliable data quality: (1) historical warning device types for each crossing for 30 years are not readily available and (2) inventory data contains some quality issues and results of warning device type could not be cross-validated. This could be the reason why not many long-term studies exist in the literature and some counterintuitive countermeasure effectiveness results in the literature [72]. Moreover, the focus of the study is to demonstrate the model's capability to model crash and severity likelihoods and its interpretive capability to provide contributors' long-term and instantaneous effects; contributors can be easily included in the model when they become available.

55% decrease in PDO crash likelihood associated with crossings intersected with an unpaved highway compared to a paved highway. The numerical value of HR can be any positive number with an HR of 1, indicating lack of association (likelihood of change is no different than zero), an HR greater than 1, suggesting an increased risk, and an HR less than 1, indicating a reduced risk.

According to Table 2, in general, intercity passenger train service (compared to freight train service), constant warning time (CWT) train detection system (compared to no train detection system), daytime train traffic, train speed, roadway traffic, and number of roadway lanes have positive impacts on crash likelihood while nighttime train traffic, direct current (DC) train detection system (compared to no train detection system), no commercial power available (compared to commercial power being available), and truck percentage have negative crash frequency impacts. Traffic exposure factors, such as daytime train traffic, roadway traffic, and train speed, all have positive impact on the likelihood of HRGC crashes, which meet expectations and agree with previous research results. However, nighttime traffic's negative impact on the likelihood of HRGC crashes is an interesting finding in the research. The potential rationales need further research investigation. This negative impact may be caused by the operating changes. More nighttime idling trains switching to nighttime operating trains would reduce the daytime train traffic and reduce the conflict probability at grade crossings since, in general, roadway traffic is concentrated during the daytime.

One can see from Table 2 that some factors have significant impact on certain crash severity likelihoods but not on others. These results can be underestimated because of the independent censoring assumption. However, according to equation (5), CIF is estimated without the assumption of competing risks independency. Consequently, calculated competing events marginal likelihood indicates the competing events' dependency and is more accurate. Table 1 indicates that one factor has significant impact on all three severity likelihoods: night train traffic. Night train traffic has a negative impact on PDO and injury crashes but has a positive impact on fatality crashes. The potential rationale for the positive impact on fatality crashes and negative impact on PDO and injury crashes could be because nighttime drivers are less aware of the environment and existence of a HRGC because of reduced visibilities. In addition, nighttime drivers are more likely to operate their

Table 2: Estimated coefficient and hazard ratio.

Variable	PDO		Injury		Fatality		Crash	
	Coef	HR (CI: 95%)	Coef	HR (CI: 95%)	Coef	HR (CI: 95%)	Coef	HR (CI: 95%)
Type of train service (reference: freight)								
Intercity passenger	0.6**	1.82 (1, 3)	−0.2	0.82 (0.4, 2)	0.7	2.01 (0.8, 5)	0.4**	1.50 (1, 2)
Train detection (reference: none)								
CWT	0.2	1.22 (0.7, 2)	0.5**	1.65 (1, 3)	1***	2.72 (1, 8)	0.4***	1.50 (1.2, 2)
DC	−0.7***	0.50 (0.3, 0.8)	−0.6	0.55 (0.2, 1)	−2.0	0.14 (0.03, 2)	−0.7***	0.5 (0.3, 0.7)
Is commercial power available? (reference: yes)								
No	−0.1	0.90 (0.6, 1)	−0.7***	0.50 (0.3, 0.8)	0.3	1.35 (0.80, 2)	−0.2*	0.80 (0.6, 1)
Is roadway/pathway paved? (reference: yes)								
No	−0.8***	0.44 (0.3, 0.7	−0.4	0.67 (0.4, 1)	−0.2	0.82 (0.4, 2)	−0.6***	0.50 (0.4, 0.7)
Total daytime through trains	0.2***	1.22 (1, 1)	0.1	1.11 (1, 1)	−0.4	0.67 (0.4, 1)	0.2***	1.22 (1.1, 1.3)
Total nighttime through trains	−0.2***	0.82 (0.7, 0.9)	−0.2*	0.82 (0.7, 1)	0.5*	1.65 (0.9, 3)	−0.1***	0.90 (0.8, 1)
Total switching trains	0.01	1.01 (0.8, 1)	0.03	1.03 (0.8, 1)	0.5**	1.65 (1, 3)	0.03	1.03 (0.9, 1.2)
Maximum train speed	0.004	1 (1, 1)	0.05**	1.05 (1, 1)	0.03	1.03 (1, 1)	0.02***	1 (1, 1)
Annual average daily traffic	0.00009**	1 (1, 1)	0.00004	1 (1, 1)	0.0001	1 (1, 1)	0.00008***	1 (1, 1)
Percent trucks	−0.08**	0.92 (0.9, 1)	−0.1***	0.90 (0.8, 1)	0.04	1.04 (0.9, 1)	−0.09***	0.91 (0.9, 1)
Distance to nearby intersecting roadway	−0.001**	1 (1, 1)	0.0002	1 (1, 1)	0.0005	1 (1, 1)	−0.0003	1 (1, 1)
Smallest crossing angle	−0.002	1 (1, 1)	−0.01**	1 (1, 1)	0.006	1 (1, 1)	−0.004	1 (1, 1)
Number of traffic lanes	0.4*	1.50 (1, 2)	0.1	1.11 (0.7, 2)	−0.01	1 (0.4, 2)	0.3**	1.35 (1.1, 1.6)

*Significant at a 90% confidence level. **Significant at a 95% confidence level. ***Significant at a 99% confidence level.

vehicles at higher speeds; thus more severe crashes could result from the increase in nighttime train traffic [40], making it less likely that nighttime crashes will be PDO or injury-only crashes.

It is found that, with a one-unit increase in train speed, crash likelihood increased by 5% for injury crashes and by 3% for fatality crashes but the likelihood of a change in PDO and overall crashes is not significantly different than zero.

Another interesting finding from the research is that the distance between the crossings to the nearest intersection has negative impact on the likelihood of PDO crashes. HRGCs with longer road distances to the nearest intersections are less likely to have PDO crashes, which is expected. This is because the longer distance provides a larger vehicle storage capability, which reduces the potential impact of roadway intersection operations and reduces potential PDO crashes. However, this longer distance has no significant impact on more severe crashes such as injury or fatality crashes.

Table 3 summarizes the importance ranking information based on the cause-specific HR. The instantaneous risk changes associated with HR and contributors are defined as "% impact," and the contributors are ranked based on this value. Table 3 indicates that "train service" has the highest impact on PDO, and "train detection" has the highest impact on injury crash, fatality crash, and crash occurrence likelihood.

As indicated earlier, the hazard ratio reflects critical information regarding the contributor's influence on instantaneous crash and severity probabilities. However, the significance of the contributors can be underestimated because of the independent censoring assumption. For example, the effect for one risk such as a PDO crash may reflect the influence of competing risks such as in injury crash or a fatality crash. To truly understand contributors' effects on hazard ratio and long-term crash probabilities when considering competing characteristics as in this case, a crossing that failed by PDO could have potentially failed by injury or fatal crash too. To capture the unobserved dependence among failure types, the cumulative-incidence-based effect analysis should be conducted.

4.3. Cumulative Probability over Time. Hazard ratio is a direct isolated influential indicator to a specific failure event. The isolated influential effects do not consider the same contributor's effects on other competing failure events. In other words, they tend to underestimate the contributor's impact when analyzing the marginal probability for cause-specific events considering the competing nature of multiple causes to the same event. Such analysis results can be very sensitive to different modeling techniques and provide very different contributing effects [74]. One of the advantages that the competing risks model can provide is to evaluate contributors' long-term robust influence. Previous research [75] verified and indicated that a covariate has no effect on a

Table 3: Ranking variables based on hazard ratio.

Variable	PDO		Injury		Fatality		Crash	
	Rank	% impact	Rank	% impact	Rank	% impact	Rank	% impact
Type of train service:								
Intercity passenger	1	82%	5	18%	2	101%	2	49%
Train detection:								
CWT	5	22%	1	65%	1	172%	2	49%
DC	3	50%	3	45%	3	86%	1	50%
Is commercial power available?								
No	8	10%	2	50%	6	35%	6	18%
Is roadway/pathway paved?								
No	2	55%	4	33%	8	18%	3	45%
Total daytime through trains	6	22%	7	11%	7	33%	5	22%
Total nighttime through trains	7	18%	6	18%	4	65%	7	10%
Total switching trains	10	1%	11	3%	5	65%	9	3%
Maximum train speed	11	0.4%	10	5%	10	3%	10	2%
Annual average daily traffic	14	0.01%	14	0.004%	14	0.01%	13	0.01%
Percent trucks	9	8%	9	10%	9	4%	8	9%
Distance to nearby intersecting roadway	13	0.1%	13	0.02%	13	0.05%	12	0.03%
Smallest crossing angle	12	0.2%	12	1%	12	1%	11	0.4%
Number of traffic lanes	4	49%	8	11%	11	1%	4	35%

Note. % impact = $|\mathrm{HR} - 1| \times 100$.

competing-event failure risk based on the cause-specific hazard analysis, but it still can indicate a significant impact on the cumulative risk probability of the competing event. A contributor, which has no direct influence on one type of failure event, may still be significantly associated with the cumulative probability (incidence) of that failure event, if that contributor has influences on a competing failure event. The estimated marginal probability of a certain failure event can be calculated with its cause-specific probability and the overall cumulative survival probability. Estimating cumulative probability of the failure events as the cumulative incidence function with equation (9) depends on HR for both the event of interest (crash occurrence) and the competing events (PDO, injury, and fatality crashes).

In the following study, "train service" and "train detection" are selected to perform the cumulative probability analysis due to the limited journal paper space and the facts that they are ranked as the top impact factors for PDO, injury, fatality crashes and crash occurrence likelihood. Moreover, the two-sample t-test method is used to quantify the significance of contributors' effects on cumulative probability for crash occurrence and each severity level.

Predicted cumulative crash probabilities of each crash severity are estimated base on equations (8) and (9). To compare the change in predicted cumulative probability by changing train service types, two subsamples are created: one sample with freight train service only and the other with passenger train service only. The rest of the contributors' values are controlled at a fixed level, mode value. Figure 2 indicates the 30-year predictions of cumulative crash severity and crash occurrence probabilities for "train service" in parts *a*, *b*, *c*, and *d*, respectively. Figure 3 describes the cumulative probability analysis results for train detection systems.

From Figure 2, one can see the following: (1) Cumulative crash probabilities are increasing over time at different rates under various conditions. (2) In general, crossings with intercity freight train service have higher increasing rate of cumulative probabilities for all risks except for injury compared to the crossings with freight train services. (3) The increase in cumulative crash probabilities is faster for PDO crash occurrence than for injury and fatality crashes for both types of train service. (4) The increasing rate differs significantly between the two types of train service for all crash risks except for injury. (5) According to part *c*, even the absolute magnitude of the increasing rate is relatively small between the two services but the proportion of increased fatal probability is almost doubled for crossings with passenger train services compared to crossings with freight train services.

From Figure 3, one can see that crossings with DC systems have reduced crash likelihood on all crashes and all severities, but crossings with CWT systems have higher values compared to those with no detection systems. The absolute probability differences are all very small and around 0.1%. The fatal crash probability is more than doubled for crossings with CWT detection system compared to crossings with no detection system, 224% over 30 years. Note that crossings with a DC train detection system installed are 79% less likely in cumulative fatal probability compared to the crossings with no automatic detection system.

The results for CWT may sound counterintuitive. Detailed safety countermeasure effectiveness with regard to train detection systems should be conducted in the future with actual before- and after-installation data. Train detection warning systems are designed to warn crossing users of an approaching train through certain automatic train detection methods. For the DC system, current flows from a battery through a fixed rail segment to the coil of a relay. The locations of the battery and relay define the location of the warning-trigger rail segment.

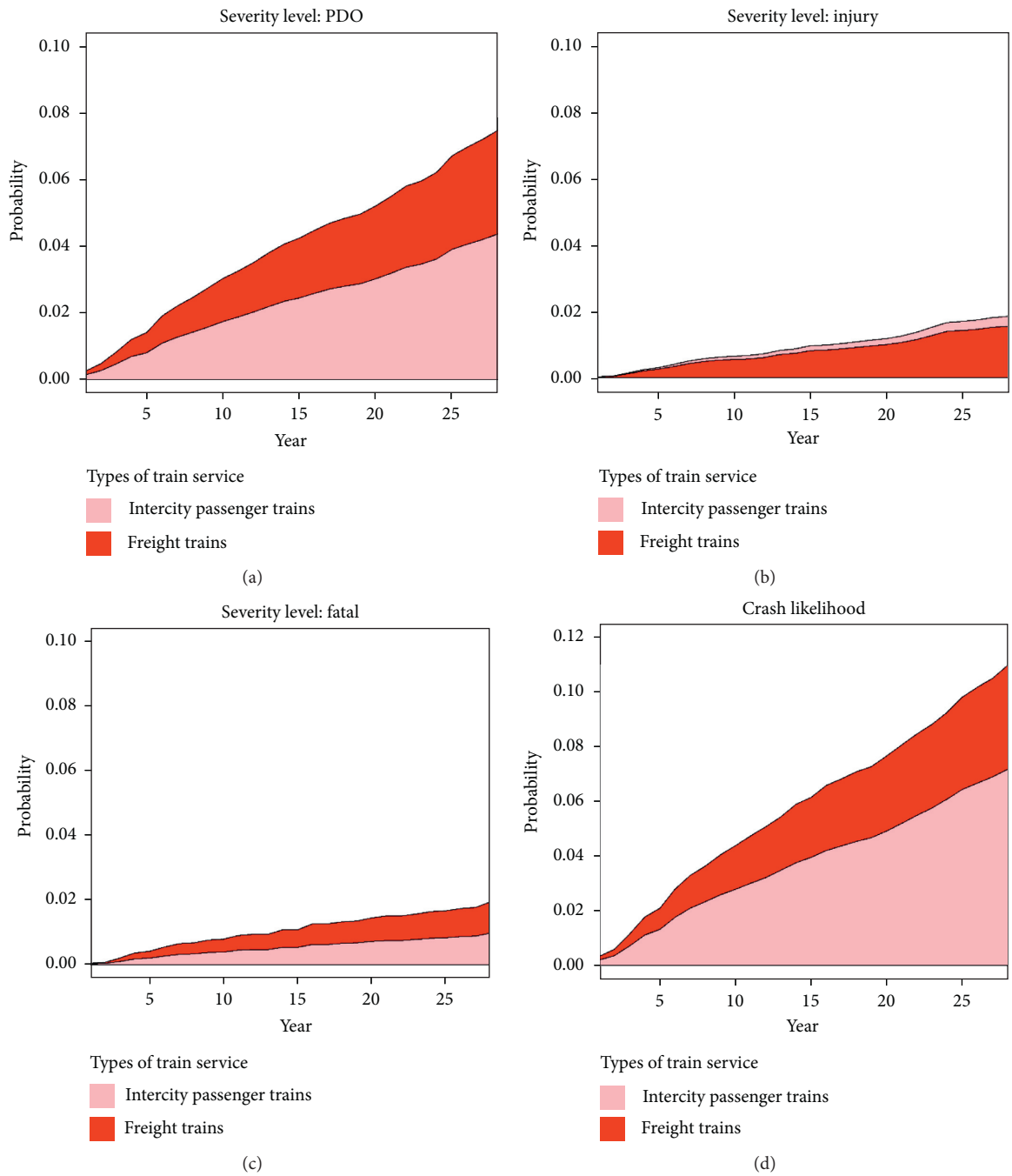


Figure 2: Cumulative probabilities of crash severity and crash occurrence for train service.

The DC method uses the rail as an energy conductor. When a train enters the segment, the axles short/shunt the circuit, which activates the crossing warning system to warn users of an approaching train. This method generates a warning based on the track occupation status, which has a fixed predefined distance from the crossing, usually between 1500 and 2000 feet. The CWT system is a smart system that is capable of determining the speed and location of an approaching train. Thus, it is able to predict when the train will arrive at the crossing. With the CWT system, a warning signal is activated intentionally to provide a constant preselected warning time, usually 25 seconds. So, for a slow-moving train, the distance between the train and the crossing could be much closer than that for a faster-moving train. However, a CWT system is not able to measure a change in speed accurately, which results in variability in the actual warning time. For example, if a CWT system predicts the warning-activation time for a slow train and then the approaching train accelerates towards to the crossing, it will result in a less-than-desired warning time. This could be why crossings with CWT systems tend to have more crash results.

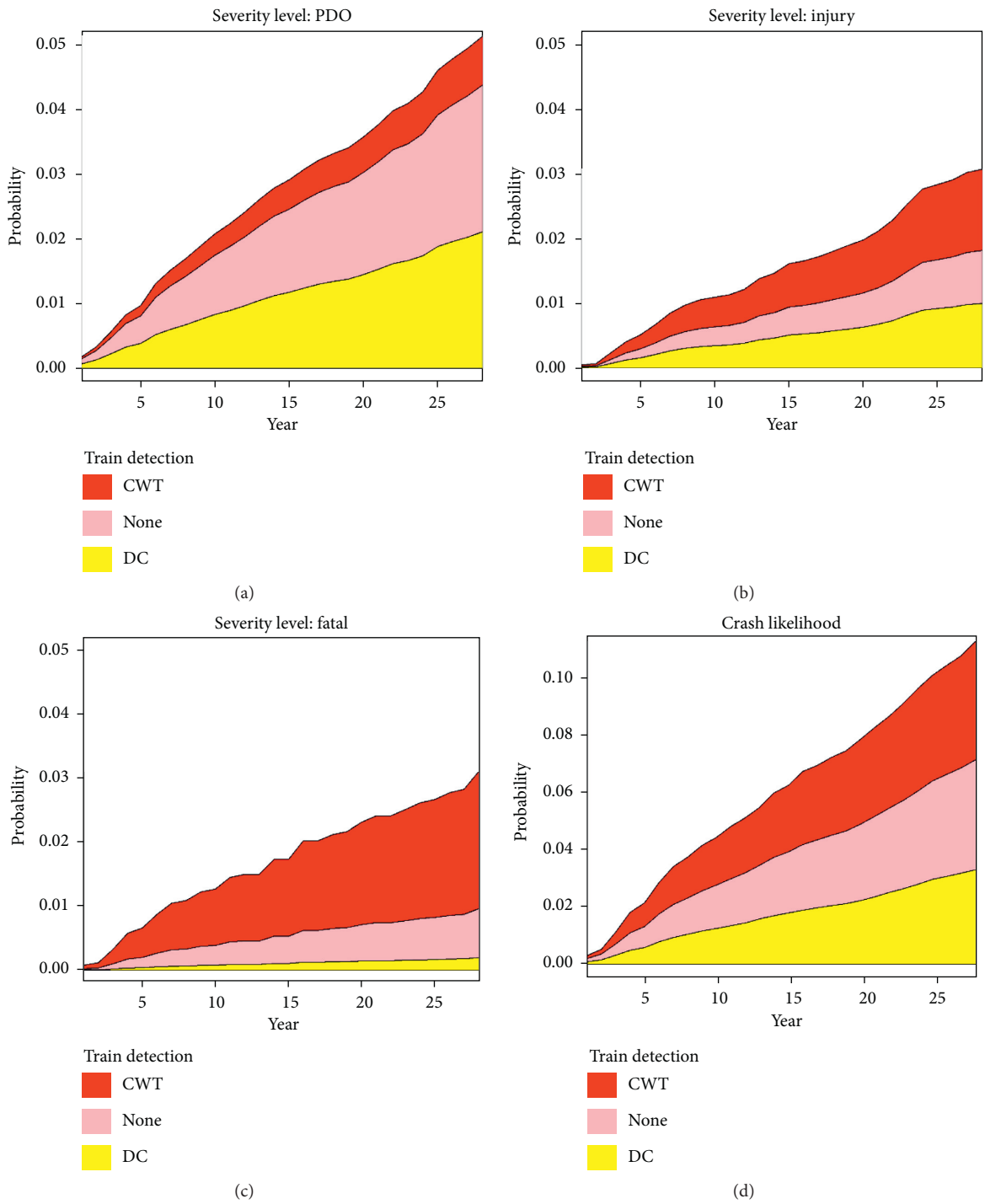


Figure 3: Cumulative probabilities of crash severity and crash occurrence for train detection.

The average annual probability increasing rates during the 30-year study period and the two-sample t-test results are summarized in Table 4. To apply the two-sample t-test, the average annual crash probability increasing rates of all records in the study dataset are estimated based on their cumulative incidence, and the test is applied for the selected categorical variable for each group level. One can see that, on average, PDO crash likelihood increased by about 0.26% each year for the crossings with passenger train service compared to 0.152% for crossings with freight train services. On average, PDO crash likelihood at crossings with passenger train service increases by about 0.108% annually, which is 71% increase compared to freight train service category. Based on the t-test, train service change significantly impacts PDO crash likelihood at a 99% significance level.

Compared to Table 2, one can also see that crossings with intercity passenger service were identified as not significant

Table 4: Annual average crash probability increasing percentages for train service and detection system.

Variables	PDO % and change		Injury % and change		Fatality % and change		Crash % and change	
Train service								
Freight	0.152	NA	0.062	NA	0.033	NA	0.247	NA
Intercity passenger	0.26	0.108*** 71%	0.052***	−0.01 −16%	0.066***	0.033 100%	0.378***	0.131 53%
Train detection								
None	0.152	NA	0.062	NA	0.033	NA	0.247	NA
CWT	0.178***	0.026 17%	0.105***	0.043 69%	0.107***	0.074 224%	0.39***	0.143 58%
DC	0.073*	−0.079 −52%	0.034***	−0.028 −45%	0.007***	−0.026 −79%	0.114***	−0.133 −54%

to instantaneous injury crash likelihood compared to crossings with freight train services without considering competing risks. However, crossings with intercity passenger service were identified to significantly impact cumulative injury probability compared to freight train services when considering competing risks. The same change is observed for CWT systems for PDO crashes and for DC systems for injury and fatality crashes. As mentioned before, estimating coefficients and HRs through cause-specific hazard function are based on the assumption of independent censoring, which estimates coefficient and HR associated with each severity level separately. However, CIF is able to indicate whether a crossing is at the risk of one of crash severity levels (e.g., injury) and also would be at risk of the other crash severity levels (e.g., PDO or fatality), which means that CIF estimation is free of any events independency assumptions. Correspondingly, the significance difference between the two tables originates from the dependence of contributors' effects on competing events.

5. Conclusions and Discussion

The competing risks model is applied to examine crash frequency and severity simultaneously at public highway-rail grade crossings in North Dakota from 1989 to 2018. The competing risks model demonstrates its capability to identify specific contributors and to simultaneously model crash occurrence and crash severity likelihoods. It is able to produce easy-to-interoperate results such as the estimated coefficients, hazard ratios, and cumulative probabilities. It also demonstrates its ability to capture the dependence of contributors' effects on competing events.

A few important findings are identified with the research:

(1) Train services, train detection system, commercial power availability, roadway surface, train traffic, roadway traffic, train speed, truck percentage, and number of road lanes are identified as having significant impact on crash occurrence likelihood.

(2) Crossings with passenger train services have higher PDO crash risk compared to crossings with freight train service, but they perform no differently in terms of instantaneous injury and fatal crash risk. Crossings with increased night train traffic are more likely to have fatality crashes and are less likely to have PDO and injury crashes. Increasing the distance between the crossing and the nearest roadway intersection will decrease PDO crash risk but will have no impact on instantaneous injury or fatality crash risk. Increasing the smallest acute crossing angle may decrease injury crash risk. However, it will not impact instantaneous PDO and fatality crash risk.

(3) Except for night train traffic, which is found to significantly impact all three crash severities, contributors are found to only have direct cause-specific impacts on certain crash severities but not on all of them.

(4) Contributors, such as a CWT train detection system for PDO crashes, may not be significant on instantaneous risks, but they can be identified as significant contributors from a cumulative probability perspective when considering dependence competing risks.

(5) Based on cumulative probabilities, the annual PDO crash probability increase is 0.108% for crossings with passenger train service compared to crossings with freight train service, which accounts for about 76% of the increase. Fatal probability increasing rate reduction is 0.033% for crossings with passenger train service compared to freight train service. However, this accounts for a 100% reduction. Crossings with DC detection systems tend to have lower crash increasing rates compared to crossings with no detection systems, but crossings with CWT detection systems tend to have higher crash increasing rates compared to those with no detection systems. Crossings with freight train service have about a 0.131% reduction in crash occurrence probability annual increasing rate compared to crossings with passenger train service.

The competing risks modeling demonstrates its capability to identify risk factors and the marginal probability for crash severity and crash occurrence simultaneously. The model is designed to identify and summarize the contributors and risk probabilities. However, the study does not suggest countermeasure effectiveness. A detailed countermeasure effectiveness analysis should be conducted with available supporting data.

Data Availability

The data that support the findings of this study are available from the authors upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest associated with this publication.

Authors' Contributions

Research idea and study concept initiation were contributed by Dr. Pan Lu and Dr. Denver Tolliver. Study design was proposed by Dr. Pan Lu, Dr. Denver Tolliver, and Amin Keramati. Data collection was performed by Amin Keramati and Dr. Pan Lu. Model construction was carried out by Amin Keramati and Dr. Pan Lu. Lab/software resource support was provided by Dr. Denver Tolliver. Analysis and interpretations of results were performed by Amin Keramati, Xiaoyi Zhou, Dr. Pan Lu, and Dr. Denver Tolliver. Draft manuscript preparation was executed by Amin Keramati, Dr. Pan Lu, and Xiaoyi Zhou. All the authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

The authors express their gratitude to North Dakota State University and the Mountain-Plains Consortium (MPC), a University Transportation Center funded by the U.S. Department of Transportation, for their support.

References

[1] Q. Cai, M. Abdel-Aty, and J. Lee, "Macro-level vulnerable road users crash analysis: a bayesian joint modeling approach of frequency and proportion," *Accident Analysis & Prevention*, vol. 107, pp. 11–19, 2017.

[2] K. Geurts, I. Thomas, and G. Wets, "Understanding spatial concentrations of road accidents using frequent item sets," *Accident Analysis & Prevention*, vol. 37, no. 4, pp. 787–799, 2005.

[3] Y. Hao, L. Xu, B. Qi, T. Wang, and W. Zhao, "A machine learning approach for highway intersection risk caused by harmful lane-changing behaviors," in *Proceedings o the CICTP 2019*, pp. 5623–5635, Nanjing, China, July 2019.

[4] H. Huang, H. Zhou, J. Wang, F. Chang, and M. Ma, "A multivariate spatial model of crash frequency by transportation modes for urban intersections," *Analytic Methods in Accident Research*, vol. 14, pp. 10–21, 2017.

[5] S. Islam and J. Brown, "A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama," *Accident Analysis & Prevention*, vol. 108, pp. 163–171, 2017.

[6] S. Kumar, D. Toshniwal, and M. Parida, "A comparative analysis of heterogeneity in road accident data using data mining techniques," *Evolving Systems*, vol. 8, no. 2, pp. 147–155, 2017.

[7] J. Lee, M. Abdel-Aty, and Q. Cai, "Intersection crash prediction modeling with macro-level data from various geographic units," *Accident Analysis & Prevention*, vol. 102, pp. 213–226, 2017.

[8] L. Li, S. Shrestha, and G. Hu, "Analysis of road traffic fatal accidents using data mining techniques," in *Proceedings of the 2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 363–370, IEEE, London, UK, June 2017.

[9] M. Paul, "Safety assessment at unsignalized intersections using post-encroachment time's threshold—a sustainable solution for developing countries," in *Lecture Notes in Civil Engineering*, pp. 117–131, Springer, Berlin, Germany, 2019.

[10] X. Qin, J. N. Ivan, and N. Ravishanker, "Selecting exposure measures in crash rate prediction for two-lane highway segments," *Accident Analysis & Prevention*, vol. 36, no. 2, pp. 183–191, 2004.

[11] M. B. Ulak, E. E. Ozguven, R. Moses et al., "Assessment of traffic performance measures and safety based on driver age and experience: a microsimulation based analysis for an unsignalized t-intersection," *Journal of Traffic and Transportation Engineering*, vol. 6, no. 5, 2019.

[12] N. Veeramisti, A. Paz, M. Khadka, and C. Arteaga, "A clusterwise regression approach for the estimation of crash frequencies," *Journal of Transportation Safety & Security*, pp. 1–31, 2019.

[13] X. Wang and M. Abdel-Aty, "Temporal and spatial analyses of rear-end crashes at signalized intersections," *Accident Analysis & Prevention*, vol. 38, no. 6, pp. 1137–1150, 2006.

[14] Z. Zheng, P. Lu, and B. Lantz, "Commercial truck crash injury severity analysis using gradient boosting data mining model," *Journal of Safety Research*, vol. 65, pp. 115–124, 2018.

[15] H. Cho and L. R. Rilett, "Modeling signalized intersections near highway-railroad grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1973, no. 1, pp. 149–156, 2006.

[16] H. Ghomi, M. Bagheri, L. Fu, and L. F. Miranda-Moreno, "Analyzing injury severity factors at highway railway grade crossing accidents involving vulnerable road users: a comparative study," *Traffic Injury Prevention*, vol. 17, no. 8, pp. 833–841, 2016.

[17] K. Haleem, "Investigating risk factors of traffic casualties at private highway-railroad grade crossings in the United States," *Accident Analysis & Prevention*, vol. 95, pp. 274–283, 2016.

[18] A. Khattak, M. Gao, and Z. Luo, "The impact of removing centerline barriers from highway-rail grade crossings on drivers' violations," *Journal of Transportation Safety & Security*, vol. 4, no. 4, pp. 295–307, 2012.

[19] P. Lu and D. Tolliver, "Accident prediction model for public highway-rail grade crossings," *Accident Analysis & Prevention*, vol. 90, pp. 73–81, 2016.

[20] L.-W. Tung and A. Khattak, "Distracted motor vehicle driving at highway-rail grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2476, no. 1, pp. 77–84, 2015.

[21] H. Yue and E. Jones, "Archiving capability of spatio-temporal data in different highway railroad grade crossing (HRGC) databases," in *Proceedings of the Annual Intelligent Transportation System Conference*, Madeira, Portugal, September 2010.

[22] S. Zhao and A. J. Khattak, "Factors associated with self-reported inattentive driving at highway-rail grade crossings," *Accident Analysis & Prevention*, vol. 109, pp. 113–122, 2017.

[23] Z. Zheng, P. Lu, and D. Pan, "Predicting highway-rail grade crossing collision risk by neural network systems," *Journal of*

Transportation Engineering, Part A: Systems, vol. 145, no. 8, Article ID 4019033, 2019.

[24] Z. Zheng, P. Lu, and D. Tolliver, "Decision tree approach to accident prediction for highway-rail grade crossings: empirical analysis," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2545, no. 1, pp. 115–122, 2016.

[25] R. D. Austin and J. Carson, "An alternative accident prediction model for highway-rail interfaces," *Accident Analysis & Prevention*, vol. 34, no. 1, pp. 31–42, 2002.

[26] R. Guadamuz-Flores and J. Aguero-Valverde, "Bayesian spatial models of crash frequency at highway-railway crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2608, no. 1, pp. 27–35, 2017.

[27] S. Heydari, L. Fu, L. Thakali, and L. Joseph, "Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: applications to grade crossing safety analysis," *Analytic Methods in Accident Research*, vol. 19, pp. 33–48, 2018.

[28] S. Heydari and L. Fu, "Developing safety performance functions for railway grade crossings: a case study of Canada," in *Proceeings of the 2015 Joint Rail Conference American Society of Mechanical Engineers (ASME)*, Article ID V001T06A017, San Jose, CA, USA, March 2015.

[29] S.-R. Hu, C.-S. Li, and C.-K. Lee, "Model crash frequency at highway-railroad grade crossings using negative binomial regression," *Journal of the Chinese Institute of Engineers*, vol. 35, no. 7, pp. 841–852, 2012.

[30] S. Hu and J. Lin, "Effect of time to train arrive on crash frequency at highway-railroad grade crossings: a general classification regression model," in *Proceedings of the 91st Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 2012.

[31] A. Khattak, A. Sharma, and Z. Luo, "Implications of using annual average daily traffic in highway-rail grade crossing safety models," 2012.

[32] A. Khattak and Z. Luo, "Pedestrian and bicyclist violations at highway-rail grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2250, no. 1, pp. 76–82, 2011.

[33] J. Lee, D. Nam, and D. Moon, "A zero-inflated accident frequency model of highway-rail grade crossing," in *Proceedings of the Transportation Research Board Annual Meeting*, Washington, DC, USA, January 2004.

[34] J. C. Medina and R. F. Benekohal, "Macroscopic models for accident prediction at railroad grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2476, no. 1, pp. 85–93, 2015.

[35] H. Millegan, X. Yan, S. Richards, and L. Han, "Evaluation of effectiveness of stop-sign treatment at highway-railroad grade crossings," *Journal of Transportation Safety & Security*, vol. 1, no. 1, pp. 46–60, 2009.

[36] J. Oh, S. P. Washington, and D. Nam, "Accident prediction model for railway-highway interfaces," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 346–356, 2006.

[37] F. F. Saccomanno, P. Y.-J. Park, and L. Fu, "Estimating countermeasure effects for reducing collisions at highway-railway grade crossings," *Accident Analysis & Prevention*, vol. 39, no. 2, pp. 406–416, 2007.

[38] F. F. Saccomanno and X. Lai, "A model for evaluating countermeasures at highway-railway grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1918, no. 1, pp. 18–25, 2005.

[39] X. Yan, S. Richards, and X. Su, "Using hierarchical tree-based regression model to predict train-vehicle crashes at passive highway-rail grade crossings," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 64–74, 2010.

[40] N. Eluru, M. Bagheri, L. F. Miranda-Moreno, and L. Fu, "A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings," *Accident Analysis & Prevention*, vol. 47, pp. 119–127, 2012.

[41] W. Fan, M. R. Kane, and E. Haile, "Analyzing severity of vehicle crashes at highway-rail grade crossings: multinomial logit modeling," *Journal of the Transportation Research Forum*, vol. 54, pp. 39–56, 2015.

[42] K. Haleem and A. Gan, "Contributing factors of crash injury severity at public highway-railroad grade crossings in the U.S," *Journal of Safety Research*, vol. 53, pp. 23–29, 2015.

[43] W. Hao and J. Daniel, "Driver injury severity related to inclement weather at highway-rail grade crossings in the United States," *Traffic Injury Prevention*, vol. 17, no. 1, pp. 31–38, 2016.

[44] W. Hao and J. Daniel, "Motor vehicle driver injury severity study under various traffic control at highway-rail grade crossings in the United States," *Journal of Safety Research*, vol. 51, pp. 41–48, 2014.

[45] W. Hao and J. R. Daniel, "Severity of injuries to motor vehicle drivers at highway-rail grade crossings in the United States," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2384, no. 1, pp. 102–108, 2013.

[46] S.-R. Hu, C.-S. Li, and C.-K. Lee, "Investigation of key factors for accident severity at railroad grade crossings by using a logit model," *Safety Science*, vol. 48, no. 2, pp. 186–194, 2010.

[47] Y. Kang and A. Khattak, "Cluster-based approach to analyzing crash injury severity at highway-rail grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2608, no. 1, pp. 58–69, 2017.

[48] J. Liu and A. J. Khattak, "Gate-violation behavior at highway-rail grade crossings and the consequences: using geo-spatial modeling integrated with path analysis," *Accident Analysis & Prevention*, vol. 109, pp. 99–112, 2017.

[49] C. Ma, W. Hao, W. Xiang, and W. Yan, "The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossings accidents," *Journal of Advanced Transportation*, vol. 2018, Article ID 9841498, 10 pages, 2018.

[50] P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus, "The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1666–1676, 2011.

[51] S. Zhao, A. Iranitalab, and A. J. Khattak, "A clustering approach to injury severity in pedestrian-train crashes at highway-rail grade crossings," *Journal of Transportation Safety & Security*, vol. 11, no. 3, pp. 305–322, 2018.

[52] S. Zhao and A. Khattak, "Motor vehicle drivers' injuries in train-motor vehicle crashes," *Accident Analysis & Prevention*, vol. 74, pp. 162–168, 2015.

[53] Z. Ye, Y. Xu, and D. Lord, "Crash data modeling with a generalized estimator," *Accident Analysis & Prevention*, vol. 117, pp. 340–345, 2018.

[54] A. Iranitalab and A. Khattak, "Comparison of four statistical and machine learning methods for crash severity prediction," *Accident Analysis & Prevention*, vol. 108, pp. 27–36, 2017.

[55] D. Lee, J. Warner, and C. Morgan, "Discovering crash severity factors of grade crossing with a machine learning approach," in *Proceedings of the Joint Rail Conference American Society of*

Mechanical Engineers Digital Collection, Salt Lake City, UT, USA, April 2019.

[56] C. Yang, E. Trudel, and Y. Liu, "Machine learning-based methods for analyzing grade crossing safety," *Cluster Computing*, vol. 20, no. 2, pp. 1625–1635, 2017.

[57] M. Abdel-Aty and J. Keller, "Exploring the overall and specific crash severity levels at signalized intersections," *Accident Analysis & Prevention*, vol. 37, no. 3, pp. 417–425, 2005.

[58] X. Zhou, P. Lu, Z. Zheng, D. Tolliver, and A. Keramati, "Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree," *Reliability Engineering & System Safety*, vol. 200, Article ID 106931, 2020.

[59] M. Abdel-Aty and P. Nawathe, "A novel approach for signalized intersection crash classification and prediction," *Advances in Transportation Studies*, vol. 9, 2006.

[60] D. A. Zalinger, B. A. Rogers, and H. P. Johri, "Calculation of hazard indices for highway-railway crossings in Canada," *Accident Analysis & Prevention*, vol. 9, no. 4, pp. 257–273, 1977.

[61] X. Ye, R. M. Pendyala, V. Shankar, and K. C. Konduri, "A simultaneous equations model of crash frequency by severity level for freeway sections," *Accident Analysis & Prevention*, vol. 57, pp. 140–149, 2013.

[62] P. K. Andersen, L. S. Hansen, and N. Keiding, "Non-and semi-parametric estimation of transition probabilities from censored observation of a non-homogeneous markov process," *Scandinavian Journal of Statistics*, vol. 18, pp. 153–167, 1991.

[63] M. Fiocco, H. Putter, C. J. H. Van De Velde, and J. C. Van Houwelingen, "Reduced rank proportional hazards model for competing risks: an application to a breast cancer trial," *Journal of Statistical Planning and Inference*, vol. 136, no. 5, pp. 1655–1668, 2006.

[64] M. Fiocco, H. Putter, and J. C. Van Houwelingen, "Reduced rank proportional hazards model for competing risks," *Biostatistics*, vol. 6, no. 3, pp. 465–478, 2005.

[65] R. B. Geskus, "On the inclusion of prevalent cases in hiv/aids natural history studies through a marker-based estimate of time since seroconversion," *Statistics in Medicine*, vol. 19, no. 13, pp. 1753–1769, 2000.

[66] R. B. Geskus, F. A. Miedema, J. Goudsmit, P. Reiss, H. Schuitemaker, and R. A. Coutinho, "Prediction of residual time to aids and death based on markers and cofactors," *JAIDS Journal of Acquired Immune Deficiency Syndromes*, vol. 32, no. 5, pp. 514–521, 2003.

[67] T. A. Gooley, W. Leisenring, J. Crowley, and B. E. Storer, "Estimation of failure probabilities in the presence of competing risks: new representations of old estimators," *Statistics in Medicine*, vol. 18, no. 6, pp. 695–706, 1999.

[68] R. P. Van Rij, A.-M. De Roda Husman, M. Brouwer, J. Goudsmit, R. A. Coutinho, and H. Schuitemaker, "Role of ccr2 genotype in the clinical course of syncytium-inducing (si) or non-si human immunodeficiency virus type 1 infection and in the time to conversion to si virus variants," *The Journal of Infectious Diseases*, vol. 178, no. 6, pp. 1806–1811, 1998.

[69] J. Liu, A. J. Khattak, S. H. Richards, and S. Nambisan, "What are the differences in driver injury outcomes at highway-rail grade crossings? untangling the role of pre-crash behaviors," *Accident Analysis & Prevention*, vol. 85, pp. 157–169, 2015.

[70] H. Putter, M. Fiocco, and R. B. Geskus, "Tutorial in biostatistics: competing risks and multi-state models," *Statistics in Medicine*, vol. 26, no. 11, pp. 2389–2430, 2007.

[71] L. C. De Wreede, M. Fiocco, and H. Putter, "The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models," *Computer Methods and Programs in Biomedicine*, vol. 99, no. 3, pp. 261–274, 2010.

[72] E. H. Farr, "Summary of DOT Rail-Highway Crossing Resource Allocation Procedure-Revised," Federal Railroad Administration, Washington, DC, USA, 1987.

[73] B. R. Logan, M.-J. Zhang, and J. P. Klein, "Regression models for hazard rates versus cumulative incidence probabilities in hematopoietic cell transplantation data," *Biology of Blood and Marrow Transplantation*, vol. 12, no. 1, pp. 107–112, 2006.

[74] J. J. Dignam, Q. Zhang, and M. Kocherginsky, "The use and interpretation of competing risks regression models," *Clinical Cancer Research*, vol. 18, no. 8, pp. 2301–2308, 2012.

[75] M. Wolbers, M. T. Koller, V. S. Stel et al., "Competing risks analyses: objectives and approaches," *European Heart Journal*, vol. 35, no. 42, pp. 2936–2941, 2014.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 3463287, 15 pages
https://doi.org/10.1155/2020/3463287

Research Article

Examining the Impact of Different Periodic Functions on Short-Term Freeway Travel Time Prediction Approaches

Xu Miao,[1] Bing Wu,[1] Yajie Zou,[1] and Lingtao Wu[2]

[1]*The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China*
[2]*Texas A&M Transportation Institute, Texas A&M University System, College Station, TX 77843-3135, USA*

Correspondence should be addressed to Yajie Zou; yajiezou@hotmail.com

Received 1 December 2019; Revised 8 July 2020; Accepted 11 July 2020; Published 1 August 2020

Academic Editor: Zhigang Xu

Freeway travel time prediction is a key technology of Intelligent Transportation Systems (ITS). Many scholars have found that periodic function plays a positive role in improving the prediction accuracy of travel time prediction models. However, very few studies have comprehensively evaluated the impacts of different periodic functions on statistical and machine learning models. In this paper, our primary objective is to evaluate the performance of the six commonly used multistep ahead travel time prediction models (three statistical models and three machine learning models). In addition, we compared the impacts of three periodic functions on multistep ahead travel time prediction for different temporal scales (5-minute, 10-minute, and 15-minute). The results indicate that the periodic functions can improve the prediction performance of machine learning models for more than 60 minutes ahead prediction and improve the over 30 minutes ahead prediction accuracy for statistical models. Three periodic functions show a slight difference in improving the prediction accuracy of the six prediction models. For the same prediction step, the effect of the periodic function is more obvious at a higher level of aggregation.

1. Introduction

Travel time can effectively measure roadway traffic conditions [1]. Thus, accurate prediction of freeway travel time is important for traffic management agencies to provide better traffic guidance. However, it is challenging for researchers to predict travel time accurately due to the complex changes in traffic states [2]. A large number of algorithms have been proposed to improve the prediction accuracy of travel time. The existing short-term traffic forecasting algorithms were reviewed by Vlahogianni et al. [3] and Vlahogianni et al. [1]. Existing short-term traffic forecasting algorithms can be categorized into two major strands: statistical models and machine learning models. Linear regression analysis method [4, 5], time series method [6–8], and space time prediction methods [9, 10] are statistical models. Kalman filtering model [2, 11, 12], support vector regression model [13–15], and neural network model [16–20] are machine learning models. A series of combination models [21–25] are proposed in recent years.

Some researchers compared the performance of statistical models and machine learning models. For example, Stathopoulos et al. [26] found that fuzzy neural network outperformed Autoregressive Integrated Moving Average Model (ARIMA) in prediction performance. Vlahogianni [27] suggested that the advanced Neural Network (NN) structure can perform better than the ARIMA model. Jiang et al. [28] examined the prediction performance of different models under multiple steps ahead, and their results indicated that the machine learning models are superior to the two statistical models (i.e., vector autoregressive models and ARIMA).

Traffic data usually exhibit periodic characteristics during weekdays. Thus, considering the periodicity of data can improve the prediction performance. Up to date, three different approaches have been proposed to capture the periodic characteristics. Zou et al. [29] found that a synthetic prediction model consisting of statistical models and trigonometric polynomial function (TPF) can achieve higher prediction accuracy when the forecasting horizon is greater

than half hour with 5 minutes as the aggregation level. Tang et al. [30] applied a double exponential smoothing method (DES) to describe the weekly similarities of traffic data. In the course of the study, Chen et al. [31] utilized the prediction model in accordance with the original traffic flow series compared with the intraday trend removed the by simple average (SA) approach. It is found that the accuracy of the prediction could be considerably improved by using the residual time series.

Regarding the prediction interval (steps), some existing studies have investigated the impact of data resolution on model prediction performance, but there are no definitive results. For example, Park et al. [32] considered the aggregation level from 2 minutes to 60 minutes of the ARIMA model based on travel time data. They concluded that forecasting route travel time required higher concentration levels than link travel time prediction. Vlahogianni et al. [33] found that time clustering may distort critical traffic flow information, and we need further research to determine the optimal concentration level. Some studies found that higher data resolution usually shows larger noise [34, 35].

Based on the previous studies, some studies have compared statistical models and machine learning models, and some scholars have proposed the improvement of periodic functions on travel time prediction. However, few studies have comprehensively evaluated the effects of different periodic functions on the two types of models under different prediction steps. Thus, this study focuses on multistep ahead travel time prediction by considering different periodic functions. The periodic characteristics of the travel time are captured by SA, TPF, and DES models. The residual part is modeled by the statistical models (ARIMA, space time (ST) model, vector autoregressive (VAR) model) and machine learning models(support vector machine (SVM), back propagation neural network (BPNN), multilinear regression (MLR)). In total, 18 hybrid prediction models were established and compared. In addition, the performance of prediction models was evaluated under different scenarios: multistep ahead prediction (1, 3, 6, and 12 steps ahead predictions) with different aggregation levels (5-minute, 10-minute, and 15-minute).

The remainder of the paper is organized as follows. In Section 2, we introduce the travel time data in the study. We describe the data collection site and analyze the temporal and spatial correlation as well as the diurnal pattern observed in the data. In Section 3, we introduce periodic functions and two main methodologies used in this study: statistical models and machine learning approaches. We also discuss the evaluation measures and determine the appropriate training periods. In Section 4, we evaluate the prediction performance of the six models and compare the impacts of different periodic functions on prediction models under different scenarios. In Section 5, we provide the conclusions and some future works.

2. Travel Time Data

This study analyzed the travel time data of US-290 between IH-610 and FM-1960 in Houston, Texas. The total length is approximately 12 miles. The segment is divided into five links by six automatic vehicle identification (AVI) readers (Figure 1). Vehicles with toll tags passed through the AVI readers will be recorded with their ID and timestamps. Travel time of the link enclosed by this pair of AVI readers is the difference in the timestamps. The length of link A to link E is 0.8, 2.6, 3.0, 1.5, and 4.1 miles, respectively. The data collection duration is from January 2008 to August 2008, a total of 174 days. The travel times were initially collected once every 30 seconds, 24 hours per day. We calculated the arithmetic mean of travel time and aggregated the travel time into 5-minute, 10-minute, and 15-minute intervals for each link. The missing data for the five links are all less than 1%, and historical averaged based data imputation method have been implemented to ensure the selected travel time data are appropriate for model validation and evaluation in this study. This study only focuses on the weekday (Monday-Friday) travel time prediction.

2.1. Temporal and Spatial Correlation of the Travel Time. We calculate the historical average travel time per mile of the five links (Monday to Friday, January to August 2008) (Figure 2). It can be found that the peak time of traffic occurs in the afternoons of all these five links, and there are mainly three types of travel time patterns. For link A, travel time increases after 12:00, peaks at about 16:30, and finishes later than 20:00. For links B and C, traffic congestion starts at 12:00, peaks around 17:35, and returns to usual after 20:00. For links D and E, traffic congestion often occurs before 16:00, peaks around 17:50, and dissipates after 20:00. In our study, link D is chosen as the target link.

Changes in traffic flow have certain temporal and spatial characteristics. Autocorrelation and cross-correlation functions were calculated to examine the temporal and spatial correlation. The equation adopted here follows that of Zou et al. [23], as shown in equations (1)–(3).

$$ar_{k_a} = \frac{ac_{k_a}}{ac_0},$$
$$ac_{k_a} = \frac{1}{N_a}\sum_{t=1}^{N_a-k_a}\left(x_t - \overline{x}\right)\left(x_{(t+k_a)} - \overline{x}\right), \tag{1}$$

where ar_{k_a} is the sample autocorrelation function; k_a is the time lag; $ac_0 = ac_{k_a=0}$; N_a is the number of observations; x_t is the sample observation; and $\overline{x}$ is the sample means of the series.

In this case, the cross-correlation function measures the temporal and spatial correlation between the travel time data pairs recorded on two selected links. For travel time data pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, an estimate of the cross-covariance function $c_{xy}(k_c)$ is

$$c_{xy}(k_c) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{t=1}^{n_c-k_c}\left(x_t - \overline{x}\right)\left(y_{t+k_c} - \overline{y}\right), & k_c = 0, 1, 2, \ldots, \\ \dfrac{1}{n}\displaystyle\sum_{t=1}^{n_c+k_c}\left(y_t - \overline{y}\right)\left(x_{t-k_c} - \overline{x}\right), & k_c = 0, -1, -2, \ldots, \end{cases} \tag{2}$$

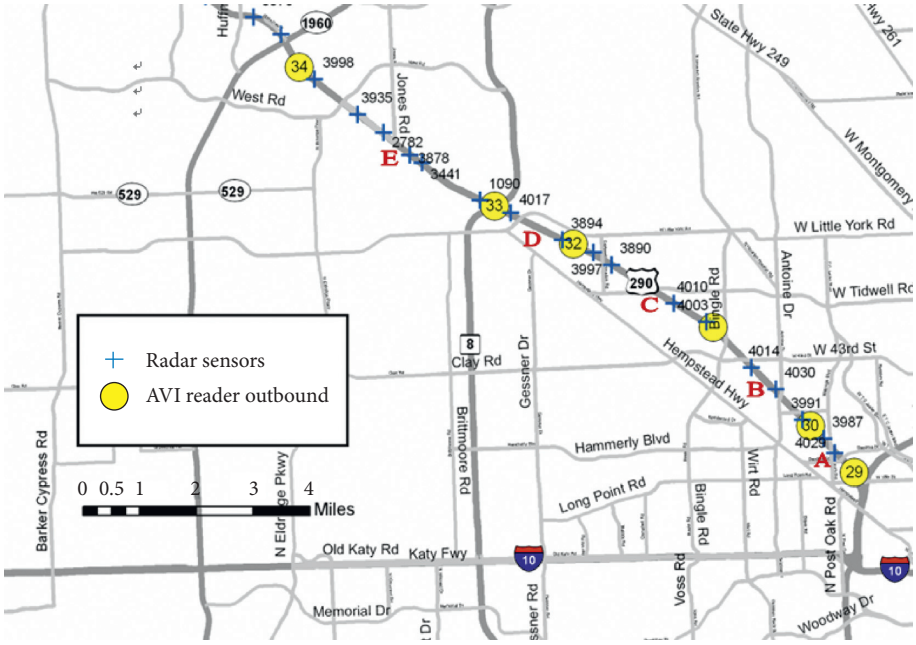


Figure 1: Study links along the US-290 [23].

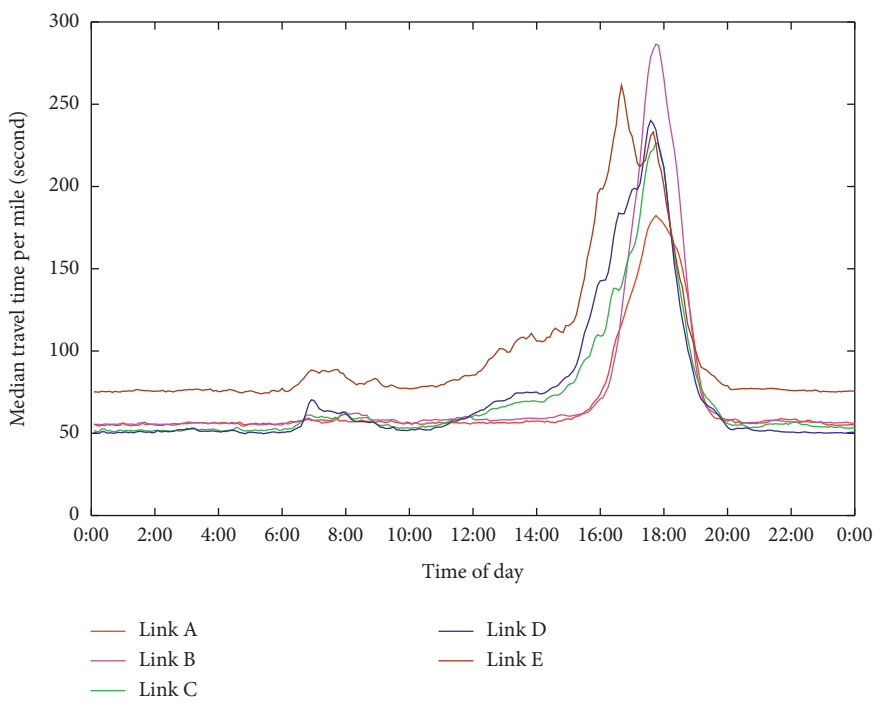


Figure 2: Historical median travel times on the five links.

where $\overline{x}$ and $\overline{y}$ are the sample means of the x_t series and y_t series, respectively. n is the number of travel time data pairs, and an estimate of the lag k_c cross-correlation function is

$$cr_{xy}(k_c) = \frac{c_{xy}(k_c)}{s_x s_y}, \quad k_c = 0, \pm 1, \pm 2, \ldots, \tag{3}$$

where $s_x = \sqrt{c_{xx}(0)}$ and $s_y = \sqrt{c_{yy}(0)}$, and $cr_{xy}(k_c)$ is the sample cross-correlation function.

We found that the autocorrelation function of travel time shows a downward trend with time lag (Figure 3). Cross-correlation functions between link D and links A, B, C, and E peak at the lag of −9 and −4, 0 and 0, respectively (Figure 4). As can be seen from previous analysis (Figure 2), the peak times of five links are not same. The peak of links A and B occurs earlier than link D, so the cross-correlation functions between links A and B and link D peak reach the peak at lags of −9 and −4. The traffic state of links A and B

changed 45 minutes and 20 minutes earlier than link D. Links C and E are directly connected to link D, the traffic congestion state and the peak time are more similar, the cross-correlation functions between links C, E and link D peak at lags of 0. Furthermore, the maximum cross-correlation values between link D and its adjacent links A, B, C, and E are 0.547, 0.720, 0.822, and 0.904.

2.2. Periodic Pattern of Travel Time. Previous research showed that travel time exhibits periodic characteristics during the weekdays. Similar periodic characteristics were found by Kamarianakis et al. [36] in occupancy, speed, and flow data. Because the periodic trend may affect the travel time prediction, this study proposed the hybrid prediction models to accommodate the periodic trend components as well as the temporal and spatial correlation observed in the data. Specifically, periodic characteristics are modeled using TPF, SA, and DES methods.

3. Methodology

3.1. Periodic Functions

3.1.1. Simple Average Method. Simple average method is one of the commonly used methods to describe the periodic characteristics [31]. During the study, the researchers set the hypothesis that the sampling travel time data of M consecutive working days could be written as a series of one-dimensional vectors Y_d, as shown in equation (4). The intraday trend $\overline{Y}$ is calculated by simple average method as equation (5).

$$\begin{aligned} Y_1 &= [y_1(1), y_1(2), ..., y_1(m)], \ldots, \\ Y_M &= [y_M(1), y_M(2), ..., y_M(m)], \end{aligned} \tag{4}$$

$$\overline{Y} = [\overline{y}(1), \ldots, \overline{y}(m)] = \left[\frac{1}{M}\sum_{d=1}^{d=M} y_d(1), \ldots, \frac{1}{M}\sum_{d=1}^{d=M} y_d(m)\right], \tag{5}$$

where $y_d(m)$ stands for travel time data collected at time m on day d. M indicates the number of sampled days, m is sampling data points per day. In this study, $M = 30$, $m = 288$.

3.1.2. Trigonometric Polynomial Function. Trigonometric polynomial adopts the sinusoids and cosinusoids to describe the periodic pattern. Equation (6) was used to calculate the average daily travel time at each station. Trigonometric polynomial function is represented in the following equation:

$$\begin{aligned} S_t = a_0 &+ a_1 \sin\left(\frac{2\pi t}{m}\right) + a_2 \cos\left(\frac{2\pi t}{m}\right) + \cdots \\ &+ a_{2n_r-1} \sin\left(\frac{2n_r \pi t}{m}\right) + a_{2n_r} \cos\left(\frac{2n_r \pi t}{m}\right), \end{aligned} \tag{6}$$

where S_t is the estimated periodic component at time t; m indicates the number of samples per day; n_r is the number of trigonometric polynomials; and $a_0, \ldots, a_{2n_r}$ are the coefficients.

Regarding the selection of optimal number of trigonometric series functions, Zou et al. [29] claimed that the number of trigonometric polynomials might have an impact on the prediction accuracy of the hybrid model. At the same time, they found that 15 or more trigonometric polynomials should be included in the periodical component. Therefore, in this study, the researchers set the value of n_r in equations (6)–(15).

3.1.3. Double Exponential Smoothing. Double exponential smoothing is one widely used method for both smoothing and forecasting time series. This approach builds the prediction in accordance of the levels mean M_t and the trend T_t. The model can be expressed as

$$\begin{aligned} M_t &= \alpha Y_t + (1-\alpha)(M_{t-1} + T_{t-1}), \\ T_t &= \beta(M_t - M_{t-1}) + (1-\beta)T_{t-1}, \\ Y_{t+1} &= M_t + T_t, \\ M_1 &= Y_1, \\ T_1 &= \frac{((Y_2 - Y_1) + (Y_3 - Y_2) + (Y_4 - Y_3))}{3}, \end{aligned} \tag{7}$$

where Y_t is the observed travel time at time t; M_t stands for the estimate of level of series at time t; T_t indicates the estimate of slope of series at time t; α and β are smoothing parameters, the two parameters can be estimated using the Levenberg-Marquardt algorithm.

3.2. Travel Time Prediction Models

3.2.1. Statistical Models

(1) Autoregressive Integrated Moving Average (ARIMA) Models. The ARIMA model transforms nonstationary time series into stationary time series after *di* differences, and then stationary sequence can be predicted by the ARMA model. From the view of mathematics, the demonstration of an ARMA (p, q) procedure is as

$$y(t) = \varepsilon(t) + \sum_{i=1}^{i=p} \varphi_i y(t-i) + \sum_{j=1}^{j=q} \theta_i \varepsilon(t-j), \tag{8}$$

where $y(t)$ stands for the future travel time at time t; φ_i and θ_i are the parameters of pattern; $\varepsilon(t)$ indicates white Gaussian noise with mean zero and variance σ_w^2; p is the number of autoregressive terms; q is the amount of lagged forecast errors. Let

$$B^i y(t) = y(t-i). \tag{9}$$

ARIMA model is as equation (10):

$$\left(1 - \sum_{i=1}^{i=p} \varphi_i B^i\right)(1-B)^{di} y(t) = \left(1 + \sum_{j=1}^{j=p} \theta_i B_i\right)\varepsilon(t), \tag{10}$$

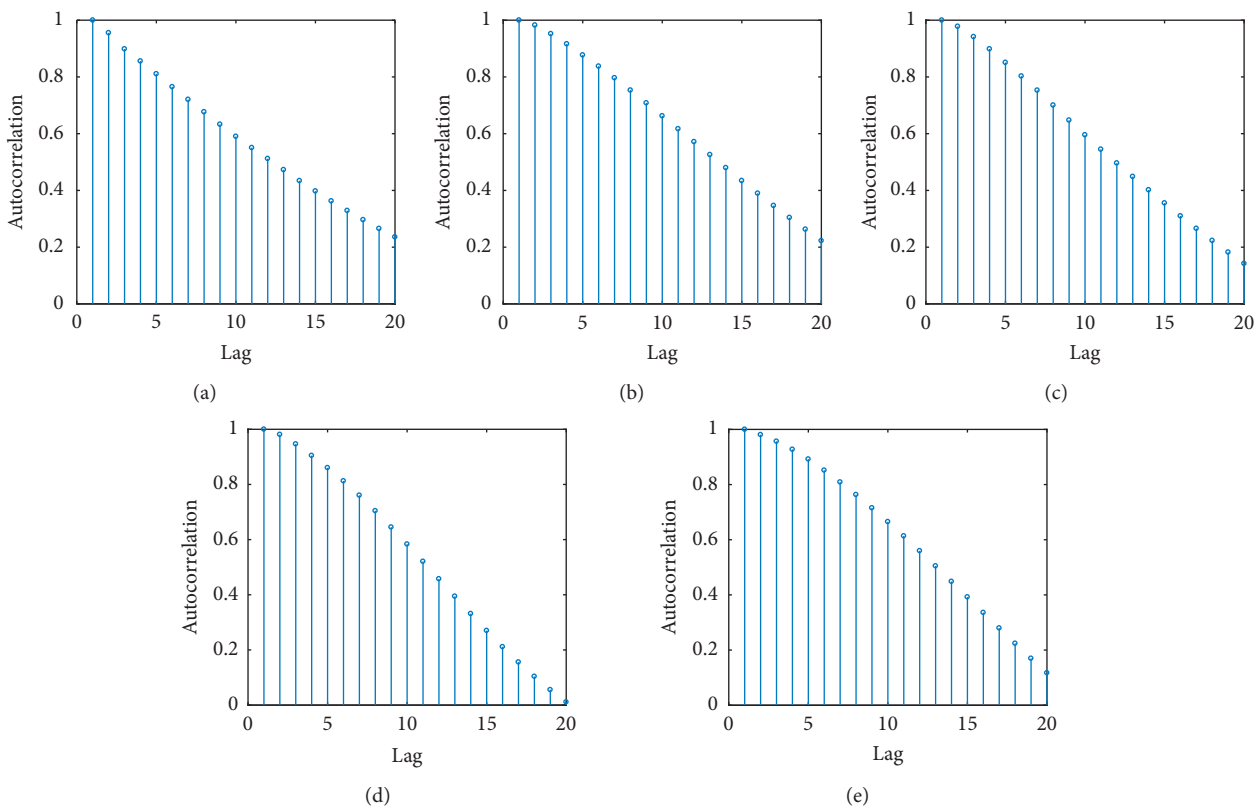


Figure 3: Autocorrelation functions of 5-minute travel time on links A, B, C, D, and E. (a) Link A. (b) Link B. (c) Link C. (d) Link D. (e) Link E.

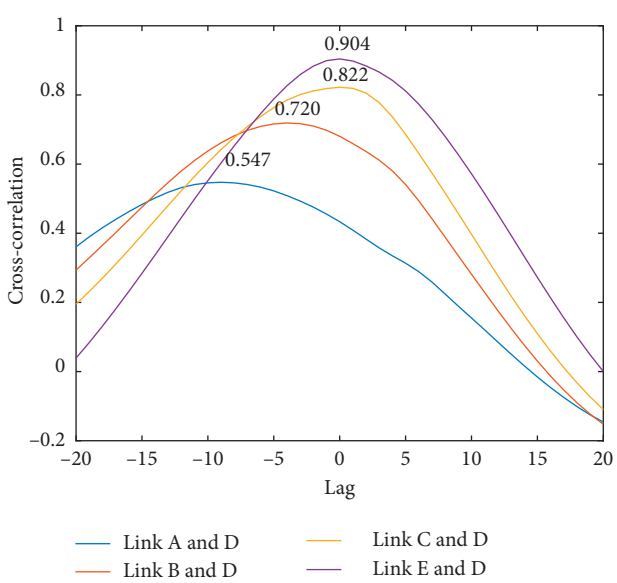


Figure 4: Cross-correlation functions of 5-minute travel time on link D.

where di is a nonnegative integer, which stands for the number of nonseasonal differences. If $di = 0$, the ARMA model could be obtained. When predicting each future travel time value, the best order of the ARIMA model is decided by Akaike information criterion (AIC).

(2) Space Time Model. As a probabilistic modeling method, ST model can provide point prediction and corresponding prediction intervals. The normal distribution is used to describe travel time in this study. The point prediction D_{t+p} is D_{t+p} given by

$$Z_\tau = u_{t+pr} + \sigma_{t+pr} \times \Phi^{-1}[\tau], \tag{11}$$

where $pr = 1, 2, 3, \ldots, 12$; u_{t+pr} and σ_{t+pr} are the location parameter and scale parameter of $N(u_{t+pr}, \sigma^2_{t+pr})$; Φ is the cumulative density function of a standard normal distribution. The u_{t+pr} is modeled through a linear combination of current and previous values of all travel time series on all links. When choosing the predictive variables, different combinations of predictive variables need to be considered. Therefore, researchers begin with the most complex models and gradually subtract predictive variables until no further improvement was obtained. For instance, if $pr = 1, 9$ variables were selected,

$$\begin{aligned} u_{t+1} = {} & \alpha_0 + \alpha_1 A_t + \alpha_2 B_t + \alpha_3 B_{t-1} + \alpha_4 C_t + \alpha_5 C_{t-1} \\ & + \alpha_6 D_t + \alpha_7 D_{t-1} + \alpha_8 E_t + \alpha_9 E_{t-1}, \end{aligned} \tag{12}$$

where A_t, B_t, C_t, D_t, E_t are the travel time at links A, B, C, D, and E at time t and $\alpha_0, \alpha_1, \ldots, \alpha_9$ are model coefficients.

To build a model for the predictive spread, σ_{t+p}, the ST model allows for conditional heteroscedasticity by modeling σ_{t+pr} as a linear function of the volatility value v_t,

$$\sigma_{t+pr} = b_0 + b_1 v_t. \tag{13}$$

The coefficients b_0 and b_1 are nonnegative, and their volatility values could be modeled as

$$v_t = \left(\frac{1}{10} \sum_{i=0}^{i} \left((A_{t-i} - A_{t-i-1})^2 + (B_{t-1} + B_{t-i-1})^2 + (C_{t-1} + C_{t-i-1})^2 + (D_{t-1} + D_{t-i-1})^2 + (E_{t-1} + E_{t-i-1})^2 \right) \right)^{1/2}. \tag{14}$$

(3) Vector Autoregressive Models. VAR model is regarded as one of the most widespread methods which utilize statistical methods in time series prediction. The model can include many factors consisting of the impact of upstream and downstream links on predicting future travel time. During the process of this research, a 5-equation VAR model is utilized, and it can be expressed as follows:

$$X_{t+1} = c_0 + c_1 X_t + c_2 X_{t-1} + \cdots + c_k X_{t-k-1} + f_{t+1}, \tag{15}$$

where $X_{t+1} = (A_{t+1}, B_{t+1}, C_{t+1}, D_{t+1}, E_{t+1})^T$; $c_0 = 5 \times 1$ constant term; $c_1, \ldots, c_k = 5 \times 5$ coefficient matrices; and f_{t+1} = the corresponding 5×1 independently and identically distributed random vector, $E(f_{t+1}) = 0$.

The stability of the VAR model could be guaranteed through the characteristic polynomial

$$\det(I_5 - c_1 z - \ldots - c_o z^o) \neq 0, \quad \text{for } |z| \leq 1, \tag{16}$$

where I_5 stands for a 5×5 identity matrix. It is a necessary and sufficient condition that all characteristic roots are located outside the unit circle for stability.

3.2.2. Machine Learning Models

(1) Support Vector Machine Model. The SVM approach is a method which could be used to map the sample space into high or even infinite dimension feature space (Hilbert space) by nonlinear mapping to construct linear regression in a new space. Given a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$ for regression, N is the number of training samples. Normally, the objective of SVM is to find a function

$$g(x) = w \cdot \Phi(x) + b, \tag{17}$$

where $\Phi(\cdot)$ = the kernel function that maps input x into the feature space g; w is the weighting vector; b is a constant bias.

A λ-insensitive loss function is assumed as

$$L_\lambda(x, y, g) = |y - g(x)|_\lambda = \max(0, |y - g(x)| - \lambda). \tag{18}$$

Then, it could be estimated that w and b by working out this optimization problem:

$$R(f) = C \sum_{i=1}^{i=N} L_\lambda(x_i, y_i, g) + \frac{1}{2}\|w\|^2, \tag{19}$$

where λ is the maximum deviation permitted; $L_\lambda(x, y, g)$ is loss function, C indicates the related penalty for stating deviation within the training process that assesses the tradeoff between the empirical risk and the smoothness of the model. The relaxation variables β_i and β_i^* are used to indicate the optimization objective into the optimization issue stated as

$$\begin{aligned} \min \quad & C \sum_{i=1}^{i=N} (\beta_i + \beta_i^*) + \frac{1}{2}\|w\|^2 \\ & y_i - w \cdot \Phi(x) - b \leq \lambda + \beta_i \\ \text{s.t.} \quad & w \cdot \Phi(x) + b - y_i \leq \lambda + \beta_i^* \\ & \beta_i, \beta_i^* \geq 0. \end{aligned} \tag{20}$$

That issue mentioned above is worked out by utilizing the Lagrange equation. The regression function is demonstrated as

$$\begin{aligned} f(x) &= \sum_{x_i \in SV} (\alpha_i - \alpha_i^*) E(s_i, s_j) + b, \\ E(s_i, s_j) &= \exp\left(-\gamma \left\| s_i, s_j \right\|^2\right), \end{aligned} \tag{21}$$

where $E(s_i, s_j)$ is the kernel function. α_i and α_i^* are the solutions to dual problem. In our study, the grid analysis and cross validation are used to optimize the parameters C and γ. The cross-validation method divides data into three groups, among which one subset is the validation group and the other two subsets are used as the training set; 3 models are obtained. Grid analysis is a method of programming enumeration to compare the performance of models with different parameters C and γ. In this paper, all combinations of Log $2C$ and Log 2γ parameters between -5 and 5 were traversed. The parameter combination with minimum mean square error was selected.

(2) Back Propagation Neural Network Model. In short, the BPNN model is a multilayer feed forward neural network which consists of many parallel nonlinear computing elements. As we all know, initialization network is composed of input layer, hidden layer, and output layer. Within the neural network, the weights between the most important parameter connection layers can be calculated by error back propagation algorithm. When a neural network model acquires the mapping relationship between input and output variables through continuous learning, it can predict the output according to the given input variables.

First, equation (22) can be used to calculate the value of the predicted hidden layer:

$$H_j = S\left(\sum_{i=1}^{i=\text{num}} v_{ij} x_j - e_j \right), \quad j = 1, 2, \ldots, h, \tag{22}$$

where H_j stands for the production of hidden layer and S is the incentive function of neurons, h stands for the neuron number of hidden layers, num refers to the neuron number of the input-layers, v_{ij} stands for the weight element between input-layer and hidden-layer, e_j stands for the bias value of hidden layer.

Second, predicting value of the output layer could be calculated through

$$Q = \sum_{j=1}^{j=h} H_j v_{j\zeta} - q_\zeta, \quad \zeta = 1, 2, \ldots, k_1, \tag{23}$$

where Q stands for the actual output of output layer, q_ζ refers to the bias value of output layer, and k_1 stands for the neuron number of the output layer. In our study, the empirical formula combined with the trial and error method was used to determine the number of nodes in the hidden layer. 4 nodes with the best performance were selected finally.

(3) Multilinear Regression Model. Compared with the above two supervised algorithms, the construction of multiple linear regressions is simpler and belongs to regression learning category. In MLR, the prediction values can be calculated by the following equation:

$$y(t) = r_0 + \sum_{j=1}^{j=lr} r_j y(t-j), \tag{24}$$

where $y(t)$ represents the prediction value at time t. The independent variable $y(t-j)$ means the travel time data at the previous $t-j$ period, lr is the number of historical travel times considered in MLR model, and r_0 , . . ., r_j are the regression parameters which can be optimized by training samples. lr is chosen on basis of an analysis of the travel data from January to April 2008. Different numbers of lr are considered.

3.3. Hybrid Prediction Models. As mentioned in Section 2, freeway travel time has a daily periodic characteristic. Therefore, it can be assumed that the travel time has two parts. One of the two parts is the deterministic component; the other is the irregular component. In such a hypothesis, the hybrid prediction model can be used to describe or calculate the freeway travel time:

$$P_t = D_t + y_t^r, \tag{25}$$

where P_t is the travel time at time t at station D; D_t is the periodic component; and y_t^r represents the residual part after removing the periodic component.

Periodic component can be described by three kinds of functions (TPF, SA, and DES), and the residual part is modeled by six prediction models. We compare the impacts of different periodic functions on multistep ahead freeway travel time prediction models using travel time data with different aggregation levels.

3.4. Measures and Training Period. To evaluate the multistep prediction performance of all prediction models, three indicators, mean absolute error (MAE), mean absolute prediction error (MAPE) and root mean square error (RMSE) are considered comprehensively. The equations for calculating three indexes are as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right|, \\ \text{MAPE} &= \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| * .100\%, \\ \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n}}, \end{aligned} \tag{26}$$

where n is the number of observations; y_i represents the actual travel time at time i on link D; and $\hat{y}_i$ refers to the predicted travel time.

So far, there is no automatic way to calculate and evaluate the model training period. This study considered different training periods of 15, 20, 25, 30, 40, 50 and 60 days. For comparison, the travel time data in August (21 weekdays) were used as the test set. Figure 5 shows MAE, MAPE, and RMSE values of six travel time prediction models under different lengths of training periods. It is observed that the prediction performance of statistical models and MLR model changed slightly as the number of training period increases. The performance of SVM and BPNN has been greatly improved as training period increases when the training period was less than 30 days. If the training period was more than 30 days, the prediction accuracy of SVM and BPNN models changed slightly. Longer training period usually requires larger computational time for each model. For example, the computational time of the SVM model is 5 minutes when 10-day travel time data was used for model training, and the computational time can be as high as 68 minutes when 60 days was chosen as training period. The calculation time and prediction accuracy were considered comprehensively in our study, and a 30-day (July (23 days) and June (7 days)) training period is chosen for models.

4. Results and Discussion

In this part, the multistep ahead prediction performance of SVM, BPNN, MLR, ARIMA, ST, VAR under different aggregation levels (i.e., 5-minute, 10-minute, and 15-minute) are evaluated using the travel time data observed on link D. In addition, we explored the impacts of different periodic functions on statistical models and machine learning models under different aggregation levels for the input data. The testing period is 15:30 to 19:30 from 1 August to 31 August (21 weekdays).

4.1. The Performance of Six Models. The study provides the MAE, MAPE, and RMSE values of SVM, BPNN, MLR, ARIMA, ST, and VAR models for different forecasting horizons under different aggregation levels (5-minute, 10-minute, and 15-minute) for the input data (Tables 1–3). Tests on travel time data indicate the following findings. First, the prediction accuracy deteriorates as the forecasting step increases for all models. Second, the higher the data aggregate level, the higher the accuracy of short-term travel time

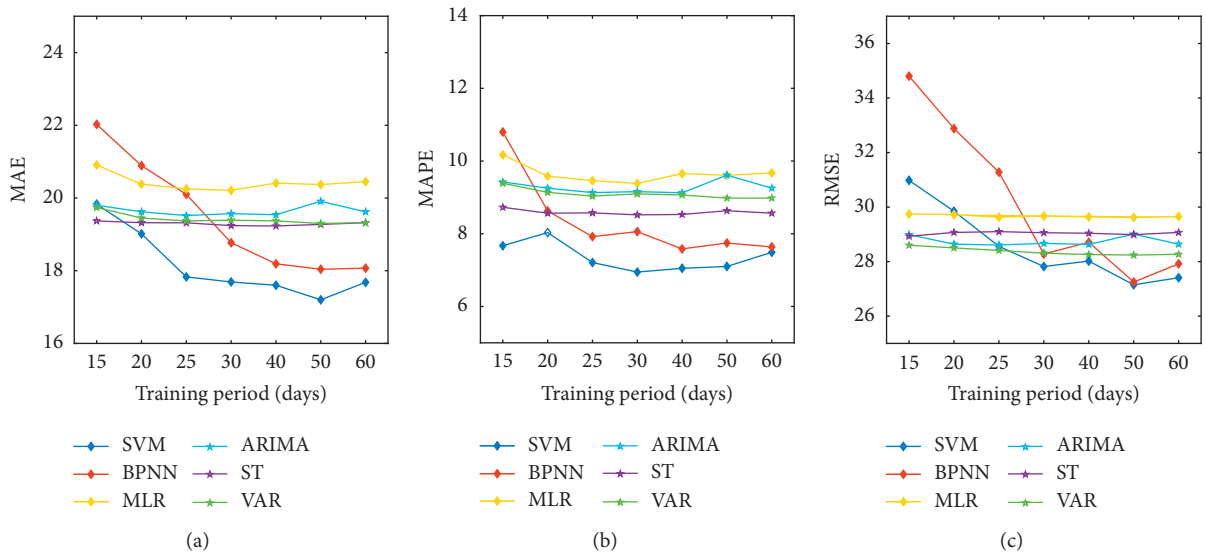


Figure 5: Comparison of training period for the different models.

Table 1: MAE, MAPE, and RMSE of six models for different minutes ahead predictions with 5-minute as aggregating level.

Time scale: 5 min		Minutes ahead predictions					
MAE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min
	SVM	**17.69**	**25.84**	**32.52**	**46.33**	**53.78**	**60.73**
Machine learning models	BPNN	18.10	26.10	34.87	48.02	54.07	60.94
	MLR	20.21	34.01	46.32	74.53	95.71	111.05
	ARIMA	19.80	31.13	40.56	57.02	65.98	77.50
Statistical models	ST	**19.37**	**29.86**	**38.40**	**51.22**	**63.41**	**73.79**
	VAR	19.39	30.74	39.31	56.48	68.82	78.16
MAPE (%)	Model	5 min	10 min	15 min	30 min	45 min	60 min
	SVM	**6.95**	**10.18**	**12.69**	**19.90**	**25.14**	**29.73**
Machine learning models	BPNN	7.20	10.57	14.73	23.77	26.28	30.82
	MLR	9.38	17.66	25.16	43.13	55.28	64.55
	ARIMA	9.42	15.90	21.66	32.78	38.26	45.21
Statistical models	ST	**8.73**	**14.30**	**19.21**	**25.38**	**32.34**	38.60
	VAR	9.10	15.15	19.72	28.19	33.16	**38.06**
RMSE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min
	SVM	**27.82**	41.08	**52.33**	**72.33**	83.07	91.32
Machine learning models	BPNN	28.38	**40.93**	53.60	73.15	**81.75**	**89.47**
	MLR	29.67	46.96	61.95	92.99	118.52	134.65
	ARIMA	28.99	44.47	56.80	77.89	**92.55**	108.58
Statistical models	ST	28.93	**43.57**	55.60	**73.85**	94.27	108.50
	VAR	**27.98**	43.62	**55.18**	78.74	98.34	**105.69**

prediction results. For example, when we predict the 30-minute ahead travel time of link D, the prediction result is better with 15-minute data as the aggregation level than that with 10-minute and 5-minute data as the aggregation level. Third, for machine learning models, the prediction accuracy of MLR is lower than that of BPNN and SVM. For statistical models, the prediction accuracy of ARIMA is lower than that of ST and VAR. The possible reason is that the SVM, BPNN, ST, and VAR models use spatial and temporal information from neighboring links to predict the future travel time value at time $t+p$. While MLR and ARIMA models use the travel time data collected on the target link D only to predict travel time values at time $t+p$. Fourth, the prediction accuracy of two machine models (SVM and BPNN) is better than that of statistical models. However, the prediction accuracy of MLR model is significantly lower than that of statistical models.

4.2. Impacts of Periodic Functions. To investigate whether the proposed periodic functions improve the performance of six prediction models, hybrid models and prediction models are considered to predict travel time values at link D for the same testing period (15:30 to 19:30 from 1 August to 31 August). The periodic function shows a consistent rule on improving the prediction model, whether it is based on MAE index, MAPE or RMSE index. The figure shows the

Table 2: MAE, MAPE, and RMSE of six models for different minutes ahead predictions with 10-minute as aggregating level.

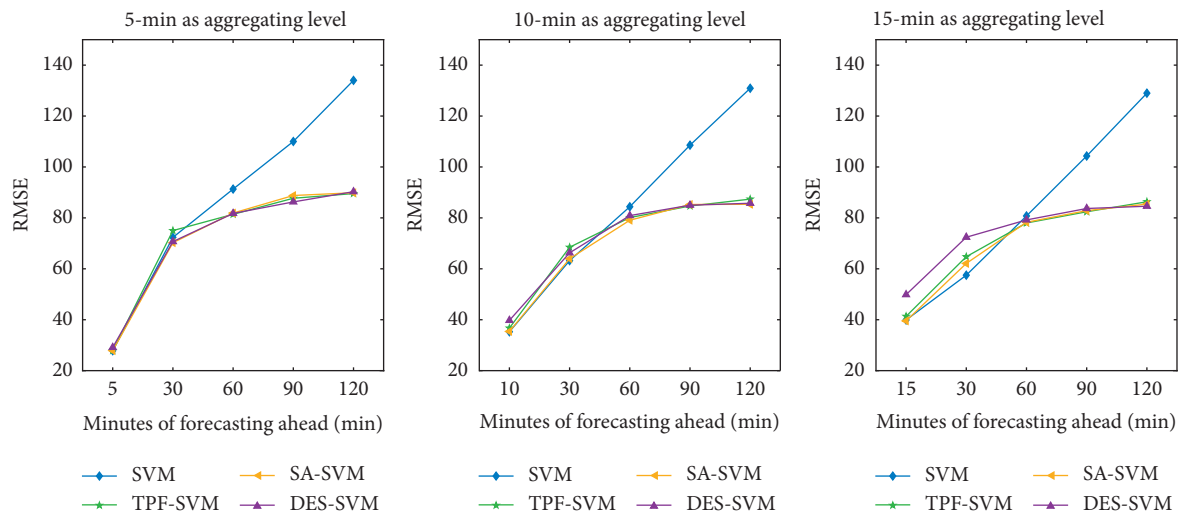
Time scale: 10 min		Minutes ahead predictions					
MAE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min
	SVM	—	21.79	—	**39.29**	—	**54.90**
Machine learning models	BPNN	—	**21.60**	—	44.41	—	59.36
	MLR	—	30.32	—	71.39	—	109.57
	ARIMA	—	29.32	—	62.79	—	86.27
Statistical models	ST	—	**26.88**	—	**51.55**	—	71.44
	VAR	—	26.89	—	51.77	—	**67.50**
MAPE (%)	Model	5 min	10 min	15 min	30 min	45 min	60 min
	SVM	—	**8.88**	—	**17.20**	—	**25.55**
Machine learning models	BPNN	—	8.92	—	20.81	—	30.91
	MLR	—	15.51	—	41.05	—	64.00
	ARIMA	—	14.93	—	36.55	—	50.69
Statistical models	ST	—	**12.44**	—	**26.19**	—	37.60
	VAR	—	13.49	—	26.79	—	**34.04**
RMSE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min
	SVM	—	**35.27**	—	**63.21**	—	**84.38**
Machine learning models	BPNN	—	35.67	—	70.50	—	87.33
	MLR	—	41.98	—	89.53	—	132.75
	ARIMA	—	41.49	—	83.54	—	115.66
Statistical models	ST	—	40.09	—	**72.82**	—	104.54
	VAR	—	**38.66**	—	73.83	—	**97.83**

Table 3: MAE, MAPE, and RMSE of six models for different minutes ahead predictions with 15-minute as aggregating level.

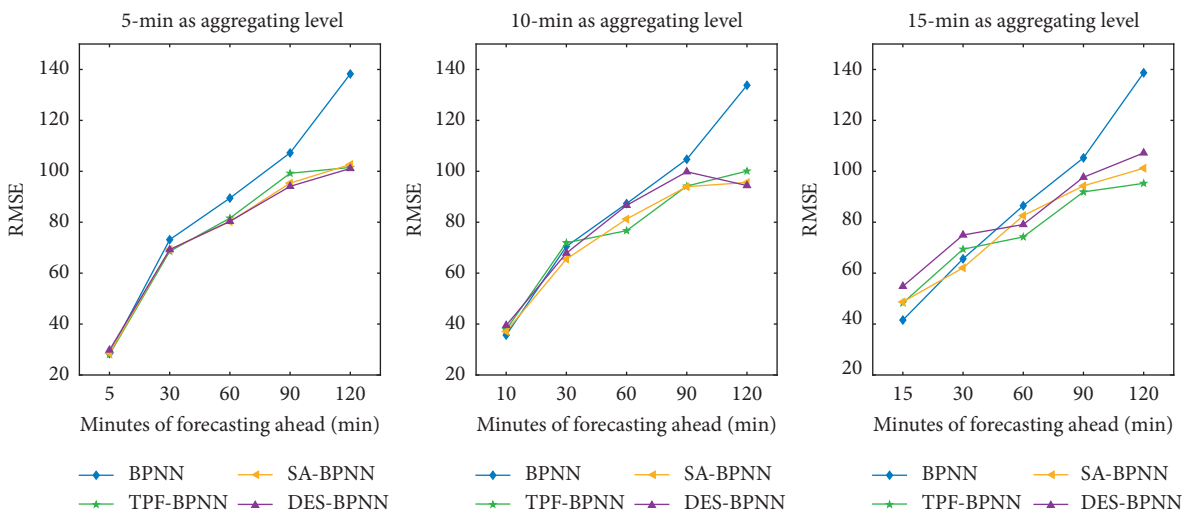
Time scale: 15 min		Minutes ahead predictions					
MAE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min
	SVM	—	—	**24.30**	**35.98**	45.79	**52.72**
Machine learning models	BPNN	—	—	25.41	38.48	**45.73**	61.38
	MLR	—	—	37.14	68.42	88.75	106.79
	ARIMA	—	—	32.64	56.16	65.82	79.20
Statistical models	ST	—	—	**30.47**	49.01	60.88	68.28
	VAR	—	—	30.87	**48.66**	**58.80**	**65.82**
MAPE (%)	Model	5 min	10 min	15 min	30 min	45 min	60 min
	SVM	—	—	**9.70**	**15.43**	**21.39**	**24.81**
Machine learning models	BPNN	—	—	11.09	16.25	21.66	28.05
	MLR	—	—	20.99	39.86	51.47	62.47
	ARIMA	—	—	18.19	33.01	37.99	48.02
Statistical models	ST	—	—	**15.55**	**24.54**	31.12	35.59
	VAR	—	—	16.48	25.19	**29.74**	**33.13**
RMSE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min
	SVM	—	—	**39.90**	**57.49**	**70.21**	**80.76**
Machine learning models	BPNN	—	—	41.57	65.63	72.81	95.45
	MLR	—	—	49.56	85.06	112.00	130.03
	ARIMA	—	—	45.50	75.91	90.63	106.43
Statistical models	ST	—	—	45.06	71.97	91.30	100.50
	VAR	—	—	**43.86**	**70.56**	**87.05**	**96.55**

Bold values indicate the smallest MAE, MAPE, and RMSE values in machine learning models and statistical models, respectively.

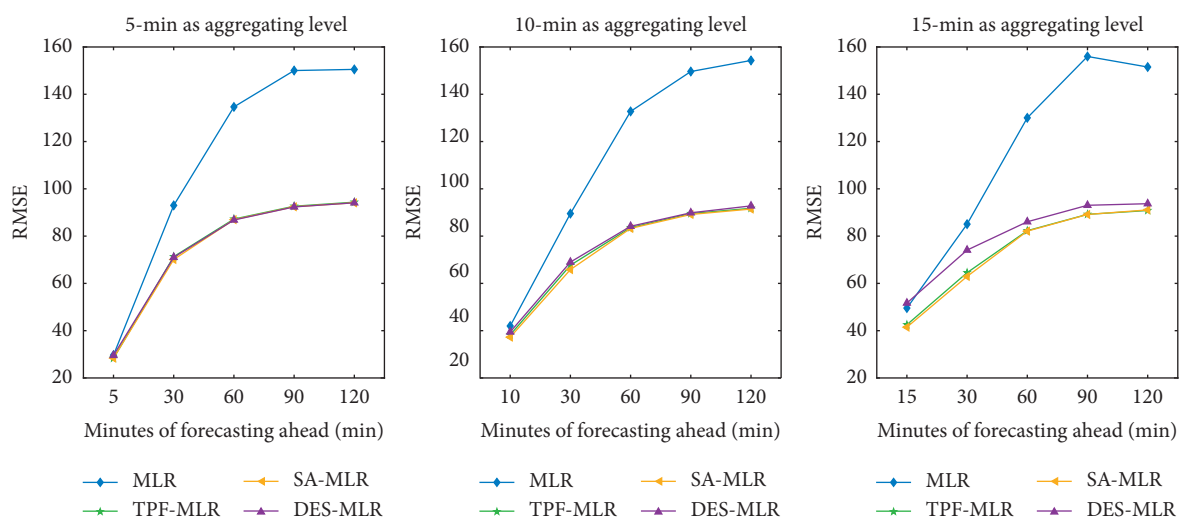
RMSE results of the six models and 18 hybrid models for different forecasting horizons under different aggregation levels (5-minute, 10-minute, 15-minute) (Figure 6). Based on the observation of the results, several interesting conclusions can be drawn. First, period functions have similar impacts on SVM and BPNN models. The periodic functions have a definite improvement for more than 60 minutes ahead prediction under three data aggregation levels. Second, three periodic functions have improved the prediction performance of MLR model in multistep ahead prediction for three data aggregation levels. Third, three periodic functions can improve the prediction accuracy of travel time over 30 minutes ahead for all statistical models. Fourth, with the increase of aggregation level, the difference of prediction results of the comprehensive prediction model considering periodic functions increases gradually. For example, when the aggregation level is 5-minute, the prediction results of the three SVM comprehensive models have little difference, while when the aggregation level is 15 minutes, the prediction results are significantly different.



(a)



(b)



(c)

Figure 6: Continued.

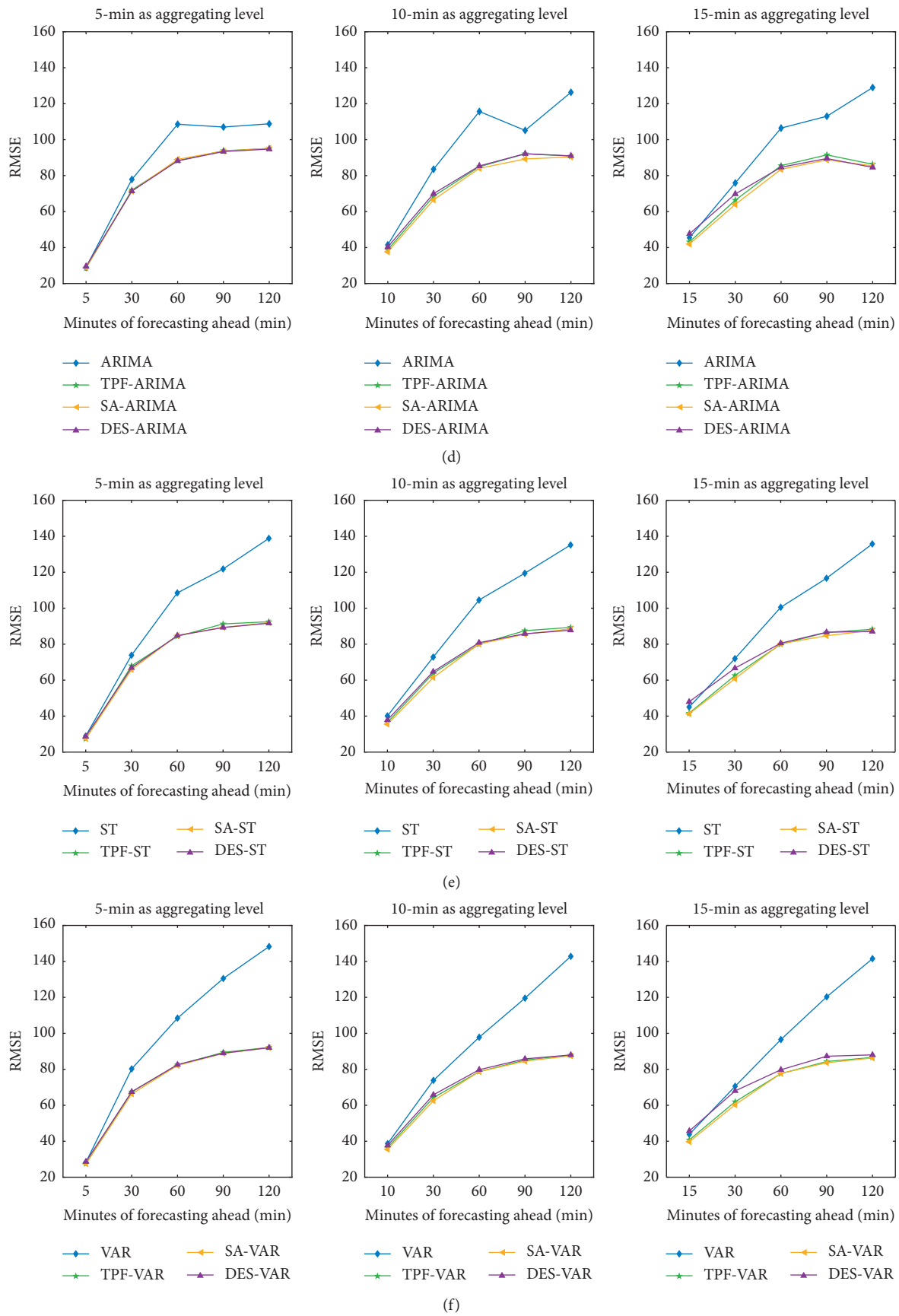
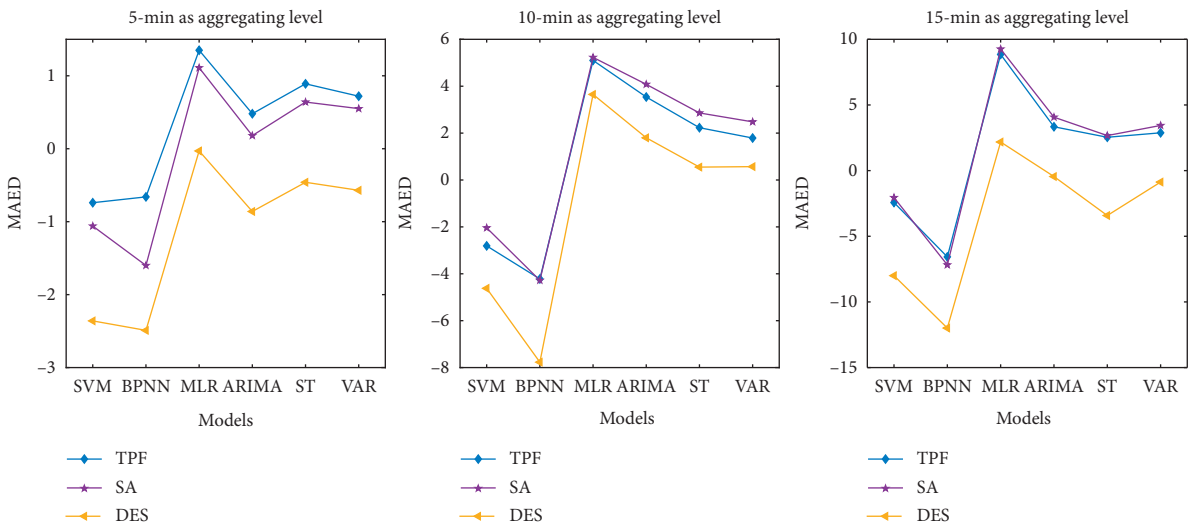
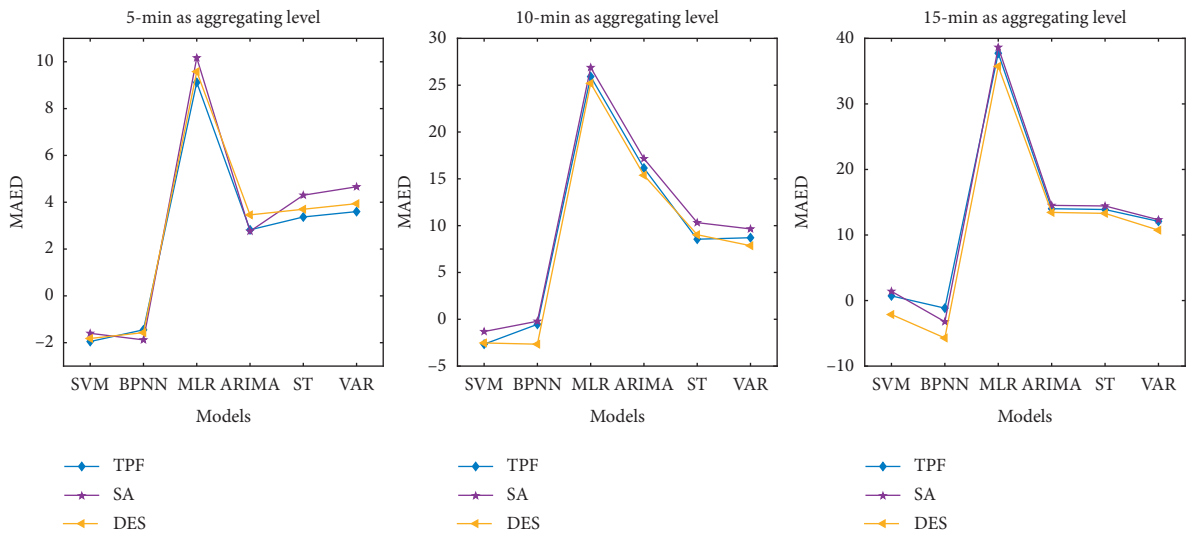


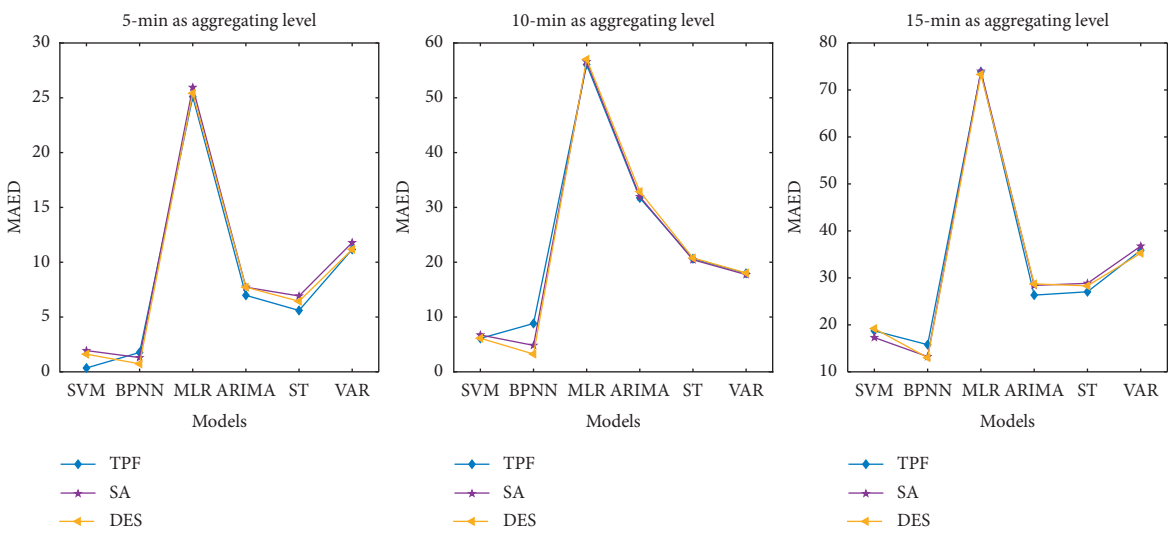
Figure 6: RMSE values of six models for different minutes ahead predictions. (a) RMSE values of SVM models for different minutes ahead predictions. (b) RMSE values of BPNN models for different minutes ahead predictions. (c) RMSE values of MLR models for different minutes ahead predictions. (d) RMSE values of ARIMA models for different minutes ahead predictions. (e) RMSE values of ST models for different minutes ahead predictions. (f) RMSE values of VAR models for different minutes ahead predictions.



(a)

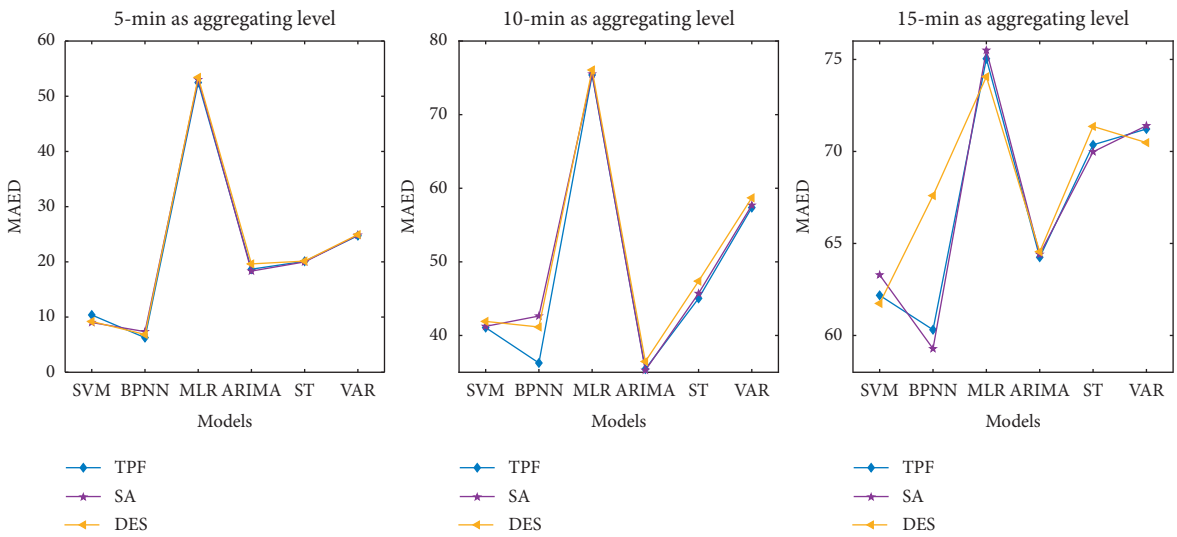


(b)



(c)

Figure 7: Continued.



(d)

Figure 7: MAED for multistep ahead prediction with different aggregation levels (5-minute, 10-minute, and 15-minute). (a) MAED for 1-step ahead prediction. (b) MAED for 3-step ahead prediction. (c) MAED for 6-step ahead prediction. (d) MAED for 12-step ahead prediction.

Table 4: Aggregation level suggestions and periodic function suggestions.

Models	Minutes forecasting ahead (min)	Aggregating level suggestions (min)	Periodic function suggestions
SVM, BPNN	5	5	—
	10	10	—
	15	15	—
	30	15	—
	45	15	—
	60	15	TPF
	90	15	TPF
	120	15	TPF
ARIMA, ST, VAR, MLR	5	5	TPF
	10	10	SA
	15	15	SA
	30	15	SA
	45	15	SA
	60	15	SA
	90	15	SA
	120	15	SA

From the above analysis, we can conclude that the periodic functions obviously improve the prediction accuracy of the six prediction models for multistep ahead prediction. Then we analyze the impact degree of different periodic functions on prediction models based on mean absolute error difference (MAED). The equation of the MAED is as follows:

$$\text{MAED} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| - \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_{if} - y_i\right|, \tag{27}$$

where n is the number of observations; y_i represents the actual travel time at time i on link D; $\hat{y}_i$ refers to the predicted travel time based on traditional prediction models, and $\hat{y}_i$ refers to the predicted travel time based on models considering periodic functions. If the MAED is greater than zero, this periodic function improves the prediction accuracy of the traditional prediction model; otherwise, it reduces the prediction accuracy of the traditional prediction model. The result shows the MAED values for 18 hybrid models from 1-step to 12-step ahead forecasting with 5-minute, 10-minute and 15-minute as aggregation levels (Figure 7). First, periodic functions can significantly improve the MLR model in multistep ahead prediction for three data aggregation levels. Second, for 1-step and 3-step ahead prediction, three periodic functions reduce the prediction accuracy of SVM and BPNN models. Third, for 1-step ahead prediction, both TPF and SA improve the performance of statistical models in multistep ahead prediction

for three data aggregation. When the aggregation level is 5-minute, TPF can obviously improve the prediction accuracy. While when the aggregation level increases to 10-minute, SA periodic function performs better. For 3-step ahead prediction, three periodic functions improve the forecasting results of the statistical models obviously, and three periodic functions have slight difference in improving prediction accuracy of the statistical models. Fourth, when 6-step and 12-step prediction ahead, three periodic functions improve the forecasting results of the six models obviously, and three periodic functions have slight difference in improving prediction accuracy of the prediction model. Fifth, for the same prediction step, the improvement of periodic function is more obvious with the increase of data aggregation level. For example, for 12-step ahead prediction, hybrid models perform better with 15-minute as aggregated level than that of with 5-minute as aggregated level.

In this section, we discuss the aggregation level and periodic function suggestions. According to the conclusion of Table 1 and Figure 7, the higher the data aggregate level, the higher the accuracy of short-term travel time prediction results. When prediction time was greater than 15 minutes, the highest accuracy could be obtained by using the aggregation level of 15 minutes. According to Figures 6 6(a) and 6(b), the periodic functions cannot improve the prediction performance of SVM and BPNN models when prediction time was less than 60 minutes, it is recommended that no periodic function should be considered. When the prediction time is greater than 60 minutes, TPF periodic functions are recommended. As can be seen from Figure 7, both TPF and SA have improved the performance of statistical models and MLR model in multistep ahead prediction for three data aggregation and SA performed better. For different minutes ahead prediction, aggregation level and periodic function suggestions are shown in Table 4.

5. Conclusions

This paper evaluated the multistep ahead prediction performance of SVM, BPNN, MLR, ARIMA, ST, and VAR models using the freeway travel time data collected from vehicle identification readers along US-290 in Houston, Texas. The performances of the six prediction models under different aggregation levels (5-minute, 10-minute, and 15-minute) were compared. The impacts of different periodic functions on machine learning and statistical models under different aggregation levels (5-minute, 10-minute, and 15-minute) are also investigated. Several important conclusions can be drawn based on the results. First, the periodic functions can improve the prediction performance of machine learning models for more than 60 minutes ahead prediction and improve the over 30 minutes ahead prediction accuracy of all statistical models. Second, the considered three periodic functions have slight difference in improving prediction accuracy of the six prediction models during multistep ahead prediction. Third, with the increase of prediction steps, the impact of periodic function on the prediction model becomes obvious. Fourth, for the same prediction step, the effect of periodic function is more obvious with the increase of data aggregation level. For future work, since nonrecurrent events (incidents, special events, etc.) may disturb the cyclical pattern of travel time, it will be interesting to analyze and compare the impacts of periodic functions on prediction models under nonrecurrent traffic conditions. In addition, artificial intelligence has greatly promoted the development of traffic science. Especially deep learning algorithms, such as deep residual networks, cyclic neural networks and convolutional neural networks, have been rapidly developed in transportation field. It is also interesting to examine the impact of different periodic functions on deep learning algorithms.

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 research was funded by the National Key Research and Development Program of China (grant no. 2018YFE0102800).

References

[1] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, 2014.

[2] J. Xia, M. Chen, and W. Huang, "A multistep corridor travel-time prediction method using presence-type vehicle detector data," *Journal of Intelligent Transportation Systems*, vol. 15, no. 2, pp. 104–113, 2011.

[3] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: overview of objectives and methods," *Transport Reviews*, vol. 24, no. 5, pp. 533–557, 2004.

[4] M. Elhenawy and H. A. Rakha, *Expected Travel Time and Reliability Prediction Using Mixture Linear Regression*, Transportation Research Board, Washington DC, USA, 2016.

[5] X. Zhang and J. A. Rice, "Short-term travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 3-4, pp. 187–210, 2003.

[6] A. Guin, "Travel time prediction using a seasonal autoregressive integrated moving average time series model," in *Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference*, September 2006.

[7] S. Ishak and H. Al-Deek, "Performance evaluation of short-term time-series traffic prediction model," *Journal of Transportation Engineering*, vol. 128, no. 6, pp. 490–498, 2002.

[8] Z. Tongyu, K. Xueping, L. Weifeng, Z. Yuan, and D. Bowen, "Travel time prediction for float car system based on time series," in *Proceedings of the 2010 12th International Conference on Advanced Communication Technology (ICACT)*, IEEE, Phoenix Park, South Korea, February 2010.

[9] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.

[10] Q. F. Yang, X. W. Wei, C. Y. Lin, Z. L. Li, and X. Y. Liu, "Reliability prediction of travel time based on spatio-temporal

bayesian model," *Journal of South China University of Technology (Natural Science Edition)*, vol. 44, pp. 115–122, 2016.

[11] S. I. Chien and C. M. Kuchipudi, "Dynamic travel time prediction with real-time and historic data," *Journal of Transportation Engineering*, vol. 129, pp. 609–616, 2003.

[12] L. Chu, J.-S. Oh, and W. Recker, *Adaptive Kalman Filter Based Freeway Travel Time Estimation*, Transportation Research Board, Washington DC, USA, 2005.

[13] W.-C. Hong, "Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm," *Neurocomputing*, vol. 74, no. 12-13, pp. 2096–2107, 2011.

[14] W. Chun-Hsin, H. Jan-Ming, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, pp. 276–281, 2004.

[15] K. L. Li, Y. H. Ma, Y. M. Tian, and J. Xie, "An improved LS-SVR ensemble learning in internet traffic prediction," *Applied Mechanics and Materials*, vol. 121-126, pp. 3794–3798, 2011.

[16] J. W. C. V. Lint, "Reliable real-time framework for short-term freeway travel time prediction," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 132, pp. 921–932, 2006.

[17] J. W. C. V. Lint, "Online learning solutions for freeway travel time prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, pp. 38–47, 2008.

[18] J. Y. Wang, K. I. Wong, and Y. Y. Chen, "Short-term travel time estimation and prediction for long freeway corridor using NN and regression," in *Proceedings of the 2012 15th International IEEE Conference on Intelligent Transportation Systems*, September 2012.

[19] X. Zeng and Y. Zhang, "Development of recurrent neural network considering temporal-spatial input dynamics for freeway travel time modeling," *Computer-aided Civil and Infrastructure Engineering*, vol. 28, no. 5, pp. 359–371, 2013.

[20] J. Wang, I. Tsapakis, and C. Zhong, " A space -time delay neural network model for travel time prediction," *Engineering Applications of Artificial Intelligence*, vol. 52, pp. 145–160, 2016.

[21] X. Yang, Y. Zou, J. Tang, J. Liang, and M. Ijaz, "Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models," *Journal of Advanced Transportation*, vol. 2020, Article ID 9628957, 16 pages, 2020.

[22] J. Tang, L. Zheng, C. Han et al., "Statistical and machine-learning methods for clearance time prediction of road incidents: a methodology review," *Analytic Methods in Accident Research*, vol. 27, Article ID 100123, 2020.

[23] Y. Zou, X. Zhu, Y. Zhang, and X. Zeng, "A space-time diurnal method for short-term freeway travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 33–49, 2014.

[24] N. Zou, J. Wang, and G. L. Chang, "A reliable hybrid prediction model for real-time travel time prediction with widely spaced detectors," in *Proceedings of the 2008 11th International IEEE Conference on Intelligent Transportation Systems*, October 2008.

[25] C. M. Kuchipudi and S. I. J. Chien, "Development of a hybrid model for dynamic travel-time prediction," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1855, no. 1, pp. 22–31, 2003.

[26] A. Stathopoulos, L. Dimitriou, and T. Tsekeris, "Fuzzy modeling approach for combined forecasting of urban traffic flow," *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 7, pp. 521–535, 2008.

[27] E. I. Vlahogianni, "Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics," *Journal of Intelligent Transportation Systems*, vol. 13, no. 2, pp. 73–84, 2009.

[28] H. Jiang, Y. Zou, S. Zhang, J. Tang, and Y. Wang, "Short-term speed prediction using remote microwave sensor data: machine learning versus statistical model," *Mathematical Problems in Engineering*, vol. 2016, Article ID 9236156, 13 pages, 2016.

[29] Y. Zou, X. Hua, Y. Zhang, and Y. Wang, "Hybrid short-term freeway speed prediction methods based on periodic analysis," *Canadian Journal of Civil Engineering*, vol. 42, no. 8, pp. 570–582, 2015.

[30] J. Tang, H. Wang, Y. Wang, X. Liu, and F. Liu, "Hybrid prediction approach based on weekly similarities of traffic flow for different temporal scales," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2443, no. 1, pp. 21–31, 2014.

[31] C. Chen, Y. Wang, L. Li, J. Hu, and Z. Zhang, "The retrieval of intra-day trend and its influence on traffic prediction," *Transportation Research Part C: Emerging Technologies*, vol. 22, pp. 103–118, 2012.

[32] D. Park, L. R. Rilett, B. J. Gajewski, C. H. Spiegelman, and C. Choi, "Identifying optimal data aggregation interval sizes for link and corridor travel time estimation and forecasting," *Transportation*, vol. 36, no. 1, pp. 77–95, 2009.

[33] E. Vlahogianni and M. Karlaftis, "Temporal aggregation in traffic data: implications for statistical characteristics and model choice," *Transportation Letters*, vol. 3, no. 1, pp. 37–49, 2011.

[34] F. Qiao, X. Wang, and L. Yu, *Optimizing Aggregation Level for Its Data Based on Wavelet Decomposition*, Transportation Research Board, Washington D.C., 2003.

[35] X. Liu, X. Fang, Z. Qin, C. Ye, and M. Xie, "A short-term forecasting algorithm for network traffic based on Chaos theory and SVM," *Journal of Network and Systems Management*, vol. 19, no. 4, pp. 427–447, 2011.

[36] Y. Kamarianakis, H. Oliver Gao, and P. Prastacos, "Characterizing regimes in daily cycles of urban traffic using smooth-transition regressions," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 821–840, 2010.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 1360491, 14 pages
https://doi.org/10.1155/2020/1360491

Research Article

A Generalized Dynamic Potential Energy Model for Multiagent Path Planning

Liu He,[1,2] Haoning Xi,[3] Tangyi Guo,[1,2] and Kun Tang[1,2]

[1]*Department of Automation, Nanjing University of Science and Technology, Jiangsu 210094, China*
[2]*MIIT Key Lab of Traffic Information Fusion & System Control, Nanjing, China*
[3]*Research Center for Integrated Transport Innovation (RCITI), School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia*

Correspondence should be addressed to Haoning Xi; haoning.xi@unsw.edu.au

Received 3 December 2019; Revised 24 April 2020; Accepted 15 June 2020; Published 24 July 2020

Academic Editor: Lu Gao

Path planning for the multiagent, which is generally based on the artificial potential energy field, reflects the decision-making process of pedestrian walking and has great importance on the field multiagent system. In this paper, after setting the spatial-temporal simulation environment with large cells and small time segments based on the disaggregation decision theory of the multiagent, we establish a generalized dynamic potential energy model (DPEM) for the multiagent through four steps: (1) construct the space energy field with the improved Dijkstra algorithm, and obtain the fitting functions to reflect the relationship between speed decline rate and space occupancy of the agent through empirical cross experiments. (2) Construct the delay potential energy field based on the judgement and psychological changes of the multiagent in the situations where the other pedestrians have occupied the bottleneck cell. (3) Construct the waiting potential energy field based on the characteristics of the multiagent, such as dissipation and enhancement. (4) Obtain the generalized dynamic potential energy field by superposing the space potential energy field, delay potential energy field, and waiting potential energy field all together. Moreover, a case study is conducted to verify the feasibility and effectiveness of the dynamic potential energy model. The results also indicate that each agent's path planning decision such as forward, waiting, and detour in the multiagent system is related to their individual characters and environmental factors. Overall, this study could help improve the efficiency of pedestrian traffic, optimize the walking space, and improve the performance of pedestrians in the multiagent system.

1. Introduction

Multiagent system used in the simulation reflects the psychological and physical properties of pedestrians. Walking is a hot topic in traffic research in recent years [1]. Walking is a sustainable mode in transportation with little space, no resource consumption, and no environmental pollution [2]. A growing number of recent studies have focused on improving the sustainability of transportation systems by routinely converting motorized travel to walking [3]. With the general change of the concept from vehicle-oriented to people-oriented traffic, related studies on pedestrian traffic are gradually increasing. In order to identify effective strategies for improving the efficiency of pedestrians in a multiagent system, planners need to identify how the current personal characters affect walking in transportation. Literature studies [1–4] conducted quantitative analysis to optimize the walking space. Literature studies [5–7] analysed the importance and effectiveness of walking in transportation through empirical case studies and system simulations. Eady and Burtt investigated the role of walking and access to suburban shopping centres will affect people traveling around Melbourne [5]. Bangcheng et al. considered the stability of the walking system and the rationality of its structural design in the study of dynamics for a humanoid robot [6]. Yamazaki et al. proposed an evacuating agent walking model and the estimated distribution of the evacuation based on the real evacuation data [7].

Existing quantitative research on pedestrian traffic mainly focuses on system simulation, which can be divided into discrete and continuous models. Cellular automaton (CA) [8] model is a type of discrete model, which divides pedestrian walking areas into cells. Social force model is a type of continuous model, which defines pedestrians in a continuous space and describes the pedestrians' behaviours through various forces. Both models are easy to understand, but cannot reflect the difference among individual characteristics of pedestrians. Dynamic pedestrian planning is the core of pedestrian simulation, and the main method of path planning is to establish a potential energy field [9]. The robot moves in the gradient direction, in which the potential energy drops fastest to avoid obstacles. Deepak Subramani and Lermusiaux combined the decision theory with the essential stochastic time-optimal path planning to establish the uncertain, strong, and dynamic risk-optimal path planning scheme based on partial differential equations [10]. Antonio Sedeño-noda and Colebrook extended the Dijkstra algorithm to the two-objective shortest path problem [11]. Rui Song et al. established a smooth A $*$ model for path planning of unmanned vehicles (UVs) [12]. Prahlad et al. combined the genetic algorithm with the global search function and the artificial potential energy field to construct the EAPF (evolutionary artificial potential field) algorithm; this method is affected by the population number and algebra of the genetic algorithm. The operation efficiency is low [13]. Ulises Orozco-Rosas et al. constructed a membrane evolution potential energy field based on a multiprocessor, which improves the computational efficiency of robot path planning [14]. Deepak N. Subramani et al. addressed the optimal path planning problem in a stochastic dynamic environment by combinatorial decision-making and time optimization theory [10], ant colony algorithm [15], simulated annealing method [16], firefly algorithm [17], and Q-learning [18]. The studies mentioned above mainly focus on the optimal path planning in various dynamic scenes under certain perceptual conditions from the view of system control, and they are not suitable for pedestrian simulation with physiological, psychological, and social uncertainties.

Pieces of research on pedestrian path planning in the multiagent system belong to the artificial intelligence field. Multiagent system is composed of multiple interacting elements, known as agents. They can decide how to satisfy their objectives and interact with other agents by engaging in analogues of the kind of social activity: cooperation, coordination, and negotiation [19]. Since Kinny et al. introduced the multiagent into the air traffic control system [20], the multiagent has been a research focus in the field of transportation. Compared with vehicles, pedestrians have the characteristics of blindness (blindness means when the crowd is highly concentrated, especially when the situation is unclear, the pedestrian tends to follow the crowd blindly due to the herd mentality), purpose (purpose means when pedestrians have a clear destination, they will care about facilities, guidance systems, etc., so as to arrive as soon as possible; otherwise, it is easier to be attracted by the surrounding environmental factors), randomness, commonality, and so on, Hence, it is very important to highlight the process of independent decision-making and group game in path planning in a multiagent system. These research studies proposed several prototype techniques for agent systems in transportation management, including methodologies for agent-oriented analysis and design, formal specification and verification methods for agent systems, and techniques for implementing agent specifications [21–23].

Existing studies only analyze the speed characteristics of pedestrian flow in the multiagent system [24–27] or study the large-scale macrocollection and distribution capacity of pedestrians [28–30]. Few research studies focus on the mechanism of path planning for the multiagent at mesolevel and consider different factors such as gender and luggage. Moreover, the deadlock problem always arises in traditional cellular automata simulations. In order to fill in these research gaps, this paper constructed the generalized dynamic potential energy field with large cells and small time segments to improve the discrete artificial potential energy field.

To the best of the author's knowledge, few studies establish a dynamic simulation model at the mesolevel to reflect the personal characteristics of each individual in a multiagent system. Hence, based on the disaggregation characteristics of the multiagent system, we establish a generalized dynamic potential energy model (DPEM) efficiently and accurately at mesolevel to strengthen the disaggregation characteristic of the multiagent system and reflect the characteristics such as gender and carry-on luggage by considering the characteristics of each individual as key parameters. The results of the case study validate the effectiveness of the proposed dynamic potential energy model (DPEM) and indicate that each agent's path planning decision such as forward, waiting, and detour in the multiagent system is related to their characters and environmental factors.

The rest of the paper is organized as follows: Section 2 presents the procedure of establishing a generalized dynamic penitential energy model (DPEM) through 4 steps. Section 3 conducts a case study to illustrate the feasibility and effectiveness of the proposed DPEM. Section 4 concludes with a summary of main findings in this research and suggestions for future study.

2. Methodology

2.1. Notations and Preliminaries. Table 1 summarizes the parameters, variables, and abbreviations used in this paper to describe the mathematical models.

A pedestrian in the simulation system can be regarded as a multiagent; HCM2010 defines a pedestrian as an ellipse of $0.46\,\mathrm{m} \times 0.61\,\mathrm{m}$, with a contact area of $0.28\ \mathrm{m}^2$ and a noncontact area of $0.66\ \mathrm{m}^2$. In Figure 1, we measured the space occupancy of pedestrians of different genders in five states of luggage: none, carry a bag, carry a knapsack, suitcase (stationary), and suitcase (moving). To simplify the model, the pedestrians defined in this paper do not include the elderly and children.

In CA models, the length of a cell is 0.1 m to 0.5 m. Whether the cell is occupied by a pedestrian or obstacles is denoted as a {0, 1} variable. In this paper, the length of the

Table 1: Parameters, variables, and abbreviations.

q	The ID of an agent	(i, j)	The location of the cell
k	Time step of simulation	dir	The direction of the agent walking ahead
t_k	Time of kth time step	$\tau_{ij}(\rho_q, \mu)$	The velocity decline of agent q with space occupation ρ walking through the cell with μ
ρ	Space occupation by the agent	$G(\text{dir})$	The distance to the adjacent cell by direction dir
μ	Space occupied by the obstacle	α_q	The waiting potential energy dissipation coefficient of agent q
$\Delta\mu$	Change rate of μ during a time step	β_q	The waiting potential energy enhancement coefficient of agent q
v_q	The expected speed of agent q.	$V_{ij}(q)$	Space potential energy of agent q in cell (i, j)
SO	Stationary obstacle	$T_{ij}(q, t_k)$	Delay potential energy of agent q in cell (i, j) at t_k
MG	Male group	$W_{ij}(q, t_k)$	Waiting potential energy of agent q in cell (i, j) at t_k
FG	Female group	$U_{ij}(q, t_k)$	Generalized dynamic potential energy of agent q in cell (i, j) at t_k

Figure 1: Pedestrians of different genders in five states of luggage.

cell is set as 0.8 m, which is larger than the normal cell defined in CA models, and a cell can accommodate multiple pedestrians or obstacles according to its space occupancy. The pedestrian space of pedestrians is shown in Table 2.

In the discrete simulation models, an agent can only reach the cells in his neighbourhood; this paper chooses the Moore neighbourhood and assumes that pedestrians can only reach one of the eight adjacent directional cells.

According to the relevant literature, the pedestrian walking speed is between 0.4 m/s and 1.4 m/s [31]. So, it took 0.6 s~2 s for a cell to reach its adjacent vertical cell and 0.8 s~2.8 s to the oblique adjacent cell. Therefore, the simulation time step of the model is set as the maximum common divisor: 0.2 s. The actual walking time between the cell is commonly determined by the speed of the pedestrian and the number of rows in the cell. In this paper, we focus on a new type of cell, whose length is set as 0.8 m and time step is set as 0.2 s, and is vividly defined as the "pancake cell."

2.2. Problem Description. Walking is a dynamic process of pedestrian path-planning decisions and displacements. Ideally, pedestrians determine their forward direction based on the shortest distance or time while avoiding other pedestrians and obstacles. Figure 2(a) shows an agent's walking area, in which the yellow part is an isolation bar. There is a gap (11, 13) on the west side of the isolation bar, which is the bottleneck of the pedestrian flow [32]. An agent travels from the (2, 2) to the blue cell (24, 8). During walking, gap (10, 13) is temporarily occupied by other agents. According to the path planning principle , the agent falls into the local minimum point of the artificial potential field. At this point, the agent faces three options:

(a) Delay: if the delay time is much lower than the detour time, the agent will wait and increase the delay time in the obstacle cell

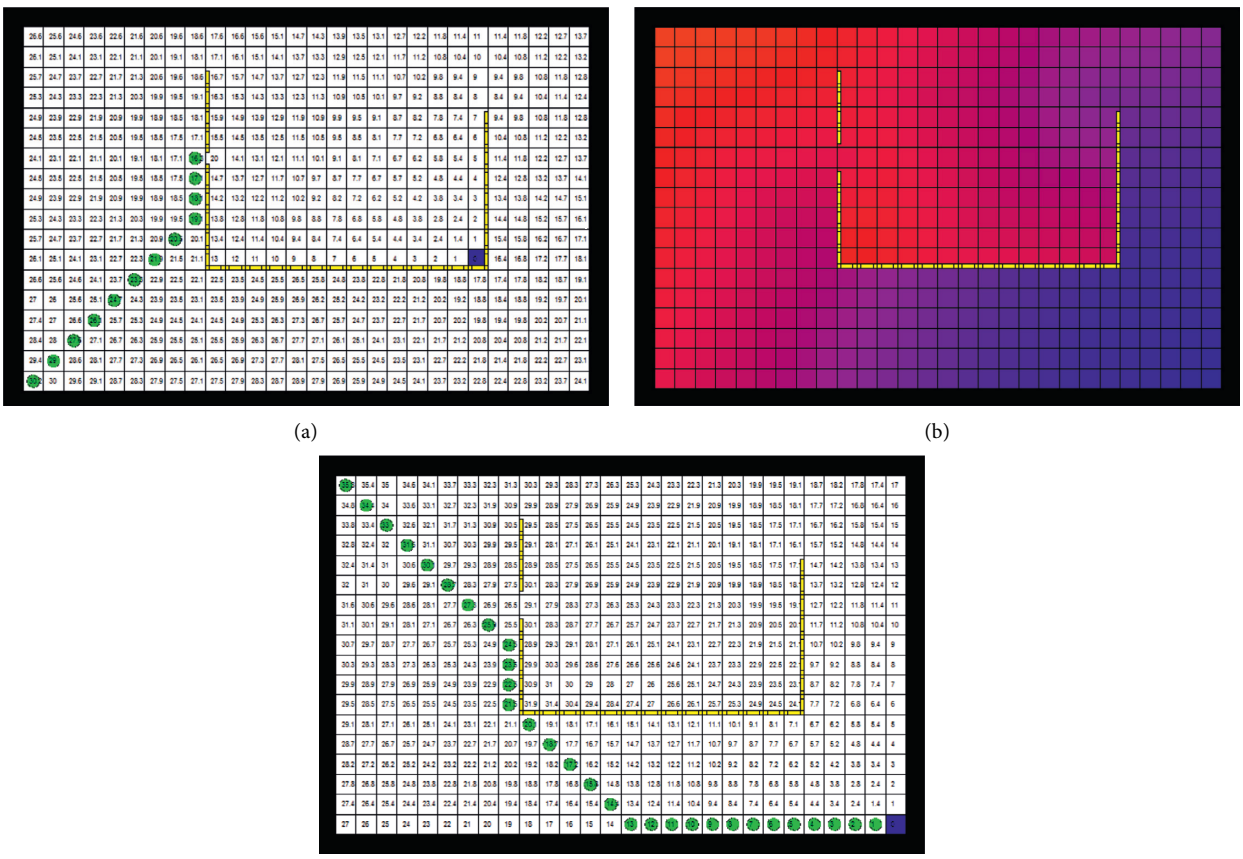
(b) Detour: if the delay time is larger than the detour time, the agent will change other routes

(c) Waiting: an agent will wait in place for a certain period before reselection

Therefore, the dynamic potential energy model (DPEM) for the local minimum problem should reflect three characteristics: (1) an agent will make choices among delay detour and waiting; (2) an agent will jump out once he falls into the local minimum; and (3) an agent will not return

Table 2: Space occupancy of different pedestrians.

Gender	Luggage	Width (m)	Distance (m)	Area (m^2)	Space occupancy (ρ) (%)
Male	None	0.55	0.4	0.17	27
	Carry a bag	0.6	0.45	0.21	33
	Carry a knapsack	0.55	0.5	0.22	34
	Suitcase (stationary)	0.8	0.5	0.31	48
	Suitcase (moving)	0.8	0.95	0.60	94
Female	None	0.45	0.3	0.11	17
	Carry a bag	0.55	0.4	0.17	27
	Carry a knapsack	0.45	0.4	0.14	22
	Suitcase (stationary)	0.75	0.5	0.29	45
	Suitcase (moving)	0.75	0.95	0.56	88



(a) (b) (c)

Figure 2: Static artificial potential energy field and walking trajectory.

back after jumping out of the local minimum cell. The effect of delay and waiting cost should be considered besides the static space in the potential energy field.

2.3. Model Formulation. In order to reflect the actual behaviours of pedestrians such as delay, detour, and waiting, we obtain the generalized dynamic potential energy by superimposing the delay potential energy, waiting potential energy, and static space potential energy all together. Then, we investigate the functional relationship between individual attributes and potential energy of the multiagent to reflect pedestrians' different decisions for path planning.

2.3.1. Space Potential Energy. In the space potential energy field V, the spatial potential energy is determined by the distance from the current position to the destination. Since the space potential energy is only related to the Euclidean distance and is static, it is called the static space potential energy.

Consider the path planning problem of the multiagent to multidestinations; if there are m agent origins from different cells O_i, $i \in m$, to the destination cell D_j, $j \in n$, the traditional path search algorithm will calculate $m \times n$ times in each time step. In order to improve the computational efficiency, we can start from the destination set $\{D\}$ to search the cells in all possible directions and record the distance to any cell.

In each step, when pedestrians update the current cell coordinates, at least one cell that is nearer to the destination can be found in the neighbourhood. Pedestrians will choose the nearest one as the forward direction and finally reach the nearest cell in the destination set. Since all of the points in this region are search targets, the number of searches is fixed, and the Dijkstra algorithm can be improved directly.

Here is the procedure of the improved Dijkstra algorithm:

(i) Step 1: set all cells which meet $\mu > 1 - \rho$ to be -1, and the others have no potential energy.

(ii) Step 2: traverse all the cells to find the cell b satisfying the following conditions:

Cell b has no potential energy; there is at least one cell with positive potential energy in the neighbourhood of the cell. The two cells are connectable. The cell is assigned a temporary value $\widehat{V}_b = V_a + G_{ab}$, where V_a is the potential energy of the adjacent cell and G_{ab} is the distance between two adjacent cells, whose value is 1 or 1.414. If there is more than one cell in the neighbourhood which satisfies the above conditions, choose a smaller value, namely, $\widehat{V}_b = \min(V_{a_i} + G_{a_i b}), i = 1, 2, \ldots, 8$; then, all the satisfied cells will create a set {B}.

(iii) Step 3: replace the potential energy value with the smallest temporary value in $\{B\}$, $V_b = \widehat{V}_b$.

(iv) Step 4: if all the cells have potential energy, end; otherwise, return to Step 2.

The simulation process of the improved Dijkstra algorithm is shown in Figure 3.

Here, we consider a case where an agent will walk from the upper left corner to the lower right corner. Figure 2(b) displays the distribution of the static potential energy field, where blue colour represents the lower potential energy value and red colour represents the higher value. Figure 2(c) displays the potential energy and walking trajectory of an agent.

2.3.2. Delay Potential Energy

(1) Obstacle Delay Function. The open space cell has no constraints on an agent's walking path choice; an agent can enter from any direction in the open space cell and walk through with the desired speed, and thus, the initial space occupancy of an open space cell is zero. Moreover, the obstacles such as walls, foundation columns, isolation bars, and other pedestrians have impacts on an agent's path choice. The agent will initiatively stay away from the walls, bypass the columns and facilities, and keep their distance from each other. From the perspective of the space occupation of the cell, the walking environment can be divided into four categories.

In Table 3, ρ and μ represent the space occupancy of an agent and obstacle, respectively. The cell with no obstacles does not occupy space and has no effects on path planning of the agent. The complete obstacle cell is repulsive to the agent from any direction. A channelized obstacle, such as an isolation bar, is considered to be at the edge of a specified direction of the cell to prevent an agent from walking toward in the specified direction, but there are no constraints on pedestrians in other directions.

The delay potential energy reflects the time consumption of pedestrians while passing through the cell. In this paper, a continuous variable $\mu \in [0, 1]$ is used to indicate whether the cell is occupied or not, instead of 0-1 variables in the traditional methods. When an agent with a body size of ρ walks through a cell occupying the space of μ, there will be delays due to lateral avoidance. Delays are affected by the following two factors: (1) the "body type" of an agent is determined by the gender and the number of carry-on luggage; (2) the space occupation of the cell.

Then, a cross-observation experiment is carried out to explore the regulation of speed decline rate. 40 student participants from the author's college were evenly divided into two groups by gender. Place cone buckets of different sizes or stand on other participant groups to change the occupancy of the cell every time. Each group of participants is required to carry different luggage through the cell in turn. Record the time consumed by each participant from entering to leaving the cell. Divide the time passed by empty hands by the time passed by this round to get the speed decline rate. If the participant cannot pass the cell, it is set as 0. Table 4 shows the results of the experiment.

Through increasing the space occupation of the obstacles from 0% to 100%, the functional relationship between speed decline rate and space occupation for both males and females is fitted and shown in Figure 4.

The experimental results (Figure 4) show that the speed decline rate decreases with the obstacle occupancy (μ) for the agent with different occupancy (ρ). If $\mu + \rho \longrightarrow 1$, the rate of speed decline rate will dramatically increase. Different types of obstacles have different impacts on the speed of the agent, among which the stationary obstacle has the least impact on obstacles. Through the experimental observation, a male agent has a higher spatial tolerance to obstacles and can adjust their posture and luggage position more flexibly to get through obstacles quickly. As a result, although the value of ρ for the male agent is larger than that for the female agent, the speed of a male agent is less affected by obstacles compared with that of the female agent. Moreover, since the space of the agent with a suitcase is more flexible, it is possible for the agent to walk through the cell when $\mu + \rho > 1$.

Figure 5 displays the curve of τ for male and female agents, and the fitting formulas of the velocity decline rate $\tau(\rho, \mu)$ for the male agent are given as

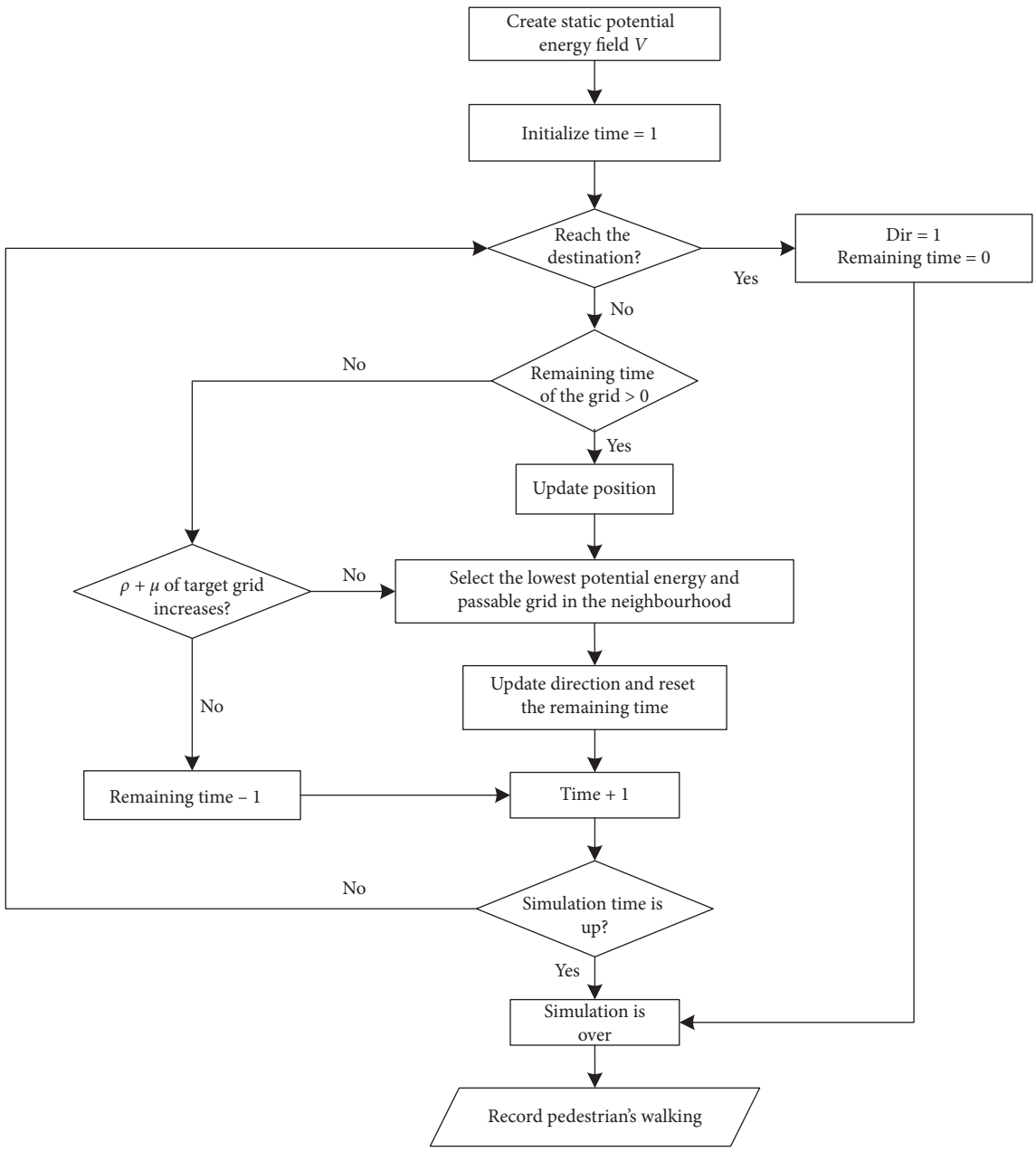


Figure 3: Flowchart of the improved Dijkstra algorithm.

Table 3: Obstacle classification.

Type	Space occupation μ	Impact on the cell	Example
No obstacle	$\mu = 0$	None	Open space
Complete obstacle	$\mu \geq 1 - \rho$	Prevents entry from any direction	Walls
Canalized obstacle	$\mu = 0$	Prevents entry from a particular direction	Isolation bar
Partial obstacle	$0 < \mu < 1 - \rho$	Slows down the walking speed	Dustbin

Table 4: Relationship between speed decline rate and space occupation.

Sex	Luggage	ρ (%)	Obstacle type	Obstacle occupation μ										
				0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Male	None	27	Stationary obstacle (SO)	100%	98%	95%	84%	76%	61%	40%	10%	1%	0%	0%
	Bag	33		100%	93%	90%	78%	67%	52%	21%	3%	1%	0%	0%
	Knapsack	34		100%	95%	91%	80%	71%	57%	33%	7%	1%	0%	0%
	Suitcase	94		100%	61%	24%	11%	3%	2%	0%	0%	0%	0%	0%
Female	None	17		100%	96%	90%	81%	72%	54%	35%	12%	2%	0%	0%
	Bag	27		100%	92%	88%	75%	62%	47%	15%	2%	0%	0%	0%
	Knapsack	22		100%	94%	86%	78%	65%	50%	22%	3%	1%	0%	0%
	Suitcase	88		100%	35%	5%	2%	1%	1%	0%	0%	0%	0%	0%
Male	None	27	Male group (MG)	100%	95%	88%	75%	63%	49%	33%	10%	3%	1%	0%
	Bag	33		100%	90%	80%	69%	54%	38%	15%	5%	2%	0%	0%
	Knapsack	34		100%	91%	82%	71%	58%	42%	20%	6%	2%	0%	0%
	Suitcase	94		100%	40%	12%	5%	2%	1%	1%	0%	0%	0%	0%
Female	None	17		100%	90%	78%	47%	21%	7%	2%	0%	0%	0%	0%
	Bag	27		100%	85%	70%	41%	18%	5%	1%	0%	0%	0%	0%
	Knapsack	22		100%	88%	72%	42%	18%	6%	1%	0%	0%	0%	0%
	Suitcase	88		100%	40%	10%	5%	4%	3%	2%	1%	0%	0%	0%
Male	None	27	Female group (FG)	100%	95%	87%	69%	47%	30%	8%	0%	0%	0%	0%
	Bag	33		100%	92%	82%	60%	38%	15%	5%	0%	0%	0%	0%
	Knapsack	34		100%	92%	83%	65%	41%	15%	2%	0%	0%	0%	0%
	Suitcase	94		100%	33%	8%	5%	3%	0%	0%	0%	0%	0%	0%
Female	None	17		100%	94%	82%	72%	54%	41%	30%	12%	5%	0%	0%
	Bag	27		100%	88%	75%	58%	41%	29%	12%	2%	1%	0%	0%
	Knapsack	22		100%	88%	76%	60%	43%	32%	18%	2%	1%	0%	0%
	Suitcase	88		100%	35%	10%	3%	2%	1%	0%	0%	0%	0%	0%

$$\tau(\rho,\mu) = \begin{cases} 1, & \mu = 0, \\ 0, & \rho + \mu \geq 1, \\ -1.452(\rho+\mu)^2 + 0.5893(\rho+\mu) + 1, \text{ type} = \text{SO}, & \rho + \mu < 1, \\ -1.297(\rho+\mu)^2 + 0.343(\rho+\mu) + 1, \text{ type} = \text{MG}, & \rho + \mu < 1, \\ -1.616(\rho+\mu)^2 + 0.4191(\rho+\mu) + 1, \text{ type} = \text{FG}, & \rho + \mu < 1. \end{cases} \tag{1}$$

The fitting formulas of the velocity decline rate $\tau(\rho,\mu)$ for the female agent are given as

$$\tau(\rho,\mu) = \begin{cases} 1, & \mu = 0, \\ 0, & \rho + \mu \geq 1, \\ -1.1947(\rho+\mu)^2 + 0.1656(\rho+\mu) + 1, \text{ type} = \text{SO}, & \rho + \mu < 1, \\ -0.795\ln(\rho+\mu) - 0.0748, \text{ type} = \text{MG}, & \rho + \mu < 1, \\ -0.6275(\rho+\mu)^2 - 0.3902(\rho+\mu) + 1, \text{type} = FG, & \rho + \mu < 1. \end{cases} \tag{2}$$

As shown in Figure 5, all five curves had R^2 fitting values greater than 0.9, except for the curve representing women crossing the male group (MG), and this may be because women are more repelled by the opposite gender. Therefore, the time that the agent reaches any adjacent cell can be expressed as

$$T_{ij}(q) = \frac{G(\text{dir})}{v_q \tau_{ij}(\rho_q, \mu)}, \tag{3}$$

where $G(\text{dir})$ represents the distance to the adjacent cell and is related to the direction of the agent and v_q denotes the expected speed of agent q.

(2) Delay Potential Energy Function. The static potential energy value is determined by the spatial distance between the cell and the destination. The pedestrian chooses the cell with the lower potential energy value, which is closer to the destination. Assuming that the pedestrians will walk at the

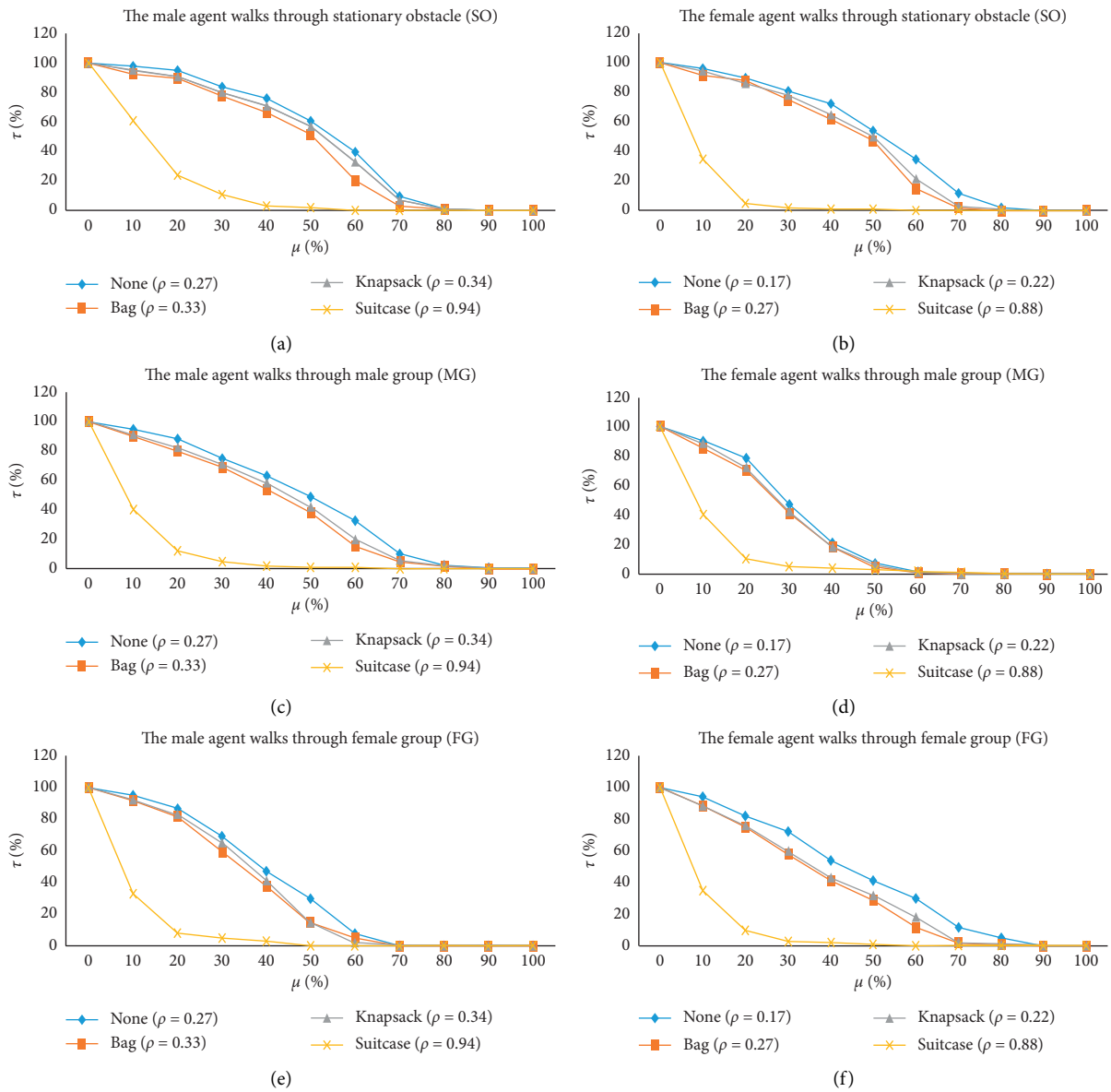


Figure 4: Experimental curve of speed drop and space occupation.

expected speed, the shortest path is considered both the shortest distance and time to unify the time and space.

After considering the influence of obstacles, the actual speed of the agent is gradually lower than the expected speed, and the relationship between walking distance and time is nonlinear (Figure 6). In this case, shortening the walking distance will increase the time, the agent needs to walk for more distance to save the time, and thus, the multipath problem will arise. Therefore, the microscopic path-planning problem can be transformed into a multiobjective optimization problem considering the influence of partial obstacles. There are two methods to solve the multiobjective problem: (1) to transform the multiobjective problem into a single objective problem by weighting; (2) to solve the problem based on the Pareto optimal solutions, such as data envelopment analysis (DEA) and heuristic algorithm. Since the two objectives of time and space distance can be converted through velocity, the first method is adopted in this paper.

The agent often makes different choices on the time and space distance. From experience, the agent in a hurry will choose the most time-saving path, while others will choose a relatively short path. We explain this phenomenon as the "degree of laziness," and the generalized potential energy of the agent is that spatial distance and time are weighted by "laziness coefficient." The weighting factor of the distance is larger for the "lazier" agent, and the weighting factor of the time is larger for the diligent agent, so the spatial potential energy can be transformed into the generalized potential energy. In this way, the choice of path for the agent is more objective and diverse.

Then, we explain the "laziness coefficient" from the view of opportunity cost, where the difference lies in the value of time (VOT) of the agent. Moreover, the opportunity cost of

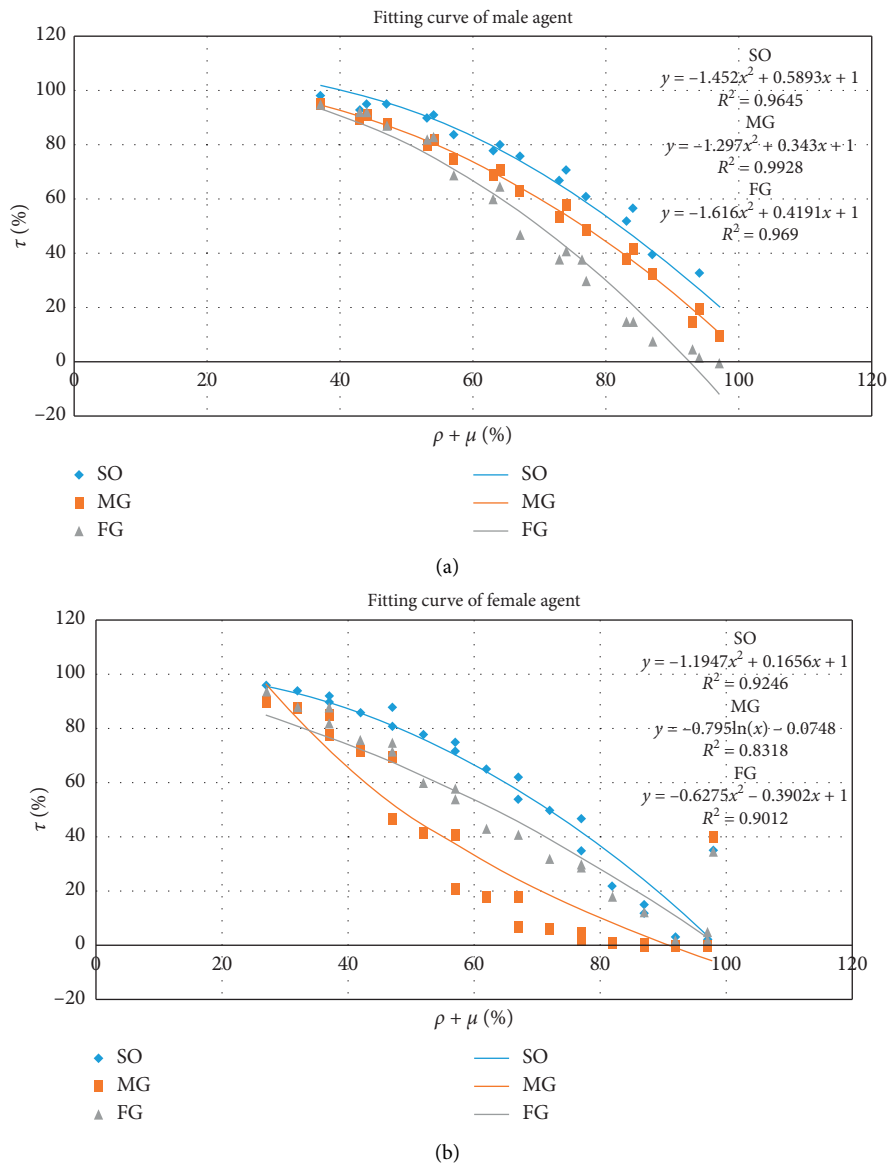


Figure 5: Fitting curve of τ. (a) The relationship between speed decline rate (τ) and (b) space occupation ($\mu + \rho$) for male and female agents, respectively.

time is reflected by the maximum distance that the agent can walk during the delayed time period.

Thus, waiting potential energy $T_{ij}(q)$ can be expressed as

$$T_{ij}(q) = \left(\frac{G(\text{dir})}{\tau_{ij}(\rho_q, \mu) v_q} - \frac{G(\text{dir})}{v_q}\right) v_q = \frac{1 - \tau_{ij}(\rho_q, \mu)}{\tau_{ij}(\rho_q, \mu)} G(\text{dir}), \tag{4}$$

where v_q represents the expected speed of agent q and $\tau_{ij}(\rho_q, \mu)$ represents the coefficient of the velocity decline rate for the partial obstacle in the cell (i, j).

2.3.3. Waiting Potential Energy. Since the agent will lose patience and choose other paths after waiting if other pedestrians have occupied the target cell, the waiting potential energy is introduced in this paper. The waiting potential energy is affected by the characteristics of the agent and changes with time and has two opposite characteristics of dissipation and enhancement.

(1) Dissipation. If the pedestrian has left the cell (i, j), the waiting potential energy W will dissipate over time. At first, the agent still remembers that he has passed through the cell (i, j), and thus, the waiting potential energy can keep pedestrians from returning. When the waiting potential energy dissipates to 0, the agent will forget whether he has passed through the cell.

When agent q leaves the cell (i, j) at t_0, the waiting potential energy of the cell (i, j) can be expressed as $W_{ij}(q, t_k)$ after k simulation time steps:

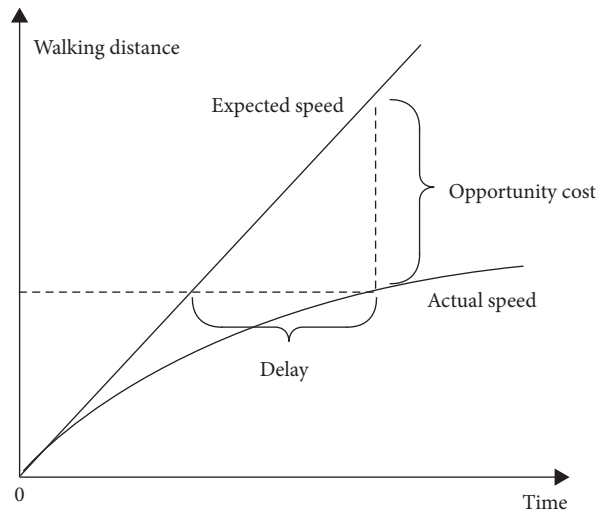


Figure 6: Relationship between time and walking distance.

Table 5: Simulation process of multiagent path planning.

Step	Description
1	Model initialization. Set the generation rate of the agent as α_o and the number of agents as k.
2	Obtain the static spatial potential energy field V for destination set through the improved Dijkstra algorithm.
3	Generalize the delayed potential energy field T for all agents; set the potential energy field $W = 0$ and simulation time $t = 1$.
4	Wait for the dissipation of potential energy W.
5	Set the agent number as $k = 1$.
6	Select the kth agent and jump to Step 13 if the agent has not departed or left the simulation area. Otherwise, move to Step 7.
7	If the agent is not waiting and the remaining time of the cell is greater than 0, then the remaining time will be −1, and jump to Step 13. Otherwise, move on to Step 8.
8	If the agent reaches the destination, set the agent waiting, subtract the space occupation of the agent from the space occupation of the cell; if the waiting potential energy field W becomes 0, then jump to Step 13; otherwise, move on to Step 9.
9	If the current time equals the initial entry time of the agent, strengthen the space occupation in the origin cell. Otherwise, if the direction of the agent is 1, the waiting potential energy will be strengthened.
10	Strengthen the basic waiting potential energy, and update 3 types of potential energy in all directions to obtain the generalized potential energy U by superposition.
11	Determine the target cell according to the cell probability selection model.
12	If the cell is a target cell, the agent will choose to stay, and set the agent waiting. Otherwise, update the direction and remaining time.
13	If $k = K$, move on to the next step. Otherwise, $k = k + 1$, and return to Step 6.
14	If $t = t_{\max}$, end. Otherwise, $t = t + 1$, and return to Step 4.

$$W_{ij}(q, t_k) = \alpha_q W_{ij}(q, t_{k-1}) = \alpha_q^2 W_{ij}(q, t_{k-2}) = \alpha_q^k W(q, t_0), \quad 0 < \alpha_q < 1, \tag{5}$$

where $W(q, t_0)$ is the basic waiting potential energy and α_q is the waiting potential energy dissipation coefficient of agent q. The smaller α_q is, the faster the waiting potential energy will dissipate.

The waiting potential energy $W(q, t_k)$ is mainly affected by three variables: basic waiting potential energy $W(q, t_0)$, dissipation coefficient α_q, and enhancement coefficient β_q.

If the agent waits for a time step, he will give up the movement during this time step at the expected speed, and this is also defined as the opportunity cost:

$$W(q, t_0) = 0.2 \cdot v_q(t_k). \tag{6}$$

The dissipation coefficient α_q reflects the perception of the agent. If $\alpha_q \longrightarrow 1$, it will dissipate slowly, and it is impossible for the agent to return. If $\alpha_q \longrightarrow 0$, the agent will return to the same local minimum cell and fall into a dead cycle. From this point of view, α_q should be as large as possible. However, the local minimum of the cell may disappear after a short period; since the agent blocking at the door will leave, it will take a long time for the agent to return. Therefore, the dissipation coefficient also reflects how far the agent is willing to search after falling into a local minimum and successfully jumping out. According to observation experiments, the interval between pedestrian path-planning decisions is about 1 second, and it is related to the gender, age, and luggage of the agent. To simplify the model, let $\alpha_q = 0.9$; after 1 second (5 time steps), the waiting potential energy can be reduced to the original $0.9^5 = 0.59$ times, namely, it will dissipate at a rate of 60% per second.

(2) Enhancement. If the agent stays in the cell (i, j), the waiting potential energy W will increase over time, and

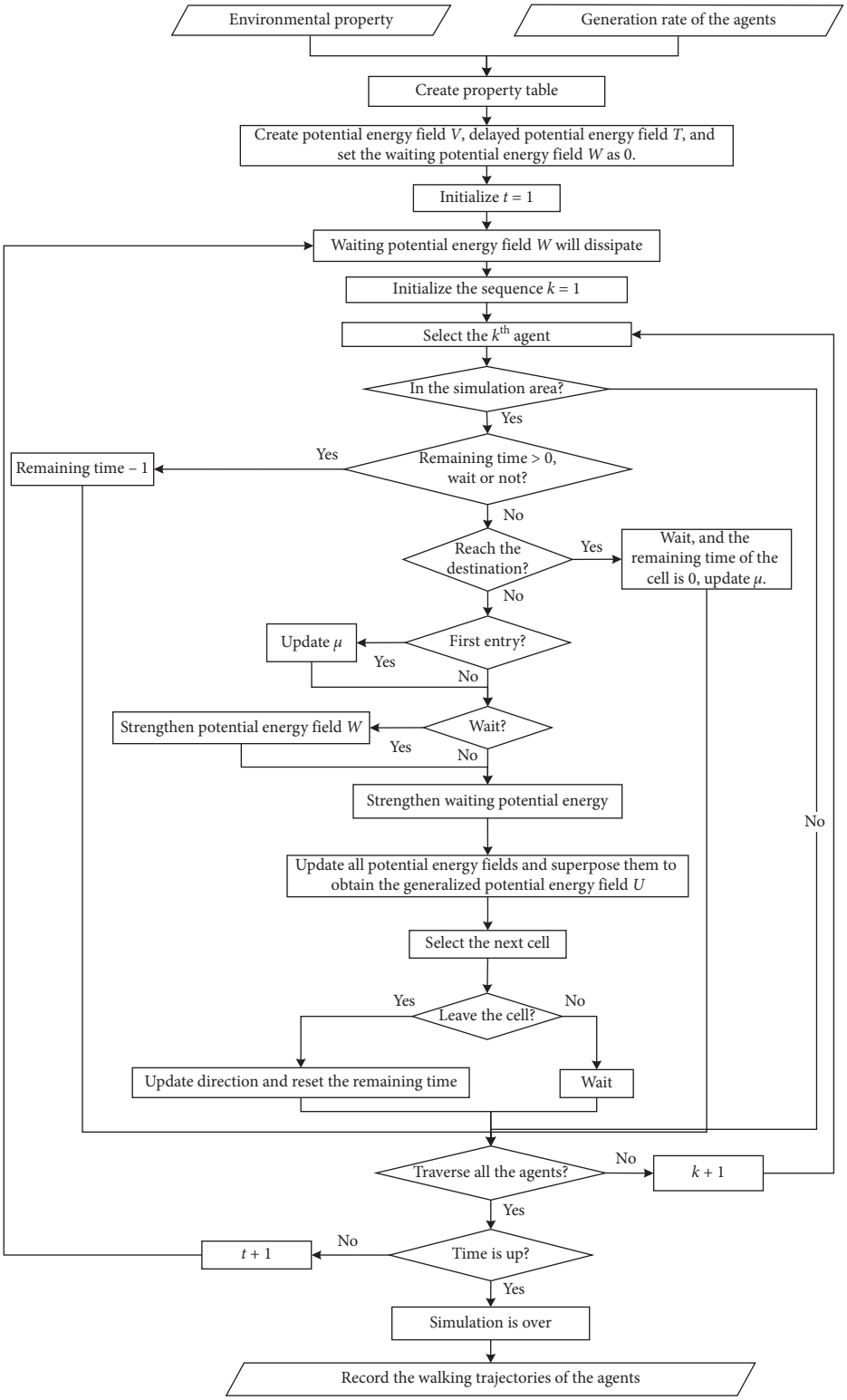
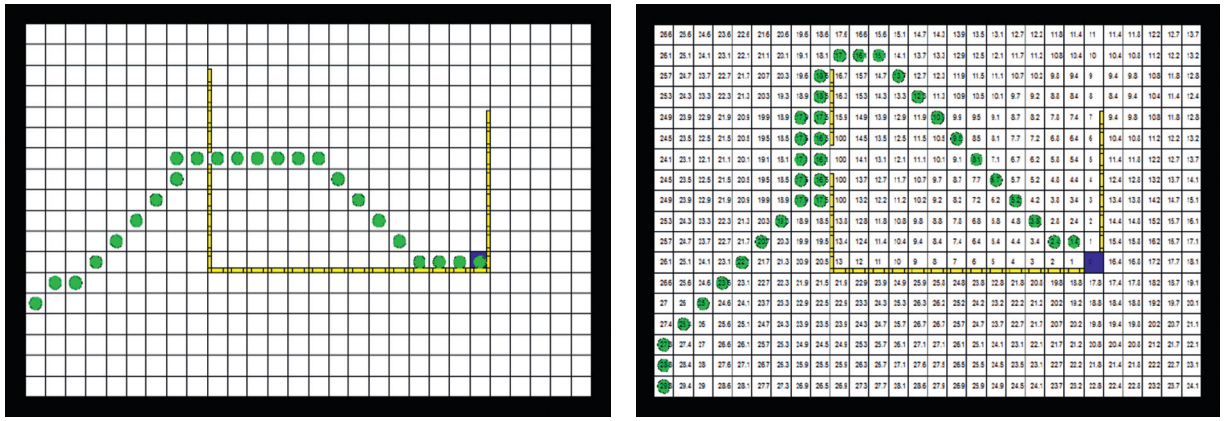


Figure 7: Flowchart of multiagent path planning.

the agent will wait before falling into a local minimum. The longer the waiting time is, the greater the waiting potential energy of the cell will be. When the superimposed potential energy of the cell is higher than that of other cells in the neighbourhood, the agent will successfully jump out of the local minimum. On the contrary, after jumping out of the local minimum, the waiting potential energy of the cell is large and dissipates slowly, so pedestrians will not return in a short time. When pedestrian q enters and stays in the cell (i, j) at t_0,



(a) (b)

Figure 8: Testing the local minimum detour (a) and (b) walking trajectories of the agent, where $\mu = 0.2$ and $\mu = 0.9$, respectively.

after k simulation time steps, the waiting potential energy of the cell (i, j) is expressed as

$$W_{ij}(q, t_k) = \beta_q W_{ij}(q, t_{k-1}) + W_{ij}(q, t_0) = \begin{cases} \dfrac{1-\beta_q^{k+1}}{1-\beta_q} W(q, t_0), & \beta_q > 1, \\ kW(q, t_0), & \beta_q = 1, \end{cases} \tag{7}$$

where β_q is the waiting potential energy enhancement coefficient of agent q. The larger value of β_q represents that the agent is reluctant to wait. When $\beta_q = 1$, the waiting potential energy increases linearly with time; when $\beta_q > 1$, the longer the pedestrians wait, the faster the waiting potential energy increases. When the total potential energy of the superimposed cell is greater than that of the neighbourhood cell, pedestrians successfully jump out of the local minimum. This can also reflect the process that pedestrians gradually lose patience and look for other paths after waiting for a period.

The enhancement coefficient β_q reflects the patience of the agent. The lower value of the coefficient means that the agent is more patient. The decisive factor of β_q is the agent's judgement on velocity decline rate of the occupation in the bottleneck cell, which is also related to the gender, age, and travel purpose of the agent. In order to simplify this model, it is assumed that β_q is only related to the variation of occupation μ in the next time step.

Impatient agent judges that the cell occupation of obstacles is constant and will leave at the next time step after jumping into the local minimum cell, and thus, when $\Delta\mu = \mu(t_{k+1}) - \mu(t_k) = 0$, β_q will reach the maximum $\beta_q = +\infty$. The patient agent thinks that the occupancy of the obstacle cell will be reduced to zero, and thus, when $\Delta\mu = \mu(t_{k+1}) - \mu(t_k) = \mu$, β_q reaches the minimum value $\beta_q = 1$. Therefore, the logarithmic function can be established:

$$\beta_q = \ln \frac{\mu}{\Delta\mu} + 1, \quad \mu \geq \Delta\mu. \tag{8}$$

2.3.4. Generalized Dynamic Potential Energy. The generalized dynamic potential energy $U_{ij}(q, t_k)$ can be obtained by superposing the space potential energy $V_{ij}(q)$, delay potential energy $T_{ij}(q, t_k)$, and waiting potential energy $W_{ij}(q, t_k)$. Moreover, three types of potential energy are unified by distance:

$$U_{ij}(q, t_k) = V_{ij}(q) + T_{ij}(q, t_k) + W_{ij}(q, t_k). \tag{9}$$

The simulation process of the multiagent is shown in Table 5 and Figure 7.

3. Case Study

After superposing the delay potential energy field T and waiting potential energy field W, consider the path planning problem of a male agent without luggage ($\rho = 0.27$) under different space occupation of obstacles (μ). The results obtained from the DPEM state that the agent will make different choices while facing the local minimum cell with different values of μ.

If $\mu = 0.2$, the agent will choose to go through the cell; if $\mu = 0.9$, the agent will choose to detour affected by conflicts between the potential energy and the inertia potential energy of static obstacles.

In Figure 8(a), the agent falls into the local minimum cell (10, 13) at the 75th time step. After delaying 4 time steps, the agent will jump out at the 83rd time step and delay passing through the cell (11, 13), where the obstacle is located, and reach the destination at the 158th time step.

In Figure 8(b), the agent falls into the local minimum cell (10, 13) at the 75th time step. After waiting for 5 time steps, the agent jumps out at the 80th time step and choose to detour, reaching the destination at the 214th time step.

4. Conclusions

This paper first proposes a dynamic potential energy model (DPEM) under the spatial-temporal simulation environment with large cells and small time segments, where the deadlock problem arisen in traditional cellular automata simulations can be avoided. Secondly, this paper concludes

that the agent's selection such as forward, waiting, and detour while facing the obstacles is a dynamic decision-making process. Each agent will decide on opportunity cost with the objective of utility maximization and is affected by their personal characters and other environmental factors. Thirdly, through the empirical experiments, it is showed that the sum of the agent and obstacles' space occupation ($\rho + \mu$) is the main factor affecting the velocity decline rate in the multiagent system. The velocity decline rate is also related to the obstacle types, such as the stationary objective (SO), male group (MG), and female group (FG). Moreover, it is showed that the trajectory of each agent in the multiagent system has two characteristics: dissipation and enforcement. The dissipation characteristic can guarantee the condition in which each agent will not return, and the enforcement characteristic can guarantee the condition in which each agent will first wait and then make detours while facing the obstacles. Then, the relationship between each agent's patience while waiting and the space occupation change rate of obstacles is showed to be logarithmic. Finally, we conduct a case study to verify the effectiveness of the proposed DPEM. The research results of this paper will be helpful for the construction of a mesoscopic pedestrian traffic simulation model accurately and efficiently and reflect different characters of each agent, such as the gender and carry-on baggage. In the future research, we will introduce more parameters and traffic control measures, such as travel companion, traffic signals, pedestrians' selection between stairs and escalators to describe the environmental factors, and group characters of the multiagent based on the technology of data mining and decision-making game theory.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The Fundamental Research Funds for the Central Universities (30919011290).

References

[1] E. Porter, S. H. Hamdar, and W. Daamen, "Pedestrian dynamics at transit stations: an integrated pedestrian flow modeling approach," *Transportmetrica A: Transport Science*, vol. 14, no. 5-6, pp. 468–483, 2018.

[2] T. Saghapour, S. Moridpour, and R. Thompson, "Sustainable transport in neighbourhoods: effect of accessibility on walking and bicycling," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 849–871, 2019.

[3] I. M. Lee and D. M. Buchner, "The importance of walking to public health," *Medicine & Science in Sports & Exercise*, vol. 40, no. 7, pp. S512–S518, 2008.

[4] R. Gerike, A. de Nazelle, R. Wittwer, and J. Parkin, "Special issue "walking and cycling for better transport, health and the environment," *Transportation research Part A: Policy and practice*, vol. 123, 2019.

[5] J. Eady and D. Burtt, *Walking and Transport in Melbourne Suburbs*, Victoria Walks Incorporated, Victoria, Australia, 2019.

[6] B. Zhang, C. Shao, Y. Li, H. Tan, and D. Jiang, "Dynamic simulation analysis of humanoid robot walking system based on ADAMS," *Journal of Shanghai Jiaotong University (Science)*, vol. 24, no. 1, pp. 58–63, 2019.

[7] T. Yamazaki, N. Kobayashi, Y. Owada, and G. Sato, "Agent walking model construction in urban disaster simulation," in *Proceedings of the 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*, IEEE, Nara, Japan, pp. 355-356, October 2018.

[8] C. Dias and R. Lovreglio, "Calibrating cellular automaton models for pedestrians walking through corners," *Physics Letters A*, vol. 382, no. 19, pp. 1255–1261, 2018.

[9] O. Khatib, *Real-Time Obstacle Avoidance For Manipulators And Mobile Robots*, pp. 396–404, Springer, New York, NY, USA, 1986.

[10] D. N. Subramani and P. F. J. Lermusiaux, "Risk-optimal path planning in stochastic dynamic environments," *Computer Methods in Applied Mechanics and Engineering*, vol. 353, pp. 391–415, 2019.

[11] A. Sedeño-noda and M. Colebrook, "A biobjective Dijkstra algorithm," *European Journal of Operational Research*, vol. 276, no. 1, pp. 106–118, 2019.

[12] R. Song, Y. Liu, and R. Bucknall, "Smoothed $A*$ algorithm for practical unmanned surface vehicle path planning," *Applied Ocean Research*, vol. 83, pp. 9–20, 2019.

[13] P. Vadakkepat, K. C. Tan, and W. Ming-Liang, "Evolutionary artificial potential fields and their application in real time robot path planning," in *Proceedings of the 2000 congress on evolutionary computation*, vol. 1, IEEE, La Jolla, CA, USA, pp. 256–263, 2000, July.

[14] U. Orozco-Rosas, O. Montiel, and R. Sepúlveda, "Mobile robot path planning using membrane evolutionary artificial potential field," *Applied Soft Computing*, vol. 77, pp. 236–251, 2019.

[15] Z. Zhishui, "Ant colony algorithm based on path planning for mobile agent migration," *Procedia Engineering*, vol. 23, pp. 1–8, 2011.

[16] H. Miao and Y.-C. Tian, "Dynamic robot path planning using an enhanced simulated annealing approach," *Applied Mathematics and Computation*, vol. 222, pp. 420–437, 2013.

[17] B. K. Patle, A. Pandey, A. Jagadeesh, and D. R. Parhi, "Path planning in uncertain environment by using firefly algorithm," *Defence Technology*, vol. 14, no. 6, pp. 691–701, 2018.

[18] E. S. Low, P. Ong, and K. C. Cheah, "Solving the optimal path planning of a mobile robot using improved Q-learning," *Robotics and Autonomous Systems*, vol. 115, pp. 143–161, 2019.

[19] M. Wooldridge, *An Introduction to Multiagent Systems*, John Wiley & Sons, Hoboken, NY, USA, 2009.

[20] D. Kinny and M. Georgeff, "Modelling and design of multiagent systems," in *Proceedings of the International Workshop on Agent Theories, Architectures, and Languages*, Springer, Berlin, Germany, pp. 1–20, 1996 August.

[21] M. Wooldridgey and P. Ciancarini, "Agent-oriented software engineering: the state of the art," in *Proceedings of the International Workshop on Agent-Oriented Software Engineering*, Springer, Berlin, Heidelberg, pp. 1–28, 2000 June.

[22] L. Okdinawati, T. M. Simatupang, and Y. Sunitiyoso, "Multi-agent reinforcement learning for collaborative transportation management (ctm)," in *Proceedings of the Agent-Based Approaches in Economics and Social Complex Systems IX*, Springer, Singapore, pp. 123–136, 2017.

[23] A. Baykasoglu, V. Kaplanoglu, and C. Sahin, "Route prioritisation in a multi-agent transportation environment via multi-attribute decision making," *International Journal of Data Analysis Techniques and Strategies*, vol. 8, no. 1, pp. 47–64, 2016.

[24] C. Liao, H. Guo, K. Zhu, and J. Shang, "Enhancing emergency pedestrian safety through flow rate design: bayesian-nash equilibrium in multi-agent system," *Computers & Industrial Engineering*, vol. 137, p. 106058, 2019.

[25] C. Yu, G. Ren, and T. Zhang, "Simulation of pedestrian flow at a crosswalk by a multi-agent system with a pre-avoid forces model," in *Proceedings of the CICTP 2016*, pp. 2303–2313, Shanghai, China, July 2016.

[26] J. Ren, W. Xiang, Y. Xiao, R. Yang, D. Manocha, and X. Jin, "Heter-Sim: heterogeneous multi-agent systems simulation by interactive data-driven optimization," *IEEE Transactions on Visualization and Computer Graphics*, 2019.

[27] P. Stone and M. Veloso, "Multiagent systems: a survey from a machine learning perspective," *Autonomous Robots*, vol. 8, no. 3, pp. 345–383, 2000.

[28] M. Hussein and T. Sayed, "Validation of an agent-based microscopic pedestrian simulation model in a crowded pedestrian walking environment," *Transportation Planning and Technology*, vol. 42, no. 1, pp. 1–22, 2019.

[29] S. Tak, S. Kim, and H. Yeo, "Agent-based pedestrian cell transmission model for evacuation," *Transportmetrica A: Transport Science*, vol. 14, no. 5-6, pp. 484–502, 2018.

[30] K. R. Rozo, J. Arellana, A. Santander-Mercado, and M. Jubiz-Diaz, "Modelling building emergency evacuation plans considering the dynamic behaviour of pedestrians using agent-based simulation," *Safety Science*, vol. 113, pp. 276–284, 2019.

[31] C. Y. Cheung and W. H. K. Lam, "Pedestrian route choices between escalator and stairway in MTR stations," *Journal of Transportation Engineering*, vol. 124, no. 3, pp. 277–285, 1998.

[32] O. Handel and A. Borrmann, "Service bottlenecks in pedestrian dynamics," *Transportmetrica A: Transport Science*, vol. 14, no. 5-6, pp. 392–405, 2018.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 7096767, 13 pages
https://doi.org/10.1155/2020/7096767

Research Article

Location Design of Electrification Road in Transportation Networks for On-Way Charging

Yue Qiu,[1] Yuchuan Du,[1] Shanchuan Yu,[2] and Shengchuan Jiang[1]

[1]*Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China*
[2]*China Merchants Chongqing Communications Research & Design Institute Co. Ltd., Chongqing 400067, China*

Correspondence should be addressed to Yuchuan Du; ycdu@tongji.edu.cn

Received 18 January 2020; Revised 14 March 2020; Accepted 26 June 2020; Published 23 July 2020

Academic Editor: Wei (David) Fan

Electric vehicles tend to be a great mobility option for the potential benefits in energy consumption and emission reduction. On-way charging (OWC) has been recognized to be a promising solution to extend driving range for electric vehicles. Location of the electrification road (ER) is a critical issue for future urban traffic management to accommodate the new mobility option. This paper proposes a mathematical program with equilibrium constraint (MPEC) approach to solve this problem, which minimizes the total travel time with a limited construction budget. To describe the drivers' routing choice, a path-constrained network equilibrium model is proposed to minimize their travel time and prevent running out of charge. We develop a modified active set algorithm to solve the MPEC model. Numerical experiments are presented to demonstrate the performance of the model and the solution algorithm and analyze the impact of charging efficiency, battery size, and comfortable range.

1. Introduction

Electrification systems based on renewable energy power sources are introduced in the urban transportation system for positive environmental effect and carbon reduction [1, 2]. Recently, with the increasing concern about sustainable transportation [3], electric vehicles (EVs) are widely adopted in urban travel. Electrification road (ER) is considered as a promising domain for sustainable electrical energy harvesting to support on-way charging (OWC) for EVs, thanks to the advancement of application of photovoltaic, piezoelectric, and pyroelectric materials converting solar energy, physical pressure, and thermal energy into electrical energy, e.g., [4–8]. OWC provides a new mode of charging for EV drivers to extend driving range, who are suffering from range anxiety of running out of power on the way [9, 10] and long charging time ranging from 0.5 to 2 hours for a full charge [11]. OWC has raised the interests of organizations and is being tested all over the world. ER deployed on a network has been tested in South Korea [12]. UK has conducted a series of OWC recharging tested on highways [13]. In Sweden, the Volvo Group and the Swedish Transport Association planned to build 500-meter ER for wireless power transfer [14]. Such tests and experiments show that OWC would come true in future. Owing to the high investment in the ER, which reaches $4 million per lane mile [15], it is imperative to find the optimal location.

Most previous studies related to the location design of charging facilities are based on user equilibrium (UE) problems with EVs which investigate how the limited driving range of EVs and location of charging facilities affect the routing choice and subsequently the flow equilibrium distribution on road networks. Among these studies, Jiang et al. [16] first proposed a path-constrained assignment model where lengths of selected routes are within the driving ranges of EVs. The model is further extended to consider mixed gasoline and electric vehicle flows and their combined routing choices under different problem settings [17, 18]. He et al. investigated the optimal prices of electricity and the integrated prices of electricity and roads based on the multiclass spatial distribution of electric vehicles across the transportation network [19, 20]. All these studies assumed the energy consumption of EVs is flow-independent.

The aforementioned studies do not consider the location of charging facilities in networks. Problems related to the optimal location of charging stations or battery swapping stations have received much attention. He et al. [21] first proposed a UE-based framework for locating charging stations to maximize social welfare, considering route choices of electric vehicles, and further developed three models of different flow dependencies and energy consumption [22, 23]. Xu et al. [24] defined a network with battery swapping stations where swapping time is independent with charging capacity, considering the impact of queuing to find the best charging strategy. These existing relevant studies are intended for charging stations or battery swapping stations, where EVs reach the station, stop, and get replenished. Such models cannot address issues for OWC directly.

OWC, unlike charging stations or battery swapping stations, can charge EVs during the ride. OWC technology has been implemented in a network in South Korea [12]. A limited number of research studies have been done on the OWC problem to select the optimal ER locations so far. Riemann et al. [25] proposed a mixed-integer nonlinear program to describe the flow-capturing location model considering driver's routing behavior. Chen et al. [26] investigated the optimal deployment of charging lanes based on the speed control of EV drivers on the charging lanes. Chen et al. [27] proposed the integrated deployment of charging lanes and charging stations along a long traffic corridor, which cannot be applied in the transportation network directly. Liu and Song [28] adopted a robust optimization methodology to provide robust location optimization of wireless charging facilities for the electric bus system to reduce total cost of batteries and charging facilities with different uncertainty levels. Liu et al. [29] developed a mixed-integer linear program for OWC location and battery size optimization and battery size for an electric bus system with overlapping bus lines. Ahmed et al. [30] introduced a method to find the best combination of battery capacity and wireless charger characteristics to solve the tradeoff between maximizing charge sustaining, minimum battery capacity, and minimizing the initial investment. Bi et al. [31] adopted a genetic algorithm to optimize the rollout of OWC infrastructure both spatially and temporally in order to minimize life cycle costs and energy burdens. Zhao et al. [32] proposed a biobjective optimization problem for integrated EV location and on-board battery size design of electric bus systems to minimize deployment cost and reduce energy consumption of electrified systems.

The above research studies can solve the ER location optimization to some extent, but they tend to assume that charging amount is only related to the charging time. Under such assumption, drivers tend to deliberately slow down to extend the charging time, and even EV may revisit the same charging lanes. All the vehicles behind the EV that is moving slowly to get charged are delayed. A road contains more than two lanes, one of which is the ER. Vehicles without charging intention would perform lane-changing, which reduces the road capacity. Relationship between the cost of building the ER and width is seldom considered.

In this paper, charging amount is assumed to be related with the driving distance, while charging and the candidates of the ER have to satisfy the minimum energy supply. The width of the ER is considered to affect ER building cost. The link performance function is set as the BPR function in numerical examples. Therefore, the route choice of drivers will be complicated, and the UE flows of EVs in the transportation network will be different.

This paper investigates the location design problem of the ER based on the analysis of the routing behavior of EV drivers, which is different from the previous studies with the charging station or swapping station. Our model inherits from the paper written by He et al. [21], but charging behaviors and routing behaviors are different tracking back to the specific situation of OWC in contrast to charging at station, which leads to a totally different location design. Assuming that the energy consumption rate and recharging rate are flow-independent, this paper proposes a path-constrained network equilibrium model (PCNE) considering the routing and recharging behavior of EV drivers as well as the constraint of driving range. Then, we develop an iterative solution framework to solve the PCNE problem. Lastly, we investigate a mathematical program with equilibrium constraint (MPEC) model to optimize the ER location to minimize the system total travel time under a given construction budget limit. With the proposed model, discussions about the effect of different recharging rates, comfortable ranges, and battery sizes on the location design and system performance are conducted.

The remainder of the paper is organized as follows. The notation, assumption, formulation of the PCNE problem, and a solution framework are elaborated in Section 2. Section 3 formulates the location design problem as a MPEC program. Section 3.1 proposes a modified active set method to solve the problem. Section 4 presents numerical cases based on the Sioux Falls network, and Section 5 offers the conclusions of this study and implications for future research on the transportation network of the ER.

2. Path-Constrained Network Equilibrium Model for On-Way Charging

The section proposes a PCNE model to describe the routing and charging behaviors of EV drivers under a given location design, which determine the network flow pattern. Decision variables of the model include equilibrium flow and charging amount of the EV.

2.1. Notation and Assumption. Following assumptions are made before the model formulation:

Assumption 1. Drivers do not need to revisit the same ER to guarantee that the EV will not be out of charge.

Assumption 2. If the EV is full of charge, charging process stops while running on the ER.

Assumption 3. Given an ER location design, a physical path from the origin to the destination may not be feasible for EV drivers due to the limit for the EV battery.

Assumption 4. All vehicles in networks are EVs. This assumption can be relaxed as the model can be extended to accommodate both EVs and regular vehicles.

Assumption 5. Recharging rate is not affected by speed of the EV.

When EV drivers travel from their origins to destinations, they are assumed to select routes to minimize their travelling cost. Travelling cost includes electricity cost and travel time cost. Electricity cost is much smaller than travel time cost [23]. Therefore, we simply assumed that EV drivers select routes of the least travelling time. While running on the ER, EV drivers can decide whether to charge and charging amount. Charging amount is not greater than the maximum charging amount that equals charging rate multiplied by the distance of the ER.

Consider a transportation network containing ER running EVs, denoted by $G = (N, A)$, where N and A are sets of nodes and link, respectively. The link $(i, j) \in A$ is directional and emanates from the starting node $i \in N$ to the ending node $j \in N$ with the length D_{ij} and the road grade G_{ij}. Travel demands are between a set of O-D pairs W. q^{od} represents the travel demand between the O-D pairs, $od \in W$, and R^{od} represents the set of paths between the O-D pairs.

2.2. Formulation. Following definitions are introduced here to describe the PCNE problem.

Definition 1. A path is usable if an EV is able to complete the path with the remaining charge state never falling below the comfortable range.

EV drivers are assumed to select routes of the least travel time among all the usable paths. At least one usable path exists for each O-D pair here for model generality. As the energy consumption rate and recharging rate are assumed to be flow-independent, the usability of a path is independent of the drivers' routing choice, which can be predetermined before flow distribution given the location design of the ER. The following path-constrained network equilibrium proposed by He et al. [23] is adopted here.

Definition 2. At the path-constrained network equilibrium, all utilized paths are usable, and travel time of all utilized paths of a given O-D pair is the same, which is not more than that of any unutilized usable paths of the same O-D pair.

The formulation of the path-constrained network equilibrium is as follows:

PCNE:

$$\underset{f}{\text{minimize}} \quad \sum_{(i,j)\in A} \int_0^{\sum_{od\in W}\sum_{r\in \widehat{R}^{od}} f_r^{od}\chi_{ij,r}^{od}} t_{ij}(z)\mathrm{d}z, \tag{1}$$

subject to

$$\sum_{r\in R^{od}} f_r^{od} = q^{od}, \quad \forall od \in W, \tag{2}$$

$$f_r^{od} \geq 0, \quad \forall od \in W,\ r \in \widehat{R}^{od}, \tag{3}$$

where $\widehat{R}^{od}$ represents the set of all usable paths between a given O-D pair, $od \in W$. Energy consumption rate and charging rate depend on the travel time, which are flow-independent. Therefore, the usability of any path is flow-independent, which can be calculated. $f = (\ldots, f_r^{od}, \ldots)$ is a vector of path flow. f_r^{od} represents the traffic flow on path $r \in R^{od}$ of O-D pair, $od \in W$. Constraint (2) demonstrates the flow balance for each O-D pair. Constraint (3) ensures nonnegativity of flow for each usable path.

Here, $\chi_{ij,r}^{od}$ is a binary variable to represent whether path r traverses link $(i, j) \in A$, which equals 1 if path r traverses link $(i, j) \in A$, and 0 otherwise. v_{ij} represents the traffic flow on the link $(i, j) \in A$. The travel time of link $t_{ij}(v_{ij})$, $(i, j) \in A$, is a strictly increasing function of the flow on link $(i, j) \in A$. Here, assume that link travel time takes the following form of Bureau of Public Roads (BPR):

$$t_{ij} = t_{ij}^0\left[1 + 0.15\left(\frac{v_{ij}}{c_{ij}}\right)^4\right], \quad \forall (i, j) \in A, \tag{4}$$

where t_{ij}^0 represents the free-flow travel time of link (i, j) and c_{ij} represents the capacity of link (i, j).

2.3. Solution Procedure. If we can enumerate all usable paths beforehand, the proposed PCNE problem is a regular nonlinear optimization which can be solved easily by commercial nonlinear solvers such as CONOPT. However, enumeration of usable paths is a time-consuming work. Here, we adopt the solution procedure proposed by He et al. [22] to obtain the solution. The procedure starts with a subset of $\widehat{R}^{od}$, $od \in W$, and solves a restricted version of PCNE defined upon the subset. Then, the feasibility of the solution to the restricted PCNE is tested. If not, a new usable path will be generated and added to the subset. Iteration proceeds until termination.

Some new variables are introduced in the subproblem. Here, $l_{\max}$ and l_0 are the battery size and the initial state of charge. For an EV travelling between the O-D pair, $od \in W$, the state of charge at node i is l_i^{od}. δ_{ij} is a binary parameter representing whether link $(i, j) \in A$ is equipped with the ER to support OWC. The variable equals 1 if the street (i, j) is equipped with the ER, and 0 otherwise. $A_p \subseteq A$ denotes the set of all links equipped with the ER. Let ω_{ij} denote the energy consumption rate on link $(i, j) \in A$ and ϖ denote the recharging rate of the ER. Therefore, link energy consumption is denoted as $c_{ij} = (\omega_{ij} - \delta_{ij}\varpi)D_{ij}$. m^{od} is the minimum charge state within the comfortable range for drivers of the O-D pair, $od \in W$. $N_t(n)$ is the set of tails of those links heading to node n, and $N_h(n)$ is the set of heads of those links emanating from node n. K and M are sufficient large constants. x_{ij}^{od} is a binary variable indicating utilization incidence, which equals 1 if link (i, j) is utilized for travel demands between the O-D pair, $od \in W$, and 0 otherwise.

Accordingly, ρ_{ij}^{od} is a variable that equals 0 if link (i, j) is utilized for travel demands between the O-D pair, $od \in W$, and is unrestricted otherwise. ξ_{nd} is a binary variable, which equals 1 when $n = d$, and 0 otherwise; ξ_{on} is a binary variable, which equals 1 when $n = o$, and 0 otherwise; and e_{ij}^{od} represents the recharging amount at link $(i, j) \in A$.

For an O-D pair, $od \in W$, the link flow obtained $(\ldots, v_{ij}^{[k]}, \ldots)$ at the kth iteration of PCNE is adopted at the shortest usable path solution framework (denoted as SUPF-k). The formulation of SUPF-k is shown as follows:

SUPF-k:

$$\left(x^{[k]}, e^{[k]}\right) \in \underset{x,e\in\Omega}{\arg\min} \sum_{(i,j)\in A} t_{ij}\left(v_{ij}^{[k]}\right)x_{ij}^{od}, \tag{5}$$

where Ω consists of the following conditions:

$$\sum_{i\in N_t(n)} x_{in}^{od} - \sum_{j\in N_h(n)} x_{nj}^{od} - \delta_{nd} + \delta_{on} = 0, \quad \forall n \in N, \tag{6}$$

$$l_j^{od} - l_i^{od} + \omega_{ij}D_{ij} - e_{ij}^{od} = \rho_{ij}^{od}, \quad \forall (i, j) \in A, \tag{7}$$

$$0 \le e_{ij}^{od} \le D_{ij}\delta_{ij}\varpi, \quad \forall (i, j) \in A, \tag{8}$$

$$0 \le l_n^{od} \le l_{\max}, \quad \forall n \in N. \tag{9}$$

$$-K\left(1 - x_{ij}^{od}\right) \le \rho_{ij}^{od} \le K\left(1 - x_{ij}^{od}\right), \quad \forall (i, j) \in A, \tag{10}$$

$$l_i^{od} - \left(\omega_{ij}D_{ij} - e_{ij}^{od}\right) \ge -M\left(1 - x_{ij}^{od}\right) + m^{od}, \quad \forall (i, j) \in A, \tag{11}$$

$$l_o^{od} = l_0, \tag{12}$$

$$\omega_{ij} = \varphi\left(G_{ij}\right), \quad \forall (i, j) \in A, \tag{13}$$

$$x_{ij}^{od} \in \{0, 1\}, \quad \forall (i, j) \in A. \tag{14}$$

In the above, the objective function is to minimize the travel time. Constraint (6) guarantees the balance of traffic flow; constraint (7) specifies the relation between the states of the charge of EV batteries travelling from node i to node j, link $(i, j) \in A$. Constraint (8) ensures the recharging amount of electricity on the ER at link $(i, j) \in A_P$ does not exceed the maximum charging quantity that equals charging rate multiplied by the distance of the link. Constraint (9) sets the upper and lower bounds of the states of the charge of the EV. Constraint (10) suggests that the EV driver can only recharge when travelling on the ER. Constraint (11) sets the lower bounds of the states of charge of the EV during the trip to satisfy comfortable range for drivers. Constraint (12) specifies the initial state of charge. Constraint (13) is a given electricity consumption function with respect to the road grade. Constraint (14) requires x_{ij}^{od} to be binary.

The formulation assumes that EV drivers can control the recharging process on the ER. They can determine the quantity of recharging on the ER. SUPF-k is a mixed-integer linear program, which can be easily solved by commercial solvers like CPLEX 12.8 for small- or medium-sized problems.

The solution to SUPF-k for each O-D pair, denoted as $(\ldots, x_{ij}^{od[k]}, \ldots\ldots, e_{ij}^{[k]}, \ldots)$, is adopted in the set construction of the shortest usable path. The iterative steps of PCNE is shown as follows:

Step 0: set the iteration variable $k = 0$ and traffic flow $(\ldots, v_{ij}^{[k]}, \ldots) = (\ldots, 0, \ldots)$ for each O-D pair, $od \in W$. Solve SUPF-k to obtain the optimal solution $(\ldots, x_{ij}^{od[k]}, \ldots\ldots, e_{ij}^{[k]}, \ldots)$. Construct $\widehat{R}^{od} = \widehat{R}_S^{od[k]}$.

Step 1: solve the restricted NE upon $\widehat{R}^{od}$. Denote $(\ldots, v_{ij}^{[k+1]}, \ldots)$ and $(\ldots, \mu^{od}, \ldots)$ as the optimal solutions and multipliers associated with constraint (2).

Step 2: for each O-D pair, $od \in W$, solve SP- $(k + 1)$ and obtain the optimal solution, $(\ldots, x_{ij}^{od[k+1]}, \ldots\ldots, e_{ij}^{[k+1]}, \ldots)$. For $\widetilde{od} \in W$, if $\mu^{\widetilde{od}} > \sum_{(i,j)\in A} t_{ij}(v_{ij}^{[k+1]}) x_{ij}^{\widetilde{od}[k+1]}$, $\widehat{R}^{od} = \widehat{R}^{od} \cup \widehat{R}_S^{\widetilde{od}[k+1]}$, and go to Step 1, $k = k + 1$. If $\mu^{od} \le \sum_{(i,j)\in A} t_{ij}(v_{ij}^{[k+1]}) x_{ij}^{od[k+1]}$ for all O-D pairs, terminate and $(\ldots, v_{ij}^{[k+1]}, \ldots)$ is the equilibrium link flow distribution.

The above procedure terminates in a finite number of steps.

3. Location Design of the ER

This section investigates the location design under a given budget limit to maximize social welfare (i.e., minimize the total travel time of all EVs in our paper). The following assumptions are introduced here.

Assumption 5. Construction cost of the ER is related to length, width of the street.

Assumption 6. When a street is designed to be equipped with the ER, vehicles driving on any lane of the street can get charged. That is, streets in the network are not mixed with charging lanes and regular lanes.

When the ER location is determined, the equilibrium traffic flow can be derived, as well as the recharging and routing behavior. The problem of the location design for electrification road (LDER) is formulated as follows:

LDER:

$$\underset{\delta}{\text{minimize}} \quad \sum_{(i,j)\in A} t_{ij}\left(v_{ij}^*\right)v_{ij}^*, \tag{15}$$

subject to

$$D_{ij}\delta_{ij}\varpi \ge -G\left(1 - \delta_{ij}\right) + l_{\max}, \quad (i, j) \in A, \tag{16}$$

$$\sum_{(i,j)\in A} W_{ij}D_{ij}\delta_{ij}B \le \Theta, \tag{17}$$

$$\delta_{ij} \in \{0,1\}, \quad (i,j) \in A, \tag{18}$$

$$(f^*, v^*) \in \underset{f \in \Psi(\delta)}{\arg\min} \sum_{(i,j)\in A} \int_0^{\sum_{od\in W}\sum_{r\in \widehat{R}^{od}(\delta)} f_r^{od}\lambda_{ij,r}^{od}} t_{ij}(z)\mathrm{d}z, \tag{19}$$

where $\Psi(\delta) = \{f \mid \sum_{r\in R^{od}} f_r^{od} = q^{od}, \forall od \in W; f_r^{od} \geq 0, \ \forall od \in W, r \in \widehat{R}^{od}(\delta)\}$; W_{ij} is the width of link $(i,j) \in A$; B is the construction cost of the unit-area ER, which is the function of the recharging rate ϖ; Θ is the budget limit; and G is a sufficiently large constant.

In the above, the objective function represents the total driving time. Constraint (16) presents the shortest length of the street chosen as the ER, guaranteeing that Assumption 1 always holds. Constraint (17) specifies the budget limit for the ER deployment. Constraint (18) requires δ_{ij} to be binary. Constraint (19) states the traffic flow following the result predicted by the PCNE model.

3.1. Active Set Method. LDER is a typical bilevel discrete network design problem with NE at lower level. Many existing solutions from the literature can be explored, for example, branch and bound [33], simulated annealing [34], and SO relaxation [35] with consensus of reformulating an equivalent single-level model with equivalent optimality conditions of the lower-level problem like KKT conditions to find the solution of the bilevel program. Problems with the nonconvex feasible region generally contain many local optimal solutions and typically require a time-consuming branch-and-bound scheme to search for a globally optimal solution. Such algorithms are not readily applicable considering the nonlinear and complementary constraints, adding enormous amount of calculation.

To solve the LDEP, we utilize the active set method [36], which terminates after a finite number of iterations ending with a strongly stationary solution. The active set method has been proved to have the potential for solving larger network design problems. The active set method assigns nonnegative variables in each pair as zero to construct the location design problem as a regular nonlinear program. Adopting multipliers associated with the constraints forcing nonnegative variables to be zero, the proposed algorithm formulates the binary knapsack problem to solve the zero-value assignments to decrease the total travel time.

The active set algorithm starts by setting an initial feasible solution $U = \left\{\delta_{ij}^0 \mid (i,j) \in A\right\}$ and considering two sets:

$$\Omega_0 = \left\{(i,j), \delta_{ij} = 0\right\}, \tag{20}$$

$$\Omega_1 = \left\{(i,j), \delta_{ij} = 1\right\}, \tag{21}$$

where $\Omega_0 \cup \Omega_1 = A$ and $\Omega_0 \cap \Omega_1 = \phi$. Under the location design for the ER corresponding to the initial feasible solution U^0, the NE problem can be solved easily by the path-based iteration procedure proposed in Section 2. Let $V^* = \left\{v_{ij}^*, (i,j) \in A\right\}$ be the solution set of NE, which is unique because of the monotonically increasing BPR function.

The basic principle of the active set algorithm is to exchange elements between Ω_0 and Ω_1 in order to reduce the overall travel time in each iteration until the system optimal solution is found. In each iteration, the active sets are adjusted to meet the following theorems proved by Zhang et al. [36].

Theorem 1. *Given that Ω_0 and Ω_1 are feasible solution sets, if $\lambda_{ij} < 0$ for some $(i,j) \in \Omega_0$, switching (i,j) from Ω_0 to Ω_1 yields less system travel time. If $\mu_{ij} > 0$ for some $(i,j) \in \Omega_1$, switching (i,j) from Ω_1 to Ω_0 yields less system travel time.*

Theorem 2. *The active set method converges after a finite number of iterations.*

λ_{ij} and μ_{ij} are the Lagrangian multipliers associated with equations (20) and (21), respectively.

To ensure that budget limit and shortest-length limit are satisfied in each iteration according to constraints (16) and (17), an embedded program is proposed as follows:

$$\text{minimize} \quad \sum_{(i,j)\in\Omega_0} \lambda_{ij} g_{ij} - \sum_{(i,j)\in\Omega_1} \mu_{ij} h_{ij}, \tag{22}$$

subject to

$$\sum_{(i,j)\in\Omega_0} D_{ij} g_{ij} \varpi + \sum_{(i,j)\in\Omega_1} D_{ij}\left(1 - h_{ij}\right)\varpi \geq \sum_{(i,j)\in\Omega_0} \left(-G\left(1 - g_{ij}\right)\right) + \sum_{(i,j)\in\Omega_0} \left(-Gh_{ij} + l_{\max}\right), \quad \forall (i,j) \in A, \tag{23}$$

$$\sum_{(i,j)\in\Omega_0} W_{ij} D_{ij} g_{ij} B + \sum_{(i,j)\in\Omega_1} W_{ij} D_{ij}\left(1 - h_{ij}\right)B, \quad \forall (i,j) \in A, \tag{24}$$

$$\sum_{(i,j)\in\Omega_0} \lambda_{ij} g_{ij} - \sum_{(i,j)\in\Omega_1} \mu_{ij} h_{ij} \geq \theta, \quad (i,j) \in A, \tag{25}$$

$$g_{ij}, h_{ij} \in \{0,1\}, \quad \forall (i,j) \in A, \tag{26}$$

where $g_{ij} = 1$ records a shift in $(i,j) \in \Omega_0$ to Ω_1 and $h_{ij} = 1$ records a shift in $(i,j) \in \Omega_1$ to Ω_0. The objective of programming (22) is to minimize the estimated decrease, the negative value of which implies a potential reduction of the objective value. The process terminates when equation (22) is zero.

Constraint (23) is consistent with constraint (16). Constraint (24) is consistent with constraint (17). Parameter θ is set to guarantee a decrease in the objective function. In the first iteration, $\theta = -\infty$. In the next iteration, θ can be calculated by the following equation:

$$\theta = \varepsilon + \sum_{(i,j)\in\Omega_0} \lambda_{ij} g_{ij} - \sum_{(i,j)\in\Omega_1} \mu_{ij} h_{ij}, \tag{27}$$

where ε is a sufficiently small constant. Equation (27) prevents the solution from degenerating back to the previous iteration by increasing θ by ε.

Set the iteration variable $\eta = 1$, $\mathrm{SO}^0 = +\infty$, $\mathrm{TD}^0 = -\infty$, and the initial feasible fixed solutions $(\Omega_0^\eta, \Omega_1^\eta)$

Solve (1)–(14) with $(\Omega_0^\eta, \Omega_1^\eta)$ and derive λ_{ij}^η and μ_{ij}^η. Calculate system total travel time SO^η. If $\mathrm{SO}^\eta \geq \mathrm{SO}^\eta$, then $\eta = \eta - 1$; go to step 3.

Solve (22)–(26) and $\theta = \varepsilon + \mathrm{TD}^{\eta-1}$

If the optimal objective value is zero, $(\Omega_0^\eta, \Omega_1^\eta)$ is the best solution and the iteration ends;

If not, derive $(g_{ij}, h_{ij}))$ and the objective value TD^η and go to step 4.

Obtain a new solution to $(\Omega_0^{\eta+1}, \Omega_1^{\eta+1})$:

$$\Omega_0^{\eta+1} = \Omega_0^\eta - \left\{(i,j) \in \Omega_0^\eta\colon g_{ij} = 1\right\} + \left\{(i,j) \in \Omega_1^\eta\colon h_{ij} = 1\right\}$$

$$\Omega_1^{\eta+1} = \Omega_1^\eta - \left\{(i,j) \in \Omega_1^\eta\colon h_{ij} = 1\right\} + \left\{(i,j) \in \Omega_0^\eta\colon g_{ij} = 1\right\}$$

$\eta = \eta - 1$; go to step 2.

Algorithm 1

Steps of the algorithm are shown in the following. Here, SO denotes the objective value of programming (15), and TD denotes the objective value of programming (22). (Algorithm 1).

4. Numerical Examples

4.1. Basic Settings. In this section, numerical examples on the Sioux Falls network in South Dakota are presented to demonstrate the performance of the proposed model as shown in Figure 1, which consists of 24 nodes, 76 links, and 14 OD pairs. All links are available to be equipped with the ER. Table 1 reports the link characteristics, including free-flow travel time (FFTT), width, length, and capacity. The OD demand is given in Table 2.

We assume that the battery size $l_{\max}$ is 25 kWh. The initial state of the charge l_0 is set as 25% $l_{\max}$. For simplicity, it is assumed that energy consumption rate ω at all links in the network is set as 0.3 (kWh/mile), and the recharging rate ϖ is set as 2.5 (kWh/mile).The comfortable range for all drivers m^{od} is set as zero. Assume that the cost of building unit-area ER is $((\$2 \times 10^5)/(\text{mile} \times \text{meter}))$ at the recharging rate $\varpi = 2.5$ (kWh/mile). The budget limit Θ is set as $\$200,000,000$. All the above values are chosen for illustrative purpose. We adopt CPLEX12.8 to solve the PCNE, SP, and LDER.

4.2. Base Case Result. The LDER problem is solved under the aforementioned basic setting, as shown in Figure 2(a). Figure 2(b) depicts the convergence performance of the SP-k algorithm, indicating the potential of the SP-k algorithm in solving the PCNE problem. The equilibrium link flows are reported in Table 3, with the total travel time of 9.20×10^4 h · veh and the average travel time 0.575 h per vehicle.

4.3. Strategy Comparison under Different Budget Limits. In this section, we consider different location designs of the proposed model under different budget limits to demonstrate the system performance, as shown in Figure 3(a). Figure 3 compares the total travel time, cost of the ER, and the number of ERs under different budget limits. As expected, total travel time of the network decreases with an increase in the budget. More ERs are built with an increasing budget when the budget

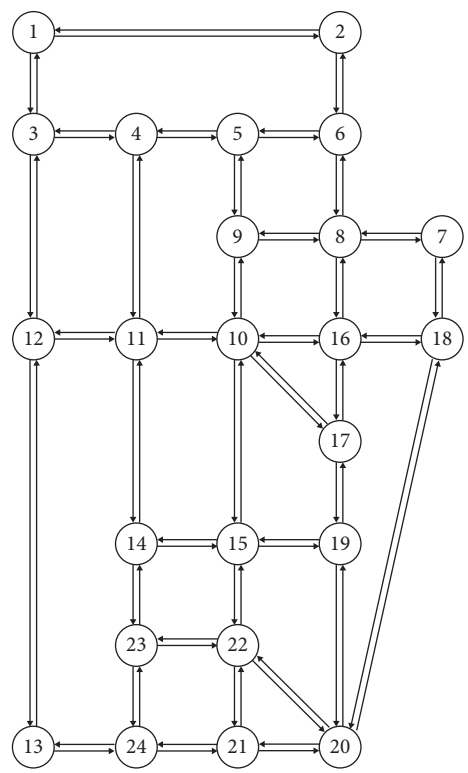


Figure 1: Sioux Falls network.

ranges from $\$1.3 \times 10^8$ to $\$2 \times 10^8$, generating more usable route choice for the EV to complete the tour, which promotes average distribution of traffic flow leading to reduction in the system travel time. Nevertheless, more ERs are built under the budget of $\$1.2 \times 10^8$, comparing with the budget of $\$1.3 \times 10^8$ and $\$1.4 \times 10^8$. It can be observed from Figure 3(a) that when the budget exceeds $\$1.4 \times 10^8$, link (20,18) is always chosen to be equipped with ER. The usability of link (20,18) has a significant impact on EV drivers' routing behavior due to its special location. However, ER on link (20,18) of the 15 m width is an expensive project, which has to be weighed according to the budget limit. Figure 3(d) compares the equilibrium link flow of different location designs under different budgets. It can be observed that, under a relatively low budget, the routing choice is limited, which leads to

Table 1: Link characteristics: FFTT (min), width (meter), length (mile), and capacity (103 veh/h).

Link	FFTT	Width	Length	Capacity
1-2	6	15	15	25.90
1-3	4	15	10	23.40
2-1	6	15	15	25.90
2-6	5	3.5	12.5	4.96
3-1	4	15	10	23.40
3-4	4	11.25	10	17.11
3-12	4	15	10	23.40
4-3	4	11.25	10	17.11
4-5	2	11.25	5	17.78
4-11	6	3.5	15	4.91
5-4	2	11.25	5	17.78
5-6	4	3.5	10	4.95
5-9	5	7.5	12.5	10.00
6-2	5	3.5	12.5	4.96
6-5	4	3.5	10	4.95
6-8	2	3.5	5	4.90
7-8	3	7.5	7.5	7.84
7-18	2	15	5	23.40
8-6	2	3.5	5	4.90
8-7	3	7.5	7.5	7.84
8-9	10	3.5	25	5.05
8-16	5	3.5	12.5	5.05
9-5	5	7.5	12.5	10.00
9-8	10	3.5	25	5.05
9-10	3	7.5	7.5	13.92
10-9	3	7.5	7.5	13.92
10-11	5	7.5	12.5	10.00
10-15	6	11.25	15	13.51
10-16	4	3.5	10	4.85
10-17	8	3.5	20	4.99
11-4	6	3.5	15	4.91
11-10	5	7.5	12.5	10.00
11-12	6	3.5	15	4.91
11-14	4	3.5	10	4.88
12-3	4	15	10	23.40
12-11	6	3.5	15	4.91
12-13	3	15	7.5	25.90
13-12	3	15	7.5	25.90
13-24	4	3.4	10	5.09
14-11	4	3.5	10	4.88
14-15	5	3.5	12.5	5.13
14-23	4	3.5	10	4.92
15-10	6	11.25	15	13.51
15-14	5	3.5	12.5	5.13
15-19	3	11.25	7.5	14.56
15-22	3	7.5	7.5	9.60
16-8	5	3.5	12.5	5.05
16-10	4	3.5	10	4.85
16-17	2	3.5	5	5.23
16-18	3	15	7.5	19.68
17-10	8	3.5	20	4.99
17-16	2	3.5	5	5.23
17-19	2	3.5	5	4.82
18-7	2	15	5	23.40
18-16	3	15	7.5	19.68
18-20	4	15	10	23.40
19-15	3	11.25	7.5	14.56
19-17	2	3.5	5	4.82
19-20	4	3.5	10	5.00

Table 1: Continued.

Link	FFTT	Width	Length	Capacity
20-18	4	15	10	23.40
20-19	4	3.5	10	5.00
20-21	6	3.5	15	5.06
20-22	5	3.5	12.5	5.08
21-20	6	3.5	15	5.06
21-22	2	3.5	5	5.23
21-24	3	3.5	7.5	4.89
22-15	3	7.5	7.5	9.60
22-20	5	3.5	12.5	5.08
22-21	2	3.5	5	5.23
22-23	4	3.5	10	5.00
23-14	4	3.5	10	4.92
23-22	4	3.5	10	5.00
23-24	2	3.5	5	5.08
24-13	4	3.5	10	5.09
24-21	3	3.5	7.5	4.89
24-23	2	3.5	5	5.08

Table 2: Network O-D demand (103 veh/h).

O	D	Demand
1	20	16
20	1	16
1	13	8
13	1	8
1	7	12
7	1	12
2	13	10
13	2	10
2	20	8
20	2	8
7	13	14
13	7	14
7	20	12
20	7	12

extremely high traffic flow on some links accompanying considerable travel time, like link 7 and link 35. As budget is increasing, EV drivers have more usable path to select. Equilibrium flow on such links decreases.

4.4. Strategy Comparison with Different Recharging Rates. In order to investigate the impact of the charging rate on location design and equilibrium flow (see Figure 4), we solve the DLEP and NE with charging rate ω set as 2kWh/mile (LEVEL1), 2.5kWh/mile (LEVEL2), and 3.5kWh/mile (LEVEL3). Cost of the ER is set as $\$1 \times 10^5$, $\$2 \times 10^5$, and $\$3 \times 10^5$, respectively. Budget limit is set as $\$1.9 \times 10^8$. Interestingly, the minimum total travel time can be obtained under this budget in LEVEL 1 and LEVEL 2 cases. That is, even when the budget limit increases, the total travel time will not be reduced. To explore the minimum travel time in LEVEL 3, we repeat to solve the DLEP and NE in the LEVEL 3 case with the budget limit increasing gradually and find the minimum travel time of $9.13 \times 10^4 \text{h} \cdot \text{veh}$ at the budget of $\$2.73 \times 10^8$, which is less than the travel time of $9.20 \times 10^4 \text{h} \cdot \text{veh}$ in the LEVEL 2 case. Figure 4(b) compares the total

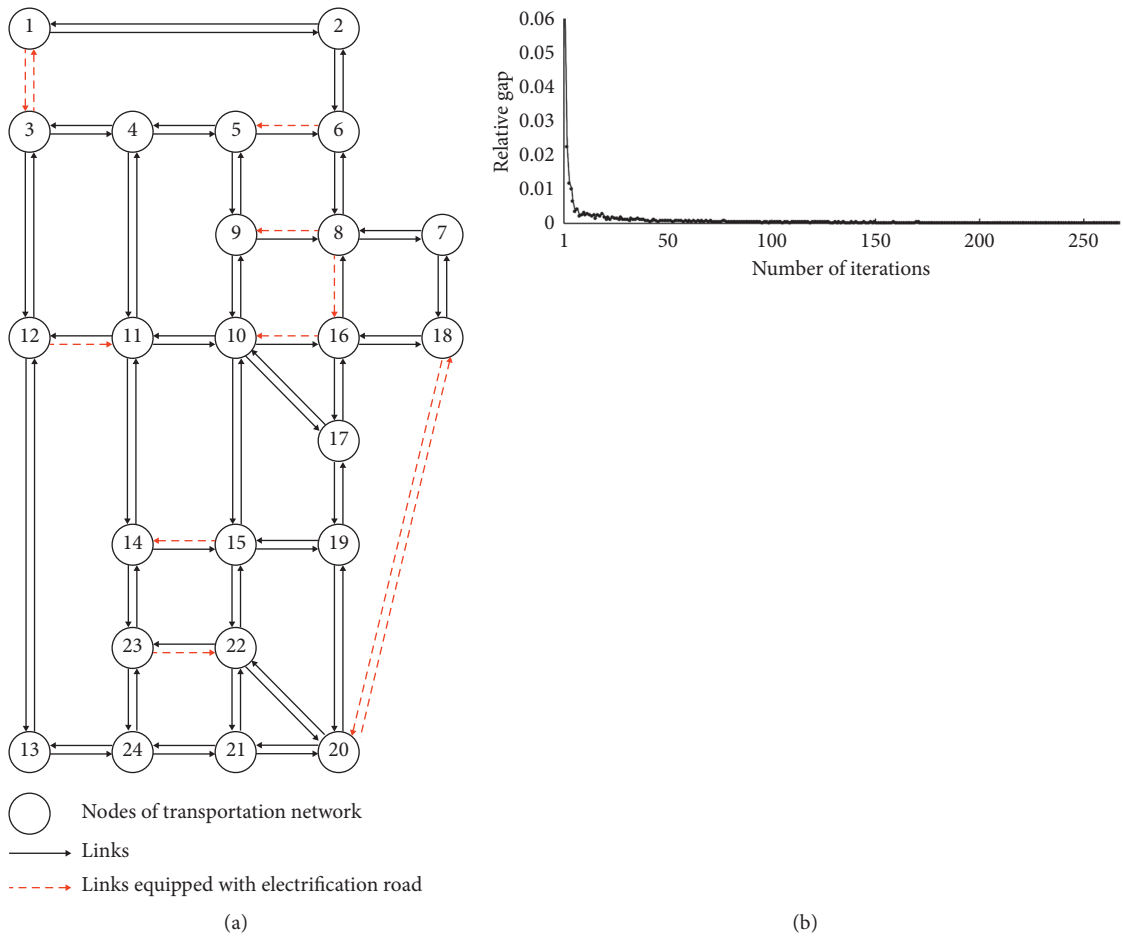


Figure 2: Base case result in the Sioux Falls network. (a) Location design of the ER. (b) Performance of the SP algorithm.

Table 3: Equilibrium link flow (10^3 veh/h).

Link	Flow	Link	Flow	Link	Flow	Link	Flow
1-2	10.15	8-7	12.05	13-12	26.68	19-15	6.58
1-3	46.04	8-9	5.99	13-24	10.72	19-17	2.56
2-1	11.36	8-16	6.75	14-11	7.59	19-20	7.78
2-6	7.96	9-5	13.75	14-15	5.34	20-18	31.25
3-1	44.82	9-8	6.10	14-23	0.13	20-19	6.62
3-4	22.77	9-10	6.53	15-10	0.00	20-21	6.09
3-12	23.33	Link	Flow	15-14	7.59	20-22	5.36
4-3	23.45	10-9	7.76	15-19	10.30	21-20	5.33
4-5	17.25	10-11	5.39	15-22	3.91	21-22	0.00
4-11	5.52	10-15	6.35	16-8	5.67	21-24	6.46
5-4	17.47	10-16	8.58	16-10	8.93	22-15	3.52
5-6	5.85	10-17	0.25	16-17	4.22	22-20	6.91
5-9	12.62	11-4	5.98	16-18	17.93	22-21	0.38
6-2	9.18	11-10	8.66	17-10	4.22	22-23	3.98
6-5	4.94	11-12	6.99	17-16	2.81	23-14	0.00
6-8	12.53	11-14	5.47	17-19	0.00	23-22	5.52
7-8	13.33	12-3	21.44	18-7	29.68	23-24	3.98
7-18	28.40	12-11	8.61	18-16	18.60	24-13	10.44
8-6	12.84	12-13	26.96	18-20	29.30	24-21	5.33
						24-23	5.39

travel time and cost in these four cases. It can be observed that, under the budget limit of $\$1.9 \times 10^8$, the ER of LEVEL 2 gets the best performance, which is much better than the other two cases. Three explanations are offered here. First, due to constraint (16), shortest length of the ER is different as recharging rate changes. In the case of LEVEL 1, only 26

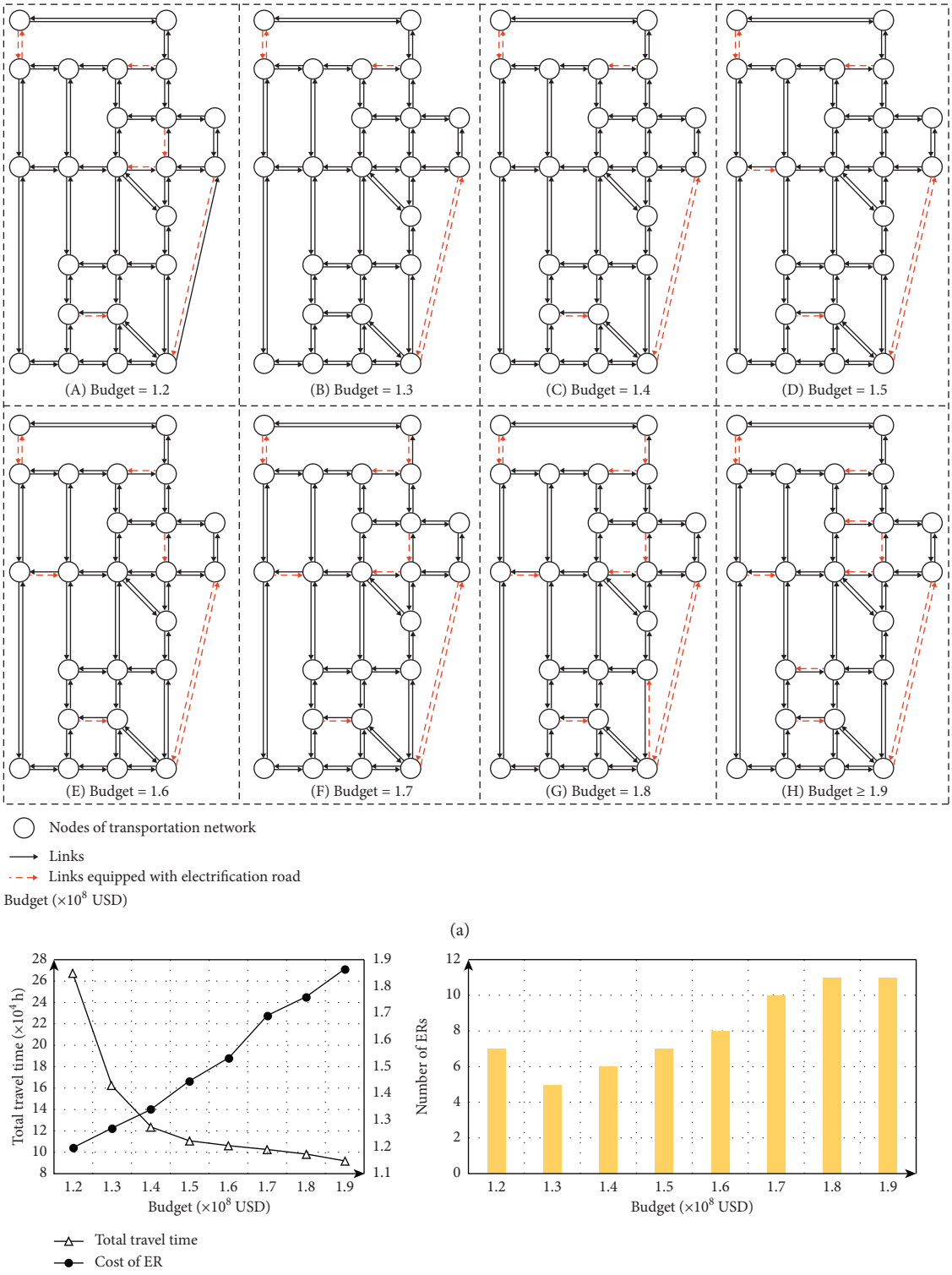
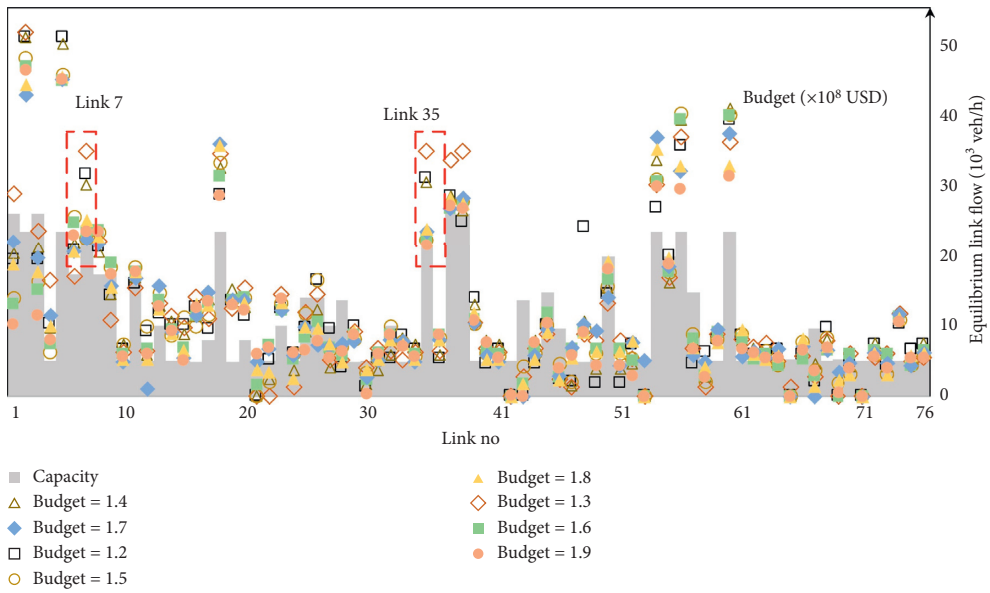
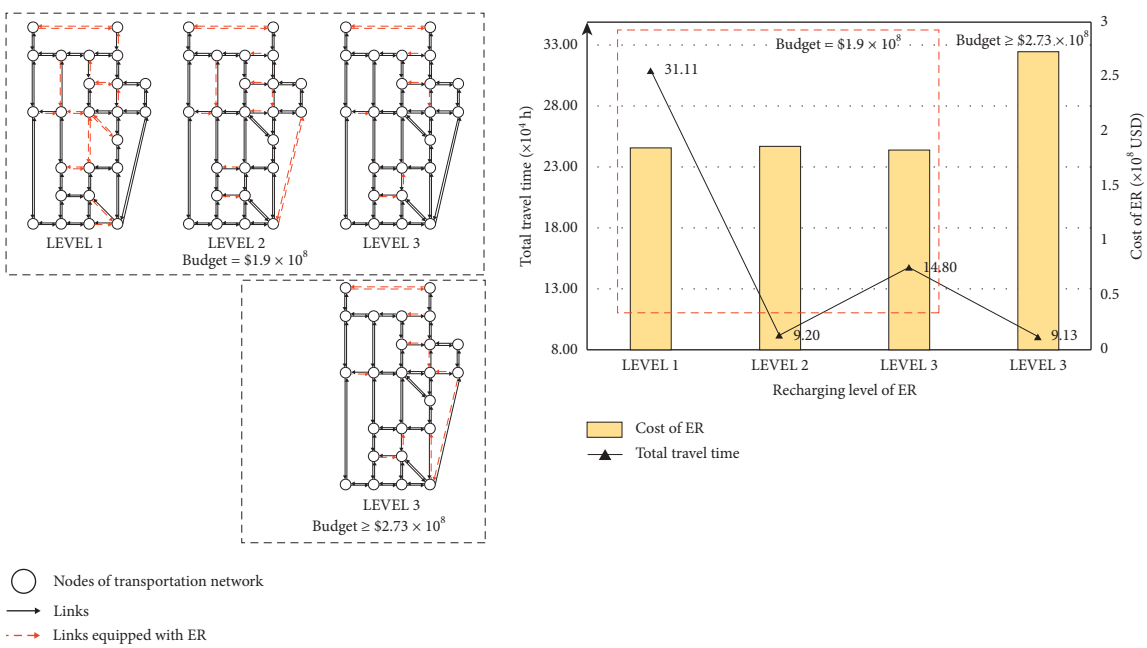


Figure 3: Continued.



(d)

Figure 3: Comparison under different budget limits. (a) Comparison of location designs under different budgets. (b) Total travel time and ER cost. (c) The number of ERs under different budgets. (d) Equilibrium flow of different location designs under different budgets.

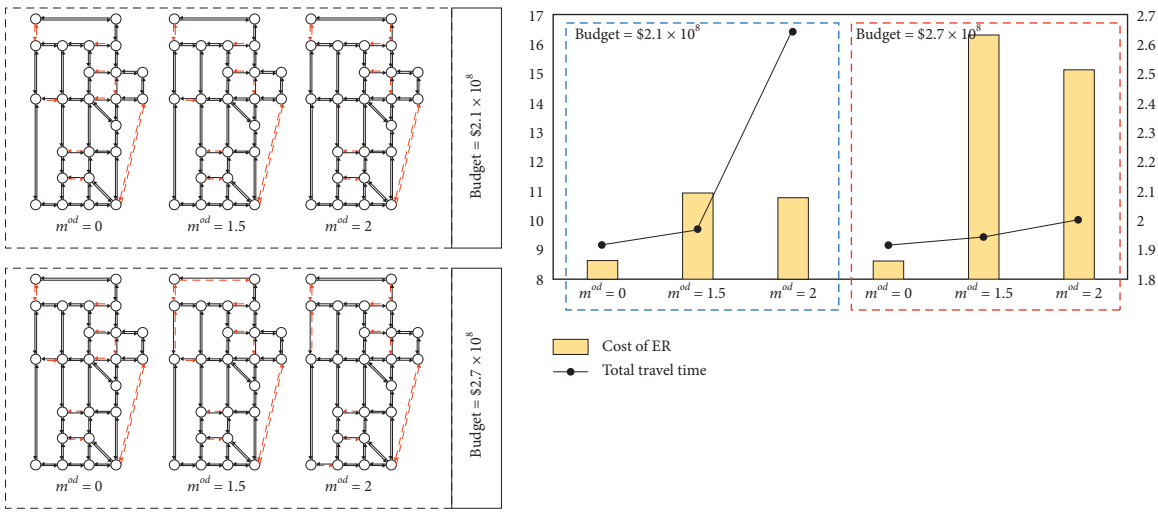


(a) (b)

Figure 4: Comparison with the ER of different recharging rates. (a) Location designs. (b) Total travel time and cost of the ER.

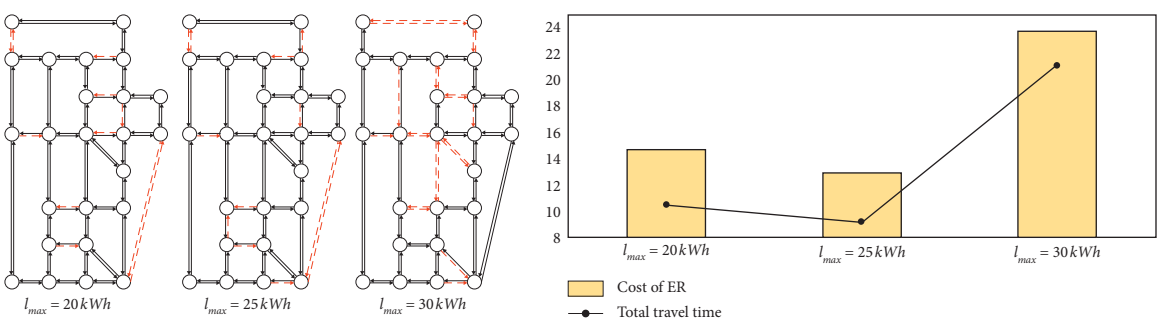
links of the length not less than 12.5 miles are qualified to be equipped with the ER. In the case of LEVEL 3, 62 links can be chosen as the candidate location of the ER. Thus, usable paths of ER drivers and total travel time are affected. Second, the cost per unit ER increases with the rise of the recharging rate, resulting in less ER constructed giving a budget constraint. Third, the state of charge and charging behavior are affected by the recharging rate, bringing about further impact on routing behavior. When the budget exceeds $\$2.73 \times 10^8$, total travel time of the network equipped with the ER of LEVEL 3 attains the minimum. As expected, more usable paths are generated as recharging rate increases, which may reduce the total travel time with more probable links sharing the traffic loads.

4.5. Impacts of Comfortable Range on Location Design and System Performance. Variations in comfortable range change the usable path sets, and thereby hinder routing choice and



(a) (b)

Figure 5: Comparison with different comfortable ranges ($\mathbf{m^{od}}$: **kWh**). (a) Location designs (m^{od}: kWh). (b) Total travel time and cost of the ER.



(a) (b)

Figure 6: Comparison of the ER with different battery sizes. (a) Location designs. (b) Total travel time and cost of the ER.

the ER location design. In this section, the impact of comfortable range on location design and system performance is tested with DLEP of different comfortable ranges $m^{od} = 1.5$ kWh and $m^{od} = 2$kWh under the budget limit of $\$2.1 \times 10^8$ and $\$2.7 \times 10^8$ solved, respectively, comparing with $m^{od} = 0$. Location designs of the ER are shown in Figure 5(a). Figure 5(b) compares the total travel time and cost in these six cases. Under the budget limit of $\$2.1 \times 10^8$, change of m^{od} from 0 to 2kWh increases the total travel time by 78%. To further explore the impacts of comfortable range, we raise the budget limit to $\$2.7 \times 10^8$ and find that the variations of the total travel time with the change in comfortable range are not obvious. That is to say, under a given budget of $\$2.1 \times 10^8$, increased comfortable range leads to the increment in the total travel time. When the budget limit is set as $\$2.7 \times 10^8$, the location design of $m^{od} = 0$ is the same as that under the budget limit of $\$2.1 \times 10^8$, with the construction cost of $\$1.865 \times 10^8$. The location designs of $m^{od} = 1.5$ kWh and $m^{od} = 2$kWh change to yield the minimum travel time, which is close to the number when $m^{od} = 0$, with the construction cost increasing to $\$2.63 \times 10^8$ and $\$2.5 \times 10^8$, respectively.

4.6. Impacts of the Battery Size on Location Design and System Performance. Different battery sizes affect the routing choice and location design in two ways. First, the shortest length of the ER is different according to constraint (16). Second, the state of charge and charging behavior are different according to constraints (7) and (9). In this section, budget limit is set as $\$4 \times 10^8$ to explore the location design and system performance under different battery sizes $l_{\max} = 20, 25$, and 30 kWh. Initial state of charge is set as $l_0 = 25\% \, l_{\max}$. As revealed in Figure 6, location design with the battery size of $l_{\max} = 25$ kWh has the least total travel time and construction cost of the ER. On the one hand, as commonly considered, a larger battery size allows more charging amount on the ER, which may generate more usable path and reduce the total travel time. On the other hand, constraint (16) sets the

minimum length of the ER which is related to the battery size. When $l_{\max} = 3$ kWh, only 26 links whose length is more than 12 miles can be the candidate of the ER, which affects the set of usable paths and the total travel time.

From the above discussion, we can find that

(1) System total travel time of the network decreases with an increase in the budget. More ERs are built with an increasing budget, generating more usable route choice for the EV to complete the tour, which promotes average distribution of traffic flow leading to reduction in the system travel time.

(2) Recharging rate sets the bottom boundary for the length of the candidate of the ER, which affects the EV drivers' routing choice. Besides, recharging rate is related to the construction cost, which should be balanced with the budget limit.

(3) Increase in comfortable range changes the sets of usable path. Therefore, system travel time increases.

(4) Battery sizes affect the shortest length of the ER as well as the initial state of charge and charging behavior, which should be balanced under different budgets.

5. Conclusion

To enhance the transportation network performance as ER is equipped, this paper proposes a LDER model to optimize the location of the ER, considering the routing and charging behavior of EV drivers by a PCNE model. EV drivers are assumed to decide their routing and recharging plan to minimize their travel time and prevent running out of charge. We develop a modified active set algorithm to obtain the optimal location of the ER. Numerical examples are conducted to demonstrate the performance of the model in comparison with the location designs in different budget limits, recharging rates, comfortable range, and battery sizes.

The energy consumption rate and the recharging rate of the proposed path-constrained network equilibrium are assumed to be flow-independent. Our future study will extend our model to investigate the effect of the traffic congestion on the energy consumption rate and the recharging rate and try to find the solution algorithm. Besides, the transportation network can be extended to accommodate both EVs and regular vehicles, with impact of the mixed traffic flow investigated.

Data Availability

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

Disclosure

The authors are responsible for all views and opinions expressed in this paper.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the Joint Science Foundation of Ministry of Education of China & CMCC (Grant No. 2018202004). The first author was supported by the Program for Changjiang Scholars and Innovative Research Team in University. The third author was supported by the National Natural Science Foundation for Young Scholars of China (Grant No. 71901190).

References

[1] C. W. Yang and Y.-L. Ho, "Assessing carbon reduction effects toward the mode shift of green transportation system," *Journal of Advanced Transportation*, vol. 50, no. 5, pp. 669–628.

[2] P. Vorobiev and Y. Vorobiev, "About the possibilities of using the renewable energy power sources on railway transport," *Journal of Advanced Transportation*, vol. 47, no. 8, pp. 681–691, 2013.

[3] M. D. Meyer, "Transport planning for urban areas: a retrospective look and future prospects," *Journal of Advanced Transportation*, vol. 34, no. 1, pp. 143–171, 2000.

[4] A. Shukla and S. A. Ansari, "Energy harvesting from road pavement: a cleaner and greener alternative," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 2, 2018.

[5] A. Northmore and S. Tighe, "Innovative pavement design: are solar roads feasible?" in *Proceedings of the Conference of the Transportation Association of Canada*, Fredericton, Canada, October 2012.

[6] A. S. Dezfooli, F. M. Nejad, H. Zakeri, and S. Kazemifard, "Solar pavement: a new emerging technology," *Solar Energy*, vol. 149, pp. 272–284, 2017.

[7] X. Yu, S. Sandhu, S. Beiker, R. Sassoon, and S. Fan, "Wireless energy transfer with the presence of metallic planes," *Applied Physics Letters*, vol. 99, no. 21, pp. 1230–1610, 2011.

[8] P. Ning, J. M. Miller, O. C. Onar, C. P. White, and L. D. Marlino, "A compact wireless charging system development," in *Proceedings of the Twenty-Eighth Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*, IEEE, Long Beach, CA, USA, March 2013.

[9] D. U. Eberle and D. R. Von Helmolt, "Sustainable transportation based on electric vehicle concepts: a brief overview," *Energy & Environmental Science*, vol. 3, no. 6, pp. 689–699, 2010.

[10] N. S. Pearre, W. Kempton, R. L. Guensler, and V. V. Elango, "Electric vehicles: how much range is required for a day's driving?" *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1171–1184, 2011.

[11] I. E. Agency and E. V. Global, "Global ev outlook. understanding the electric vehicle landscape to 2020," 2013.

[12] P. Bansal, "Charging of electric vehicles: technology and policy implications," *The Journal of Science Policy & Governance*, vol. 6, no. 1, p. 38, 2015.

[13] J. U. K. Tidy, "Test wireless charging for electric car. 2015," http://news.sky.com/story/.

[14] J. Ramsey, "Volvo studying test of electric roads in sweden," 2014, http://www.autoblog.com/.

[15] M. Fuller, "Wireless charging in California: range, recharge, and vehicle electrification," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 343–356, 2016.

[16] N. Jiang, C. Xie, and T. Waller, "Path-constrained traffic assignment: model and algorithm," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2283, no. 1, pp. 25–33, 2012.

[17] N. Jiang and C. Xie, "Computing and analyzing equilibrium network flows of gasoline and electric vehicles," in *Proceedings of the Transportation Research Board 92nd Annual Meeting*, Washington, DC, USA, January 2013.

[18] N. Jiang, C. Xie, J. C. Duthie, and S. T. Waller, "A network equilibrium analysis on destination, route and parking choices with mixed gasoline and electric vehicular flows," *EURO Journal on Transportation and Logistics*, vol. 3, no. 1, pp. 55–92, 2014.

[19] F. He, Y. Yin, and J. Zhou, "Integrated pricing of roads and electricity enabled by wireless power transfer," *Transportation Research Part C: Emerging Technologies*, vol. 34, no. 9, pp. 1–15, 2013.

[20] F. He, Y. Yin, J. Wang, and Y. Yang, "Sustainability SI: optimal prices of electricity at public charging stations for plug-in electric vehicles," *Networks and Spatial Economics*, vol. 16, no. 1, pp. 131–154, 2016.

[21] F. He, D. Wu, Y. Yin, and Y. Guan, "Optimal deployment of public charging stations for plug-in hybrid electric vehicles," *Transportation Research Part B: Methodological*, vol. 47, no. 1, pp. 87–101, 2013.

[22] F. He, Y. Yin, and J. Zhou, "Deploying public charging stations for electric vehicles on urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 227–240, 2015.

[23] F. He, Y. Yin, and S. Lawphongpanich, "Network equilibrium models with battery electric vehicles," *Transportation Research Part B: Methodological*, vol. 67, no. 3, pp. 306–319, 2014.

[24] M. Xu, Q. Meng, and K. Liu, "Network user equilibrium problems for the mixed battery electric vehicles and gasoline vehicles subject to battery swapping stations and road grade constraints," *Transportation Research Part B: Methodological*, vol. 99, pp. 138–166, 2017.

[25] R. Riemann, D. Z. W. Wang, and F. Busch, "Optimal location of wireless charging facilities for electric vehicles: flow-capturing location model with stochastic user equilibrium," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 1–12, 2015.

[26] Z. Chen, F. He, and Y. Yin, "Optimal deployment of charging lanes for electric vehicles in transportation networks," *Transportation Research Part B: Methodological*, vol. 91, pp. 344–365, 2016.

[27] Z. Chen, W. Liu, and Y. Yin, "Deployment of stationary and dynamic charging infrastructure for electric vehicles along traffic corridors," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 185–206, 2017.

[28] Z. Liu and Z. Song, "Robust planning of dynamic wireless charging infrastructure for battery electric buses," *Transportation Research Part C: Emerging Technologies*, vol. 83, pp. 77–103, 2017.

[29] Z. Liu, Z. Song, and Y. He, "Optimal deployment of dynamic wireless charging facilities for an electric bus system," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2647, no. 1, pp. 100–108, 2017.

[30] A. S. Ahmed, A. M. Mohamed, and Z. H. U. Lei, "System design and optimization of in-route wireless charging infrastructure for shared automated electric vehicles," *IEEE Access*, vol. 7, pp. 79968–79979, 2019.

[31] Z. Bi, G. A. Keoleian, Z. Lin et al., "Life cycle assessment and tempo-spatial optimization of deploying dynamic wireless charging technology for electric cars," *Transportation Research Part C: Emerging Technologies*, vol. 100, pp. 53–67, 2019.

[32] T. Zhao, Z. Wu, and Y. Zhang, "Bi-objective optimization of electric bus dynamic wireless charging facilities location and onboard battery size considering both facility cost and energy consumption," in *Proceedings of the Transportation Research Board 98th Annual Meeting*, Washington, DC, USA, January 2019.

[33] L. J. Leblanc, "An algorithm for the discrete network design problem," *Transportation Science*, vol. 9, no. 3, pp. 183–199, 1975.

[34] T. L. Friesz, H.-J Cho, N. J. Mehta, R. L. Tobin, and G. A. Nandalingam, "A simulated annealing approach to the network design problem with variational inequality constraints," *Informs*, vol. 26, no. 1, pp. 18–26, 1992.

[35] S. Wang, Q. Meng, and H. Yang, "Global optimization methods for the discrete network design problem," *Transportation Research Part B Methodological*, vol. 50, pp. 42–60.

[36] L. Zhang, S. Lawphongpanich, and Y. Yin, *An Active-Set Algorithm for Discrete Network Design Problems*, Springer Science+Business Media, Berlin, Germany, 2009.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 4905059, 22 pages
https://doi.org/10.1155/2020/4905059

Research Article

Modeling the Curb Parking Price in Urban Center District of China Using TSM-RAM Approach

Yan Wan,[1,2] **Jibiao Zhou**,[3,4] **Wenqiang He**,[1] **and Changxi Ma**[5]

[1]*School of Civil and Transportation Engineering, Ningbo University of Technology, Ningbo 315211, China*
[2]*Universität Duisburg-Essen, Duisburg 47057, Germany*
[3]*Department of Transportation Engineering, Tongji University, Shanghai 201804, China*
[4]*Intelligent Transport System (ITS) R & D Center, Shanghai Urban Construction Design and Research Institute (Group) Co., Ltd., Shanghai 201804, China*
[5]*School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China*

Correspondence should be addressed to Jibiao Zhou; zhoujibiao@tongji.edu.cn and Changxi Ma; machangxi@mail.lzjtu.cn

Received 8 December 2019; Revised 16 June 2020; Accepted 23 June 2020; Published 17 July 2020

Academic Editor: Zhigang Xu

Parking demand forecasting is an important part of urban parking planning and is also an important basis for the development of parking facilities. The primary objective of this study was to explore multiple factors that affect the curb parking price (CPP) and the changing rules of the curb parking price (CPP) with these factors and to predict the CPP in terms of urban mobility. The data were collected through a statistical survey that was administered in 81 cities in China. The cities were divided into three categories: rich cities (RCs), poor cities (PCs), and tourist cities (TCs). Both the time series method (TSM) and regression analysis method (RAM) were developed to simultaneously examine the factors associated with the CPP among parking users. The results showed that TSM and RAM can account for common urban curb parking prices. The prediction results showed that the CPP is affected by the number of urban dwellers (UD), the prevalence of car ownership (CO), and the per capita disposable income (PCDI) of urban residents; the CPP can be predicted by a model built on the basis of the above three influencing factors. The results can enhance our understanding of the factors that affect CPP. Based on the results, some suggestions regarding the use of the CPP range in parking policy planning were discussed.

1. Introduction

Curb parking is a public resource [1]. At present, the average ratio of cars to parking spaces in the entire urban district of large cities in China is approximately 1 : 0.8; compared with the ratio of 1 : 1.3 in developed countries, the proportion is seriously low, and the national parking space shortfall is of more than 50 million spaces [2]. For example, there are 5.64 million vehicles in Beijing but only 1.93 million parking spaces; an average of 2.92 vehicles share each parking space, and the parking space gap is 3.71 million [2]. The number of cars in Shenzhen is 3.22 million, and the total number of parking spaces is 1.11 million. An average of 2.90 vehicles share each parking space, and the parking space gap is 2.11 million [2]. Although the number of motor vehicles in Shanghai is only 3.59 million, less than 60% that of Beijing, the total number of parking spaces is only 600,000, and an average of 5.98 vehicles are allocated to one parking space. [2]. Since people are more likely to travel by private motor cars, road traffic congestion is becoming increasingly serious due to the imbalance between parking supply and demand. Statistical data show [3, 4] that the search time for a curb parking space in CBDs (central business districts) accounts for 40% of the total travel time, which also proves that the lack of parking spaces aggravates road congestion to some extent.

Given the current grim parking situation in Chinese cities such as Beijing, Shanghai, and Shenzhen, the government has put forward restrictive measures to address the imbalance between parking supply and demand. For

example, Beijing has used the experience of Japan and other countries to predict the increase in the number of parking places, but this has not been effective. Many cities have strict parking rules, but violations continue, and the expected effect of the policy has not been achieved [5]. Research on intelligent parking systems has focused on parking induction and berth prediction for more than ten years [6] but has been unable to effectively address the imbalance of parking supply and demand. To alleviate the negative impact of parking problems, a common method is to induce adaptation to parking supply by reducing parking demand, and controlling the parking price is an effective means of adjusting the balance between the two [7]. In many cities in the UK, parking management measures are often combined with changes in parking prices [8–10]. Drivers are more sensitive to changes in the curb parking price (CPP) than to changes in public transportation or fuel prices because it represents a direct user cost [11]. A CPP that is too low or too high can make parking demand unreasonable in terms of the spatial distribution, leading drivers to spend a great deal of time searching for available parking spaces [12–14]. Therefore, CPP management can be used as a parking management policy to reduce parking problems in certain areas [15, 16]. Furthermore, if the CPP in some areas can be predicted, it will provide references for the formulation and implementation of parking policies.

The current research on the curb parking pricing (CPP) mostly uses discrete selection models, game theory, nonlinear decision analysis models, and other methods to evaluate the rationality of on-street parking pricing, while there are few studies on the curb parking pricing (CPP) prediction. The independent predictors selected by the above model are less predictable and are only applicable to the status quo evaluation. Therefore, it is difficult to help policy makers have a clearer understanding of the city's curb parking prices (CPP) in the future years before formulating relevant policies, which is likely to cause short-sightedness and irrationality of parking policies.

To fill this gap, we propose a mixed forecast method for curb parking price forecasting, combining the time series method (TSM) and regression analysis method (RAM). By collecting historical parking prices, we can predict the future CPP for the core areas of Ningbo, Yancheng, and Kunming. Finally, the practical application shows that the forecast method can be applied to other cities in China. This prediction of the curb parking price can help decision makers develop better parking strategies to balance parking supply and demand and thoroughly solve the problem of urban parking.

This research makes the following three contributions. First, we develop a curb parking price model that combines the TSM approach and RAM approach, using historical data on the number of urban dwellers (UD), car ownership (CO), and per capita disposable income (PCDI) of urban residents to estimate the 2 h CPP in the core city for the next year. Second, we divide China's cities into the rich cities (RCs), poor cities (PCs), and tourist cities (TCs) and analyze the CPP for the cities under different scenarios. Third, we classify the prediction results for the future year as optimistic, aggressive, or conservative to ensure the accuracy of the outcomes.

The remainder of this paper is organized as follows. Section 2 reviews existing studies on the TSM and the RAM. Section 3 details the data collection method. Section 4 introduces the urban curb parking pricing model, including the model assumptions, construction, and testing. Section 5 describes an application of the curb parking pricing (CPP) model. Section 6 presents the prediction results of the model. Section 7 discusses the results obtained and draws conclusions.

2. Literature Review

2.1. Time Series Method Analysis. The time series method (TSM) is a method for establishing mathematical models based on time series data obtained by systematic observation. It is generally performed using curve fitting and parameter estimation methods (i.e., nonlinear least squares) and is widely used in the fields of economics [17], geography [18], and electrical engineering [19].

In economics, Zhu [20] used the seasonal time series model to analyze the compensation data of China's insurance industry from January 2002 to November 2008 and established the Box-Jenkins seasonal model. The results show that the model offers good prediction results, which can provide a reference for the supervision of and decisions on insurance payments in China. While the prediction error increases as the forecast period is extended, data can be continuously added to achieve dynamic prediction after the model is built, which maintains high prediction accuracy. Rabindra and Nirash [21] examined the interrelationships between energy consumption, output, and carbon emissions in a developing economy using an augmented vector autoregression model. Time series data for the period 1975–2013 in Nepal were studied using the population and gross fixed capital formation as additional variables. The authors found that the government of Nepal can address energy poverty by accelerating the adoption of energy conservation policies such as rationing energy consumption and energy efficiency improvements to narrow the energy supply-demand gap. The results remained robust across different estimators and contribute to an emerging literature on the nexus connecting energy consumption, income, and carbon emissions in developing economies.

In the field of geography, Yin et al. [18] assessed the vegetation utilization rate, determined through high-resolution remote sensing inversion, in the Inner Mongolia autonomous region from 1999 to 2009. The time series analysis method was used to evaluate the desertification development trend with reference to a discussion of the desertification reversal problem. The results showed that the ecological environment in the Inner Mongolia autonomous region had generally improved over the 11 years of the sample period, but the development trend of desertification was not obvious. Reiser and Kutie [22] detected significant trends in the uncertainty of total annual rainfall, the number of rain spells, rain-spell yields, and rainy season length from the time series data of 41 weather stations across the Mediterranean region from 1931 to 2006. In addition, they detected significant temporal changes in the

occurrence of extreme events on these parameters. The study found that despite the general assumption of tremendous changes in the rainfall regime, there were no significant temporal or uncertainty trends for total annual rainfall, number of rain spells, rain-spell yields, or rainy season length at most of the stations. However, in a few cases, a significant trend was detected.

In electrical engineering, Ding et al. [23] proposed a wind speed prediction model based on time series analysis and used an information criterion to test the model. The results showed that the time series model was suitable for the prediction of wind speeds on wind farms and could reflect future wind speed distribution characteristics. Abdelaal and Algarni [24] used data on monthly domestic electric energy consumption in the eastern province of Saudi Arabia for 5 years to develop autoregressive integrated moving average models and evaluated forecast data for the sixth year. The results showed the optimum ARIMA model forecast monthly data for the evaluation year with an average percentage error of 3.8%, compared to 8.1% and 5.6% for the best multiple-series regression and mechanism models, respectively. In other words, the ARIMA model reduced the mean-square forecasting error by factors of 3.2 and 1.6, respectively. Time series analyses in the parking field, on the other hand, have been limited to assessments of the evolution of the number of parking places, and there has been little attention paid to the exploration of parking prices.

2.2. Regression Analysis Method. Since the regression analysis method (RAM) is simpler and more convenient for analyzing multifactor models, its application in many fields, such as mathematics [25], chemistry [26], computer science and technology [27], medicine [28], and other fields, is quite mature, and the results of such analysis can be accurate. The RAM can measure the degree of correlation between various factors and the goodness of fit of the regression.

In the field of mathematics, Lin et al. [29] discussed the estimation and testing of unknown parameters in multiple linear regression models and offered an example thereof. The study found that the accuracy and reliability of the recorded data and anomalies in the data set affect the predictive analysis of the dependent variable. Val et al. [30] evaluated the gait parameters and activity of 87 healthy subjects between the ages of 21 and 84, and they constructed a model of each gait parameter and activity measure by regression analysis. The results showed that normalizing gait parameters and activity metrics through a linear regression model can enhance the ability to compare objects with different anthropometric values.

In the chemistry field, Yu et al. [31] used near-infrared spectral data of water molecules to establish a support vector regression model for moisture content. The study found that the predicted root mean-square error of the set of support vector regression models was 2.930%, the correlation coefficient was 0.994, and the relative analysis error was 9.473. Kaneko et al. [32] combined independent component analysis and regression analysis methods to extract the significant components, verified the superiority of ICA-MLR (multiple linear regression) over partial least squares with simulation data, and tried to apply this method to a quantitative structure-property relationship analysis of aqueous solubility. The results showed that ICA-MLR achieved higher predictive accuracy than PLS. The study also found that ICA-MLR could extract the effective components from explanatory variables and construct the regression.

In the computer science field, Liu et al. [33] established a linear regression model between TPC-C server performance and the hardware index by using the mathematical statistics method after analyzing various factors affecting the performance of TPC-C. The basic data of the model were derived from the TPC-C server test results from 2008 to 2013. The results showed that the optimized model estimation accuracy was over 95% and could explain the causal relationship between the hardware index of the server and TPC-C performance to a certain extent. Peng et al. [34] extracted residual images by using multiple linear regressions to discriminate between natural images and computer-generated graphics and then investigated the fitting degree of the regression model. Experimental results and analysis show that it can achieve an average identification accuracy of 98.69%, and it is robust against JPEG compression, rotation, additive noise, and image resizing.

In the medical field, Ma et al. [35] performed multiple linear regression analysis of X-ray measurements and WOMAC scores of knee osteoarthritis and analyzed their relationship with clinical and biomechanical concepts. The results showed the statistical significance of AP X-ray values and WOMAC scores ($P < 0.05$) but no statistical significance of lateral X-ray values and WOMAC scores ($P > 0.05$). Kondo et al. [36] proposed a logistic group method of data handling (GMDH)-type neural network and applied it to medical image diagnostics for lung cancer, using principal component-regression analysis to estimate the parameters of the neural network. The identification results show that the logistic GMDH-type neural network algorithm is useful for medical image diagnostics for lung cancer since the optimum neural network architecture is automatically organized to fit the complexity of the medical images.

Turning our attention to our research setting, we note that parking prices are actually related to a number of external factors [37], and the RAM approach is easier to analyze multifactor models; hence, the RAM approach can also be applied in the field of parking prices. In related research, Kelly and Clinch [38] used questionnaires to investigate the difference in price sensitivity between trips made for business purposes and those made for nonbusiness purposes. Ordered probit regression analysis was used on survey responses from travelers who used parking spaces to a series of suggested increases in localized on-street parking tariffs. The results showed a progressively widening gap in price sensitivity as the suggested parking prices increased. The results highlighted that the potentially varied impact of pricing measures on specific subsets of the market are often overlooked in the policymaking process. Albalate and Gragera [39] explored the determinants of garage prices by drawing on a new self-constructed database for all garages in the city of Barcelona. The results indicated that prices are mainly influenced by drivers' fixed and variable costs, the dominant

position of the garage in its surrounding market and the garage's interaction with curbside parking. The study also found that the garage prices react to the scarcity of street parking spaces and to the curbside price fixed by the public authority. Kobus et al. [40] introduced probabilistic regression analysis to estimate the effect of parking prices on car drivers' choice between street and garage parking. The methodology was applied to daytime parking data for an area where cruising for parking is absent, street parking is ubiquitous, and garage parking is discretely located over space. The study found that drivers were willing to pay a premium for street parking that ranged from €0.37 to €0.60, and the demand for street parking was price elastic. The results showed that even small reductions in street parking prices induced a strong increase in the stock of cars parked on-street, and a policy that imposed a premium on-street parking reduced the total number of parking places. The curb parking prices (CPP) are largely a manifestation of the relationship between parking supply and demand. Increasing parking supply will induce higher parking demand, which will lead to a vicious circle; increasing parking prices can increase parking supply while suppressing parking demand. Therefore, the study of curb parking prices (CPP) is actually to solve the problem of parking demand, such as parking turnover rate and parking duration.

In order to explore multiple factors that affect the curb parking price (CPP) and the changing rules of the CPP with these factors and to predict the CPP in terms of urban mobility, this article reviews the domestic and foreign research related to time series and regression analysis. From the analysis results of the literature review, time series methods and regression analysis methods can be used to achieve the purpose of this article.

3. Data Collection

It should be noted that the number of urban dwellers (UD), the prevalence of car ownership (CO), and per capita disposable income (PCDI) are macrolevel influencing factors of curb parking prices. Because the topic of this study is mainly curb parking prices in urban core areas, the spatial scope is relatively wide, and the degree of influence of microlevel factors such as major and minor roads and road length on curb parking prices in the sample area is assumed to be relatively negligible.

The basic data of the article come from the China Statistical Yearbook [2]. The reliability of data sources can be guaranteed. In the process of data source selection, this article compares the national economic and social development statistical bulletins of various cities to avoid potential errors in data sources. Among them, the number of urban dwellers (UD) and per capita disposable income (PCDI) can be directly obtained, while car ownership (CO) needs to be determined through further calculation of the data in statistical yearbooks.

Except for the RCs and PCs, cities can be divided by their PCDI, and the TCs appear in this section in a separate category because of their special urban functions and positioning. Some RCs, such as Suzhou and Xiamen, and PCs, such as Guiyang, need to be attentive to the impact of tourism, which accounts for a large proportion of their gross domestic product (GDP). Therefore, some RCs and PCs can also be classified as TCs.

TCs, for their part, can be considered a city classification between RCs and PCs. This study collects data on the following variables for the core areas of 36 RCs, 26 PCs, and 31 TCs: 2 h CPP, UD, CO, and PCDI. The statistical results are shown from Tables 1 to 3.

4. Urban Curb Parking Pricing Model

4.1. Model Assumptions. The core area of the city is the main part of the urban public activity system. It displays a certain agglomeration effect and is an important place for urban residents to carry out various activities and exchanges. Therefore, the parking problem is an important focus in such areas. The basic hypotheses of the CPP prediction model in this paper are as follows.

4.1.1. The CPP (Y) Has a Clear Relationship with UD (X_1), CO (X_2), and PCDI (X_3), Which Are Significant at $P \leq 0.05$. The CPP is generally common to entire urban areas. Therefore, the UD and PCDI are selected as the predictor variables. At the same time, the parking prices basically apply only to cars, so CO is selected as another predictor variable. In a study, Humphrey and Swingley [30] took PCDI and motor vehicle ownership as predictor variables, but the study's accuracy has still not been established.

4.1.2. The Three Variables, UD (X_1), CO (X_2), and PCDI (X_3), Increase with Time (t). To obtain the future 2 h CPP in the core area of the city, it is necessary to calculate UD, CO, and PCDI for the coming years. If UD, CO, and PCDI have an increasing trend in terms of time, future variation in the three parameters can be obtained by using the TSM. Thus, the CPP in the core area of the city can be predicted by means of this method.

4.1.3. The Absolute Values of the Prediction Error of the Three Parameters, UD (X_1), CO (X_2), and PCDI (X_3), Are below 0.05, i.e., $\Delta \leq 0.05$. The prediction accuracy refers to the degree of density or dispersion in the prediction error distribution, that is, the dispersion between the actual and the corresponding predicted values. If the prediction error Y_{VC} is small, it indicates that the prediction accuracy is high, and if the prediction error is large, it indicates that the prediction accuracy is low. Therefore, the prediction accuracy values of UD, CO, and PCDI are related to the CPP prediction accuracy. In this paper, an absolute value of the prediction error below 0.05 is considered to be within an acceptable range.

4.2. Parking Pricing Prediction Model Construction. The relationship between (Y) and each of the two variable combinations of the three variable types (X_1, X_2, X_3) is fitted

Table 1: The curb parking price in the core areas in RCs (2017).

Cities	2 h CPP (yuan)	UD (10 thousand)	CO (10 thousand)	PCDI (yuan)
Beijing	40	1876.6	564	62406
Shanghai	70	2120.88	359	62596
Suzhou	18	1068.36	355	58806
Shenzhen	21	1249.57	322	52938
Dongguan	13	749.66	263	42944
Wuhan	7	853.65	261	43409
Qingdao	16	625.25	246	50817
Hangzhou	22	727.14	244	56276
Guangzhou	32	1248.9	240	55400
Nanjing	28	685.89	239	54538
Ningbo	18	579.56	229	55656
Foshan	18	727.11	228	46849
Changsha	15	614.38	217	46948
Shenyang	17	668.2	210	41359
Fuzhou	16	510.5	118	40973
Xiamen	28	357.3	123	50019
Dalian	12	417.7	140	40587
Wenzhou	18	574.68	183.2	51866
Shaoxing	16	328.2	149.52	54445
Jiaxing	20	300.31	119.51	53057
Wuxi	14	498.03	176.45	52659
Taizhou	11	380.54	148.3	51374
Changzhou	12	338.7	122.8	49955
Tianjin	16	1291.11	287	37022
Harbin	10	463.8	162	35546
Hefei	9	587.4	169.74	37972
Nanchang	10	289.78	167	37575
Wuhu	9	240.42	147.68	35175
Jinan	20	483.75	195	46642
Nantong	9	482.4	187.3	42661
Quanzhou	7	568.3	167	42696
Tangshan	14	486.8	185	36415
Yantai	8	451.31	187.31	41837
Zhuhai	11	157.8	88.4	46826
Xi'an	12	679.26	271	38636
Kunming	16	467.7	215	39788

Table 2: The curb parking price in the core areas in PCs (2017).

Cities	2 h CPP (yuan)	UD (10 thousand)	CO (10 thousand)	PCDI (yuan)
Yancheng	12	489.19	81.67	33115
Xingtai	3	379.12	78.63	26179
Tieling	4	127.8	43.2	28337
Liaoyuan	3	59.61	12.91	30267
Tongliao	3	95.69	46.2	29667
Lu Liang	4	154.04	35.3	28704
Zhumadian	3	162.02	81.6	26340
Linyi	4	305.88	105	33266
Suzhou	4	66.76	44.4	32392
Suqian	3	187.5	44.85	28118
Huanggang	3	234.1	43.2	26884
Huaihua	3	128.9	51.97	29498
Hanzhong	4	129.66	60.16	28812
Guiyang	6	389.19	141.4	32186
Lanzhou	8	226.05	101.7	31071
Xining	5	167.53	100.1	32043
Yinchuan	4	171.56	75.88	32981
Ganzhou	5	343.38	79.86	29567

Table 2: Continued.

Cities	2 h CPP (yuan)	UD (10 thousand)	CO (10 thousand)	PCDI (yuan)
Taiyuan	9	370.97	143.63	33469
Kaifeng	3	215.73	54.2	29864
Nanning	6	378.44	104.29	33217
Jingdezhen	6	109.8	18	34283
Huainan	4	221.29	59.6	32405
Qinhuangdao	3	180.1	62.1	32795
Changchun	7	438.3	83.75	33168
Fushun	4	146	28.9	30346

in MATLAB, and it was found that all display quadratic curve relationships, as shown in Figures 1–3.

The purpose of the data visualization step was to determine the relationship and trend between the curb parking price (CPP) and the three independent variables to create a theoretical basis for the construction of the parking pricing model.

4.2.1. RAM Model. All the basic data were fitted again, and all three city types showed the highest fit with the quadratic curve. Therefore, the following ternary quadratic function can be established to describe the relationships among Y, X_1, X_2, and X_3:

$$Y = a_1X_1^2 + a_2X_2^2 + a_3X_3^2 + a_4X_1X_2 + a_5X_1X_3 + a_6X_2X_3 + C, \tag{1}$$

where $a_1, a_2, \ldots, a_6$ are the regression coefficients and C is a constant value.

For convenience, equation (1) was converted to a six-element linear regression as follows:

$$Y = a_1z_1 + a_2z_2 + a_3z_3 + a_4z_4 + a_5z_5 + a_6z_6 + C. \tag{2}$$

The independent variable is changed from the original variables X_1, X_2, X_3 to $z_1, z_2, z_3, z_4, z_5, z_6$. Y is output as the dependent variable and $z_1, z_2, z_3, z_4, z_5, z_6$ are input as the independent variables. The CPP models of RC (Y_{RC}), PC (Y_{PC}), and TC (Y_{VC}) are obtained as follows:

$$\begin{aligned}
Y_{\mathrm{RC}} &= 4.2979 * 10^{-5}X_1^2 + 9.2284 * 10^{-4}X_2^2 + 9.8863 * 10^{-9}X_3^2 - 3.7286 * 10^{-4}X_1X_2 \\
&\quad + 6.0596 * 10^{-7}X_1X_3 - 3.3812 * 10^{-6}X_2X_3 + 1.6041, \\
Y_{\mathrm{PC}} &= 1.1069 * 10^{-4}X_1^2 + 2.7739 * 10^{-3}X_2^2 + 6.2943 * 10^{-9}X_3^2 - 8.3403 * 10^{-4}X_1X_2 \\
&\quad + 4.9973 * 10^{-7}X_1X_3 - 2.1672 * 10^{-6}X_2X_3 - 2.3045, \\
Y_{\mathrm{VC}} &= 6.9089 * 10^{-5}X_1^2 + 6.6710 * 10^{-4}X_2^2 + 7.4237 * 10^{-9}X_3^2 - 4.3752 * 10^{-4}X_1X_2 \\
&\quad - 1.9094 * 10^{-7}X_1X_3 + 7.7826 * 10^{-7}X_2X_3 - 2.1381.
\end{aligned} \tag{3}$$

4.2.2. TSM Model. A time series is a sequence of successive observations of the same phenomenon at different times. Here, t is used to indicate the time of the observation, X is the observed value, and X_i $(i = 1, 2, \ldots, m)$ is the observed value at time t_i.

Exponential smoothing is a method of predicting the weighted average of past observations, which makes the predicted value of the $t + 1$ period equal to the weighted average of the actual observation value in period t and the predicted value in period t. An exponential smoothing method is a special form of a weighted average. At longer observation time horizons, the weight of the index is decreased. The single exponential smoothing method has only one smoothing coefficient and uses the linear combination of the predicted value and the observed value for a period as the predicted value for period $t + 1$. The prediction model is as follows:

$$F_{t+1} = \alpha X_t + (1 - \alpha)F_t, \tag{4}$$

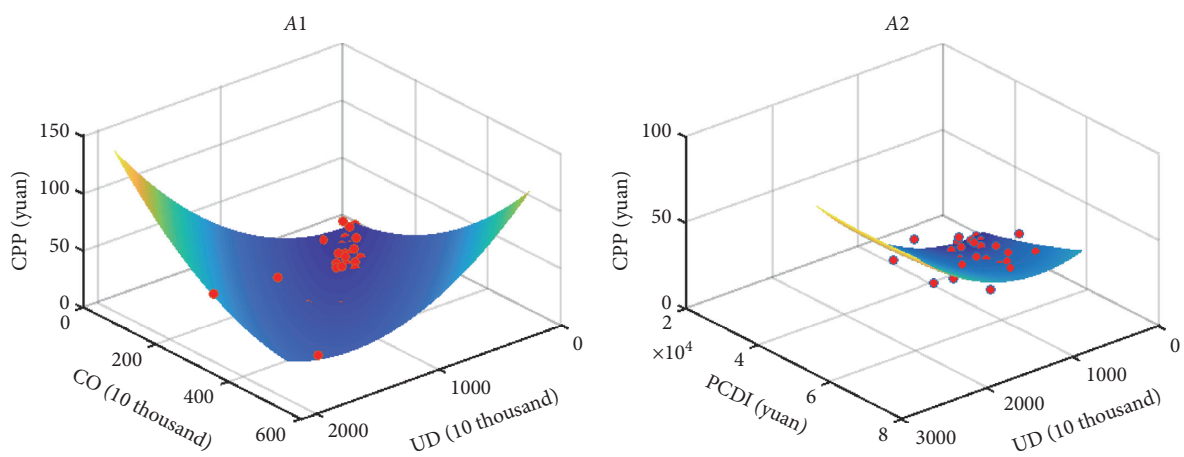
where R^2 is the predicted value for period R^2, R^2 is the observed value for period R^2, and X_1 is the smoothing coefficientX_2.

It can be seen from the above equation that the predicted value of period X_3 is a weighted average of the actual observed value in period t and the predicted value in period t. At the beginning of the calculation, there is no prediction F_1 for the first period, so we assume that $F_1 = X_1$.

By analogy, it can be seen that any predicted value F_{t+1} is a weighted average of all previous actual observations. For

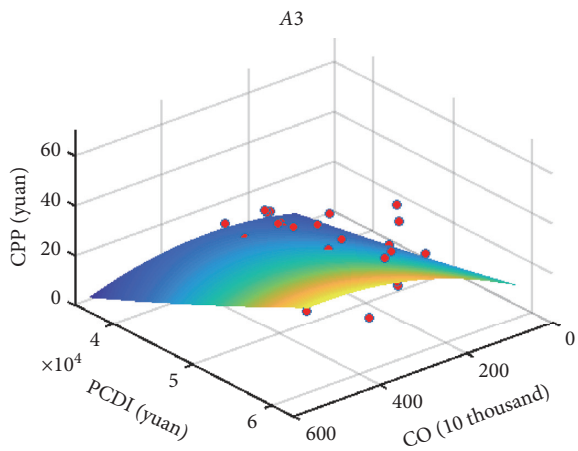
Table 3: The curb parking price in the core areas in TCs (2017).

Cities	2 h CPP (yuan)	UD (10 thousand)	CO (10 thousand)	PCDI (yuan)
Tianjin	16	1291.11	287	37022
Chengdu	16	1152.81	452	38918
Kunming	16	467.7	215	39788
Suzhou	18	1068.36	355	58806
Zhuhai	11	157.8	88.4	46826
Yantai	8	451.31	187.31	41837
Xiamen	28	357.3	123	50019
Yangzhou	7	197.77	67	38828
Dalian	12	417.7	140	40587
Qingdao	16	625.25	246	50817
Haikou	16	103.95	77.25	33320
Xi'an	12	679.26	271	38636
Hangzhou	22	727.14	244	56276
Shenzhen	21	1249.57	322	52938
Nanjing	28	685.89	239	54538
Guangzhou	32	1248.9	240	55400
Shenyang	17	668.2	210	41359
Harbin	10	463.8	162	35546
Changchun	7	438.3	143.75	33168
Jinan	20	516.36	191.1	46642
Huangshan	4	70.45	18.48	30821
Guilin	5	247.34	56.15	32534
Weihai	4	187.79	66.16	27898
Qinhuangdao	3	180.1	62.1	32795
Sanya city	7	443.62	112	33638
Xianyang	4	219.94	83.4	34246
Dunhuang	4	12.79	10.2	31322
Taian	4	342.3	69.39	32739
Lijiang	4	46.4	35.2	30403
Jingdezhen	6	109.8	18	34283
Guiyang	6	389.19	131.4	32186



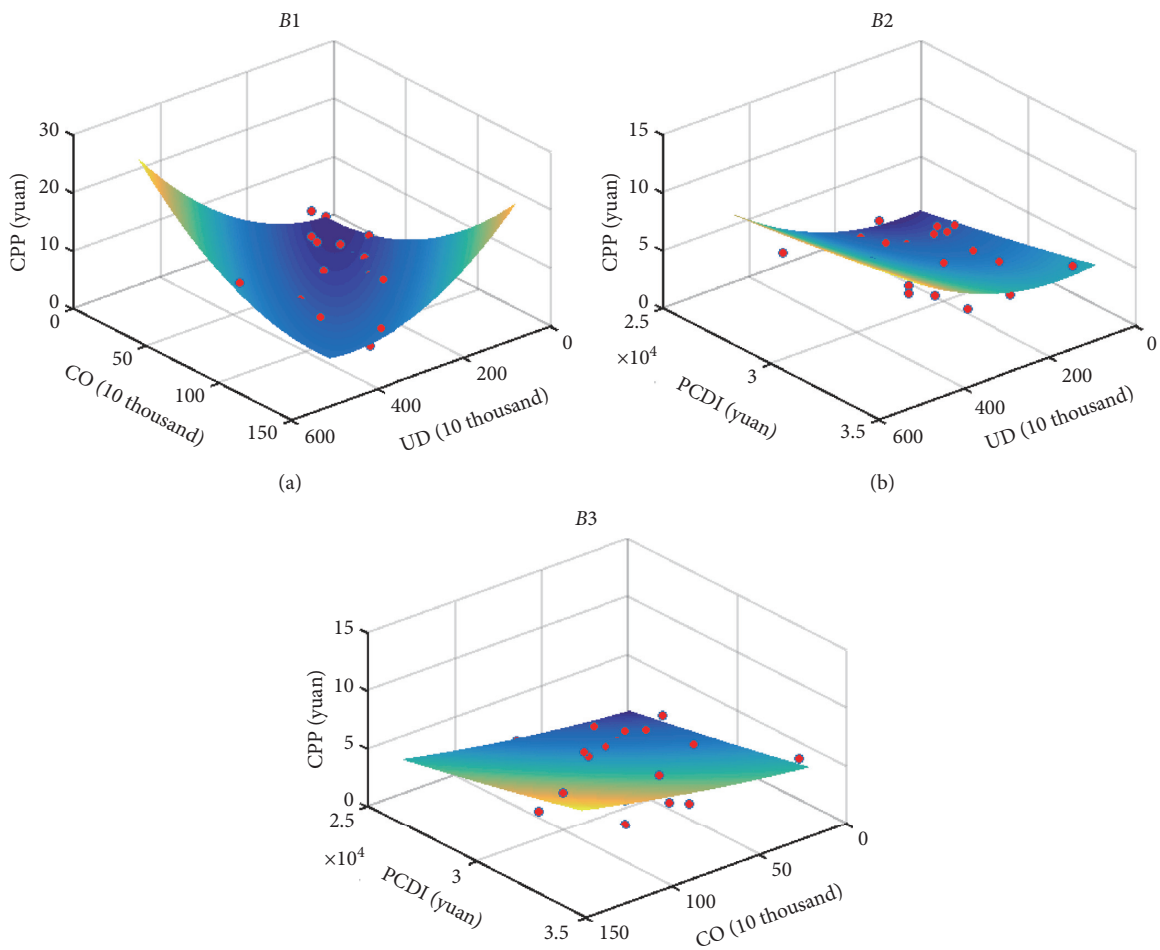
(a) (b)

Figure 1: Continued.



(c)

Figure 1: 3D data visualization for rich cities: (a) the effect among CPP-UD-CO, (b) the effect among CPP-UD-PCDI, and (c) the effect among CPP-CO-PCDI.



(c)

Figure 2: 3D data visualization for poor cities: (a) the effect among CPP-UD-CO, (b) the effect among CPP-UD-PCDI, and (c) the effect among CPP-CO-PCDI.

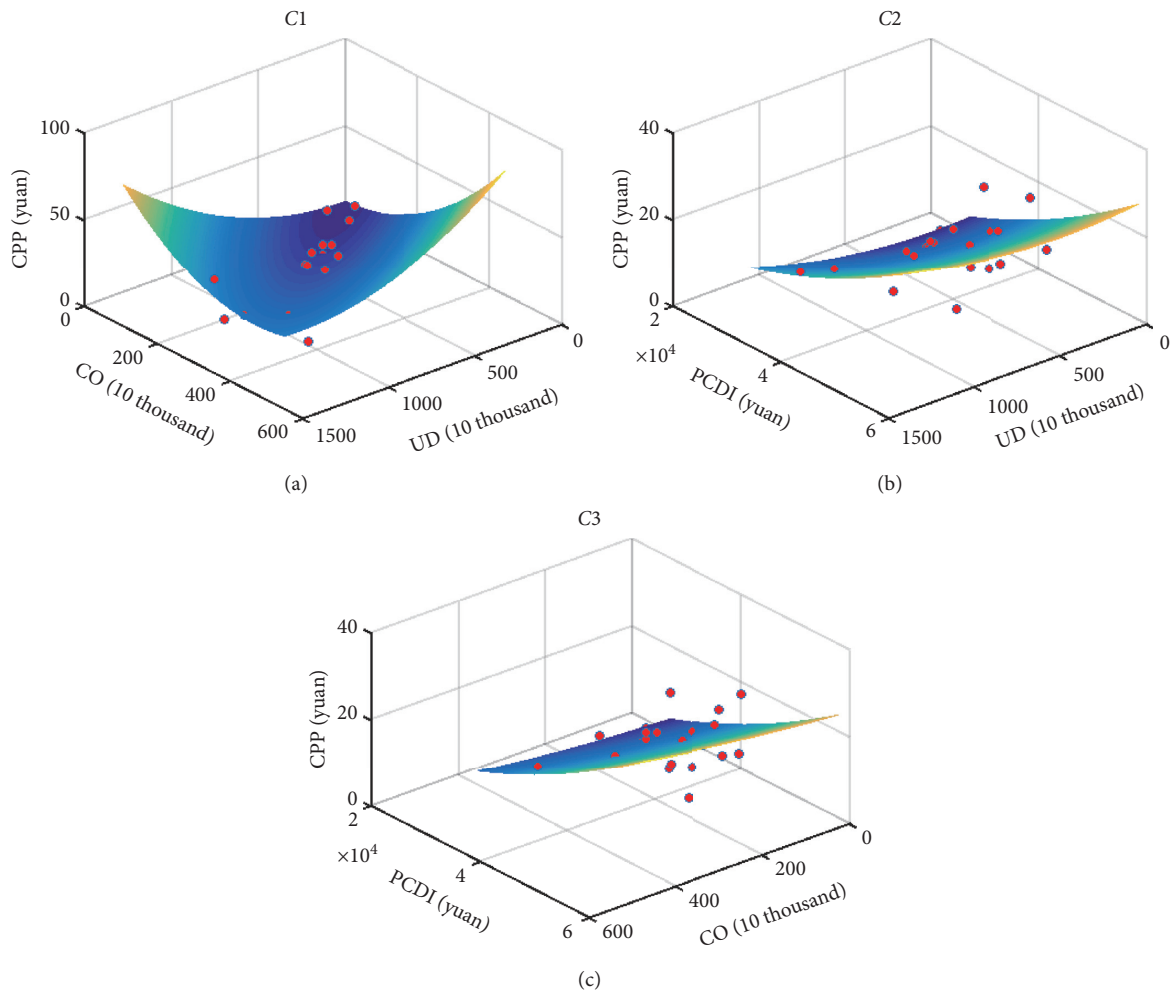


Figure 3: 3D data visualization for touristed cities: (a) the effect among CPP-UD-CO, (b) the effect among CPP-UD-PCDI, and (c) the effect among CPP- CO-PCDI.

this reason, the formula for exponential smoothing can be written as follows:

$$\begin{aligned} F_{t+1} &= \alpha X_t + (1-\alpha)F_t \\ &= \alpha X_t + F_t - \alpha F_t \\ &= F_t + \alpha(X_t - F_t). \end{aligned} \tag{5}$$

It can be seen that F_{t+1} is the sum of the predicted value F_t of period t and the prediction error T.

4.2.3. TSM-RAM Model. According to the above, the combined time series and regression analysis model should be as follows:

$$\begin{aligned} Y_{t+1} &= \sum_{t}^{n} a_n[F_{nt} + \alpha(X_{nt} - F_{nt})] + C \\ &= a_1[F_{1t} + \alpha(X_{1t} - F_{1t})] + a_2[F_{2t} + \alpha(X_{2t} - F_{2t})] \\ &\quad + a_3[F_{3t} + \alpha(X_{3t} - F_{3t})] \\ &\quad + a_4[F_{4t} + \alpha(X_{4t} - F_{4t})] + a_5[F_{5t} + \alpha(X_{5t} - F_{5t})] \\ &\quad + a_6[F_{6t} + \alpha(X_{6t} - F_{6t})] + C, \end{aligned} \tag{6}$$

where X_{nt} is the observed value of X_n in period t, F_{nt} is the predicted X_{nt} value in period t, α is the smoothing coefficient $(0 < \alpha < 1)$, and $n \in (1, 2, \ldots, 6)$.

In this paper, the TSM-RAM method is used to calculate the parking price for future years. Due to the higher data frequency of the time series for the independent variables, the time series method is used to calculate the change in the value of the independent variables in future years. At the same time, according to the 3D data visualization results, the relationship between the parking price and the three independent variables is a quadratic function, so the regression analysis method can be used to estimate the relationship between the independent and dependent variables.

4.3. Testing the Parking Price Model

4.3.1. Goodness-of-Fit Test. The R^2 is the coefficient used to assess the goodness of fit of the regression line with the observations, and its maximum value is 1. The larger the R^2 is, the better the fit. In contrast, the smaller the R^2 is, the worse the fit.

Table 4: Goodness-of-fit test results.

City type	R	R^2	Adjusted R^2	Standard error
RC	0.9360	0.8762	0.8505	4.4401
PC	0.9062	0.8212	0.7647	1.1696
TC	0.8819	0.7777	0.7221	4.2778

Table 5: Variance analysis.

City type	F	Significance F
RC	34.1967	0.0000
PC	14.5449	0.0000
TC	13.9906	0.0000

Table 6: Regression coefficient test.

City type	Coefficient	Value	T	Significance
RC	a_1	$4.2979*10^{-5}$	3.3815	0.0021
	a_2	$9.2284*10^{-4}$	2.2091	0.0352
	a_3	$8.8863*10^{-9}$	3.5412	0.0014
PC	a_1	$1.1069*10^{-4}$	3.9411	0.0009
	a_2	$1.7739*10^{-3}$	3.5259	0.0023
	a_3	$6.2943*10^{-9}$	3.1700	0.0050
TC	a_1	$6.9089*10^{-5}$	2.4022	0.0244
	a_2	$6.6710*10^{-4}$	2.1513	0.0417
	a_3	$7.4237*10^{-9}$	2.5686	0.0169

According to the goodness-of-fit test results, the R^2s of the parking price model for RCs, PCs, and TCs are 0.8762, 0.9050, and 0.7777, respectively. This means that the accuracy of UD (X_1), CO (X_2), and PCDI (X_3) in predicting the 2 h CPP in the RCs, PCs, and TCs is 88.14%, 90.50%, and 77.53%, respectively, which indicates that the variables selected in this model are accurate and appropriate, as shown in Table 4.

4.3.2. F Test. An F test is used to test whether the variance of the two samples is significantly different. According to the results of a variance analysis after model fitting, the F values of the three types of cities are 34.1967, 14.5449, and 13.9996, as shown in Table 5. The corresponding significance levels are 0.0000, 0.0000, and 0.0000 ($P < 0.05$). This indicates that the ternary quadratic nonlinear model in equation (1) is appropriate. It also proves that there is a significant relationship between the 2 h CPP in the RCs, PCs, and TCs and the three explanatory variables.

4.3.3. T-Test. The T-test evaluates the significance of the relationship between the three variables, UD, CO, and PCDI, and the 2 h CPP. According to the regression results, the significance ofX_1, X_2, and X_3 in the RCs is 0.0017, 0.0327, and 0.0013, respectively. All of these values are less than the critical value of 0.05, which indicates that the impact of these three variables on the 2 h CPP in the core area of developed cities is significant. Similarly, the significance levels ofX_1, T, and X_3 in the PCs and TCs are less than the critical value of 0.05, as shown in Table 6. Thus, the impact of the three independent variables selected in this paper on the 2 h CPP in the core area in PCs and TCs is also significant.

5. Application

5.1. Case Study. Ningbo is a subprovincial city with municipalities that have an independent planning status under national social and economic development. It is also the economic center of the Yangtze River Delta and of Zhejiang Province. Since 2000, the city's economy has undergone sustained and rapid growth. The living standards of residents have increased substantially, and the amount of car ownership has also increased year by year. By 2018, it had increased 28-fold in 18 years, resulting in increasing parking pressure. The government is now focusing on how to manage parking demand through parking fees. We take Ningbo as a representative RC and collect data on UD, CO, and PCDI in Ningbo to conduct an empirical analysis of the period from 2000 to 2018.

Yancheng is located in the central part of China's eastern coast, in the central and eastern part of Jiangsu Province in the north wing of the Yangtze River Delta. It is the largest prefecture-level city in Jiangsu Province, with a city area of 17,000 square kilometers. The city is flat and resource-rich. The rivers run north-south and east-west. Constrained by

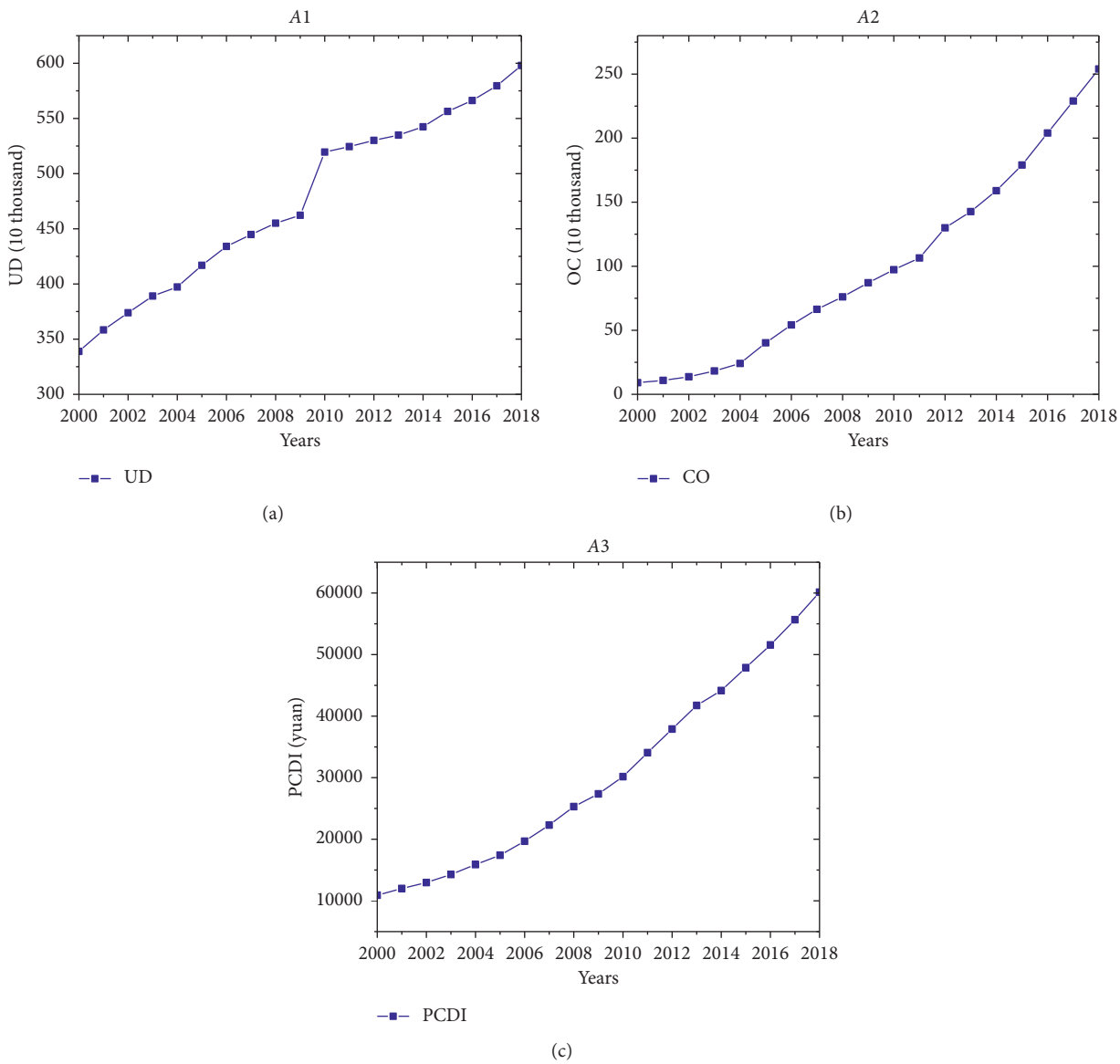


Figure 4: Changes in the three data series for Ningbo from 2000 to 2018: (a) changes in the UD historical data, (b) changes in the CO historical data, and (c) changes in the PCDI historical data.

historical conditions, infrastructure, production factors, etc., Yancheng has had a low level of regional economic development for a long time. The problem of high input and low output is obvious. Its car ownership figures fall in the middle to lower levels of the distribution for Jiangsu Province overall, and the development of its parking fees system lags behind that of other provinces. We take Yancheng as a representative PC and collect UD, CO, and PCDI data in the city to conduct an empirical analysis of the period from 2000 to 2018.

Kunming is the capital of Yunnan Province. It is located in the southwestern part of China. It is warm year-round and also known as the "Spring City." Its booming tourism industry has brought enormous opportunities to Kunming. The disposable income and car ownership of urban residents in Kunming can be compared with those of some developed cities. Therefore, parking problems have gradually become an urgent issue in Kunming. We take Kunming as a representative TC and collect UD, CO, and PCDI data for the city to conduct an empirical analysis of the period from 2000 to 2018.

5.2. Data Processing. We analyze the three data series for each city, as shown in Figures 4–6.

The three data series for Ningbo show an increasing trend over time. Among them, the increasing trends in CO (Figure 4(b)) and PCDI (Figure 4(c)) are relatively stable; the increasing trend of UD (Figure 4(a)) is stronger in 2009-2010, resulting from a surge in the migrant population [41]. Overall, the three data series show common characteristics among them.

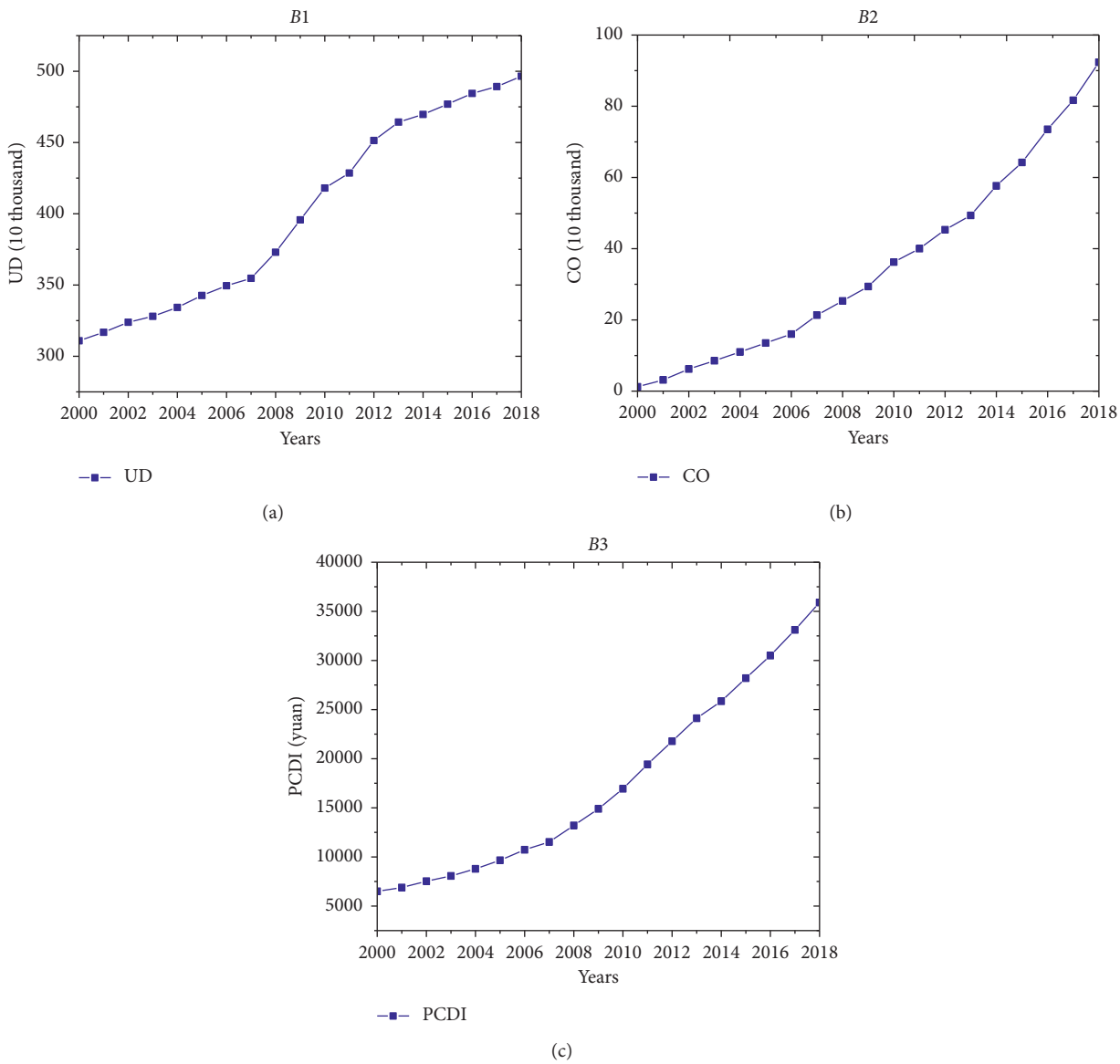


Figure 5: Changes in the three data series for Yancheng from 2000 to 2018: (a) changes in the UD historical data, (b) changes in the CO historical data, and (c) changes in the PCDI historical data.

Yancheng, which has the largest area of a prefecture-level city in Jiangsu Province, China, has a large number of permanent residents, and the proportion of urbanization is relatively low. The growth rate of UD in Yancheng city from 2007 to 2013, which appears in Figure 5(a), is due to the acceleration of urbanization in the city [42]. Correspondingly, the levels of CO (Figure 5(b)) and PCDI (Figure 5(c)) also steadily increase each year. Overall, the UD, CO, and PCDI data series for Yancheng city show common characteristics among them.

The development of tourism has brought enormous economic benefits to Kunming, and its urbanization level exceeded 60% in 2008, while the number of permanent urban residents (Figure 6(a)) reached 3 million in 2000 and has maintained a rapid growth trend. Similarly, the improvement of the city's economic level is also reflected in the rapid increase in CO (Figure 6(b)) and PCDI (Figure 6(c)). Overall, the UD, CO, and PCDI data series for Kunming show common characteristics among them.

5.3. Parameter Prediction

5.3.1. Time Series Prediction. The prediction results are divided into optimistic, aggressive, and conservative scenarios. The TSM is used to predict the changes in UD, CO, and PCDI in Ningbo, Yancheng, and Kunming from 2019 to 2021. The statistical results are shown in Table 7, and the overall changes are shown in Figures 7–9. At the same time, the long-term evolution (for 2025 and 2030) of the three data series is also estimated (see Tables 7–9).

The results of the optimistic, aggressive, and conservative estimations of the trends in UD, CO, and PCDI are ordered as follows: aggressive > optimistic > conservative. The changing characteristics of the data confirm the rigor and distinctiveness

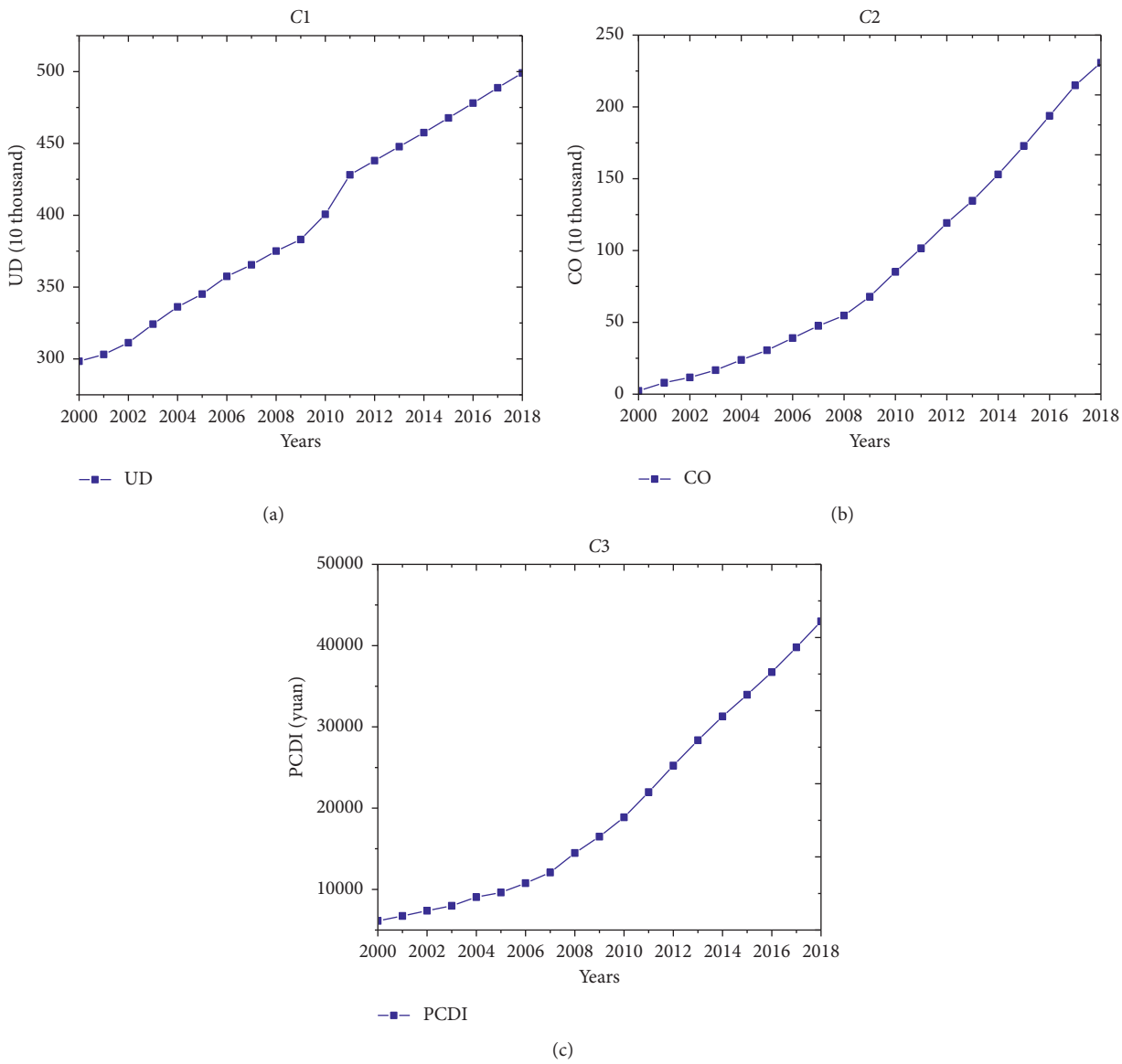


Figure 6: Changes in the three data series for Kunming from 2000 to 2018: (a) changes in the UD historical data, (b) changes in the CO historical data, and (c) changes in the PCDI historical data.

Table 7: Time series prediction for 2019–2021 (optimistic).

Cities	Term	Year	UD (10 thousand)	CO (10 thousand)	PCDI (yuan)
NB	Short-term	2019	611.19	278.87	64455
		2020	625.51	303.75	68756
		2021	639.82	328.63	73097
	Long-term	2025	698.95	428.17	90958
		2030	770.60	552.59	112977
YC	Short-term	2019	503.22	101.92	38818
		2020	509.98	113.06	41880
		2021	516.73	123.84	45083
	Long-term	2025	543.77	172.87	59303
		2030	577.56	244.09	80241
KM	Short-term	2019	510.17	246.50	46341
		2020	521.31	262.25	49846
		2021	532.46	278.00	53504
	Long-term	2025	577.05	341.00	69661
		2030	632.79	419.75	93290

Note. NB denotes Ningbo city, YC denotes Yancheng city, and KM denotes Kunming city.

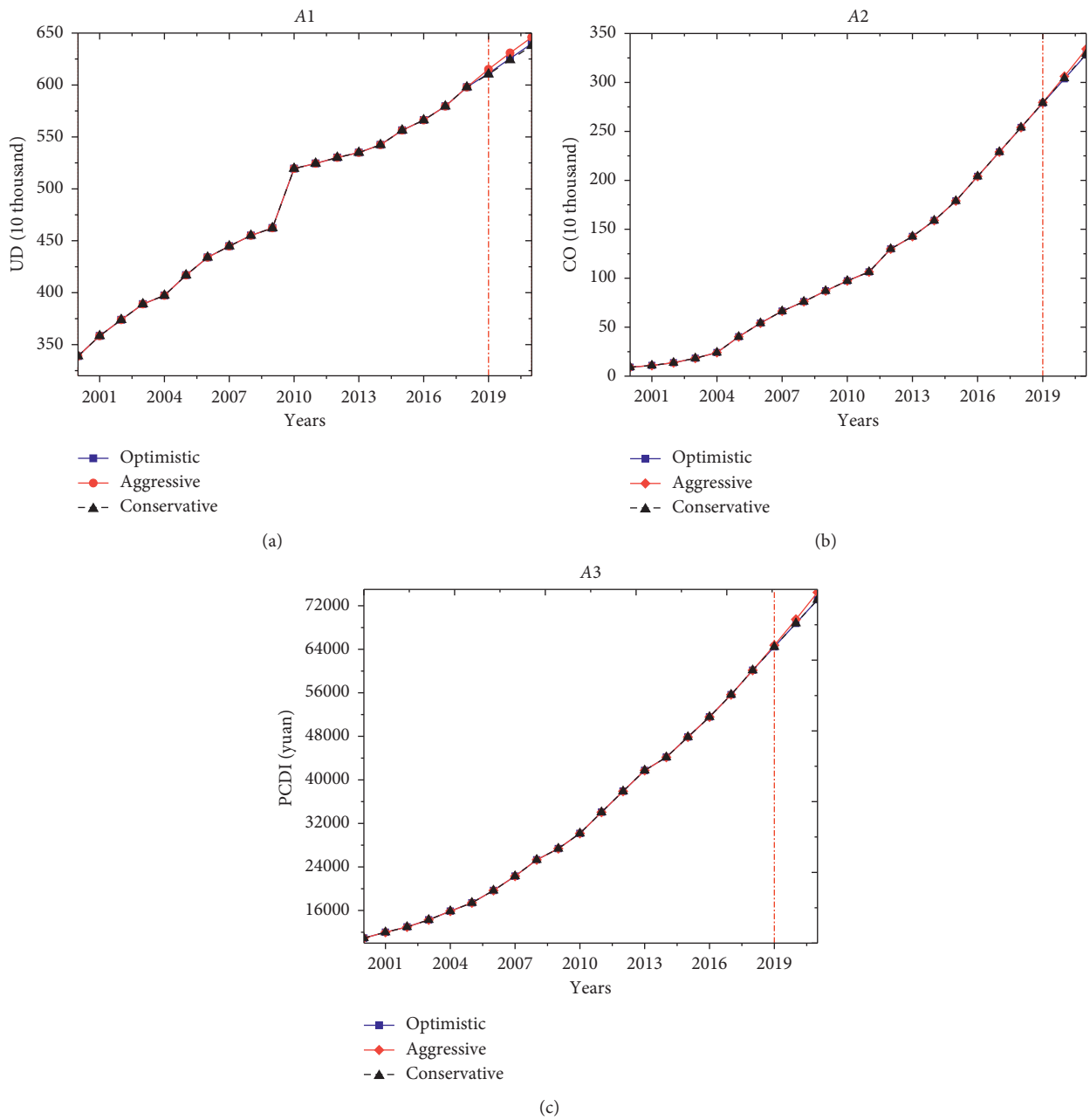


Figure 7: Future changes in the three data series for Ningbo (2019–2021): (a) future changes in UD, (b) future changes in CO, and (c) future changes in PCDI.

of the TSM. The three cities of Ningbo, Yancheng, and Kunming are ordered, in terms of overall economic strength, as follows: Ningbo > Kunming > Yancheng. Moreover, the levels of CO and PCDI in future years in Table 7 are consistent with this ranking. For example, the predicted PCDI for 2030 is 112,977 yuan for Ningbo, 93,290 yuan for Kunming, and 80,241 yuan for Yancheng city. These projections are also consistent with the results of the aggressive and conservative predictions in Tables 8-9.

5.3.2. Error Test. The prediction error refers to the difference between the prediction result and the real result of the development of the predicted variable and is divided into the predicted relative error and the predicted absolute error. The absolute error is the absolute difference between the predicted value and the actual observed value, and the relative error is the percentage difference relative to the observed value. Here, the absolute error is selected to characterize the error of the three data series predicted by the TSM.

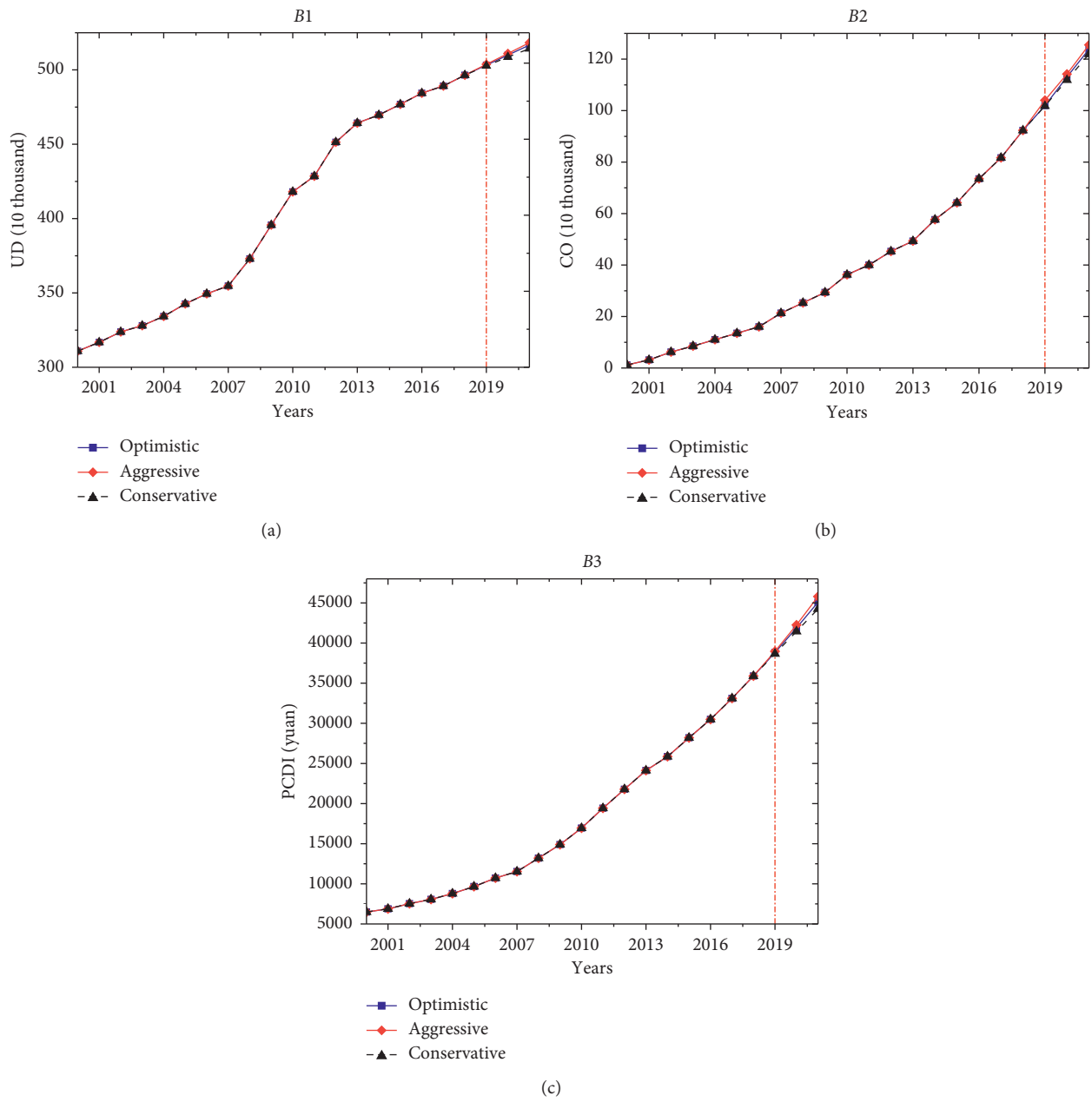


Figure 8: Future changes in the three data series for Yancheng (2019–2021): (a) future changes in UD, (b) future changes in CO, and (c) future changes in PCDI.

Taking 2017 and 2018 as examples, the actual values for the UD, CO, and PCDI series of Ningbo, Yancheng and Kunming are shown in Tables 10–12.

It can be seen from the Tables 11–13 that the three data series estimated by the TSM from 2017 to 2018 for the cities of Ningbo, Yancheng, and Kunming have good results: the maximum error is 3.76%, the minimum error is 0.005%, and the absolute error is below 0.05 (5%), which is within the acceptable range. At the same time, the average absolute error values of the optimistic, aggressive, and conservative scenarios are 0.532%, 1.075%, and 1.080%, respectively—all less than 0.05 (5%).

5.3.3. Goodness-of-Fit Test. According to the test results, the goodness of fit of the TSM for the optimistic, aggressive, and

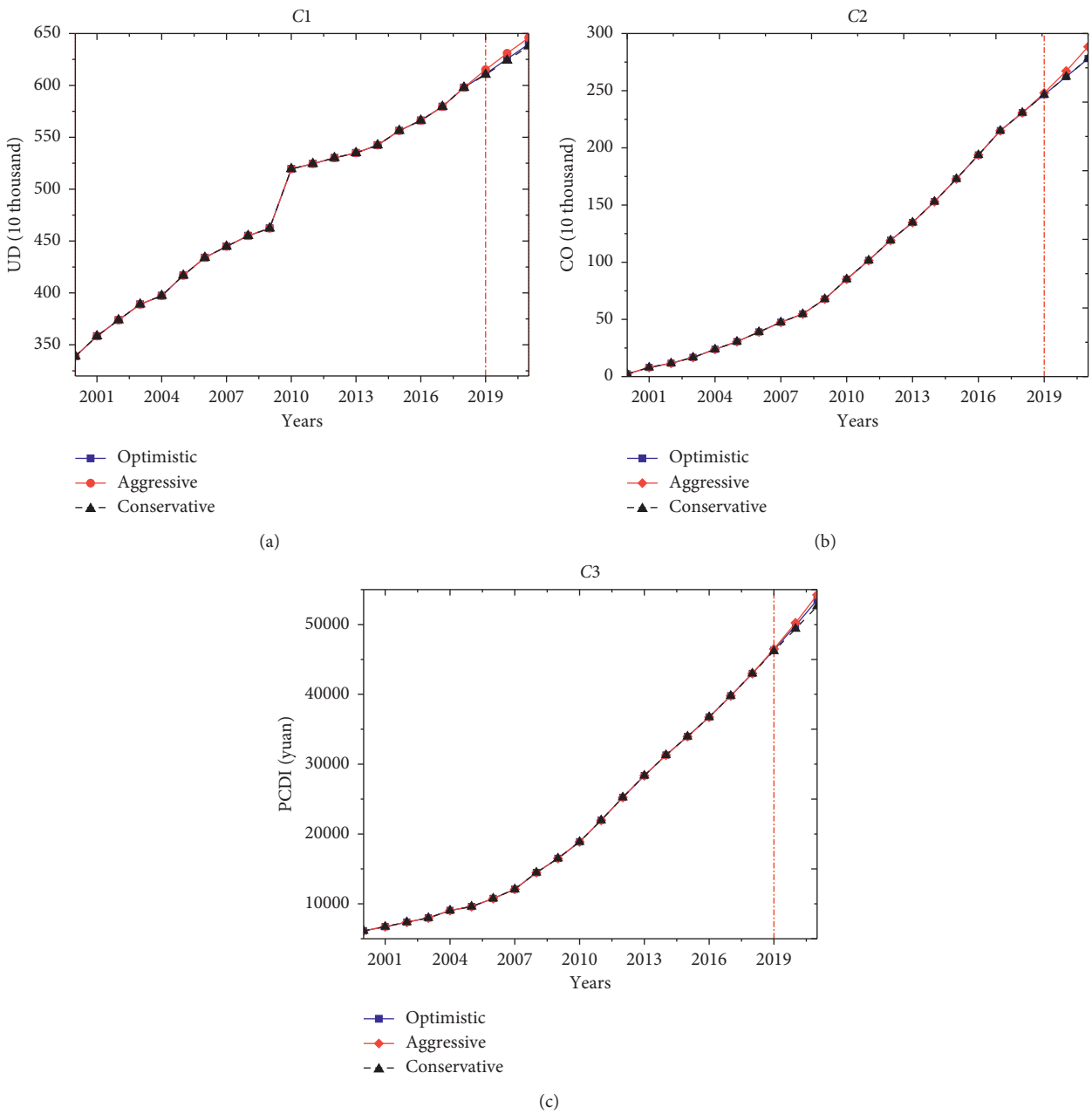


Figure 9: Future changes in the three data series for Kunming (2019–2021): (a) future changes in UD, (b) future changes in CO, and (c) future changes in PCDI.

conservative scenarios for the three cities is above 0.9 in all cases: the maximum is 0.999 and the minimum is 0.939, which indicates that the data estimates obtained by the TSM are highly reliable. In terms of the level of goodness of fit, the conservative prediction is the highest, the optimistic prediction falls in the middle, and the aggressive prediction is the lowest, which is also consistent with the characteristics of the results shown in Table 13.

6. Results

The UD, CO, and PCDI data series for future years predicted by the three TSM models are substituted into the corresponding ternary quadratic equation presented in the first part of this paper. The 2 h CPP in the core areas of NB, YC, and KM for the short term (2019–2021) and long term (2025, 2030) appear in Table 14.

Table 8: Time series prediction for 2019–2021 (aggressive).

Cities	Term	Year	UD (10 thousand)	CO (10 thousand)	PCDI (yuan)
NB	Short-term	2019	615.13	279.51	64730
		2020	630.91	306.19	69500
		2021	645.97	334.11	74463
	Long-term	2025	704.09	458.23	96252
		2030	775.97	641.38	127848
YC	Short-term	2019	503.80	104.02	38967
		2020	511.02	114.21	42252
		2021	518.33	125.53	45800
	Long-term	2025	548.76	183.55	62991
		2030	590.91	278.32	92705
KM	Short-term	2019	510.51	247.98	46496
		2020	525.06	267.18	50237
		2021	538.36	288.43	54244
	Long-term	2025	592.17	389.33	71951
		2030	661.89	535.62	97913

Note. NB denotes Ningbo city, YC denotes Yancheng city, and KM denotes Kunming city.

Table 9: Time series prediction for 2019–2021 (conservative).

Cities	Term	Year	UD (10 thousand)	CO (10 thousand)	PCDI (yuan)
NB	Short-term	2019	610.32	279.12	64454
		2020	624.05	304.24	68773
		2021	637.77	329.33	73093
	Long-term	2025	692.50	429.43	90369
		2030	760.62	554.00	111960
YC	Short-term	2019	502.75	101.80	38664
		2020	508.67	111.86	41454
		2021	514.29	121.93	44243
	Long-term	2025	534.01	162.18	55398
		2030	553.45	212.48	69339
KM	Short-term	2019	509.42	246.49	46179
		2020	519.80	262.22	49389
		2021	530.19	277.93	52600
	Long-term	2025	571.75	340.68	65442
		2030	623.69	418.85	81495

Note. NB denotes Ningbo city, YC denotes Yancheng city, and KM denotes Kunming city.

Table 10: Comparison of the predicted and actual UD values (unit (10 thousand)).

Cities	Type	Year	Predicted value	Actual value	Error (%)
Ningbo	Optimistic	2017	583.11	579.56	0.613
		2018	595.29	597.93	0.442
	Aggressive	2017	589.24	579.56	1.670
		2018	601.22	597.93	0.550
	Conservative	2017	576.21	579.56	0.578
		2018	596.33	597.93	0.268
Yancheng	Optimistic	2017	491.88	489.19	0.550
		2018	494.80	496.50	0.342
	Aggressive	2017	496.44	489.19	1.482
		2018	504.23	496.5	1.557
	Conservative	2017	484.17	489.19	1.026
		2018	493.22	496.5	0.661
Kunming	Optimistic	2017	489.17	488.72	0.092
		2018	499.87	499.02	0.170
	Aggressive	2017	496.28	488.72	1.547
		2018	503.27	499.02	0.852
	Conservative	2017	484.32	488.72	0.900
		2018	498.21	499.02	0.162

Table 11: Comparison of the predicted and actual CO values (unit (10 thousand)).

Cities	Type	Year	Predicted value	Actual value	Error (%)
Ningbo	Optimistic	2017	226.82	229.00	0.010
		2018	253.43	254.00	0.002
	Aggressive	2017	232.41	229	1.489
		2018	259.34	254	2.102
	Conservative	2017	227.42	229	0.690
		2018	253.84	254	0.063
Yancheng	Optimistic	2017	81.57	81.67	0.122
		2018	91.44	92.33	0.964
	Aggressive	2017	82.73	81.67	1.298
		2018	92.09	92.33	0.26
	Conservative	2017	80.11	81.67	1.910
		2018	92.07	92.33	0.282
Kunming	Optimistic	2017	214.67	215.00	0.153
		2018	236.23	230.75	2.374
	Aggressive	2017	215.06	215	0.028
		2018	235.32	230.75	2.145
	Conservative	2017	213.21	215	0.833
		2018	228.07	230.75	1.161

Table 12: Comparison of the predicted and actual PCDI values (unit (yuan)).

Cities	Type	Year	Predicted value	Actual value	Error (%)
Ningbo	Optimistic	2017	55178	55656	0.860
		2018	59609	60134	0.875
	Aggressive	2017	55785	55656	0.233
		2018	60824	60134	1.147
	Conservative	2017	55024	55656	1.136
		2018	58145	60134	3.308
Yancheng	Optimistic	2017	32933	33115	0.550
		2018	35875	35896	0.059
	Aggressive	2017	33728	33115	1.851
		2018	36023	35896	0.354
	Conservative	2017	32909	33115	0.622
		2018	35372	35896	1.460
Kunming	Optimistic	2017	39677	39788	0.279
		2018	42990	42988	0.005
	Aggressive	2017	39924	39788	0.342
		2018	43178	42988	0.442
	Conservative	2017	39542	39788	0.618
		2018	41372	42988	3.760

Table 13: Testing the goodness of fit of the time series prediction results.

Types	Cities	UD	CO	PCDI
Optimistic	Ningbo	0.980	0.992	0.994
	Yancheng	0.955	0.997	0.999
	Kunming	0.969	0.999	0.999
Aggressive	Ningbo	0.943	0.962	0.971
	Yancheng	0.951	0.983	0.986
	Kunming	0.958	0.939	0.967
Conservative	Ningbo	0.977	0.985	0.994
	Yancheng	0.970	0.994	0.989
	Kunming	0.973	0.981	0.992

Table 14: CPP prediction results.

Cities	Term	Year	CPP (yuan)		
			Optimistic	Aggressive	Conservative
Ningbo (NB)	Short-term	2019	25.89	26.04	25.92
		2020	30.18	30.74	30.26
		2021	35.06	36.38	35.19
	Long-term	2025	60.56	70.65	60.79
		2030	105.23	149.64	105.86
Yancheng (YC)	Short-term	2019	12.05	11.86	11.96
		2020	12.52	12.63	12.38
		2021	13.42	13.73	13.05
	Long-term	2025	21.07	23.98	18.32
		2030	43.97	62.15	31.12
Kunming (KM)	Short-term	2019	21.68	21.99	21.58
		2020	26.36	27.3	26.05
		2021	31.63	33.89	30.94
	Long-term	2025	59.18	74.44	54.65
		2030	110.67	164.02	93.64

Table 15: Results for the confidence intervals of three types of data in Ningbo (optimistic).

Year	Confidence interval	UD (10 thousand)	CO (10 thousand)	PCDI (yuan)
2019	Highest	635.12	289.24	65827.4
	Lowest	587.26	268.49	63082.2
2020	Highest	653.41	323.71	71483.7
	Lowest	597.60	283.79	66067.6
2021	Highest	671.20	359.89	77362.5
	Lowest	608.44	297.38	68830.5

Table 16: The 2 h CPP range in the core area of Ningbo city in future years (optimistic).

Year	CPP range (yuan)
2019	(24.70, 27.11)
2020	(27.36, 33.26)
2021	(29.93, 40.93)

7. Discussion and Conclusion

7.1. Discussion. According to the *Goldilocks principle* of [43], parking demand and supply are best balanced by setting the parking price appropriately. At present, the CPP in different cities in China is not the same, but it remains essential to balance demand and supply through price-setting [44]. On the basis of the research in this paper, parking pricing has a very high correlation with UD, CO, and PCDI. Therefore, the predicted development of the three data series can reflect the CPP in the core area of a city. This fact indicates that the research results in this paper have a certain predictive power.

This paper divides domestic cities into RCs, PCs, and TCs according to the differences in their parking fee systems and level of economic development. The 2 h CPP data for almost all urban core areas, as well as the UD, CO, and PCDI data for most prefecture-level cities and above, were collected to fit the models. The final calculation results also reveal future changes in the 2 h CPP of the urban core areas of Ningbo, Yancheng, and Kunming. Likewise, it would be possible to collect historical UD, CO, and PCDI data and then calculate the future CPP for other cities to which the model is fitted. For cities for which these data have not yet been collected, the model can be refitted to ensure that the ternary quadratic regression function is in accordance with the actual situation of the city, and then the CPP can be predicted after the parameters are modified.

As shown in Table 14, the CPP of the three cities of Ningbo, Yancheng, and Kunming shows a rapid growth trend predicted for the coming years. This is explained by the increases in UD, CO, and PCDI, which are also an inevitable result of accelerating urbanization. Among these variables, the increase in CO determines that the parking problem will continue to be an area of focus for policymakers in future years. Considering the combination of this increase and the current curb parking pricing policy adopted by most cities, the main problems are as follows:

(1) Parking demand that has not been effectively regulated [45]: the key principle for regulating parking demand is that off-street parking should be preferred over on-street parking; among the different types of off-street parking, indoor parking should be preferred as much as possible over open-air parking [46]. At present, most cities implement a time-based CPP, divided into spells of less than 2 h and more than 2 h. Generally, managers force parked cars to transfer to off-road parking after 2 h spells, so the

unit parking fee for spells beyond 2 h is higher than that for spells under 2 h. The CPPs in the core areas of Ningbo, Yancheng, and Kunming are low. Although the parking price of Ningbo is only 3 yuan, it is higher than that of the other two cities. This charging mechanism can neither effectively highlight the differences between on-street and off-street parking nor regulate the transfer of parked vehicles from on-street to off-street parking spaces.

(2) Pricing to encourage long-term parking [47]: the standard daytime CPP for parking spaces in 2017 is free parking for spells of less than 30 min and 2 yuan per 30 mins thereafter in Wanda Square, Ningbo, and the daytime parking fee per unit of time up to 12 h does not change. This price level is clearly within the tolerance of most urban residents because a car user pays only 51 yuan for 24 h of parking. Before the implementation of the intelligent parking project, the long-term parking utilization in this area was close to 50%. There is a similar case in Ningbo city's Tianyi Square, where the current CPP standard is free parking for spells of up to 15 min and 3 yuan per 20 mins within 2 h; this setup does not incentivize car users to transfer the vehicle to a car park because the fees for the first and second hours are the same. This situation is also common in Yancheng city, where the CPP is 1.5 yuan per 15 mins in the first-level area. In addition, the CPP in some cities decreases as the parking time increases; this is the case in Chengdu, which charges 10 yuan for the first hour of parking in the core area and 6 yuan for each hour thereafter up to a certain limit, which encourages long-term on-street parking and restricts the development of high churn in on-street parking.

The CPP in the core area of Kunming is predicted to be 5.44 yuan higher than that of Ningbo in 2035, as seen in Table 12. In addition to the differences caused by the goodness of fit of the models for the RCs and TCs, Kunming's overall economic strength may be higher than that of Ningbo in future years.

In the past, when research has evaluated the current parking pricing problem, the final goal was always to obtain the *optimal parking price* [48–53]. We argue that the solution to the parking pricing problem should also consider the optimal parking price range, which spans the maximum value of the CPP that is acceptable to travelers and the minimum value of the CPP that is acceptable to decision makers. Similarly, the CPP problem is also likely to remain an issue in future years, so our findings can also be considered to obtain the CPP range for future years.

In this study, the confidence interval is set in the model fitting and variable prediction process (the confidence level is 95%). When the TSM is used to evaluate trends in UD, CO, and PCDI, the output contains the results corresponding to the highest and lowest confidence levels for the three data series over future years. Examples of the predictions for the three data series in the optimistic TSM scenario for Ningbo city are shown in Table 15.

By substituting the data in the above table into the nonlinear regression in sequence, the corresponding parking pricing fee interval for future years can be obtained, as shown in Table 16.

7.2. Conclusion. At present, the issue of imbalance between parking supply and demand for urban development is still a major challenge. It is of great practical importance to accurately determine the CPP for future years so as to quickly address the imbalance between parking supply and demand and provide theoretical support to decision makers.

The present study applied a TSM-RAM model to predict the CPP and solve the traffic problem caused by the imbalance between parking supply and demand. The data were obtained through the China Statistical Yearbook. The results showed the effectiveness of the TSM-RAM model for making parking price forecasts. At the same time, we paid special attention to dividing the results into optimistic, aggressive, and conservative estimates when applying the TSM to data series. In addition, the prediction of the curb parking price (CPP) was also based on the level of urbanization, with Chinese cities divided into RCs, PCs, and TCs, and case studies of Ningbo, Yancheng, and Kunming, which were selected as representative cities corresponding to each category. The diversity of the results also provides extra information to help policymakers respond to future parking problems.

The conclusion of this article can be attributed to the following three parts. Firstly, in terms of data, we found that the goodness-of-fit test results of the curb parking prices (CPP) and the number of urban dwellers (UD), car ownership (CO), and per capita disposable income of urban residents (PCDI) are all above 0.9, indicating that the selected very high correlation between independent and dependent variables. Secondly, in terms of models and methods, we found that time series methods are used to predict the number of urban dwellers (UD), car ownership (CO), and per capita disposable income of urban residents (PCDI) results have extremely low errors, all of which are below 0.05. Combining the three types of data with a very high degree of fit for on-street parking prices can prove that the TSM-RAM method is suitable for the CPP prediction. Finally, in terms of policies, we recommend regulating parking demand that off-street parking is encouraged between on-street parking and off-street parking; on the other hand, indoor parking is encouraged as much as possible between open-air parking and indoor parking for off-street parking and then determining prices that encourage short-term parking.

However, there are some limitations to this study. First, the model in this paper only fits data for 36 RCs, 26 PCs, and 31 TCs. If sample data for more cities are properly added, the goodness of fit of the model could be improved. Second, this paper selects only three variables related to the CPP, namely, UD, CO, and PCDI. After researchers solve the difficult problems of data collection and prediction, factors such as road congestion can be added to the initial data to better improve the model fit and the accuracy of the model.

Furthermore, extensions of this work should examine the categories of urban areas. This paper divides cities in China into RCs, PCs, and TCs according to their economic level, but it may also be a good choice to categorize cities according to their administrative level. This work has provided the framework for a TSM-RAM predictive curb parking price model. Under this framework, the curb parking price as affected by UD, CO, PCDI, or various other factors can be estimated to address parking problems in central urban districts.

In large cities, curb parking pricing (CPP) policies must differentiate parking charges by region. This article focuses on the selection of influencing factors of curb parking prices in core areas. Therefore, the article does not consider whether differentiated parking fees by region has an impact on curb parking pricing in a single region. This issue will be studied in detail in the next phase of the study by considering the impact of differentiated parking pricing policies on individual curb parking price.

In addition, the TSM-RAM method proposed in this paper has certain errors, but the results of the goodness-of-fit test, T-test, F test, and error test are adequate, indicating that the error of the prediction result is within a reasonable range. Although a panel data model may have obvious advantages in recognizing measurement time series and cross-sectional data, describing individual behavioral differences, and constructing more complex behavioral models, there are also shortcomings related to the short time sequence and difficulties of variable design and data collection. Therefore, in the next step, we will conduct a panel data model study based on parking price research.

Data Availability

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

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the Natural Science Foundation of Zhejiang Province (No. LQ19E080003), Philosophy and Social Science Program of Ningbo (G20-ZX07 and G20-ZX37), the Natural Science Foundation of Ningbo (No. 2018A610127), the National Natural Science Foundation of China (No. 71861023), the Program of Humanities and Social Science of Education Ministry of China (No. 18YJC630118), and the Foundation of a Hundred Youth Talents Training Program of the Lanzhou Jiaotong University.

References

[1] T. Lin, G.-L. Lyu, F. Tian, and Y. Lu, "A pilot study of on-street parking charge in Shenzhen," *Urban Transport of China*, vol. 14, no. 4, pp. 30–39, 2016.

[2] China Statistical Bureau, *Statistical Communique of National Economic and Social Development of China City*, China Statistical Bureau, Beijing, China, 2000–2018, http://www.stats.gov.cn/tjsj/ndsj/.

[3] D. C. Shoup, "The ideal source of local public revenue," *Regional Science and Urban Economics*, vol. 34, no. 6, pp. 753–784, 2004.

[4] Y. Guo, A. Osama, and T. Sayed, "A cross-comparison of different techniques for modeling macro-level cyclist crashes," *Accident Analysis & Prevention*, vol. 113, pp. 38–46, 2018.

[5] P. A. Barter, "Off-street parking policy surprises in Asian cities," *Cities*, vol. 29, no. 1, pp. 23–31, 2012.

[6] Z.-S. Yang and X.-D. Chen, "Research on the estimation for effective parking space of the intelligentized parking guidance system," *Journal of Transportation Systems Engineering and Information Technology*, vol. 3, no. 4, pp. 12–15, 2003.

[7] A. Mondschein, D. A. King, C. Hoehne, Z. Jiang, and M. Chester, "Using social media to evaluate associations between parking supply and parking sentiment," *Transportation Research Interdisciplinary Perspectives*, vol. 4, pp. 100085–100088, 2020.

[8] T. Rye, T. Cowan, and S. Ison, "Expansion of a controlled parking zone (CPZ) and its influence on modal split: the case of Edinburgh," *Transportation Planning and Technology*, vol. 29, no. 1, pp. 75–89, 2006.

[9] F. Kirschner and M. Lanzendorf, "Parking management for promoting sustainable transport in urban neighbourhoods. A review of existing policies and challenges from a German perspective," *Transport Reviews*, vol. 40, no. 1, pp. 54–75, 2020.

[10] Z. Zhao, Y. Zhang, and Y. Zhang, "A comparative study of parking occupancy prediction methods considering parking type and parking scale," *Journal of Advanced Transportation*, vol. 2020, no. 10, 12 pages, 2020.

[11] C. Ma, W. Hao, A. Wang, and H. Zhao, "Developing a coordinated signal control system for urban ring road under the vehicle-infrastructure connected environment," *IEEE Access*, vol. 6, pp. 52471–52478, 2018.

[12] W.-D. Middleton, "Transportation for livable cities," *Railway Age*, vol. 30, pp. 1–10, 2000.

[13] G. Marsden, "The evidence base for parking policies-a review," *Transport Policy*, vol. 13, no. 6, pp. 447–457, 2006.

[14] D. Shoup, "The price of parking on a great street," *Parking Today*, vol. 14, no. 2, pp. 22-23, 2006.

[15] T. Litman, "Pricing for traffic safety: how efficient transport pricing can reduce roadway crash risk," *Transportation Research Record Journal of the Transportation Research Board*, vol. 8, no. 2318, pp. 16–22, 2012.

[16] N. Terui and H. K. Van Dijk, "Combined forecasts from linear and nonlinear time series models," *International Journal of Forecasting*, vol. 18, no. 3, pp. 421–438, 2002.

[17] W.-G. Xiao and X.-M. Lan, "Research on the nonlinear mechanism of the influence of the RMB exchange rate expectation on real estate price in China," *Soft Science*, vol. 31, no. 12, pp. 129–133, 2017.

[18] H. Yin, Z.-G. Li, Y.-L. Wang, and F. Cai, "Assessment of desertification using time series analysis of hyper-temporal vegetation indicator in inner Mongolia," *Acta Geographica Sinica*, vol. 66, no. 5, pp. 653–661, 2011.

[19] X.-B. Huang, H.-B. Li, Y.-C. Zhu, Y.-X. Wang, X.-X. Zhang, and Y.-G. Wang, "Short-term forecast for transmission line icing by time series analysis and kalman filtering," *High Voltage Engineering*, vol. 43, no. 6, pp. 1943–1949, 2017.

[20] Z.-Y. Zhu, "Time series analysis of Chinese insurance compensation-modeling and forecasting," *Journal of Applied Statistics and Management*, vol. 29, no. 4, pp. 698–704, 2010.

[21] N. Rabindra and P. Nirash, "A multivariate time series analysis of energy consumption, real output and pollutant emissions in a developing economy: new evidence from Nepal," *Economic Modeling*, vol. 77, pp. 164–173, 2019.

[22] H. Reiser and H. Kutie, "Rainfall uncertainty in the Mediterranean: time series, uncertainty, and extreme events," *Theoretical and Applied Climatology*, vol. 104, no. 3-4, pp. 357–375, 2011.

[23] M. Ding, L.-J. Zhang, and Y.-C. Wu, "Wind speed forecast model for wind farms based on times series analysis," *Electric Power Automation Equipment*, vol. 25, no. 8, pp. 32–34, 2005.

[24] R. E. Abdelaal and A. Z. Algarni, "Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis," *Energy*, vol. 22, no. 11, pp. 1059–1069, 2014.

[25] J. Hu, J.-G. Lin, and Y.-Y. Zhao, "The influence of generalized linear model based on the pena distance," *Mathematica Applicata*, vol. 30, no. 3, pp. 539–546, 2017.

[26] Z.-H. Wang, G.-Y. Shan, and Y.-W. Lin, ""Multi-linear regression analysis using Microsoft excel and it's application in Chemistry," *Computers and Applied Chemistry*, vol. 22, no. 9, pp. 121–124, 2005.

[27] H. Sun, Z.-M. Zhang, and H.-J. Ge, "Application of PSO to improve multiple linear regression," *Computer Engineering and Applications*, vol. 43, no. 3, pp. 43-44, 2007.

[28] H. Zhou, C.-W. Jiang, W. Fang, M.-W. Liang, J.-W. Huang, and T.-Y. Hou, "Verification for performance of anti-dsDNA antibody quantitative ELISA kit," *Chinese Journal of Clinical Laboratory Science*, vol. 35, no. 7, pp. 533–537, 2017.

[29] B. Lin, "Multiple linear regression analysis and its application," *China Science and Technology Information*, vol. 9, no. 3969, pp. 60-61, 2010.

[30] M. Val, Y. Shih-Cheng, T. Arthur et al., "Regression analysis of gait parameters and mobility measures in a healthy cohort for subject-specific normative values," *PLoS One*, vol. 13, no. 6, Article ID e0199215, 2018.

[31] Q.-B. Yu, D. Su, Z.-J. Jiao, and C. Li, "Enhancement of near infrared spectroscopy determination of water contents in oils using oil-water stabilization and support vector regression," *Chinese Journal of Analytical Chemistry*, vol. 42, no. 9, pp. 1364–1368, 2014.

[32] H. Kaneko, M. Arakawa, and K. Funatsu, "Development of a new regression analysis method using independent component analysis," *Journal of Chemical Information and Modeling*, vol. 48, no. 3, pp. 534–541, 2008.

[33] D. Liu, J.-D. Zhai, and W.-G. Chen, "A regression-based predication model for tpc-c performance of high-end fault-tolerant computer," *Chinese Journal of Computers*, vol. 36, no. 6, pp. 1267–1279, 2013.

[34] F. Peng, D.-L. Zhou, M. Long, and X.-M. Sun, "Discrimination of natural images and computer generated graphics based on multi-fractal and regression analysis," *AEU—International Journal of Electronics and Communications*, vol. 71, pp. 72–81, 2017.

[35] Y. Ma, Q. Wang, Z. Chen et al., "Multiple linear regression analysis of X-ray measurement and WOMAC scores of knee osteoarthritis," *China Journal of Orthopaedics and Traumatology*, vol. 25, no. 2, pp. 373–376, 2012.

[36] T. Kondo, J. Ueno, and S. Takao, "Logistic GMDH-type neural network using principal component-regression analysis and its application to medical image diagnosis of lung cancer," *Artificial Life and Robotics*, vol. 20, no. 2, pp. 137–144, 2015.

[37] J. Qu and X. Chen, "Study of curb parking price in Chengdu," *Journal of Transportation Engineering and Information*, vol. 10, no. 1, pp. 111–114, 2012.

[38] J. A. Kelly and J. P. Clinch, "Influence of varied parking tariffs on parking occupancy levels by trip purpose," *Transport Policy*, vol. 13, no. 6, pp. 487–495, 2006.

[39] D. Albalate and A. Gragera, "The determinants of garage prices and their interaction with curbside regulation," *Transportation Research Part A: Policy and Practice*, vol. 101, pp. 86–97, 2017.

[40] M. B. W. Kobus, E. Gutiérrez-i-Puigarnau, P. Rietveld, and J. N. Van Ommeren, "The on-street parking premium and car drivers' choice between street and garage parking," *Regional Science and Urban Economics*, vol. 43, no. 2, pp. 395–403, 2013.

[41] X. Huang, "Remodeling and adjustment: study on county floating population identity in the process of new urbanization," *Journal of Ningbo Radio & Tv University*, vol. 15, no. 1, pp. 22–27, 2017.

[42] X. Mao, C. Yuan, J. Gan, and J. Zhou, "Optimal evacuation strategy for parking lots considering the dynamic background traffic flows," *International Journal of Environmental Research and Public Health*, vol. 16, no. 12, pp. 2194–2205, 2019.

[43] T. Rye, *The Implementation and Effectiveness of Transport Demand Management Measures: An International Perspective*, Rout Ledge: Taylor & Francis Group, Abingdon, UK, 2016.

[44] Y.-Y. Guo, P. Liu, Y. Wu, and J.-X. 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. 1–38, 2018.

[45] M.-C. He, J.-H. He, J.-L. Li, and Y. Li, "Curb parking pricing model based on time series and regression analysis," *Journal of Transportation Engineering and Information*, vol. 16, no. 1, pp. 79–84, 2018.

[46] C.-X. Ma, W. Hao, R.-C. He et al., "Distribution path robust optimization of electric vehicle with multiple distribution centers," *PLoS One*, vol. 13, no. 3, pp. 1–16, 2018.

[47] Z. Wang, "An important carrier for building urban business center," *Ningbo Communications*, vol. 18, no. 2, pp. 46-47, 2012.

[48] A. Filipovitch and E. Frimpong Boamah, "A systems model for achieving optimum parking efficiency on campus: the case of Minnesota State University," *Transport Policy*, vol. 45, pp. 86–98, 2016.

[49] Y. Guo, T. Sayed, and M. H. Zaki, "Evaluating the safety impacts of powered two wheelers on a shared roadway in China using automated video analysis," *Journal of Transportation Safety & Security*, vol. 11, no. 4, pp. 414–429, 2019.

[50] Y. Guo, Z. Li, P. Liu, and Y. Wu, "Modeling correlation and heterogeneity in crash rates by collision types using full Bayesian random parameters multivariate Tobit model," *Accident Analysis & Prevention*, vol. 128, pp. 164–174, 2019.

[51] Y. Guo, L. Zheng, and T. Sayed, "A comparison of collision-based and conflict-based safety evaluation of left-turn bay extension," *Transportmetrica A: Transport Science*, vol. 16, no. 3, pp. 676–694, 2020.

[52] Y.-Y. Guo, P. Liu, Y. Wu, and M.-L. Yang, "Traffic conflict model based on bayesian multivariate Poisson-lognormal normal distribution," *China Journal of Highway and Transport*, vol. 31, no. 1, pp. 101–109, 2018.

[53] G. Antolín, B. Alonso, and R. Cordera, "The effect of introducing parking policies on managing mobility to beaches in touristic coastal towns," *Sustainability*, vol. 11, no. 3528, pp. 1–13, 2019.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 2529856, 11 pages
https://doi.org/10.1155/2020/2529856

Research Article

A Novel On-Ramp Merging Strategy for Connected and Automated Vehicles Based on Game Theory

Haigen Min,[1,2] **Yukun Fang**,[1] **Runmin Wang**,[2] **Xiaochi Li**,[3] **Zhigang Xu**,[1,2] **and Xiangmo Zhao**[1,2]

[1]*School of Information & Engineering, Chang'an University, Xi'an 710064, China*
[2]*The Joint Laboratory for Internet of Vehicles, Ministry of Education, China Mobile Communications Corporation, Chang'an University, Xi'an 710064, China*
[3]*Henan Transport Investment Group Co., Ltd., Zhengzhou 450016, China*

Correspondence should be addressed to Haigen Min; hgmin@chd.edu.cn and Zhigang Xu; xuzhigang@chd.edu.cn

Received 3 February 2020; Revised 5 April 2020; Accepted 27 May 2020; Published 8 July 2020

Academic Editor: Eneko Osaba

Connected and automated vehicles (CAVs) have attracted much attention of researchers because of its potential to improve both transportation network efficiency and safety through control algorithms and reduce fuel consumption. However, vehicle merging at intersection is one of the main factors that lead to congestion and extra fuel consumption. In this paper, we focused on the scenario of on-ramp merging of CAVs, proposed a centralized approach based on game theory to control the process of on-ramp merging for all agents without any collisions, and optimized the overall fuel consumption and total travel time. For the framework of the game, benefit, loss, and rules are three basic components, and in our model, benefit is the priority of passing the merging point, represented via the merging sequence (MS), loss is the cost of fuel consumption and the total travel time, and the game rules are designed in accordance with traffic density, fairness, and wholeness. Each rule has a different degree of importance, and to get the optimal weight of each rule, we formulate the problem as a double-objective optimization problem and obtain the results by searching the feasible Pareto solutions. As to the assignment of merging sequence, we evaluate each competitor from three aspects by giving scores and multiplying the corresponding weight and the agent with the higher score gets comparatively smaller MS, i.e., the priority of passing the intersection. The simulations and comparisons are conducted to demonstrate the effectiveness of the proposed method. Moreover, the proposed method improved the fuel economy and saved the travel time.

1. Introduction

Congestion has caused many problems such as excessive fuel consumption and increased travel time in the real transportation system. According to a survey, the total cost of congestion in urban areas in the United States was estimated at 160 billion dollars and an extra 3.1 billion gallons of fuel consumed in 2014 [1]. Collaborative control of connected autonomous vehicles (CAVs) in a networked environment enables vehicles to cooperate with each other through information interaction, which can improve the road traffic efficiency and reduce the energy consumption while ensuring safety. To achieve the cooperative control of vehicles, advanced localization [2] and communication technologies [3, 4] (including V2V (vehicle to vehicle) and V2I (vehicle to infrastructure) communication, etc.), are required to assist the autonomous vehicle making decisions. Control of CAVs can be applied to many issues related to transportation optimization, and one of them is on-ramp merging of vehicles. In fact, on-ramp merging is one of the main causes of traffic congestion and the bottleneck of the traffic efficiency [5] since vehicles attempting to merge may initially slow down even and stop on the on-ramp to await a proper opportunity to merge. CAVs controlled by algorithms with high efficiency are expected to be able to reduce traffic congestion and enhance vehicle safety [6–9]. More key technologies and algorithms for CAVs control have been surveyed in [10].

The essence of the on-ramp merging problem is the competition for the priority of passing the merging area and reflected on the assignment of merging sequence (MS) for CAVs under the control of a centralized controller. We can regard such competition as a game in which every agent (i.e., the connected autonomous vehicle) competes for the prior merging sequence. Contribution of this paper just lies in the game framework we developed. In this paper, we propose a centralized approach based on game theory to control the on-ramp merging of all agents without any collisions and reduce the cumulative fuel consumption and total travel time. As to the game, we regard vehicles on the same road as a group and they will collaborate to compete with another group. Benefit, loss, and rules are three basic components in a game, and in our framework, benefit is the priority of passing the merging point and loss is the cost of fuel consumption and the travel time. Rules are the core for a game since agents in each group take actions based on the game rules to maximize the benefits and minimize the loss, and we proposed three basic rules in accordance with the Traffic Density Principle, FIFO (First In First Out) Principle, and Wholeness Principle, respectively. Importance of three rules varies, and we formulate the problem as a double-objective optimization problem of cumulative fuel consumption and total travel time, searching the feasible Pareto solutions to ascertain the weight of each rule.

Simulations and comparisons are conducted to validate the effectiveness of the proposed framework. Contribution of this paper mainly lies in the (1) construction of the framework for global optimal merging of CAVs based on the game theory and (2) approaches to searching the Pareto solutions and ascertaining the optimal weights of three rules via back search through formulating the problem as a double-objective optimization problem.

The structure of the paper is depicted as follows. In Section 2, we will introduce the related work. Section 3 describes the problem framework. Modeling and solution will be illustrated in Section 4, and simulation results and analysis will be displayed in Section 5. We end the manuscript with conclusions in Section 6.

2. Related Work

Current researches of on-ramp merging mainly focus on the scheduling algorithm of passing vehicles per unit time and control of merging at the expressway to improve the traffic efficiency [11–15]. A series of related control algorithms, control strategies, and scheduling algorithms are proposed under security principles. For example, in 2004, Dresner and Stone [16] proposed an automatic intersection control method based on retention algorithm. Generally, research work of CAVs on-ramp merging can be divided into centralized methods and decentralized methods [17], and both of these two categories have been studied [18–21]. Ntousakis et al. [20] proposed a decentralized automatic merging algorithm, in which each vehicle makes use of the information received from other agents to ascertain the appropriate sequence to merge on the ramp, and experiments show that the algorithm is safely executed and the traffic stability is well maintained. In comparison to the decentralized methods, centralized approaches have also been widely discussed. Cao et al. [21] have proposed the concept of cooperative merging, in which only the vehicles on the main road get the information of the vehicles on ramp road from the centralized controller and then adjust their speed to optimize the passing efficiency. In [22], both centralized and decentralized methods are adopted and a trajectory planning method is proposed to optimize engine efficiency and passenger comfort by adding jerk (derivative of acceleration) in the objective function. The analytical solution is obtained using the optimal control theory and the linear quadratic regulator method, and the model predictive control scheme is then used to compensate for potential interference in the trajectory.

However, most of the methods discussed focus on the optimization of vehicle trajectories with little or no emphasis on the calculation of the best merge sequence (MS), and few studies have discussed the creation of MS assignment [23]. Besides, game theory is also seldom seen in the literature related to the research about on-ramp merging. In essence, issue of vehicles merging at intersection can be regarded as the competition for a prior MS from the angle of an individual agent, which means that small MS represents the priority of passing the merging point. In [24], Jing et al. proposed a cooperative multiplayer game-based optimization framework to coordinate vehicles and achieve minimum values for the global payoff conditions. To simplify the problem, the multiplayer games were decomposed into multiple two-player games and finally formulated as an optimization problem and got an analytic solution.

We also apply game theory to deal with the on-ramp merging issue but the differences between our work and [24] primarily lie on the following: (1) cooperation is allowed among vehicles on the same road, and the competition is between two groups; (2) more aspects, like traffic intensity, fairness, and wholeness principle, are considered when designing the game rules.

3. Problem Framework

The scenario of vehicle merging consists of a main road and a ramp road and both of them are single lane. We assume there exists a centralized controller that can communicate with all vehicles in the control area without any time delay. The control area is divided into *Game Area*, where each agent adopts its optimal strategy according to the game rules, and *Adjusting Area*, where each agent adjusts their states to cooperatively pass the merging point O without any collisions. O is the origin of the coordinates, and the two roads are noted as X_m and X_r, representing the main road and ramp road, respectively, and the length of *Game Area* and *Adjusting Area* is G and M, as shown in Figure 1.

Assume that there are W vehicles passing the merging point during the research time and the centralized controller will assign the passing sequences i ($i = 1,2, \ldots, W$) to each agent at the moment when the first vehicle V_1 reaches the *Adjusting Area*, in accordance with the states of all vehicles at

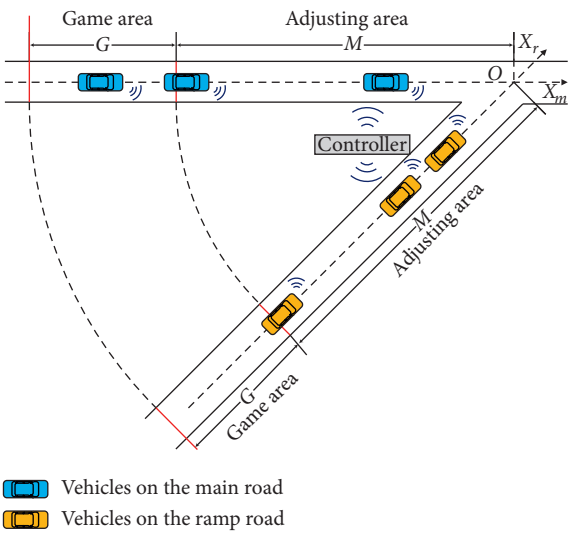


Figure 1: General framework of the proposed system.

that moment. Each vehicle V_i is modeled as a point mass and its state is described as

$$\chi_i(t) = [p_i(t), v_i(t), a_i(t)], \tag{1}$$

where $p_i(t), v_i(t), a_i(t)$ represent the position (or coordinate value), velocity, and acceleration of the vehicle V_i at the time t and $t = 0$ is the moment when the W th vehicle V_W enters the *Game Area* (we assume that length of the *Game Area* is long enough and the first vehicle V_1 is still in *Game Area* at $t = 0$).

4. Modeling and Solution

Game theory is the cornerstone of the proposed model, and in this section, we will illustrate our model in detail and formulate the problem and then give a numeric solution based on searching Pareto solutions.

4.1. Priority of Passing Game. In a game, each player tends to make the best decision for himself/herself based on the obtained information and game rules. In *Game Area*, each agent can obtain the information of other agents from the central controller and try to gain the maximum benefits with minimum cost based on a series of rules and the information obtained. Here, $t = 0$ is the time when the game begins, and the game ends at the moment when the first vehicle reaches the *Adjusting Area*. And, in this part, we will illustrate the benefits, cost, and rules in detail.

During the gaming period, vehicles on the same road take the strategy of cooperation since overtaking is not allowed on a single lane road. We regard all vehicles on the main road as a group to compete with another group consisting of all vehicles on the ramp road, and each group tends to strive for the maximum benefits with least possible cost from the holistic perspective. Benefits are the merging sequences (MS, also known as passing sequences) assigned to each vehicle, and small MS represents a senior right to pass the merging point. The cost is originated from state changing of the vehicles, represented via fuel consumption, and the rules are given empirically and listed as below:

Rule 1: vehicles in the group that contains more vehicles are more prior to pass

Rule 2: vehicles that are closer to the merging point O, i.e., have greater coordinate values, are more prior to pass

Rule 3: vehicles with smaller mean space gap from its preceding agent and its following agent are more prior to pass

These rules can be explained via three principles, i.e., Traffic Density Principle, FIFO Principle, and Wholeness Principle, respectively.

Rule 1 takes the Traffic Density Principle into consideration. If one of the roads has more vehicles, it means the traffic density of this road is higher and gives passing priority to the vehicles on this road which will improve the traffic efficiency (represented by the passing time of all vehicles).

Rule 2 is in accordance with the FIFO Principle, and this rule is related to fairness. Assume there are two roads, i.e., Road A and Road B, converging at an intersection, and the traffic density of Road A is much greater than Road B. The traffic efficiency will be improved if vehicles on Road B stop before the merging area to await the vehicles on another road passing through the merging point. But such a situation is unfair to the vehicles on Road B because they suffer from a jam at the same time.

Rule 3 is related to the Wholeness Principle. Considering the scenario shown in Figure 2, x_j ($j = 1,2,\ldots, 6$) represents the coordinate values of vehicle V_j and $x_1 > x_2 > x_3 > x_4 > x_5 > x_6$. Among these vehicles, V_2, V_3, and V_5 are on the main road and V_1, V_4, and V_6 are on the ramp road, and the spacing gap between V_2 and V_3 is much smaller than that between V_1 and V_4. Assume V_1 and V_2 will have a collision when merging if none of them change their states, and in this case, adjusting V_1 pays less cost than adjusting V_2 and V_3. That means V_2 and V_3 share more "wholeness" than V_1 and V_4 and should have the priority to pass the merging point as a whole. Similar analysis can be done to V_4, V_5, and V_6. In fact, this rule is the extension of *Rule* 1 embodied in a specific local area and represented by measuring the mean space gap of a vehicle between its preceding one and its following one (the first vehicle on a road uses the gap between itself and its following one to describe this attribute, while the last vehicle on a road uses the gap between itself and its preceding one to describe this attribute). For example, L_{23} depicts such attribute to V_2 and $(L_{23} + L_{35})/2$ to V_3.

Based on these three rules aforementioned, each vehicle will adjust their states during gaming period with two motivations, i.e.,

(a) Gaining a coordinate value as greater as possible (O is the origin of coordinates and the positive direction is the moving direction of vehicles, as shown in Figure 1.)

(b) Narrowing the space gap between successive vehicles under the safe vehicle space distance constraint

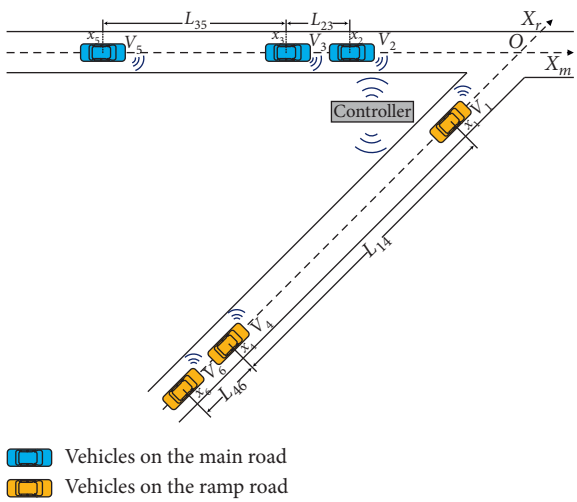


Figure 2: Scenario illustrating *Rule 3*.

We use IDM (intelligent driver model) to depict following vehicles in a group, and the leader vehicle for each group will adopt the strategy of speeding up. Once there is a vehicle reaching the *Adjusting Area*, the game ends and we note this moment as t_f. And, the controller assigns MS to each vehicle according to the states of all vehicles at t_f moment. IDM [25] is formulated as follows:

$$v(t) = v(t-1) + \Delta t \cdot \alpha\left[1 - \left(\frac{v(t-1)}{v_0}\right)^4 - \left(\frac{S_0 + v(t-1)\cdot T_h + ((v(t-1)\cdot \Delta v)/2\sqrt{\alpha\beta})}{S}\right)\right], \tag{2}$$

where $v(t)$ is the velocity at moment t, which can be iteratively obtained from $v(t-1)$, Δt is the interval of two successive time frames, v_0 is the velocity that the vehicle expects to achieve, S is the real distance between the vehicle and its preceding one, S_0 is the minimal safety distance between two vehicles, T_h is the time headway, Δv is the velocity difference between two successive vehicles (note: $v_p(t)$ is the velocity of the preceding vehicles at moment t, $\Delta v = v(t-1) - vp(t-1)$), and α and β are two parameters.

Equation (2) is a linear control law to maintain the stability of a group of vehicles (also called "a platoon"), but this model may generate unstable acceleration, especially at the initialization and state switching moment. We will illustrate such cases in detail in the next part of this section.

As mentioned above, we evaluate each vehicle from 3 aspects, i.e., number of vehicles in its group, distance from the merging point O, and mean space distance of a vehicle between its preceding one and its following one, via scoring at the moment t_f. For each item, the vehicle which performs the best gets 10 scores and the worst one gets 1 scores, and the others linearly get the corresponding scores. For example, to get the score that evaluates the distance from the merging point (assuming the coordinate value for each vehicle V_j is $x_j\,(j = 1, 2, \ldots, n)$, we firstly set the vehicle with the maximum coordinate value max (x_j) as 10 and the vehicle with the minimum coordinate value min (x_j) as 1 and then obtain the linear coefficient vector $K_2 = [k_2, b_2]^T$:

$$K_2 = [k_2, b_2]^T = \begin{bmatrix} \max(x_j) & 1 \\ \min(x_j) & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 10 \\ 1 \end{bmatrix}. \tag{3}$$

Each vehicle V_j can be scored as

$$S_2 = \left[s_2^1, s_2^2, \ldots, s_2^n\right]^T = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \\ 1 & 1 & \ldots & 1 \end{bmatrix}^T \cdot K_2. \tag{4}$$

Similar technique can be applied to get the score S_3 that evaluates the mean space distance of a vehicle between its preceding one and its following one:

$$\begin{aligned} K_3 &= [k_3, b_3]^T = \begin{bmatrix} \min(d_1) & 1 \\ \max(d_n) & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 10 \\ 1 \end{bmatrix}, \\ S_3 &= \left[s_3^1, s_3^2, \ldots, s_3^n\right]^T = \begin{bmatrix} d_1 & d_2 & \ldots & d_n \\ 1 & 1 & \ldots & 1 \end{bmatrix}^T \cdot K_3, \end{aligned} \tag{5}$$

where K_3 is the linear coefficient and $d_k\,(k = 1, 2, \ldots, n)$ is the mean space gap of vehicle V_k between its preceding one and its following one.

The scores $S_1 = [s_1^1, s_1^2, \ldots, s_1^n]^T$ are used to evaluate the number of vehicles in a group, and the vehicles in the same group will get the same score. Note Num_m as the number of vehicles on the main road and Num_r as the number of vehicles on the ramp road, and then all vehicles on the main road is scored as $Num_m/(Num_m + Num_r)$ and $Num_r/(Num_m + Num_r)$ for all vehicles on the ramp road.

The weight vector $w = [w_1, w_2, w_3]^T$ is assigned to describe the importance of S_1, S_2, and S_3 correspondingly. We can get the final score S as

$$S = [\,S_1 \;\; S_2 \;\; S_3\,] \cdot w. \tag{6}$$

As the method of giving the optimal weight vector w, we use the searching approach that will be illustrated in Section 3. And, the controller assigns the MS according to the final score. The higher score a vehicle gets, the smaller sequence (means it is more prior for the vehicle to pass the merging point O) the vehicle obtained.

However, it may occur that the vehicle in front is assigned a bigger MS than the vehicle behind in a group. In this case, vehicles in the same group should adjust their MS in accordance with their distance to the merging point O (the closer to the merging point, the smaller sequence is obtained) since overtaking is not allowed on a single lane road.

4.2. State Adjusting. Once the game ends at moment t_f, the state adjusting process starts and we assume the length of *Adjusting Area* is long enough for state adjusting. We map all vehicles into one dimension, and all vehicles get a sequence based on their coordinate values, called position sequence (PS). The bigger the coordinate value of a vehicle, the smaller PS it gets. There are 3 possible relationships between MS and PS, i.e., MS = PS, MS > PS, and MS < PS. By comparing the MS and PS, each vehicle will respond differently, as shown below.

4.2.1. MS = PS. MS = PS means the merging sequence for a vehicle is the same to the position sequence, and maintaining the current state will consume the least fuel. In this circumstance, the vehicle will simply obey the IDM model. but the reference vehicle may not on the same road. As shown in Figure 3, V_1 and V_3 are on the same road and V_2 is on another road, while the relationship of their coordinate values is $x_1 > x_2 > x_3$ and the MS relationship is $\mathrm{MS}_1 < \mathrm{MS}_2 < \mathrm{MS}_3$. For V_3, its reference vehicle is V_2, not V_1.

As we mentioned in the previous part, IDM may generate acceleration jumping especially at the state switching moment. When we map all vehicles into one dimension, the reference vehicle for an agent may change, which is equivalent to that its "preceding" vehicle's state abruptly changes and it causes the acceleration jumping. To solve this problem, we constrain acceleration within certain range to avoid the unexpected jumping.

4.2.2. MS > PS. MS > PS means the merging sequence for a vehicle is more prior than its real position sequence. Therefore, it should adopt the strategy of accelerating only if the space distance of the vehicle and its preceding one is greater than the minimal safe distance.

4.2.3. MS < PS. MS < PS means the vehicle should make concessions to other vehicles. Therefore, it should adopt the strategy of decelerating until it receives the signal which indicates that the PS of the vehicle is already in accordance with its MS from the controller. Then, the vehicle just simply adopts the IDM model illustrated in the part "MS = PS."

4.3. Optimization for Fuel Consumption and Travel Time. In the previous section, we discussed the possible actions that each vehicle can take in different periods (i.e., gaming period and adjusting period). We can conclude that the strategy a vehicle adopts in gaming period is in accordance with the game rules, and the weight vector $w = [w_1, w_2, w_3]^T$ measures the importance of different rules and its value will influence the assigning of the MS to each vehicle. Therefore, we try to find an optimal w to get a reasonable MS so that the fuel consumption and total travel time can be optimized. This is a typical optimization problem of two objectives and can be formulated as

$$\begin{aligned} &\min, \quad J = \mu F + \gamma T = g\left(w = \left[\, w_1 \;\; w_2 \;\; w_3 \,\right]^T\right), \\ &\text{s.t,} \quad a_i(t) \in \left[a_{\min}, a_{\max}\right], v_i(t) \in \left[0, v_{\max}\right], \sum_{j=1}^{3} w_j = 1, \end{aligned} \tag{7}$$

where F is the total fuel consumption, T is the total travel time, g is the function that denote the relationship between weight vector w and the optimization object, $a_{\min}$ and $a_{\max}$ are the minimum deceleration and maximum acceleration, respectively, $v_{\max}$ is the maximum speed limit, and μ and γ are two parameters that can adjust the object function.

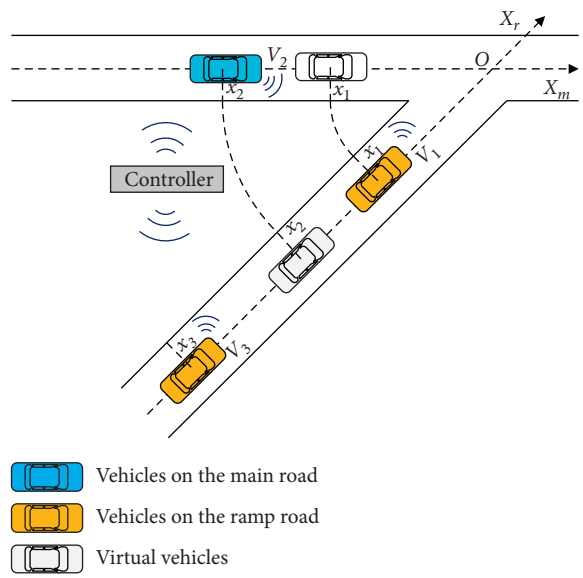


Figure 3: Explanation to the reference vehicle.

As to the fuel consumption, we refer to [26, 27] and adopt the estimation as

$$\begin{aligned} f &= f_v + f_a, \\ f_v &= q_0 + q_1 v(t) + q_2 v^2(t) + q_3 v^3(t), \\ f_a &= a(t) \cdot \left(r_0 + r_1 v(t) + r_2 v^2(t)\right), \end{aligned} \tag{8}$$

where f is the instantaneous fuel consumption consisting of f_v and f_a, representing instantaneous fuel consumption caused by velocity and acceleration, respectively. If the acceleration value is negative, we use the absolute value instead. Coefficient vectors $q = [q_0, q_1, q_2]$ and $r = [r_0, r_1, r_2]$ are obtained from the experiment [28], and the values are $q = (0.1569,\ 2.45 \times 10^{-2},\ -7.415 \times 10^{-4})$ and $r = (0.07224,\ 9.681 \times 10^{-2},\ 1.075 \times 10^{-3})$. The total fuel consumption is the sum of instantaneous fuel consumption in each moment for all vehicles from $t = 0$ (the moment when the last vehicle arrives to the *Game Area*) to the moment when the last vehicle passes the merging point O, noted as t_F.

About the total travel time T, we start timing at $t = 0$ and end timing at $t = t_F$, and $T = t_F - 0 = t_F$.

From Formula (7), we can see that F and T are functions of w, but we do not know the explicit expression of g; therefore, we cannot solve the optimization problem analytically based on (7). Besides, for a double-objective optimization problem, there usually does not exit such a solution where both objectives are optimized; we should make a trade-off between objectives according to practical requirement. Here, we will find a feasible Pareto solution [29] for this problem. A Pareto solution is the solution where further optimization for an objective is always at the cost of the deterioration of other objectives. For example, as shown in Figure 4, A and B are two objectives we dedicate to minimize and (x_1, y_1) is a Pareto solution, because if we continue decrease the value near x_1 to optimize objective A, the performance of objective B will deteriorate. Situation for (x_3, y_3) is similar to (x_1, y_1), while (x_2, y_2) is not a Pareto solution since both objectives can be further minimized near this point. There

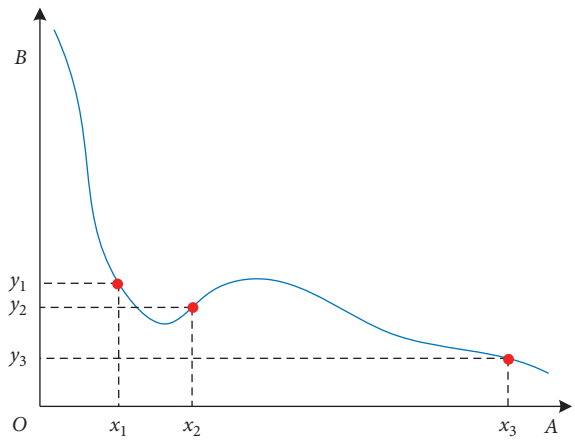


Figure 4: Illustration of the Pareto solution.

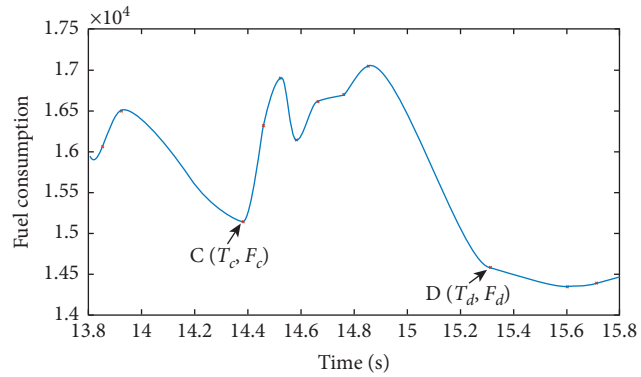


Figure 5: *F-T* curve in one experiment. The red dot is the simulation results, and the blue curve is the shape preserving interpolation based on the red dot.

are many Pareto solutions in a double (or multiple) objective optimization problem, and these solutions consist a solution set (called Pareto front), as shown in Figure 4. There are feasible solutions in such sets and we select the best fit in accordance with the trade-off between different objectives.

4.4. Searching for the Pareto Solution. From previous part, we conclude that possible acceptable solutions can be found in the Pareto front set. In our problem, two optimization objectives are cumulative fuel consumption F and total travel time T, and in this part, we give the steps of searching Pareto solutions for this specific problem:

Step 1: divide w with comparatively large step size and compute the cumulative fuel consumption F and total travel time T.

In this step, we divide each component of w with stride of 0.05, under the constraint that the sum of all components is 1. For each node, compute F and T for N times (N is the repeated computing times for a node, for example, $N = 100$) and average the results.

Step 2: subdivide w in the sparse area.

By analyzing the results in Step 1, the distribution of the results in some range may happen to be sparse. We further subdivide w in these sparse range with stride of 0.01, even much smaller, and compute F and T for N times at these subdivided nodes.

Step 3: find the "inflection points" on the *F-T* curve.

Figure 5 is a *F-T* curve we get in one of our simulation experiments and we call points C and D "inflection" points where optimizing one of the objectives will lead to large deterioration of another objective. In acceptable range of F and T, we regard these points as the best trade-off of different objectives according to practical requirements and choose one of them as the "best" fit for the problem.

Step 4: back search for the corresponding w.

We choose point $C(T_c, F_c)$ as the best trade-off of our two objectives and search back to find the w that correspond to (T_c, F_c). There can be several proper values for w and any one of them can be chosen as the solution.

5. Simulation Results and Analysis

To display the simulation results of the trajectory, velocity, acceleration, and fuel consumption during gaming and adjusting periods, we should firstly ascertain a weight vector $w = [w_1, w_2, w_3]$. Based on the steps of searching Pareto solutions described in the previous section, we select $w = [0.15, 0.45, 0.4]$ as the best fit, and all case studies are under such condition. Besides, we set a "sliding window" parameter W to decide the number of agents per game.

5.1. Case Study 1: "Sliding Window" Parameter $W = 5$. This simulation is to verify the effectiveness under the condition that the number of vehicles is comparatively small in one turn of a game, in this case, 5 CAVs driving on the two merging roads. The entry time and the number of vehicles on main road and ramp road are random. The value of the time-headway can be in the range of [1.2 s, 2.4 s] in different scenarios [30] and here we choose $T_h = 1.2$ s. As to the length of *Game Area* and *Adjusting Area*, we set $G = 100$ m and $M = 200$ m. Each vehicle enters the *Game Area* with a random initial speed around 15 m/s, and the maximum speed limitation is 30 m/s. And, the acceleration of the vehicles is constrained from $-3\,\mathrm{m/s^2}$ to $3\,\mathrm{m/s^2}$.

The simulation of the position, velocity, and acceleration is shown in Figures 6–8. The gaming period ends at the moment when the first vehicle reaches the *Adjusting Area*, and in this simulation, the moment is at 3.12 s. Since saltation of acceleration is not practical, we set the change rate of acceleration (also known as jerk) as $1\,\mathrm{m/s^3}$. Therefore, there is no abrupt change of acceleration, and the profiles for velocity and trajectory are smooth except the transition moment.

The time-headway of each vehicle passing the merging point is close but not strictly equal, shown in Figure 6, since the velocity of each one at the merging point is different.

As to fuel consumption, we adopt the polynomial model given in formula (8), and it manifests that the instantaneous

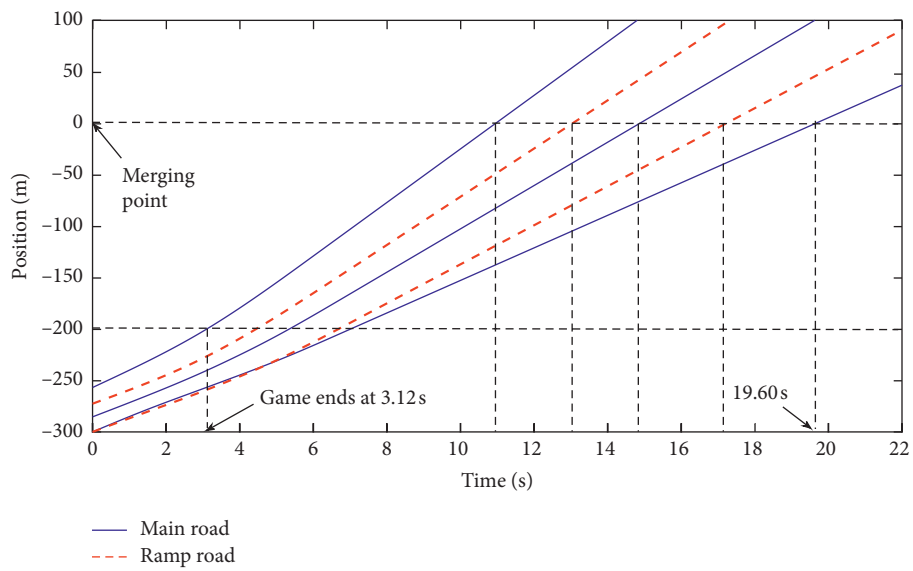


Figure 6: Position of vehicles for Case Study 1.

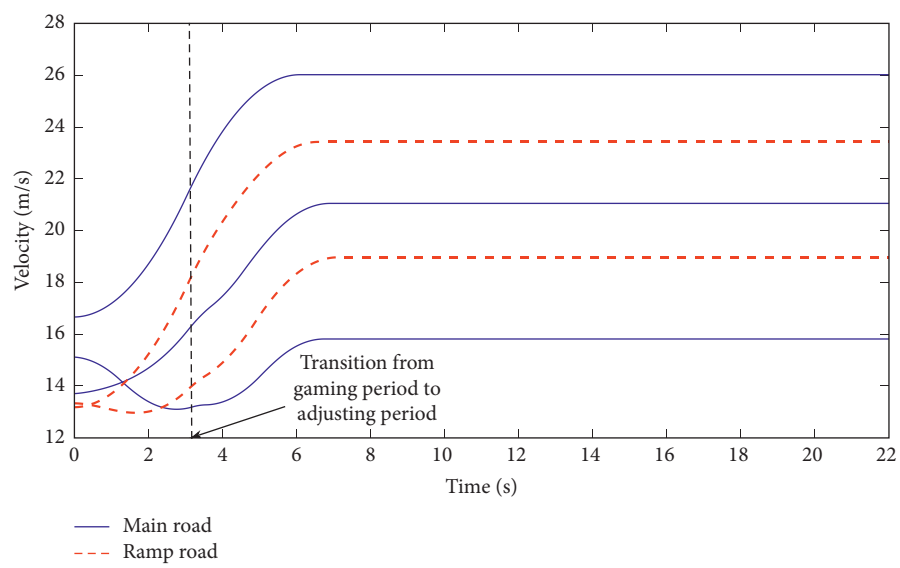


Figure 7: Velocity of vehicles for Case Study 1.

fuel consumption depends on the instantaneous velocity and acceleration. We try to compare the fuel consumption between our model and (1) FIFO; (2) Density Prior, and (3) No Control, as shown in Figure 9. Here, "FIFO" is the situation where vehicles that are closer to the merging point are prior to pass the merging point. "Density Prior" is the situation where vehicles on the road with higher traffic density (reflected on the number of vehicles on that road) are prior to pass the merging point and vehicles on the other road should stop to wait until all vehicles on the road with higher traffic density pass the merging point. "No Control" is the situation where vehicles on the ramp road have to stop to wait for vehicles on the main road as they cross the merging point and then start accelerating to reach the maximum speed to pass the merging point.

For FIFO, we initialize all vehicles with the velocity of 15 m/s, and the MS is assigned at $t = 0$ completely based on their distance to the merging point, and other settings are the same as our model. It is, in fact, a special situation of our model where the length of *Game Area* is zero. For "No Control," we also set the initial velocity of each vehicle as 15 m/s, and in this situation, vehicles on the ramp road must stop to wait until vehicles on the main road passing the merging point. For "Density Prior," we set the same initial velocity for each vehicle. Since, in this case study, the number of vehicles on the main road is greater than that on the ramp road, the vehicles on the ramp road should also stop to wait until vehicles on the main road pass the merging point. However, in Case Study 2, we discuss and display the situation where number of vehicles on the ramp road is greater than that on the main road.

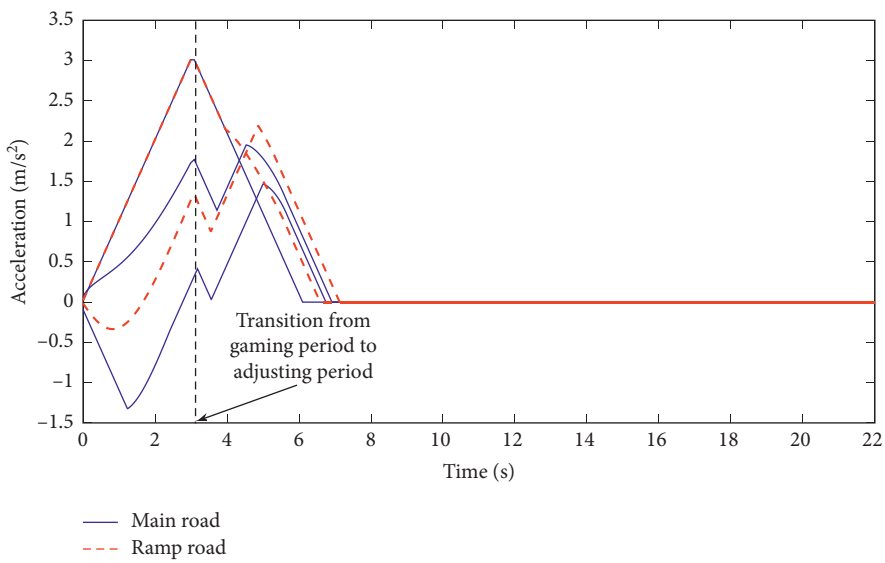


Figure 8: Acceleration of vehicles for Case Study 1.

From Figure 9, we can see that the proposed model consumes the least cumulative fuel at the end. However, fuel consumption of the "No Control" model is much lower at the beginning stage since vehicles do not change their states during this period. While fuel consumption of "No Control" model markedly increases in two stages, i.e., the stage in which vehicles on the ramp road slow down and stop and the stage in which vehicles on the ramp road restart and accelerate. Fuel consumption of "Density Prior" is very similar to that of "No Control," since, in this case study, the number of vehicles on the main road is greater than that on the ramp road, and "Density Prior" and "No Control" are equivalent in this situation.

As to the comparison of the fuel consumption between proposed model and FIFO model, the difference is that there is no *Game Area* in the FIFO model. FIFO is the model that address fairness at the cost of efficiency. Therefore, we can see that the total fuel consumption of FIFO is higher than that of proposed model.

The moment when the last vehicle passing the merging point is the end of the process, and the value of this moment equals the total travel time. And, for this simulation, the total travel time is 19.60 s, shown in Figure 6. The total travel time for the FIFO model is close to the proposed model, but for the "No Control" and "Density Prior" model, it is remarkably higher (about 12% higher) than the other two.

5.2. Case Study 2: "Sliding Window" Parameter $W = 10$. In this case study, 10 vehicles are involved in one turn of the game and the number of vehicles on the main road and ramp road is randomly set. Constraints for the vehicles are the same as those in Case Study 1, and some parameters are adaptively changed. We set $G = 300$ and $M = 200$, and the simulation of the position, velocity, and acceleration is shown in Figures 10–12.

From Figures 10–12, we can see that the vehicles on the main road pay more cost, reflected from the amplitude of variation of acceleration and velocity; because in this situation, the number of the vehicles on the ramp road is remarkably greater than that on the main road, and the vehicles on the ramp road are dominant as a whole based on the game rules. However, acceptable sacrifice of the disadvantaged group brings the improvement of the efficiency and stability for the whole, and we can see that the variation of the acceleration and velocity of the vehicles on the ramp road (the dominant group) is relatively smoother.

Comparison of fuel consumption among three different models in Case Study 2 is shown in Figure 13. We can see the proposed model still consumes the least cumulative fuel at the end. However, because there exist more vehicles on the ramp road in this case, fuel consumption of the "No Control" model remarkably increased especially during the stage in which vehicles on the ramp road slow down to stop and the stage in which vehicles on the ramp road accelerate. As to the "Density Prior," vehicles on the main road slow down or stop to avoid collision in this case because number of CAVs on the ramp road is greater than that on the main road. And, we can see that the shape of the fuel consumption curve for "Density Prior" is similar to that of "No Control," but the cumulative fuel consumption is less since fewer vehicles change their states in the "Density Prior" model than those in the "No Control" model under this case study. Analysis to the comparison between proposed model and FIFO model is similar to that of Case Study 1 and the former still takes advantages than the latter in Case Study 2.

As to total travel time in Case Study 2, it is 35.90 s, shown in Figure 10, and for the FIFO model, it is near to the proposed model. The "No Control" model spends much more time in this case study (about 17% higher than the other two) since the number of vehicles on the ramp road is much greater than that on the main road and the process of slowing down and accelerating to restart takes much time. The proposed model does not take advantage in the total travel time compared with the "Density Prior" model in this case, since the number of vehicles on different roads is unbalanced. "Density Prior" is a model that emphasizes the

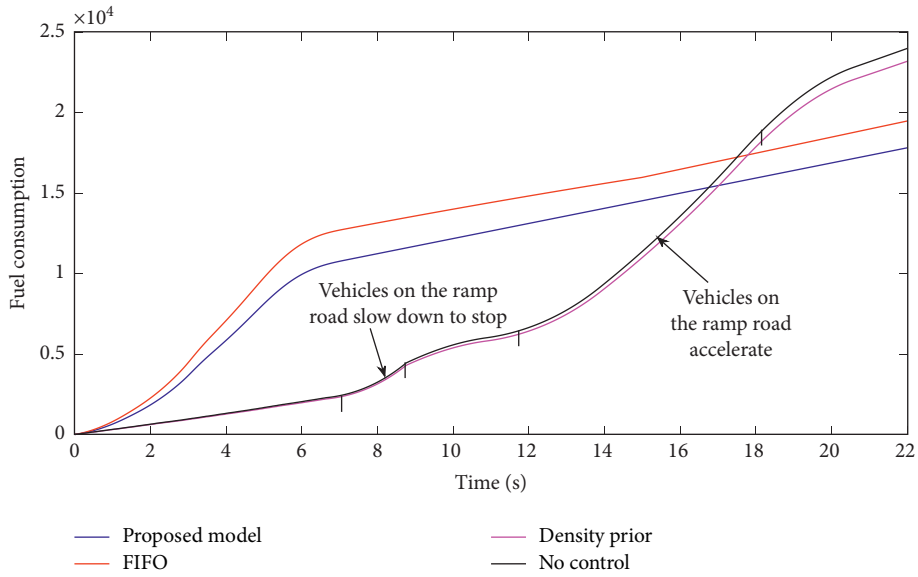


Figure 9: Comparison of fuel consumption for different models in Case Study 1.

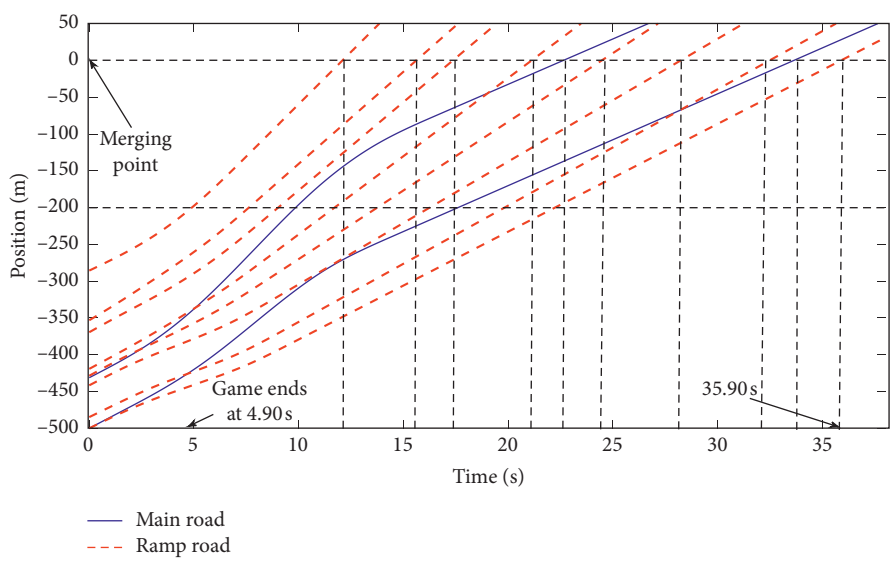


Figure 10: Position of vehicles for Case Study 2.

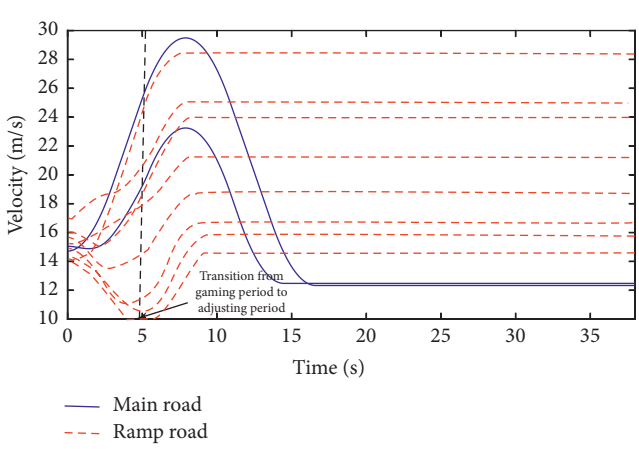


Figure 11: Velocity of vehicles for Case Study 2.

efficiency, and it performs well for those circumstances where the traffic density for different roads is extremely unbalanced.

5.3. Brief Summary of the Simulation Results. We simulate two cases and validate the effectiveness and efficiency of the proposed model. Case Study 1 mainly focused on the scenario where the number of connected autonomous vehicles is comparatively small in one turn of a game, while Case Study 2 mainly researched the situation where much more agents are involved per round in the game. As mentioned above, FIFO is the model that addresses fairness, while "Density Prior" is a model that emphasizes the efficiency. The proposed model gives consideration to both aspects and performs well.

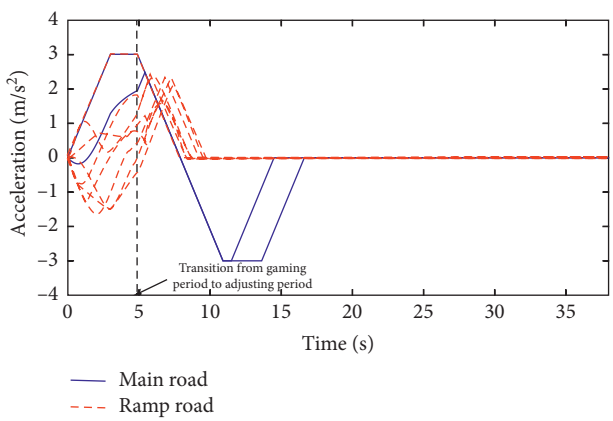


Figure 12: Acceleration of vehicles for Case Study 2.

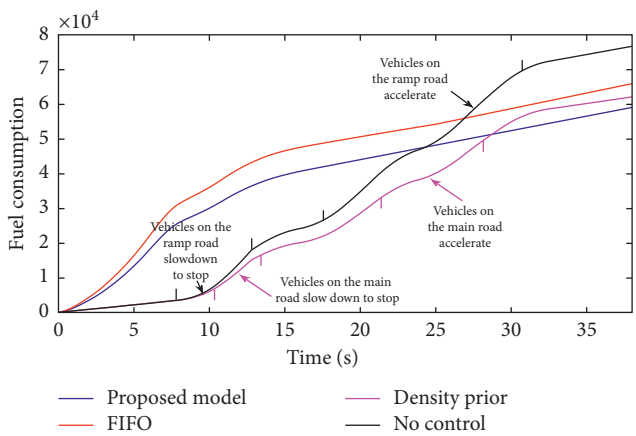


Figure 13: Comparison of fuel consumption for different models in Case Study 2.

One thing that should be pointed out is the setting of the parameters M and G. Although we regard the vehicles as mass point, we cannot set M or G as a very small value and settings of these two parameters varies, given different "sliding window" parameter W.

6. Conclusion

In this paper, we focus on the scenario of CAV on-ramp merging where each vehicle is equipped with V2I equipment and can communicate with the centralized controller without any time delay according to the assumption.

The core of the CAVs on-ramp merging problem lies in the assignment of the MS, and different MS leads to the difference on cumulative fuel consumption and total travel time. In our paper, we regard the competition of MS as a game and vehicles on the same road are regarded as a group taking the strategy of cooperation to strive for as small MS as possible against another group. As to the game, we proposed three basic rules in accordance with the Traffic Density Principle, FIFO Principle, and Wholeness Principle, respectively. Further, the control area is divided into *Game Area*, where vehicles on each road change their states cooperatively, attempting to get a better MS, and *Adjusting Area*, where all vehicles adjusting their states according to the MS are assigned. The game starts at the moment when the last vehicle reaches the *Game Area* (we assume length of the *Game Area* is long enough that all vehicles are still in the *Game Area* when game starts) and ends at the moment when there is a vehicle reaching the *Adjusting Area*.

In both gaming period and adjusting period, IDM is adopted to describe the behavior of the following vehicle and we modify this model, such as adding the constraints of acceleration and velocity, to avoid the abrupt state change at the transition moment. As to the assignment of MS, we evaluate each competitor from three aspects based on the three given rules. To ascertain an optimal weight vector which depicts the influence of each rule, we formulate the problem as a double-objective optimization problem of cumulative fuel consumption and total travel time. Since the relationship between the optimization objectives and the weight vector is unknown, we give different weight vectors as input to search the Pareto Solutions and choose the "inflection point" on the F-T curve as the best fit of our problem. And finally, we search back for the corresponding weight vector as the solution.

Simulation and comparison of the proposed model and other models are conducted and the effectiveness of our model is validated. However, all the results are obtained based on a series of ideal assumption, including no communication, vehicle dynamic delay, etc. Further exploration is needed on the use of feedback control to improve the accuracy, considering the environmental noise and disturbance. Besides, lateral control to the vehicle should also be considered when extending the model to the practical implementation in the future work.

Data Availability

The original codes of the numerical tests 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 interests.

Authors' Contributions

Xiangmo Zhao and Haigen Min conceived and designed the research. Xiaochi Li and Runmin Wang performed the experiments. Haigen Min and Yukun Fang analysed the data. Haigen Min and Yukun Fang wrote the manuscript. All authors read and approved the final manuscript.

References

[1] B. Schrank, B. Eisele, T. Lomax, and J. Bak, *Urban Mobility Scorecard*, Texas A&M Transp. Inst., College Station, TX, USA, 2015.

[2] L. Xu, C. Feng, V. R. Kamat, and C. C. Menassa, "An occupancy grid mapping enhanced visual SLAM for real-time locating applications in indoor GPS-denied environments," *Automation in Construction*, vol. 104, pp. 230–245, 2019.

[3] M. Yang, B. Ai, R. He, C. Huang, J. Li, L. Chen, and X. Li, "Influence of different antenna locations on channel characterization for V2V communications," in *Proceedings of the 2018 IEEE Antennas and Propagation Society International Symposium and USNC/URSI National Radio Science Meeting*, pp. 377-378, Boston, MA, USA, July 2018.

[4] I. Kabashkin, "Reliable v2x communications for safety-critical intelligent transport systems," in *Proceedings of the 2017 Advances in Wireless and Optical Communications (RTUWO)*, pp. 251–255, Riga, Latvia, November 2017.

[5] L. C. Davis, "Effect of cooperative merging on the synchronous flow phase of traffic," *Physica A-Statistical Mechanics and its Applications*, vol. 361, no. 2, pp. 606–618, 2006.

[6] E. Moradi-Pari, H. N. Mahjoub, H. Kazemi, A. Tahmasbi-Sarvestani, and Y. P. Fallah, "Utilizing model-based communication and control for cooperative automated vehicle applications," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 1, pp. 38–51, 2017.

[7] S. E. Li, Y. Zheng, K. Li et al., "Dynamical modeling and distributed control of connected and automated vehicles: challenges and opportunities," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 3, pp. 46–58, 2017.

[8] C. M. Filho, M. H. Terre, and D. F. Wolf, "Safe optimization of highway traffic with robust model predictive control-based cooperative adaptive cruise control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 11, pp. 3193–3203, 2017.

[9] O. D Altan, G. Wu, M. J. Barth, K. Boriboonsomsin, and J. A. Stark, GlidePath: eco-friendly automated approach and departure at signalized intersections," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 4, pp. 266–277, 2017.

[10] D. Bevly, X. Cao, M. Gordon et al., "Lane change and merge maneuvers for connected and automated vehicles: a survey," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 105–120, 2016.

[11] H. Hadj-Salem, J. M. Blosseville, and M. Papageorgiou, "ALINEA: a local feedback control law for on-ramp metering; a real-life study," in *Proceedings of the Third International Conference on Road Traffic Control*, pp. 194–198, London, UK, 1990.

[12] M. Papagcorgiou and A. Kotsialos, "Frecway ramp metering: an overview," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 271–281, 2002.

[13] E. Smaragdis, M. Papageorgiou, and E. Kosmalopoulos, "A flow-maximizing adaptive local ramp metering strategy," *Transportation Research Part B: Methodological*, vol. 38, no. 3, pp. 251–270, 2004.

[14] I. Papamichail and M. Papageorgiou, "Traffic-responsive linked ramp-metering control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 111–121, 2008.

[15] R. C. Carlson, L. Papamichail, and M. Papageorgiou, "Local feedkack-based mainstream traffic flow control on motorways using variable speed limits," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1261–1276, 2011.

[16] K. Dresner and P. Stone, "Multiagent traffic management: a reservation-based intersection control mcchanism," in *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 530–537, New York, NY, USA, July 2004.

[17] J. Rios-Torres and A. A. Malikopoulos, "A survery on the coordination of Connected and automated vehicles at intersections and merging at Highway on-ramps," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1066–1077, 2017.

[18] K. Dresner and P. Stone, "A multiagent approach to autonomous inter-section management," *Journal of Artificial Intelligence Research*, vol. 31, no. 1, pp. 591–653, 2008.

[19] T.-C. Au, P. Stone, and M. Hausknecht, "Motion planning algorithms for autonomous intersection management," in *Proceedings of the AAAI Workshop Bridging Gap BTAMP*, pp. 1–8, Seattle, WA, USA, 2010.

[20] I. A. Ntousakis, K. Porfyri, I. K. Nikolos, and M. Papageorgiou, "Assessing the impact of a coperative merging system on highway traffic using amicroscopic flow simulator," in *Proceedings of the International Mechanical Engineering Congress and Exposition*, pp. 1–10, Montreal, Quebec, Canada, November 2014.

[21] W. Cao, M. Mukai, T. Kawabe, H. Nishira, and N. Fujiki, "Coperative vehicle path generation during merging using model predictive control with real time optimization," *Control Engineering Practice*, vol. 34, pp. 98–105, 2015.

[22] I. A. Ntousakis, I. K. Nikolos, and M. papagcorgiou, "Optimal vehicle trajectory planning in the context of cooperative merging on highways," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 464–488, 2018.

[23] T. Awal, L. kulik, and K. Ramamohanrao, "Optimal traffic merging strategy for communication-and sensor-enabled vehicles," in *Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pp. 1468–1474, The hague, Netherlands, October 2013.

[24] S. Jing, F. Hui, X. Zhao, J. Rios-Torres, and A. J. Khattak, "Cooperative game approach to optimal merging sequence and on-ramp merging control of connected and automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 11, pp. 4234–4244, 2019.

[25] D. Sun, Y. He, M. Zhao, and S. Cheng, "Cooperative driving modelling in the vicinity of traffic signals based on intelligent driver model," *IET Intelligent Transport Systems*, vol. 12, no. 10, pp. 1236–1242, 2018.

[26] J. Rios-Torres and A. A. Malikopoulos, "Automated and cooperative vehicle merging at highway on-ramps," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 780–789, 2017.

[27] M. Kamal, M. Mukai, J. Murata, and T. Kawabe, "Model predictive control of vehicles on urban roads for improved fuel economy," *IEEE Transactions on Control Systems Technology*, vol. 21, no. 3, pp. 831–841, 2013.

[28] W. Zhao, N. Dong, S. Simon, 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.

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 1st edition, 2004.

[30] H. Liu, H. Wei, T. Zuo, Z. Li, and Y. J. Yang, "Fine-tuning ADAS algorithm parameters for optimizing traffic safety and mobility in connected vehicle environment," *Transportation Research Part C: Emerging Technologies*, vol. 76, pp. 132–149, 2017.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 5896739, 13 pages
https://doi.org/10.1155/2020/5896739

Research Article

Forecasting Urban Rail Transit Vehicle Interior Noise and Its Applications in Railway Alignment Design

Yifeng Wang,[1] Ping Wang,[1] Zihan Li,[1] Zhengxing Chen,[1] and Qing He[1,2]

[1]*Key Laboratory of High-Speed Railway Engineering of the Ministry of Education, School of Civil Engineering, Southwest Jiaotong University, Chengdu 610031, China*
[2]*Department of Civil, Structural and Environmental Engineering and Department of Industrial and Systems Engineering, University at Buffalo, The State University of New York, 313 Bell Hall, Buffalo, NY 14260, USA*

Correspondence should be addressed to Qing He; qinghe@buffalo.edu

Received 12 December 2019; Revised 3 May 2020; Accepted 29 May 2020; Published 23 June 2020

Academic Editor: Xingju Wang

In this study, a data-driven interior noise prediction model is developed for vehicles on an urban rail transit system based on random forest (RF) and a vehicle/track coupling dynamic model (VTCDM). The proposed prediction model can evaluate and optimize the sustainability of railway alignment from the perspective of interior noise. First, a data collection framework via embedded sensors of onboard smartphones was developed. Then, for establishing the mapping relationship between the dynamic responses of the car body and interior noise, the collected dataset was fed to the RF. Parameter, error distribution, and feature importance analyses were conducted for evaluating and optimizing the performance of the RF. With the optimized parameters, the probability of prediction errors being within 5 dB was 86.9%. Next, the VTCDM was established using an existing industry multibody simulation tool and verified through a comparison between the simulated and field dynamic responses. Finally, a case study that extends the application of this interior noise prediction model to railway alignment design is presented.

1. Introduction

By the end of 2018, 35 cities in mainland China had opened 180 urban rail transit (URT) lines, with a total operation mileage of more than 5700 km. Moreover, the length of subway lines was approximately 4350 km, accounting for 75.6% of all URT lines. The rapidly expanded URT networks offered great convenience to the citizen and were highly advantageous in solving traffic congestions. As a critical URT component, the subway is crucial in people's daily life. However, owing to the unreasonable design of railway alignments and the degradation of track infrastructures, the running quality of trains worsens with the increase in service time. Furthermore, these issues caused abnormal interior noise and vibration, which severely affected passengers' ride comfort. The interior noise of subway vehicles significantly affected passengers' experience; therefore, the design, construction, and operation of subway lines must be emphasized [1].

Generally, train noise can be divided into two categories: external and interior noises [2]. The interior noise is a complex sound field resulting from external acoustics and mechanical excitation sources transmitting, attenuating, and radiating inward through the carriage structure [3]. The interior noise of trains primarily comes from the electrical equipment, aerodynamic, and wheel-rail noises of the train [2]. The operation speed of subway trains is typically within the range of 30–80 km/h, and at such speeds, wheel-rail noises such as rolling noise, squeal, and impact noise are dominant in the generation of interior noise [4]. Curve squeal noise is an intense noise generated when railway vehicles pass sharp curves [5]. Owing to the limitations in cities, many sharp curves appear in subway lines; therefore, the interior noise caused by squeal is serious. Furthermore, defects on the treads of wheels and rails, such as rail corrugation [4], spall of railheads, and polygonal wear of wheels [6], significantly affect the generation of interior noise.

Therefore, wheel-rail noise can be considered a main source of vehicle interior noise.

Wheel-rail noise, including rolling noise, impact noise, and curve squeal, is primarily affected by the wheel-rail relationship when trains move along a track. However, the wheel-rail interaction not only affects the generation of noise but also affects the dynamic responses of vehicles. For example, when rail corrugation occurs, a series of dynamic problems as well as serious train interior noise are likely to appear. Hence, onboard devices to detect track faults from cabin vibrations and interior noise have been developed [7–9]. Because noise and dynamic responses are both related to wheel-rail relationships, it is assumed that interior noise can be forecasted with the dynamic responses of the car body. As for subways, most trains move in a tunnel, the sound field inside of which is not easily interfered by external environmental factors. Therefore, we propose an idea to establish a mapping relationship between vehicle interior noise and the dynamic responses of trains with RF, which is suitable for nonlinear problems. The RF algorithm was first proposed by [10]; it is widely used owing to its high accuracy and strong robustness. Although the dynamic responses and interior noise between individual vehicles differed even through the same section of the alignment, the strong robustness rendered it capable of learning from complex data.

Because RF is a supervised learning algorithm, a large amount of training datasets is required when modeling. For acquiring these datasets, an onboard smartphone application was developed to collect the dynamic responses and interior noise of the subway vehicle. With the development of microelectromechanical system technologies, smartphones with built-in sensors exhibit excellent properties in terms of portability and practicability. Moreover, several applications regarding transport infrastructure condition monitoring and vehicle motion pattern recognition have been reported [11–14]. These applications and our previous studies indicate that the sensing capabilities of the smartphone can satisfy the requirement of this study [15, 16].

The goal of this study was to develop a data-driven method for forecasting vehicle interior noise for the optimization of URT railway alignments. Hence, a large number of train operation status data were collected using the embedded sensors of an Android smartphone. Subsequently, the relationship between car body dynamic responses and interior noise of subway trains was established using RF. For obtaining the simulated dynamic responses of the car body under specific railway alignments, we built a vehicle/track coupling dynamic model (VTCDM) via the Universal Mechanism (UM) software.

Finally, based on the mapping relationship learning from data and dynamic responses from the simulated model, the interior noise under specific railway alignments can be predicted. Further, the prediction results can be applied for optimizing railway alignments at the design stage. The key contributions of this study are as follows:

(1) Propose a data-driven prediction framework for subway vehicle interior noise based on the smartphone-collected dataset, RF, and VTCDM

(2) Establish a VTCDM via the UM software based on multibody system dynamics

(3) Validate the VTCDM and the proposed interior noise prediction model using the field test data from Chengdu Subway Line 7

(4) Develop an application scenario for optimizing railway alignments with the prediction model of vehicle interior noise

The remainder of this study is organized as follows. Section 2 briefly reviews previous studies regarding methods to predict interior noise in the railway domain. In Section 3, a comprehensive overview of the proposed methodology is provided. Furthermore, in Section 4, the data and features used in our model are introduced. Section 5 describes the modeling approach of RF and VTCDM in detail. The analysis results and discussion are elaborated in Section 6. Finally, conclusions are presented in Section 7.

2. Literature Review

Because train interior noise directly affects passengers' ride comfort, noise control becomes a challenging technology in the design and construction of railway lines. Many factors have an impact on vehicle interior noise. The potential sources of interior noise can be categorized into two groups according to the characteristics of sound: the airborne and structure-borne noise sources [17]. The airborne noise denotes the noise generated by the wheel-rail track system and aerodynamic noise that transmit into the cabin via air, and the structure-borne one refers to the noise radiated from the vibration of vehicle structures inside the cabin. The results of partial coherence analysis of the vibroacoustical signals indicated that the structural vibrations, the cause of structure-borne noise, contribute more to interior noise than airborne noise [18]. Figure 1 describes the composition of vehicle interior noise and some critical influencing factors. The wheel-rail contact force, originating from the roughness of the wheel-rail contact surface, excites the vibration of the wheel and rail to radiate noise. On the other hand, it transmits along with the bogie and suspension system and finally becomes the suspension force to excite vibration of the carriage cabin bottom, generating noise at low frequency [19]. Travel speed of the train, having a relationship with both the wheel-rail force and aerodynamic noise, is one of the most pivotal influencing factors of vehicle interior noise. Besides, the roughness of the tread of rail and wheel, parameters of the wheel-rail and suspension system, and structures of the carriage cabin all have an impact on the characteristics of vehicle interior noise.

Currently, numerous investigations have been conducted regarding the analysis of noise characteristics [20], sound quality evaluations [21], and noise level prediction [22]. For forecasting vehicle interior noise, typically used methods include the finite element method (FEM) [23],

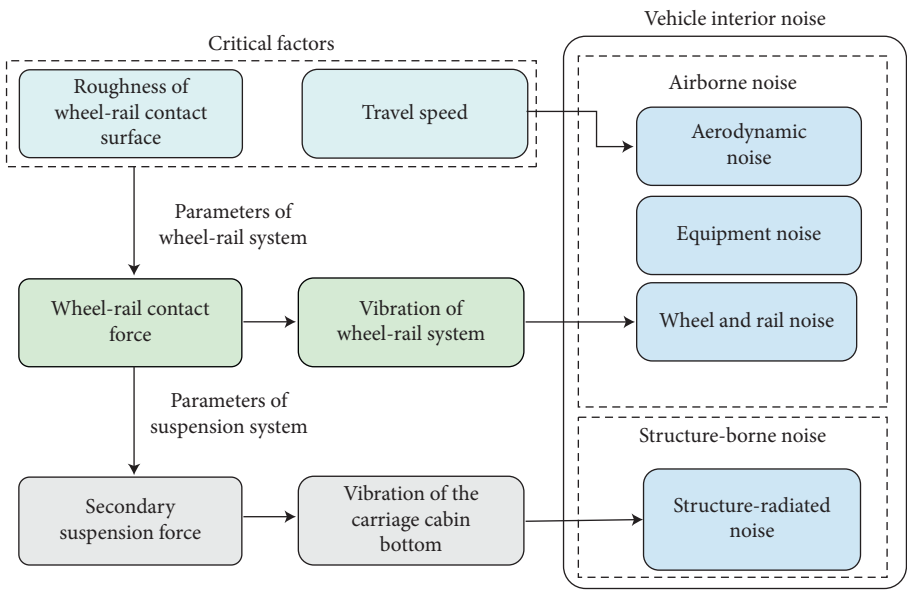


Figure 1: The composition of vehicle interior noise and some critical influencing factors.

boundary element method (BEM) [24], and statistical energy analysis method (SEAM) [3]. As each method has its limitations, these approaches can only be applied to specific situations. To achieve higher prediction accuracy and more exceptional generalization ability, hybrid models comprising two or three of the above numerical methods have been widely implemented in the prediction of vehicle interior noise.

In 1966, Gladwell et al. first proposed formulating acoustic problems in terms of displacements, similar to a continuous elastic structure [25], which is the beginning of applying the FEM for acoustic analysis. The FEM was first used to predict the interior noise of a car supported by General Motors Corporation. Owing to the low running speed of the train, the vehicle interior noise was dominated by low-frequency structural noise, and the FEM performed well [26]. However, when Lucas et al. attempted to predict high-frequency aerodynamic noise using the FEM, satisfactory results were unavailable. They indicated that the FEM could not predict high-frequency noise; that is, as the frequency increased, the modes of the structure became denser and could not be identified by the current identification technologies [27]. Therefore, the FEM was primarily used to predict the structural and wheel-rail noise of the car body.

Unlike the FEM, the BEM does not require the acoustic space to be discretized; only the boundary conditions require discretization, which can reduce the calculation time and minimize the calculation error. Owing to the superiority of the boundary element method, it has been widely promoted recently. Similar to the FEM, the BEM was also first used to predict the interior noise of cars [28]. Subsequently, Letourneaux established a train interior noise within a low-frequency range prediction model based on the BEM, and the high-frequency noise prediction problem was investigated [29]. For a long time, the BEM has only been applied for predicting the vibration and radiated noise of components. In 2011, Soltani established an entire train vehicle model using the BEM and systematically studied the structural noise of the car body [30]. However, the fundamentals of the BEM and the FEM are the same, both of which are based on the method of modal analysis. Therefore, the BEM yields a significant prediction error in the high-frequency range, which has been verified by other researchers [31]. In summary, although the BEM is improved compared with the FEM, it still cannot accurately predict the full-spectrum train interior noise.

The SEAM is entirely different from the two methods above. It utilizes energy as a variable to describe the state of a system, and its core is energy conservation. By constructing the energy flow equation between the cavity subsystems, the energy stored in the relevant acoustic cavity subsystem is available, which can be used to predict the acoustic and dynamic responses. In 1997, Radcliff used the SEAM to predict the interior noise caused by automobile engines. In 2003, James predicted the aerodynamic noise of high-speed trains based on the SEAM, and the feasibility and accuracy of the SEAM for predicting the high-frequency train interior noise were verified by comparison with wind tunnel test results [32]. Givargis and Forssén used the SEAM to predict the high-speed train interior noise in their respective countries and proved the validity of the SEAM in predicting interior noise in high-speed trains [33, 34]. However, the accuracy of the SEAM depends on the damping coefficient matrix, which is difficult to obtain.

The methods above are numerical methods. Although they can yield more accurate results under specific conditions, the requirements for complex model parameters and a large amount of computing capacity render them challenging for practical applications. Recently, data-driven approaches have provided insight for different areas in railway transportation fields [35]. However, to our best

knowledge, studies regarding interior noise prediction based on data-driven methods are few.

3. Research Methodology

Herein, a data-driven interior noise prediction method is proposed, which is different from the traditional numerical model, as shown in Figure 2. Vehicle dynamics responses are key parameters to be considered in the design of URT alignments [36]. Generally, engineers evaluate the designed railway alignments with the simulated vehicle dynamic responses from the VTCDM. Subsequently, the interior noise can serve as a reference in the design of railway alignments using the forecasting model for the vehicle interior noise of the train.

4. Data Collection and Description

Figure 3 shows the setup of the field test when collecting data with an Android smartphone (Huawei Honor FRD-AL00). The smartphone was placed on the cabin floor above the bogie to measure the condition of the vehicle-track system when trains move along the track. The embedded inertia measurement unit (LSM6DS3, manufactured by STMicroelectronics) and microphone sensor (MP34DB02, manufactured by STMicroelectronics) of the smartphone were used to acquire the vibration acceleration, angular velocity, and audio data during the test. More detailed information about the smartphone and sensors is shown in Table 1. According to the comparative experiment, laying the smartphone on the floor directly without securement slightly affects the amplitude of the vertical acceleration, but it performs well in other directions and frequency domains, which have no significant impact on our analysis [16]. Additionally, in this study, the measure accelerations were transformed from the coordinate system of the smartphone to the vehicle for eliminating errors caused by different coordinate systems [37]. However, the accelerations in different positions within the car were different; therefore, in this study, the smartphone was placed in the same position to overcome this issue.

By reading the embedded sensors of the smartphone with the developed application, the data required, that is, seven types of signals, were readily available. The audio signal was acquired using the microphone sensors of the smartphone, and the dynamic responses of the car body including the vibration accelerations (horizontal, vertical, and longitudinal accelerations) and rotational angular velocities (pitch angular, yaw angular, and roll angular velocities) were available with the built-in inertial sensors. In our study, the sampling frequency of the audio signal was 22,050 Hz and that of all the inertial sensors was set as 100 Hz.

The sound pressure recorded by using the microphone differed from that perceived by the human ear (even in the same field) owing to various factors, such as psychological effects, presence of outer ear, and cochlear health status. To objectively reflect the passenger's hearing experience, we employed the A-weighting sound pressure level (SPL(A)) in this study to evaluate the interior noise. The SPL(A) can be calculated as follows:

$$\mathrm{SPL}(\mathrm{A}) = 20 * \log\left(\frac{P_{\mathrm{A}}}{P_{\mathrm{ref}}}\right), \tag{1}$$

where P_{A} is the A-weighting sound pressure (in Pa) and P_{ref} is the reference pressure (in Pa), which is typically set as 2×10^{-6} Pa. The data, including horizontal, vertical, and longitudinal accelerations and pitch angular, yaw angular, and roll angular velocities collected by the smartphone, were used to describe the dynamic responses of the car body. However, using only the data collected by the vehicle-carried smartphone may not be sufficient to explain the causes of the interior noise. For better prediction results, we considered the rotational angular accelerations of the car body obtained by calculating the first derivative of the rotational angular velocity. The effects of the trains' running speed on the generation of interior noise were nonnegligible. As a crucial parameter, the train speed was introduced to the prediction model. Because the subway tunnels are a GPS-free environment, the location and running speed information were not available through the GPS module embedded in the smartphones. The first-order integration of the longitudinal acceleration was used to overcome the problem. The running velocity of the train can be calculated by the following equation [37]:

$$v(t) = \int_0^t a_L \mathrm{d}x + v_0, \tag{2}$$

in which $v(t)$ corresponds to the running velocity of the train at time t, a_L is the longitudinal acceleration of the car body, and v_0 is the initial velocity of the train. Because the stopping state of the train can be easily recognized using a smartphone, v_0 is typically regarded as 0. Because the velocity cannot be validated directly, by regarding the interval length between adjacent stations as the ground truth, the error between the integral displacement and real interval length is 9.5%. [37]. The data collected and used in this model are presented in Table 2.

A disadvantage of using raw signals collected by smartphones is that they show track quality problems in only single points, which complicates the generation of a reasonable section length for optimization. Additionally, the sampling frequencies of the inertial sensors and microphone are different, which renders it challenging to build a point-to-point relationship directly. Hence, a moving time window method was employed in this study. For confirming the optimal time window, the size of the time window was varied from 0.5 to 10 s, and the hop length of the window was half the size of the time window. Additionally, a series of features was selected to reflect the characteristics of the signals inside the window. We selected the root mean square (RMS) of the SPL(A) in the windows as the index of the interior noise. As for the dynamic responses of the car body, the following features were used: (1) mean value, (2) RMS, (3) variance, (4) standard deviation, (5) peak value, (6) skewness, (7) kurtosis, (8) shape factor, (9) crest factor, and (10) clearance factor. In each frame of the dynamic response signals of the car body,

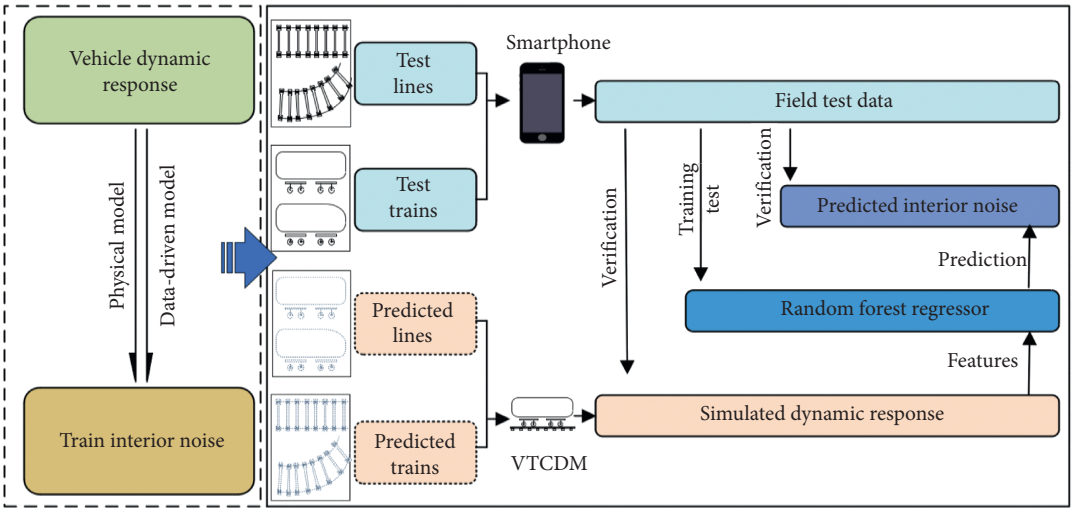


Figure 2: Research methodology.

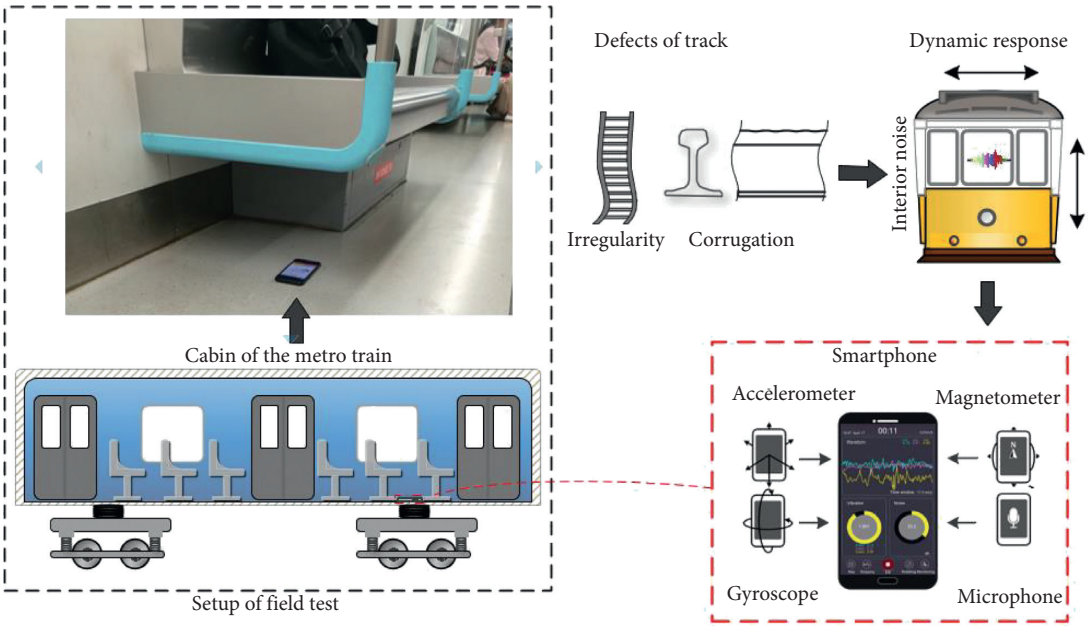


Figure 3: Data collection framework for this study.

Table 1: Parameters about the smartphone and sensors.

Device	Item	Parameter
Smartphone	Model	FRD-AL00
	Operating system	Android 8.0
	Chipset	HiSilicon Kirin 950
	CPU	Octa-core (4 × 2.3 GHz and 4 × 1.8 GHz)
	GPU	Mali-T880 MP4
	Memory	4 GB RAM
Inertial measurement unit	Model	LSM6DS3
	Acceleration range	±2/±4/±8/±16 g
	Angular rate range	±25/±250/±500 dps
	Analog supply voltage	1.71 V to 3.6 V
	Power consumption	1.25 mA

Table 1: Continued.

Device	Item	Parameter
Microphone sensors	Model	MP34DB02
	Frequency range	20 Hz to 20 kHz
	Supply voltage range	1.64 V to 3.6 V
	Acoustic overload point	120 dBSPL
	Signal-to-noise ratio	62.6 dB
	Sensitivity	−26 dBFS

Table 2: Features involved in the prediction model.

	Collected signal	Used signal	Selected features
1	Audio signal	SPL(A)	RMS
2	Horizontal acceleration	Horizontal acceleration	Mean value
3	Vertical acceleration	Vertical acceleration	RMS
4	Longitudinal acceleration	Longitudinal acceleration	Variance
5	Pitch angular velocity	Pitch angular velocity	Standard deviation
6	Yaw angular velocity	Yaw angular velocity	Peak value
7	Roll angular velocity	Roll angular velocity	Skewness
8		Pitch angular acceleration	Kurtosis
9		Yaw angular acceleration	Shape factor
10		Roll angular acceleration	Crest factor
11		Operating velocity of trains	Clearance factor

these features were calculated as input parameters for our prediction model.

5. Modeling Approach

5.1. RF. RF is an ensemble learning approach comprising hundreds of decision trees for performing classification or regression tasks independently. Using the average of all the trees' results as the final output can significantly improve the predictive accuracy. RF was developed based on the decision tree structure; however, two additional characteristics, including bagging and random subspace methods, were added to improve the accuracy and robustness. First, by creating a series of bootstrap samples, the bagging algorithm is fundamental for improving the accuracy and stability of machine-learning models. The introduction of the bagging method helps reduce variance and control overfitting. Next, the random subspace method is designed to increase tree independence by generating trees with a random sample of features rather than the entire feature set [38]. Additionally, the collected raw data can be directly fed into the RF model without any preprocessing [39], which renders the RF method easier to implement. Furthermore, the RF method can be used for feature importance analysis.

As a data-driven method, RF relies more on a tremendous amount of data to perform interior noise prediction. Compared with mathematical and physical models, data-driven prediction models do not require complex parameters, strict conditions, or hypotheses. Furthermore, this method can potentially reduce computational time and require less computer memory compared with numerical models. The mapping relationship between the dynamic responses of the car body and interior noise exhibits a strong nonlinear characteristic. Because the RF is a combination of a series of decision trees, it exhibits an outstanding performance for fitting nonlinear relationships. Therefore, RF was selected in this study to fit the mapping relationship between the dynamic responses of the car body and the interior noise.

In RF, one dependent variable exists, that is, the RMS of SPL(A) for each moving window. Moreover, 100 independent variables (10 dynamic response signals × 10 selected features; 10 dynamical response signals are items 2 to 11 of Column 3, Table 2; 10 selected features are the items 2 to 11 of Column 4, Table 2) were input to the model. In the model study, the effects of the number of decision trees, window size, and modes of the maximum amount of features were investigated. The computational time of different max feature modes was also considered. Using the feature selection method provided by RF, we performed the feature importance analysis.

5.2. VTCDM. Using RF, we established the mapping relationship between the dynamic response of the car body and the subway vehicle interior noise. However, it is difficult to exert its value using only such a mapping relationship. To expand its application scenario, we propose applying the mapping relationship for evaluating the design of the URT railway alignments based on the simulated dynamic responses of the vehicles. Hence, a VTCDM was constructed, with which the required simulated dynamic response signals could be obtained.

Vibrations of the vehicle can be transmitted to the track via the wheel-rail contact and cause vibrations of the track structure. However, the vibrations of the vehicle are also affected by the track in reverse [40]. Therefore, the vibrations of the track and the vehicle couple with each other, which allows us to evaluate the track state with the dynamic

responses of the train. Using the parameters of a type A subway train, we established the vehicle submodel in UM, which comprised 15 rigid bodies, such as the car body, bogies, wheelsets, and axle boxes. Moreover, the detailed parameters of the type A subway train used in our model are shown in Table 3. In the model, the flexibility and mass of the rail were ignored. Additionally, the rail was regarded as a massless block connected to the foundation with spring and damping force elements. Such a track model could be used to analyze the dynamic response in a low-frequency range. The effect of high-frequency vibration on the wheel-rail contact behavior was neglected. However, this method provided a high calculation efficiency that satisfied the general design requirements. In this study, the fifth-grade American spectrum was adapted as an excitation to input the model. Finally, a nonelliptical contact model was used to couple the two subsystems.

Table 3: Key parameters of type A subway vehicle.

	Parameters	Value
1	Length between bogie pivot centers (m)	15.7
2	Wheelbase (m)	2.5
3	Lateral distance of the rolling circle (m)	1.493
4	Nominal rolling circle radius (m)	0.84
5	Mass of car body (kg)	50 877
6	Mass of bogie (kg)	2 721.5
7	Mass of wheel (kg)	1900
8	Wheel tread type	LMA

6. Results and Discussion

After RF and the VTCDM were established, a series of analyses were implemented. First, the performance of the RF regression model was analyzed in terms of parameter tuning, feature importance, and error distribution. Next, the VTCDM and the interior noise prediction model were verified through the field test data. Subsequently, an application scenario for evaluating the railway alignments from the perspective of interior noise was presented.

6.1. Analysis of the RF Regression Model

6.1.1. Parametric Analysis. To obtain the optimal parameters of RF, the effects of the number of decision trees, size of time windows, and mode of the max feature numbers on the performance were studied. It is noteworthy that the Out-of-Bag R-squared (OOB R^2) was employed as a critical parameter to evaluate the performance of the RF-based regression model [41–43]. Figure 4(a) presents the effects of the number of decision trees on the performance of this forecasting model with different time windows. As one can see, OOB R^2 increases gradually as the number of trees increases. However, when there are more than 150 trees, OOB R^2 remains stable. The relationship between OOB R^2 and the size of time windows is presented in Figure 4(b). It is indicated that, as the window size changes from 0.5 to 10 s, OOB R^2 shows an overall downward trend, and when the window size is 0.5 s, OOB R^2 reaches its maximum value of 0.776. Meanwhile, local fluctuations occur when the window size is 4 s.

The number of features considered when forming the random forests significantly affects the goodness of fit. Three max feature modes, "auto," "sqrt," and "log2," are discussed from the perspective of computation time and goodness of fit, as shown in Figure 4(c). If the "auto" mode was selected, then the number of features to consider when forming the trees was equal to the number of all features input to the model. If "sqrt" was selected, the number of features considered was the square root of the number of all features. When mode "log2" was selected, the number of features considered was the base-two logarithm of the number of total features. In terms of goodness of fit, the "auto" mode performed the best, followed by "sqrt" and "log2." Furthermore, the calculation time of the three methods increased linearly with the increase in the number of decision trees. Under the same condition, the "auto" mode consumed the most time, whereas the "sqrt" mode consumed slightly more time than the "log2" mode. It is clear from the analysis that although the "auto" mode consumed the most time, its fitting effect was the best; the "log2" mode computed faster than the other two methods but yielded the worst performance.

From the analysis above, the optimal parameters of the RF were confirmed. The RF with 200 trees and "auto" max feature mode was selected in our study to obtain better fitting results with less computing capacity. Figure 4(d) shows the comparison between the predicted interior noise by RF and the field test results. This figure shows that the interior noise by RF agrees well with the field test results. However, when the train stopped at the stations, significant errors occurred therein. This was because when the train stopped or started, the primary source of interior noise of the subway vehicles changed from wheel-rail noise to electromechanical equipment noise and broadcast sound.

6.1.2. Prediction Error Distribution Analysis. Additionally, we studied the distribution characteristics of prediction errors of the RF regression model (RFRM). The prediction error distribution of the RFRM was compared with that of the linear regression model (LRM), radial-basic-function-based support vector regression model (rbf-SVRM), and gradient boosting regression model (GBRM). Figure 5(a) shows the density function curves of prediction errors of different models. Figure 5(b) shows the cumulative distribution function curves of different models. First, the figure shows that the prediction errors of the RF have the highest occurrence probability in the area nearby 0, which indicates that the RFRM is more accurate than the other models, that is, GBRM, rbf-SVRM, and LRM, in that order. For the RFRM, prediction errors within 5 and 10 dB(A) constitute 86.9% and 98.1% of the 13,000 test samples, respectively. It is noteworthy that the accuracy of this prediction model requires further improvement.

6.1.3. Feature Importance Analysis. The essence of feature importance assessment in RF is to calculate the average of each feature's contribution to all trees in the forest and subsequently compare the contributions among those

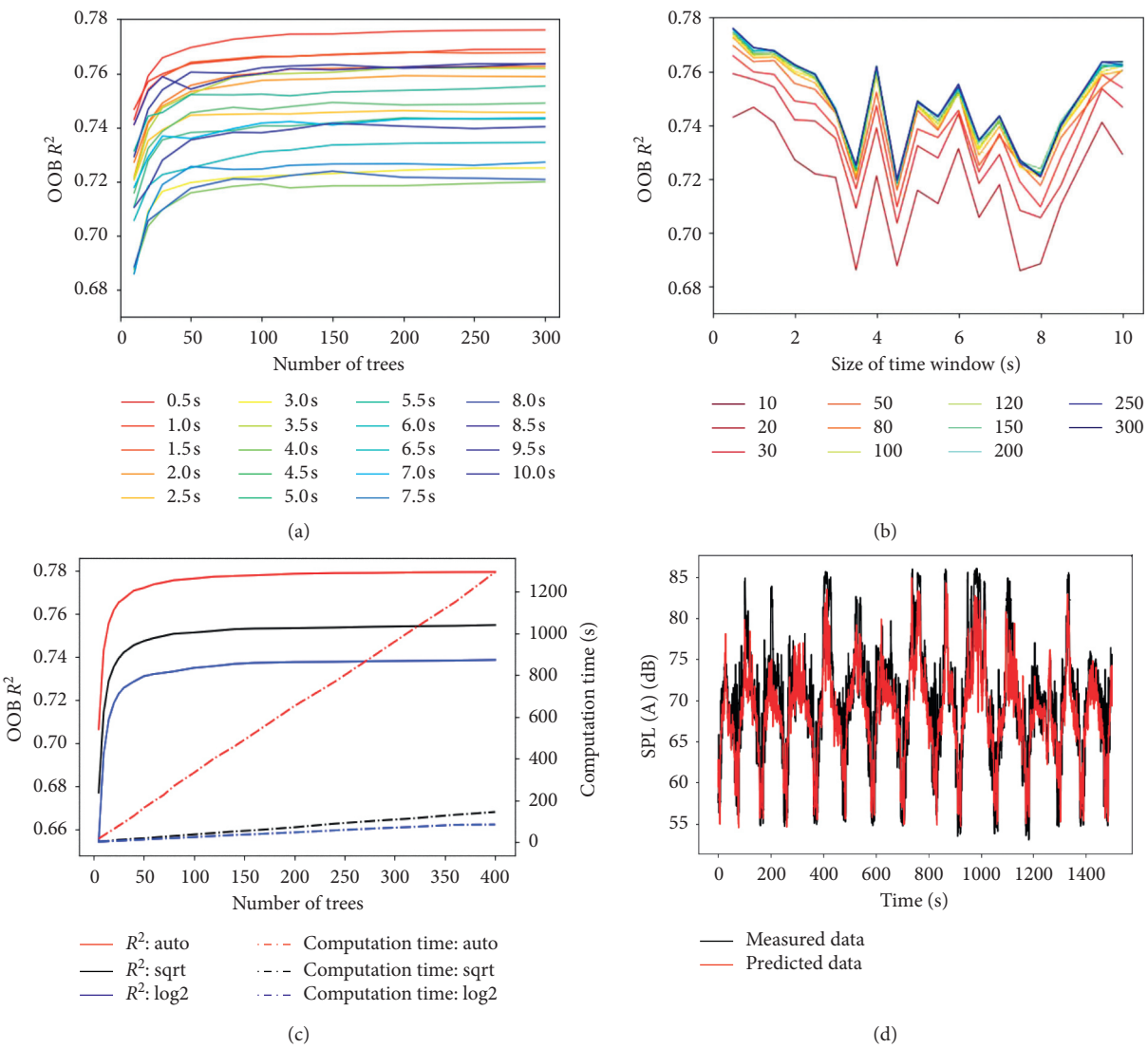


Figure 4: Effect of number of estimators on OOB R^2.

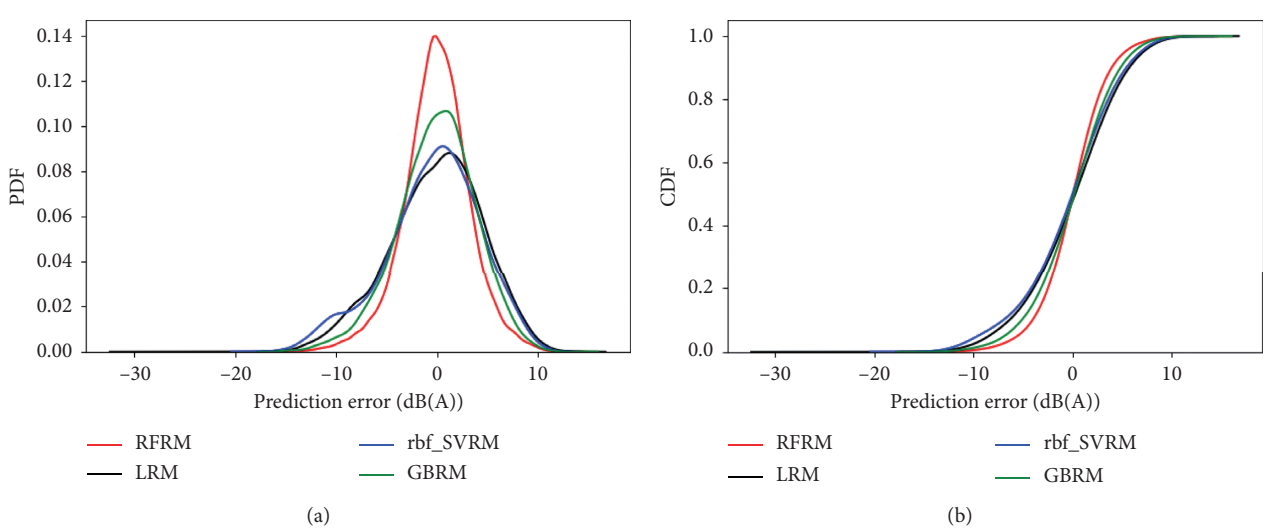


Figure 5: Comparison of the prediction error distributions.

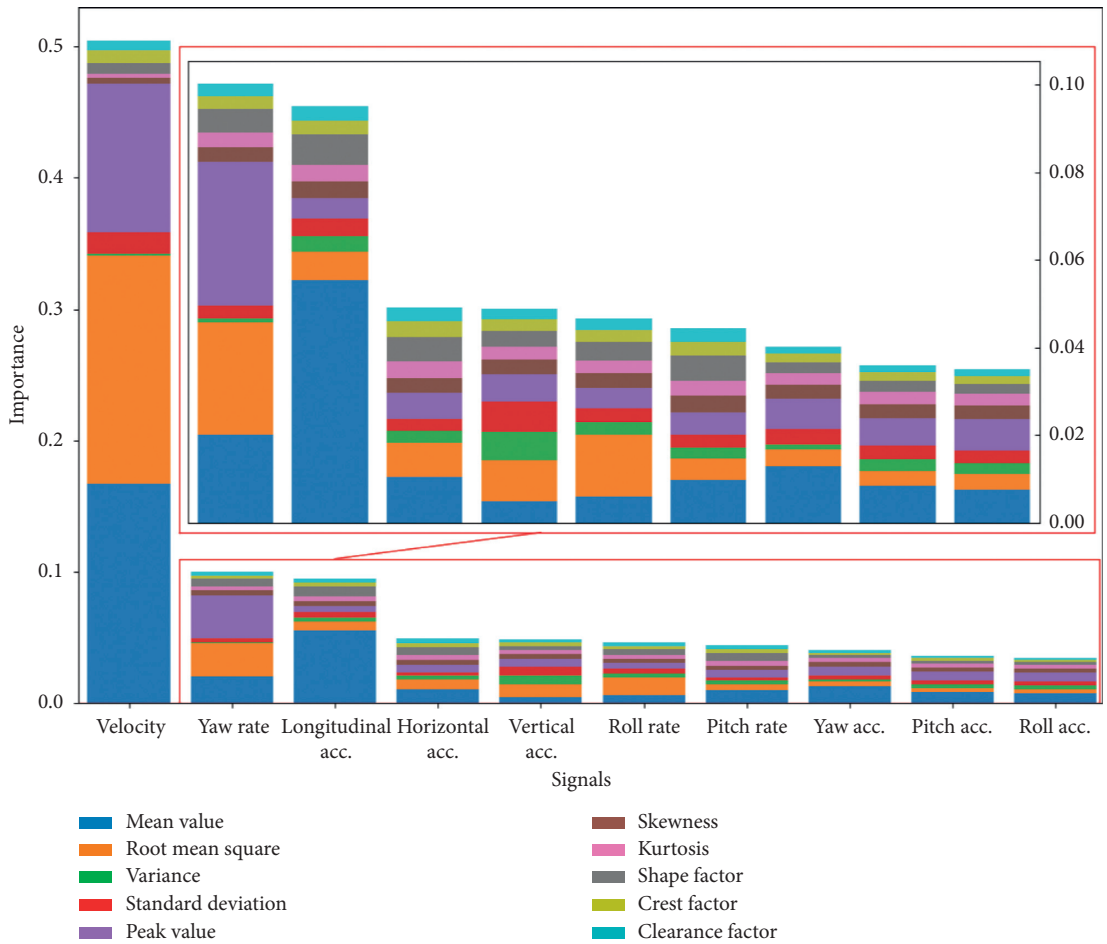
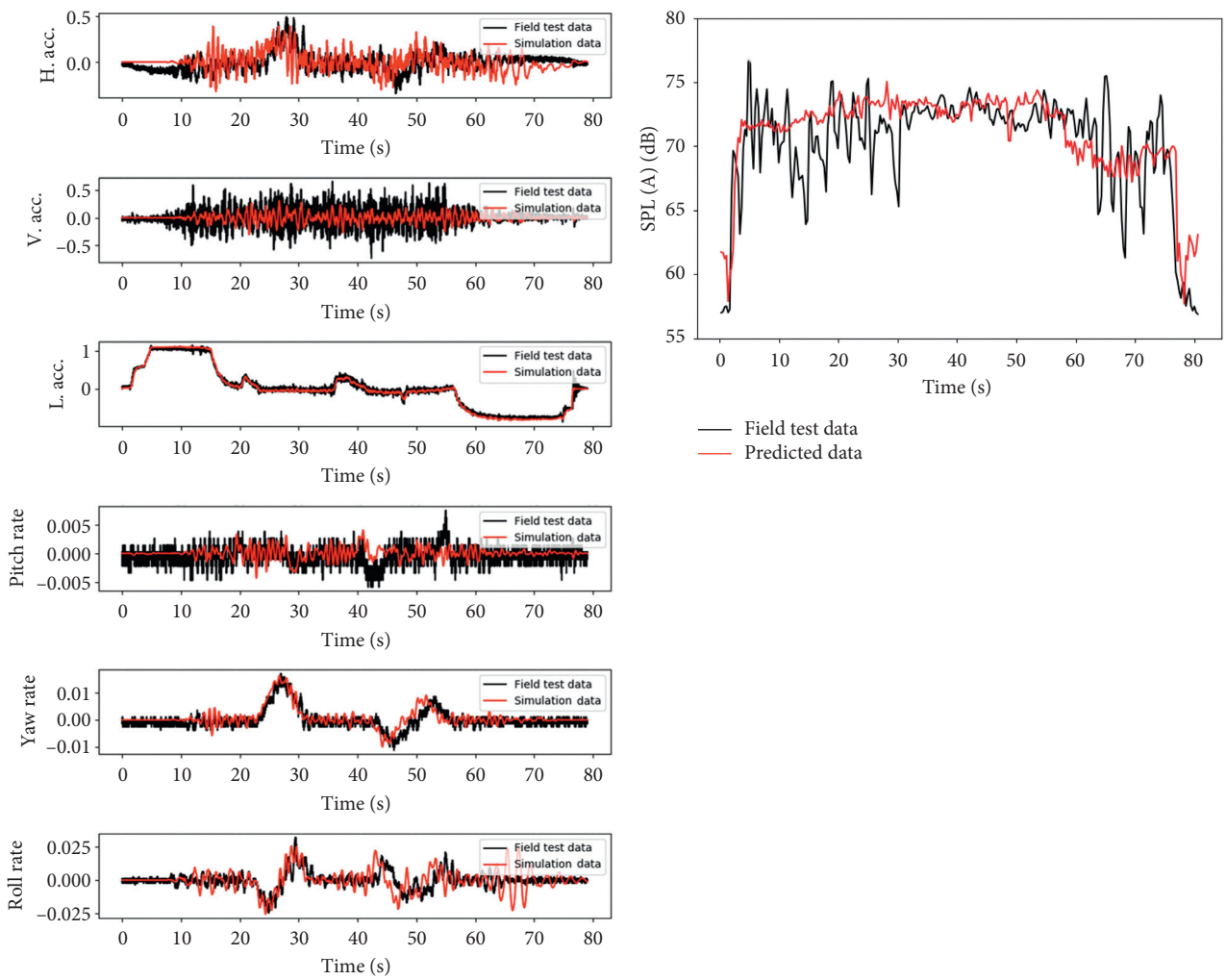


Figure 6: Feature importance analysis.

features. The mean decrease impurity method was adopted in this study to perform feature importance analysis and variance was used as the impurity measure [44]. The results of the feature importance analysis are presented in Figure 6. For showing the specific details clearly, an enlarged view of the feature importance of the partial signals is included. A total of 100 features (10 signals × 10 indexes) are presented, and the summary of these features' importance coefficient is 1. Figure 6 shows that the importance of velocity is the highest, reaching nearly 0.5. This suggests that the running speed is the most crucial factor affecting vehicle interior noise. The importance of yaw rate and longitudinal acceleration ranks second and third with 0.10 and 0.096, respectively. This is because the yaw rate relates well with the curve radius, where rail corrugation and squeal often occur. Among all the indexes, the mean value, RMS, and peak value are the most important ones.

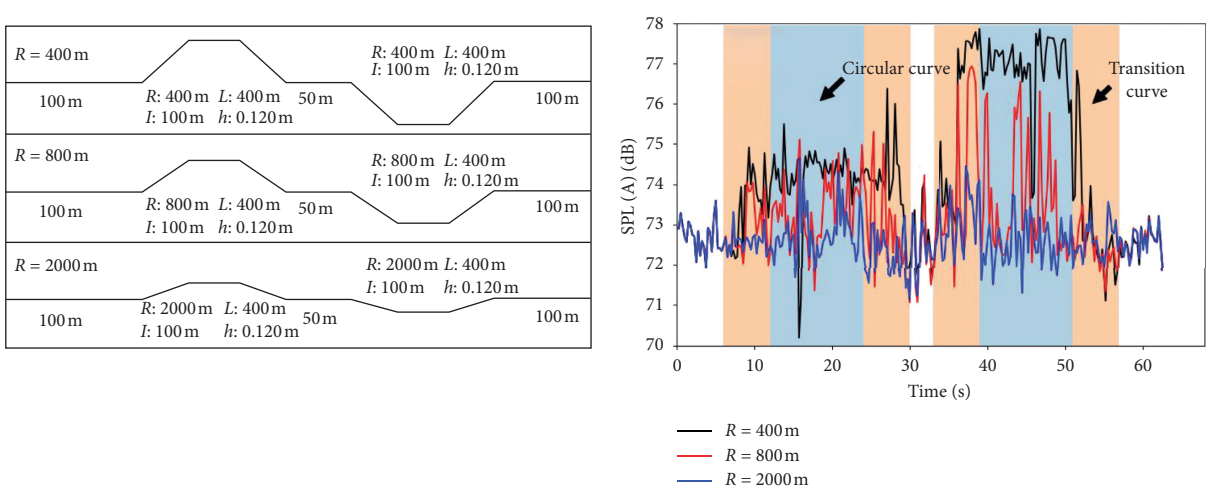
6.2. Verifications of VTCDM and the Interior Noise Prediction Model. Not only the RF but also the VTCDM significantly affects the performance of the interior noise prediction model. In Figure 7(a), the simulated dynamic responses, including the horizontal acceleration (H. acc.), vertical acceleration (V. acc.), longitudinal acceleration (L. acc.), pitch rate, yaw rate, and roll rate, are compared with the field measured data. The figure shows that the simulated data agree well with the measured data. However, the simulated vertical acceleration of the car body is lower than the measured data, which may be caused by the simplification of the car body into a rigid body. The comparison between the measured and forecasted vehicle interior noise is shown in Figure 7(b). The predicted SPL(A) by our model is similar to the measured data. However, our model cannot predict the fluctuation of the interior noise caused by the broadcast sound when the train starts and brakes.

6.3. Case Study for Optimizing Railway Alignment Designs. We developed an application case of the interior noise prediction model for evaluating railway alignments in the design stage. In this study, we designed three types of railway alignments, as shown in Figure 8(a). These three railway alignments have the same parameters except the radius of the circular curve. The radii of the circular curve were 400, 800, and 2000 m, which are typical in subway lines. Subsequently, the designed railway alignments were input to the VTCDM for obtaining the simulated dynamic responses of the car body. The simulated train moved at a constant speed of 60 km/h under the three alignments above. Subsequently,



(a) (b)

Figure 7: Verification of the VTCDM and interior noise prediction model.



(a) (b)

Figure 8: Evaluation of railway alignments considering interior noise.

the predicted interior noise (SPL(A)) was obtained by inputting the simulated dynamic responses to the trained RF.

In Figure 8(b), the predicted interior noise under different railway alignments is presented (the black line refers to the SPL(A) with circular curves of radius 400 m; the red line denotes the SPL(A) with circular curves of radius 800 m; the blue line represents the SPL(A) with circular curves of radius 2000 m). The findings obtained from the figure are as follows: first, the predicted SPL(A) in the curved segments, especially in the circular sections, is higher than that in the straight-line segments; next, the predicted SPL(A) increases as the radius of the curve decreases. Therefore, sharp curves must be avoided in the design of railway alignments for reducing the interior noise of the train. The purpose of this case study is to verify the feasibility of using this interior noise prediction model for evaluating railway alignments. Regulations for designing rail alignments consider more about the economy and the running stability of trains, such as the lateral acceleration [45–47], whereas, as one of the critical indices for the ride comfort, the noise level is not sufficiently considered in designing rail alignments. These regulations usually provide a general limit value of noise level, which cannot be directly applied to designing rail alignments. Comparing the predicted noise levels corresponding to different alignments with those limit values allows us to evaluate rail alignments from the perspective of noise. The combination of the predicted noise level and regulations in force is able to make the designed rail alignments more reasonable.

In this case, we did not consider different vehicles, as only one type of vehicle was selected for a specific subway line in general. For simplifying the model of the cases, the impact of travel speeds of the train on the noise level was not considered, whereas in practical operation, trains usually repeatedly go through the process of starting-speeding up—keeping a high travel speed—speeding down-stopping. In the section with small-radius curves, the higher the speed of the train is, the more likely it causes noise issues, such as squeal. Therefore, the sections where trains have a high travel speed should avoid small-radius curves or reverse curves as much as possible in the design of rail alignments.

7. Conclusions

The interior noise of subway vehicles based on RF and VTCDM was forecasted herein using data collected from onboard smartphones. Through parameter study, we confirmed the optimal parameters of the RF model: number of trees, 200; size of time window, 0.5 s; "auto" max feature mode selected. With such parameters, R^2 reached its highest value, 0.78. Error analysis showed that RF had a higher prediction accuracy than the other models. The probability of a prediction error within 5 dB was 86.9%.

Feature importance analysis demonstrated that the running velocity of the train affected the interior noise the most. Additionally, the yaw rate and longitudinal acceleration of the car body affected the vehicle interior noise significantly. The effectiveness of the VTCDM was verified through comparisons between the simulated and measured dynamic responses of the car body. From the comparison with the field-test vehicle interior noise, it was evident that the proposed prediction method could accurately predict the trend of subway train interior noise. However, the local fluctuations of the interior noise caused by some nonwheel-rail noise, such as the sounds of broadcast and electrical devices, could not be forecasted correctly. The case study demonstrated that the proposed interior noise forecasting method could be used for evaluating and optimizing railway alignment designs in the early stage from the perspective of interior noise.

Next, to improve the performance of the prediction model, we could categorize the training data in detail according to different characteristics, such as track type, train type, and service time of the track system. Because the simplification of the vehicle track into a multirigid-body system in the current model reduced the forecasting accuracy, the flexibility of the car body and track shall be considered in future studies.

Data Availability

Noise data will be available upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was funded by the China Scholarship Council (ID: 201907000077) and National Natural Science Foundation of China (Grant no. 51878576).

References

[1] J. Zhang, X. Xiao, X. Sheng, Z. Li, and X. Jin, "A systematic approach to identify sources of abnormal interior noise for a high-speed train," *Shock and Vibration*, vol. 2018, Article ID 5085847, 12 pages, 2018.

[2] B. T. Atmaja, M. F. Puabdillah, M. N. Farid, and W. A. Asmoro, "Prediction and simulation of internal train noise resulted by different speed and air conditioning unit," *Journal Of Physics*, vol. 1075, Article ID 012038, 2018.

[3] W. Dai, X. Zheng, L. Luo, Z. Hao, and Y. Qiu, "Prediction of high-speed train full-spectrum interior noise using statistical vibration and acoustic energy flow," *Applied Acoustics*, vol. 145, pp. 205–219, 2019.

[4] J. Han, X. Xiao, Y. Wu, Z. Wen, and G. Zhao, "Effect of rail corrugation on metro interior noise and its control," *Applied Acoustics*, vol. 130, pp. 63–70, 2018.

[5] D. J. Thompson, G. Squicciarini, B. Ding, and L. Baeza, "A state-of-the-art review of curve squeal noise: phenomena, mechanisms, modelling and mitigation," *Noise and Vibration Mitigation for Rail Transportation Systems*, vol. 3–41, 2018.

[6] J. Zhang, G. Han, X. Xiao, R. Wang, Y. Zhao, and X. Jin, "Influence of wheel polygonal wear on interior noise of high-speed trains," *Journal of Zhejiang University Science*, vol. 15, no. 12, pp. 1002–1018, 2014.

[7] H. Tsunashima, Y. Naganuma, A. Matsumoto, T. Mizuma, and H. Mori, "Condition monitoring of railway track using

in-service vehicle," *Reliability and Safety in Railway*, vol. 12, pp. 334–356, 2012.

[8] Y. Hayashi, T. Kojima, H. Tsunashima, and Y. Marumo, "Real time fault detection of railway vehicles and tracks," in *Proceedings of the Institution of Engineering & Technology International Conference on Railway Condition Monitoring*, Birmingham, UK, 2007.

[9] M. Odashima, S. Azami, Y. Naganuma, H. Mori, and H. Tsunashima, "Track geometry estimation of a conventional railway from car-body acceleration measurement," *Mechanical Engineering Journal*, vol. 16, Article ID 00498, 2017.

[10] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[11] Y. Yu, R. Han, X. Zhao et al., "Initial validation of mobile-structural health monitoring method using smartphones," *International Journal of Distributed Sensor Networks*, vol. 11, no. 2, Article ID 274391, 2015.

[12] S. H. de Frutos and M. Castro, "Using smartphones as a very low-cost tool for road inventories," *Transportation Research Part C: Emerging Technologies*, vol. 38, pp. 136–145, 2014.

[13] A. Paixão, E. Fortunato, and C. Rui, "Smartphone's sensing capabilities for on-board railway track monitoring: structural performance and geometrical degradation assessment," *Advances in Civil Engineering*, vol. 2019, Article ID 1729153, 13 pages, 2019.

[14] S. Kanarachos, S.-R. G. Christopoulos, and A. Chroneos, "Smartphones as an integrated platform for monitoring driver behaviour: the role of sensor fusion and connectivity," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 867–882, 2018.

[15] A. Ghose, P. Biswas, C. Bhaumik, M. Sharma, A. Pal, and A. Jha, "Road condition monitoring and alert application: using in-vehicle smartphone as internet-connected sensor," in *Proceedings of the 2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, IEEE, Lugano, Switzerland, pp. 489–491, March 2012.

[16] P. Wang, Y. Wang, Li Wang, R. Chen, and J. Xiao, "Measurement of carbody vibration in urban rail transit using smartphones," 2017.

[17] H.-M. Noh, "Contribution analysis of interior noise and floor vibration in high-speed trains by operational transfer path analysis," *Advances in Mechanical Engineering*, vol. 9, no. 8, Article ID 1687814017714986, 2017.

[18] H. B. Huang, X. R. Huang, M. L. Yang, T. C. Lim, and W. P. Ding, "Identification of vehicle interior noise sources based on wavelet transform and partial coherence analysis," *Mechanical Systems and Signal Processing*, vol. 109, pp. 247–267, 2018.

[19] J. Zhang, X. B. Xiao, X. Z. Sheng, R. Fu, D. Yao, and X. S. Jin, "Characteristics of interior noise of a Chinese high-speed train under a variety of conditions," *Journal of Zhejiang University-Science A*, vol. 18, no. 8, pp. 617–630, 2017.

[20] Y. Sun and Y. Zhao, "Characteristics of interior noise in MonoRail and noise control," in *Proceedings of the Inter-Noise and Noise-Con Congress and Conference*, vol. 258, Institute of Noise Control Engineering, Chicago, IL, USA, pp. 1461–1467, August 2018.

[21] K. Hu, Y. Wang, H. Guo, and H. Chen, "Sound quality evaluation and optimization for interior noise of rail vehicle," *Advances in Mechanical Engineering*, vol. 6, p. 820875, 2014.

[22] L. G. Kurzweil and R. Lotz, *Prediction and Control of Noise and Vibration in Rail Transit Systems*, Urban Mass Transportation Administration, Washington, DC, USA, 1978.

[23] W. U. Di and J. Ge, "Analysis of the influence of racks on high speed train interior noise using finite element method," *Applied Mechanics & Materials*, vol. 675–677, pp. 257–260, 2014.

[24] L. Franzoni, J. Rouse, and T. Duvall, "A broadband energy based boundary element method for predicting vehicle interior noise," *The Journal of the Acoustical Society of America*, vol. 115, no. 5, p. 2538, 2004.

[25] G. M. L. Gladwell and G. Zimmermann, "On energy and complementary energy formulations of acoustic and structural vibration problems," *Journal of Sound and Vibration*, vol. 3, no. 3, pp. 233–241, 1966.

[26] S. H. Sung and D. J. Nefske, "A coupled structural-acoustic finite element model for vehicle interior noise analysis," *Journal of Vibration and Acoustics*, vol. 106, no. 2, pp. 314–318, 1984.

[27] C. Lucas, D. Shahmirzadi, and M. N. Bahrami, "A FEM-based quasi-static neuro-model for acoustic noise in switch reluctance Motors," *Journal of Computational Acoustics*, vol. 12, no. 1, pp. 85–97.

[28] E. Ghafoorifard, M. Claessens, and X. Hu, "An analytical-experimental method for analysing the low-frequency interior acoustics of a passenger car," in *Proceedings of the International Seminar on Modal Analysis*, vol. 3, KU Leuven, Orlando, FL, USA, pp. 1331–1338, February 1998.

[29] F. Létourneaux, S. Guerrand, and F. Poisson, "Low-frequency acoustic transmission of high-speed trains: simplified vibroacoustic model," *Journal of Sound and Vibration*, vol. 231, no. 3, pp. 847–851, 2000.

[30] A. Soltani and M. K. Demmeh, "Analyzing of noise inside a simple vehicle cabin using boundary element method," *World Academy of Science, Engineering and Technology*, vol. 49, pp. 630–634, 2011.

[31] R. H. Lyon and G. Maidanik, "Power flow between linearly coupled oscillators," *The Journal of the Acoustical Society of America*, vol. 34, no. 5, pp. 623–639, 1962.

[32] B. James, H. Andrew, and D. Sam, "Aero-acoustic tests and analysis to predict internal and external railway vehicle aerodynamically-generated noise," in *Proceedings of the Tenth International Congress on Sound and Vibration*, Stockholm International Institute of Acoustics and Vibration, Stockholm, Sweden, pp. 1697–1704, July 2003.

[33] S. Givargis and H. Karimi, "Mathematical, statistical and neural models capable of predicting LA,max for the Tehran-Karaj express train," *Applied Acoustics*, vol. 70, no. 7, pp. 1015–1020, 2009.

[34] J. Forssén, S. Tober, A. C. Corakci, and A. Frid, "Modelling the interior sound field of a railway vehicle using statistical energy analysis," *Applied Acoustics*, vol. 73, no. 4, pp. 307–311, 2012.

[35] F. Kropp, Q. He, R. M. P. Goverde, and X. Liu, "Recent applications of big data analytics in railway transportation systems: a survey," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 226–246, 2018.

[36] X. Lai and S. Paul, "Optimization of rail transit alignments considering vehicle dynamics," *Transportation Research Record*, vol. 2275, no. 2275, pp. 77–87, 2012.

[37] Y. Wang, J. Cong, H. Tang, X. Liu, T. Gao, and P. Wang, "A data fusion approach for speed estimation and location calibration of a metro train in underground environment based on low-cost sensors in smartphones," *IEEE Sensors Journal*, vol. 19, no. 22, pp. 10744–10752, 2019.

[38] A. Falamarzi, S. Moridpour, M. Nazem, and S. Cheraghi, "Development of random forests regression model to predict track degradation index: melbourne case study," in

Proceedings of the Australasian Transport Research Forum (ATRF), 40th, 2018, Darwin, Australia, October 2018.

[39] Y. Li, C. Zou, M. Berecibar et al., "Random forest regression for online capacity estimation of lithium-ion batteries," *Applied Energy*, vol. 232, pp. 197–210, 2018.

[40] W. Zhai, K. Wang, and C. Cai, "Fundamentals of vehicle-track coupled dynamics," *Vehicle System Dynamics*, vol. 47, no. 11, pp. 1349–1376, 2009.

[41] U. Grömping, "Variable importance assessment in regression: linear regression versus random forest," *The American Statistician*, vol. 63, no. 4, pp. 308–319, 2009.

[42] D. Liakhovitski, Y. Bryukhov, and M. Conklin, "Relative importance of predictors: comparison of random forests with johnson's relative weights," *Model Assisted Statistics and Applications*, vol. 5, no. 4, pp. 235–249, 2010.

[43] M. Vreugdenhil, W. Wagner, B. Marschallinger, I. Pfeil, I. Teubner, and C. Rüdiger, "Sensitivity of sentinel-1 backscatter to vegetation dynamics: an Austrian case study," *Remote Sensing*, vol. 10, no. 9, p. 1396, 2018.

[44] G. Strauss, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," *Advances In Neural Information Processing Systems*, vol. 1, pp. 431–439, 2013.

[45] EU Commission, "Commission regulation (EU) no 1304/2014," 2014, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32014R1304.

[46] GB 50157-2013 and Mohurd, *Code for Design of Metro*, Scientific Research, Wuhan, China, 2013.

[47] GB 14892-2006 and Mohurd, *Noise Limit and Measurement for Train of Urban Rail Transit Standards*, Scientific Research, Wuhan, China, 2006.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 6751728, 10 pages
https://doi.org/10.1155/2020/6751728

Research Article

A Gradient Boosting Crash Prediction Approach for Highway-Rail Grade Crossing Crash Analysis

Pan Lu,[1] **Zijian Zheng**,[2] **Yihao Ren**,[3] **Xiaoyi Zhou**,[3] **Amin Keramati**,[3] **Denver Tolliver**,[3] **and Ying Huang**[4]

[1]*Department of Transportation, Logistics, and Finance, Upper Great Plains Transportation Institute, North Dakota State University NDSU, Dept 2880, P. O. Box 6050, Fargo, ND 58108-6050, USA*
[2]*Gates Corporation, 1144 Fifteenth St. Suite 1400, Denver, CO 80202, USA*
[3]*Upper Great Plains Transportation Institute, North Dakota State University NDSU, Dept 2880 P. O. Box 6050, Fargo, ND 58108-6050, USA*
[4]*Department of Civil and Environmental Engineering, North Dakota State University NDSU, Dept 2470, P. O. Box 6050, Fargo, ND 58108-6050, USA*

Correspondence should be addressed to Pan Lu; pan.lu@ndsu.edu

Received 13 February 2020; Revised 28 April 2020; Accepted 25 May 2020; Published 19 June 2020

Academic Editor: Lu Gao

Highway-rail grade crossing (HRGC) crashes continue to be the major contributors to rail causalities in the United States and have been intensively researched in the past. Data-mining models focus on prediction while dominant general linear models focus on model and data fitness. Decision makers and traffic engineers rely on prediction models to examine at-grade crash frequency and make safety improvement. The gradient boosting (GB) model has gained popularity in many research areas. In this study, to fully understand the model performance on HRGC accident prediction performance, the GB model with functional gradient descent algorithm is selected to analyze crashes at highway-rail grade crossings (HRGCs) and to identify contributor factors. Moreover, contributors' importance and partial-dependent relations are generated to further understand the relationship of identified contributors and HRGC crash likelihood to concur "black box" issues that most machine learning methods face. Furthermore, to fully demonstrate the model's prediction performance, a comprehensive model prediction power assessment based on six measurements is conducted, and the prediction performance of the GB model is verified and compared with a decision tree model as a reference due to their popularity and comparable data availability. It is demonstrated that the GB model produces better prediction accuracy and reveals nonlinear relationships among contributors and crash likelihood. In general, HRGC crash likelihood is significantly impacted by several traffic exposure factors: highway traffic volume, railway traffic volume, and train travel speed and others.

1. Introduction

Crashes between motor vehicles and trains at highway-rail grade crossings (HRGCs) often have severe consequences [1]. Of all crashes at HRGCs in the U.S. (2000 to 2014), 12% resulted in fatalities [2]. Numerous models have been developed to identify major contributing factors and explore relationships between crashes and explanatory variables to better understand safety performance and be able to apply effective countermeasures to reduce crash rates at HRGCSs.

Since crash data have random, discrete, and nonnegative characteristics, generalized linear models (GLMs) [3] have been commonly selected to investigate the relationship between crashes and contributing factors. However, Lord and Mannering [4] pointed out that these models face various data challenges which stem from crash data distribution and inappropriately fitted GLMs. As indicated by Lu and Tolliver [5] and Oh et al. [6], HRGC crash data often show underdispersion distribution where sample variance is less than the sample mean, and less common GLMs are

suitable for such datasets. Moreover, the available crash dataset often contains a large portion of missing data and outliers. GLMs are very sensitive to noise data [7].

In this study, to fully demonstrate the model application and its capabilities to analyze safety data, a robust data-mining technique, the gradient boosting (GB) model is selected to analyze crashes at HRGCs. Unlike GLMs, it requires no predefined underlying relationship between dependent and independent variables. Thus, underdispersed HRGC data are not an issue. Moreover, to better understand the model forecasting performance, a comprehensive model forecasting accuracy evaluation system including six measurements is proposed and evaluated.

2. Literature Review

GLMs or other statistical models have been commonly adopted by transportation safety decision makers and researchers to capture relationships among many factors to allow assessment of transportation safety risk. Poisson and negative binomial (NB) models have been widely applied in crash frequency studies [8–14]. Zero-inflated Poisson and zero-inflated negative binomial models have been developed as extensions of Poisson and NB models to try to overcome poor model performance for rare-event data [6, 15–18]. However, these models still need to meet the data distribution assumptions. As indicated in the literature, these parametric regression models have severe limitations [19–22]. Researchers have difficulty upholding the various required assumptions for these models in many applications. Failure to satisfy these assumptions results in numerous errors. Lu and Tolliver [5] stated that underdispersion exists in HRGC crash data. However, the negative binomial model as the most popular GLM models assumes overdispersion distribution. The gamma model [6], Conway–Maxwell–Poisson, and Bernoulli [5] are recommended to handle underdispersed crash data. However, those models are hard to implement because of the model's complexity. Moreover, all GLMs assume linear relationship between the transformed response in terms of the link function and the explanatory variables, which make it hard to capture the real dynamic nonlinear relationship.

Tree-based machine learning models and other nonparametric methods were recently recognized by transportation safety researchers [23], and Karlaftis and Golias [21] conducted safety research to model crash frequency on rural roads with a decision tree model. Yan and Radwan [24] studied the influential factors of rear-end crashes. Qin and Han [25] classified intersection crashes. Yan et al. [26] applied a decision tree (DT) model to predict train-vehicle crashes at HRGCs. Keramati et al. [27] adopted a survival analysis to evaluate geometric effects on HRGC safety performance. All the existing literature adopted GB as a method to conduct safety research only focuses on one part of the application; some focuses on its capability to better handle the missing value and to provide better overall accuracy [28], and others only focus on its simplified version, the DT method, to produce the interpretable relationship [24, 26].

A nonparametric tree-based model, the gradient boosting model, is used to examine the relationship between HRGC crashes and the contributing factors. The GB model is extremely powerful in understanding the structure of complex datasets and exploring potential relationships between dependent variables and independent variables. GB models are widely used in various transportation research areas [29, 30] and [31]. Unlike linear models, the GB model requires no statistical assumptions. The GB model is believed to be superior to simple DT models because of its techniques for handling missing data, robustness with data noise, and resistance to overfitting [32]. In this study, the authors tend to conduct a complete application and performance assessment to demonstrate the GB model's full learning and prediction aspects including (1) building the predictive model from the data, (2) producing interpretable relationship, and (3) demonstrating sound prediction power through a complete forecasting performance analysis.

3. Methodology

3.1. Data. Data used in this study came from two public sources: (1) the FRA's (Federal Railroad Administration) Office of Safety accident/incident database and (2) the FRA's Office of Safety highway-rail crossing inventory. A new combined database is generated based on HRGC identification numbers in both databases, and it includes information regarding the crash and a description of the highway-rail crossing. The database contains 19 years of historical crash information at HRGCs in North Dakota. There were 5,713 HRGCs in North Dakota, of which 354 have historical crash records in 19 years. With the intent to study crash-associated factors, a binary target variable (CRASH) is defined with two levels: a value of 1 indicates that a crash happened, while a value of 0 represents a crossing with no crash. Table 1 lists all screened variables, including one target variable, one ID variable, and thirty potential contributing variables.

3.2. Gradient Boosting. The gradient boosting method can be viewed as multiple additive trees (MATs) and is a machine-learning data-mining technique for regression and classification problems proposed by [33, 34] at Stanford University. The GB method theoretically extends and improves the simple DT model using stochastic gradient boosting [33]. GB produces a predictive model in the form of an ensemble of several simple decision tree models [35]. Therefore, the GB model inherits all of the advantages of decision tree models while improving in other aspects, such as robustness and accuracy [34]. Moreover, several other features make the GB model attractive, including capability of handling large datasets without preprocessing, resistance to outliers, capability to handle missing data, robustness to complex data, and resistance to overfitting [34, 36].

A GB model is a series expansion approximating the true functional relationship [36]. In general, the GB model starts by fitting the data with a simple decision tree model, which has certain level of error in terms of fitness with the data. A

Table 1: Input variable description.

Variable	Property	Description
CRASH	Target variable	1 = crash happened, 0 = no crash
ID	ID variable	Crossing identification
AADT_N	Numeric	Annual average daily traffic
AVERAGE_TRAIN_SPEED	Numeric	Average train speed
DAYSWT	Numeric	Day switching trains
DAYTHRU	Numeric	Day through trains
NGHTSWT	Numeric	Night switching trains
NGHTTHRU	Numeric	Night through trains
SCHLBUS	Numeric	Average number of school bus passing over the crossing
Highway_Paved	Category	Is highway paved or not? 1 = yes, 0 = no
Highway_Stop	Category	Highway stop sign presence: 1 = yes, 0 = no
HWYSYS	Category	Highway system the highway belongs to: 1 = National Interstate System, 2 = other national highway systems, 3 = Federal-Aid Highway System, 4 = non-Federal-Aid Highway System
TRUCKLN	Category	Are truck pull-out lanes present? 1 = yes, 0 = no
ADVWARN	Category	Railroad advance warning signal presence: 1 = yes, 0 = no
COMPOWER	Category	Commercial power availability: 1 = yes, 0 = no
DOWNST	Category	Does track run down a street? 1 = yes, 0 = no
FLASHMAS	Numeric	Number of mast-mounted flashing lights in pairs
FLASHNOV	Numeric	Number of cantilevered flashing lights not over traffic lane
FLASHOV	Numeric	Number of cantilevered flashing lights over traffic lane
FLASHPAI	Numeric	Number of flashing lights in pairs
GATES	Numeric	Number of gates
Near_City	Category	The crossing is in or near city? 1 = near city, 0 = in city
PAVEMRK	Category	Distance of pavement markings from crossing: 1 = less than 75 ft, 2 = 75 to 200 ft, 3 = 200 to 500 ft, 4 = N/A
SGNLEQP	Category	Is track equipped with train signals? 1 = yes, 0 = no
SPSEL	Category	Train detection: 1 = constant warning time system (CWT), 2 = direct current audio frequency overlay (DC/AFO), 3 = N/A
STOPSTD	Numeric	Number of highway stop signs
TOTAL_NUMBER_TRACK	Numeric	Number of rail tracks
TRAFICLN	Numeric	Number of traffic lanes crossing railroad
WHISTBAN	Category	Quiet zone: 1 = 24 hours, 2 = partial, 3 = unknown, 4 = no
WIGWAGS	Numeric	Number of wigwags
XANGLE	Category	Smallest crossing angle: 1 = 0–29, 2 = 30–59, 3 = 60–90
XBUCK	Numeric	Number of cross bucks

detailed description of the simple decision tree algorithm and data selection can be referred to Zheng et al. [28]. Considering the errors having the same correlation with outcome value, the GB model then develops another decision tree model on the errors or the residuals of the previous tree. This sequential process will repeat until errors are minimized.

The detailed algorithm of GB is described as follows [29, 37]: $f(x) = \sum_n f_n(x) = \sum_n \beta_n g(x, \gamma_n)$,where x is a set of predictors and $f(x)$ is the approximation of the response variable. $g(x, \gamma_n)$ are single decision trees with the parameter $\gamma_n$indicating the split variables. $\beta_n (n = 1, 2, \ldots, n)$ are the coefficients and determine how each single tree is to be combined. Friedman [38] proposed a functional gradient descent optimization algorithm to find the final optimal GB model, which is an iterative tree-building process. The method keeps adding trees until all observations are perfectly fitted. It will result in an overfitted model that only performs perfect for the trained data (low bias) but has very low prediction accuracy with a different dataset (high variance). To avoid overfitting, the model then needs to be tested by fitting a testing dataset. Iterative training will stop when the performance of the model reaches a point where the model predicts well for both the training and testing datasets.

Regularization parameters are critical for avoiding overfitting and improving model performance. Regularization parameters usually include two parameters: learning rate and tree complexity. Learning rate is also called shrinkage rate [39]. It controls how fast the model is updated or improved after each stage. The value of learning rate ranges from 0 to 1. A small value of learning rate yields great improvement and minimizes loss function but requires more iterations and computational time [29]. Higher values, close to 1, result in overfitting and poor performance [39]. Tree complexity represents the number of nodes per single simple tree [37]. A higher number of nodes will introduce lower bias in the training set while increasing the variance in the testing set. This phenomenon is recognized as the bias-variance trade-off. As a result, both learning rate and tree complexity rate must be balanced to satisfy this trade-off and avoid overfitting.

4. Results

This section presents findings of this study. R is the software to conduct all the analysis for the research. First, the selection of an optimal model based on model performance is presented. Then, based on the optimal model, influential variables are ranked by their importance to the target variable. After that, impacts of top influential variables on crash prediction are analyzed. Finally, a model performance comparison is conducted between the DT model and the GB model.

4.1. Model Setup. To detect interactions between variables and to take full advantage of the GB model, a higher level of tree complexity and a low learning rate are suggested for experimentation [39]. In this study, the model is tested under three values of learning rate: 0.05, 0.01, and 0.005, and five levels of tree complexity: 2, 4, 6, 8, and 15. Table 2 shows how the model performs with various learning rate and tree complexity levels. "Class" represents crash levels with 0 indicating no crash and 1 indicating crash. Columns "Training data error percent" and "Testing data error percent" show the percentage of prediction error for training data and testing data, respectively. "Number of trees" indicates the number of trees needed to train for an optimal model under corresponding learning rate and tree complexity. It is clear that for a lower learning rate or a lower tree complexity, more trees are needed to achieve the optimal model.

The optimal model should predict well for both training and testing data; in addition, accurate prediction of event level is also critically important. Moreover, the number of trees required to achieve the optimal performance model indicates computing time and should be considered when selecting regularization parameters. By balancing model performance in terms of training error, testing error, event forecasting error, nonevent forecasting error, number of trees needed to obtain optimization model, and computing resource requirement, this research selected the model with a learning rate of 0.01, tree complexity of 8, and an ensemble of 1,092 trees as the optimal setting. Prediction accuracy for the optimal model is 85.7% for nonevent level (CRASH = 0) and 83.9% for event level (CRASH = 1). Variable importance and their impacts on crash at HRGCs will be generated based on the optimal model using 1,092 simple decision trees.

4.2. Variable Importance. The importance of a variable in a simple single tree is measured by the number of times the variable is used as the splitter and the squared improvement attributed to the tree due to the splits by the variable. After summing the number of times used as the splitter and the squared improvement over the ensemble of trees, the average value of the summation is regarded as the variable importance in the model. A high value of variable importance indicates a high level of contribution to the prediction [34]. In this study, the authors use the same algorithm to measure variable importance.

Table 3 presents the relative variable importance of each contributor based on the selected optimal GB model. The "Relative importance" column shows the importance value of the corresponding variable with 100 being assigned to the most important variable and a relative percentage values to all other contributors regard to the top variable. "Influence percent (%)" is an absolute importance factor which indicates how much each variable contributes to the prediction. Twenty-eight factors out of thirty are identified as having impacts on crashes at HRGCs, and the top ten factors contribute about 60% of the total crash influence power. Single factor influence power ranges from 1% to 11%. Highway traffic and railroad daytime traffic alone contribute about 20% influence on crash, and the majority of them are less than 5%. In other words, crashes at HRGCs are complicated and cannot be explained by only a few factors, but highway traffic and daytime rail traffic do play the most influential impact on crash likelihood.

Among all 28 influential factors, average annual daily highway traffic, daily through-train traffic, train detection type, nightly through-train traffic, average train speed, and the number of traffic lanes are the top six contributors to crash prediction. Among these six variables, four of them are traffic-related variables describing highway and railway traffic exposures and contribute about 30% to the prediction. Most of the predictors are crossing characteristics (17 out of 28), such as SPSEL, ADVWARN, and PAVEMRK, which provide information about warning systems and train-detecting systems, and they cumulatively contribute to about 50% of the impacts.

4.3. Marginal Effect of Contributing Variables. One of the criticisms frequently found in the literature for newer predictive modeling approaches such as gradient boosting is the difficulty of interpretation relative to linear regression models. For that reason, partial-dependent plot analysis is conducted in this study. Partial-dependent plots can be viewed as a graphical representation of contributor coefficients for each individual independent variable. Essentially, partial plots are model-based simulations [36]. The values appearing in the y-axis are the modeled values of the response variable. A positive y value indicates that the contributing variable at the corresponding value has a positive influence on the classification in the model. In this study, all other contributors hold at their mean values while researching on the influence changes of a target variable. Figure 1 illustrates the use of partial-dependence plots to characterize the marginal effects of the three types of contributors: traffic, highway, and crossing characteristics.

4.3.1. Traffic Characteristic Variables. Figures 1(a)–1(d) present the effects of AADT, DAYTHRU, NGHTTHRU, and AVG_TRAIN_SPEED on crash likelihood at HRGCs, respectively. The nonmonotonic relationship indicates a clear nonlinear, dynamic, and complex relationship between the target contributor and crash likelihood. Note that the impact indicates the effect of the target variable on crash likelihood while all other contributors hold at their mean

Table 2: Misclassification rate vs learning rate and complexity of trees.

Learning rate	Tree complexity	Class	Training data error percent	Testing data error percent	Number of trees
0.05	2	0	0.1737	0.1847	1439
		1	0.1691	0.1829	
	4	0	0.1486	0.1611	468
		1	0.1544	0.2927	
	6	0	0.1404	0.1574	394
		1	0.1397	0.2927	
	8	0	0.1216	0.141	181
		1	0.1287	0.3171	
	15	0	0.112	0.1292	181
		1	0.114	0.3171	
0.01	2	0	0.18	0.1938	2853
		1	0.1949	0.2073	
	4	0	0.1491	0.1629	2461
		1	0.1581	0.2683	
	6	0	0.143	0.1565	1424
		1	0.1507	0.2683	
	8	0	0.1399	0.1547	1092
		1	0.1324	0.2561	
	15	0	0.1153	0.1338	711
		1	0.1176	0.3049	
0.005	2	0	0.1808	0.192	5317
		1	0.2132	0.2073	
	4	0	0.1514	0.1656	4528
		1	0.1618	0.2561	
	6	0	0.1446	0.1611	2745
		1	0.1544	0.2561	
	8	0	0.1277	0.1383	1986
		1	0.1287	0.2927	
	15	0	0.1164	0.1328	1387
		1	0.1103	0.3049	

influential levels. However, it is clear that a roughly increasing pattern exists in all traffic exposure variables except nighttime train volume. In Figure 1(a), crash rate suddenly reaches a peak value when AADT is about 500, which indicates that a "crash" is very likely to occur when AADT is 500. With AADT greater than 500, one can tell in general, crash likelihood is increasing gradually with AADT increasing; however, the relationship is not monotonic, and there are other two peaks at 2,500 and 10,000 AADT. The reason for the nonmonotonic is the influence of all other contributors hold at their mean level. In Figure 1(b), crash likelihood stays roughly constant when DAYTHRU is between 7 and 20; however, it starts increasing after daytime train traffic increasing from 20 trains per day. As in Figure 1(c), crash likelihood fluctuates at a low rate before NGHTTHRU reaches 11, beyond which, a sudden dramatic increase is observed which indicates that crash likelihood increases dramatically if nighttime through-train traffic increases from 11 to 13 and remains high when nightly through-train volume is greater than 13. Observations from this figure indicate the other controlled contributors at their mean level dominating the influence on crash likelihood compared to nighttime train traffic while its impact is around zero before the volume reaches 11. Figure 1(d) suggests that a crash is less likely to happen when the train speed is less than 30 mph and increases dramatically when train speeds increase up to 35 mph while other contributors hold at their mean influence level. As shown, crossings with trains travelling at speeds between 3 and 13 are less likely to have a crash.

4.3.2. Highway Characteristic Variables. Figure 1(e) shows the effect of one of the highway characteristic variables: HWYSYS. It is found that crashes tend to happen at crossings intersecting Federal-aid highways (coded as 3). In contrast, crossings intersecting with non-Interstate highways (coded as 2) or non-Federal-aid highways (coded as 4) are less likely to have a crash.

4.3.3. Crossing Characteristic Variables. Effect of crossing characteristic variables is critical for HRGC design. Figures 1(f) and 1(g) show the effects of two crossing characteristic variables, SPSEL and TRAFICLN, respectively. It indicates that a direct current audio frequency overlay (SPSEL = 2) installed at a HRGC helps to reduce the likelihood of crashes. It also suggests that crashes tend to happen at HRGCs with constant warning time (CWT) systems. With the CWT system, a warning signal is activated intentionally to provide a constant preselected warning time, usually 25 seconds. So, for a slow-moving train, the distance between the train and the crossing could be much closer than for a faster-moving train. However, a CWT system is not able to measure a change in speed accurately which can result in variability in the actual warning time oftentimes, less than

Table 3: Variable importance based on the GB model.

Variable	Relative importance	Influence percent (%)	Cumulative influence
AADT_N	100.00	11.00	11.00
DAYTHRU	73.33	8.07	19.07
SPSEL	67.73	7.45	26.52
NGHTTHRU	58.28	6.41	32.93
AVERAGE_TRAIN_SPEED	54.93	6.04	38.98
TRAFICLN	45.62	5.02	44.00
HWYSYS	37.98	4.18	48.18
ADVWARN	37.19	4.09	52.27
TOTAL_NUMBER_TRACK	35.75	3.93	56.20
PAVEMRK	34.15	3.76	59.96
XANGLE	32.85	3.61	63.57
FLASHPAI	31.24	3.44	67.01
COMPOWER	27.52	3.03	70.04
HIGHWAY_STOP	27.36	3.01	73.05
NEAR_CITY	26.31	2.89	75.94
XBUCK	25.73	2.83	78.77
HIGHWAY_PAVED	21.93	2.41	81.19
FLASHMA	21.42	2.36	83.54
SCHLBUS	19.78	2.18	85.72
SGNLEQP	19.45	2.14	87.86
GATES	18.00	1.98	89.84
TRUCKLN	16.70	1.84	91.68
STOPSTD	14.24	1.57	93.24
DOWNST	13.58	1.49	94.74
WHISTBAN	13.46	1.48	96.22
DAYSWT	13.19	1.45	97.67
NGHTSWT	11.80	1.30	98.97
FLASHOV	9.37	1.03	100.00
FLASHNOV	0.00	0.00	100.00
WIGWAGS	0.00	0.00	100.00

desired warning time. As shown in Figure 1(f), highways with no more than 2 lanes have a negative impact on crash. Moreover, it is noticeable that a highway with 4 lanes has the highest positive impact on crash prediction. This is potentially caused by the more complicate traffic condition which involves lane-changing activities and will block the driver's vision of incoming crossing and trains.

4.4. Model Forecasting Accuracy Assessment and Comparison. Prediction results can be summarized in a classification table (Table 4), and based on which, model prediction accuracy measurements are computed. The observed event class is represented by "Present" in Table 4, while the observed nonevent class is represented by "Absent" in Table 4. If an observation is predicted to be event class, it is indicated as "Positive" in Table 4, otherwise "Negative." The number of true positive (TP) and true negative (TN) indicates the number of correct predictions. The number of false positive (FP) and that of false negative (FN) indicate the number of wrong predictions against observed conditions.

Even though the prediction performance of a model is a critical indicator, only a limited number of researchers published their prediction performance results in their studies [40–43]. For the ones who do report prediction assessment results, they only tend to selectively evaluate the accuracy rather than provide a full picture of the model's forecasting skills because in their study, the prediction performance is used as a validation tool rather than a full assessment focus. Most commonly selected measurements are described in equations (1)–(3) for event class, nonevent class, and overall prediction, respectively:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{1}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{2}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}. \tag{3}$$

Sensitivity and specificity compute the number of correct predictions given the number of observed conditions. Sensitivity in some other research is also referred as recall or true positive rate while specificity is referred as true negative rate. Equation (1) indicates that among all the observed present conditions (TP + FN), the model makes a number of TP correct predictions. However, it ignores the number of false positive (FP) predictions. The issue with sensitivity ignoring FP is that sensitivity can be high by sacrificing a great amount of false positive forecasts; in return, it could waste limited safety improvement budget allocation if decision makers rely on model forecasting result to allocate budget. Traditional selected prediction accuracy parameters,

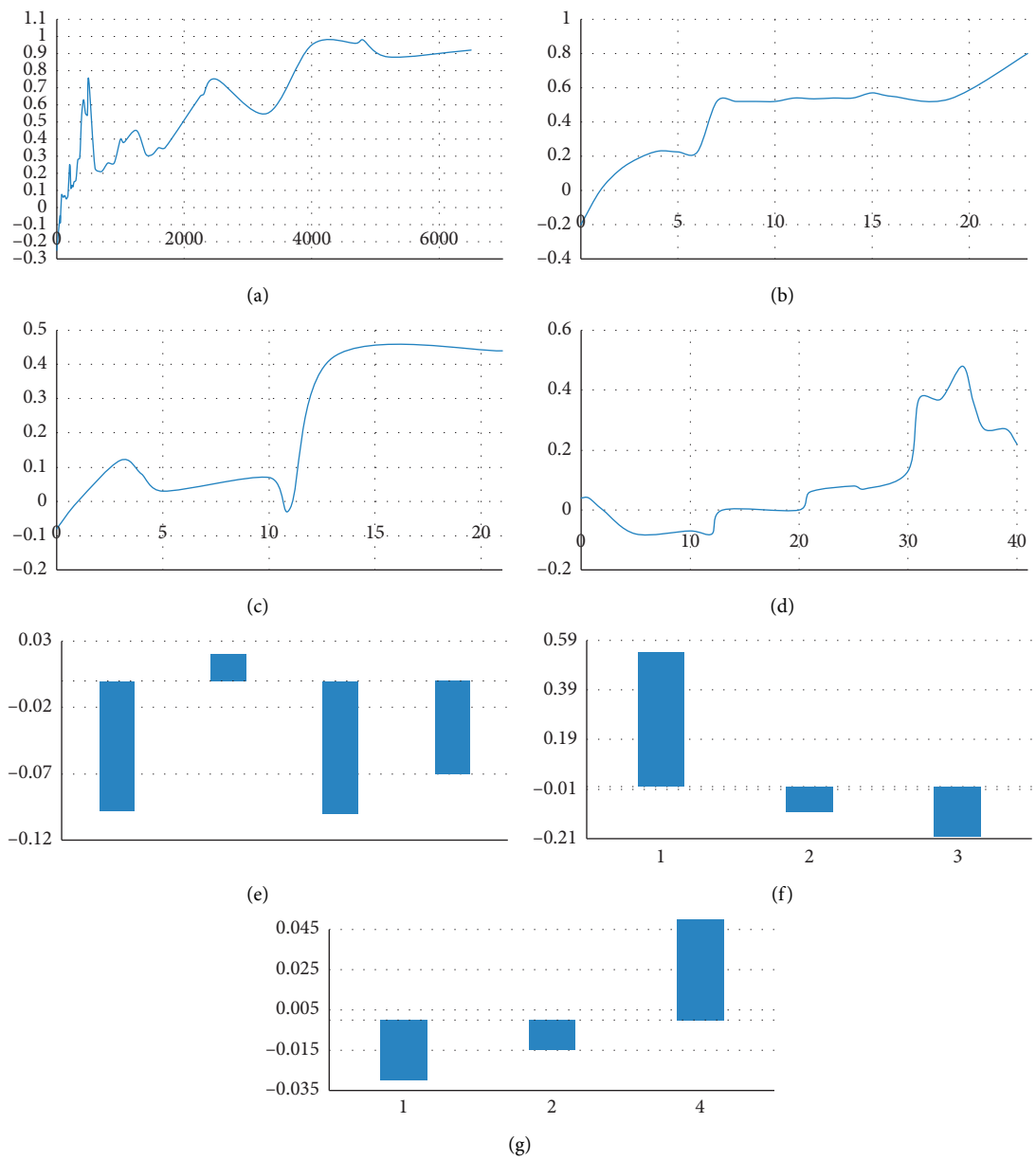


Figure 1: Partial-dependent plots. (a) AADT. (b) Day through train. (c) Night through train. (d) Average train speed. (e) Highway systems. (f) Train detection system. (g) Number of highway lanes.

Table 4: Classification table.

		Observed condition	
		Present	Absent
Predicted condition	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

sensitivity, specificity, and accuracy, only partially represent a model's prediction accuracy power. To draw the full picture of a model's prediction accuracy, three additional prediction accuracy measurements should also include in the analysis.

For event and nonevent class, respectively, positive predictive ratio (PPR) and negative predictive ratio (NPR) show true positive and true negative prediction rate over predicted condition for each case shown in equations (4) and (5). Forecasted accuracy rates are calculated by equations (3) and (6) for accurate class (accuracy) and failed class (false rate (FR)), respectively:

$$\mathrm{PPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{4}$$

$$\mathrm{NPR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}}, \tag{5}$$

$$\mathrm{FR} = \frac{\mathrm{FP} + \mathrm{FN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}}. \tag{6}$$

Table 5: Comparison of three predictive accuracy measurements.

	DT			GB		
	Training (%)	Testing (%)	Overall (%)	Training (%)	Testing (%)	Overall (%)
Accuracy	77.7	77.1	77.6	84.4	82.1	84.1
False rate	22.3	22.9	22.4	15.6	17.9	15.9
Sensitivity	84.1	88.9	84.6	88.6	88.6	88.6
Specificity	77.2	76.3	77.1	84.1	81.6	83.8
Positive predictive ratio	20.0	20.1	20.0	27.3	23.8	26.9
Negative predictive ratio	98.6	99.0	98.7	99.1	99.1	99.1

The greater the value of all the indicators except FR, the better the forecasting power indicated by the model .

In this study, all 6 measurements are evaluated for the proposed gradient boosting model and compared with decision tree model results [28] which serve as a reference level. The result is shown in Table 5.

Several interesting findings can be drawn from Table 5:

(1) Forecasting power based only on accuracy, sensitivity, and specificity can overestimate the model's forecasting power. The high value of those measurements can be biased by the high volume of nonevent forecasts for an imbalanced dataset such as crash dataset because the crash event is relatively rare compared to the noncrash event.

(2) Note that relative low positive predictive ratios for both models are caused by the imbalanced datasets with low separability. Positive predictive ratio also known as precision indicates false alarm rate. This value is often low for imbalanced data because models try to improve accuracy and sensitivity by sacrificing more false positive estimates. In this study, even though the GB model overperforms than the DT model in terms of PPR/sensitivity, both models still fail to provide a sound sensitivity for imbalanced data such as the crash data used in this study, which can be very troublesome because when decision makers depend on the model's estimates to allocate budget, they can increase unnecessary spending and waste scares resources.

(3) By looking at comprehensive accuracy assessment as the proposed six measurements, one can have a complete picture of the model's true forecasting capability and the model's forecasting trade-offs.

(4) The GB model is supreme than the DT model by simultaneously improving sensitivity and precision. In all measurements, the GB model is overperformed than the DT model.

(5) For GB model's training dataset, comparing sensitivity, precision, NPR and specificity, only 27.3% of the GB model's event estimates are correct even though this accounts for 88.6% of actual observed events. Moreover, 99.1% of noncrash forecasts are correct, which accounts for 84.1% of actual observed nonevents. The same pattern can be found in DT models. It indicates that both models overestimate crash (more false alarms compared to true event estimates) and underestimate noncrash (less misses compared to true nonevent estimates).

(6) In most of measurements, both GB and DT models perform better in the training dataset and relatively less good in the testing dataset.

(7) It is also worth to mention that accuracy is greater than 84% in the GB model, which is an outstanding improvement, compared with previous studies [44–46].

5. Research Summary

As demonstrated, the GB model can accurately identify contributing variables and determine an optimal model to avoid overfitting with regularization parameter simulation analysis. More importantly, it can also provide easy-to-interpret relative importance of influential variables and partial-dependent plots which present influential variables' marginal effect on crash prediction.

The GB model overcomes many shortcomings of common statistical models such as limited capability to model underdispersed crash and poor forecasting performance of rare event (crash) by using both training data and testing data. Comparing with the DT model result, the GB model demonstrates its ability to successfully identify contributing variables and their relative importance levels.

Moreover, the proposed evaluation approach based on six measurements can assess the model prediction accuracy more thoroughly and comprehensively in a classification study. The GB model is superior to the DT model in terms of its improved forecasting accuracies in both training and testing datasets. Both of the current GB and DT data-mining models overestimate events (false alarming) to increase its accurate event coverage forecasting rate (sensitivity) which can be misleading to decision makers to waste safety improvement resources on the unnecessary crossings. Handling imbalanced datasets in machine learning and improve forecasting performance need future research attention.

Data Availability

The data that support the findings for this study are available from the authors on the request.

Disclosure

The content of this paper reflects the views of the authors, who are responsible for the facts and accuracy of the information presented.

Conflicts of Interest

The authors wish to confirm that there are no known conflicts of interest associated with this publication.

Authors' Contributions

Dr. Pan Lu, Dr. Denver Tolliver, and Dr. Ying Huang are responsible for the research idea and study concept initiation. Dr. Pan Lu, Dr. Denver Tolliver, and Dr. Zijian Zheng designed the study. Dr. Zijian Zheng and Amin Keramati collected the data. Dr. Zijian Zheng and Dr. Pan Lu performed the model construction. Dr. Ying Huang is responsible for lab/software resource support. Dr. Zijian Zheng, Dr. Pan Lu, Xiaoyi Zhou, and Amin Keramati carried out the analysis and interpretations of results. Dr. Zijian Zheng, Dr. Pan Lu, and Dr. Denver Tolliver prepared the draft manuscript. All the authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

The authors would like to express their deep gratitude to the North Dakota State University and the Mountain-Plains Consortium (MPC), a University Transportation Center funded by the U.S. Department of Transportation.

References

[1] C. Ma, W. Hao, W. Xiang, and W. Yan, "The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossings accidents," *Journal of Advanced Transportation*, vol. 2018, Article ID 9841498, 10 pages, 2018.

[2] Federal Railroad Administration, "Highway-rail grade crossings overview," 2015, https://railroads.dot.gov/program-areas/highway-rail-grade-crossing/highway-rail-grade-crossings-overviewlast.

[3] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman & Hall/CRC, London, UK, 2nd edition, 1989.

[4] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.

[5] P. Lu and D. Tolliver, "Accident prediction model for public highway-rail grade crossings," *Accident Analysis & Prevention*, vol. 90, pp. 73–81, 2016.

[6] J. Oh, S. P. Washington, and D. Nam, "Accident prediction model for railway-highway interfaces," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 346–356, 2006.

[7] L. Yee Peng, H. Midi, S. Rana, and A. Fitrianto, "Identification of multiple outliers in a generalized linear model with continuous variables," *Mathematical Problems in Engineering*, vol. 2016, Article ID 5840523, 9 pages, 2016.

[8] R. D. Austin and J. L. Carson, "An alternative accident prediction model for highway-rail interfaces," *Accident Analysis & Prevention*, vol. 34, no. 1, pp. 31–42, 2002.

[9] F. Guo, X. Wang, and M. A. Abdel-Aty, "Modeling signalized intersection safety with corridor-level spatial correlations," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 84–92, 2010.

[10] M. J. Maher and I. Summersgill, "A comprehensive methodology for the fitting of predictive accident models," *Accident Analysis & Prevention*, vol. 28, no. 3, pp. 281–296, 1996.

[11] S.-P. Miaou and H. Lum, "Modeling vehicle accidents and highway geometric design relationships," *Accident Analysis & Prevention*, vol. 25, no. 6, pp. 689–709, 1993.

[12] J. Milton and F. Mannering, "The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies," *Transportation*, vol. 25, no. 4, pp. 395–413, 1998.

[13] Y.-J. Park and F. F. Saccomanno, "Evaluating factors affecting safety at highway-railway grade crossings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1918, no. 1, pp. 1–9, 2005.

[14] W. F. W. Yaacob, M. A. Lazim, and Y. B. Wah, "Evaluating spatial and temporal effects of accidents likelihood using random effects panel count model," in *Proceedings of the 2010 International Conference on Science and Social Research (CSSR 2010)*, IEEE, Kuala Lumpur, Malaysia, pp. 960–964, December 2010.

[15] S. R. Hu and C. K. Lee, "Analysis of Accident Risk at Railroad Grade Crossing (No. 08-1426)," 2008.

[16] S. S. P. Kumara and H. C. Chin, "Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros," *Traffic Injury Prevention*, vol. 4, no. 1, pp. 53–57, 2003.

[17] J. Lee and F. Mannering, "Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis," *Accident Analysis & Prevention*, vol. 34, no. 2, pp. 149–161, 2002.

[18] D. Nam and J. Lee, "Accident frequency model using zero probability process," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1973, no. 1, pp. 142–148, 2006.

[19] M. A. Hadi, J. Aruldhas, L. F. Chow, and J. A. Wattleworth, "Estimating safety effects of cross-section design for various highway types using negative binomial regression," *Transportation Research Record*, vol. 1500, p. 169, 1995.

[20] M. G. Karlaftis and A. P. Tarko, "Heterogeneity considerations in accident modeling," *Accident Analysis & Prevention*, vol. 30, no. 4, pp. 425–433, 1998.

[21] M. G. Karlaftis and I. Golias, "Effects of road geometry and traffic volumes on rural roadway accident rates," *Accident Analysis & Prevention*, vol. 34, no. 3, pp. 357–365, 2002.

[22] Y. M. Mohamedshah, J. F. Paniati, and A. G. Hobeika, "Truck accident models for interstates and two-lane rural roads," *Transportation Research Record*, vol. 1407, pp. 35–41, 1993.

[23] X. Zhou, P. Lu, Z. Zheng, D. Tolliver, and A. Keramati, "Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree," *Reliability Engineering & System Safety*, vol. 200, Article ID 106931, 2020.

[24] X. Yan and E. Radwan, "Analyses of rear-end crashes based on classification tree models," *Traffic Injury Prevention*, vol. 7, no. 3, pp. 276–282, 2006.

[25] X. Qin and J. Han, "Variable selection issues in tree-based regression models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2061, no. 1, pp. 30–38, 2008.

[26] X. Yan, S. Richards, and X. Su, "Using hierarchical tree-based regression model to predict train-vehicle crashes at passive highway-rail grade crossings," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 64–74, 2010.

[27] A. Keramati, P. Lu, D. Tolliver, and X. Wang, "Geometric effect analysis of highway-rail grade crossing safety performance," *Accident Analysis & Prevention*, vol. 138, Article ID 105470, 2020.

[28] Z. Zheng, P. Lu, and D. Tolliver, "Decision tree approach to accident prediction for highway-rail grade crossings: empirical analysis," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2545, no. 1, pp. 115–122, 2016.

[29] G. De'Ath, "Boosted trees for ecological modeling and prediction," *Ecology*, vol. 88, no. 1, pp. 243–251, 2007.

[30] Z. Zheng, P. Lu, and B. Lantz, "Commercial truck crash injury severity analysis using gradient boosting data mining model," *Journal of Safety Research*, vol. 65, pp. 115–124, 2018.

[31] S. R. Mousa, P. R. Bakhit, O. A. Osman, and S. Ishak, "A comparative analysis of tree-based ensemble methods for detecting imminent lane change maneuvers in connected vehicle environments," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 42, pp. 268–279, 2018.

[32] H. Trevor, "Gradient boosting machine learning," 2014, https://www.youtube.com/watch?v=wPqtzj5VZus.

[33] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.

[34] J. H. Friedman and J. J. Meulman, "Multiple additive regression trees with application in epidemiology," *Statistics in Medicine*, vol. 22, no. 9, pp. 1365–1381, 2003.

[35] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999.

[36] Salford-Systems, "Frequently asked questions and answers about TreeNet," 2017, http://www.salfordsystems.com/doc/TreeNetFAQ.pdf.

[37] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Science & Business Media, Berlin, Germany, 2009.

[38] J. H. Friedman, "Machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[39] Salford Systems, "SPM user guide: introducing TreeNet," 2017, http://media.salford-systems.com/pdf/spm7/IntroTN.pdf.

[40] S. Dhruvit, H. R. Varia, and P. M. Shah, "Road accident analysis and severity prediction model on state highway-5 (Halol Shamlaji section)," *International Journal of Science & Engineering Development Research*, vol. 1, no. 5, pp. 346–351, 2016.

[41] G. M. McCollister and C. C. Pflaum, "A model to predict the probability of highway rail crossing accidents," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 221, no. 3, pp. 321–329, 2007.

[42] G. Prati, L. Pietrantoni, and F. Fraboni, "Using data mining techniques to predict the severity of bicycle crashes," *Accident Analysis & Prevention*, vol. 101, pp. 44–54, 2017.

[43] F. F. Saccomanno, L. Fu, and L. F. Miranda-Moreno, "Risk-based model for identifying highway-rail grade crossing blackspots," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1862, no. 1, pp. 127–135, 2004.

[44] L.-Y. Chang and W.-C. Chen, "Data mining of tree-based models to analyze freeway accident frequency," *Journal of Safety Research*, vol. 36, no. 4, pp. 365–375, 2005.

[45] L.-Y. Chang and J.-T. Chien, "Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model," *Safety Science*, vol. 51, no. 1, pp. 17–22, 2013.

[46] L.-Y. Chang and H.-W. Wang, "Analysis of traffic injury severity: an application of non-parametric classification tree techniques," *Accident Analysis & Prevention*, vol. 38, no. 5, pp. 1019–1027, 2006.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 2469436, 10 pages
https://doi.org/10.1155/2020/2469436

Research Article

Compressive Strength Gain Behavior and Prediction of Cement-Stabilized Macadam at Low Temperature Curing

Yongli Xu, Guang Yang, and Hongyuan Zhao

School of Civil Engineering, Northeast Forest University, Harbin 150040, China

Correspondence should be addressed to Yongli Xu; xuyongli77@163.com

Received 3 December 2019; Accepted 17 February 2020; Published 30 March 2020

Guest Editor: Wei (David) Fan

For cement-based materials, the curing temperature determines the strength gain rate and the value of compressive strength. In this paper, the 5% cement-stabilized macadam mixture is used. Three indoor controlled temperature curing and one outdoor natural curing scenarios are designed and implemented to study the strength development scenario law of compressive strength, and they are standard temperature curing (20°C), constant low temperature curing (10°C), day interaction temperature curing (varying from 6°C to 16°C), and one outdoor natural temperature curing (in which the air temperature ranges from 4°C to 20°C). Finally, based on the maturity method, the maturity-strength estimation model is obtained by using and analyzing the data collected from the indoor tests. The model is proved with high accuracy based on the validated results obtained from the data of outdoor tests. This research provides technical support for the construction of cement-stabilized macadam in regions with low temperature, which is beneficial in the construction process and quality control.

1. Introduction

Cement-stabilized macadam is low-dose cement-base-stabilized mixture and its cement dosage is 5% or so; it is generally used for base layer of pavement construction in China [1]. Whether or not the compressive strength of cement-based materials largely depends upon the curing process is well known, in which both the curing temperature and time are particularly important [2, 3]. For conventional laboratory test of compressive strength, the curing is typically carried in constant temperature environment at 20°C in many national specifications [4–6]. But for the pavement construction project, the actual outdoor curing temperature varies with weather. Specification requires that more than 5°C temperature can be carried out for construction [4]. However, in the northern seasonal frozen areas such as China's Heilongjiang Province, although the temperature is more than 5°C in the month of April, the temperature changes a lot and is very unstable. Due to the large temperature difference between day and night and the fact that usually does not reach 20°C during curing time, the compressive strength sometimes cannot meet the requirements, causing the coring to be loose. Since the strength cannot be confirmed, one cannot reasonably arrange the next process [7]. Based on this special temperature condition, there is a strong need to study the compressive strength gain laws under such different low temperature curing conditions. In that regard, several indoor and outdoor experiments are designed accordingly in this paper to conduct such study.

There are many research efforts made to study the impact of curing temperature on cement-based materials, such as Portland cement-stabilized soil, light weight cemented soil, sand, coal fly ash, and lime blends [8–10]. As for the curing temperature, many research studies have been reported for the high temperature, and most results showed that high temperature curing can increase the initial compressive strength [11, 12]. The compressive strength and tensile strength of cement-stabilized marine soils which were used as road construction materials were studied under cured temperatures varying from 40°C to 60°C in Wang's research work [13]. Escalante-Garcia et al. [14] tested the compressive strength of hydration at five temperatures ranging from 10°C to 60°C, and the results showed that high temperature can improve the initial compressive strength, but it can actually

decrease the strength in the long term. Wang et al. [15] conducted tests of calcium sulphoaluminate cement at different curing temperatures (i.e., 0°C–80°C) in order to study the influence of hydration evolution on compressive strength. The results indicated the early-age compressive strength increases with increasing temperature but decreases at the temperature ranging from 40°C to 80°C, and the compressive strength is mainly affected by hydration degree.

About low temperature curing, several studies have been reported in the literature. Price [16] showed that the strength of concrete mixed developed significantly slower at low temperature than that at room temperature. Husem et al. [17] tested the compressive strength of ordinary and high performance concrete under standard curing (at 23 ± 2°C) and other low temperature curing (at 10, 5, 0, and −5°C, resp.). The results indicated that the strength at 10°C and less than 10°C was lower than that in the standard curing. Kim et al. [18] investigated the strength development for curing histories with 5°C, 20°C, and 40°C temperature, which indicated that the concrete strength at low temperature was less than that at standard temperature initially but was almost the same with time. Marzouk et al. [19] performed the tests at five temperatures ranging from −10°C to 20°C over 3 months and found that there was a proportional relationship between compressive strength and temperature.

In addition, in terms of strength prediction, many literatures have shown that maturity theory is appropriate and better in the strength prediction than some other methods [20, 21]. In 1951, Saul et al. [22] first proposed the "maturity" concept, which was defined as the product of the curing time and temperature. It was pointed out in the famous Nurse-Saul maturity function that when the maturity is the same, the strength will also be approximately the same. It is well noticed that the Nurse-Saul maturity model has been constantly improved and modified later, and different mathematical models have been adopted to predict the strength. For example, in Chitambira's model, the equivalent age as an index was proposed which combined both the curing age and curing temperature [23]. There was a linear relationship between double logarithmic strength and logarithmic maturity under different curing temperatures. Jeong et al. [24] calibrated the relationship of relative strength and maturity by the moisture factor.

A review of the existing literature revealed that although there have been many studies on other cement-based materials, less research efforts were made to the 5% cement-stabilized macadam. Many studies were devoted to the effect of curing temperature on strength. However, most of them focused on high temperatures, and, furthermore, almost all the curing (be they under either high temperature or low temperature) was made under varied constant controllable temperature at the laboratory chamber. It is important to note that such curing failed to account for the alternate changes in the temperatures during actual days and nights (like the construction project), and there have been no tests conducted under outdoor natural conditions. As such, this purpose of this study is to focus on the strength gain law of 5% cement-stabilized macadam mixture at low temperature, which meets the actual temperature of the construction project. The maturity theory will be employed to predict the compressive strength. Appropriate function will be selected and relevant parameters will be calibrated and obtained by using and analyzing the experimental data. The research results will provide technical support for the construction of cement-stabilized macadam in regions of low temperature, which is beneficial to the construction quality and process control.

2. Descriptive Analysis of Temperatures in Harbin Area

Harbin City, Heilongjiang Province, China, is located in the north latitude 44°04′ ~ 46°40′, mainly plain, belonging to the north temperate zone continental monsoon climate, and the temperature changes fast in the spring and autumn. Annual precipitation reaches 400 mm–600 mm, moisture coefficient is in the range of 0.25–1.25, and the average maximum permafrost is 120 cm–240 cm.

The temperature distribution from the 15th to 30th of April from 2012 to 2014 at Harbin City is shown in Figure 1. The trend of the high temperature and the low temperature during the construction period is basically similar. Most of the high temperatures are distributed in the range from 15°C to 20°C, and most of the low temperatures fall in the range from 5°C to 10°C. The average high temperature is 16°C and the average low temperature is 6°C.

Figure 2 showed the daily temperature data from the 15th to 30th of April, 2014, at Harbin City. The data of other years follow similar pattern. At about 2 am–4 am the temperatures were the lowest, from 5 am the temperatures began to rise consistently for 9 hours at a high rate, at 12 pm–14 pm the temperatures reached the highest, and then the temperatures started to decrease continuously for 15 hours at a relatively low rate.

3. Indoor and Outdoor Testing Plans

According to the law of temperature variation, three indoor testing cases and one outdoor test were designed. The temperatures of three indoor tests were determined according to nearly 3 years of data at Harbin, as shown in Figure 3, and the outdoor tests started at the 17th of April, 2015.

The cylinder specimens of 150 mm × 150 mm size with 5% cement-stabilized macadam were prepared according to the stabilized macadam mix design. Under three different curing temperatures, unconfined compressive strength tests were carried out on a daily basis.

Case 1 (standard temperature curing): the standard curing was in full accordance with the requirements of the specification operation in which the temperature was 20°C. The unconfined compressive strength test was carried out from the 3rd day to the 7th day. The compressive strength of the 7th day (i.e., the 7th standard strength) was used as a standard for reference.

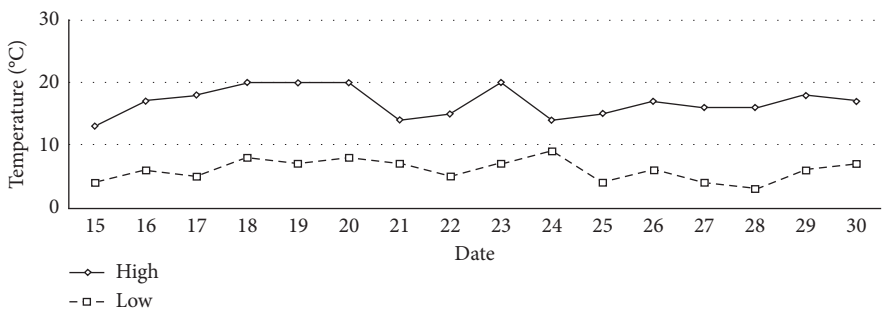


Figure 1: Temperature distribution in April from 2012 to 2014 at Harbin City.

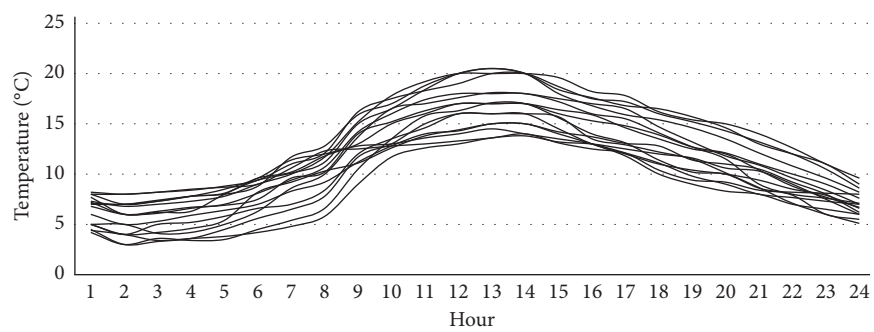


Figure 2: Daily temperatures from the 15th to 30th of April, 2014, at Harbin City.

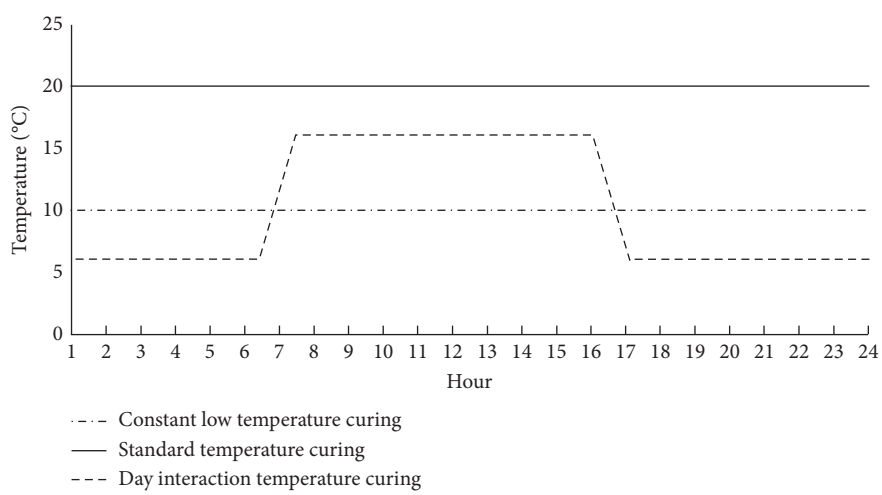


Figure 3: Temperature schematic diagram of three indoor testing cases.

Case 2 (constant low temperature curing): the curing temperature was 10°C that was determined according to the average high and average low temperatures weighted by time during nearly three years. The compressive strength was tested and the tests would not stop until the compressive strength became more than the 7th standard strength.

Case 3 (day interaction temperature curing): the temperatures were changed in the test chamber to simulate the large variations in the day and night temperatures. As shown in Figure 3, the high temperature was kept at 16°C from 7 am to 15 pm for 8 hours, and the low temperature was 6°C from 16 pm to 6 am for 14 hours. From 6 am to 7 am, the temperatures were increased from 6°C to 16°C, and from 15 pm to 16 pm, the temperature was reduced from 16°C to 6°C. Also, the compressive strength would continue to be tested beyond the 7th day until the strength was more than the 7th standard strength.

Case 4 (outdoor natural temperature curing): according to the weather forecast data, the test started on April 17, 2015. The specimens were placed in test pit. The pavement base layer and curing methods were simulated and the compressive strength was tested from the 7th day until the strength was more than the 7th standard strength. The specific operating procedure and temperature measuring method are discussed as follows.

First, a depth of 15 cm pit was dug and the bottom was flattened. The specimens were then neatly placed into pit, and the gap was filled with fine aggregate and compact. The upper was covered with a white geotextile to keep moisture, and the water was sprayed on the surface at noon every day. Specimen placement pictures were shown in Figure 4.

Three specimens were used for temperature measurement. On each specimen, four temperature sensors were embedded in the upper, the middle external, the bottom, and the central parts of the body, which were used to measure the temperature of different parts of each specimen. Figure 5 is the schematic diagram showing the location of temperature sensors, among which the central sensor was embedded in the specimen production process and the external three sensors were fixed on the surface later. The pictures showing the central sensors and middle external sensors were given in Figure 6. During the outdoor curing period, the hand-held thermometer was used to measure the temperatures and the measurement frequency was 1 reading/hour.

4. Material Performance and Test Methods

4.1. Cement Performance. The cement used in the experiment was Harbin TIANE 425 #. The technical indexes of cement are shown in Table 1. Note that the cement dosage is 5% of the aggregate weight.

4.2. Aggregate Grade. The aggregates used were of four sizes: 2 cm–3 cm, 1 cm–2 cm, 0.5 cm–1 cm, and 0 cm–0.5 cm. The gravel used was in line with the "Road Pavement Construction Technical Specifications (JTJ034-2000)" requirements. The aggregate composite grade is shown in Table 2.

4.3. Compaction Test. To prepare for the specimen making, the maximum dry density and the optimum water content of the mixture were determined through compaction tests. According to the procedures described in "Test Procedure for Stabilized Materials for Highway Engineering Inorganic Binder (JTG E51-2009)", the optimum water content was 6.8% and the maximum dry density was 2.144 g/cm^3.

4.4. Unconfined Compressive Strength Test. The specimens were made and kept in the curing chamber. In accordance with the requirements, the curing temperatures in three cases were controlled at 20°C and 10°C and in the range from 6°C to 16°C. The specimens were subjected to the unconfined compressive strength tests according to the designed testing plan.

5. Results and Discussion

5.1. Indoor Test Results. Figure 7 is the compressive strength gain law of three-case indoor tests. As for the standard curing temperature of 20°C (Case 1), the strength increases as the curing time increases, and the gain rate is initially high but gradually decreases until the 7th day. The strength is 3.5 MPa which can meet the standard requirements. Under the constant low temperature of 10°C conditions (Case 2), the compressive strength increases continuously with the increase of the curing time, but the gain rate is less than that under the standard curing condition. The compressive strength is 2.2 Mpa at the 7th day, accounting for only 62.9% of the 7th standard strength. The compressive strength does not achieve the 7th standard strength until the 14th day. In the day interaction temperature curing of 6°C–16°C (Case 3), the compressive strength also increases with the increase of the curing time, but the gain rate is less than that for the standard curing and is also slightly less than that under the constant low temperature curing condition. The compressive strength is 2.1 Mpa at the 7th day, which is only 60% of the 7th standard strength under the standard curing condition. The compressive strength does not reach the 7th standard strength until the 14th day.

5.2. Outdoor Test Results

5.2.1. Temperature Transfer Law of the Specimens in Outdoor Natural Environments. Figure 8 shows the day temperature curve at each position of the specimens on April 20, 2015. It can be seen that the temperature variation in the specimens was similar to that of the air temperature, and the fluctuation range in the upper part was larger than that in the middle and the lower parts. The difference between the central and the middle external was small, which indicated that the temperature transfer was small in the horizontal direction. Temperature transfer law of the specimens in outdoor natural environments is presented as follows.

(1) From 6 am, the temperature started to increase, and the temperature difference between the upper, the middle, and the bottom parts also gradually increased.

(2) At 11 am~14 pm, the temperature difference between the upper and bottom part reached the maximum of 8°C, while the upper and middle temperature difference was about 6°C and the middle and bottom temperature difference was about 2°C. This clearly indicated that the temperature demonstrated a nonlinear pattern in the depth direction. In other words, the heat received by the surface was the most significant; then the heat reduced noticeably when it was transferred to the middle and was almost nonexisting until the bottom.

(3) At 13 pm, the upper temperatures reached the maximum in the day, and at 14 pm, the middle and the bottom temperatures reached the maximum in the day. After that, the temperature of all parts gradually decreased in which the upper part temperature dropped at the fastest speed and the middle and lower temperatures slowly went down.

(4) From 20 pm to nearly 5 am or so, the temperatures at each position were basically the same in which the temperature difference between the upper, middle, and bottom parts is within 2°C.

The "Temperature × Time" data was used as an index to analyze the curing status at each position of the specimens.

(a) (b)

Figure 4: Specimen placement pictures. (a) Specimens were placed into pit. (b) Gap was filled and covered with geotextile.

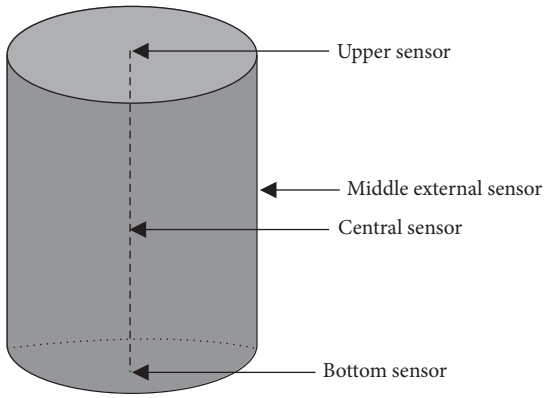
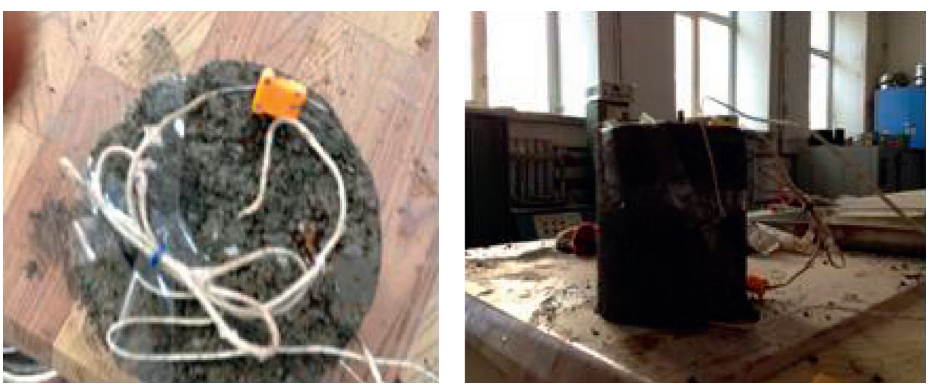


Figure 5: Schematic diagram showing the location of temperature sensors.



(a) (b)

Figure 6: Sensors embedded in specimen. (a) Central sensors. (b) Middle external sensor.

The "Temperature × Time" cumulative sum at each position of the specimens in outdoor natural environments was calculated from the 7th day and was shown in Table 3. The "Temperature × Time" of the 7th day standard curing was calculated to be 3360°C·h.

As can be seen from Table 3, when the curing continued to the 12th day, the "Temperature × Time" value at the upper position reached 3569°C·h, which exceeded the standard curing at the 7th day of 3360°C·h. However, it was only 2498°C·h at the bottom position and 2979°C·h at the central position. Based on maturity theory, it could be considered that the compressive strength at the upper position has achieved the 7th standard strength, while that at the middle and bottom positions did not reach the 7th standard strength. This can also be a good explanation as to why coring in the construction site can sometimes fail in which only the upper part is solid and the bottom part is fairly loose, as shown in Figure 9.

Table 1: Technical indexes of cement.

Index	Initial setting time	Final setting time	3D strength (MPa)	
			Compressive strength	Flexural strength
Value	1 h 3 min	2 h 40 min	21.3	4.8

Table 2: Composite grade of concrete aggregate.

Screen size (mm)	26.5	19	9.5	4.75	2.36	0.6	0.075
Composite grade	97.7	77.0	48.0	28.6	21.0	10.5	2.2

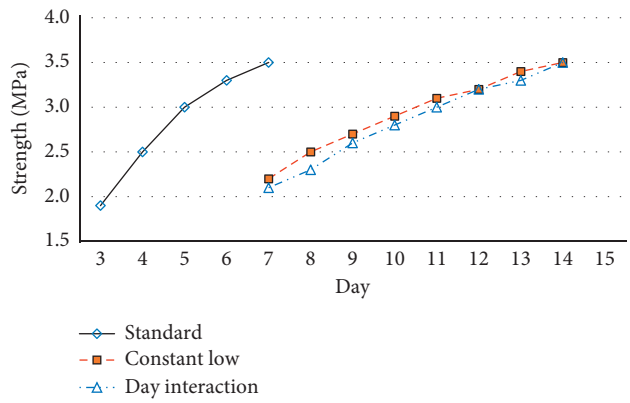


Figure 7: Strength gain curve of three case indoor tests.

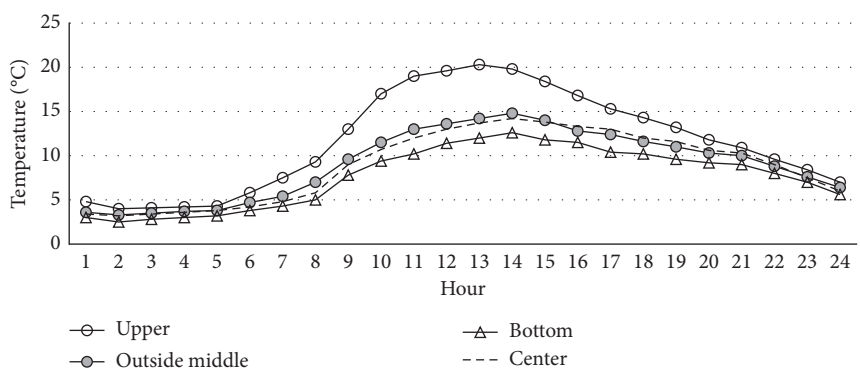


Figure 8: The day temperature curve at each position of the specimens.

5.2.2. Strength Gain Law under Outdoor Natural Temperature Curing. Figure 10 shows the strength gain law under outdoor nature temperature curing. The compressive strength gained with increasing curing days. The strength at day 7 was 2.2 MPa accounting for only 62.9% of standard curing and achieved the 7th standard strength when the number of days reached 13.

6. Comparisons of Strength Gain Law and Establishing the Maturity-Strength Model

6.1. Comparison of Strength Gain Law under Four Curing Conditions. Figure 11 provides the comparisons of compressive strength gain curves under different curing conditions. The following conclusions can be made.

(1) In all four cases, the compressive strength increased with the increasing curing time. The gain rate of low temperature curing was lower than that under the standard curing temperature curing. The gain rates can be sorted in descending (from high to low) order: standard temperature curing > outdoor natural temperature curing > constant low temperature curing > day interactive temperature curing, in which the difference between the last two was insignificant.

(2) The strength gain curves under four cases were in accordance with the logarithmic curve with the function form being $f = a\ln(M) - b$. After the model calibration, it was found that the average gain rate for the standard temperature was

Table 3: The "Temperature × Time" data for each position (°C·h).

Location curing days (d)	Upper	Middle	Bottom	Central
7	2057	1727	1427	1690
8	2360	1987	1641	1946
9	2660	2247	1853	2200
10	2965	2515	2068	2462
11	3265	2779	2280	2719
12	3569	3045	2498	2979
13	3877	3315	2720	3246

Figure 9: Coring loose picture.

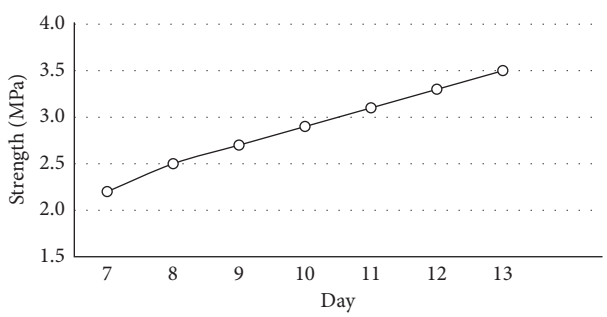


Figure 10: Strength gain curve under outdoor natural temperature curing.

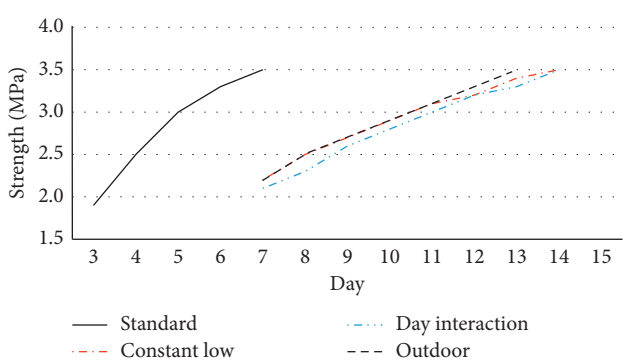


Figure 11: Comparison of strength gain law in four cases.

$a = 1.0152$, for the constant low temperature of 10°C it was $a = 1.4635$, for the day interactive temperature it was $a = 1.5106$, and for the outdoor natural temperature the average gain rate was $a = 1.6107$.

(3) To achieve the same strength of 3.5 MPa, the number of days required under each of these four cases was shown as follows: 7 days for standard temperature, 14 days for both the constant low and day interaction temperatures, and 13 days for the outdoor temperature.

(4) On the 7th day, the standard strength reached 3.5 MPa, while the other three were 2.2 MPa, 2.1 MPa, and 2.2 MPa, respectively, which accounted for only 62% or so.

(5) Among three low temperature curing cases, the curves of the constant low temperature and the outdoor natural temperature were the same until the 11th day, both of which were also very close to the day interaction temperature case although the day interactive gain was the slowest among these three cases. The maturity theory will be used to explain this result in the next section.

6.2. Maturity-Strength Model Estimation and Forecasting. The cement-stabilized macadam mixture consists of mainly cement, graded macadam, and water. The composition is similar to that of the cement concrete. The only difference lies in the cement dosage. Maturity theory has been widely used to predict the strength of cement concrete. As such, from the material composition point of view, the prediction function can be established based on maturity theory to forecast the compressive strength of 5% cement-stabilized macadam mixture. Because cement-stabilized macadam can be seen as cement concrete with low-dose cement, there are four functions that can be used based on the existing studies of cement concrete, including

Table 4: The maturity data for the standard curing (Case 1) (°C·h).

Days	3 d	4 d	5 d	6 d	7 d
Stand curing	1440	1920	2400	2880	3360

Table 5: The maturity data in Case 2 and Case 3 (°C·h).

Days	7 d	8 d	9 d	10 d	11 d	12 d	13 d	14 d
Case 2	1680	1920	2160	2400	2640	2880	3120	3360
Case 3	1638	1872	2106	2340	2574	2808	3042	3276

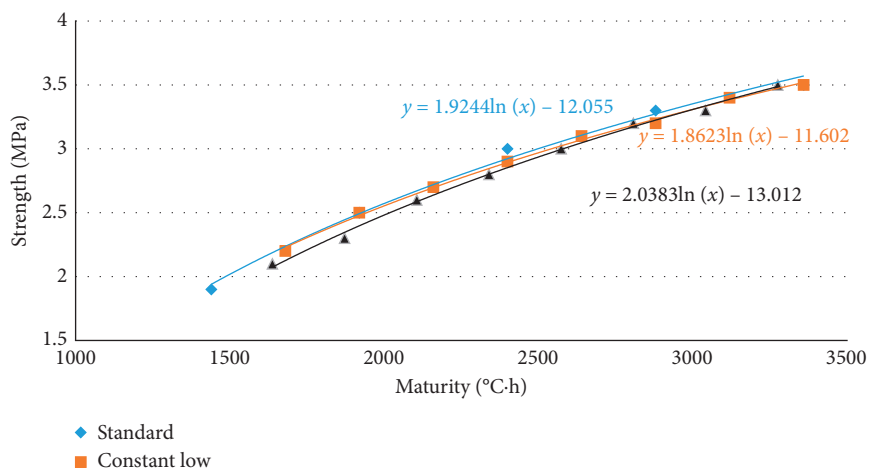


Figure 12: Relationship between maturity and strength.

Table 6: The compressive strength of tested and predicted outdoor curing.

Days	7 d	8 d	9 d	10 d	11 d	12 d	13 d
Maturity (°C·h)	1690	1946	2200	2462	2719	2979	3246
Tested value (MPa)	2.200	2.500	2.700	2.900	3.100	3.300	3.500
Predicted value (MPa)	2.205	2.478	2.715	2.933	3.125	3.302	3.468

the power function, logarithmic function, exponential function, and hyperbolic function [25].

The maturity of three indoor experiments was calculated and shown in Tables 4 and 5. The relationship between the maturity and strength in three cases is shown in Figure 12. It seems that the logarithmic functions $f = a\ln(M) - b$ are the best predictive curves under all three cases, and, therefore, it was used as the preferred function for cement-stabilized macadam mixture. In addition, by combining the data under all three cases and developing a single predictive model, the parameters $a = 1.9358$ and $b = 12.183$ were obtained by fitting the compressive strength and maturity data, and the correlation coefficient was $R^2 = 0.9907$. In short, the Maturity-Strength prediction model of 5% cement-stabilized macadam mixture was $f = 1.9358\ln(M) - 12.183$.

For the outdoor natural curing cases, the central position data was used for maturity calculation, It should be noted that one hour was used as the temperature range, was then accumulated into one day, and was again accumulated across days to get maturity value. Using the obtained function $f = 1.9358\ln(M) - 12.183$ to predict the compressive strength under the outdoor curing case, the results were shown in Table 6. Note that these results were very close to the tested strength and the correlation coefficient was as high as 99.865%, which clearly indicated the high accuracy of the model. According to the model, the compressive strength of low temperature curing can be predicted with maturity, which provides a reference to calculate the strength and determine the schedule of the construction project for engineering applications.

7. Conclusion

The present study discusses the compressive strength gain law of 5% cement-stabilized macadam at low temperature curing, with a particular focus on the varied temperature curing which is similar to the varied air temperature in the real world.

In this paper, experiments under three indoor temperature curing cases and one outdoor natural curing were

conducted. Experimental results showed that the compressive strength increased with the increasing curing time under all four cases and that the gain rate at low temperature was smaller than that at standard temperature. The gain rates can be sorted in descending order: standard temperature curing > outdoor natural temperature curing > constant low temperature curing > day interactive temperature curing. The standard strength reached 3.5 MPa on the 7th day, while the others accounted for only 62% or so. Numerical results also indicated that to achieve the same strength of 3.5 MPa, the number of days required under each low temperature case was 14 days for both the constant low and day interaction temperatures and 13 days for the outdoor temperature.

According to the temperature data and the strength information collected by several indoor tests, an estimated model $f = 1.9358 \ln(M) - 12.183$ was established to predict the strength based on the maturity theory. The model is proved to have the ability to predict with high accuracy based on the validated results obtained from the data of outdoor tests.

As the line of research matures in the future, the characteristics associated with the compressive strength in the long term can also be investigated with more data collected over time.

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

The authors would like to express their sincere gratitude to the National Key R&D Program of China (2017YFF0205600) for financial support.

References

[1] H. C. Wang, X. M. Huang, and Z. Fu, "Experimental study on the performance of semi-rigid base coarse materials," *Journal of Highway and Transportation Research and Development*, vol. 22, no. 11, pp. 45–49, 2005.

[2] ACI Committee, *Cold-Weather Concreting (ACI 306R-88): Reapproved 2002*, American Concrete Institute, Farmington Hills, MI, USA, 1988.

[3] D. P. Bentz, P. E. Stutzman, and F. Zunino, "Low-temperature curing strength enhancement in cement-based materials containing limestone powder," *Materials and Structures*, vol. 50, no. 3, p. 173, 2017.

[4] X. D. Wang, L. Z. F. Zeng, and Q. Xiao, *Technical Guidelines for Construction of Highway Road Bases (JTG/T F20-2015)*, Highway & Transportation Industry Standard, Beijing, China, 2015.

[5] ACI Committee, *Standard Practice for Curing Concrete (ACI 308-92): Reapproved 1997*, American Concrete Institute, Farmington Hills, MI, USA, 1997.

[6] German Committee for Reinforced Concrete, *Recommendation on the Heat Treatment of Concrete*, Deutscher Ausschuss fuer Stahlbeton, Berlin, Germany, 1989.

[7] X. Y. Zhao, L. L. Chen, and Y. Zhao, "Investigation on coring time of road base course in cold region," *Low Temperature Architecture Technology*, vol. 44, no. 8, pp. 99-100, 2011.

[8] K. W. Oh and T. H. Kim, "Dependence of the material properties of lightweight cemented soil on the curing temperature," *Journal of Materials in Civil Engineering*, vol. 26, no. 7, Article ID 06014008, 2013.

[9] I. L. Howard and T. Cost, *Curing Temperature Effects on Soils Stabilized with Portland Cement Having Different Sulfate Contents*, Geo-Congress, Atlanta, GA, USA, 2014.

[10] N. C. Consoli1, C. G. D. Rocha, and C. Silvani, "Effect of curing temperature on the strength of sand, coal fly ash, and lime blends," *Journal of Materials in Civil Engineering*, vol. 26, no. 8, 2014.

[11] R. G. Burg, *The Influence of Casting and Curing Temperature on the Properties of Fresh and Hardened Concrete: PCA Research and Development Bulletin RD113*, Portland Cement Association, New York City, NY, USA, 1996.

[12] ACI Committee, *"Accelerated Curing of Concrete at Atmospheric Pressure (ACI 517.2R-87): Revised 1992*, American Concrete Institute, Farmington Hills, MI, USA, 1992.

[13] D. X. Wang, Z. R. Zentar, and N. E. Abriak, "Temperature-accelerated strength development in stabilized marine soils as road construction materials," *Journal of Materials in Civil Engineering*, vol. 29, no. 5, 2017.

[14] J. I. Escalante-Garci'a and J. H. Sharp, "Effect of temperature on the hydration of the main clinker phases in Portland cements: part II, blended cements," *Cement and Concrete Research*, vol. 28, no. 9, pp. 1259–1274, 1998.

[15] P. Wang, N. Li, and L. Xu, "Hydration evolution and compressive strength of calcium sulphoaluminate cement constantly cured over the temperature range of 0 to 80°C," *Cement and Concrete Research*, vol. 100, pp. 203–213, 2017.

[16] W. H. Price, "Factors influencing concrete strength," *ACI Journal Proceedings*, vol. 47, no. 2, pp. 417–432, 1951.

[17] M. Husem and S. Gozutok, "The effects of low temperature curing on the compressive strength of ordinary and high performance concrete," *Construction and Building Materials*, vol. 19, no. 1, pp. 49–53, 2005.

[18] J.-K. Kim, Y.-H. Moon, and S.-H. Eo, "Compressive strength development of concrete with different curing time and temperature," *Cement and Concrete Research*, vol. 28, no. 12, pp. 1761–1773, 1998.

[19] H. Marzouk and A. Hussein, "Effect of curing age on high-strength concrete at low temperatures," *Journal of Materials in Civil Engineering*, vol. 7, no. 3, pp. 161–167, 1995.

[20] T. Kim and K. L. Rens, "Concrete maturity method using variable temperature curing for normal-strength concrete mixs. II: theoretical study," *Journal of Materials in Civil Engineering*, vol. 20, no. 12, 2008.

[21] R. O. Rasmussen, J. K. Cable, D. J. Turner, and G. F. Voigt, "Strength prediction by using maturity for Portland cement concrete pavement construction at airfields," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1893, no. 1, pp. 18–25, 2004.

[22] A. G. A. Saul, "Principles underlying the steam curing of concrete at atmospheric pressure," *Magazine of Concrete Research*, vol. 2, no. 6, pp. 127–140, 1951.

[23] B. Chitambira, *Accelerated Ageing of Cement Stabilised/Solidified Contaminated Soils with Elevated Temperatures*, Ph.D. thesis, Cambridge University Geotechnical Engineering Group, Cambridge, UK, 2004.

[24] J. H. Jeong, S. Lim, and D. G. Zollinger, "Development of a moisture-modified maturity model for Portland cement

concrete pavements," *Baltic Journal of Road and Bridge Engineering*, vol. 3, no. 1, 2008.

[25] N. J. Carino and H. S. Lew, *The Maturity Method: From Theory to Application: The Structures Congress*, American Society of Civil Engineers, Washington, DC, USA, 2001.