Journal of Advanced Transportation

Management and Control Methods of Mixed Traffic Flow with Connected Automated Vehicles

Lead Guest Editor: Zhihong Yao
Guest Editors: Yang Zhou, Gen Li, and Yanyan Qin

Management and Control Methods of Mixed Traffic Flow with Connected Automated Vehicles

Journal of Advanced Transportation

Management and Control Methods of Mixed Traffic Flow with Connected Automated Vehicles

Lead Guest Editor: Zhihong Yao
Guest Editors: Yang Zhou, Gen Li, and Yanyan Qin

Contents

Contents

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 5879568, 13 pages
https://doi.org/10.1155/2022/5879568

Research Article

Differentiated Speed Planning for Connected and Automated Electric Vehicles at Signalized Intersections considering Dynamic Wireless Power Transfer

Lan Yang,[1] Mengjie Han,[1] Shan Fang,[1] Guoyuan Wu,[2] He Sheng,[1] Heng Wei,[3] and Xiangmo Zhao[1]

[1]*School of Information Engineering, Chang'an University, Xi'an, Shaanxi, China*
[2]*Center for Environmental Research and Technology, University of California at Riverside, Riverside, CA, USA*
[3]*ART-Engines Transportation Research Laboratory, College of Engineering and Applied Science, University of Cincinnati, OH, USA*

Correspondence should be addressed to Mengjie Han; han114336@163.com

Received 13 April 2022; Revised 25 July 2022; Accepted 15 September 2022; Published 10 October 2022

Academic Editor: Zhihong Yao

Deploying the dynamic wireless power transfer (DWPT) equipment at intersections can improve the transportation efficiency and decrease the energy consumption. It can easily turn the disadvantages of waiting time for red phase into the advantages of charging benefits for electric vehicles (EVs). This study develops a multiobjective speed planning model with differentiated charging strategy to optimize the EV's trajectory at signalized intersections with DWPT. To mitigate the negative impact of DWPT, the proposed model is divided into charging priority driving (C-eco-driving) mode and time priority driving (T-eco-driving) mode for connected and automated electric vehicles (CAEVs) to apply different scenarios. Meanwhile, a widely applied car-following model, i.e., the Intelligent Driver Model (IDM), has been calibrated with the ground-truth dataset to meet the intersection characteristic for EVs in mixed traffic. The efficiency and robustness of two eco-driving modes were validated in single-vehicle and mixed traffic based on MATLAB simulation. For single vehicle, the electricity benefits of the C-eco-driving mode and T-eco-driving mode increase about 0.0867 kWh and 0.0532 kWh, respectively. T-eco-driving mode reduces 2.5 s of travel time. For mixed traffic, the C-eco-driving mode provides more charging benefits about 51~73% than no-control strategy. In contrast, the T-eco-driving mode provides significant time benefits of 2.6 s and slight electricity benefits of 12~30% than no-control strategy. The increment of market penetration rate and wireless charging length can significantly improve the charging benefits. When charging length has the same value, the closer the DWPT facilities to the intersection, the more obvious the charging benefits.

1. Introduction

EVs have been recognized as an effective means of achieving energy conservation and sustainable development in recent years [1–4]. Governments have implemented various policies and regulations to encourage the popularization of EVs, including a green-car subsidy program and license plate bias [5]. With this momentum, EVs will phase out fuel vehicles in the near future.

Due to the limited battery capacity and short driving range, EVs require frequent charging. The widespread adoption of EVs may result in an increase in charging demand. As a result, many charging stations, battery swap stations, and wireless charging equipment have been constructed to meet the demands for charging [6–8]. The construction of charging station has low cost, but charging waiting time for EVs is long, and the fast-charging technology can decrease the battery life [9]. The advantage of the battery

swap station is the short charging time; the disadvantage is the unstandardized battery [10]. For wireless charging, EVs can achieve electricity supplements when traveling at the wireless charging lanes. It is more convenient and safer than other charging infrastructures [11–13].

A few studies have concentrated on the competitive analysis of different types of charging infrastructures [14–16]. Based on the charging demands of electric public transit system, Chen et al. [17] established mathematical models to achieve the optimal deployment of various charging facilities, and the DWPT technology is more competitive than charging stations and battery swapping stations with low service frequency and high operating speed. To address the range and recharge issues for EVs, Fuller [18] concluded that dynamic wireless charging can be a more cost-effective approach to extending driving range than increasing battery capacity. Thus, it can be seen that DWPT will play a significant role in future charging technology field.

When deploying the DWPT along ingress lanes at intersections or arterial, it can easily achieve electricity benefits for EVs [19–21]. Therefore, some researches focused on the influence of wireless charging lanes on traffic system. He et al. [22, 23] investigated the impacts of the dynamic charging lane on EV's motion behavior, and the results show that the each EV could run slowly on the charging lane with causing low traffic efficiency. To improve the traffic efficiency, He et al. [24] established a model to optimize the location of the charging lane by considering their adverse effects on road capacity. To address the optimal deployment problem at signalized arterial, Li et al. [25] proposed a biobjective model considering both traffic efficiency and charging facilities' utilization rate. The results show that DWPT equipment is more suitable to be installed near intersections. Based on this theory, Mohrehkesh and Nadeem [26] proposed a wireless charging scheme for the battery electric vehicles and investigated how to integrate control strategies at intersections but have not established a specific speed guidance model for EVs. Subsequently, Zhang et al. [27] proposed an eco-driving control strategy for CAEVs at signalized intersections with charging lane.

To sum up, the DWPT technology can decrease partially traffic operation efficiency but can increase the charging efficiency. To mitigate the negative impacts of DWPT technology, this paper proposed a new scheme for CAEVs at intersections. Two driving control strategies are investigated to apply different scenarios. The charging priority control strategy is more suitable for the following scenarios: (1) long red phase waiting time, (2) low traffic flow, and (3) low state of charge (SOC) for EVs. The time priority control strategy is more suitable for the following scenarios: (1) high traffic flow and (2) high SOC for EVs. The scheme of differentiated strategy can easily reduce disadvantage of DWPT in traffic efficiency, meanwhile significantly improving the advantage of DWPT in electricity benefits. As a result, the speed planning strategy for CAEVs should consider a variety of optimization factors, such as traffic efficiency, electricity consumption, and driver comfort.

In this study, a multiobjective speed planning model based on differentiated control strategy at a signalized intersection with DWPT is proposed. Based on the SOC of EVs entering the control segment, the model is divided into C-eco-driving and T-eco-driving by adjusting three weight coefficients. The Next-Generation Simulation (NGSIM) dataset is used to calibrate the car-following model for the EV in mixed traffic. Numerical studies of single-vehicle and mixed traffic are used to validate the efficiency and adaptability of the speed planning model.

The remainder of this paper is composed as follows. Section 2 provides the scenario description for EVs traversing intersections with DWPT. Section 3 describes the differentiated speed planning model and the calibrated IDM. Next, the single-vehicle and mixed-traffic examples are carried out to demonstrate the efficiency and adaptability of the proposed model in Section 4. Finally, the concluding remark is presented in Section 5.

2. Scenario Description

As shown in Figure 1, a signalized intersection with DWPT is established, and vehicle queues consisting of CAEVs and human-driven EVs pass through the intersection. The total control segment's start position is O, and its length is S. $\mathrm{WCL}(S_1, S_2)$ denotes that the start position of deploying the wireless charging transfer is S_1, and the terminal location is S_2. The length of wireless charging is $(S_2 - S_1)$. The position of the intersection stop line is S_3. When the CAEV enters the control segment of a signalized intersection, it is assumed that it can obtain certain information via V2I and V2V communications, such as the distance to the stop bar, SPaT information, the location and length of the charging area, and the states of preceding vehicles, which are not available to regular EVs. When mixed vehicles enter the control segment, the CAEV will follow the proposed model's trajectory planning advice to navigate the intersection smoothly and avoid stopping. In contrast, the regular EV will follow the designated car-following model. Two distinct planning strategies for CAEV are considered in this study in order to meet varying demands. As illustrated in Figure 1, low-battery CAEVs may prefer the charging priority driving strategy, whereas high-battery CAEVs may prefer the time priority driving strategy. The former is concerned with electricity, whereas the latter is concerned with time and driving comfort.

3. Two Models at a Signalized Intersection with DWPT

3.1. Differentiated Speed Planning Model for the CAEV

3.1.1. Multiobjective Optimization. To guide CAEVs passing through signalized intersections safely and efficiently, the model optimization objective sets three factors: traffic efficiency, electricity consumption, and driver comfort. Since three factors have different units and dimensions and orders of magnitude, when three factors are directly added together, the role of the factor with a larger value in the objective function will be highlighted. Therefore, this paper standardizes them to eliminate the influence of different dimensions.

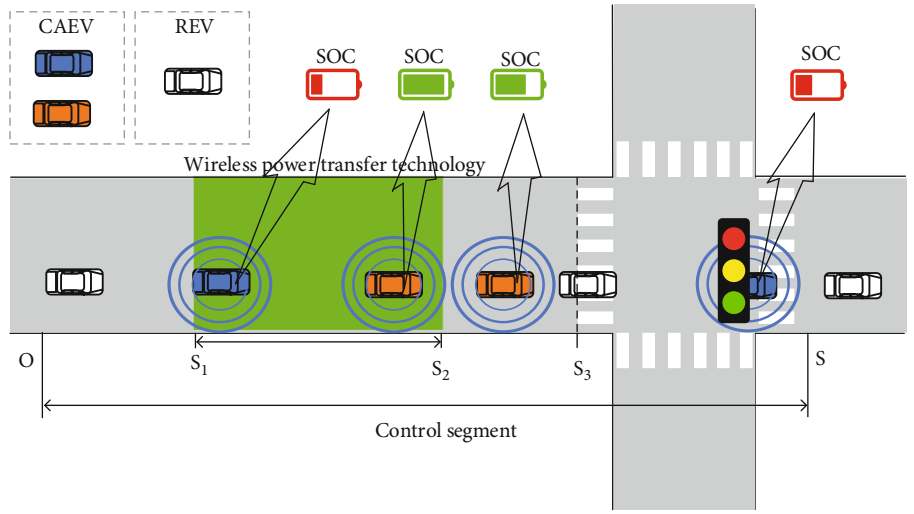


Figure 1: The scenario with differentiated driving strategy at the signalized intersection with the DWPT.

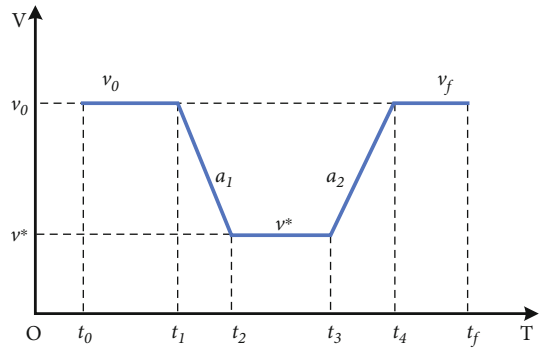


Figure 2: Entire process of speed planning.

Table 1: Weighting parameters of two driving modes.

Parameters	C-eco-driving	T-eco-driving
ω_1	0.1~0.2	0.4~0.6
ω_2	0.6~0.8	0.1~0.2
ω_3	0.1~0.2	0.2~0.4

(1) Traffic efficiency optimization

This study improves the efficiency of traffic flow by reducing the guidance time of vehicles in the control segment, while avoiding the behavior of stop-go at signalized intersections. Therefore, the efficiency optimization formula is

$$F_t = \left(t_f - t_0 - T_{\min}\right)/\left(T_{\max} - T_{\min}\right), \tag{1}$$

$$T_{\max} = \max_{i=1\cdots n}\left\{t_f(i) - t_0(i)\right\}, \tag{2}$$

$$T_{\min} = \min_{i=1\cdots n}\left\{t_f(i) - t_0(i)\right\}, \tag{3}$$

where F_t is the efficiency objective value; t_0 and t_f are the initial time and terminal time of the control segment; $T_{\min}$ and $T_{\max}$ are the shortest and longest travel times in all historical trajectories, respectively; and n is the number of historical trajectories.

(2) Electricity consumption optimization

The vehicle's SOC is not only influenced by the electricity loss of the traction but also supplemented by DWPT facility and the regenerative braking system (RBS). It can be formulated as follows:

$$\begin{aligned} F_e &= \frac{\left(E_{\mathrm{NC}} - E_{\mathrm{Charge}} - E_{\min}\right)}{\left(E_{\max} - E_{\min}\right)}, \\ E_{\mathrm{NC}} &= E_{\mathrm{Consumption}} - E_{\mathrm{Recovery}}, \\ E_{\mathrm{Charge}} &= P_{\mathrm{Charge}} T_{\mathrm{Charge}}, \\ E_{\max} &= \max_{i=1\cdots n}\left\{E_{\mathrm{NC}}(i) - E_{\mathrm{Charge}}(i)\right\}, \\ E_{\mathrm{mix}} &= \min_{i=1\cdots n}\left\{E_{\mathrm{NC}}(i) - E_{\mathrm{Charge}}(i)\right\}, \end{aligned} \tag{4}$$

where F_e is the electricity objective value; E_{NC} is the electricity consumption that has considered RBS; and E_{Charge} is the charging electricity with DWPT. $E_{\min}$ and $E_{\max}$ are the minimum and maximum total electricity consumption in all historical trajectories, respectively.

(3) Driver comfort optimization

This study optimizes the driver comfort by reducing the speed fluctuation of CAEVs during the control segment. The formula is shown as follows:

$$F_s = \frac{\left(\int_{t_0}^{t_f} a^2(t)dt - U_{\min}\right)}{\left(U_{\max} - U_{\min}\right)}, \tag{5}$$

1 Step 1. Data preprocessing
2 Remove exception data and out-of-bound data from the original data.
3 Select continuous four state vectors for FVs: speed, space difference, speed difference, and acceleration.
4 Unit conversion, such as ft, ft/s to m, and m/s.
5 Step 2. Initialization
6 Randomly generate the initial values for five calibrated parameters within the bound, a, b, s_0, T, and v_0.
7 Initialize GA basic parameters, such as maximum iteration, population size, and crossfactor.
8 Step 3. Fitness function
9 Estimate the acceleration (predictive value) for FV at the next time point using Equations (17) and (18).
10 Evaluate RMSE between predictive value and true value with Equation (19).
11 Step 4. Update five calibrated parameters
12 Update five parameters with fitness function.
13 Step 5. Termination condition
14 If the number of iteration is larger than the maximum value or RMSE is not larger than the specified gap, stop the algorithm. Otherwise, go back to Step 3.

Algorithm 1: IDM parameter calibration algorithm based on GA.

Table 2: IDM parameter calibration results.

Parameters	Description	Bounds	Calibrated
a (m/s^2)	Acceleration	[0.1, 5]	4.1
b (m/s^2)	Deceleration	[0.1, 5]	3.7
s_0 (m)	Minimum distance	[0.1, 10]	5.68
T (s)	Time gap	[0.1, 5]	1.5
v_0 (km/h)	Desired speed	[1, 150]	72.3

Table 3: Parameters used in simulations.

Parameters	Values	Parameters	Values
$v_{\max}$ (m/s)	22	η_d	0.92
$v_{\min}$ (m/s)	2.00	η_m	0.91
$a_{\max}$ (m/s^2)	4.88	η_b	0.90
$a_{\min}$ (m/s^2)	-3.41	α	0.0411
m (kg)	1521	P_{Charge} (kW)	20
θ	0	W (kWh)	60
C_r	1.75	S_3 (m)	500
c_1	0.0328	S (m)	550
c_2	4.575	T_{red} (s)	35
A_f (m^2)	2.3316	T_{green} (s)	45
C_D	0.28	t_{entry} (s)	0
Threshold	0.5		

$$U_{\max} = \max_{i=1\cdots n}\left\{\int_{t_0(i)}^{t_f(i)} a^2(t)dt\right\}, \tag{6}$$

$$U_{\min} = \min_{i=1\cdots n}\left\{\int_{t_0(i)}^{t_f(i)} a^2(t)dt\right\}, \tag{7}$$

where F_s is the stability objective value and $U_{\min}$ and $U_{\max}$ are the minimum and maximum speed fluctuations in all historical trajectories, respectively.

(4) Three-objective optimization function

Based on the above, we integrate three optimizations by applying three weighting coefficients:

$$\begin{aligned} \min_{a(t)} L &= \omega_1 F_t + \omega_2 F_e + \omega_3 F_s, \\ \omega_1 + \omega_2 + \omega_3 &= 1, \\ 0 < \omega_i < 1, &\quad i = 1, 2, 3, \end{aligned} \tag{8}$$

where L is the objective value and ω_1, ω_2, and ω_3 are weights for different objectives, presenting the importance of time, electricity, and stability, respectively. Three weighting coefficients can be dynamically adjusted to achieve differentiated travel demands.

3.1.2. Dynamic Constraints. The objective function for CAEV is subject to the following constraints:

(1) Motion state definitions and initial state declarations are formulated as follows:

$$\begin{aligned} \dot{x}(t) \triangleq [v(t), a(t)]^T = f(x(t), u(t)), &\quad t_0 \le t \le t_f, \\ s(t_0) = 0, v(t_0) = v_0, & \end{aligned} \tag{9}$$

where $x(t)$ is the state vector of the EV at time t and $s(t)$, $v(t)$, and $a(t)$ are distance, speed, and acceleration at time t, respectively. Because the detector is set at the starting point of the control segment, the initial coordinate for CAEVs is 0.

(2) The vehicle distance constraint is defined as the distance characteristic that the vehicle satisfies within the control segment:

$$s(t_{s1}) = S_1 \quad s(t_{s2}) = S_2, \tag{10}$$

$$s(t_{s3}) = S_3 \quad s(t_f) = S, \tag{11}$$

Table 4: Parameter setting of eight scenarios.

Scenario	Initial speed (m/s)	WCL (m)	Scenario	Initial speed (m/s)	WCL (m)
A	16	(50, 250)	E	20	(50, 150)
B	18	(50, 250)	F	20	(50, 250)
C	20	(50, 250)	G	20	(50, 350)
D	22	(50, 250)	H	20	(50, 450)

$$s_k(t) < s_{k-1}(t), \tag{12}$$

where t_{s1} and t_{s2} are the entry time and exit time in the wireless charging area, respectively, and t_{s3} is the arrival time at the stop line S_3. Equation (12) is the safe distance constraint.

(3) Vehicle kinematic constraints declare the physical boundaries of vehicle speed and acceleration:

$$\begin{aligned} v_{\min} \le v(t) \le v_{\max}, \\ a_{\min} \le a(t) \le a_{\max}, \end{aligned} \tag{13}$$

where $v_{\min}$ and $v_{\max}$ are the maximum speed, respectively, and $a_{\min}$ and $a_{\max}$ are minimum and maximum accelerations. The maximum speed is the legal limit on the road, and the minimum speed is a reasonable threshold to avoid the vehicle traveling at very low speeds in the charging area.

(4) Traffic state constraint that ensures the EV can pass through the signalized intersection during the green time:

$$\begin{aligned} t_{\min} \le t_{s3} \le t_{\max}, \\ t_{\min} = T_r + kT, t_{\max} = T_r + G + kT \text{ if signal}_{t_0} = \text{red}, \\ t_{\min} = 0, t_{\max} = T_g \text{ if signal}_{t_0} = \text{green}, \\ t_{\min} = T_g + R + kT, t_{\max} = T_g + (k+1)T \text{ if signal}_{t_0} = \text{green}, \end{aligned} \tag{14}$$

where $k = 0, 1, 2 \cdots$, $t_{\min}$ and $t_{\max}$ are the shortest time and the longest time, respectively; T_r and T_g are the red light or green light duration when the vehicle enters the segment; R and G are the red and green periods, respectively; T is the traffic phase period; and signal_{t_0} is the phase state at the initial time.

3.1.3. Model Regularization. To achieve high computational efficiency for real-time control, this study numerically discretizes time and space dimensions into a series of collocation points using metaheuristics or gradient-based methods and regularizes the speed profile of the entire process by consulting reference [27, 28]. The entire speed planning process can be divided into five phases, as illustrated in Figure 2.

(1) Cruise at the initial speed v_0 from t_0 to t_1

(2) Decelerate at a constant rate a_1 from t_1 to t_2

(3) Cruise at a specific speed v^* from t_2 to t_3

(4) Accelerate at a constant rate a_2 from t_3 to t_4

(5) Cruise at the terminal speed v_f from t_4 to t_f

The initial speed v_0 is assumed to be the same value with v_f in the control segment, which contents the characteristics of the wireless charging scheme.

$$\begin{aligned} t_0 \le t_1 \le t_2 \le t_3 \le t_4 \le t_f, \\ v_0 = v_f, \\ a_1 = \frac{2(S_3 - v_0(t_{s3} - t_0))}{(t_2 - t_1)(t_4 + t_3 - t_2 - t_1)}, \\ a_2 = -\frac{t_2 - t_1}{t_4 - t_3} a_1, \\ v^* = v_0 + (t_2 - t_1)a_1, \end{aligned} \tag{15}$$

where v^* is the minimum speed of the control segment. a_1 and a_2 are deceleration and acceleration, respectively. Therefore, Equation (8) can be simplified, as shown in Equation (16).

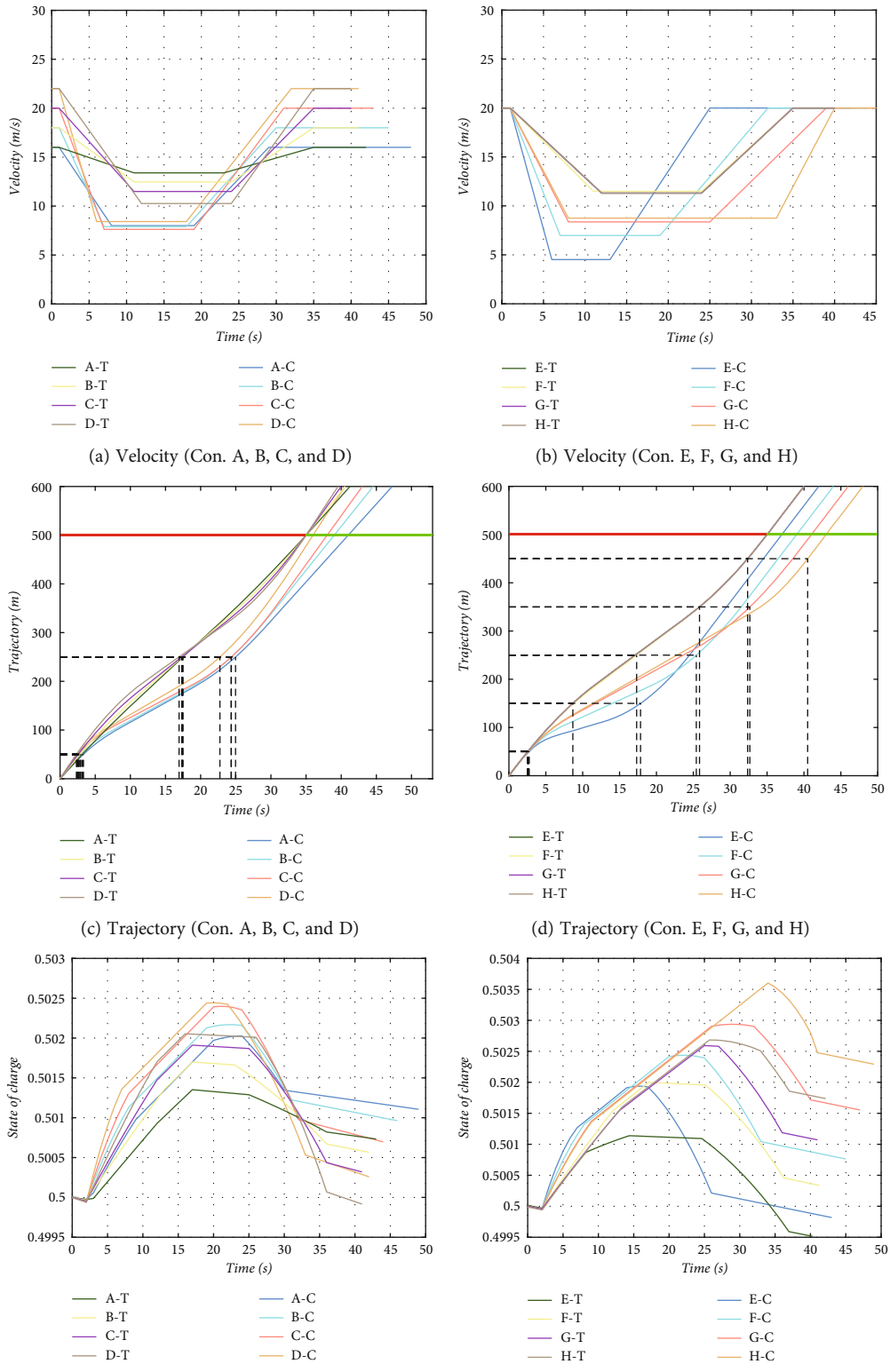
$$\begin{cases} \min\limits_{t_1, t_2, t_3, t_4} L = \omega_1 \left(\dfrac{t_f - t_0 - T_{\min}}{T_{\max} - T_{\min}} \right) + \omega_2 \left(\dfrac{E_{\text{NC}} - E_{\text{Charge}} - E_{\min}}{E_{\max} - E_{\min}} \right) + \omega_3 \left(\dfrac{\int_{t_0}^{t_f} a^2(t)dt - U_{\min}}{U_{\max} - U_{\min}} \right) \\ \text{s.t.Equations (9) – (15)} \end{cases} \tag{16}$$

The proposed model is solved by the Particle Swarm Optimization (PSO) algorithm [29]. The comprehensive power-based energy consumption model is used to estimate the energy consumption [30].

3.1.4. Differentiated Travel Mode. The value of weighting coefficient represents different driving strategies. By adjusting the values of ω_1, ω_2, and ω_3, the proposed model is divided into C-eco-driving mode and T-eco-driving mode. Table 1 shows the value range of two modes. The coefficient is set to a range of values rather than fixed values to ensure dynamicity.

For C-eco-driving mode, when the SOC of CAEV entering to the segment is below the threshold, the strategy focuses on the charging efficiency. Thus, ω_2 is set as primary weighting coefficient (0.6~0.8), as the upper bound of C-eco-driving mode; ω_1 and ω_3 are set as secondary weighting coefficient (0.1~0.2), as the lower bound of C-eco-driving mode. Only considering electricity ($\omega_1\omega_2, \omega_3 = 0, 1, 0$) is likely to cause the vehicle to remain stationary in the charging area, which makes the model unsolvable.

For the T-eco-driving mode, when the SOC of CAEV entering to the segment is above the threshold, the strategy focuses on the traffic efficiency. ω_1 is set as the first



(a) Velocity (Con. A, B, C, and D)

(b) Velocity (Con. E, F, G, and H)

(c) Trajectory (Con. A, B, C, and D)

(d) Trajectory (Con. E, F, G, and H)

(e) State of charge (Con. A, B, C, and D)

(f) State of charge (Con. E, F, G, and H)

FIGURE 3: Results of eight scenarios.

weighting coefficient (0.4~0.6) to achieve high traffic efficiency. ω_3 is set as the second weighting coefficient (0.2~0.4) to avoid the occurrence of large speed fluctuations. ω_2 is set as the third weighting coefficient to ensure that speed fluctuation is adjusted in the direction of increasing electricity benefits.

Table 5: Numerical results under eight scenarios.

Con.	t_f (s)	E_{cha} (Ws)	E_{con} (Ws)	E_{rec} (Ws)	Con.	t_f (s)	E_{cha} (Ws)	E_{con} (Ws)	E_{rec} (Ws)
A-C	45.0	430612	274757	96390	A-T	39.0	286859	146650	26886
B-C	42.0	425701	360127	131887	B-T	38.0	291260	245521	75750
C-C	43.0	471664	473915	180412	C-T	38.0	295516	354387	128464
D-C	39.0	404445	543897	206228	D-T	38.0	291191	481342	184875
E-C	43.0	365223	535568	189656	E-T	38.0	123058	356885	129217
F-C	43.0	471664	473915	180412	F-T	38.0	295516	354387	128464
G-C	43.0	576075	416312	161602	G-T	38.0	465161	356885	129217
H-C	43.0	691393	378857	145336	H-T	38.0	595830	356885	129217

Table 6: Attribute settings in mixed traffic.

Attributes	Value
Initial speed of the EV	16~22 m/s
Initial speed of the CAEV	18~20 m/s
Initial headway	30~60 m
Penetration rate	0~100%
Wireless charging length	100~400 m

Table 7: Parameter settings of six numerical examples.

Condition	MPR (%)	WCL (m)	Condition	MPR (%)	WCL (m)
I	0	(50, 250)	IV	60	(50, 250)
II	20	(50, 250)	V	80	(50, 250)
III	40	(50, 250)	VI	100	(50, 250)

ω_3 is not set as the first weighting coefficient as the comfort priority mode. The reason is that only considering driver comfort without considering travel time and SOC does not satisfy the real demands.

3.2. Calibrated Car-Following Model for the EV. The IDM [31] is an accident-free model capable of producing realistic acceleration profiles, and it can accurately simulate various traffic flow scenarios. Therefore, this study applies IDM to plan the trajectory for EVs. As defined in Equation (17) and Equation (18):

$$\dot{v} = a\left[1 - \left(\frac{v}{v_0}\right)^{\delta} - \left(\frac{s^*(v,\Delta v)}{s}\right)^2\right], \tag{17}$$

$$s^*(v,\Delta v) = s_0 + \max\left(0, vT + \frac{v\Delta v}{2\sqrt{ab}}\right), \tag{18}$$

where v, a, and b are the speed, acceleration, and deceleration of the following vehicle, respectively; s, s_0, and s^* are the current distance, minimum distance, and desired distance, respectively; v_0 is the desired speed; δ is the acceleration; Δv is the speed difference between the preceding vehicle and following vehicle; and T is the time gap that means following vehicle's reaction time.

Model calibration refers to parameter calibration, and different parameters typically reflect different driving characteristics. To meet the intersection characteristic, the NGSIM dataset [32], collected from a segment of Lankershim-Boulevard in Los Angeles, California, is used to calibrate the IDM. A genetic algorithm [33] is used to calibrate the five major parameters of IDM, such as a, b, s_0, T, and v_0. The Root Mean Square Error (RMSE) was used evaluate the difference between predicted values and true values, whose formula is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{a}_i - a_i)^2}, \tag{19}$$

where $\hat{y} = \{\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n\}$ are predicted values; $y = \{a_1, a_2, \cdots, a_n\}$ are true values; and n denotes the number of all data points.

Based on the preceding discussion, Algorithm 1 presents the pseudocode for the calibration method. Table 2 shows the calibration results for the IDM. The bounds of calibrated parameters refer to the literature [34].

4. Numerical Studies

To validate the effectiveness and adaptability of the proposed model, single-vehicle and mixed traffic is performed under various numerical examples based on MTALB simulation. Table 3 shows the basic parameters of numerical simulations.

4.1. Single-Vehicle Simulation under Different Initial Speeds and Wireless Charging Lengths. The initial speed of vehicles and wireless charging length (WCL) are used as variables in single-vehicle simulation. This paper built eight scenarios to evaluate the traffic benefits of the proposed model. As shown in Table 4, scenarios A, B, C, and D are used to compare each other with different initial speeds when the WCL is fixed. Scenarios E, F, G, and H are used to compare each other with different WCLs when the initial speed is fixed.

Figure 3 depicts the velocity, trajectory, and SOC of CAEVs. According to Figure 3(a), the vehicle with the T-eco-driving mode has higher speed than that with the C-eco-driving mode in total, and the CAEV with the T-eco-driving mode has less speed fluctuation than that with the

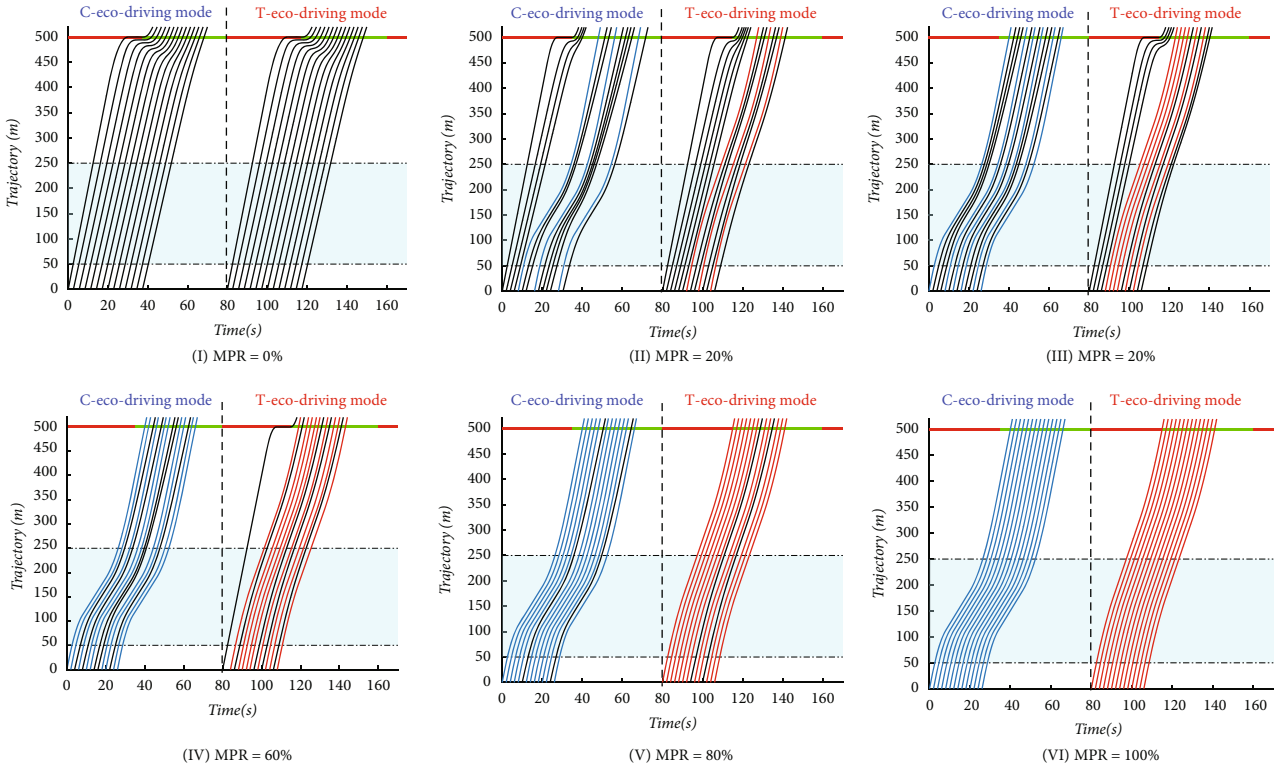


Figure 4: Spatial-temporal trajectories of six conditions (I~VI).

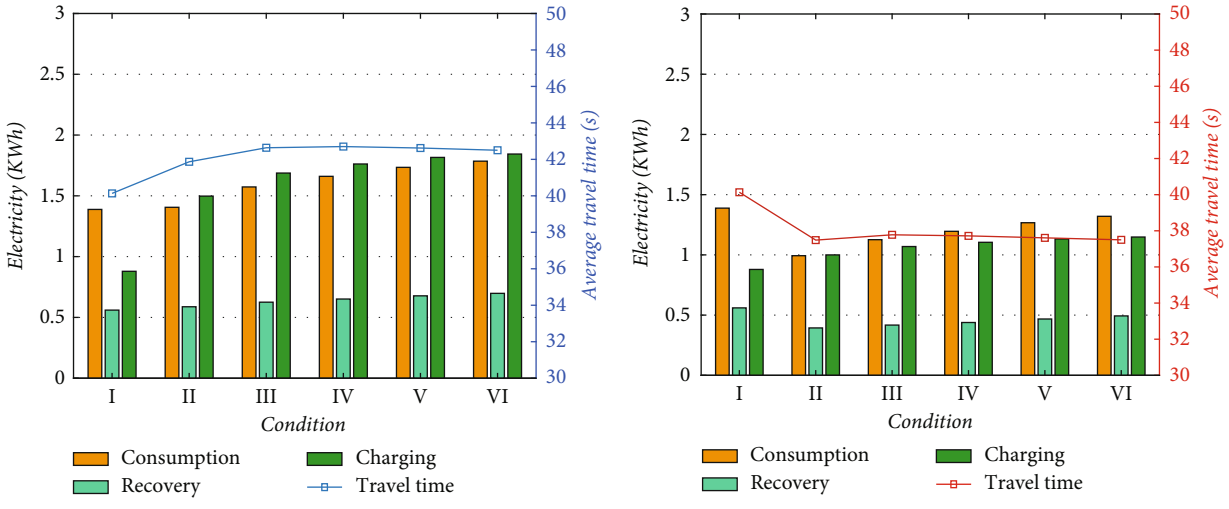


Figure 5: Electricity usage (consumption, recovery, and charging) and average travel time of six conditions (I~VI).

Table 8: Parameter settings of four numerical examples.

Condition	MPR (%)	WCL (m)	Condition	MPR (%)	WCL (m)
VII	40	(50, 150)	IX	40	(50, 350)
VIII	40	(50, 250)	X	50	(50, 450)

C-eco-driving mode. It demonstrates the significance of time and stability for the T-eco-driving mode. More specifically, for the T-eco-driving mode, the minimum speed v^* within the control segment declines when the initial speed continuously increases. However, the minimum speed v^* remains relatively stable as the initial speed increases for the C-eco-driving mode. The main reason is that the weighting coefficient ω_3 is different. From Figure 3(b), for the T-eco-driving mode, the speed of four scenarios basically overlaps. For the

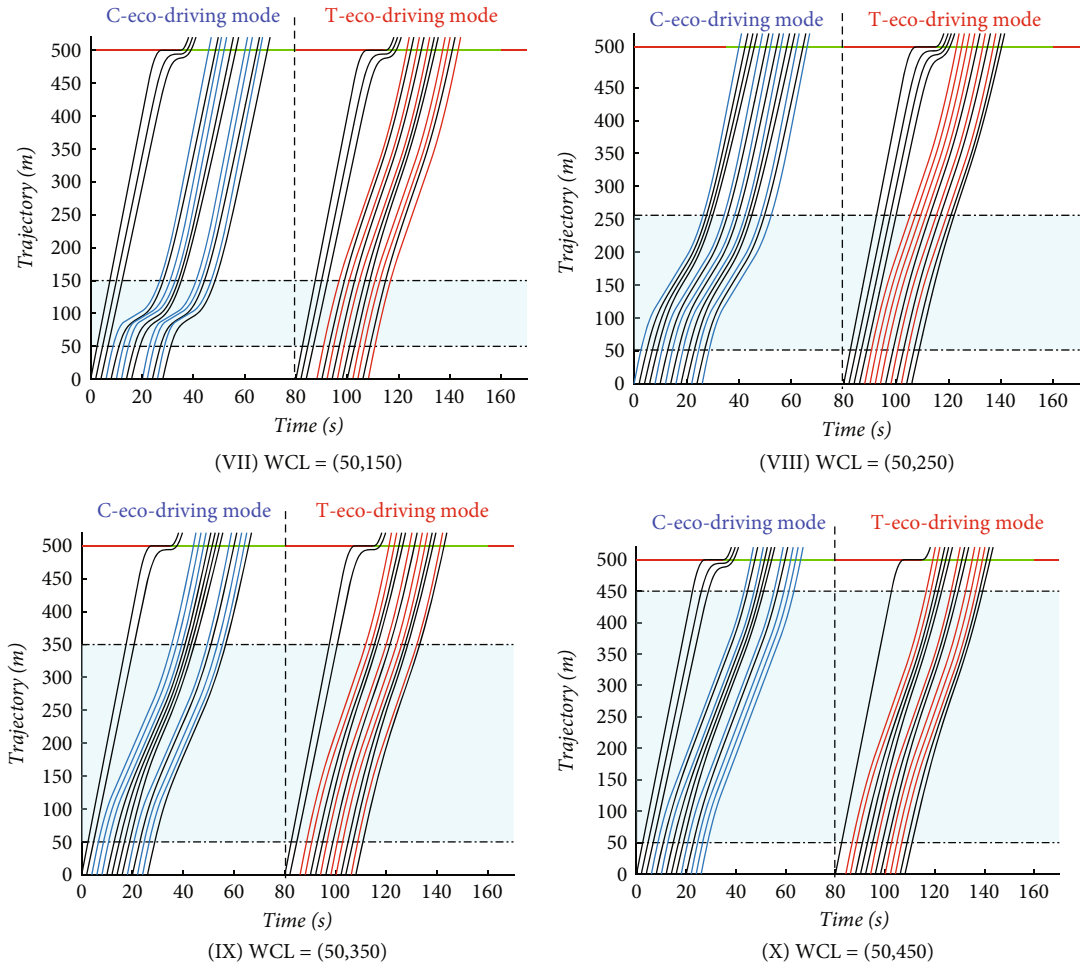


Figure 6: Spatial-temporal trajectories of four conditions (VII~X).

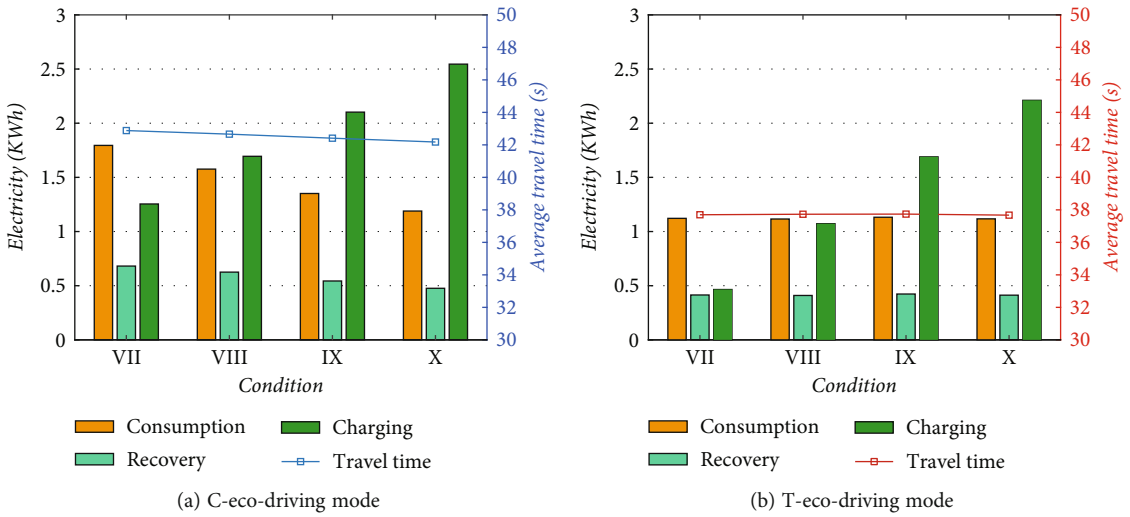
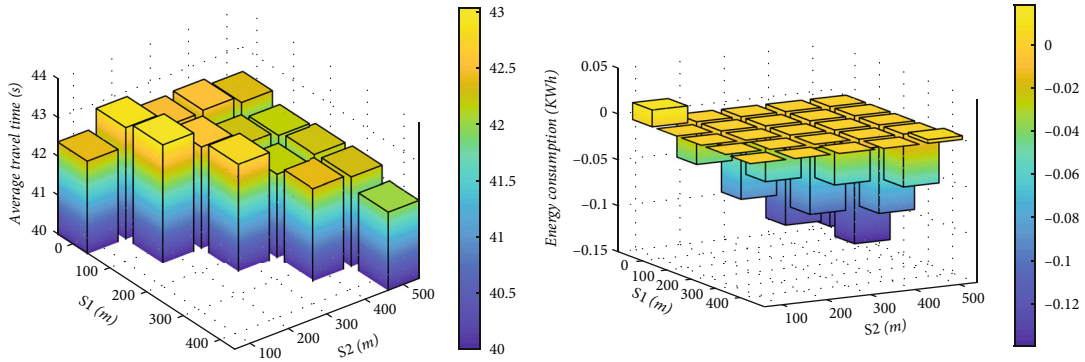


Figure 7: Electricity usage (consumption, recovery, and charging) and average travel time under four conditions (VII~X).

C-eco-driving mode, the value of v^* maintains an upward trend and cruise time at the speed v^* becomes longer with the extension of the wireless charging area (WCA). It is easily inferred that the upward is limited.

Figures 3(c) and 3 (d) depict the trajectory results of eight scenarios for the CAEV. The dotted lines indicate the point of entering and exiting the charging area. As shown in Figure 3(c), for the T-eco-driving mode, the vehicle



(a) Average travel time (b) Total electricity consumption

FIGURE 8: Traffic benefits of different charging locations in mixed traffic.

arrives at the intersection in 35 seconds, and the total time spent traversing the control segment is approximately 40 seconds. For the C-eco-driving mode, the arrival time at the intersection is 36 s, 39 s, 41 s, and 43 s, respectively. The total time is approximately 45 seconds. It implies that the vehicle with the T-eco-driving mode takes less time to pass through an intersection than the C-eco-driving mode. According to Figure 3(d), trajectories under four scenarios are consistent due to speed profile overlap for the T-eco-driving mode. The charging duration of the CAEV increases with the extension of the charging lane for the C-eco-driving mode. Similarly, the CAEV with the T-eco-driving mode takes less time than the C-eco-driving mode.

Furthermore, Figures 3(e) and 3 (f) depict the SOC of the CAEV. The battery's maximum capacity is assumed to be 60 kWh, and the initial SOC is 0.5. Overall, the CAEV with the C-eco-driving mode has more electricity than the T-eco-driving mode, indicating the importance of charging electricity for the C-eco-driving mode. Based on the RBS system, the SOC of the CAEV regularly rises when the vehicle operates rapidly deceleration or drives in a wireless charging area, and SOC falls when the vehicle accelerates rapidly. According to Figure 3(e), regardless of which eco-driving mode the CAEV uses, the terminal SOC rises with the increase of the initial speed. It implies that the initial speed has a significant impact on the model's performance. As illustrated in Figure 3(f), the longer the WCL, the greater the charging benefits. The DWPT technology has the potential to significantly promote electricity for two eco-driving modes. Numerical results show that the benefit in electricity replenishment for the C-eco-driving mode is 0.114~0.216 kWh, and the benefit for the T-eco-driving mode is 0.066~0.162 kWh.

Detailed numerical results under eight scenarios have been shown in Table 5. Four metrics are used for evaluation. t_f is the terminal time; E_{cha} is the recharge energy within the charging area; E_{con} is the energy consumption; E_{rec} is the recovered energy due to RBS.

4.2. Mixed-Traffic Simulation with Different Market Penetration Rates, Charging Lengths, and Locations. In mixed-traffic simulation, the market penetration rate (MPR) of CAEV, charging lengths, and locations are used as variables to analyze the adaptability of the proposed model. As shown in Table 6, the fluctuation of the initial status is considered to better evaluate the robustness of the model.

4.2.1. Adaptability Analysis of Different MPRs. To analyze the traffic benefits of different MPRs, six numerical examples are used to validate the model adaptability. As shown in Table 7, when charging area is set to (50, 250) as the constant value, the MPR is set to the variable value ranging from 0% to 100%.

Figure 4 depicts the spatial-temporal trajectories of approximately 28 vehicles under six conditions (I~VI). The trajectories of vehicles controlled by the C-eco-driving mode, T-eco-driving mode, and calibrated IDM are shown in blue, red, and black lines, respectively. In general, at a higher penetration rate, many regular EVs are more likely to drive on suboptimal trajectories due to car-following characteristics, and these EVs could take advantage of DWPT and travel on a suboptimal trajectory with a low delay. However, the benefit of a moderate penetration rate is highly fluctuant. For instance, in the left of Figure 4 (III), the lead vehicle is a CAEV which can advantageously guide rear vehicles. In contrast, the lead vehicle is a regular EV which cannot benefitly guide following vehicles in the right of Figure 4 (III). As a result, benefit performance is strongly related to the type of the lead vehicle.

Figure 5 shows the average travel time and electricity usage under six conditions for two eco-driving modes. As shown in Figure 5(a), for the C-eco-driving mode, as MPR increases, electricity consumption slowly rises from 1.39 kWh to 1.84 kWh, and electricity recovery remains stable at 0.56~0.69 kWh. Compared to scenario I, electricity charging in scenarios II~VI increases dramatically, with a growth rate ranging from 70.4% to 100.1%. It demonstrates the advantage of the C-eco-driving mode when recharging electricity. Under scenarios II~VI, the average travel time increases slightly; it is unavoidable to sacrifice a little travel time due to the slow speed of vehicles in exchange for more charge. For the T-eco-driving mode from Figure 5(b), compared to scenario I, the electricity consumption and recovery under scenarios II~VI decrease certainly, by approximately 4.9~28.2% and 11.9~30.1%, respectively; the consumption

is proportional to the recovery due to the coexistence of acceleration and deceleration. Electricity charging rises slightly about 14.3~30.5% with the MPR. Then, average travel time under scenarios II~VI decreases by approximately 2.6 s when compared to scenario I. It is the benefit of the T-eco-driving mode in terms of traffic efficiency.

4.2.2. Adaptability Analysis of Different Charging Lengths. Similarly, four numerical examples are used to analyze the model adaptability of different charging lengths, as shown in Table 8. When MPR is set to 40% as the constant value, charging length is set to the variable value.

The spatial-temporal trajectories under four conditions are shown in Figure 6. With the increase of WCL in the C-eco-driving mode, the vehicle's trajectory becomes completely smooth, and the fluctuation gradually decreases. The shorter charging area could cause the vehicle to drive slowly in a short period. In other words, a larger charging area can provide more charging benefits as well as a more pleasant driving experience. The trajectory in the T-eco-driving mode has not changed significantly with the WCL.

Figure 7 illustrates the electricity consumption and average travel time for two eco-driving modes under four different conditions (VII~X). For the C-eco-driving mode, the electricity consumption drops steadily with the extension of the charging lane, and the reason is that the longer the WCL, the less fluctuation in speed, as illustrated in Figure 6. The recovery of electricity is also slightly reduced due to the coexistence of acceleration and deceleration. Additionally, charging capacity increases dramatically from 1.25 kWh to 2.56 kWh as a result of the WCL extension. When the WCL is increased from 100 m to 400 m, the average travel time remains stable at about 42.5 s. This demonstrates that increasing the WCL can significantly improve energy efficiency while maintaining the same travel time. For the T-eco-driving mode, electricity consumption and recovery remain basically stable as the WCL changes in response to the trajectory's fluctuation. Similarly, according to the WCL, electricity charging increases by approximately 0.46~2.21 kWh, and the average travel time is nearly 38 s. Additionally, we compare the electricity charging and travel time between two eco-driving modes. On the one hand, while both have an upward trend in terms of charging, the C-eco-driving mode has a greater increase than the T-eco-driving mode. The T-eco-driving mode, on the other hand, has a shorter travel time than the C-eco-driving mode. Thus, both C-eco-driving and T-eco-driving modes successfully accomplish their respective objectives.

4.2.3. Adaptability Analysis of Different Charging Locations. To analyze the impact of different locations of the DWPT facilities on traffic benefits, four charging lengths (100, 200, 300, and 400 m) are used to validate the adaptability of the proposed model. This study moves the same charging length by a distance of 100 m, which simulates 15 numerical examples.

Traffic benefits of different charging locations in mixed traffic are shown in Figure 8. Firstly, for the average travel time of vehicles, when the charging area is (100, 200), the average travel time is 43.03 s. When the charging area is (400, 500), the average travel time is about 42.01 s. It can be seen that when the length has the same value, the closer the charging area is to the intersection, the shorter the average travel time. Secondly, for total electricity consumption, only when the charging area is (0, 100) and (400, 500), the electricity consumption is the positive value, and EVs passing through signalized intersections do consume additional battery electricity. Finally, the results of the three-dimensional histogram show that when the lengths are the same, the closer the DWPT facilities to the intersection, the more obvious the electricity benefits of EVs. Both evaluation indicators show that the DWPT is more profitable when it is set close to the intersection.

5. Conclusion

In this paper, a multiobjective speed planning model for CAEVs with the C-eco-driving model and T-eco-driving model is proposed, which can provide a differentiated speed guidance advice for meeting the driving strategy at signalized intersections with consideration of DWPT. The IDM calibrated by the NGSIM dataset is then used to calculate the trajectory for regular EVs. Numerical examples are used to demonstrate the proposed model's effectiveness and robustness at various scenarios of signalized intersections. For single vehicle, the T-eco-driving mode produces less speed fluctuation and travel time than the C-eco-driving mode. The C-eco-driving mode produces a more competitive charging benefit than the T-eco-driving mode. We note that adjusting the wireless charging length does not influence the results of the T-eco-driving mode. For mixed traffic, the C-eco-driving mode significantly outperforms the no guidance model in terms of charging benefit. The T-eco-driving mode significantly outperforms the no guidance model in terms of electricity consumption and average travel time. It is worth noting that the benefits of the moderate MPR have certain fluctuations and are influenced by the type of the leading vehicle. The simulation results indicate that the proposed model has a significant improvement of traffic benefits. When charging length has the same value, the closer the DWPT facilities to the intersection, the more obvious the charging benefits. Additionally, the coexistence of two eco-driving modes and human driving will result in a mutual game effect in mixed traffic, which requires extensive discussion and is the direction of future research. How to improve the cost competitive of DWPT and encourage the government as well as private enterprises to construct more DWPT facilities is also a challenge for future work.

Data Availability

The Lankershim dataset, provided by the Federal Highway Administration's NGSIM program, is used to calibrate the IDM model at signalized intersections in this study. The dataset includes detailed vehicle and road information for micro traffic flow research (http://ops.fhwa.dot.gov/traffiffifficanalysistools/ngsim.htm 2006).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (No. 61703053), the National Key Research and Development Program of China (2021YFB2501200), Shaanxi Provincial Key R&D Program (2022GY-300), and the Joint Laboratory for Internet of Vehicles and Ministry of Education-China Mobile Communications Corporation.

References

[1] J. Coria, J. Bonilla, M. Grundström, and H. Pleijel, "Air pollution dynamics and the need for temporally differentiated road pricing," *Transportation Research Part A: Policy and Practice*, vol. 75, pp. 178–195, 2015.

[2] G. J. Offer, D. Howey, M. Contestabile, R. Clague, and N. P. Brandon, "Comparative analysis of battery electric, hydrogen fuel cell and hybrid vehicles in a future sustainable road transport system," *Energy Policy*, vol. 38, no. 1, pp. 24–29, 2010.

[3] J. Du and D. Ouyang, "Progress of Chinese electric vehicles industrialization in 2015: a review," *Applied Energy*, vol. 188, pp. 529–546, 2017.

[4] X. Qi, M. J. Barth, G. Wu, K. Boriboonsomsin, and P. Wang, "Energy impact of connected eco-driving on electric vehicles," in *In Road Vehicle Automation 4*, pp. 97–111, Springer, Cham, 2018.

[5] F. Song, "The impact of China's electric vehicle subsidy policy on the new energy automobile industry——based on the research on the 5 pilot cities. Wireless charging of electric vehicles in electricity and transportation networks," *IEEE Transactions on Smart Grid*, vol. 9, pp. 4503–4512, 2016.

[6] S. Lukic and Z. Pantic, "Cutting the cord: static and dynamic inductive wireless charging of electric vehicles," *IEEE Electrification Magazine*, vol. 1, no. 1, pp. 57–64, 2013.

[7] R. Tavakoli and Z. Pantic, "Analysis, design, and demonstration of a 25-kW dynamic wireless charging system for roadway electric vehicles," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 6, no. 3, pp. 1378–1393, 2018.

[8] H. Zhang, Z. Hu, Z. Xu, and Y. Song, "An integrated planning framework for different types of PEV charging facilities in urban area," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2273–2284, 2016.

[9] A. Y. Lam, Y. W. Leung, and X. Chu, "Electric vehicle charging station placement: formulation, complexity, and solutions," *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2846–2856, 2014.

[10] F. Ahmad, M. Saad Alam, I. Saad Alsaidan, and S. M. Shariff, "Battery swapping station for electric vehicles: opportunities and challenges," *IET Smart Grid*, vol. 3, no. 3, pp. 280–286, 2020.

[11] 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.

[12] A. Ahmad, M. S. Alam, and R. Chabaan, "A comprehensive review of wireless charging technologies for electric vehicles," *IEEE Transactions on Transportation Electrification*, vol. 4, no. 1, pp. 38–63, 2018.

[13] T. M. Fisher, K. B. Farley, Y. Gao, H. Bai, and Z. T. H. Tse, "Electric vehicle wireless charging technology: a state-of-the-art review of magnetic coupling systems," *Wireless Power Transfer*, vol. 1, no. 2, pp. 87–96, 2014.

[14] 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.

[15] X. Sun, Z. Chen, and Y. Yin, "Integrated planning of static and dynamic charging infrastructure for electric vehicles," *Transportation Research Part D: Transport and Environment*, vol. 83, article 102331, 2020.

[16] S. Jeong, Y. J. Jang, and D. Kum, "Economic analysis of the dynamic charging electric vehicle," *IEEE Transactions on Power Electronics*, vol. 30, no. 11, pp. 6368–6377, 2015.

[17] Z. Chen, Y. Yin, and Z. Song, "A cost-competitiveness analysis of charging infrastructure for electric bus operations," *Transportation Research Part C: Emerging Technologies*, vol. 93, pp. 351–366, 2018.

[18] M. Fuller, "Wireless charging in California: range, recharge, and vehicle electrification," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 343–356, 2016.

[19] Y. D. Ko and Y. J. Jang, "The optimal system design of the online electric vehicle utilizing wireless power transmission technology," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1255–1265, 2013.

[20] F. Musavi, M. Edington, and W. Eberle, "Wireless power transfer: a survey of EV battery charging technologies," in *2012 IEEE Energy Conversion Congress and Exposition*, pp. 1804–1810, Raleigh, NC, USA, 2012.

[21] T. Theodoropoulos, A. Amditis, J. Sallán et al., "Impact of dynamic EV wireless charging on the grid," in *2014 IEEE International Electric Vehicle Conference*, pp. 1–7, Florence, Italy, 2014.

[22] J. He, H. J. Huang, H. Yang, and T. Q. Tang, "An electric vehicle driving behavior model in the traffic system with a wireless charging lane," *Physica A: Statistical Mechanics and its Applications*, vol. 481, pp. 119–126, 2017.

[23] J. He, H. Yang, H. J. Huang, and T. Q. Tang, "Impacts of wireless charging lanes on travel time and energy consumption in a two-lane road system," *Physica A: Statistical Mechanics and its Applications*, vol. 500, pp. 1–10, 2018.

[24] J. He, H. Yang, T. Q. Tang, and H. J. Huang, "Optimal deployment of wireless charging lanes considering their adverse effect on road capacity," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 171–184, 2020.

[25] M. Li, X. Wu, Z. Zhang, G. Yu, Y. Wang, and W. Ma, "A wireless charging facilities deployment problem considering optimal traffic delay and energy consumption on signalized arterial," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 12, pp. 4427–4438, 2019.

[26] S. Mohrehkesh and T. Nadeem, "Toward a wireless charging for battery electric vehicles at traffic intersections," in *2011 14th international IEEE conference on intelligent transportation systems*, pp. 113–118, Washington, DC, USA, 2011.

[27] J. Zhang, T. Q. Tang, Y. Yan, and X. Qu, "Eco-driving control for connected and automated electric vehicles at signalized intersections with wireless charging," *Applied Energy*, vol. 282, article 116215, 2021.

[28] X. He, H. X. Liu, and X. Liu, "Optimal vehicle speed trajectory on a signalized arterial with consideration of queue,"

Transportation Research Part C: Emerging Technologies, vol. 61, pp. 106–120, 2015.

[29] F. Marini and B. Walczak, "Particle swarm optimization (PSO). A tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 149, pp. 153–165, 2015.

[30] C. Fiori, K. Ahn, and H. A. Rakha, "Power-based electric vehicle energy consumption model: model development and validation," *Applied Energy*, vol. 168, pp. 257–268, 2016.

[31] 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.

[32] US Federal Administration, "Next Generation Simulation Program," 2006, http://ops.fhwa.dot.gov/traffiffifficanalysistools/ngsim.htm.

[33] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, 2021.

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

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 4092011, 13 pages
https://doi.org/10.1155/2022/4092011

Research Article

Synchronous Optimization for Demand-Driven Train Operation Plan in Rail Transit Network Using Nondominated Sorting Coevolutionary Memetic Algorithm

Zhenyu Han, Dewei Li, Baoming Han, and Han Gao

School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Baoming Han; bmhan@bjtu.edu.cn

Received 22 April 2022; Revised 24 June 2022; Accepted 16 July 2022; Published 5 October 2022

Academic Editor: Zhihong Yao

In many cities and regions, decision makers independently develop Train Operation Plan (TOP) for each line in the rail transit network, resulting in a lack of TOP Synchronization (TOPS). Considering the entire network as a whole, researchers have realized that synchronous optimization is of great significance. In this paper, we formulate two Mixed-Integer Linear Programming (MILP) models to optimize demand-driven TOP in the network. The former is an Asynchronous TOP Optimization (ATOPO) model, while the latter is a Synchronous TOP Optimization (STOPO) model. The bi-objective models simultaneously determine train frequency, train timetable, and rolling stock circulation under small-granularity passenger demand to minimize trains' total cost and passengers' total time. Then, we propose the Nondominated Sorting Coevolutionary Memetic Algorithm (NSCMA) to solve the combinatorial optimization problems. The hybrid heuristic algorithm incorporates Coevolutionary Memetic Algorithm (CMA) into Advanced and Adaptive Nondominated Sorting Genetic Algorithm II (AANSGA-II) to ameliorate the evolution process for elite individuals. On this basis, we study the case of Shenyang Metro to verify the models and the algorithm. The results demonstrate that the STOPO model is better than the ATOPO model in reducing trains' total cost and passengers' total time. In addition, NSCMA is better than AANSGA-II in obtaining elite individuals.

1. Introduction

As carbon emissions become a global issue, governments have paid more and more attention to energy consumption [1]. In recent decades, rail transit has developed rapidly around the world as an environmentally friendly mode of public transport. The rail transit systems in many cities and regions have entered the network era. However, lines in most networks are connected only by transfer stations, and trains on most lines are organized independently.

Since researchers know the independence of lines, ATOPO has become a hot issue. In these problems, lines are separated from the network, and TOPs are separately optimized for each line. Several studies have made significant progresses on demand-driven ATOPO problems [2–11]. These studies considered nontransfer passengers but omitted transfer passengers. Therefore, the asynchronously optimized TOPs are probably optimal for nontransfer passengers but probably not optimal for transfer passengers.

Since researchers understand the importance of transfer passengers, TOPS has become another hot issue. In these problems, transfer stations are separated from the network, and TOPs are synchronized for the network. Several studies have made many contributions to TOPS problems. Wong et al. [12] combined a heuristic algorithm with CPLEX to solve a TOPS model aiming at minimizing passengers' transfer waiting time. Wu et al. [13] presented a TOPS model to minimize the maximal transfer waiting time while limiting passengers' transfer waiting time equitably over any transfer station. Guo et al. [14] constructed a TOPS model to maximize the number of transfer synchronization events for the transitional period (from peak to off-peak hours or vice versa). A hybrid heuristic algorithm combined particle swarm optimization with simulated annealing to obtain near-optimal solutions efficiently. Liu et al. [15] built a TOPS model for minimizing passengers' transfer waiting time. Tian and Niu [16] developed a TOPS model to minimize passengers' transfer waiting time and maximize the number

of connections. A novel sequential search algorithm solved the bi-objective model. Cao et al. [17] proposed a Genetic Algorithm (GA for short) with a Local Search (LS for short) strategy to solve a TOPS model by maximizing the number of connections. These studies considered transfer passengers but omitted nontransfer passengers. Therefore, the synchronized TOPs are probably optimal for transfer passengers but probably not optimal for nontransfer passengers.

Since researchers realize the importance of all passengers, STOPO has become an acknowledged challenge. STOPO overcomes the drawbacks of ATOPO and TOPS. In these problems, the network is seen as a whole, and TOPs are simultaneously optimized for the network. Several studies have made some attempts at STOPO problems. Niu et al. [18] presented a demand-driven STOPO model aimed at minimizing passengers' waiting time and crowding disutility. Robenek et al. [19] constructed a demand-driven STOPO model to maximize companies' profit and passengers' satisfaction. Shang et al. [20] built a demand-driven STOPO model for minimizing passengers' travel time. Wang et al. [21] developed a demand-driven STOPO model to minimize passengers' waiting time and the number of passengers with failed transfers. Han et al. [22] formulated a demand-driven STOPO model to minimize trains' operation cost and passengers' total time. AANSGA-II obtained an approximate Pareto Optimal Solution Set (POSS for short) efficiently. These studies considered train formation, train frequency, and train timetable but omitted rolling stock circulation. Therefore, the synchronously optimized TOPs lack practicality.

This paper focuses on an integrated demand-driven TOP Optimization (TOPO for short) problem in the rail transit network. Two bi-objective MILP models called the ATOPO model and the STOPO model simultaneously determine train frequency, train timetable, and rolling stock circulation under small-granularity passenger demand to minimize trains' total cost and passengers' total time. A hybrid heuristic algorithm called NSCMA efficiently solves the bi-objective problem by ameliorating the evolution process for elite individuals based on AANSGA-II. A case study of Shenyang Metro verifies that the STOPO model is better than the ATOPO model and that NSCMA is better than AANSGA-II.

This paper is organized as follows: Section 2 states the demand-driven TOPO problem. Section 3 formulates the ATOPO and STOPO models. Section 4 proposes NSCMA. Section 5 studies the case of Shenyang Metro. Section 6 presents the conclusions.

2. Problem Statement

We focus on a network formed by a set of bidirectional lines $S_l\{1, 2, \cdots, L\}$. Each line $l \in S_l$ contains a set of stations $S_i^l\{1, 2, \cdots, I_l, \cdots, 2I_l\}$ and a set of transfer stations $S_j^l\{1, 2, \cdots, J_l, \cdots, 2J_l\}$, as illustrated in Figure 1. Each physical station on line l refers to i_l and $2I_l + 1 - i_l$ in both directions, and each physical transfer station on line l refers to j_l and $2J_l + 1 - j_l$ in both directions.

We use station $i(j_l)$ to reindex transfer station j_l. Each transfer station $j_l \in S_j^l$ refers to station $i(j_l) \in S_i^l$. On this basis, the set of transfer stations $S_j^l\{1, 2, \cdots, J_l, \cdots, 2J_l\}$ corresponds to a set of stations $S_{li}^0\{i(1), i(2), \cdots, i(J_l), \cdots, i(2J_l)\}$. Binary parameter $q_{jj'}^{ll'}$ equals to one if transfer corridor $(j_l, j_{l'}')$ is valid.

We schedule a set of trains $S_k^l\{1, 2, \cdots, K_l^0, \cdots, K_l\}$ with capacity c_l on each line $l \in S_l$. Limited by the maximum transport capacity, at most K_l trains run on line l with headways of at least $h_{\min}^l$. Required by the minimum service level, at least K_l^0 trains run on line l with headways of at most $h_{\max}^l$. Each active train $k_l \in S_k^l$ on line l runs from station 1 to station $2I_l$. We preset dwell time d_j^l (at transfer station j_l), travel time e_i^l (from station 1 to station i_l), the earliest departure time $g_{\min}^{lk}$ and the latest departure time $g_{\max}^{lk}$ of each train, essential cycle time e_l^0 (on line l) of each connection as well as transfer walking time $f_{jj'}^{ll'}$ (in transfer corridor $(j_l, j_{l'}')$) of each transfer passenger.

Assumption 1. We assume that no line adopts overtaking, skip-stop, cross-line, multi-routing, and multi-marshalling strategies.

We construct a set of time slices $S_t^l\{T_l^0, T_l^0 + 1, \cdots, T_l\}$ (also known as a set of time points) with length τ to express the time period of line l. Combining the time periods of all lines, the time period of the network is expressed as $S_t\{T^0, T^0 + 1, \cdots, T\}$. Notably, the time periods of all stations on line l are normalized. The normalization of time periods reduces the dimension of variables effectively [22].

We describe dynamic passenger demand as small-granularity cumulative number of arriving passengers p_{it}^l (at station i_l at time t), alighting ratio of loaded passengers a_{it}^l (at station i_l at time t) and transferring ratio of alighting passengers $o_{jj't}^{ll'}$ (to transfer corridor $(j_l, j_{l'}')$ at time t). The data are processed from small-granularity origin-destination matrices for simplicity [22].

Assumption 2. We assume that passengers from the outside arrive evenly during each time slice.

We design three binary variables and three integer variables for the demand-driven TOPO problem, as represented in formulas (1)–(6). Binary variable x_{kt}^l equals to one if train k_l departs at time t. Binary variable $w_{kk'}^l$ equals to one if train k_l connects train k_l'. Binary variable $y_{jj'kt}^{ll'}$ equals to one if transfer passengers in transfer corridor $(j_l, j_{l'}')$ from train k_l arrive at time t. Integer variable u_{ik}^l indicates the number of loaded passengers on train k_l in section $(i_l, i_l + 1)$. Integer variable b_{it}^l denotes the number of boarding passengers at station i_l at time t. Integer variable $v_{jj't}^{ll'}$ represents the number of transfer passengers in transfer corridor $(j_l, j_{l'}')$ at time t.

$$x_{kt}^l \in \{0, 1\}, \forall l \in S_l, \forall k_l \in S_k^l, \forall t \in S_t^l, \tag{1}$$

$$w_{kk'}^l \in \{0, 1\}, \forall l \in S_l, \forall k_l \in S_k^l, \forall k_l' \in S_k^l, \tag{2}$$

$$\begin{aligned} &y_{jj'kt}^{ll'} \in \{0, 1\}, \forall l \in S_l, \forall l' \in S_l, \forall j_l \in S_j^l, \\ &\quad \forall j_{l'}' \in S_j^{l'}, \forall k_l \in S_k^l, \forall t \in S_t, \end{aligned} \tag{3}$$

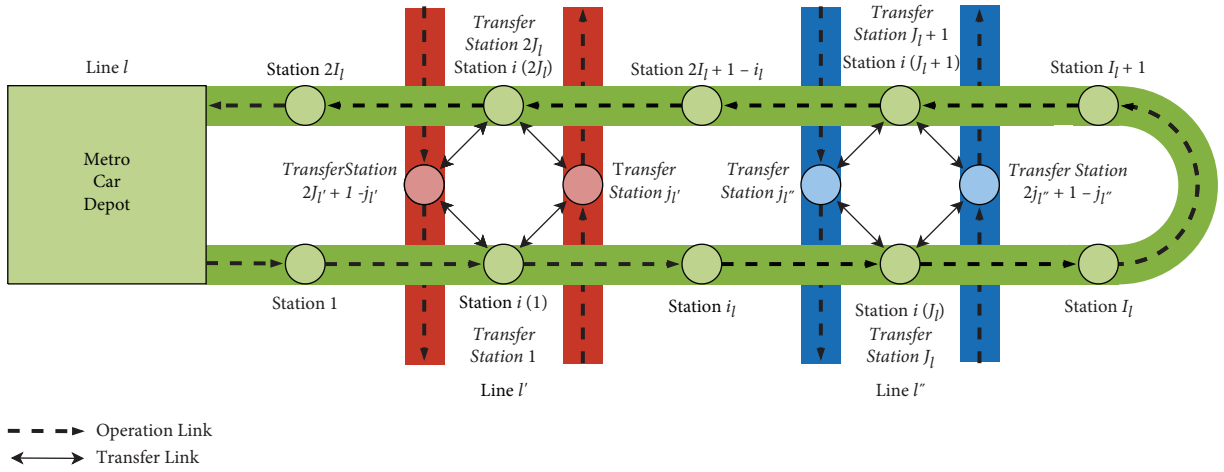


Figure 1: Sketch map of each line.

$$u_{ik}^{l} \in \mathrm{N}, \forall l \in S_l, \forall i_l \in S_i^l, \forall k_l \in S_k^l, \tag{4}$$

$$b_{it}^{l} \in \mathrm{N}, \forall l \in S_l, \forall i_l \in S_i^l, \forall t \in S_t^l, \tag{5}$$

$$v_{jj't}^{ll'} \in \mathrm{N}, \forall l \in S_l, \forall l' \in S_l, \forall j_l \in S_j^l, \forall j_{l'}' \in S_j^{l'}, \forall t \in S_t. \tag{6}$$

3. Mathematical Models

3.1. Asynchronous Train Operation Plan Optimization Model. The ATOPO model separately optimizes the TOP for each line in the network. It is formulated as a MILP model. Although binary variables are more than integer variables in formulating the same problem, the MILP model is linear without processing [23].

3.1.1. Objective Function. The ATOPO model aims to minimize generalized cost Z_l of line l for companies and passengers in (7). On the one hand, we select trains' total cost Z_{TTC}^l to express companies' cost of line l. On the other hand, we select passengers' total time Z_{PTT}^l under weight μ to represent passengers' cost of line l.

$$\min Z_l = Z_{TTC}^l + \mu * Z_{PTT}^l, \forall l \in S_l. \tag{7}$$

Trains' total cost Z_{TTC}^l of line l includes trains' operation cost Z_{TOC}^l and trains' depreciation cost Z_{TDC}^l, as represented in (8).

$$Z_{TTC}^l = Z_{TOC}^l + Z_{TDC}^l, \forall l \in S_l. \tag{8}$$

Trains' operation cost Z_{TOC}^l of line l equals to unit operation cost m_l multiplied by the number of trains, as shown in (9). The number of trains on line l is accumulated by binary variable x_{kt}^l.

$$Z_{TOC}^l = \sum_{k_l \in S_k^l} \sum_{t \in S_t^l} m_l * x_{kt}^l, \forall l \in S_l. \tag{9}$$

Trains' depreciation cost Z_{TDC}^l of line l equals to unit depreciation cost m_l^0 multiplied by the number of rolling stocks, as displayed in (10). The number of rolling stocks on line l equals to the number of trains minus the number of connections. The number of connections on line l is accumulated by binary variable $w_{kk'}^l$.

$$Z_{TDC}^l = \sum_{k_l \in S_k^l} \sum_{t \in S_t^l} m_l^0 * x_{kt}^l - \sum_{k_l \in S_k^l} \sum_{k_l' \in S_k^l} m_l^0 * w_{kk'}^l, \forall l \in S_l. \tag{10}$$

Passengers' total time Z_{PTT}^l of line l includes passengers' waiting time Z_{PWT}^l and passengers' penalty time Z_{PPT}^l, as represented in (11).

$$Z_{PTT}^l = Z_{PWT}^l + Z_{PPT}^l, \forall l \in S_l. \tag{11}$$

Passengers' waiting time Z_{PWT}^l of line l consists of passengers' basic waiting time and passengers' additional waiting time, as shown in (12). Passengers' basic waiting time of line l equals to half τ multiplied by the number of arriving passengers from the outside. Passengers' additional waiting time of line l equals to τ multiplied by the number of waiting passengers.

$$Z_{PWT}^l = \tau * \left(\frac{1}{2} \sum_{i_l \in S_i^l} p_{iT}^l + \sum_{i_l \in S_i^l} \sum_{t=T_l^0}^{T_l - 1} p_{it}^l - \sum_{i_l \in S_i^l} \sum_{t=T_l^0}^{T_l - 1} \sum_{t'=T_l^0}^{t} b_{it'}^l \right), \quad \forall l \in S_l. \tag{12}$$

Passengers' penalty time Z_{PPT}^l of line l equals to unit penalty time ε multiplied by the number of finally stranded passengers, as displayed in (13).

$$Z_{PPT}^l = \varepsilon * \left(\sum_{i_l \in S_i^l} p_{iT}^l - \sum_{i_l \in S_i^l} \sum_{t \in S_t^l} b_{it}^l \right), \forall l \in S_l. \tag{13}$$

3.1.2. Constraints. The ATOPO model is subject to train constraints, connection constraints, and passenger constraints.

(1) Train Constraints. Train constraints stipulate the uniqueness, priority, departure time, and headway of each train, as well as the number of trains.

Constraints (14) and (15) specify the uniqueness of each train. Each time point can only correspond to at most one active train. Meanwhile, each train can only correspond to at most one time point.

$$\sum_{k_l \in S_k^l} x_{kt}^l \le 1, \forall l \in S_l, \forall t \in S_t^l, \tag{14}$$

$$\sum_{t \in S_t^l} x_{kt}^l \le 1, \forall l \in S_l, \forall k_l \in S_k^l. \tag{15}$$

Constraint (16) states the priority of each train. A train should be inactive if the previous train is inactive.

$$\sum_{t \in S_t^l} x_{kt}^l \le \sum_{t \in S_t^l} x_{k-1,t}^l, \forall l \in S_l, \forall k_l \in \left(S_k^l \backslash \{1\}\right). \tag{16}$$

Constraints (17) and (18) limit that the departure time of each active train should be between $g_{\min}^{lk}$ and $g_{\max}^{lk}$.

$$\sum_{t \in S_t^l} t * x_{kt}^l \ge \frac{g_{\min}^{lk}}{\tau} * \sum_{t \in S_t^l} x_{kt}^l, \forall l \in S_l, \forall k_l \in S_k^l, \tag{17}$$

$$\sum_{t \in S_t^l} t * x_{kt}^l \le \frac{g_{\max}^{lk}}{\tau} * \sum_{t \in S_t^l} x_{kt}^l, \forall l \in S_l, \forall k_l \in S_k^l. \tag{18}$$

Constraints (19) and (20) limit that the headway of each active train should be between h_{min}^l and h_{max}^l.

$$\sum_{t \in S_t^l} t * x_{kt}^l - \sum_{t \in S_t^l} t * x_{k-1,t}^l \ge \frac{h_{\min}^l}{\tau} - 2T_l * \left(1 - \sum_{t \in S_t^l} x_{kt}^l\right), \forall l \in S_l, \forall k_l \in \left(S_k^l \backslash \{1\}\right), \tag{19}$$

$$\sum_{t \in S_t^l} t * x_{kt}^l - \sum_{t \in S_t^l} t * x_{k-1,t}^l \le \frac{h_{\max}^l}{\tau} + 2T_l * \left(1 - \sum_{t \in S_t^l} x_{kt}^l\right), \forall l \in S_l, \forall k_l \in \left(S_k^l \backslash \{1\}\right). \tag{20}$$

Constraints (21) and (22) limit that the number of trains should be between K_l^0 and K_l.

$$\sum_{k_l \in S_k^l} \sum_{t \in S_t^l} x_{kt}^l \ge K_l^0, \forall l \in S_l, \tag{21}$$

$$\sum_{k_l \in S_k^l} \sum_{t \in S_t^l} x_{kt}^l \le K_l, \forall l \in S_l. \tag{22}$$

(2) Connection Constraints. Connection constraints stipulate the uniqueness and cycle time of each connection.

Constraints (23) and (24) clarify the uniqueness of each connection. Each active train can only connect at most one active train. Meanwhile, each active train can only be connected by at most one active train.

$$\sum_{k_l' \in S_k^l} w_{kk'}^l \le \sum_{t \in S_t^l} x_{kt}^l, \forall l \in S_l, \forall k_l \in S_k^l, \tag{23}$$

$$\sum_{k_l' \in S_k^l} w_{k'k}^l \le \sum_{t \in S_t^l} x_{kt}^l, \forall l \in S_l, \forall k_l \in S_k^l. \tag{24}$$

Constraint (25) claims that the cycle time of each connection should be at least e_l^0.

$$\sum_{t \in S_t^l} t * x_{kt}^l - \sum_{t \in S_t^l} t * x_{k't}^l \ge \frac{e_l^0}{\tau} - 2T_l * \left(1 - w_{kk'}^l\right), \quad \forall l \in S_l, \forall k_l \in S_k^l, \forall k_l' \in S_k^l. \tag{25}$$

(3) Passenger Constraints. Passenger constraints stipulate the number of loaded and boarding passengers.

Constraint (26) limits that the number of loaded passengers should not exceed c_l if the train is active and equals to zero otherwise.

$$u_{ik}^l \le c_l * \sum_{t \in S_t^l} x_{kt}^l, \forall l \in S_l, \forall i_l \in S_i^l, \forall k_l \in S_k^l. \tag{26}$$

Constraint (27) limits that the number of boarding passengers should not exceed c_l if an active train departs at the time point and equals to zero otherwise.

$$b_{it}^l \le c_l * \sum_{k_l \in S_k^l} x_{kt}^l, \forall l \in S_l, \forall i_l \in S_i^l, \forall t \in S_t^l. \tag{27}$$

Constraints (28)–(31) declare the quantitative relationship between loaded passengers and boarding passengers. Constraints (28) and (29) are only for station 1, while Constraints (30) and (31) are only for the other stations. The number of loaded passengers should correspond to the number of boarding passengers and the number of alighting passengers. The number of alighting passengers should be equal to the alighting ratio multiplied by the number of loaded passengers.

$$u_{1k}^l \ge b_{1t}^l - c_l * \left(1 - x_{kt}^l\right), \forall l \in S_l, \forall k_l \in S_i^l, \forall t \in S_t^l, \tag{28}$$

$$u_{1k}^l \le b_{1t}^l + c_l * \left(1 - x_{kt}^l\right), \forall l \in S_l, \forall k_l \in S_i^l, \forall t \in S_t^l, \tag{29}$$

$$u_{ik}^{l} \geq b_{it}^{l} + u_{i-1,k}^{l} * \left(1 - a_{it}^{l}\right) - c_l * \left(1 - x_{kt}^{l}\right), \quad \forall l \in S_l, \forall i_l \in \left(S_k^l \backslash \{1\}\right), \forall k_l \in S_k^l, \forall t \in S_t^l, \tag{30}$$

$$u_{ik}^{l} \leq b_{it}^{l} + u_{i-1,k}^{l} * \left(1 - a_{it}^{l}\right) + c_l * \left(1 - x_{kt}^{l}\right), \quad \forall l \in S_l, \forall i_l \in \left(S_k^l \backslash \{1\}\right), \forall k_l \in S_k^l, \forall t \in S_t^l. \tag{31}$$

Constraint (32) specifies the quantitative relationship between boarding passengers and arriving passengers. The cumulative number of boarding passengers should not exceed the cumulative number of arriving passengers.

$$\sum_{t'=T_l^0}^{t} b_{it'}^{l} \leq p_{it}^{l}, \forall l \in S_l, \forall i_l \in S_i^l, \forall t \in S_t^l. \tag{32}$$

In summary, the ATOPO model consists of the objective function and constraints as follows:

$$\text{ATOPO model} \begin{cases} \text{Objective function: (7) – (13);} \\ \text{Train constraints: (14) – (22);} \\ \text{Connection constraints: (23) – (25);} \\ \text{Passenger constraints: (26) – (32);} \\ \text{Variable definitions: (1) – (2), (4) – (5).} \end{cases} \tag{33}$$

3.2. Synchronous Train Operation Plan Optimization Model. The STOPO model simultaneously optimizes the TOPs for all lines in the network. It is also formulated as a MILP model.

3.2.1. Objective Function. The STOPO model aims to minimize generalized cost Z of the network in (34).

$$\min Z = Z_{TTC} + \mu * Z_{PTT}. \tag{34}$$

Trains' total cost Z_{TTC} of the network includes trains' operation cost Z_{TOC} and trains' depreciation cost Z_{TDC}, which are accumulated by trains' operation cost Z_{TOC}^{l} and trains' depreciation cost Z_{TDC}^{l} of each line, respectively, as represented in (35)-(37).

$$Z_{TTC} = Z_{TOC} + Z_{TDC}, \tag{35}$$

$$Z_{TOC} = \sum_{l \in S_l} Z_{TOC}^{l}, \tag{36}$$

$$Z_{TDC} = \sum_{l \in S_l} Z_{TDC}^{l}. \tag{37}$$

Passengers' total time Z_{PTT} of the network includes passengers' waiting time Z_{PWT} and passengers' penalty time Z_{PPT}, which are accumulated by passengers' waiting time Z_{PWT}^{l} and passengers' penalty time Z_{PPT}^{l} of each line, respectively, as represented in (38)–(40). Since arriving passengers from the transfer corridors also wait for trains, passengers' waiting time Z_{PWT} and passengers' penalty time Z_{PPT} should also consider arriving passengers from the transfer corridors.

$$Z_{PTT} = Z_{PWT} + Z_{PPT}, \tag{38}$$

$$Z_{PWT} = \sum_{l \in S_l} Z_{PWT}^{l} + \tau * \sum_{l \in S_l} \sum_{l' \in S_l} \sum_{j_l \in S_j^l} \sum_{j_{l'} \in S_j^{l'}} \sum_{t=T_l^0}^{T_l - 1} \sum_{t'=T_l^0}^{t} v_{jj't}^{ll'}, \tag{39}$$

$$Z_{PPT} = \sum_{l \in S_l} Z_{PPT}^{l} + \varepsilon * \sum_{l \in S_l} \sum_{l' \in S_l} \sum_{j_l \in S_j^l} \sum_{j_{l'} \in S_j^{l'}} \sum_{t \in S_t^l} v_{jj't}^{ll'}. \tag{40}$$

3.2.2. Constraints. The STOPO model not only inherits all the constraints in the ATOPO model but also supplements several new constraints.

Constraints (41) and (42) declare the quantitative relationship between boarding passengers and arriving passengers to replace constraint (32). Constraint (41) is only for the nontransfer stations, while constraint (42) is only for the transfer stations.

$$\sum_{t'=T_l^0}^{t} b_{it'}{}^{l} \leq p_{it}^{l}, \forall l \in S_l, \forall i_l \in \left(S_i^l \backslash S_{li}^0\right), \forall t \in S_t^l, \tag{41}$$

$$\sum_{t'=T_l^0}^{t} b_{i(j)t'}{}^{l} \leq p_{i(j)t}^{l} + \sum_{l' \in S_l} \sum_{j_{l'} \in S_j^l} \sum_{t'=T_l^0}^{t} v_{jjt}^{ll'}, \forall l \in S_l, \forall j_l \in S_j^l, \forall t \in S_t^l. \tag{42}$$

In addition, passenger constraints also stipulate the uniqueness and arriving time of each transfer passenger, and the number of transfer passengers.

Constraint (43) states the uniqueness of each transfer passenger. Each transfer passenger should alight from an active train, walk in a valid transfer corridor, and arrive at a time point.

$$\sum_{t \in S_t} y_{jj'kt}^{ll'} = q_{jj'}^{ll'} * \sum_{t \in S_t^l} x_{kt}^{l}, \forall l \in S_l, \forall l' \in S_l, \forall j_l \in S_j^l, \forall j_{l'} \in S_j^{l'}, \forall k_l \in S_k^l. \tag{43}$$

Constraint (44) clarifies that the arrival time of each transfer passenger should correspond to the departure time of the train from which the transfer passenger alights.

$$\sum_{t\in S_t} t * y_{jj'kt}^{ll'} = q_{jj'}^{ll'} * \left(\sum_{t\in S_t^l} t * x_{kt}^l + \frac{\left(e_{i(j)}^l - e_{i(j')}^{l'} + f_{j'j}^{ll'} - d_j^l\right)}{\tau} * \sum_{t\in S_t^l} x_{kt}^l \right), \tag{44}$$

$$\forall l \in S_l, \forall l' \in S_l, \forall j_l \in S_j^l, \forall j_{l'}' \in S_j^{l'}, \forall k_l \in S_k^l.$$

Constraint (45) limits that the number of transfer passengers should not exceed c_l if an active train departs at the time point and equals to zero otherwise.

$$v_{jj't}^{ll'} \le c_l * q_{jj'}^{ll'} * \sum_{k_l\in S_k^l} y_{jj'kt}^{ll'}, \forall l \in S_l, \forall l' \in S_l, \forall j_l \in S_j^l, \forall j_{l'}' \in S_j^{l'}, \forall l \in S_t. \tag{45}$$

Constraints (46) and (47) claim the quantitative relationship between transfer passengers and loaded passengers. The number of transfer passengers should be equal to the transferring ratio multiplied by the number of alighting passengers.

$$v_{jj't}^{ll'} \ge u_{i(j)-1,k}^l * a_{i(j)t}^l * o_{jj't}^{ll'} - c_l * \left(1 - y_{jj'kt}^{ll'}\right), \forall l \in S_l, \forall l' \in S_l, \forall j_l \in S_j^l, \forall j_{l'}' \in S_j^{l'}, \forall k_l \in S_k^l,$$

$$\forall t \in S_t^l, \forall t' \in S_t,\ t' = t + \frac{\left(e_{i(j)}^l - e_{i(j')}^{l'} + f_{j'j}^{ll'} - d_j^l\right)}{\tau}, \tag{46}$$

$$v_{jj't}^{ll'} \le u_{i(j)-1,k}^l * a_{i(j)t}^l * o_{jj't}^{ll'} + c_l * \left(1 - y_{jj'kt}^{ll'}\right), \forall l \in S_l, \forall l' \in S_l, \forall j_l \in S_j^l, \forall j_{l'}' \in S_j^{l'}, \forall k_l \in S_k^l, \forall t \in S_t^l, \forall t' \in S_t,\ t'$$

$$= t + \frac{\left(e_{i(j)}^l - e_{i(j')}^{l'} + f_{j'j}^{ll'} - d_j^l\right)}{\tau}. \tag{47}$$

In summary, the STOPO model consists of the objective function and constraints as follows:

$$\text{STOPOmodel} \begin{cases} \text{Objective function: (9) – (10), (12) – (13), (34) – (40);} \\ \text{Train constraints: (14) – (22);} \\ \text{Connection constraints: (23) – (25);} \\ \text{Passenger constraints: (26) – (31), (41) – (47);} \\ \text{Variable definitions: (1) – (6).} \end{cases} . \tag{48}$$

4. Nondominated Sorting Coevolutionary Memetic Algorithm

4.1. Algorithm Scheme. With the progress in artificial intelligence, machine learning algorithms have been widely used in transportation research [24–26]. However, heuristic algorithms are still the most common method to solve TOPO problems as they have been proven to be NP-hard problems [27].

The demand-driven TOPO models in this paper aim to minimize generalized cost for companies and passengers, in which weight μ is undetermined. Since decision makers expect different μ in different situations, a novel heuristic algorithm that can obtain an approximate POSS is appropriate. We select AANSGA-II to build the algorithm scheme.

AANSGA-II originates from NSGA-II. NSGA-II is a classical heuristic algorithm for solving multi-objective problems [28]. AANSGA-II improves NSGA-II in three

aspects for TOPO problems. In local sorting, AANSGA-II proposes neighborhood distance instead of crowding distance to sequence the individuals in each frontier appropriately. In crossover and mutation, AANSGA-II introduces a scoring mechanism and alternative operators to produce the offspring population effectively. In population initialization, AANSGA-II adopts boundary individuals to generate the initial population reasonably. [22].

Despite these improvements, AANSGA-II also shows a drawback. Due to the nature of local sorting, AANSGA-II tends to pursue the first frontier rather than the elite individuals. We select CMA to improve the algorithm mechanism.

CMA originates from MA. MA combines GA with LS to improve computation efficiency and solution quality. Excellent individuals generated by LS participate in GA instead of original individuals [29, 30]. CMA improves MA by encoding LS settings (i.e., position, direction, step, strategy, and other parameters) as memes for coevolution [31, 32].

Inspired by MA based on NSGA-II, NSCMA incorporates CMA into AANSGA-II [33, 34]. It consists of population initialization, local search, nondominated sorting, local sorting, tournament selection, crossover and mutation, population combination, population replacement, POSS extraction, operator scoring, and termination judgement, as demonstrated in Table 1.

Table 1: Pseudo code of NSCMA.

	Nondominated sorting coevolutionary memetic algorithm
1	**Start**
2	Number of generations for final termination: Λ;
3	Number of generations for early termination: Λ';
4	Number of generations for operator scoring: Λ^0;
5	Size of population: Θ;
6	Representative weights of passengers' total time: μ;
7	Probabilities of operators: P, P_π, P_ϖ;
8	Tolerance factor of reference point: γ;
9	Objective values of benchmark individual: Z^0_{TTC}, Z^0_{PTT};
10	Index of current generation: $\lambda \leftarrow 1$;
11	$S_{\text{parent}} \leftarrow$ **Population initialization** ();
12	$S_{\text{parent}} \leftarrow$ **Local search** (S_{parent}, μ);
13	$S_{\text{parent}} \leftarrow$ **Nondominated sorting** (S_{parent});
14	$S_{\text{parent}} \leftarrow$ **Local sorting** (S_{parent}, γ, Z^0_{TTC}, Z^0_{PTT});
15	**While** $\lambda \leq \Lambda$ **do**
16	$S_{\text{elite}} \leftarrow$**Tournament selection** (S_{parent});
17	$S_{\text{offspring}} \leftarrow$**Crossover and mutation** (S_{elite}, P, P_π, P_ϖ);
18	$S_{\text{hybrid}} \leftarrow$**Population combination** (S_{parent}, $S_{\text{offspring}}$);
19	$S_{\text{hybrid}} \leftarrow$**Local search** ($S_{\text{hybrid}}$, μ);
20	$S_{\text{hybrid}} \leftarrow$**Nondominated sorting** (S_{hybrid});
21	$S_{\text{hybrid}} \leftarrow$**Local sorting** ($S_{\text{hybrid}}$, γ, Z^0_{TTC}, Z^0_{PTT});
22	$S_{\text{parent}} \leftarrow$**Population replacement** (S_{hybrid});
23	$S^\lambda_{\text{Pareto}} \leftarrow$**POSS extraction** ($S_{\text{parent}}$);
24	P, P_π, $P_\varpi \leftarrow$**Operator scoring** (S_{parent}, S_{hybrid}, Λ^0);
25	**If** $S^\lambda_{\text{Pareto}} = S^{\lambda-\Lambda}_{\text{Pareto}}$ **then**
26	**Break**;
27	**End if**
28	$\lambda \leftarrow \lambda + 1$;
29	**End while**
30	**Return** $S^\lambda_{\text{Pareto}}$;
31	**End**

4.2. Algorithm Improvements. We only focus on the improvements in NSCMA since AANSGA-II was introduced comprehensively in our previous work [22]. The improvements are reflected in chromosome construction, local search, and local sorting.

4.2.1. Improvements in Chromosome Construction. AANSGA-II encodes real variables h^l_k (i.e., headway of train k_l) instead of binary variables x^l_{kt} as genes to express the GA information. The chromosome in AANSGA-II is $(h^1_1, h^1_2, \cdots, h^1_K, h^2_1, h^2_2, \cdots, h^2_K, \cdots, h^L_1, h^L_2, \cdots, h^L_K)$.

Differing from AANSGA-II, NSCMA encodes integer variable n_1 and binary variables n_2, n_3, and n_4 as memes to express the LS settings. Integer variable n_1 represents the position (i.e., line) of LS, which is between 1 and L. Binary variable n_2 expresses the direction of LS. $n_2 = 0$ means to broad the headway, while $n_2 = 1$ means to narrow the headway. Binary variable n_3 represents the step of LS. $n_3 = 0$ means to move τ, while $n_3 = 1$ means to move 2τ. Binary variable n_4 expresses the strategy of LS. $n_4 = 0$ means overall movement, while $n_4 = 1$ means local movement. The chromosome in NSCMA is $(h^1_1, h^1_2, \cdots, h^1_K, h^2_1, h^2_2, \cdots, h^2_K, \cdots, h^L_1, h^L_2, \cdots, h^L_K, n_1, n_2, n_3, n_4)$, as illustrated in Figure 2.

As a module of the chromosome, the LS settings participate in crossover and mutation just like the GA information. In each crossover operation, the LS settings of the two individuals are partially exchanged. In each mutation operation, the LS settings of the individual are partially replaced.

4.2.2. Improvements in Local Search. AANSGA-II adopts all individuals in the hybrid population without processing. Due to the nature of crossover and mutation, the child population is generated without an optimization guarantee. The optimization mechanism in AANSGA-II is entirely nondeterministic.

Differing from AANSGA-II, NSCMA introduces LS to enhance the optimization guarantee. Several representative weights μ are preset to obtain all elite individuals in the first frontier since only elite individuals are valuable for decision makers. Each elite individual performs LS according to its LS settings. The processed individual is accepted if an improvement is achieved, and the original individual is retained otherwise. As the nature of LS, the processed population is optimized with an optimization guarantee. The optimization mechanism in NSCMA is partially deterministic.

4.2.3. Improvements in Local Sorting. AANSGA-II sequences all individuals in each frontier by local sorting. Local sorting is based on neighborhood distance. The individuals in AANSGA-II have descending original neighborhood distances.

Differing from AANSGA-II, NSCMA proposes a reference point to improve local sorting. The reference point is

Figure 2: Structure of chromosome.

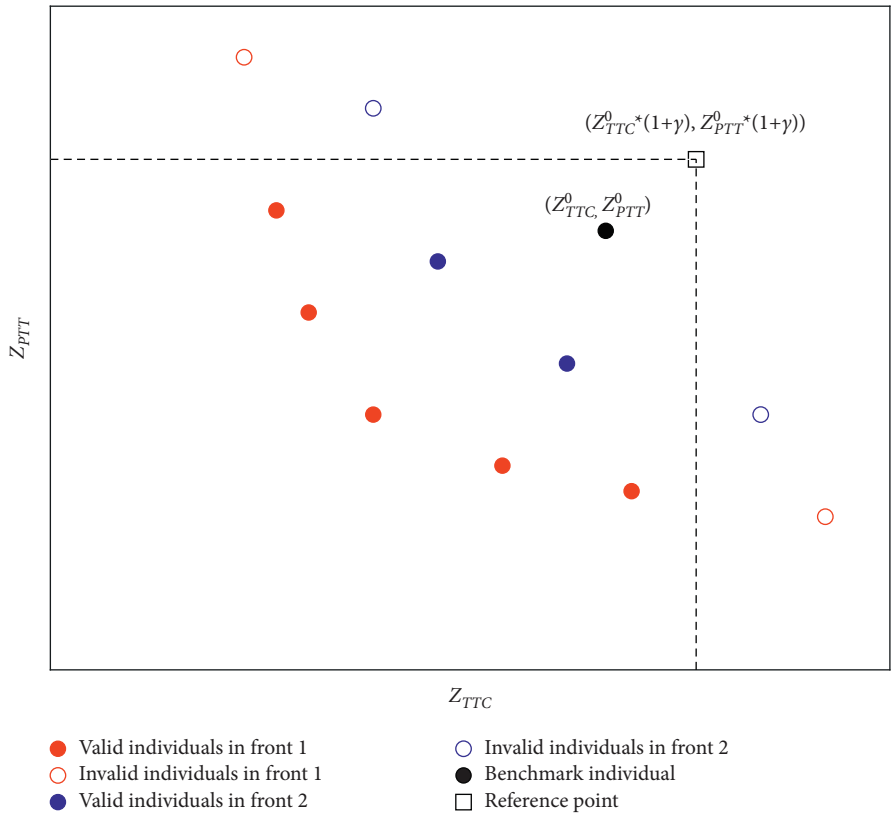


Figure 3: Classification of valid and invalid solutions.

composed of the maximum acceptable objective values for decision makers. An individual is valid if it dominates the reference point and invalid otherwise. On this basis, all individuals in each frontier are classified into valid individuals and invalid individuals. Since valid individuals are more valuable than invalid individuals, the neighborhood distances of valid individuals are retained, while the neighborhood distances of invalid individuals are cleared. The individuals in NSCMA have descending processed neighborhood distances.

Since the benchmark individual with trains' total cost Z^0_{TTC} and passengers' total time Z^0_{PTT} may be too excellent to serve as the reference point, NSCMA proposes tolerance factor γ to expend the maximum acceptable objective values to $Z^0_{TTC} * (1+\gamma)$ and $Z^0_{PTT} * (1+\gamma)$, as illustrated in Figure 3.

5. Case Study

5.1. Case Setup. Shenyang Metro in Northeast China operated a cruciform rail transit network from December 30, 2013, to April 7, 2018, as illustrated in Figure 4.

Dynamic passenger demand is strictly processed from the historical data on December 9, 2016. The distribution of passenger demand is bimodal, as illustrated in Figure 5.

The actual TOPs of the two lines are encoded together as the Benchmark Solution (BS for short). The time period of each line is normalized to [4 : 50, 22 : 20]. Table 2 demonstrates the parameters of the case.

The case focuses on the approximate POSS based on the STOPO model (SS for short) and the approximate POSS based on the ATOPO model (AS for short). Notably, AS of Line 1 (AS-1 for short) and AS of Line 2 (AS-2 for short) are optimized independently. Each solution in AS-1 and each solution in AS-2 form an integrated solution together. AS refers to the POSS in all integrated solutions.

The case applies NSCMA and AANSGA-II for comparisons. Both heuristic algorithms are encoded in MATLAB R2019a. All computations were performed on a personal computer. Table 3 demonstrates the parameters of the heuristic algorithms.

5.2. Case Results. Firstly, we solve the STOPO model with NSCMA and AANSGA-II. NSCMA obtains SS at generation 8491 within 123 min, while AANSGA-II obtains SS^* at

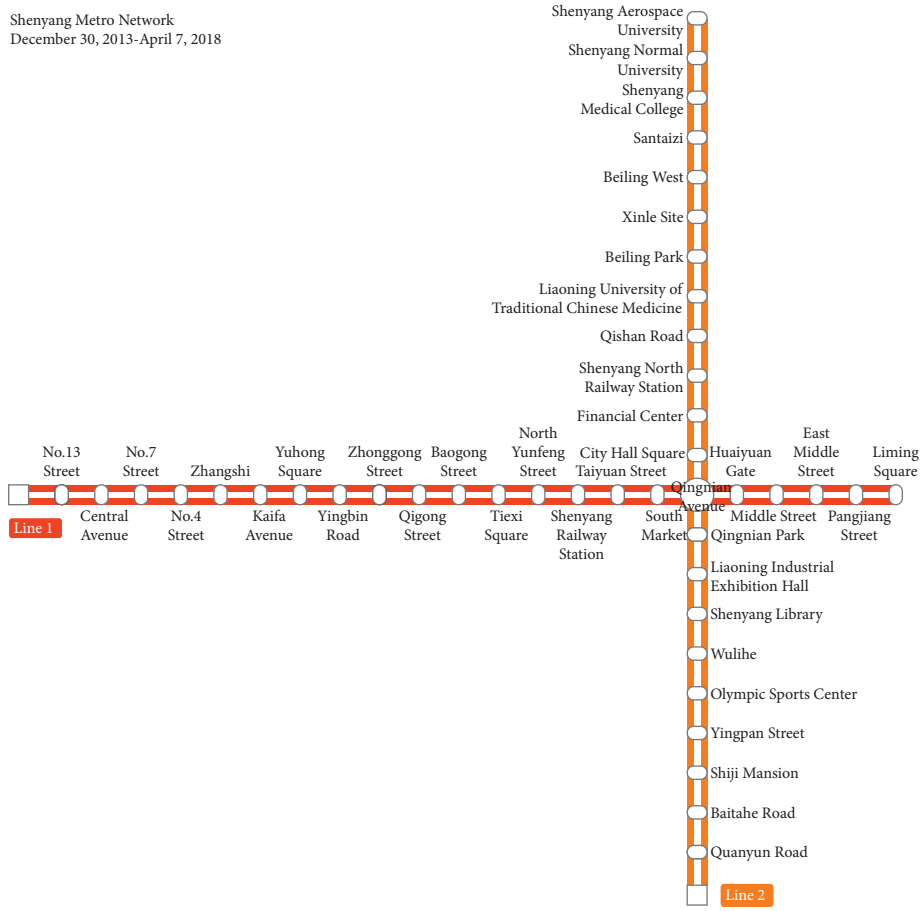


Figure 4: Sketch map of Shenyang Metro.

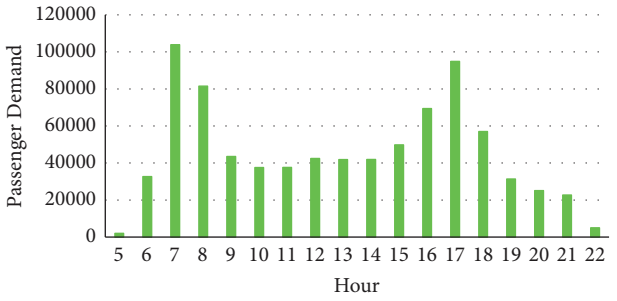


Figure 5: Distribution of passenger demand on December 9, 2016.

generation 9773 within 127 min. We use trains' total cost Z_{TTC} and passengers' total time Z_{PTT} as indicators to compare the performances of SS, SS*, and BS, as illustrated in Figure 6. Notably, only the solutions better than BS in both costs, also known as the valid solutions, are drawn in Figure 6.

According to Figure 6, SS outperforms SS* in progressiveness, while SS does not outperform SS* in diversity. SS dominates SS*, which means that NSCMA is better than AANSGA-II in the evolution process for the elite individuals. However, SS* presents more small cracks but fewer large cracks than SS, which means that NSCMA is not better than AANSGA-II in the evolution process for the first frontier.

Secondly, we solve the ATOPO model by NSCMA. NSCMA obtains AS-1 at generation 1142 within 7 min and AS-2 at generation 304 within 2 min. AS is selected from all combinations of AS-1 and AS-2 within 1 min. We use trains' total cost Z_{TTC} and passengers' total time Z_{PTT} as indicators to compare the performances of SS, AS, and BS, as illustrated in Figure 7.

In the light of Figure 7, SS contains thirty-four solutions, while AS contains fifteen solutions. SS dominates AS, obviously, which means that the STOPO model is better than the ATOPO model.

Thirdly, we list the Elite Solutions in SS (SS-Es for short) and the Elite Solutions in AS (AS-Es for short) under specific weights. Inspired by our previous work, the representative weights are $\mu \in \{0.05, 0.072, 0.1037, 0.1493, 0.215, 0.3096\}$ [22]. These six numbers form a proportional sequence with a common ratio of 1.44. We use trains' total cost Z_{TTC}, passengers' total time Z_{PTT}, generalized cost Z, and improvement rates over BS Δ_{TTC}, Δ_{PTT}, and Δ_Z as indicators to compare the performances of SS-Es, AS-Es, and BS under different μ, as demonstrated in Table 4. Notably, all elitist solutions are marked in Figure 7.

According to Table 4, both elite solution sets consist of five distinct solutions, respectively. Compared with BS, SS-Es improve Z_{TTC} by 4.35% and Z_{PTT} by 5.51% on average, while AS-Es improve Z_{TTC} by 1.57% and Z_{PTT} by 1.39% on

Table 2: Parameters of the case.

Parameter	Description	Value
L	The number of lines	2
I_l	The number of stations in each direction	22 (line 1, line 2), 44 (total)
J_l	The number of transfer stations in each direction	1 (line 1, line 2), 2 (total)
K_l^0	The minimum number of trains	153 (line 1), 136 (line 2), 289 (total)
K_l	The maximum number of trains	180 (line 1), 160 (line 2), 340 (total)
T_l^0	The start time of the time period	1 (line 1, line 2)
T_l	The end time of the time period	2100 (line 1, line 2)
c_l	The capacity of each train	2016 (line 1, line 2)
$h_{\min}^l$	The minimum headway of each train (min)	4 (line 1, line 2)
$h_{\max}^l$	The maximum headway of each train (min)	10 (line 1, line 2)
e_l^0	The essential cycle time of each connection (min)	112 (line 1), 111 (line 2)
τ	The length of each time slice (min)	0.5
m_l	The unit operation cost of each train (CNY)	852 (line 1), 834 (line 2)
m_l^0	The unit depreciation cost of each rolling stock (CNY)	3696 (line 1, line 2)
ε	The unit penalty time of each finally stranded passenger (min)	1000

Table 3: Parameters of the heuristic algorithms.

Parameter	Description	Value
Λ	The number of generations for final termination	10000
Λ'	The number of generations for early termination	50
Λ^0	The number of generations for operator scoring	50
Θ	The size of population	50 (SS), 20 (AS)
μ	The representative weights of passengers' total time	{0.05, 0.072, 0.1037, 0.1493, 0.215, 0.3096}
P	The initial probability of crossover operation	0.5
P_π	The initial probability of each crossover operator	0.5
P_ω	The initial probability of each mutation operator	0.0625
γ	The tolerance factor of reference point	0.03
Z_{TTC}^0	The trains' total cost of benchmark individual (CNY)	447552
Z_{PTT}^0	The passengers' total time of benchmark individual (min)	3163228.25

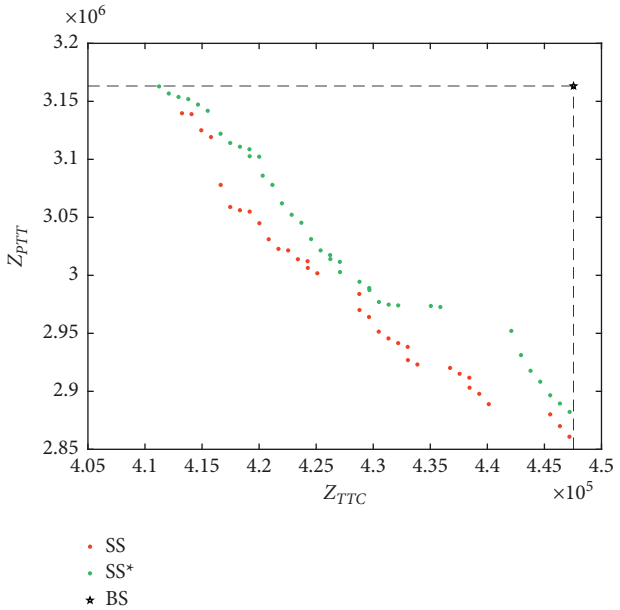


Figure 6: Performances of SS, SS^*, and BS.

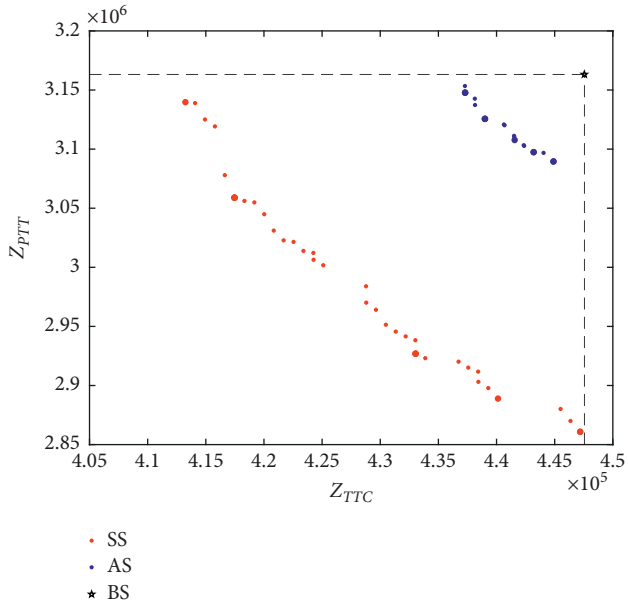


Figure 7: Performances of SS, AS, and BS.

Table 4: Performances of SS-Es, AS-Es and BS under different μ.

μ	Solution	Z_{TTC}	Δ_{TTC}	Z_{PTT}	Δ_{PTT}	Z	Δ_Z
0.05	SS-E-1	413244	7.67%	3139740.25	0.74%	570231.01	5.86%
	AS-E-1	437298	2.29%	3147741.75	0.49%	594685.09	1.82%
	BS	447552	—	3163228.25	—	605713.41	—
0.072	SS-E-2	417468	6.72%	3058919.75	3.30%	637710.22	5.57%
	AS-E-1	437298	2.29%	3147741.75	0.49%	663935.41	1.68%
	BS	447552	—	3163228.25	—	675304.43	—
0.1037	SS-E-2	417468	6.72%	3058919.75	3.30%	734616.80	5.27%
	AS-E-2	439002	1.91%	3125601.25	1.19%	763064.34	1.61%
	BS	447552	—	3163228.25	—	775515.50	—
0.1493	SS-E-3	433038	3.24%	2926909.25	7.47%	870023.21	5.41%
	AS-E-3	441558	1.34%	3107683.75	1.76%	905532.70	1.55%
	BS	447552	—	3163228.25	—	919819.45	—
0.215	SS-E-4	440124	1.66%	2888932.25	8.67%	1061217.99	5.89%
	AS-E-4	443190	0.97%	3097419.75	2.08%	1109106.90	1.64%
	BS	447552	—	3163228.25	—	1127617.12	—
0.3096	SS-E-5	447192	0.08%	2860860.25	9.56%	1332876.63	6.59%
	AS-E-5	444894	0.59%	3089543.75	2.33%	1401376.03	1.79%
	BS	447552	—	3163228.25	—	1426845.78	—
Average	SS-E	428089	4.35%	2989046.92	5.51%	867779.31	5.86%
	AS-E	440540	1.57%	3119288.67	1.39%	906283.41	1.68%
	BS	447552	—	3163228.25	—	921802.62	—

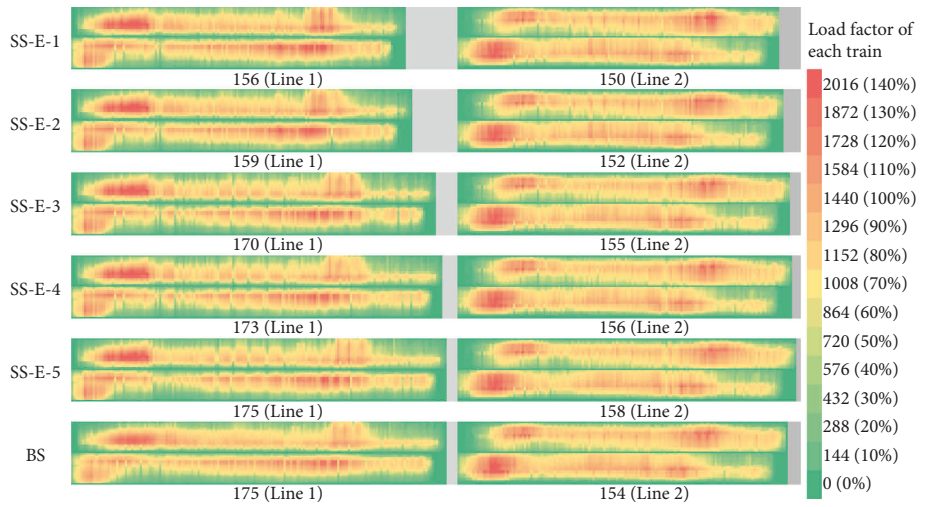


Figure 8: Utilization of trains in SS-Es and BS.

average. Compared with AS-Es, SS-Es improve Z_{TTC} by 2.83% and Z_{PTT} by 4.18% on average.

Fourthly, we use the load factor of each train as an indicator to analyze the utilization of trains in SS-Es and BS, as illustrated in Figure 8.

In the light of Figure 8, most SS-Es utilize trains more effectively than BS. Overall, SS-E-1 uses twenty-three fewer trains than BS, SS-E-2 uses eighteen fewer trains than BS, SS-E-3 uses four fewer trains than BS, SS-E-4 uses the same number of trains as BS, and SS-E-5 uses four more trains than BS. Specifically, no SS-E uses more trains on Line 1 than BS, while three SS-Es use more trains on Line 2 than BS. The difference demonstrates that the actual TOP of Line 1 has more space for optimization than that of Line 2. Besides, fewer trains generally mean higher average load factors.

Fifthly, we use the task order of each rolling stock as an indicator to analyze the utilization of rolling stocks in SS-Es and BS, as illustrated in Figure 9.

According to Figure 9, all SS-Es utilize rolling stocks more effectively than BS. Overall, SS-E-1 and SS-E-2 use four fewer rolling stocks than BS, SS-E-3 uses three fewer rolling stocks than BS, SS-E-4 uses two fewer rolling stocks than BS, and SS-E-5 uses one fewer rolling stocks than BS. Specifically, all SS-Es use fewer rolling stocks on Line 1 than BS, while no SS-E uses fewer rolling stocks on Line 2 than BS. The difference proves again that the actual TOP of Line 2 matches passenger demand better than that of Line 1.

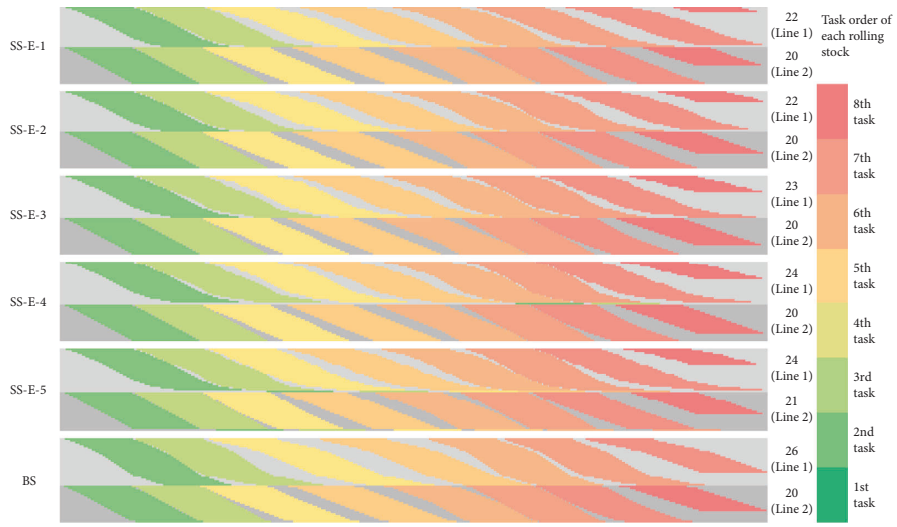


Figure 9: Utilization of rolling stocks in SS-Es and BS.

Besides, fewer rolling stocks generally mean more tasks per rolling stock.

In summary, the superiority of the STOPO model over the ATOPO model explains that TOPS is an important element in TOPO for the network. Besides, the superiority of NSCMA over AANSGA-II suggests that LS is a powerful complement to evolutionary algorithms.

6. Conclusion

This paper researched the demand-driven TOPO problem in the rail transit network. The bi-objective MILP models, the ATOPO model, and the STOPO model minimize trains' total cost and passengers' total time by simultaneously determining train frequency, train timetable, and rolling stock circulation. The hybrid heuristic algorithm, NSCMA, ameliorates the evolution process for elite individuals by incorporating CMA into AANSGA-II.

According to the case of Shenyang Metro, the STOPO model is better than the ATOPO model. The elite synchronous TOPs reduce trains' total cost by 2.83% and passengers' total time by 4.18% on average compared to the elite asynchronous TOPs. Besides, NSCMA is better than AANSGA-II. NSCMA outperforms AANSGA-II in obtaining the elite individuals, while NSCMA does not outperform AANSGA-II in obtaining the first frontier.

In the future, we will focus on the difficulties as follows. Firstly, the weight of passengers' total time is discussed with an alternative set. It is meaningful to integrate all objectives into a unified dimension. Secondly, NSCMA does not show comprehensive improvement over AANSGA-II. It is optional to apply other heuristic algorithms.

Data Availability

The passenger demand data used to support the findings of this study have not been made available because of the secrecy agreement.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported by the National Natural Science Foundation of China (71971019), the Fundamental Research Funds for the Central Universities (2020JBZD007), the Frontiers Science Center for Smart High-Speed Railway System, and the Beijing Natural Science Foundation (L201013).

References

[1] Y. Yang, Z. Yuan, J. Chen, and M. Guo, "Assessment of osculating value method based on entropy weight to transportation energy conservation and emission reduction," *Environmental Engineering and Management Journal*, vol. 16, no. 10, pp. 2413–2423, 2017.

[2] H. Niu and X. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 212–230, 2013.

[3] E. Barrena, D. Canca, L. C. Coelho, and G. Laporte, "Single-line rail rapid transit timetabling under dynamic passenger demand," *Transportation Research Part B: Methodological*, vol. 70, pp. 134–150, 2014.

[4] H. Niu, X. Zhou, and R. Gao, "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: nonlinear integer programming models with linear constraints," *Transportation Research Part B: Methodological*, vol. 76, pp. 117–135, 2015.

[5] T. Zhang, D. Li, and Y. Qiao, "Comprehensive optimization of urban rail transit timetable by minimizing total travel times under time-dependent passenger demand and congested conditions," *Applied Mathematical Modelling*, vol. 58, pp. 421–446, 2018.

[6] Y. Yin, D. Li, N. Bešinović, and Z. Cao, "Hybrid demand-driven and cyclic timetabling considering rolling stock circulation for a bidirectional Railway line," *Computer-Aided*

Civil and Infrastructure Engineering, vol. 34, no. 2, pp. 164–187, 2019.

[7] P. Mo, L. Yang, Y. Wang, and J. Qi, "A flexible Metro train scheduling approach to minimize energy cost and passenger waiting time," *Computers & Industrial Engineering*, vol. 132, pp. 412–432, 2019.

[8] S. Yang, F. Liao, J. Wu, H. J. P. Timmermans, H. Sun, and Z. Gao, "A Bi-objective timetable optimization model incorporating energy allocation and passenger assignment in an energy-regenerative Metro system," *Transportation Research Part B: Methodological*, vol. 133, pp. 85–113, 2020.

[9] X. Dong, D. Li, Y. Yin, S. Ding, and Z. Cao, "Integrated optimization of train stop planning and timetabling for commuter railways with an extended adaptive large neighborhood search metaheuristic approach," *Transportation Research Part C: Emerging Technologies*, vol. 117, p. 102681, 2020.

[10] Z. Cao, A. A. Ceder, D. Li, and S. Zhang, "Robust and optimized urban rail timetabling using a marshaling plan and skip-stop operation," *Transportmetrica: Transportation Science*, vol. 16, no. 3, pp. 1217–1249, 2020.

[11] R. Yang, B. Han, Q. Zhang, Z. Han, and Y. Long, "Integrated optimization of train route plan and timetable with dynamic demand for the urban rail transit line," *Transportation Business: Transport Dynamics*, vol. 26, pp. 1–34, 2022.

[12] R. C. W. Wong, T. W. Y. Yuen, K. W. Fung, and J. M. Y. Leung, "Optimizing timetable synchronization for rail mass transit," *Transportation Science*, vol. 42, no. 1, pp. 57–69, 2008.

[13] J. Wu, M. Liu, H. Sun, T. Li, Z. Gao, and D. Z. W. Wang, "Equity-based timetable synchronization optimization in urban subway network," *Transportation Research Part C: Emerging Technologies*, vol. 51, pp. 1–18, 2015.

[14] X. Guo, H. Sun, J. Wu, J. Jin, J. Zhou, and Z. Gao, "Multi-period-based timetable optimization for Metro transit networks," *Transportation Research Part B: Methodological*, vol. 96, pp. 46–67, 2017.

[15] X. Liu, M. Huang, H. Qu, and S. Chien, "Minimizing Metro transfer waiting time with AFCS data using simulated annealing with parallel computing," *Journal of Advanced Transportation*, vol. 75, pp. 1–17, 2018.

[16] X. Tian and H. Niu, "A Bi-objective model with sequential search algorithm for optimizing network-wide train timetables," *Computers & Industrial Engineering*, vol. 127, pp. 1259–1272, 2019.

[17] Z. Cao, A. A. Ceder, D. Li, and S. Zhang, "Optimal synchronization and coordination of actual passenger-rail timetables," *Journal of Intelligent Transportation Systems*, vol. 23, no. 3, pp. 231–249, 2019.

[18] H. Niu, X. Tian, and X. S. Zhou, "Demand-driven train schedule synchronization for high-speed rail lines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2642–2652, 2015.

[19] T. Robenek, Y. Maknoon, S. S. Azadeh, J. Chen, and M. Bierlaire, "Passenger centric train timetabling problem," *Transportation Research Part B: Methodological*, vol. 89, pp. 107–126, Jul 2016.

[20] P. Shang, R. Li, Z. Liu, K. Xian, and J. Guo, "Timetable synchronization and optimization considering time-dependent passenger demand in an urban subway network," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2672, no. 8, pp. 243–254, 2018.

[21] Y. Wang, D. Li, and Z. Cao, "Integrated timetable synchronization optimization with capacity constraint under time-dependent demand for a rail transit network," *Computers & Industrial Engineering*, vol. 142, p. 106374.

[22] Z. Han, B. Han, D. Li, S. Ning, R. Yang, and Y. Yin, "Train timetabling in rail transit network under uncertain and dynamic demand using advanced and adaptive NSGA-II," *Transportation Research Part B: Methodological*, vol. 154, pp. 65–99, 2021.

[23] Z. Yao, H. Jiang, Y. Cheng, Y. Jiang, and B. Ran, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 1841–1851, 2022.

[24] C. S. Ying, A. H. F. Chow, and K. S. Chin, "An actor-critic deep reinforcement learning approach for Metro train scheduling with rolling stock circulation under stochastic demand," *Transportation Research Part B: Methodological*, vol. 140, pp. 210–235, 2020.

[25] Y. Yang, K. He, Y. p. Wang, Z. z. Yuan, Y. h. Yin, and M. z. Guo, "Identification of dynamic traffic crash risk for cross-area freeways based on statistical and machine learning methods," *Physica A: Statistical Mechanics and Its Applications*, vol. 595, p. 127083.

[26] Y. Yang, Z. Yuan, and R. Meng, "Exploring traffic crash occurrence mechanism toward cross-area freeways via an improved data mining approach," *Journal of Transportation Engineering Part A: Systems*, vol. 148, no. 9, 2022.

[27] X. Cai and C. J. Goh, "A fast heuristic for the train scheduling problem," *Computers & Operations Research*, vol. 21, no. 5, pp. 499–510, 1994.

[28] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: nsga-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[29] H. Wang, D. Wang, and S. Yang, "A memetic algorithm with adaptive hill climbing strategy for dynamic optimization problems," *Soft Computing*, vol. 13, no. 8-9, pp. 763–780, 2008.

[30] D. Tang, Z. Liu, J. Yang, and J. Zhao, "Memetic frog leaping algorithm for global optimization," *Soft Computing*, vol. 23, no. 21, pp. 11077–11105, 2018.

[31] J. E. Smith, "Co-evolving memetic algorithms: a learning approach to robust scalable optimisation," in *Proceedings of the IEEE Congress on Evolutionary Computation*, pp. 498–505, Canberra, AUSTRALIA, 2003.

[32] E. Özcan, J. H. Drake, C. Altıntaş, and S. Asta, "A self-adaptive multimeme memetic algorithm Co-evolving utility scores to control genetic operators and their parameter settings," *Applied Soft Computing*, vol. 49, pp. 81–93, 2016.

[33] M. Frutos, A. C. Olivera, and F. Tohmé, "A memetic algorithm based on a NSGAII scheme for the flexible job-shop scheduling problem," *Annals of Operations Research*, vol. 181, no. 1, pp. 745–765, 2010.

[34] C. Cobos, C. Erazo, and J. Luna, "Multi-objective memetic algorithm based on NSGA-II and simulated annealing for calibrating CORSIM micro-simulation models of vehicular traffic flow," *Advances in Artificial Intelligence*, vol. 43, pp. 468–476, 2016.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 7248854, 6 pages
https://doi.org/10.1155/2022/7248854

Research Article

Impacts of Cooperative Adaptive Cruise Control Links on Driving Comfort under Vehicle-to-Vehicle Communication

Yongchun Li and Chuanping Shan

China Railway Changjiang Transport Design Group Co., Ltd, Chongqing 401121, China

Correspondence should be addressed to Yongchun Li; lyc5136661@163.com

Received 21 January 2022; Revised 3 March 2022; Accepted 27 August 2022; Published 10 September 2022

Academic Editor: Yanyan Qin

Although automated vehicles could release drivers from the driving task, there are still passengers sitting in the vehicle. It is required that the driving comfort of passengers should be guaranteed. Cooperative adaptive cruise control (CACC) vehicle is of the one important type of automated vehicles using vehicle-to-vehicle (V2V) communications with various communication links. Different V2V communication links might have different driving comfort. Then, this paper focuses on exploring which link type for CACC vehicles is better from the perspective of improving driving comfort. To deal with this, car-following models of manual-driven vehicles (MDV) and CACC vehicles were first described. Then, simulations were performed using these car-following models, in which various CACC feedback link types, CACC penetration rates, and flow scenarios were taken into consideration. Simulations outputted microcosmic trajectory data of vehicles, based on which the driving comfort was evaluated using the comfort index described by the International Organization for Standardization (ISO) 2631-1. From the driving comfort perspective, simulation results suggest that CACC should monitor the immediately preceding vehicle and the third vehicle ahead when CACC penetration rates are less than approximately 50%. Additionally, if CACC penetration rates exceed 50%, the better choice is that CACC receives feedback links from two immediately successive vehicles ahead.

1. Introduction

In transportation systems, passenger comfort is arising great interest with the increase in transit time and consumer expectations [1]. However, human drivers have limitations with response time and driving mistakes to deal with disturbances downstream, thereby usually causing discomfort. Fortunately, cooperative adaptive cruise control (CACC) systems are developed for smooth vehicular flow, which is also helpful for improving passenger driving comfort [2]. CACC vehicles need vehicle-to-vehicle (V2V) communication to monitor multiple vehicles ahead [3]. Therefore, as in the previous studies [4–6], this paper also assumes that manual-driven vehicles (MDV) are equipped with V2V communication devices and can send their motion information to CACC vehicles, regardless of whether the MDV can receive information from other vehicles or not. Under such an assumption, the percentage of active CACC vehicles is equal to CACC penetration rate in the mixed traffic flow [7–10].

Although there are many studies conducted on the impacts of CACC on traffic flow, such as capacity [11–14], safety [15, 16], emissions [17, 18], and stability [19–23], only a few literature focused on driving comfort of CACC. Generally speaking, the existing literature can be divided into two categories. On the one hand, some literature [24–26] developed various vehicle speed control algorithms to smooth vehicular dynamics in order to improve driving comfort. On the other hand, Ref [1] reviewed the state-of-the-art of CACC driving comfort and pointed out a research gap in evaluating passenger comfort for the traffic flow mixed with different CACC penetration rates. This paper focuses on the driving comfort of the mixed CACC traffic flow.

To avoid confusion, the CACC feedback link is defined as the connectivity information from the vehicle ahead to the tail CACC in this paper. Therefore, the feedback link means

the CACC can monitor the corresponding vehicle ahead. Although more feedback links for CACC are more helpful for anticipations of traffic flow dynamics downstream, it really also complicates CACC control design [2]. Therefore, within the V2V communication range, all available feedback links from vehicles ahead to the tail CACC are not necessarily the best choice. It needs to compare and determine which type of CACC feedback link is better from the perspective of improving driving comfort. However, to the best of our knowledge, little research was conducted to evaluate the impacts of CACC feedback link types on driving comfort for the mixed CACC traffic flow. Then, the objective of this paper is to compensate for the research in this area. To deal with this, the objective and car-following models are first described. Then, simulations are performed on the highway with an on-ramp using the car-following models. The comfort index is finally used to evaluate the driving comfort situations under various CACC feedback link types, CACC penetration rates, and traffic flow demands, thereby providing suggestions for the design of CACC feedback links with the increase of CACC penetration rates.

2. Objective and Models

2.1. Feedback Link Types of CACC. With V2V communication, CACC can monitor multiple vehicles ahead. Meanwhile, the maximum number of vehicles that one CACC can monitor is also limited by the V2V communication range. At the present stage, the V2V communication range is approximately 300 m, which means the feedback information can be sent from the fourth vehicle ahead to the tail CACC under any driving speed [4]. Moreover, it is found that multianticipation leads to significant improvements in stability and comfort for two to four leaders but hardly any further improvement for further leaders. Therefore, this paper defines that CACC can receive feedback information from one, two, three, or four vehicles ahead. Although more feedback links may be more helpful for anticipations of vehicular flow dynamics, it also complicates CACC control design [4]. Hence, all available feedback links from all four vehicles ahead for CACC are not necessarily the best choice. The previous study [4] suggested that two feedback links for CACC could satisfy the required driving tasks. Moreover, the feedback link from the immediately preceding vehicle to the tail CACC is essential to response traffic flow dynamics. Then, another feedback link comes from the second, third, or fourth vehicle downstream, respectively, as shown in Figure 1. Figure 1 shows three types of feedback links for CACC, named link type I, link type II, and link type III defined in this paper. Link type I means the tail CACC monitors feedback signals from the immediately preceding vehicle and the second vehicle ahead. The feedback is sent from the immediately preceding vehicle and the third vehicle ahead in the case of link type II. Besides, link type III contains one link from the immediately preceding vehicle and another one from the fourth vehicle ahead. The three feedback link types of CACC in Figure 1 are considered as the objective of this paper, whose respective impacts on driving comfort will be compared in order to determine which one is better from the driving comfort perspective.

2.2. Car-Following Models. Car-following models [27–31] are essential for microscopic simulations to evaluate impacts on driving comfort. In the case of the MDV model, the intelligent driver model (IDM) [32, 33] with response time is used. The model equation is written as

$$\dot{v}_n(t+\tau_1) = a\left[1-\left(\frac{v_n(t)}{v_0}\right)^{\delta}-\left(\frac{d+v_n(t)T-v_n(t)\left[v_{n-1}(t)-v_n(t)\right]/2\sqrt{ab}}{s_n(t)}\right)^2\right], \quad (1)$$

where $v_n(t)$ is the speed of vehicle n at time t, $v_{n-1}(t)$ is the speed of vehicle $n-1$ at time t, $s_n(t)$ is the distance gap between vehicle n and its preceding vehicle $n-1$, τ_1 is the response time of human drivers, $\dot{v}_n(t+\tau_1)$ is the acceleration of vehicle n after the delay time τ_1, a is the maximum acceleration, v_0 is the free flow speed, δ is the exponent coefficient, d is the minimum gap, T is the safety time gap, and b is the comfort deceleration. Based on real data, the calibration results of IDM parameters are [34] $\tau_1 = 1.3575$ s, $a = 1.2681$ m/s^2, $v_0 = 30.0$ m/s, $\delta = 3.0244$, $d = 9.6312$ m, $T = 1.7031$ s, and $b = 2.8638$ m/s^2.

In the case of the CACC car-following model, intelligent property can be described by the IDM model. Then, this paper adopts the model which is written as follows [35]:

$$\dot{v}_n(t+\tau_2) = a\left[1-\left(\frac{v_n(t)}{v_0}\right)^{\delta}-\left(\frac{d+v_n(t)T-v_n(t)\left[v_{n-1}(t)-v_n(t)\right]/2\sqrt{ab}}{s_n(t)}\right)^2\right]+\sum_{i=1}^{m} r_i\left(\theta_{n-i}(t)-\theta_n(t)\right), \quad (2)$$

where τ_2 is the V2V communication delay time that is set as 0.1 s, r_i is the feedback gain from the i-th vehicle ahead, m is the maximum number of vehicles that the CACC can monitor (set as 4 in this paper), and $\theta_n(t)$ is called the electronic throttle angle of vehicle n at time t, while $\theta_{n-i}(t)$ is that of vehicle $n-i$ at time t. Based on the previous studies [35–37], the feedback information $\theta_{n-i}(t)-\theta_n(t)$ is an integration of speeds and accelerations, and $\theta_n(t)$ is modeled as follows:

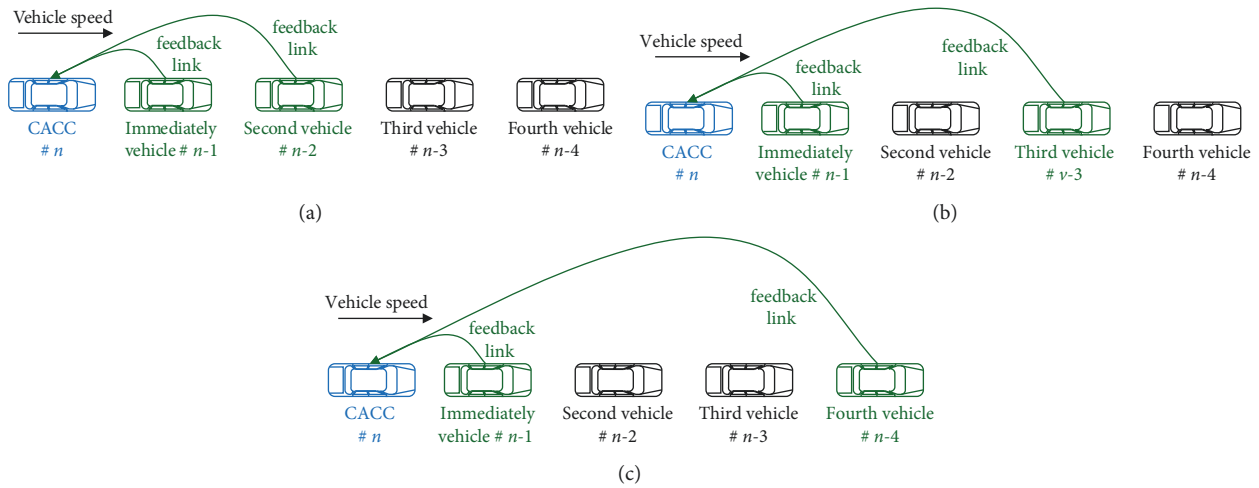


Figure 1: CACC feedback link types studied as the objective of this paper. (a) Link type I. (b) Link type II. (c) Link type III.

$$\theta_{n-i}(t) - \theta_n(t) = \frac{1}{\beta}\left[\dot{v}_{n-i}(t) - \dot{v}_n(t) + \alpha\left(v_{n-i}(t) - v_n(t)\right)\right], \tag{3}$$

where v_e is the equilibrium speed, θ_e is the electronic throttle angle at the equilibrium state, α and β are sensitivity coefficients, which are set as 0.8 and 0.27, respectively [36].

The feedback gain r_i of the CACC model ranges from 0 to 1. Because low-frequency oscillations usually occur with too low values of feedback gain and too high gains may cause high-frequency instability, all the gains shall be kept around 0.5 [4]. Therefore, the four feedback gains of CACC are all considered as 0.5 in this paper.

3. Numerical Simulations

Numerical simulations are important and necessary before CACC implements in larger-scale real vehicles. The car-following models of MDV and CACC are used to perform the simulations, based on which simulation trajectory data of vehicles can be obtained. Then, a driving comfort measure will be used to evaluate driving comfort situations under different CACC link types and penetration rates.

3.1. Design and Measure. Because car-following models describe longitudinal movements of vehicles, the simulation segment is assumed to be a one-lane highway with an on-ramp located in the middle of the segment [20, 38]. The length of the simulation segment is 6.5 km, which is a straight road without a gradient. Different flow scenarios should be considered in simulations. Because driving comfort problems will not easily arise in light traffic flow, the main-line flow should not be too low in simulations. According to the previous studies [20, 38], the main-line flow is set as 1800 veh/hr and 1200 veh/hr, respectively, while the ramp flow is set as 360 veh/hr. The simulations are performed in Matlab software using car-following models, in which the total simulation time is 1 hour and the time step is 0.1 s. In simulations, the desired speed of vehicles is 30 m/s, the maximum acceleration is set as 4 m/s^2, and the emergency deceleration is considered to be −6 m/s^2. The vehicles randomly have a speed of 25–30 m/s when entering the simulation segment. Because what we are concerned about is the main-line flow [20, 38]., the simulation data of vehicles travelling on the main-line road are obtained after simulations. Different feedback links of CACC are illustrated in Figure 1 and various CACC penetration rates are taken into consideration in simulations. For each simulation, the vehicle orders are random. Hence, each simulation is repeated ten times to calculate the average value as the result.

Based on the simulation trajectory data of vehicles under each traffic flow scenario, such as CACC link types and penetration rates, an appropriate driving comfort measure should be used to evaluate the corresponding driving comfort. Here, we use the comfort index (CI) [39] as the driving comfort measure, which is described by the International Organization for Standardization (ISO) 2631-1. The value of CI is calculated by using instantaneous accelerations of vehicles and the calculation equation is as follows:

$$\mathrm{CI} = \left(\frac{1}{N}\sum_{i=0}^{N} a_i^2\right)^{1/2}, \tag{4}$$

where a_i is the ith acceleration obtained in simulations and N is the total number of accelerations that are used for the calculation. The smaller value of CI stands for a better comfort situation.

4. Results

The CI values of 0% CACC penetration rate are considered as the benchmark, compared with which the percentage reductions are calculated for different CACC penetration rates. As noted before, the main-line flow is set as 1800 veh/hr and 1200 veh/hr, respectively, and the three CACC link types illustrated in Figure 1 are considered in the calculations. The results are shown in Tables 1 and 2, in which Table 1 shows the percentage reductions of CI for the three link types with an increase of CACC penetration rates under main-line flow 1800 veh/hr, while Table 2 shows those underflow 1200 veh/hr. According to Tables 1 and 2, the

Table 1: Percentage reductions of CI under 1800 veh/hr flow.

CACC penetration rates (%)	Average reducing under 1800 veh/hr flow (%)		
	Link type I	Link type II	Link type III
0	0.00	0.00	0.00
10	13.32	23.45	27.12
20	30.51	38.57	36.91
30	32.62	41.39	38.95
40	34.87	43.99	42.44
50	38.68	45.60	42.51
60	53.83	47.01	43.87
70	59.55	56.65	56.71
80	61.41	62.03	58.08
90	61.93	62.76	61.35
100	62.19	63.13	61.62

Table 2: Percentage reductions of CI under 1200 veh/hr flow.

CACC penetration rates (%)	Average reducing under 1200 veh/hr flow (%)		
	Link type I	Link type II	Link type III
0	0.00	0.00	0.00
10	2.04	5.38	6.12
20	4.83	19.43	14.51
30	9.27	21.35	18.09
40	10.85	24.83	18.40
50	22.18	26.55	21.56
60	27.91	26.86	23.63
70	30.42	31.61	28.20
80	31.62	32.62	31.92
90	34.43	34.58	32.82
100	34.95	36.36	35.37

values of CI decrease with the increase of the CACC penetration rate, which means an improvement in driving comfort. Compared with the 0% CACC rate, 100% CACC vehicles can enhance driving comfort by 60% for all three link types under 1800 veh/hr flow, while the improvement is approximately 35% under 1200 veh/hr. This indicates that CACC vehicles are more helpful to improve driving comfort in more congestion situations of traffic flow because a larger main-line flow is apt to result in more congestion.

What is more concerning about this paper is the comparison among different CACC link types on improving driving comfort. To deal with this, we plot figures for better visualization, as shown in Figures 2 and 3. Figure 2 shows the comparison among CACC link types under 1800 veh/hr flow, while the comparison of CACC link types under 1200 veh/hr flow is illustrated in Figure 3. It can be seen that the CACC link types II and III have more improvement in driving comfort than link type I when the CACC penetration rate is less than approximately 50%. However, the CACC link type I is apt to have more driving comfort enhancement if the CACC penetration rate exceeds 50%. It should be noted that the control design is more complex if CACC monitors the farther vehicle ahead. Therefore, it is suggested that the CACC link type II should be chosen when the CACC penetration rate is below 50%, while the CACC link type I might be the better one if the CACC penetration rate is

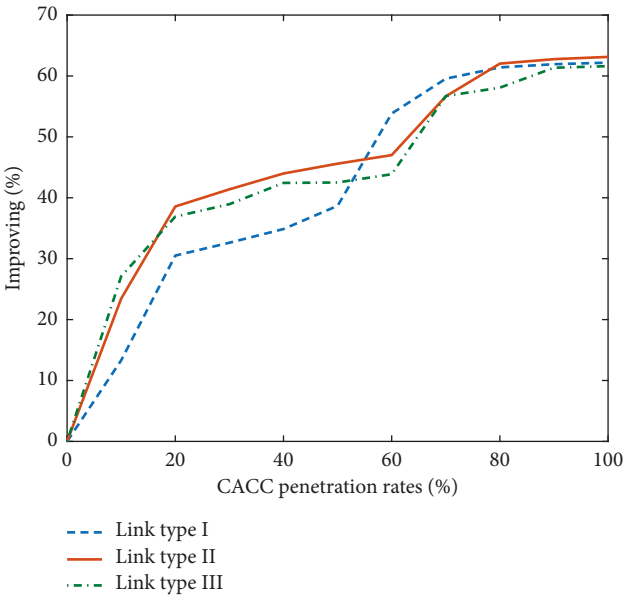


Figure 2: Comparison among CACC link types under 1800 veh/hr flow.

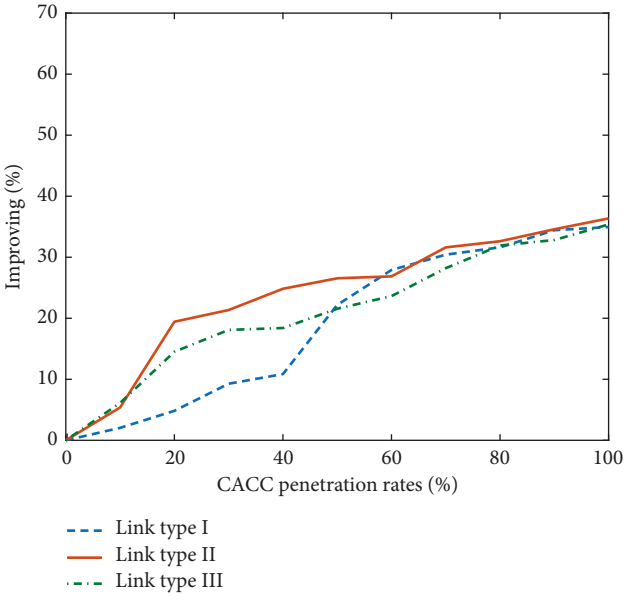


Figure 3: Comparison among CACC link types under 1200 veh/hr flow.

more than 50%. As described in Figure 1, link type I means the CACC monitors two immediately successive vehicles ahead, while link type II requires CACC monitors the third farther vehicle ahead. Therefore, the implements based on this paper's suggestion can reduce the CACC control complexity when CACC penetration rates are high for large-scale flow.

5. Conclusions

By using V2V communication, CACC can receive feedback information from multiple vehicles ahead to improve driving comfort. Although more feedback links for CACC

might be useful for anticipations of traffic flow dynamics, it also complicates the control design work of CACC. Previous studies [4] indicated that two feedback links shall be enough to deal with vehicular flow dynamics downstream. However, two feedback links can be further divided into several types, which distinguish the specific two vehicles ahead. Moreover, little research has been conducted on determining which feedback link type for CACC is a better choice from the perspective of improving driving comfort. This paper makes an effort to deal with this within the farthest fourth vehicle ahead, taking into consideration the V2V communication range. The simulations are performed using car-following models, in which the highway with an on-ramp is considered as the simulation segment [20, 38]. The driving comfort is evaluated under various CACC feedback link types, CACC penetration rates, and flow scenarios. From the perspective of improving driving comfort, simulation results suggest that CACC should receive feedback information from the immediately preceding vehicle and the third vehicle ahead if CACC penetration rates are below approximately 50%. However, when CACC penetration rates are more than 50%, the better design is that CACC monitors two immediately successive vehicles ahead. This indicates that CACC might not need to monitor farther vehicles ahead when its penetration rate is high enough. Because the control design will be relatively more complex if CACC monitors the farther vehicle ahead, the suggestion proposed in this paper can also simplify the CACC control complexity with higher CACC penetration rates.

We should also mention that the conclusions obtained in this paper may depend on the models used. However, the study in this paper indicates that all available feedback links from vehicles ahead to the tail CACC are not necessarily the best choice. This makes us aware that we should compare and determine which type of CACC feedback links is better from the perspective of improving driving comfort. We are aware that real experimental tests are necessary to validate the conclusions, while implementations of large-scale CACC flow with different penetration rates are not easy in the present stage. Therefore, this paper provides an important insight into driving comfort from the perspective of CACC links before the real experimental test.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] M. Elbanhawi, M. Simic, and R. Jazar, "In the passenger seat: investigating ride comfort measures in autonomous cars," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 3, pp. 4–17, 2015.

[2] S. E. Shladover, C. Nowakowski, X. Y. Lu, and R. Ferlis, "Cooperative adaptive cruise control: definitions and operating concepts," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1, pp. 145–152, 2015.

[3] Y. Qin, H. Wang, and D. Ni, "Lighthill-Whitham-Richards model for traffic flow mixed with cooperative adaptive cruise control vehicles," *Transportation Science*, vol. 55, no. 4, pp. 883–907, 2021.

[4] J. I. Ge and G. Orosz, "Dynamics of connected vehicle systems with delayed acceleration feedback," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 46–64, 2014.

[5] J. I. Ge, S. S. Avedisov, C. R. He, W. B. Qin, M. Sadeghpour, and G. Orosz, "Experimental validation of connected automated vehicle design among human-driven vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 91, pp. 335–352, 2018.

[6] J. Lioris, R. Pedarsani, F. Y. Tascikaraoglu, and P. Varaiya, "Platoons of connected vehicles can double throughput in urban roads," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 292–305, 2017.

[7] D. Chen, A. Srivastava, S. Ahn, and T. Li, "Traffic dynamics under speed disturbance in mixed traffic with automated and non-automated vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 113, pp. 293–313, 2020.

[8] E. Van Nunen, J. Reinders, E. Semsar-Kazerooni, and N. Van De Wouw, "String stable model predictive cooperative adaptive cruise control for heterogeneous platoons," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 2, pp. 186–196, 2019.

[9] A. Ghiasi, O. Hussain, Z. S. Qian, and X. Li, "A mixed traffic capacity analysis and lane management model for connected automated vehicles: a Markov chain method," *Transportation Research Part B: Methodological*, vol. 106, pp. 266–292, 2017.

[10] S. Jin, D. H. Sun, M. Zhao, Y. Li, and J. Chen, "Modeling and Stability Analysis of Mixed Traffic with Conventional and Connected Automated Vehicles from Cyber Physical Perspective," *Physica A: Statistical Mechanics and Its Applications*, vol. 551, Article ID 124217, 2020.

[11] A. Kesting, M. Treiber, M. Schönhof, and D. Helbing, "Adaptive cruise control design for active congestion avoidance," *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 6, pp. 668–683, 2008.

[12] A. Olia, S. Razavi, B. Abdulhai, and H. Abdelgawad, "Traffic capacity implications of automated vehicles mixed with regular vehicles," *Journal of Intelligent Transportation Systems*, vol. 22, no. 3, pp. 244–262, 2018.

[13] H. Liu, X. D. Kan, S. E. Shladover, X. Y. Lu, and R. E. Ferlis, "Impact of cooperative adaptive cruise control on multilane freeway merge capacity," *Journal of Intelligent Transportation Systems*, vol. 22, no. 3, pp. 263–275, 2018.

[14] S. M. Weaver, S. A. Balk, and B. H. Philips, "Merging into strings of cooperative-adaptive cruise-control vehicles," *Journal of Intelligent Transportation Systems*, vol. 25, no. 4, pp. 401–411, 2021.

[15] D. Milakis, B. Van Arem, and B. Van Wee, "Policy and society related implications of automated driving: a review of literature and directions for future research," *Journal of Intelligent Transportation Systems*, vol. 21, no. 4, pp. 324–348, 2017.

[16] M. S. Rahman and M. Abdel-Aty, "Longitudinal safety evaluation of connected vehicles' platooning on expressways," *Accident Analysis & Prevention*, vol. 117, pp. 381–391, 2018.

[17] B. Zhao, Y. Lin, H. Hao, and Z. Yao, "Fuel consumption and traffic emissions evaluation of mixed traffic flow with connected automated vehicles at multiple traffic scenarios,"

Journal of Advanced Transportation, vol. 2022, Article ID 6345404, 14 pages, 2022.

[18] Z. Li, L. Chen, S. Xu, and Y. Qian, "Analytical studies of CO_2 emission in a mixed traffic flow with different vehicles," *Physica A: Statistical Mechanics and Its Applications*, vol. 413, pp. 320–328, 2014.

[19] X. Chang, H. Li, J. Rong, X. Zhao, and A. Li, "Analysis on traffic stability and capacity for mixed traffic flow with platoons of intelligent connected vehicles," *Physica A: Statistical Mechanics and its Applications*, vol. 557, Article ID 124829, 2020.

[20] A. Talebpour and H. S. Mahmassani, "Influence of connected and autonomous vehicles on traffic flow stability and throughput," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 143–163, 2016.

[21] J. Sawant, U. Chaskar, and D. Ginoya, "Robust control of cooperative adaptive cruise control in the absence of information about preceding vehicle acceleration," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 9, pp. 5589–5598, 2021.

[22] M. Montanino and V. Punzo, "On string stability of a mixed and heterogeneous traffic flow: a unifying modelling framework," *Transportation Research Part B: Methodological*, vol. 144, pp. 133–154, 2021.

[23] Y. J. Zhou, H. B. Zhu, M. M. Guo, and J. L. Zhou, "Impact of CACC vehicles' cooperative driving strategy on mixed four-lane highway traffic flow," *Physica A: Statistical Mechanics and Its Applications*, vol. 540, Article ID 122721, 2020.

[24] H. Bellem, T. Schönenberg, J. F. Krems, and M. Schrauf, "Objective metrics of comfort: developing a driving style for highly automated vehicles," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 41, pp. 45–54, 2016.

[25] R. Dang, J. Wang, S. E. Li, and K. Li, "Coordinated adaptive cruise control system with lane-change assistance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2373–2383, 2015.

[26] R. Jayachandran and S. Krishnapillai, "Modeling and optimization of passive and semi-active suspension systems for passenger cars to improve ride comfort and isolate engine vibration," *Journal of Vibration and Control*, vol. 19, no. 10, pp. 1471–1479, 2013.

[27] H. U. Ahmed, Y. Huang, and P. Lu, "A review of car-following models and modeling tools for human and autonomous-ready driving behaviors in micro-simulation," *Smart Cities*, vol. 4, no. 1, pp. 314–335, 2021.

[28] S. Ding, X. Chen, Z. Fu, and F. Peng, "An extended car-following model in connected and autonomous vehicle environment: perspective from the cooperation between drivers," *Journal of Advanced Transportation*, vol. 2021, Article ID 2739129, 17 pages, 2021.

[29] M. F. Aycin and R. F. Benekohal, "Comparison of car-following models for simulation," *Transportation Research Record*, vol. 1678, no. 1, pp. 116–127, 1999.

[30] A. Kesting and M. Treiber, "Calibrating car-following models by using trajectory data: methodological study," *Transportation Research Record*, vol. 2088, no. 1, pp. 148–156, 2008.

[31] M. Brackstone and M. McDonald, "Car-following: a historical review," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 2, no. 4, pp. 181–196, 1999.

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

[33] M. Treiber and A. Kesting, "The intelligent driver model with stochasticity-new insights into traffic flow oscillations," *Transportation Research Procedia*, vol. 23, pp. 174–187, 2017.

[34] H. Wang, W. Wang, J. Chen, and M. Jing, "Using trajectory data to analyze intradriver heterogeneity in car-following," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1, pp. 85–95, 2010.

[35] Y. Qin and H. Wang, "Analytical framework of string stability of connected and autonomous platoons with electronic throttle angle feedback," *Transportmetrica: Transportation Science*, vol. 17, no. 1, pp. 59–80, 2021.

[36] Y. Li, L. Zhang, S. Peeta, X. He, T. Zheng, and Y. Li, "A car-following model considering the effect of electronic throttle opening angle under connected environment," *Nonlinear Dynamics*, vol. 85, no. 4, pp. 2115–2125, 2016.

[37] K. Li and P. Ioannou, "Modeling of traffic flow of automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 2, pp. 99–113, 2004.

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

[39] G. S. Paddan and M. J. Griffin, "Evaluation of whole-body vibration in vehicles," *Journal of Sound and Vibration*, vol. 253, no. 1, pp. 195–213, 2002.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 2604150, 22 pages
https://doi.org/10.1155/2022/2604150

Research Article

Approximating Dynamic Equilibrium Analysis in Multi-Region Network Based on Macroscopic Fundamental Diagram

Bo Liu,[1,2] **Yanqing Cen**,[1] **and Xianghui Song**[1]

[1]*Research Institute of Highway Ministry of Transport, Beijing 100088, China*
[2]*Department of Automation, Tsinghua University, Beijing 100083, China*

Correspondence should be addressed to Bo Liu; boliu2015@163.com

Received 20 January 2022; Revised 7 July 2022; Accepted 9 July 2022; Published 26 August 2022

Academic Editor: Gen Li

Modeling and control of road traffic in large-scale urban networks present considerable challenges. The traffic equilibrium phenomena, with the question of route choice behavior in case of heterogeneous urban networks, has not been thoroughly investigated in parsimonious and classical models due to the limitation, like large network size, spatiotemporal propagation of congestion, and the interaction between driver decisions, etc. In this paper, we propose a bi-level approximating dynamic equilibrium model (BLADEM) for the approximating dynamic equilibrium analysis in multi-region network based on macroscopic fundamental diagram (MFD). The proposed model combines the region-based model and the internal-region model. With the information from region MFD, the region-based model is used to implement the time-dependent regional route choice estimation. Traffic equilibrium condition (dynamic user equilibrium, DUE) is considered in an internal-region model with time-aggregated regional OD demand from the region level. Furthermore, the complexity of the proposed model is derived. Then, the comparative analysis of the algorithm complexity between the proposed model and the DUE model is given. The proposed model is evaluated based on the high-resolution vehicle trajectory data (or connected vehicles trajectory data) from the DiDi platform collected in Chengdu, China with more than 3,000,000 GPS points during a typical workday. The evaluation results show that the proposed model can obtain the approximating traffic dynamics compared with the DUE algorithm. Pleasantly, the improved calculation efficiency is between 21% and 42%. The results indicate the promising potential of using the proposed model to analyze approximating dynamic equilibrium in the multi-region heterogeneous network.

1. Introduction

Improving mobility and accessibility in large-scale urban networks presents substantial challenges in the development of modeling, estimation, evaluation and control techniques. As the research foundations of estimation and evaluation in the network, traffic assignment is the main factor for traffic managers to estimate the expected state of the network and the development direction of the planning. However, due to some limitations, including large network size, unpredictability of travelers' behavior, spatiotemporal propagation of congestion, and the interaction between driver decisions and so on [1], the dynamic traffic assignment (DTA) with high complexity is difficult problem. Considering severe model uncertainty, and excessive computational burden associated with detailed link-level modeling and control methods, such the DTA approaches appear to be practically inefficient under congested conditions in large-scale urban networks. As an alternative to these link-level approaches, network-level methods employing aggregated modeling and control approaches using the parameter characteristics of regions and correlation between regions, receive increasing attention as practicable approaches for city-wide traffic control. Therefore, developing an aggregated model for large-scale urban networks is essential for both feasibility and efficiency reasons.

Macroscopic fundamental diagram (MFD) has been widely used for aggregate modeling of urban traffic network dynamics to tackle the dimensionality problem of microscopic approaches. The MFD provides a unimodal, low-

scatter, and demand-insensitive relationship between network vehicle density and space-mean flow in homogeneous urban areas (with small spatial link density heterogeneity) [2]. Furthermore, based on the emerging high-resolution vehicle trajectory data from the floating cars or CVs (connected vehicles), a range of advanced methods are adopted to measure the real-time MFD information of the network or the region. The MFD as a concept was first proposed in Godfrey (1969) [3] with an optimum accumulation and similar approaches were introduced later by [4–6]. The idea of network-level traffic control with an MFD-based model was originally proposed by Daganzo (2007) [6] for a single region. The work first proposes a steady-state approximation to establish the relation between regional outflow (i.e. trip completion rate) and accumulation. Based on this work, numerous MFD-based modeling and control methods have been developed for multi-region urban networks, like optimal control [7, 8], robust control [9, 10], model predictive control (MPC) [11–15], model-free adaptive control [16, 17], demand management [18].

In addition to purely control efforts, using the MFD framework to describe the dynamics of large-scale urban traffic network, has been further excavated. Recent research works have been dedicated to combining the DTA and the MFD framework by incorporating route choices or preferences for networks modeled by multiple MFD regions. Different from the existing vehicle dispatching frameworks, this fusion framework has analytically tractable nature [19] and gives rise to a promising solution to the challenge of spatial dimensionality in the meanwhile. Yildirimoglu and Geroliminis (2014) [20] proposed an aggregated DTA procedure for MFD-based dynamic stochastic user equilibrium (DSUE) conditions at the network level. Then, they developed the procedure with dynamic user equilibrium (DUE) for route choice behaviors [21]. Keyvan Ekbatani et al. (2015) [22] proposed a DTA procedure which was used in the multiple concentric gating traffic control method. To extend the MFD-based perimeter control to heterogeneous regions, Ramezani et al. (2015) [23] proposed a hierarchical (bi-level) perimeter control method. Yildirimoglu et al. (2018) [24] built a two-level route guidance system based on the MPC scheme to minimize the total travel time. These works proposed the region-based model with user equilibrium conditions and achieved a good effect. However, the limitations are obvious because there are no dispatching models within the region. The sub-regions divided by regions are the smallest control objects. As is well-known, although we use MFD to analyze the aggregate dynamics at network level, the detailed description of strategy more likely to be accepted for travelers and managers at link level. Therefore, in this paper, we developed the model of Yildirimoglu et al. (2015) [21] to establish links as the smallest control objects. Moreover, the results of modeling at link level can be approximately equal to that at the sub-region level on a larger scale. Nevertheless, ignoring the probability that the path selection behavior obeys the wardrop principle [25] and other disturbances, the sub-region level model has drawbacks in a simple sub-region (only few links inside).

Recently, several researches introduced some points worth paying attention, including modeling the boundary dynamics, the network separability and the DUE application principle. For modeling the boundary dynamics, a mechanism of dynamically rescaling the MFD after excluding the queued vehicle was proposed by Ni and Cassidy (2020) [26]. The mechanism assumes that the region's capacity to the within-region traffic diminishes in proportion to the import lanes spaces occupied by the queued vehicles. However, this proportion varies with boundary node topology and control mode so that it is difficult to measure in practical. Guo and Ban (2020) [27] assumes that the region is not completely jammed so that the queued vehicles in its buffer zone will not severely affect the overall congestion of the entire region. Obviously, for the congested region, the state of buffer zone can affect the overall congestion of the entire region based on the transmissibility of traffic flow. Not to mention, the buffer zone is inside the region, not really an independent zone. In this paper, we focused directly on link-level within the region so that the dynamics of boundary nodes can be obtained from the dynamics of the links in all directions. For the network separability, Aghamohammadi and Laval (2019) [28] verify that the large-scale network or region can partition into smaller "cells" (sub-regions) and assume that congestion is homogeneously distributed in each cell. We adopt this point in this paper as an assumption to divide the experimental network into four regions (see in Section 3) and the region is further divided into sub-regions (see in Section 4.3). For the DUE application principle, Huang et al. (2020) [29] proposed the DUE model through the differential variational inequality based on the multi-region MFD dynamics with saturated state and inflow constraints. Then, a methodological framework for estimating traffic-dependent distributions of trip lengths was proposed by Batista et al. (2021) [30], which incorporated in the R-DTA proposed by Batista and Leclercq (2019) [31]. However, different from the proposed model in this paper, the DUE models established by these studies are limited to the region level and cannot guarantee DUE conditions within regions.

To sum up, this paper provides contributions in DUE conditions on the following directions: (i) developing the approximate traffic equilibrium conditions to be integrated in a large-scale network which is modeled with multiple MFDs for different regions, (ii) collaborating the DUE model between the region and within region based on the consistency of the boundary nodes dynamics, instead of the DUE condition which is limited to the region level. The remainder of the paper is organized as follows; in Section 2, we introduce the dynamics of the region-based model and the internal-region model, including the calculation procedure and complexity analysis. In Section 3, the detailed experimental design is given based on the high-resolution vehicle trajectory data from the DiDi platform. Section 4 presents and discusses results of the experimental design with different analysis perspectives. Finally, Section 5 concludes the paper with future work directions.

2. Method

In this study, a methodological framework, called bi-level approximating dynamic equilibrium model (BLADEM), that combines region-based model and internal-region model is introduced, which can integrate aggregated route choice into advanced traffic management strategy (e.g., perimeter control) in real time and optimize the network operation performance. The urban network is partitioned into n regions with a low-scatter MFD in each region (that means each region has a homogeneous distribution of congestion). With the travel production (veh.m/s) and region accumulation (veh) from MFD, the region-based model is used to implement the time-dependent regional route choice estimation. Traffic equilibrium condition (dynamic user equilibrium, DUE) is considered in an internal-region model with time-aggregated regional *O-D* demand from region level. Based on the regional path and network topology within a region, the internal-region traffic dynamics is re-described by one-to-all time-dependent shortest path algorithm depending on DUE. Furthermore, the region accumulation and average trip length are also updated, which will cause a change in the estimated result of regional route choice. Therefore, there is an iterative process until the convergence conditions are reached between the BLADEM model.

From the above, the bi-level modeling framework enables to measure and control traffic state at different layers and incorporate heterogeneity effect in the urban network dynamics, see Figure 1. Methodology section is structured as follows. The next subsection introduces the aggregated network dynamics with the newly proposed dynamic aggregated region route choice. The following subsection provides a detailed DUE process which expounds internal-region traffic dynamics. The last subsection presents traffic equilibrium analysis flowchart in a bi-level model and methods for updating aggregate route choice parameters during the process.

2.1. Aggregated Route Choice in Region-Based Model. Let us assume that a large-scale network $G = (V, E, n)$ is partitioned into n regions with a heterogeneous distribution of congestion, $\mathbb{R} = \{R_1, R_2, \ldots, R_n\}$. In network topology with multi-regions, V and E represent nodes (intersections) and edges (links), respectively. For internal-region, the traffic dynamics is well described by a low-scatter MFD $F(\mathrm{Pr}_{R_n}(N_{R_n}), N_{R_n})$, where $\mathrm{Pr}_{R_n}(N_{R_n}(t))$ is defined as the travel production[veh.m/s] at time t corresponding region R_n accumulation $N_{R_n}(t)$ (veh). What is more, according to the characteristics of MFD, the average trip length $L_{R_n}(t)$ is constant if there is no change for the route choice and traffic control scheme. Then, we can get the trip completion rate $M_{R_n}(t) = (\mathrm{Pr}_{R_n}(N_{R_n}(t))/L_{R_n}(t))$ (veh/s) and the average speed $v_{R_n}(t) = (\mathrm{Pr}_{R_n}(N_{R_n}(t))/L_{R_n}(t))$, which have been reported in [7, 8, 11].

In an urban network $G = (V, E, n)$, any given traditional OD demand (represented by a sequence of nodes or links) can be transformed into a region path $p_{U,K}$ from region U to the final destination K. As shown in Figure 2, a region trip is denoted as $p^{B\&D}_{A,E}\{R_A, R_B, R_D, R_E\}$. Accordingly, $p^{B}_{A,E}$ is a path containing the sequence of regions to reach E starting from A and through the region B. Obviously, $p^{B\&D}_{A,E}$ is a form of

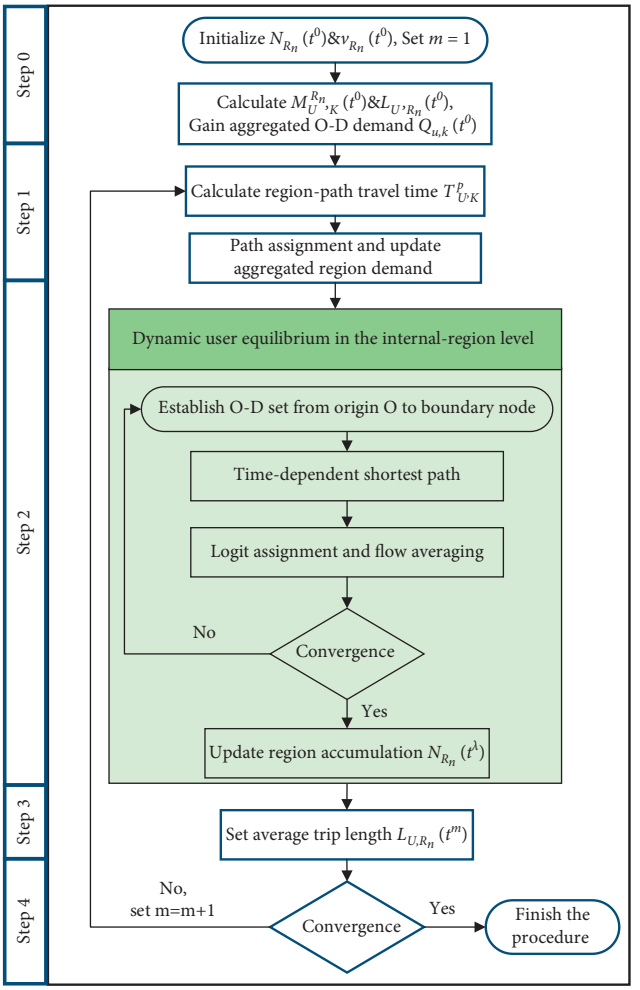


Figure 1: Methodological framework.

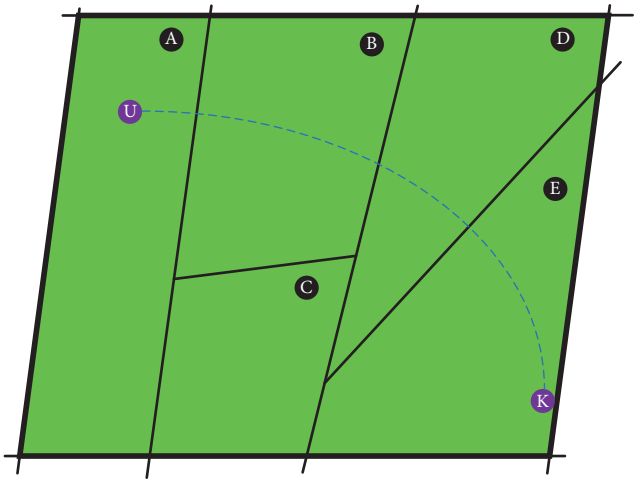


Figure 2: The schematic of a multi-region urban network and a path example.

$p^B_{A,E}$. Let $Q_U(t)$ (veh/s) denote the total traffic demand generated in the region U at the time t, $Q_{U,K}(t)$ be the exogenous traffic flow demand generated in the region U with the final destination K. Particularly, $Q_{U,U}(t)$ is defined as internal traffic flow demand in the region U without going through another region. For $U, K \in \mathbb{R}$, we have $Q_U(t) = \sum_{K\in\mathbb{R}} Q_{U,K}(t)$.

Approximately, $N_U(t)$ [veh] is the total accumulation in the region U at the time t and $N_{U,K}(t)$ is the vehicle accumulation from region U to the final destination K. In traffic flow conversion process in n regions, more than one route collaboratively to complete the transfer trips for a given OD pair. Let $N^{R_n}_{U,K}(t)$ denote the outflow from the region U to the final destination K through the region R_n; $U, K \in \mathbb{R}$, $N_U(t) = \sum_{R_n\in H_U}\sum_{K\in\mathbb{R}} N^{R_n}_{U,K}(t)$, where H_U is set of next regions(directly adjacent regions) for the region U. Let $q^{p^{R_n}_{U,K}}$ be the demand of $p^{R_n}_{U,K}$. Consequently, the instantaneous accumulation is computed as follow:

$$N_{R_n}(t) = \sum_{U\in\mathbb{R}}\sum_{K\in\mathbb{R}}\sum_{p\in W^{R_n}_{U,K}} q^{p^{R_n}_{U,K}}(t), \tag{1}$$

$$\begin{aligned} N^{R_n}_{U,K}(t) &= Q_{U,R_n}(t) \\ &= \sum_{p\in W^{R_n}_{U,K}} q^{p^{R_n}_{U,K}}(t), \end{aligned} \tag{2}$$

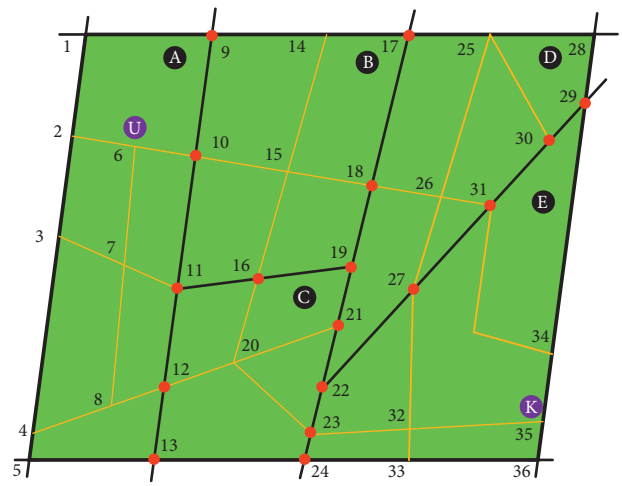


FIGURE 3: The schematic of a multi-region network with internal topology.

Where $W^{R_n}_{U,K}$ is the set of paths which reach K starting from U and through region R_n. The trip completion rate $M^{R_n}_{U,K}(t)$ for the vehicles from region U to the final destination K through the next region $R_n \in H_U$ is calculated by the following equation.

$$M^{R_n}_{U,K}(t) = \min\left[v_U(t)\cdot\frac{N^{R_n}_{U,K}(t)}{L_{U,R_n}(t)}, C_{U,R_n}\left(N_{R_n}(t)\right)\cdot\frac{N^{R_n}_{U,K}(t)}{\sum_{I\in\mathbb{R}}\left(N^{R_n}_{U,I}(t)\right)}\right], \tag{3}$$

Where $L_{U,R_n}(t)$ [m] is the average trip length and $C_{U,R_n}(N_{R_n}(t))$ is the boundary capacity corresponding region trips from region U to adjacent region $R_n \in H_U$. (3) indicates that the transfer flow between the consecutive regions is the minimum of two terms: (i) the transfer flow which depends on the traffic dynamics in region U and (ii) the boundary capacity between consecutive regions and the proportion of vehicles from region U to destination K through the region $R_n \in H_U$, among all the vehicles that cross the same boundary, i.e. $\sum_{I\in\mathbb{R}}(N^{R_n}_{U,I}(t))$. More details about the boundary capacity can be founded in previous publications (e.g., [21, 23]).

In a n regions MFD system, the traffic flow dynamic equations are listed below.

$$\frac{dN_U(t)}{dt} = Q_U(t) - \sum_{R_n\in H_U} M^{R_n}_{U,R_n}(t) + \sum_{R_n\in H_U} M^{U}_{R_n,U}(t), \tag{4}$$

$$\frac{dN_{U,K}(t)}{dt} = Q_{U,K}(t) - \sum_{R_n\in H_U} M^{R_n}_{U,K}(t) + \sum_{R_n\in H_U} M^{U}_{R_n,K}(t). \tag{5}$$

The traffic flow conservation of region is illuminated in (4), while the transfer flow dynamic is described by (5). It is noteworthy that, the vehicle paths to include more than one getting through the same boundaries is permitted, i.e., a region path $p^{C\&D}_{B,B}\{R_B, R_C, R_D, R_B\}$. For each region path p, p, ppp, p between origin O and destination D, the travel time $T^p_{O,D}$ is calculated based on total trip length $L^p_{O,D}$ and region average speed v_{R_n}. For example, to a region path $p^{B\&D}_{A,E}$ (see Figure 2), we have:

$$L^{p^{B\&D}_{A,E}}_{A,E}(t) = L_{A,B}(t) + L_{B,D}(t) + L_{D,E}(t), \tag{6}$$

$$\begin{aligned} T^{p^{B\&D}_{A,E}}_{A,E}(t) = &\frac{L_{A,B}(t)\cdot(N_A + N_B)}{v_A\cdot N_A + v_B\cdot N_B} \\ &+ \frac{L_{B,D}(t)\cdot(N_B + N_D)}{v_B\cdot N_B + v_D\cdot N_D} \\ &+ \frac{L_{D,E}(t)\cdot(N_D + N_E)}{v_D\cdot N_D + v_E\cdot N_E}. \end{aligned} \tag{7}$$

(7) indicates that average trip time is obtained by integrating average trip length and average speed between adjacent regions. Note that, in the region-based model, region path defines the sequence of regions from an origin node to a destination node, not the detailed sequence of links. However, the DUE condition is considered in an internal-region model that represents the connectivity about traffic flow distribution of links. Also, traffic equilibrium is based on the accumulation of the vehicles in the

corresponding region. In other words, the accumulation is extracted and updated from DUE in the internal-region model. Finally, the region path re-decision based on the changed average trip length and average speed in the region. The trip length between two adjacent regions, like $L_{A,B}(t)$ etc., can be calculated by the (12) (the initial length calculation method) and (14) (the length calculation method during iterative process). The DUE process and the estimation of traffic dynamic parameters $L_{U,R_n}(t)$, $T^p_{O,D}(t)$ and $C_{U,R_n}(N_{R_n}(t))$ in region level are discussed in the following subsection.

2.2. Traffic Dynamics in an Internal-Region Model. Consider time-aggregated regional OD demand whose dynamic variation trend is well described by aggregated route choice in the previous subsection. In other words, the traffic demand $Q_{U,R_n}(t), R_n \in H_U$ in consecutive regions is calculated in real time. Let $V^i_{U\cap R_n}$ denote the node i which belongs to the set of boundary nodes between the region U and adjacent region R_n, $V_{U\cap R_n} = \{V^1_{U\cap R_n}, V^2_{U\cap R_n}, \ldots, V^i_{U\cap R_n}\}$, like red nodes as shown in Figure 3. At internal-region level, the adjacent region trip p_{U,R_n} is taken into account as two form of sub-trip: (i) origin O or internal node in the region U to boundary node $V^i_{U\cap R_n}$ and (ii) boundary node $V^i_{U\cap R_n}$ to internal node in the region R_n or destination D.

Consequently, the traffic demand $Q_{U,R_n}(t)$ of the adjacent region trip p_{U,R_n} is conducted by two-time DTA (dynamic traffic assignment) algorithm. Furthermore, a DUE condition which requires minimal and equal travel times on alternative paths at the same time t [6], is considered for the internal-region traffic dynamics. Note that, in the region trip transformation process, there may be more than one path in each form of sub-trip and application of all boundary nodes in consecutive regions is permitted. That is, for each origin node o, we apply one-to-all time-dependent shortest path algorithm [32] depending on DUE to all boundary nodes, corresponding all-to-one time-dependent shortest path algorithm [33] is used to destination node d from all boundary nodes. For example, a region trip $p^B_{A,D}$ with origin node 6 to destination node 26 can be transferred as three-stage sub-trip, including origin node 6 to boundary nodes $V_{A\cap B} = \{V^9_{A\cap B}, V^{10}_{A\cap B}, V^{11}_{A\cap B}\}$; up-boundary nodes $V_{A\cap B}$ to down-boundary nodes $V_{B\cap D} = \{V^{17}_{B\cap D}, V^{18}_{B\cap D}, V^{19}_{B\cap D}\}$; boundary nodes $V_{B\cap D}$ to destination node 26 (see Figure 3).

The boundary capacity $C_{U,R_n}(N_{R_n}(t))$ as mentioned in the previous subsection can be regarded as the integration of the capacity of the boundary nodes on the driving direction of vehicles. In an urban network, boundary node generally represents the signal control intersection and have the different capacity at different directions, where the capacity of the node $V^i_{U\cap R_n}$ at direction com denotes as $C^{\text{com}}(V^i_{U\cap R_n})$.

$$C_{U,R_n}\left(N_{R_n}\right) = \sum_{V^i_{U\cap R_n}\in V_{U\cap R_n}} \left(C^{\text{com}}\left(V^i_{U\cap R_n}\right)\cdot \delta_{\text{com},r}\right), A. \tag{8}$$

Where $\delta_{\text{com},r}$ is an indicator function with a value equal to 1 if the direction com is the same as the path r, otherwise zero. Let $c^r_{o,d}$ represent the experienced travel time of route r from origin node o to destination node d, corresponding $\widehat{q}^r_{o,d}$ denotes as the assigned demand of route r from Q_{U,R_n} and l_r is the experienced trip length of the route r. The travel time of link t_e is calculated by BPR (Bureau of Public Roads) function, and we have $c^r_{o,d} = \sum_{e\in E} t_e \delta_{er}$, where δ_{er} is an indicator function with a value equal to 1 if the route r passes through the link e, otherwise zero. Time t is omitted from the following equations for the sake of notational simplicity. The traffic flow conservation equation at the boundary of regions is as follow:

$$\begin{aligned} Q_{U,R_n}|_{R_n\in H_U} &= \sum_{o\in U}\sum_{r}\sum_{V^i_{U\cap R_n}\in V_{U\cap R_n}} \widehat{q}^{\,r}_{o,V^i_{U\cap R_n}} \\ &= \sum_{d\in R_n}\sum_{r}\sum_{V^i_{U\cap R_n}\in V_{U\cap R_n}} \widehat{q}^r_{V^i_{U\cap R_n},d}. \end{aligned} \tag{9}$$

The results taken from DUE are processed through a well-known heuristic solution called a method of successive averages (MSA). It is effective and highly implemented in DTA [34]. In internal-region model, MSA is employed in each iteration to project future traffic information as part of the direction-finding mechanism in searching for a solution. The establishing DUE state can be summarized as follows:

Step 0. Initialization.

(a) Set iteration number $\lambda = 1$.

(b) Extract the traffic region demand $Q_{U,R_n}(t^\lambda)$ at a time t^λ from region-based level for $\forall U \in \mathbb{R}, R_n \in H_U$.

Step 1. Calculating assignment ratios

$$\eta_{o,V^i_{U\cap R_n}}\left(t^\lambda\right) = \frac{e^{-c_{o,V^i_{U\cap R_n}}}}{\sum_i e^{-c_{o,V^i_{U\cap R_n}}}}. \tag{10}$$

(a) For each origin internal-region node o, apply one-to-all time-dependent shortest path algorithm to boundary nodes and reserve the travel time $c_{o,V^i_{U\cap R_n}}(t^\lambda)$ which represents the lowest travel time origin o to boundary node $V^i_{U\cap R_n}$ at a time t^λ.

(b) Based on the lowest travel time $c_{o,V^i_{U\cap R_n}}(t^\lambda)$, compute the radio of the demand from origin o to boundary node $V^i_{U\cap R_n}$ with the following logit formula:

(c) Calculate the internal-region OD demand: $\widehat{q}_{o,V^i_{U\cap R_n}}(t^\lambda) = \eta_{o,V^i_{U\cap R_n}}(t^\lambda)\cdot Q_{U,R_n}(t^\lambda)$.

(d) For each boundary nodes, which demand is calculated by (10) based on from origin node o, apply all-to-one time-dependent shortest path algorithm to destination node d. Note that, implement a)-c) process in Step 1 for up-boundary nodes to down-boundary nodes. For example, the travel time $c_{V^i_{A\cap B},V^i_{B\cap D}}(t^\lambda)$ can be calculated to compute the radio of the demand for $\forall i$.

Step 2. DUE implementation

$$x_e^{\lambda+1} = x_e^{\lambda} + \left(\frac{1}{\lambda}\right)\left(y_e^{\lambda} - x_e^{\lambda}\right). \tag{11}$$

(a) Based on **Step 1**, well-informed internal-region OD demand can be calculated, and we have traffic flow $x_e(t^{\lambda}) = \sum_w \sum_r \widetilde{q}_w^r(t^{\lambda}) \cdot \underset{\sim}{\delta}_{er}$ for each internal-region OD pair $w \in W$, where W is the set of internal-region OD pairs.

(b) Apply shortest path algorithm for each OD pair and perform an all-or-nothing assignment for the chosen shortest path. Let $y_e^{\lambda}(t^{\lambda})$ represents the updated link flow after an all-or-nothing assignment.

(c) Find the iteration-direction based on MSA and time t^{λ} is omitted from the following equation:

Step 3. Stopping test

(a) For each (o, d, r) trio, evaluate $\varepsilon = [y_e^{\lambda} - x_e^{\lambda}]$.

(b) If $\varepsilon \geq \mathfrak{T}$, where $\mathfrak{T}$ is a pre-defined threshold, set $\lambda = \lambda + 1$ and go to Step 1. Otherwise, finish the procedure and return the updated region accumulation $N_{R_n}(t^{\lambda})$.

Note that the DUE process in internal-region model includes two main steps; Calculate well-informed internal-region OD demand based on region demand from region-based level and DUE implementation by MSA. The former is key to connecting region-based level and internal-region level. For the logit model, which manages the travelers' perception of travel time, can be replaced with C-logit or a cross-nested logit model if necessary. On the other hand, because fixed step size $\alpha = (1/\lambda)$, traffic assignment by MSA is general for small "toy" networks, corresponding slow convergence problems in large-scale networks. However, in this study, MSA is only used in the internal-region level assignment. Therefore, there are no limitations to use MSA because internal-region level can be regarded as small "toy" networks. Even if a more intricate is an alternative, it is not expected to improve the results and will bring additional time cost.

2.3. Traffic Equilibrium Analysis in the Bi-Level Model. Traditional traffic equilibrium analysis is complex and difficult to tackle in a large-scale network with a large number of nodes and links because it is difficult to get full path alternatives and determining the complex effective-path sets is time-consuming. As an alternative, the problem of traffic equilibrium analysis in this study is to provide the region path (met the DUE condition) in the region-based model and establish a DUE state in the internal-region model. Different from traditional traffic equilibrium analysis, this study produces a faster process and considers traffic dynamics in a rolling horizon framework, meaning that the method establishes a DUE state at the internal-region level and optimizes the region path based on updated region accumulations.

The region-based model employs time-dependent aggregated parameters to compute region path with demand assignment, i.e. $N_{U,K}^{R_n}(t)$, $M_{U,K}^{R_n}(t)$ and $L_{U,R_n}(t)$. Such a region path represents the behavior of the overall travelers for route choice. Based on the aggregated route choice at the macroscopic level, the DUE state is established at the internal-region level, and actual traffic flow distribution and chosen path are obtained. In this way, the results of the traffic equilibrium analysis are approximate to traditional models, and the corresponding numerical test is provided in section 4.

For an initialized urban network (traffic flow has been loaded), based on traffic information database of ITS (intelligent transportation system), accumulation of region and distribution of traffic flow are well known before the network optimization based on detector data, i.e. the vehicle trajectory data from equipped GPS (Global Position System) on vehicle and fixed loop detector or camera at the intersection. Therefore, the computing region path is not a complicated project. What is more, compared with the traditional model, reducing the application scope DUE to internal-region level makes the calculation time greatly reduced.

Let $N_{U,K}(t) = N_{U,K}(t^0)$ and $M_{U,K}^{R_n}(t) = M_{U,K}^{R_n}(t^0)$ for $U, K \in \mathbb{R}, R_n \in H_U$ at the initial time t^0. Consequently, we can get the initialized average trip length $L_{U,R_n}(t^0)$ as follow.

$$L_{U,R_n}\left(t^0\right) = v_U\left(t^0\right) \cdot \frac{N_{U,K}^{R_n}\left(t^0\right)}{M_{U,K}^{R_n}\left(t^0\right)}. \tag{12}$$

Note that, we assume that complete information about MFD is taken from ITS of an initialized urban network. The traffic equilibrium analysis flowchart in the bi-level model is presented in Figure 1, while the corresponding algorithm is summarized as follows:

Step 0. Initialization.

(a) Set iteration number $m = 1$.

(b) Initialize region accumulation $N_{R_n}(t^0)$ and speeds $v_{R_n}(t^0)$ based on traffic data acquisition from the network.

(c) Apply (1)–(5) to gain the required inputs, i.e. $M_{U,K}^{R_n}(t^0)$, $N_{U,K}(t^0)$. The initialized average trip length $L_{U,R_n}(t^0)$ is calculated by (12).

(d) Based on the initialized *O-D* demand, get the conversion of the exogenous traffic flow demand $Q_{U,K}(t^0)$ (aggregated region *O-D* demand). Note that, the demand which needs to be assigned $Q_{U,K}(t^1) = Q_{U,K}(t^0)$ because no optimization has been made at the initial time t^0.

Step 1. Calculating aggregate route choice parameters

$$\phi_{U,K}^p(t^m) = \frac{e^{-T_{U,K}^p}}{\sum_p e^{-T_{U,K}^p}}. \tag{13}$$

(a) For each origin region U to the final destination K, apply (6) and (7) to calculate the travel time $T_{U,K}^p$ of effective region path p for $\forall p \in W_{U,K}$.

(b) Based on the travel time $T_{U,K}^p$, compute path assignment ratios with the following logit formula:

(c) Calculate the path assignment demand: $q^p_{U,K}(t^m) = \phi^p_{U,K}(t^m) \cdot Q_{U,K}(t^m)$.

(d) Update aggregated region demand $Q_{U,R_n}(t^m)$ by (2) for $\forall U \in \mathbb{R}, R_n \in H_U$.

Step 2. Establishing DUE state

(a) Apply Step 1-3 from the DUE algorithm.

(b) Return the new accumulation $N_{R_n}(t^\lambda)$ based on the DUE result in the internal-region model.

Step 3. Update aggregate route choice parameters

$$L_{U,R_n}(t^m) = \frac{\sum_{o\in U}\sum_r \sum_{V^i_{U\cap R_n}\in V_{U\cap R_n}} \left(\widetilde{q}^r_{o,V^i_{U\cap R_n}} \cdot l_r\right) + \sum_{d\in R_n}\sum_r \sum_{V^i_{U\cap R_n}\in V_{U\cap R_n}} \left(\widetilde{q}^r_{V^i_{U\cap R_n},d} \cdot l_r\right)}{N^{R_n}_{U,K}(t^m)}. \tag{14}$$

(a) Update the aggregate parameters: let the new region accumulation $N_{R_n}(t^m) = N_{R_n}(t^\lambda)$ in region level and implement a region-based model (1)–(5) to get region level dynamics, i.e. $M^{R_n}_{U,K}(t^m)$, $N_{U,K}(t^m)$.

(b) Calculate the new average trip length as follow:

Step 4. Stopping criteria

(a) Evaluate $M = \sum_{R_n\in\mathbb{R}}(N_{R_n}(t^m) - N_{R_n}(t^{m-1}))^2$.

(b) If $M \geq \aleph$, where $\aleph$ is a pre-defined threshold, set $m = m + 1$ and go to Step 1. Otherwise, finish the procedure.

It is noteworthy that the iteration number (m and λ) is different in the bi-level model, meaning that each iteration in the region-based model corresponds to a complete DUE process in internal-region level. In the rolling horizon framework, all parameters can be obtained or calculated in real time. Therefore, the proposed traffic equilibrium analysis in this study is applicable for network performance analysis or route guidance with high timeliness. Particularly, the aggregated region OD demand $Q_{U,K}(t)$ can be regarded as a constant without elastic demand in the network. In other words, the proposed model be able to work well in the day-to-day network because the aggregated region OD demand is dynamic in each iteration.

2.4. The Complexity of Approximating Dynamic Equilibrium Analysis. The input information of the proposed model of approximating dynamic equilibrium analysis is mainly based on the initial network flow distribution and macroscopic traffic parameter information ($N^{R_n}_{U,K}(t)$, $M^{R_n}_{U,K}(t)$, $L_{U,R_n}(t)$) at time t provided by the MFD of each region. The complexity of the proposed model is as follows:

Step 0.The complexity of establishing the initialized network:

(a) -b) Obtain the MFD and MFD parameters according to the average network flow and network density based on the network flow distribution. Set the number of links in the region R_n as E_{R_n} and the total of the nodes as V_{R_n}. There are g pairs in the directly connected regions. Calculating the average network flow is a weighted average of the link flow. The complexity is $O(\sum_{R_n} E_{R_n})$. Similarly, the complexity of the average network density is $O(\sum_{R_n} E_{R_n})$;

(c) Get the MFD of the regions and calculate $M^{R_n}_{U,K}(t^0)$, $N_{U,K}(t^0)$, $L_{U,R_n}(t^0)$. The complexity is $O(3 \cdot n\wedge 2)$. It is worth noting that, the value of $M^{R_n}_{U,K}(t^0)$ is zero for indirectly connected regions, however, it still needs to be calculated to form a matrix list of the parameter information of each region;

(d) Construct the regional OD matrix: only allocate the traffic demand to all the OD pairs at time t, and add them to the corresponding regional OD matrix.

Step 1. The complexity of calculating the regional route choice parameters:

$$\begin{aligned} O &= n \cdot O(n \cdot \log n + (K + g) \cdot \log g) \\ &= O(n\wedge 2 \cdot \log n + n \cdot (K + g) \cdot \log g). \end{aligned} \tag{15}$$

(a) Calculate the travel cost of the route between regions. The total of routes needs to be calculated is $W_{U,K}$. The maximum complexity is $O(n \cdot (n-2)!)$. However, because the number of routes between regions is small and the proposed model has an iterative process, all routes between regions will be involved in the iterative process. Therefore, taking the k shortest paths between the origin region and the destination region in each iteration is enough. The complexity is as follows:

(b) -c) The complexity of calculating the probability of all route choices is $O(k \cdot n)$ and allocating the region demand is $O(k \cdot n)$;

(d) According to the traffic demand allocated by the regional route, to determine the traffic flow between various regions only needs to traverse the regional route of the allocated traffic demand and superimpose the traffic of the same regional pair, which is similar to the flow conversion between routes and links in a standard traffic network. If the regional route contains h_{R_n} inter-regional links on average, the complexity is $O(h_{R_n} \cdot k \cdot n)$.

Step 2. The complexity of DUE in region a) According to the traffic demand $Q_{U,R_n}(t^m)$ between regions and the inter-regional traffic demand, solve the DUE by MSA algorithm. Assume that any shortest path has $h^{in}_{R_n}$ links and DUE needs to be iterated λ times in the region, the complexity is as

follows according to the analysis of DUE in Appendices.

$$O\left(\lambda \cdot \sum_{R_n}\left[E_{R_n} + V^i_{U\cap R_n} \cdot V_{R_n} + \left(V_{R_n} + h^{in}_{R_n} + 1\right) \cdot V_{R_n} \wedge 2\right]\right). \quad (16)$$

(b) Update the average traffic flow of each region according to the link flow calculated by the DUE in the region. The complexity is $O(\sum_{R_n} E_{R_n})$.

Step 3. The complexity of calculating the macroscopic parameter

(a) The complexity of updating the macroscopic parameter $M^{R_n}_{U,K}(t^m)$ and $N_{U,K}(t^m)$ is $O(2 \cdot n \wedge 2)$;

(b) Based on the updating formula of $L_{U,R_n}(t^m)$, count the number of combinations of internal-region nodes and boundary nodes between the regions. The complexity is $O(2V^i_{U\cap R_n} \cdot V_{R_n})$.

Step 4. The complexity of stopping criteria

The complexity of evaluating M is $O(n)$.The iterative process is Step 1 to Step 4, and the number of iterations is m times. In summary, the complexity of the proposed model is as follows:

$$O(\text{BLADEM}) = \left\{ 2 \cdot O\left(\sum_{R_n} E_{R_n}\right) + O(3 \cdot n \wedge 2) + O(V \wedge 2) + m \cdot \left[\begin{array}{c} O(n \cdot (n-2)!) + O(n \wedge 2 \cdot \log n + n \cdot (k+g) \cdot \log g) + 2 \cdot O(k \cdot n) + \\ O\left(h_{R_n} \cdot k \cdot n\right) + O\left(\lambda \cdot \sum_{R_n}\left[2E_{R_n} + V^i_{U\cap R_n} \cdot V_{R_n} + \left(V_{R_n} + h^{in}_{R_n} + 1\right) \cdot V_{R_n} \wedge 2\right]\right) + \\ O\left(\sum_{R_n} E_{R_n}\right) + O(2 \cdot n \wedge 2) + O\left(2\sum_{R_n} V^i_{U\cap R_n} \cdot V_{R_n}\right) + O(n) \end{array} \right] \right\}. \quad (17)$$

Based on the qualitative analysis of the network topology, the complexity of the proposed model can be simplified. First of all, the number of regions that the urban transportation network is divided into is limited. Therefore, the general value is $n = [3, 6]$ and $\log n \approx [1, 2]$. In the k shortest search algorithm, k can be selected according to actual needs. In the proposed model, the value is 3. Due to the limited layout of the regional locations, there is less direct communication between regions. Therefore, the value of g is lower and set $\log g \approx [1, 3]$. $h^{in}_{R_n}$ is the average number of links contained in the shortest path within the region. Since any region is much smaller than the entire transportation network, $h^{in}_{R_n} \ll h$ can be speculated. $V^i_{U\cap R_n}$ indicates the number of boundary nodes between regions and $\sum_{R_n} V^i_{U\cap R_n} \ll V$. Through ignoring the smaller complexity, the simplified expression of the complexity of BLADEM is as follows:

$$O(\text{BLADEM}) \approx \left\{ (2m \cdot \lambda + m + 2) \cdot O(E) + O(V \wedge 2) + m \cdot \left[\begin{array}{c} O\left(\lambda \cdot \sum_{R_n}\left[V^i_{U\cap R_n} \cdot V_{R_n} + \left(V_{R_n} + h^{in}_{R_n} + 1\right) \cdot V_{R_n} \wedge 2\right]\right) \\ + O\left(2\sum_{R_n} V^i_{U\cap R_n} \cdot V_{R_n}\right) \end{array} \right] \right\}. \quad (18)$$

2.5. Comparative Analysis of the Algorithm Complexity. According to the complexity of the DUE (see in Appendices) model $O(\text{DUE})$ and the proposed model $O(\text{BLADEM})$, it is possible to evaluate the pros and cons of the two models in terms of operational efficiency and applicability in the actual traffic environment by the difference comparison method. The complexity difference function $f(O(\Delta))$ of the two models is shown below:

$$\begin{aligned} f(O(\Delta)) &= O(\text{DUE}) - O(\text{BLADEM}) \\ &= O\left(J \cdot m_{\text{DUE}} \cdot [2E + (V + h + 1) \cdot V \wedge 2]\right) - V \wedge 2 \\ &\quad - \left(2m_{\text{BLADEM}} \cdot \lambda + m_{\text{BLADEM}} + 2\right) \cdot E - 2m_{\text{BLADEM}} \sum_{R_n} V^i_{U\cap R_n} \cdot V_{R_n} \\ &\quad - m_{\text{BLADEM}} \cdot \left(\lambda \cdot \sum_{R_n}\left[V^i_{U\cap R_n} \cdot V_{R_n} + \left(V_{R_n} + h^{in}_{R_n} + 1\right) \cdot V_{R_n} \wedge 2\right]\right). \end{aligned} \quad (19)$$

The complexity difference function can be adjusted as follows:

$$\begin{aligned} f(O(\Delta)) = {} & O\left[\left(J \cdot m_{\mathrm{DUE}} - m_{\mathrm{BLADEM}} \cdot \lambda - 0.5 m_{\mathrm{BLADEM}} - 1\right) \cdot 2E\right] \\ & - O\left[\left(J \cdot m_{\mathrm{DUE}} - 1\right) \cdot V \wedge 2\right] - O\left(2 m_{\mathrm{BLADEM}} \sum_{R_n} V^i_{U \cap R_n} \cdot V_{R_n}\right) \\ & + O\left[\left(J \cdot m_{\mathrm{DUE}} \cdot (V + h) \cdot V \wedge 2\right)\right] \\ & - O\left[m_{\mathrm{BLADEM}} \cdot \left(\lambda \cdot \sum_{R_n}\left[V^i_{U \cap R_n} \cdot V_{R_n} + \left(V_{R_n} + h^{in}_{R_n} + 1\right) \cdot V_{R_n} \wedge 2\right]\right)\right]. \end{aligned} \tag{20}$$

In order to facilitate the analysis, $f(O(\Delta))$ can be split into four parts $f(O(\Delta_1))$, $f(O(\Delta_2))$, $f(O(\Delta_3))$ and $f(O(\Delta_4))$:

$$\begin{aligned} f\left(O\left(\Delta_1\right)\right) &= O\left[\left(J \cdot m_{\mathrm{DUE}} - m_{\mathrm{BLADEM}} \cdot \lambda - 0.5 m_{\mathrm{BLADEM}} - 1\right) \cdot 2E\right], \\ f\left(O\left(\Delta_2\right)\right) &= O\left[\left(J \cdot m_{\mathrm{DUE}} - 1\right) \cdot V \wedge 2\right] - O\left(2 m_{\mathrm{BLADEM}} \sum_{R_n} V^i_{U \cap R_n} \cdot V_{R_n}\right), \\ f\left(O\left(\Delta_3\right)\right) &= O\left(J \cdot m_{\mathrm{DUE}} \cdot V \wedge 2\right) - O\left(\lambda \cdot m_{\mathrm{BLADEM}} \cdot \sum_{R_n} V^i_{U \cap R_n} \cdot V_{R_n}\right), \\ f\left(O\left(\Delta_4\right)\right) &= O\left(J \cdot m_{\mathrm{DUE}} \cdot (V + h - 1) \cdot V \wedge 2\right) - O\left(\sum_{R_n} \lambda \cdot m_{\mathrm{BLADEM}}\left(V_{R_n} + h^{in}_{R_n} + 1\right) \cdot V_{R_n} \wedge 2\right]\right). \end{aligned} \tag{21}$$

(a) For $f(O(\Delta_1))$, because m_{DUE} is the number of iterations of traffic assignment for the overall network, and λ is the number of iterations of traffic assignment within the regional network, there is $m_{\mathrm{DUE}} > \lambda$. The number of iterations m_{BLADEM} of traffic assignment between each regional network is generally less than 10. However, any vehicle entering the link at the beginning of Δt cannot leave the link before the end of Δt. Therefore, the value of is generally not small. Therefore, the value of J is generally not small. For example, the total research time is only 30 minutes, however, J is generally controlled at $[25, 35]$ to meet the constraints of DUE. To sum up, $f(O(\Delta_1)) \in (0, +\infty)$.

(b) For $f(O(\Delta_2))$, In generally, the number of nodes on the boundary is less than the total number of nodes in the region $V^i_{U \cap R_n} < V_{R_n}$. Existing $V = \sum_{R_n} V_{R_n}$ and $J \cdot m_{\mathrm{DUE}} \gg 2 m_{\mathrm{BLADEM}}$, the following can be obtained:

$$V \wedge 2 = \left(\sum_{R_n} V_{R_n}\right) \wedge 2 \gg \sum_{R_n}\left(V_{R_n}\right) \wedge 2 > \sum_{R_n} V^i_{U \cap R_n} \cdot V_{R_n}. \tag{22}$$

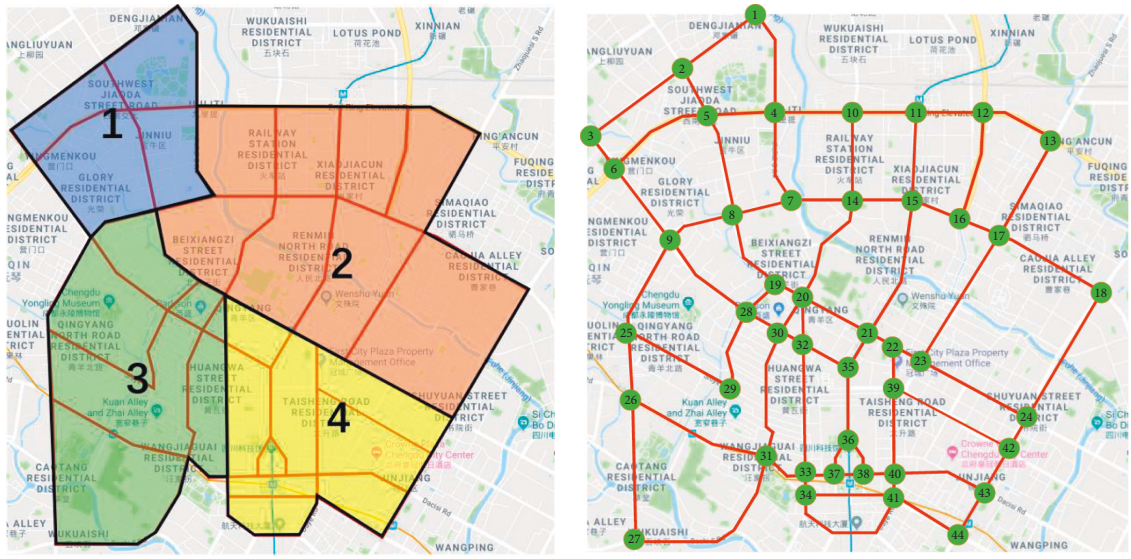
To sum up, $f(O(\Delta_2)) \in (0, +\infty)$.

(c) Obviously, for $f(O(\Delta_3))$, there is $f(O(\Delta_3)) \in (0, +\infty)$. Due to $V \wedge 2 \gg \sum_{R_n}(V_{R_n}) \wedge 2$, the following can be obtained:

$$J \cdot m_{\mathrm{DUE}} \cdot (V + h - 1) > \lambda \cdot m_{\mathrm{BLADEM}}\left(V_{R_n} + h^{in}_{R_n} + 1\right). \tag{23}$$

To sum up, $f(O(\Delta_4)) \in (0, +\infty)$. Note that, h and $h^{in}_{R_n}$ are the average number of links contained in the shortest path between the regions and within the region respectively. Therefore, there must be $h > h^{in}_{R_n}$ for a given network.

In summary, $f(O(\Delta)) = [O(\mathrm{DUE}) - O(\mathrm{BLADEM})] \gg 0$ can be concluded. That is, the complexity of BLADEM model must be lower than the DUE model. Moreover, based on the analysis of $f(O(\Delta_1))$, $f(O(\Delta_2))$, $f(O(\Delta_3))$ and $f(O(\Delta_4))$, when the topology structure of the network is complicated, the difference between the number of nodes in the network and the number of nodes in the regional network is greater, leading to the value of $f(O(\Delta))$ to be larger. Therefore, the BLADEM model has more obvious advantages in computational efficiency.



(a) (b)

Figure 4: The topology of the experimental network. (a) Regional representation (b) Link-level representation.

3. Experimental Design

The experimental network presented in Figure 4, including 44 signal-controlled intersections (nodes) and 70 links (edges), consists of four regions, where region 1 is a university area, region 2 is the workspace with a transportation hub, region 3 is tourist areas within the city, and region 4 is a representative CBD (Central Business District) area. To analyze the high-level traffic demand and time-varying traffic conditions in every region, the detailed traffic information (as inputs of the model) such as traffic flow, speed, the density of links need to be real time updated based on fixed-location detectors. However, subjecting to research conditions, only high-precision trajectory data of the floating car can be obtained through the GAIA opening data plan by DiDi. The trajectories data were collected from 10 : 00 a.m. to 12 : 00 a.m. on one workday (11/25/2016) and the total number of trajectory data points is 3,330,870. Each trajectory data point includes information on Vehicle_ID, GPS coordinates, and instantaneous speed at the frequency of 3 s.

The scheme of experimental design is as follows:

(i) Get the traffic flow of links. An analytical method for traffic flow estimation based on high-resolution vehicle trajectories can be adopted to get the traffic flow of links in the experimental network, seeing literature [35] for details.

(ii) Estimate the OD matrix. Since the purpose of this experimental design is to verify the efficiency and applicability of the BLADEM model, the actual traffic demand of network is only the basic input information during each iteration of the model. Therefore, the estimation of the OD matrix only requires approximate accuracy. In the experimental design, the method of estimating the OD matrix is adopted, seeing literature [36] for details.

(iii) Set the OD matrix as the initial input information of all models (including DUE model, BLADEM model and the previous model) in the experimental design.

(iv) The regional OD pairs are aggregated based on the distribution of origin and destination points across the regions. The boundary capacity of regions can be regarded as the integration of the capacity of intersection on the driving direction.

(v) Analyze results from three different perspectives: (1) Comparison of the traffic dynamics with the DUE algorithm; (2) Comparison of the running time with the DUE algorithm; (3) Comparison of the traffic dynamics with the previous algorithm.

Regarding the DUE state, the MSA method is adopted in the entire network or the internal-region (presented in Section 2.2) to achieve that the actual travel times experienced by travelers departing at the same time are equal and minimal for each OD pair [37]. However, different application scopes of the DUE model lead to significant differences in computing efficiency. Theoretical analysis is given in detail (presented in Section 2.5), and experimental results are presented in the following section.

It's worth noting that, the reason of choosing the real urban network as the experimental network is two-fold: to observe the application of the proposed model in the complex traffic network and to provide qualitative analysis of traffic dynamics. Regarding the former, due to the inputted traffic information is updated in real time and the topology of the network is extracted from the real urban network, the experimental design is more convenient for practice. For the other, based on the characteristics of regions, the rationality of regional dynamics can be discussed. The purpose of experimental design is elaborated as follows:

(1) Verify the accuracy of the proposed model in traffic assignment. Regarding the dynamic user equilibrium

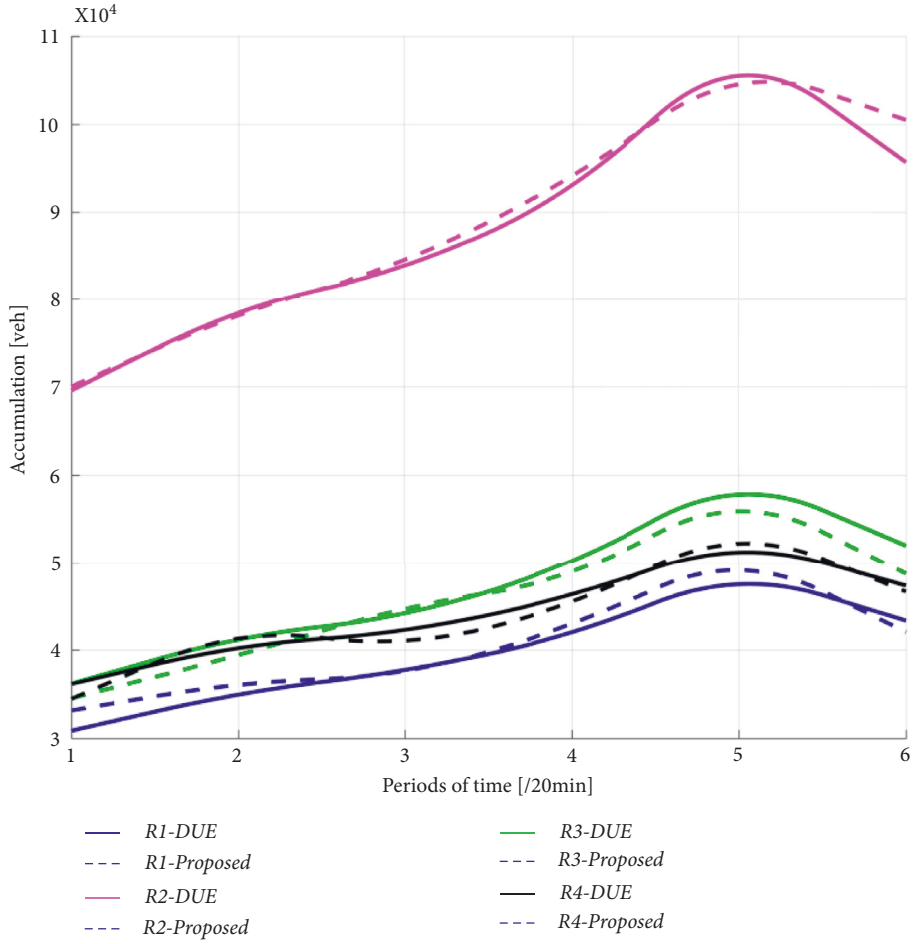


Figure 5: Accumulation of regions with different models.

assignment, we incorporate the MFD framework into the bi-level model. Through the macro-control at the regional level and micro-assignment at the link-level, DTA in a new profile is studied to satisfy DUE conditions. Compared to the DUE method, the bi-level model is valid if the difference between the results of traffic assignment is less than the threshold value under the DUE state.

(2) Verify the degree of improvement of computational efficiency. Due to multi-MFDs and dynamic regional paths are incorporated into the proposed model, the computational complexity is lower than the DUE algorithm. The difference in computational complexity is proved by comparing the running time of the proposed model and the traditional model to reach the DUE state under the same traffic demand in the same period.

In this paper, a large-scale network can be represented as a structure: network-region-link. There is an assumption: the network can be divided into several regions (with small spatial link density heterogeneity) with stable MFD. On the principle of partitioning, the fewer divided regions is the better under the condition that the MFD of the divided region is stable. This is because too many regions can greatly affect the computational complexity of network management control strategies. The proposed model seeks for approximate dynamic equilibrium after the region division is completed. Therefore, under the condition that the results of region division are consistent, region division does not affect the comparative analysis of the proposed model with traditional DUE model and other models. Moreover, the partition structure shown in Figure 4 is confirmed according to the partition principle. That is, the stability of the region MFD has been verified and the range of region is as large as possible.

4. Results and Discussion

This section consists of three subsections. Section 4.1 investigates how well the proposed model developed in section 2.3 can approximate the dynamics of the DUE algorithm. The results of the running time from the proposed model comparing the DUE algorithm with different traffic demand are presented in Section 4.2. Furthermore, comparison of the traffic dynamics with the previous algorithm (has the similar model structure) is discussed in Section 4.3. It should be clearly noted that the proposed model analyses approximate dynamic user equilibrium. The object that proposed models need to compare is DUE model. Therefore, we compared the

degree of approximation to DUE model from various perspectives in section of results and discussion, including accumulations of 4 regions, path travel time, MFD dynamics of each region, the average speed of vehicles in the regions, as shown in Figures 5–7. Figures 5–7 and Table 1 mainly demonstrates approximation and computational efficiency. The difference between proposed model and MRRGM model is mainly DUE application layer, not approximation and computational efficiency. As a matter of fact, the algorithm structure of the link level is more complex than the sub-region level, and the proposed model should be lower than the MRRGM model in computational efficiency.

4.1. Comparison of the Traffic Dynamics with the DUE Algorithm. This section introduces the comparison of the dynamic equilibrium conditions in the experimental network with the DUE algorithm. The purpose of the comparison is to prove the proposed model can obtain approximate traffic equilibrium analysis with the DUE algorithm. The comprehensive comparison includes multilevel analysis, such as region accumulation, the experienced travel time of path, the MFD characteristics of each region, the average speed of vehicles in the regions, etc.

The 2 hours of floating car trajectory data on the day of 11/25/2016 is divided into 6 sub-datasets. Every 2 minutes, the MFD information of the region is calculated and the OD matrix is estimated for each sub-dataset. That is, the BLADEM model and the DUE model are run in each sub-dataset (in each period). It is easy to get the traffic dynamics of each region to investigate how well the BLADEM model can approximate the dynamics of the DUE algorithm.

Figure 5 compares the accumulations of 4 regions with different models. Intuitively, the proposed model fits well with the evolution trend of accumulations from the DUE algorithm. Due to the limit of the topology of region network (the minimum arterials and area), the lowest region accumulation is region 1. However, similar the topology of region network does not necessarily have the same level of region accumulation, like region 2 and region 3. The commuting behavior of travelers will affect the distribution of traffic demand. Therefore, the region accumulation of region 2 is significantly higher than region 3. To some extent, it proves that the trajectory data extracted from the real urban network can help to analyze the rationality of regional dynamics. Note that approximate region accumulation does not imply complete agreement because a little deviation is inevitable. However, it can be applied well to the perimeter control of network flow with less elaborate requirement on macroscopic level.

Another important question is how reliable the proposed model is in dynamic user equilibrium states. Based on DUE conditions, for each OD pair, if the actual travel times experienced by travelers departing at the same time are equal and minimal. As shown in Figure 6, the random selected paths with the same origin node and destination node (node 5 and node 24) have approximately uniform path travel time. In other word, any path has the same chosen probability for each OD pair. Because the calculation of travel time within each region by internal-region model has some deviations, the aggregated deviations are inevitable in cross-region paths. However, these minor aggregated deviations are allowed within the margin of error. What is more, Figure 6 provides indirect evidence that the proposed model achieves approximate equilibrium conditions. As the blue line and pink line in Figure 6, the DUE-Path (calculated by the DUE algorithm) and the Proposed-Path (calculated by the proposed model) with same node transition sequence have the similar path travel time.

In order to analyze multi-region MFD dynamics, the travel production and region accumulation are calculated inside each region. It is worth noting that the similar MFD shapes are obtained for each region as shown in Figure 7. Based on the intuitive observation, the MFD dynamics of the region 2 have the best consistency between the DUE model and the proposed model, subjectively. On the contrary, some fluctuations were detected in the region 1 and 4. In general, the travel choices are increasingly unpredictable in densely populated areas, like the university area of region 1 and the CBD area of region 4. Therefore, the method of time-aggregated regional OD demand analysis based on the aggregated route choice in the region-based model is more different from the DUE algorithm in densely populated areas. It is worth noting that, since the experimental data are taken from the real traffic environment during 10 : 00 AM to 12 : 00 AM, the MFD dynamics of each region does not include the low traffic flow stage, such as the origin where the accumulation of regions is 0.

To verify the consistency of the MFD dynamics with the DUE algorithm, *K-S* test (Kolmogorov-Smirnov test) is employed as a numerical test, and related indicators are illustrated in Figure 8. In statistics, the *K-S* test [38] is a nonparametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare two samples (two-sample *K-S* test). The two-sample *K-S* test is one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples. Cumulative probability distribution as an indicator, is applied to check the evolution trend. The null hypothesis, which assumes the MFD dynamics are subject to the same distribution as the DUE algorithm, is accepted when probability (*K-S* test returns) is above 0.05. Then set the logical value of the hypothesis h to zero, otherwise equal to 1.

It is evident that the MFD dynamics based on the proposed model are subject to the same distribution as the DUE algorithm because all returned probabilities are more than 0.05. In other words, the proposed model can follow the evolution of MFD dynamics from the DUE algorithm, which is very important in the perimeter flow control for multi-region networks and can help traffic engineers to develop better traffic management projects. The optimal consistency in region 2 is confirmed due to the supreme returned probability $P = 0.909$. It is same with intuitive observation in Figure 7.

According to the above formula and the traffic distribution of the road network at each time, the average vehicle speed of the region can be calculated, as shown in Figure 9.

Table 1: Notation list.

Variable	Specification
G	Represents the large-scale network and $G = (V, E, n)$
$\mathbb{R}$	Represents the set of multi-regions, $\mathbb{R} = \{R_1, R_2, \ldots, R_n\}$
$F(\mathrm{Pr}_{R_n}(N_{R_n}), N_{R_n})$	Represents the MFD of region R_n
$\mathrm{Pr}_{R_n}(N_{R_n}(t))$	The travel production (veh.m/s) at time t of region R_n
$N_{R_n}(t)$	The accumulation (veh) at time t of region R_n
$L_{R_n}(t)$	The average trip length at time t of region R_n
$M_{R_n}(t)$	The trip completion rate at time t of region R_n
$v_{R_n}(t)$	The average speed at time t of region R_n
$p_{U,K}^{R_n}$	Represents a region path from region U to the final destination K through the region R_n
$Q_U(t)$	The total traffic demand generated in the region U at the time t
$Q_{U,K}(t)$	The exogenous traffic flow demand generated in region U with the final destination K
$N_{U,K}^{R_n}(t)$	The outflow from the region U to the final destination K through the region R_n
$q^{p_{U,K}^{R_n}}$	The demand of the region path $p_{U,K}^{R_n}$
H_U	The set of next regions (directly adjacent regions) for the region U
$W_{U,K}^{R_n}$	The set of paths which reach K starting from U and through region R_n
$M_{U,K}^{R_n}(t)$	The trip completion rate for the vehicles from region U to the final destination K through the next region $R_n \in H_U$
$C_{U,R_n}(N_{R_n}(t))$	The boundary capacity corresponding region trips from region U to adjacent region $R_n \in H_U$
$c_{o,d}^r$	Represent the experienced travel time of route r from origin node o to destination node d
$\widehat{q}_{o,d}^r$	Denotes as the assigned demand of route r from Q_{U,R_n}
l_r	The experienced trip length of the route r
t_e	The travel time of the link e

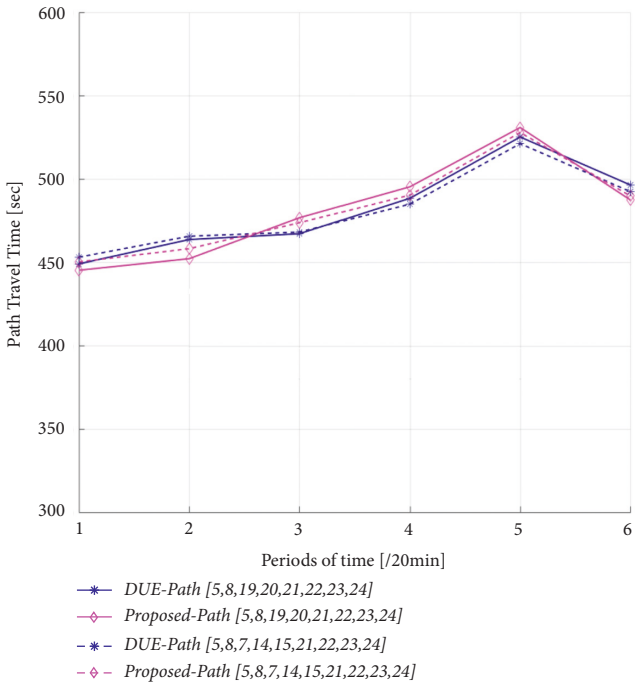


Figure 6: The travel time of cross-region path for the same OD pair.

The average speed of vehicles in the four regions is dynamic. To some extent, the average speed of vehicles in the road network indirectly reflects the degree of congestion of the road network. For example, in the initial period of the study period 0–54 min, the average speed of vehicles in region 4 is the lowest, indicating that the congestion level in region 4 is relatively high during this time range. In the 54–76 min period, the average speed of vehicles in region 2 is the lowest. From 76 minutes to the end of the study period, compared with other regions, the congestion degree of region 1 becomes the highest.

In addition, there are two key points in Figure 9:

(1) Although the time when the average vehicle speed of each region reaches the lowest point is different, it is concentrated between 30 and 50 minutes, that is, within the time period of 10 : 50–11 : 10. It shows that the congestion degree of the region is more obvious during the peak period;

(2) The average speed of vehicles in the regions is similar at any time. The reason for this phenomenon is that the upper model of BLADEM model reasonably allocates the region OD matrix based on the average length of region paths. At the region level, as shown in Figure 4(a), the travel costs of all region paths are approximately equal when the BLADEM model converges. When the difference of region OD matrix is not very large, the congestion degree of each region in a balanced state is not very different. This point can indirectly prove that the road network has approached the equilibrium state.

4.2. Comparison of the Running Time with the DUE Algorithm. In the same calculation and analysis environment, the BLADEM model and the DUE model are respectively applied to calculate and analyze the equilibrium state of the network based on the 6 sub-datasets. The advantages of the two models in calculation efficiency are compared according to the calculation running time. In experimental design, the configuration of the computer is: Intel(R) Core (TM) i7-8650U CPU; 16.0 GB RAM; MATLAB version is R2017a.

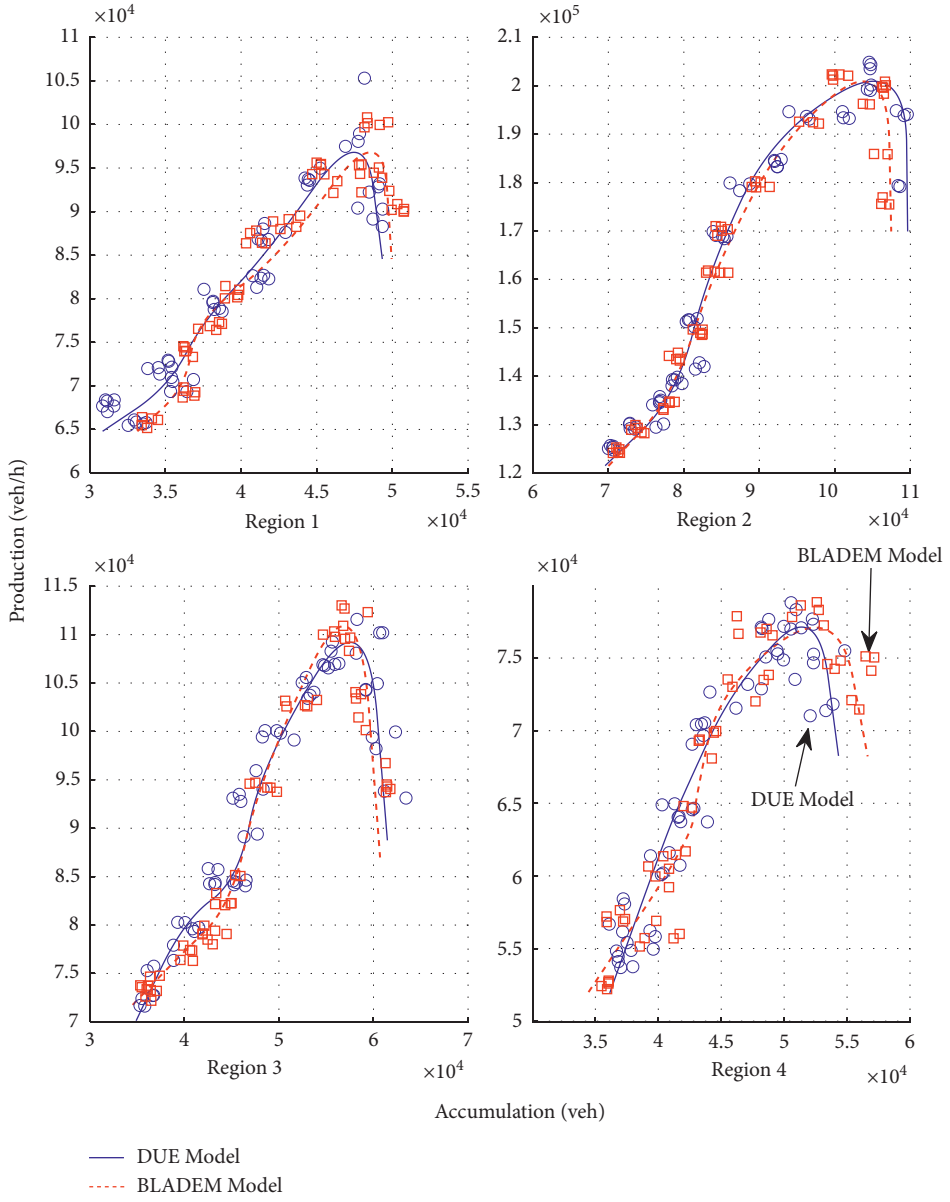


Figure 7: Consistency validation of dynamic characteristics in MFD.

As shown in Table 2, under the same calculation and analysis environment, the calculation time of the BLADEM model is generally lower than the DUE model. Therefore, combined with the verification and analysis in Section 4.1, compared with the DUE model, the BLADEM model can effectively improve the calculation efficiency under the premise of obtaining approximate calculation results. The improved calculation efficiency is between 21% and 42%, which means that the average calculation time is saved by 35.59%. It is worth noting that although the data types of the 6 sub-datasets are the same, there are still differences in computing time under the same computing environment using the same model. The reason for this difference is not only the impact of different CPU temperatures during the calculation and analysis, but also the impact of the difference in the distribution of traffic demand in the initial state of the road network in 6 groups and 4 regions. However, the calculation time of the BLADEM model is relatively more stable.

Compared with other periods, the traffic flow of the traffic network is the lowest in the periods "10 : 00–10 : 20", as shown in Figure 7. This means that there are the less regional OD pairs when the OD matrix is estimated. Because the path of region OD is too long, it will increase the computational complexity of the DUE model. Therefore, the computational time of the DUE model will be reduced. However, for the proposed model, due to the bi-level distribution processing, there is no significant difference in computational complexity between the regional OD pairs and the OD pairs within regions. In other words, without increasing regions, the change of network traffic flow does not affect the computational complexity of the proposed model.

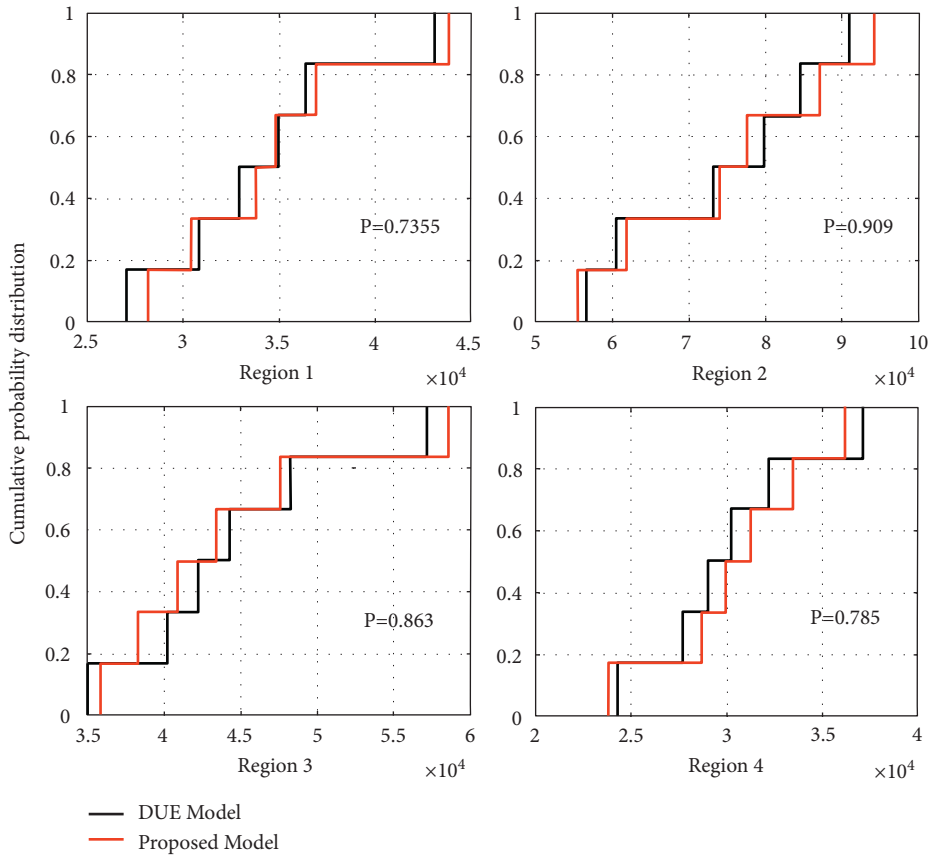


Figure 8: *K-S* test results for different MFD dynamics.

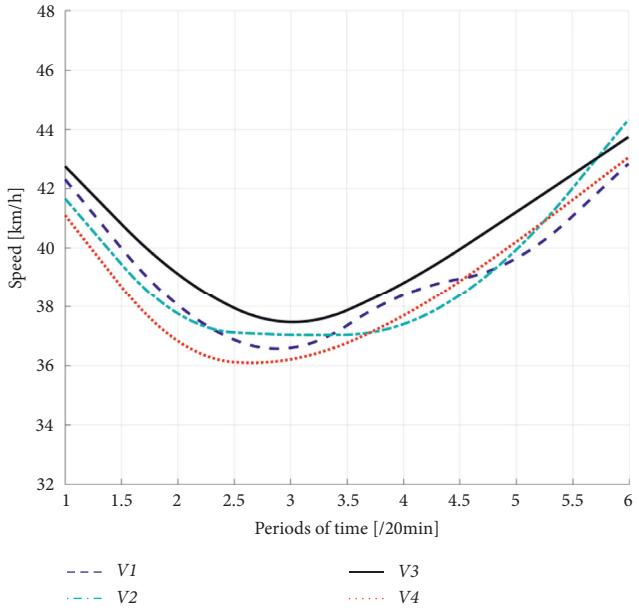


Figure 9: The average speed of vehicles in the regions.

4.3. Comparison of the Traffic Dynamics with the Previous Algorithm. The allocation of region OD matrix based on the MFD information of the region has been carried out by many researchers. The most prominent one is the Multi-Region Route Guidance Model (MRRGM) based on dynamic MFD proposed by Yildirimoglu et al. [21]. The main research approach is as follows:

(1) Divide multiple sub-regions within the region, assuming that each sub-region has a stable MFD;

(2) Based on the MFD information of the sub-region, calculate the average travel speed of the vehicle in the sub-region, and then calculate the travel impedance in the sub-region;

(3) According to the travel impedance in the sub-region, calculate the dynamic traffic assignment of the sub-region route in each region, and update the MFD of the sub-region;

(4) Recalculate the MFD information of the region according to the updated MFD of the sub-region, and readjust the traffic demand assignment result of the region route.

The MRRGM model provides travel guidance information for region route, which is mainly expressed in the form of region sequences. For the large-scale urban transportation network, there are two shortcomings when facing the control of the actual transportation network:

(1) The travel guidance information of the region route can only facilitate managers to control the traffic flow distribution of the regions. But for travelers, the travel guidance information is still the best

Table 2: Comparison of the running time.

Periods of time	DUE model (sec)	BLADEM model (sec)	Degree of improvement (%)
10 : 00–10 : 20	311.16	245.57	21.08
10 : 20–10 : 40	394.27	239.38	39.29
10 : 40–11 : 00	385.12	241.65	37.25
11 : 00–11 : 20	412.61	241.03	41.58
11 : 20–11 : 40	384.57	234.53	39.02
11 : 40–12 : 00	401.65	259.83	35.31

constructed by the links, so that the travelers follow the guidance information to change the travel route. Therefore, the travel route guidance provided by the MRRGM model has a limited range of actual traffic applications.

(2) The MRRGM model assumes that the region can be divided into multiple sub-regions with stable MFD, but the reasonable division of the network itself is a complex traffic problem. The reasonable division of multiple sub-regions will be too much time and resources are consumed, which reduces the application efficiency of the model. In addition, in the actual traffic environment, the region may not be exactly divided into a limited number of sub-regions with stable MFD to completely cover the region.

The BLADEM model is different from the MRRGM model in that: first, the BLADEM model only needs to calculate the MFD of region; secondly, the BLADEM model provides a standard path composed of links.

For the same network (the topology is regions-subregions-links-nodes, as shown in Figure 10), the output of the MRRGM model of is the equalization of sub-region flow. The output of the BLADEM model is the equalization of link flow. However, the traffic flow of each sub-region can be deduced inversely based on the traffic distribution of the link calculated by the BLADEM model. Comparing the sub-region flow from the BLADEM model and the MRRGM model, it can be found that the distribution of traffic in sub-regions is sometimes consistent and sometimes inconsistent. The BLADEM model verifies the approximation to DUE. The deduced inversely sub-region flow meets the DUE requirements. Therefore, the BLADEM model can obtain the sub-region flow more in line with DUE.

In order to verify this point, reconstruct the topology of the network shown in Figure 4 to obtain the network as shown in Figure 10: the network topology is $G = (\{R_n\}, \{\text{Sub} - R_n\}, \{E\}, \{V\})$, that is, the network contains V nodes, E edges (links), n regions R_n and n_{sub} sub-regions $\text{Sub} - R_n$.

As shown in Figure 10, different colors represent different regions, and each region is divided into 2–3 sub-regions. The sub-region is divided according to the nature of the land and the level of surrounding links. The model of the region divided into sub-regions is not used for calculation, but the stability of the sub-region MFD has been verified based on the traffic distribution calculated by the BLADEM model. It proves the rationality of the divided sub-region. The comparison of the accumulations in the sub-region of

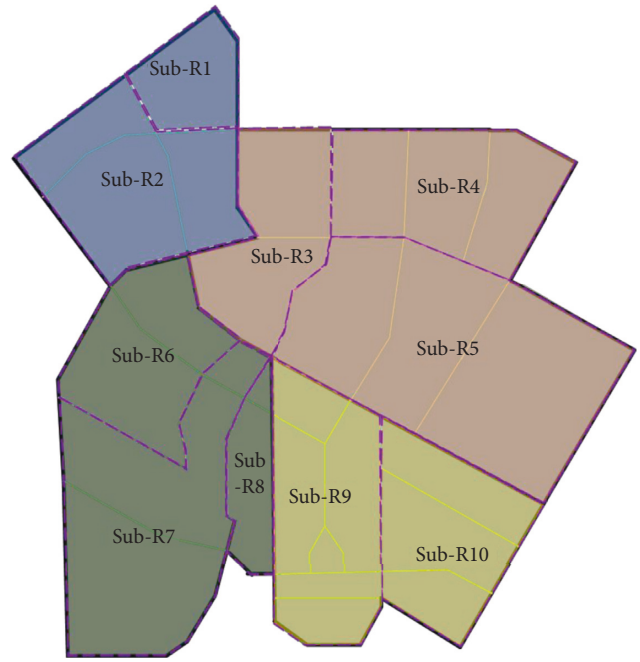
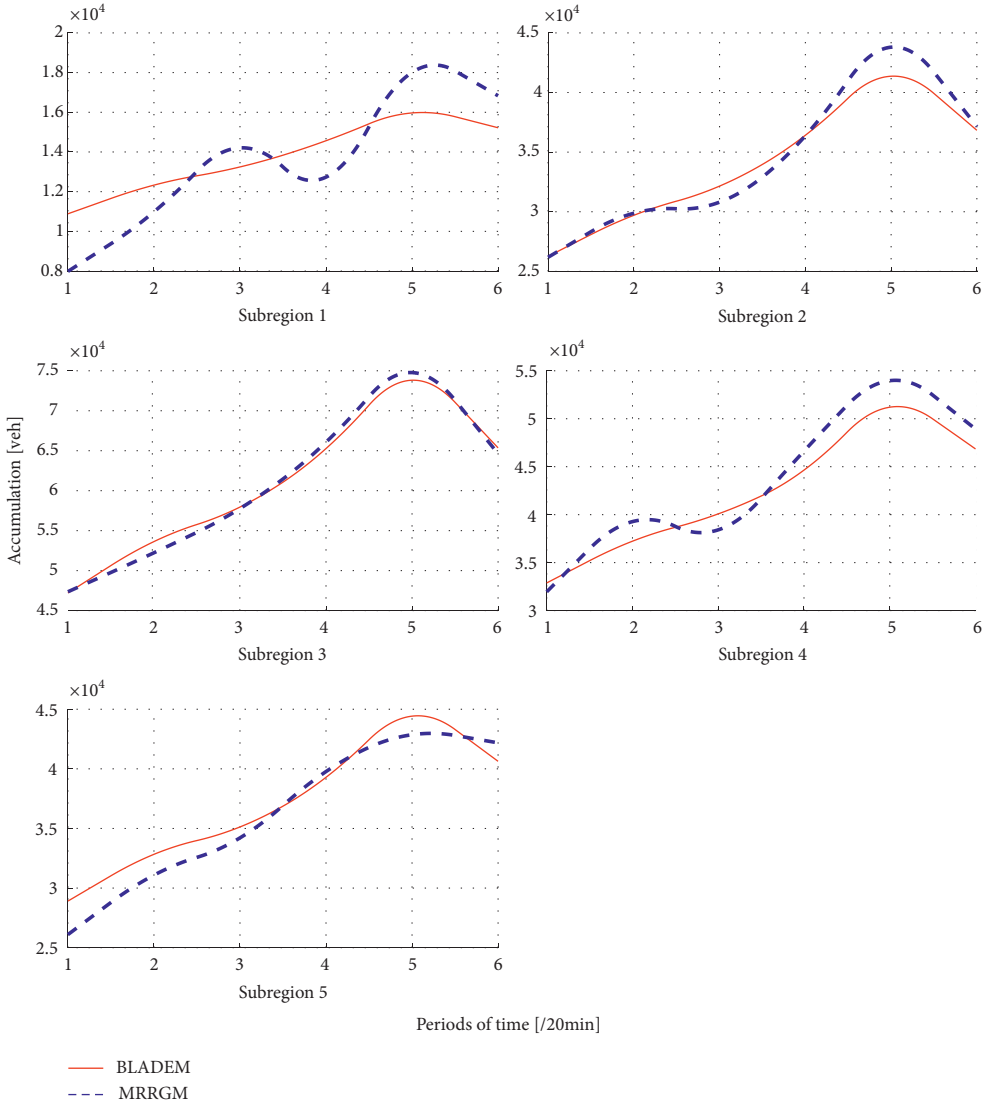


Figure 10: The topology of the network at the sub-region level.

the BLADEM model and the MRRGM model are shown in Figure 11. From the accumulation curve, the similar results of both models were achieved in most sub-region.

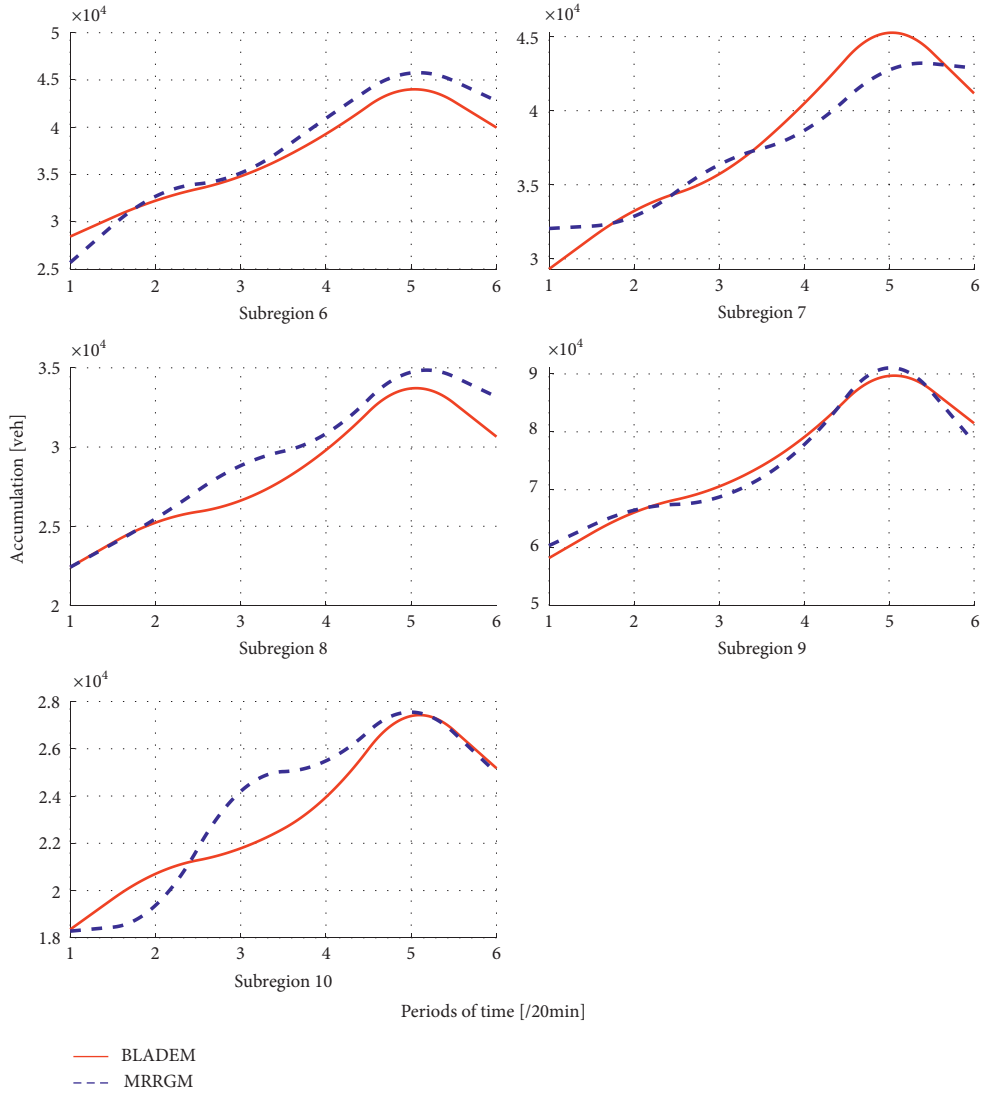
The results of sub-region 3 and sub-region 9 are the most similar, and the difference in sub-region 1 is the largest. The reason for this deviation is the topology of sub-region 1. Except for the boundary nodes of the network, there are no other nodes and links in the sub-region 1. Therefore, when the travel path needs to pass through the sub-region 1, the traveler makes travel route decisions between the boundary nodes of the sub-region 1, and does not need to enter the region. However, the boundary is shared by adjacent sub-regions. Like sub-region 1, if the path decision passing only at the boundary, it is not appropriate to use only MFD information as route decision information at the sub-region level of the MRRGM model. Obviously, the BLADEM model uses links as a medium to connect internal nodes and boundary nodes. Even if it is from one boundary node to another, the BLADEM model is possible to adjust the path reasonably for travelers.

It should be clearly noted that the differences between link level and sub-region level is a normal phenomenon, because there are essential differences in the details of the algorithm. Figure 11 is aim to show the difference between



(a)

Figure 11: Continued.



(b)

Figure 11: The accumulation curve in sub-region. (a) Subregion 1~5. (b) Subregion 6~10.

the results at the segment level and the results at the sub-region level, rather than showing that the two can be similar. When fluctuation is allowed, the sub-regions with some differences are not analyzed, such as sub-regions 4,7,8,10. Therefore, Only the most prominent differences (sub-region 1) are properly analyzed in this paper. As for the reasons for the existence of fluctuations, this paper does not carry out theoretical derivation, which will be carried out in the next stage of research.

To sum up, when the MRRGM model is applied, the sub-region divided within the region requires stable MFD information. Furthermore, the MRRGM model requires the sub-region to have a certain scale (at least one link inside the sub-region). Before the application of the MRRGM model, the work of dividing sub-regions is more complicated and has many constraints. Therefore, the BLADEM model, which takes the links as the research object within the region, is more suitable for the actual traffic environment.

5. Conclusions

The emerging high-resolution vehicle trajectory data from the floating cars or CVs provide invaluable opportunities to measure the real-time MFD information of the network or the region. Such data can be collected through V2C communication to generate dynamics of regions both in real time and offline.

In this paper, the authors developed an innovative approach to get approximating dynamic equilibrium state in multi-region network based on MFD, considering the impact of complex traffic conditions between the regions. In the proposed approach, the network is divided into some regions with the stable MFD, and the DUE model is applied within the region to assign the internal-region O-D demand. For the region level, the length of region path (defines the sequence of regions) is established and the aggregated route choice is analyzed in region-based model. Furthermore, the

complexity of the proposed model is derived. Then, the comparative analysis of the algorithm complexity between the proposed model and the DUE model is given.

The proposed model is evaluated by using real-world vehicle trajectory data from the DiDi platform. Based on the comprehensive comparison, including region accumulation, the experienced travel time of path, the MFD characteristics of each region and the average speed of vehicles in the regions, we verify that the proposed model can obtain the approximate traffic equilibrium analysis result. Moreover, the proposed model shows superior performance on calculation efficiency. Comparing with the DUE model, the improved calculation efficiency is between 21% and 42% under the same calculation and analysis environment. In addition, because the proposed model is more suitable for the actual traffic environment than the MRRGM model, it has the potentials of supporting the management, route guidance, and performance monitoring.

However, this work still has some limitations:

(1) We assume that the network can be divided into some regions with the stable MFD. However, in practical, only a reasonable division of the network can ensure the stability of the region MFD. Namely, the network partitioning method need to be studied before the proposed model applying to the large urban network.

(2) A dynamic route choice mechanism for travelers is incorporated into the model framework. We assume that the travelers preferentially choose the route with the lowest generalized cost. However, in the actual traffic environment, the probability of travelers being completely rational is low, and there will generally be information perception errors. Therefore, the information perception error factors need to be considered in the follow-up research to further optimize the model.

Appendix

A. The Complexity of the DUE Model

The rate of traffic flow of link e at time t denote as $x_e(t)$. In the process of DTA, the traffic flow of link e will be transmitted to the subsequent links on the driving route at all times till the destination. Therefore, the following formula can be derived:

$$\frac{\mathrm{d}x_e(t)}{\mathrm{d}t} \neq 0. \tag{A.1}$$

Therefore, the travel impedance of the link will change with the traffic flow. Then, the shortest path between OD pairs is time-varying. Although DTA is difficult problem due to its complexity, the analysis results of DTA are closer to the real traffic network operating environment.

In the time T, the actual impedance of the link is time-varying according to the OD demand matrix. When the optimal state is reached, for any OD pair at time t, the actual impedance of travelers will be uniform and the smallest of all feasible paths. Furthermore, the impedance of the path that is not chosen by the traveler will not be less than it. Therefore, the traffic flow of the link will not change even if the traveler changes the route. In other words, the traffic flow distribution in the network has reached an equilibrium state, which is called Dynamic User Equilibrium (DUE).

Set the research period as [0, T] and divide it into J segments. The length of each segment is Δt, $T = J \cdot \Delta t$. Significantly, any vehicle entering the link at the beginning of Δt cannot leave the link before the end of Δt. Set $t \in 1, 2, \ldots, J$, then the actual departure time of the traffic flow can be expressed as $t \cdot \Delta t$. The dynamic user equilibrium problem under discrete conditions can be described as:

B. Objective Function:

$$\text{Min: } Z(x_e) = \min \sum_{t=0}^{T} \sum_{o,d} \sum_{e} \int_{0}^{x_e^{o,d}(t)} c_e^{o,d}(t). \tag{B.1}$$

C. Constraint:

$$\begin{aligned}
&x_e^d(t)\\
&x_e^d(t+1) = x_e^d(t) + \int_t^{t+1} \left(u_{\mathrm{in},e}^d(t) - u_{\mathrm{out},e}^d(t)\right), \quad \forall d, e, t,\\
&\sum_{e \in v+} u_{\mathrm{in},e}^d(t) = \sum_{e \in v-} u_{\mathrm{out},e}^d(t) + u_v^d(t), \quad \forall d, e, v, t,\\
&x_e^{o,d}(t=0) = 0, \quad x_e^{o,d}(t) \geq 0, \ \forall o, d, e, t,
\end{aligned} \tag{C.1}$$

Where, $c_e^{o,d}(t)$ is generalized travel cost of link e between the origin node o and the destination node d at time t. Similarly, $x_e^{o,d}(t)$ is the number of vehicles of link e. The number of vehicles of link e which the destination is node d at time t is denoted as $x_e^d(t)$. $u_{\mathrm{in},e}^d(t)$ and $u_{\mathrm{out},e}^d(t)$ respectively indicates the traffic flow entering and leaving link e which the destination is node d at time t. $e \in v+$ indicates the link e that belongs to the upstream of node v. Obviously, $e \in v-$ indicates the link e that belongs to the downstream of node v. In development, $u_v^d(t)$ indicates the traffic flow which its destination is node d is generated by node v at time t.

The DUE model can be solved based on the framework of the Method of Successive Averages (MSA) algorithm. The time-varying traffic demand is allocated to the network through the generalized cost of the route between ODs in different time periods. At the initial stage of each time period, according to the recalculation and analysis of the traffic flow of the link, the generalized cost of the route can be obtained again. Then, the traveler's travel path decision can be made on this basis. The process is repeated until all travelers obtain the lowest travel cost of the current network. The dynamic user equilibrium model solution steps can be summarized as follows:

Step 1. Initialize the network, set $t = 0$, load network related parameters $x_e^{o,d}(0), u_{\text{in},e}^d(0), u_{\text{out},e}^d(0), u_v^d(0)$;

Step 2. Calculate travel costs of the link $\{C_e\}$ in the initial network, and establish adjacency matrix;

Step 3. Find the shortest path according to the Dijkstra algorithm. Set $m = 1$. According to the OD matrix (traffic demand) at the time t, the all-or-nothing assignment is performed to obtain the additional load traffic flow of the initial link, and superimpose it with initialize the network flow to obtain the new network flow $\{y_e^m\}$;

Step 4. Set the iteration step $\theta = (1/m)$. Determine the starting point of a new iteration: $x_e^{m+1} = x_e^m + \theta(y_e^m - x_e^m)$

Step 5. If the following formula true (ε is pre-determined small positive number), stop the calculation and update $x_e^{o,d}(t), u_{\text{in},e}^d(t), u_{\text{out},e}^d(t), u_v^d(t)$ at time t. Else, set $m = m + 1$, then recalculate the travel cost of the link $\{C_e\}^m$ according to the new flow x_e^{m+1}, and go to step 2;

$$\frac{\sqrt{\sum_e \left(x_e^{m+1} - x_e^m\right)^2}}{\sum_e x_e^m} < \varepsilon. \tag{A.4}$$

Step 6. If $t = J$, stop iteration and finish the solution steps. Else, set $t = t + 1$ and go to step 2.

Solving the shortest path from the origin node n_o to other nodes in the transportation network by the dijkstra algorithm can be achieved by the following methods:

S denotes the set of nodes where the shortest path has been found and $T = Q - S$ denotes the set of nodes where the shortest path has not yet been found. Initially, set $S = \{n_o\}$ and $T = \{\text{other nodes}\}$. Calculate the distance between the origin node n_o and the directly connected nodes in T. Noteworthy, the distance between the nodes that are not directly connected is recorded as infinity. Select the node with the smallest distance and put it into S. Then, update the distance between the added node and the remaining nodes in T. If the added node is used as an intermediate node and the distance from the origin node n_o to any node n_i is less than the route without this node, then modify this distance value. Repeat the above steps until all nodes are included in S. The above process searches the smallest element in the distance matrix, and the complexity of the algorithm is $O(V \wedge 2)$. Therefore, the complexity of the algorithm for calculating the shortest path between all OD pairs in the network is $O(V \wedge 3)$ based on the Dijkstra algorithm.

According to the above analysis, the complexity of the DUE model can be obtained. The detailed analysis is as follows:

Step 1. is to initialize the network and defining variables. There is no calculation process in the computer, so the complexity is $O(0)$;

Step 2. calculates the adjacency matrix involving impedance according to the BPR function. The calculating all impedances of links complexity is $O(E)$;

Step 3. obtains the shortest path between all OD pairs based on the Dijkstra algorithm. The complexity is $O(V \wedge 3)$. Then allocate the traffic demand of the corresponding OD pair on all the shortest paths. Since any OD pair has obtained a shortest path, the total of the shortest paths is $V \wedge 2$. The complexity is $O(V \wedge 2)$. Finally, the shortest path is selected arbitrarily. Then allocate the traffic demand and superimpose it with the original flow of the link to obtain the new link flow. Assume that the shortest path contains h edges (links) on average, the complexity is $O(h \cdot V \wedge 2)$;

Step 4. mainly use the iterative update framework of the MSA algorithm to construct the new traffic flow of link. The complexity is $O(E)$;

Step 5. is the process of iterating. The complexity of the algorithm is related to the number of iterations. So far, the algorithm complexity is shown in the following formula:

$$\begin{aligned} O &= m \cdot [O(0) + O(E) + O(V \wedge 3) \\ &\quad + O(V \wedge 2) + O(h \cdot V \wedge 2) + O(E)] \\ &= O(m \cdot [2E + (V + h + 1) \cdot V \wedge 2]). \end{aligned} \tag{A.5}$$

Step 6. is the accumulation of research time periods. There are the total of J periods, so the complexity of the DUE model is:

$$O(\text{DUE}) = O(J \cdot m \cdot [2E + (V + h + 1) \cdot V \wedge 2]). \tag{A.6}$$

To sum up, it can be seen from the expression of the complexity of the DUE that the complexity is closely related to the topological structure of the network. The urban transportation network contains too many nodes, so the conventional dynamic user equilibrium state is difficult to solve. Furthermore, there are many restrictions when providing traffic control decision information based on this.

Data Availability

The high-resolution vehicle trajectory data used to support the findings of this study were supplied by DiDi platform-GAIA Open Dataset under license and so cannot be made freely available. Requests for access to these data should be made to didioutreach@didichuxing.com.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research is partially funded by the Central Public-Interest Scientific Institution Basal Research Fund (2021–9081 & 2020–9018). The authors would like to thank the DiDi for providing the data through their open-data platform.

References

[1] Z. Yao, H. Jiang, Y. Cheng, Y. Jiang, and B. Ran, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 1841–1851, 2022.

[2] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008.

[3] J. W. Godfrey, "The mechanism of a road network," *Traffic Engineering and Control*, vol. 7, no. 11, pp. 323–327, 1969.

[4] R. Herman and I. Prigogine, "A two-fluid approach to town traffic," *Science*, vol. 204, no. 4389, pp. 148–151, 1979.

[5] H. Mahmassani, J. Williams, and R. Herman, "Investigation of network-level traffic flow relationships: some simulation results," *Transportation Research Record*, vol. 971, pp. 121–130, 1984.

[6] C. F. Daganzo, "Urban gridlock: macroscopic modeling and mitigation approaches," *Transportation Research Part B: Methodological*, vol. 41, no. 1, pp. 49–62, 2007.

[7] J. Haddad, "Optimal coupled and decoupled perimeter control in one-region cities," *Control Engineering Practice*, vol. 61, pp. 134–148, 2017.

[8] J. Haddad, "Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics," *Transportation Research Part B: Methodological*, vol. 96, pp. 1–25, 2017b.

[9] R. Zhong, C. Chen, Y. Huang, A. Sumalee, W. Lam, and D. Xu, "Robust perimeter control for two urban regions with macroscopic fundamental diagrams: a control-lyapunov function approach," *Transportation Research Part B: Methodological*, vol. 117, pp. 687–707, 2018.

[10] R. Mohajerpoor, M. Saberi, H. L. Vu, T. M. Garoni, and M. Ramezani, "H∞ robust perimeter flow control in urban networks with partial information feedback," *Transportation Research Part B: Methodological*, vol. 137, pp. 47–73, 2020.

[11] M. Hajiahmadi, J. Haddad, B. De Schutter, and N. Geroliminis, "Optimal hybrid perimeter and switching plans control for urban traffic networks," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 2, pp. 464–478, 2015.

[12] I. I. Sirmatel and N. Geroliminis, "Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1112–1121, 2018.

[13] I. I. Sirmatel and N. Geroliminis, "Nonlinear moving horizon estimation for large-scale urban road networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 12, pp. 4983–4994, 2020.

[14] J. Haddad and Z. Zheng, "Adaptive perimeter control for multi-region accumulation-based models with state delays," *Transportation Research Part B: Methodological*, vol. 137, pp. 133–153, 2020.

[15] I. I. Sirmatel, D. Tsitsokas, A. Kouvelas, and N. Geroliminis, "Modeling, estimation, and control in large-scale urban road networks with remaining travel distance dynamics," *Transportation Research Part C: Emerging Technologies*, vol. 128, pp. 103–157, 2021.

[16] T. Lei, Z. Hou, and Y. Ren, "Data-driven model free adaptive perimeter control for multi-region urban traffic networks with route choice," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 7, pp. 2894–2905, 2020.

[17] Y. Ren, Z. Hou, I. I. Sirmatel, and N. Geroliminis, "Data driven model free adaptive iterative learning perimeter control for large-scale urban road networks," *Transportation Research Part C: Emerging Technologies*, vol. 115, Article ID 102618, 2020.

[18] M. Yildirimoglu and M. Ramezani, "Demand management with limited cooperation among travellers: a doubly dynamic approach," *Transportation Research Part B: Methodological*, vol. 132, pp. 267–284, 2020.

[19] M. Ramezani and M. Nourinejad, "Dynamic modeling and control of taxi services in large-scale urban networks: a macroscopic approach," *Transportation Research Part C: Emerging Technologies*, vol. 94, pp. 203–219, 2018.

[20] M. Yildirimoglu and N. Geroliminis, "Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams," *Transportation Research Part B: Methodological*, vol. 70, pp. 186–200, 2014.

[21] M. Yildirimoglu, M. Ramezani, and N. Geroliminis, "Equilibrium analysis and route guidance in large-scale networks with MFD dynamics," *Transportation Research Part C: Emerging Technologies*, vol. 59, pp. 404–420, 2015.

[22] M. Keyvan-Ekbatani, M. Yildirimoglu, N. Geroliminis, and M. Papageorgiou, "Multiple concentric gating traffic control in large-scale urban networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2141–2154, 2015.

[23] M. Ramezani, J. Haddad, and N. Geroliminis, "Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control," *Transportation Research Part B: Methodological*, vol. 74, pp. 1–19, 2015.

[24] M. Yildirimoglu, I. I. Sirmatel, and N. Geroliminis, "Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance," *Transportation Research Part B: Methodological*, vol. 118, pp. 106–123, 2018.

[25] F. Ordóñez and N. E. Stiermoses, "Robust wardrop equilibrium," *Lecture Notes in Computer Science*, vol. 4465, pp. 247–256, 2007.

[26] W. Ni and M. Cassidy, "City-wide traffic control: modeling impacts of cordon queues," *Transportation Research Part C: Emerging Technologies*, vol. 113, pp. 164–175, 2020.

[27] Q. Guo and X. J. Ban, "Macroscopic fundamental diagram based perimeter control considering dynamic user equilibrium," *Transportation Research Part B: Methodological*, vol. 136, pp. 87–109, 2020.

[28] R. Aghamohammadi and J. A. Laval, "A continuum model for cities based on the macroscopic fundamental diagram: a semi-Lagrangian solution method," *Transportation Research Procedia*, vol. 38, pp. 380–400, 2019.

[29] Y. P. Huang, J. H. Xiong, A. Sumalee et al., "A dynamic user equilibrium model for multi-region macroscopic fundamental diagram systems with time-varying delays," *Transportation Research Part B: Methodological*, vol. 131, pp. 1–25, 2020.

[30] S. F. A. Batista, L. Leclercq, and M. Menéndeza, "Dynamic Traffic Assignment for regional networks with traffic-dependent trip lengths and regional paths," *Transportation Research Part C: Emerging Technologies*, vol. 127, pp. 76–103, Article ID 103076, 2020.

[31] S. F. A. Batista and L. Leclercq, "Regional dynamic traffic assignment framework for macroscopic fundamental diagram

multi-regions models," *Transportation Science*, vol. 53, no. 6, pp. 1563–1590, 2019.

[32] B. V. Cherkassky, A. V. Goldberg, and T. Radzik, "Shortest paths algorithms: theory and experimental evaluation," *Mathematical Programming*, vol. 73, no. 2, pp. 129–174, 1996.

[33] I. Chabini, "Discrete dynamic shortest path problems in transportation applications: complexity and algorithms with optimal run time," *Transportation Research Record Journal of the Transportation Research Board*, vol. 1645, no. 1, pp. 170–175, 1998.

[34] S. Peeta and H. S. Mahmassani, "Multiple user classes real-time traffic assignment for online operations: a rolling horizon solution framework," *Transportation Research Part C: Emerging Technologies*, vol. 3, no. 2, pp. 83–98, 1995.

[35] X. Luo, Bo Liu, P. J. Jin, Y. Cao, and W. Hu, "Arterial traffic flow estimation based on vehicle-to-cloud vehicle trajectory data considering multi-intersection interaction and coordination," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2673, no. 6, pp. 68–83, 2019.

[36] P. Krishnakumari, H. V. Lint, T. Djukic, and O. Cats, "A data driven method for OD matrix estimation," *Transportation Research Procedia*, vol. 38, pp. 139–159, 2019.

[37] M. Kuwahara and T. Akamatsu, "Decomposition of the reactive dynamic assignments with queues for a many-to-many origin-destination pattern," *Transportation Research Part B: Methodological*, vol. 31, no. 1, pp. 1–10, 1997.

[38] C. Q. Shao and X. M. Liu, "Estimation of saturation flow rates at signalized intersections," *Discrete Dynamics in Nature and Society*, vol. 1802, pp. 1–9, 2012.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 3866042, 15 pages
https://doi.org/10.1155/2022/3866042

Research Article

Characteristics Analysis and Equilibrium Optimization of Mixed Traffic Flow considering Connected Automated and Human-Driven Vehicles

Zhaoming Zhou,[1,2,3] **Jianbo Yuan**,[1] **Shengmin Zhou**,[4] **Qiong Long**,[2] **Jianrong Cai**,[2] **and Lei Zhang**[2]

[1]*School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha, China*
[2]*College of Civil Engineering, Hunan City University, Yiyang, China*
[3]*Engineering Research Center of Catastrophic Prophylaxis and Treatment of Road and Traffic Safety of Ministry of Education, Changsha University of Science and Technology, Changsha, China*
[4]*Xiangtan Technology Research Center of Urban Planning Information, Xiangtan, China*

Correspondence should be addressed to Jianbo Yuan; yuanjb01@163.com

Received 18 April 2022; Revised 27 June 2022; Accepted 6 July 2022; Published 12 August 2022

Academic Editor: Zhihong Yao

Considering the impact of informatization condition, vehicles on the road network are divided into connected automated vehicles (CAVs) and human-driven vehicles (HDVs), which follow the principle of system optimization and stochastic user equilibrium, respectively. Taking the road network reserve capacity maximization model under the condition of road capacity constraint as the upper-level programming and the traffic assignment model under heterogeneous flow environment as the lower level programming, then a bilevel programming model is constructed. Among them, the nonuniform demand growth multiplier is adopted for each OD pair to reflect the inconsistency of traffic demand structure growth, and the calculation of link capacity is related to the market penetration of CAVs. The incremental method, method of successive averages, and simulated annealing algorithm are used to solve the model, and the effects of different market penetration on road network capacity, travel time, and saturation are analyzed through a numerical example. The relevant data under different weights are normalized and the optimal deployment scheme of CAVs and HDVs in different periods is obtained by comprehensive evaluation. Meanwhile, the mixed equilibrium flow state is explored under the premise of given market penetration to verify the feasibility of the model and algorithm.

1. Introduction

In the background of increasing traffic pressure and continuous progress of science and technology, the intelligent transportation system emerges at the historic moment and has achieved unprecedented progress in the field of transportation. High-end chips, 5G communication, and other new generations of information technology have been booming. The prospect of unmanned driving is so promising that it is bound to be an upcoming travel necessity. In recent years, vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) technologies have matured [1] and connected automated vehicle (CAV) technology has witnessed unprecedented development. There is an emerging trend of CAVs related studies. Milakis et al. [2] did a comprehensive review to explore the potential effects of automated driving on policy and society. Noruzoliaee et al. [3] determined the optimal purchase price of autonomous vehicles from the perspective of the manufacturer, considering vehicle type and route choice behavior.

Compared with human-driven vehicles (HDVs), CAVs can obtain extensive benefits in traffic capacity [4], travel

safety [5], energy consumption [6], and exhaust emission. However, before HDVs are completely eliminated, the traffic flow in the road network will still be heterogeneous traffic flow composed of HDVs and CAVs. It is necessary to analyze how to make traffic flow reach equilibrium under the mutual influence. The concept of mixed traffic equilibrium has been mentioned in earlier literature. Haurie and Marcotte [7] illustrated the coexistence of competitive and cooperative traffic flows. Wang et al. [8] formulated two bilevel programs to study the network reserve capacity with the SUE principle and resolved it via sensitivity analysis. Ryu et al. [9] launched traffic assignment under capacity constraints in combination with stochastic utility theory and settled the model by combining the equilibrium iteration method and column generation method. Huang et al. [10] established a stochastic assignment model for travelers' path selection on the basis of the disequilibrium theory, which was verified by model simulation. Bagloee et al. [11] established a mixed equilibrium model in which connected vehicles and other vehicles adhere to the system optimal principle and the user equilibrium principle, respectively, while the practical features, for instance, road capacity, elastic demand, and travel time are explicitly considered.

The concept of mixed traffic equilibrium is not new in the literature, scholars have conducted extensive and in-depth research on mixed traffic flow, and different types of vehicles follow corresponding principles to select paths and achieve equilibrium under mutual influence. However, most studies only consider two types of vehicles (CAVs and HDVs). In fact, for the existing HDVs, travelers can obtain comprehensive, efficient, and real-time traffic information via advanced traveler information systems (ATIS) devices and assist travelers in choosing travel time and route. With the emergence of ATIS, scholars have researched its impact on traffic flow. Vuren and Watling [12] upheld that the system optimum (SO) principle can be realized through ATIS route guidance. Yang and Meng [13] proposed a modified logistic growth model to investigate the adoption rate of ATIS. Yin and Yang [14] categorized travelers according to with or without ATIS, selected paths in conformity with stochastic user equilibrium (SUE), and analyzed market penetration rate and compliance rate. Bifulco et al. [15] constructed an assignment model for ATIS to analyze the stability of traffic equilibrium. Dell'Orco et al. [16] proposed the dynamic model of driver's information compliance to represent the dynamic selection behavior of the driver in ATIS. The above studies are only for HDVs, the adoption rate of ATIS devices is analyzed, and the traffic assignment model is constructed.

Under the background of the intelligent network, a well-designed planning and operation strategy for CAVs based on estimated network flows is of crucial importance during the coexistence period of CAVs and HDVs, so as to realize maximum operation efficiency. Relevant research is mainly focused on topics such as road network capacity, market penetration, and lane management. First, traffic congestion, primarily caused by the mismatch between traffic demand and supply [17], has become a thorny problem in metropolises around the world [18]. Traffic congestion can be alleviated from the level of traffic supply. As one of the important basic data of traffic supply, road network capacity is a key index of traffic planning and management. It could be regarded as the maximum sum of throughput of all OD pairs under the constraint of link capacities [19]. Modeling and analysis of road network capacity can predict the maximum number of trips that the road network can be satisfied and evaluate the network performance of the road system. Based on this, scholars regard road network capacity as a key indicator to measure the performance of traffic system under CAV environment. Levin and Boyles [20] presented a multiclass traffic assignment model by considering the influence of the proportion of CAVs on road capacity. Van den Berg and Verhoef [21] adopt a dynamic equilibrium model to investigate the effect of CAVs on congestion and confirm that CAVs can improve road capacity and reduce bottleneck externality. Noruzoliaee et al. [3] presented that CAVs can reduce the value of travel time by making in-vehicle time more efficiently and increase road capacity through smaller headway. Second, despite the rapid technological development, there is still a long time to go before the CAVs dominate the full market. In this context, considering the market penetration, heterogeneous traffic flow consisting of CAVs and HDVs is more plausible for the foreseeable future. As a critical factor affecting the road network capacity in the mixed traffic, many scholars have conducted researches based on the market penetration. Lavasani et al. [22] constructed a market penetration model for CAVs based on existing technologies and conducted sensitivity analysis from two aspects of scale and price. Li et al. [23] deliberated the nonlinear change in road capacity with increasing penetration rate of autonomous vehicles to investigate the impacts of mixed flow conditions. Chen et al. [24] proposed formulations for the capacity of mixed traffic in equilibrium state that takes autonomous vehicle market penetration rate into account. Furthermore, several researches also pointed out that through lane management, the operation efficiency of CAV can be better improved by transforming traditional lanes into CAV dedicated lanes. Ghiasi et al. [25] developed a compact lane management model to efficiently determine the optimal number of CAV exclusive lanes to maximize the HDVs and CAVs mixed flow. Chen et al. [26] presented a framework for the optimal design of CAV zones to adapt to and further promote the deployment of CAV technology. Amirgholy et al. [27] designed the optimal lane management strategy for the heterogeneous traffic condition to reduce the experienced delay.

To sum up, scholars have explored the mixed traffic flow of CAVs and HDVs, followed the principles of SO, UE, or SUE, and discussed the optimization design under the mixed equilibrium state. However, few scholars further subdivided HDVs according to the presence of ATIS devices. Meanwhile, most of the literature shows that the application of CAVs will affect road capacity but usually only analyze the increasing effect of CAVs on lane capacity, without considering the impact on the entire road network capacity. Some scholars have revealed the change process of road network capacity, but generally all OD pairs adopt the

uniform demand growth multiplier, and the impact of each OD pair with different demand growth multiplier on the road network reserve capacity is rarely studied. Therefore, it is necessary to fully research the road network capacity under different traffic demands and market penetration rates.

On the basis of the above work, the vehicles on the road network in this paper are divided into CAVs and HDVs, which assumed that the two types of vehicles follow the principles of SO and SUE, respectively. Among them, HDVs are further subdivided considering whether the ATIS device is available, where HDVs-I and HDVs-II represent HDVs with ATIS devices and HDVs without ATIS devices, respectively. A bilevel programming model of mixed traffic flow with nonuniform demand growth multiplier is constructed to maximize the reserve capacity of road network under the condition of the link capacity constraints. The Nguyen-Dupuis network is adopted to explore the influence of diverse market penetration on road network capacity, travel time, and saturation. The optimal deployment scheme of CAVs and HDVs is formulated by analyzing the characteristics of road network, and the mixed equilibrium state of road network is verified.

The remainder of this paper is organized as follows. Section 2 formulates a bilevel model to maximize the travel demand considering link capacity with equilibrium constraints. Section 3 describes the solution procedures for the bilevel programming model. In Section 4, a numerical example is used to discuss the influence of market penetration on transportation equilibrium assignment and illustrate the effectiveness of the proposed model. Section 5 summarizes the finding in this paper and points out the potential future work.

2. The Bilevel Planning Model

2.1. Notation. In Table 1, we summarize the notations commonly used in this paper, and other notations are explained when they are used.

2.2. The Upper Level Model. The upper-level optimization objective is to maximize the traffic travel volume of the road network, different OD pairs adopt nonuniform demand growth multiplier, and the constraint condition is the link capacity, predicting the maximum demand that the road network can accommodate.

$$\begin{aligned} &\max \sum_{w\in W} \mu_w q_w, \\ &\text{s.t. } x_a(\boldsymbol{\mu}\mathbf{q}) \le C_a, \quad \forall a \in A. \end{aligned} \tag{1}$$

2.3. The Lower Level Model

2.3.1. Traffic Assignment Model of Connected Automated Vehicles. The planning model of CAVs is constructed according to the SO assignment model, and the system optimization can be achieved when the marginal travel time function is adopted to carry out the user optimal flow assignment. The specific planning model is as follows:

$$\min Z_1(X_1) = \sum_{a\in A} x_{a,1} t_a\left(x_{a,1}\right) = \sum_{a\in A} \int_0^{x_{a,1}} \bar{t}_a(\omega)\mathrm{d}\omega, \tag{2}$$

$$\text{s. t. } q_{w,1} = \alpha\mu_w q_w, \tag{3}$$

$$\sum_k f_k^{w,1} = q_{w,1}, \quad \forall w \in W, \tag{4}$$

$$x_{a,1} = \sum_w \sum_k f_k^{w,1} \delta_{a,k}^w, \quad \forall a \in A, \tag{5}$$

$$f_k^{w,1} \ge 0, \quad \forall w \in W, k \in K. \tag{6}$$

2.3.2. Traffic Assignment Model of Human-Driven Vehicles. The planning model of HDVs is constructed according to the SUE assignment model, and considering whether the ATIS device is available, the corresponding information quality level is different. The specific planning model is as follows:

$$\begin{aligned} \min Z_2(X_2) = &\sum_{a\in A} \int_0^{x_{a,2}} t_a(\omega)\mathrm{d}\omega + \frac{1}{\theta_1} \sum_w \sum_k f_k^{w,3} \ln f_k^{w,3} \\ &+ \frac{1}{\theta_2} \sum_w \sum_k f_k^{w,4} \ln f_k^{w,4}, \end{aligned} \tag{7}$$

$$\begin{aligned} \text{s.t. } q_{w,2} &= (1-\alpha)\mu_w q_w, \\ q_{w,3} &= \beta q_{w,2}, \\ q_{w,4} &= (1-\beta) q_{w,2}, \end{aligned} \tag{8}$$

$$\begin{aligned} \sum_k f_k^{w,2} &= q_{w,2}, \\ \sum_k f_k^{w,3} &= q_{w,3}, \\ \sum_k f_k^{w,4} &= q_{w,4}, \quad \forall w \in W, \end{aligned} \tag{9}$$

$$x_{a,2} = \sum_w \sum_k f_k^{w,2} \delta_{a,k}^w, \quad \forall a \in A. \tag{10}$$

$$\begin{aligned} f_k^{w,2} &\ge 0, \\ f_k^{w,3} &\ge 0, \\ f_k^{w,4} &\ge 0, \\ \forall w &\in W, \\ k &\in K, \end{aligned} \tag{11}$$

where θ_1 represents the quality level of travel information received by drivers with ATIS device, and θ_2 represents the familiarity of drivers without ATIS device to road networks. In general, θ_1 is greater than θ_2.

Table 1: Notation summarization.

Notations	Definitions
Sets	
N	Set of nodes
A	Set of links
K	Set of paths
T	Set of vehicle types: type 1 denotes CAVs, type 2 denotes HDVs, type 3 denotes HDVs-I, type 4 denotes HDVs-II
W	Set of origin-destination (OD) pairs
R	Set of origin nodes and $R \subset N$
S	Set of destination nodes and $S \subset N$
Parameters	
a	Index of link, $a \in A$
k	Index of path, $k \in K$
τ	Index of vehicle types, $\tau \in T$
w	Index of OD pairs, $w \in W$
$\delta_{a,k}^{w}$	Link–path incidence that equals 1 if link a belongs to path k between OD pair $w \in W$ and 0 otherwise.
C_a	Capacity of link a
L_a	Length of link a
α	Market penetration of CAVs among all vehicles
β	Market penetration of ATIS vehicles among human-driven vehicles
Variables	
μ_w	Multiplier of travel demand growth between OD pair $w \in W$
q_w	Basic travel demand between OD pair $w \in W$
$q_{w,\tau}$	Basic travel demand of vehicle types $\tau \in T$ between OD pair $w \in W$
x_a	Traffic flow on link a
$x_{a,\tau}$	Traffic flow of vehicle types $\tau \in T$ on link $a \in A$
f_k^w	Traffic flow on path k between OD pair $w \in W$
$f_k^{w,\tau}$	Traffic flow of vehicle types $\tau \in T$ on path k between OD pair $w \in W$
$t_a(x_a)$	Travel time on link a
c_k^w	Actual travel time on path k between OD pair $w \in W$, i.e. $c_k^w = \sum_{a \in A} t_a(x_a) \delta_{a,k}^w$
$\overline{t}_a(x_a)$	Marginal travel time on link a

2.3.3. Traffic Assignment Model of Mixed Equilibrium. CAVs and HDVs are mixed in the road network according to different path selection principles. Among them, the former follows the principle of system optimization, and the flow of HDVs is taken as the background flow, while the latter follows the principle of stochastic user equilibrium and the flow of HDVs is regarded as the background flow. Considering the influence of information conditions, the SO-SUE mixed equilibrium traffic assignment model is constructed as follows:

$$\begin{aligned} \min Z(X) = & \sum_{a \in A} \int_0^{x_{a,1}} \overline{t}_a\left(x_{a,2} + \omega\right) \mathrm{d}\omega \\ & + \sum_{a \in A} \int_0^{x_{a,2}} t_a\left(x_{a,1} + \omega\right) \mathrm{d}\omega \\ & + \frac{1}{\theta_1} \sum_w \sum_k f_k^{w,3} \ln f_k^{w,3} \\ & + \frac{1}{\theta_2} \sum_w \sum_k f_k^{w,4} \ln f_k^{w,4}, \end{aligned} \tag{12}$$

s.t., (3)–(6) and (8)–(11).

3. Solution Algorithm of the Programming Model

Different demand growth multipliers are adopted in each OD pair and the upper model is a nonconvex function. It is difficult to obtain the optimal solution by conventional numerical optimization methods. The incremental method, method of successive averages (MSA), and simulated annealing algorithm (SAA) are adopted to solve the established bilevel programming model. The concept of SAA is based on the principle of solid annealing. Under the action of high temperature, the particles move relatively freely and disorderly, and the probability of accepting the inferior solution is large. With the continuous decrease of temperature, an ordered state is gradually formed, the probability of receiving a poor solution is reduced, and the global optimal solution of the objective function is randomly searched. The temperature is an important factor affecting the global search performance of SAA. The initial temperature should be set sufficiently large to ensure a wide search range of the algorithm, and the minimum temperature usually takes a small decimal. The temperature change rate is mainly used to control the annealing speed. In practical applications, it is generally taken as 0.95~0.99. In addition, the Metropolis criterion accepts the nonoptimal solution with a certain probability, which is the key for SAA to jump out of the local

optimality and converge to the global optimality. The specific steps of the algorithm are as follows.

Step 1. Set the initial temperature T_0, minimum temperature $T_{\min}$, temperature change rate ΔT, let the current temperature $T = T_0$, take the initial travel demand growth multiplier as the initial solution **S**, and take the total initial travel demand as the objective function $F(\mathbf{S})$, the vector of nonuniform demand growth multiplier step is set as $\Delta\mu$, the number of iterations $k = 1$.

Step 2. Randomly select the value in vector $\Delta\mu$ as the demand growth multiplier step for each OD pair $\Delta\mu_w$, then generate a new solution S'.

Step 3. According to the known road network information, the basic travel demand of OD pair $w \in W$ is given as q_w, the nonuniform demand growth multiplier step is set as $\Delta\mu_w$, and initialize μ_w, i.e. $\mu_w(1) = \mu_0$, $n = 1$.

Step 4. In accordance with the given $\mu_w(n)$, the lower mixed equilibrium traffic assignment model is solved, and the link flow $x_a(n)$ is obtained.

(4.1) The travel flow is divided according to the market penetration of different vehicles, and the travel flow of CAVs, HDVs, HDVs-I, and HDVs-II can be obtained as $q_{w,1} = \alpha\mu_w q_w$, $q_{w,2} = (1-\alpha)\mu_w q_w$, $q_{w,3} = \beta q_{w,2} = (1-\alpha)\beta\mu_w q_w$, $q_{w,4} = (1-\beta)q_{w,2} = (1-\alpha)(1-\beta)\mu_w q_w$, $\forall w \in W$, respectively.

(4.2) Let the flow on each link be zero, and the free flow travel time of each link is obtained as $t_a^{(0)} = t_a(0)$. The all or nothing traffic flow assignment is carried out on $q_{w,1}$, and the stochastic traffic flow assignment is carried out on $q_{w,3}$ and $q_{w,4}$. Then the flow of CAVs, HDVs-I and HDVs-II can be, respectively, obtained as $x_{a,1}^{(m)}$, $x_{a,3}^{(m)}$ and $x_{a,4}^{(m)}$, respectively, so as to obtain the link flow $x_a^{(m)} = x_{a,1}^{(m)} + x_{a,2}^{(m)} = x_{a,1}^{(m)} + x_{a,3}^{(m)} + x_{a,4}^{(m)}$, and let the iteration number $m = 1$.

(4.3) The new actual travel time of the link is calculated according to the link flow, i.e. $t_a^{(m)} = t_a(x_a^{(m)})$, $\forall a \in A$.

(4.4) According to the marginal cost function and in combination with the link travel time calculated in step 2.3, the marginal travel time of the link is calculated as $\overline{t}_a^{(m)} = t_a^{(m)} + x_{a,1}^{(m)}(\mathrm{d}t_a^{(m)}/\mathrm{d}x_{a,1}^{(m)})$. The all or nothing traffic flow assignment is carried out on $q_{w,1}$, the additional traffic flow of each link $y_{a,1}^{(m)}$ is obtained, and then the search direction $d_{a,1}^{(m)} = y_{a,1}^{(m)} - x_{a,1}^{(m)}$ is determined. The weighted average method is adopted to update the link traffic flow, i.e. $x_{a,1}^{(m+1)} = x_{a,1}^{(m)} + 1/m(y_{a,1}^{(m)} - x_{a,1}^{(m)})$.

(4.5) The stochastic traffic flow assignment is carried out on $q_{w,3}$ and $q_{w,4}$ in accordance with the link travel time calculated in step 2.3, the additional traffic flow of each link $y_{a,3}^{(m)}$ and $y_{a,4}^{(m)}$ are obtained. The search direction is determined as $d_{a,3}^{(m)} = y_{a,3}^{(m)} - x_{a,3}^{(m)}$ and $d_{a,4}^{(m)} = y_{a,4}^{(m)} - x_{a,4}^{(m)}$, respectively. The weighted average method is adopted to update the link traffic flow, i.e., $x_{a,3}^{(m+1)} = x_{a,3}^{(m)} + 1/m(y_{a,3}^{(m)} - x_{a,3}^{(m)})$ and $x_{a,4}^{(m+1)} = x_{a,4}^{(m)} + 1/m(y_{a,4}^{(m)} - x_{a,4}^{(m)})$.

(4.6) The total link flow is calculated based on the link flow of CAVs and HDVs, i.e., $x_a^{(m+1)} = x_{a,1}^{(m+1)} + x_{a,2}^{(m+1)} = x_{a,1}^{(m+1)} + x_{a,3}^{(m+1)} + x_{a,4}^{(m+1)}$.

(4.7) The convergence test is performed by the predetermined iteration accuracy ε, and if $\sqrt{\sum_a (x_a^{(m+1)} - x_a^{(m)})^2 / \sum_a x_a^{(m)}} \le \varepsilon$, then terminate the circulation; Otherwise, let $m = m + 1$ and return to step 2.3 and continue the iterative calculation.

Step 5. If $\forall a \in A$, $x_a(\mu_w(n)) \le C_a$, then let $\mu_w(n+1) = \mu_w(n) + \Delta\mu_w$, $n = n + 1$, and return to step 4. Otherwise, calculate the maximum reserve capacity of road network as the objective function $F(S') = \sum_{w \in W}[\mu_w(n) - \Delta\mu_w]q_w$.

Step 6. Calculate the difference in objective function $\Delta F = F(S') - F(\mathbf{S})$, and judge whether to accept the new solution S' based on the metropolis criterion. If $\Delta F > 0$, S' is accepted as the new current solution **S**; otherwise, the acceptance probability of the new solution is calculated, i.e. $P = \exp(-\Delta F/T)$, and a random number R uniformly distributed in the interval $(0, 1)$ is randomly generated, S' is accepted as the new current solution **S** while $P > R$.

Step 7. If the termination condition is satisfied, the current solution **S** is output, which is the optimal travel demand growth multiplier for each OD pair, and the algorithm is terminated; otherwise, the annealing process is performed according to the attenuation function, i.e. $T = T \times \Delta T$, return to step 2 after the temperature is reduced, and let $k = k + 1$. The termination condition is usually set as the temperature T reaches the minimum temperature $T_{\min}$ or several consecutive new solutions are not accepted.

4. Numerical Example Analysis

4.1. Basic Information of Road Network. The test road network includes 13 nodes, 19 links, and 4 OD pairs, as shown in Figure 1. The OD initial travel demand is given in Table 2. Assuming initial temperature $T_0 = 1000$, minimum temperature $T_{\min} = 1 \times 10^{-3}$, temperature change rate $\Delta T = 0.98$, and the termination condition is set as the temperature T reaches the minimum temperature $T_{\min}$ or fifty consecutive identical solutions are accepted. The vector of nonuniform demand growth multiplier step is set as $\Delta\mu = [0.01, 0.02, 0.03, 0.04]$, and the value can be randomly selected from this interval vector as the increasing factor of an OD pair. The initial traffic demand growth multiplier μ_0 is assumed to be 1. The iteration accuracy ε is assumed to be 1×10^{-4}. The multiplier θ_1 and θ_2 are assumed to be 10 and

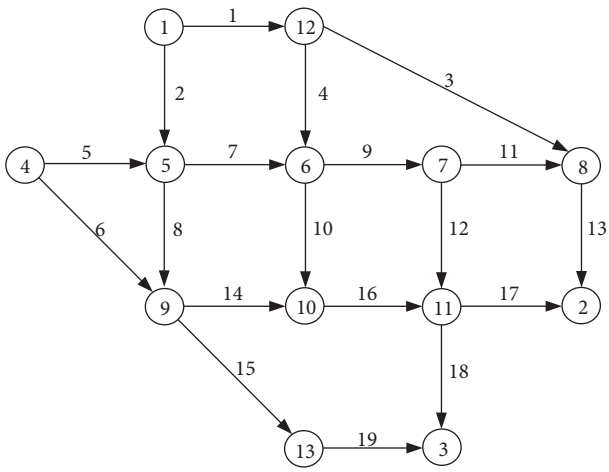


Figure 1: Test road network.

Table 2: Initial travel demand of OD pair.

OD pair number	Origin number	Destination number	Demand (veh/h)
1	1	2	400
2	1	3	800
3	4	2	600
4	4	3	200

0.5, respectively. The Bureau of Public Roads (BPR) function is applied as the link impedance function.

$$t_a(x_a) = t_a^0\left[1 + 0.15\left(\frac{x_{a,1} + x_{a,2}}{C_a}\right)^4\right], \quad (13)$$

where t_a^0 is the free flow travel time on link a, C_a is the capacity of the link a. In this study, C_a is a function of the proportion of CAVs, which can be formulated as follows:

$$C_a = \frac{1}{\left(x_{a,1}/x_a\right)\cdot\left(1/C_{a,1}\right) + \left(1 - \left(x_{a,1}/x_a\right)\right)\cdot\left(1/C_{a,2}\right)}, \quad (14)$$

where $C_{a,1}$ and $C_{a,2}$ are the link capacity for pure CAV flow and pure HDV flow, respectively.

Compared with HDVs, CAVs have shorter reaction times and can respond faster when encountering problems during driving. CAVs can follow the vehicle ahead more closely than HDVs, leading to increased link capacity. Therefore, it is assumed that the link capacity of pure CAV flow is twice than that of pure HDV flow. The characteristic parameters of each link, including free travel time, capacity, and length, are shown in Table 3. The paths and corresponding links contained in each OD pair are shown in Table 4.

4.2. Market Penetration Analysis. The flow of different types of vehicles is determined according to the market penetration of CAVs among all vehicles and ATIS vehicles among HDVs. The impact of different market penetration on road network capacity, travel time, and saturation is analyzed.

4.2.1. Impact on Road Network Capacity. Table 5 demonstrates the road network reserve capacity of different market penetration rates under a nonuniform mode. Note that the market penetration of HDVs-I here is based on human-driven vehicles, not all vehicles in the road network. It can be concluded on the basis of the relevant data that:

(1) The reserve capacity of the road network and the market penetration of CAVs change in the same direction. When the market penetration of CAVs $\alpha < 0.3$, network reserve capacity shows a trend of slow growth before it booms and then shows a trend of rapid growth. When all the vehicles on the road network are CAVs, i.e., $\alpha = 1$, the road network reserve capacity is maximum, up to 9436 veh/h. The car-following capacity of CAVs is better than HDVs, such as the safety distance and headway, thereby significantly increasing the reserve capacity of the road network when α is large enough.

(2) There is basically a reverse change relationship between the road network reserve capacity and the market penetration of HDVs-I, but the change is small. There is a slight increase when the market penetration of CAVs and HDVs-I is very small. When the market penetration of CAVs is large enough, it is almost unaffected by HDVs-I. The road network reserve capacity is a minimum of 2682 veh/h when all vehicles on the road network are HDVs-I.

(3) Compared with HDVs, the market penetration of CAVs has a greater impact on the road network reserve capacity, indicating that the market penetration of CAVs is a more sensitive factor. Especially when the proportion of CAVs becomes larger, the road network reserve capacity is almost unaffected by the market penetration of HDVs-I.

Table 6 and Figure 2 show the travel demand growth multiplier of each OD pair and the distribution of road network capacity under different market penetrations. Each OD pair adopts different travel demand growth multiplier steps to maximize road network reserve capacity under the constraint of link capacity. It can be seen that affected by the market penetration of different types of vehicles, the change trend of the travel demand multiplier for each OD pair is different. However, on the whole, when all vehicles in the road network are human-driven vehicles, with the increase of HDVs-I market penetration, the road network capacity first increases and then decreases, but the change is small. The emergence of CAVs has improved the road network capacity, especially when the market penetration of CAVs reaches a certain value. The improvement effect is more obvious.

4.2.2. The Impact on Travel Time. Considering that the total travel time increases with the increase of road network flow, the average travel time is adopted as the evaluation index. The change of average travel time under different market penetration is shown in Table 7. According to the relevant data, it can be obtained that

Table 3: Characteristic parameters of road network link.

Links	Free travel time (min)	Capacity of HDVs (veh/h)	Capacity of CAVs (veh/h)	Link length (km)
1	8	1400	2800	2.3
2	7	1800	3600	2.5
3	15	1400	2800	4.9
4	7	1000	2000	1.8
5	9	1400	2800	2.6
6	14	1800	3600	4.5
7	5	1600	3200	1.7
8	9	1200	2400	2.4
9	5	1800	3600	1.9
10	13	1000	2000	2.7
11	5	1000	2000	1
12	9	1200	2400	2.3
13	10	1400	2800	2.8
14	10	1400	2800	3
15	9	1200	2400	2
16	8	1400	2800	2.3
17	9	1400	2800	2.6
18	8	1400	2800	2.5
19	11	1400	2800	3.2

Table 4: The index of road network path.

OD pair number	Path number	Included links
1	1	1, 3, 13
	2	1, 4, 9, 11, 13
	3	1, 4, 9, 12, 17
	4	1, 4, 10, 16, 17
	5	2, 7, 9, 11, 13
	6	2, 7, 9, 12, 17
	7	2, 7, 10, 16, 17
	8	2, 8, 14, 16, 17
2	9	1, 4, 9, 12, 18
	10	1, 4, 10, 16, 18
	11	2, 7, 9, 12, 18
	12	2, 7, 10, 16, 18
	13	2, 8, 14, 16, 18
	14	2, 8, 15, 19
3	15	5, 7, 9, 11, 13
	16	5, 7, 9, 12, 17
	17	5, 7, 10, 16, 17
	18	5, 8, 14, 16, 17
	19	6, 14, 16, 17
4	20	5, 7, 9, 12, 18
	21	5, 7, 10, 16, 18
	22	5, 8, 14, 16, 18
	23	5, 8 ,15, 19
	24	6, 14, 16, 18
	25	6, 15, 19

(1) With the increase of the market penetration of CAVs, more and more traffic flow is distributed on the link, the driving time of the link is prolonged, and the average travel time will rise, showing a trend from slow to fast and then to stable.

(2) With the increase of the market penetration of HDVs-I, limited by the link capacity, the traffic flow on the link will gradually decrease, the link travel time will shorten, and the average travel time will decline, but with a small overall decline range.

(3) When all vehicles on the road network are HDVs-I, i.e., $\alpha = 0\beta = 1$, the average travel time at this time is a minimum of 35.54 min; when the market penetration of CAVs and HDVs-I is 0.9 and 0.2, respectively, the average travel time at this moment is a maximum of 38.78 min.

4.2.3. The Impact on Saturation. Saturation is an important index to evaluate the service level of the road network and reflects the matching of traffic supply and demand. Based on various market penetration, the flow distribution and corresponding saturation of each link are obtained, and the average saturation ratio $\overline{S}$ is determined. The calculation formula is as follows:

$$\overline{S} = \frac{\sum_{a\in A} S_a L_a}{\sum_{a\in A} L_a} = \frac{\sum_{a\in A} x_a L_a}{\sum_{a\in A} C_a L_a}. \tag{15}$$

Table 8 shows the changes of average saturation at different market penetration. It can be seen that

(1) When all vehicles are HDVs-I, and the average saturation is a minimum of 0.401 at this time. The average saturation is a maximum of 0.734 when all vehicles are CAVs.

(2) On the one hand, the average saturation increases with the market penetration of CAVs and presents a growth from slow to fast and then gentle. On the other hand, the average saturation presents a small decrease with the increase in the market penetration of HDVs-I.

(3) The value of average saturation should be moderate. When the average saturation is small, it indicates that the road congestion degree is low, the road traffic is smooth, and the driving comfort degree is high, but

Table 5: Road network reserve capacity under different market penetration (veh/h).

β	α										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	3156	3178	3196	3218	3120	3064	2986	2900	2840	2760	2682
0.1	3368	3406	3408	3400	3344	3260	3188	3120	3058	2992	2930
0.2	3672	3710	3674	3632	3568	3496	3426	3380	3334	3248	3196
0.3	4014	3980	3980	3914	3848	3794	3728	3656	3584	3536	3458
0.4	4618	4622	4484	4450	4300	4160	4070	4016	3976	3924	3872
0.5	5220	5230	5240	5348	5186	5168	5220	5120	4970	4964	4340
0.6	6000	6074	6032	5990	5906	5840	5828	5808	5774	5712	5720
0.7	6740	6740	6680	6620	6680	6662	6588	6514	6482	6424	6488
0.8	7624	7550	7500	7476	7502	7500	7434	7390	7328	7328	7254
0.9	8480	8480	8468	8468	8468	8468	8468	8468	8468	8468	8468
1.0	9436	9436	9436	9436	9436	9436	9436	9436	9436	9436	9436

Table 6: Distribution of demand growth multiplier under different market penetration.

Vehicle type	OD pair number	The market penetration									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
CAVs and HDVs-II	1	2.44	2.76	3.12	3.31	3.80	4.00	4.16	4.04	4.60	3.86
	2	1.36	1.44	1.53	1.77	2.40	3.00	3.37	4.04	4.60	5.29
	3	1.36	1.44	1.53	1.77	1.70	2.00	2.58	3.28	3.70	5.29
	4	2.44	2.76	3.12	4.08	3.80	4.00	4.16	4.04	3.70	2.43
CAVs and HDVs-I	1	1.44	1.60	2.04	2.08	2.44	2.56	2.86	3.04	3.84	3.94
	2	1.44	1.60	1.52	1.81	2.08	1.78	3.48	3.72	3.84	4.92
	3	1.11	1.15	1.26	1.27	1.36	1.78	2.24	3.04	3.13	3.94
	4	1.44	1.60	2.04	2.08	2.08	4.12	2.24	2.36	3.84	2.96
HDVs-I and HDVs-II	1	2.02	2.24	2.04	1.84	1.80	1.76	1.68	1.60	1.56	1.80
	2	1.34	1.31	1.52	1.84	1.80	1.76	1.68	1.60	1.56	1.20
	3	1.34	1.31	1.26	1.21	1.20	1.19	1.17	1.15	1.14	1.20
	4	2.36	2.24	2.04	1.42	1.20	1.19	1.34	1.45	1.42	1.80

the road network resources are not fully utilized, and the road benefit is low. When the average saturation is large, the utilization rate of road resources is high, and the roads benefit more, but the overall congestion of the road network is high, and the service level is low.

4.2.4. Comprehensive Evaluation. Due to different dimensions of road network reserve capacity, the average travel time, and the average saturation ratios, a simple direct addition of indexes of different properties cannot correctly reflect the comprehensive results. In order to solve the comparability between data indexes, data have to be processed by the Method of Normalization, rendering each index normalized in the same order of magnitude for a better comprehensive analysis. In this thesis, deviation standardization is adopted to make the result value map between [0, 1] through linear processing of original data, and the conversion function is as follows:

$$x_{ij}' = \frac{x_{ij} - \min\{x_{ij}\}}{\max\{x_{ij}\} - \min\{x_{ij}\}}. \tag{16}$$

The average travel time is a backward indicator and the smaller, the better. Render it positive through $y_{ij} = -x_{ij}$. The average saturation is a moderate index, and its value should be neither too large nor too small. Through the forward transformation of the index by $y_{ij} = -|x_{ij} - k|$, where the moderate value k is the mean of saturation, i.e., $k = \sum_i \sum_j x_{ij}/n$, the normalized processing results of road network reserve capacity, average travel time, and average saturation are shown in Tables 9–11, respectively.

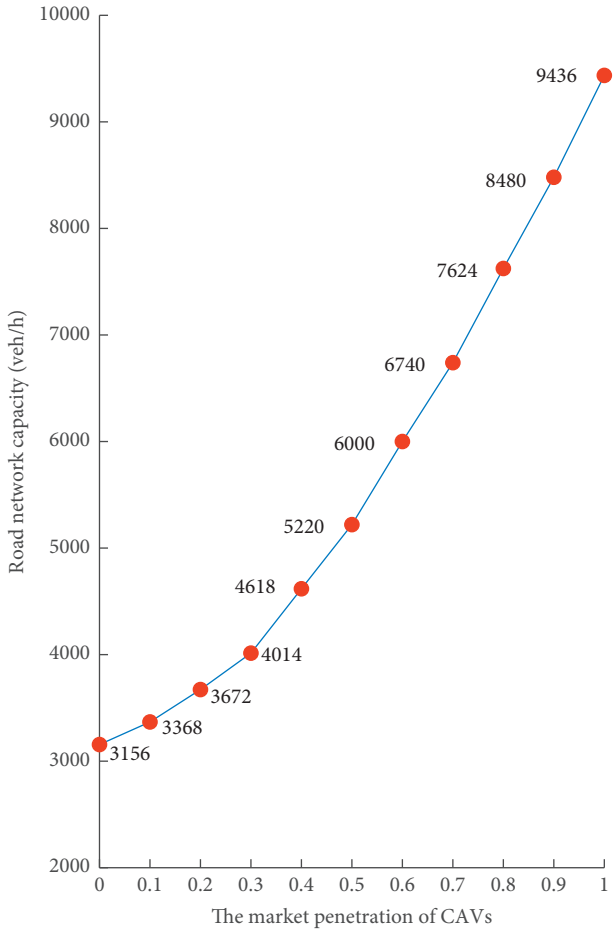
By normalizing the data of three indicators, the comprehensive score value D can be calculated as follows:

$$D = \lambda_1 C^* + \lambda_2 T^* + \lambda_3 S^*, \tag{17}$$

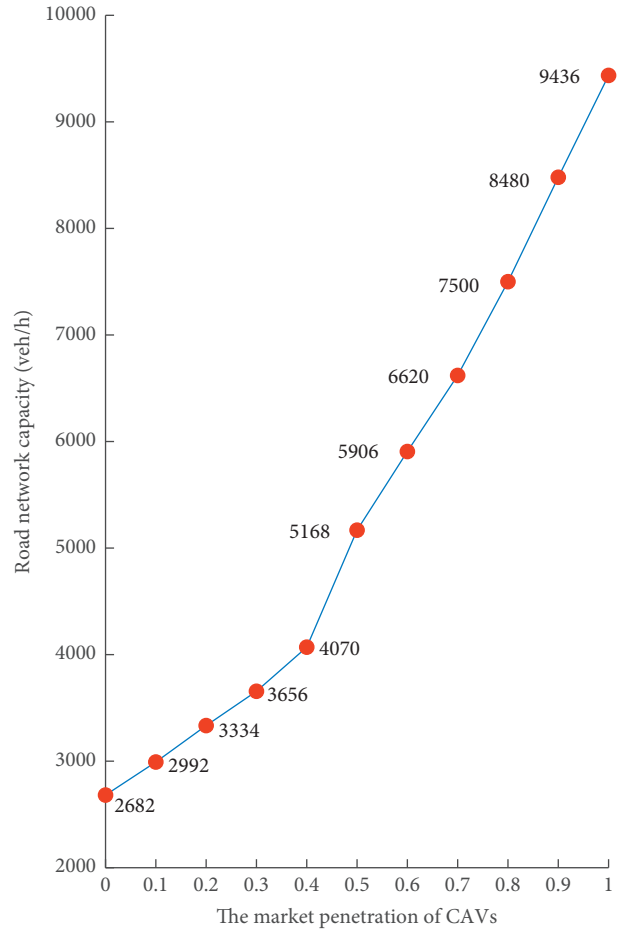
where C^*, T^* and S^* are the normalized matrix of road network reserve capacity, average travel time, and average saturation, respectively. λ_1, λ_2 and λ_3 are the weight of road network reserve capacity, average travel time, and average saturation, respectively.

Among them, the index weight can be determined according to the road network, traffic demand, road utilization degree, etc. The value range of the index weight is [0, 1], and the sum of the weights is 1.

The corresponding comprehensive score values are calculated according to the weights of different indicators, and the location distribution of the maximum comprehensive score values in different weights is shown in Figure 3. The numbers in Figure 3 are position numbers, indicating the market penetration distribution of CAVs and HDVs-I, numbered in order from top to bottom and



(a)



(b)

Figure 2: Continued.

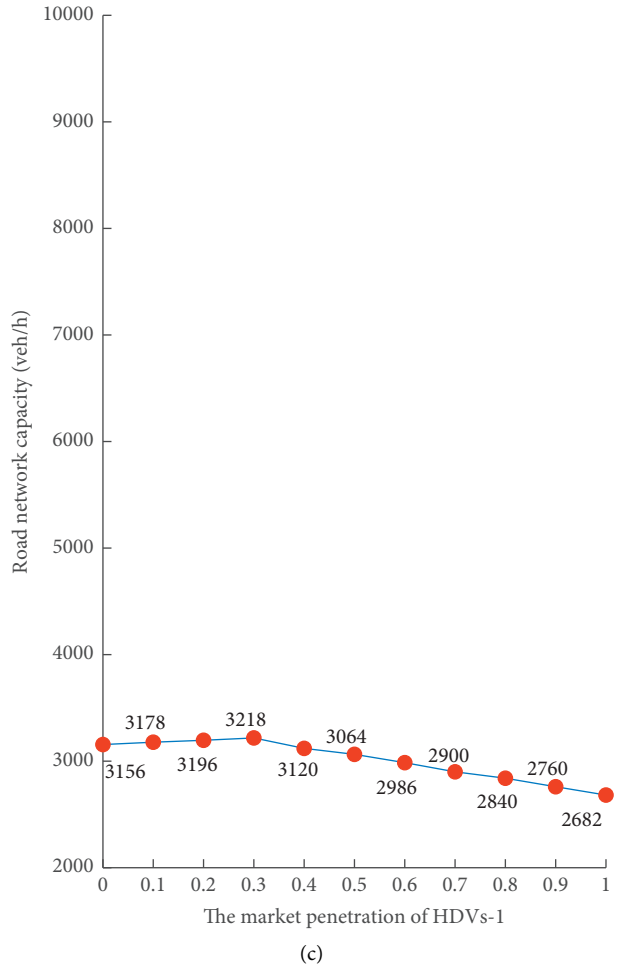


Figure 2: Changing diagram of road network capacity under different market penetration. (a) CAVs and HDVs-II. (b) CAVs and HDVs-I. (c) HDVs-I and HDVs-II.

Table 7: Average travel time under different market penetration (min).

β	α										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	36.61	36.58	36.57	36.70	36.53	36.38	36.17	35.96	35.84	35.74	35.54
0.1	36.55	36.55	36.58	36.70	36.56	36.28	36.06	36.08	35.83	35.71	35.60
0.2	36.59	36.60	36.59	36.67	36.56	36.25	36.06	35.98	35.91	35.81	35.73
0.3	36.64	36.66	36.63	36.70	36.62	36.21	36.35	36.00	36.12	36.04	35.71
0.4	36.88	36.95	36.75	36.96	36.72	36.60	36.15	36.35	36.11	36.03	35.95
0.5	37.20	37.43	37.42	37.57	37.30	37.68	37.62	37.45	37.44	37.52	35.97
0.6	37.76	37.79	37.71	37.64	37.50	37.88	37.77	37.67	37.60	37.90	37.80
0.7	38.01	38.00	37.91	38.09	38.09	38.01	37.91	37.80	37.78	37.66	38.00
0.8	38.46	38.36	38.19	38.24	38.15	38.16	38.07	38.01	38.17	38.03	37.93
0.9	38.56	38.56	38.78	38.77	38.77	38.76	38.76	38.76	38.76	38.76	38.75
1.0	38.67	38.67	38.67	38.67	38.67	38.67	38.67	38.67	38.67	38.67	38.67

left to right, as shown in Table 12. Through analysis and calculation, the reasonable market penetration of different types of vehicles under different characteristic requirements can be obtained so as to determine the optimal deployment scheme of CAVs and HDVs in different periods.

(1) When the weight of road network reserve capacity is larger, the optimal position number of the comprehensive score value is 11, and the market penetration of CAVs is 1. It can be seen that if the priority is to increase the reserve capacity of the road network, the putting into use of CAVs should be

Table 8: Average saturation under different market penetration.

β	α										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.475	0.478	0.481	0.484	0.469	0.460	0.448	0.434	0.425	0.411	0.401
0.1	0.483	0.488	0.489	0.488	0.480	0.467	0.455	0.446	0.434	0.425	0.414
0.2	0.501	0.506	0.502	0.497	0.488	0.475	0.464	0.456	0.448	0.436	0.422
0.3	0.519	0.517	0.516	0.510	0.501	0.489	0.481	0.466	0.457	0.446	0.433
0.4	0.548	0.547	0.529	0.524	0.507	0.501	0.493	0.484	0.478	0.465	0.437
0.5	0.583	0.582	0.581	0.592	0.571	0.562	0.566	0.551	0.529	0.521	0.451
0.6	0.634	0.642	0.636	0.631	0.620	0.605	0.602	0.597	0.590	0.578	0.562
0.7	0.672	0.672	0.665	0.656	0.661	0.657	0.648	0.638	0.632	0.621	0.615
0.8	0.700	0.693	0.698	0.686	0.695	0.693	0.685	0.679	0.677	0.670	0.651
0.9	0.721	0.720	0.717	0.717	0.717	0.716	0.716	0.715	0.715	0.714	0.714
1.0	0.734	0.734	0.734	0.734	0.734	0.734	0.734	0.734	0.734	0.734	0.734

Table 9: Normalization of road network reserve capacity.

β	α										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.070	0.073	0.076	0.079	0.065	0.057	0.045	0.032	0.023	0.012	0.000
0.1	0.102	0.107	0.107	0.106	0.098	0.086	0.075	0.065	0.056	0.046	0.037
0.2	0.147	0.152	0.147	0.141	0.131	0.121	0.110	0.103	0.097	0.084	0.076
0.3	0.197	0.192	0.192	0.182	0.173	0.165	0.155	0.144	0.134	0.126	0.115
0.4	0.287	0.287	0.267	0.262	0.240	0.219	0.206	0.198	0.192	0.184	0.176
0.5	0.376	0.377	0.379	0.395	0.371	0.368	0.376	0.361	0.339	0.338	0.245
0.6	0.491	0.502	0.496	0.490	0.477	0.468	0.466	0.463	0.458	0.449	0.450
0.7	0.601	0.601	0.592	0.583	0.592	0.589	0.578	0.567	0.563	0.554	0.564
0.8	0.732	0.721	0.713	0.710	0.714	0.713	0.704	0.697	0.688	0.688	0.677
0.9	0.858	0.858	0.857	0.857	0.857	0.857	0.857	0.857	0.857	0.857	0.857
1.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 10: Normalization of average travel time.

β	α										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.670	0.679	0.682	0.642	0.694	0.741	0.806	0.870	0.907	0.938	1.000
0.1	0.688	0.688	0.679	0.642	0.685	0.772	0.840	0.833	0.910	0.948	0.981
0.2	0.676	0.673	0.676	0.651	0.685	0.781	0.840	0.864	0.886	0.917	0.941
0.3	0.660	0.654	0.664	0.642	0.667	0.793	0.750	0.858	0.821	0.846	0.948
0.4	0.586	0.565	0.627	0.562	0.636	0.673	0.812	0.750	0.824	0.849	0.873
0.5	0.488	0.417	0.420	0.373	0.457	0.340	0.358	0.410	0.414	0.389	0.867
0.6	0.315	0.306	0.330	0.352	0.395	0.278	0.312	0.343	0.364	0.272	0.302
0.7	0.238	0.241	0.269	0.213	0.213	0.238	0.269	0.302	0.309	0.346	0.241
0.8	0.099	0.130	0.182	0.167	0.194	0.191	0.219	0.238	0.188	0.231	0.262
0.9	0.068	0.068	0.000	0.003	0.003	0.006	0.006	0.006	0.006	0.006	0.009
1.0	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034

vigorously promoted, and the market penetration of CAVs should be improved as much as possible.

(2) When the weight of average travel time is larger, the best position number of the comprehensive score value is 111. At this time, the market penetration of CAVs and HDVs-I is 0 and 1, respectively. It can be seen that if the travel time of travelers is to be reduced as much as possible, the market penetration of HDVs-I should be increased, and more travelers can travel in the shortest path through ATIS devices.

(3) When the weight of average saturation is larger, the best position numbers of the comprehensive score value are 50. At this time, the market penetration of CAVs and HDVs-I is 0.5 and 0.4, respectively. Therefore, the market penetration of CAVs and HDVs-I should be increased to a certain level at the same time if the traffic load of the road network is considered to ensure that the appropriate traffic flow is relatively smooth to drive in the road network.

(4) When the weight of the three indicators is moderate, the best position is different according to the focus of

Table 11: Normalization of average saturation.

β	α										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	0.435	0.453	0.471	0.488	0.400	0.347	0.276	0.194	0.141	0.059	0.000
0.1	0.482	0.512	0.518	0.512	0.465	0.388	0.318	0.265	0.194	0.141	0.076
0.2	0.588	0.618	0.594	0.565	0.512	0.435	0.371	0.324	0.276	0.206	0.124
0.3	0.694	0.682	0.676	0.641	0.588	0.518	0.471	0.382	0.329	0.265	0.188
0.4	0.865	0.859	0.753	0.724	0.624	0.588	0.541	0.488	0.453	0.376	0.212
0.5	0.968	0.974	0.979	0.915	1.000	0.947	0.971	0.882	0.753	0.706	0.294
0.6	0.668	0.621	0.656	0.685	0.750	0.838	0.856	0.885	0.926	0.997	0.947
0.7	0.444	0.444	0.485	0.538	0.509	0.532	0.585	0.644	0.679	0.744	0.779
0.8	0.279	0.321	0.291	0.362	0.309	0.321	0.368	0.403	0.415	0.456	0.568
0.9	0.156	0.162	0.179	0.179	0.179	0.185	0.185	0.191	0.191	0.197	0.197
1.0	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079	0.079

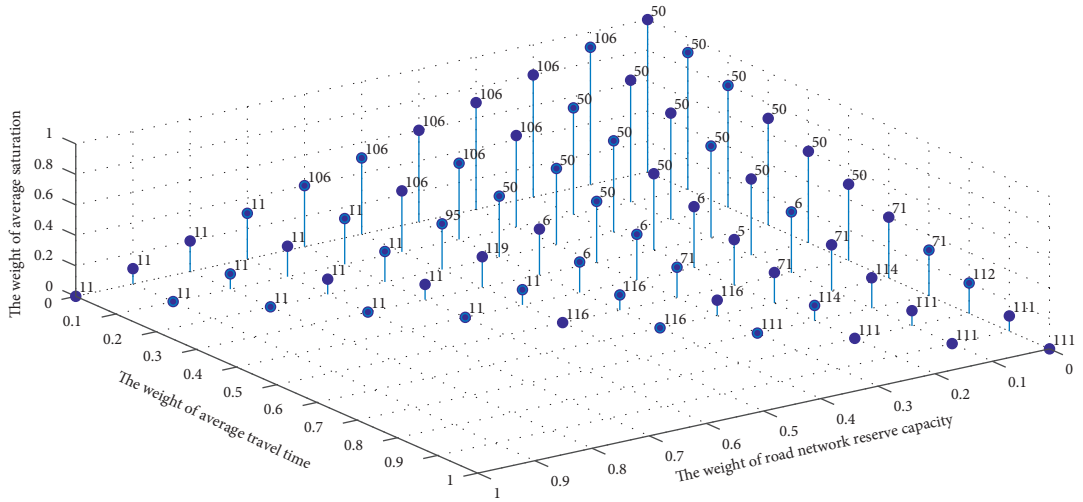


Figure 3: The optimal location distribution of comprehensive score values under different index weights.

Table 12: Position numbers corresponding to different market penetration.

β	α										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	1	12	23	34	45	56	67	78	89	100	111
0.1	2	13	24	35	46	57	68	79	90	101	112
0.2	3	14	25	36	47	58	69	80	91	102	113
0.3	4	15	26	37	48	59	70	81	92	103	114
0.4	5	16	27	38	49	60	71	82	93	104	115
0.5	6	17	28	39	50	61	72	83	94	105	116
0.6	7	18	29	40	51	62	73	84	95	106	117
0.7	8	19	30	41	52	63	74	85	96	107	118
0.8	9	20	31	42	53	64	75	86	97	108	119
0.9	10	21	32	43	54	65	76	87	98	109	120
1.0	11	22	33	44	55	66	77	88	99	110	121

Table 13: Travel demand of OD pair.

OD pair number	Demand growth multiplier	Travel demand (veh/h)
1	3.36	1344
2	2.77	2216
3	1.59	954
4	3.36	672

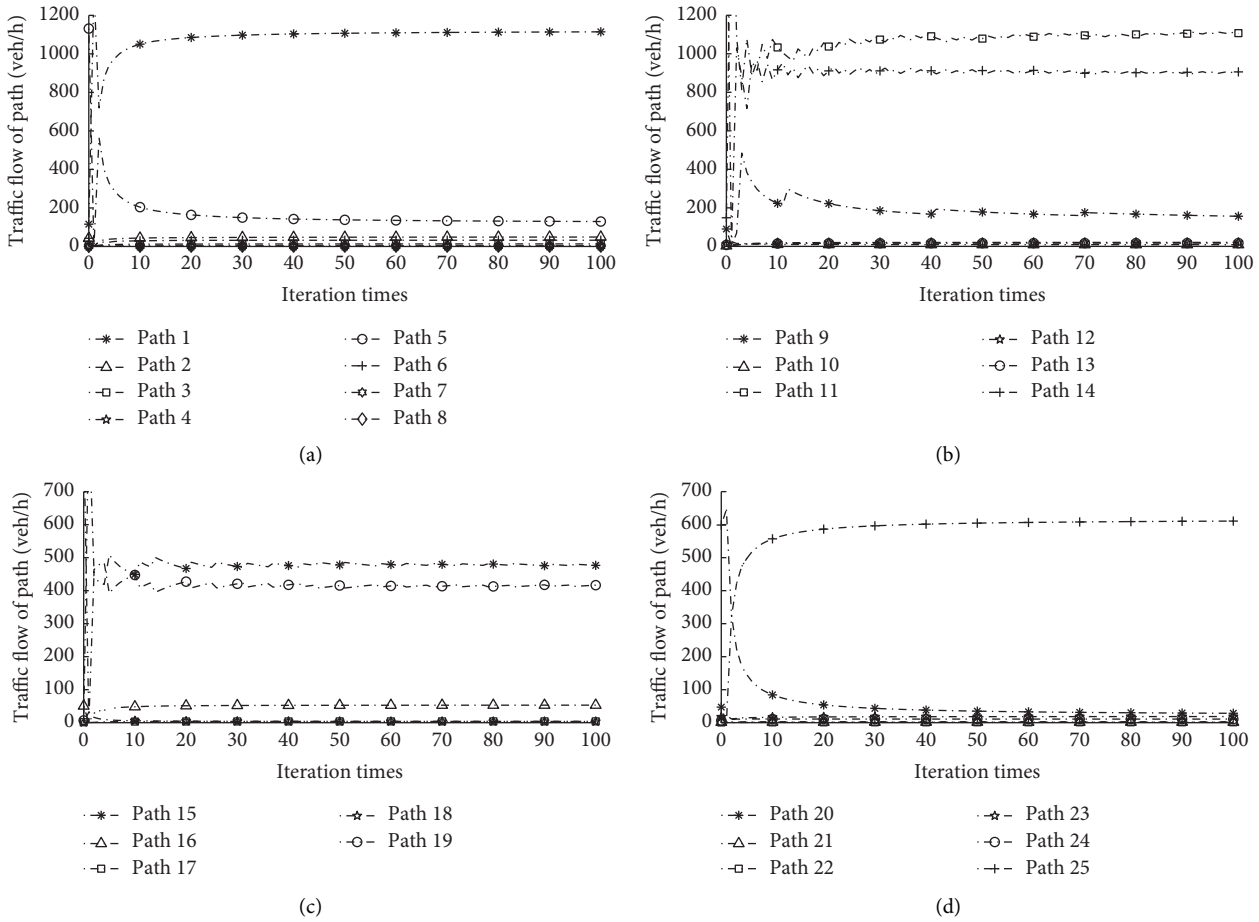


Figure 4: Evolution of path flow with iterations. (a) OD pair 1. (b) OD pair 2. (c) OD pair 3. (d) OD pair 4.

Table 14: Path data of OD pair 2 in equilibrium state.

Path number	Traffic flow of CAVs (veh/h)	Traffic flow of HDVs-I (veh/h)	Traffic flow of HDVs-II (veh/h)	Total traffic flow (veh/h)	Average travel time (min)
9	37	0	113	150	39.65
10	0	0	9	9	44.79
11	430	419	271	1120	37.90
12	0	0	21	21	43.04
13	0	0	17	17	43.39
14	641	24	234	899	38.19

attention but mainly concentrated in 50. At this time, without the excessive pursuit of the comprehensive popularization of CAVs, different types of vehicles reach a certain proportion, orderly mixed in the road network.

4.3. Analysis of the Bilevel Programming Model. The previous section has given the optimal comprehensive score value under different weights and the corresponding market penetration. Among them, position number 50 has the largest proportion. Meanwhile, considering that CAVs cannot be fully popularized for a long time in the future, different types of vehicles will be mixed in the road network. This section assumes that the market penetration of CAV and HDVs-I are 0.5 and 0.4, respectively. The demand growth multiplier and travel demand of each OD pair can be obtained as shown in Table 13. According to the bilevel programming model established in this thesis, the mixed traffic flow is analyzed, and the test road network can rapidly converge to the equilibrium state, as shown in Figure 4. When it gets into the equilibrium state, the road network reserve capacity is 5186 veh/h, the total travel time is 193426 min, the average travel time is 37.30 min, and the average saturation is 0.571. For the OD pair 1, more than 83.4% of the traffic flow is concentrated in path 1, where the

flow of this path is 1121 veh/h, and the path travel time is the smallest, which is 34.99 min. The flow of OD pair 2 is mainly distributed in paths 11 and 14, of which 50.5% is distributed in paths 11, up to 1120 veh/h, and the travel time of this path is 37.90 min, while the flow of path 14 is 899 veh/h and the travel time is 38.19 min. The flow of OD pair 3 is mainly concentrated in paths 15 and 19, with selection probabilities of 50.1% and 43.5%, respectively. The travel time of path 15 is 37.02 min and the flow is 478 veh/h, while the travel time of path 19 is 41.05 min and the flow is 415 veh/h. In addition, the traveling time of path 16 is 39.64 min, which is smaller than that of path 19. However, link 7 contained in path 16 is saturated, and the flow allocated by path 16 is relatively small due to the constraints of the capacity of links. For OD pair 4, more than 91.7% of the traffic flow selects path 25, and the travel time is 35.34 min.

OD pair 2 is selected to analyze the path selection of different types of vehicles in the equilibrium state to verify the traffic assignment model. The relevant data are shown in Table 14. The traffic flow of CAVs is distributed on paths 9, 11, and 14, of which path 14 is not the shortest path but has the most distributed traffic flow, indicating that CAVs follow the SO principle to select paths to minimize the total travel time of the system, and the traffic flow is not concentrated on the shortest path. The traffic flow of HDVs is allocated according to the SUE principle. On the one hand, HDVs-I has a high level of information quality, and the traffic is only distributed on paths 11 and 14. Among them, path 11 is the shortest path and has the most distributed traffic, but some traffic is also distributed in the second shortest path. If the level of information quality is further improved, users can fully grasp the road information, the traffic flow can be concentrated on the shortest path, and ultimately realize user optimization. On the other hand, the traffic flow of HDVs-II is distributed on all paths, indicating that HDVs-II has certain randomness in selecting the path, and the smaller the travel time, the greater the probability of being selected.

5. Conclusion

Note that smart and unintelligent vehicles unavoidably converge on the road network with the development of increasingly maturing autonomous driving technology. This study focuses on the development of a model and algorithm for traffic assignment problems under link capacity constraints, which aims to formulate a reasonable CAVs deployment scheme by analyzing the impact of market penetration of different types of vehicles on road network performance so as to facilitate the promotion and application of CAVs.

In this paper, SO-SUE mixed equilibrium traffic assignment model is constructed, which maximizes the road network reserve capacity for the sake of the link capacity constraint. The effects of different market penetration on road network reserve capacity, average travel time, and average saturation are analyzed using the Nguyen-Dupuis network. On the one hand, the three indicators all increase with the market penetration of CAVs and present a trend from slow to fast; On the other hand, all indicators generally decrease with the increase of the market penetration of HDVs-I, and the decreased amplitude is small. There is a slight increase when the market penetration of CAVs and HDVs-I is very small. Through the normalization of three index data, the optimal comprehensive score and the corresponding market penetration in different index weights are analyzed herein. Thus, the optimal deployment scheme of CAVs and HDVs can be determined according to different planning objectives. The bilevel programming model is solved by a numerical example to analyze the equilibrium state of mixed traffic flow, and the convergence of the model and algorithm is verified. It is worth noting that the nonuniform demand growth multiplier adopted by each OD pair in this paper not only improves the reserve capacity of the road network but also reflects the inconsistency of traffic demand structure growth, which is more in line with the characteristics of actual travel demand. Considering that CAVs have small headway, the link capacity is not a fixed value, and its size depends on the market penetration of CAVs. In addition, SAA is innovatively proposed to solve the model to obtain the global optimal solution.

Beyond the above research, there are several directions for further study. First, there are some research theories about the dedicated area of CAVs, and it is necessary to study the impact of the dedicated area setting on road network capacity. Second, the next step can consider that the travelers in the road network follow the system optimal, user optimal, and stochastic user optimal to select the path and analyze the influence of mixed traffic flow characteristics on the mixed equilibrium traffic assignment mechanism, which will be more in line with the actual situation. In addition, from the perspective of comfort, the influence of turning times, service level, number of signal lights, and other factors on path selection will be considered, and the model will be improved so as to satisfy the actual travel demand and describe the distribution pattern of road network traffic flow more realistically.

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 supported by the National Natural Science Foundation of China (51578080 and 51978082), Scientific Research Foundation of Hunan Provincial Education Department (20C0375 and 20A093), Postgraduate Scientific Research Innovation Project of Hunan Province (CX20210749), Open Fund of Engineering Research Center of Catastrophic Prophylaxis and Treatment of Road and Traffic Safety of Ministry of Education (kfj190406), Transportation Science and Technology Project of Hunan

Province (202039), and Hunan Provincial Natural Science Foundation of China (2021JJ40025).

References

[1] J. Lioris, R. Pedarsani, F. Y. Tascikaraoglu, and P. Varaiya, "Platoons of connected vehicles can double throughput in urban roads," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 292–305, 2017.

[2] D. Milakis, B. van Arem, and B. van Wee, "Policy and society related implications of automated driving: a review of literature and directions for future research," *Journal of Intelligent Transportation Systems*, vol. 21, no. 4, pp. 324–348, 2017.

[3] M. Noruzoliaee, B. Zou, and Y. Liu, "Roads in transition: integrated modeling of a manufacturer-traveler-infrastructure system in a mixed autonomous/human driving environment," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 307–333, 2018.

[4] P. Tientrakool, Y. C. Ho, and N. F. Maxemchuk, "Highway capacity benefits from using vehicle-to-vehicle communication and sensors for collision avoidance," in *Proceedings of the Seventy Fourth IEEE Vehicular Technology Conference*, pp. 5–8, San Francisco, CA, USA, September 2011.

[5] A. Assidiq, O. O. Khalifa, R. Islam, and S. Khan, "Real time lane detection for autonomous vehicles," in *Proceedings of the International Conference on Computer and Communication Engineering*, pp. 82–88, Kuala Lumpur, Malaysia, July 2008.

[6] S. Eben Li, K. Li, and J. Wang, "Economy-oriented vehicle adaptive cruise control with coordinating multiple objectives function," *Vehicle System Dynamics*, vol. 51, no. 1, pp. 1–17, 2013.

[7] A. Haurie and P. Marcotte, "On the relationship between Nash-Cournot and Wardrop equilibria," *Networks*, vol. 15, no. 3, pp. 295–308, 1985.

[8] J. Wang, W. Deng, and J. Zhao, "Road network reserve capacity with stochastic user equilibrium," *Transport*, vol. 30, no. 1, pp. 103–116, 2015.

[9] S. Ryu, A. Chen, X. Xu, and K. Choi, "A dual approach for solving the combined distribution and assignment problem with link capacity constraints," *Networks and Spatial Economics*, vol. 14, no. 2, pp. 245–270, 2014.

[10] Z. Huang, J. Wu, A. Kuang, S. Zhang, and Y. Xu, "Modeling and simulation of travelers' route choice behavior based on disequilibrium theory," *Journal of System Simulation*, vol. 30, no. 11, pp. 4067–4078, 2018.

[11] S. A. Bagloee, M. Sarvi, M. Patriksson, and A. Rajabifard, "A mixed user-equilibrium and system-optimal traffic flow for connected vehicles stated as a complementarity problem," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 7, pp. 562–580, 2017.

[12] T. V. Vuren and D. Watling, "A multiple user class assignment model for route guidance," *Transportation Research Record*, vol. 1306, pp. 22–32, 1991.

[13] H. Yang and Q. Meng, "Modeling user adoption of advanced traveler information systems: dynamic evolution and stationary equilibrium," *Transportation Research Part A: Policy and Practice*, vol. 35, no. 10, pp. 895–912, 2001.

[14] Y. Yin and H. Yang, "Simultaneous determination of the equilibrium market penetration and compliance rate of advanced traveler information systems," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 2, pp. 165–181, 2003.

[15] G. N. Bifulco, G. E. Cantarella, F. Simonelli, and P. Velona, "Advanced traveller information systems under recurrent traffic conditions: network equilibrium and stability," *Transportation Research Part B: Methodological*, vol. 92, pp. 73–87, 2016.

[16] M. Dell'Orco and M. Marinelli, "Modeling the dynamic effect of information on drivers' choice behavior in the context of an Advanced Traveler Information System," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 168–183, 2017.

[17] D. Xu, Y. Dong, P. Peng, L. Lin, and Y. Liu, "The evaluation of urban road network based on complex network," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, 2021.

[18] T. Li, A. Ni, C. Zhang, G. Xiao, and L. Gao, "Short-term traffic congestion prediction with Conv-BiLSTM considering spatio-temporal features," *IET Intelligent Transport Systems*, vol. 14, no. 14, pp. 1978–1986, 2020.

[19] T. Akamatsu and O. Miyawaki, "Maximum network capacity problem under the transportation equilibrium assignment," *Infrastructure Planning Review*, vol. 12, pp. 719–729, 1995.

[20] M. W. Levin and S. D. Boyles, "Effects of autonomous vehicle ownership on trip, mode, and route choice," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2493, no. 1, pp. 29–38, 2015.

[21] V. A. C. van den Berg and E. T. Verhoef, "Autonomous cars and dynamic bottleneck congestion: the effects on capacity, value of time and preference heterogeneity," *Transportation Research Part B: Methodological*, vol. 94, pp. 43–60, 2016.

[22] M. Lavasani, X. Jin, and Y. Du, "Market penetration model for autonomous vehicles on the basis of earlier technology adoption experience," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2597, no. 1, pp. 67–74, 2016.

[23] T. Li, F. Guo, and R. Krishnan, A. Sivakumar and J. Polak, Right-of-way reallocation for mixed flow of autonomous vehicles and human driven vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 115, Article ID 102630, 2020.

[24] D. Chen, S. Ahn, M. Chitturi, and D. A. Noyce, "Towards vehicle automation: roadway capacity formulation for traffic mixed with regular and automated vehicles," *Transportation Research Part B: Methodological*, vol. 100, pp. 196–221, 2017.

[25] A. Ghiasi, O. Hussain, Z. S. Qian, and X. Li, "A mixed traffic capacity analysis and lane management model for connected automated vehicles: a Markov chain method," *Transportation Research Part B: Methodological*, vol. 106, pp. 266–292, 2017.

[26] Z. Chen, F. He, Y. Yin, and Y. Du, "Optimal design of autonomous vehicle zones in transportation networks," *Transportation Research Part B: Methodological*, vol. 99, pp. 44–61, 2017.

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

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 1438630, 10 pages
https://doi.org/10.1155/2022/1438630

Research Article

Effects of Implementing Night Operation Signal Coordination on Arterials

Rui Yue,[1] **Guangchuan Yang**,[2] **Yichen Zheng**,[3] **Yang Yang**,[4] **and Zong Tian**[5]

[1]*School of Traffic and Transportation, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District, Beijing 100044, China*
[2]*Institute for Transportation Research and Education (ITRE), North Carolina State University, 909 Capability Drive, Raleigh, NC 27695, USA*
[3]*Beijing Nebula Link Technology Co., Ltd, Suite A9, No. 8 Xueqing Rd, Haidian District, Beijing 10083, China*
[4]*School of Trasportation Science and Engineering, Beihang University, No. 37 Xueyuan Road, Haidian District, Beijing 100091, China*
[5]*Department of Civil and Environmental Engineering, University of Nevada, Reno, 1664 N. Virginia Street, MS 258, Reno, NV 89557, USA*

Correspondence should be addressed to Rui Yue; yuerui@bjtu.edu.cn

Received 22 April 2022; Accepted 13 June 2022; Published 2 August 2022

Academic Editor: Gen Li

Traffic signal coordination, which connects a series of signals along an arterial by various coordination methodologies, has been proven as one of the most cost-effective means for alleviating traffic congestion. Various metropolitan planning organizations (MPO) or transportation management centers (TMC) have included signal timing updates in their strategic plans. However, in practice, signal coordination is usually implemented when traffic volume is heavy (i.e., during peak hours). For the rest of the day, the free operation strategy is usually used to reduce the waiting time of uncoordinated phases. However, this free operation strategy may result in the loss of operational efficiency on the major street. Currently, implementing signal coordination during off-peak hours is rare in the U.S. since there is lack of an efficient method that considers traffic operations for both the major and the minor streets. Therefore, this research provides a novel method that balances the control delays between the major street and the minor street. The procedure is to optimize the splits of the major street while also using the reservice strategy in the signal controller for the minor street. Microsimulation modeling was employed to assess the performance of traffic signal coordination during off-peak periods. Results show that, under reasonable splits, the coordination effect on the major street can be achieved and protected with an acceptable delay to minor street traffic. The strategy can be immediately implemented to reduce travel time for major street traffic.

1. Introduction

Over the years, signal coordination has been adopted as an effective measure for alleviating congestion, and thus signal coordination plans are typically implemented during peak hours. In current practice, when traffic volume is low, signal coordination is usually deactivated from the controller. Instead, a free operation strategy, a fully actuated coordination that requires detecting all movements, is typically employed. This is mainly because signal coordination often requires more green time assigned to the central street to achieve a guaranteed mainline capacity. However, when mainline traffic flow is not heavy, the extra time for mainline cannot be used by the side street, resulting in the loss of intersection capacity. Hence, it has been a common sense that signal coordination may not be necessary when traffic volume is relatively low, such as at night. Nevertheless, although free operation can balance the waiting time of each approach, it may break the mainline traffic flow at any time once vehicles are on the side street, which may make drivers experience frequent stops and produce more emissions.

There has been a sustained debate about the suitable signal control strategy that should be used during nighttime. The key point is to balance the tradeoffs between the delays

in the coordinated and non-coordinated phases. Currently, the free operation strategy has been commonly adopted in practice. However, even though most agencies prefer to use the free operation under low volume conditions, no evidence indicates that running the free operation is a better solution compared to coordinated plans from the perspective of operational efficiency. Moreover, the use of signal reservice strategy enables flexible switches between side street (non-coordinated) and main street (coordinated) phases, which avoids the conflict between the main street coordination delay to the side street. With this consideration, this research investigated the operation difference by comparing the simulated performance of two signal operation strategies under a series of low volume conditions in VISSIM.

The remainder of this paper is divided into the following parts: first, a comprehensive literature review regarding the prevalent signal timing strategies was conducted. Then, a feasible methodology that can balance the mainline progression effects and the minor street waiting time under a low volume condition was proposed. After that, the effects of the proposed strategy by comparing it with the free operation using field data and microsimulation modeling were evaluated, and finally, conclusions and recommendations were summarized.

2. Literature Review

The concept and strategy of signal coordination have been proposed and developed in the last century; currently, it is one of the critical methods in protecting cities from congestion. Bandwidth progression [1] was developed and gradually became a major optimization method. Furthermore, two algorithms based on bandwidth optimization were sequentially proposed: half-integer algorithm [2] and the mixed-integer linear programming [3], which can be used to derive maximum bandwidth. While simply achieving the maximum bandwidth is not enough, the interference elimination function [4] was later developed, and then phase sequences and offsets were also becoming adjustable. These parameters can effectively influence the signal coordination effect.

Based on these parameters, some algorithms and useful strategies were carried out to benefit the signal coordination. A software PASSER-2 [5] was developed, which enabled flexible phase adjustment. Little et al. [6] proposed the MAXBAND algorithm which involved offsets, sequences, speeds, and a series of parameters. Gartner [7] further refined MAXBAND to a multiband model, which can accommodate variable bandwidths. Abbas [8] et al. developed a real-time offset transition algorithm, which has the potential to cater to the traffic flow. Tian investigated a series of signal coordination strategies for practical use, including the system partition technique [9], the split-phasing operation [10], and the lead-lag phasing operation [11]. After that, studies focusing on improving intersection [12] and interchange control [13, 14] were continuously studied by different researchers.

As for low volume conditions' signal coordination, to the best knowledge of the research team, little is mentioned in the existing federal or regional signal timing manual since the coordination is originally for conditions where significant volumes appear. Although several researchers mentioned how to determine time-of-day (TOD) [15–17] breakpoints, what plan should be used after peak hour remains a problem. There are several alternative operational strategies for low volume conditions.

In the MnDOT Traffic Signal Timing and Coordination Manual [18], traffic actuated control was recommended as the countermeasure for low volume conditions, in other words, the coordination plan was not recommended. Abdelghany et al. [19] also mentioned that the fully actuated operation is preferred under low volume conditions; however, they also have concerns that not every intersection is capable of fully actuated operation due to intersections lacking detectors.

Another practice is to mimic stop/yield operation when the volume is low by using the flashing mode in the signal controller. Bonneson et al. [20] mentioned that, during low volume periods, the flash operation can be used. However, safety issues were raised by others since data show that this strategy increased the right-angle crash rate. Therefore, the pretimed signal with a plan suitable for the low volume condition is also recommended. Messer et al. [21] mentioned that, at low volumes, any reasonable signal timing strategy works well as long as the detectors work and traffic signals are coordinated.

However, the aforementioned research did not state what strategy should be used under different low-volume conditions. Later, Andalibian [22] mentioned that if the percentage of stops along a non-coordinated arterial exceeds 50, then coordination is recommended. When the percentage of stops is 20 or lower, the actuated operation is recommended. He also raised another criterion [23] based on volumes; signals should be coordinated when traffic volumes in the peak direction for major street reaches 750 vph for arterials with two lanes in each direction and 850 vph for three lanes in each direction. Another practice from the Washington Department of Transportation [24] stated that 400 vph should be the threshold for whether to use signal coordination or not. However, these studies still did not provide solid evidence that their threshold can balance the operational efficiency and minor street waiting time. The engineers of the City of Laguna Niguel [25] mentioned that, during nonpeak times, drivers are less willing to accept the minor street and left turn delays that are part of coordination during quieter/lower volume times of the day. Therefore, a novel strategy that can better balance coordination effects and the minor street delay is necessary. Besides, recent research indicates users following traffic rules will have a lower crash rate under coordination, and safety concerns regarding coordination were also examined and cleared [26] for using signal coordination.

3. Methodology

Although operational efficiency derived from the bandwidth optimization is an important consideration in the determination of whether or not to use a coordination plan, it is

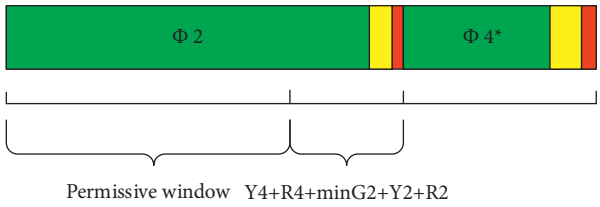


Figure 1: Permissive window of two phase control. Y4: yellow time of phase 4; R4: all red time of phase 4; minG2: the minimum green of phase 2; Y2: yellow time of phase 2; R2: all red time of phase 2.

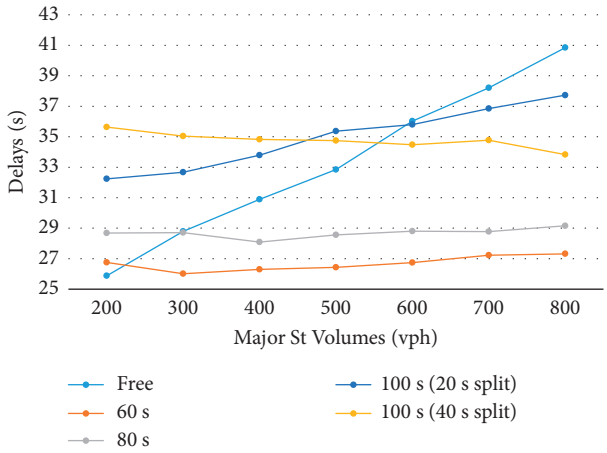


Figure 2: Average delay under minor street volume of 50 vph.

unable to directly reflect minor street drivers' expectations since minor street drivers may still experience long waiting times even if the overall intersection operational efficiency is good. However, in practice, driver expectation also plays an important role in operational performance assessment, such as minor street drivers may get frustrated if the coordination phases do not gap out especially when no vehicles on the major street. Therefore, to satisfy drivers' expectations, this research proposes an analytical model to calculate the probability of stopping using basic signal timing parameters. According to field observations and through expert interviews, a 20-second waiting time (control delay) was selected as the maximum acceptable tolerance time for drivers if no conflict vehicles were present.

At a signalized intersection, the probability of a car stopping more than 20 s at the intersection can be calculated according to the geometric distribution, as follows.

$$p = \max\left(0, 1 - \frac{C - G_e + 4 + 20}{C}\right), \quad (1)$$

where p is the probability of a car experiencing a stop of more than 20 s; G_e is the effective green time of the major street; and C is the cycle length. The reason for adding 4 seconds is because most drivers tend to choose to release the accelerator or use a light break once they see the red light.

With the above model, it is capable to estimate the probability of waiting for more than 20 s under different conditions, which is critical for determining the acceptable signal timing parameters.

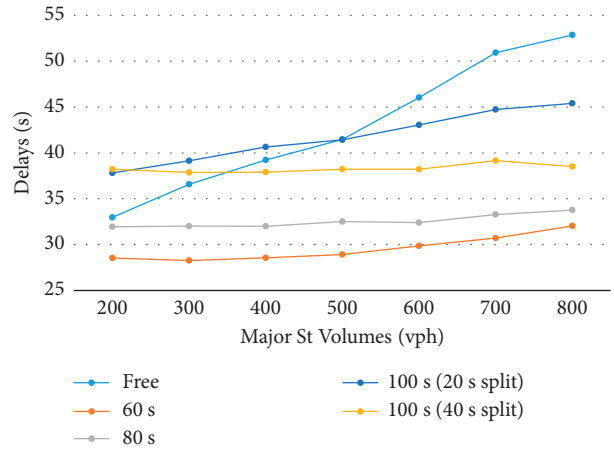


Figure 3: Average delay under minor street volume of 100 vph.

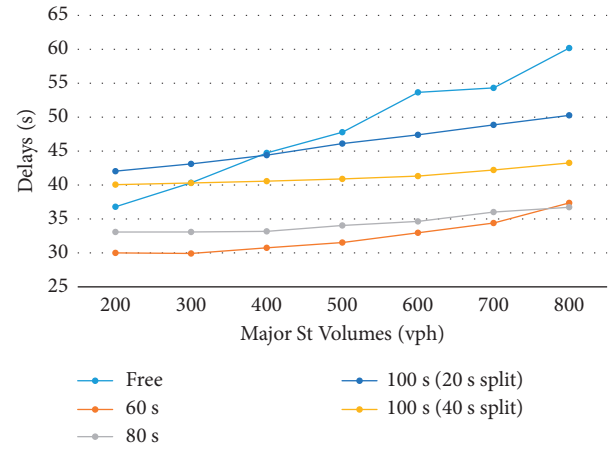


Figure 4: Average delay under minor street volume of 150 vph.

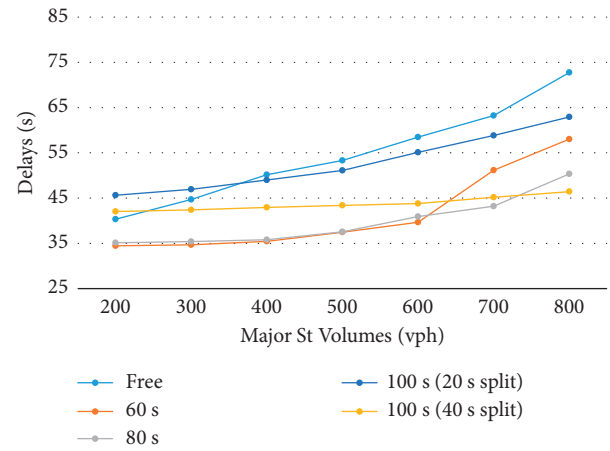


Figure 5: Average delay under minor street volume of 200 vph.

Recently, with the evolution of signal controllers, some new features and technologies have been applied to aid in the operation of intersections. To better satisfy drivers' expectations (i.e., minimum waiting time), a reservice strategy can be applied as it is capable of better assigning the capacity to

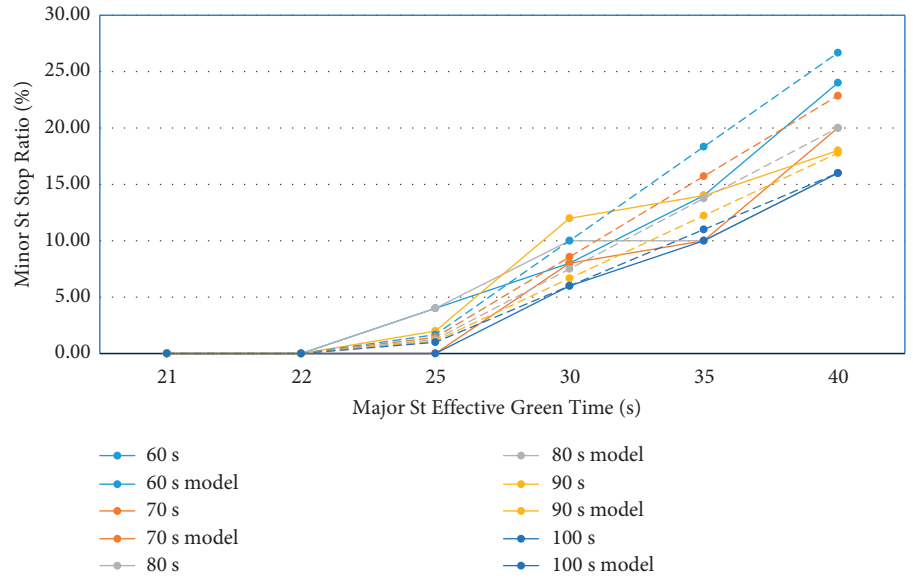


Figure 6: Minor St tolerance stop ratio of coordination plans under the volume of 50 vph.

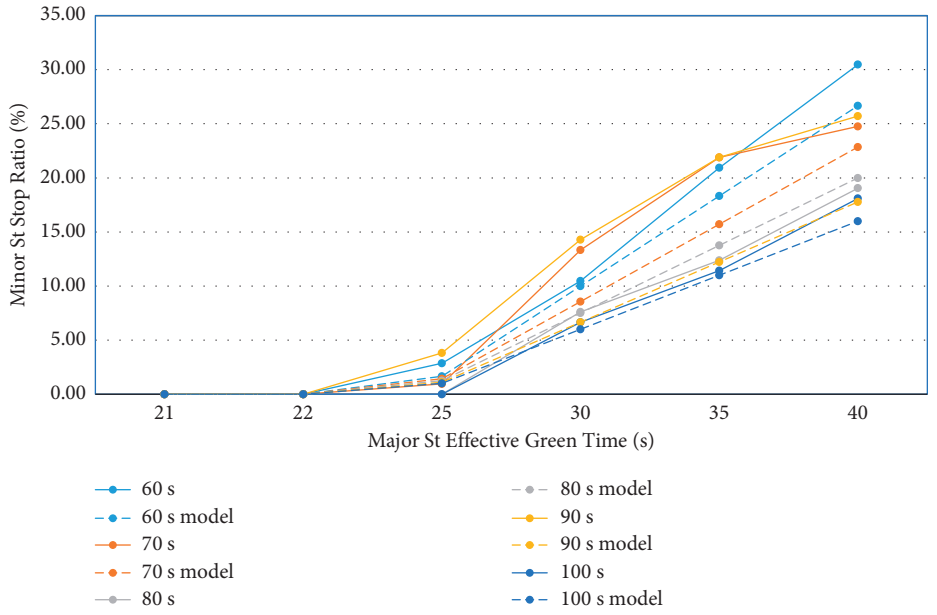


Figure 7: Minor St tolerance stop ratio of coordination plans under the volume of 100 vph.

different movements and improving the operational efficiency. The concept of this strategy is when the major street traffic flow is not platooned, the coordinated phase can be terminated and temporarily transfer right-of-way to non-coordinated phases. After serving the non-coordinated phase, if the coordinated phase is still in the permissive window and has vehicles to be served, the remaining green time will be returned . Therefore, the reservice strategy is considered to benefit operational efficiency. The feature was intentionally designed to reduce the delay in the movements with nonplatoon vehicles. However, insufficient attention was given to its potential for coordination during nonpeak periods. With the reservice strategy, coordination during nonpeak periods becomes feasible. The reservice strategy can happen to some of the in-use phases in their permissive windows. The details of the reservice logic may vary by case, and an example of the permissive window of a two-phase signal control strategy is shown in Figure 1.

The coordinated phase is phase 4. From the figure, phase 2 occupied most of the cycle time. During the permissive window, if phase 2 has no vehicles to serve and phase 4 has vehicles, phase 2 can gap out and switch to phase 4. After serving phase 4, if phase 2 has vehicles waiting and the current time is still within the permissive window, the green signal can be switched back to phase 2 until all phase 2 vehicles are served or until the end of the phase.

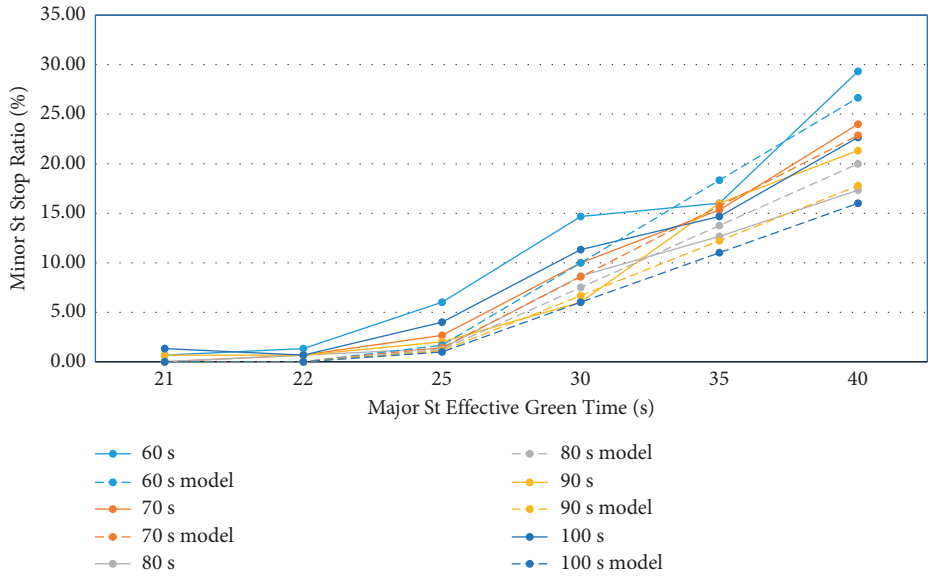


Figure 8: Minor St tolerance stop ratio of coordination plans under volume of 150 vph.

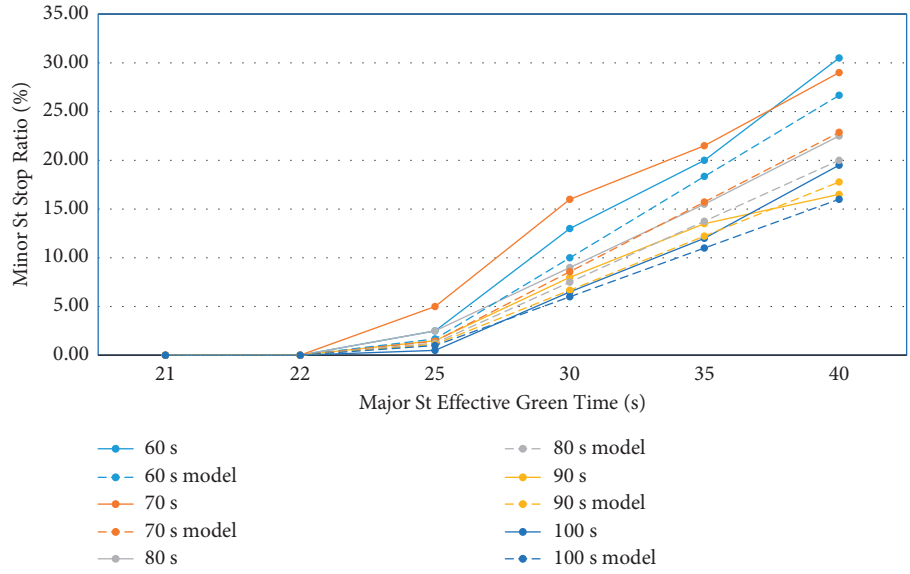


Figure 9: Minor St tolerance stop ratio of coordination plans under volume of 200 vph.

4. Operational Efficiency Assessment

To make sure the strategy is effective, it is necessary to compare the operational efficiency between the free operation strategy and the coordination strategy. The simulation modeling was conducted in the PTV VISSIM. The reservice option was selected before simulation, and all the phases were placed with no recalls. The purpose was to ensure the reservice function was applied. The North McCarran Blvd in Reno, Nevada, was selected as the simulation testbed. Coordination plans under three-cycle lengths (i.e., 60 s, 80 s, and 100 s) and free operations were tested under various major street traffic volumes (i.e., 200 vph–800 vph with an iteration interval of 100vph) and minor street traffic volumes (i.e., 50 vph–200vph with an iteration interval of 50vph). The 100 s cycle length has two conditions: major street split of 20 s and 40 s. Simulated delays are summarized in Figures 2, 3, 4, and 5.

Results show that, under low minor street volume conditions, the average delay under free operation is the lowest when major streets are also under low volume conditions. However, the delay of free operation increases sharply as major street volume increases, which suggests the performance of the free operation is unstable. On the other hand, the delay under the 60 s cycle plan is almost the same as the free operation when the major street volume is 200vph, as shown in Figure 2. While for the rest cases, it significantly outperforms free operation. From Figures 2 to 5, although the minor street volume increases, the 60 s cycle plan always has the lowest delay in most of the cases, with the

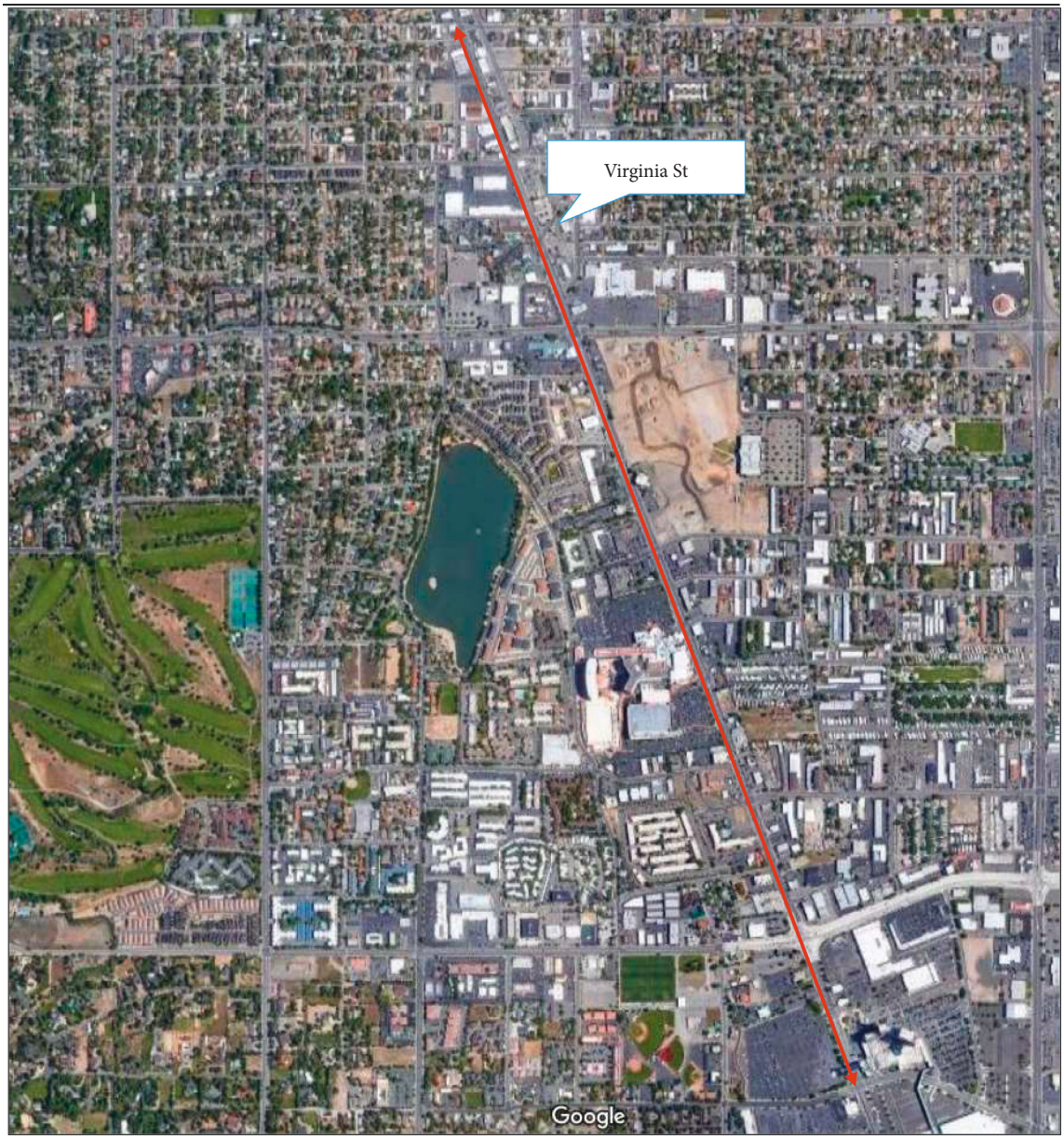


Figure 10: The scope of coordinated segment of Virginia St.

80 s cycle plan being the second optimal plan. Therefore, it can be concluded that, with an appropriate cycle length, the overall intersection efficiency under coordination is better than free operation for low volume conditions.

5. Stop Rate

In addition to delay, stop rate and stop time are also crucial, especially for the side street vehicles. To simulate vehicle stopping conditions under different volume scenarios, a single-lane minor street was constructed in VISSIM which crosses another single-lane major street. Since a vehicle is assumed to arrive randomly following the uniform distribution during a period (within a cycle or during the whole analysis period), the probability that a vehicle arrives at each timestamp is equal. Therefore, the probability that a vehicle makes or does not make at least a stop that is longer than 20 s can be calculated via equation (1).

To validate whether the calculated stop ratio matches the actual stop ratio, the simulation process was conducted under four volume levels and five cycle lengths. Stop ratios are collected and illustrated in Figures 6, 7, 8, and 9. The solid lines indicate simulation results, and dash lines indicate calculation results from the model.

Results show that stop ratios under different volume levels have a similar increasing trend. In general, the simulated stop ratios are highly consistent with the model estimations. From all figures, it can be found that the minor street stop ratio can maintain a relatively low level (about 20%) when the major street split is less than or equal to 35 s. For minor street efficiency, according to calculated probability and stop ratio simulations, a larger cycle length will

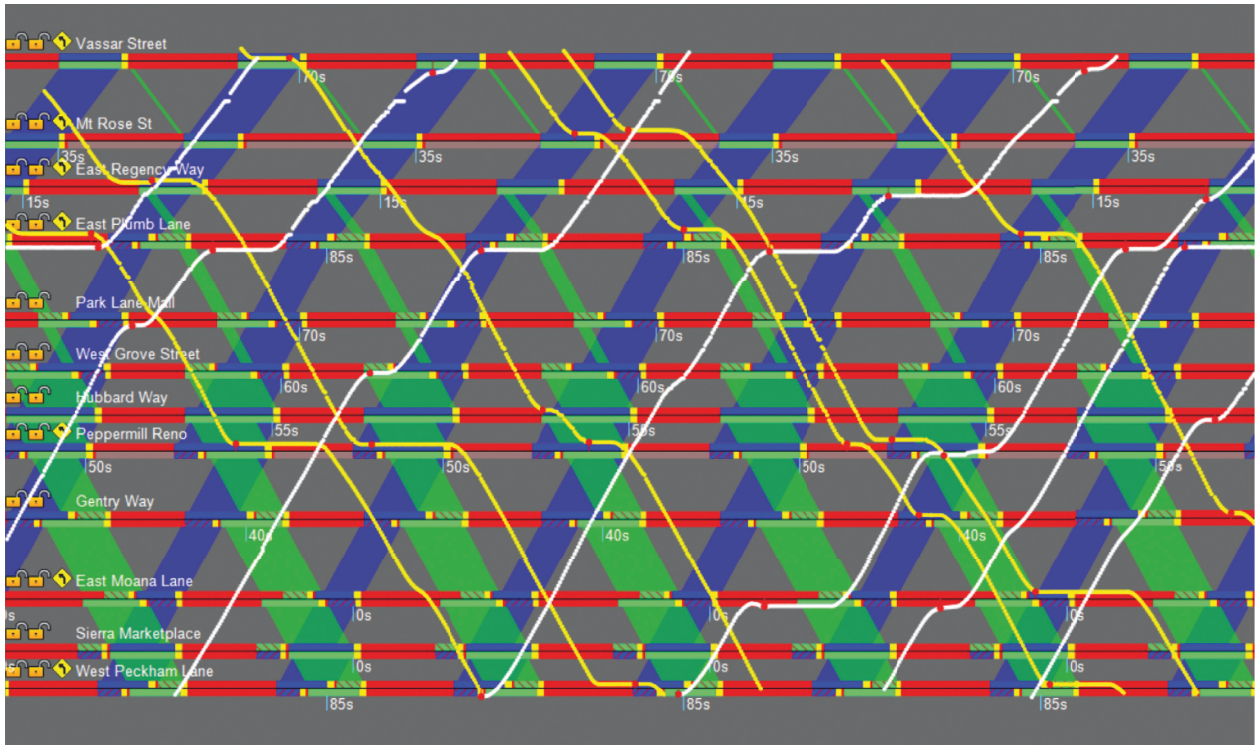


Figure 11: Vehicle trajectories under free operation.

better benefit the minor street. However, according to the North McCarran Blvd practice, with a shorter cycle length, the overall intersection performance becomes better, as indicated in Figures 2 to 5. The reason is that a shorter cycle length could benefit the design of the coordination bandwidth. On the other hand, the reservice strategy has already reduced minor street delays regardless of the cycle length that was used. Based on all the above, using a shorter cycle length is a compromise that displays benefits from each method.

6. Field Validation

To provide a conclusive recommendation, field implementation and validation are necessary. A signal coordination plan for the night condition was developed and implemented along Virginia Street in Reno, Nevada. Twelve intersections were selected between Peckham Lane and Vassar Street, as shown in Figure 10. This segment was chosen due to the relatively significant side street volumes reported that access Midtown and Downtown via this route. Since the number of signals in coordination is more than 10 and the pedestrian movements are relatively heavy, the splits of coordinated phases for these intersections range from 27 s to 56 s; therefore, some of the splits are larger than 35 s. The planned operation time is from 8:00 pm to 10:00 pm. Vehicle trajectory data were collected through TranSync-M software during the operation time. The trajectories of the probe vehicle on the major street under free operation and coordinated operation were extracted, as shown in Figures 11 and 12, respectively. The background time-space diagram (TSD) reflects the proposed signal timing plan, the TSD of the free operation varies by time, in Figure 11, and while the free operation background is kept the same as the proposed plan for the convenience of comparison between their trajectories.

From the figures, most of the trips experienced two or three stops during the data collection period, indicating that, on average, the probe vehicle only stops one time out of five intersections. This suggests the performance of both the free operation and coordination has good results. While it is hard to capture the differences simply from visual observation, to validate the difference, the trajectories information was analyzed by the incorporated performance evaluation algorithm: signal performance index [27]. Table 1 shows the detailed performance parameters under free operation and coordination strategies. Results show that the overall speed increased by approximately 3 mph after the coordination; the improvements are significant in both the southbound and northbound directions. For the number of tops, it is also reduced according to the stop score. The quality of signal timing generally indicates the two directions was also upgraded from B+ to A and from B to A−. Therefore, the coordination plan provided better performance than free operation on the mainline.

Both trajectories and videos of vehicles passing minor streets were collected for evaluating the side street performance. Based on the 100 samples collected from the field, this research found that vehicle waiting times vary from 0 s to 72 s. However, most of the waiting experiences did not make the minor street drivers become impatient since there were vehicles on the major street. Therefore, the maximum times

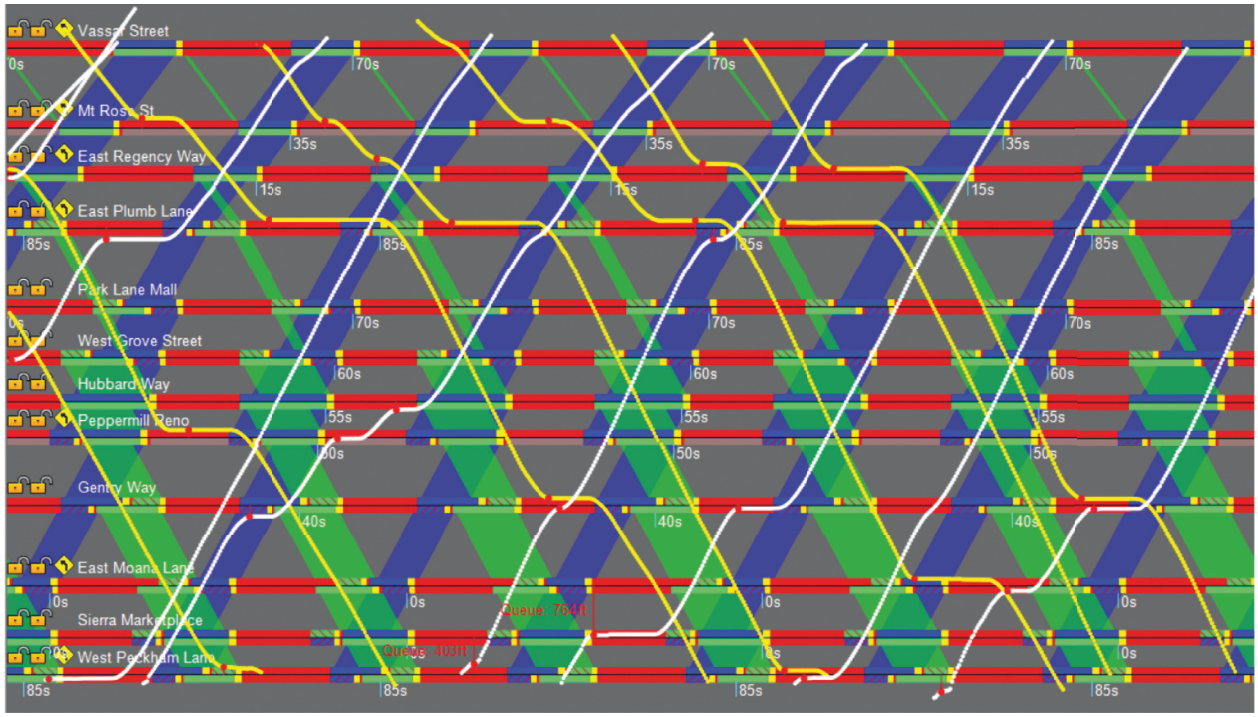


Figure 12: Vehicle trajectories under signal coordination.

Table 1: Performance results under free operation and signal coordination.

Signal strategy	Free operation strategy					Signal coordination strategy				
Performance measure	Avg. speed (mph)	Avg. travel time (s)	Avg. delay (s)	Stop time (s)	Signal quality	Avg. speed (mph)	Avg. travel time (s)	Avg. delay (s)	Stop time (s)	Signal quality
Evening plan avg.	20.6	320	110	82	B	23.4	282	72	53	A
Evening plan NB	20.2	327	116	78	B+	25.1	263	54	35	A
Evening plan SB	21.1	313	103	87	B	21.8	303	92	73	A-

without major street vehicle appearances were collected, and they varied from 0 to 35 s. However, only three samples exceed 20 s, among which two of them had pedestrian calls. Therefore, the minor street driver is less likely to become impatient when the night plan is incorporated if appropriate signal coordination and reservice are incorporated.

7. Concluding Remarks

A critical concern in the implementation of a coordination plan during nonpeak hours is that minor street vehicles may unnecessarily wait for the major street operation to terminate particularly when there are no vehicles on the central street. This research provided a feasible approach to keep the progression on the major street without having the minor street vehicles waiting for a long time. The method is to set the major street splits as small as possible so that minor street vehicles can have a larger possibility to proceed the intersection with an acceptable stop time. The reservice strategy was applied to eliminate the situation where minor street vehicles cannot pass once the early return happens (i.e., the green signal will always hold to the end of coordinated phases). A 20 second threshold was used as the acceptable waiting time, which has been recommended by several professional engineers from the Regional Transportation Commission of Northern Nevada. Engineers may also adopt other values as the threshold according to different regions and driver behaviors and use the proposed model to calculate the appropriate splits corresponding with the desired stop probability. Although pedestrian movement may also contribute to a longer wait time, the delay caused by pedestrians can be tolerated by drivers since this is usually considered as expected delay. In terms of the influence of pedestrian crossing on mainline traffic, it can also be reduced by turning on the stop-in-walk feature in the signal controller. Therefore, in general, the proposed approach solves the problem of unnecessarily minor street vehicles waiting with the preservation of the major street coordination. However, there remain several issues not solved in this research such as the extra waiting time caused by the queue or the pedestrian crossing impact on the coordination. In the current model, extra waiting time was not considered since

the low volume condition is expected to have little probability of encountering the queueing issue; future research may further explore the effects of the pedestrian crossing on major street traffic operations, and the field performance comparison between the traditional coordination plan and the proposed coordination plan could be conducted.

Data Availability

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

Conflicts of Interest

The authors declare there are no conflicts of interest in this research.

Authors' Contributions

RY and ZT contributed to study conception and design; RY, YZ, and GY were responsible for experiment design and data collection; RY and YZ took part in experiment design and data collection; RY, GY, and YY prepared the draft article. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

The authors would like to thank Andrew Jayankura in the Regional Transportation Commission of Washoe County, Nevada, for discussion with the tolerance time and basic signal timing data. This research was funded by the Fundamental Research Funds for the Central Universities 2022RC021.

References

[1] D. A. Bowers, "Progressive timing for traffic signals," *Proc. Inst. Traff. Engrs*, vol. 51, no. 4, pp. 93–100, 1947.

[2] J. T. Morgan and J. D. Little, "Synchronizing traffic signals for maximal bandwidth," *Operations Research*, vol. 12, no. 6, pp. 896–912, 1964.

[3] J. D. C. LittleLittle, "The synchronization of traffic signals by mixed-integer linear programming," *Operations Research*, vol. 14, no. 4, pp. 568–594, 1966.

[4] W. D. Brooks, "Vehicular Traffic Control–Designing Arterial Progressions Using a Digital Computer," *Data Processing IBM*, 1965.

[5] C. J. Messer, R. H. Whitson, C. L. Dudek, and E. J. Romano, "A variable-sequence multiphase progression optimization program," *Highway Research Record*, vol. 445, pp. 24–33, 1973.

[6] J. D. Little, M. D. Kelson, and N. H. Gartner, "MAXBAND: A Versatile Program for Setting Signals on Arteries and Triangular Networks," *Transportation Research Record Journal of the Transportation Research Board*, vol. 795, pp. 40–46, 1981.

[7] N. H. Gartner, S. F. Assman, F. Lasaga, and D. L. Hou, "A multi-band approach to arterial traffic signal optimization," *Transportation Research Part B: Methodological*, vol. 25, no. 1, pp. 55–74, 1991.

[8] M. Abbas, D. Bullock, and L. Head, "Real-time offset transitioning algorithm for coordinating traffic signals," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1748, no. 1, pp. 26–39, 2001.

[9] Z. Tian and T. Urbanik, "System partition technique to improve signal coordination and traffic progression," *Journal of Transportation Engineering*, vol. 133, no. 2, pp. 119–128, 2007.

[10] Z. Tian, V. Mangal, and H. Liu, "Effectiveness of lead-lag phasing on progression bandwidth," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2080, no. 1, pp. 22–27, 2008.

[11] Z. Z. Tian, T. Urbanik, R. Engelbrecht, and K. Balke, "Pedestrian timing alternatives and impacts on coordinated signal systems under split-phasing operations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1748, no. 1, pp. 46–54, 1748.

[12] Z. Yang, P. Liu, Z. Z. Tian, and W. Wang, "Effects of left-turn waiting areas on capacity and level of service of signalized intersections," *Journal of Transportation Engineering*, vol. 139, no. 11, pp. 1076–1085, 2013.

[13] Z. Tian, C. Messer, K. Balke, and T. Urbanik, "Integration of diamond interchange and ramp metering operations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1925, no. 1, pp. 100–111, 2005.

[14] R. Yue, G. Yang, Z. Tian, H. Xu, D. Lin, and A. Wang, "Microsimulation analysis of traffic operations at two diamond interchange types," *Journal of Advanced Transportation*, vol. 2019, pp. 1–11, 2019.

[15] L. Wan, C. Yu, L. Wang, and W. Ma, "Identification of time-of-day breakpoints based on trajectory data of probe vehicles," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 5, pp. 538–547, 2019.

[16] B. B. Park and J. Lee, "A procedure for determining time-of-day break points for coordinated actuated traffic signal systems," *KSCE Journal of Civil Engineering*, vol. 12, no. 1, pp. 37–44, 2008.

[17] B. B. ParkPark, P. Santra, I. Yun, and D. H. Lee, "Optimization of time-of-day breakpoints for better traffic signal control," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1867, no. 1, pp. 217–223, 2004.

[18] Minnesota Department of Transportation, *MnDOT Traffic Signal Timing and Coordination Manual*, Minnesota Department of Transportation, Minneapolis, MN, USA, 2013.

[19] A. F. Abdelghany and B. Connor, *Guidelines for Operating Traffic Signals during Low-Volume Conditions*, vol. 5, Washington, DC, USA, 2006.

[20] J. Bonneson, M. Pratt, and K. Zimmerman, "Development of a Traffic Signal Operations Handbook," *Texas Transportation Institute*, TX, USA, 2009.

[21] C. J. Messer and R. V. Nageswara, *Improved Traffic Signal Coordination Strategies for Actuated Control*, Southwest Region University Transportation Center, Center for Transportation, University of Texas, USA, 1996.

[22] R. Andalibian and Z. Tian, "Signal Timing and Coordination Strategies under Varying Traffic Demands," *Nevada. Dept. of Transportation*, vol. 236, pp. 11–803, 2012.

[23] R. Andalibian, *Volume-based Probabilistic Approaches to Determining when to Turn on and off Signal Coordination Plans*, TRB 2016, Washington, DC, USA, 2015.
[24] Washington Department of Transportation, *Signal Timing Review and Optimization*, 2019.
[25] City of Laguna Niguel, *Traffic Signal Timing and Coordination FAQ's*, Laguna Niguel, CA, USA, 2021.
[26] R. Yue, G. Yang, Y. Zheng, Y. Tian, and Z. Tian, "Effects of traffic signal coordination on the safety performance of urban arterials," *Computational Urban Science*, vol. 2, no. 1, p. 3, 2022.
[27] R. Yue, G. Yang, D. Lin, A. Wang, and Z. Tian, "Traffic signal retiming to improve corridor performance," *Journal of Transportation Engineering, Part A: Systems*, vol. 147, no. 1, Article ID 05020009, 2021.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 8970695, 16 pages
https://doi.org/10.1155/2022/8970695

Research Article

Real-Time Incident-Responsive Signal Control Strategy under Partially Connected Vehicle Environment

Kancharla K. K. Chandan,[1] **Álvaro J. M. Seco**,[2] **and Ana M. C. Bastos Silva**[2]

[1]*Centre for Territory, Transport and Environment (CITTA), Department of Civil Engineering, University of Coimbra, Coimbra 3030-788, Portugal*
[2]*CITTA, Department of Civil Engineering, University of Coimbra, Coimbra 3030-788, Portugal*

Correspondence should be addressed to Kancharla K. K. Chandan; kamalkeerthichandan@gmail.com

Received 19 October 2021; Revised 29 April 2022; Accepted 23 May 2022; Published 22 July 2022

Academic Editor: Yanyan Qin

The performance of the traffic system can drastically drop when nonrecurrent congestion caused by incidents occurs. Early detection and clearing of traffic incidents will enable the mitigation of the congestion and early restoration of normal traffic conditions. The research in this paper utilized the vehicle information from the recent technological advancement in transportation systems, connected vehicles (CV), and loop-detector information for nonconnected vehicles (NCVs) and developed a novel algorithm to (1) control traffic signals for normal traffic conditions in the absence of incidents, (2) detect traffic incidents using CV/NCV information, and (3) control traffic signals during the occurrence and dissipation of incidents. All the 3 strategies were integrated into one algorithm, which runs as per the real-time traffic conditions, in the presence or absence of incidents. Space-mean speeds of the vehicles on nonincident lanes and throughput maximization criteria were taken as the indicators for the activation of specific signal timings directed at the incident-affected approach. Diverse incident scenarios were tested on a four-legged isolated intersection using the VISSIM simulation tool. Incident detection results showed a higher detection rate and lower mean detection time at higher CV penetration and higher traffic volumes, and at the incident locations nearer to the stop-line. The proposed incident-responsive signal control strategy at 40% and higher CV penetration showed better performance over EPICS adaptive signal control solution, in reducing average travel time delay and the average number of stops per vehicle.

1. Introduction and Background

Nonrecurrent congestion caused by traffic incidents remains a critical problem in urban areas. Traffic incidents such as vehicle collisions, vehicle breakdowns, unscheduled road maintenance, and illegally parked vehicles temporarily disrupt the traffic flow on the roadway, resulting in frustrating delays to the road users and even extending the queues to the neighbouring roads if not compensated in a timely fashion [1–3]. More than 25% of congestion is caused by nonrecurring incidents that block a part of the road section from use [4, 5]. Prompt and accurate detection of traffic incidents in real time can lead to mitigation of the congestion and early restoration of normal traffic operations of the affected roadway.

Usually, traffic on arterial streets features a greater dynamic and is subject to greater disruptions due to the influence of traffic signal control, pedestrian crossings, parking manoeuvres, bus stops, road works, etc., which can create incident-like traffic patterns [6, 7]. Due to these factors, from the past three decades, research has been dedicated to developing incident detection algorithms specifically for signalized arterials. With the availability of new vehicle detection technologies, automatic incident detection algorithms (AIDA) were developed in recent years. AIDAs use real-time traffic data, from traffic sensors such as inductive loop detectors (ILD), video-based sensors, and probe vehicles to automatically detect incidents, when abnormal changes in traffic data are identified [8].

Ahmed and Hawas [7] developed a general linear regression model, which uses a threshold value to predict the incident status. If the estimated dependent variable is higher than the threshold value, an incident is indicated. Results showed that the detection rates and false alarm rates ranged from 23% to 87% and 0% to 20%, respectively. Ghosh and Smith [9] customized three artificial neural network (ANN)-based and one support vector machine (SVM)-based freeway algorithms and applied them to signalized arterial roads to detect incidents. Volume, speed, and occupancy data from ILDs were used as input parameters to the four chosen AIDAs. Asakura et al. [10] used travel time data from probe vehicles to predict the time and location of traffic incidents, using shockwave theory. Yu et al. [11] developed a comparative incident detection algorithm using arterial travel time data from a Bluetooth-based vehicle re-identification system and one year of corresponding incident data. Historical travel time data were compared with real-time samples and unusual deviations from normal traffic behaviour were used to identify potential incidents. Oskarbski et al. [6] modified an existing highway AIDA (TRISTAR) for signalized intersections. Travel times obtained from Bluetooth and Wi-Fi detectors were compared with historical reference values, and alerts were raised using an algorithm based on a Kalman filter when significant differences were identified. Ren et al. [12] proposed a video-based method to detect and locate an incident on a road segment. Each lane was divided into a cluster of cells, and traffic states in the cells were judged by a fuzzy-identification method, and then, an SVM classifier was applied to the congested cells to detect an incident and locate its position. Gu et al. [13] combined the data sources from Twitter and incident records and employed semi-Naive Bayes classification to detect five major types of incidents. D'Andrea and Marcelloni [14] detected traffic incidents using real-time data from GPS trackers and drivers' smartphones. Initially, a traffic state was assigned to each road segment based on the speeds of the vehicles, and then, based on the spatiotemporal analysis of the road segments, traffic state (e.g., incident, slowed traffic, and blocked traffic) and the estimated speed of vehicles were determined. Results showed a detection ratio of 91.6% and an average detection time of 7 minutes. Using data from ILDs, Rizvi et al. [15] proposed a data-driven framework to detect traffic incidents in real time and to estimate road capacity and incident location. Incident was detected when variation in speed and occupancy of the traffic flow exceeds predefined threshold limits. Sheikh et al. [16] proposed an AIDA based on a lane-changing speed mechanism for highway traffic using V2I communication. When the average change in vehicle speeds during the lane-changing process takes longer than the time taken during nonincident conditions, an incident was assumed to be detected. Simulation results showed 30% faster detection compared with other methods. Zaitouny et al. [17] proposed a Quadrant Scan methodology as a tool to analyse traffic volume data to detect incidents by integrating multiple sensors' data. The proposed method also distinguishes nonrecurrent traffic congestion caused by incidents from recurrent congestion. Liang et al. [18] used surveillance video stream to detect traffic objects and proposed an algorithm to detect traffic incidents based on a spatiotemporal map of vehicle trajectories. According to the vehicle trajectory and vehicle position in each frame, the vehicle is re-identified across frames to associate the same vehicle between different frames and a global spatiotemporal map of the trajectory was generated under the current road segment. Traffic incidents such as traffic congestion, vehicle speeding, and illegal parking were analysed using the trajectory state.

Besides detecting incidents in real time, a few researchers have focussed on methods to curtail incident-induced congestion through traffic signal control, which is essential for an effective incident management system. Sheu [3] developed a stochastic optimal control-based algorithm to tackle lane-blocking incidents at isolated intersections. The proposed method controlled traffic signal timings by minimizing a time-varying cost function, which was measured based on comparing the real-time estimates of interlane and intralane traffic states with their ideal values. Lu et al. [19] developed an incident-responsive optimal signal control solution using historical and real-time volume and occupancy data collected from the sensors. Total delays around the incident location were minimized by the optimization model. Simulation results from VISSIM showed a 39.5% reduction in delay under heavy flow situations. Long et al. [20] extended a cell transmission model (CTM) and applied it to simulate incident-based jam propagation in two-way rectangular grid networks. Four control strategies were proposed based on the concept of a vehicle movement ban for dispersing incident-based traffic jam. Qi et al. [1] developed a traffic-light control system at a signalized intersection to reduce incident-based traffic congestion. By adopting additional traffic warning lights, ban signal and warning signal strategies were used to disperse the traffic congestion. Timed Petri nets were used to describe the cooperation between traffic lights and warning lights. Mao et al. [21] developed a signal control optimization method for urban networks affected by incidents. Genetic algorithm (GA) was used to minimize the total travel time of a four-intersection testbed network. Simulation results in AIMSUN showed a 40.7% reduction in total travel time compared with the incident scenario without GA traffic control optimization. Wang et al. [22] proposed a traffic signal optimization strategy to maximize the throughput of a two-phased signal control isolated intersection while restricting the queue length caused by traffic incidents. Numerical results showed significant improvement in intersections' throughput when compared to a fixed-time control. Yao and Chen [23] proposed an adaptive traffic signal control strategy that responds to traffic disruptions at an isolated intersection. Dynamic phase selection is applied to adjust the traffic signal control plan adaptively during the incident stage, while the queue length dissipation algorithm is adopted to carry out optimal green time calculation during the incident recovery stage. The proposed signal strategy was found to improve the resiliency of a typical intersection against disruptions by clearing queues faster and reducing overall traffic loss time over conventional fixed and actuated traffic signal plans.

In summary, most of the incident-related state-of-the-art literature work was focussed on developing the methods to identify the occurrence and location of incidents using fixed sensors or probe vehicle data. Quite limited research was available on real-time incident-responsive signal control for isolated intersections. Furthermore, there seems to be an opportunity for testing the potential of the latest technological advancements such as connected vehicles (CV) to tackle incident-related traffic problems.

In recent years, CV technology has been getting attention as a step towards the next generation's transportation system. Traffic signal systems under CV environment can use wireless information transactions between vehicles (V2V), vehicles and infrastructure (V2I and I2V), and vehicles and handheld devices (V2D), which are collectively referred to as V2X (vehicle to anything) communication, thus enabling the access to detailed and instantaneous vehicle information such as its speed and location [24–26]. Such real-time information has the potential to be used to detect incidents in a timelier fashion than traditional detection techniques and to support the design and operation of signal control strategies to efficiently mitigate incident-based congestion. The current main drawback of this technology is that its low penetration market makes it less beneficial to exchange CV information. According to the forecasts of Capgemini Invent [27], in 2023, 352.9 million connected cars will be on the road, which corresponds to 24% of all cars globally. At 97%, the United States is anticipated to have the highest penetration rate of CVs in 2035 [28]. Therefore, it is a forthcoming challenge to utilize this emerging technology in its early developmental stage to improve the existing transportation systems, during a time when the connected vehicle penetration takes more than a decade to reach 100% [24].

Recently, Chandan et al. [29] proposed a connected vehicle-based signal control algorithm (CVSC) for an isolated intersection, which utilized the speeds and positions of CVs, assuming 100% connected vehicle penetration (CVP). Furthermore, to tackle the real-life situations where the CVP is less than 100%, Chandan et al. [30] also presented an algorithm that utilized both ILD and CV information, to estimate the speeds and positions of the nonconnected vehicles (NCV) (those not equipped with V2X communication devices), and then fed the estimated NCV data to the CVSC algorithm, to adjust the traffic signal timings dynamically and more efficiently.

In this research work, a novel algorithm is proposed for isolated intersections that integrate a signal control strategy for normal traffic (nonincident) conditions, an incident detection method, and a traffic signal control strategy during the occurrence and dissipation of incidents. The integrated algorithm is also a distinctive feature of the subject area, since, to the best of the authors' knowledge, studies tackling both the aspects of incident detection and incident-responsive signal control were not found in the literature. The objectives of the research presented in this paper are threefold: (1) improve the signal control strategy (for non-incident conditions) developed by Chandan et al. [29] with minor upgrades, (2) develop an incident detection technique using real-time CV/NCV information, and (3) develop an incident-responsive signal control strategy under a partially CV environment. Further, we evaluate the incident detection accuracy and the performance of the incident-responsive signal control strategy on an isolated intersection using the VISSIM simulation tool.

2. Incident Types and Durations

Traffic Incident Management Handbook [31] defines an incident as "any nonrecurring event that causes a reduction of roadway capacity or an abnormal increase in demand." "Under this definition, events such as traffic crashes, disabled vehicles, spilled cargo, highway maintenance and reconstruction projects, and special nonemergency events (ball games and concerts) are classified as incidents" [32]. Most researchers obtain incident data from various traffic incident record systems from both freeways and arterial roads, where different types of incidents and their durations were recorded. Studies on freeway incidents include various incident types, such as accident, breakdown, and debris [33]; accident and disabled vehicles [34]; accident, congestion, and reckless driving [35]; collision, debris, disabled vehicle, and abandoned vehicle [36, 37]; and crashes, hazard, and stationary vehicles [38, 39].

Compared with freeway traffic, traffic on arterial roads features greater dynamic and disruptions due to the influence of signalized intersections, conflicting movements among left and right and through traffic, pedestrian crossings, frequent vehicle manoeuvres (lane change and on-street parking), public transport stops, road maintenance works, recurrent or nonrecurrent queues, loading/unloading stops, and traffic signal malfunction [6, 11, 40]. Some studies have analysed the influence of on-street parking on traffic performance [41, 42]. On-street parking manoeuvres can often block traffic for short time causing temporary bottlenecks [43] or sometimes create start-stop traffic flow behaviour on the lanes adjacent to the parking lane [44] and can reduce the road capacity by 20–30% [45]. Some studies have also analysed the impacts of roadside bus stops, which can also disrupt the traffic flow at signalized intersections [46, 47]. Buses dwelling at the bus stops to load and unload passengers may occupy a portion of a traffic lane or a bus bay and while moving into or out of the bus stop can block the traffic flow on the adjacent lane, causing excessive delays to the road users and reduction in the capacity of the intersection [13, 48–50]. Therefore, such events, which cause minor disruptions to traffic flow, are also considered as incidents in this paper.

The duration of incidents is a major factor affecting nonrecurring congestion. According to Li et al. [51], the traffic incident duration time is "the time difference between the occurrence of an incident and the clearance of the incident site." According to Highway Capacity Manual, the total incident duration is divided into four distinct time intervals, namely, detection, response, clearance, and recovery times [52, 53]. Many researchers have analysed the duration of incidents (obtained from various incident record systems) in their case studies on both freeways and urban

arterials. Few studies include total incident duration, while others focussed on clearance time, a combination of response and clearance times. From the vast literature, several relevant studies corresponding to various incident types and durations are presented in Table 1.

The incident durations from various databases varied from 2 seconds (parking manoeuvre) to 120 minutes (crashes), indicating a wide distribution of traffic disruption times. Hence, the research presented in this paper developed (1) an incident detection algorithm that can detect any type of incident with any incident duration that can cause minor or major disruptions to traffic flow and (2) a CV-based traffic signal control strategy to respond to any kind of traffic incident.

3. Modelling Approach

Real-time CV information such as speeds and positions of individual vehicles is a valuable resource to detect incidents upstream of the traffic signals and develop better signal control strategies during such situations. However, the current real-life CV penetration rate will take some years to reach 100%. Under such low penetration situations, it is, thus, important to be able to detect the incidents as early as possible after their occurrence and to act under that knowledge, to minimize its impact. This section presents strategies to detect incidents and control the traffic signal times in real time, during the presence and absence of incidents. The modelling approach is divided into 3 parts, namely, incident absence connected vehicle signal control strategy (IA-CVSC), identification method of incidents' occurrence and clearance, and incident-responsive connected vehicle signal control strategy (IR-CVSC).

3.1. Incident Absence Connected Vehicle Signal Control Strategy (IA-CVSC). In this section, a CV-based signal control strategy is presented for normal traffic without incidents (therefore, the acronym IA-CVSC). The strategy is adopted from Chandan et al. [29] with minor upgrades as presented below. The objective of this strategy was to develop a traffic signal timing algorithm for an isolated intersection, which minimizes average delay and stops, considering the arrival and departure flows of all vehicles on all the approaches of the intersection at every time-step. This strategy utilizes CV data, assuming that all vehicles would share their information (such as acceleration, speed, and location) at every time-step to the signal controlling algorithm, through V2X communication devices. The algorithm obtains CV data at every time-step, from 300m before the stop-line (which is within the DSRC range of 100 m–1000 m [58]) until it crosses the stop-line. For real-world situations, where CVP is less than 100%, NCV information can be obtained through ILDs that typically exist at the upstream of an intersection. The forward movement of NCVs towards the intersection can be estimated using Gipps' car-following model, assuming that the NCVs do not change their lanes. Utilizing the CV and ILD information, Chandan et al. [30] developed a method to estimate the speeds and positions of NCVs at signalized intersections using Gipps' car-following model and it is adopted in the current modelling approach. Detectors can be placed within 100–300 m from the stop-line depending on the general queue length of the intersection, but the closer to the stop-line they are placed, the less the occurring lane changes, and, thus, the better the NCV information estimation accuracy. Here, ILDs are placed on each lane at 150 m before the stop-line. The IA-CVSC strategy (see Figure 1) is explained step-wise as below.

Step -3.1.1: CV and NCV data collection—obtain and update the arrival and departure information of CVs and NCVs from V2X devices and ILDs, respectively.

Step -3.1.2: NCV information estimation—estimate and update the forward movement of NCVs (speed and position) applying Gipps' car-following model.

Step -3.1.3: Speed ratio check—during the green phase, initially green time is provided (limiting to maximum green) until the space-mean speed of all vehicles in the current green phase reaches 90% of their desired speed. This is defined by a metric called speed ratio in equation (1). The average desired speed of the vehicles is to be calibrated in advance for any intersection. Speed ratio check ensures that all the vehicles that were in queue during the red interval are served, and the approaching vehicles are moving close to their free-flow speed, at which time the green phase is re-evaluated.

$$\text{Speed ratio} = \frac{\text{Actual space mean speed of all vehicles at current time step} \times 100}{\text{Desired space mean speed of all vehicles at current time step}}. \quad (1)$$

note that the maximum green time is taken as 1.3 times the average actual green time of the past five cycles [59]. This factor accommodates well most fluctuations in vehicle arrival rates.

Step -3.1.4: Reserve-time and throughput ratio—At the time-step when speed ratio reaches 90%, the strategy aims at maximizing the throughput of the intersection considering the cumulative arrival and departure flows on all the approaches of the intersection. This is done during the time that is remaining between the current time-step and the maximum green period, termed as reserve-time period, as shown in equation (2). Throughput ratio is calculated as the ratio of the sum of departure flows of all the phases to the sum of arrival flows of all the phases, as shown in equation (3). Throughput ratio minimizes the differences between the arrival and departure flows on all approaches of the intersection, by maintaining its value as near as possible

Table 1: Studies on incident types and duration time analysis.

Study	Incident type	Duration
Alkaabi et al. [54]	Accidents	Clearance time—26 minutes
Ahmed and Hawas [7]	Simulated incidents	Incident duration—6, 8, and 10 minutes
Hou et al. [55]	Disabled vehicles, abandoned vehicles, debris, collision, and others	Response time—9, 7, 8, 8, and 8 minutes, respectively
Tavassoli Hojati et al. [38]	Crash, hazard, and stationary vehicle	Incident duration—120, 103, and 76 minutes, respectively
Ghosh et al. [56]	Abandoned vehicle, flat tire, out of gas, clearing debris, directing traffic, towing, mechanical problems, and others	Clearance time—9 minutes
Ding et al. [57]	Disabled vehicles, abandoned vehicles, debris, collision, and others	Response time—10 minutes; clearance time—28 minutes
D'Andrea and Marcelloni [14]	Simulated incidents	Incident duration—15, 20, and 30 minutes
Shang et al. [33]	Accidents, breakdowns, and debris/pedestrian	Incident duration—27, 25, and 10 minutes, respectively
Tang et al. [37]	Disabled vehicles, abandoned vehicles, debris, and others	Clearance time—13 minutes
Chavis and Christofa [48]	Bus stop	Dwell time—40 seconds (mean); 30 seconds (standard deviation)
Ghasemlou et al. [50]	Bus stop	Dwell time—20, 30, and 40 seconds
Liang and Ma [46]	Bus stop	Dwell time—10 to 20 seconds
Liu and Jian [47]	Simulated incident—bus stop	Dwell time—35 to 55 seconds
Cao and Menendez [43]	Parking manoeuvre	Incident duration—2 to 32 seconds

to 1.0. Since throughput ratio considers the combined effect of vehicle arrivals and departures on all approaches of all phases, if the green time is extended (limiting to the maximum green period) as long as the throughput ratio is nearer to 1.0, the performance of the entire intersection can be improved reducing the congestion.

$$\text{Reserve} - \text{time} = \text{Maximum green time} - \text{Time when the speed ratio reaches 90\%}, \tag{2}$$

$$\text{Throughput} - \text{ratio} = \frac{\sum_{p=1}^{P} (\text{Cumulative departure flow})_p}{\sum_{p=1}^{P} (\text{Cumulative arrival flow})_p}, \tag{3}$$

where p = phase, P = total number of phases.

Step -3.1.5: Green phase termination check—during the reserve-time period, using the updated CV and NCV information, throughput ratio is calculated as a rolling horizon at every time-step and the green time is extended until the throughput ratio values stop increasing in comparison with the previous one or until the maximum green period is reached, whichever is earlier. Throughput ratio stops increasing as the arrival flow builds up in the other phases, at which point the current green phase is terminated and switched to the next phase.

3.2. Incidents' Occurrence and Clearance Identification Method. When an incident occurs on a lane, vehicles on that lane queue up behind the incident location, and the queue length increases, and the average speed of the vehicles on that approach decreases, or sometimes the vehicles come to a halt. If a halted vehicle does not move from its position after some threshold time, a potential incident can be supposed. During the red interval, as all the vehicles come to a halt, it is difficult to identify if the vehicles stopped due to the influence of an incident or the red signal. Hence, the incident's occurrence is identified from the start of the green interval. The step-wise incident detection procedure (see Figure 1) is explained below.

Step -3.2.1: CV and NCV information at halt—at every time-step from the beginning of the green phase, we obtain the positions of all the CVs whose speeds are 0 kmph, in each lane, and those of NCVs, if stopped on ILDs with 0 kmph.

Step -3.2.2: Incident identification—we calculate the time required for the halted CVs/NCVs to cross the stop-line, as below

$$T = L + h \cdot N, \tag{4}$$

where T is the time required for the Nth vehicle in the queue to cross the stop-line, L is the start-up lost time, and h is the saturation headway in seconds. The values

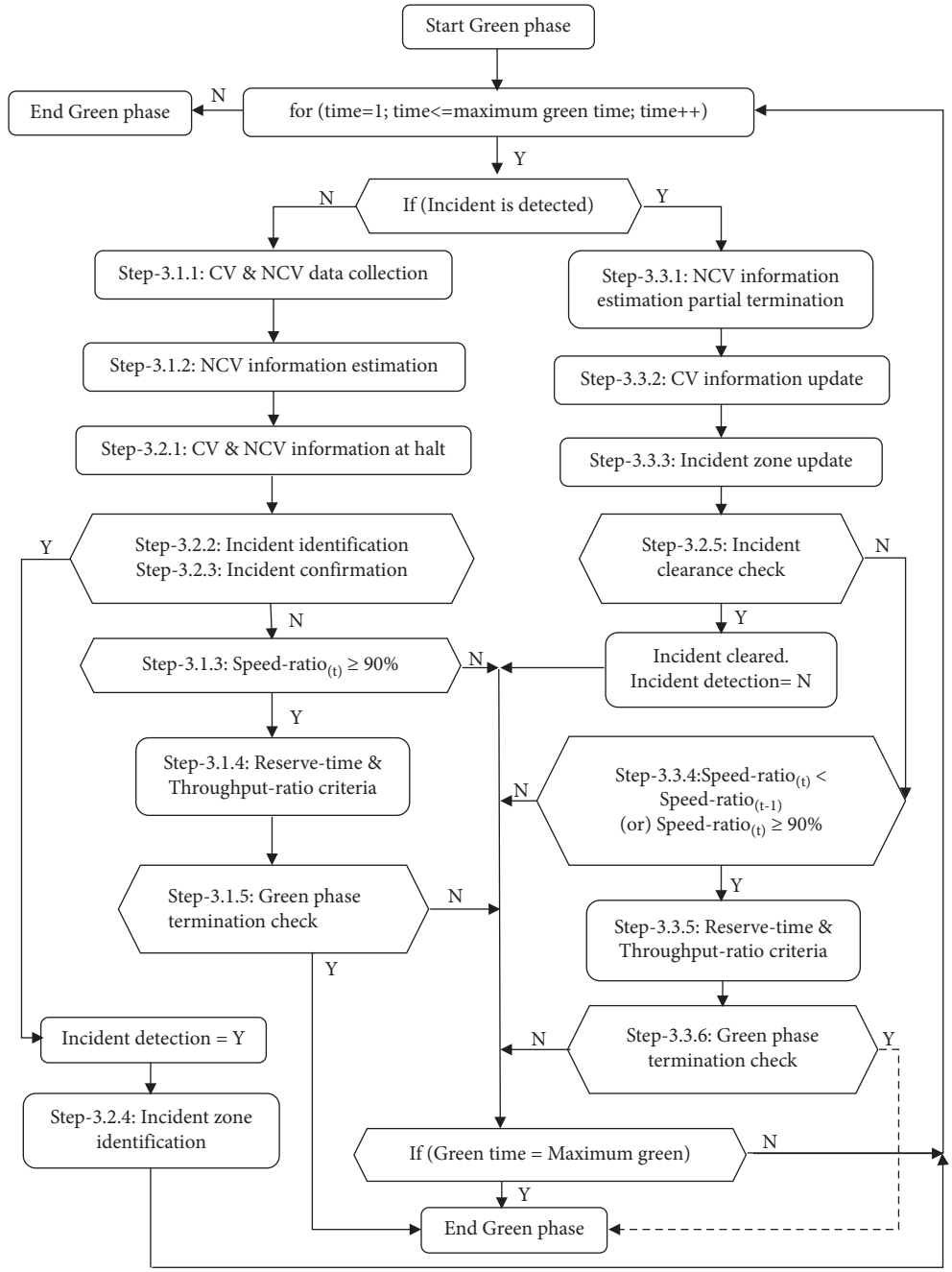


Figure 1: Integrated algorithm of IA-CVSC, incident detection, and IR-CVSC strategies.

of L and h are to be calibrated in advance for any intersection.

It is assumed that if a vehicle at a halt (during red or green interval), which can take at least T seconds to cross the stop-line, fails to move from its position within some threshold time (to be calibrated based on real-world data), incident's occurrence can be presumed on the corresponding lane. Here, a threshold time of $T/3$ seconds is assumed as an upper limit for a CV/NCV to move from its stopped position, if the traffic is normal without any incident. A longer threshold time implies higher incident detection time and vice versa, for a shorter threshold time. The halted CV's position (or NCV's position on an ILD) is checked for its forward movement, at every time-step until the threshold time irrespective of the signal state (green/amber/red). Figure 2 shows the movements and positions of different vehicles considered during the incident identification process.

If an incident occurs during the amber or red interval, the algorithm continues to obtain the information of all the vehicles and updates their positions at every time-step. During these intervals as all the vehicles slow down and come to a halt, based on the updated vehicles'

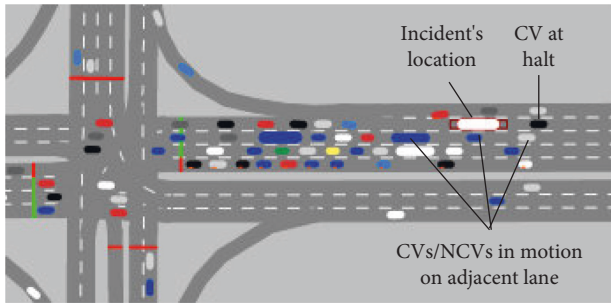


Figure 2: Vehicles' movements and positions during an incident.

positions, the algorithm calculates the stop-line crossing time (T) of the halted vehicles. As the halted vehicles start to move from the beginning of the green phase, the algorithm applies the threshold-time ($T/3$) criteria to check for the incidents' occurrence. If the incident occurs in the middle or end of the green interval, threshold-time criteria are applied from the time-step when a vehicle comes to a halt until the threshold time is reached irrespective of amber/red interval.

Step -3.2.3: Incident confirmation—this step is used to endorse the confirmation of the incidents' occurrence in the previous step, as well as to reduce the false alarm rate. If a CV/NCV at a halt (on the incident lane) fails to move from its position within the predefined threshold time, and if there is at least one CV (beside or behind the subject CV/NCV at the halt) in motion on the adjacent lane (see Figure 2), incidents' occurrence is confirmed.

Step -3.2.4: Incident zone identification—in this step, the proximity of the incident location, referred to as the "incident zone," is identified. Generally, when an incident occurs, the vehicles behind the incident location move to the adjacent lane and travel towards the intersection either on the adjacent lane or by moving back to the incident lane after crossing the incident's location. Therefore, a certain portion of the incident-affected lane remains unused due to vehicles' lane changes or traffic diversions by the incident response teams. The unused portion of the traffic lane near the incident location is referred to as the incident zone (see Figure 3). In other words, the incident zone is the empty portion or gap in front of the incidents' location, where vehicles do not traverse due to the influence of the incident.

To identify the incident zone in real time using CV information, first, the portion of the incident-affected lane that is between the stop-line and the location of the subject CV/NCV (through which incidents' occurrence was identified) is divided into segments of 1-metre length. Second, at every time-step, with the updated location information of all CVs, each 1-metre segment is checked for the presence of vehicle movement. The length of the road segments, which are not traversed (between the stop-line and the location of the subject CV/NCV) by any vehicle, is taken as the updated incident zone.

Step -3.2.5: Incident clearance check—as more vehicles traverse the road segments in the incident zone, the length of the incident zone decreases (see Figures 3(a) and 3(b)). At every time-step, the incident zone length is updated based on the updated CV's location and the length of the road segments traversed. An incident is cleared, when all the 1-metre road segments in the incident zone are traversed by the CVs.

3.3. Incident-Responsive Connected Vehicle Signal Control Strategy (IR-CVSC). When an incident occurs on a lane, all the vehicles behind the incident's location gradually slow down and come to halt, while the vehicles on the adjacent nonincident lanes continue to move to cross the intersection (see Figure 4). During the green phase, the platoon of halted vehicles on the incident lane try to change their lane as soon as they find a gap on the adjacent lane and then pick up their speeds to cross the stop-line (see Figures 5(a) and 5(b)). This lane-changing pattern can decrease the space-mean speed of the vehicles on the nonincident lane. The time taken by the halted vehicles behind the incident location during the process of lane-changing and attaining the speed to cross the stop-line can increase the delay time of the vehicles that are waiting for the green signal on the other approaches. At that situation, the green time for the incident phase can be terminated and the traffic of the next phase can be served. Hence, for the incident-occurring phase, the space-mean speed of the vehicles on the nonincident lane and the ones in front of the incident's location is taken into consideration for the extension/termination of green interval. The step-wise signal control procedure of the IR-CVSC strategy (see Figure 1) is explained below.

Step -3.3.1: NCV information estimation partial termination—during the incident, as the queued-up vehicles behind the incident location try to change their lane, the NCV information estimation is aborted for the incident-occurring phase, as the car-following algorithm does not predict the lane changes of the NCVs, and it is resumed only after the incident is cleared. However, the IA-CVSC strategy with the NCV's information estimation procedure is continued to be in effect for all the other phases that are not affected by the incident.

Step -3.3.2: CV information update—we obtain and update the arrival, departure, speed, and position information of CVs on each lane of the incident approach, using V2X devices.

Step -3.3.3: Incident zone update—from the updated CV's position information, the length of the incident zone is updated (as in Step -3.2.4).

Step -3.3.4: Speed-ratio check—for the incident's phase, initially green time is provided until the vehicles in front of the incident zone and the ones on the adjacent (nonincident) lane are served. During this period, the halted vehicles behind the incident location try to find a gap on the adjacent lane and interrupt the traffic while entering the lane (see Figure 5). In this situation, the space-mean speed of the vehicles on the nonincident

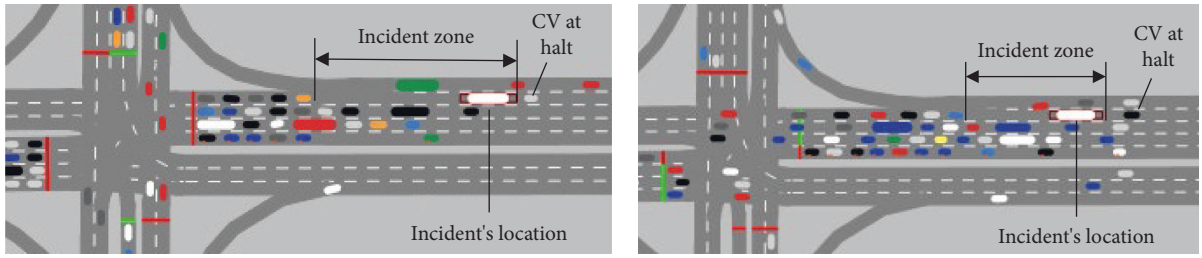
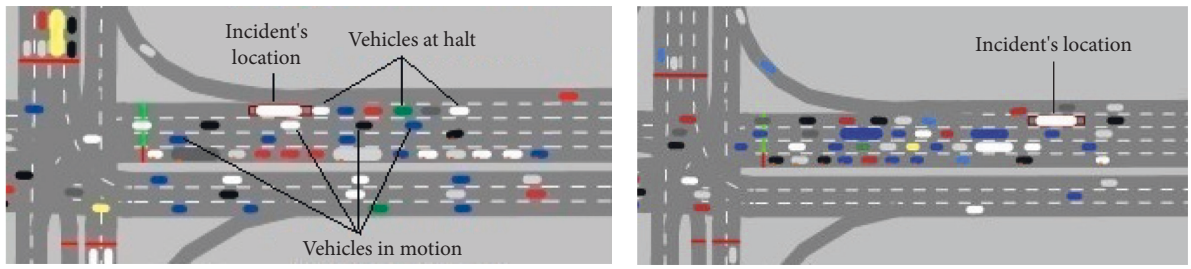
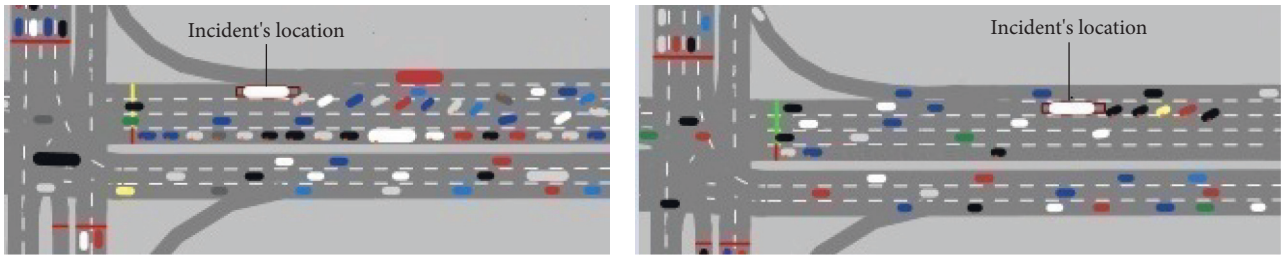


Figure 3: Depiction of the incident zone.



(a) (b)

Figure 4: Traffic flow depiction during the occurrence of an incident.



(a) (b)

Figure 5: Vehicles' lane-change behaviour and during an incident.

lane decreases, indicating the need to re-evaluate the current green phase. During undersaturated traffic conditions, the space-mean speed of the vehicles on the nonincident lane might not get affected when the halted vehicles change their lane.

In another situation, the halted vehicles behind the incident location might not be able to find a gap due to a continuous demand on the nonincident lane (see Figure 4). In this situation, the space-mean speed of the vehicles on the nonincident lane increases, indicating the need for the green time extension.

Therefore, to tackle the above situations, green time is provided until the speed ratio of the vehicles on the nonincident lane decreases when compared to the previous time-step or the speed ratio reaches 90%, at which point, the algorithm moves to the next step.

Step -3.3.5: Reserve-time and throughput ratio—depending on the cumulative arrival and departure flows on all the approaches, throughput ratio criteria define whether to extend or not further the green period. As in the IA-CVSC strategy, reserve-time is calculated, and throughput ratio criteria are applied as rolling-horizon at every time-step to extend the green time until the throughput ratio no longer increases in comparison with its value in the previous time-step.

Figures 5(a) and 5(b) show the lane-changing behaviour of the halted CVs when they find a gap on the nonincident lanes. The difference between figures is that the throughput ratio criteria in Figure 5(a) led to the termination of the green phase (therefore, the signal light is amber) as more vehicles are waiting for the green in other phases when compared to the incident's phase, while in Figure 5(b), the situation led to the extension of green (therefore, the signal light is green) until throughput ratio value stops increasing compared with the previous time-step.

Step -3.3.6: Green phase termination check—when the throughput ratio reaches its maximum value, the green

period for the incident phase is terminated and switched to the next phase.

It should be noted that the green time extension at every time step is subject to the maximum green time limit in both IA-CVSC and IR-CVSC strategies. The difference between the signal control criteria of IA-CVSC and IR-CVSC strategies is that IA-CVSC provides green time until the speed ratio of all the CVs and NCVs reaches 90% and then applies throughput ratio criteria to extend or terminate the green phase, whereas IR-CVSC strategy provides green time until the speed ratio of the vehicles in front of the incident zone and adjacent to the incident lane stops increasing in comparison with the previous one (as the halted CVs behind the incident location move to the nonincident lane, causing the decrease in speed ratio) and then applies the throughput ratio criteria to extend or terminate the green phase. Figure 1 shows the flowchart that includes all the steps of IA-CVSC, incident detection, and IR-CVSC strategies integrated into one algorithm, which is applied during the green phase. During the red and amber intervals, the algorithm obtains and updates the CV and NCV information and estimates the forward movements of NCVs.

4. Testing and Evaluation of the Proposed Strategy

4.1. Tested Incidents. A variety of incident scenarios such as vehicle breakdown, road maintenance works, bus stop, illegal parking, and vehicle loading/unloading can be tackled by the integrated algorithm. For a robust evaluation of the algorithm, three types of incidents were tested (see Table 2), and each applied in different contexts and within a wide range of specifications, including different locations of incidents' occurrence, with varied frequencies and duration times, which can disrupt the normal traffic flow and allows the possibility to study the effects of incidents and the vehicles' behaviour in such situations.

4.2. Intersection Layout and Traffic Demand Patterns. Simulations of different incident scenarios were carried out in VISSIM 8 software [60], on a four-legged isolated intersection, along Castle Downs Road and 97 Street, Edmonton (see Figure 6). The case study was based on the existing layout and traffic movements obtained from the database of Edmonton [61].

Vehicle volumes were converted to approximate different intersection saturation rates using the Intersection Capacity Utilization (ICU) metric [62]. The field recorded volumes produced 0.65 ICU across the intersection. These basic, saturated, volumes were altered by uniform factors to generate volumes of 0.35, 0.50, 0.80, and 0.95 ICU, respectively (see Table 3). To calculate the ICUs, the lost-time per phase, minimum green time, and the reference cycle length were assumed as 4, 5, and 120 seconds, respectively.

A basic, constant, split between different types of vehicles was used in the simulations: cars (95%), buses (3%), and heavy goods vehicles (HGVs) (2%). Minimum and maximum desired speeds ranges in VISSIM were set as 48–58 kmph, 40–45 kmph, and 40–45 kmph, respectively, for cars, buses, and HGVs. The average desired speed of these ranges was used in the speed ratio calculation in equation (1).

A four-phase traffic signal plan was applied with turning movements allocated to Phase 1 (NBT, NBR, SBT, and SBR), Phase 2 (NBL and SBL), Phase 3 (EBT, EBR, WBT, and WBR), and Phase 4 (EBL and WBL). The movements EBR, WBR, NBR, and SBR on the intersection have dedicated right turns.

4.3. Strategy's Evaluation Process. The testing and evaluation of the IR-CVSC strategy were performed by comparison with EPICS [63], a reference, state-of-the-art, adaptive signal control solution, that is embedded in the VISSIM 8 simulation tool, which provides a platform for the comparison procedures. The proposed algorithm was coded in C++ language and was integrated with VISSIM COM API, which enables the users to access the vehicles' speed and position, which were, then, used in the IA-CVSC and IR-CVSC strategies to control the signal timings.

Each incident event was tested for the selected set of scenarios as shown in Table 2, with varying traffic demand (ICUs—0.35, 0.50, 0.65, 0.80, and 0.95), varying CV penetration rates (100%, 80%, 60%, 40%, and 20%) to reflect different near-future possible traffic conditions, and varying incident locations from the stop-line (0 m, 25 m, 50 m, 100 m, 150 m, and 200 m). The simulation was run for 1 hour in which the first 15 minutes was for warming-up, and the rest for the incident duration and congestion dissipation. The results of the different scenarios were based on the average of 10 random seeds. A total of 29,700 simulations were run on the intersection with results being obtained for all the signal control solutions (IR-CVSC, IA-CVSC, and EPICS) analysed.

4.4. Performance Indicators. As the measure of the quality of success of an incident detection algorithm, the following measures of effectiveness (MOEs) were used:

(a) Detection rate (DR)—ratio of the number of detected incidents over the total number of actual incidents,

(b) False alarm rate (FAR)—ratio of the number of "false" detected incidents over the total number of detection algorithm decisions,

(c) Mean time to detect (MTTD)—average time interval from the moment of incident's occurrence/clearance to the moment of its detection.

To compare the relative performance of different signal control solutions (IR-CVSC, IA-CVSC, and EPICS), the MOEs used were the average travel time delay and the average number of stops per vehicle, which represent the difference between ideal travel time/stops (computed assuming no other vehicles on the network and no delays/stops at signal controls) and actual travel time/stops.

Table 2: Tested incident scenarios.

Incident	Bus breakdown	Bus stop						Illegal car parking	
	Type 1	Type 2.1	Type 2.2	Type 2.3	Type 2.4	Type 2.5	Type 2.6	Type 3.1	Type 3.2
Frequency	Once	2 min	2 min	5 min	5 min	10 min	10 min	10 min	10 min
Duration	20 min	20 s	40 s	20 s	40 s	20 s	40 s	60 s	120 s

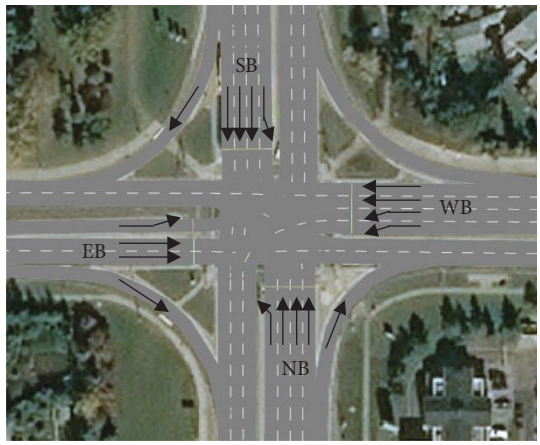


Figure 6: Snapshot of test intersection (Source: Bing maps).

Table 3: Tested volumes along the intersection.

ICU	EBL (8%)	EBT (69%)	EBR (23%)	WBL (33%)	WBT (46%)	WBR (21%)	NBL (25%)	NBT (52%)	NBR (23%)	SBL (13%)	SBT (55%)	SBR (32%)
0.35	35	296	99	159	225	103	244	514	239	108	452	268
0.50	50	423	141	228	322	148	348	735	341	154	645	383
0.65	65	550	183	296	418	192	453	955	443	200	839	498
0.80	80	677	225	364	514	236	558	1175	509	246	1033	613
0.95	95	804	267	433	611	281	662	1396	605	292	1226	728

5. Results

This section presents the simulation results of all the incident scenarios of IR-CVSC, IA-CVSC, and EPICS solutions. Evaluation of the results is divided into two sections: (1) efficiency of IR-CVSC to detect incidents, where the performance indicators of incident detection, for each incident scenario, are presented; (2) relative performance of IR-CVSC strategy, where the MOEs of IR-CVSC and its performance in comparison with IA-CVSC and EPICS are presented.

5.1. Efficiency of IR-CVSC Strategy to Detect Incidents. To evaluate the efficiency of the IR-CVSC strategy to detect the incident's occurrence (IO) and incidents' clearance (IC), the parameters MTTD-IO, MTTD-IC (95th percentile), and DR-IO, DR-IC (5th percentile) were chosen. The evaluation results comparing all the incident types at different CVPs, ICUs, and incident locations are presented in Figure 7.

The following basic conclusions can be drawn from the detection efficiency's results:

(1) With the decrease in CVP and decrease in ICU, DR-IO and DR-IC decreased, and MTTD-IO and MTTD-IC increased (due to fewer vehicles, more space, and more lane changes).

(2) With incidents' location away from the stop-line, DR-IO and DR-IC decreased, and MTTD-IO and MTTD-IC increased (due to more space availability for the vehicles to traverse in front of incidents' location).

(3) For a given frequency, as the incidents' duration increased (as in type 2 and type 3), DR-IO and DR-IC increased, and MTTD-IO and MTTD-IC decreased.

(4) For a given duration, as the incidents' frequency increased (as in type 2), congestion increased causing more stopping time, resulting in the increase in DR-IO and DR-IC, and decrease in MTTD-IO and MTTD-IC.

(5) For type 2 and type 3 incidents whose frequency of occurrence is more than once, DR-IO is always greater than or equal to DR-IC. This is because until the algorithm detects the incidents' clearance, the next incidents' occurrence is not detected.

(6) FAR of incidents' occurrence and clearance was 0 in all cases. This is due to the applied detection method (as in Section 3.2).

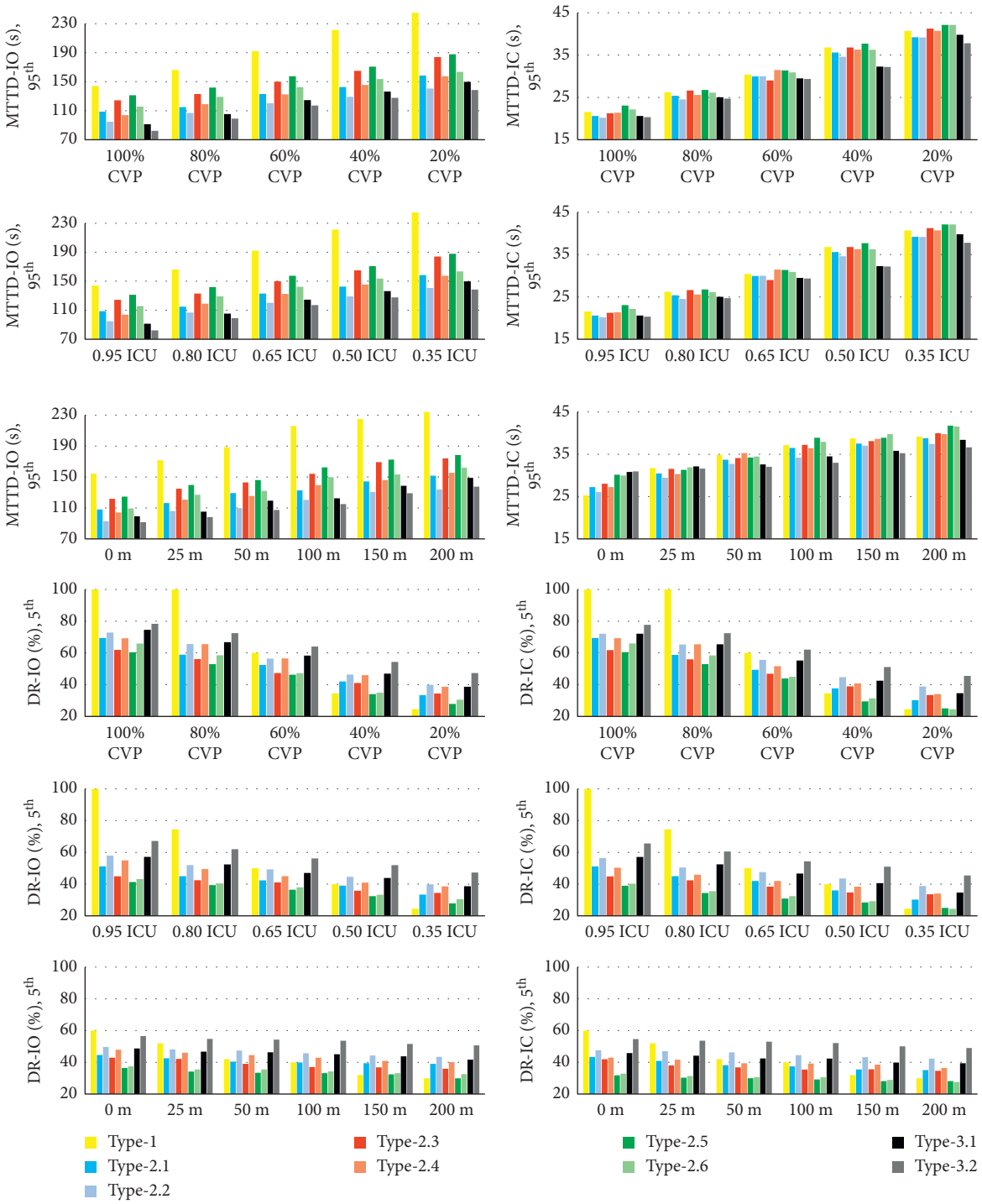


Figure 7: MTTD (95th percentile) and DR (5th percentile) comparison of all incident types at different CVPs, ICUs, and incident locations from the stop-line. Each bar represents an incident type.

5.2. Relative Performance of IR-CVSC Strategy. The signal control evaluation of IR-CVSC, IA-CVSC, and EPICS strategies and comparing all the incident types at all CVPs, ICUs, and incident locations are presented in Figures 8 and 9 and Table 4. It should be noted that IA-CVSC and EPICS solutions do not detect the incidents. They are, however, compared with the IR-CVSC strategy to analyse the differences in their working principles during the occurrence of incidents and also to check whether the new strategy can improve the intersections' performance in case of incidents' occurrence.

The following basic conclusions can be drawn from the signal control evaluation results:

(1) For a given frequency, as the incidents' duration increased (as in type 2 and type 3), delays and stops

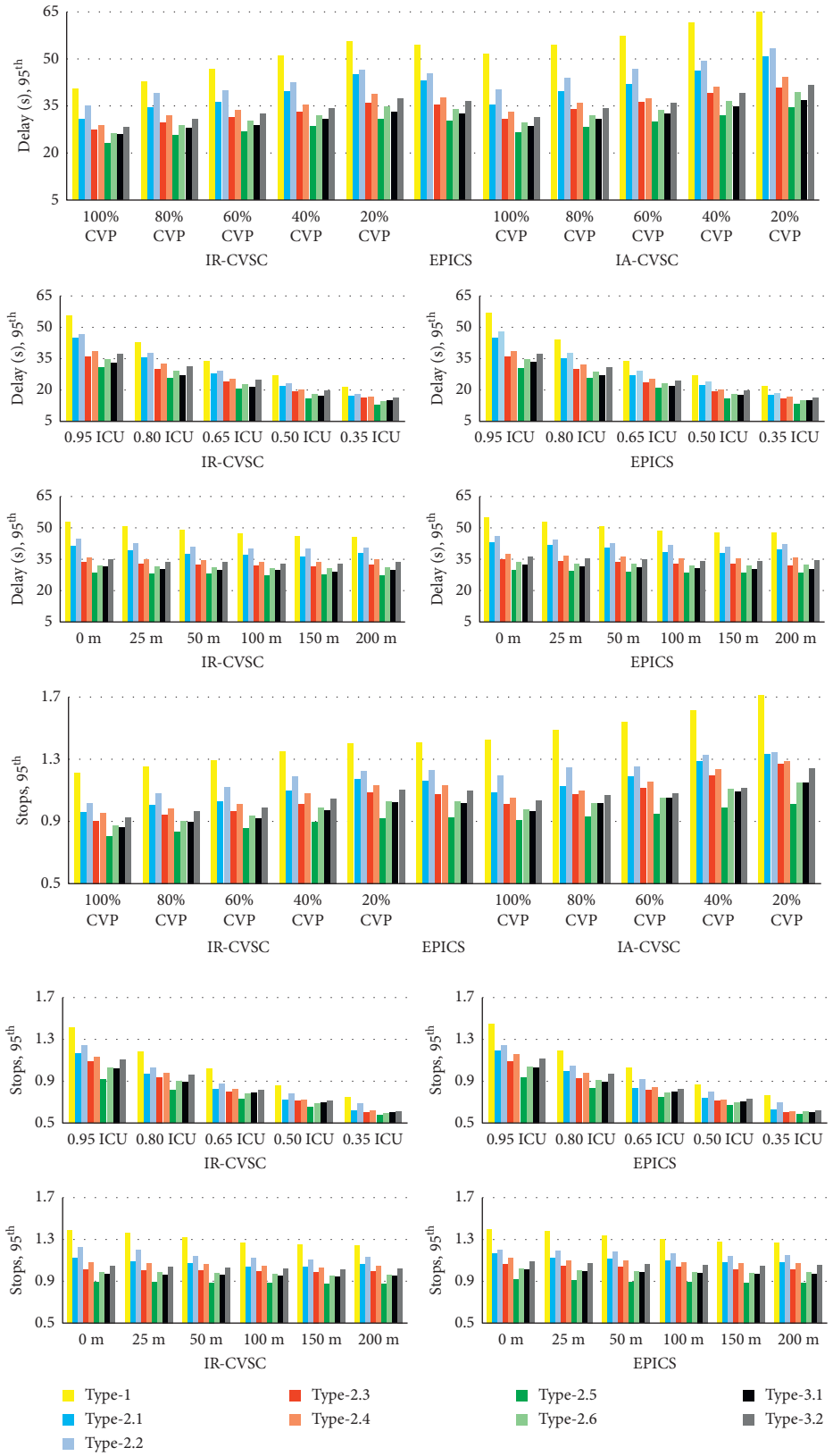


Figure 8: Signal control MOEs of IR-CVSC, IA-CVSC, and EPICS strategies of all incidents. Each bar represents the MOE of an incident type.

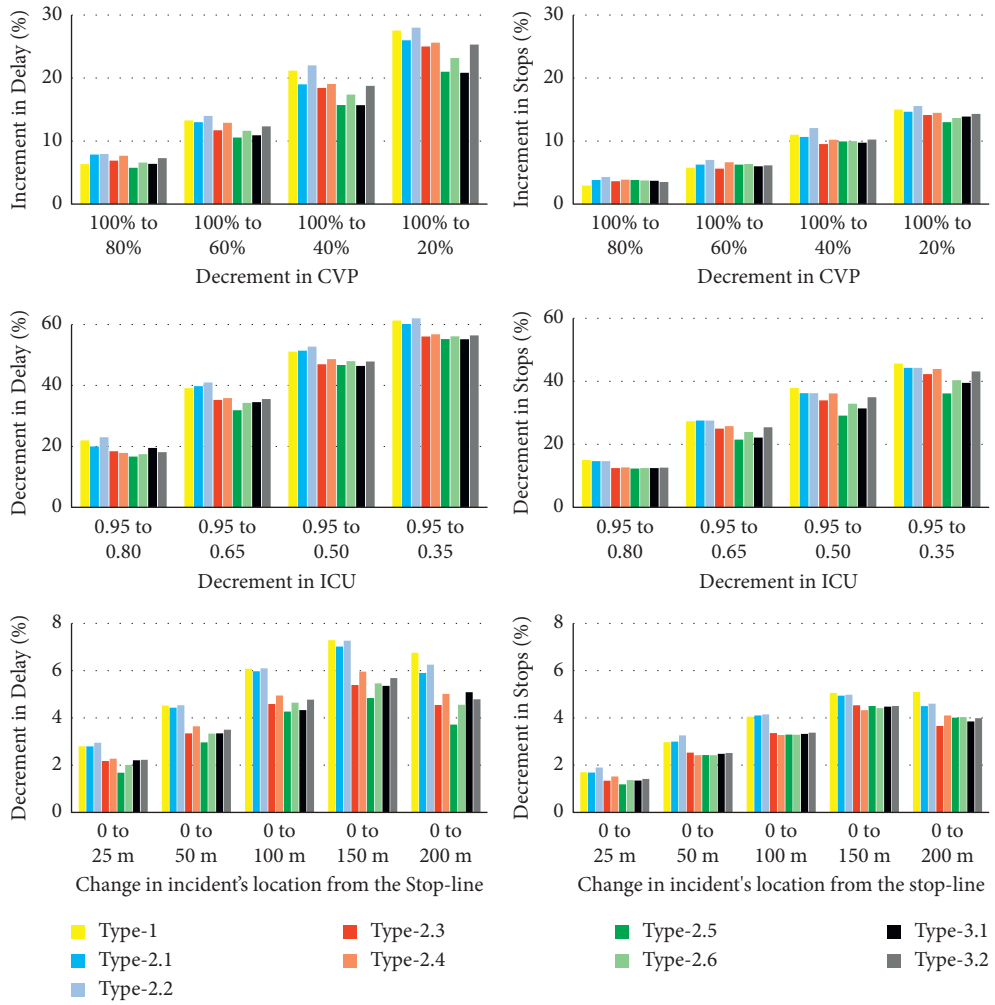


Figure 9: Percent change in IR-CVSC delays and stops with change in CVP, ICU, and incident's location from the stop-line.

Table 4: Signal control MOEs' evaluation at all CVPs, ICUs, and incident locations.

Incident	Type 1	Type 2.1	Type 2.2	Type 2.3	Type 2.4	Type 2.5	Type 2.6	Type 3.1	Type 3.2
Frequency	Once	2 min	2 min	5 min	5 min	10 min	10 min	10 min	10 min
Duration	20 min	20 s	40 s	20 s	40 s	20 s	40 s	60 s	120 s
Performance range of IR-CVSC over IA-CVSC at all CVPs, ICUs, and incident locations									
Delay (%)	13 to 26	10 to 24	8 to 27	7 to 17	9 to 21	6 to 24	8 to 21	5 to 20	6 to 16
Stops (%)	7 to 20	7 to 18	4 to 16	7 to 15	8 to 17	6 to 13	8 to 13	8 to 14	5 to 20
Performance range of IR-CVSC over EPICS at all CVPs, ICUs, and incident locations									
Delay (%)	−3 to 29	−8 to 34	−5 to 29	−6 to 25	−3 to 29	−7 to 25	−3 to 25	−3 to 22	−3 to 27
Stops (%)	−1 to 16	−2 to 19	−2 to 19	−5 to 17	−3 to 17	−1 to 16	−1 to 15	−2 to 16	−1 to 16

increased in all the signal control strategies, as expected.

(2) For a given duration, as the incidents' frequency increased (as in type 2), congestion increased, resulting in higher delays and stops.

(3) Delays and stops increased with the decrease in CVP and increase in ICU.

(4) As the incidents' location moved from 0m to 150m before the stop-line, delays and stops decreased. As expected, the closer the incidents' occurrence to the stop-line, the higher the delays and stops are.

(5) As the incidents' location moved from 150 m to 200 m before the stop-line, delays and stops increased (above 90% of the time) in all strategies. This is because the incidents' location at 200 m from the stop-line is far for the signal control strategy to serve the vehicles with lower MOEs compared with the location at 150 m. Changes in road geometry and the number of lanes might also be the influential factors for this pattern.

Conclusions on the performance of IR-CVSC strategy are as follows:

(1) IR-CVSC showed 5–27% lower delay and 4–20% lower stops compared with IA-CVSC strategy at all incident scenarios, thus, proving the necessity for an incident-responsive signal control strategy.

(2) At 40% and higher CVPs, at all types of incidents, IR-CVSC showed better performance over EPICS adaptive signal control, in reducing delays and stops by 1–34% and 1–19%, respectively.

(3) At 20% CVP, at all ICUs and incident locations, IR-CVSC showed 1–8% higher delays and 1–5% higher stops compared with EPICS.

(4) IR-CVSC strategy performed better with higher DR for incidents with longer duration. DR-IO for type-1 incidents was 21–36% and 12–19% higher than type-2 and type-3 incidents, respectively.

(5) For type-2 and type-3 incidents, as the frequency of incidents' occurrence is more than once and the incident duration is much shorter than type-1 incidents, IR-CVSC showed lower DR compared with type 1. This might be because the incident detection method was applied only during the green interval; thus, shorter duration incidents that might have occurred and cleared within a red interval remained undetected, or the incident might have occurred and cleared within the predefined threshold time (T/3 seconds).

(6) At higher ICUs, due to increased congestion with incidents' occurrence, IR-CVSC showed higher DRs and lower MTTDs. DR-IO and MTTD-IO at higher ICUs (0.95 and 0.80) were 17–25% higher and 40–66% lower than that of lower ICUs (0.50 and 0.35), respectively.

6. Conclusions and Future Work

This paper presented a novel algorithm that tackles incident-related traffic congestion problems for isolated intersections, by integrating 3 different strategies, namely, (1) a CV-based signal control strategy for normal traffic conditions without incidents, (2) an incident detection technique using CV/NCV information, and (3) a CV-based signal control strategy during the occurrence and dissipation of incidents. The integrated algorithm applies the corresponding strategy in real time, according to the incident/nonincident traffic conditions.

During nonincident conditions, the integrated algorithm applies the IA-CVSC strategy, where the speed and location information of CVs and NCVs at every time-step was obtained and green time was provided based on the ratio of space-mean speed to the desired speed of vehicles, as well as intersections' throughput maximization criteria. Incidents' occurrence is detected when a stopped vehicle during any green interval fails to move within a predefined threshold time. When an incident is detected, the integrated algorithm switches from IA-CVSC to IR-CVSC strategy, where green time is provided based on the space-mean speeds of vehicles on the nonincident lane, lane-changing pattern of the CVs behind the incident location, and throughput maximization criteria. Incidents' clearance is identified when the road portion between the stop-line and the location of the subject vehicle (which is responsible for the incidents' identification) on the incident lane is traversed. When the incident is cleared, IR-CVSC switches to the IA-CVSC strategy.

The incident detection accuracy and the signal control performance of the IR-CVSC strategy were evaluated by performing simulations on a real-field four-legged isolated intersection using the VISSIM software tool. Different incident scenarios were selected, with varied incident durations and frequencies of occurrence, and tested at different CVPs, ICUs, and incident locations. From the simulation results, the range of DR-IO and DR-IC (5^{th} percentile), and MTTD-IO and MTTD-IC (95^{th} percentile) for long-duration incidents (type 1) was 25–100%, 25–100%, 119–246 s, and 22–41 s, respectively, and for short-duration incidents with different frequencies of occurrence (type 2 and type 3) was 28–78%, 24–78%, 82–188 s, and 20–42 s, respectively. Higher DRs and lower MTTDs have resulted at higher CVPs, higher ICUs, and closer incident locations to the stop-line. In relation to the signal control strategy, at 40% and higher CVPs, at all types of incidents, IR-CVSC showed better performance over EPICS, in reducing delays and stops. Delays and stops of IR-CVSC strategy increased with a decrease in CVP and increase in ICU, and at incidents' locations nearer to the stop-line.

Findings from the results of incident detection and incident-responsive signal control strategies showed the modelling approach's potential in utilizing the emerging connected vehicle technology to mitigate the impact of incident-related traffic congestion significantly compared with an adaptive signal control solution. The integrated algorithm framework also provides more opportunities to study, build, and test several new heuristics to tackle incident-related traffic problems on both isolated intersections and intersection networks.

It is necessary to mention a few limitations related to the proposed algorithm. Currently, the algorithm is applicable for isolated intersections that have at least one adjacent lane available to serve the vehicles during the occurrence of incidents. IR-CVSC strategy must be improved for the situation where, if the incident occurs on a left-turning lane that does not belong to the phase of straight and right-turning movements, the phase plan must be altered in real time by including all the traffic movements on the incident-affected lanes under one phase, thus giving the possibility for the left-turning vehicles to use an adjacent lane during the occurrence of incidents. Future work should also focus on a method to estimate the speeds and positions of NCVs during a lane change and integrate it into the IR-CVSC strategy for more benefits.

Data Availability

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

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

This research was funded by the Portuguese Foundation for Science and Technology (FCT), under Grant SFRH/BD/51536/2011, MIT Portugal Program. The authors would like to acknowledge the PTV Group for providing the EPICS adaptive signal control software.

References

[1] L. Qi, M. Zhou, and W. Luan, "A two-level traffic light control strategy for preventing incident-based urban traffic congestion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 13–24, 2016.

[2] E. Parkany and C. Xie, *A Complete Review of Incident Detection Algorithms & Their Deployment: What Works and what Doesn't (No. NETCR 37, NETC 00-7)*, 2005.

[3] J.-B. Sheu, "A stochastic optimal control approach to real-time, incident-responsive traffic signal control at isolated intersections," *Transportation Science*, vol. 36, no. 4, pp. 418–434, 2002.

[4] Fhwa, *Reducing Non-recurring Congestion*, Federal Highway Administration, U.S. Department of Transportation, Washington DC, 2021.

[5] H. Park and A. Haghani, "Real-time prediction of secondary incident occurrences using vehicle probe data," *Transportation Research Part C: Emerging Technologies*, vol. 70, pp. 69–85, 2016.

[6] J. Oskarbski, M. Zawisza, and K. Żarski, "Automatic incident detection at intersections with use of telematics," *Transportation Research Procedia*, vol. 14, pp. 3466–3475, 2016.

[7] F. Ahmed and Y. E. Hawas, "A threshold-based real-time incident detection system for urban traffic networks," *Procedia - Social and Behavioral Sciences*, vol. 48, pp. 1713–1722, 2012.

[8] S. Hireche and A. Dennai, "Machine learning techniques for road traffic automatic incident detection systems: a review, Lecture Notes in Networks and Systems," *In International Conference in Artificial Intelligence in Renewable Energetic Systems*, pp. 60–69, 2019.

[9] B. Ghosh and D. P. Smith, "Customization of automatic incident detection algorithms for signalized urban arterials," *Journal of Intelligent Transportation Systems*, vol. 18, no. 4, pp. 426–441, 2014.

[10] Y. Asakura, T. Kusakabe, N. X. Long, and T. Ushiki, "Incident detection methods using probe vehicles with on-board GPS equipment," *Transportation Research Procedia*, vol. 6, pp. 17–27, 2015.

[11] W. Yu, S. Park, D. S. Kim, and S. S. Ko, "Arterial road incident detection based on time-moving average method in bluetooth-based wireless vehicle reidentification system," *Journal of Transportation Engineering*, vol. 141, no. 3, pp. 401–408, 2015.

[12] J. Ren, Y. Chen, L. Xin, J. Shi, B. Li, and Y. Liu, "Detecting and positioning of traffic incidents via video-based analysis of traffic states in a road segment," *IET Intelligent Transport Systems*, vol. 10, no. 6, pp. 428–437, 2016.

[13] W. Gu, V. V. Gayah, M. J. Cassidy, and N. Saade, "On the impacts of bus stops near signalized intersections: models of car and bus delays," *Transportation Research Part B: Methodological*, vol. 68, pp. 123–140, 2014.

[14] E. D'Andrea and F. Marcelloni, "Detection of traffic congestion and incidents from GPS trace analysis," *Expert Systems with Applications*, vol. 73, pp. 43–56, 2017.

[15] S. M. A. Rizvi, A. Ahmed, and Y. Shen, "Real-time incident detection and capacity estimation using loop detector data," *Journal of Advanced Transportation*, vol. 2020, Article ID 8857502, 14 pages, 2020.

[16] M. S. Sheikh, J. Liang, and W. Wang, "An improved automatic traffic incident detection technique using a vehicle to infrastructure communication," *Journal of Advanced Transportation*, vol. 2020, Article ID 9139074, 14 pages, 2020.

[17] A. Zaitouny, A. D. Fragkou, T. Stemler et al., "Multiple sensors data integration for traffic incident detection using the quadrant scan," *Sensors*, vol. 22, no. 8, p. 2933, 2022.

[18] H. Liang, H. Song, X. Yun, S. Sun, Y. Wang, and Z. Zhang, *Traffic Incident Detection Based on a Global Trajectory Spatiotemporal Map*, , pp. 1–20, Complex & Intelligent Systems, 2021.

[19] B. Lu, R. Xu, L. Hou, and Z. Zhang, "Signal control optimization for urban traffic against incident-induced congestion," in *Proceedings of the 2011 Chinese Control and Decision Conference (CCDC)*, pp. 1910–1914, Mianyang, China, May 2011.

[20] J. Long, Z. Gao, P. Orenstein, and H. Ren, "Control strategies for dispersing incident-based traffic jams in two-way grid networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 469–481, 2011.

[21] T. Mao, A. S. Mihăiță, and C. Cai, "Traffic signal control optimisation under severe incident conditions using genetic algorithm," in *ITS World Congress 2019 (ITSWC2019)*Singapore, 2019.

[22] J. Wang, J. Hang, and X. Zhou, "Signal timing optimization model for intersections in traffic incidents," *Journal of Advanced Transportation*, vol. 2020, Article ID 1081365, 9 pages, 2020.

[23] K. Yao and S. Chen, "Resilience-based adaptive traffic signal strategy against disruption at single intersection," *Journal of Transportation Engineering, Part A: Systems*, vol. 148, no. 5, 04022018 pages, 2022.

[24] Z. Huang, *Traffic Signal Control at Connected Vehicle Equipped Intersections*, Doctoral dissertation, Mississippi State University, 2016.

[25] Ite, "Connected/Automated Vehicles" , 2018, https://www.ite.org/technical-resources/topics/connected-automated-vehicles/.

[26] Itsjpo, "Connected vehicle pilot deployment Program- connected vehicle project reveals ambiguity in communication standard," 2021, https://www.its.dot.gov/pilots/communication_standard.htm.

[27] C. Invent, "Connected vehicle trend radar," 2020, https://www.capgemini.com/pt-en/resources/connected-vehicle-trend-radar/.

[28] Statista, "Connected vehicle fleet from 2025 to 2035, by key region," 2020, https://www.statista.com/statistics/442529/connected-car-market-size-by-region/.

[29] K. Chandan, A. M. Seco, and A. B. Silva, "Real-time traffic signal control for isolated intersection, using car-following logic under connected vehicle environment," *Transportation Research Procedia*, vol. 25, pp. 1610–1625, 2017.

[30] K. K. K. Chandan, Á. J. M. Seco, and A. M. C. Bastos Silva, "A real-time traffic signal control strategy under partially

connected vehicle environment," *Promet - Traffic & Transportation*, vol. 31, no. 1, pp. 61–73, 2019.

[31] N. Owens, A. Armstrong, and P. Sullivan, *Traffic Incident Management Handbook (No. FHWA-HOP-10-013)*, 2010.

[32] Fhwa, "Emergency transportation operations," 2017, https://ops.fhwa.dot.gov/eto_tim_pse/docs/incident_mgmt_perf/section2.htm.

[33] Q. Shang, D. Tan, S. Gao, and L. Feng, "A hybrid method for traffic incident duration prediction using BOA-optimized random forest combined with neighborhood components analysis," *Journal of Advanced Transportation*, vol. 2019, Article ID 4202735, 11 pages, 2019.

[34] M. Elhenawy, H. A. Rakha, and H. I. Ashqar, "Joint impact of rain and incidents on traffic stream speeds," *Journal of Advanced Transportation*, vol. 2021, Article ID 8812740, 12 pages, 2021.

[35] Z. Iqbal, M. I. Khan, S. Hussain, and A. Habib, "An efficient traffic incident detection and classification framework by leveraging the efficacy of model stacking," *Complexity*, vol. 2021, Article ID 5543698, 17 pages, 2021.

[36] Y. Zou, B. Lin, and X. Yang, "Application of the bayesian model averaging in analyzing freeway traffic incident clearance time for emergency management," *Journal of Advanced Transportation*, vol. 2021, Article ID 6671983, 9 pages, 2021.

[37] J. Tang, L. Zheng, C. Han, F. Liu, and J. Cai, "Traffic incident clearance time prediction and influencing factor analysis using extreme gradient boosting model," *Journal of Advanced Transportation*, vol. 2020, Article ID 6401082, 12 pages, 2020.

[38] A. Tavassoli Hojati, L. Ferreira, S. Washington, P. Charles, and A. Shobeirinejad, "Modelling total duration of traffic incidents including incident detection and recovery time," *Accident Analysis & Prevention*, vol. 71, pp. 296–305, 2014.

[39] R. Jiang, M. Qu, and E. Chung, "Traffic incident clearance time and arrival time prediction based on hazard models," *Mathematical Problems in Engineering*, vol. 2014, Article ID 508039, 11 pages, 2014.

[40] H. Dia and K. Thomas, "Development and evaluation of arterial incident detection models using fusion of simulated probe vehicle and loop detector data," *Information Fusion*, vol. 12, no. 1, pp. 20–27, 2011.

[41] Y. Zhu, X. Ye, J. Chen, X. Yan, and T. Wang, "Impact of cruising for parking on travel time of traffic flow," *Sustainability*, vol. 12, no. 8, p. 3079, 2020.

[42] Y. Cao, Z. Z. Yang, and Z. Y. Zuo, "The effect of curb parking on road capacity and traffic safety," *European Transport Research Review*, vol. 1, no. 9, pp. 1–10, 2016.

[43] J. Cao and M. Menendez, "Generalized effects of on-street parking maneuvers on the performance of nearby signalized intersections," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2483, no. 1, pp. 30–38, 2015.

[44] S. Wijayaratna, "Impacts of on-street parking on road capacity," *In Australasian Transport Research Forum*, vol. 21, pp. 1–15, 2015.

[45] O. Hrytsun, O. Lanets, O. Lanets, and S. Solodkyy, "Impact of street parking on delays and the average speed of traffic flow," *Transport Technologies*, vol. 2020, no. 1, pp. 33–44, 2020.

[46] S. Liang and M. Ma, "Analysis of bus bunching impact on car delays at signalized intersections," *KSCE Journal of Civil Engineering*, vol. 23, no. 2, pp. 833–843, 2019.

[47] Z. Liu and M. Jian, "Traffic impacts analysis of bus stops near signalized intersections based on an optimal velocity model," *Advances in Mechanical Engineering*, vol. 11, no. 5, pp. 1–11, 2019.

[48] C. Chavis and E. Christofa, "A real-time signal control strategy for mitigating the impact of bus stops at urban signalized intersections," *Journal of Intelligent Transportation Systems*, vol. 21, no. 5, pp. 349–363, 2017.

[49] W. Ranasinghe, J. Bunker, and A. Bhaskar, "Saturation headway variation at a signalised intersection approaches with a downstream bus stop and bicycle lane," *In Australasian Transport Research Forum*, vol. 54, pp. 27–29, 2017.

[50] K. Ghasemlou, M. Mutlu Aydın, and M. Sinan Yıldırım, "An investigation on lane blockage effects at signalized intersections," *International Journal of Traffic and Transportation Engineering*, vol. 6, no. 3, pp. 289–302, 2016.

[51] R. Li, F. C. Pereira, and M. E. Ben-Akiva, "Overview of traffic incident duration analysis and prediction," *European transport research review*, vol. 10, no. 2, pp. 1–13, 2018.

[52] Hcm, *Highway Capacity Manual. Transportation Research Board*, National Research Council, Washington, DC, 2010.

[53] Fhwa, "Guide for highway capacity and operations analysis of active transportation and demand management strategies," *Appendix F: Speed/Capacity for Incident Duration Reductions*, 2020.

[54] A. M. S. Alkaabi, D. Dissanayake, and R. Bird, "Analyzing clearance time of urban traffic accidents in Abu Dhabi, United Arab Emirates, with hazard-based duration modeling method," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2229, no. 1, pp. 46–54, 2011.

[55] L. Hou, Y. Lao, Y. Wang, Z. Zhang, Y. Zhang, and Z. Li, "Modeling freeway incident response time: a mechanism-based approach," *Transportation Research Part C: Emerging Technologies*, vol. 28, pp. 87–100, 2013.

[56] I. Ghosh, P. T. Savolainen, and T. J. Gates, "Examination of factors affecting freeway incident clearance times: a comparison of the generalizedFmodel and several alternative nested models," *Journal of Advanced Transportation*, vol. 48, no. 6, pp. 471–485, 2014.

[57] C. Ding, X. Ma, Y. Wang, and Y. Wang, "Exploring the influential factors in incident clearance time: disentangling causation from self-selection bias," *Accident Analysis & Prevention*, vol. 85, pp. 58–65, 2015.

[58] K. Yang, S. I. Guler, and M. Menendez, "Isolated intersection control for various levels of vehicle technology: conventional, connected, and automated vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 109–129, 2016.

[59] P. Koonce and L. Rodegerdts, *Traffic Signal Timing Manual (No. FHWA-HOP-08-024)*, Federal Highway Administration, United States, 2008.

[60] VISSIM. Ptv, *PTV VISSIM 8 User Manual*, PTV Planung Transport Verkehr AG, Karlsruhe, Germany, 2015.

[61] Edmonton, "Traffic volumes and turning movements," *Transportation Data for the City of Edmonton*, 2015.

[62] D. Husch and J. Albeck, *Intersection Capacity Utilization: Evaluation Procedures for Intersections and Interchanges*, Trafficware, 2003.

[63] EPICS. Ptv, *PTV EPICS User Manual*, PTV Planung Transport Verkehr AG, Karlsruhe, Germany, 2015.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 3582355, 17 pages
https://doi.org/10.1155/2022/3582355

Research Article

Automatic Vehicles' Trajectories Optimization on Highway Exclusive Lanes

Lingjuan Chen,[1] **Yangbo Ruan**,[1] **and Yiquan Gou**[2]

[1]*School of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan, China*
[2]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China*

Correspondence should be addressed to Lingjuan Chen; chenlingjuan@wust.edu.cn

Received 15 November 2021; Revised 18 April 2022; Accepted 3 June 2022; Published 29 June 2022

Academic Editor: Alain Lambert

The rapid development of V2X communication has made it possible to optimize and control the trajectories of vehicles from the whole traffic flow's perspective and improve traffic performance. Therefore, this paper discusses the trajectories management problem on highway facilitated with lanes exclusively for autonomous vehicles (AVs). The paper proposes a model that aims to search for optimal trajectories and minimize total travel time for AVs with multiple initial and target states while averting crashes and conforming to vehicles' kinetic. Dividing the time zone into discrete pieces, the model is analyzed as a large-scale discrete problem influenced by the randomness of the sequence of vehicles. A two-phase algorithm combined with upper evolution strategies and lower dynamic programming is developed to diminish stochastics and reduce computation step by step and solve the trajectories optimization model. Numerical experiments validate that the proposed method is capable of generating optimal trajectories for multiple AVs and approaching to system optimum by simultaneously solving all the spatial and temporal values of the trajectories. The two-phase algorithm can be applied efficiently in practice to obtain a feasible approximate solution for trajectories optimization by presetting appropriate algorithm parameters.

1. Introduction

Autonomous vehicles (AVs) are likely to create a revolutionary paradigm shift in the near future for real-time traffic system automation and control. The AVs can improve the transportation system's performance and reduce congestion, emissions, accidents, and time consumption by delivering system optimum travel strategies to vehicles with V2X communication. Although it may take a long time to realize a popular market occupancy, an intermediate step could be achieved and bring significant improvements that AVs travel on exclusive lanes in a special zone, such as autonomous vehicle lanes [1–4] or autonomous vehicle areas [4–6]. It is envisaged that while AVs enter into the specialized area [5], the vehicle's control is handed over to a central agent where the cyber component (e.g., data and shared information through vehicle-to-vehicle and vehicle-to-infrastructure communication) can aim to optimally control the physical entities (e.g., CAVs and non-CAVs); see [6]. The agent would guide it through the area (presumably by sending detailed trajectories to the vehicle's onboard computer).

Therefore, AV technologies allow vehicles to conform to uniform and global optimum trajectories aided by the cloud system's central computation. For traditional human drivers, numerous studies on travel behaviors or trajectories choice have been conducted, including UE or SUE, pre-departure route choice or en route revision, static or dynamic flow assignment, and macroscopic or microscopic vehicles' trajectories with various information [7–12]. The above-mentioned studies are based on the foundation that travelers are selfish to reduce their consumption, and all travelers are unfamiliar with each other. While the drivers benefit much from traffic guidance to reduce congestion with the development of information systems, the travel cost still remains at a high level due to the users being unable to observe others' reflections on multiple information. The AV technologies could change this mode with the evolution of communication and guide vehicles to respond immediately

to preplanned trajectories. The route choice or trajectory solving is transformed into a mathematic optimization problem with several constraints similar to railway or aviation systems and different from the former choice behavior models. In recent years, significant attention has been given to AVs' trajectories optimization, and a review can be concluded as follows.

1.1. Review of AV Developments in Traffic Management. Microsimulation models (cellular automata model and following model) have been used to develop the mixed traffic flow basic diagram and the influence of automatic driving on traffic flow [13]. Due to the difficulty in carrying out real vehicle experiments, the research on the traffic flow theory of connected and automated vehicles (CAV) is mainly focused on the longitudinal control technology of assisted driving, such as adaptive cruise and coadaptive cruise systems [14], as well as the road intersection management of AVs [15]. Studies have also been conducted to smooth highway traffic by controlling individual vehicles [16–18]. Since traffic management focuses on the interactions between the vehicles and resolving the conflicts that result from them, the studies that exclusively deal with trajectory planning for a single vehicle are not considered [19–23]. Liu et al. [24] solved an optimal trajectory problem for one single vehicle and used this trajectory as a template to control multiple vehicles with variable speed limits. Ahn et al. [25] also proposed a rolling-horizon individual CAV control strategy that only considered the road geometries. To find optimal trajectories for multiple AVs and improve the computational efficiency, Lu et al. [26] subsequently developed a specialized algorithm based on the rolling horizon approach (RHA) to obtain the best approximate solution.

Despite relatively homogenous constraints and complex algorithms, the studies mentioned above demonstrated great potential in traffic management. Therefore, according to multiple performance indicators (such as distance, time, and energy), an optimal trajectory is generated from the initial state to the final state, which has important research significance in the field of intelligent vehicle motion planning.

1.2. Review of Vehicle Trajectory Optimization Models. Vehicle trajectory optimization has been extensively studied in a broader domain. According to the summary presented by Betts [27], the path planning problem is discretized in time by allowing the vehicle to only make decisions at discrete time intervals. It is further discretized in space by only allowing the vehicle to make a limited number of choices at each time step. For multivehicle trajectories, the vehicles move from the initial-boundary states to the final states, which can be solved directly by using linear programming or integer programming solver tools (CPLEX). He et al. [20] introduced vehicle queue constraints into a multiphase optimal control model to construct an approximation formulation that contained fewer decision variables and was easier to solve. Furthermore, Wu et al. [21] applied the model and the algorithm to obtain the trajectories across intersections. In terms of car-following behavior, Chen et al. [1, 5] proposed a time-dependent model to optimally deploy AV lanes on a general network consisting of both CVs and AVs. In the network [28], AVs and HDVs (human-driving vehicles) in the road links were managed to use exclusive lanes (common lane and AV lane). In this case, no interference between the two different types of vehicles existed in the links, so the advantage of AV technology would be fully utilized. Ghiasi et al. [29] obtained the optimal number of AV exclusive lanes under some common vehicle spacing settings. Kakimoto et al. [3] also studied the influence of CAVs on single-lane expressways based on different time intervals. Actually, there will be a long period of the mixed traffic flow by AVs and HDVs temporarily. For lane changing, Zhang [30] decomposed complex maneuvers into two submaneuvers, that is, lane change and lane keeping. Thus, the trajectory planning was simplified mainly based on lane-change maneuvers. Luo et al. [22] proved that vehicles could perform real-time calculations and update the lane-changing track before completing lane changing. Li et al. [15] studied the problem of simultaneous lane changing of multiple vehicles through the cooperation of multiple vehicles. Lu et al. [26] subsequently developed a mathematical model with safety and car-following constraints. This paper also assumes that the vehicle's lane changing is instantaneous. The assumption would hold if AVs have sufficiently high autonomy and maneuverability; the travel time reduction of en route lane change could be offset.

1.3. Review of Vehicle Trajectory Optimization Methods. From the perspective of optimization methods, these problems are all nondeterministic polynomial-time-hard (NP-hard). The traditional methods for trajectory optimization include analytical approaches that can only solve simple problems with special structures and numerical approaches [31]. A vehicle trajectory is essentially an infinite-dimensional object in which the state (e.g., location, speed, acceleration) at each time spot can be varied. In the optimization model established by [32, 33], the traditional genetic algorithm has a small processing scale and is difficult to effectively deal with optimization problems with higher dimensions.

It is challenging to obtain one single vehicle's trajectory, particularly under nonlinear constraints. Therefore, Barnier and Brisset [34] adopted a new optimization method combining the genetic algorithm with the constraint satisfaction solving technology. The main idea was to deal with the subdomains of variables through the genetic algorithm, which is used for combinatorial optimization problems. Zhou et al. [35] devised a heuristic algorithm satisfying the need for formulating high-dimensional objects or complex system constraints, which could efficiently construct a smooth feasible trajectory vector with limited control parameters. Several studies have attempted to improve the algorithms by reducing the computation time. Gong's [36] numerical experiments showed that the original algorithms in Koshal et al. [37] and their convergence analysis often led to small step lengths and slow convergence. While the two

algorithms achieve the same numerical accuracy and convergence, the computation time of the modified algorithm was nearly reduced by half. Several recent studies have improved the results by approximating the trajectories with a simple piecewise quadratic function [35–38]. However, these approaches are heuristics and cannot guarantee the exact optimal solutions.

Through Lu's mixed-integer program (MIP) formulation, the feasible set may be tightened by rewriting some of the constraints or adding more valid constraints [26]. The RHA seems promising but requires further development to address the scheduling and equity constraints. Rios-Torres and Malikopoulos [39] presented an optimization framework and an analytical closed-form solution to find the optimal sequence and trajectory of vehicles. Wang et al. [40] proposed a distributed consensus protocol approach. However, the approach manages CAV based on relevant rules and does not use the optimization method to optimize the vehicle timing and trajectory as a whole.

As concluded above, significant progress has been made in resolving the trajectories issues. One approach is establishing multiple models in various circumstances, such as planning tracks at sections, ramps, working zones, and other special areas. The goals are aimed to maximize capacity, reduce time cost, and fuel emissions with diversiform restrain. The other approach is adopting a mathematical algorithm or car-following simulation to decrease the computation time. However, there are several problems:

(1) Models of trajectory optimization for multiple vehicles operating in distinct lanes are typically simplified for two reasons: first, it is challenging to solve the resulting nonlinear optimization problems involving a large number of decision variables (note that we often). Second, it is generally accepted that modeling errors do not significantly contribute to the loss of objective values when long planning time horizons are considered [41]. However, this central issue has not been fully addressed in the literature [42–45], and conventional research suffers from two limitations. First, numerical methods alone (e.g., Liu et al. and Ahn et al.) do not provide sufficient analytical knowledge, intuitive understanding, or fundamental insight into the structure of problems and solutions, which may obstruct the discovery of some potentially useful management insights in real-world applications. Second, the trajectory generated by these methods [20, 21] may not be smooth and comfortable enough for the vehicle to follow in practice. Compared to previous research, to address this gap, this paper focuses on determining the optimal trajectory for each vehicle over the entire setting time horizon with a variety of initial and final states. In this paper, the spontaneous idea is to decompose the problem into a two-phase model. In the first phase, a commonly used strategy, evolutionary strategy (ES) algorithm is designed to define the lane-change slot and the lane occupation for each interval combined with lane-change rules. In the second phase, a mixed integrated linear model constituted by constraint conditions is solved, and the objective function is the fitness of the first phase. The algorithm can solve optimum trajectories satisfying all the constraints in an acceptable running time. The trajectories are relatively smooth while simultaneously conforming to car-following, lane-changing behaviors, and maintaining steady speeds.

(2) A modified algorithm is established to reduce the computation time: most existing studies [30, 46, 47] either oversimplified the trajectories' constraints, adopted ad hoc heuristics [35, 38] without optimality assurance, or relied on complicated numerical procedures [48–50] that are unsuitable for real-time applications. In this paper, the ES can reduce the large scale of lane-change spots and rapidly search for optimum results by gathering a series of individuals' advantages. In the second phase, cutting invalid vehicles' orders reserves sufficient optimum results and speeds up the solving process in the dynamic programming for fitness computation. The evolution strategy is capable of reducing the large scale of lane-change spots and searching for optimum results. Cutting invalid vehicles' orders at each stage reserves sufficient optimum results.

The remainder of this paper is organized as follows. Section 2 discusses the hypothesis of automatic traffic flow in an exclusive area and the objectives and constraints of trajectory management. Section 3 introduces and analyzes a trajectory optimization model. Section 4 proposes a two-phase solution algorithm that utilizes upper evolution strategies and lowers dynamic programming to reduce the stochasticity associated with lane-change slot and vehicle order. Section 5 summarizes the results of a numerical experiment and presents the results of a controlled experiment designed to verify the algorithm's effectiveness. Finally, Section 6 summarizes the findings, implications, and limitations and suggests some future research directions.

2. Analysis of Automatic Traffic Management

This section proposes an AV trajectory optimization problem on a single exclusive highway segment. All the vehicles on the segment consisting of three lanes are automatic, which could be comprehensively organized, aiming to uniform objectives. Mixed traffic management with HDVs is the next research topic. Unlike the existing HDV management theories, the AV trajectory problem is a travel behavior optimization program instead of human-driven habits' approximate simulation. The main purpose is to determine each vehicle's optimum trajectory along the whole setting time horizon with various initial and final states.

2.1. Hypothesis of Automatic Traffic Flow. The research problem and the details of the hypothesis are presented as follows:

(1) Trajectory optimization is strictly a time-continuous problem. However, in reality, the velocity, acceleration, or deceleration are impossible to change continuously due to the constraints of vehicles' mechanical properties and reaction time. Therefore, in this paper, the time horizon is initially divided into identical segments denoted by k and Δt is each interval's duration ($k = 1, 2, \ldots, K$). For the consideration of optimization efficiency and computation consumption, the interval's duration should be set as a proper value to obtain a satisfying result and reduce the computational complexity as much as possible.

(2) The driven road is set as a highway or freeway and not as an urban network, as depicted in Figure 1. Vehicle characteristic parameters include entry time, velocity, acceleration, entry lane, and objective exit lane. If the entry does not conform with the exit lane, the vehicle must change the lane at the optimum interval. Some rules for lane-changing are explained as follows:

① The objective lane of the lane-changing vehicle must be the adjacent lane close to the final exit lane. Specifically, the vehicles are forbidden to travel across two lanes in one interval.

② In this paper, compulsory lane-change induced by nonconformity of entry and exit lanes is the sole concern. However, en route lane change due to lanes' travel time difference is not considered. This assumption could be reasonable in automatic traffic management. For AVs, the trajectories are controlled and optimized comprehensively, and the travel time reduction of en route lane change could be offset. This explanation is different from the HDVs, whose trajectories are decided by the drivers' self-strategy, and lane-change for shorter travel cost is obvious for individuals.

③ Lane changes of vehicles can be completed during a time interval. In other words, if a vehicle's lane-change time spot is k, then the vehicle will drive on the objective lane during the whole interval k. It will hold since the lane-change preparation and space adjustment are already accomplished before.

(3) In reality, incoming flow to the highway segment is continuous along the time axis. Therefore, a discrete-time horizon is set in this paper for the optimization section. The spatial-temporal segmentation and rolling promotion are credible and efficient for traffic management complex systems. Therefore, the proposed model and theory can be utilized in a continuous system.

2.2. Trajectory Management Objectives and Constraints. Under the above assumptions, the proposed solution aims to obtain the vehicle's velocity, acceleration/deceleration, position for each interval, and travel time on the segment. The optimal objective includes but is not limited to all vehicles' total time cost, fuel consumption, emission, and driven comfort. These indexes are all vital and closely correlate to the vehicle's maneuvers. In the proposed model, a minimum travel time is simplistically established for the unique objective of efficiency. Similarly, others could be utilized just by adopting various index calculation methods.

Vehicles' trajectories are subject to many constraints: (1) ensuring no collisions and safe space for each lane during the time horizon, (2) guaranteeing the vehicles travel on the exact lane before exiting from the road segment, and (3) restricting the lane-change maneuvers to the above hypothesis and avoiding crashes on the current and objective lanes simultaneously. The following section will explicitly explain the optimization problem and establish a mathematical model based on the above analysis.

3. Trajectory Optimization Model

3.1. Model Notations and Variables. The road segment is denoted as $R(x_o, x_d)$; the symbols x_o, x_d are location stamped as origin and terminus; and road length is L. Notation $i\,(1 \le i \le C)$ represents vehicle's number ordered by entering time, where C is the total number of vehicles. The artery consists of three lanes denoted as $l = 1, 2, 3$. l_{si}, l_{ei} are set as the initial and final lanes for vehicle i, respectively. Other variables are listed in Table 1.

3.2. Objectives and Constraints. In this paper, vehicles' total travel cost traversing the road is the unique objective for trajectory optimization.

$$Z = \min \sum_{i \in C} \left(k_i^{\text{out}} - k_i^{\text{in}}\right). \tag{1}$$

A binary variable θ_{ik} is induced to indicate whether the vehicle i still travels on the segment. If $x_{ik} - x_d \ge 0$, $\theta_{ik} = 0$; otherwise, $\theta_{ik} = 1$. θ_{ik} is subject to the constraints as described below:

$$\left(\theta_{ik} - 1\right) \times M < x_d - x_{ik} \le \theta_{ik} \times M, \tag{2}$$

$$\text{if } k < k_i^{\text{in}},\ \theta_{ik} = 0,\ x_{ik+1} = 0, \tag{3}$$

where M is a large positive real number. If $\theta_{ik} = 1$, equation (2) reads $0 < x_d - x_{ik} \le M$ and means $x_d > x_{ik}$; otherwise, if $\theta_{ik} = 0$, equation (2) reads $x_d < x_{ik}$. Thus, formulas (2)~(3) explain the variable δ_{ik} exactly, and objective Z could be reconstructed as follows:

$$Z = \min \sum_{i \in C} \sum_{1 \le k \le K} \theta_{ik} \times \Delta t. \tag{4}$$

The vehicles' movement on the segment is constrained to a series of conditions:

3.2.1. Velocity and Acceleration Constraints

$$-d_i^{\max} \le a_{ik} \le a_i^{\max},\ 1 \le i \le C, \tag{5}$$

$$0 \le v_{ik} \le v_i^{\max},\ 1 \le i \le C, \tag{6}$$

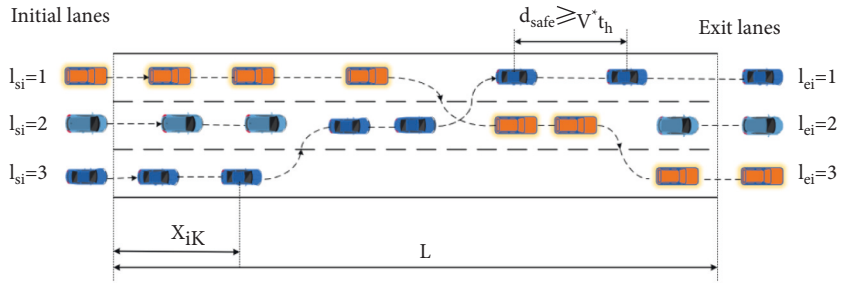


Figure 1: Illustration of vehicle trajectories.

Table 1: Description of variables.

Variables	Description
x_{ik}	Position of vehicle i at interval k
a_{ik}	Acceleration/deceleration of vehicle i at interval k
d_{ik}	Deceleration of vehicle i at interval k
v_{ik}	Velocity of vehicle i at interval k
l_{ik}	Occupied lane of vehicle i at interval k
$N_{i1}, N_{i2}, \ldots$	Lane-change intervals of vehicle i
$k_i^{\text{in}}, k_i^{\text{out}}$	Entering and exiting time of vehicle i

$$v_{ik} = v_{ik} + a_{ik} \times \Delta t, \tag{7}$$

$$x_{ik} = x_{ik-1} + v_{ik} \times \Delta t + \frac{1}{2} \times a_{ik} \times \Delta t^2. \tag{8}$$

$d_i^{\max}$, $a_i^{\max}$ and $v_i^{\max}$ are the limited deceleration, acceleration, and velocity for vehicle i, respectively. Equations (7) and (8) explain the recurrence relation among variables a_{ik}, v_{ik}, and x_{ik} and can hold in a short time duration.

3.2.2. Collision Avoidance Constraints. Two binary auxiliary variables, δ_{ijk} and θ_{ik}, are defined to construct the safety constraints. If vehicle i is the former vehicle of j on the identical lane k, $\delta_{ijk} = 1$; otherwise, $\delta_{ijk} = 0$. The position relationship can be imposed as follows:

$$0 \leq \delta_{ijk} - \frac{x_{ik} - x_{jk}}{M} < 1, \tag{9}$$

where $x_{ik} - x_{jk} > 0$ and variable $\delta_{ijk} = 1$; otherwise, $\delta_{ijk} = 0$. Therefore, equation (9) is consistent with the meaning of δ_{ijk}.

Constraint 1. Safety constraints for vehicles in the same lane. If vehicle i is the former one, i and j are not intended to change lanes:

$$x_{ik+1} - x_{jk+1} \geq t_h \times v_{jk+1}. \tag{10}$$

such that $l_{ik} = l_{jk}$, $k \neq N_{i1}, N_{i2}, N_{j1}, N_{j2}$, where t_h is the reaction time for driving straight. $N_{i1}, N_{i2}, N_{j1}, N_{j2}$ are vehicles' lane-change intervals. If $1 \leq k \leq N_{i1}$, $l_{ik} = l_{si}$; if $N_{i1} + 1 \leq k \leq N_{i2}$, $l_{ik} = l_{si} \pm 1$. To indicate the vehicle sequence and lane occupation, equation (10) becomes

$$\left(\left(1 - \delta_{ijk}\right) - \theta_{ik} - \theta_{jk} + 2\right) \times M + \left(x_{ik+1} - x_{jk+1}\right) \geq t_h \times v_{jk+1}. \tag{11}$$

It can be proved that equation (11) is equivalent to equation (10). If $\theta_{ik} = 0$ or $\theta_{jk} = 0$, equation (11) is not validated. When $\theta_{ik} = 1$, $\theta_{jk} = 1$; if $\delta_{ijk} = 0$, vehicle i is not following vehicle j, and equation (11) is not validated; if $\delta_{ijk} = 1$, equation (11) is active and consistent with equation (10).

Constraint 2. Safety constraints at a lane-change time interval. When vehicle i intends to the objective lane at slots N_{i1}, N_{i2}, a shorter reaction time $t_{h'}$ and safety space $t_{h'} \times x_{ik+1}$ could be accepted. Meanwhile, collisions on the current and objective lanes must be avoided simultaneously.

For the vehicles in the current lane, equation (11) can be amended as follows:

$$\left(\left(1 - \delta_{ijk}\right) - \theta_{ik} - \theta_{jk} + 2\right) \times M + \left(x_{ik+1} - x_{jk+1}\right) \geq t_{h'} \times v_{jk+1} \tag{12}$$

such that $l_{ik} = l_{jk}$, $k = N_{i1}, N_{i2}$.

For the vehicles on the target lane l'_{ik}, equation (12) can be repeated such that $l'_{ik} = l_{ik} \pm 1; l_{jk} = l'_{ik}; k = N_{i1}, N_{i2}$.

3.2.3. Compulsory Lane-Change Constraints. When vehicle i leaves out the road segment, it is ensured that the current lane is identical to the final objective lane. In other words, lane change is guaranteed to accomplish before the destination. The following constraint should be satisfied:

$$x_{iN_{im_i}} < x_d, \tag{13}$$

where m_i is the total lane-change number and N_{im_i} represents the last lane-change interval.

3.3. Model Properties

Claim 1. The established model is an NP-hard problem. The size of the model can be estimated as follows:

$$\prod_{i=1}^{C'} \left(T - N_{C'_i} - \left(C' - i + 1\right)\right) * \left(C' - i\right) * \left(\frac{C}{3} + 1\right) * C * T, \tag{14}$$

where C and C' represent the total numbers of vehicles and "lane-changing" vehicles, respectively; C_i' is the i-th lane-change vehicle with ascending order; and $N_{C_i'}$ is the i-th lane-change interval.

The proposed model is equivalent to the traveling salesman problem (TSP). The variable described as "lane-changing time" is the critical element bringing out the uncertainties of vehicles' order in the same lane. For safety constraints in equations (9)–(12), the adjacent vehicles i, j cannot be determined that deduces the computational complexity of the model.

In equation (14), for any C_i', summed as $(C' - i)$, the feasible lane-change interval can take any integer value between $(N_{C_{i-1}'} + 1)$ and $(T - (C' - i))$. The minimum value coincides with Assumption (2), and the maximum value guarantees time possibility for the subsequent vehicles. Thus, just considering lane-changing time, the complexity can be expressed as $((T - (C' - i)) - (N_{C_{i-1}'} + 1) * (C' - i))$.

For considering entering into multiple vehicles' gaps, roughly $C/3$ vehicles can be counted on the target lane with $((C/3) + 1)$ gaps. Thus, the total number of vehicle orders can be expressed as follows:

$$\prod_{i=1}^{C'} \left(T - N_{C_i'} - \left(C' - i + 1\right)\right) * \left(C' - i\right) * ((C/3) + 1). \quad (15)$$

For a stationary vehicle order, the model employs $C * T$ variables to define acceleration for every vehicle at each time interval. Therefore, the size of the model is $\prod_{i=1}^{C'} (T - N_{C_i'} - (C' - i + 1)) * (C' - i) * (C/3 + 1) * C * T$. It is obvious that the computation complexity increases exponentially with C' and T and quadratically with C.

From the above analysis, for vehicle i, lane-change intervals $N_{i1}, N_{i2}, \ldots, N_{im_i} N_{im_i}$ are the core variables to generate uncertainty of lane occupation and vehicle order, resulting in large model size. The spontaneous idea is to decompose the problem into a two-phase model to reduce the computational complexity.

In the first phase, a commonly used strategy, ES is designed to define the lane-change interval and lane occupation combined with lane-change rules.

For example, if vehicle i on the initial lane l_{si} needs to change lane at $k = N_{i1}, N_{i2}\,(N_{i2} > N_{i1})$ to enter into the final lane l_{ei}, the following can be defined:

$$\text{If } k_i^{\text{in}} \le k < N_{i1}, \text{ then } l_{ik} = l_{si}, \quad (16)$$

$$\text{If } N_{i1} \le k < N_{i2}, \text{ then } l_{ik} = l_{i1} \pm 1, \quad (17)$$

$$\text{If } N_{i2} \le k_i^{\text{out}}, \text{ then } l_{ik} = l_{ei}. \quad (18)$$

In the second phase, a mixed integrated linear model constituted by equations (1)–(13)is solved, and equation (4) is the fitness of the first phase. With the determination of lane occupation's time, the mixed integrated linear model's complexity is reduced drastically. However, the vehicle orders dependent on chosen gap are also discrete random variables that can be further reduced in the solution space and illustrated in Section 4 explicitly.

The proposed strategy focuses on the lane occupation and vehicle sequence and decomposes the complex problem into two phases. On the one hand, the maneuver reduces the size of variables and constraints. On the other hand, the lane-change define solution is a commonly used and maturely developed algorithm that can rapidly obtain a satisfactory scheme.

4. Two-Phase Algorithm for Trajectory Management

4.1. The First Phase. Lane-change interval optimization. Evolution strategies are search procedures aiming to mimic the natural evolution of the large-scale stochastic and reduce the feasible solutions using elimination mechanisms based on objective function and mutation, crossover, and selection operations. This method is used for the upper level to determine the optimal lane-changing interval.

4.1.1. Coding for Individual Representation. It should be noted that the multidimensional coding combined with lane-change time and vehicles' order is not used. Although the multidimensional coding has the advantage of a simple points search operation, the critical defect is that the points' feasibility is not guaranteed. In the evolution process, a large number of infeasible solutions cause the loss of solutions' diversity. Therefore, the ES methods just eliminate one random element, and the vehicle orders' diversity is remained to solve in the second phase.

Implementing the ES for the trajectory optimization model requires the representation of the potential solution, which is a point in the feasible search space. Each solution is a vector consisting of integer variables, denoted as $N_{11}, N_{12}, \ldots, N_{1m_1} \cdots N_{n1}, N_{n2}, \ldots, N_{nm_n} N_{1m_1} N_{n1}, N_{n2} N_{nm_n}$. In the expression, N_{ij} represents the j-th lane-change slot of vehicle i, and m_i is the number of total maneuvers. The lane-change slots coding by integer representations can be defined with a series of requirements as follows:

(1) *Coding Approach* Coding approach: for each vehicle i, the coding variables $N_{i1}, N_{i2}, \ldots, N_{im_i} N_{im_i}$ satisfy $N_{i1} < N_{i2} < \cdots < N_{im_i}$. It is obvious that the posterior lane change is later than the former one. Spontaneously, two coding methods can be utilized to realize variables' feasibility. Table 2 introduces and compares the two approaches. In this paper, Approach 1 is chosen for variable coding by generating new individuals to eliminate duplicate ones and augment diversity. Approach 2 is prone to cause the latter variable to be out of range due to a large frontier value, and the individual representation is an unfeasible solution. Therefore, Approach 2 is abandoned.

(2) Variables range constraints: $N_{ij} < N_i\,(j = 1, 2, \ldots, m_i)$, where N_i is preset to ensure that vehicle i completes changing lanes before driving out.

Table 2: Reranking and construction coding methods.

	Approach 1: reranking	Approach 2: construction
Coding process	Rank coding variables $N_{i1}, N_{i2}, \ldots, N_{im_i} N_{im_i}$ with ascending sort for each i to constitute a new feasible individual.	Denote $N_{ij+1} = N_{ij} + \varepsilon_{ij}$, where ε_{ij} is a random integer number, and generate a valid individual by coding ε_{ij}.
Advantage	Easily practicable and code value restricted in variable range by simple constraints.	Guarantee individuals' diversity, which is the shortage of approach 1.
Disadvantage	Reduce diversity. For example, code $N_{i1}, N_{i2}, \ldots, N_{im_i} N_{i1}, N_{i2} N_{im_i}$ is identical with $N_{im_i}, N_{i2}, \ldots, N_{i1} N_{i1}$.	Difficult to satisfy variables' value range

(3) Initialization: a random value in the boundary range is assigned to each decision variable for every individual of the initial population.

$$N_{ij}^0 = k_i^{\text{in}} + r \text{ and } (0,1) \times \left(N_i - k_i^{\text{in}}\right). \tag{19}$$

4.1.2. Mutation. At generation g, for each parent vector $N_{11}^g, N_{12}^g, \ldots, N_{1m_1}^g, \ldots, N_{n1}^g, N_{n2}^g, \ldots, N_{mm_n}^g N_{1m_1}^g N_{n1}^g$, N_{n2}^g, $N_{nm_n}^g$ in the population, introduce indexes r_1, r_2 $(r_1 \neq r_2)$ to stamp different locations of vectors. The corresponding mutational vector is translated by three common strategies as follows:

First strategy: interconverting values at two random points

$$\begin{aligned} N_{r_1}^{g+1} &= N_{r_2}^g, \\ N_{r_2}^{g+1} &= N_{r_1}^g. \end{aligned} \tag{20}$$

Second strategy: permute the variables located in the interval $[r_1, r_2]$ in reverse order

$$N_j^{g+1} = N_{r_1+r_2-j}^g,\ r_1 \leq j \leq r_2. \tag{21}$$

Third strategy: insert variable located at the spot r_1 into spot r_2

$$N_{r_2}^{g+1} = N_{r_1}^g, \tag{22}$$

$$N_j^{g+1} = N_{j+1}^g,\ r_1 \leq j < r_2. \tag{23}$$

Random numbers r_1, r_2 are required to ensure that the two variables correspond to disparate vehicles. Reranking the new created vector is a subsequent maneuver to obtain a feasible solution.

The above three strategies can be repeated at stated times, aiming to search for better individuals in the neighbor region for each parent vector.

4.1.3. Crossover. Introduce indexes r_1, r_2, r_3 $(r_1 \neq r_2 \neq r_3)$. Two random parent vectors $N_1^g, N_2^g, \ldots, N_{\sum_{\forall i} m_i}^g N_{\sum_{\forall i} m_i}^g$ and $M_1^g, M_2^g, \ldots, M_{\sum_{\forall i} m_i}^g M_{\sum_{\forall i} m_i}^g$ at generation g are selected to generate new vectors using the following scheme:

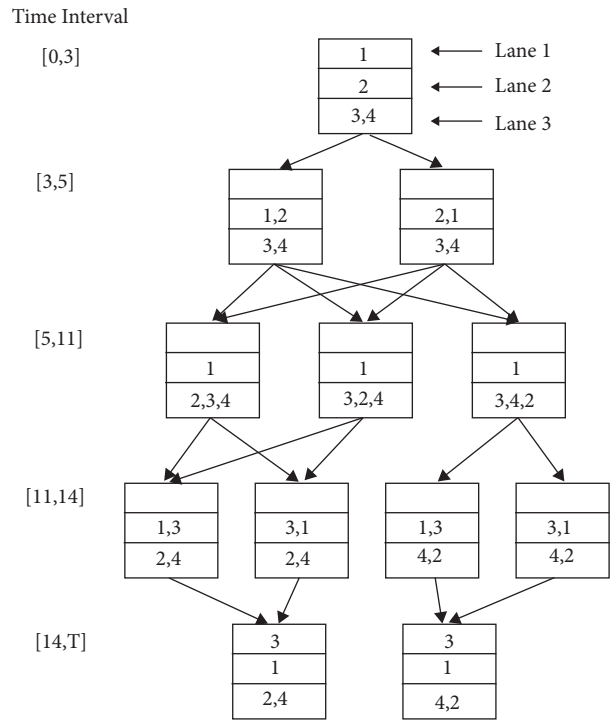


Figure 2: Demonstration example of order states on lanes.

$$\begin{aligned} N_j^{g+1} = N_j^g,\ M_j^{g+1} = M_j^g,\ 1 \leq j < r_1, \\ \text{or } r_2 \leq j < r_3, \end{aligned} \tag{24}$$

$$\begin{aligned} N_j^{g+1} = M_j^g,\ M_j^{g+1} = N_j^g,\ r_1 \leq j < r_2, \\ \text{or } r_3 \leq j \leq \sum_{\forall i} m_i. \end{aligned} \tag{25}$$

4.1.4. Selection. Only if the new-created vector yields a better fitness value, then update the parent individual. Otherwise, retain it.

4.1.5. Stopping Criterion. The search process is to break up when one of the two conditions is satisfied: (1) a maximum consuming time of iterations is reached and (2) improvement of the fitness value is not found for a preset number of generations. $|F^{g+\Delta g} - F^g| < \tau$, where τ is a small positive number.

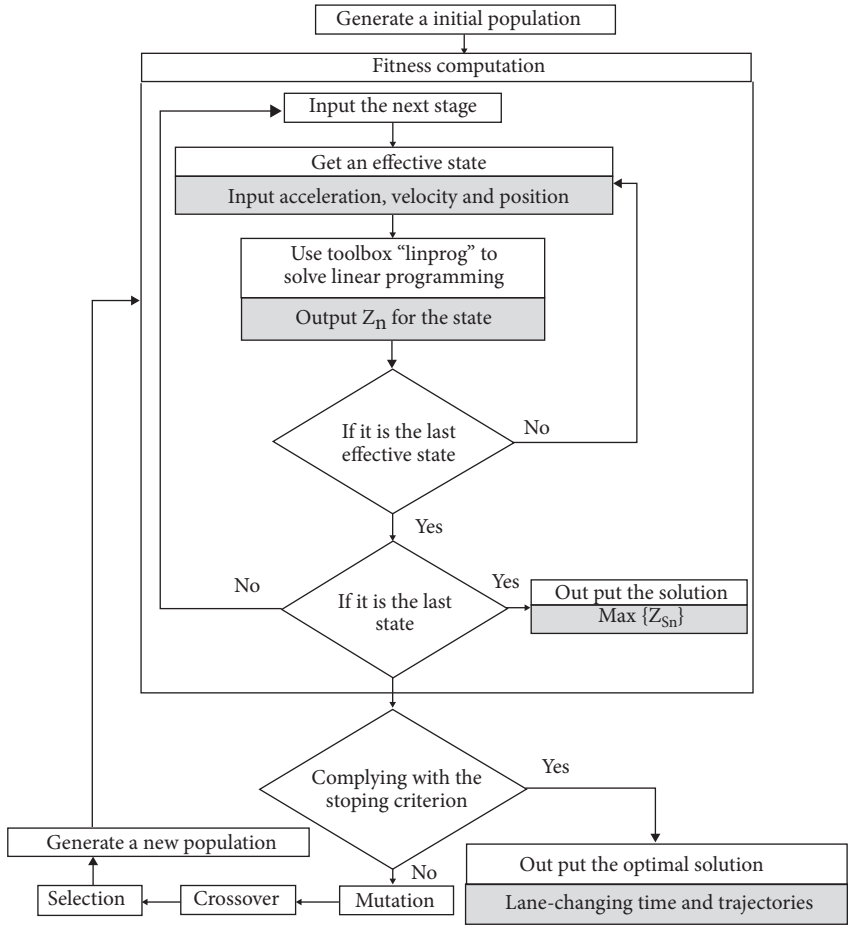


Figure 3: Computational flowchart.

(1) **Input:** initial velocity, acceleration, entering time, occupied lane, terminal lane, maximum acceleration and deceleration for each vehicle; road's length;
(2) **Output:** trajectory including acceleration, velocity, position, and an occupied lane for each interval;
(3) **Main loop:**
(4) **Outer iteration: Evolution strategy**
(5) **Initialization**: set $k = 1$, generating the first population of random individuals;
(6) **While** the stopping criterion is not satisfied, **do**:
(7) **Fitness computation**: for each individual, enter into inner iteration to obtain the individual's fitness;
(8) **Inner iteration**: Fitness computation
(9) **Divide** time zone T into D segments, $d = 1$;
(10) **While** $d \leq D$, for each terminal vehicles' order, do:
(11) **Update** position, velocity, acceleration, occupied lane, and vehicles' sequence using results of stage $d - 1$;
(12) **Use** MATLAB toolbox "lapdog" to solve linear programming;
(13) **Repeat** solving linear programming for each state and select effective states to enter into stage $d + 1$;
(14) **Then** update $d = d + 1$; repeat steps 10 to 13;
(15) **When** $d = D + 1$, iteration ends, output: individual's fitness $\max\{Z_{S_n}\}$; break inner iteration; enter into outer iteration;
(16) Perform mutation, crossover, and selection strategies on the kth generation of individuals to obtain the $(k + 1)$th generation; then set $k = k + 1$; repeat Steps 6 to 15;
(17) **When** the stopping criterion is satisfied, the two-phase algorithm ends, output: the optimum individual corresponding to vehicles' lane-changing time, acceleration, velocity, and position at each time interval.

Algorithm 1: Two-phase algorithm.

Table 3: Parameters values.

Variables	Description	Value
$[l_{e1}, l_{e2}, \ldots, l_{e10}]$	Initial lane vector	[1 2 3 1 2 3 1 1 2 3]
$[l_{s1}, l_{s2}, \ldots, l_{s10}]$	Terminal lane vector	[1 2 3 1 3 1 2 2 2 3]
$a_i^{\max}$	The maximum acceleration	3 m/s^2
$v_i^{\max}$	The maximum velocity	30 m/s
$d_i^{\max}$	The maximum deceleration	3 m/s^2
$t_h/t_{h'}t_{h^{\cdot}}$	Reaction time/short reaction time	0.3 s/0.2 s

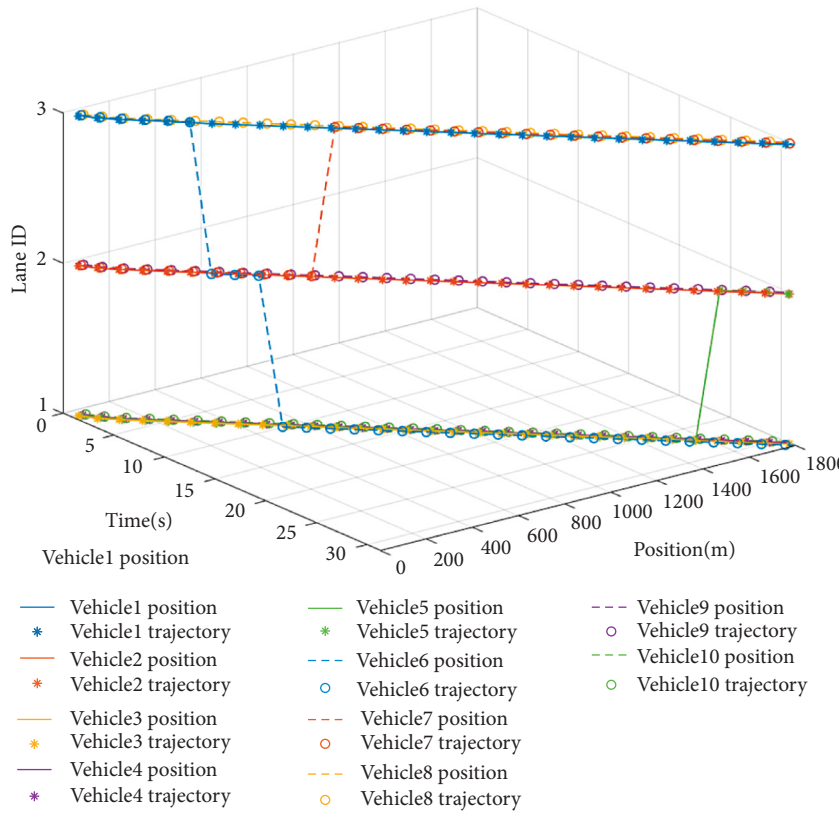


Figure 4: Optimal vehicles trajectories.

4.2. The Second Phase. Fitness Computation. To diminish the vehicle sequence's stochastics on the identical lane, compute fitness for each coding individual. Dynamic programming is introduced as follows:

(1) Reorder the variables $N_{11}, N_{12}, \ldots, N_{1m_1}, \ldots$ $N_{n1}, N_{n2}, \ldots, N_{nm_n} N_{nm_n}$ in ascending order and determine vehicles' sequence on each lane. To explain the entire process and list all the order states, Figure 2 presents a simple example of four vehicles coded as [3 5 11 14], corresponding to vehicles [1 2 3 3] with initial lanes [1 2 3 3] and terminal lanes [2 3 1 3]. It can be observed that the entire process has 48 different states, and the fitness for coding individuals [3 5 11 14] is the optimum one in the various states. Noticeably, the order state remains stable in the time range between two sequential lane-change spots.

(2) The second-part model will be solved as dynamic programming. Divide the entire process into a few stages according to lane-change spots. The status of vehicle order and the status transition for each stage can be confirmed as shown in Figure 2. Solving a maximization problem will obtain other statuses, including velocity, acceleration, and position. In equation (1), the objective is to minimize the total travel time. Conversely, the benefit function $Z_{s\in S_d}$ of state s at stage d can be defined as follows:

$$Z_{s\in S_d} = \max \sum_{i\in C} \sum_{k\in d} x_{ik}. \tag{26}$$

The above expression aims to maximize vehicles' total travel distance at each stage. It can be regarded as equal to the original objective, which can be guaranteed by setting proper lane-changing boundaries that vehicles cannot drive out at stage d.

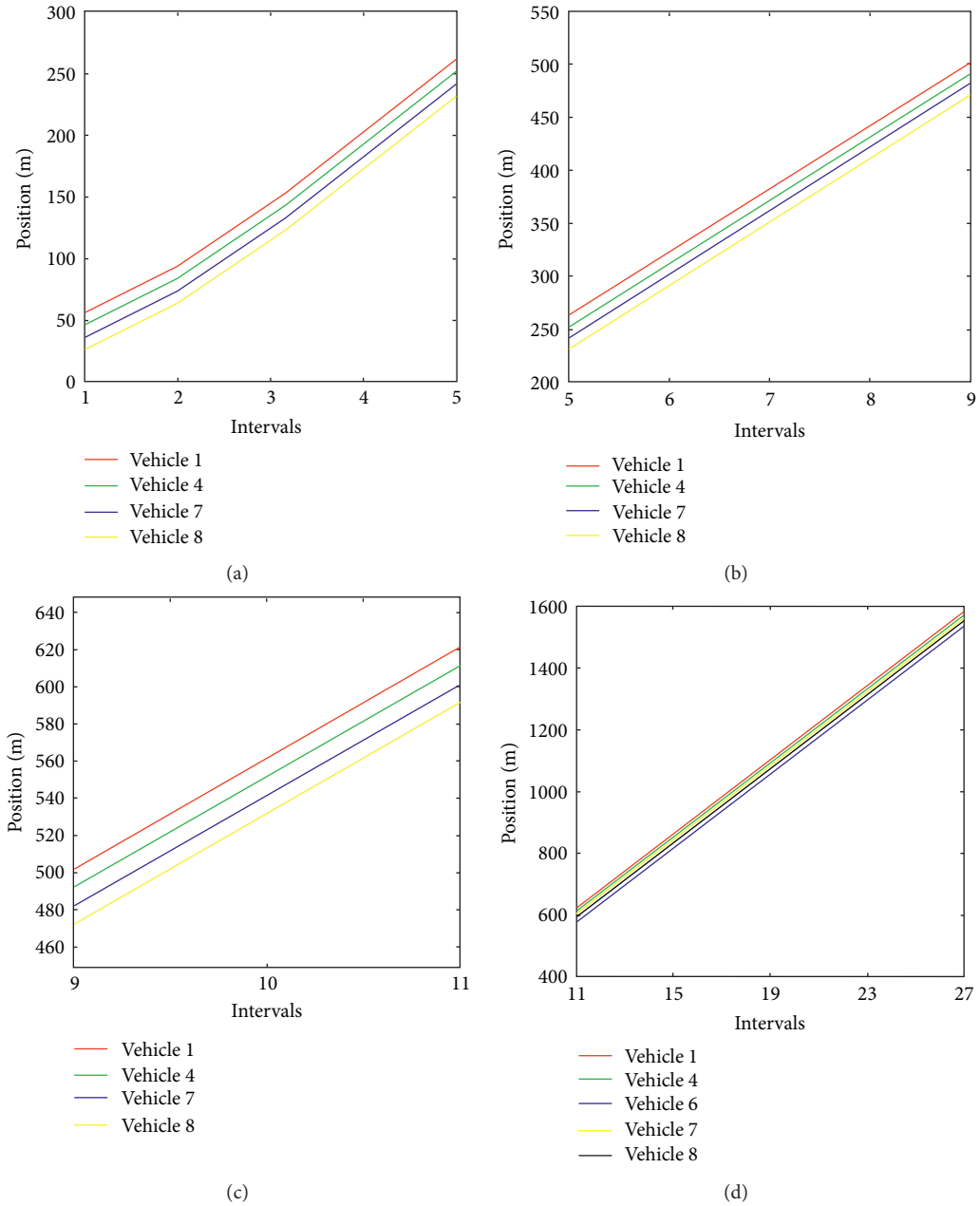
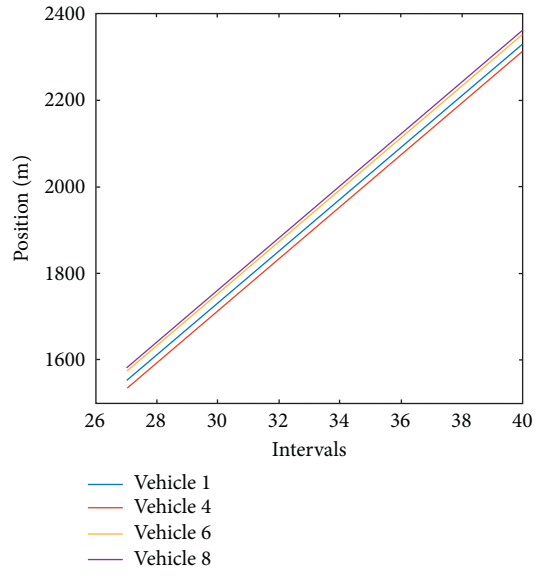


Figure 5: Continued.



(e)

Figure 5: Vehicles trajectories on lane 1: (a) 2–10 s, (b) 10–18 s, (c) 18–22 s, (d) 22–54 s, and (e) 54–80 s.

(3) At stage d, for each state $s \in S_d$, the terminal states set of stage d (the initial state of stage $d+1$) can be assured, and stage benefit with variables of each interval is to be computed by solving linear programming with objective (26) and constraints (2)–(13). The binary variables δ_{ijk} can be removed, and the constraints significantly decrease for the determined vehicle order.

(4) Define effective states to reduce the computation complexity. As illustrated in Figure 2 and equation (14), if C', T, and C are large, the states will dramatically increase. Theoretically, the lane-change vehicle is likely to insert into the space between any two adjacent vehicles at the end of the stage. However, it is time-consuming for several states; for example, in the condition that vehicle i located as x_i chooses to insert into the gap $[x_j, x_{j-1}]$, $x_i \ll x_j$ or $x_i \gg x_{j-1}$, maximum velocity deceleration must be enforced to satisfy the lane-changing criteria.

Definition of effective states: For state $s_{d_i} \in S_d$, S_d is states set in descending order by states' fitness at stage d. For $kk = \max(i | Z_{s_1} - Z_{s_{d_i}} \le \varepsilon_d)$ or kk is a preset value, $s_{d_i}\,(i \le kk)$ is defined as effective states at stage d, and the other states can be removed from S_d.

(5) For each coding $N_{11}, N_{12}, \ldots, N_{1m_1}, \ldots$

$N_{n1}, N_{n2}, \ldots, N_{nm_n}$, the time zone T is divided into $D = \sum_{i=1}^{n} m_i$ segments. At stage d, the linear programming is solved by the simplex algorithm to compute each effective initial state's objective. Individual's fitness is the maximum value among all the effective states at the terminal stage. The pseudocode of the two-phase algorithm for trajectory optimization is described in Algorithm 1. And a computational flowchart is presented in Figure 3 to display the algorithm more clearly.

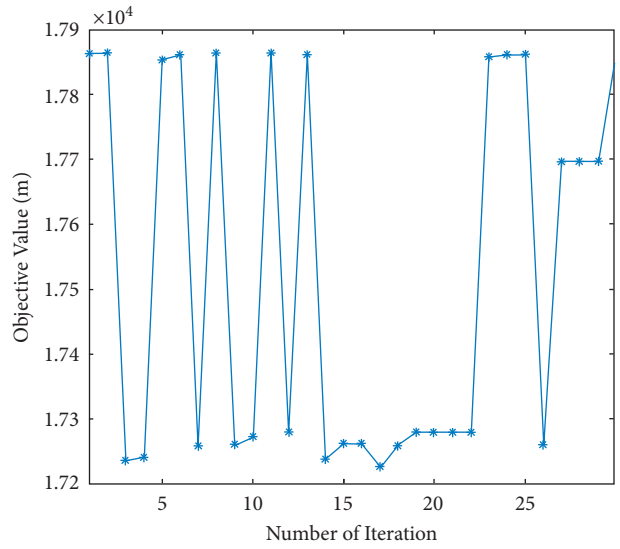


Figure 6: Optical objective value for the repeated running of the algorithm.

5. Numerical Example and Results

In this section, experiments are conducted to validate the proposed model and algorithm on a road segment consisting of three lanes, similar to Figure 1. The main targets are as follows: (1) optimize vehicles' trajectory on a numerical example and (2) compare with the no optimized case. All experiments were implemented using MATLAB and conducted on a PC workstation running Windows 10 with an Intel Xeon e5 3.5 GHz processor and 64 GB of main memory.

5.1. Optimized Vehicles' Trajectory. A road length $L = 1800$ m, 10 vehicles with an initial entering time $k_i^{\text{in}} = [0$

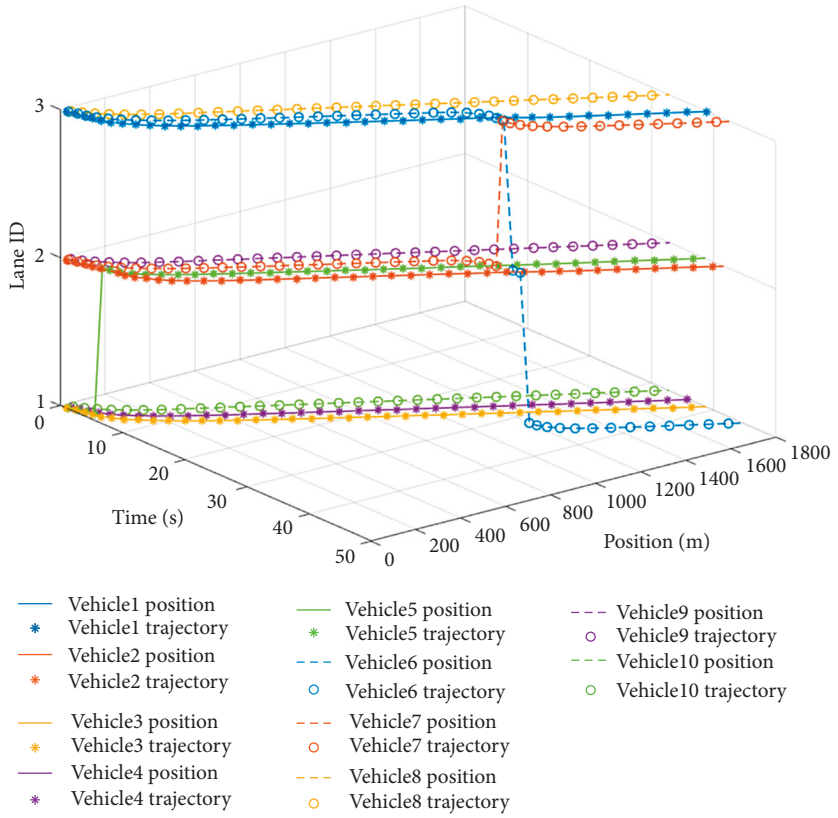


Figure 7: Optimal vehicle trajectories.

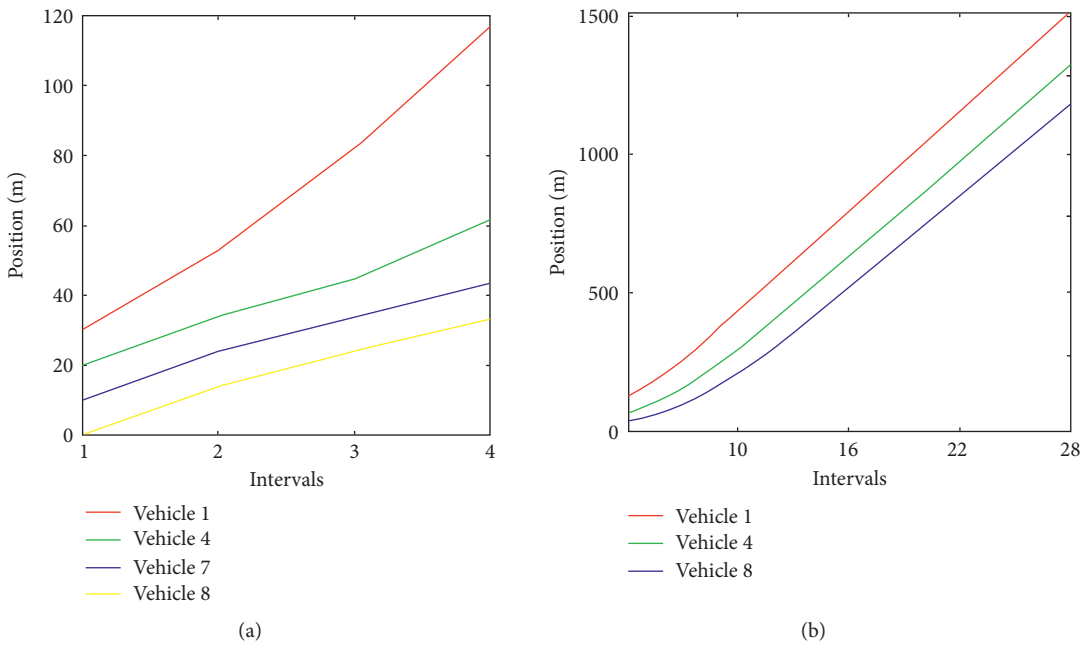


Figure 8: Continued.

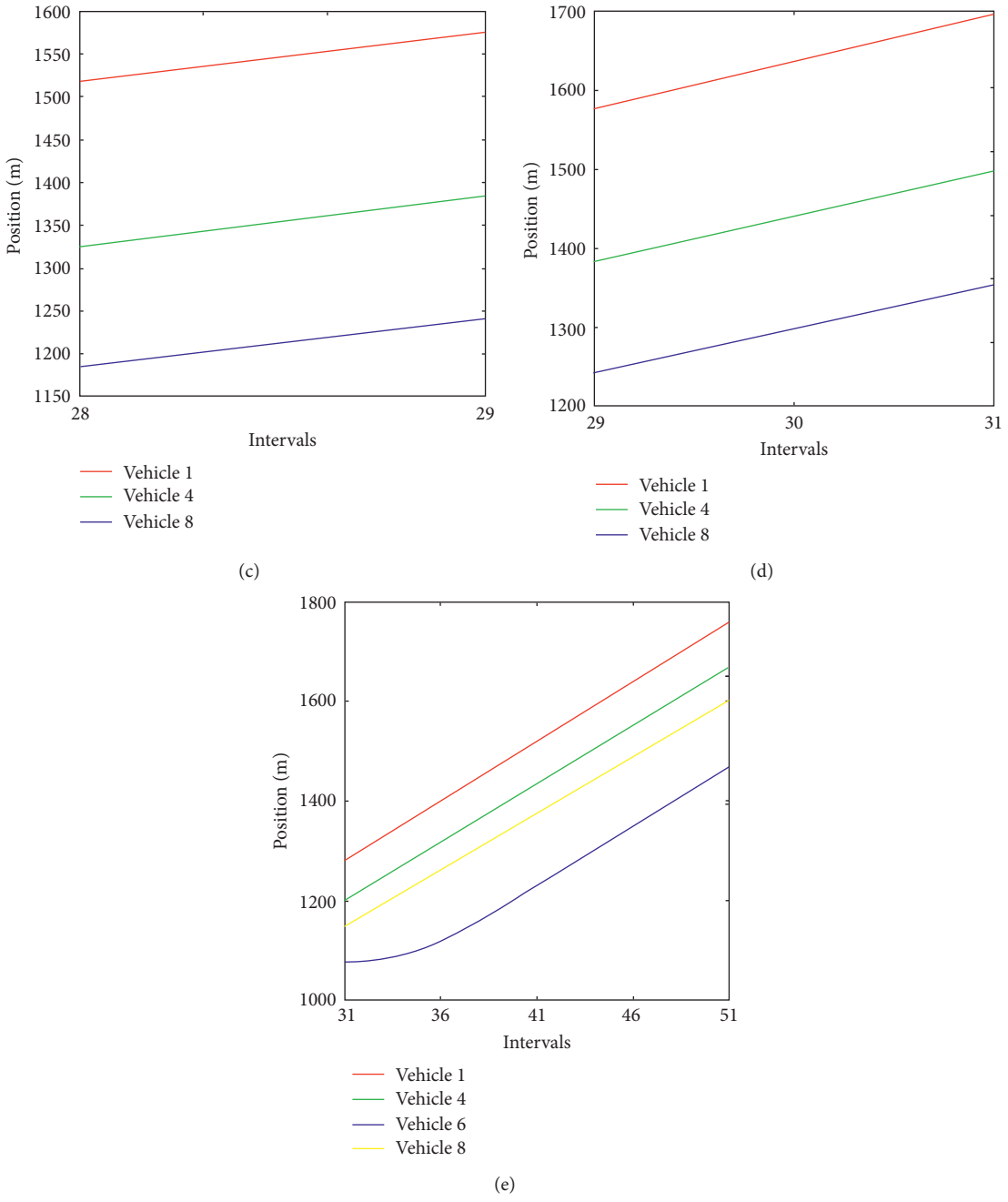


Figure 8: Comparison of vehicle trajectories on lane 1: (a) 2–8 s, (b) 8–56 s, (c) 56–58 s, (d) 58–62 s, and (e) 62–100 s.

0 0 2 4 4 4 6 6 6] ss will be optimized. To ensure that all vehicles can drive out, the time horizon is $T = 80$ s, and coding boundary in the ES algorithm is preset 60 s (which is estimated by maximum velocity), discretized into 40 and 30 intervals of equal duration, where each interval is 2 seconds due to precision deficiency and warranty of lane-change completeness. A reasonable spatial and temporary division is necessary for continuous traffic flow to realize discretization.

The values of other parameters are listed in Table 3. In each generation, 10 individuals are produced, and the entire algorithm is executed in the 60 s. To select effective states in the inner dynamic trajectories programming, $kk = 2$. The results are shown in Figures 4–6 .

Vehicles trajectories are displayed in Figure 4. Lane-change times are 22 s (vehicle 5), 10 s (vehicle 6), 18 s (vehicle 6), and 54 s (vehicle 7). It is observable that all vehicles drive on the target lane before the exit positioned at 1,800 m. The safe distance between the adjacent vehicles along the whole time zone is guaranteed (the vehicle's gap is larger than the rear vehicle's velocity multiple reaction time).

Figure 5 presents vehicles trajectories on lane 1 as a demonstration sample and shows the occupied lanes variation with time interval. All the vehicles are able to exit

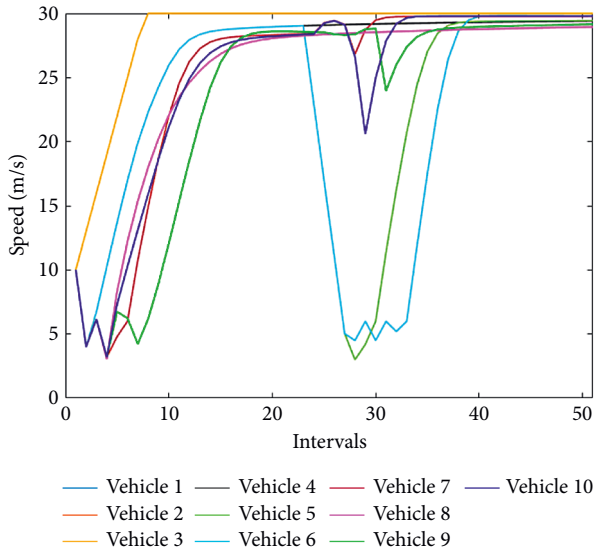


Figure 9: The vehicles' speeds.

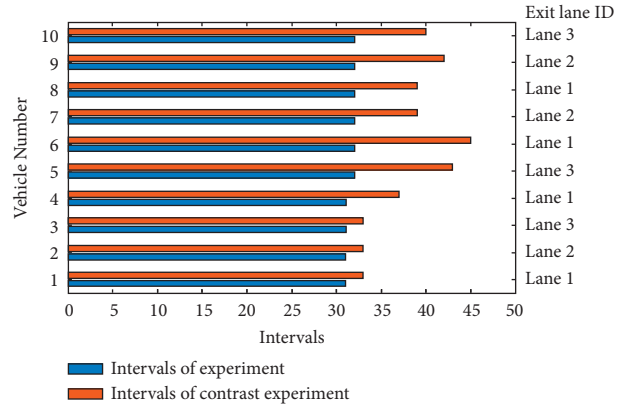


Figure 10: Total travel time of the vehicles.

before interval 33, which indicates that the coding range [0 30] is reasonable. The trajectory lines are smooth, and just several inflection points exist at lane-changing intervals. This phenomenon validates the results availability ensuring the vehicles are driven steadier. To analyze the influence of ES randomness, the algorithm was rerun 28 times repetitiously, setting the programming running time as 1 min. Figure 6 shows that the optimum fitness values' variation range remains within 600 m for 10 vehicles, and the randomness is acceptable in practical application.

5.2. Comparison with Unoptimizable Situation. In this section, the proposed trajectory programming is compared with an unoptimizable situation in which drivers follow the preceding car with intelligent driver model (IDM) and change lanes by self-decision as follows:

The IDM details can be expressed as follows:

$$a_{ik} = a_i^{\max}\left[1 - \left(\frac{v_{ik}}{v_i^{\max}}\right)^{\delta} - \left(\frac{s_k'}{D_{ik}}\right)^2\right], \tag{27}$$

$$s_k' = s_0 + \max\left(0,\ t_h v_{ik} + \frac{v_{ik}\Delta v_{ik}}{2\sqrt{a_i^{\max} b}}\right), \tag{28}$$

$$x_{ik+1} = x_{ik} + v_{ik}\Delta t + 0.5 a_{ik}\Delta t^2, \tag{29}$$

where $D_{ik} = x_{ik} - x_{i+1k}$ is the headway gap to the preceding vehicle (m), s_0 is the minimum safe gap for congested flow (m) computed by reaction time t_h, Δv_{ik} is the velocity difference (m/s), and b is the comfortable deceleration (m/s^2). In this part, $b = 3$ m/s^2; other parameters are identical to the values in Section 5.1. The self-decision lane-change strategy can be described as follows:

(1) If the following constraints are simultaneously satisfied, change to the adjacent lane directly:

$$x_{ik} - x_{jk} \geq x_{jk} \times t_{h'}, \tag{30}$$

$$x_{j-1k} - x_{ik} \geq x_{ik} \times t_{h'}, \tag{31}$$

where x_{j-1k} and x_{jk} are the positions of the preceding and the subsequent vehicles on the target lane, respectively.

(2) If the gap conditions in equations (30) and (31) are not satisfied, the following two cases can exist:

Case 1: If $x_d - x_{ik} > L'$ $(L' < L)$, a long distance is acceptable to await safe gap; vehicles choose IDM to follow cars

Case 2: If $x_d - x_{ik} < L'$, a speed adjustment mode will be activated to enlarge the vehicles' gap

If equation (30) is not satisfied,

$$a_{jk+1} = \max\left(\frac{\left(-a_{\max}, \left(v_{\min} - v_{jk}\right)\right.}{\Delta t}\right). \tag{32}$$

If equation (31) is not satisfied,

$$a_{ik+1} = \max\left(\frac{\left(-a_{\max}, \left(v_{\min} - v_{ik}\right)\right)}{\Delta t}\right). \tag{33}$$

The common parameters and the initial settings are assigned with identical values as in Section 5.1, executing the entire process in the same running environment. Vehicles' trajectories and other meaningful results are shown in Figures 7–9. Consistent with Section 5.1, the vehicle trajectories are presented in Figure 7, and the trajectories on lane 1 are illustrated in Figure 8. It can be seen that lane-change time spots are 58 s (vehicle 5), 56 s (vehicle 6), 62 s (vehicle 6), and 8 s (vehicle 7), different from the results in Section 5.1. Similar conclusions can be found that a safe gap is provided throughout the process, and all vehicles exist out of the target lanes.

Compared with Figure 4, more inflection points and longer periods of speed adjustment appear in the trajectory lines in Figure 8. The speeds of human-driven vehicles are depicted in Figure 9. As illustrated in Figure 9, speeds

fluctuate significantly throughout the time zone, particularly around lane-change intervals. In contrast, the speeds are significantly smoother in the AV trajectories depicted in Figure 5. This indicates that the proposed algorithm may improve driving stability, thereby lowering fuel consumption and gas emissions.

The reason is that when the HDVs intend to change lanes, they slow their speed to satisfy the safe gap in the shortest time, simultaneously inducing a significant speed oscillation range. But, for AVs, speed adjustment is dispersed in the whole time zone through global optimization.

Figure 10 shows that each vehicle spends more travel time on the whole journey in the contrastive example. The total time cost is saved by 20%, and the average speed is increased by 18%. The road length is longer, and the trajectory algorithm will obtain more significant optimum results for the entire process.

6. Discussion and Conclusions

This study studies the trajectory optimization problem for autonomous vehicles on particular roads with communication facilities. The objective of the proposed model is to search for optimal trajectories for autonomous vehicles. Meanwhile, the trajectories are restrained by some incorporating constraints, including safe gaps for collision avoidance, vehicle kinematic requirements, and occupied lane variations. Due to the diversity of lane-change spots and the order of the vehicles in each lane, a two-phase algorithm is proposed to diminish the uncertainties and improve the computation efficiency. A numerical example and a comparative experiment indicate the following meaningful findings:

(1) The proposed two-phase algorithm is able to solve optimum trajectories satisfying all the constraints in an acceptable running time. The trajectories are relatively smooth while simultaneously conforming to car-following, lane-changing behaviors, and maintaining steady speeds.

(2) Solving the trajectory problem has high computation complexity, and the required running time is relative to the total vehicles, time scope, and section's length. The evolution strategy can reduce the large scale of lane-change spots and search for optimum results. Cutting invalid vehicles' orders at each stage reserves sufficient optimum results and speeds up the solving process in the dynamic programming for fitness computation.

(3) The proposed two-phase algorithm demonstrates favorable computational performance, while some optimal solutions are lost. Due to the large scale and high stochastic in reality, it is reasonable and practical to locate a satisfactory solution in a limited time range.

(4) The comparative experimental results show that vehicles' lane-change spots positions are close to the road exit unless the target lane has low-density flow and provides a suitable vehicle gap. It can be explained that the drivers are opposed to speed adjustment and tend to wait for an opportunity for an adequate gap until driving near the exit. Also, the solution efficiency has been significantly improved by decreasing the total travel time in the proposed algorithm. It is observable that the AVS' trajectories approach system optimum by global optimization. Otherwise, the comparative experimental results are users' manual decisions through communication and just realize user optimum at a discrete-time interval.

As a result of the preceding discussions and conclusions, it can be concluded that the proposed model and algorithm can be used to manage traffic in an AV-only zone. A cost-effective and smooth trajectory scheme will be computed and transmitted to vehicles via a central system in a timely manner. It is anticipated that it will be used in congested areas. There are, however, some limitations. Notably, if the vehicles are not evenly distributed relative to one another in each lane at the entrance, the algorithm will be combined with a pretrip lane selection maneuver to achieve satisfactory results. Moreover, the ES algorithm yields an approximation result. Parameters and stopping criteria should be dynamically adjusted.

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 study was financially supported by the National Social Science Fund of China (20CGL018).

References

[1] Z. Chen, F. He, L. Zhang, and Y. Yin, "Optimal deployment of autonomous vehicle lanes with endogenous market penetration," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 143–156, 2016.

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

[3] Y. Kakimoto, M. Iryo-Asano, E. Orhan, and H. Nakamura, "A study on the impact of av-hdv mixed traffic on flow dynamics of single-lane motorway," *Transportation Research Procedia*, vol. 34, 2018.

[4] L. Alvarez and R. Horowitz, "Safe platooning in automated highway systems Part I: safety regions design," *Vehicle System Dynamics*, vol. 32, no. 1, pp. 23–55, 1999.

[5] Z. Chen, F. He, Y. Yin, and Y Du, "Optimal design of autonomous vehicle zones in transportation networks," *Transportation Research Part B: Methodological*, vol. 99, pp. 44–61, 2017.

[6] C. G. Cassandras, "Automating mobility in smart cities," *Annual Reviews in Control*, vol. 44, pp. 1–8, 2017.

[7] S. Bekhor and T. Toledo, "Investigating path-based solution algorithms to the stochastic user equilibrium problem," *Transportation Research Part B: Methodological*, vol. 39, no. 3, pp. 279–295, 2005.

[8] S. Bekhor, L. Reznikova, and T. Toledo, "Application of cross-nested logit route choice model in stochastic user equilibrium traffic assignment," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2003, no. 1, pp. 41–49, 2007.

[9] E. Köhler and M. Strehler, "Combining static and dynamic models for traffic signal optimization inherent load-dependent travel times in a cyclically time-expanded network model," *Procedia - Social and Behavioral Sciences*, vol. 54, pp. 1125–1134, 2012.

[10] S. Bekhor, C. Chorus, and T. Toledo, "Stochastic user equilibrium for route choice model based on random regret minimization," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2284, no. 1, pp. 100–108, 2012.

[11] L. Li and X. Chen, "Vehicle headway modeling and its inferences in macroscopic/microscopic traffic flow theory: a survey," *Transportation Research Part C: Emerging Technologies*, vol. 76, pp. 170–188, 2017.

[12] 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.

[13] L. Ye and T. Yamamoto, "Modeling connected and autonomous vehicles in heterogeneous traffic flow," *Physica A: Statistical Mechanics and Its Applications*, vol. 490, pp. 269–277, 2018.

[14] C. H. Hsu, "Cellular automata simulation for mixed manual and automated control traffic," *Mathematical and Computer Modelling*, vol. 51, no. 7-8, pp. 1000–1007, 2010.

[15] J. Li, C. Wang, and S. He, "Dynamic traffic shockwave speed estimation in connected vehicle environment," in *Proceedings of the 2017 4th International Conference on Transportation Information and Safety (ICTIS)*, pp. 54–59, IEEE, Banff, AB, Canada, August 2017.

[16] J. N. Hooker, "Optimal driving for single-vehicle fuel economy," *Transportation Research Part A: General*, vol. 22, no. 3, pp. 183–201, 1988.

[17] T. Tang, W. Shi, H. Shang, and Y Wang, "A new car-following model with consideration of inter-vehicle communication," *Nonlinear Dynamics*, vol. 76, no. 4, pp. 2017–2023, 2014.

[18] T.-Q. Tang, W.-F. Shi, H.-Y. Shang, and Y.-P. Wang, "An extended car-following model with consideration of the reliability of inter-vehicle communication," *Measurement*, vol. 58, pp. 286–293, 2014.

[19] M. Werling, S. Kammel, J. Ziegler, and L Gröll, "Optimal trajectories for time-critical street scenarios using discretized terminal manifolds," *The International Journal of Robotics Research*, vol. 31, no. 3, pp. 346–359, 2012.

[20] X. He, H. X. Liu, and X. Liu, "Optimal vehicle speed trajectory on a signalized arterial with consideration of queue," *Transportation Research Part C: Emerging Technologies*, vol. 61, pp. 106–120, 2015.

[21] X. Wu, X. He, G. Yu, A. Harmandayan, and Y. Wang, "Energy-optimal speed control for electric vehicles on signalized arterials," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2786–2796, 2015.

[22] Y. Luo, Y. Xiang, K. Cao, and K Li, "A dynamic automated lane change maneuver based on vehicle-to-vehicle communication," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 87–102, 2016.

[23] B. Gutjahr, L. Gröll, and M. Werling, "Lateral vehicle trajectory optimization using constrained linear time-varying MPC [J]," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1586–1595, 2016.

[24] B. Liu, D. Ghosal, C.-N. Chuah, and H. M. Zhang, "Reducing greenhouse effects via fuel consumption-aware variable speed limit (FC-vsl)," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 111–122, 2012.

[25] K. Ahn, H. A. Rakha, and S. Park, "Ecodrive application," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2341, no. 1, pp. 1–11, 2013.

[26] G. Lu, Y. Nie, X. Liu, and D Li, "Trajectory-based traffic management inside an autonomous vehicle zone," *Transportation Research Part B: Methodological*, vol. 120, pp. 76–98, 2019.

[27] J. T. Betts, "Survey of numerical methods for trajectory optimization [Review]," *Journal of Guidance, Control, and Dynamics: A Publication of the American Institute of Aeronautics and Astronautics Devoted to the Technology of Dynamics and Control*, vol. 21, no. 2, pp. 193–207, 1998.

[28] N. Jiang, "Optimal signal design for mixed equilibrium networks with autonomous and regular vehicles," *Journal of Advanced Transportation*, vol. 2017, Article ID 5649823, 13 pages, 2017.

[29] A. Ghiasi, O. Hussain, Z. Qian, and X. Li, "A mixed traffic capacity analysis and lane management model for connected automated vehicles: a Markov chain method," *Transportation Research Part B: Methodological*, vol. 106, pp. 266–292, 2017.

[30] S. Zhang, W. Deng, and Q. Zhao, "Dynamic trajectory planning for vehicle autonomous driving," in *Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4161–4166, IEEE, Manchester, UK, October 2013.

[31] O. Von Stryk and R. Bulirsch, "Direct and indirect methods for trajectory optimization," *Annals of Operations Research*, vol. 37, no. 1, pp. 357–373, 1992.

[32] K. Ghoseiri and S. F. Ghannadpour, "Multi-objective vehicle routing problem with time windows using goal programming and genetic algorithm," *Applied Soft Computing*, vol. 10, no. 4, pp. 1096–1107, 2010.

[33] W. Ho, G. T. S. Ho, P. Ji, and H. C. W. Lau, "A hybrid genetic algorithm for the multi-depot vehicle routing problem," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 4, pp. 548–557, 2008.

[34] N. Barnier and P. Brisset, "Optimization by hybridization of a genetic algorithm with constraint satisfaction techniques," in *Proceedings of the 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98TH8360)*, pp. 645–649, IEEE, AK, USA, May 1998.

[35] F. Zhou, X. Li, and J. Ma, "Parsimonious shooting heuristic for trajectory design of connected automated traffic part I: theoretical analysis with generalized time geography," *Transportation Research Part B: Methodological*, vol. 95, pp. 394–420, 2017.

[36] S. Gong, J. Shen, and L. Du, "Constrained optimization and distributed computation based car following control of a connected and autonomous vehicle platoon," *Transportation Research Part B: Methodological*, vol. 94, pp. 314–334, 2016.

[37] J. Koshal, A. Nedić, and U. V. Shanbhag, "Distributed multiuser optimization: algorithms and error analysis," in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) Held Jointly with 2009 28th Chinese Control Conference*, pp. 4372–4377, IEEE, Shanghai, China, January 2009.

[38] J. Ma, X. Li, F. Zhou, J. Hu, and B. B. Park, "Parsimonious shooting heuristic for trajectory design of connected automated traffic part II: computational issues and optimization," *Transportation Research Part B: Methodological*, vol. 95, pp. 421–441, 2017.

[39] 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.

[40] Z. Wang, G. Wu, and M. Barth, "Distributed Consensus-Based Cooperative Highway On-Ramp Merging Using V2X communications," SAE Internationals, Pennsylvania, USA, SAE Technical Paper, 2018.

[41] Z. Yao, B. Zhao, T. Yuan, H. Jiang, and Y. Jiang, "Reducing gasoline consumption in mixed connected automated vehicles environment: a joint optimization framework for traffic signals and vehicle trajectory," *Journal of Cleaner Production*, vol. 265, Article ID 121836, 2020.

[42] L. Li and F. Y. Wang, *Advanced Motion Control and Sensing for Intelligent Vehicles*, Springer Science & Business Media, Berlin, Germany, 2007.

[43] C. Katrakazas, M. Quddus, W.-H. Chen, and L. Deka, "Real-time motion planning methods for autonomous on-road driving: state-of-the-art and future research directions," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 416–442, 2015.

[44] D. Gonzalez, J. Perez, V. Milanes, and F. Nashashibi, "A review of motion planning techniques for automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1135–1145, 2016.

[45] W. Liu, Z. Li, L. Li, and F.-Y. Wang, "Parking like a human: a direct trajectory planning solution," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 12, pp. 3388–3397, 2017.

[46] Z. Yao, H. Jiang, and Y. Cheng, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, 2020.

[47] X. T. Yang, K. Huang, Z. Zhang, Z. A. Zhang, and F. Lin, "Eco-driving system for connected automated vehicles: multi-objective trajectory optimization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 12, pp. 7837–7849, 2021.

[48] Y. Wang, P. Cai, and G. Lu, "Cooperative autonomous traffic organization method for connected automated vehicles in multi-intersection road networks," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 458–476, 2020.

[49] H. Yao and X. Li, "Lane-change-aware connected automated vehicle trajectory optimization at a signalized intersection with multi-lane roads," *Transportation Research Part C: Emerging Technologies*, vol. 129, Article ID 103182, 2021.

[50] X. Liu, J. Liang, and J. Fu, "A dynamic trajectory planning method for lane-changing maneuver of connected and automated vehicles," *Proceedings of the Institution of Mechanical Engineers - Part D: Journal of Automobile Engineering*, vol. 235, no. 7, pp. 1808–1824, 2021.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 6093217, 18 pages
https://doi.org/10.1155/2022/6093217

Research Article

Evaluating the Performance of Connected and Automated Vehicles in Fixed Signal-Controlled Conventional Intersections and Superstreets with Platooning-Based Trajectory Planning

Shaojie Liu[1] **and Wei David Fan**[2]

[1]*USDOT Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE), Department of Civil and Environmental Engineering, University of North Carolina at Charlotte, 9201 University City, Charlotte 28223, NC, USA*
[2]*USDOT Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE), Department of Civil and Environmental Engineering, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte 28223, NC, USA*

Correspondence should be addressed to Wei David Fan; wfan7@uncc.edu

Received 5 March 2022; Revised 6 May 2022; Accepted 14 May 2022; Published 2 June 2022

Academic Editor: Yanyan Qin

Connected and autonomous vehicles (CAVs) are emerging technology that attracts the interests of many transportation professionals and computational scientists. Several recent studies have investigated different model frameworks of CAVs in different transportation environments, such as on freeways and at conventional intersections. Nevertheless, few efforts have been made to investigate the performances of CAVs at innovative intersections, and the lack of knowledge can result in an inaccurate prediction of CAVs performances in the existing transportation network. This research intends to mitigate this research gap by studying the traffic delay and fuel consumption of CAVs in the environment of the superstreet and its equivalent conventional intersection through simulation-based experiments. A real-world superstreet in Leeland, NC, is selected and used. A conventional intersection with equivalent road designs is established in the simulation platform to make a comparison with the selected superstreet. This research develops both platooning and trajectory planning modeling frameworks to examine the implications of CAVs with different capabilities. The Intelligent Driver Model (IDM) is selected and applied to model the CAV behaviors, while Wiedemann 99 (W99) is used to model Human-Driven Vehicles (HDVs). The simulation results demonstrate the efficiency of both platooning and trajectory planning, respectively. Different effects of CAVs in the superstreet and its equivalent conventional intersection are observed. The findings from this research can provide an important reference for transportation planners and policymakers in predicting the influence of CAVs on the existing transportation infrastructure.

1. Introduction

CAVs are promising technology and can yield significant impacts in various transportation environments. Extensive efforts have been devoted to exploring the potential effects of CAVs in various transportation environments, including freeways [1–3], roundabouts [4, 5], and conventional intersections [6, 7]. In these studies, assumptions were often made for CAVs' capabilities, such as trajectory planning, shorter headways, accurate controls, shorter reaction times, and communication with other vehicles or infrastructure, such as traffic signals [8, 9]. However, few efforts have been devoted to investigating the impact of CAVs on the operational performance of innovative intersections. Among the numerous innovative intersection designs, superstreet is one of them that has been successfully implemented in several places in the US [10].

Therefore, this research intends to mitigate the research gap by exploring the different performances of CAVs in the environments of the superstreet and conventional intersection. The research uses Simulation of Urban Mobility (SUMO) as the simulation platform due to its rich

Application Programming Interfaces (APIs) to build proper behavior models of CAVs and HDVs. To be specific, different car-following models are considered for CAVs and HDVs, respectively. To account for the car-following characteristics of HDVs, W99 is selected for its consideration of both mechanical features of vehicles and the randomness of human drivers. As for CAVs, IDM is selected and used due to its intuitive measurable parameters and accurate controls of vehicle movements. This research also takes account of popular advanced features of CAVs including platooning and trajectory planning strategies. A superstreet from North Carolina in the real world is selected for the case study for its typical superstreet design and traffic flow characteristics. An equivalent conventional intersection with a similar road configuration is also designed and used in the same simulation platform to make a fair comparison. The performance evaluation also explicitly accounts for different traffic demands and market penetration rates.

The rest of the paper is laid out as follows: First, this research gives a brief literature review on the superstreet and CAVs technology, respectively. Second, this paper introduces the modeling framework for CAVs and HDVs, especially for the behavior models of CAVs. Third, this paper presents and discusses relevant results generated from the designed scenarios. Then conclusions are drawn based on the research findings.

2. Literature Review

Superstreet is one of the popular innovative intersection designs which has been applied in numerous places, especially in the states of North Carolina and Maryland in the US [10]. By allocating more green time for vehicles from the main road, superstreets could yield less traffic delay compared to conventional intersections, especially when there are significant unbalanced traffic volumes between the main road and the minor road [11, 12]. Hummer et al. [13] conducted a study investigating a superstreet corridor situated in North Carolina. Conclusions were made that the superstreet could potentially yield travel time and safety benefits. In addition to the studies on the performance's comparison between superstreet and its equivalent conventional intersection, Xu et al. [14, 15] investigated the optimal design of U-turn offset length and signal timing plan using an analytical approach and optimization methods, respectively.

In recent years, CAVs have become promising solutions to reduce congestion at intersections due to their shorter headways and trajectory guidance capabilities. The performances of CAVs are robust in different transportation scenarios, and they can consistently yield less travel time, traffic delay, and fuel consumption [7, 16, 17]. Table 1 provides a summary of recent existing studies on CAVs categorized by transportation environments.

CAVs and HDVs are often modeled with different car-following models. On the one hand, W99 is the popular HDV model for its rich parameters that capture the randomness of human driver behaviors [1, 28]. On the other hand, CAVs can be modeled with IDM, the adaptive cruising control (ACC) model, or the cooperative adaptive cruising control (CACC) model. These models all have intuitive measurable parameters for CAVs. CACC model is developed based on the ACC model by adding two-way communication between the preceding vehicles and the following vehicles. With CACC, the CAVs can travel on the roads with shorter headways (0.6s) compared to ACC vehicles [29].

Platooning is a unique feature for CAVs for its communication capabilities with other vehicles. A platoon of CAVs can travel on the roads with homogenous speeds and shorter headways [30]. Moreover, when it comes to planning trajectories, researchers only need to determine the trajectories of leading vehicles, and the rest of the vehicles inside the platoon can follow the trajectory of the leading vehicle. Different studies may have different definitions for the platoon. Feng et al. [31] and Yu et al. [32] defined the platoon as a group of vehicles that can travel through the intersection within one phase, while Ye et al. [33] defined the platoon as a group of vehicles that share similar speeds and have a close distance in between. In addition, some studies specifically defined the car-following behaviors of the vehicles inside the platoon [34, 35].

With trajectory planning, CAVs can follow a calculated velocity or acceleration/deceleration rate at each time step when approaching the intersection. The calculated velocity/acceleration profiles may achieve certain optimal traffic performance measures such as minimal travel time, fuel consumption, or emissions [31–33]. When the objectives are fuel consumption or carbon dioxide emissions, the objective functions often result in a nonlinear function form whose analytical solutions are hard to obtain. Hence, nonlinear programming, dynamic programming, and other meta-heuristic methods like the genetic algorithm (GA) are popular approaches to solving such problems [32, 33, 36, 37]. Nevertheless, computational efficiency is an inevitable great challenge in the real world when these approaches are in deployment. Notably, trajectory planning that is combined with signal optimization can eliminate vehicle stops [31, 32]. Adaptive signal controls with knowledge of vehicle arrival patterns may perform significantly better than trajectory planning standalone. Nevertheless, the combined optimization of trajectory planning and signal timing has limitations in the requirement of the full market penetration rate of CAVs and considerable investment in the hardware installation and maintenance of traffic lights. Thus, this joint coordination between the traffic light and CAVs is likely to be implemented in the latter stages of CAV development. In an environment where adaptive signals are not available, Green Light Optimal Speed Advisory (GLOSA) was proposed in the fixed signal-controlled intersections [38, 39]. With GLOSA, an optimal speed could be determined for CAVs based on the remaining green time or the green phase initiation time to reduce the number of stops or total waiting time.

Through literature review, a general modeling framework of CAVs and HDVs can be identified. This research adopts concepts including car-following models, platooning, and trajectory planning to model the behaviors of CAVs. Also, it is found that few studies have been conducted on investigating the CAVs' performances in innovative intersections. By simulating the HDVs and CAVs in the environment of the superstreet and equivalent conventional

Table 1: Recent studies on the CAVs categorized by transportation environments.

Transportation environments	Authors	Year	CAV features
Freeway	Guo et al. [3]	2020	CACC, platoons, cooperative merging
	Adebisi et al. [18]	2020	CACC models
	Liu and Fan [1]	2020	CAVs with revised intelligent driver model
	Chityala et al. [19]	2020	CAVs with shorter headways
	Hu and Sun [20]	2019	Cooperative lane changing control, cooperative merging control
Conventional intersection	Han et al. [7]	2020	Platooning-based trajectory planning with optimal control framework
	Pourmehrab et al. [21]	2020	CAVs with an intelligent intersection control algorithm (IICA) and hybrid autonomous intersection management (H-AIM)
	Guo et al. [22]	2019	Joint optimization of vehicle trajectory and intersection controller with combined dynamic programming and shooting heuristic approach
	Li and Zhou [23]	2017	Signal timing optimization with the brand and bound algorithm considering mixed traditional vehicles
	Zhou et al. [24]	2017	Parsimonious shooting heuristics algorithm to construct vehicle trajectories on a signalized highway segment
	Jiang et al. [6]	2017	Eco approaching at an isolated signalized intersection under partially connected and automated vehicles environment
	Dresner and Stone [25]	2008	Reservations based algorithm in the lightless intersection
Roundabout	Mohebifard and Hajbabaie [4]	2021a	Optimization on trajectory and merging sequence; customized solution technique that transforms the two-dimensional optimization problem into a combination of easier one- and two-dimensional subproblems
	Mohebifard and Hajbabaie [5]	2021b	Trajectory control in a roundabout with a mixed fleet of automated and human-driven vehicles
	Martin-Gasulla and Elefteriadou [26]	2021	Roundabout management algorithm for trajectory planning of CAVs
	Chalaki et al. [27]	2020	Trajectory planning control framework for roundabout

intersection, this research can close the identified research gap and provide an important reference for policymakers.

3. Methodology

This section illustrates the methodology of modeling HDVs and CAVs procedures. The researchers utilize the W99 model that is calibrated by GA to model the HDVs in superstreets, while CAVs are modeled with IDM. Moreover, platooning and trajectory planning schemes are also developed for CAVs.

3.1. Information on the Selected Location for Case Study. A superstreet situated in Leeland, NC is identified for the case study. This superstreet is selected for its typical geometric design and traffic flow characteristics. The traffic characteristic information on the selected superstreet is available in Hummer et al. [13]. Figure 1 shows the selected superstreet and signal locations in Google Maps, and Table 2 provides the traffic characteristics information. The maximum speed limits are set as 29 m/s (i.e., 65mph) for the main road and 15.6 m/s (i.e., 35 mph) for the minor road. The four minor intersections in the superstreet system are all signal controlled with a cycle length of 120 s.

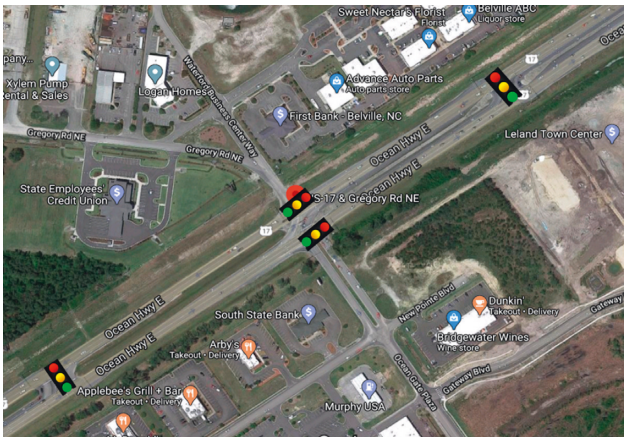


Figure 1: Selected superstreet for the case study and signal locations (adapted from the screenshot of google maps).

3.2. Car-Following Models

3.2.1. IDM Model. IDM was developed by Treiber et al. [40]. It is a collision-free model with intuitively measurable parameters. Due to these advantages, the IDM has been popularly used in modeling CAVs [1, 41, 42]. The acceleration rate in IDM is a function of the velocity of the subject vehicle, the gap to the preceding vehicle, and the velocity difference to the preceding vehicle, as (1) and (2) show below:

$$a(s, v, \Delta v) = a_m \left(1 - \left(\frac{v}{v_d}\right)^{\propto} - \left(\frac{s^*(v, \Delta v)}{s}\right)^2\right), \quad (1)$$

$$s^*(v, \Delta v) = s_0 + vT + \frac{v \times \Delta v}{2\sqrt{a_m b}}, \quad (2)$$

Table 2: Traffic characteristic information on the superstreet at Leeland, NC.

Approach	Average speed (m/s)	Peak hour demand	Average stops	Travel time (minutes)
EBL	5.99	18	3	2.45
EBR	6.93	20	2	1.38
EBT	5.68	9	2	2.25
NBL	8.00	20	1	1.17
NBR	14.08	71	0	0.64
NBT	14.75	2029	0	0.81
SBL	5.72	321	1	1.26
SBR	14.26	38	0	0.4
SBT	19.58	1637	0	0.58
WBL	8.09	66	2	2
WBR	7.69	345	1	0.89
WBT	5.05	11	2	2.09

where a indicates the desired acceleration rate; a_m is the maximum acceleration rate; v denotes the current speed; v_d indicates the desired speed (assumed equal to the speed limit in this research); v represents the speed difference between the subject vehicle and its preceding vehicle; $\propto$ means the acceleration exponent, which is set as 4 in this research; s is the current gap distance between the subject vehicle and its preceding vehicle; s_0 denotes the standing distances (2.5 m); T represents the desired headway (1s); b is the maximum deceleration rate.

3.2.2. W99 Model. W99 is a microsimulation psychophysical model which has ten parameters available for calibration. These ten parameters are intuitively consistent with human driver behaviors with certain randomness. To ensure that the W99 can represent the local traffic accurately, the ten parameters are calibrated to ensure that the average speeds on each approach in the simulation are matched with the ones that were from the field survey, according to Hummer [13].

Considering the data availability, this research selects the minimal difference between simulated average speeds and observed average speed for each approach as the objective function used in the calibration process. An overall difference within 15% is regarded as acceptable performance.

$$\frac{\left(\sum_i^N \left|v_{o,i} * - v_{s,i}\right| / v_{o,i}\right)}{N}, \tag{3}$$

where $v_{o,i}$ and $v_{s,i}$ are the observed and simulated average speed for approach i, respectively, and N indicates the total number of approaches.

A genetic algorithm is utilized to minimize the difference between the observed average speeds and simulated average speeds for each approach. GA is a popular and efficient approach to calibrating the car-following model parameters. For brevity, this research skips the introduction of GA. Readers may refer to existing studies of GA calibration for more details [43]. The population size and the maximum number of generations are set as 10 and 20, respectively. The final difference becomes stable at 11%, which is recognized as an acceptable difference. The obtained parameter values from GA are presented in Table 3. The lane changing movement is controlled by the default car-following model in SUMO, i.e., LC2013 [44].

3.3. Platooning Schemes. Vehicle platooning is one of the advanced features of CAVs. It can only be achieved with CAVs that have communication capabilities with other vehicles. Two assumptions were often made with CAVs platooning. One is shorter headways for vehicles inside a defined platoon, and the other is homogenous speeds. With shorter headways and homogenous speeds, the vehicles inside the same platoon can be regarded as a single unit to travel on the road, which can increase the capacity of the roads and also reduce the computational complexity when trajectory planning is involved. This research has also adopted these concepts to fully release the potentiality of CAVs.

3.3.1. Platoon Formulation and Splitting. The platoon control system in this research iterates all active vehicles in the simulation environment and checks whether the neighboring vehicles meet the requirements for the platoon formulation. The requirements are that the vehicles

(1) are in the same lane

(2) stay within the range of a certain distance

(3) share the same path

If the requirements above are met, then the system can define such a group of vehicles as a platoon and thus share the same speed with the leading vehicle. However, if any of the vehicles inside the platoon fail to suffice these requirements, then the platoon splits up and switches back to the default car-following model.

There is one more condition guaranteeing platoon splitting. When the platoon is approaching an intersection, the remaining green time g_p may not be sufficient for all vehicles in a platoon to pass the intersection together, especially when the platoon size is large. Thus, to make the platoon system practical, the vehicles with platooning are assumed to have the knowledge of remaining green time. With the information on the remaining green time g_p, the platooning system checks whether all vehicles inside a platoon can pass the intersection or not through,

Table 3: GA calibrated W99 parameter values.

Parameters	Interpretation	Default values	Calibrated values
CC0	Average standstill distance (meter)	1.4	1.287251
CC1	Headway (seconds)	1.2	1.569918
CC2	Longitudinal oscillation (meters)	8	1.28187
CC3	Start of deceleration process (seconds)	−12	−12.3849
CC4	Minimal closing Δv(m/s)	−1.5	−2.398
CC5	Minimal opening Δv(m/s)	2.1	0.324976
CC6	Speed dependency of oscillation (10^{-4} rad/s)	6	4.047425
CC7	Oscillation acceleration-m/s^2	0.25	0.29111
CC8	Acceleration rate when starting (m/s2)	2	4.582238
CC9	Acceleration behavior at 80 km/h (m/s^2)	1.5	4.261776

$$g_p^w \geq \frac{D_t^i}{v_t^i}, i \in P, \tag{4}$$

where g_p^w denotes the remaining green time for the platoon **P**, D_t^i and v_t^i indicate the remaining distance toward the intersection and the speed of the i^{th} vehicle inside the platoon **P** at the time step t. In the platoon **P**, when i^{th} the vehicle cannot pass the intersection and the vehicle directly ahead of the i^{th} vehicle, i.e., $i-1^{th}$, can pass the intersection, then the platoon **P** disbands from the $i-1^{th}$ vehicle. The vehicles after the $i-1^{th}$ vehicle in the platoon **P** would reform a platoon to decelerate together. When the platoons are approaching the intersection, the platooning system checks (4) for each vehicle in the platoons at every time step. In this manner, the platoon system can avoid situations where the platoon runs a red light because of its large platoon size.

3.3.2. Platooning Behaviors. The vehicles inside a platoon share the same speed and keep a constant close distance in between. The platoon speed is naturally determined by the leading vehicle's speed. The platoon attempts to set the following vehicles' speeds the same as that of the leading vehicle within acceleration capacity in every time step. If the speed difference between the leading vehicle and the following vehicle exceeds the acceleration/deceleration capacity, the speeds of the following vehicles will execute the boundary speeds to match the leading vehicle's speed as close as possible, as shown in the following:

$$v_t^{\text{following}} = \begin{cases} \max\left(v_{t-1}^{\text{following}} - a_L, v_{t-1}^{\text{leading}}\right), \text{if } v_{\text{leading}} \leq v_{\text{following}}, \\ \min\left(v_{t-1}^{\text{following}} + a_U, v_{t-1}^{\text{leading}}\right), \text{if } v_{\text{leading}} > v_{\text{following}}, \end{cases} \tag{5}$$

where $v_t^{\text{following}}$ and $v_{t-1}^{\text{following}}$ indicate the speed of the following vehicle at the time step t and time step $t-1$, respectively, and v_{t-1}^{leading} denotes the speed of the leading vehicle at the time step $t-1$.

Indeed, in this system, the distance that guarantees a platoon formulation may have an important influence on the performance of the platooning system. Hence, this research also conducts a sensitivity analysis of this parameter. The selection of distance boundaries ranges from 5m to 31m with an increment of 4m. Each distance boundary has 5 simulation runs and each simulation lasts for 900s (15 minutes). This research obtains the traffic delay and fuel consumption to determine the optimal searching distance. Figure 2 provides the average traffic delay and fuel consumption results for each distance boundary. According to Figure 2, it can be observed that both traffic delay and fuel consumption reach relatively low values at a distance of $21m$, and thus this research selects $21m$ as the distance boundary for further analysis.

3.4. Trajectory Planning

3.4.1. Optimal Trajectory Based on Accumulated Absolute Acceleration Rates. CAVs can plan their trajectories based on the signal information obtained to achieve a certain objective, such as minimizing fuel consumption or traffic delay. The popular approach is to formulate trajectory planning as an optimal control problem whose objective can be a certain traffic performance measure. When the goal is to minimize fuel consumption or emissions, the objective function often takes a nonlinear form and requires nonlinear programming to obtain an optimal solution. Significant computational resources may be required in the real world. A substitute approach to achieving the optimal fuel consumption or emissions benefit is to minimize accumulated absolute acceleration rates along the trajectories, according to [31]. First, a generalized trajectory planning problem of CAVs can be formulated with the objective of minimizing cost **C**.

$$\begin{aligned} &\min \mathbf{C}(a, v), \\ &\begin{cases} \dot{x}(t) = v(t) \\ \dot{v}(t) = a(t) \end{cases}, \\ &\begin{cases} x(t_0) = 0 \\ v(t_0) = v_0 \end{cases}, \\ &\begin{cases} x(t_f) = D \\ v(t_f) = v_f \end{cases}, \\ &-a_L \leq a(t) \leq a_{U,}, \\ &0 < v(t) < v_{\max}, \\ &t_0 \leq t \leq t_f, \\ &t_f \text{ fixed}. \end{aligned} \tag{6}$$

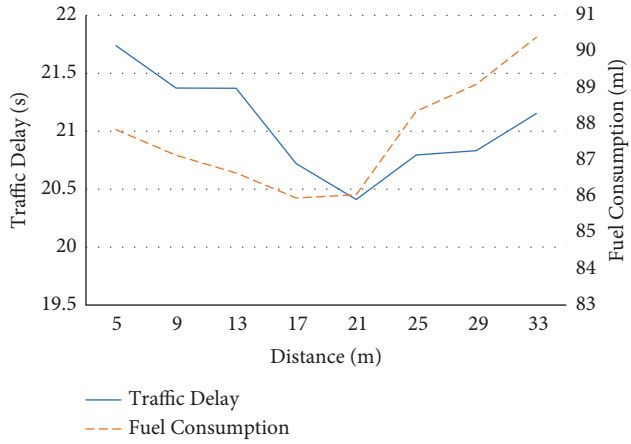


FIGURE 2: Performances of platooning with different values of distance boundaries.

where $\mathbf{C}(a, v)$ represents the cost function, $x(t)$ and $v(t)$ are control variables that indicate the traveled distance and instant speed value at the time step t respectively. $a(t)$ is the control variable that represents the acceleration rate at time step t. t_0 and t_f are the time steps when the CAVs start and finish the trajectory, respectively. a_L and a_U represent the absolute values for the maximum deceleration rate and acceleration rates. $v_{\max}$ indicates the maximum speed (speed limits) while v_f denotes the final speed when the vehicle arrives at the intersection. v_0 represents the initial speed at which the vehicle enters the communication range between CAVs and signalized intersections. D is the target distance that the subject vehicle needs to travel, which often is the distance between the vehicle and the intersection. The fuel consumption or emission is known to be significantly related to the acceleration rates. [31] developed a trajectory planning strategy to minimize fuel consumption based on Pontryagin's Minimum Principle (PMP). Through analysis of PMP, a generalized solution can be achieved with the objective of minimizing the accumulated absolute acceleration rates along the trajectory, which is as follows:

$$\min \mathbf{C} = \int_{t_0}^{t_f} |a(t)| dt. \tag{7}$$

The solution to the optimal trajectory generally results in a three-segment trajectory, in which vehicles remain at a constant speed at the second segment. The first and the third segment have a constant either maximum acceleration or deceleration rate according to the relationships between the [31]initial speed and final speed, as Figure 3(a) shows. Figure 3(b) provides an example comparison of when CAVs are enabled with and without such trajectory planning feature.

The transition time steps t_1 and t_2 can be determined given the following equations in the deceleration case $(v_0 > v_f)$ and acceleration case $(v_0 < v_f)$, respectively, where v_c indicates the constant speed in the second segment and the other variables are defined earlier.

$$\frac{v_0 + v_c}{2} * t_1 + v_c * (t_2 - t_1) + \frac{v_f + v_c}{2} * (t_f - t_2) = D,$$

$$v_c = \begin{cases} v_0 - a_L * t_1 = v_f + a_L(t_f - t_2), v_0 > v_f, \\ v_0 + a_U * t_1 = v_f - a_U(t_f - t_2), v_0 < v_f. \end{cases} \tag{8}$$

Additionally, one can obtain the lower and upper travel time boundary to guarantee that a feasible solution will be obtained as shown below:

$$v_0 > v_f \begin{cases} t_L = \dfrac{D}{v_0} + \dfrac{(v_0 - v_f)^2}{2 * v_0 * a_L}, \\ t_U = \dfrac{D}{v_f} - \dfrac{(v_0 - v_f)^2}{2 * v_f * a_L}, \end{cases} \tag{9}$$

$$v_0 < v_f \begin{cases} t_L = \dfrac{D}{v_f} + \dfrac{(v_0 - v_f)^2}{2 * v_f * a_U}, \\ t_U = \dfrac{D}{v_0} - \dfrac{(v_0 - v_f)^2}{2 * v_0 * a_U}. \end{cases} \tag{10}$$

Notably, a feasible three-segment trajectory solution only exists when the vehicle arrival time t_f is strictly within the boundary of t_L and t_U, i.e.,

$$t_L < t_f < t_U. \tag{11}$$

When $t_f = t_L$ or $t_f = t_U$, the three-segment trajectory solution collapses into the two-segment trajectory. The lower/upper-time boundaries indicate two-segment trajectories in acceleration and deceleration, respectively, as shown in Figure 4.

In a deceleration scenario, the lower boundary indicates that the vehicle keeps its current speed in the first segment and then decelerates to its final speed in the second segment. The upper boundary indicates that the vehicle first decelerates to the target final speed, then keeps the target final speed until it arrives at the intersection. On the other hand, in an acceleration scenario, the lower boundary indicates that the vehicle first accelerates the final speed v_f and then cruises at the target speed until arriving at the intersection. When the final speed v_f is equal to the maximum speed $v_{\max}$ such trajectory type can yield the minimum travel time t_{minimum} and thus is referred to as the minimum travel time trajectory. The upper boundary in an acceleration scenario means that the vehicle first keeps its initial speed and then accelerates to its target speed. Intuitively, when the travel time is strictly within the lower- and upper-time boundaries, then an optimal three-segment trajectory exists. When the travel time is equal to one of the two boundary values, then a two-segment trajectory introduced above can be applied. Nevertheless, when travel time exceeds the boundary, then no feasible solution exists with the given distance,

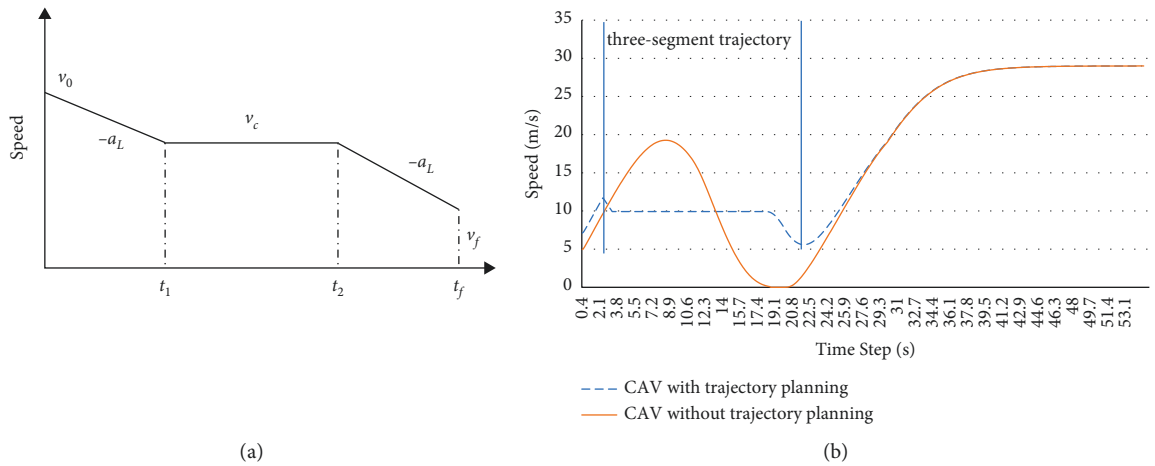


FIGURE 3: A general optimal trajectory for the deceleration scenario. (a) Theoretical three-segment trajectory in deceleration case (revised from (b) Three-segment trajectory in simulation.

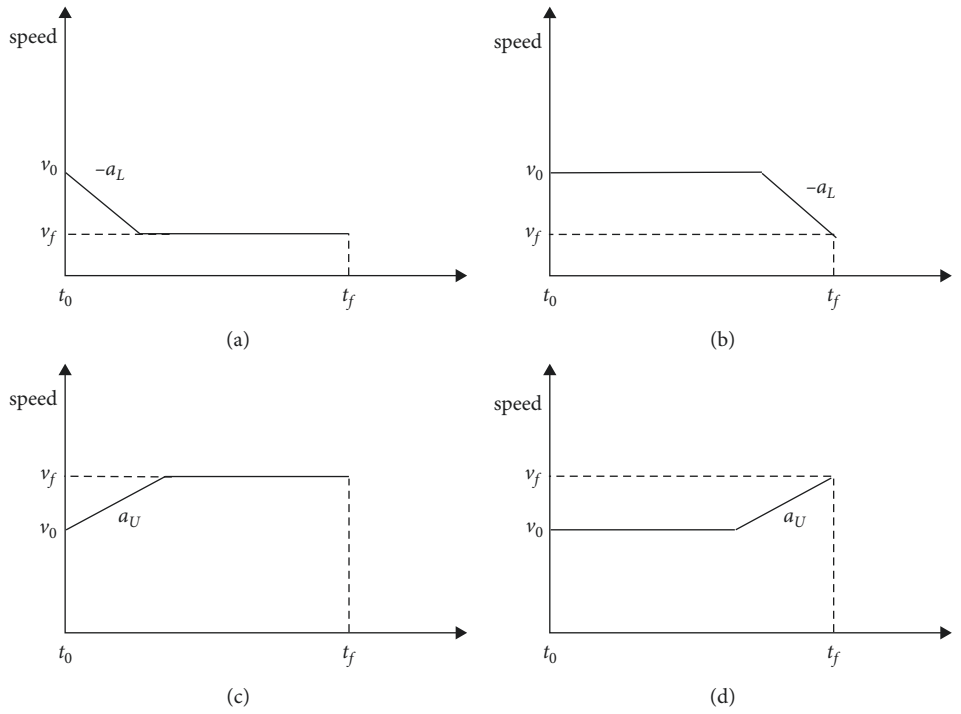


FIGURE 4: The two-segment trajectory when the travel time is equal to the boundary values. (a) $t_f = t_U$ and $v_0 > v_f$. (b) $t_f = t_L$ and $v_0 > v_f$. (c) $t_f = t_L$ and $v_0 < v_f$. (d) $t_f = t_U$ and $v_0 < v_f$.

acceleration rate, and initial speeds. This reflects real-world scenarios. For example, a vehicle may not be able to decelerate to a speed of zero if the remaining distance to the intersection is too short or the initial speed is too high.

Demonstrated that this trajectory planning strategy could successfully [31] reduce traffic delay and fuel consumption in a standard isolated conventional intersection with a joint adaptive signal optimization algorithm. With the adaptive signal control, Equation (11) can stand for most cases and the vehicle can avoid stops under certain traffic conditions. Nevertheless, this strategy cannot be directly transferred to a fixed signal-controlled intersection. In a fixed signal-controlled intersection, the final travel time t_f is largely dependent on the initiation time or remaining time of the target green phase in a fixed signal timing plan, where vehicles cannot avoid stopping entirely. To apply this trajectory planning scheme in a fixed signal-controlled intersection, this research also considers a constant deceleration trajectory when (11) cannot be sufficed. For a constant deceleration trajectory, the vehicle will keep a constant deceleration rate until it arrives at the intersection with a speed of 0, as shown in Figure 5. The deceleration rate a_{dec}

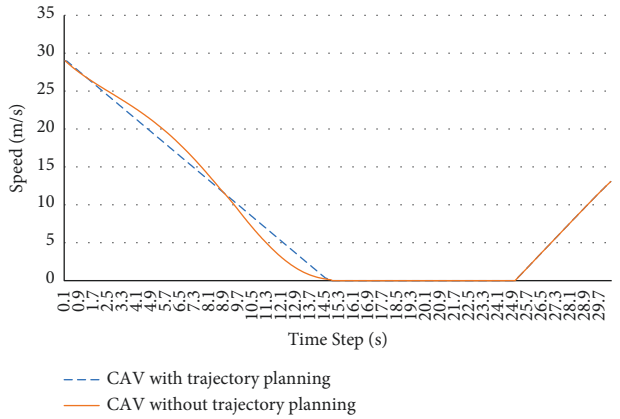


Figure 5: Constant deceleration trajectory.

can be easily obtained through the basic kinetic law, which is described by (11).

$$a_{dec} = \frac{v_0}{2 * D/v_0}. \tag{12}$$

Based on the signal status and the next signal switch time t_{switch}, the vehicle can choose different speed trajectories as introduced above.

(1) Red Signal. When the upcoming signal status for the subject vehicle is red, the signal switch time t_{switch} indicates the initiation of green time. The lower time boundary obtained through (10) is equal to the minimum travel time t_{minimum} when the given final speed $v_f = v_{\max}$. If the switch time t_{switch} is less than or equal to the t_{minimum}, then the vehicle can meet a green signal with a two-segment trajectory as shown in Figure 6 to achieve minimal traffic delay.

If the switch time is greater than the minimum travel time, i.e., $t_{\text{switch}} \geq t_{\text{minimum}}$, then the vehicle with a minimum travel time trajectory will meet a red signal. In this situation, it is assumed that $t_f = t_{\text{switch}}$. From (9) and (10), one may obtain t_L and t_U given a final speed v_f. Hence, researchers may simply enumerate all possible final speeds $[0, v_{\max})$ to obtain a feasible speed candidate list $\mathbf{V_f}$ so that (12) stands.

$$t_L < t_{\text{switch}} < t_U. \tag{13}$$

This research selects the $\max(\mathbf{V_f})$ so that the subject vehicle can travel through the intersection with maximum final speed to minimize the traffic delay, where the max function returns the maximum value among the feasible final speed list $\mathbf{V_f}$.

(2) Green Signal. If the ahead signal status is green, then $t_{\text{signalswitch}}$ indicates the remaining green time for the subject vehicle. This research mainly considers two cases based on the relationship between signal switch time $t_{\text{signalswitch}}$ and minimum travel time t_{minimum} of the subject vehicle.

Case 1. When the subject vehicle can traverse through the intersection with minimum travel time t_{minimum} (i.e., $t_{\text{signalswitch}} \geq t_{\text{minimum}}$), then the vehicle may accelerate its maximum speed to pass the intersection to achieve the minimal traffic delay. However, this strategy may potentially increase the average fuel consumption as the fuel consumption is closely related to the acceleration rate. In some circumstances, if the $t_{\text{signalswitch}} \geq t_{\text{currentspeed}}$, where $t_{\text{currentspeed}}$ is the travel time to the intersection when the vehicle keeps its current speed, then the decision-makers who assign a higher priority to fuel consumption may let the vehicle keep its current speed to avoid increasing fuel consumption with acceptable compromise on the traffic delay.

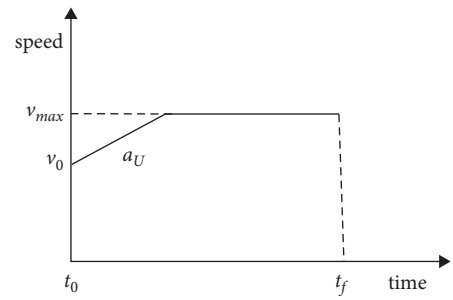


Figure 6: Speed trajectory with minimum travel time.

Case 2. $t_{\text{signalswitch}} \leq t_{\text{minimum}}$ means the subject vehicle cannot arrive at the intersection within the given remaining green time even if the vehicle accelerates to maximum speed. In such a situation, a constant deceleration trajectory introduced above may be executed. The subject vehicle may need to check whether the vehicle can meet the second green with a given final speed within $[0\, v_{\max})$ when the distance D is large.

3.4.2. Encountering Preceding Vehicles during Trajectory Planning. In the real world, the vehicles may have close preceding vehicles on the road, and following the predetermined trajectories may lead to collisions with the preceding vehicles. Therefore, to avoid these collisions in this research, when a vehicle has preceding vehicles that are within a 3*s* headway, the vehicle will stop executing the planned trajectory and switch to the predefined car-following model, which is the IDM model in this research. Note that the system constantly checks each vehicle's distance to the preceding vehicles at each time step. When the distance to the preceding vehicle is greater than 3*s* and there is an upcoming signalized intersection, then the system will plan the vehicle trajectory again for the subject vehicle to follow. With this function, the vehicles following the planned trajectory can successfully avoid collision with not only the close preceding vehicles but also the queueing vehicles in front of the intersection because of the red signal.

3.5. Simulation Scenarios and Relevant Settings. An equivalent conventional intersection with the same road segment length, lane configuration, and maximum speed is designed in the simulation platform. The cycle length is also set the same as the superstreet in the real world, i.e., 120s, to make a fair comparison. The green splits for each approach are determined

by their volume ratios. To account for different traffic conditions, this research tests four different traffic scales including 25%, 50%, 75%, and 100% of peak hour traffic volumes from Table 2. Furthermore, a market penetration analysis is conducted on the 100% peak hour traffic volumes. 25%, 50%, and 75% of CAV market penetration rates are considered in the simulation. Every scenario is run five times with different random seeds to account for the randomness. To make the system more robust and increase calculation accuracy, the simulation resolution is set as 10HZ, which means the simulation runs 10 time steps every second. Once the vehicle enters the roadway network, the vehicle is assumed to enter the Vehicle-to-Infrastructure (V2I) communication range, which is reasonable since the selected superstreet has a rather short road segment length in all approaches before the traffic signals (less than 300 m). Average traffic delay (delay per vehicle) and fuel consumption (fuel consumption per vehicle) are the performance indicators that are used for this research. Traffic delay is measured by the ideal travel time (free-flow speed without any stop) minus actual travel time. Fuel consumption is measured by the default emission model from SUMO, i.e., HBEFT.3 [45]. The maximum acceleration rates and deceleration rates for IDM are set as 2.5 m/s^2. Considering drivers' comfort, the maximum acceleration rate and deceleration rate in CAV trajectory planning are 2 m/s^2.

4. Results and Discussion

4.1. The Performance of CAVs in Conventional Intersections

4.1.1. Traffic Delay. To provide an initial understanding of the performance of CAVs, this research first obtains the simulation results of CAVs from the equivalent conventional intersection. The traffic delay results are presented in Figure 7. From Figure 7, it can be observed that the developed platooning, trajectory planning, and platooning-based trajectory planning can reduce the traffic delay in most scenarios. The exception is CAVs with platooning at 25%. When CAVs are enabled with platooning, the speed of the following vehicles is influenced by the leading vehicle in the same platoon and may not be able to achieve their maximum speeds even in light traffic volume. This may potentially explain that no benefit is gained for platooning in the traffic demand of 25% and 50% peak hour traffic volume scenarios. The traffic delay improvements for CAV with platooning increase as the traffic demand increases.

Trajectory planning can reduce traffic delay to a larger extent in light traffic volume scenarios, and the improvement magnitudes shrink as the traffic volumes increases. These results can be explained by the trajectory planning modeling framework. As mentioned in the methodology section, to avoid collisions with preceding vehicles and queueing vehicles in front of the intersection, CAVs with trajectory planning may switch to the default car-following model frequently in high traffic demand scenarios. For CAV with platooning-based trajectory planning, the traffic delays share a similar trend as the ones from CAV with platooning. Notably, platooning-based trajectory planning also successfully reduces the traffic delay in low traffic demand scenarios.

4.1.2. Fuel Consumption. From Figure 8, it can be observed that platooning could provide larger benefits in terms of fuel consumption in high traffic volume scenarios. The improvement magnitudes are also consistent with existing studies on platooning [46]. The proposed trajectory planning framework reduces the average fuel consumption to a certain extent in low traffic volume scenarios. However, the fuel consumption benefits from trajectory planning are less significant compared to platooning. In addition, the trajectory planning framework may produce adverse effects on fuel consumption in high traffic volume scenarios, as observed in 100% peak hour traffic volume scenarios. In high traffic volume scenarios, CAVs with trajectory planning capability change to the car-following model frequently because of the presence of preceding vehicles, which may produce speed fluctuations and higher fuel consumption. CAV with platooning-based trajectory planning produces the optimal fuel consumption results in most traffic demand levels.

4.2. Comparison between CAVs and HDVs with Calibrated W99.

This research first examines the performance of the calibrated W99 model, IDM model, IDM with platooning, IDM with trajectory planning, and IDM with platooning-based trajectory under 100% peak hour traffic volume, respectively.

Although it is expected that CAVs outperform HDVs, it may not be necessarily always true in the real world. For instance, when the vehicle travels through a congested intersection, HDVs are likely to have shorter headways and practice emergency deceleration or acceleration to achieve minimal travel time or avoid collisions, while CAVs cannot exceed the predetermined boundary of safe headway and acceleration rates. According to Figure 9, the results from calibrated W99 and IDM prove this assumption since they have similar average delays and fuel consumption.

However, when CAVs are enabled with platooning and trajectory planning, the CAVs may be superior to HDVs. For the proposed platooning model, compared to the IDM model, the traffic delay decreases from 23.42 to 20.49 (around 13%), while the fuel consumption decreases from 95.79 to 85.87 (around 10% reduction). Since HDVs with calibrated W99 have similar traffic delay and fuel consumption, similar improvements can be found when comparing CAV with platooning against HDVs with calibrated W99. Table 4 presents the comparison results between CAVs with different features and HDVs with calibrated W99.

The outstanding performance of platooning performances may be related to the large traffic volume in this scenario. On the other hand, IDM with trajectory planning has few benefits in terms of both traffic delay and fuel consumption compared to IDM only. As described in the previous section, CAVs will change into the car-following model when they detect vehicles that are within a 3s headway. In a congested traffic condition such as 100% peak hour traffic volume, the advantages of trajectory planning are significantly compromised. As for CAVs with platooning-based trajectory planning, the traffic delay decreases and reaches the lowest traffic delay (19.80s) among all scenarios,

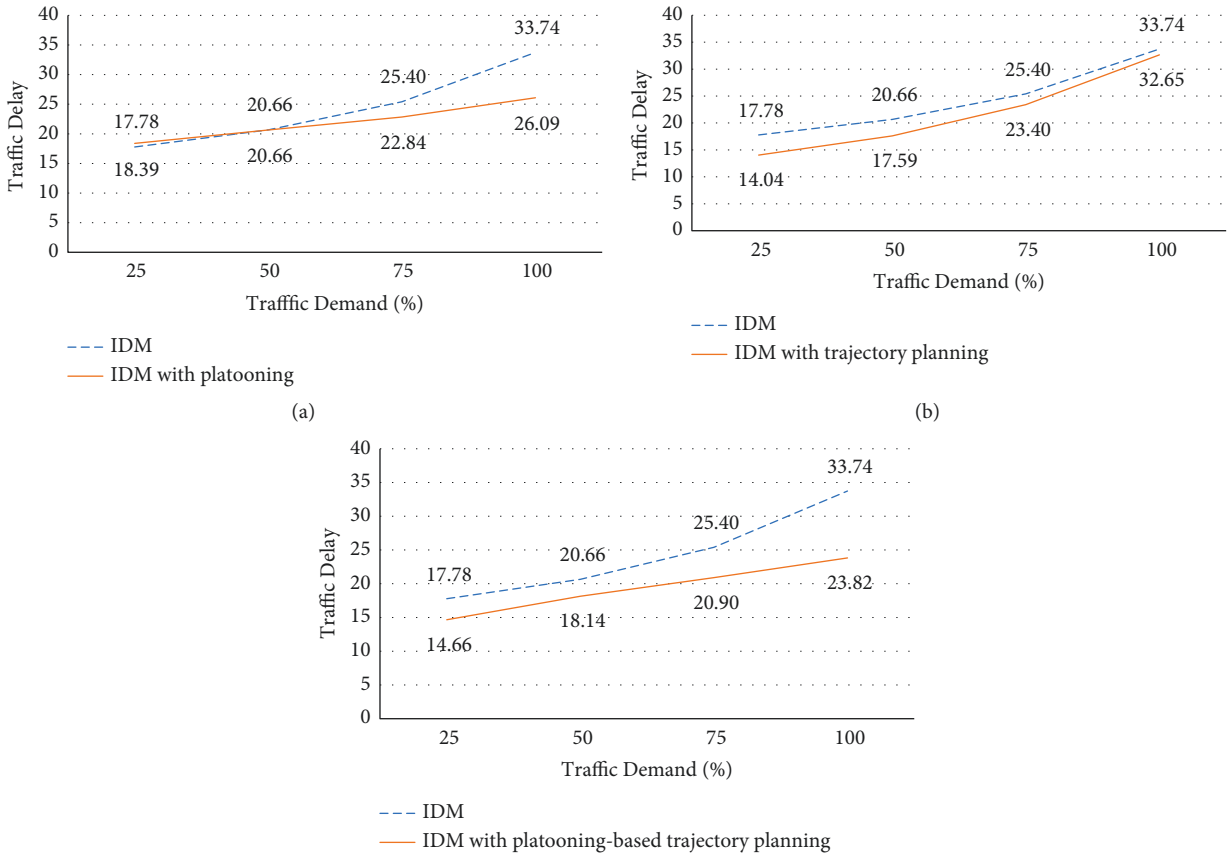


Figure 7: Average traffic delay(s) of CAVs in the equivalent conventional intersection. (a) Average traffic delay for IDM and IDM with platooning (b) Average traffic delay for IDM and IDM with trajectory planning. (c) Average traffic delay for IDM and IDM with platooning-based trajectory planning.

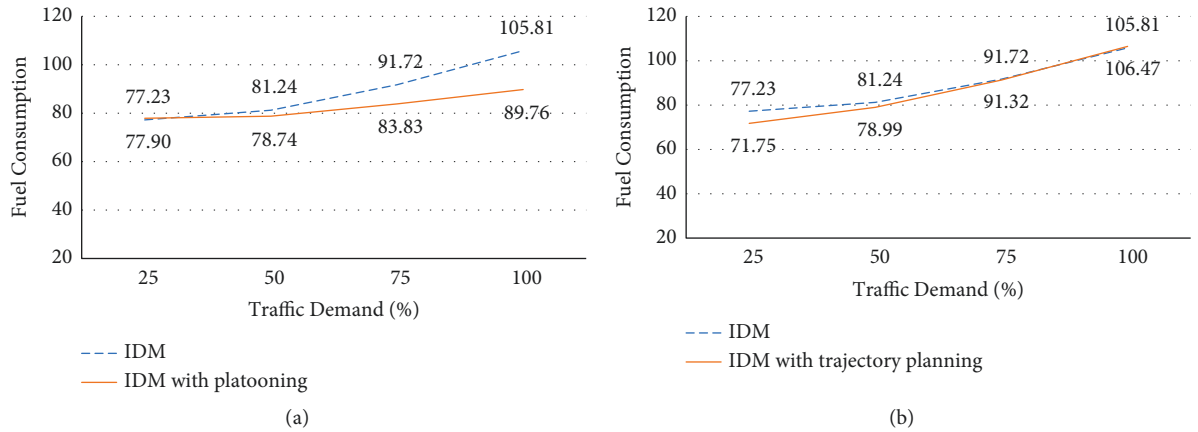


Figure 8: Continued.

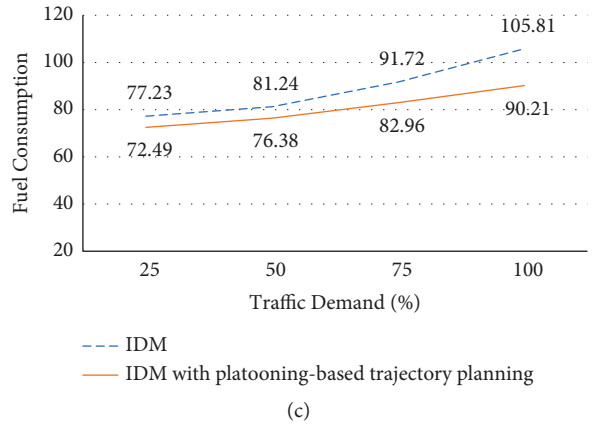


Figure 8: Average fuel consumption(ml) of CAVs in the equivalent conventional intersection. (a) Average fuel consumption for IDM and IDM with platooning. (b) Average fuel consumption for IDM and IDM with trajectory planning. (c) Average fuel consumption for IDM and IDM with platooning-based trajectory planning.

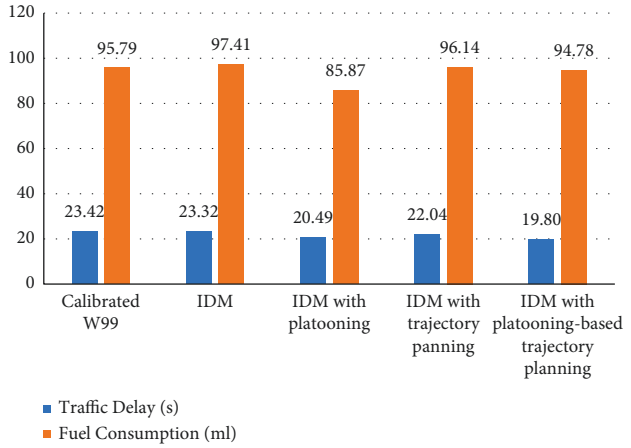


Figure 9: Traffic performances with different scenarios.

Table 4: CAV performances compared to HDVs with calibrated W99.

	TD[1]	Improvement[2] (%)	FC[3]	Improvement[4]
IDM	23.32	0	97.41	-2%
IDM with platooning	20.49	13	85.87	10%
IDM with trajectory panning	22.04	6	96.14	0%
IDM with platooning-based trajectory planning	19.80	15	94.78	1%

[1] average traffic delay in seconds;[2] benchmark is 23.42;[3] average fuel consumption in milliliter;[4] benchmark is 95.79.

while the fuel consumption is lower compared to CAVs with trajectory planning but higher compared to CAVs with platooning. CAVs with platooning and trajectory planning, when vehicles are close to each other, form a platoon so that trajectory planning can be executed, which explains the greater traffic delay reduction in CAVs with platooning-based trajectory planning. The fuel consumption of platooning-based trajectory planning is higher than the ones of platooning but lower than the ones of trajectory planning.

4.3. The Performances of CAVs in Superstreets

4.3.1. Traffic Delay. Figure 10 presents the average traffic delay when CAVs are enabled with platooning, trajectory planning, and platooning-based trajectory planning. CAVs with platooning have similar performances as they did in the equivalent conventional intersection. When the traffic scale is at 25% peak hour traffic volume, the CAVs with platooning fails to reduce the average traffic delay. Nevertheless,

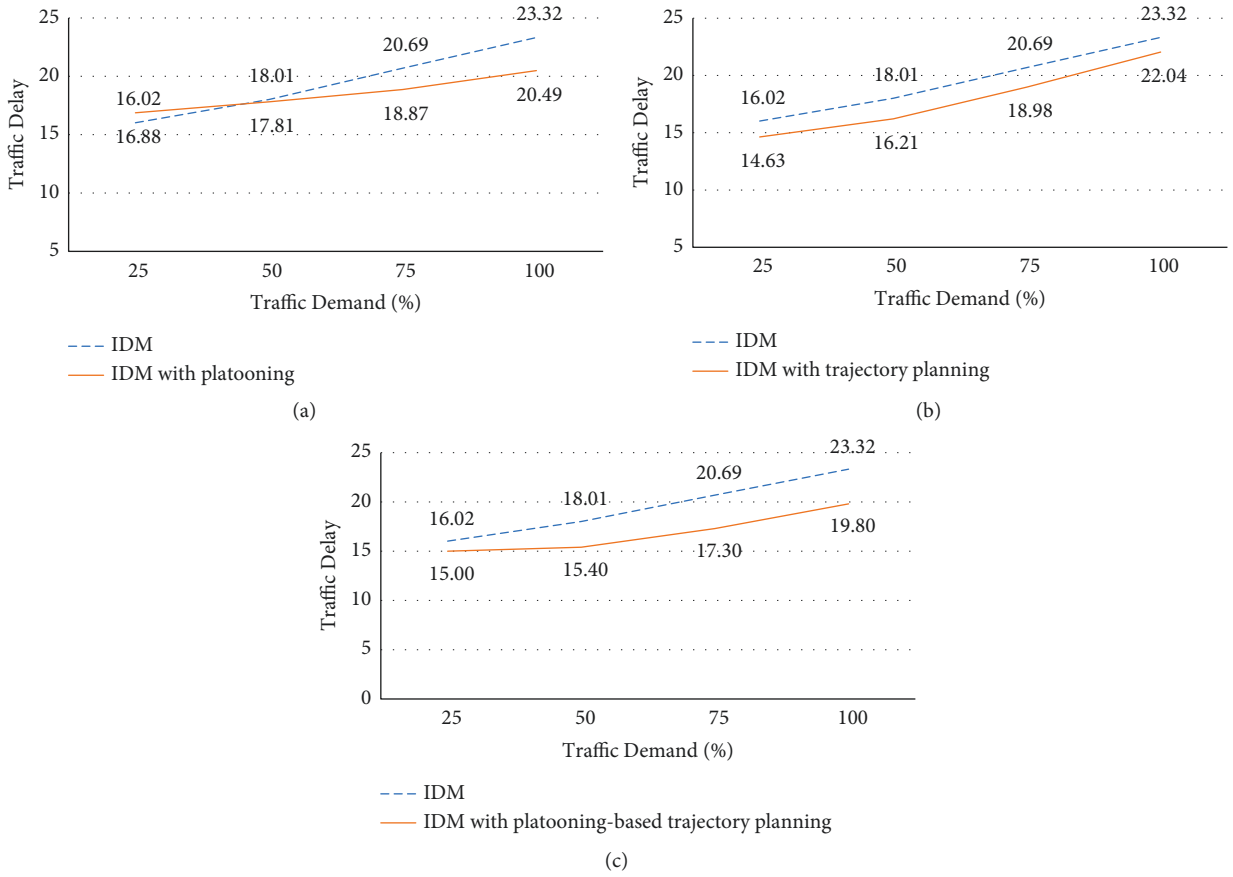


Figure 10: Average traffic delay(s) of CAVs in the superstreet. (a) Average traffic delay for IDM and IDM with platooning. (b) Average traffic delay for IDM and IDM with trajectory planning. (c) Average traffic delay for IDM and IDM with platooning-based trajectory planning.

when the traffic demand is greater or equal to 50% of peak hour traffic volume, the CAVs start to reduce the traffic delay in the superstreet.

As for trajectory planning, the reductions of traffic delay in different demands are relatively constant compared to the ones in the conventional intersection. In superstreet, the road capacity often is larger than the equivalent conventional intersection. Therefore, CAVs might not have to switch to the car-following model frequently as they did in the equivalent conventional intersection in 100% peak hour traffic volume demand, which explains the relevant constant traffic delay reduction.

CAVs with platooning-based trajectory planning still produce minimal traffic delays in nearly all demand levels (except for 25% peak hour traffic demand). The general trend of traffic delays is similar to that in platooning scenarios as in the equivalent conventional intersection.

4.3.2. Fuel Consumption. Figure 11 presents the fuel consumption of CAVs in the superstreet. Platooning yields similar fuel consumption trends as it did in the traffic delay results. Nevertheless, CAVs with trajectory planning produce higher average fuel consumption, especially in the lower traffic demand scenarios. The increased average fuel consumption is potentially attributed to two reasons: (1) the acceleration behavior of CAVs with trajectory planning in order to catch the remaining green or initiation green time; (2) CAVs with trajectory planning may stop at the second consecutive intersection after passing the first intersection with acceleration in the superstreet system. In high traffic volume scenario, the adverse effects of fuel consumption are alleviated since CAVs with trajectory planning do not have much freedom of accelerating before the intersection. This result demonstrates the necessity of incorporating two consecutive signal information in designing a trajectory planning framework when two signals are closely spaced. The adverse effects on fuel consumption are alleviated when CAVs are enabled with platooning-based trajectory planning.

4.3.3. CAVs with Different Market Penetration Rates. The dominance of CAVs on the road is a gradual process in which technology, political and legal challenges continuously remain. The policymakers may be interested in the performances of CAVs with different levels of market penetration rates. Therefore, this research also conducts a market penetration analysis where HDVs controlled by calibrated W99 and CAVs controlled by IDM with

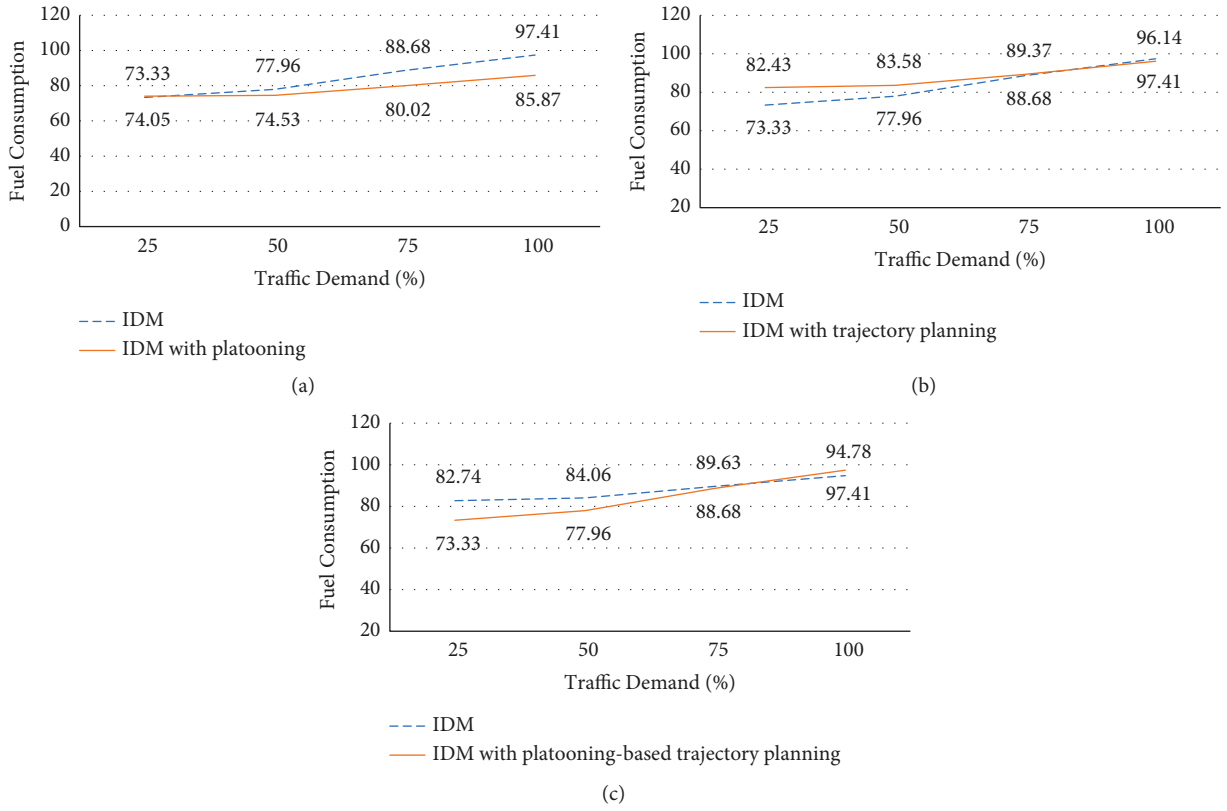


Figure 11: Average fuel consumption (ml) of CAVs in the superstreet. (a) Average fuel consumption for IDM and IDM with platooning. (b) Average fuel consumption for IDM and IDM with trajectory planning. (c) Average fuel consumption for IDM and IDM with platooning-based trajectory planning.

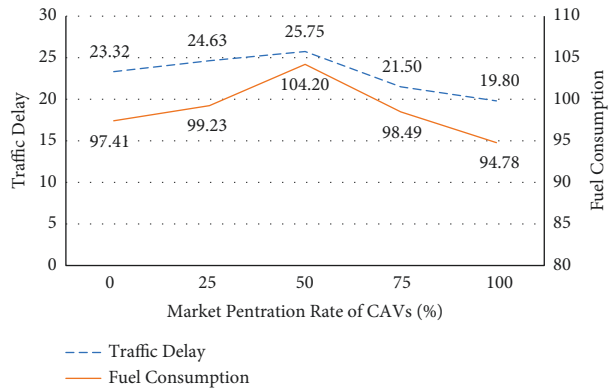


Figure 12: Analysis for different CAV market penetration rates.

platooning-based trajectory planning coexist. 25%, 50%, and 75% CAV market penetration rates are tested under 100% peak hour traffic volume. When CAVs follow HDVs, CAVs are often assumed to have larger headways [47]. Therefore, when CAVs follow HDVs, the CAV headway is set the same as HDVs, i.e., 1.6s. Figure 12 provides the results of the market penetration analysis. Based on Figure 12, it can be observed that traffic delay starts to fall at the market penetration of 75% CAVs, where the fuel consumption is similar to that of 0% CAV. The fuel consumption and traffic delay are highest when the market penetration rate of CAVs is at the 50% level. Overall, the more mixed the vehicle types are (i.e., equal market penetration rate of CAVs and HDVs), the worse the traffic performance is.

4.4. A Comparison between Conventional Intersection and Superstreet. Figures 13 and 14 compare the average traffic delay and fuel consumption of CAVs in the equivalent conventional intersection and superstreet, respectively. Based on Figure 13, with IDM vehicles, the superstreet can consistently outperform the equivalent conventional

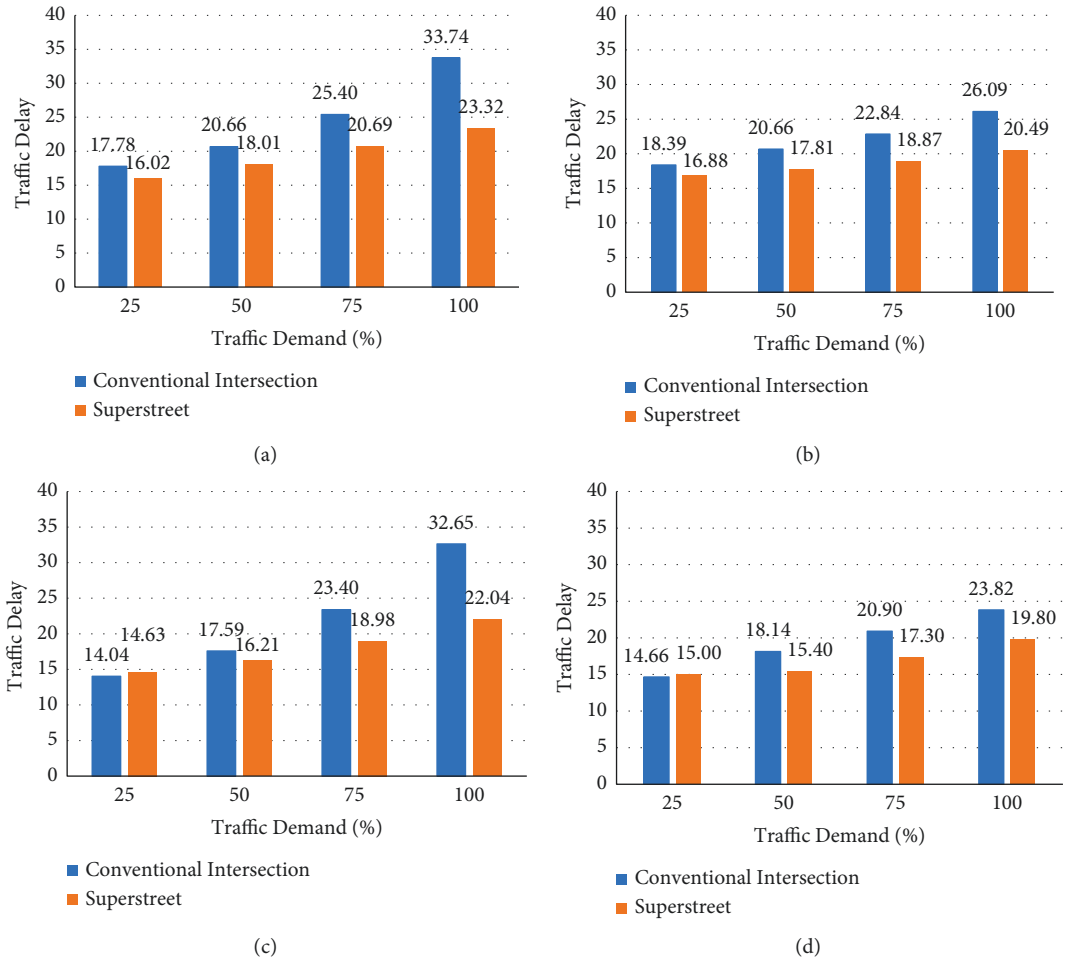


Figure 13: Average traffic delay(s) comparison of CAVs between the conventional intersection and superstreet. (a) IDM. (b) IDM with platooning. (c) IDM with trajectory planning. (d) IDM with platooning-based trajectory planning.

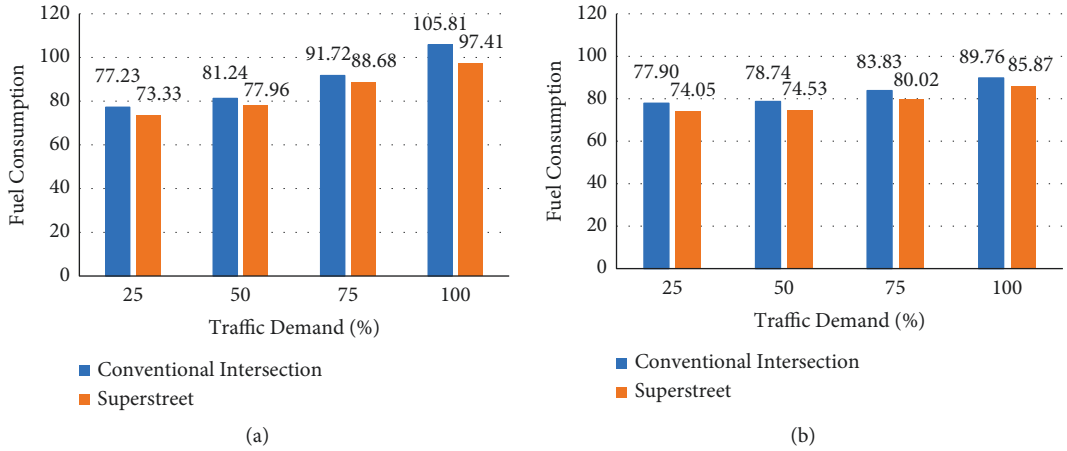


Figure 14: Continued.

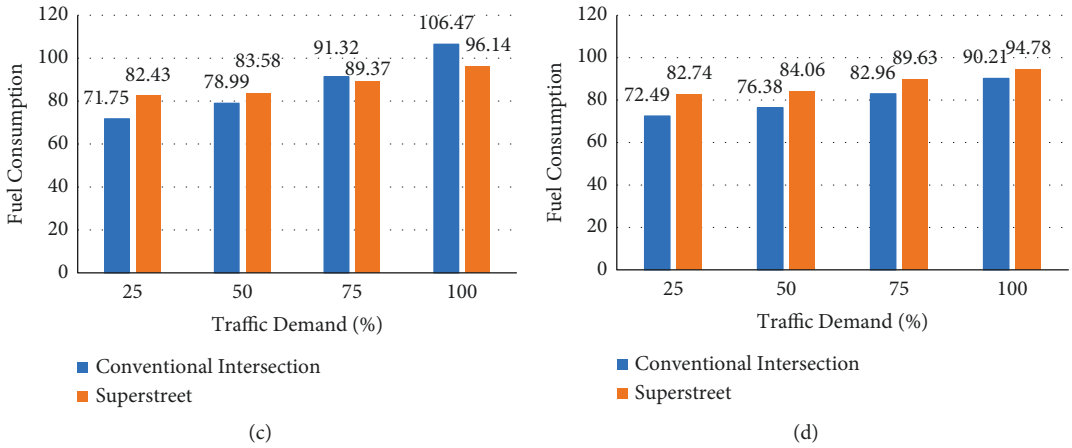


Figure 14: Average fuel consumption(ml) comparison of CAVs between the conventional intersection and superstreet. (a) IDM. (b) IDM with platooning. (c) IDM with trajectory planning. (d) IDM with platooning-based trajectory planning.

Table 5: Average traffic delay improvement magnitudes of superstreet compared to the equivalent conventional intersection with different CAV Features.

CAV features	Peak demand scale			
	25%	50%	75%	100%
IDM	10%	13%	19%	31%
IDM with platooning	8%	14%	17%	21%
IDM with trajectory planning	−4%	8%	19%	32%
IDM with platooning-based trajectory planning	−2%	15%	17%	17%

Table 6: Average fuel consumption improvement magnitudes of superstreet compared to the equivalent conventional intersection with different CAV Features.

CAV features	Peak demand scale			
	25%	50%	75%	100%
IDM	5%	4%	3%	8%
IDM with platooning	5%	5%	5%	4%
IDM with trajectory planning	−15%	−6%	2%	10%
IDM with platooning-based trajectory planning	−14%	−10%	−8%	−5%

intersection regarding average traffic delay. However, it could also be observed that the average traffic delay differences between the conventional intersection and superstreet are reduced in platooning and platooning-based trajectory planning scenarios compared to CAVs controlled by IDM only.

As for fuel consumption, Figure 14 shows that the average fuel consumptions of CAVs with trajectory planning are higher when they are in the superstreet under 25% and 50% peak hour traffic volume. When CAVs are enabled with platooning-based trajectory planning, they have higher average fuel consumption on all demand levels in the superstreet. As explained in the previous section, this may potentially result from the lack of consideration of two closely spaced signalized intersections when developing the trajectory planning control framework. Tables 5 and 6 provide improvement magnitudes of superstreet compared to the conventional intersection in terms of the average traffic delay and fuel consumption, respectively.

5. Conclusions

This research investigated the performances of CAVs and HDVs in the environments of the superstreet and conventional intersection. CAVs were modeled with the IDM car-following model, while HDVs were modeled with the W99 car-following model. A real-world superstreet situated in Leeland, NC, was replicated in the simulation platform to test the performances of CAVs and HDVs under different traffic conditions. In addition, to fully examine the potentiality of CAVs, a simple platooning scheme and trajectory planning strategy were developed for CAVs, respectively. In this research, the W99 model was calibrated with GA so that the W99 model can better represent the local drivers' behaviors. Different traffic demands and market penetration rates were taken into consideration in the designed scenarios.

The simulation results indicated that without platooning and trajectory planning, CAV modeled by IDM did not have significant improvement compared to HDVs modeled by

W99. The developed platooning strategy can successfully reduce the traffic delay and fuel consumption at relatively high traffic demand scenarios (50%, 75%, and 100% peak hour volume) in both the superstreet and the conventional intersection. Trajectory planning could reduce the traffic delay in both superstreet and conventional intersection environments but with different impacts on fuel consumption. CAVs with trajectory planning produced higher fuel consumption in the superstreet in the lower traffic demand scenarios, especially in traffic demand 25% and 50% of peak hour traffic volume. A potential reason is that CAVs which accelerate to pass the first intersection may fail to pass the consecutive second intersection in the environment of superstreet. In the market penetration rate analysis of CAVs, it was found that the mixed traffic environment can compromise the benefit when the CAVs market penetration rates were at 25% and 50% peak hour traffic volume. CAVs have better performances when the market penetration rate is about 75% and above.

This research also compared the traffic performances of CAVs in the conventional intersection and superstreet. A notable finding was that the proposed trajectory planning control strategy can successfully reduce the average traffic delay without increasing the average fuel consumption in the conventional intersection. This was different from superstreet where CAVs enabled with trajectory planning increase the fuel consumption. This demonstrated the efficiency of the proposed trajectory planning strategy in an isolated intersection. However, this result also indicated that the trajectory planning without considering special features of two closely spaced signalized intersections may suffer adverse effects on fuel consumption. Overall, the improvement magnitude of platooning and trajectory planning was larger in the conventional intersection.

Based on these research findings, the future research directions could be the adaptive signal control strategy that takes arrival information on CAVs into consideration, which may reduce the adverse effects of trajectory planning on fuel consumption identified in this research. Also, a more sophisticated trajectory planning algorithm that takes into account two consecutive signalized intersections can be developed.

Data Availability

All data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors want to express their deepest gratitude for the financial support by the United States Department of Transportation, University Transportation Center through the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE) at the University of North Carolina at Charlotte (Grant no. 69A3551747133).

References

[1] P. Liu and W. D. Fan, "Exploring the impact of connected and autonomous vehicles on freeway capacity using a revised Intelligent Driver Model," *Transportation Planning and Technology*, vol. 43, no. 3, pp. 279–292, 2020.

[2] X. Guo, Y. Peng, S. Ashraf, and M. W. Burris, "Performance analyses of information-based managed lane choice decisions in a connected vehicle environment," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2674, no. 11, pp. 120–133, 2020.

[3] J. Guo, S. Cheng, and Y. Liu, "Merging and diverging impact on mixed traffic of regular and autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1639–1649, 2021.

[4] R. Mohebifard and A. Hajbabaie, "Connected automated vehicle control in single lane roundabouts," *Transportation Research Part C: Emerging Technologies*, vol. 131, Article ID 103308, 2021.

[5] R. Mohebifard and A. Hajbabaie, "Trajectory Control in Roundabouts with a Mixed Fleet of Automated and Human-driven Vehicles," *Computer?Aided Civil and Infrastructure Engineering*, 2021.

[6] H. Jiang, J. Hu, S. An, M. Wang, and B. B. Park, "Eco approaching at an isolated signalized intersection under partially connected and automated vehicles environment," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 290–307, 2017.

[7] X. Han, R. Ma, and H. M. Zhang, "Energy-aware trajectory optimization of CAV platoons through a signalized intersection," *Transportation Research Part C: Emerging Technologies*, vol. 118, Article ID 102652, 2020.

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

[9] C. Mu, L. Du, and X. Zhao, "Event triggered rolling horizon based systematical trajectory planning for merging platoons at mainline-ramp intersection," *Transportation Research Part C: Emerging Technologies*, vol. 125, Article ID 103006, 2021.

[10] J. Hummer, B. Ray, A. Daleiden, P. Jenior, and J. Knudsen, "Restricted Crossing U-Turn: Informational Guide," Report\No. FHWA-SA-14-070, Federal Highway Administration Office of Safety, Washington, DC, USA, 2014.

[11] R. L. Haley, S. E. Ott, J. E. Hummer, R. S. Foyle, C. M. Cunningham, and B. J. Schroeder, "Operational effects of signalized superstreets in North Carolina," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2223, no. 1, pp. 72–79, 2011.

[12] H. H. Naghawi, "Analyzing delay and queue length using microscopic simulation for the unconventional intersection design Superstreet," *Journal of the South African Institution of Civil Engineering= Joernaal van die Suid-Afrikaanse Instituut van Siviele Ingenieurswese*, vol. 56, no. 1, pp. 100–107, 2014.

[13] J. E. Hummer, R. L. Haley, S. E. Ott, R. S. Foyle, and C. M. Cunningham, "Superstreet Benefits and Capacities," Report\No. FHWA/NC/2009-06, Institute for Transportation Research and Education, Raleigh, NC, USA, 2010.

[14] L. Xu, X. Yang, and G.-L. Chang, "Computing the minimal U-turn offset for an unsignalized superstreet," *Transportation*

Research Record: Journal of the Transportation Research Board, vol. 2618, no. 1, pp. 48–57, 2017.

[15] L. Xu, X. Yang, and G.-L. Chang, "Two-stage model for optimizing traffic signal control plans of signalized Super-street," *Transportmetrica: Transport Science*, vol. 15, no. 2, pp. 993–1018, 2019.

[16] Y. Qin, H. Wang, and B. Ran, "Stability analysis of connected and automated vehicles to reduce fuel consumption and emissions," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 11, Article ID 04018068, 2018.

[17] J. Rios-Torres and A. A. Malikopoulos, "Impact of partial penetrations of connected and automated vehicles on fuel consumption and traffic flow," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 4, pp. 453–462, 2018.

[18] A. Adebisi, Y. Liu, B. Schroeder et al., "Developing highway capacity manual capacity adjustment factors for connected and automated traffic on freeway segments," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2674, no. 10, pp. 401–415, 2020.

[19] S. Chityala, J. O. Sobanjo, E. Erman Ozguven, T. Sando, and R. Twumasi-Boakye, "Driver behavior at a freeway merge to mixed traffic of conventional and connected autonomous vehicles," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2674, no. 11, pp. 867–874, 2020.

[20] X. Hu and J. Sun, "Trajectory optimization of connected and autonomous vehicles at a multilane freeway merging area," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 111–125, 2019.

[21] M. Pourmehrab, P. Emami, M. Martin-Gasulla, J. Wilson, L. Elefteriadou, and S. Ranka, "Signalized intersection performance with automated and conventional vehicles: a comparative study," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 9, Article ID 04020089, 2020.

[22] Y. Guo, J. Ma, C. Xiong, X. Li, F. Zhou, and W. Hao, "Joint optimization of vehicle trajectories and intersection controllers with connected automated vehicles: combined dynamic programming and shooting heuristic approach," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 54–72, 2019.

[23] P. Li and X. Zhou, "Recasting and optimizing intersection automation as a connected-and-automated-vehicle (CAV) scheduling problem: a sequential branch-and-bound search approach in phase-time-traffic hypernetwork," *Transportation Research Part B: Methodological*, vol. 105, pp. 479–506, 2017.

[24] F. Zhou, X. Li, and J. Ma, "Parsimonious shooting heuristic for trajectory design of connected automated traffic part I: Theoretical analysis with generalized time geography," *Transportation Research Part B: Methodological*, vol. 95, pp. 394–420, 2017.

[25] K. Dresner and P. Stone, "A multiagent approach to autonomous intersection management," *Journal of Artificial Intelligence Research*, vol. 31, pp. 591–656, 2008.

[26] M. Martin-Gasulla and L. Elefteriadou, "Traffic management with autonomous and connected vehicles at single-lane roundabouts," *Transportation Research Part C: Emerging Technologies*, vol. 125, Article ID 102964, 2021.

[27] B. Chalaki, L. E. Beaver, and A. A. Malikopoulos, "Experimental validation of a real-time optimal controller for coordination of cavs in a multi-lane roundabout," in *Proceedings of the 2020 IEEE Intelligent Vehicles Symposium (IV)*, pp. 775–780, IEEE, Las Vegas, NV, USA, October 2020.

[28] R. M. James, B. E. Hammit, and S. D. Boyles, "Methods to obtain representative car-following model parameters from trajectory-level data for use in microsimulation," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 7, pp. 62–73, 2019.

[29] L. Xiao, M. Wang, W. Schakel, and B. van Arem, "Unravelling effects of cooperative adaptive cruise control deactivation on traffic flow characteristics at merging bottlenecks," *Transportation Research Part C: Emerging Technologies*, vol. 96, pp. 380–397, 2018.

[30] S. Tsugawa, S. Jeschke, and S. E. Shladover, "A review of truck platooning projects for energy savings," *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 68–77, 2016.

[31] Y. Feng, C. Yu, and H. X. Liu, "Spatiotemporal intersection control in a connected and automated vehicle environment," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 364–383, 2018.

[32] C. Yu, Y. Feng, H. X. Liu, W. Ma, and X. Yang, "Integrated optimization of traffic signals and vehicle trajectories at isolated urban intersections," *Transportation Research Part B: Methodological*, vol. 112, pp. 89–112, 2018.

[33] Q. Ye, X. Chen, R. Liao, and L. Yu, "Development and evaluation of a vehicle platoon guidance strategy at signalized intersections considering fuel savings," *Transportation Research Part D: Transport and Environment*, vol. 77, pp. 120–131, 2019.

[34] S. Bang and S. Ahn, "Platooning strategy for connected and autonomous vehicles: transition from light traffic," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2623, no. 1, pp. 73–81, 2017.

[35] F. Ma, Y. Yang, J. Wang et al., "Eco-driving-based cooperative adaptive cruise control of connected vehicles platoon at signalized intersections," *Transportation Research Part D: Transport and Environment*, vol. 92, Article ID 102746, 2021.

[36] J. N. Hooker, A. B. Rose, and G. F. Roberts, "Optimal control of automobiles for fuel economy," *Transportation Science*, vol. 17, no. 2, pp. 146–167, 1983.

[37] X. He, H. X. Liu, and X. Liu, "Optimal vehicle speed trajectory on a signalized arterial with consideration of queue," *Transportation Research Part C: Emerging Technologies*, vol. 61, pp. 106–120, 2015.

[38] K. Katsaros, R. Kernchen, M. Dianati, and D. Rieck, "July). Performance study of a Green Light Optimized Speed Advisory (GLOSA) application using an integrated cooperative ITS simulation platform," in *Proceedings of the 2011 7th International Wireless Communications and Mobile Computing Conference*, pp. 918–923, IEEE, Istanbul, Turkey, July 2011.

[39] R. Stahlmann, M. Möller, A. Brauer, R. German, and D. Eckhoff, "Exploring GLOSA systems in the field: technical evaluation and results," *Computer Communications*, vol. 120, pp. 112–124, 2018.

[40] M. Treiber, A. Hennecke, and D. Helbing, "Microscopic simulation of congested traffic," *Traffic and Granular Flow '99*, Springer, Salmon, NY, USA, pp. 365–376, 2000.

[41] W. Do, O. M. Rouhani, and L. Miranda-Moreno, "Simulation-based connected and automated vehicle models on highway sections: a literature review," *Journal of Advanced Transportation*, vol. 2019, Article ID 9343705, 14 pages, 2019.

[42] Z.-w. Yi, W.-q. Lu, L.-h. Xu, X. Qu, and B. Ran, "Intelligent back-looking distance driver model and stability analysis for connected and automated vehicles," *Journal of Central South University*, vol. 27, no. 11, pp. 3499–3512, 2020.

[43] T. Ma and B. Abdulhai, "Genetic algorithm-based optimization approach and generic tool for calibrating traffic

microscopic simulation parameters," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1800, no. 1, pp. 6–15, 2002.

[44] J. Erdmann, "SUMO's lane-changing model," in *Modeling Mobility with Open Data*, pp. 105–123, Springer, Salmon, NY, USA, 2015.

[45] D. Krajzewicz, M. Behrisch, P. Wagner, R. Luz, and M. Krumnow, "Second generation of pollutant emission models for SUMO," *Modeling Mobility with Open Data*, pp. 203–221, Springer, Berlin, Germany, 2015.

[46] A. Alam, B. Besselink, V. Turri, J. Mårtensson, and K. H. Johansson, "Heavy-duty vehicle platooning for sustainable freight transportation: a cooperative method to enhance safety and efficiency," *IEEE Control Systems Magazine*, vol. 35, no. 6, pp. 34–56, 2015.

[47] 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.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 1373391, 18 pages
https://doi.org/10.1155/2022/1373391

Research Article

CARSP: A Smart Parking System Based on Doubly Periodic Rolling Horizon Allocation Approach

Shangbin Ning[1,2], **Zhenyu Han**[1], **Yang Yang**[3,4], **Zhenzhou Yuan**[1,2] and **Xianyu Wu**[1,2]

[1]*School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*
[2]*Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing 100044, China*
[3]*School of Transportation Science and Engineering, Beihang University, Beijing 100191, China*
[4]*Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety Control, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Zhenzhou Yuan; zzyuan@bjtu.edu.cn

Received 1 April 2022; Revised 18 April 2022; Accepted 25 April 2022; Published 28 May 2022

Academic Editor: Zhihong Yao

Blind search for available parking space is accountable for most traffic congestion, accident, and pollution in cities, which severely impact people's life. Parking management based on an online smart parking system is practical to alleviate parking problems in which parking allocation is the core. However, existing researches are weak at satisfying allocation effect and speed simultaneously when solving large-scale dynamic parking allocation problem. To address this problem, we firstly construct an online "Collection-Allocation-Response" smart parking system (CARSP) to offer parking services to users and rent parking spaces from owners so as to obtain revenue for system managers. We then propose a novel Doubly Periodic Rolling Horizon allocation approach (DPRH) that circularly conduct allocation within a short period and reallocation within a long period. We formulate a narrow allocation model (without reallocation) and broad allocation model (with reallocation), both of which are binary integer programming models with the objective of maximizing system integrated benefit. We design seven performance metrics to evaluate the overall allocation effect and speed of CARSP based on DPRH. According to the three-day district-level instance in Beijing, CARSP based on DPRH performs excellently in balancing allocation effect and speed. This study is meaningful for constructing and optimizing an online smart parking system.

1. Introduction

With the swift development in urbanization and the great growth in living standards, the conflict between the increasing parking demands and scarce parking resources is becoming severer [1]. Besides parking resource shortage, inefficient parking resource usage also leads to parking difficulties, especially the increasing amounts of the blind search for available parking. Constrained by the limited land space in the urban area, it is difficult to solve the parking problem solely by constructing more parking resources. Therefore, how to efficiently utilize the limited parking resources has become a significant issue.

In practice, more and more commercial companies worldwide have built online smart parking systems with the progress in Internet and communication technology, such as Airparking (China), Pavemint (the United States), and Nokisaki Parking (Japan). Applications of the above smart parking systems indicate that appropriate dynamic allocation approaches are the key to effectively maximizing system integrated benefit and maintaining allocation speed.

Most online smart parking systems adopt First-Book-First-Serve (FBFS), and some studies proposed event-driven allocation approaches based on FBFS. In FBFS, each unallocated demand has a unique allocation priority, and each allocation process uses all unallocated resources to match one unallocated demand at demand submission time. Obviously, FBFS has an extremely high allocation speed since each demand is allocated immediately. However, FBFS sacrifices the allocation effect since each demand is

allocated independently and myopically without any overall planning.

As the research scale expands, researchers recognize that a large number of demands will be submitted to the system concurrently, while FBFS will result in unnecessary waiting of users. To improve the allocation effect, researchers proposed time-driven allocation approaches based on Rolling Horizon (RH). Based on Rolling Horizon allocation approaches, parking system collects demands in each allocation interval and allocates them simultaneously at the end time point of the interval. Demands participating in each allocation have the same priority. Naturally, unallocated demands could collectively participate in each allocation time point to achieve a better allocation effect. The unallocated demands refer to both newly submitted and previously submitted but unallocated demands.

On this basis, some researchers noticed that users have different demand submission preferences: some prefer submitting demands a long time in advance before the trip, while some prefer submitting demands a short time in advance during the trip. Since the parking demands and resources are submitted to the system dynamically as time rolls on, the system managers can reallocate the allocated but unoccupied demands and meanwhile allocate the unallocated demands at each allocation time point to improve the resource utilization so as to improve the system's integrated benefit. The allocated but unoccupied demands and unallocated demands are collectively regarded as unoccupied demands.

The allocation is called narrow allocation (NA for short) when only unallocated demands and resources participate, while the allocation is called broad allocation (BA for short) when unoccupied demands and resources participate. In other words, NA is allocation, while BA is the combination of allocation and reallocation. BA is practical and able to optimize the allocation effect as long as the system promises that these allocated but unoccupied demands can be successfully reallocated in each BA. Therefore, a rolling horizon allocation approach conducting NA at each allocation time point is Rolling Horizon Narrow Allocation (RHN), and a rolling horizon allocation approach conducting BA at each allocation time point is Rolling Horizon Broad Allocation (RHB).

Under the same condition, i.e., the same overall time horizon, allocation period, demands and resources, RHN and RHB have different allocation effects and allocation speeds. In each allocation, the scales of demands and resources in RHN are smaller than that in RHB. Thus, RHN leads to higher allocation speed, while RHB leads to a better allocation effect. Especially, the larger the research scale is, the longer the allocation period is, and the closer the parking peak is, the larger the demand and service scale difference is, hence leading to the greater allocation effect gap and allocation speed gap between RHN and RHB. In this way, both RHN and RHB have unique drawbacks.

In this paper, we firstly present an online smart parking system named CARSP to imply demands and resources collection, allocation, and response in the dynamic parking environment. CARSP not only fully considers the heterogeneity of demands and resources but also achieves accurate allocation to parking spaces rather than parking facilities, which are different from most existing smart parking systems.

To mitigate the drawbacks of FBFS, RHN, and RHB, we creatively propose the Doubly Periodic Rolling Horizon allocation approach that circularly conduct NA within a short period and BA within a long period. DPRH is a brand new attempt at a dynamic allocation approach since it combines the superiorities of RHN and RHB, i.e., optimizing allocation effect and guaranteeing allocation speed. NA and BA models are formulated to optimize the integrated benefit for both system managers and CPLEX solvers are applied to solve the model at each allocation time point to obtain the optimal solution.

The rest of this study is structured as follows. Section 2 reviews the relevant studies. In Section 3, the structure and reaction scheme of "Collection-Allocation-Response" smart parking system are presented. In Section 4, the novel Doubly Periodic Rolling Horizon allocation approach is described. Meanwhile, the narrow and broad allocation models are formulated. In Section 5, seven performance metrics are designed to evaluate the allocation effect and speed. Meanwhile, a three-day district-level instance in Beijing, China, is studied. Section 6 gives the conclusions.

2. Literature Review

In existing studies on dynamic parking allocation problems, some focused on proposing event-driven allocation approaches. Raichura and Padhariya [2] proposed a smart parking allocation system containing a static allocation process and a dynamic allocation process. The dynamic allocation process was event-driven, which meant only when a user arrived at the allocated parking facility earlier would the user be reallocated to a more suitable parking facility. Nugraha and Tanamas [3] proposed a dynamic allocation subsystem based on an event-driven approach to reallocate FBFS results when trigger events happen. To maximize the user's comfort level and the owner's revenue simultaneously, Hassija et al. [4] proposed an event-driven parking space allocation framework based on the Virtual Voting and Adaptive Pricing Algorithm.

To obtain better allocation effects, some studies focused on proposing time-driven allocation approaches to allocate parking demands and parking services group-by-group at each allocation time point, in which RHN, RHB, and static allocation models are improved.

2.1. Developments on Rolling Horizon Narrow Allocation. In the studies on time-driven allocation approaches based on RH, some focused on Narrow Allocation and proposed a rolling horizon narrow allocation approach.

Zou et al. [5] firstly formulated a static allocation model from the perspective of society. Afterward, a dynamic model with the objective of maximizing social welfare was formulated and solved at each decision time interval as time rolled on. Both static and dynamic models were improved

with a payment scheme to align users' selfish intents with the system managers' intents. Lei and Ouyang [6] regarded the problem as a Stackelberg leader-follower game and formulated a multiperiod bilevel model. The upper level decided the dynamic parking price of each parking facility and aimed at maximizing system profits, while the lower level decided the dynamic allocation and aimed at minimizing users' disutility. An approximate dynamic programming approach was proposed to solve the model at each decision time point. He et al. [7] proposed an RHN to solve the dynamic parking allocation problem. A binary integer programming model with the objective of minimizing users' costs was formulated. At each decision time point, the static model was solved by ILOG CPLEX. Yan et al. [8] focused on the dynamic parking allocation problem under uncertain demand and supply. Based on RHN, a mixed-integer programming model with the objective of minimizing users' costs was formulated and solved by an iterative two-stage heuristic algorithm.

2.2. Developments on Rolling Horizon Broad Allocation. In the studies on time-driven allocation approaches based on RH, some focused on Broad Allocation and proposed a rolling horizon broad allocation approach (RHB for short).

Geng and Cassandras [9] and Geng and Cassandras [10] firstly defined users' costs function as the weighted sum of the total monetary costs and walking distance between the parking space and actual destination. Then mixed-integer linear programming models with the objective of minimizing users' costs were formulated. Afterward, the RHB was proposed to allocate users including unallocated users and allocated users at each decision time point. However, users who were away from their destination were kept in a waiting queue and forbidden to be allocated, which not only increased users' waiting costs but also failed to respond timely to users on providing parking guarantees. Based on the above studies, Kotb et al. [11] combined dynamic and static parking space allocation. A mixed-integer programming model aiming at minimizing users' costs was formulated, in which users' costs function was similar to the above studies. Under RHB, users in the dynamic reservation would participate in BA at each decision time point until they reached destination zones, while users in static reservation would not participate in subsequent allocation once they were successfully allocated. Mladenović et al. [12] proposed a four-layer RH framework to tackle the real-time updates of parking demands and parking spaces. During each time interval, users that had not arrived at their parking will be reallocated. A binary integer programming model aiming to minimize users' costs was formulated. A heuristic algorithm and an exact algorithm were implied to solve the allocation model at each decision time point. Zhao et al. [13] researched the large-scale parking allocation problem in a dynamic parking environment of mixed automated and human-driven vehicles. A mixed-integer programming model aiming at minimizing users' costs was formulated and solved by Monte Carlo Tree Search at each decision time point.

2.3. Improvements in Static Allocation Models. Studies focused on improving static allocation models are from the perspectives of system managers, users, and society.

From the perspective of system managers, the objective of allocation models was generally maximizing system profits. Xu et al. [14] researched the private parking space sharing problem with the market design theory and extended the parking space allocation mechanism with money flow. Static allocation mechanisms named TTCD and PC-TTCC for lessor-like agents and lessee-like agents, respectively, were proposed. Yang et al. [15] supposed the system received parking supply and demand before a certain time and formulated a binary integer programming model aiming at maximizing system profits. User costs containing walking time cost and rejection cost were also considered and transferred into penalties for the system managers in the objective function. Han et al. [16] focused on the sharing of residential parking spaces and the improvements in parking resource utilization. A binary integer programming model with the objective of maximizing system profits was formulated. Ning et al. [17] focused on the private shared parking spaces allocation and formulated a binary integer programming model aiming at maximizing system profits, in which costs for failing allocating demands were also considered in the objective function as the penalty for the system managers. Jiang and Fan [18] focused on the users' parking unpunctually and formulated a binary stochastic linear programming model with the objective of maximizing system profits. The stochastic programming model was then transformed into an expectation model by formulating the parking probability function of users and owners.

From the perspective of users, the objective of allocation models was generally minimizing user costs. When constructing user cost functions, traveling time for parking, walking distance or walking time from parking facilities to destinations, and parking price were major considerations. Arellano-Verdejo and Alba [19] focused on the available parking space allocation in a city according to users' preferences. The user costs function was formulated by combining the driving distance with the preference deviation distance. The driving distance was the distance between users' current positions and the allocated parking spaces, and the preference deviation distance was the distance between the allocated parking spaces and the parking spaces desired by users. Meanwhile, an evolutionary algorithm based on Steady-State Evolutionary Algorithm was designed to solve the problem. To solve the problem brought by unpunctuality, Li et al. [20] formulated a shared parking allocation optimization model considering user's default, which possibilities were measured by credit value. The objective was to maximize the shared parking rate for all users. The problem was transformed into a vertex coloring problem and solved by an improved ant colony algorithm.

From the perspective of society, the objective of allocation models was generally maximizing shared parking resource utilization. Shao et al. [21] formulated an allocation model to embrace private residential parking space sharing between residents and public users. The objective of the model was to maximize resource utilization under given

Table 1: Summary of the studies on rolling horizon allocation approach in the online smart parking system.

Study	Participants	Allocation approach	Case study		
			Scale (total demands-total spaces)	Time horizon	Allocation period
Zou et al. [5]	U-S	RHN	200-100	20 periods	NM
Lei and ouyang [6]	U-S	RHN	NM-295	360 min	30 min
He et al. [7]	U-S	RHN	1500-1000	1080 min	3 min
Yan et al. [8]	U-S-O	RHN	300-200	720 min	10 min
Geng and cassandras [9]	U-S	RHB	NM-30	300 min	10/15/20/25/30 s
Geng and cassandras [10]	U-S	RHB	NM-2611	3000 min	1 min
Kotb et al. [11]	U-S	RHB	NM-112	720 min	NM
Mladenović et al. [12]	U-S	RHB	1000-480	1440 min	1 min
Zhao et al. [13]	U-S	RHB	6944-NM	1667 min	10 s
This study	U-S-O	DPRH	31494-1800	4320 min	0.5/1/5/10/15 min

Note. U: users; S: system managers; O: owners; RHN: rolling horizon narrow allocation approach; RHB: rolling horizon broad allocation approach; DPRH: doubly periodic rolling horizon allocation approach; NM: not mentioned.

demand and supply time windows and preset parking prices.

Some researchers considered all three perspectives, and therefore, formulated allocation models with multiple objectives. Jiang et al. [22] constructed a two-stage method to allocate private parking spaces. The first stage screened the available parking spaces according to demand walking distance and parking price. The second stage allocated spaces to demands by a formulated multiobjective allocation model aiming to maximize users' satisfaction, private idle parking space owners' satisfaction, and the system profits. The model was solved by an improved nondominated sorting genetic algorithm II (INSGA II).

Table 1 shows the summary of the studies on rolling horizon allocation approaches in smart parking systems, in which this study is included.

Apparently, existing studies on rolling horizon allocation approaches were either RHN or RHB. Few had proposed a rolling horizon allocation approach that integrated the superiorities of RHN and RHB.

3. "Collection-Allocation-Response" Smart Parking System

In the dynamic parking environment, "Collection-Allocation-Response" smart parking system (CARSP for short) collects enough demand and service data at one time, allocates demands and services without unnecessary interaction processes, and responds to users and owners with necessary notifications. Under this consideration, CARSP is composed of a collection center (CC for short), allocation center (AC for short), and response center (RC for short). CC and RC are responsible for data transmission through human-machine interaction in the front end, while AC is accountable for data storing, allocating, and updating in the back end. The framework of CARSP is shown in Figure 1. All notations of this paper are summarized in Table 2.

3.1. Collection Center

3.1.1. Demand Collection. CC collects parking demands from either leisure or urgent users. Leisure users refer to users with a long planning time. Urgent users refer to users with a short planning time. Planning time is the duration between demand submission time and demand start time.

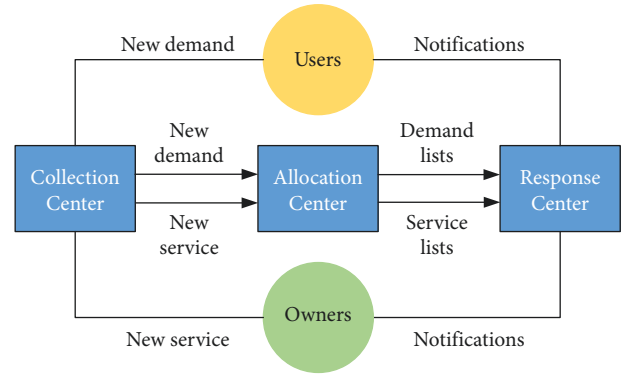


Figure 1: Framework of CARSP.

The overall time horizon is discretized into a set of parking time intervals with length τ^0. Each user can submit several parking demands. Each parking demand corresponds to a unique demand time window composed of continuous parking time intervals.

Each parking demand $i \in S_i\{1, 2, \ldots, I\}$ contains the following information: submission time t_i^P, time window $[t_i^S, t_i^E]$, destination location $[\text{long}_i, \text{lat}_i]$, the maximum acceptable walking distance $d_i^{\max}$, the maximum acceptable parking price $p_i^{\max}$, and the maximum acceptable waiting time $w_i^{\max}$.

3.1.2. Service Collection. CC collects parking spaces in either public or private parking facilities from owners. Public parking facilities refer to the parking facilities that are fully open to the public, such as the public parking garages and the curbside parking spaces. Private parking facilities refer to the parking facilities that are personal belongings, such as the residential parking garages. In a certain area, both public and private parking facilities compose the set of parking facilities $S_f\{1, 2, \ldots, F\}$.

Based on the renting schemes, parking spaces are classified into short-term renting and long-term renting parking spaces. Owners can choose either renting scheme according

Table 2: Notations of this paper.

Set	Description
$S_t\{1, 2, \ldots, T\}$	The overall time horizon
$S_t^A\{T_A^1, T_A^2, \ldots, T_A\}$	The set of allocation time points
$S_t^B\{T_B^1, T_B^2, \ldots, T_B\}$	The subset of BA time points
$S_t^N\{T_N^1, T_N^2, \ldots, T_N\}$	The subset of NA time points
$S_i\{1, 2, \ldots, I\}$	The set of total parking demands
$S_j\{1, 2, \ldots, J\}$	The set of total parking services
$S_f\{1, 2, \ldots, F\}$	The set of parking facilities
$S_i^U(t)$	The unallocated parking demand set at time point t
$S_i^{UN}(t)$	The newly submitted parking demand set at time point t
$S_i^{UP}(t)$	The previously submitted (but unallocated) parking demand set at time point t
$S_i^A(t)$	The allocated (but not occupied) parking demand set at time point t
$S_i^{AR}(t)$	The arriving parking demand set at time point t
$S_i^{ARN}(t)$	The allocated but not arriving parking demand set at time point t
$S_i^{AP}(t)$	The approaching parking demand set at time point t
$S_i^{APN}(t)$	The allocated but not approaching parking demand set at time point t
$S_i^O(t)$	The occupied (but not terminated) parking demand set at time point t
$S_i^T(t)$	The terminated parking demand set at time point t
$S_i^F(t)$	The failed parking demand set at time point t
$S_i^N(t)$	The NA parking demand set at time point t
$S_i^B(t)$	The BA parking demand set at time point t
$S_i^B(t)$	The unallocated parking service set at time point t
$S_j^A(t)$	The allocated (but not occupied) parking service set at time point t
$S_j^{AR}(t)$	The arriving parking service set at time point t
$S_j^{ARN}(t)$	The allocated but not arriving parking service set at time point t
$S_j^{AP}(t)$	The approaching parking service set at time point t
$S_j^{APN}(t)$	The allocated but not approaching parking service set at time point t
$S_j^O(t)$	The occupied (but not terminated) parking service set at time point t
$S_j^T(t)$	The terminated parking service set at time point t
$S_j^N(t)$	The NA parking service set at time point t
$S_j^B(t)$	The BA parking service set at time point t
	Parameter description
t_i^P	The submission time of parking demand i
t_i^R	The response time of parking demand i
$[t_i^S, t_i^E]$	The time window of parking demand i
$[\text{long}_i, \text{lat}_i]$	The destination location of parking demand i
$d_i^{\max}$	The maximum acceptable walking distance of parking demand i
$p_i^{\max}$	The maximum acceptable parking price of parking demand i
$w_i^{\max}$	The maximum acceptable waiting time of parking demand i
$[t_j^S t_j^E]$	The time window of parking service j
$[\text{long}_j, \text{lat}_j]$	The parking space location of parking service j
p_j	The parking price of parking service j
r_j	The short-term renting price of parking service j 0 - if parking service j is long-term renting
r_j	The long-term renting price of parking service j 0 - if parking service j is short-term renting
q	The compensation price
τ^0	The length of parking time interval
τ	The allocation period
τ'	The BA period
T_{AR}	The time threshold to determine whether demands and services are arriving
T_{AP}	The time threshold to determine whether demands and services are approaching
$I_N(t)$	The size of NA parking demand set at time point t
$I_B(t)$	The size of BA parking demand set at time point t
I_j	The number of parking demands served by parking space j
d_{ij}	The walking distance between parking space location of parking service j and destination location of parking demand i
c_{ij}	The time window relationship between parking demand i and parking service j 1 - if the time window of parking demand i is within that of parking service j 0 - otherwise
$c_{ii'}$	The time window relationship between parking demand i and parking demand i' 1 - if the time window of parking demand i is not in conflict with that of parking demand i' 0 - otherwise
b_{jf}	The belonging relationship between parking service j and parking facility f 1 - if parking service j belongs to parking facility f 0 - otherwise

Table 2: Continued.

Set	Description
a_i	The current status of passenger demand i 1 - if passenger demand i is arriving 0 - otherwise
$\tilde{x}_{ij}$	The last allocation result between parking demand i and parking service j 1 - if parking service j was allocated to parking demand i 0 - otherwise
Variable description	
x_{ij}	The allocation result between parking de mand i and parking service j 1 - if parking service j is allocated to parking demand i 0 - otherwise.

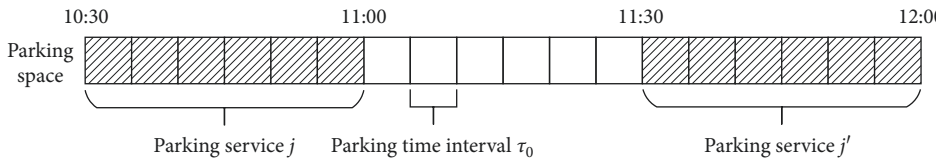


Figure 2: Example of one parking space providing two parking services.

to their situations. Though system managers totally decide the parking prices, owners can mainly decide the renting prices.

For each short-term renting parking space, the owner provides short-term usage authority to CARSP according to real-time parking arrangements. For example, commuters can provide their own parking spaces to CARSP for nine-to-five sharing on weekdays. Under the short-term renting price, the rent will be paid once the parking space is occupied for a specific period.

For each long-term renting parking space, the owner provides long-term usage authority to CARSP according to historical parking characteristics. For example, shopping mall managers can provide long-term idle parking spaces to CARSP for twenty-four-hour sharing. Under the long-term renting price, the rent will be paid in advance whether the parking space is occupied or not.

Each owner can submit several parking spaces, and each parking space can provide several parking services. Each parking service corresponds to a unique service time window composed of continuous parking time intervals.

Figure 2 shows a typical example of one parking space providing two parking services. The Time window of parking space [10: 30, 12: 00] is divided into eighteen parking time intervals with a length of 5 minutes. Time window of parking service j is [10: 30, 11: 00], and the time window of parking service j' is [11: 30, 12: 00]. The parking space is not for sharing in the time window [11: 00, 11: 30].

Each parking service $j \in S_j\{1, 2, \ldots, J\}$ contains the following information: time window $[t_j^S, t_j^E]$, parking space location $[\text{long}_j, \text{lat}_j]$, parking price p_j, short-term renting price r_j , and long-term renting price r_j'.

3.2. Allocation Center

3.2.1. Demand and Service Storing. Once CC collects a parking space or a parking demand, AC receives the data from CC immediately. All data is stored in AC by category.

At time point t, the demand data stored in AC includes the unallocated parking demand set $S_i^U(t)$, the allocated (but not occupied) parking demand set $S_i^A(t)$, the occupied (but not terminated) parking demand set $S_i^O(t)$, the terminated parking demand set $S_i^T(t)$, and the failed parking demand set $S_i^F(t)$. Furthermore, the unallocated parking demand set $S_i^U(t)$ includes both the newly submitted parking demand set $S_i^{UN}(t)$ and the previously submitted (but unallocated) parking demand set $S_i^{UP}(t)$. Meanwhile, the allocated (but not occupied) parking demand set $S_i^A(t)$ includes both the arriving parking demand set $S_i^{AR}(t)$ and the allocated but not arriving parking demand set $S_i^{ARN}(t)$. Given an appropriate time threshold T_{AR}, parking demand i in $S_i^A(t)$ belongs to the arriving parking demand set $S_i^{AR}(t)$ if $t_i^S \le t + T_{AR}$, while it belongs to the allocated but not arriving parking demand set $S_i^{ARN}(t)$ otherwise.

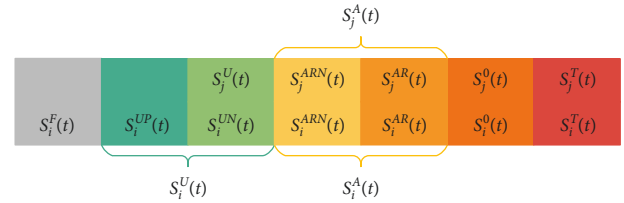


Figure 3: Smart parking spectrum graph.

Simultaneously, the service data stored in AC includes the unallocated parking service set $S_j^U(t)$, the allocated (but not occupied) parking service set $S_j^A(t)$, the occupied (but not terminated) parking service set $S_j^O(t)$, and the terminated parking service set $S_j^T(t)$. Moreover, the allocated (but not occupied) parking service set $S_j^A(t)$ includes both the arriving parking service set $S_j^{AR}(t)$ and the allocated but not arriving parking service set $S_j^{ARN}(t)$. Given the same time threshold T_{AR}, parking service j in $S_j^A(t)$ belongs to the arriving parking service set $S_j^{AR}(t)$ if $t_j^S \le t + T_{AR}$ and the allocated but not arriving parking service set $S_j^{ARN}(t)$ otherwise.

Inspired by Spectrum, a parking spectrum graph is introduced to demonstrate the transition of all demand and service sets. In Figure 3, except for the failed parking demand set $S_i^F(t)$, the further to the left, the more necessary demands and services are to be allocated. Conversely, the further to

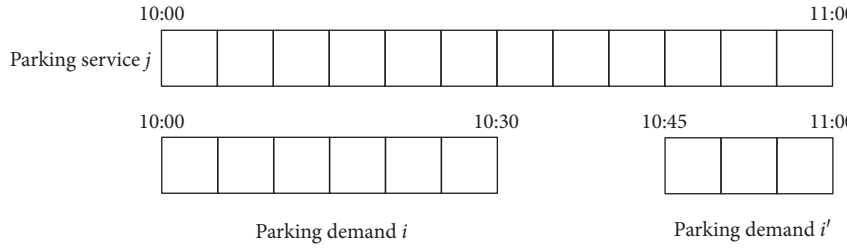


Figure 4: Example of one parking service serving two parking demands.

the right, the more parking details are required by relevant users and owners.

3.2.2. Demand and Service Allocating. At each allocation time point, AC selects specific demand and service to participate in the allocation process.

Parking service j can serve parking demand i if the following three criteria are achieved. Firstly, parking price p_j is within the maximum acceptable parking price $p_i^{\max}$. Secondly, walking distance d_{ij} (calculated by destination location $[\text{long}_i, \text{lat}_i]$ and parking space location $[\text{long}_j, \text{lat}_j]$) is within the maximum acceptable walking distance $d_i^{\max}$. Thirdly, time window $[t_i^S, t_i^E]$ is within the time window $[t_j^S, t_j^E]$. On this basis, parking service j can serve parking demand i' at the same time if time window $[t_{i'}^S, t_{i'}^E]$ is not in conflict with time window $[t_i^S, t_i^E]$.

Figure 4 shows a typical example of one parking service serving two parking demands. Time window of parking service j[10: 00, 11: 00] is divided into twelve parking time intervals with length of 5 minutes. Time window of parking demand i is [10: 00, 10: 30], and time window of parking demand i' is [10: 45, 11: 00]. Parking demand i has no time conflicts with parking demand i' so that they can be allocated to parking service j simultaneously.

It is preset that users cannot cancel or modify parking demands and owners cannot cancel or modify parking services. That is, users will accept the parking services allocated by CARSP, and owners will accept the parking demand allocated by CARSP.

3.2.3. Demand and Service Updating. Once the allocation process is complete, AC updates demand and service immediately according to the allocation results.

Unallocated parking demand i will be moved to $S_i^{\text{UP}}(t)$ to wait for the next allocation process if waiting time $t - t_i^P$ is within the maximum acceptable waiting time $w_i^{\max}$. Unallocated parking demand i will be moved to $S_i^F(t)$ and added to failed demand list if waiting time $t - t_i^P$ exceeds the maximum acceptable waiting time $w_i^{\max}$. Allocated parking demand i will be moved to $S_i^{AR}(t)$ (if $t_i^S \le t + T_{AR}$) or $S_i^{ARN}(t)$ (otherwise) and added to the allocated demand list.

Unallocated parking service j will be moved to $S_j^U(t)$ to wait for the next allocation process. Allocated parking service j will produce I_j totally allocated parking services and at most $I_j + 1$ totally unallocated parking services. All totally allocated parking services will be moved to $S_j^{AR}(t)$ (if $t_j^S \le t + T_{AR}$) or $S_j^{ARN}(t)$ (otherwise). All totally unallocated parking services will be moved to $S_j^U(t)$ to wait for the next allocation process.

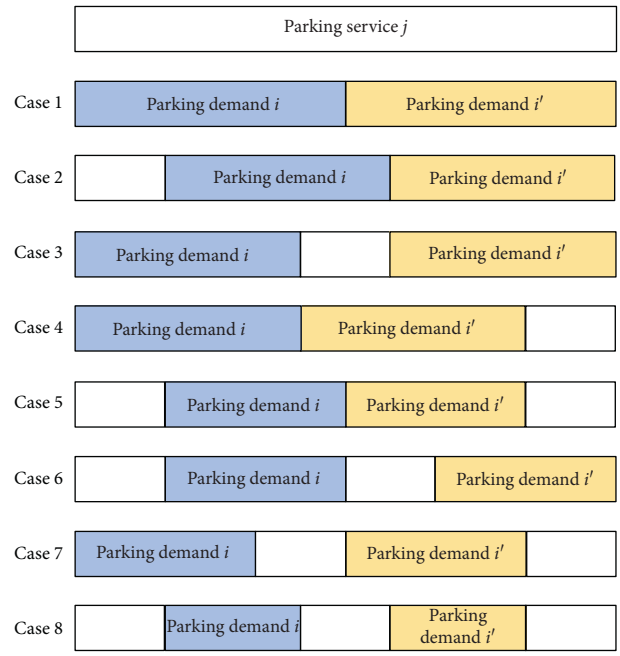


Figure 5: Example of one parking service producing several parking services.

Figure 5 shows a typical example of one parking service producing several parking services. Since parking service j serves parking demand i and parking demand i', it produces two totally allocated parking services and, at most, three totally unallocated parking services. There are eight cases due to the different relationships of time windows $[t_j^S, t_j^E]$, $[t_i^S, t_i^E]$, and $[t_{i'}^S, t_{i'}^E]$. Parking service j produces two parking services in Case 1. Parking service j produces three parking services in Cases 2, 3, and 4. Parking service j produces four parking services in Cases 5, 6, and 7. Parking service j produces five parking services in Case 8.

The overall time horizon is discretized into a set of time units $S_t\{1, 2, \ldots, T\}$. Each time unit corresponds to its end time point, so that S_t is also a set of time points. At each time point, AC also updates demand and service periodically.

Parking demand i in $S_i^{ARN}(t)$ will be moved to $S_i^{AR}(t)$ and added to arriving demand list if $t_i^S = t + T_{AR}$. Parking demand i in $S_i^{AR}(t)$ will be moved to $S_i^O(t)$ and added to the occupied demand list if $t_i^S = t$. Parking demand i in $S_i^O(t)$ will be added to leaving demand list if $t_i^E = t + T_{AR}$. Parking

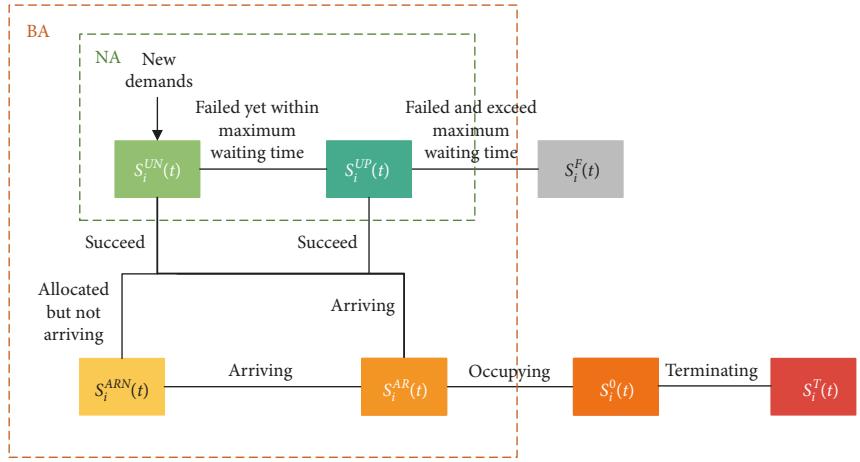


Figure 6: State transitions of demands.

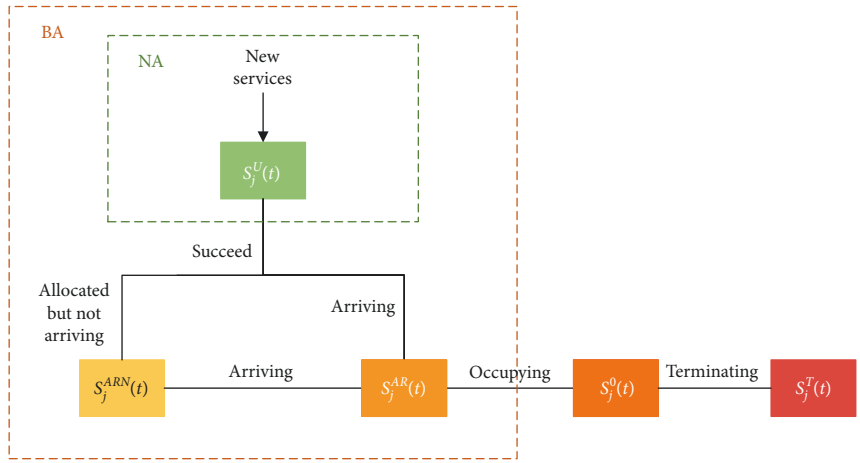


Figure 7: State transitions of services. After each update, AC sends the demand and service lists to RC.

demand i in $S_i^O(t)$ will be moved to $S_i^T(t)$ and added to terminated demand list if $t_i^E = t$.

Parking service j in $S_j^{ARN}(t)$ will be moved to $S_j^{AR}(t)$ if $t_j^S = t + T_{AR}$. Parking service j in $S_j^{AR}(t)$ will be moved to $S_j^O(t)$ and added to occupied service list if $t_j^S = t$. Parking service j in $S_j^O(t)$ will be moved to $S_j^T(t)$ and added to terminated service list if $t_j^E = t$.

Figures 6 and 7 show the state transitions of demands and services, respectively.

3.3. Response Center. Once receiving the demand and service lists from AC, RC sends specific notifications to relevant users and owners.

After each allocation process, RC will send "allocation success" notifications to users in the allocated demand list and send "allocation failure" notifications to users in failed demand list. "Allocation success" notification is the parking promise, but it does not contain a specific number and location of the parking facility and parking space. Specific number and location of parking facilities will be sent to users whose demands are in $S_i^{AR}(t)$ and the details of parking spaces will be sent to users in the occupied demand list. This kind of notification scheme provides a possibility for integrating all available parking services.

At each time point, RC will send "parking start" notifications to users in the occupied demand list and owners in the occupied service list and send "parking end" notifications to users in terminated demand list and owners in the terminated service list. [23] "Parking start" notifications should contain accurate information of parking facility and parking space to users and owners, and "parking end" notifications should contain bill information to users and owners.

Besides, RC will send "arriving reminder" notifications to users in arriving demand list and send "leaving reminder" notifications to users in leaving demand list. "Arriving reminder" notifications should prompt users to drive to the allocated parking facility on time, and "leaving reminder" notifications should prompt users to leave the parking facility on time.

It is assumed that users and owners are punctual. That is, users will arrive at and leave parking spaces on time, and owners will guarantee that parking services are available.

4. Doubly Periodic Rolling Horizon Allocation Approach

4.1. Allocation Approach Description. As a classical theory, the rolling horizon has been widely used in optimization, estimation, control, and other fields [24, 25]. Referring to the relevant studies, time-driven rolling horizon allocation (RH for short) is introduced to replace event-driven First-Book-First-Serve (FBFS for short). In a large-scale dynamic parking allocation problem, RH aims at realizing the overall optimization in each period but not the overall optimization throughout the entire time horizon.

As mentioned before, CARSP executes either NA or BA at each allocation time point. Thus, CARSP is based on Rolling Horizon Narrow Allocation if it conducts NA at each allocation time point, while CARSP is based on Rolling Horizon Broad Allocation if it conducts BA at each allocation time point. Under the same condition, RHN shows superiority in allocation speed while RHB shows superiority in allocation effect and both RHN and RHB have unique advantages and cannot be replaced.

In this way, we propose a Doubly Periodic Rolling Horizon allocation approach in CARSP to combine the superiorities of RNA and RHB.

Three principles are clarified to ensure allocation effect and allocation speed in CARSP based on DPRH.

Firstly, the unallocated parking service set S_j^U and the allocated parking service set S_j^A cannot be simply merged but should be deeply integrated before each BA process. Referring to Figure 5, the parking service is split into several parking services. Then, the several parking services will be integrated into one parking service to participate in the BA process.

Secondly, the parking guarantee for each parking demand $i \in S_i^A$ cannot be broken, which means that each allocated parking demand must be reallocated to one parking space in each BA process.

Thirdly, the parking facility for each parking demand $i \in S_i^{AR}$ cannot be changed, which means that each arriving parking demand must be reallocated to one parking space within the parking facility in each BA process. Given this, parking facility information will be sent in "Arriving reminder" notifications, and parking service information will be sent in "Parking start" notifications.

The overall time horizon is discretized into a set of allocation time points $S_t^A\{T_A^1, T_A^2, \ldots, T_A^n\}$ with period τ. $S_t^A\{T_A^1, T_A^2, \ldots, T_A^n\}$ can be divided into a subset of BA time points $S_t^B\{T_B^1, T_B^2, \ldots, T_B^n\}$ with period $\tau\prime$ and a subset of NA time points $S_t^N\{T_N^1, T_N^2, \ldots, T_N^n\}$. In RHN, $S_t^N = S_t^A$, $S_t^B = \varnothing$. In RHB, $\tau\prime = \tau$, $S_t^B = S_t^A$, $S_t^N = \varnothing$. Thus, in DPRH, we describe the relation of $\tau\prime$ and τ as $\tau\prime = m * \tau (m > 1)$ and $S_t^B \subsetneq S_t^A$, $S_t^N \subsetneq S_t^A$, $S_t^B \cup S_t^N = S_t^A$.

Figure 8 shows the process of DPRH, in which the black arrow points to the direction of time rolling, demands and services participating in NA are indicated by green rectangles, and the demands and services participating in BA are denoted by orange rectangles. The green and orange rectangles are also marked in Figures 6 and 7.

At each NA time point $t \in S_t^N$, the unallocated parking demand set $S_i^U(t)$ is the NA parking demand set $S_i^N(t)$, and the unallocated parking service set $S_j^U(t)$ is the NA parking service set $S_j^N(t)$. The NA model allocates parking demands in $S_i^N(t)$ and parking services in $S_j^N(t)$.

At each BA time point $t \in S_t^B$, the unallocated parking demand set $S_i^U(t)$ and the allocated parking demand set $S_i^A(t)$ are simply merged into the BA parking demand set $S_i^B(t)$, and the unallocated parking service set $S_j^U(t)$ and the allocated parking service set $S_j^A(t)$ are deeply integrated into the BA parking service set $S_j^B(t)$. The BA model allocates parking demands in $S_i^B(t)$ and parking services in $S_j^B(t)$.

DPRH can theoretically reach the ideal allocation effect and speed. However, in practice, especially in a large city, BA demand and service sets $S_i^B(t)$ and $S_j^B(t)$ should be controlled within reasonable scales due to the limitation of computing power. Considering the differences in allocation urgency, the unallocated demand and service sets $S_i^U(t)$ and $S_j^U(t)$ should be totally reserved in BA, while the allocated demand and service sets $S_i^A(t)$ and $S_j^A(t)$ should be partly erased from BA. Given an appropriate time threshold T_{AP} ($T_{AP} > T_{AR}$), parking demand i in $S_i^A(t)$ belongs to the approaching parking demand set $S_i^{AP}(t)$ if $t + T_{AR} < t_i^S \le t + T_{AP}$, while it belongs to the allocated but not approaching parking demand set $S_i^{APN}(t)$ if $t_i^S > t + T_{AP}$. Simultaneously, parking service j in $S_j^A(t)$ belongs to the approaching parking service set $S_j^{AP}(t)$ if $t + T_{AR} < t_j^S \le t + T_{AP}$ and the allocated but not approaching parking service set $S_j^{APN}(t)$ if $t_j^S > t + T_{AP}$. Obviously, the approaching demand and service sets $S_i^{AP}(t)$ and $S_j^{AP}(t)$ should be reserved to participate in BA.

4.2. Allocation Models. In DPRH, NA and BA are conducted at specific allocation time points. Both models are formulated to maximize revenue for system managers and minimize waiting time for users. We integrate user waiting time into system cost so that the objectives of both models are maximizing integrated benefit for system. The integrated benefit consists of system revenue and system cost. System revenue comes from the parking fees paid by allocated users. Under parking price p_j, parking fees are positively correlated with occupancy time of parking demands. System cost refers to not only the renting cost paid to owners but also the compensation cost paid to users waiting to be allocated. Under short-term renting price r_j, renting cost is positively correlated with occupancy time of parking demands. Under long-term renting price r_j', renting cost is not correlated with occupancy time of parking demands. Under compensation price q, compensation cost is positively correlated with waiting time of parking demands.

Five binary parameters are designed in NA and BA models: (1) c_{ij} expresses the time window relationship between parking demand i and parking service j. $c_{ij} = 1$ if time window of parking demand i is within that of parking service j, while $c_{ij} = 0$ otherwise. (2) $c_{ii'}$ expresses the time window relationship between parking demand i and parking demand

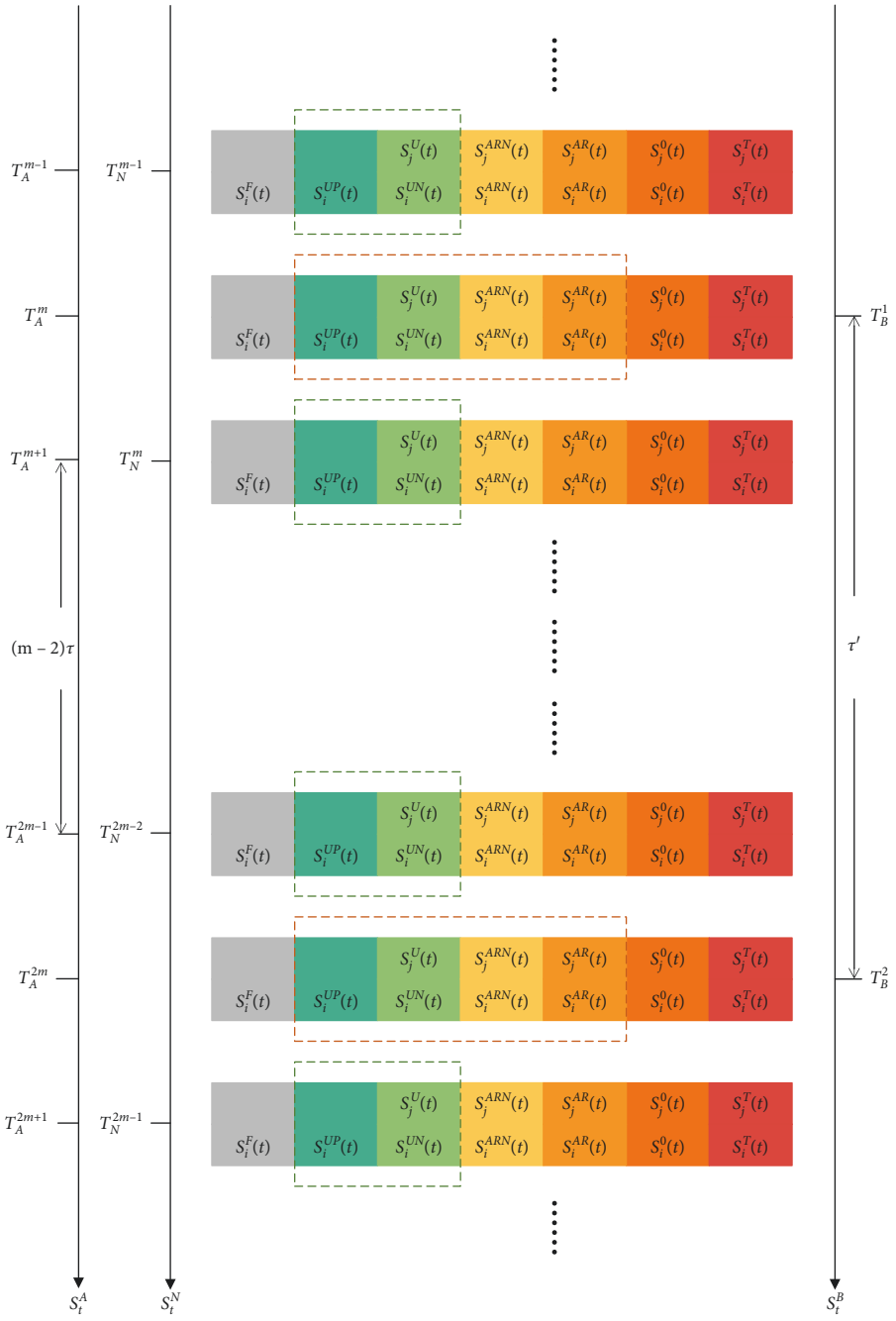


Figure 8: Process of DPRH.

i'. $c_{ii'} = 1$ if time window of parking demand i is not in conflict with that of parking demand i', while $c_{ii'} = 0$ otherwise. (3) b_{jf} indicates the belonging relationship between parking service j and parking facility f. $b_{jf} = 1$ if parking service j belongs to parking facility f, while $b_{jf} = 0$ otherwise. (4) a_i denotes the current status of passenger demand i. $a_i = 1$ if passenger demand i is arriving, while $a_i = 0$ otherwise. (5) $\tilde{x}_{ij}$ represents the last allocation result between parking demand i and parking service j. $\tilde{x}_{ij} = 1$ if parking service j was allocated to parking demand i in the last allocation, while $\tilde{x}_{ij} = 0$ otherwise.

Besides, a binary variable x_{ij} is introduced to represent the allocation result between parking demand i and parking service j. $x_{ij} = 1$ if parking service j is allocated to parking demand i, while $x_{ij} = 0$ otherwise.

4.2.1. Narrow Allocation Model. Narrow allocation model is formulated as a binary integer programming model with the objective of maximizing system integrated benefit at each NA time point $t \in S_t^N$, as formula (1) shows and it is subject to constraints (2)-(7).

$$\max Z_N(t) = \sum_{i \in S_i^N(t)} \sum_{j \in S_j^N(t)} \left(p_j - r_j\right) \cdot \left(t_i^E - t_i^S\right) \cdot x_{ij} - q \cdot \tau \cdot \sum_{i \in S_i^N(t)} \left(1 - \sum_{j \in S_j^N(t)} x_{ij}\right), \tag{1}$$

$$\sum_{j \in S_j^N(t)} x_{ij} \le 1, \quad \forall i \in S_i^N(t), \tag{2}$$

$$x_{ij} \le c_{ij}, \quad \forall i \in S_i^N(t), \forall j \in S_j^N(t), \tag{3}$$

$$x_{ij} + x_{i\prime j} \le c_{ii\prime} + 1, \quad \forall i, i\prime \in S_i^N(t), \quad i \ne i\prime, \quad \forall j \in S_j^N(t), \tag{4}$$

$$d_{ij} \cdot x_{ij} \le d_i^{\max}, \quad \forall i \in S_i^N(t), \quad \forall j \in S_j^N(t), \tag{5}$$

$$p_j \cdot x_{ij} \le p_i^{\max}, \quad \forall i \in S_i^N(t), \quad \forall j \in S_j^N(t), \tag{6}$$

$$x_{ij} \in \{0, 1\}, \quad \forall i \in S_i^N(t), \forall j \in S_j^N(t). \tag{7}$$

Constraint (2) restricts that each parking demand cannot be allocated to more than one parking service. Constraint (3) claims that each parking demand cannot be allocated to the conflicting parking service. Constraint (4) implies that every two conflicting parking demands cannot be allocated to the same parking service. Constraint (5) ensures that each parking demand cannot be allocated to parking services with too long walking distances. Constraint (6) assures that each parking demand cannot be allocated to parking services with too high parking prices. Constraint (7) defines variable x_{ij}.

4.2.2. Broad Allocation Model. Broad allocation model is formulated as a binary integer programming model with the objective of maximizing system integrated benefit at each BA time point $t \in S_t^B$, as formula (8) shows.

To ensure allocation effect, different from NA model, BA model must guarantee that all unoccupied demands participating in BA are successfully allocated. Meanwhile, BA must guarantee that all approaching demands will not be reallocated to another parking facility. Constraints are shown in (9)-(16).

$$\max Z_B(t) = \sum_{i \in S_i^B(t)} \sum_{j \in S_j^B(t)} \left(p_j - r_j\right) \cdot \left(t_i^E - t_i^S\right) \cdot x_{ij} - q \cdot \tau \cdot \sum_{i \in S_i^B(t)} \left(1 - \sum_{j \in S_j^B(t)} x_{ij}\right), \tag{8}$$

$$\sum_{j \in S_j^B(t)} x_{ij} \le 1, \quad \forall i \in S_i^B(t), \tag{9}$$

$$x_{ij} \le c_{ij}, \quad \forall i \in S_i^B(t), \forall j \in S_j^B(t), \tag{10}$$

$$x_{ij} + x_{i\prime j} \le c_{ii\prime} + 1, \quad \forall i, i\prime \in S_i^B(t), \quad i \ne i\prime, \quad \forall j \in S_j^B(t), \tag{11}$$

$$d_{ij} \cdot x_{ij} \le d_i^{\max}, \quad \forall i \in S_i^B(t), \forall j \in S_j^B(t), \tag{12}$$

$$p_j \cdot x_{ij} \le p_i^{\max}, \quad \forall i \in S_i^B(t), \forall j \in S_j^B(t), \tag{13}$$

$$\sum_{j \in S_j^B(t)} x_{ij} \ge \sum_{j \in S_j^B(t)} \widetilde{x}_{ij}, \quad \forall i \in S_i^B(t), \tag{14}$$

$$\sum_{j \in S_j^B(t)} b_{jf} \cdot x_{ij} \ge \sum_{j \in S_j^B(t)} a_i \cdot b_{jf} \cdot \widetilde{x}_{ij}, \quad \forall i \in S_i^B(t), \forall f \in S_f, \tag{15}$$

$$x_{ij} \in \{0, 1\}, \quad \forall i \in S_i^B(t), \forall j \in S_j^B(t). \tag{16}$$

Constraints (9)-(14) are derived from constraints (2)-(7), respectively. Constraint (15) specifies that each allocated parking demand must be allocated to one parking space. Constraint (16) stipulates that each arriving parking demand can only be reallocated to another parking space within the same parking facility.

4.3. Performance Metrics. RHN, RHB, and DPRH are all based on RH and it is meaningless to compare the total value of the objective function or the value of the objective in each allocation. Therefore, we set seven performance metrics to assess the actual allocation performances throughout the overall time horizon rather than simply comparing the values of the objective [26]. Performance metrics (1) to (5) describe allocation effect, in which (1) represents system integrated benefit throughout the overall time horizon, (2) and (3) indicate resource utilization, (4) and (5) denote user satisfaction. Performance metric (6) expresses allocation speed. Performance metric (7) shows total computing time.

4.3.1. Total Integrated Benefit (TIB for Short). TIB refers to total integrated benefit throughout the overall time horizon, which is calculated in (3). A higher TIB means that system managers can guarantee more sustainable operation.

$$TIB = \sum_{i \in S_i} \sum_{j \in S_j} \left(p_j - r_j\right) \cdot \left(t_i^E - t_i^S\right) \cdot x_{ij} - T \cdot \sum_{j \in S_j} r_j' - q \cdot \sum_{i \in S_i} \left(t_i^R - t_i^P\right). \tag{17}$$

4.3.2. Service Temporal Utilization (STU for Short). STU refers to the ratio of total occupancy time to total available

parking time, which is calculated in (4). A higher STU means that system managers can utilize parking services more adequately.

$$STU = \frac{\sum_{i \in S_i} \left(t_i^E - t_i^S\right) \cdot x_{ij}}{\sum_{j \in S_j} \left(t_j^E - t_j^S\right)}. \tag{18}$$

4.3.3. Effective Service Temporal Utilization (ESTU for Short). ESTU refers to the ratio of total occupancy time to total effective parking time, which is calculated in (5). Effective parking time is the time of the parking service that has been allocated to at least one parking demand. A higher ESTU means that system managers can utilize the allocated parking services more efficiently.

$$ESTU = \frac{\sum_{i \in S_i} \left(t_i^E - t_i^S\right) \cdot x_{ij}}{\sum_{j \in S_j} \left(t_j^E - t_j^S\right) \cdot y_j}. \tag{19}$$

4.3.4. Allocation Success Probability (ASP for Short). ASP refers to the ratio of allocated parking demands to total parking demands, which is calculated in (6). A higher ASP means that users can obtain parking services more easily.

$$ASP = \frac{\sum_{i \in S_i} \sum_{j \in S_j} x_{ij}}{I}. \tag{20}$$

4.3.5. Average Planning Time (APT for Short). APT refers to the average duration between demand submission time and demand start time, which is calculated in (7). A shorter APT means that urgent users can obtain parking services more easily.

$$APT = \frac{\sum_{i \in S_i} \sum_{j \in S_j} \left(t_i^S - t_i^P\right) \cdot x_{ij}}{\sum_{i \in S_i} \sum_{j \in S_j} x_{ij}}. \tag{21}$$

4.3.6. Average Waiting Time (AWT for Short). AWT refers to the average duration between demand submission time and response time, which is calculated in (8). A longer AWT means that users can obtain responses faster.

$$AWT = \frac{\sum_{i \in S_i} \left(t_i^R - t_i^P\right)}{I}. \tag{22}$$

4.3.7. Total Computing Time (TCT for Short). TCT refers to the total computing time throughout the overall time horizon. Under the same computing environment, a shorter TCT means that the approach is more efficient and more practical.

5. Large-Scale Real-World Instance

5.1. Instance Setup. Figure 9(a) shows the district in Chaoyang District, Beijing, China, where the historical parking data come from. Figure 9(b) shows the normalized district, of which the length and width are both 1 km. There are five adjacent blocks in the district. Blocks with the same color have the same land use. The green block is for business, the yellow block is for commercial, and the orange block is for residential. The solid circle in each block represents a parking facility. The numbers beneath each solid circle represent the coordinates of each parking facility. The five parking facilities rent parking spaces to CARSP for sharing. The sizes of blocks are also set as Figure 9(b) shows. Table 3 shows the parameters of the instance.

The instance selects historical parking data from 00 : 00 on April 20th, 2018, to 00 : 00 on April 23rd, 2018, so that the overall time horizon is 3 days. The overall time horizon is divided into 4320 time units with length of 1 minute. The parking time interval is set to be 5 minutes.

The total number of historical parking demands is 31494. Figures 10 and 11 show the start time and end time distribution of parking demands [27]. For each parking demand, the submission time varies within [5, 1440] minutes before its start time. The maximum acceptable walking distance varies within [100, 700] meters. The maximum acceptable parking price varies within [0.5, 1.2] CNY per 5 minutes according to Beijing Parking Charge Standard. The maximum acceptable waiting time varies within [1, 10] minutes.

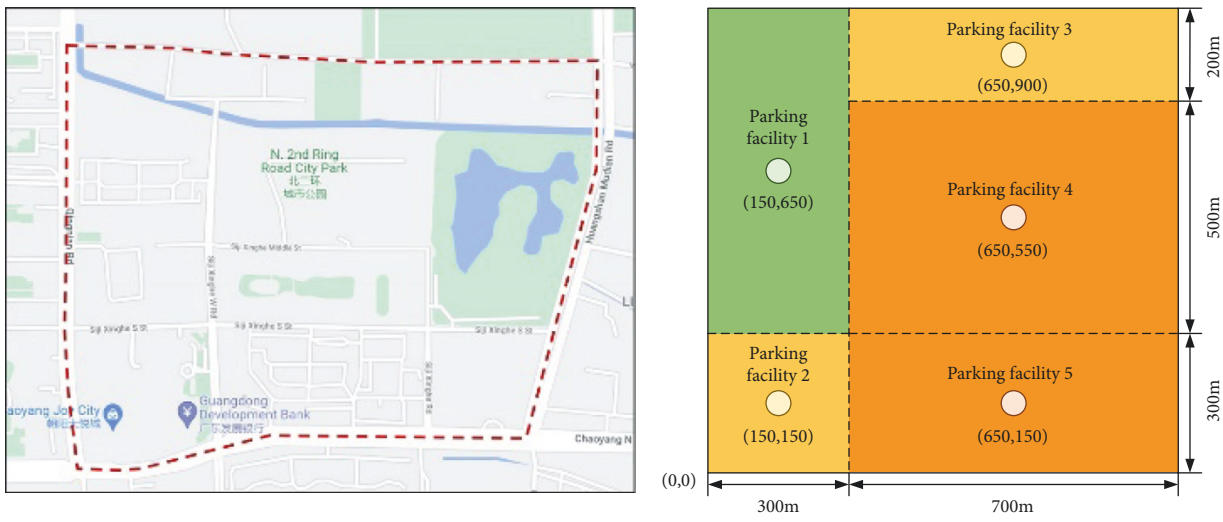
The total number of parking services is 1800. Table 4 shows the details of parking services. According to land use, parking facilities 1, 2, and 3 provide long-term renting parking services, while parking facilities 4 and 5 provide short-term renting parking services. For each short-term renting parking service, the time window varies within [1, 4320]. The renting price is 0.5 CNY per 5 minutes. The parking price varies within [0.6, 1.2] CNY per 5 minutes. For each long-term renting parking service, the time window is [1, 4320]. The renting price is 0.1 CNY per 5 minutes. The parking price is 0.5 or 0.7 CNY per 5 minutes.

Besides, the compensation price is set to be 0.025 CNY per minute. The arrival time threshold is 15 minutes. The approaching time threshold is 30 minutes.

The instance sets thirteen groups to test the performances of different allocation approaches with different allocation models under different allocation periods. Table 5 shows the details of the test groups. Group 1 is the FBFS group. Groups 2 to 5 are the RHN groups, in which $\tau \in \{0.5, 1, 5, 10\}$ and $\tau' = 0$. Groups 6 to 9 are the RHB groups, in which $\tau \in \{0.5, 1, 5, 10\}$ and $\tau' = \tau$. Groups 10 to 13 are the DPRH groups, in which $\tau \in \{0.5, 1\}$ and $\tau' \in \{5, 10\}$.

The CARSP programs based on FBFS, RHN, RHB, DPRH allocation approaches are all encoded in Visual Studio 2017 using C# language. The software used to solve the allocation problem at each allocation time point is IBM ILOG CPLEX Solver 12.6.3. All computations were performed on a personal computer with Intel Core i5-7200U CPU, 8G RAM, and a 64 bit Windows 10 Operating System.

5.2. Instance Results. We set two tests to study the performance of CARSP based on FBFS, RHN, RHB, and DPRH



(a) (b)

Figure 9: District of the instance.

Table 3: Parameters of the instance.

Parameters	Descriptions	Value
T	The number of time units	4320
I	The number of parking demands	31494
J	The number of parking services	1800
F	The number of parking facilities	5
$d_i^{\max}$	The maximum acceptable walking distance of parking demand i (m)	[100, 700]
$p_i^{\max}$	The maximum acceptable parking price of parking demand i (CNY/5 min)	[0.5, 1.2]
$w_i^{\max}$	The maximum acceptable waiting time of parking demand i (min)	[1, 10]
p_j	The parking price of parking service j (CNY/5 min)	[0.5, 1.2]
r_j	The short-term renting price of parking service j (CNY/5 min)	0.5
r_j'	The long-term renting price of parking service j (CNY/5 min)	0.1
q	The compensation price (CNY/min)	0.025
ρ	The split penalty coefficient (CNY/min)	0.001
τ^0	The length of parking time interval (min)	5
T_{AR}	The time threshold to determine whether demands and services are arriving (min)	15
T_{AP}	The time threshold to determine whether demands and services are approaching (min)	30

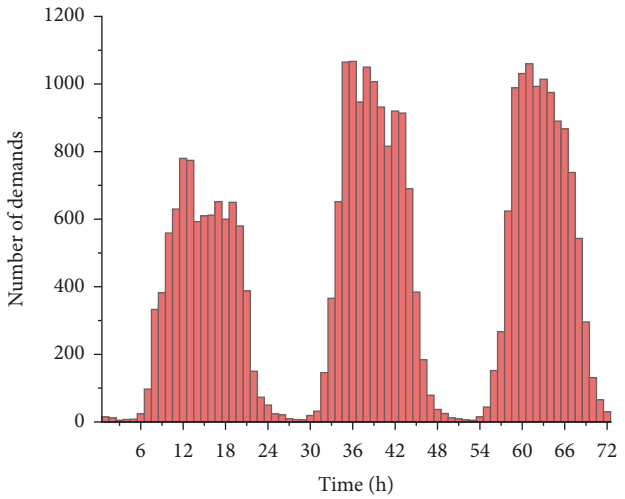


Figure 10: Start time distribution of parking demand.

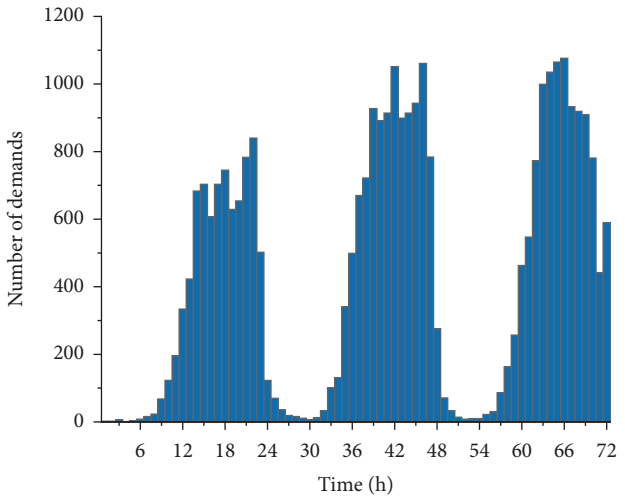


Figure 11: End time distribution of parking demand.

Table 4: Details of parking services.

Parking facility	Number of parking services	Renting type	Renting price (CNY/5 min)	Parking price (CNY/5 min)
1	400	Long-term	0.1	0.5
2	600	Long-term	0.1	0.7
3	400	Long-term	0.1	0.7
4	200	Short-term	0.5	[0.6, 1.2]
5	200	Short-term	0.5	[0.6, 1.2]

Table 5: Details of the test groups.

Group index	Group code	Allocation approach	Allocation models	τ (min)	τ' (min)
1	FBFS	FBFS	—	—	—
2	RHN-0.5	RHN	NA model	0.5	0
3	RHN-1			1	0
4	RHN-5			5	0
5	RHN-10			10	0
6	RHB-0.5	RHB	BA model	0.5	0.5
7	RHB-1			1	1
8	RHB-5			5	5
9	RHB-10			10	10
10	DPRH-0.5-5	DPRH	NA model BA model	0.5	5
11	DPRH-0.5-10			0.5	10
12	DPRH-1-5			1	5
13	DPRH-1-10			1	10

Table 6: Performances of different groups over three days.

Group code	TIB (CNY)	STU	ESTU	ASP	APT (min)	AWT (s)	TCT (h)
FBFS	276239.40	0.450	0.529	0.580	801.05	0.06	0.29
RHN-0.5	277690.72	0.452	0.532	0.573	805.28	15.15	0.47
RHN-1	277370.12	0.452	0.531	0.571	806.58	30.37	0.42
RHN-5	278382.68	0.454	0.533	0.566	808.52	158.74	1.11
RHN-10	276771.09	0.452	0.531	0.566	809.70	317.21	0.99
RHB-0.5	283237.34	0.458	0.538	0.583	790.82	24.85	26.84
RHB-1	284196.22	0.459	0.539	0.582	792.15	41.18	12.75
RHB-5	285339.14	0.460	0.541	0.581	791.09	172.57	4.89
RHB-10	284600.37	0.459	0.540	0.582	790.15	337.12	3.66
DPRH-0.5-5	282422.16	0.457	0.537	0.581	791.82	16.61	3.85
DPRH-0.5-10	280797.81	0.455	0.536	0.577	795.10	15.28	2.10
DPRH-1-5	283141.95	0.458	0.538	0.582	790.34	32.95	3.73
DPRH-1-10	282115.44	0.457	0.538	0.578	792.69	31.65	2.04

when applying to large-scale real-world instance. Test 1 aims to explore the effects of different allocation approaches. Test 2 aims to explore the effects of different allocation periods.

The performances of different groups over three days are shown in Table 6.

5.2.1. Test 1: Performance of Different Allocation Approaches. Test 1 aims to seek the effects of different allocation approaches. Min-Max normalization is applied to normalize the performance metric values of thirteen groups. In this way, under each performance metric, the strengths and weaknesses of each allocation approach can be demonstrated more clearly. The allocation approach is better if TIB, STU, ESTU, and ASP are higher while APT, AWT, and TCT are lower. Therefore, the normalization equation for the former kind of performance metrics is shown as (23). The normalization equation for the latter kind of performance metrics is shown as (24). Under each performance metric, Max is the maximum of thirteen performance metric values while Min is the minimum. Current is the performance metric value of the current allocation approach, while Current* is the normalized performance metric value. Therefore, the allocation approach is better if the normalized performance metric value is higher.

$$\text{Current}^* = \frac{\text{Current} - \text{Min}}{\text{Max} - \text{Min}}, \tag{23}$$

$$\text{Current}^* = \frac{\text{Max} - \text{Current}}{\text{Max} - \text{Min}}. \tag{24}$$

Figure 12. Comparisons of FBFS, RHN, RHB, and DPRH.

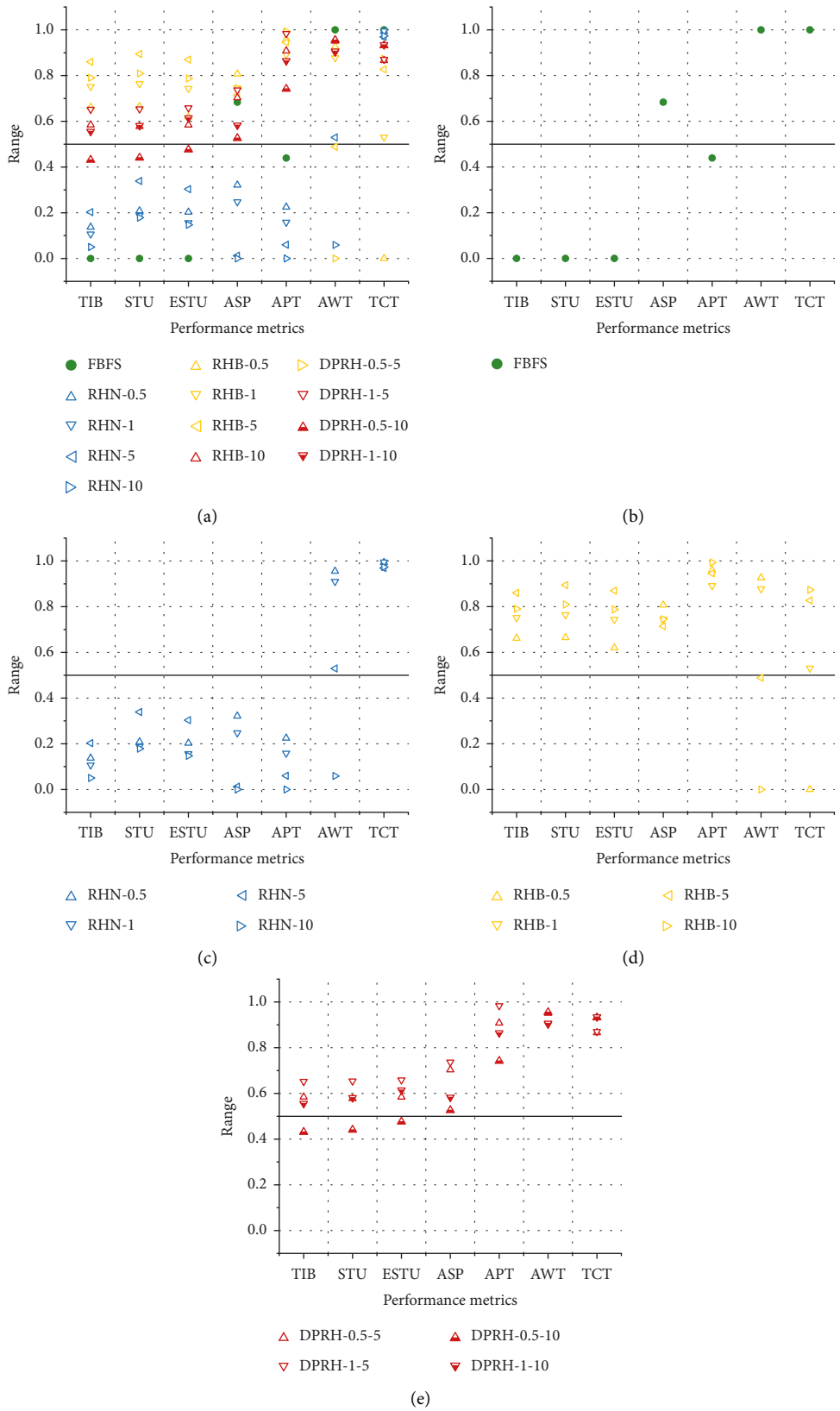


Figure 12: Shows the comparisons of FBFS, RHN, RHB, and DPRH.

For FBFS, the lowest AWT indicates excellent allocation speed. However, the lowest TIB reflects dreadful integrated benefit, the lowest STU and ESTU reflects dreadful resource utilization, and the moderate ASP and APT reflect poor user satisfaction. Meanwhile, FBFS requires the shortest computing time.

For RHN, the normalized TIB of all RHN Groups ranges widely, but on the whole, RHN leads to poorly integrated benefit. Similarly, the normalized STU and normalized ESTU of all RHN Groups range widely, but on the whole, RHN results in poor resource utilization. The normalized ASP and normalized APT of all RHN Groups are centralized in the lower half, thus, RHN leads to dreadful user satisfaction. The normalized AWT of all RHN Groups ranges widely, but on the whole, RHN results in good allocation speed. Meanwhile, RHN requires the second shortest computing time.

For RHB, the normalized TIB of all RHB Groups are centralized in the upper half, which indicates that RHB results in well integrated benefit. Similarly, the normalized STU and normalized ESTU of all RHB Groups are centralized in the upper half, which indicates that RHB results in good resource utilization. The normalized ASP and normalized APT of all RHB Groups are centralized in the upper half. Meanwhile, RHB results in the highest ASP and APT. Thus RHB results in excellent user satisfaction. The normalized AWT of all RHB Groups ranges widely, but on the whole, RHB leads to dreadful allocation speed. Meanwhile, RHB requires the longest computing time.

For DPRH, all normalized performance metrics values are centralized in the upper half. Comparing the thirteen Groups, DPRH results in the highest TIB, STU, and ESTU. Compared with FBFS, DPRH raises TIB by 3.83%, raises STU by 2.65%, and raises ESTU by 2.72%. Therefore, DPRH results in excellent integrated benefits, excellent resource utilization, good user satisfaction, and good allocation speed. Meanwhile, DPRH requires the second-longest computing time.

5.2.2. Test 2: Effects of Different Allocation Periods. Test 2 aims to seek the effects of different allocation periods in the same allocation approach and the same allocation model. Allocation periods $\tau \in \{0.5, 1, 5, 10\}$. The normalization (9) and (10) are also applied to demonstrate the strengths and weaknesses of each allocation period more clearly.

Figure 13 shows the comparisons of different allocation periods in RHN. Figure 14 shows the comparisons of different allocation periods in RHB. Figure 15 shows the comparisons of different allocation periods in DPRH.

In the light of Figure 13, in RHN, when allocation period $\tau = 0.5$ (i.e., RHN-0.5), TIB, STU, and ESTU are good. Meanwhile, ASP and APT are the best. AWT is the best and TCT is good. When allocation period $\tau = 1$ (i.e., RHN-S-1), TIB, STU, and ESTU are poor. ASP and APT are good. AWT is good and TCT is the best. When allocation period $\tau = 5$ (i.e., RHN-5), TIB, STU, and ESTU are the best. ASP and APT are poor. AWT is poor and TCT is the worst. When allocation period $\tau = 10$ (i.e., RHN-10), TIB, STU, ESTU,

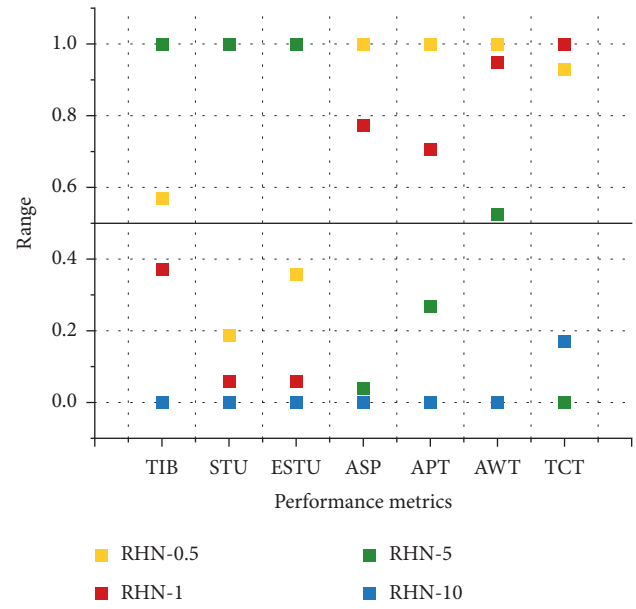


Figure 13: Comparisons of RHN with different allocation periods.

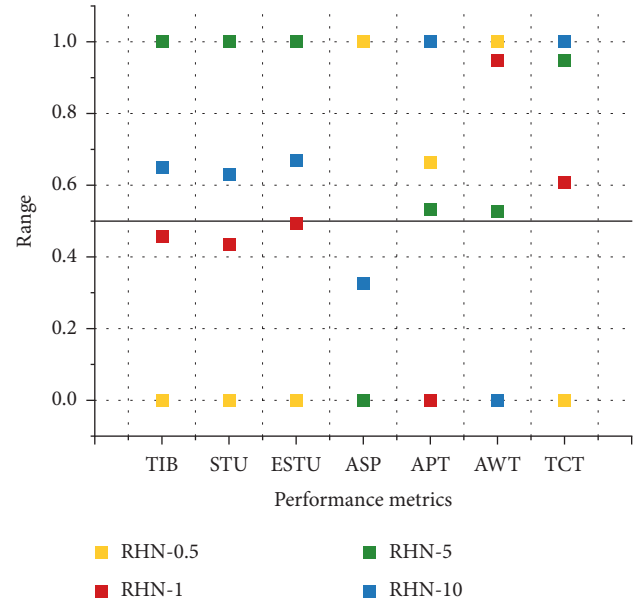


Figure 14: Comparisons of RHB with different allocation periods.

ASP, APT, and AWT are the worst while TCT is poor. Clearly, RHN-0.5 performs the best in user satisfaction and allocation speed, while RHN-5 performs the best in integrated benefit and resource utilization. From the perspective of system managers, the integrated benefit is the most concern. However, allocation speed is also vital since it is the representation of system service quality. In light of Table 6, the TIB of RHN-5 is raised by 0.25%, while the AWT of RHN-5 is ten times that of RHN-0.5.

In the light of Figure 14, in RHB, when allocation period $\tau = 0.5$ (i.e., RHB-0.5), TIB, STU, and ESTU are the worst. ASP is the best and APT is good. AWT is the best, but TCT is the worst. When allocation period $\tau = 1$ (i.e., RHB-1), TIB,

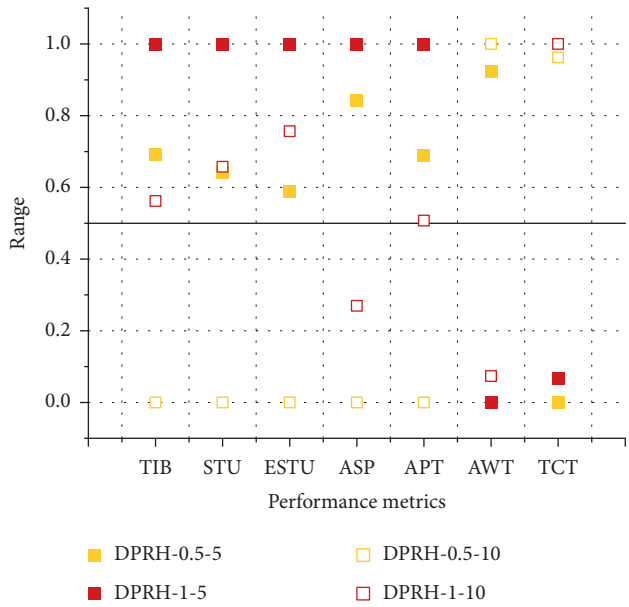


Figure 15: Comparisons of DPRH with different allocation periods.

STU, and ESTU are poor. ASP is good but APT is the worst. AWT is good but TCT is poor. When allocation period $\tau = 5$ (i.e., RHB-5), TIB, STU, and ESTU are the best. ASP is the worst and APT is poor. AWT is poor while TCT is good. When allocation period $\tau = 10$ (i.e., RHB-10), TIB, STU, and ESTU are good. ASP is good, while APT is poor. AWT is poor while TCT is good. Though RHB-5 and RHB-10 perform better than RHN; AWT is a defect that cannot be ignored.

In the light of Figure 15, in DPRH, when allocation period $\tau = 0.5$ and $\tau\prime = 5$ (i.e., DPRH-0.5-5), TIB is good, STU and ESTU are poor. ASP and APT are good. AWT is good while TCT is the worst. When allocation period $\tau = 0.5$ and $\tau\prime = 10$ (i.e., DPRH-0.5-10), TIB, STU, ESTU, ASP, and APT are the worst. AWT is the best and TCT is good. When allocation period $\tau = 1$ and $\tau\prime = 5$ (i.e., DPRH-1-5), TIB, STU, ESTU, ASP, and APT are the best. AWT is the worst and TCT is poor. When allocation period $\tau = 1$ and $\tau\prime = 10$ (i.e., DPRH-1-10), TIB is poor. STU and ESTU are good. ASP and APT are poor. AWT is poor while TCT is the best. In summary, it is clear that when τ is the same, DPRH with $\tau\prime = 0.5$ achieves higher integrated benefit, higher resource utilization, and higher user satisfaction but a lower allocation speed than DPRH with $\tau\prime = 1$. When $\tau\prime$ is the same, DPRH with $\tau = 1$ achieves higher integrated benefit, higher resource utilization, and higher user satisfaction but lower allocation speed than DPRH with $\tau = 0.5$.

6. Conclusion

This study researched the large-scale dynamic parking allocation problem by proposing a smart parking system CARSP based on a novel rolling horizon allocation DPRH.

The "Collection-Allocation-Response" smart parking system (CARSP) is constructed to amply describe trilateral requirements, deeply portray trilateral relationships, and thoroughly realize trilateral data transmission in the dynamic smart parking environment. The Doubly Periodic Rolling Horizon allocation approach (DPRH) is proposed to combine the superiorities of RHN and RHB. The NA and BA models integrate the user's waiting time and system revenue so as to maximize system integrated benefit from the perspective of both users and system managers.

According to the large-scale real numerical instance study, DPRH is superior to FBFS, RHN, and RHB in balancing allocation effect and allocation speed: Though FBFS is unparalleled in allocation speed, DPRH achieves better allocation effect, i.e., the total integrated benefit is raised by 3.83%, resource utilization is raised by 2.7%, user satisfaction is raised by 1.53% with acceptable allocation speed. DPRH reaches nearly the same allocation speed as RHN but leads to a better allocation effect. DPRH achieves nearly the same allocation effect as RHB but improves allocation speed.

Future research will focus on the following aspects. Firstly, the allocation approach can be improved to achieve nearly the same allocation speed as FBFS. Secondly, the allocation approach and allocation model can be improved to deal with the unpunctuality of users and owners in a realistic smart parking environment. Thirdly, dynamic pricing can be considered in the allocation models to further improve the allocation effects.

Data Availability

The parking data used to support the findings of this study are currently under embargo, while the research findings are commercialized. Requests for data 12 months after publication of this article will be considered by the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Beijing Natural Science Foundation (J210001), which support is gratefully acknowledged. Besides, the authors are very grateful to the referees for their valuable comments and suggestions, which helped to improve this study.

References

[1] X. Chen, S. Wu, and C. Shi, "Sensing data supported traffic flow prediction via denoising schemes and ann: a comparison," *IEEE Sensors Journal*, vol. 20, no. 23, pp. 14317–14328, 2020.

[2] K. Raichura and N. Padhariya, "Edpas: event-based dynamic parking allocation system in vehicular networks," in *eds of the 2014 IEEE 15th International Conference on Mobile Data Management*, pp. 79–84, Brisbane, Australia, July 2014.

[3] B. Nugraha and R. Tanamas, "Off-street parking space allocation and reservation system using event-driven algorithm," in *Proceedings of the 6th International Conference on Electrical Engineering and Informatics (ICEEI) - Sustainable Society*

through Digital Innovation, Langkawi, Malaysia, November 2017.

[4] V. Hassija, V. Saxena, V. Chamola, and R. Yu, "A parking slot allocation framework based on virtual voting and adaptive pricing algorithm," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 5945–5957, 2020.

[5] B. Zou, N. Kafle, O. Wolfson, and J. Lin, "A mechanism design based approach to solving parking slot assignment in the information era," *Transportation Research Part B: Methodological*, vol. 81, pp. 631–653, 2015.

[6] C. Lei and Y. Ouyang, "Dynamic pricing and reservation for intelligent urban parking management," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 226–244, 2017.

[7] H. He, Z. Zhang, and P. Yan, "A real-time reservation service for smart parking system," in *Proceedings of the 15th International Conference on Service Systems and Service Management.(ICSSSM), Hangzhou, China*, July 2018.

[8] P. Yan, X. Cai, D. Ni, F. Chu, and H. He, "Two-stage matching-and-scheduling algorithm for real-time private parking-sharing programs," *Computers & Operations Research*, vol. 125, Article ID 105083, 2021.

[9] Y. Geng and G. Cassandras, "Dynamic resource allocation in urban settings: a "smart parking" approach," in *Proceedings of the 2011 IEEE International Symposium on Computer-Aided Control System Design (CACSD)*, pp. 1–6, Institute of Electrical and Electronics Engineers Inc, Denver, CO, USA, September 2011.

[10] Y. Geng and G. Cassandras, "New "smart parking" system based on resource allocation and reservations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1129–1139, 2013.

[11] O. Kotb, C. Shen, X. Zhu, and Y. Huang, "iParker-A new smart car-parking system based on dynamic resource allocation and pricing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 9, pp. 2637–2647, 2016.

[12] M. Mladenović, T. Delot, G. Laporte, and C. Wilbaut, "A scalable dynamic parking allocation framework," *Computers & Operations Research*, vol. 125, Article ID 105080, 2021.

[13] C. Zhao, M. Chen, X. Li, and Y. Du, "Urban parking system based on dynamic resource allocation in an era of connected and automated vehicles," in *Proceedings of the IEEE Intelligent Transportation Systems Conference (IEEE-ITSC)*, pp. 3094–3099, Auckland, New Zealand, 2019.

[14] X. Xu, M. Cheng, R. Kong, H. Yang, and Q. Huang, "Private parking slot sharing," *Transportation Research Part B: Methodological*, vol. 93, pp. 596–617, 2016.

[15] B. Yang, Z. Yuan, Y. Yang, R. Yin, and Y. Wang, "The study on allocation model of shared parking slots in multi-parking lots," in *Proceedings of the 5th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering (ICMMCCE)*, vol. 141, pp. 457–465, Chongqing, China, January 2017.

[16] Y. Han, Q. Zhang, and W. Huang, "Optimal parking space allocation model based on the shared parking in residential areas," in *Proceedings of the 19th COTA International Conference of Transportation Professionals (CICTP) - Transportation in China 2025*, pp. 2588–2599, Nanjing, China, July 2019.

[17] S. Ning, Z. Yuan, Z. Han, and Y. Yang, "A novel reservation-based allocation mechanism of private parking slots sharing," in *Proceedings of the International Conference on Transportation and Development 2020: Traffic and Bike/Pedestrian Operations*, pp. 227–238, American Society of Civil Engineers, Seattle, WA, USA, May 2020.

[18] B. Jiang and P. Fan, "Optimal allocation of shared parking slots considering parking unpunctuality under a platform-based management approach," *Transportation Research Part E: Logistics and Transportation Review*, vol. 142, Article ID 102062, 2020.

[19] J. Arellano-Verdejo and E. Alba, "Optimal allocation of public parking slots using evolutionary algorithms," in *Proceedings of the 2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, Ostrawva, Czech Republic, September 2016.

[20] C. Li, Y. Tao, and S. Liu, "A shared parking space optimization model to alleviate China's parking problem considering traveler's tiered credit risk," *Transportation Letters*, vol. 13, no. 1, pp. 45–52, 2019.

[21] C. Shao, H. Yang, Y. Zhang, and J. Ke, "A simple reservation and allocation model of shared parking lots," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 303–312, 2016.

[22] P. Jiang, R. Shao, C. Song, "Matching model between private idle parking slots and demanders for parking slot sharing," *Journal of Transportation Engineering, Part A: Systems*, vol. 147, no. 6, Article ID 04021028, 2021.

[23] Y. Yang, K. He, Y. Wang, Z. Yuan, Y. Yin, and M. Guo, "Identification of dynamic traffic crash risk for cross-area freeways based on statistical and machine learning methods," *Physica A: Statistical Mechanics and Its Applications*, vol. 595, Article ID 127083, 2022.

[24] A. Bischi, L. Taccari, and E. Martelli, "A rolling-horizon optimization algorithm for the long term operational scheduling of cogeneration systems," *Energy*, vol. 184, pp. 73–90, 2019.

[25] Z. Yao, H. Jiang, Y. Cheng, Y. Jiang, and B. Ran, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 1841–1851, 2022.

[26] Y. Yang, Z. Yuan, J. Chen, and M. Guo, "Assessment of osculating value method based on entropy weight to transportation energy conservation and emission reduction," *Environmental Engineering and Management Journal*, vol. 16, no. 10, pp. 2413–2423, 2017.

[27] Y. Yang, Z. Yuan, and R. Meng, "Exploring traffic crash occurrence mechanism towards cross-area freeways via an improved data mining approach," *Journal of Transportation Engineering Part A Systems*, vol. 148, 2022.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 6184790, 15 pages
https://doi.org/10.1155/2022/6184790

Research Article

Spatially Formulated Connected Automated Vehicle Trajectory Optimization with Infrastructure Assistance

Ran Yi,[1] **Yang Zhou**,[1] **Xin Wang**,[2] **Zhiyuan Liu**,[3] **Xiaotian Li**,[1] **and Bin Ran**[1]

[1]*Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, USA*
[2]*Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA*
[3]*Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing, China*

Correspondence should be addressed to Yang Zhou; zhou295@wisc.edu

Received 5 October 2021; Accepted 5 April 2022; Published 20 May 2022

Academic Editor: Alessandro Severino

This paper presents a constrained connected automated vehicles (CAVs) trajectory optimization method on curved roads with infrastructure assistance. Specifically, this paper systematically formulates trajectory optimization problems in a spatial domain and a curvilinear coordinate. As an alternative of temporal domain and Cartesian coordinate formulation, our formulation provides the constrained trajectory optimization flexibility to describe complex road geometries, traffic regulations, and road obstacles, which are usually spatially varying rather than temporal varying, with assistances vehicle to infrastructure (V2I) communication. Based on the formulation, we first conducted a mathematical proof on the controllability of our system, to show that our system can be controlled in the spatial domain and curvilinear coordinate. Further, a multiobjective model predictive control (MPC) approach is designed to optimize the trajectories in a rolling horizon fashion and satisfy the collision avoidances, traffic regulations, and vehicle kinematics constraints simultaneously. To verify the control efficiency of our method, multi-scenario numerical simulations are conducted. Suggested by the results, our proposed method can provide smooth vehicular trajectories, avoid road obstacles, and simultaneously follow traffic regulations in different scenarios. Moreover, our method is robust to the spatial change of road geometries and other potential disturbances by the road curvature, work zone, and speed limit change.

1. Introduction

The CAVs equipped with advanced sensing and communication technologies enable their self-driving tasks and connection with other vehicles as well as infrastructures. Under such a connected and automated environment, CAVs can obtain precise ambient information via sensing and communication to make decisions and control more efficiently and safely than human-driven vehicles. Due to the potentials, CAV technologies have experienced a fast development in very recent years [1–5]. Among these technologies, trajectory planning serves as a critical component to plan the vehicles' future movement to avoid hazards and make the CAVs operate in a safe, comfortable, and efficient manner. Rather than treating the highway as a straight line, there are many curved roads in the real world such as mountain area that requires ego-CAV trajectory planning algorithm to render both longitudinal and lateral movements to make CAVs pass safely.

Due to the importance, two-dimensional CAVs trajectory planning algorithms have been widely researched in recent years [6–14]. The state of art of CAV trajectory planning algorithms can be generally divided into four different approaches: (i) the graph search-based approach [14, 15]; (ii) the interpolating curve planner [16–18]; (iii) the sampling-based approach [19, 20]; and (iv) the function optimization approach [21, 22]. The graph search-based approach divides the feasible vehicle travel region into multiple grids, and by that, finding the optimal trajectory can be equivalently treated as finding the shortest path among

nodes of feasible grids. By the shortest path formulation, this type of approach (e.g., [23, 24]) usually applied dynamic programming and A-star algorithm. By the nature of the shortest path problem, this type of approach needs to tradeoff between computation and precision. When the size of lattice is small and number of lattices are big, this type of algorithm can be slow computing. Whereas when the size of the grid is big, the trajectory can be discontinuous and nonsmooth. On the other hand, the interpolating curve planner, which uses predefined curve functions to generate reference points from the available space [18, 25, 26], is comparatively computational less expensive. The above-mentioned approaches largely ignore vehicle kinematics by assuming the physical positions are attainable regardless of vehicle acceleration and angular speed constraints that may induce implementation infeasibility is the real-world application. The sampling-based approach executes a random search over the vehicle feasible state space (e.g., position, speed, and acceleration) and finds an optimal sequence of state heuristically by comparing these randomly sampled state spaces according to a predefined objective function (e.g., minimize travel delay and intensive acceleration or brake). Though effective, these approaches usually render suboptimal solutions suggested by [27] due to the searching heuristics nature. Further, if the state dimension is large, the computation complexity will also increase significantly. As an opposite, the function optimization approach formulates the trajectory planning of CAV as a constrained optimization problem that is flexible in handling constraints from the environment [8, 13, 28, 29], which can minimize a multiobjective cost function meanwhile satisfying the vehicle dynamics and hazard avoidances constraints. Further, the constrained optimization approach is usually implemented in a rolling horizon fashion to be against unexpected disturbances, and it is also known as MPC in the control theory field, which is widely applied in CAV car following control [30–32].

Though promising, the constrained optimization-based CAVs trajectory planning still faces some challenges. One challenge is to describe the road geometries in the formulation explicitly. Differed in the utilized coordinates, they can be further categorized as (i) CAV trajectory constrained optimization on a Cartesian coordinate [8, 13, 28, 29] and (ii) CAV trajectory constrained optimization on a curvilinear coordinate [9, 10, 33, 34]. As a most widely applied coordinate, the Cartesian coordinate is more suited to describe open spaces. However, road geometry is usually composed of complex and composite curves, which means using Cartesian coordinates will be technically challenging to formulate the road boundaries as constraints. On the other hand, the curvilinear coordinate system is born to describe geometries formed by curved lines [35]. Especially for the road geometry description, one axle of the curvilinear coordinate can be the lane centerline, while the other axle can be perpendicular to the centerline tangent. By that, the road geometry and boundaries can be simply formulated, reducing a great formulation and computation complexity.

Nevertheless, most of the vehicle trajectory optimization in the literature is vastly built by a temporal formulation [6–11], assuming that the road geometry characteristics (e.g., curvature, lane tangent directions) remain unchanged over time, that has not fully exploited the potentials of the curvilinear coordinate. The road geometries characteristics, traffic regulations (e.g., speed limit), and hazards are spatially varying rather than temporally varying, which serve as an unobservable exogenous disturbance impacting the performances that further exert a negative impact on the desired planning objectives. Though the real-time information of road attributes can be conveyed to CAVs via V2I, with the increasing maturity of the vehicle to infrastructure (V2I) communication, explicitly incorporating these spatially varying characteristics in the temporal domain is still forbiddingly hard. However, the very recent works by [36, 37], which utilize the spatial formulation for one-dimensional CAV car following control and curvilinear coordinate based human driven vehicles modeling in an intersection, respectively, provide a new angle to formulate the trajectory planning problems in a spatial domain. Inspired by their works, we found that the spatial domain formulation can explicitly incorporate the optimal formulation of spatially varying attributes on a curvilinear coordinate. Hence, an infrastructure assisted spatially constrained optimal ego-CAV trajectory planning algorithm based on a curvilinear coordinate is provided in this paper, which can deal with any types of curved road safely and efficiently. Especially, we contribute to design a spatially formulated constrained CAV trajectory optimization in a curvilinear coordinate that is capable of: (i) greatly simplifying the formulation compared with ones formulated in the time domain and Cartesian coordinates; (ii) quickly converging to the desired vehicle operation state (e.g., target speed, following lane centerline), and robust to the spatially disturbances (e.g., lane curvature change) by the infrastructure communication and assistance and (iii) providing a multi-objective optimization (e.g., smoothness of vehicle control, less deviation from desired state) framework which explicitly incorporates the spatial characteristics (e.g., lane curvature) and constraints (e.g., speed limit, obstacles, and lane width) while avoiding obstacles.

The organization of this paper is as follows. Section 2 provides detailed problem descriptions and gives assumptions of our paper. Section 3 describes the system by a state space and presents the design of MPC for the system. Section 4 presents the designed simulation experiments to verify the effect of the proposed model. The conclusion of the paper and future work is summarized and discussed in Section 5.

2. Problem Description

We provide an illustrative example to motivate our work on trajectory planning first. For our modeling, we focus on two-dimensional ego-CAV trajectory planning on a highway without any intersections. Specifically, our planning target is to ensure ego-CAV travel as fast as possible and meanwhile keep to the targeted road center line. When a CAV drives on a complex traffic scenario such as on a winding lane with or without a potential obstacle, as shown in Figure 1, an infinite number of paths are possible for a CAV, but limited time is

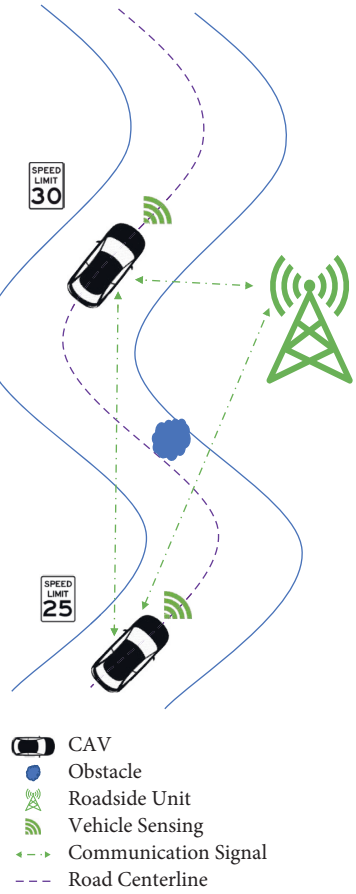


Figure 1: Problem description.

given for the decision. It is assumed that the CAV can collect vehicle dynamics from its own sensors. However, road attributes (road geometry, traffic control devices) are obtained from communicating with roadside units (RSUs) [38], by which CAVs know exact traffic environment information in advance. We provide detailed geometries via V2I communication because road geometries are relatively static and can be easily measured by infrastructure. If we rely on vehicular sensing, we will be unable to obtain and utilize road geometries that are not within the sensing range. Additionally, when a CAV is traveling at a high speed, the road geometry information obtained via vehicular sensing may be noisy or inaccurate. The purpose of the trajectory planning algorithm is to find the optimal path from an infinite number of possible paths based on the CAV's location in a very short time to ensure safety, control efficiency, comfort requirements, and traffic regulation obeyance. Before the detailed modeling, we made the following assumptions for the involved system.

Assumption 1: the vehicle is treated as a point.

Assumption 2: all actuation and communication delays are negligible.

Assumption 3: the vehicle is above automation level 3 according to Society of Automotive Engineering standard [39].

Assumption 4: all vehicle dynamics effects such as suspension movement, road inclination, and aerodynamic forces are negligible.

Assumption 5: the road geometry characteristics, spatially traffic regulations and control (e.g., speed limit sign), and obstacle position information can be measured by RSU and communicate to CAVs.

3. Methodology

3.1. State Space Formulation. In this section, we describe the system by a spatial domain state space in a curvilinear coordinate and prove its controllability. One axle of a curvilinear coordinate for a path is usually defined as its centerline, while the other axle can be perpendicular to the lane tangent. By that, road geometry and its boundaries can be simply formulated.

Let C represents the centerline of the lane as shown in Figure 2. The lane centerline, with respect to the global frame $\{G\}$ that is represented by the $[x_{\text{des}}(s), y_{\text{des}}(s), \theta_{\text{des}}(s)]^T \in R^3$, where its position is given by $[x_{\text{des}}(s), y_{\text{des}}(s)]^T$ and its orientation in the global frame is represented by $\theta_{\text{des}}(s)$. Let s be the mapped length of the curvature along the lane centerline that vehicle has traveled. Based on that, we define $\{V\}$ represents the vehicle body-fixed reference frame with its pace $(p_v(s))$, the reciprocal of the vehicular velocity, represents the vehicle's forward speed in space domain, and its orientation in the global frame is represented by $\theta_v(s)$.

To describe the vehicle dynamics over the established curvilinear coordinate, we construct a state space system and define the system state $X(s)$ for each space s as

$$X(s) = \begin{bmatrix} r(s) \\ \psi(s) \\ p(s) \end{bmatrix}, \tag{1}$$

where $r(s)$ is the lateral deviation, which equals to signed orthogonal distance from the CAV to the closest point on the lane centerline C, in meters; $\psi(s) = \theta_v(s) - \theta_{\text{des}}(s)$, where $\psi(s)$ is the angular deviation, θ_v is the angle between the CAV heading and the x-axis in the global frame $\{G\}$ and $\theta_{\text{des}}(s)$ is the angle between the tangent of the lane centerline and the x-axis in the global frame $\{G\}$, in radians; $p(s) = p_v(s) - p_{\text{des}}(s)$, where $p(s)$ is the pace deviation, $p_{\text{des}}(s)$ is the reciprocal of the road speed limit ($p_{\text{des}}(s)$ is a nondifferentiable function of s, which can be treated as a constant), in second per meter. Note that, $\theta_{\text{des}}(s)$ and $p_{\text{des}}(s)$ are spatially varying and got in a real-time manner by the infrastructure. By V2I, $\theta_{\text{des}}(s)$ and $p_{\text{des}}(s)$ can be directly obtained by vehicle and used for modeling. Specifically, the

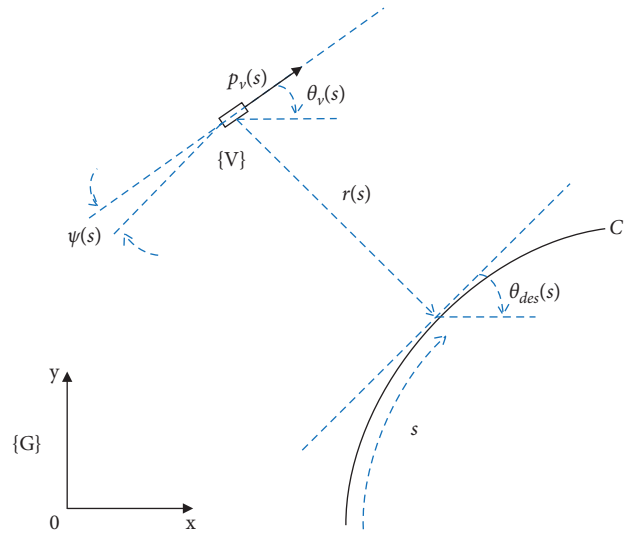


FIGURE 2: Vehicle modelling on the defined curvilinear coordinate.

vehicle dynamics are modeled as a nonlinear state space system as

$$\begin{aligned}\frac{dX(s)}{ds} &= \frac{d}{ds}\begin{bmatrix} r(s) \\ \psi(s) \\ p(s) \end{bmatrix} \\ &= \begin{bmatrix} \sin(\psi(s)) \\ k(s) \\ \alpha(s) \end{bmatrix} \\ &= f[X(s), U(s)],\end{aligned} \tag{2}$$

where $k(s) = k_v(s) - k_{\text{des}}(s)$, which is the relative angular spatial change rate that controlled by the steering wheel. $k_v(s)$ is the curvature of the vehicle trajectory; $k_{\text{des}}(s)$ is the curvature of the lane centerline, in radian per meter; and $\alpha(s)$ is relative moderation that indicates acceleration of the CAV, which can be treated as the control of the brake or throttle pedal (negative value indicates accelerating), in second per meter squared. $k(s)$ and $\alpha(s)$ compose our "road characteristics compensated" control input $U(s)$, given as $U(s) = [k(s), \alpha(s)]^T$.

However, the road speed limit can be changed based on geometric road design in a real-world situation, which means $p_{\text{des}}(s)$ cannot be treated as a spatially invariant constant when speed limit change occurs. Moreover, the speed limit change for the road is a unit jump, and $p_{\text{des}}(s)$ is still a nondifferentiable function of s. To solve this problem, we introduce a "buffer zone" in front of the speed limit changing area where a changed speed limit is posted at the same length as our MPC model's prediction horizon. The "buffer zone" section divides speed limit difference by the "buffer zone" length to smoothly form a discrete velocity transit. With Δs is sufficiently small, the velocity change in the "buffer zone" is close enough to be treated as a continuous system. Thus, $p_{\text{des}}(s)$ becomes a differentiable function of s inside the "buffer zone," and $\alpha_{\text{des}}(s)$ is the parameter indicating acceleration of the road is the derivative of $p_{\text{des}}(s)$, in second per meter squared. The new nonlinear function with the "buffer zone" introduced is $(dX(s)/ds) = (d/ds)\begin{bmatrix} r(s) \\ \psi(s) \\ p(s) \end{bmatrix} = \begin{bmatrix} \sin(\psi(s)) \\ k(s) \\ \alpha(s) \end{bmatrix}$, where $\alpha(s) = \alpha_v(s) - \alpha_{\text{des}}(s)$, where $\alpha_v(s)$ is the parameter indicating acceleration of the CAV.

Table 1 shows brief equations and their associated units of measurement

Given the state space defined above, we first analyzed its controllability that describes the ability of any control variables that can move the state of a system from any initial state to any final state in a finite time interval. Hence, we applied the following definitions and theorems to approve the controllability.

Theorem 1. *Nonlinear system small-time locally controllability [40]: the linear test: if a nonlinear $\dot{y} = f(y, u)$, whose linearized control system over an equilibrium point (y_e, u_e): $\dot{y} = Ay + Bu$ is controllable, then it is small-time locally controllable at (y_e, u_e).*

Definition 1. Equilibrium point: for a nonlinear differential equation $(dX(s)/ds) = f[X(s), U(s)]$, where f is a function mapping $R^n \times R^m \longrightarrow R^n$. A point $X_e \in R^n$ is called an equilibrium point if there is a specific $U_e \in R^m$ such that $f(X_e, U_e) = 0_n$.

By the Definition 1, it is trivial to find that the equilibrium point of the system as equation (2) occurs if and only if $X_e = [0, 0, 0]^T$, $U_e = [0, 0]^T$

Theorem 2. *Linear time-invariant system controllability [40]: For a linear time-invariant system with the form $\dot{x} = Ax + Bu$, the controllability matrix can be written as*

Table 1: Table of variables.

	Variables	Equation	Unit
Lateral deviation	$r(s)$	(N/A)	m
Angular deviation	$\psi(s)$	$\theta_v(s) - \theta_{\text{des}}(s)$	rad
Pace deviation	$p(s)$	$p_v(s) - p_{\text{des}}(s)$	(s/m)
Relative angular spatial change rate	$k(s)$	$k_v(s) - k_{\text{des}}(s)$	(rad/m)
Relative moderation	$\alpha(s)$	$\alpha_v(s) - \alpha_{\text{des}}(s)$	(s/m^2)

$$G = G(A, B) = \underbrace{\left[B, AB, A^2B, \ldots, A^{n-1}B\right]}_{n\times(mn)\text{ matrix}}. \tag{3}$$

If $\text{rank}(G) = n$, this linear system is controllable.

Based on the Definition 1, Theorems 1 and 2, we can have the following proposition:

Proposition 1. *The state space formulated as equation (2) is small-time locally controllable at the equilibrium point $X_e = [0, 0, 0]^T$, $U_e = [0, 0]^T$.*

Proof. With Taylor series, equation (2) at the equilibrium point (X_e, U_e) can be rewritten as

$$\frac{dX(s)}{ds} = \sum_{i=1}^{n}\left[\frac{\left(d^i f(X(s), U(s))_{X(s)=X_e}/d^i X(s)\right)(X(s) - X_e)^n}{n!} + \frac{\left(d^i f(X(s), U(s))_{U(s)=U_e}/d^i U(s)\right)(U(s) - U_e)^n}{n!}\right]. \tag{4}$$

According to the Taylor series, a representation of a function as an infinite sum of terms that are calculated from the values of the function's derivatives at a single point. In our case, only 1st order derivative needs to be considered for the linearized control system. Moreover, the equilibrium in our system occurs when the CAV drives along the lane centerline, which means that there is no deviation between the CAV and the lane centerline. The 1st order derivative of equation (4) at the equilibrium point (X_e, U_e) can be written as

$$\frac{dX(s)}{ds} = \begin{bmatrix} 0 & \cos(\psi) & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{\psi=\psi_e}(X(s) - X_e) + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}(U(s) - U_e), \tag{5}$$

and equation (5) can be simplified as

$$\frac{dX(s)}{ds} = AX(s) + BU(s), \tag{6}$$

where $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Based on equation (3), the controllability matrix of equation (6) can be written as

$$G(A, B) = \left[B, AB, A^2B\right] = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{7}$$

which gives us the rank of the controllability matrix:

$$\text{rank}\, G(A, B) = 3. \tag{8}$$

□

Remark 1. The above linearization process can also be proved by the small-angle approximation. Specifically, when the angle is relevant small and $\cos(\psi) \approx 1$, the small-angle approximation can be applied. The same state-space model, as equation (6), can be derived based on the small-angle approximation from equation (2).

To reflect control frequency in the real world, we applied the zero[th]-order hold (ZOH) for control input for discretization. Specifically, the control input is assumed to be a constant during each update spatial interval Δs, and when Δs is sufficiently small, the discretization process can be treated as a continuous system [41]. The discrete version of equation (5), according to equation (6), is shown as

$$X_{(m+1)\Delta s} \approx A_d X_{m\Delta s} + B_d U_{m\Delta s}, \forall m > 0, \tag{9a}$$

$$\text{s.t. } A_d = e^{A\Delta s}, \tag{9b}$$

$$B_d = \int_0^{\Delta s} e^{A\gamma} d\gamma B. \tag{9c}$$

For notation brevity, we use X_m to represent $X_{m\Delta s}$ for the rest of the paper.

3.2. MPC Formulation. In this section, we provide an MPC formulation due to its great capability to systematically deal with system state and control constraints and meanwhile handling multiobjectives. Further, it is robust to be against system disturbances due to its rolling horizon implementation. Based on the discretized control Equation (9a), a linear MPC formulation is formulated in our study. In our MPC framework as illustrated by Figure 3, at each current space step m, we solved a constrained trajectory optimization problem over a fixed finite prediction horizon with spatial length m_p to calculate the optimal control input and state sequences within the horizon. The controller only implements the first step optimal control input at space step m, and the algorithm continues this process repetitively until the end of algorithm, as shown in Figure 3.

To better illustrate the algorithm, we use $\mathcal{U}_m^p = [U_m^{p,m}, U_{m+1}^{p,m}, \ldots, U_{m+m_p-1}^{p,m}]$ to denote the optimal control vectors for CAV obtained at space m for the prediction horizon, m to $m+m_p$; $\mathcal{X}_m^p = [X_m^{p,m}, X_{m+1}^{p,m}, \ldots, X_{m+m_p}^{p,m}]$ to denote the predicted future states for CAV obtained at space m for the prediction space horizon, and m to $m+m_p$; $\mathcal{X}_m^r = [X_0^r, X_1^r, \ldots, X_m^r]$ to denote the realized states for CAV by space m, which can be seen as the optimal solution where X_0^r is the initial state when the control introduces.

By carefully considering the control efficiency and driving smoothness, an optimal control strategy can be obtained by solving the optimal control problem given as

$$(\mathcal{X}_m^{p*}, \mathcal{U}_m^{p*}) = \arg\min F\left(X_{m+m_p}^{p,m}\right) + \sum_{i=m}^{m+m_p-1} L\left(X_i^{p,m}, U_i^{p,m}\right), \tag{10a}$$

$$\text{s.t. } X_{i+1}^{p,m} = A_d X_i^{p,m} + B_d U_i^{p,m}, \quad \forall i \in \{m, m+1, m+2, \ldots, m+m_p-1\}, \tag{10b}$$

$$X_m^{p,m} = X_m^r, \tag{10c}$$

$$X_i^{p,m} \in \mathbb{X}_i \forall i \in \{m, m+1, m+2, \ldots, m+m_p\}, \tag{10d}$$

$$U_i^{p,m} \in \mathbb{U}_i \forall i \in \{m, m+1, m+2, \ldots, m+m_p\}, \tag{10e}$$

where m_p is the prediction space horizon; $L(X_i^{p,m}, U_i^{p,m})$ is the running cost consists of the CAV penalties on the deviation from equilibrium point and driving discomfort; and $F(X_{m+m_p}^{p,m})$ is the terminal cost that refers to the final stage of the prediction horizon. Equation (10c) is the initial state for the prediction horizon at space step m; and equation (10d) is the state constraint used to guarantee the vehicle states at each space point. Equation (10e) is the control constraint used to regulate that the steering wheel rotation and acceleration/deceleration are within a reasonable range of $U(s)$.

For the objective function formulated in equation (10a), the running cost is specified as follows:

$$L(X_{m+n}^{p,m}, U_{m+n}^{p,m}) = (X_{m+n}^{p,m})^T P X_{m+n}^{p,m} + (U_{m+n}^{p,m})^T Q U_{m+n}^{p,m} + 2q_1 k_{v,m+n} k_{\text{des},m+n} - q_1 k_{\text{des},m+n}^2, \tag{11a}$$

where P and Q are diagonal positive definite matrices, usually designed as the diagonal matrix below:

$$P = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}, p_1, p_2, p_3 > 0, \tag{11b}$$

$$Q = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}, q_1, q_2 > 0. \tag{11c}$$

Especially, if we want to regulate that the vehicles stick to the lane centerline for safety concern, we can predefine p_1 and p_2 with large values, while on the other hand, if we are willing to provide the trajectories with flexibility, we can set $p_{1,2} \longrightarrow 0$.

We further specify the terminal cost as:

$$F\left(X_{m+m_p}^{p,m}\right) = \left(X_{m+m_p}^{p,m}\right)^T S X_{m+m_p}^{p,m}, \tag{12}$$

where $S = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$, $s_1, s_2, s_3 > 0$.

In equations (10d) and (10e), by considering vehicle's physical limits, CAV's state constraint determined at space m is formulated as

$$r_{m+n}^L \le C X_{m+n}^{p,m} \le r_{m+n}^U \forall n \in \{0, 1, 2, \ldots, m_p\}, \tag{13a}$$

$$\psi_{\min} \le D X_{m+n}^{p,m} \le \psi_{\max} \forall n \in \{0, 1, 2, \ldots, m_p\}, \tag{13b}$$

$$0 \le T X_{m+n}^{p,m} \le +\infty \forall n \in \{0, 1, 2, \ldots, m_p\}, \tag{13c}$$

$$-\frac{1}{R_{\min}} - k_{\text{des},m+n} \le E U_{m+n}^{p,m} \le \frac{1}{R_{\min}} - k_{\text{des},m+n} \quad \forall n \in \{0, 1, 2, \ldots, m_p - 1\}, \tag{13d}$$

$$\alpha_{\min} \le I U_{m+n}^{p,m} \le \alpha_{\max} \forall n \in \{0, 1, 2, \ldots, m_p - 1\}. \tag{13e}$$

Equation (13a) is the constraint to make sure that the CAV drives within the drivable lane, where $C = [1, 0, 0]$, and $r_{m+n}^L = \begin{cases} r_{m+n}^-, \text{if} O_b \le m+n \le O_e \\ -r_{\max}, \text{Otherwise} \end{cases}$ and $r_{m+n}^U = \begin{cases} r_{m+n}^+, \text{if} O_b \le m+n \le O_e \\ r_{\max}, \text{Otherwise} \end{cases}$. Where r_{m+n}^- and r_{m+n}^+ are lower bound and upper bound that CAV can pass through within the obstacle zone, respectively; O_b and O_e are spatial positions where obstacle start and end, which can be obtained from infrastructure through V2I in advance. $r_{\max}$ is the half of the lane width; equation (13b) is the constraint that the physical limits of allowable angular deviation for CAV to make sure that the CAV's driving direction will not deviate much from the lane centerline direction, where $D = [0, 1, 0]$, $\psi_{\min}$ and $\psi_{\max}$ indicates the lower and upper

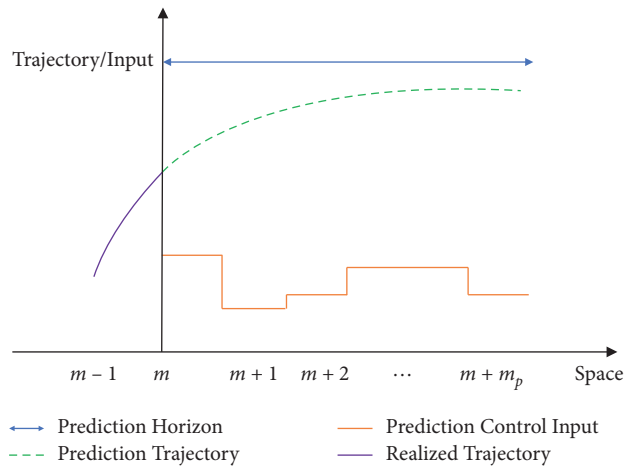


Figure 3: Illustration of model predictive control.

bound of allowable angular deviation; equation (13c) is the constraint to ensure that the CAV does not exceed the speed limit, where T = [0, 0, 1], and this constraint also makes sure that the CAV cannot drive backward; equation (13d) is the constraint that the realized moving path of CAV is bounded by a given minimum turning radius, where $E = [1, 0]$, $R_{\min}$ is the minimum turning radius of the CAV; equation (13e) is the constraint based on the physical limits of vehicle's acceleration/deceleration rate a. In this constraint, $I = [0, 1]$ and $\alpha_{\min} = -a_{\max} p_s^3$, where $a_{\max}$ is the CAV's maximum acceleration limit. Similarly, $\alpha_{\max,s} = -a_{\min} p_s^3$, where $a_{\min}$ is the CAV's maximum deceleration limits.

To be noted that the above formulation can be expanded in the scenario with speed limit change or stop sign by introducing the concept of "buffer zone," and correspondingly, equation (13e) changes to $\alpha_{\min} - \alpha_{\text{des}} \le IU_{m+n}^{p,m} \le \alpha_{\max} - \alpha_{\text{des}} \forall n \in \{0, 1, 2, \ldots, m_p - 1\}$. The information $k_{\text{des},m+n}$, $\theta_{\text{des},m+n}$ and $p_{\text{des},m+n}$ can be hardly measured by vehicle due to the limit of sensing range, whereas this static information can be readily measured and communication by the infrastructure. As can be found that, differed from the method merely based on vehicular sensors, the infrastructure-assisted approach can provide necessary information to further improve control. Furthermore, the information that is transmitted by vehicle-to-vehicle (V2V) communication is extremely dependent on traffic volume and the distance between the controlled vehicle and its leading vehicle. If the volume of traffic is low, there may be no other vehicles to provide geometric data. Although the volume of traffic is not necessarily low, V2V communication cannot provide information if the distance between vehicles is large, as the vehicle will not save much geometric road information as it passes.

4. Experiment and Results

To test our optimal control model, we create a one-lane road segment with a series of curves as the numerical simulation environment shown in Figure 4. The road comprises three continuous curves with a total length of 1600 m and 3.6 m lane width, which means a 1.8 m width from lane centerline to left/right lane boundary. The first two curves have the same radius of 300 m, and the radius of the third curve is 200 m. The speed limit of the road is set as 54 (km/h) (15 (m/s)).

To validate our proposed method, simulation experiments of multiple traffic scenarios are performed on MATLAB since the field test is expensive and beyond the scope of this paper. The parameter setting for the CAV trajectory optimization as equations (10a) to (10e) is given in Table 2, according to [37, 42].

Five different simulation experiments of multiple traffic scenarios are summarized in Table 3.

4.1. Scenario 1: A Continuous Curvy Road Segment with the Constant Speed Limit. For the initial condition, we set the vehicle's lateral deviation from the lane centerline to be 1 m, the angular deviation to be $-(\pi/6)$ rad, and the pace deviation to be $(1/30)$ (s/m) (equivalently desired speed difference 5 (m/s) in our case). To better analyze the convergence behavior of the proposed algorithm, we plot the proposed controller performance of the first 50 m in Figure 5, which demonstrates the proposed trajectory optimization method's performance without the "buffer zone" introduced. To be noted that system states and control inputs maintain very closely to the system equilibrium point with nearly no oscillation from 50 m to 1600 m. Figures 5(a)–5(e) indicate the system state evolution, including lateral deviation, angular deviation, and pace deviation, and how the control inputs change in the space domain as the CAV moves by our control, respectively. As can be seen, similar trends are observed for different system states and the control input values. To be more specific, the lateral deviation exhibits some reasonable fluctuation initially: CAV moves 1.386 m laterally from lane centerline's left side to the right side in the first 6 m along the road, where the positive sign indicates that the CAV has been off the track and is at the left side of the lane centerline and the negative sign indicates that the CAV has off the track and is at the right side of the lane centerline. Then it quickly turns back from the right side to the lane centerline in the next 6 m and keeps driving on the lane centerline afterward. Angular deviation and relative angular spatial change rates converge in a nearly similar fashion as lateral deviation. As for the pace deviation and relative moderation converges to zero gradually. These results show that the algorithm quickly finds the difference between the vehicle states and the lane centerline and calculates the optimal control inputs to achieve the system equilibrium dynamically. Moreover, even though the $\theta_{\text{des}}(s)$ and $k_{\text{des}}(s)$ evolve spatially, the proposed controller shows great robustness to these disturbances mentioned above. Rather than the time-domain approach (e.g., [43]), the proposed method can better handle the space varying $\theta_{\text{des}}(s)$ and $k_{\text{des}}(s)$, which shows the superiority of proposed algorithm by incorporating the road geometric attributes via infrastructure to vehicle communication.

To gain further insight into the proposed trajectory optimization method, we convert $p_v(s)$, $\alpha(s)$, and $k_v(s)$ in

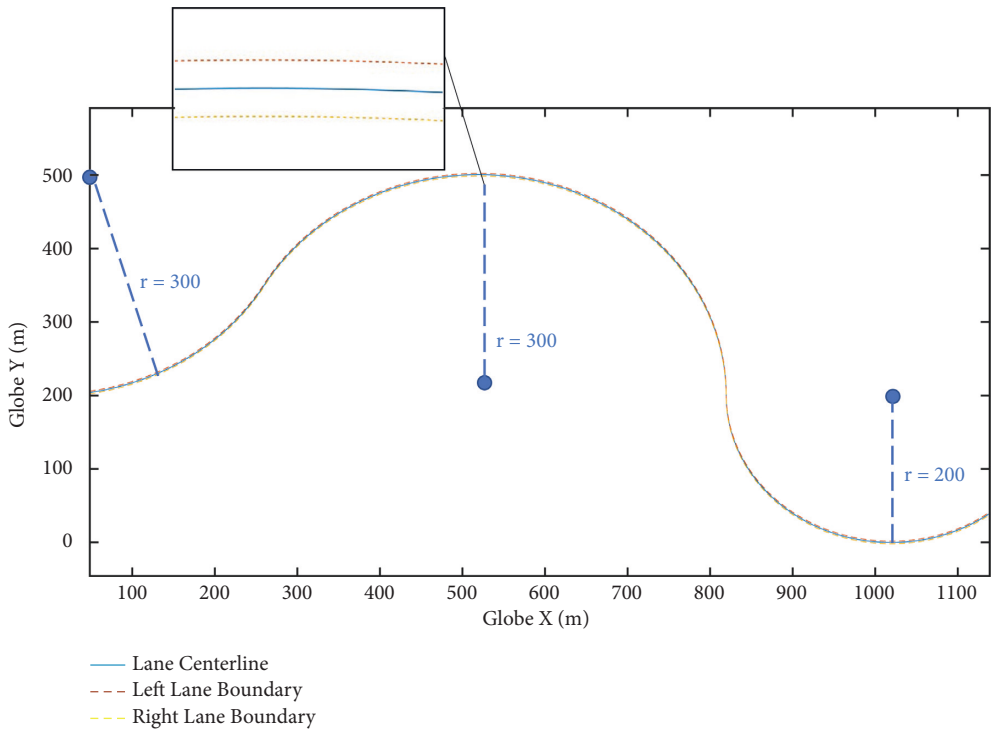


Figure 4: Illustration of road trajectory.

Table 2: Default parameters for model validation.

Parameters	Value
ψ_{min} and ψ_{max}	$-(\pi/6)$ rad and $(\pi/6)$ rad
$\mathbf{a}_{min}$ and $\mathbf{a}_{max}$	-5 (m/s^2) and 3 (m/s^2)
$\mathbf{r}_{max}$	1.8 m
$\mathbf{p}_1$, $\mathbf{p}_2$ and $\mathbf{p}_3$	0.33, 0.1 and 10
$\mathbf{q}_1$ and $\mathbf{q}_2$	1 and 500
$\mathbf{s}_1$, $\mathbf{s}_2$ and $\mathbf{s}_3$	$5p_1$, $5p_2$ and $5p_3$
$\Delta\mathbf{s}$	2 m
$\mathbf{m}_p$	80 m

Table 3: Scenarios settings.

Scenarios	Speed limit change	Driving behavior change	Obstacle	System disturbances
1	No	No	No	No
2	Yes	No	No	No
3	No	Yes	No	No
4	No	Yes	Yes	No
5	No	Yes	Yes	Yes

the spatial domain into speed, acceleration, and angular deviation rate in the time domain, as shown in Figure 6. The CAV starts with and keeps a high acceleration for the first 15 m and then it decreased gradually until it achieves equilibrium points, as shown in Figure 6(a). These acceleration changes lead to a speed increase in the first 50 m and make the speed converge to the desired speed, as shown in Figure 6(b). We can also find that the angular deviation rate is within a reasonable range initially and gradually converges to zero.

4.2. Scenario 2: A Continuous Curvy Road Segment with Speed Limit Change. The results in the previous sections show that the proposed trajectory optimization method is efficient and stable without speed limit change to be considered. However, in order to test how the proposed trajectory optimization method works in a real-world situation, we change the speed limit of the road segment from 1104 m to 1600 m to be 10 (m/s) and create a "buffer zone" of the same length as m_p. In other words, we now pay our attention to a new

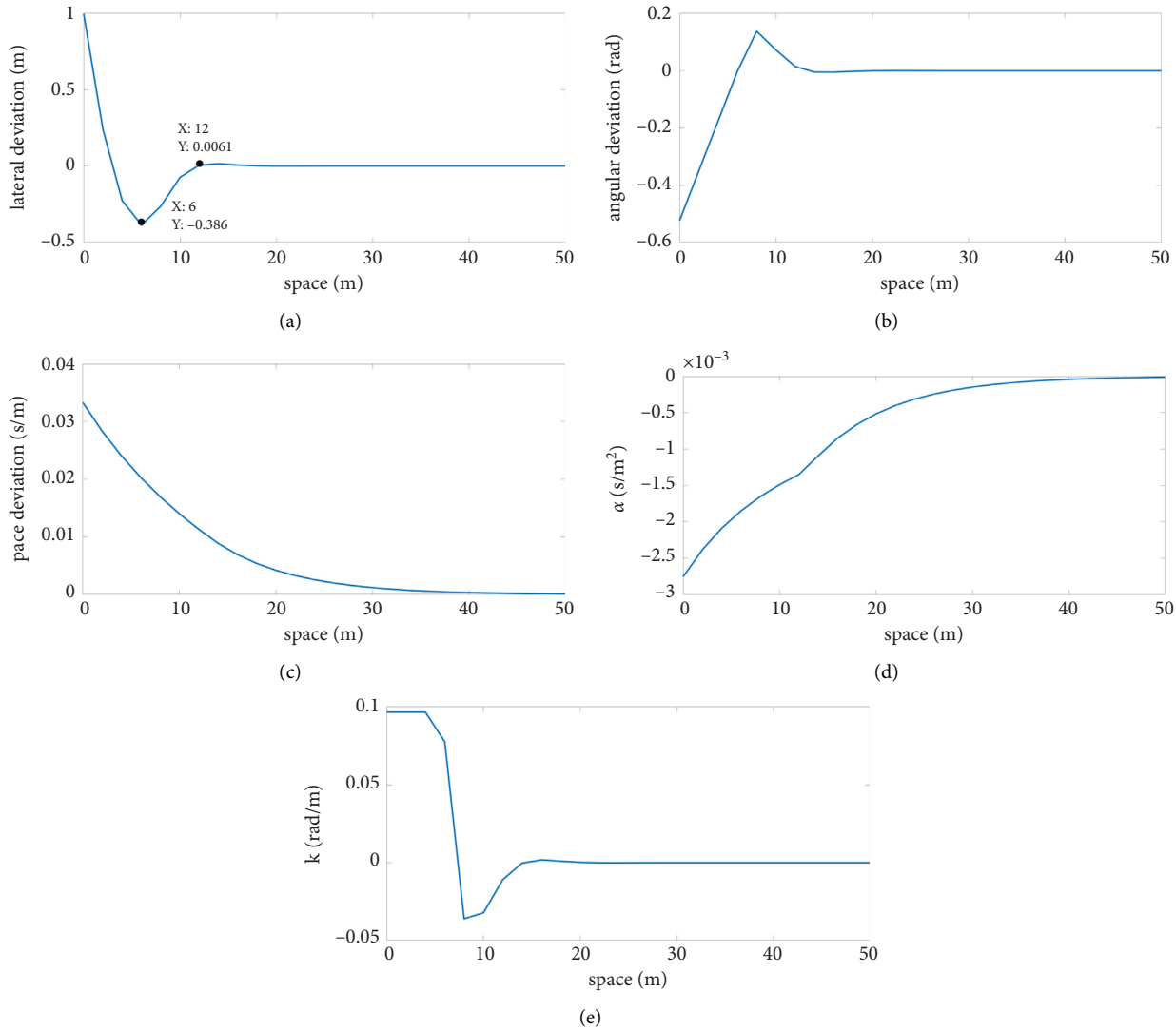


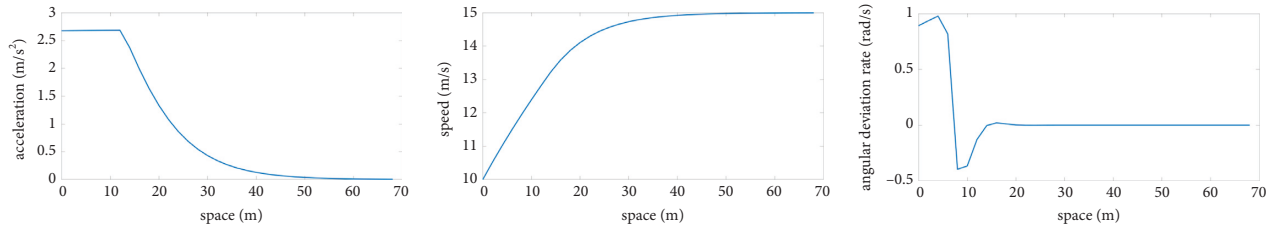
FIGURE 5: CAV state and control results of the first 50 meters: (a) lateral deviation $r(s)$; (b) angular deviation $\psi(s)$; (c) pace deviation $p(s)$; (d) relative moderation $\alpha(s)$; (e) relative angular spatial change rate $k(s)$.

situation where the road segment from 1024 m to 1104 m is replaced by the "buffer zone" with α_{des} equals $(1/1200)(s/m^2)$. The illustration for the road trajectory with a "buffer zone" is given in Figure 7.

The results in Figure 8 give us a generalized illustration of how the proposed trajectory optimization method performs with the "buffer zone." Although the "buffer zone" is introduced, the system state evolution and how the control input change in the space domain, as shown in Figures 8(a)–8(e) respectively, is stable and show great robustness to the real-world disturbance. The change of angular deviation rate that indicates the relative turning speed and direction of the CAV shown in Figure 8(h) shows a similar trend as that in Figure 6(c). Furthermore, as expected, CAV's speed and acceleration oscillate around the "buffer zone."

Figure 9 gives us a detailed illustration of CAV's speed and acceleration changes inside the "buffer zone." Figure 9(a) shows that the algorithm quickly detects the speed limit change in the "buffer zone" and makes the deceleration decision to achieve the system equilibrium dynamically, which leads to a smooth speed transition, as shown in Figure 9(b).

4.3. Scenario 3: A Curved Road with Two Different Desired Driving Behaviors. The previous sections demonstrated the performance of our proposed trajectory optimization method. In this section, we further conduct a comparison to see the p_1 and p_2 weight impact on the obedience to the lane centerline. To see the flexibility of CAV's decision, we set an obedient driving behavior with p_1 and p_2 as the default case and a flexible one with weights approaches to zero. To better visualize, we create a new one-lane 150 m curvy road with the same CAV initial conditions and adjust the reference line's weight. The dash-dotted line in Figure 10 shows the result with original weight, and the dotted line shows the result of all q_1, q_2, s_1, and s_2 changed to 0 which means that there is no reference line to follow.



(a) (b) (c)

FIGURE 6: Speed, acceleration, and angular deviation rate of CAV without the "buffer zone": (a) acceleration; (b) speed; (c) angular deviation rate.

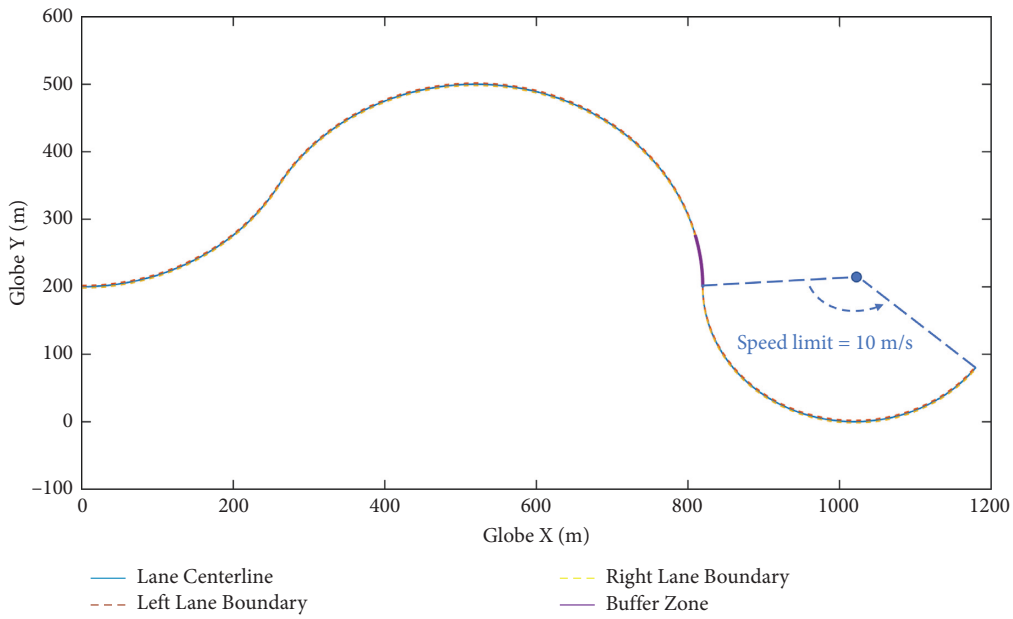
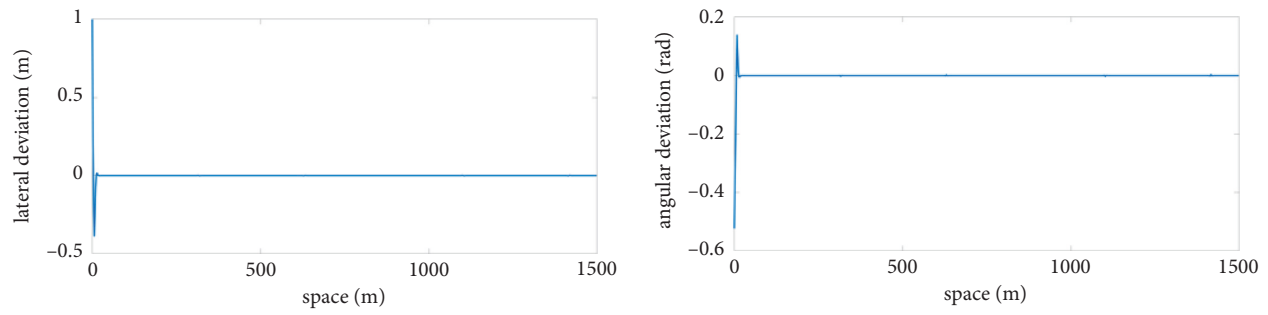


FIGURE 7: Illustration of road trajectory with a "buffer zone."



(a) (b)

FIGURE 8: Continued.

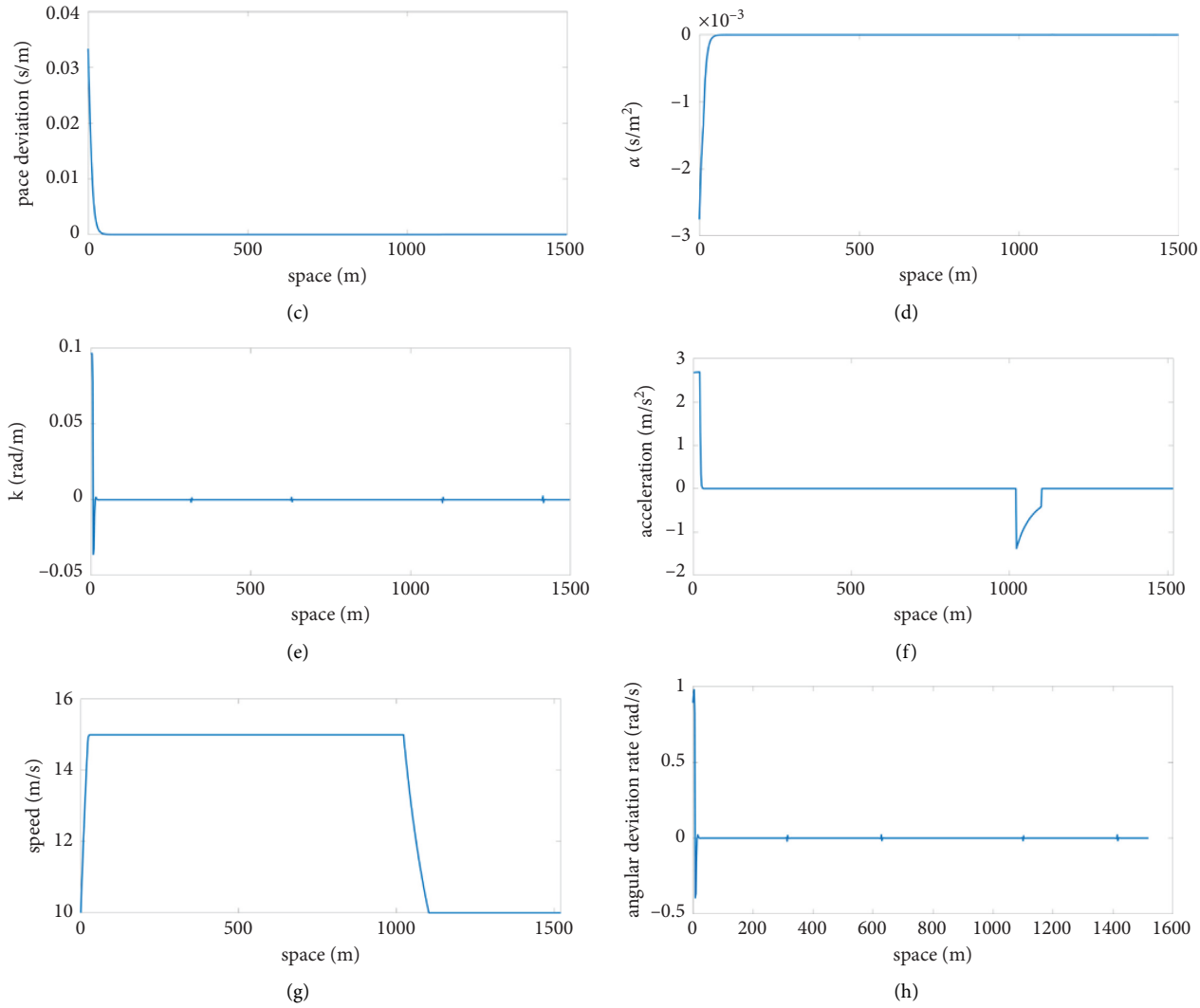


FIGURE 8: CAV state and control results with the "buffer zone": (a) lateral deviation $r(s)$; (b) angular deviation $\psi(s)$; (c) pace deviation $p(s)$; (d) relative moderation $\alpha(s)$; (e) relative angular spatial change rate $k(s)$; (f) acceleration; (g) speed; (h) angular deviation rate.

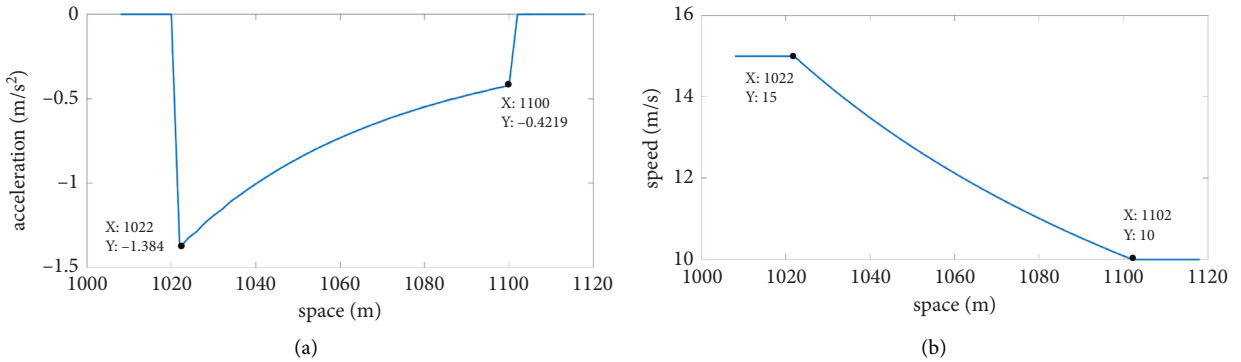


FIGURE 9: Speed and acceleration inside the "buffer zone": (a) acceleration; (b) speed.

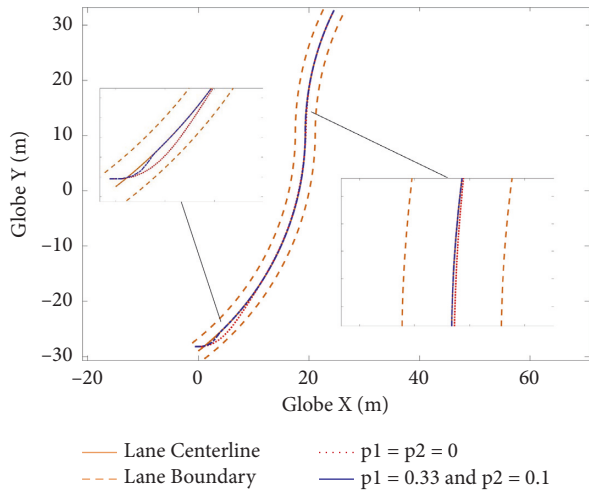


FIGURE 10: Comparison of different p_1 and p_2 values.

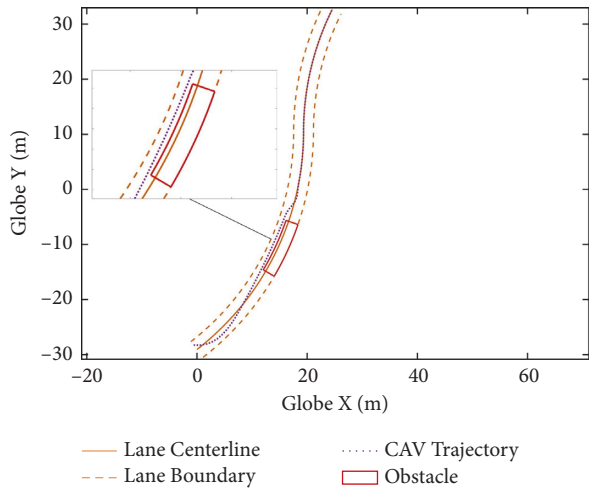


FIGURE 11: CAV trajectory with obstacle introduced.

The results shown in Figure 10 illustrate that CAV makes much more aggressive decisions with the obedient driving behavior than those with the flexible driving behavior. As we can see, CAV with the obedient driving behavior makes sharp turning decisions and quickly turns back to the lane centerline. In contrast, CAV makes smooth turning decisions and needs more time to reach the lane centerline with the flexible driving behavior. Moreover, unlike CAV with obedient driving behavior tightly follows the reference line as long as it reaches the reference line, CAV with the flexible driving behavior can deviate from the reference line when the curve occurs.

4.4. Scenario 4: A Curved Road with an Obstacle. We further conducted a simulation experiment using the same initial condition and one-lane road as shown in Figure 7 to test our proposed trajectory optimization method with a 10 m obstacle created. The $r^L_{m+n} = 0.5m$ and $r^U_{m+n} = 1.8m$ during the obstacle section. To meet both safety and driving comfort requirements, a joint driving mode is designed, which means a flexible driving behavior before and during the obstacle and an obedient driving behavior afterward.

As we can see from Figure 11, the CAV quickly detects the obstacle and makes smooth and comfortable control decisions to avoid the obstacle. The flexible driving behavior makes the CAV drive in the center to prevent any potential collision between the CAV and the obstacle during the obstacle section. In contrast, the obedient driving behavior lets the CAV quickly turn back and keep to the lane centerline after the CAV passes the obstacle section.

4.5. Scenario 5: A Curved Road with an Obstacle and System Disturbances. To show the robustness of our proposed method, we conduct an experiment with external system disturbance [44] based on Scenario 4 to demonstrate that our method is capable of providing efficient and stable control. Specifically, we add normally distributed random generated disturbances to each of the system states. The

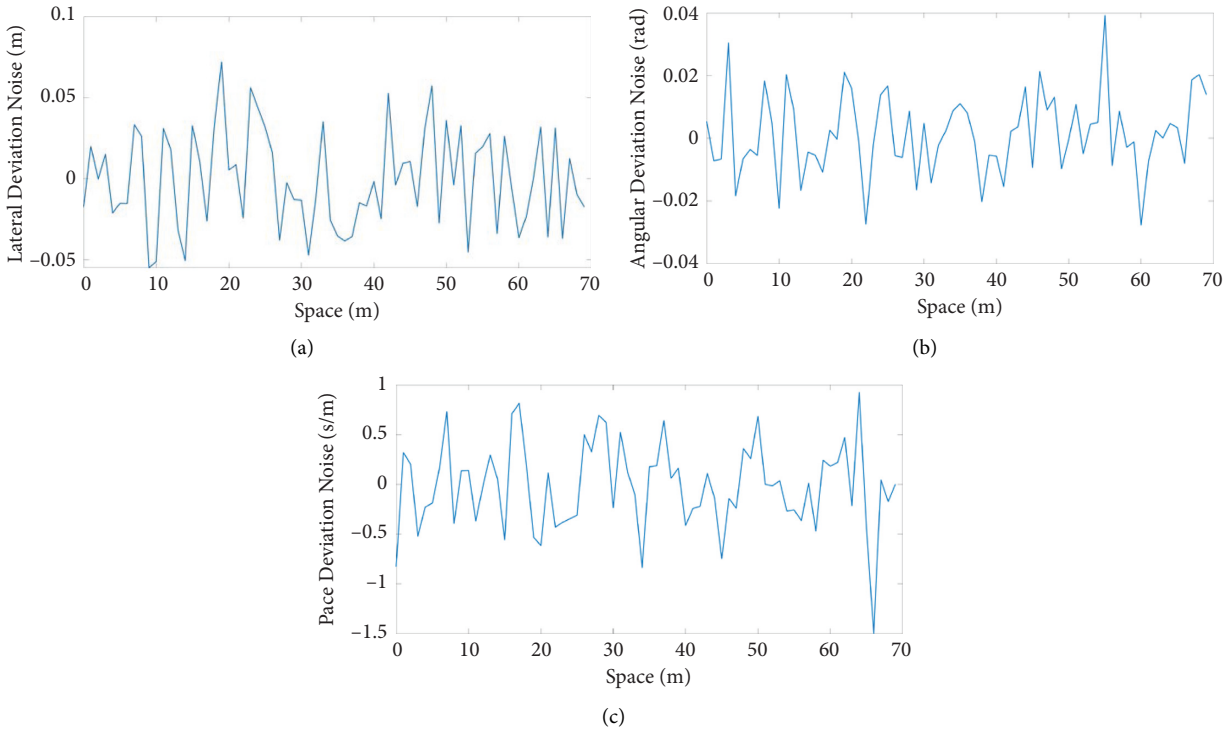


Figure 12: System disturbance: (a) lateral deviation noise; (b) angular deviation noise; (c) pace deviation noise.

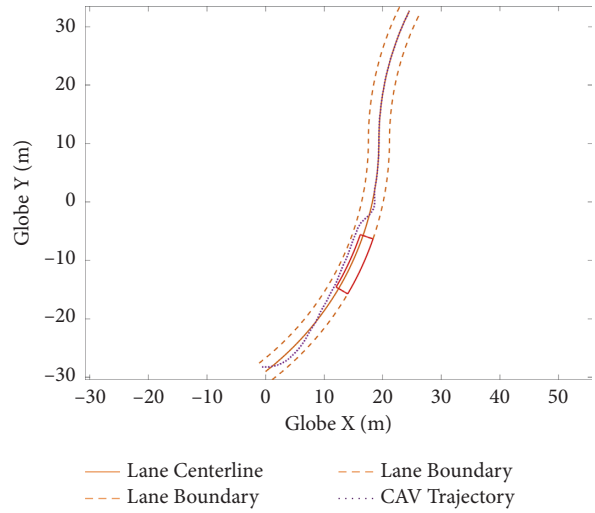


Figure 13: CAV trajectory with system disturbance.

details of disturbances are given in Figure 12, where the range of disturbance for the lateral deviation, the angular deviation, and the pace deviation is within $[-0.1, 0.1]$, $[-0.05, 0.05]$, and $[-0.0012, 0.0012]$ per meter, respectively.

As we can see from Figure 13, the resulted trajectory is also smooth and within the feasible drivable area, which suggests that our method is also efficient with system disturbance.

5. Conclusion

Vehicular trajectory optimization plays an essential role in ensuring vehicles' travel safety and efficiency. Traditional trajectory planning algorithms and methods are primarily formulated in the time domain by detecting the road information with a limited range and optimizing the trajectory

myopically, which cannot competently handle the spatially varying road geometric change, obstacles, and traffic regulations. To remedy that, this paper provided a new angle to plan long-term trajectories in a spatial domain with the help of infrastructure. Specifically, this paper systematically formulates trajectory optimization in a spatial domain and on a curvilinear coordinate, enabling our method to flexibly formulate spatially varying complex road geometries, traffic regulations, and road obstacles whose information can be obtained through V2I communication. For rigor, the controllability of the state space was mathematically proved by using both the linear test and the small-angle approximation. Considering vehicle travel efficiency and trajectory smoothness while satisfying the collision avoidance and vehicle kinematics constraint, a multiobjective MPC was constructed, which can be efficiently solved by the state of arts optimization methods.

To demonstrate the usefulness and wide applications of our proposed trajectory planning optimization algorithm, multiscenarios numerical simulations were conducted, which include five parts: (i) a continuous curvy road segment with the constant speed limit, (ii) a continuous curvy road segment with speed limit change, (iii) a curved road with two different desired driving behaviors, (iv) a curved road with an obstacle, and (v) a curved road with an obstacle and system disturbances. As the results suggested, the proposed trajectory optimization method could guarantee satisfactory performance for all scenarios under different types of disturbances. Further, the algorithm is capable of providing stable trajectory control with nearly no oscillation by utilizing the information provided by the infrastructure and newly presented formulation.

Some directions can be extended on the current framework in the future. For example, future studies could apply a more complex four-wheel vehicle control model and incorporate uncertainties in vehicle dynamics. Moreover, we will extend the MPC framework to include ambient vehicle movement and cooperatively plan the trajectories of CAVs for more applicable scenarios, such as intersections and on-ramp scenarios [45].

Data Availability

We do not use any field collected data in this paper. The numerical experiment code will be shared based on the requests.

Conflicts of Interest

The authors ensure that there are no conflicts of interest for this paper.

Acknowledgments

This research was supported by Wisconsin Traffic Operation and Safety (TOPS) Laboratory.

References

[1] D. Cao, J. Wu, J. Wu, B. Kulcsár, and X. Qu, "A Platoon Regulation Algorithm to Improve the Traffic Performance of Highway Work Zones," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, 2021.

[2] Y. Wang, S. Hou, and X. Wang, "Reinforcement Learning-Based Bird-View Automated Vehicle Control to Avoid Crossing Traffic," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 7, pp. 890–901, 2020.

[3] Z. Wang, X. Zhao, Z. Xu, X. Li, and X. Qu, "Modeling and Field Experiments on Autonomous Vehicle Lane Changing with Surrounding Human-Driven Vehicles," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 7, pp. 877–889, 2020.

[4] A. Severino, S. Curto, S. Barberi, F. Arena, and G. Pau, "Autonomous vehicles: an analysis both on their distinctiveness and the potential impact on urban transport systems," *Applied Sciences*, vol. 11, no. 8, p. 3604, 2021.

[5] F. Arena, G. Pau, A. Severino, S. Trubia, and S. Curto, "Future connected cars through the evolution of telematics and infotainment," *International Journal on Engineering Applications (IREA)*, vol. 9, no. 2, p. 49, 2021.

[6] B. Gutjahr, L. Groll, and M. Werling, "Lateral vehicle trajectory optimization using constrained linear time-varying MPC," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1–10, 2016.

[7] F. Mensing, R. Trigui, and E. Bideaux, "Vehicle trajectory optimization for application in ECO-driving," in *Proceedings of the 2011 IEEE Vehicle Power and Propulsion Conference*, pp. 1–6, Chicago, IL, USA, September 2011.

[8] H. Guo, C. Shen, H. Zhang, H. Chen, and R. Jia, "Simultaneous trajectory planning and tracking using an MPC method for cyber-physical systems: a case study of obstacle avoidance for an intelligent vehicle," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 4273–4283, 2018.

[9] S. Manzinger, C. Pek, and M. Althoff, "Using Reachable Sets for Trajectory Planning of Automated Vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 6, 2020.

[10] M. G. Plessen, "Trajectory Planning of Automated Vehicles in Tube-like Road Segments," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, Yokohama, Japan, October 2017.

[11] M. Werling, J. Ziegler, S. Kammel, and S. Thrun, "Optimal Trajectory Generation for Dynamic Street Scenarios in a Frenet Frame," in *Proceedings of the 2010 IEEE International Conference on Robotics and Automation*, pp. 987–993, Anchorage, AK, USA, May 2010.

[12] E. Frazzoli, M. A. Dahleh, and E. Feron, "Real-time motion planning for agile autonomous vehicles," *Journal of Guidance, Control, and Dynamics*, vol. 25, no. 1, pp. 116–129, 2002.

[13] F. Gritschneder, K. Graichen, and K. Dietmayer, "Fast Trajectory Planning for Automated Vehicles Using Gradient-Based Nonlinear Model Predictive Control," in *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7369–7374, Madrid, Spain, October 2018.

[14] J. Bohren, T. Foote, J. Keller et al., "Little ben: the ben franklin racing team's entry in the 2007 DARPA urban challenge," *Journal of Field Robotics*, vol. 25, no. 9, pp. 598–614, 2008.

[15] R. Kala and K. Warwick, "Multi-level planning for semi-autonomous vehicles in traffic scenarios based on separation maximization," *Journal of Intelligent and Robotic Systems*, vol. 72, no. 3-4, pp. 559–590, 2013.

[16] J. Reeds and L. Shepp, "Optimal paths for a car that goes both forwards and backwards," *Pacific Journal of Mathematics*, vol. 145, no. 2, pp. 367–393, 1990.

[17] T. Fraichard and A. Scheuer, "From reeds and shepp's to continuous-curvature paths," *IEEE Transactions on Robotics*, vol. 20, no. 6, pp. 1025–1035, 2004.

[18] L. Labakhua, U. Nunes, R. Rodrigues, and F. S. Leite, "Smooth trajectory planning for fully automated passengers vehicles: spline and clothoid based methods and its simulation," in *Informatics in Control Automation and Robotics*Springer, Berlin, Germany, 2008.

[19] Y. Kuwata, S. Karaman, J. Teo, E. Frazzoli, J. P. How, and G. Fiore, "Real-time motion planning with applications to autonomous urban driving," *IEEE Transactions on Control Systems Technology*, vol. 17, no. 5, pp. 1105–1118, 2009.

[20] S. Karaman, M. R. Walter, A. Perez, E. Frazzoli, and S. Teller, "Anytime Motion Planning Using the RRT," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation*, pp. 1478–1483, Shanghai, China, May 2011.

[21] L. B. Cremean, T. B. Foote, J. H. Gillula et al., "Alice: an information-rich autonomous vehicle for high-speed desert navigation," *Journal of Field Robotics*, vol. 23, no. 9, pp. 777–810, 2006.

[22] D. Kogan and R. Murray, "Optimization-based Navigation for the DARPA Grand Challenge," in *Proceedings of the Conference on Decision and Control (CDC)*, San Diego, CA, USA, December 2006.

[23] J. Wu, B. Kulcsár, S. Ahn, and X. Qu, "Emergency vehicle lane pre-clearing: from microscopic cooperation to routing decision making," *Transportation Research Part B: Methodological*, vol. 141, pp. 223–239, 2020.

[24] J. Wu, S. Ah, Y. Zhou, P. Liu, and X. Qu, "The Cooperative Sorting Strategy for Connected and Automated Vehicle Platoons," 2020, https://arxiv.org/abs/2003.06481.

[25] T. Berglund, A. Brodnik, H. Jonsson, M. Staffanson, and I. Soderkvist, "Planning smooth and obstacle-avoiding B-spline paths for autonomous mining vehicles," *IEEE Transactions on Automation Science and Engineering*, vol. 7, no. 1, pp. 167–172, 2010.

[26] J. Pérez, J. Godoy, J. Villagrá, and E. Onieva, "Trajectory Generator for Autonomous Vehicles in Urban Environments," in *Proceedings of the 2013 IEEE international conference on robotics and automation*, pp. 409–414, Karlsruhe, Germany, May 2013.

[27] M. Elbanhawi and M. Simic, "Sampling-based robot motion planning: a review," *IEEE Access*, vol. 2, pp. 56–77, 2014.

[28] M. Nolte, M. Rose, T. Stolte, and M. Maurer, "Model predictive control based trajectory generation for autonomous vehicles—an architectural approach," in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 798–805, Los Angeles, CA, USA, June 2017.

[29] A. Heilmeier, A. Wischnewski, L. Hermansdorfer, J. Betz, M. Lienkamp, and B. Lohmann, "Minimum Curvature Trajectory Planning and Control for an Autonomous Race Car," *Vehicle System Dynamics*, vol. 58, pp. 1–31, 2019.

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

[31] Y. Zheng, S. E. Li, K. Li, F. Borrelli, and J. K. Hedrick, "Distributed model predictive control for heterogeneous vehicle platoons under unidirectional topologies," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 899–910, 2017.

[32] S. Gong, J. Shen, and L. Du, "Constrained optimization and distributed computation based car following control of a connected and autonomous vehicle platoon," *Transportation Research Part B: Methodological*, vol. 94, pp. 314–334, 2016.

[33] T. D. Barfoot and C. M. Clark, "Motion planning for formations of mobile robots," *Robotics and Autonomous Systems*, vol. 46, no. 2, pp. 65–78, 2004.

[34] A. Khalifa, O. Kermorgant, S. Dominguez, and P. Martinet, "An observer-based longitudinal control of car-like vehicles platoon navigating in an urban environment," in *Proceedings of the 2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 5735–5741, Nice, France, December 2019.

[35] E. Héry, S. Masi, P. Xu, and P. Bonnifait, "Map-based Curvilinear Coordinates for Autonomous Vehicles," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–7, Yokohama, Japan, October 2017.

[36] Y. Zhang, Y. Bai, M. Wang, and J. Hu, "Cooperative Adaptive Cruise Control with Robustness against Communication Delay: An Approach in the Space Domain," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 9, pp. 5496–5507, 2020.

[37] J. Zhao, V. L. Knoop, and M. Wang, "Two-dimensional vehicular movement modelling at intersections based on optimal control," *Transportation Research Part B: Methodological*, vol. 138, pp. 1–22, 2020.

[38] B. Ran, T. Chen, Y. Cheng et al., "System and methods for partially instrumented connected automated vehicle highway systems," *Google Patents, Jul.*vol. 30, 2020.

[39] S. S. Shadrin and A. A. Ivanova, "Analytical review of standard Sae J3016 «taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles» with latest updates," *Avtomobil' Doroga Infrastruktura*, vol. 3, no. 21, p. 10, 2019.

[40] M. S. Mahmoud, *Advanced Control Design with Application to Electromechanical Systems*, Butterworth-Heinemann, Oxford, UK, 2018.

[41] B. M. Chen, Z. Lin, and Y. Shamash, *Linear systems theory*, Springer, Berlin, Germany, 2004.

[42] Y. Zhou, S. Ahn, M. Wang, and S. Hoogendoorn, "Stabilizing mixed vehicular platoons with connected automated vehicles: an H-infinity approach," *Transportation Research Part B: Methodological*, vol. 132, pp. 152–170, 2020.

[43] B.-C. Chen, B.-C. Luan, and K. Lee, "Design of lane keeping system using adaptive model predictive control," in *Proceedings of the 2014 IEEE International Conference on Automation Science and Engineering (CASE)*, pp. 922–926, New Taipei, Taiwan, August 2014.

[44] Y. Zhou, S. Ahn, M. Chitturi, and D. A. Noyce, "Rolling horizon stochastic optimal control strategy for ACC and CACC under uncertainty," *Transportation Research Part C: Emerging Technologies*, vol. 83, pp. 61–76, 2017.

[45] T. Chen, M. Wang, S. Gong, Y. Zhou, and B. Ran, "Connected and automated vehicle distributed control for on-ramp merging scenario: a virtual rotation approach," *Transportation Research Part C: Emerging Technologies*, vol. 133, Article ID 103451, 2021.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 3462267, 12 pages
https://doi.org/10.1155/2022/3462267

Research Article

Vehicle Path Recognition Approach Based on Incomplete Automatic Vehicle Identification

Meimei He,[1] Xianyu Wu ,[2] Zhenxi Zhan,[2] and Jingxue Guo[2]

[1]*Zhongzi Huake Traffic Construction Technology Co Ltd, Beijing, China*
[2]*Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing, China*

Correspondence should be addressed to Xianyu Wu; wuxy@bjtu.edu.cn

Received 15 March 2022; Accepted 6 April 2022; Published 9 May 2022

Academic Editor: Zhihong Yao

Vehicle path recognition is one of the key methods used in urban traffic research, such as traffic flow characteristics analysis. Automatic vehicle identification (AVI) is often used for vehicle path recognition and is suitable for mixed traffic flow with connected automated vehicles (CAVs). However, there still remain issues in overcoming the difficulty of vehicle path identification caused by the discontinuity of AVI data and solving the problem of low precision of AVI application. To model the vehicle path, this paper selects the AVI system of Yicheng Town, Linfen City, Shanxi Province, as a test bed. The travel modes of private cars and taxis are discussed, and the quantified indicators of the model are determined. By combining the analytic hierarchy process (AHP) with the entropy weight method (EWM) to get the weights of the indicators, the path recognition model under incomplete AVI data is proposed. Finally, based on the path recognition model proposed in this paper, case studies are carried out for the private car and taxi path recognition, respectively. The validity of the path identification through practical studies and the effect of the number of missing nodes of AVI equipment on the accuracy of the model are discussed. The results show that the recognition of the travel path using the proposed model is consistent with the actual travel path. The accuracy of the proposed model is more than 60% when the number of missing nodes is less than 7 in total 31 nodes. Considering the decision models for private cars and taxis, respectively, the proposed model provides a method for vehicle path recognition based on incomplete AVI data.

1. Introduction

With the prevalent traffic information and data mining technology, it becomes possible to extract the complete path information of vehicles from the collected data by the automatic vehicle identification (AVI) system in most cities. The AVI data is the most direct reflection of the urban traffic travel law, and on the other hand, it also raises attention from transportation engineers and researchers. With AVI data, the complete path information of both human-driven vehicles (HDVs) and connected automated vehicles (CAVs) could be extracted uniformly, which is the groundwork for the further study of the characteristics of mixed traffic flow with CAVs. However, due to funding limitations, the system is usually incomplete. These AVI devices had to be installed at some critical places instead of all the places that require the devices, which may result in some blind spots. The missing data then increased the difficulty of obtaining the vehicles' travel paths, whereas there is no current practical model to solve this issue. With this consideration, it is necessary to establish a vehicle path recognition method under a set of incomplete AVI data, which would help to exploit data and benefit both traffic police departments and engineers.

Most of the research on AVI data originated from the computer vision aspect, focusing on how to extract accurate license plate information from an image or video system [1, 2]. With the improvement of image recognition technology and computer computing efficiency, image processing and license plate recognition functions are gradually

added to the AVI system [3]. Ajanthan [4] built a license plate recognition system under low-resolution surveillance video, which is robust for environmental changes such as illumination. Lee [5] developed a license plate recognition method in the traffic monitoring scene and trained the license plate data set through the deep convolution network so that the license plate recognition accuracy could be higher than the existing license plate recognition method.

For the application of AVI data, prevalent studies mainly apply the data of the road network monitoring system to obtain traffic parameters such as road network OD matrix [6, 7], road traffic flow [8, 9], and road traffic time [10]. In China, some scholars used AVI data to estimate the average travel time of road segments [11], collect the traffic flow of road sections [12], and calculate the average driving speed under the signal progression [13, 14]. However, there is a lack of research on vehicle trajectory reconstruction based on AVI data. In recent years, some Chinese scholars used AVI data to recover the vehicle trajectory. By fusing the fix-point detector and signal timing data, Tang [15] invented a vehicle trajectory reconstruction method based on traffic wave theory and traffic simulation theory, which overcomes the interference of roadside vehicles. Xiang [16] invented a fusion of AVI devices and a fixed-point detector trunk trajectory reconstruction method, which can improve the accuracy of trajectory reconstruction. Yu [17] presented a vehicle identification data-based trajectory reconstruction method for signalized-link by constructing a phase-to-phase backtracking framework and using the shockwave theory to reconstruct vehicular trajectory segments involved in each backtracking step. Lin and Yang [18, 19] collected the AVI data and extracted vehicle travel trajectories. He also used the data to analyze the pollutant emission sources and emission intensity. Zhang [20] used AVI data at the signalized intersection to reconstruct the trajectory of the vehicles and extract the delay information.

The above studies are applications of AVI data in vehicle trajectory reconstruction, while these studies did not consider the problem of incomplete AVI data, which leads to a certain deviation between the traffic in the road network and the actual situation. Some shortest path algorithms are used to fill the path between the two missing nodes, but when there are more missing nodes, the vehicle does not exactly follow the shortest path, and it is impossible to depict the specific travel path of the vehicle in the road network. Rao [21] used a particle filter to estimate the probability of a vehicle's trajectory from all possible candidate trajectories based on AVI data. Li [22] used the improved Dijkstra algorithm to search for the shortest path of the first K bars between two points by distance. Mo [23] developed a Bayesian path reconstruction model to replenish the lost information resulting from the recognition error and insufficient coverage rate of the AVI system. Guo [24] extracted road travel time based on AVI data.

Some scholars use multiobjective decision-making methods to identify missing paths. Wang [25], Yang et al. [26], Yang et al. [27], and Yang et al. [28] proposed using multitarget decision-making methods to complete the vehicle's travel path, but they assumed that all vehicles use the same indicators for multitarget path decision-making, which did not consider the road network of different models of the path selection factors and did not distinguish the weight value of each indicator. Most studies use a single method to determine the index weight in calculating the weight of the indicator. The idea of AHP-EWM combining the calculation of the index combination weight has been partially applied in other fields. But the AHP-EWM study combined with path recognition is relatively scarce.

Based on the background of the above research, this paper considers the different path selection indicators of the two models (taxi and private car) under the background of the incomplete AVI data and establishes a vehicle path recognition model to solve the difficult path recognition problem caused by the discontinuity of vehicle travel records.

2. Problem Description

Vehicle travel path recognition refers to the path selection behavior at the missing point. In the set of feasible paths between the origin and the destination, the path matching the actual vehicle travel path may include multiple decision points. The schematic diagram of the path selection behavior is shown in Figure 1. There are three travel paths to choose between OD: Path 1: link 3-link 8 (O–F-D); Path 2: link 1 - link 4 - link 6 - link 7 (O-A-E-C-D); and Path 3: link 1 - link 2 - link 5 - link 7(O-A-B-C-D).

When there are more than two missing nodes between travel origin and destination, it is difficult to complete the path directly based on the neighboring intersections in the road network. In previous studies, the shortest path algorithm is mostly used to recognize the path between two missing nodes. However, vehicle travel in the actual road network is jointly influenced by a variety of factors, and different types of vehicles are correlated with different factors. So, the method of using the shortest path algorithm to fill in the missing path cannot be adapted to all vehicle types.

In small cities with incomplete public transit systems, the number of taxis in the road network is gradually increasing with the development of taxi services. The proportion of taxis in the traffic flow is gradually increasing. Private cars travel mainly for commuters, while taxis travel mainly to collect passengers for profit. Different travel purposes determine different travel paths. If the travel paths of taxis in the road network are recognized according to the path recognition model of private cars, the received travel paths often do not match the actual travel paths of taxis. If the taxis in the road network are ignored, and only the travel paths of private cars are recognized, it will cause errors in the traffic volume of the road network.

In summary, it is difficult for the existing methods to recognize all vehicles' paths in the road network. In this paper, we will determine the corresponding path decision indicators for the different travel factors of taxis and private cars. A multiobjective decision model considering different vehicle types is proposed. The missing paths of private cars and taxis in the AVI data are recognized separately using this model.

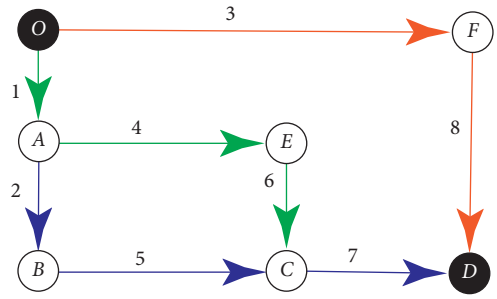


Figure 1: Path selection behavior diagram.

3. Methodology

3.1. Model Assumptions. The following assumptions are adopted to simplify the analysis:

(1) The case of vehicles traveling back and forth between ODs is not considered. That is, there is no circular path in the path.

(2) Vehicles travel only on primary roads, secondary roads, feeder roads, and some other roads with wider widths. Vehicles driving on abnormal segments (e.g., nonmotorized roads) are not considered.

3.2. Indicator Selection. To better describe the vehicle path recognition model, it is necessary to select the corresponding indicators for the path recognition model of private cars and taxis. Factors affecting vehicle routing [29] are generally divided into 3 aspects: subjective factors of travelers, such as gender, age, income, and familiarity with the road network; travel purposes, such as commuting and entertainment; environmental characteristics of roads, such as path length, road grade, the number of signalized intersections on the path, the number of turns on the path, the real-time traffic status on the path, the nature of land use, and the effective passage time of the path, etc.

All the data in this paper are based on the AVI data. There is no additional survey to the data. Therefore, the data of the traveler's characteristics cannot be known, which is temporarily not considered. Only factors of road environment characteristics are considered. Different vehicle types have different preferences to choose the travel path. 1000 path data are randomly selected from the extracted complete driving paths as the analysis sample. Combined with vehicle travel data, the Pearson correlation coefficient is used to analyze the path behavior of each indicator in private cars and taxis in SPSS.

The correlation coefficient results of private cars and taxis are shown in Figure 2. The ordinate Y-axis in Figure 2 is the correlation coefficient between each index factor and the probability of path selection. When the value is positive, it means that the indicator is positively correlated with the probability of path selection. The height of the ordinate in the diagram represents the strength of the correlation.

As can be seen from Figure 2, the correlation order of each indicator in the path selection of private cars is as follows:

Consistency between actual and ideal travel time- > Traffic operation > Left turn times > Path length- > Number of signalized intersections > Road grade.

That means, travelers are prone to choose the ideal path with the shortest path time, no congestion, and relatively smooth driving. At the same time, the correlation between the nature of land use and the probability of private car path selection is not very high, which is almost zero.

The correlation order of each indicator in the path selection of taxis is as shown in Figure 3: Consistency between actual and ideal travel time > Number of signalized intersections > Traffic operation > Land-use > Left turn times > Path length.

Different from private cars, taxis pay more attention to land use when choosing the path, which is closely related to the passenger-taking behavior of taxis. At the same time, the number of signalized intersections is also considered because taxis tend to have smaller delays. The significant correlation level of each indicator can meet the requirements ($p \leq 0.05$).

In summary, the decision-making indicator of the private car and taxi travel path is shown in Tables 1 and 2.

3.3. Modelling. The consistency between actual and ideal travel time, traffic operation, left turn times, path length, the number of signalized intersections, and road grade are selected as the decision indicators of the private car impedance model. In the other side, the decision indicators of the taxi path model include the consistency between actual and ideal travel time, the number of signalized intersections, traffic operation, land use, left turn times, and path length. The effective travel time, traffic operation, left turn times, path length, and the number of signal intersections are the common indicators of the two types of vehicles. The road grade is a special indicator for private cars, and the land use is a special indicator for taxis.

Path recognition for vehicles with missing nodes in the AVI data is a process of determining the optimal path from the set of reasonable paths obtained according to the path search algorithm. The vehicles always hope to obtain the shortest path length and the shortest travel time, which is a multiobjective decision-making problem.

Every indicator has a different degree of importance because of its different logic system. Through the no-dimension treatment of different indicators and the determination of the weight value, the multiobjective path decision-making problem is transformed into the problem of finding the optimal value of a single objective. The vehicle selects the overall most satisfactory path in the travel, and the variables (ρ_1, ρ_2) are set:

$\rho_1 = 1$ means that the traveling vehicle is a private car. Otherwise, it should show $\rho_1 = 0$. The variable $\rho_2 = 1$ indicates that the vehicle is a taxi. Otherwise, the value should be 0. The vehicle path recognition model is constructed as shown in formula (1), and formula (2) is the objective function of the model.

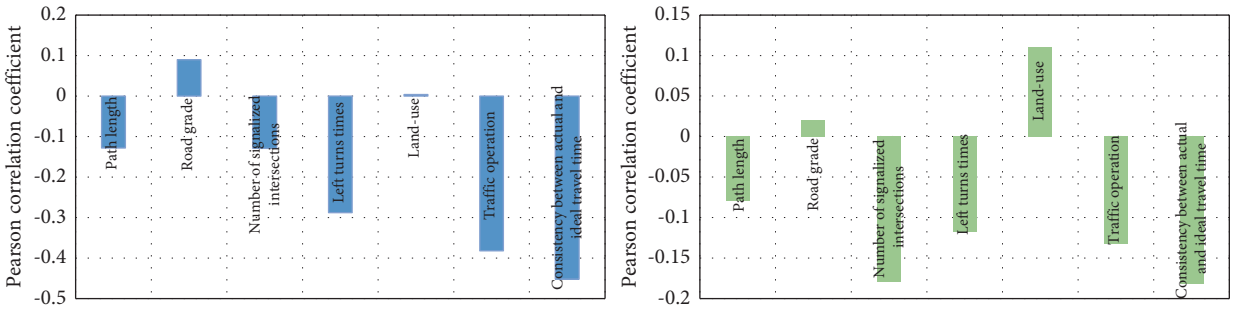


Figure 2: Correlation analysis of private car and taxi.

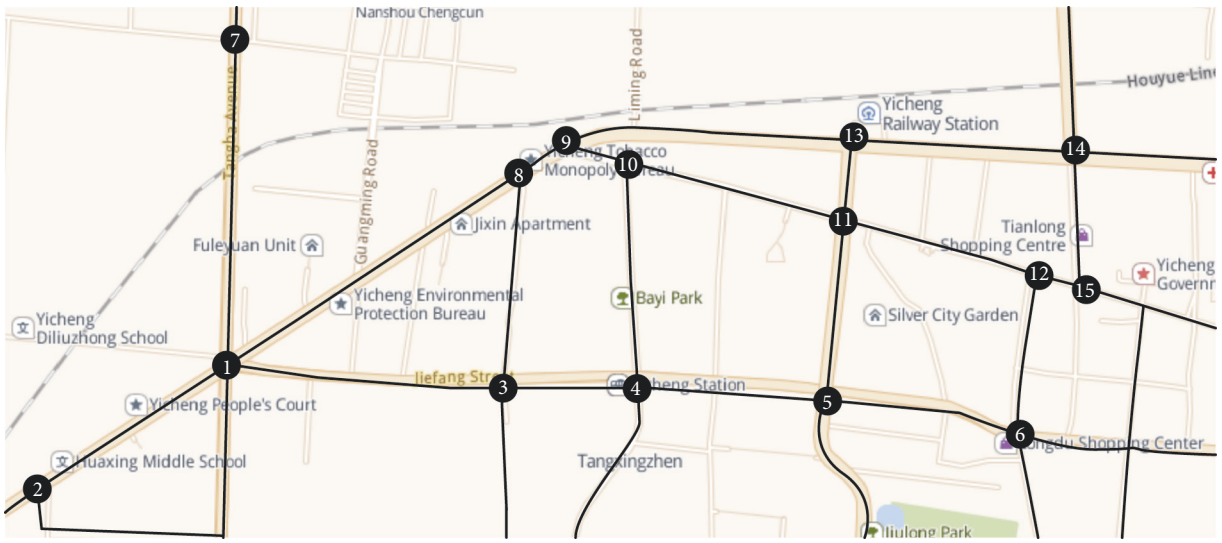


Figure 3: Road network map of Yicheng County.

Table 1: Correlation analysis of decision indicator and route selection probability of private cars.

Indicator	p value	Calculation method
Consistency between actual and ideal travel time	.000	$\tau_k^{rs} = T_k' - T_k/T_k$
Traffic operation	.002	$C_k^{rs} = \sum_{a(i,j)\in E} l_{a(i,j)} c_{a(i,j)}/L_k^{rs}$
Left turn times	.008	$\sum \psi_{rps} = P_{rp} \otimes P_{ps}$
Path length	.010	$L_k^{rs} = \sum_{a(i,j)\in E} l_{a(i,j)} \xi_{a(i,j),k}^{rs}$
Number of signalized intersections	.011	$S_k^{rs} = \sum_{a(i,j)\in E} l_{a(i,j)} s_{a(i,j)}/L_k^{rs}$
Road grade	.019	$D_k^{rs} = \sum_{a(i,j)\in E} l_{a(i,j)} d_{a(i,j)}/L_k^{rs}$

Table 2: Correlation analysis of decision indicator and route selection probability of taxis.

Indicator	p value	Calculation method
Consistency between actual and ideal travel time	.000	$\tau_k^{rs} = T_k' - T_k/T_k$
Number of signalized intersections	.001	$S_k^{rs} = \sum_{a(i,j)\in E} l_{a(i,j)} s_{a(i,j)}/L_k^{rs}$
Traffic operation	.002	$C_k = \sum_{a(i,j)\in E} l_{a(i,j)} c_{a(i,j)}/L_k^{rs}$
Land-use	.003	$N_k = \sum_{a(i,j)\in E} l_{a(i,j)} n_{a(i,j)}/L_k^{rs}$
Left turn times	.011	$\psi_{rps} = P_{rp} \otimes P_{ps} = 1$
Path length	.021	$L_k^{rs} = \sum_{a(i,j)\in E} l_{a(i,j)} \xi_{a(i,j),k}^{rs}$

$$c_{a,k} = \omega_1 L_k^{rs} + \omega_2 C_k^{rs} + \omega_3 S_k^{rs} + \omega_4 \sum \psi_{rps} + \omega_5 \tau_k^{rs} + \rho_1 \omega_6 D_k^{rs} + \rho_2 \omega_7 N_k^{rs}, \quad (1)$$

$$\min c_{a,k}, \quad (2)$$

where $c_{a,k}$ is the satisfaction value of the path k. ω_j is the combined weight value of the j indicator. L_k^{rs} is the path length of the path k between r and s.C_k^{rs} is the road operation condition of the path k between r and s.S_k^{rs} is the number of signalized intersections of the path k between r and s.$\sum \psi_{rps}$ is the number of left turns at the path k. τ_k^{rs} is the accordance rate of the theoretical transit time of path k between r and s with the actual transit time. D_k^{rs} is the road grade of the path k between r and s.N_k^{rs} is the land-use indicator of the path k between r and s.ρ_1 is the private car representative indicator. ρ_2 is the taxi representative indicator.

The constraint of the model is $\sum \omega_j = 1$, ensuring that the sum of weights is always 1.

4. Model Indicator Weight Calculation Based on AHP-EWM Method

AHP method is a method of determining weights that combines quantitative and qualitative analysis, and it is the principle that the decision maker's thinking and decision-making process can always be consistent in this paper. However, since expert scoring is used in weighting indicators in the decision-making process of the AHP method, it leads to the subjective orientation appearing in the decision-making process. The single use of the AHP method will have an impact on the evaluation results, leading to more subjective factors in the path selection. Therefore, to avoid subjective decision-making, the EWM method is introduced to correct the weight of the indicators, then the path recognition process is more consistent with the actual situation.

Assuming that there are n evaluation indicators for the k alternative paths of a certain type of vehicle. y_{ij} is the evaluation indicator j for the scheme i, where $i = 1, 2, \ldots\ldots, k$, $j = 1, 2, \ldots\ldots, n$. The calculation steps of the EWM method to correct the weight value are as follows.

Step 1. : Standardized processing. The purpose is to eliminate the interaction of dimensions and orders of magnitude between the indicators. And it is necessary to standardize the indicator y_{ij} in the original data. The common methods to eliminate the dimension are the extreme-range method and the Z-score method. The Z-score method will result in a negative entropy value due to the accumulation of errors, which is not following the entropy principle.

In this paper, the extreme-range method is considered to eliminate the dimensional problem between the indicators. The correlation analysis between the indicators and the probability of path selection shows that the two indicator values of road grade and land use are positively correlated benefit indicators, and all other indicators are negatively correlated cost indicators.

For the benefit indicator, the extreme-range formula is as follows:

$$y'_{ij} = \frac{y_{ij} - y_{\min}}{y_{\max} - y_{\min}}, \quad (3)$$

where $y_{\max}$ and $y_{\min}$ are the maximum and minimum values of the indicator j.

For cost indicators, the extreme-range formula is as follows:

$$y'_{ij} = \frac{y_{\max} - y_{ij}}{y_{\max} - y_{\min}}. \quad (4)$$

However, the extreme-range method will cause the maximum and minimum boundary values to be zero after the elimination of dimensions, which will affect the subsequent calculation of entropy. To avoid the problems, the shift operation is adopted after the standardization of the indicators so that:

$$y''_{ij} = y'_{ij} + 0.001. \quad (5)$$

Obtain the standardized decision matrix after the dimensionless processing of the indicators $Y'' = \left\{y''_{ij}\right\}_{k \times n}$, which is as follows:

$$Y'' = \begin{bmatrix} y''_{11} & y''_{12} & \cdots & y''_{1n} \\ y''_{21} & y''_{22} & \cdots & y''_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ y''_{k1} & y''_{k2} & \cdots & y''_{kn} \end{bmatrix}. \quad (6)$$

Step 2. : Calculate the proportion of each indicator value in the standardized matrix z_{ij}, which is as follows:

$$z_{ij} = \frac{y''_{ij}}{\sum_{i=1}^{k} y''_{ij}}. \quad (7)$$

Step 3. : According to the definition of information entropy, calculate the information entropy value of each indicator j, namely:

$$e_j = -\frac{1}{\ln k} \sum_{i=1}^{k} z_{ij} \ln z_{ij} \; j = 1, 2, \ldots\ldots, n. \quad (8)$$

Step4. : Determine the weight of the indicator, which is as follows:

$$\omega_{E,j} = \frac{1 - e_j}{n - \sum_{j=1}^{n} e_j} \; j = 1, 2, \ldots\ldots, n. \quad (9)$$

To make the analysis results reflect the subjective factors and objective reality of the travelers, the weight values of each index obtained by the two methods are combined and calculated by using the Lagrange multiplier method based on the principle of minimum information entropy. Based on the AHP-EWM method, the weight vector ω is as follows:

$$\omega_j = \frac{\sqrt{\omega_{A,i}\omega_{E,j}}}{\sum_{j=1}^{n}\sqrt{\omega_{A,i}\omega_{E,j}}} \quad j = 1, 2, \ldots\ldots, n. \tag{10}$$

5. Case Study

5.1. Experimental Design. To verify the accuracy of the vehicle path recognition model, take the incomplete path set (remove some nodes from the full path) as the testbed and make a comparison of real vehicle path data and experimental results. Figure 3 is a map of the actual road network in a town. The node represents the major intersection in the road network, each intersection has AVI equipment, but due to the unstable system conditions, some equipment cannot do the real-time collection to the vehicle information.

We randomly select two complete travel paths of private cars and taxis between node 3 and node 14. The diagram of the actual travel paths of private cars and taxis is shown in Figure 4. The paths chosen by the two types of vehicles are different. To obtain the incomplete paths, we erased all links between node 3 and node 14, retained node 3 and node 14, and recomplete the path according to the vehicle path recognition model.

The actual travel records of private cars are shown in Table 3.

The actual travel records of taxis are shown in Table 4.

5.2. Calculation Process

5.2.1. Generate Reasonable Path. According to the K-shortest algorithm and the deletion algorithm, the reasonable path set of private cars and taxis between node 3 and node 14 is obtained. The direction of private cars at node 14 is from West to East. It can be inferred that the upstream intersection is node 13 in the West. Search for the first five shortest paths with the fixed upstream intersection of the end node and delete the paths with loops and detours. Finally, the first four alternative paths $R_k^{car} = \{R_1^{car}, R_2^{car}, R_3^{car}, R_4^{car}\}$ for private cars are selected. The detailed alternative paths for private cars are shown in Table 5.

Similarly, the first three alternative paths $R_k^{taxi} = \{R_1^{taxi}, R_2^{taxi}, R_3^{taxi}\}$ for taxis are selected. The detailed alternative paths for taxis are shown in Table 6.

5.2.2. Calculation Process of the Path Recognition Model. The path recognition process of a private car is taken as an example to verify the calculation process. The attribute value of each indicator, according to the quantitative calculation formula of the attribute value of each path, is calculated to obtain the indicator matrix $Y = \{y_{ij}\}_{4\times6}$ (Formula (3)). The rows of the matrix represent the path number (Table 3), and the columns from left to right each represent the value of an indicator, such as the consistency of passing time, road operation, number of left turns, path length, number of signalized intersections, and road category.

$$Y = \begin{bmatrix} 1.55 & 2.76 & 1 & 2.13 & 4 & 6.37 \\ 4.37 & 4.56 & 0 & 2.07 & 3 & 5.26 \\ 8.50 & 2.95 & 2 & 2.34 & 4 & 5.85 \\ 5.12 & 3.73 & 1 & 2.27 & 4 & 5.24 \end{bmatrix}. \tag{11}$$

5.2.3. Indicator Weight Calculated by AHP Method. By constructing the indicator comparison matrix, the weight value vector of each indicator is obtained.

$$\omega_A = (\,0.3894 \quad 0.2125 \quad 0.1035 \quad 0.0335 \quad 0.0659 \quad 0.1952\,). \tag{12}$$

The maximum eigenvalue is as follows:

$$\lambda_{\max} = (\,6.58 \quad 6.3276 \quad 6.5926 \quad 6.3738 \quad 6.6484 \quad 6.2146\,). \tag{13}$$

The consistency indicator $CI = 0.074 < 0.1$, which passes the consistency test. Therefore, the indicator weight meets the requirements.

5.2.4. Calculate Indicator Weight by EWM Method. After range standardization and shifting 0.001 units to the right, the standardized matrix $Y'' = \{y_{ij}''\}_{4\times6}$ is calculated.

$$Y'' = \begin{bmatrix} 1.001 & 1.001 & 0.501 & 0.781 & 0.001 & 1.001 \\ 0.591 & 0.001 & 1.001 & 1.001 & 1.001 & 0.021 \\ 0.001 & 0.891 & 0.001 & 0.001 & 0.001 & 0.531 \\ 0.491 & 0.461 & 0.501 & 0.261 & 0.001 & 0.001 \end{bmatrix}. \tag{14}$$

The weight value vector of each indicator is obtained.

$$\omega_E = (\,0.1274 \quad 0.1472 \quad 0.1273 \quad 0.1595 \quad 0.2434 \quad 0.1951\,). \tag{15}$$

5.2.5. Calculate Combination Weight. According to formula (10), the weight values of each indicator obtained by the two methods are combined and calculated to obtain the weight vector ω based on the improved weight calculation by the AHP-EWM method.

$$\omega^{car} = (\,0.245 \quad 0.1945 \quad 0.1263 \quad 0.0804 \quad 0.1393 \quad 0.2146\,). \tag{16}$$

5.2.6. Calculate the Most Satisfactory Path. Substitute the weight calculation result into the model formula (1) to calculate the value of $c_{a,k}$.

$$\begin{aligned} c_{a,k}^{car} &= [\,3.1374 \quad 3.6713 \quad 4.9079 \quad 3.9704\,], \\ \min c_{a,k}^{car} &= 3.1374. \end{aligned} \tag{17}$$

The result means that the first path is the actual shortest path calculated by the model, as the path where the private car is finally recognized.

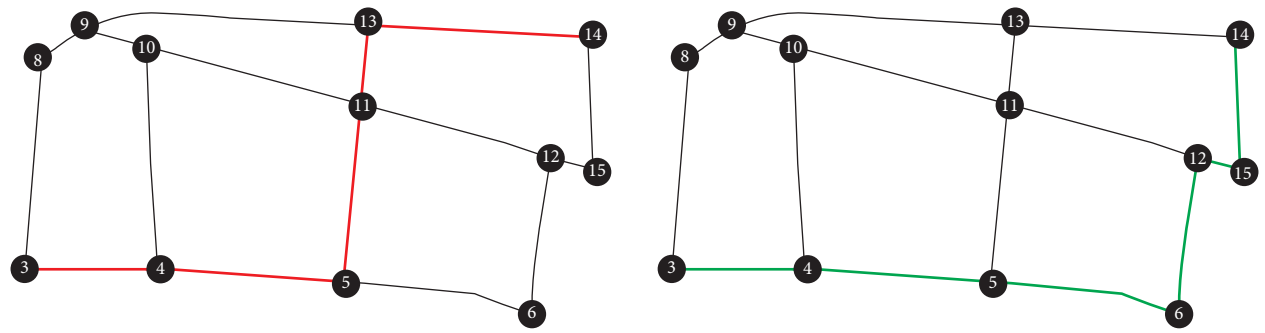


Figure 4: Diagram of the actual travel path of private cars and taxis.

Table 3: Actual travel records of private cars.

Number	Spot location	Direction	Departure time
1	3	West to east	16 : 11 : 56
2	4	West to east	16 : 12 : 33
3	5	South to north	16 : 13 : 11
4	11	South to north	16 : 15 : 56
5	13	West to east	16 : 17 : 32
6	14	West to east	16 : 18 : 34

Table 4: Actual travel records of taxis.

Number	Spot location	Direction	Departure time
1	3	West to east	11 : 15 : 11
2	4	West to east	11 : 15 : 59
3	5	West to east	11 : 17 : 12
4	6	South to north	11 : 19 : 06
5	12	West to east	11 : 22 : 32
6	15	South to north	11 : 23 : 04
7	14	South to north	11 : 24 : 07

Table 5: Alternative paths for private cars.

Alternative paths number	R_1^{car}	R_2^{car}	R_3^{car}	R_4^{car}
Path sequence	{3, 4, 5, 11, 13, 14}	{3, 8, 9, 13, 14}	{3, 4, 10, 11, 13, 14}	{3, 8, 9, 10, 11, 13, 14}

Table 6: Alternative paths for taxis.

Alternative paths number	R_1^{taxi}	R_2^{taxi}	R_3^{taxi}
Path sequence	{3, 4, 5, 6, 12, 15, 14}	{3, 4, 5, 11, 12, 15, 14}	{3, 8, 9, 10, 11, 12, 15, 14}

Similarly, the most satisfactory path for taxis is calculated.

$$c_{a,k}^{taxi} = \left[\, 3.912 \;\; 4.112 \;\; 4.832 \,\right], \quad \min c_{a,k}^{taxi} = 3.912. \tag{18}$$

The result means that the first alternative path is the path where the taxi is finally recognized.

5.3. Results and Discussion. The calculation results of the model are consistent with the actual results, and the most satisfactory path is also consistent with the actual path.

For private car: the path selection order sorted by model results is $R_1^{car} > R_2^{car} > R_4^{car} > R_3^{car}$. However, if we sort the path selection order by its length, R_2^{car} is the shortest path. Considering the traffic condition of the link R_2^{car}, sections 3-7 have lower road categories and poor traffic conditions. Congestion is likely to occur due to the increased travel activities during the evening peak. The result shows that it is consistent with the actual situation analysis and can objectively reflect the actual situation.

For taxi: the path selection order sorted by model results is $R_1^{car} > R_2^{car} > R_3^{car}$. The road conditions of sections 5-11 are better than sections 6-12. However, both paths of model results and actual results contain sections 6-12 instead of sections 5-11. Based on actual investigations, sections 6-12 are in a commercial area with large shopping malls and stores. Combining the travel time periods, the commercial area is the preferred destination for taxis, no matter whether the taxi is carrying passengers or not.

Through the above analysis, it can be concluded that the actual path chosen by travelers is not necessarily the shortest distance path, which is also compatible with the K-Shortest Paths (KSP) problem. For the same OD, the actual path results of private cars and taxis are different, as well as the path results obtained through the path recognition model are also different and consistent with the actual situation. It is proved that the model can accurately identify the vehicle travel path to a certain extent, which shows that the model established in this paper is effective for identifying the missing paths in the AVI data.

6. Influence of the Number of Missing Nodes on Model Accuracy

6.1. Experiment Design. To increase the test sample capacity, the road network is expanded to the central area of Yicheng County. The model assumes that vehicles driving in extremely narrow alleys are not considered, so the road network is simplified into a network composed of main trunk roads, secondary trunk roads and slip roads, as shown in Figure 5. The black solid nodes in the road network map are intersections with AVI device, and the white nodes are virtual. There is little AVI data at some intersections, and the phenomenon of missing vehicle detection is common. When the distribution density of the AVI devices in the road network is equal to that of the nodes in the road network, it is easy to obtain the travel records of the vehicles and analyze the traffic characteristics by mining the complete AVI data. However, due to the imperfect equipment and low coverage, the travel path cannot be obtained directly from the incomplete AVI data.

For the three paths with different distances, 100 private car paths and 50 taxi paths are randomly selected as the test data set. The actual paths of the vehicle are shown in Figures 6–8. To compare the experimental results with the real situation and analyze the error, three groups of experiments are designed according to the number of missing nodes. Remove some nodes in the complete path, and complete the path between the start and end nodes of the missing path according to the path recognition model.

The short-distance path represents the path shorter than 1 km, the medium distance path represents the path within 1 to 2 km, and the long-distance path represents the path longer than 2 km.

According to the number of missing nodes, the experiment is divided into three groups, as shown in Table 7.

The first group: 20% of the nodes are missing.

The second group: 50% of the nodes are missing.

The third group: all nodes are missing except for the origin and destination.

The method for selecting missing nodes is to generate random numbers in percentage from the sequence numbers of other nodes on the path except the start node and delete the corresponding node records.

The procedure is as follows.

Step1 : Use Dijkstra and deletion algorithm to obtain the alternative path set of the missing path in each case combined with the driving direction of the vehicle and determine the attribute indicator values on the path.

Step2 : Calculate the weight value of each attribute indicator according to the analytic hierarchy process, use MATLAB to calculate the entropy weight value of each attribute indicator, and then calculate the combined weight and the model result.

Step3 : Compare the quantity ratio of the model calculation results that are consistent with the actual path, and calculate the accuracy of the model.

Theoretically, the model calculation results should be consistent with the actual travel path. However, due to the different traffic operations in different periods, some paths may have high similarities with the theoretical path (path calculated by the proposed model). The length of the common substring $Public(R_k, R_{actual})$ represents the number of common sections of actual paths and theoretical paths, where R_k represents the k^{th} alternative path in the alternative path set and R_{actual} represents the theoretical travel path as well as the actual path. Then the accuracy of the model could be expressed as follows:

$$\gamma_j = \frac{\sum_{i=1}^{m} Public(R_k, R_{actual})}{m} \times 100\%. \tag{19}$$

6.2. Results Analysis. Through the analysis of the above experimental results, we can get the path decision-making results and accuracy of the three groups of experiments.

6.2.1. 20% of the Total Number of Nodes in the Path Were Missing. We assume that path (a) missed the data of node 3, whose front node is a six-legged intersection with complex flow directions. So, node 3, as the first intersection after turning, plays a key role in path selection. The results show that 95 percent of private car experimental paths are consistent with the actual paths; 100 percent of taxi experimental paths selections are consistent with the actual paths. Node 4 is lost in the long-distance path (c). The results show that all vehicles in the test path set of private cars and taxis choose the path consistent with the actual path. Since node 4 is between node 3 and node 5, when the number of missing nodes is low, the vehicle driving direction is clear, and there will not be too many path selection behaviors on short-distance sections, which is in line with the actual travel situation.

In conclusion, the model has good robustness when the number of missing nodes in the path is less than or equal to 20 percent of the total number of nodes on the path. The experimental results are shown in Table 8.

6.2.2. 50% of the Total Number of Nodes in the Path Were Missing. The short-distance path (a) missed nodes 3 and 4. Node 1 is a six-legged intersection, at which it is hard to know the vehicle's movement direction. So the vehicle needs to make a path decision after node 1. The results show that there are 13 vehicle paths in the private car test path set

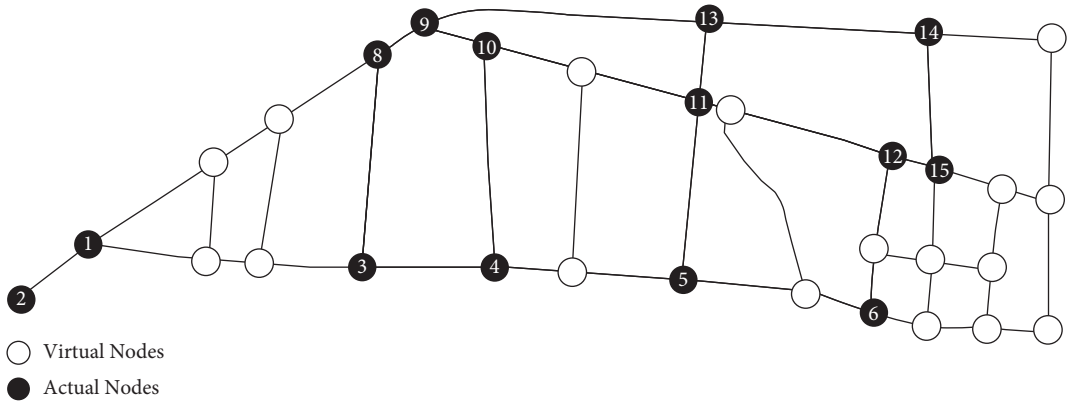


Figure 5: Simplified road network map of Yicheng County.

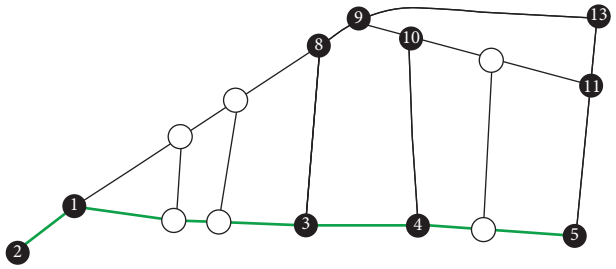


Figure 6: Short distance path a.

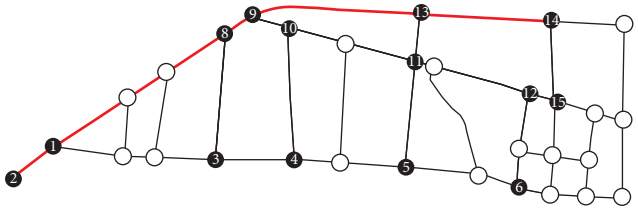


Figure 7: Medium distance path b.

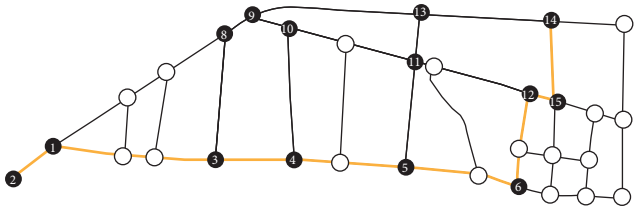


Figure 8: Long-distance path c.

inconsistent with the actual path. Besides the actual path, the most selected path is {2-1-8-9-13-11-5}. Although the distance of this path is longer, other attributes of the path have certain advantages, such as high road grade and short travel time. In addition, through the actual survey, traffic congestion occurs on the link {3-4-5} during the peak period, which is also one of the reasons for vehicles to select other paths. 16 percent of the taxi test paths chose other paths.

The medium distance path (b) missed node 9 and node 13. Node 9 is the key node of this path, where the vehicle may have decision behavior. The results show that 90 percent of the private car test paths are the same as the actual paths, as shown in Table 9. Because this path has good performance and fewer decision nodes, the alternative paths have poorer performance than the actual paths. 80 percent of the taxi test paths are consistent with the actual paths.

The long-distance path (c) missed nodes 3,4,5, and 6 as the key node of the path, the vehicle at node 1 has a greater possibility of decision making. According to the movement direction of the vehicle at node 12, the position of the previous node can be known, which can narrow down the search for alternative paths.

6.2.3. All Nodes in the Path are Missing Except for the Origin and Destination. Three key nodes are missing between the origin and destination of the short-distance path (a). Since the vehicle has node 1 as the only direction of movement after leaving node 2, this experimental condition is similar to the second set of experiments in which path (a) missed 2 nodes. The calculation results are also similar, as shown in Table 10.

Four nodes were missed between the origin and destination of the medium distance path (b). Similar to the path (a), the downstream intersection of node 2 is determined as node 1 based on the uniqueness of the direction after leaving node 2. Multiple path decisions are possible at node 1. The top 5 paths are selected based on the K-shortest path problem, and the front node position can be known according to the vehicle's moving direction at node 14, to reduce the path search range. The results are shown in Table 10.

Seven nodes were missed between the origin and destination of the long-distance path (c). Similarly, the path search range of the K-shortest path problem can be narrowed down, and the actual number of missing nodes is 5. The results are shown in Table 10. The accuracy of the results is decreased due to the accumulation of errors when the vehicle makes path decisions at nodes.

Meanwhile, the shortest path algorithm (Shortest Path, SP) is selected to compare with the path decision model in this paper, and the selected metric is the accuracy rate.

When the number of missing nodes is low, the accuracy of the results calculated by the algorithm in this paper and the shortest path algorithm is close, which indicates that

Table 7: Experimental design table.

Path number	Group 1	Group 2	Group 3
Short distance(a)	Missing node 3	Missing node 3, 4	Keeping node 2, 5
Medium distance(b)	Missing node 1	Missing node 9, 13	Keeping node 2, 14
Long distance(c)	Missing node 4	Missing node 3, 4, 5, 6	Keeping node 2, 14

Table 8: The first group of experimental results.

Path no.	Private car			Taxi		
	Sample size	Consistent number	Accuracy (%)	Sample size	Consistent number	Accuracy (%)
(A)	100	95	95	50	50	100
(B)		93	93		49	98
(C)		100	100		50	100

Table 9: The second group of experimental results.

Path no.	Private car			Taxi		
	Sample size	Consistent number	Accuracy (%)	Sample size	Consistent number	Accuracy (%)
(A)	100	87	87	50	42	84
(B)		90	93		40	80
(C)		69	77		43	86

Table 10: The third group of experimental results.

Path no.	Private car			Taxi		
	Sample size	Consistent number	Accuracy (%)	Sample size	Consistent number	Accuracy (%)
Short distance(a)	100	81	81	50	40	80
Medium distance(b)		73	73		37	74
Long distance(c)		55	69		35	70

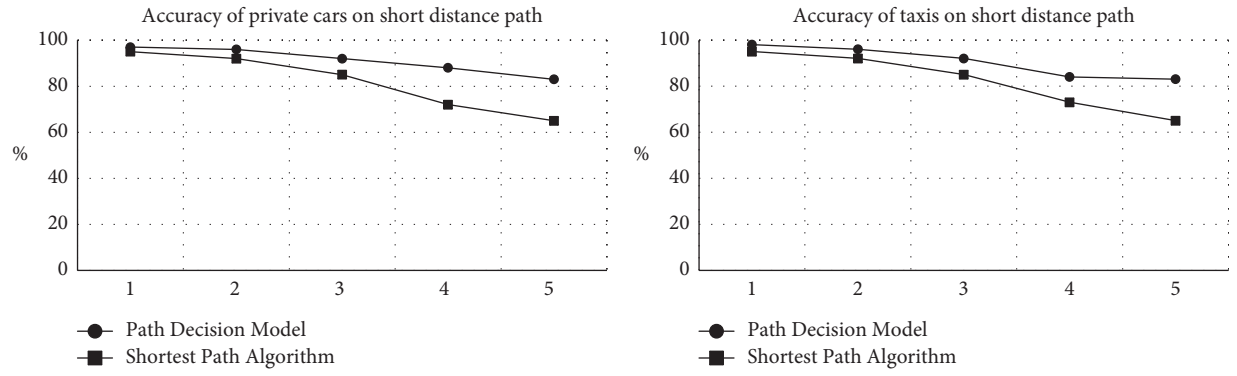


Figure 9: Accuracy of taxis and private cars on short-distance path.

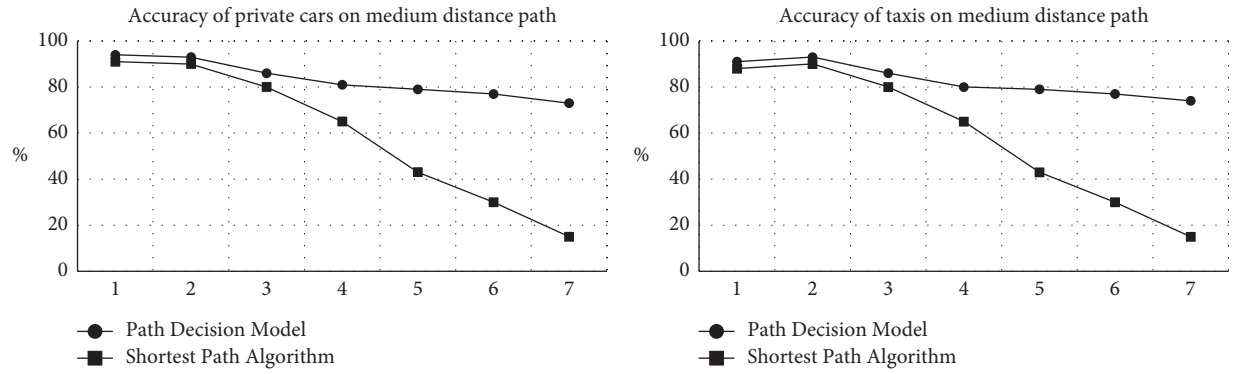


Figure 10: Accuracy of taxis and private cars on medium distance path.

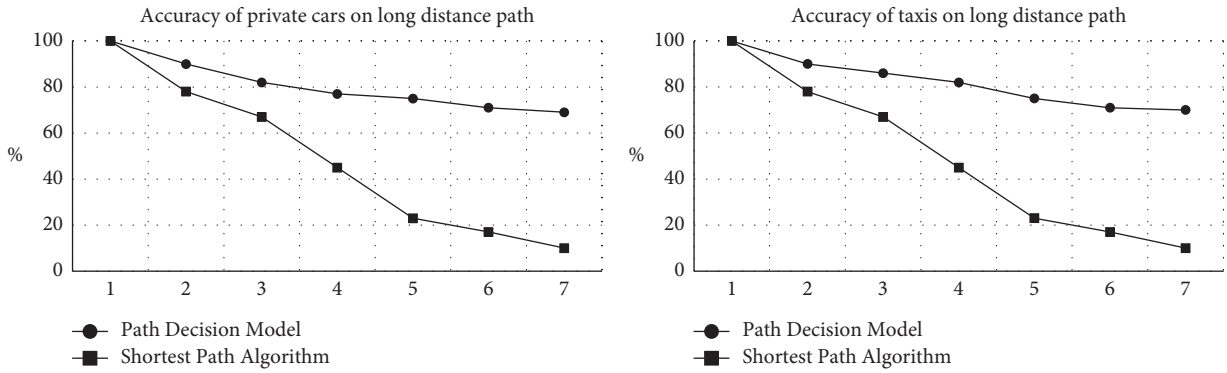


Figure 11: Accuracy of taxis and private cars on the long-distance path.

vehicles generally choose the shortest path to travel when the travel distance is short. When the number of missing nodes increases, the accuracy of both methods decreases, but the decision results of the model are significantly better than those of the shortest path algorithm.

The number of missing nodes in short-distance paths is generally less than or equal to 5. In Figure 9, it can be seen that the sensitivity of private cars and taxis to missing nodes is close in short-distance travel. The accuracy of the model is also close to the results of the shortest path algorithm.

For medium-distance paths, the performance of private cars and taxis is close, as shown in Figure 10. The accuracy of the decision results of the model in this paper has a relatively stable accuracy rate. When the number of missing nodes for medium distance and long-distance paths is over 4, the accuracy of the shortest path algorithm drops sharply, and the accuracy of the results of the model in this paper still maintains around 80%.

For the long-distance path decision, the accuracy of the proposed model decreases significantly, as shown in Figure 11. When the number of missing nodes exceeds 3, the accuracy of the shortest path algorithm starts to drop sharply, and when the number of missing nodes exceeds 4, the accuracy of the shortest path algorithm falls below 50%, but the accuracy of the proposed model is still around 80%.

In all, the accuracy of the model in this paper is better than the shortest path algorithm overall. In the short-distance missing paths, the accuracy of the model calculation results can still maintain a good level as the number of missing nodes increases. In the medium-distance missing paths, the accuracy of the model calculation results decreases slightly as the number of missing nodes increases. In the long-distance missing paths, the accuracy decreases as the number of missing nodes increases.

7. Conclusion

This paper established a vehicle path recognition model based on the road network of small cities with low coverage of AVI devices and data deficiency, considering the difference in travel path selection indicators of private cars and taxis. The method of decision indicators is defined, and the AHP-EWM combination method is used to get the indicator weights to carry out the path recognition. Firstly, the quantitative calculation methods of decision indicators are defined. The AHP-EWM combination method is used to get the indicator weights to carry out the path recognition. Through the validation of the validity and accuracy of the model, it is proved that the travel path recognized by the model is basically consistent with the actual travel paths.

However, with the increase of the number of missing nodes (when the number of missing nodes is more than 7), the accuracy of the model may further decrease.

Considering the actual deployment of the AVI system in road networks, this paper draws the following conclusions.

(1) In small cities with low coverage of AVI devices, when the missing records in the AVI data are serious, the vehicle travel paths of private cars and taxis can be identified separately according to the path recognition model in this paper. The results can effectively characterize the actual situation of vehicle travel paths, which could provide data resources for analysis of the characteristics of mixed traffic flow with CAVs.

(2) Through the analysis of the number of missing nodes in the AVI system, the following conclusions can be drawn. Within a certain area, it is not necessary to deploy the AVI devices in all sections. The model of this paper can effectively identify the travel path of vehicles, which could save system construction and equipment maintenance costs.

(3) Considering the different travel factors of taxis and private cars, the travel paths of vehicles are recognized separately and consistent with the travel behavior of vehicles in the actual road network.

Data Availability

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 study.

Acknowledgments

This study was supported by the project National Key R&D Program of China (No. 2019YFF0301401), and Beijing Natural Science Foundation (J210001).

References

[1] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, V. Loumos, and E. Kayafas, "A license plate-recognition algorithm for intelligent transportation system applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 3, pp. 377–392, 2006.

[2] C. A. Rahman, W. Badawy, and A. Radmanesh, "A real time vehicle's license plate recognition system," in *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 163–166, Miami, FL, USA, July 2003.

[3] H. A. Hegt, R. Haye, and N. A. Khan, "A high performance license plate recognition system," in *Proceedings of the IEEE International Conference on Systems*, San Diego, USA, October 1998.

[4] T. Ajanthan, P. Kamalaruban, and R. Rodrigo, "Automatic number plate recognition in low quality videos," in *Proceedings of the IEEE International Conference on Industrial & Information Systems*, Peradeniya, Sri Lanka, December 2013.

[5] Y. Lee, J. Jun, Y. Hong, and M. Jeon, "Practical license plate recognition in unconstrained surveillance systems with adversarial super-resolution," in *Proceedings of the 14th International Conference on Computer Vision Theory and Applications*, Prague, Czech Republic, 2019.

[6] Y. Asakura, E. Hato, and M. Kashiwadani, "Origin-destination matrices estimation model using automatic vehicle identification data and its application to the Han-Shin expressway network," *Transportation*, vol. 27, no. 4, pp. 419–438, 2000.

[7] Y. Yang, Z. Yuan, J. Li, and Y. Wang, "Multi-mode public transit OD prediction and scheduling model," *Advances in Transportation Studies*, vol. 3, pp. 133–146, 2018.

[8] S. Sánchez-Cambronero, E. Castillo, J. M. Menéndez, and P. Jiménez, "Dealing with error recovery in traffic flow prediction using bayesian networks based on license plate scanning data," *Journal of Transportation Engineering*, vol. 137, no. 9, pp. 615–629, 2011.

[9] Y. Yang, Z. Yuan, and R. Meng, "Exploring traffic crash occurrence mechanism towards cross-area freeways via an improved data mining approach [J]," *Journal of Transportation Engineering Part A Systems*, 2022.

[10] F. Dion and H. Rakha, "Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates," *Transportation Research Part B: Methodological*, vol. 40, no. 9, pp. 745–766, 2006.

[11] X. Li and Z. Zhai, "Travel time calculation method based on the electric police data," in *Proceedings of the 2014 9th China Intelligent Transportation Conference Paper Collection*, China Intelligent Transportation Association, Prague, Czech Republic, 2014.

[12] L. Li, W. Wang, and Y. Wang, *An Adaptive Traffic Signal Control System Based on Electric Police Data and its Working Method*, 2018.

[13] T. Xu and S. Fei, *A Method of Dynamic Green Wave Control of Traffic Signals Based on Electro-Police Data*, p. 31, 2019.

[14] Y. Yang, K. He, and Y. Wang, "Identification of dynamic traffic crash risk for cross-area freeways based on statistical and machine learning method," *Physica A: Statistical Mechanics and Its Applications*, vol. 595, p. 127083, 2022.

[15] K. Tang, T. Xu, A. Pan, and S. Li, "Signal timing and detector data-based reconstruction of vehicle trajectories on urban arterials," *Journal of Tongji University*, vol. 44, no. 10, pp. 1545–1552, 2016.

[16] J. Xiang and K. Tang, *The Method of Refactoring the Trajectory of the Main Road Vehicle Based on the Fusion of Electric Police and Fixed-Point Detector Data*, p. 24, 2018.

[17] Z. Yu, Q. Liao, and Z. He, "Vehicle trajectory reconstruction in signalized-link using vehicle identification data," *Journal of Transportation Systems Engineering and Information Technology*, vol. 19, no. 4, pp. 87–93, 2019.

[18] Y. Lin, H. Ding, Y. Liu, and X. Lin, "A study of the emission trajectory of a single vehicle based on vehicle identity detection data," *Environmental Science in China*, vol. 39, no. 12, pp. 4929–4940, 2019.

[19] Y. Yang, Z. Yuan, J. Chen, and M. Guo, "Assessment of osculating value method based on entropy weight to transportation energy conservation and emission reduction," *Environmental Engineering & Management Journal (EEMJ)*, vol. 16, no. 10, 2017.

[20] H. Zhang, X. Liu, and Y. Xu, "Delay extraction based on vehicle's trajectory reconstruction at signalized intersection," *Journal of Transportation Systems Engineering and Information Technology*, vol. 20, no. 02, pp. 237–243, 2020.

[21] W. Rao, Y. Wu, and J. Xia, "Origin-destination pattern estimation based on trajectory reconstruction using automatic license plate recognition data," *Transportation Research Part C: Emerging Technologies*, vol. 95, no. OCT, pp. 29–46, 2018.

[22] W. Li, *A Study on the Method of Vehicle Travel Path Extraction Based on License Plate Identification Data*, Guilin University of Electronic Science and Technology, Guilin, China, 2019.

[23] B. Mo, R. Li, and J. Dai, *Estimating dynamic origin–destination demand: a hybrid framework using license plate recognition data, Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 7, pp. 734–752, 2020.

[24] Y. Guo, *Research on Dynamic Ldentification and Perimeter Control Strategies of Urban Traffic Congested Regions*, Shandong University, Jinan, China, 2020.

[25] L. Wang, "Loss track reduction in vehicle travel track survey analysis," *Computer Application Research*, vol. 31, no. 01, pp. 162–165, 2014.

[26] S. Yang and H. Yu, "Vehicle travel trajectory reconstruction based on monitoring data. China intelligent transportation systems association," in *Proceedings of the Papers of the First China Intelligent Transportation Annual Conference*, China Intelligent Transportation Systems Association, 2016.

[27] B. Yang, *Motor Vehicle Track Extraction and Key Road Excavation Analysis Based on Traffic Jam Data*, Beijing Jiaotong University, Beijing, China, 2020.

[28] Y. Yang, Z. Yuan, X. Fu, Y. Wang, and D. Sun, "Optimization model of taxi fleet size based on GPS tracking data," *Sustainability*, vol. 11, no. 3, p. 731, 2019.

[29] Y. Qiu, *Analysis of the Choice Factors of Private Car Travel Path Based on RFID*, Wuhan University of Technology, Wuhan, China, 2016.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 9568723, 11 pages
https://doi.org/10.1155/2022/9568723

Research Article

Queue Length Estimation for Signalized Intersections under Partially Connected Vehicle Environment

Lu Wei,[1] **Jin-hong Li**,[1] **Li-wen Xu**,[2] **Lei Gao**,[1] **and Jian Yang**[1]

[1]*Beijing Key Lab of Urban Road Traffic Intelligent Control Technology, North China University of Technology, Beijing 100144, China*
[2]*College of Science, North China University of Technology, Beijing 100144, China*

Correspondence should be addressed to Lu Wei; wltop001@sina.com

Received 18 January 2022; Revised 21 March 2022; Accepted 21 April 2022; Published 2 May 2022

Academic Editor: Zhihong Yao

Queue length is a crucial measurement of traffic signal control at urban intersections. Conventional queue length estimation methods mostly still rely on fixed detectors. The development of connected vehicles (CV) provides massive amounts of vehicle trajectory data, and the queue length estimation based on CV data has received considerable attention in recent years. However, most existing CV-based methods require the prior knowledge of the penetration rate of CV and vehicle arrivals, but the estimation of these prior distributions has not been well studied. To address this issue, this paper proposes a cycle-based queue length estimation method under partially connected vehicle (CV) environment, with the prior vehicle arrivals being unknown. The empirical Bayes method is adopted to estimate the arrival rate by leveraging the observed queued CV information such as the number and positions. The hyperparameter estimation problem of the prior distribution is solved by the maximum likelihood estimation (MLE) method. To validate the proposed queue length estimation method, a simulation environment with partially connected vehicles is established using VISSIM and Python for data generating. The results in terms of normalized mean absolute errors (NMAE) and normalized root mean square errors (NRMSE) show that the proposed method could produce accurate and reliable estimated queue length under various CV penetration rates.

1. Introduction

Queue length is one of the most important performance measures for traffic signal control at urban road intersections. Many researchers have conducted queue length estimation based on various types of vehicle detector data [1–3]. The existing conventional methods including cumulative curve theory [4], shockwave theory [5–7], and queuing theory [8] mainly rely on fixed-location detectors (e.g., loops, magnetometer sensors, and virtual loops in cameras). The excellent data quality of these detectors could help produce well estimated results of lane-based queue length. For example, Ma et al. [9, 10] proposed a method to estimate lane-based queue length using the travel time data collected by video imaging detectors based on Ban et al. methods [11]. Two models were combined to derive the maximum queue length model. The results show that the proposed method performs a better precision compared to the existing methods in similar concepts. However, the disadvantage of these fixed-location detectors is expensive installation and maintenance costs, which leads to a low spatial coverage of urban intersections [12]. Moreover, the fixed-location detectors can only capture vehicle status at a specific point of the road section and cannot continuously track the vehicle trajectory change and running speed to support a more cost-effective, higher accuracy, and robust queue length estimation method. The installed location of these detectors is also an influence factor of the estimation results. For example, in oversaturated situation, when the queued vehicles exceed the detect zone, the total queue length usually cannot be acquired.

The rapid development of intelligent connected vehicles (CV) provides massive high-resolution vehicle trajectory data for urban traffic management and control. Different

from fixed-location detector data collecting, which requires additional equipment installation and maintained, the vehicle trajectory data is generated from the onboard GPS devices, while the moving CVs cost much less. Besides, with a larger spatial coverage, the trajectory data can capture dynamic spatial-temporal characteristics of traffic flow in the entire road network. Therefore, queue length estimation for signalized intersections based on CV data has gradually become a hot research topic and gains more and more recognition and attentions [13, 14]. Most of the existing queue length estimation methods can be classified into two categories: deterministic and stochastic methods [15]. The deterministic methods estimate queue length by reconstructing traffic flow models such as shockwaves models [16] or input-output model [17]. Based on high-precision trajectory data, Cheng et al. [18] developed a threshold-based critical point extraction algorithm to capture the changes in vehicle dynamics and then reconstruct shockwaves to estimate queue length cycle-by-cycle. The method was evaluated on both simulation and NGSIM data and the results indicated that promising outcome. Instead of using traffic volume or occupancy as the input as most queue length estimation methods based on fixed-location detectors, Ban et al. [11] estimated real-time queue length at signalized intersections using sample travel times from mobile traffic sensors. By recognizing the delay pattern changes such as nonsmoothness and discontinuities, the real-time queue length was constructed. Yin et al. [19] also proposed a queue length estimation method using low penetration vehicle trajectory data as the only input. Based on the combination of Kalman Filtering and shockwave theory, a state-space model with two state variables and the system noise was established to characterize the stochastic property of queue forming. By calculating queue-forming and queue-discharging waves, the maximum queue length can be estimated. Wang et al. [20] developed a Kalman filtering method based on the idea of input-output models to estimate the real-time queue length for an isolated intersection with CV data. The simulation results showed that the proposed method is suitable for different traffic demand levels without being affected by volume/capacity ratio. Ramezani and Geroliminis [21] integrated the traffic flow shockwave analysis and trajectory data mining techniques to extract joining and leaving critical points and then modeled queue formation process by a piecewise linear function. The significance of this study is its applicability in oversaturated conditions. These deterministic methods could produce promising estimates when the penetration rate of vehicle trajectories is sufficient high. However, due to the low penetration rates at present or even in the near future (for example, the penetration rate of vehicles with detectable trajectories from map navigation companies such as Didi, Baidu, is often less than 10% on average), these methods face the challenge of inaccurate and unstable estimation in practical applications.

Stochastic methods consider the traffic arrival as a stochastic process and estimate the distribution of queue length based on probabilistic or statistical models under sampled trajectory data environment [22]. The expectation value of the estimated distribution is then derived as total queue length. For example, Comert and Cetin [23] developed an analytical formulation based on conditional probability distributions for estimating the expected queue length and its variance from the location information of probe vehicles at signalized intersections. Various numerical results showed that only the information of the last stopped probe vehicle in the queue is sufficient for queue length estimation. The relationship between penetration rate and estimation accuracy is also analyzed in this study. To further improve the accuracy and generalization of the model estimation, Comert and Cetin incorporated the stop-line detector data [24] and the time when the last probe vehicle joined the queue [25] into the queue length estimation model. Subsequently, they further proposed a series of estimators for primary parameters of the queue length estimation model, such as the penetration rate and the arrival rate under the Poisson distribution assumption [26]. However, these models assume that vehicle arrivals follow uniform Poisson distribution, which is an indispensable a priori knowledge. In order to compensate the weakness of sparse vehicle trajectory data, Tan et al. [12] fuse license plate recognition (LRP) data and probe vehicle data for lane level queue length estimation after fully analyzing the advantages of the two data sources. Historical trajectory data and LPR data are used to calibrate two-dimensional probability density distribution for discharge headway and stop-line crossing time for queued and nonqueued vehicles based on kernel density estimation. Then, the lane-based queue length is obtained by using the Bayesian theory. To extend the queue length estimation model, Zhao et al. [27] studied the estimation of two parameters: the probe vehicle penetration and the queue length distribution using historical trajectory data. A maximum likelihood estimator is adopted, and EM algorithm is applied to solve the problem iteratively. Validation results on simulated vehicle trajectory data show that the proposed method could estimate the parameters accurately.

In summary, for stochastic methods, a priori knowledge such as penetration rate, queue length distribution, and arrival pattern is quite important. Noticeably, though some researchers have conducted the estimation of penetration rate and queue length distribution, the arrival pattern has not been well studied. Motivated by this issue and the queue length estimation model proposed by Comert and Cetin [23], this paper proposed a queue length estimation method with unknown arrival pattern under partially connected vehicle environment. The main contributions of this study include the following:

(1) By utilizing sampled CV trajectory information in the queue at signalized intersection, a method of estimating the arrival rate of all vehicles was proposed based on empirical Bayes.

(2) To estimate the hyperparameters of the prior Gamma distribution, we constructed the marginal distribution of CV arrivals and then applied maximum likelihood estimator to solve the hyperparameters estimation. Finally, the estimated queue length is derived as the expectation of conditional queue length distribution.

(3) In order to validate the proposed method, the secondary development of VISSIM simulation software based on Python is conducted to establish a partially CV data environment with adjustable penetration rates.

The rest of this paper is organized as follows. Section 2 describes the problem and provides the notations used in this paper. Section 3 introduces the queue length estimation model and presents the empirical Bayesian method for arrival rate estimation and maximum likelihood estimation of hyperparameters. In Section 4, a partially connected vehicles simulation scenario is established, and experiments analysis is also conducted in VISSIM simulation software. Conclusions and future works are summarized in Section 5.

2. Problem Statement and Notation

At urban signalized intersections, vehicles traveling on the road start to form a queue behind the stop line when the traffic signal turns red. Figure 1 shows the snapshot of the queued vehicles on an approach to the intersection. The blue rectangles indicate the connected vehicles whose position, speed, and other information can be collected in real-time. The white rectangles indicate normal vehicles, and their states cannot be collected without other additional traffic sensors. The primary objective of this research is estimating the total queue length based on the observable connected vehicle information (e.g., the number and stopped positions of the connected vehicles in the queue). Some assumptions are made to simplify the discussion. First, the intersection traffic is assumed to be undersaturated and vehicle arrivals follow Poisson distribution. Second, the signal timing of the intersection is known. Third, at least one CV exists in the queue to make sure that the proposed method can be applied.

For a better illustration the proposed method, a summary of the notation is provided in Table 1.

Take the scenario described in Figure 1 as example, the total queue length $N = 7$, the number of CVs in the queue $M = 2$, and the position of the last position of CV stopped in the queue $L = 5$.

3. Methodology

3.1. Queue Estimation Model. The process of queue length estimation under partially connected vehicle data environment can be concluded in the following steps:

(1) Deriving the number of stopped CVs in the queue based on whether the speed is less than the stopping speed threshold.

(2) Determining the position of the last stopped CV in the queue, that is, the farthest stopped CV from the stop line.

(3) Calculating the distance from the last stopped CV to the intersection stop line and then deriving an initial estimated queue length based on the standard spacing headway. The initial queue length also can be treated as the lower bound of the total queue length of all vehicles.

(4) Estimating the number of vehicles that may exist after the last stopped CV based on other prior knowledge, such as queue length and (or) vehicle arrival distribution, the penetration rate of CV, etc.

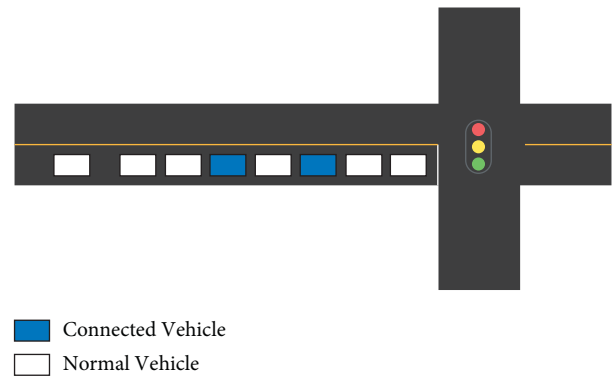


Figure 1: Illustration of queued vehicles at an intersection.

Table 1: Summary of notation.

Notation	Description
N	Total queue length
M	Number of CVs in the queue
L	Last position of stopped CVs in the queue
p	Penetration rate of CV
λ	Arrival rate of all vehicles
η	Arrival rate of CVs
t_R	Duration of red signal

Based on the observable information including the positions and number of connected vehicles, Wang et al. [14] introduce two cycle-by-cycle queue length estimators from the perspective of the probability theory: maximum likelihood estimator and expected queue length conditional on the observed partial queue, and the latter follows the model proposed by Comert and Cetin [23]. In this research, we continue to use it as the basic model for queue length estimation.

Assume that the probability of a vehicle being a connected vehicle follows Bernoulli distribution, and the probability of success is denoted as p, as the penetration rate of CV. Given the total queue length of all vehicles, denote $N = n$, and the number of CVs in the queue M follows binomial distribution and can be written as

$$P(M = m|N = n) = \binom{n}{m} p^m (1-p)^{n-m}, \quad (1)$$

where p is the penetration rate of connected vehicles, and $0 \le p \le 1$. m represents the number of CV in the queue, and the value range is $m = 0, 1, 2, \ldots, n$.

In addition, given $N = n$ and $M = m$, the probability distribution for the position of last stopped CV can be derived from the number of possible combinations [23].

$$P(L = l|N = n, M = m) = \frac{\binom{l-1}{m-1}}{\binom{n}{m}}, \quad (2)$$

where l denotes the position of last stopped CV and $l = m, m+1, m+2, \ldots, n$.

According to Bayes' rule, the total queue length distribution condition on the position of the last stopped CV and the number of CVs in the queue can be written as follows:

$$P(N = n|L = l, M = m) = \frac{P(L = l|M = m, N = n)P(M = m|N = n)P(N = n)}{P(L = l, M = m)}. \quad (3)$$

Substituting equations (1) and (2) into equation (3), and applying the Total Probability Law, equation (3) can be calculated as

$$P(N = n|L = l, M = m) = \frac{(1-p)^n P(N = n)}{\sum_{k=l}^{\infty}(1-p)^k P(N = k)}. \quad (4)$$

Equation (4) implies that the total queue length distribution only depends on the position of the last stopped CV and the penetration rate.

In this paper, the penetration rate is assumed to be known. Actually, in many engineering applications, the penetration rate usually can be obtained from the analysis of the connected vehicle trajectory data and other fixed detection data. For example, Figure 2 shows the volume comparison between license plate recognition (LPR) data collected by camera and trajectory data provided by map navigation company in an area of Lianyungang city in January 2019. The penetration rate can be estimated as the ratio of sampled trajectory volume and camera volume. In the case of Figure 2, the result of penetration rate calculation is about 0.061, which is consistent with the penetration rate provided by the map navigation data company. Therefore, this method is verified with high practicality and accuracy.

The difference between this paper and previous studies is that the arrival rate of total vehicles is unknown and needs to be estimated based on the observable CV information. Vehicle arrivals are usually assumed to be Poisson distribution with a rate of λ in undersaturated scenarios [23]. The probability density function of Poisson arrivals can be expressed as

$$P(N = n) = \frac{(\lambda t_R)^n e^{-\lambda t_R}}{n!}, \lambda > 0, n = 0, 1, 2, \ldots, \quad (5)$$

where λ is the average arrival rate, and t_R is the duration of the red signal. For the conditional probability distribution queue length estimator expressed by equation (4), with the assumption of the penetration rate is known, the rest work of queue length estimation is estimating vehicle arrival rate based on observable data.

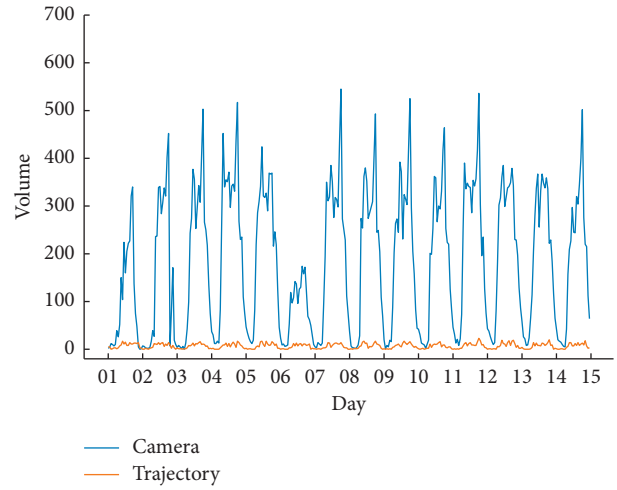


Figure 2: Volume comparison between camera and trajectory samples.

3.2. Estimation of Arrival Rate. In a partially connected vehicle data environment, only the information of connected vehicles can be observed, such as the positions and number of stopped vehicles in the queue. We further assume that the penetration rate of CV remains constant within a certain time period and space scale; in other words, the mean of the penetration rate distribution is used as estimated penetration rate. After that, under the assumption of Poisson arrivals, the probability of CV arrivals can be written as the following equation [26]:

$$P(M = m) = \frac{(\lambda p t_R)^m e^{-\lambda p t_R}}{m!}, \quad (6)$$

where p denotes the penetration rate, $0 \le p \le 1$, and m denotes the CV arrivals in t_R. Let $\eta = \lambda p t_R$, and the conditional probability density function of equation (6) can be expressed as

$$P(M = m|\eta) = \frac{\eta^m e^{-\eta}}{m!}, m = 0, 1, 2, \ldots, \quad (7)$$

where η is the unknown arrival rate of CV and needs to be estimated.

Conventional approaches for arrival rate estimating include simply computing the mean of observable samples or the Maximum Likelihood estimation. However, these approaches perform effectively only when the arrival rate sample is sufficient, and the time interval is large [28]. However, for cycle-by-cycle queue length estimation, the sample size is usually small. Also, due to cycle constraints, the time interval is not large. For this case, the Bayesian estimation approach could be used by considering that the process would benefit from prior information about the probability distribution of η. However, the assumption that the prior distribution of η is fully known is rather strong and may not be applicable in realistic scenarios. To overcome this limitation, this paper adopts the Empirical Bayes approach in which only the class of the prior is known. Compared to the classical approach, the Empirical Bayes approach improves

generality and flexibility to some extent. According to the knowledge of probability and statistics, it is known that Gamma distribution is a conjugate prior over the rate of the Poisson distributions [16]. Based on this, we assume that $\eta \sim \text{Gamma}(\nu, \sigma)$ and its probability density function can be written as

$$g(\eta; \nu, \sigma) = \frac{e^{-\eta/\sigma}\eta^{\nu-1}}{\sigma^{\nu}\Gamma(\nu)}, \eta > 0, \tag{8}$$

where ν and σ are the hyperparameters to be estimated.

The hyperparameters can be estimated via a Maximum Likelihood procedure since the arrival of the connected vehicles during the red light is observable. Therefore, the Empirical Bayes estimation of arrival rate can be carried out in the following two steps: constructing the marginal distribution and hyperparameters estimation.

3.2.1. Constructing the Marginal Distribution. First, based on the properties of the Gamma function, for any positive number c and d, we have

$$\int_0^{\infty} \eta^{c-1} e^{-\eta/d} d\eta = d^c \Gamma(c). \tag{9}$$

By using equations (6) and (7), the marginal distribution can be derived as

$$f_{\nu,\sigma}(m) = \int_0^{\infty} P(M = m|\eta) g(\eta) \mathrm{d}\eta = \frac{\int_0^{\infty} \eta^{\nu+m-1} e^{-\eta(\sigma+1)/\sigma} \mathrm{d}\eta}{\sigma^{\nu}\Gamma(\nu)m!}. \tag{10}$$

Combining the conclusion of equation (9), equation (10) can be rewritten as

$$f_{\nu,\sigma}(m) = \frac{\sigma^m \Gamma(\nu + m)}{(\sigma + 1)^{\nu+m}\Gamma(\nu)m!}. \tag{11}$$

Equation (11) shows that the hyperparameters of Gamma distribution can also be found in the marginal distribution of the connected vehicle arrivals. Therefore, the parameter estimation problem can be transformed into a Maximum Likelihood estimation procedure according to the observable connected vehicles arrivals.

3.2.2. Hyperparameter Estimation via MLE. We suppose that the arrivals of connected vehicles during red light are independent of each other. It is denoted the observed arrivals in K signal cycles as $\{m_1, m_{2,\ldots,} m_K\}$, and due to independency, the likelihood function is the product of the marginal distributions of all sampled arrivals

$$L(m_1, \ldots, m_K|\nu, \sigma) = \prod_{i=1}^{K} f_{\nu,\sigma}(m_i). \tag{12}$$

The above expression for the total probability is actually quite a pain to solve. Therefore, we work with the simpler log-likelihood instead of the original likelihood.

$$L(m_1, \ldots, m_K|\nu, \sigma) = \sum_{i=1}^{K} \ln f_{\nu,\sigma}(m_i). \tag{13}$$

Thus, the estimator $\widehat{\nu}$ of ν and $\widehat{\sigma}$ of σ can be solved by the following optimization problem:

$$\widehat{\nu}, \widehat{\sigma} = \arg\max_{\nu,\sigma} \sum_{i=1}^{K} \ln f_{\nu,\sigma}(m_i). \tag{14}$$

Take the partial derivatives with respect to ν and σ, and the following equations can be derived:

$$\left\{ \frac{\partial L}{\partial \nu} = \sum_m \sum_{i=0}^{m-1} \frac{1}{\nu + i} - \log(\sigma + 1), \frac{\partial L}{\partial \sigma} = \sum_m \left(\frac{m}{\sigma} + \frac{\nu + m}{\sigma + 1} \right). \right. \tag{15}$$

This optimization problem cannot be solved in closed-form from the above equations and requires a numerical optimization method. Therefore, this paper adopts the gradient ascent method to solve it. The gradient update equation is as follows:

$$\begin{cases} \nu_j \coloneqq \nu_j + \alpha \dfrac{\partial L}{\partial \nu}, \\ \\ \sigma_j \coloneqq \sigma_j + \alpha \dfrac{\partial L}{\partial \sigma}, \end{cases} \tag{16}$$

where α is the learning rate. After the optimal estimators of ν and σ are obtained, we can further derive the posterior distribution of η. Then, the mean of the posterior distribution $\widehat{\eta}$ is used to solve the following equation:

$$\widehat{\eta} = \lambda p t_R. \tag{17}$$

Based on equation (17), the estimated total arrival rate $\widehat{\lambda}$ can be derived as $\widehat{\lambda} = \widehat{\eta}/pt_R$.

With the derived $\widehat{\lambda}$, the arrival distribution of total vehicles can be obtained. With the known penetration rate of connected vehicles $P(N = n)$, substituting it into equation (4) and then the conditional probability distribution of total queue length can be calculated, and the mean of the distribution can be derived as the total queue length.

4. Simulation Experiments

Queue data generated from an isolated simulated intersection in VISSIM is used to validate the proposed methodology. VISSIM is a microscopic road traffic simulator based on the individual behavior of vehicles provided by PTV in German and has been widely used for various scenarios by traffic engineers in practice as well as by researchers for developments related to urban road traffic analysis, management, and control. Despite VISSIM has offered a friendly user graphical interface (GUI) for simply and quickly establishing dedicated traffic simulation networks, for the case in this paper, the GUI is not satisfied. For this end, we use Python programming language to collect the objects, manipulate the attributes of internal objects, and specify simulation parameters dynamically. VISSIM COM

interface defines a hierarchical model in which the attributes and functions of the internal objects provided by GUI can be manipulated by external programming. Yao and Jiang [29] developed an integrated connected vehicle simulation platform of VISSIM and Python to deal with the problems of difficult implementation of complex control algorithms and time-consuming. Inspired by Yao's works, a simulation evaluation platform based on Python for partially connected vehicle data environment is established in the following steps.

4.1. Simulation Settings. A pseudoisolated intersection with fixed signal timing in VISSIM is used to test the queue length estimator in the proposed method. The intersection has four approaches: and the western approach is selected to conduct on queue length estimation. In order to collect the total queue length of all vehicles as a ground truth baseline, a queue counter is set on the western approach. The random seed is set to a single fixed value (42) to ensure a trustworthy comparison and reproducibility.

Modeling connected and autonomous vehicles using VISSIM internally or externally has been stated by PTV Group [30]. Since the focus of this study is to estimate queue length from data generated by connected vehicles with different penetration rates, simultaneously considering the simplicity and convenience, this paper implements the internal way, which is to modify the VISSIM default vehicle types, traffic compositions, and driving behavior parameters [31].

To analyze the impact of different CV penetration rates, this paper establishes simulation scenarios with different penetration rates. First, a new vehicle type should be defined with the name set to CV, and then, adjust its attributes such as speed and acceleration distribution as needed. If a vehicle is set as a CV, its real-time locations and speeds can be acquired. By adding the CV vehicle type to the traffic composition and setting the percentage of each vehicle type, the scenarios with different CV penetration rates can be established. The specific process is shown in Figure 3.

4.2. Data Collection. As mentioned above, the VISSIM COM interface provides a strict object hierarchy for dynamic accessing. In this paper, we use the Python language to collect the real-time moving information of CV, including simulation timestamp, vehicle ID, position, and speed. Meanwhile, in order to evaluate the estimation accuracy of the proposed method, the real total queue length is also recorded by the queue counter cycle by cycle. The data interaction process is shown in Figure 4.

A sample of the CV moving data and the real total queue length recorded from VISSIM is shown in Table 2, where the allqueued field represents the actual total queue length, which is used as the benchmark of the proposed queue estimation method. Cvqueued represents the number of CVs stopped in the queue, cvpos is the positions of the CVs, and if there is more than one CV, it is separated by "-" to indicate their different stopping positions, and last_cv_pos denotes the position of last stopped CV, which is located farthest from the stop line. Among of them, cvqueued, cvpos, and last_cv_pos are the corresponding observable information for real-world scenarios.

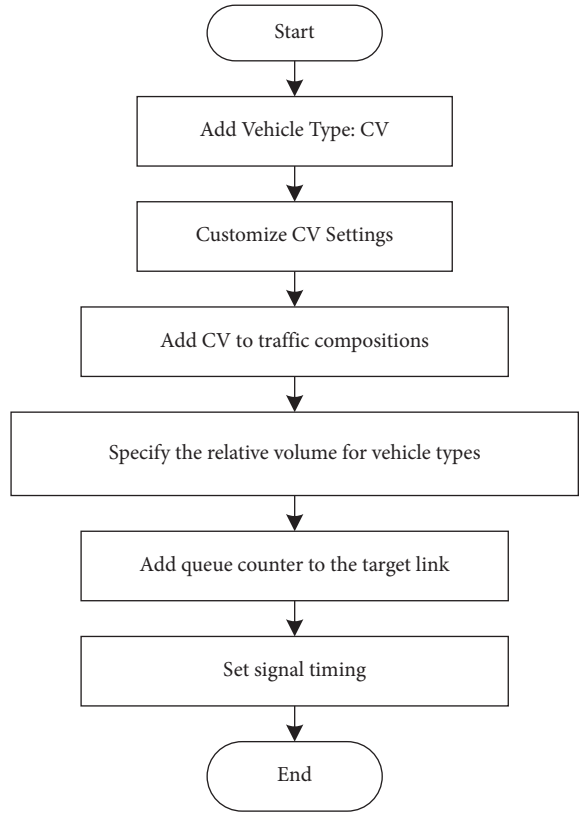


Figure 3: The process of simulation scenarios establishment.

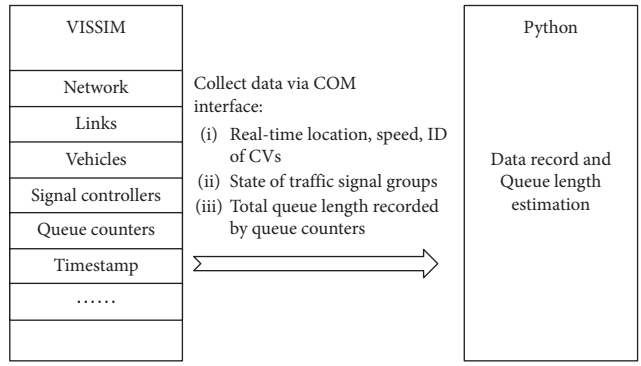


Figure 4: Collecting CV data via COM interface.

Table 2: Samples of recorded simulation data.

Time	Sigstate	Allqueued	Cvqueued	Cvpos	Last_cv_pos
20.0	Green	0	0	0	0
80.0	Green	4	3	1-2 – 4	4
140.0	Green	3	1	2	2
200.0	Green	4	2	1 – 4	4
260.0	Green	8	5	1-2 – 5-6-7	7

4.3. Results and Discussion. The estimation results can be obtained by applying the method proposed in this paper on the CV data recorded via VISSIM COM interface by Python.

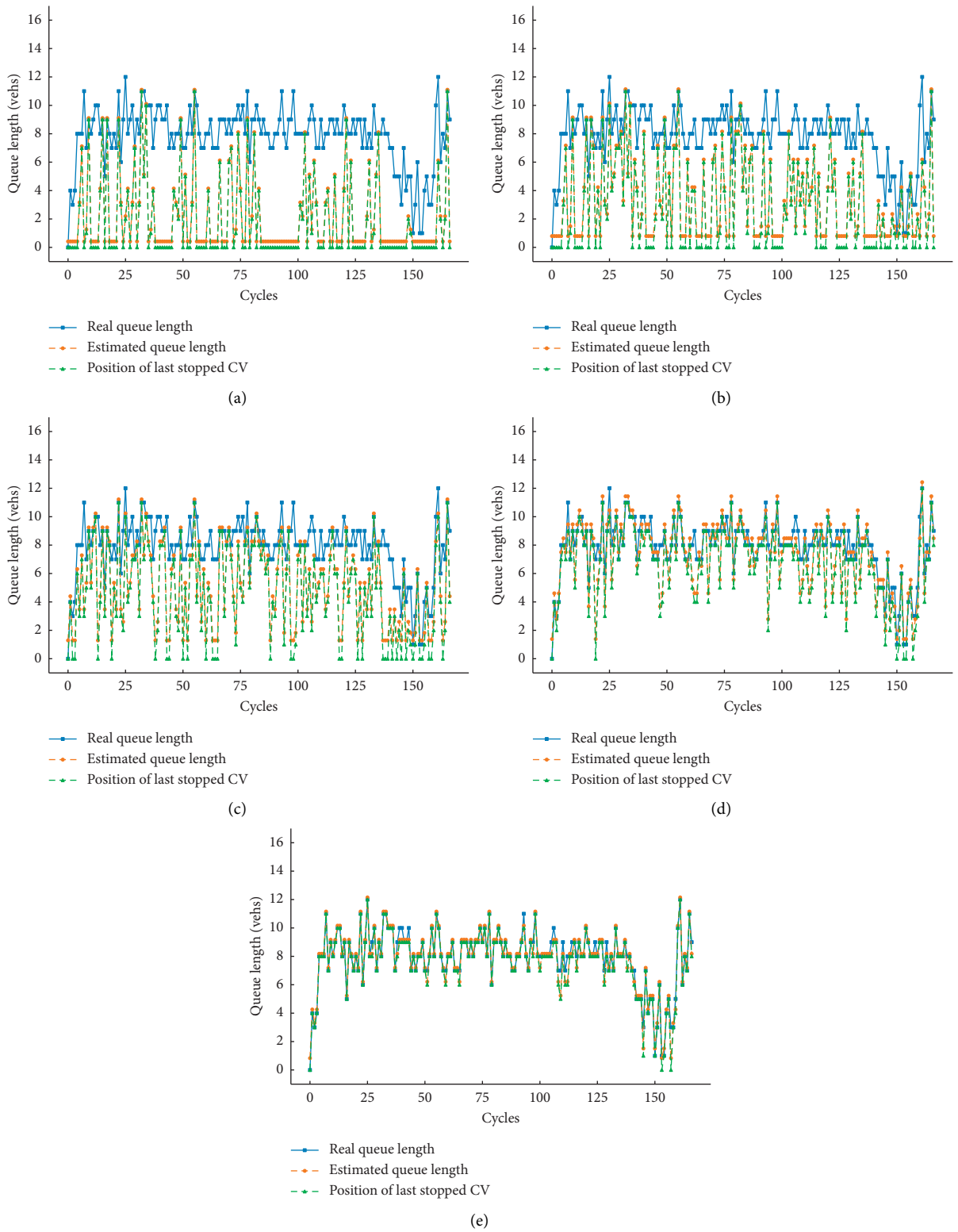


Figure 5: The estimated results with different CV penetration rates. (a) Penetration rate of 5%. (b) Penetration rate of 10%. (c) Penetration rate of 20%. (d) Penetration rate of 50%. (e) Penetration rate of 80%.

For sensitivity analysis, penetration rates of 5%, 10%, 20%, 50%, and 80% are chosen. The comparison results of the estimated queue length, the position of the last stopped CV, and the actual total queue length are presented in Figure 5. For each plot, the blue solid line with star markers represents the ground truth queue length collected by queue counter set in VISSIM, the orange dotted line with dot markers represents the estimated queue

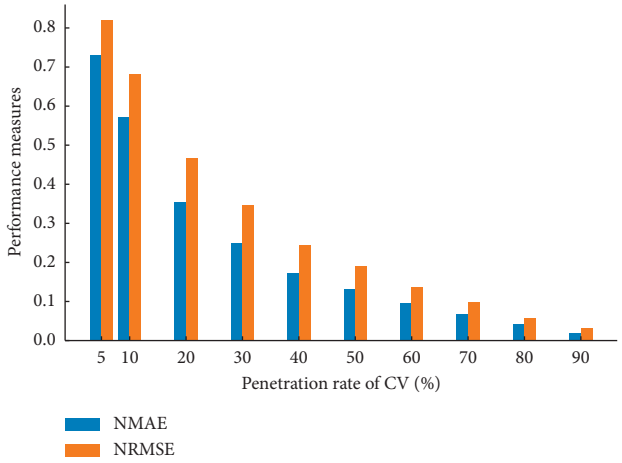


Figure 6: Performance measures under various penetration rates.

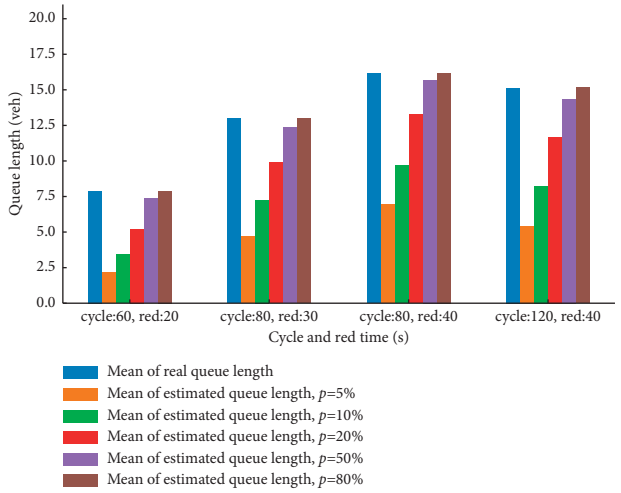


Figure 8: Estimation results under different cycle and red times.

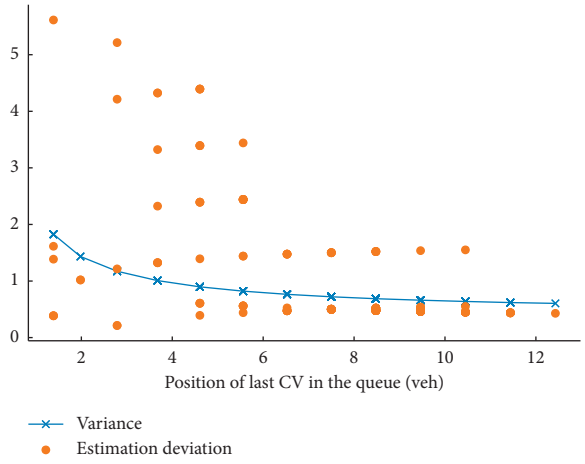


Figure 7: The relationship between the last CV in the queue and estimation error.

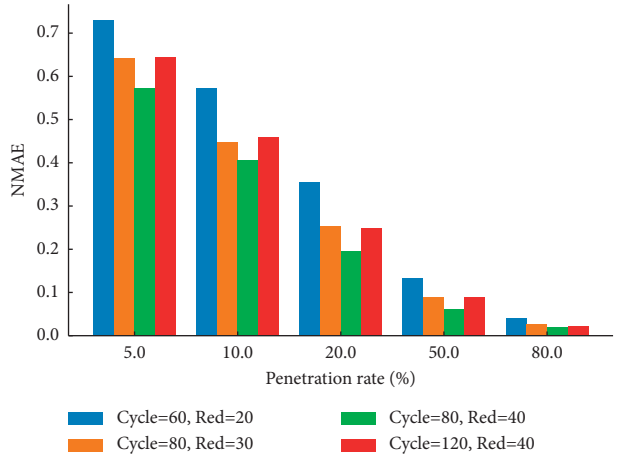


Figure 9: NMAE analysis for different cycle and red times.

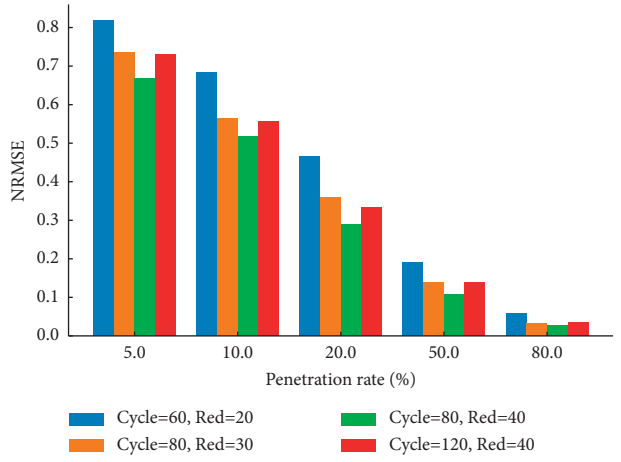


Figure 10: NRMSE analysis for different cycle and red times.

length by using the method of this study, and the green dotted line with triangle markers represents the position of the last stopped CV, which can be seen as naïve estimation of the queue length or the lower bound of the real queue length.

As shown in Figure 5, compared with other penetration rates, the estimated queue length deviates far from the ground truth queue length in the case of low penetration rates (e.g., 5% penetration rate). The reason is that when the penetration rate is low, the observed information of CV is rather limited, which results in a rough estimation of the queue length. However, when the penetration rate increases, the queue length estimated by the proposed method is approaching the real queue length. It also can be seen that the position of the last stopped CV, which represents a naïve estimation of the ground truth queue length, also performs more accurately with the increasing penetration rates.

In order to further demonstrate the results, two indicators, that is, the normalized mean absolute error (NMAE) and the normalized root mean square error (NRMSE), are used as the metrics to measure the accuracy of the proposed method. The formulas of NMAE and NRMSE are given by equations (18) and (19):

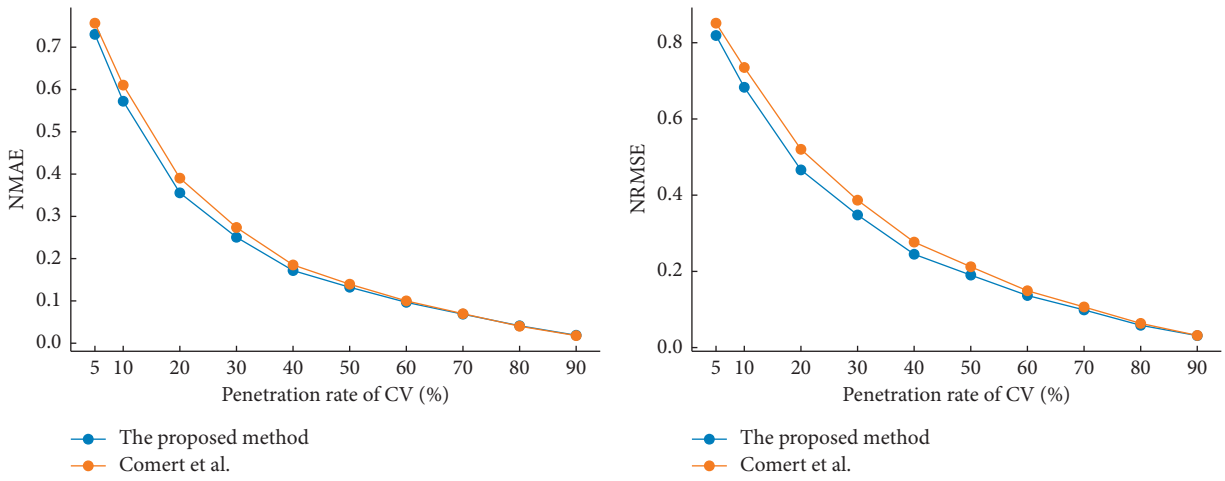


Figure 11: Comparison of different methods. (a) NMAE. (b) NRMSE.

$$\mathrm{NMAE}(n,\hat{n}) = \frac{\mathrm{MAE}(n,\hat{n})}{\mathrm{mean}(|n|)} = \frac{\sum_{i=1}^{C}|n_i-\hat{n}_i|}{\sum_{i=1}^{C}|n_i|}, \quad (18)$$

$$\mathrm{NRMSE}(n,\hat{n}) = \frac{\mathrm{RMSE}(n,\hat{n})}{\mathrm{mean}(|n|)} = \frac{\sqrt{1/C\sum_{i=1}^{C}(n_i-\hat{n}_i)^2}}{1/C\sum_{i=1}^{C}|n_i|}, \quad (19)$$

where C denotes the number of cycles, n_i denotes the ground truth of the queue length of cycle i, and $\hat{n}_i$ denotes the corresponding estimator.

As Figure 6 shows, when the penetration rate gets higher, the performance measures in terms of estimation accuracy are better, and both NMAE and NRMSE indicators are decreasing.

When the penetration rate is determined, this paper also attempts to analyze the different stop positions of the last CV in the queue with the deviation and the estimation results. As shown in Figure 7, the deviation becomes smaller when the position of the last CV stopped in the queue appearing farther from the stop line.

Figure 7 implies that when applying the proposed method in engineering, different confidence levels can be assigned according to the positions of the last stopped CV in the queue. Thus, the reliability of the estimation method could be improved practically.

To further investigate the impact of cycle time and red time on queue length estimation, four different cycle and red time scenarios are designed as follows:

(1) Cycle time: 60 s, red time: 20 s

(2) Cycle time: 80 s, red time: 30 s

(3) Cycle time: 80 s, red time: 40 s

(4) Cycle time: 120 s, red time: 40 s

The red time indicated the red signal duration of the phase for the investigated western approach. For each scenario, the penetrations of 5%, 10%, 20%, 50%, and 80% are selected, respectively. The mean value of total real queue length and the estimated results under different penetrations are calculated as Figure 8 shows. From Figure 8, it could be easily observed that the real queue length increases with the cycle time and red time becoming larger. Note that the highest real queue length appears in cycle 80 and red time 40, because the red ratio is largest among the four scenarios, which causes more stopped vehicles and delays. In addition, as the penetration rate increases, the mean value of estimated queue length gets closer to the real queue length.

NMAE and NRMSE are also calculated and analyzed to examine the performance of the proposed method for different cycle and times under the penetration rates of 5%, 10%, 20%, 50%, and 80%. As shown in Figures 9 and 10, it is obvious that the NMAE and NRMSE benefit from the increasing penetration rates for each cycle and red time scenario. Besides, for the same penetration, the scenario of cycle 80 and red time 40 performs better. Note that the mean value of real queue length for this scenario is also the highest shown in Figure 8. Therefore, the main reason may be that when the number of the queue vehicles increases, the number of CVs also increases, which leads to more CVs being detected during a signal cycle. The last stopped CV in the queue also tends to be closer to the actual end of the entire queue.

To elaborate the effectiveness of the proposed method, we compare the performance measures of the proposed method to the existing model proposed by Comert and Cetin [23], and the arrival rate estimation is implemented from Comert et al. [26]. Figures 11(a) and 11(b) show the estimation results in terms of NMAE and NRMSE provided by the two methods under different penetration rates, respectively.

From the comparison results shown in Figure 11, the proposed method could achieve more accurate estimates than the baseline method. The means of NMAE and NRMSE of the proposed method are 0.243 and 0.307, respectively, while those of the baseline method are 0.258 and 0.333. The proposed method performs better especially when the

penetration rate is not sufficient high. The potential reason for that can be attributed to the empirical Bayesian method used in this paper, which can obtain the prior distribution of arrival rate by using historical observed data compared with the baseline method, which is based on the current observation data only. However, when the penetration rates are higher than 70%, the NMAEs and NRMSEs calculated by the two methods become approximately equivalent.

5. Conclusion

Cycle-by-cycle queue length estimation methods under partially connected vehicle data environments require the knowledge of the penetration rate and the arrival rate in the existing studies. However, the estimation of these prior parameters has not been adequately and extensively studied. This paper proposes an empirical Bayes estimation method to estimate the arrival rate of total vehicles (assuming Poisson arrivals). However, we do not have any observable samples of the total queued vehicles. Fortunately, with the help of connected vehicle technology, the information such as positions and numbers of CV in the queue can be acquired in real-time. Therefore, based on the conclusions of previous research, the estimation problem can be transformed into CVs arrival rate estimation. For the hyperparameters in the prior distribution of empirical Bayes, the Maximum Likelihood estimation is applied to solve this. In order to validate the proposed methods, a partially connected vehicle environment is established by using the VISSIM COM interface and Python. The validation results have shown that the proposed method could estimate queue length accurately. The impacts of the penetration rates and the positions of the last stopped CV in the queue are also studied. We also explored the effect of cycle time and red time on the achieved outcomes and provided sensitivity analysis.

By accurately estimating queue length using the proposed method, the signal timing design and evaluation at signalized intersections could be improved. For instance, based on the estimated queue length of each direction, the optimal duration of green phase can be derived to reduce the total delay of an isolated intersection. For traffic signal coordination along arterial, queue length could help determine the offset and green wave band width. Besides, accurate queue length can also prevent queue spillovers effectively, which would lead to more serious congestion at intersections under oversaturated traffic conditions. However, some issues should be considered such as the existing GPS data that can not reach lane level when applying the proposed method into real-world engineering cases.

There are a few directions that have not been explored in this paper and may be conducted in the future work. First, the dynamic characteristics and imbalance in spatial-temporal can be explored in the queue length estimation model. Second, the nonhomogeneous Poisson process can model the vehicle arrival more accurately. Third, this paper only focuses on queue length estimation in undersaturated traffic conditions, the overflow queue estimation that considering correlations will be investigated in future work.

Data Availability

The data generated by the VISSIM simulation and recorded by Python via the VISSIM COM interface are used in this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was financially supported by the project "National Key R&D Program of China" under grant no. 2018YFB1601003. The authors are grateful to Lianyungang Jari Electronics Company, Ltd. and Baidu Map team for providing data used in this study.

References

[1] C. Tan, J. Yao, K. Tang, and J. Sun, "Cycle-based queue length estimation for signalized intersections using sparse vehicle trajectory data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 91–106, 2021.

[2] X. Zhan, R. Li, and S. V. Ukkusuri, "Lane-based real-time queue length estimation using license plate recognition data," *Transportation Research Part C: Emerging Technologies*, vol. 57, pp. 85–102, 2015.

[3] S. Lee, S. C. Wong, and Y. C. Li, "Real-time estimation of lane-based queue lengths at isolated signalized junctions," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 1–17, 2015.

[4] D. Heidemann, "Queue length and delay distributions at traffic signals," *Transportation Research Part B: Methodological*, vol. 28, no. 5, pp. 377–389, 1994.

[5] "On kinematic waves II. A theory of traffic flow on long crowded roads," *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.*, vol. 229, no. 1178, pp. 317–345, 1955.

[6] C. An, Y.-J. Wu, J. Xia, and W. Huang, "Real-time queue length estimation using event-based advance detector data," *Journal of Intelligent Transportation Systems*, vol. 22, no. 4, pp. 277–290, 2018.

[7] Z. Wang, Q. Cai, B. Wu, L. Zheng, and Y. Wang, "Shockwave-based queue estimation approach for undersaturated and oversaturated signalized intersections using multi-source detection data," *Journal of Intelligent Transportation Systems*, vol. 21, no. 3, pp. 167–178, 2017.

[8] A. C. Soh, M. H. Marhaban, M. Khalid, and R. Yusof, "Modelling and optimisation of a traffic intersection based on queue theory and markov decision control methods," in *Proceedings of the First Asia International Conference on Modelling & Simulation (AMS'07)*, pp. 478–483, Phyket, Thailand, March 2007.

[9] D. Ma, X. Luo, S. Jin, W. Guo, and D. Wang, "Estimating maximum queue length for traffic lane groups using travel times from video-imaging data," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 3, pp. 123–134, 2018.

[10] D. Ma, X. Luo, S. Jin, D. Wang, W. Guo, and F. Wang, "Lane-based saturation degree estimation for signalized intersections using travel time data," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 3, pp. 136–148, 2017.

[11] X. Ban, P. Hao, Z. Sun, and Z. Sun, "Real time queue length estimation for signalized intersections using travel times from

mobile sensors," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1133–1156, 2011.

[12] C. Tan, L. Liu, H. Wu, Y. Cao, and K. Tang, "Fuzing license plate recognition data and vehicle trajectory data for lane-based queue length estimation at signalized intersections," *Journal of Intelligent Transportation Systems*, vol. 24, no. 5, pp. 449–466, 2020.

[13] G. Comert and M. Cetin, "Queue length estimation from connected vehicles with range measurement sensors at traffic signals," *Applied Mathematical Modelling*, vol. 99, pp. 418–434, 2021.

[14] J. Wang, S. Jiang, Y. Qiu, Y. Zhang, J. Ying, and Y. Du, "Traffic signal optimization under connected-vehicle environment: an overview," *Journal of Advanced Transportation*, vol. 2021, Article ID 3584569, 16 pages, 2021.

[15] P. Hao, X. Ban, D. Guo, and Q. Ji, "Cycle-by-cycle intersection queue length distribution estimation using sample travel times," *Transportation Research Part B: Methodological*, vol. 68, pp. 185–204, 2014.

[16] M. Cetin, "Estimating queue dynamics at signalized intersections from probe vehicle data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2315, no. 1, pp. 164–172, 2012.

[17] A. Sharma, D. M. Bullock, and J. A. Bonneson, "Input-output and hybrid techniques for real-time prediction of delay and maximum queue length at signalized intersections," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2035, no. 1, pp. 69–80, 2007.

[18] Y. Cheng, X. Qin, J. Jin, and B. Ran, "An exploratory shockwave approach to estimating queue length using probe trajectories," *Journal of Intelligent Transportation Systems*, vol. 16, no. 1, pp. 12–23, 2012.

[19] J. Yin, J. Sun, and K. Tang, "A kalman filter-based queue length estimation method with low-penetration mobile sensor data at signalized intersections," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 45, pp. 253–264, 2018.

[20] Y. Wang, Z. Yao, Y. Cheng, Y. Jiang, and B. Ran, "Kalman filtering method for real-time queue length estimation in a connected vehicle environment," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2675, no. 10, pp. 578–589, 2021.

[21] M. Ramezani and N. Geroliminis, "Queue profile estimation in congested urban networks with probe data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 30, no. 6, pp. 414–432, 2015.

[22] Y. Zhao, S. Shen, and H. X. Liu, "A hidden Markov model for the estimation of correlated queues in probe vehicle environments," *Transportation Research Part C: Emerging Technologies*, vol. 128, p. 103128, 2021.

[23] G. Comert and M. Cetin, "Queue length estimation from probe vehicle location and the impacts of sample size," *European Journal of Operational Research*, vol. 197, no. 1, pp. 196–202, 2009.

[24] G. Comert, "Effect of stop line detection in queue length estimation at traffic signals from probe vehicles data," *European Journal of Operational Research*, vol. 226, no. 1, pp. 67–76, 2013.

[25] G. Comert, "Simple analytical models for estimating the queue lengths from probe vehicles at traffic signals," *Transportation Research Part B: Methodological*, vol. 55, pp. 59–74, 2013.

[26] G. Comert, "Queue length estimation from probe vehicles at isolated intersections: estimators for primary parameters," *European Journal of Operational Research*, vol. 252, no. 2, pp. 502–521, 2016.

[27] Y. Zhao, W. Wong, J. Zheng, and H. X. Liu, "Maximum likelihood estimation of probe vehicle penetration rates and queue length distributions from probe vehicle data," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–9, 2021.

[28] A. Coluccia and F. Ricciato, "Improved estimation of instantaneous arrival rates via Empirical Bayes," in *Proceedings of the 2014 13th Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET)*, pp. 211–216, Piran, Slovenia, Jun 2014.

[29] Z. Yao and Y. Jiang, "Integrated connected vehicle simulation platform of Vissim and python," *Intelligent Traffic Signal Control in a Connected Vehicle Environment*, vol. 35, pp. 143–146+405, 2018.

[30] A. Evanson, "Connected autonomous vehicle (CAV) simulation using PTV Vissim," in *Proceedings of the 2017 Winter Simulation Conference (WSC)*, p. 4420, Las Vegas, NV, USA, December 2017.

[31] S. He, X. Guo, F. Ding, Y. Qi, and T. Chen, "Freeway traffic speed estimation of mixed traffic using data from connected and autonomous vehicles with a low penetration rate," *Journal of Advanced Transportation*, vol. 2020, pp. 1–13, 2020.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 4257865, 16 pages
https://doi.org/10.1155/2022/4257865

Research Article

Predicting Freeway Traffic Crash Severity Using XGBoost-Bayesian Network Model with Consideration of Features Interaction

Yang Yang,[1,2] **Kun Wang**,[3] **Zhenzhou Yuan**,[4] **and Dan Liu**[5]

[1]*School of Transportation Science and Engineering, Beihang University, Beijing 100191, China*
[2]*Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety Control, Beihang University, Beijing 100191, China*
[3]*Beijing CenNavi Technologies Co., Ltd., Beijing 100038, China*
[4]*School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*
[5]*Freight Mobility Research Institute, Department of Civil Environmental and Geomatics Engineering, Florida Atlantic University, Boca Raton, FL 33431, USA*

Correspondence should be addressed to Yang Yang; 14114222@bjtu.edu.cn

Received 12 March 2022; Accepted 11 April 2022; Published 30 April 2022

Academic Editor: Gen Li

In the field of freeway traffic safety research, there is an increasing focus in studies on how to reduce the frequency and severity of traffic crashes. Although many studies divide factors into "human-vehicle-road-environment" and other dimensions to construct models whichshowthe characteristic patterns of each factor's influence on crash severity, there is still a lack of research on the interaction effect of road and environment characteristics on the severity of a freeway traffic crash. This research aims to explore the influence of road and environmental factors on the severity of a freeway traffic crash and establish a prediction model towards freeway traffic crash severity. Firstly, the obtained historical traffic crash data variables were screened, and 11 influencing factors were summarized from the perspective of road and environment, and the related variables were discretized. Furthermore, the XGBoost (eXtreme Gradient Boosting) model was established, and the SHAP (SHapley Additive exPlanation) value was introduced to interpret the XGBoost model; the importance ranking of the influence degree of each feature towards the target variables and the visualization of the global influence of each feature towards the target variables were both obtained. Then, the Bayesian network-based freeway traffic crash severity prediction model was constructed via the selected variables and their values, and the learning and prediction accuracy of the model were verified. Finally, based on the data of the case study, the prediction model was applied to predict the crash severity considering the interaction effect of various factors in road and environment dimensions. The results show that the characteristic variables of road side protection facility type (RSP), road section type (LAN), central isolation facility (CIF), lighting condition (LIG), and crash occurrence time (TIM) have significant effects on the traffic crash prediction model; the prediction performance of the model considering the interaction of road and environment is better than that of the model considering the influence of single condition; the prediction accuracy of XGBoost-Bayesian Network Model proposed in this research can reach 89.05%. The identification and prediction of traffic crash risk is a prerequisite for safety improvement, and the model proposed and results obtained in this research can provide a theoretical basis for related departments in freeway safety management.

1. Introduction

With the continuous improvement of motorization degree, freeway travel demand is increasing, and the freeway has become an important part of the world's highway transportation system. Due to the driving characteristics of the freeway, once a traffic crash occurs to the vehicle operating on a freeway, it usually has a consequence with a serious

casualty or property loss [1]; reducing the severity of traffic crashes is of great significance to improve the safety of the freeway operating. Simultaneously, the occurrence of traffic crashes is sudden and accidental, which is quite difficult to control. Based on this background, exploring the crash precursors of freeway and predicting the severity of freeway traffic crashes has become a popular topic in this field.

Regarding the severity of a roadway traffic crash: studies on influencing factors of crash severity have always been carried out from four dimensions of "human-vehicle-road-environment," among which the studies on environmental factors, weather and road, account for a large proportion. In order to better understand the influence of weather on traffic crash severity, Satoshi et al. [2] developed a traffic crash severity assessment model based on the ordered Probit model, which took into account traffic characteristics, road conditions, environment, and factors related to multiple vehicles, single vehicles, and bicycles. Lee et al. [3] applied the structural equation model to analyze the relationship between weather conditions and the severity of traffic crashes, and the results showed that the severity of crashes was correlated with the factors of road, traffic, environment, human, rain and water depth on the road, and other factors. Amin et al. [4] studied the impact of climate change on dangerous road crashes related to weather and analyzed the spatiotemporal relationship between weather-related explanatory variables and crash severity index by using negative binomial regression and Poisson regression models; the results showed that the surface weather conditions and weather had a strong positive correlation with the crash severity index, while the road surface form characteristics had a negative correlation with the crash severity index. For the real-time freeway traffic crash analysis, Sun et al. [5] proposed a method of crash risk assessment based on traffic safety state division and quantitatively analyzed the influence of different traffic conditions on freeway crash risk. However, traffic conditions alone may be found to constitute an elevated crash risk, and without an additional behavioral factor to help differentiate the relative risk, the predicted crash risk shall remain low, giving rise to a high proportion of false positive predictions [6]. These above studies analyzed the relationship between subjective or objective factors and the severity of the traffic crash, explained the correlation of the influencing factors by building models, and analyzed whether the influencing factors of each dimension were significantly related to the severity of a crash.

Regarding roadway traffic crash severity prediction research: predicting the severity of road traffic crashes is an important part in the study of roadway traffic safety. The model with high prediction accuracy and accurate prediction results can provide insight for relevant departments to effectively reduce the severity of traffic crashes. In addition to considering different influencing factors, scholars' research interests mainly focus on using various prediction models to figure out the severity of road traffic crashes, including regression prediction model, Decision Tree, Neural Network, and Bayesian network prediction, etc. [7–10]. Celika et al. [11] analyzed the severity of traffic crashes based on multiple Logit models; in this study, the severity of traffic crashes was divided into fatal accident, injury accident, and noninjury accident, where the results showed that the factors of driver's education level, road grade, whether there is a crosswalk, crash time, and weather all had a certain impact on the severity of traffic crashes. Jiang et al. [12] adopted the zero-expansion ordered Probit model to study the influence of traditional influencing factors such as kerb and speed limit changes on the severity of single-vehicle collision crash. Shaheed et al. [13] applied the mixed Logit model to the construction of the crash severity prediction model and analyzed whether various factors had a significant influence on the crash severity. The study was based on motorcycle crash data in Iowa State, USA; the crash severity was classified into five categories: fatal, major injury, minor injury, possible or unknown, and PDO (property damage only). Lou [14] analyzed the risk states of freeway under different time conditions and proposed a traffic crash severity prediction method based on the mixed Logit model. Alkheder et al. [15] used three data mining models, including decision tree, linear support vector machine, and Bayesian network model, to analyze the risk factors related to traffic crash severity, and the performance of the application model showed that Bayesian network predicted variables more accurately than other models. In general, the study of Bayesian network has been introduced into the field of traffic crash severity prediction, and compared with regression models, decision tree, and other models, Bayesian network prediction is typically more accurate.

Based on the above analysis, there are many existing studies which analyzed the influence of weather or road on crash severity or comprehensively analyzed the influence of road and environmental factors, but there is a lack of research considering the influence mechanism of interaction between road dimension factors and environmental dimension factors on traffic crash severity prediction. In order to achieve a more accurate prediction for the severity of freeway traffic crashes, identify the main risk factors of freeway traffic crashes, and reduce the freeway crash severity, this research takes freeway operation as the research object. In this research, XGBoost model is constructed and SHAP value is introduced to interpret the model results. The global impact and importance ranking of each characteristic factor on the severity of the crash are obtained through the interpretation results of SHAP values, and the impact degree of a single variable on the severity of the crash and the interaction effect between independent variables are obtained. The prediction model of freeway traffic crash severity is determined by the inference model based on clip-tree propagation algorithm; finally, the traffic crash severity is predicted considering the interaction effect between road and environment.

2. Data Process

2.1. Data Source. In this research, traffic crashes on freeways in Hebei Province, China, in 2018 are taken as the research object. The dataset used in this research includes a total of 567 pieces of crash data, covering 35 attribute variables. Table 1 describes the summary of attribute variables in the raw dataset.

Table 1: Summary of the attribute variables of the raw dataset.

Order	Attribute name	Order	Attribute name
1	Administrative division of province	19	Scene form
2	Prefectural administrative divisions	20	Weather
3	Road no.	21	Visibility
4	Road name	22	Pavement condition
5	Milepost	23	Surface condition
6	Meters	24	Traffic control mode
7	Week	25	Light condition
8	Month	26	Roadway type
9	Year	27	Administrative class
10	Occurred date	28	Geography
11	Occurred hour	29	Road alignment
12	Total deaths in 24 H	30	Road section type
13	Total injuries in 24 H	31	Physical road isolation
14	Location	32	Pavement structure
15	Direct property loss	33	Type of central isolation facility
16	Crash type	34	Type of roadside protection facilities
17	Crash cause	35	Number of involved
18	Crash form		

2.2. Data Variables Selection. Most of the studies involving the severity prediction of freeway traffic crashes basically include the selection of influencing factors from four aspects: human-vehicle-road-environment. However, some studies only selected some of the factors from some dimensions to predict and analyze the crash severity. For example, Ma et al. [16] selected 12 candidate objective independent variables from the aspects of time, road, and traffic operation environment in order to avoid excessive attention to human factors while ignoring the impact of objective factors on traffic crashes and established a cumulative logistic model to analyze their impact on the severity of traffic crashes. Through the goodness of fit and prediction accuracy test of the model, the model fitting effect performed better, and the modeling conclusion had a certain practical reference significance. Hence, the establishment of a prediction model based on the influence of objective factors also has a certain research significance. Consequently, based on the sample dataset, this research screens out the influencing factors of road and environment, mainly exploring the influence of road and environment factors on crash severity. Table 2 gives the influencing variables of freeway traffic crash severity in this research. Since the selected factor variables include attribute variables and continuous variables, they need to be discretized to meet the requirements of modeling.

3. Screening Influence Factors Based on XGBoost

3.1. The Fundamentals of XGBoost. Boosting library XGBoost [17, 18] is a boosting library developed by Chen in 2016, which is an improvement of the gradient lifting algorithm. Gradient lifting algorithm does not only have a gradient lifting tree; thus, the weak estimators in XGBoost algorithm can also choose linear models such as logistic regression and linear regression in addition to the tree model. However, XGBoost technically uses the tree model for integration. In many machine learning algorithms, the loss function is used to measure the generalization ability of the model, that is, the prediction accuracy of the model on unknown data, and the calculation speed of the model cannot be measured. XGBoost algorithm achieves a balance between model performance and model operation speed; it combines the prediction accuracy of the model with engineering capability. In addition to the loss function, model complexity is also introduced into the objective function of XGBoost algorithm to measure the operation efficiency of the algorithm [19, 20]. Its objective function is as follows:

$$\mathrm{Obj}^{(t)} = \sum_{t-1}^{n} l\left(y_i, y_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + \text{constant}. \quad (1)$$

For the freeway traffic crash data set, where i represents the i-th crash sample in the dataset, the first item represents the loss function of the model, which measures the difference between the real value y_i and the predicted value of the crash severity. The loss function can be selected according to the predicted demand. The second term is the regularization term, which represents the complexity of the model. It is represented by some transformation Ω of the tree model, which means that the complexity of the tree model is measured from the structure of the tree. The approximate target is obtained by Taylor expansion [19, 20], which is

$$f(x + \Delta x) \simeq f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)\Delta x^2. \quad (2)$$

Define expression:

$$g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right), h_i = \partial^2_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}^{(t-1)}\right). \quad (3)$$

The second-order Taylor expansion of the loss function can be obtained as

$$\mathrm{Obj}^{(t)} \simeq \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) + \text{constant}. \quad (4)$$

Table 2: Variables applied in this research and their values.

Variable categories	Variables	Symbols	Values
Control variable	Severity	SEV	0: Property loss 1: Injuries 2: Fatal
Road factors	Location	LOC	0: Carriageway 1: Noncarriageway
	Surface condition	RDC	0: Dry 1: Wet
	Road alignment	RDA	0: Straight 1: Not straight
	Roadway type	LAN	0: Normal section 1: Special section 2: Complex node
	Central isolation facilities	CIF	0: W-beam guardrail 1: Isolation pier 2: Concrete fence 3: Metal guardrails 4: Green belt
	Roadside protection facilities	RSP	0: W-beam guardrail 1: Isolation pier 2: Concrete fence 3: Metal guardrails 4: Green belt 5: Border tree 6: Others
	Geography	GEO	0: Plain 1: Hills 2: Mountainous
Environmental factor	Visibility	VIS	0: <50 m 1: 50~100 m 2: 100~200 m 3: >200 m
	Light condition	LIG	0: Daylight 1: Dusk/dawn 2: Lighting at night 3: No lighting at night
	Crash time	TIM	0: 0–6 am 1: 6–12 am 2: 12–18 (pm) 3: 18–24 (pm)

The second partial derivative Taylor expansion in XGBoost has a strong advantage, which makes the gradient descent faster and more accurate. The second derivative form of the function as an independent variable is obtained according to Taylor expansion; the specific form of the loss function does not have to be selected, but it only depends on the value of the input data for calculation, which increases the compatibility of the model.

3.2. Model Interpretation Based on SHAP. Compared with conventional linear models, XGBoost has improved the prediction accuracy. However, it loses the interpretability that linear models have; thus, it is almost considered as a black box model. Lundberg et al. [21] proposed the method of SHAP value in order to interpret various models (classification and regression) including the black box model.

SHAP is an additive interpretation model inspired by Shapley values. In the prediction model, each prediction sample will generate the corresponding prediction value, and the value assigned to each feature in the sample is SHAP value. Assume that the ith sample is x_i, the jth feature of the ith sample is x_{ij}, the predicted value of the model for the ith sample is y_i, and the baseline of the whole model is $\overline{y}$ (basically it is the mean of the target variable of all samples); hence the SHAP value follows the following formula [22]:

$$y_i = \overline{y} + f(x_i, 1) + f(x_i, 2) + \cdots + f(x_i, k), \tag{5}$$

where $f(x_i, 1)$ denotes the SHAP value of the first feature of the ith sample, that is, the contribution value of $f(x_i, 1)$ to the predicted value y_i. When $f(x_i, 1) > 0$, it means that this feature improves the predicted results; that is, it has a positive effect; otherwise, it has a negative effect, which means that this feature reduces the predicted value.

"Feature importance" was previously used to explain the XGBoost model. This method is applied to measure the importance of each feature in the dataset to the model and determine which features have a greater impact on the final model, but the relationship between features and predicted

results cannot be determined. Not only can SHAP values reflect the influence degree of each feature in each sample on the model, but it is able to show the positiveness and negativity of the influence, which is the advantage of using SHAP value to interpret the model [22].

3.3. XGBoost Hyperparameter Optimization. In this research, Grid Search CV is introduced to search the optimal parameter value of the model, among which, "Learning_rate" (update shrinkage step in the learning process), "max_depth" (maximum depth of tree), "*n*_estimators" (control number of weak estimators), and "subsample" (random sampling ratio), which have great influence on the model performance, are selected for optimization searching, where "learning_rate" typically ranges from 0.0 to 0.05 and the "search step" is 0.01. "Max_depth" typically ranges from 3 to 11 and the "search step" is 1. "*N*_estimators" typically ranges from 100 to 400 and the "search step" is 50. The candidate values of "subsample" are 0.7, 0.8, and 0.9.

Firstly, the data is imported, and the regression prediction model is established, and any candidate values are selected as the initial values of parameters. After the model is established, the parameter optimal solution of the model is searched by using the hyperparameter meshing optimization function, and the model optimization result is learning_rate = 0.04, max_depth = 3, *n*_estimators = 100, subsample = 0.9. The optimal parameters are reinput into the model to obtain the optimal solution model, and then the SHAP interpreter is built. The interpretation of SHAP on the model results is visualized, and the global influence and importance ranking diagram of each factor variable on the severity of the traffic crash can be output.

3.4. XGBoost Hyperparameter Optimization and Model Interpretation Analysis. First, the data is imported, and the regression prediction model is established. Then, any candidate value is selected as the initial parameter value. After the model is established, the optimal parameter solution of the model is searched by using the hyperparametric gridding optimization function. The optimization results of the model are learning_rate = 0.04, max_depth = 3, *n*_estimators = 100, subsample = 0.9. Reinput the optimal parameters into the model to obtain the optimal solution model, and then build the SHAP interpreter to visualize the interpretation of SHAP to the model results. The global impact and importance ranking of each factor variable on the traffic crash severity can be output as shown in Figure 1.

Figure 1(a) shows the contribution degree of feature factors to the severity degree of the traffic crash. Here, the importance of features is sorted according to the mean value of the absolute value of the impact degree of feature on the target variable. It can be seen from the figure that the variable RSP has the greatest global impact on the severity of the crash; that is, it has the greatest contribution to the prediction result of the freeway traffic crash severity. Additionally, LAN, CIF, LIG, and TIM also have great impact on the prediction.

Figure 1(b) shows the tendency of influencing factors to different traffic crash severity. Each row in the figure denotes a feature factor, the horizontal coordinate represents the SHAP value, each dot represents a sample, and the color of each dot denotes the value of the corresponding feature. From blue to red, the value of the feature itself is getting bigger. Smooth color transitions of SHAP values can be observed horizontally as changes in the values of the feature itself and the influence of the feature gradually alter the output of the model.

Figure 2 specifically reflects the influence of each characteristic independent variable value on the traffic crash severity of the dependent variable. The abscissa is the value of the independent variable, and the ordinate is the SHAP value, which is the contribution value of the feature to the prediction result of the crash severity. There are deputy ordinates in Figures 2(a)–2(d), which indicates that there are collinearity problems between CIF and RSP, LAN and RSP, RDA and GEO, ROC and WEA. The figure reflects the interaction effect between the two collinear independent variables and their contribution degree to the prediction results of crash severity, indicating that the model used in this research can deal with the collinearity problem well, and the collinear independent variables will not affect the fitting of the model. For example, as shown in Figure 2(a), the SHAP value of CIF does not change significantly under various CIF values, and the VALUE of CIF is large. That is, when the central isolation is metal guardrails and green belts, roadside protection facilities are generally metal guardrails, green belts, and roadside trees.

Figures 2(e)–2(h) reflect the relationship between a single variable and the severity of the traffic crash.

Figure 2(e) illustrates that when the lighting condition is daytime, dusk, or dawn, SHAP value is basically less than 0, which may reduce the value of crash severity; that is, slight severity crashes are more likely to occur. When the lighting condition is night with lighting or night without lighting, SHAP value is basically greater than 0, which has a positive promoting effect towards the occurrence of serious crashes.

Figure 2(f) illustrates that the proportion of crashes that occurred in the carriageway is large; concurrently, the severity of crashes is evenly distributed in both carriageway and noncarriageway.

Figure 2(g) shows that when the value of TIM is 3, that is, during 18:00~24:00, the corresponding SHAP value is basically concentrated on the side greater than 0, which has a positive promoting effect on the occurrence of serious crashes.

Figure 2(h) indicates that most of the crashes occurred when the visibility value is 3; that is, when the visibility is above 200 meters, SHAP value greater than 0 has a positive effect on serious crashes.

Figure 3 illustrates the interaction of multiple variables. There is an interaction effect between RSP and LAN. However, the interaction effect is not a simple linear relationship. When the red dots are mainly concentrated on the side with SHAP values less than 0 and the blue dots are mainly concentrated on the side with SHAP values greater than 0, RSP is negatively correlated with LAN. There is a

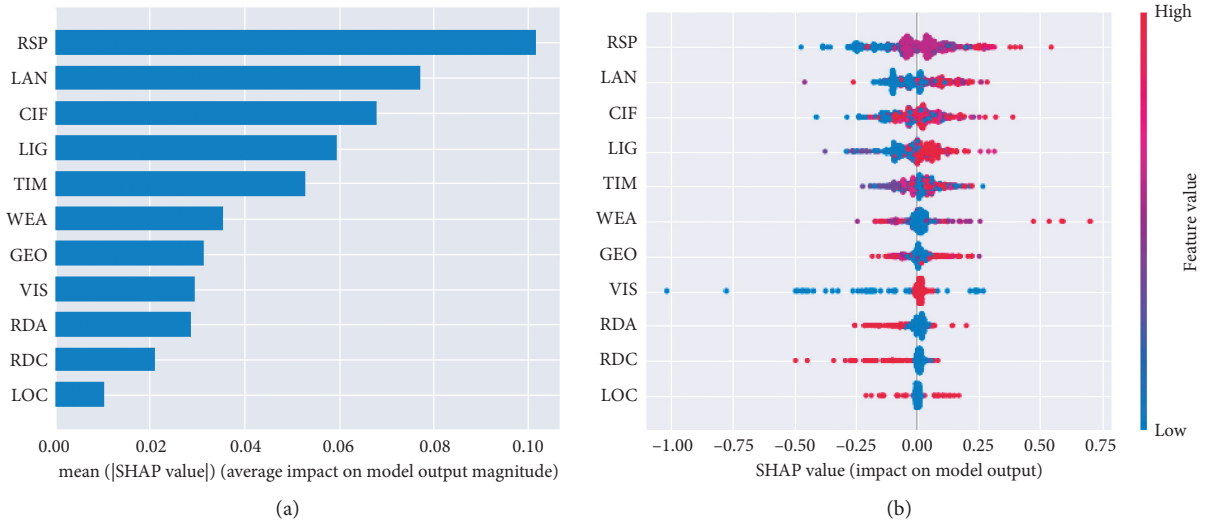


Figure 1: The relationship between the average importance of feature factors and the values. (a) Average impact on the magnitude of model output. (b) Impact on model output.

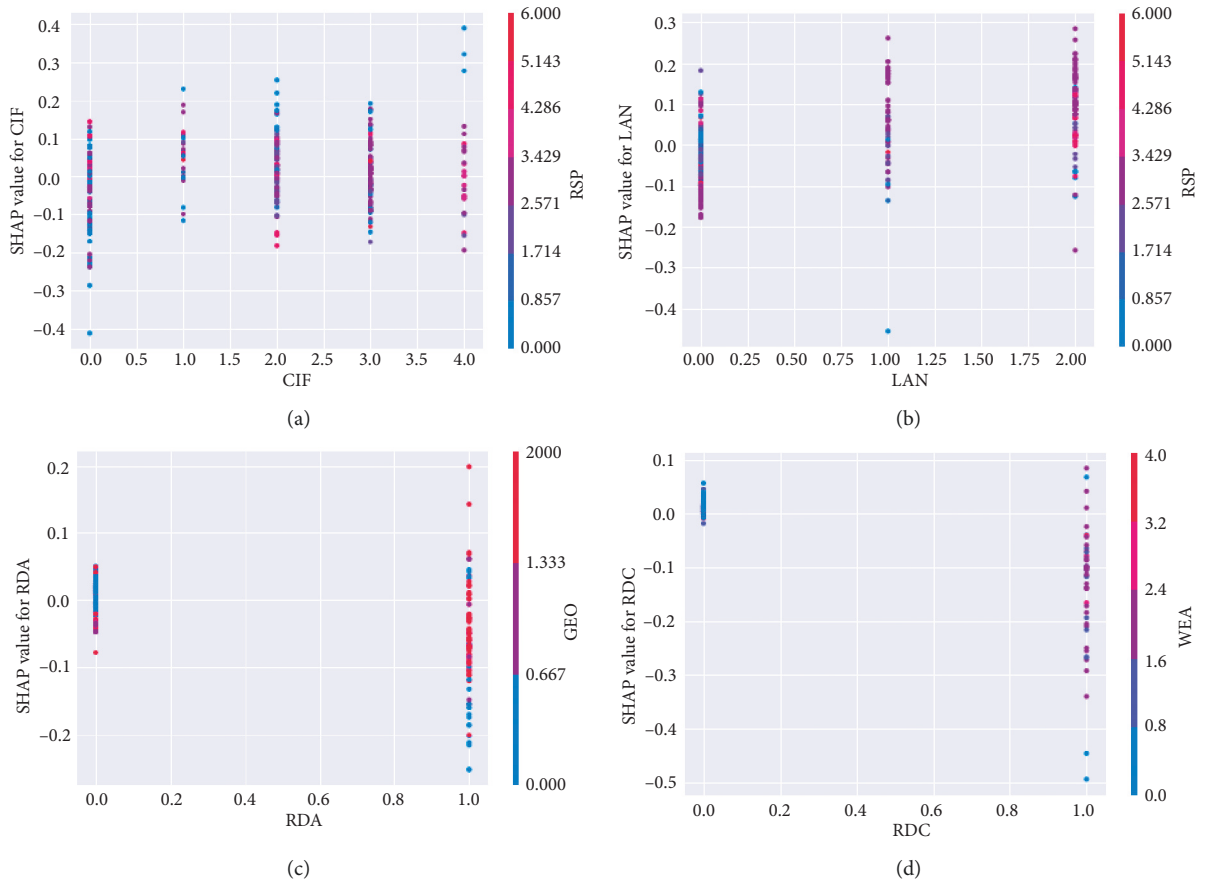


Figure 2: Continued.

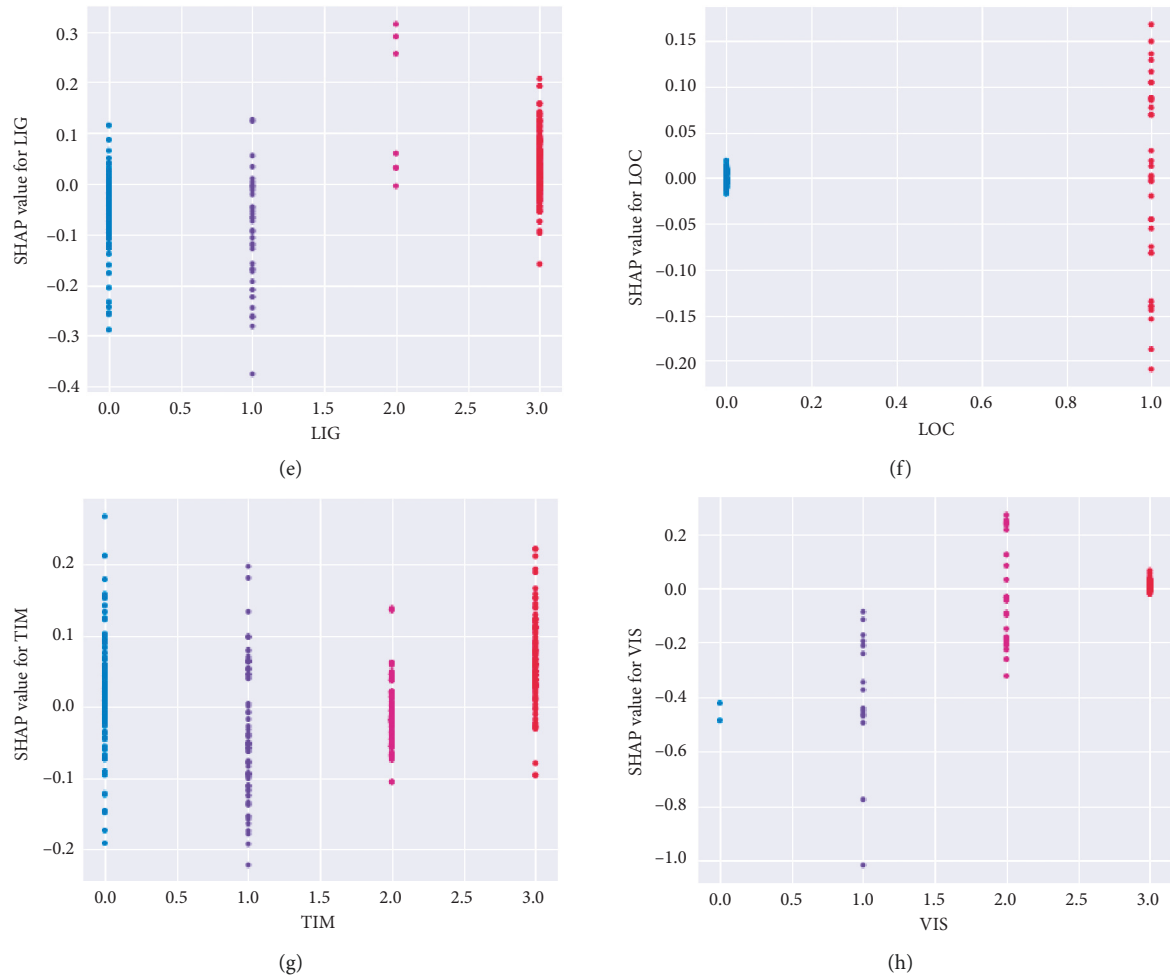


Figure 2: The dependence of independent variables on the crash severity. (a) CIF, (b) LAN, (c) RDA, (d) RDC, (e) LIG, (f) LOC, (g) TIM, (h) VIS.

certain threshold, and when the interaction between RSP and LAN exceeds this threshold, the correlation gradually shows positive. LAN and CIF have an interaction effect and weak positive correlation. The nonlinear interaction between LIG and RSP, and LIG and WEA also exists. There are interaction effects between GEO and CIF and a weak positive correlation between GEO and RDA. There are interaction effects between GEO and RDC and a weak negative correlation between LAN and RDC. RDA and LAN, and RDA and TIM have interactive effects and positive correlations. There is almost no interaction between the other two variables, and their sample distribution is random.

4. Modeling of Bayesian Network

Bayesian network has a strong probability theory foundation and strong explanatory ability through graph theory; hence some uncertain problems in many fields can be solved via Bayesian methods, such as fault diagnosis, pattern recognition, and accident prediction, etc. [23]. The content of this research is traffic crash severity prediction; since the occurrence of traffic crashes is uncertain, the content also belongs to the uncertain inference problem [24]; using Bayesian network is effective to the advantage of uncertainty inference and can build a relatively accurate inference model and realize effective prediction of the severity of traffic crashes.

4.1. Validation of Bayesian Network Learning Results. The data sample information used in this study is complete, and there are many node variables selected; therefore, the structure learning method based on scoring search is picked. The structure learning method based on score search includes two parts: one is to select an appropriate score function to evaluate the quality of the network structure; the other is to determine the appropriate search strategy for finding the highest rated network structure. The scoring function selected in this research is the BD scoring function based on Bayesian statistics, which has good accuracy and fitting effect. K2 search algorithm is a commonly used search strategy for searching high-scoring network structures, which combines the selected scoring function and search strategy to search for the optimal network structure [25].

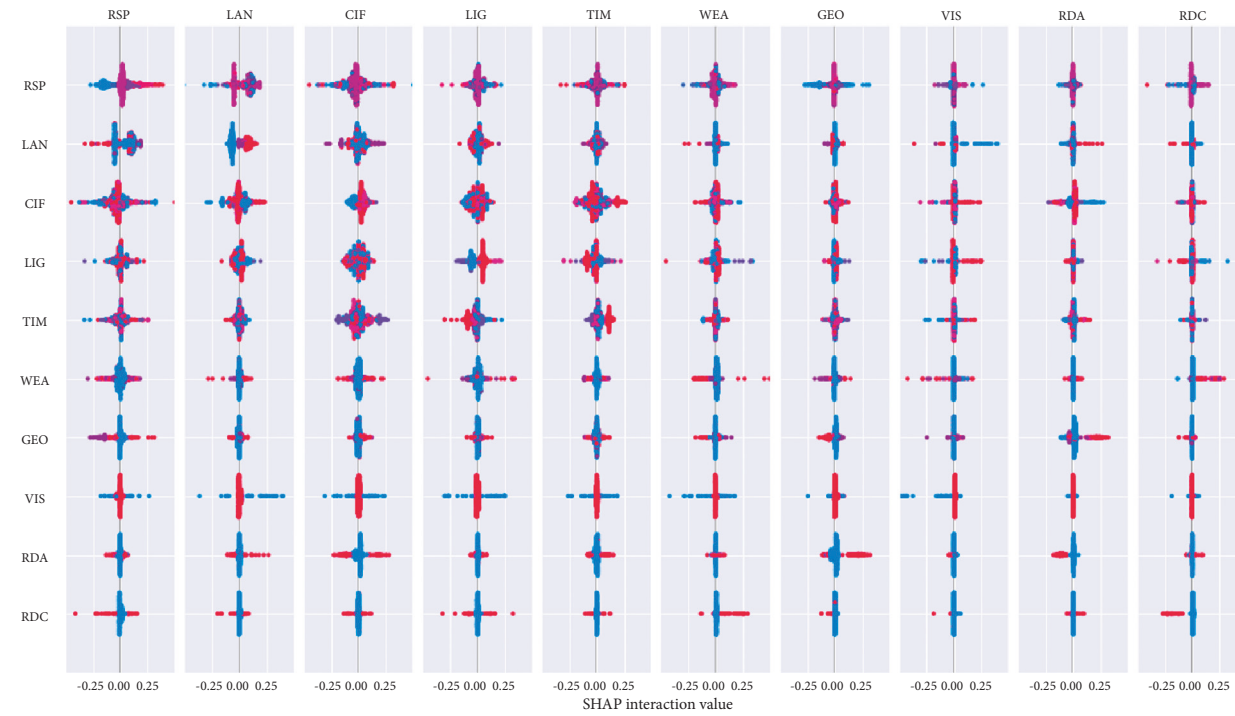


Figure 3: Independent variable interaction effect plot.

K2 search algorithm uses greedy search to obtain the maximum score function. When the K2 search algorithm is applied to learn the structure search of Bayesian network, it is necessary to determine the node order in advance. Given the order, the search range can be reduced, and the maximum number of parent nodes in the network should be given first, which is a constraint on the search optimization and can improve the search efficiency. After determining the order of nodes, consider any node in the network; if the node X_i is before node X_j, the network edge cannot exist from Xj to X_i. According to the order of network nodes and the maximum number of parent nodes, the node with the maximum value of the scoring function is selected as its parent node. When the scoring function can no longer increase, the loop is stopped. The specific optimization algorithm flow is shown in Algorithm 1.

Since the order of network nodes is required to be given when the K2 search algorithm is used for structure learning to reduce the search range, the determination of the order of network nodes is mostly based on the correlation analysis results of independent variables and dependent variables in previous studies [26]. The determination of node order is a subjective decision process, and the variables have an interaction effect with each other. It is not proper to determine the node order via analyzing the correlation between independent variables and dependent variables; the input of such network node order will affect the result of Bayesian network structure learning. Therefore, it is necessary to address this problem by the proposed approach in this research considering the interaction effect of features.

In this research, the order of network nodes is determined based on the interpretation results of XGBoost model by SHAP. XGBoost model is used to analyze the factors influencing the severity of freeway traffic crashes, and SHAP is used to explain the model to obtain the order of the importance of the global impact of each characteristic variable on the crash severity. The importance degree from large to small is type of road side protection facilities, type of road section, central isolation facilities, lighting conditions, accident time, weather, terrain, visibility, road alignment, road surface, and road section location.

Figure 4 depicts a thermal diagram of correlation coefficients between variables. According to the correlation coefficient between each independent variable and the decision variable in the figure, all variables are sorted in order from large to small: roadside protective facilities type, road type, lighting conditions, geography, road alignment, road surface conditions, weather, visibility, accident time, central isolation facilities, and road section location.

In the order obtained from the correlation of sorting, the impact of road surface conditions on the severity of crashes contributes more than that of weather; this is indeed unreasonable, because road surface conditions are largely influenced by the weather. Based on the preceding content about the interaction between the variables' effect diagram, we can see the weather and road conditions have a significant positive correlation. Therefore, the influence of weather on the severity of the crash is greater than that of the road surface condition. It can be seen that the importance ranking obtained by the proposed approach in this research is relatively reasonable.

According to the above analysis, the order of node variables is determined as follows: 1- road section location, 2- road surface condition, 3- road alignment, 4- visibility, 5- geography, 6- weather, 7- accident occurrence time, 8- lighting conditions, 9- central isolation facilities, 10- road section type,

Input: node variable set "n," node order "order," maximum number of parent nodes "Max_fan_in," observation data set "D."
Output: Optimal network structure based on K2 algorithm, that is, each point and its parent node.
(1) for $i = 1$ to n (n represents the number of total nodes);
(2) $\pi_i = \varnothing$;
(3) $p_{old} = g(i, \pi_i)$;
(4) Set variable $OKTOProceed = true$;
(5) while $OKTOProceed = true$ and $|\pi_i| < m$ do;
(6) pick one node z from $p_{red}(i) - \pi_i$, let $g(i, \pi_i \cup \{z\})$ maximize, where $p_{red}(i)$ denotes the nodes whose order is before the ith node;
(7) Calculate $P_{new} = g(i, \pi_i \bigcup \{z\})$;
(8) If $P_{new} > P_{old}$ then;
(9) $P_{old} = P_{new}$;
(10) $\pi_i = \pi_i \bigcup \{z\}$;
(11) $\pi_i = \pi_i \bigcup \{z\}$;
(12) else OKTOProceed = false;
(13) end while;
(14) return (node i, parent node π_i);
(15) end for;
(16) end K2.

Algorithm 1: K2 algorithm optimization process.

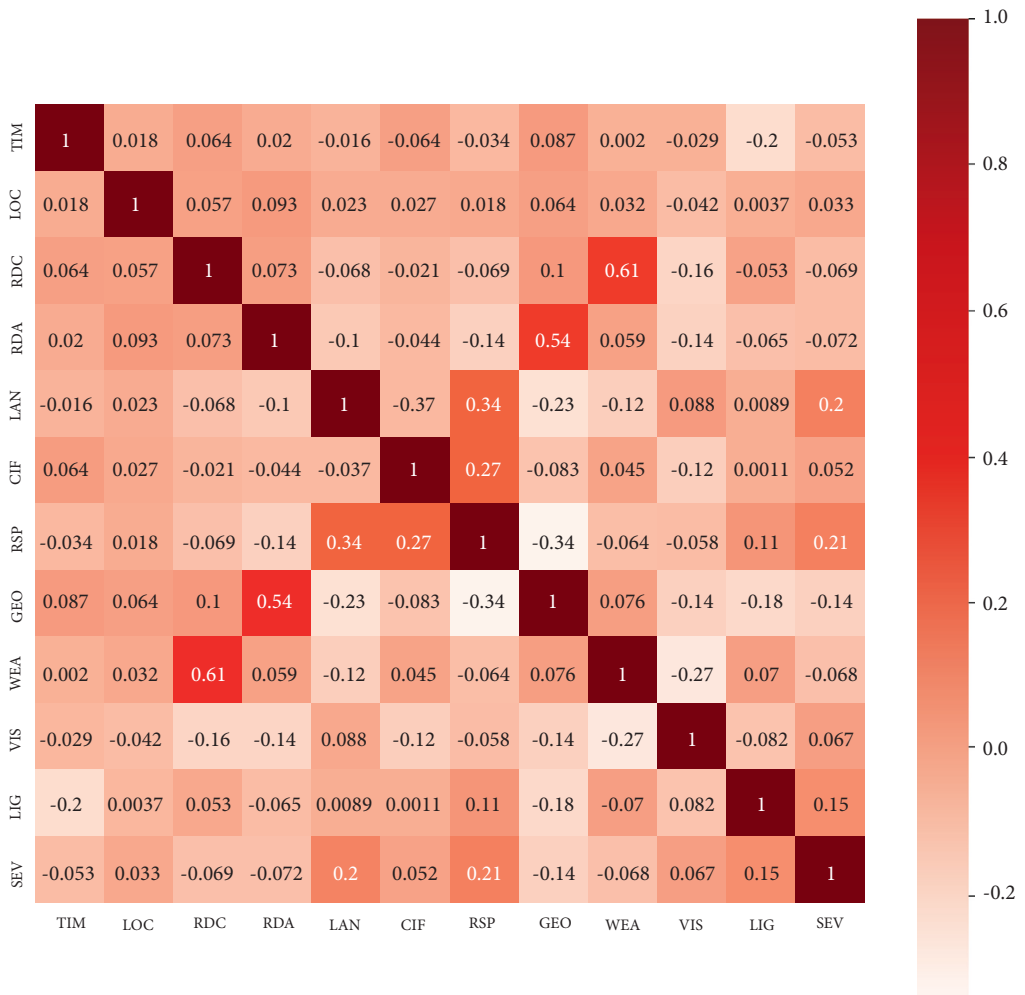


Figure 4: Correlation coefficient heat map.

11- road side protection facilities type, 12- crash severity. For the obtained Bayesian network, the validity analysis is needed to ensure that the Bayesian network inference model is accurate. Then the validity of the constructed network is verified from the aspect of learning accuracy of Bayesian network. The error of parameter learning can be obtained by comparing the statistical calculation results in the sample with the node parameter learning results in the Bayesian network. The learning accuracy of the Bayesian network can be inferred from the parameter learning error [27]. The decision variable in this research is the traffic crash severity; thus, the parameter learning results of the node of crash severity are used to verify. The learning results of node parameters of decision variables are compared with the statistical results of sample data; the absolute error and relative error between them are calculated to judge the learning effect. Due to the large number of types of roadside protection facilities directly affecting decision variables [28], the comparison results cannot be fully displayed. Therefore, the comparison results of fatal crashes are selected for analysis. Specific calculation results are shown in Table 3.

The maximum absolute error of parameter learning results is less than 6%, while the maximum relative error is mostly within 5%. Although the learning accuracy of the model network structure is not very high, the network model still has certain reference significance. The reason affecting the training accuracy probably is that the sample size of crash data is small.

4.2. Construction and Verification of Bayesian Network Inference Model. After the Bayesian network structure is obtained, the inference operation engine in the full-BNT toolbox of MATLAB software is called on the basis of this structure to obtain the Bayesian network inference model. The expected probability distribution of the desired nodes can be figured out by inputting the values of any one or more network nodes in the inference model. The node to be speculated in this research is the severity of the traffic crash, so the query variable in the model can be set to the traffic crash severity. When the values of other node variables or combinations of variables are known, the probability distribution of traffic crash severity can be calculated according to the model; that is, the severity of the crash can be predicted. The inference method selected in this research is the more commonly used and efficient clip-tree propagation algorithm, which belongs to precise reasoning, and the corresponding operation engine in the toolbox is "Jtree_INF_engine." Figure 5 shows the inference process.

4.2.1. Constructing the Moral Graph. Constructing the moral graph is the first step of Bayesian network inference based on clip-tree propagation algorithm. In this process, the direction of each directed edge in the directed acyclic graph is removed, and then each parent node in the network is combined. The undirected graph is the moral graph. After getting the initial moral graph, check whether each node is in the triangle region. If the most simplified region of the node is the polygon region, then remove the polygon by adding edges. Finally, make each node in the triangle region, and get the final moral graph of the network. Figure 6 is the moral graph of the constructed Bayesian network for freeway traffic crash severity.

4.2.2. Determining the Elimination Order of Variables. According to the moral graph, the maximum potential search method can be adopted to determine the order of variables elimination. The working principle of the maximum potential search method is that, on the basis of determining the eliminated variables, the selection of the next variable can minimize the correlation potential between variables. The specific process is as follows: firstly, select any node and mark it, then find the unmarked node that is connected to the most marked nodes, and continue marking. If there are multiple nodes with the most adjacent marked nodes, select any node to continue marking. When all nodes are marked, the nodes are sorted according to the marked order, with the first marked in the last and the last marked in the first. The resulting node order is the optimal elimination order of variables. Generally, any root node in the Bayesian network is selected as the starting marker, and the elimination order of variables according to the above search method is $12 \longrightarrow 9 \longrightarrow 2 \longrightarrow 3 \longrightarrow 10 \longrightarrow 7 \longrightarrow 5 \longrightarrow 4 \longrightarrow 6 \longrightarrow 8 \longrightarrow 11 \longrightarrow 1$.

4.2.3. Constructing the Cluster-Tree. After determining the elimination order of variables, the elimination of variables is carried out according to the elimination order starting from the Bayesian network moral graph. Before eliminating the nodes, a clique consisting of the variable to be eliminated and all variables adjacent to that variable is constructed until all variables are eliminated in order. After elimination, the resulting cliques are organized in an appropriate way to obtain a cluster-tree containing all nodes. Via this approach, a Bayesian network cluster-tree of freeway crash severity is constructed, and the results are shown in Figure 7.

4.2.4. Setting Inferential Evidence. Given the information of any node variable, the tree can effectively transfer and share the information of the node after constructing the cluster-tree according to the above way. The efficiency of cluster-tree propagation algorithm lies in its real-time information sharing, which can simplify the transfer process between cluster-tree and inference process. Before inferential analysis, the evidence variables should be given first, and then the query variables should be determined. The evidence variable and query variable can be one or a combination of multiple variables. The query process of multiple query variables uses the shared inference mechanism of cluster-tree to deduce the results of multiple query variables, based on the inference of query variables. After determining the query variable and the evidence variable, the correlation between the evidence variable and the query variable can be figured out according to the conditional probability and Bayesian theorem.

The sequence of nodes of Bayesian network of freeway traffic crash severity established in this research is as follows: [TIM] [LOC] [RDC] [RDA] [LAN] [CIF] [RSP] [GEO]

Table 3: Parameter verification of the crash severity nodes.

Variables		Parameter learning results			
		Fatal traffic crashes			
Road section type	Type of roadside protection facilities	The relative error	Absolute error	The real value	Bayes
1	1	0.0335	0.0142	0.4369	0.4227
	2	0.0002	0.0001	0.5555	0.5556
	3	0.0303	0.0187	0.6341	0.6154
	4	0.0124	0.0060	0.4866	0.4806
	5	0.0177	0.0138	0.7916	0.7778
	6	0.0000	0.0000	0.0000	0.0000
	7	0.1110	0.0666	0.6666	0.6000
2	1	0.0000	0.0000	0.3333	0.3333
	2	0.0000	0.0000	0.0000	0.0000
	3	0.1001	0.0455	0.5000	0.4545
	4	0.0689	0.0503	0.7800	0.7297
	5	0.0000	0.0000	0.0000	0.0000
	6	0.0000	0.0000	0.0000	0.0000
	7	0.0000	0.0000	1.0000	1.0000
3	1	0.0000	0.0000	0.2500	0.2500
	2	0.0000	0.0000	1.0000	1.0000
	3	0.0476	0.0286	0.5714	0.6000
	4	0.0661	0.0480	0.7731	0.7251
	5	0.0000	0.0000	0.0000	00.0000
	6	0.0554	0.0416	0.7916	0.7500
	7	0.0000	0.0000	1.0000	1.0000

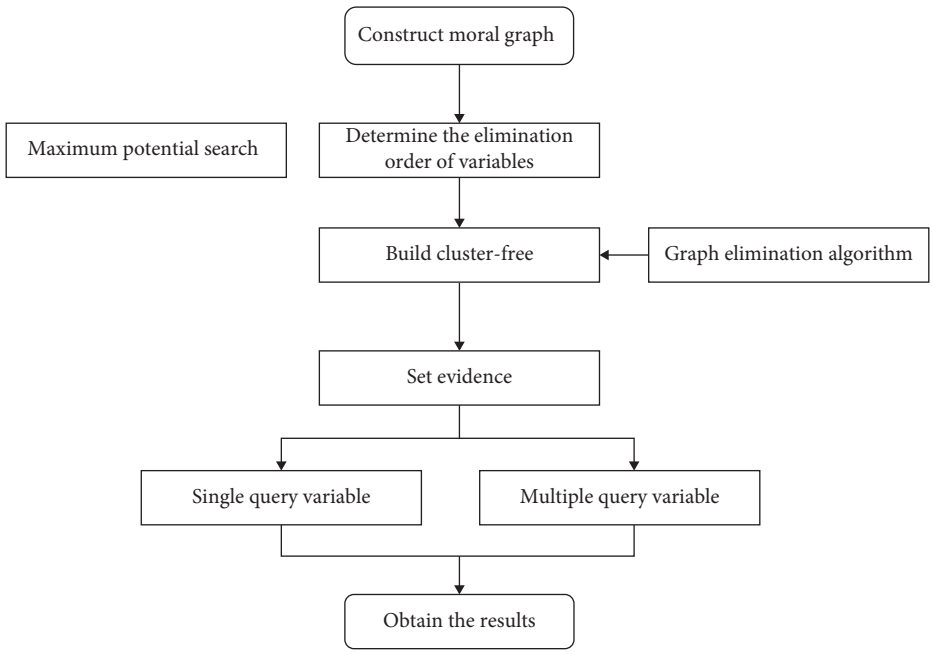


Figure 5: Traffic crash severity Bayesian network inference flow.

[WEA] [VIS] [SEV]. The purpose of constructing Bayesian network inference model is to deduce the severity of traffic crash. Therefore, "SEV" is set as the query variable in this research, which belongs to single-query variable inference. The selection of evidence variables can be any combination of the first 11 node variables.

4.2.5. Inference Learning. The inference solution process of Bayesian network is based on conditional probability and Bayesian theorem. In this research, the prediction model of crash severity is constructed via the toolbox "FULL-BNT" in MATLAB software, which integrates the learning and inferring algorithm functions of Bayesian network. The

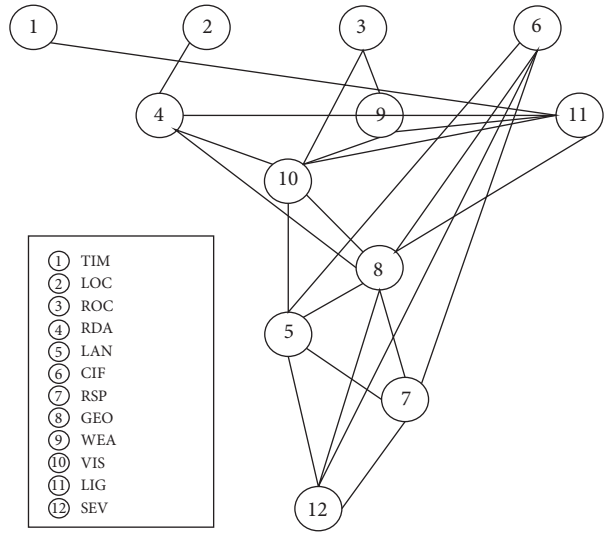


Figure 6: Bayesian network moral graph of crash severity.

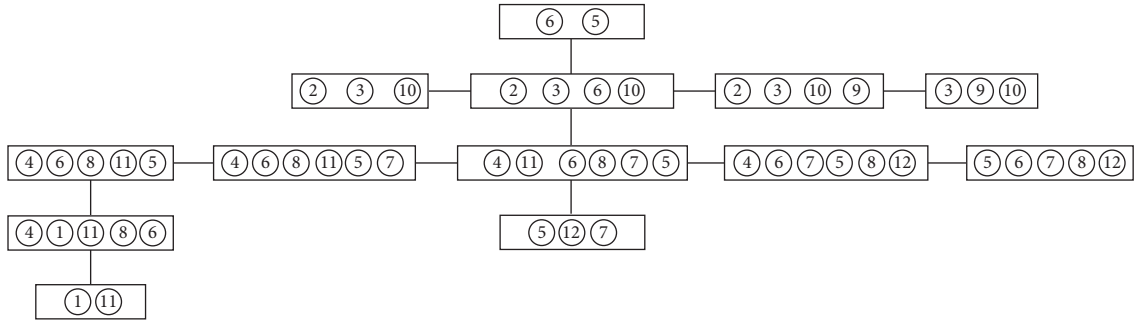


Figure 7: The cluster-tree of Bayesian network inference on crash severity.

specific inference process is as follows: Input the previously obtained Bayesian network, then use the joint tree inference engine "jtree _ inf _ engine (bnet)" in FULL-BNT toolbox to build the inference model, and finally input evidence variables and query variables to carry out inference learning. What this research expects to predict is the traffic crash severity; thus the query variable is set as SEV. After obtaining the inference model based on the above work, the probability distribution of crash severity can be figured out by the inference prediction model by inputting the values of other expected node variables into the inference program.

4.2.6. Validation of the Model Accuracy. The specific process of calculating the prediction accuracy of the model in MATLAB is as follows: First, import the existing sample data and input values of the other 11 dependent variables corresponding to each crash data into the prediction model, that is, input evidence variables; then, the predicted results of crash severity are compared with the actual severity of the corresponding crash data, and the severity of all sample data is predicted in turn; the ratio analysis is processed towards the correctly predicted quantity of traffic crashes with the quantity of all crashes; finally, the accuracy of the prediction model is carried out.

The realization for this part of the work needs to design verification programs in MATLAB and then go through 557 pieces of crash data, in which 496 crashes are accurately predicted; hence the prediction accuracy of the model is 89.05%. When the prediction accuracy of the model is greater than or equal to 80%, the prediction results of the model are relatively good [19]. Therefore, it can be seen that the traffic crash severity prediction model constructed in this research has good prediction accuracy.

5. Results and Discussion

Considering the influence of the interaction of various factors on the severity of crashes, the influence rule of these factors on the traffic crash severity is obtained by analyzing the inference results of the prediction model, which can provide a direction for the follow-up traffic safety management countermeasures.

5.1. Interaction between Weather and Road Type. Input values of weather and road type in the prediction model. Set the variables as Evidence {WEA} = I, Evidence {LAN} = j, where I ranges from 1 to 5 and j ranges from 1 to 3. The output results of the model are the prediction results of

accident severity under the corresponding weather variables and road section type variables, and the specific contents are shown in Table 4.

Road section types have a certain influence on the three types of traffic crashes, among which the influence on property loss crashes is relatively greater. When the road type is a complex node (ramp, road entrance and exit, etc.), the probability of injury crash is lower than other road types under any weather condition; synchronously, the probability of property loss and fatal crash is relatively increased. This is consistent with the inference result of single factor analysis [19].

By analyzing the inference results of the two variables on injury crashes, it can be seen that injury crashes are most likely to occur in foggy conditions of special road sections. By analyzing the inference results of property loss crashes, it can be seen that the possibility of property loss crashes in complex node sections is greater than other sections under any weather conditions; in particular it is the greatest in foggy conditions. It can be seen from the analysis of the inference results of two variables on fatal crashes that the probability of fatal crashes is the highest in rainy days at ordinary road sections, while the analysis results towards single factor influence in the previous section of this paper show that the probability of fatal is the highest in complex node crashes and the probability of fatal crashes is the highest in rainy days.

Table 4: The effect of weather and road type on the severity of crashes.

Road type	Weather	Property loss crash	Injury crash	Fatal crash
Normal section	Sunny	0.0281	0.4141	0.5578
	Overcast	0.0278	0.4137	0.5585
	Rainy	0.0277	0.4131	0.5592
	Snowy	0.0250	0.4098	0.5652
	Foggy	0.0250	0.4102	0.5647
Special section	Sunny	0.1088	0.4097	0.4815
	Overcast	0.1099	0.4122	0.4779
	Rainy	0.1098	0.4123	0.4779
	Snowy	0.1147	0.4233	0.4620
	Foggy	0.1214	0.4397	0.4389
Complex node	Sunny	0.1238	0.3530	0.5232
	Overcast	0.1253	0.3534	0.5213
	Rainy	0.1258	0.3536	0.5205
	Snowy	0.1309	0.3572	0.5118
	Foggy	0.1409	0.3564	0.5027

Table 5: The effect of weather and road alignment on the crash severity.

Road alignment	Weather	Property loss crash	Injury crash	Fatal crash
Straight	Sunny	0.0068	0.3930	0.6002
	Overcast	0.0280	0.4259	0.5461
	Rainy	0.0030	0.3892	0.6077
	Snowy	0.0014	0.3900	0.6086
	Foggy	0.0455	0.4528	0.5018
Nonstraight	Sunny	0.0021	0.4895	0.5084
	Overcast	0.0145	0.4944	0.4911
	Rainy	0.0049	0.4830	0.5121
	Snowy	0.0146	0.4665	0.5190
	Foggy	0.0597	0.5158	0.4245

5.2. Interaction between Weather and Road Alignment. Input evidence values of weather and road alignment; the settings are as follows: evidence $\{WEA\} = i$, evidence $\{RDA\} = j$, where i ranges from 1 to 5 and j ranges from 1 to 2. The output results of the model are the prediction results of crash severity under the corresponding weather variables and road alignment variables, and the specific contents are shown in Table 5.

The results in Table 5 indicate that the interaction of the two variables has an impact on all three types of crashes, among which the impact on injury and death is relatively greater. For the road alignment, the possibility of injury crash occurring in nonstraight alignment is greater than that in straight alignment, while the possibility of a fatal crash occurring in nonstraight alignment is greater than that in straight alignment, which is consistent with the law of road alignment affecting the severity of crashes by a single factor [19].

By analyzing the inference results of the two variables to the property loss crash, it can be seen that the property loss crash is most likely to occur in the foggy days of nonstraight road section. By analyzing the inference results of the two variables on injury crashes, it can be seen that injury crashes are most likely to occur in foggy conditions of nonstraight sections. It can be seen from the analysis of the inference results of fatal crashes that the possibility of fatal crashes is the greatest in the case of straight sections under snowy conditions, and the possibility of fatal crashes is relatively greater in the case of straight sections under rainy conditions, which is basically consistent with the influence rule of the single factor on the crash severity [19]. Therefore, the interaction effect of weather and road alignment on the severity of crashes is not obvious.

5.3. Interaction between Road Alignment and Crash Time. The evidence variables' values of road alignment and crash occurrence time are input into the prediction model. The settings are as follows: evidence $\{RDA\} = i$, evidence $\{TIM\} = j$, where the value of i ranges from 1 to 2 and the value of j ranges from 1 to 4. The output results of the model are the prediction results of crash severity under the corresponding road alignment variables and crash occurrence time variables, and the specific contents are shown in Table 6.

The interaction of road alignment and crash occurrence time on the severity of the accident is analyzed. Combined with the inference results in Table 6, it can be seen that the interaction of the two variables has a certain impact on the severity of the three types of crashes; among them, the impact of property loss crash and fatal crash is relatively greater. By analyzing the inference results of the two variables on property loss crash and injury crash, it can be seen that when the road alignment is not straight and the time is

Table 6: The effect of road alignment and time on the crash severity.

Road alignment	Time	Property loss crash	Injury crash	Fatal crash
Straight	0:00~6:00	0.0088	0.3686	0.6226
	6:00~12:00	0.0111	0.3892	0.5997
	12:00~18:00	0.0108	0.3888	0.6004
	18:00~24:00	0.0081	0.3675	0.6244
Nonstraight	0:00~6:00	0.0033	0.3682	0.6285
	6:00~12:00	0.0153	0.4380	0.5467
	12:00~18:00	0.0173	0.4642	0.5185
	18:00~24:00	0.0033	0.3682	0.6285

Table 7: The effect of weather and lighting conditions on the crash severity.

Lighting condition	Weather	Property loss crash	Injury crash	Fatal crash
Daytime	Sunny	0.0060	0.4221	0.5719
	Overcast	0.0186	0.4362	0.5453
	Rainy	0.0048	0.4224	0.5728
	Snowy	0.0029	0.4251	0.5720
	Foggy	0.0511	0.4725	0.4763
Dusk/Dawn	Sunny	0.0124	0.3876	0.6001
	Overcast	0.0497	0.4455	0.5049
	Rainy	0.0026	0.3886	0.6088
	Snowy	0.0007	0.4335	0.5658
	Foggy	0.0620	0.4643	0.4737
Lighting at night	Sunny	0.0171	0.2966	0.6864
	Overcast	0.0541	0.4324	0.5135
	Rainy	0.0101	0.3055	0.6844
	Snowy	0.0026	0.3736	0.6238
	Foggy	0.0879	0.5529	0.3592
Nonlight at night	Sunny	0.0065	0.4742	0.5193
	Overcast	0.0322	0.4687	0.4991
	Rainy	0.0031	0.4674	0.5294
	Snowy	0.0021	0.4390	0.5589
	Foggy	0.0522	0.4650	0.4828

during 12:00 to 18:00, the probability of property loss crash or injury crash is the largest. It can be seen from the analysis of the inference results of the two variables to the fatal crash that when the road alignment is not straight and the occurrence time is 18:00 ~ 6:00, the probability of fatal crashes accounts the largest on the nonstraight road at night.

5.4. The Interaction between Weather and Lighting Conditions. Enter the values of the evidence variables weather and lighting conditions into the prediction model; set the variables as evidence {WEA} = i, evidence {LIG} = j, where the value of i ranges from 1 to 5 and the value of j ranges from 1 to 4. The output results of the model are the prediction results of crash severity under the corresponding weather variables and lighting condition variables, and the specific contents are shown in Table 7.

The influence rules of weather and lighting conditions on the severity of crashes are analyzed. According to the inference results in Table 7, it can be seen that the interaction of the two variables has a great influence on the severity of the three types of crashes. In terms of lighting condition, the possibility of property loss crash is the greatest (including all weather conditions), and the overall probability of fatal crashes occurring when there is no lighting at night is the highest, while the overall probability of fatal crash occurring when there is lighting at night is the highest, which is basically consistent with the results of single factor analysis [19].

By analyzing the inference results of the two variables on property loss and injury crashes, it can be seen that the possibility of property loss and injury crashes on a foggy night with lighting is greater than that in other weather and lighting conditions. By analyzing the inference result of fatal crashes, the possibility of fatal crash is greater when it is sunny or rainy and there is lighting at night. The univariate analysis results show that the probability of injury crashes is high on foggy days, while the probability of fatal crash is high on rainy days [19]. Therefore, the interaction between weather and lighting conditions has no significant change on the law of traffic crash severity.

6. Conclusions

This research considers the influence of road and environmental factors on the severity of freeway traffic crashes, uses XGBoost to determine the importance of features, and establishes a Bayesian network model to analyze the prediction of traffic crash severity under the interaction of road and environmental factors. The main conclusions are as follows:

(1) Inclement weather conditions such as rain, snow, and fog occur in most of the combined conditions of fatal crashes; the nonguardrail form of roadside protection facilities, such as green belt and road trees, is more likely to cause fatal crashes; foggy days have great influence on property loss and injury crashes. Rainy days are the most likely to cause fatal crashes. The interaction between ordinary road sections and rainy days has a great influence on the crash severity. Driving on nonstraight roads at night may aggravate the severity of the crash.

(2) In this research, XGBoost-SHAP value model is introduced in the learning of Bayesian network structure to obtain the global importance ranking of each variable on decision variables. Compared with the order of importance obtained from correlation

analysis, the determination of the order of variable importance is more reasonable, which is conducive to obtaining a higher scoring Bayesian network structure.

(3) Partial results obtained by the traffic crash severity prediction method considering the features interaction effect proposed in this research consist of some of the results obtained by considering only a single factor model in previous studies, indicating that the model used in this research is reliable. Additionally, since the crash-prone variables have an interaction effect with each other, the model with the consideration of features interaction can produce more reliable results. Finally, corresponding effective measures can be put forward to prevent the occurrence of crashes or reduce the crash severity according to the combination form of road and environmental factors with higher risk coefficient, which can be applied by the relative transportation department for the freeway safety management.

(4) The amount of sample data used in this research is small, and the factor variables in the dimension of road condition in the used dataset are not complete enough. Furthermore, before learning the structure of the Bayesian network, further research into how to eliminate the influence of small datasets on Bayesian network model construction should be completed and addressed in future models.

Data Availability

Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by China Postdoctoral Science Foundation (2021M700333), Open Project of Shandong Key Laboratory of Highway Technology and Safety Assessment (SH202105), and Beijing Natural Science Foundation (J210001).

References

[1] A. T. Kashani and K. Zandi, "Influence of traffic parameters on the temporal distribution of crashes," *KSCE Journal of Civil Engineering*, vol. 24, no. 3, pp. 954–961, 2020.

[2] S. Hyodo and K. Hasegawa, "Factors affecting analysis of the severity of accidents in cold and snowy areas using the ordered probit model," *Asian Transport Studies*, vol. 7, Article ID 100035, 2021.

[3] J. Lee, J. Chae, T. Yoon, and H. Yang, "Traffic accident severity analysis with rain-related factors using structural equation modeling - a case study of Seoul city," *Accident Analysis & Prevention*, vol. 112, pp. 1–10, 2018.

[4] M. S. R. Amin, A. Zareie, and L. E. Amador-Jiménez, "Climate change modeling and the weather-related road accidents in Canada," *Transportation Research Part D: Transport and Environment*, vol. 32, pp. 171–183, 2014.

[5] D. Sun, Y. Ai, Y. Sun, and L. Zhao, "A highway crash risk assessment method based on traffic safety state division," *PLoS One*, vol. 15, 2020.

[6] S. Roshandel, Z. Zheng, and S. Washington, "Impact of real-time traffic characteristics on freeway crash occurrence: systematic review and meta-analysis," *Accident Analysis & Prevention*, 2015.

[7] Y. Ci, H. Wu, Y. Sun, and L. Wu, "A prediction model with wavelet neural network optimized by the chicken swarm optimization for on-ramps metering of the urban expressway," *Journal of Intelligent Transportation Systems*, vol. 26, no. 3, pp. 356–365, 2022.

[8] M. A. Rahim and H. M. Hassan, "A deep learning based traffic crash severity prediction framework," *Accident Analysis & Prevention*, vol. 154, no. 1, Article ID 106090, 2021.

[9] Q. Zeng, H. L. Huang, P. P. Xu, and M. Ma, "Developing an optimized artificial neural network to predict traffic crash injury severity," in *Proceedings of the Cota International Conference of Transportation Professionals*, Changsha, China, July 2014.

[10] X. Chen, H. Chen, Y. Yang et al., "Traffic flow prediction by an ensemble framework with data denoising and deep learning model," *Physica A: Statistical Mechanics and its Applications*, vol. 565, Article ID 125574, 2021.

[11] A. K. Çelik and E. Oktay, "A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars Provinces of Turkey," *Accident Analysis & Prevention*, vol. 72, pp. 66–77, 2014.

[12] X. Jiang, B. Huang, R. L. Zaretzki, S. Richards, X. Yan, and H. Zhang, "Investigating the influence of curbs on single-vehicle crash injury severity utilizing zero-inflated ordered probit models," *Accident Analysis & Prevention*, vol. 57, pp. 55–66, 2013.

[13] M. S. B. Shaheed, K. Gkritza, W. Zhang, and Z. Hans, "A mixed logit analysis of two-vehicle crash severities involving a motorcycle," *Accident Analysis & Prevention*, vol. 61, pp. 119–128, 2013.

[14] C. Lou, *Analysis of Causes and Severity of Freeway Traffic Accidents Based on Time and Space*, Beijing Jiaotong University, Beijing, China, 2021, In Chinese.

[15] S. AlKheder, F. AlRukaibi, and A. Aiash, "Risk analysis of traffic accidents' severities: an application of three data mining models," *ISA Transactions*, vol. 106, pp. 213–220, 2020.

[16] Z. Ma, C. Shao, and C. Dong, "Temporal-spatial analysis model of traffic accident severity based on cumulative logistic model," *China Safety Science Journal*, vol. 21, no. 9, pp. 94–99, 2011, In Chinese.

[17] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, August 2016.

[18] R. Hortegal, Y. Maduro, R. Cancellier et al., "Mechanical dispersion of left ventricle and left atrial reservoir strain are both superior to global longitudinal strain to predict exercise capacity in heart failure with preserved ejection fraction," *European Heart Journal - Cardiovascular Imaging*, vol. 23, 2022.

[19] C. Li, X. Wu, Z. Zhang, Z. Ma, Y. Zhu, and Y. Chen, "Freeway traffic accident severity prediction based on multi-dimensional and multi-layer Bayesian network," in *Proceedings of the 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)*, pp. 1032–1035, Shenyang, China, January 2022.

[20] S. Sun, C. Zhang, and G. Yu, "A Bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 124–132, 2006.

[21] S. M. Lundberg, G. G. Erion, and S. I. Lee, "Consistent Individualized Feature Attribution for Tree Ensembles," 2018, https://arxiv.org/abs/1802.03888.

[22] K. Futagami, Y. Fukazawa, N. Kapoor, and T. Kito, "Pairwise acquisition prediction with SHAP value interpretation," *The Journal of Finance and Data Science*, vol. 7, no. 1, 2021.

[23] A. Khosbayar, J. Valluru, and B. Huang, "Multi-rate Gaussian Bayesian network soft sensor development with noisy input and missing data," *Journal of Process Control*, vol. 105, pp. 48–61, 2021.

[24] A. Da Ntas, M. Fowler, and G. Koorey, "Effect of road network bendiness on traffic crash occurrence," in *Proceedings of the IPENZ Transportation Group Technical Conference*, Tauranga, New Zealand, October 2007.

[25] R. Aghdam, V. R. Tabar, and H. Pezeshk, "Some node ordering methods for the K2 algorithm: node Ordering Methods," *Computational Intelligence*, vol. 35, no. 5, 2018.

[26] R. Angus, D. Foreman-Mackey, and J. A. Johnson, "Systematics-insensitive periodic signal search with K2," *The Astrophysical Journal*, vol. 818, no. 2, p. 109, 2016.

[27] K. Asghari, M. Masdari, F. S. Gharehchopogh, and R. Saneifard, "A fixed structure learning automata-based optimization algorithm for structure learning of Bayesian networks," *Expert Systems*, vol. 38, 2021.

[28] Z. Yao, H. Jiang, Y. Cheng, Y. Jiang, and B. Ran, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, 2020.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 2166345, 12 pages
https://doi.org/10.1155/2022/2166345

Research Article

A Deep Learning Based Traffic State Estimation Method for Mixed Traffic Flow Environment

Fan Ding,[1] Yongyi Zhang,[1,2] Rui Chen,[3] Zhanwen Liu,[4] and Huachun Tan[1]

[1]*School of Transportation, Southeast University, Nanjing 211189, China*
[2]*Southeast University-Monash University Joint Graduate School, Suzhou, China*
[3]*State Key Laboratory of Integrated Service Networks, Xidian University, Xidian 710071, China*
[4]*School of Information Engineering, Chang'an University, Chang'an 710064, China*

Correspondence should be addressed to Huachun Tan; tanhc@seu.edu.cn

Received 12 November 2021; Revised 1 January 2022; Accepted 28 February 2022; Published 7 April 2022

Academic Editor: Alain Lambert

Traffic state estimation plays a fundamental role in traffic control and management. In the connected vehicles (CVs) environment, more traffic-related data perceived and interacted by CVs can be used to estimate traffic state. However, when there is a low penetration rate of CVs, the data collected from CVs would be inadequate. Meanwhile, the representativeness of the collected data is positively correlated with the penetration rate. This article presents a traffic state estimation method based on a deep learning algorithm under a low and dynamic CVs penetration rate environment. Specifically, we design a K-Nearest Neighbor (KNN) data filling model integrating acceleration data to solve the problem of insufficient data. This method can fuse the time feature of speed by acceleration modification and mine the distribution features of speed by KNN. In addition, to reduce the estimation error caused by penetration rate, we design a Long Short-Term Memory (LSTM) model, which uses penetration rate estimated by Macroscopic Fundamental Diagram (MFD) as one of the input factors. Finally, we use the concept of operational efficiency for reference, dividing traffic state into three categories according to the estimated speed: free flow, optimal flow, and congestion. SUMO is used to simulate traffic cases under different penetration rates to evaluate our scheme. The results suggest that our data filling model can significantly improve filling accuracy under a low penetration rate; there is also a better performance of our estimation model than that of other comparison models in both low and dynamic penetration rates.

1. Introduction

Traffic state estimation is a vitally important part of intelligent transportation management. It can improve transportation efficiency and reduce transportation costs [1]. Transportation control, management, and information services, such as congestion monitoring, ramp control, event detection, and travel time estimation, can be implemented according to state evaluation [2–4]. Traffic state estimation needs to infer real traffic conditions from incomplete and uncertain traffic information, which depends on strong traffic data support [5]. In recent years, data of traffic estimation mainly come from fixed detectors [6], probe vehicles [7, 8], cellular networks [9, 10], and so on. However, these devices have both advantages and disadvantages [11]. For example, fixed detectors are limited by excessive installation and maintenance budgets; floating vehicle data is often limited to urban roads with low proportion; cellular network data is large, but it is difficult to be used due to serious data noise and inaccurate positioning [8]. In addition, these data are difficult to meet the requirements of increasing real-time control.

Nowadays, the emergence of connected vehicles (CVs) provides new ideas and methods for traffic data acquisition and traffic state estimation [11, 12]. One of the salient features of CVs is that they can sense themselves and road operation status (such as speed, position, and acceleration) and can communicate with roadside units or central control units [13, 14]. They are equivalent to mobile sensors, which can feedback data in real-time on any road with

communication facilities. The feasibility and accuracy of using CVs data for traffic state estimation have been proved [15]. The estimation of traffic state is the estimation of three parameters of macro traffic flow, namely, speed, volume, and density [16]. However, because when there is only CVs data, only vehicle speeds can be obtained directly and can represent the overall speed of roads, while the volume and density need the support of additional sensors [11]. Therefore, this article selects speed as the index to estimate traffic state.

However, research shows that the popularization of CVs would take a long time, and research of mixed traffic, especially the low penetration rate of CVs, is still needed [13]. In the low penetration rate case, CVs could not be detected for some time, resulting in no data for a period [17]. This is one of the reasons for the poor estimation accuracy. The model proposed by Bekiaris et al. [18] needs to ensure minimum data requirements with the help of other sensor data in low penetration rate cases. Although this can ensure the accuracy of estimation, it is not suitable to be extended to most roads due to a lack of flexibility and scalability. Without drawing support of other data, scholars mainly use the time smoothing method, interpolation method, and regression method to fill insufficient data [19, 20]. By using the periodicity of data, the K-Nearest Neighbors (KNN) can apply more relevant data to explore potential laws of data [20, 21]. However, the common feature of these methods is to estimate missing data based on the time characteristics of data. The tensor completion method of rank minimization can synthesize relevant information and explore potential features [22], but its calculation is complex and consumes a lot of computer resources [23]. Besides, filling the missing data or modifying abnormal data caused by machine failure is the main purpose of the current data filling methods, and filling vacant data caused by low penetration rate has not received much attention. The CVs provide more diverse and timely data that can help find a more reliable filling method.

Based on these filling data, traffic state can be estimated. In literature, methods of state estimation can be summarized as dynamic model-based methods and data-driven methods [24, 25]. The classical dynamics models include Macroscopic Fundamental Diagram (MFD) and Cellular Transport Model (CTM). Wang et al. [26] took MFD as a time update equation and used the Kalman filter to estimate highway traffic state. Fountoulakis et al. [27] and Bekiaris et al. [15] applied the Kalman filter combined with MFD to estimate traffic state in mixed traffic scenarios. CTM can well simulate traffic flow phenomena such as queuing and dissipation [28]. Tampere et al. [29] proposed a traffic state estimation model based on CTM. However, dynamic model-based methods are difficult to use in all scenarios and conditions, while a data-driven method is much more flexible [30]. In recent years, data-driven methods have made great progress, and especially neural network such as backpropagation (BP) neural network, Artificial Neural Network (ANN), and Recurrent Neural Network (RNN) has achieved great success in traffic state estimation [30]. However, there are problems of vanishing gradient and long-term dependence in the traditional neural network [31, 32]. Long Short-Term Memory (LSTM) is a special type of RNN, which can solve these problems through some gate controls [33]. Vinayakumar et al. [34] designed traffic prediction experiments on the RNN method family, and the results showed that LSTM performed well compared with other RNN and classical methods. Cui et al. used [35] stacked bidirectional and unidirectional LSTM to predict the whole network traffic state with missing values. Du et al. [36] used LSTM to predict traffic state in the connected vehicle environment. Moreover, Khan et al. [1] integrated connected vehicle technology with deep learning to evaluate the traffic state, and the result showed that when the penetration rate exceeded 20%, the accuracy reached 85%.

Deep learning models are widely used in traffic state estimation and prediction and have achieved good results [25]. However, the model based on a single data lacks a comprehensive description of the traffic network. In addition to the features directly related to the estimated object, some external factors would also affect the accuracy of the estimated model. Therefore, current studies draw more attention to the fusion estimation of multiple data by adding additional explanatory variables [37]. Loop detector data and floating car data for state estimation are a common combination method [38]. Antoniou et al. [30] used a data-driven method, fusing traffic and other data to estimate traffic state. In this article, we would like to focus on the LSTM model that uses multiple-input factors for traffic state estimation, which can flexibly add other factors. In a mixed traffic environment, one of the biggest influencing factors in the overall state estimation by using partial connected vehicle data is the penetration rate of connected vehicles [15, 27]. When the penetration rate is higher, the accuracy of the model is higher, and vice versa. This influence could be put into the deep learning model for learning. Therefore, it is necessary to convert the penetration rate into an input factor in the estimation model.

Generally, we expand the application of connected vehicles data and propose an ACC-P-LSTM model to improve the accuracy of traffic state estimation under the low and dynamic penetration rates in mixed traffic environments. Firstly, to improve the accuracy of data filling at a low penetration rate, we proposed a KNN filling model integrating vehicle acceleration. This model can effectively combine dynamic characteristics of vehicles and extract temporal and spatial distribution of speed. Secondly, for the environment with a dynamic penetration rate, we design a penetration rate estimation method based on MFD and take the estimation result as an input factor of the state estimation model. Then, we construct an LSTM model for overall speed estimation, which takes speed and permeability as input. Finally, SUMO is used to build a mixed traffic environment for getting CVs data. In our experiment, we compared the estimation accuracy in cases with penetration rates from 0.01 to 1.0. Meanwhile, we also set the cases with irregular penetration rates to verify the sensitivity of the model to permeability.

The rest of this article is structured as follows: in the second section, the speed filling model and LSTM state estimation model are explained. In the third section, the

simulation environment is built, and CVs data are analyzed. In the fourth section, under the different CVs penetration rate, our model is compared with the benchmark models without our scheme and other existing advanced algorithms. The fifth section summarizes the full article.

2. Methods

This section explains in detail the proposed ACC-P-LSTM model including the data filling model integrating acceleration, the method of estimating penetration rate, and the LSTM model based on the first two. These three parts would be introduced, respectively, below, while the whole model is introduced first to show a logical relationship between them.

2.1. Overall Framework. The structure of the ACC-P-LSTM model is shown in Figure 1. Firstly, the original speed, acceleration, and number of CVs are acquired and preprocessed. In this model, we attempt to estimate the spatial average speed in one minute, while the data of CVs can be collected through the communication equipment every 10 seconds. The data of multiple CVs collected each time would be taken the average value as the average vehicle state. However, because the collected speed data is incomplete and poorly representative, the existing speed needs to be filled and corrected. The KNN model is used to fill the vacancy speed, and acceleration as an influencing parameter is also fused in this model. On the one hand, filling speed data would be input into the LSTM model; on the other hand, it would be input into MFD to estimate traffic volume. Then, the penetration rate of CVs can be calculated by the count of CVs and the estimated traffic volume. The estimated penetration rate is also entered into the LSTM model together with the filling speed to estimate traffic state. In short, the LSTM model would estimate the overall traffic speed based on the six instantaneous average speeds and would also learn the impact of changes in penetration rate on the accuracy of the model.

2.2. KNN Data Filling Model Integrating Acceleration. To deal with insufficient speed data, a scheme of KNN data filling model integrating acceleration is designed in this article and call it ACC-KNN. Acceleration significantly affects the speed feature of the next moment. Although the acceleration of a single vehicle is a microscopic parameter, we can characterize the macroscopic traffic flow characteristics by calculating the average acceleration of multiple vehicles at the same time, that is, the average acceleration as an indicator of the macroscopic trend. On the one hand, when the speed is missing at the next moment, the acceleration can be used to estimate the speed at the next moment. On the other hand, when a speed deviates from the surrounding speeds (maybe due to low penetration rate, a few vehicles represent the whole), the original speed can be modified by acceleration. In this study, the speed of the next state is estimated based on dynamic formula (1):

$$v'_{(i+1)} = v_{(i)} + \Delta t \cdot a_{(i)}, \tag{1}$$

where $v_{(i)}$ is the speed of state i ($i \in (0, 1, 2, \ldots, I)$) ($I$ is the number of collected speed data), $a_{(i)}$ is the acceleration of state i, $v_{(i+1)}{}'$ is the estimated speed of state $i+1$, and Δt is acceleration duration, in this article taken as 1s. Through formula (1), two sets of speed data can be obtained: one is the original speed, and the other is the estimated speed. We would not abandon the original speed but combine it with the estimated speed to eliminate the influence of deviation from the speed.

In the next step, we use the KNN model to fill the original speed and estimated speed, respectively. The KNN filling model uses correlations of different dimensions to fill and correct missing or outliers in the data. For example, if we do not obtain the measured value $v_{(i)}$ at a certain time, several measured values around it like $v_{(i-k/2)}, v_{(i-k/2+1)}, \ldots, v_{(i-1)}, v_{(i+1)}, \ldots, v_{(i+k/2)}$ can be obtained. We can use the existing speed to estimate the target speed $\tilde{v}_{(i)}$:

$$\tilde{v}_{(i)} = \frac{\sum_i w_i v_{(i)}}{\sum_i w_i}, \tag{2}$$

where w_i is the weight of state i. The weight is inversely related to the distance between adjacent points and target points.

Through formula (2), we get two filling speeds $\tilde{v}_{(i)_{knn}}$ and $\tilde{v}'_{(i)_{acc\_knn}}$. The modified filling speed $\tilde{v}_{(i)_{\text{modified}}}$ would be obtained by weighted summation of the two filling speeds.

$$\tilde{v}_{(i)_{\text{modified}}} = \gamma_i \tilde{v}_{(i)_{knn}} + (1 - \gamma_i)\tilde{v}'_{(i)_{acc\_knn}}, \tag{3}$$

where γ_i is the weight of the original speed, and $(1 - \gamma_i)$ is the weight of the estimated speed, in this article taken γ_i as 0.5.

2.3. Estimating Penetration Rate of CVs. The penetration rate is a characteristic of the local corresponding to the whole. Estimating penetration rate requires knowing the local and global features. Local features can be obtained by counting the number of CVs, while global features can be obtained from the Macroscopic Fundamental Diagram (MFD). MFD is an inherent attribute of roads [16]. It reflects the relationship between three parameters of traffic flow: speed, volume, and density. Therefore, the other two values can be calculated when one parameter is known through the formula of the fundamental diagram. The widely accepted Green Shields model [39] is used in this article.

$$Q_{(n)} = K_j \cdot \overline{v}_{s_{(n)}} - \frac{K_j \cdot \overline{v}^2_{s_{(n)}}}{V_f}, \tag{4}$$

where $\overline{v}_{s_{(n)}}$ is the spatial average speed of state n ($n \in (0, 1, 2 \ldots, N)$) ($N$ is the number of traffic state data), $Q_{(n)}$ is the volume of state n, K_j is blocking density, and V_f is the free flow speed. K_j is the traffic flow density when the vehicle speed is close to zero, which can be calculated from the vehicle length and the shortest stopping distance. V_f

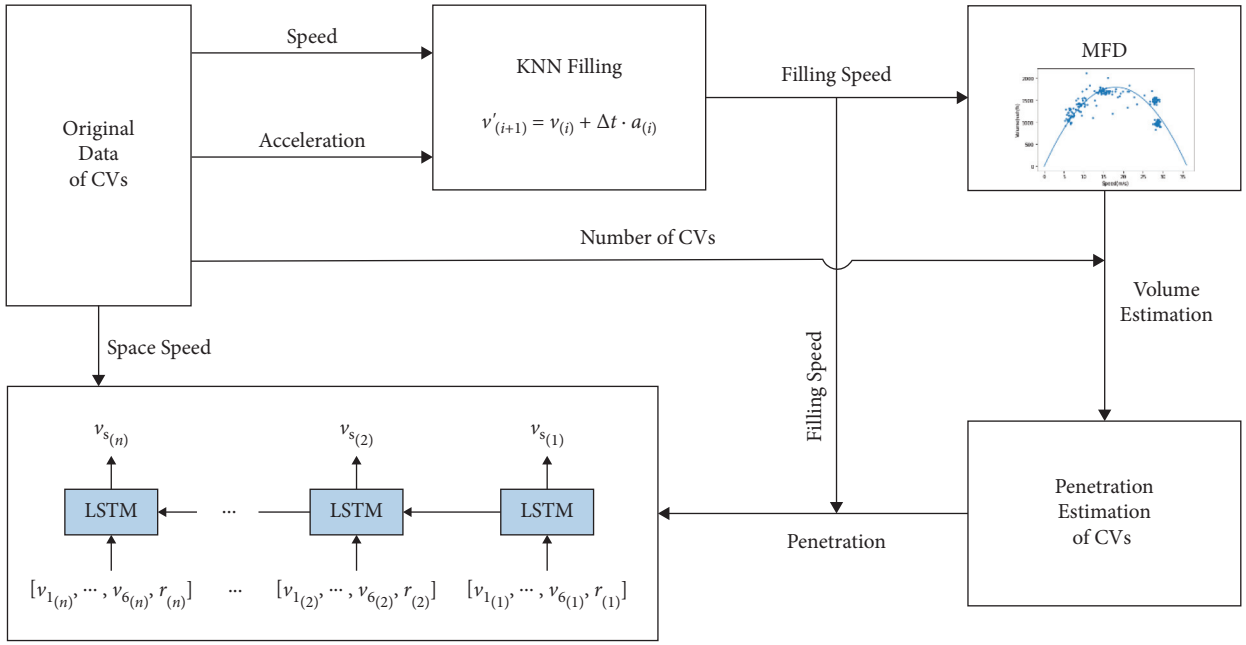


Figure 1: ACC-P-LSTM model framework.

refers to traffic flow speed not affected by upstream and downstream conditions, generally 110% of the speed limit. Therefore, MFD can be obtained on a road section for a while according to geometric characteristics of roads, speed limit, and dynamic performance of vehicles.

The speed collected in reality is location speed, which cannot be directly used for speed input in MFD. According to the conversion formula between space average speed and time average speed [40], space average speed of 1 min $\overline{v}_{s_{(n)}}$ is calculated by using the time speed $\overline{v}_{t_{(n)}}$ collected for six times, where $N = 6$.

$$\begin{aligned} \overline{v}_{t_{(n)}} &= \frac{1}{N} \sum_{i=n\cdot N+1}^{N(n+1)} \widetilde{v}_{(i)_{\text{modified}}} \\ \overline{v}_{s_{(n)}} &= \overline{v}_{t_{(n)}} - \frac{\sum_{i=n\cdot N+1}^{N(n+1)} \left(\widetilde{v}_{(i)_{\text{modified}}} - \overline{v}_{t_{(n)}} \right)}{N \cdot \overline{v}_{t_{(n)}}}. \end{aligned} \tag{5}$$

Finally, the number of CVs obtained by equipment counting is converted into the traffic flow $q_{(n)}$. Compared with the traffic volume $Q_{(n)}$, the penetration rate of CVs $r_{(n)}$ can be gotten as follows:

$$r_{(n)} = \frac{q_{(n)}}{Q_{(n)}}. \tag{6}$$

2.4. LSTM Model. LSTM includes input layers, output layers, and hidden layers. Figure 2 shows the framework and principle of the LSTM model. Our inputs include six speed data collected every 10 s in 1 min $v_{1_{(n)}}, v_{2_{(n)}}, \ldots, v_{6_{(n)}}$ and a penetration rate $r_{(n)}$. The output layer is one-minute spatial average speed of the road section. The hidden layers are realized by a carefully designed structure called 'gate' and neural network [41]. Specifically, 'forget gate' would determine the information to be discarded. The 'forget gate' reads hidden state $\mathbf{h}_{(n-1)}$ and $[v_{1_{(n)}}, v_{2_{(n)}}, \ldots, v_{6_{(n)}}, r_{(n)}]$ and outputs a forget gate weight $\mathbf{f}_{(n)}$ between 0 and 1.

$$\mathbf{f}_{(n)} = \sigma\left(\mathbf{W}_f \cdot \left[\mathbf{h}_{(n-1)}, v_{1_{(n)}}, v_{2_{(n)}}, \ldots, v_{6_{(n)}}, r_{(n)} \right] + \mathbf{b}_f \right), \tag{7}$$

where $\mathbf{W}_f$ is the input weight for the forget gate, and $\mathbf{b}_f$ is the bias for the forget gate.

The second step is to determine the information to be updated through the 'input gate' $\mathbf{i}_{(n)}$. And a new candidate value vector $\widetilde{C}_{(n)}$ is generated by 'tanh,' which could be updated into a new cell.

$$\begin{aligned} \mathbf{i}_{(n)} &= \sigma\left(\mathbf{W}_i \cdot \left[\mathbf{h}_{(n-1)}, v_{1_{(n)}}, v_{2_{(n)}}, \ldots, v_{6_{(n)}}, r_{(n)} \right] + \mathbf{b}_i \right) \\ \widetilde{C}_{(n)} &= \sigma\left(\mathbf{W}_C \cdot \left[\mathbf{h}_{(n-1)}, v_{1_{(n)}}, v_{2_{(n)}}, \ldots, v_{6_{(n)}}, r_{(n)} \right] + \mathbf{b}_C \right), \end{aligned} \tag{8}$$

where $\mathbf{W}_i$ is the input weight for the input gate, $\mathbf{b}_i$ is the bias for the input gate, $\mathbf{W}_C$ is the input weight for the current state, and $\mathbf{b}_C$ is the bias for the current state.

The third step is to update the new cell $\mathbf{C}_{(n)}$ by using a part of 'forget gate' selection and a part of 'input gate' selection.

$$\mathbf{C}_{(n)} = \mathbf{f}_{(n)} \cdot \mathbf{C}_{(n-1)} + \mathbf{i}_{(n)} \cdot \widetilde{C}_{(n)}. \tag{9}$$

Finally, 'sigmoid gate' is used to determine cell state output $\mathbf{o}_{(n)}$. The final output is obtained by multiplying $\mathbf{C}_{(n)}$ processed by 'tanh' and the output of 'sigmoid gate.'

$$\begin{aligned} \mathbf{o}_{(n)} &= \sigma\left(\mathbf{W}_o \cdot \left[\mathbf{h}_{(n-1)}, v_{1_{(n)}}, v_{2_{(n)}}, \ldots, v_{6_{(n)}}, r_{(n)} \right] + \mathbf{b}_o \right) \\ \mathbf{h}_{(n)} &= \mathbf{o}_{(n)} \cdot \tanh\left(\mathbf{C}_{(n)} \right), \end{aligned} \tag{10}$$

where $\mathbf{W}_o$ is the input weight for the output gate, and $\mathbf{b}_o$ is the bias for the output gate.

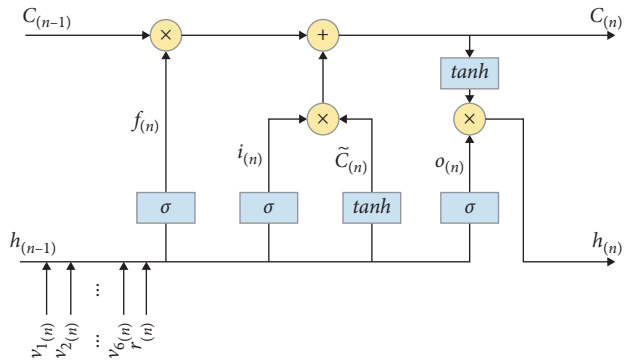


Figure 2: Framework and principle of LSTM model.

3. Simulation and Data

Since CVs are not popular on real roads now, the SUMO simulation platform is used to simulate the scene of mixed traffic. SUMO can interact with the outside world in real-time through TraCI interface, sending vehicles data. To approach the real road environment, our simulation has made the following settings:

(1) We set up a three-lane highway, in which there are ramps in the upstream and downstream, and our detection section is within 500 meters in the middle. The maximum capacity of one lane is 2000 veh/h, and the speed limit is 120 km/h (33.33 m/s);

(2) We assume that the characteristics of CVs are the same as conventional vehicles and only have the function of sensing and transmitting their own information; that is, functions of automatic driving and assisted driving are not considered;

(3) We set up a dynamic traffic flow and ramp control strategy to generate different traffic states. The traffic volume changes gradually over time, and all possible traffic states in the road environment are simulated with two peaks;

(4) The simulation time in this article is set to 7200 s, of which the first 600 s is warm-up time. Data are collected every 10 s, a total of 1380 groups, including speed, acceleration, and number of CVs.

We detect the real spatial speed and volume through sensors installed in the first place of the road section. The MFD is determined according to the characteristics of the road and the speed limit and size of the vehicle. On this simulated road, the maximum speed is 120 km/h, and the optimal traffic flow is 2000 veh/h. According to the Green Shields model, we can determine the relationship between speed and flow; that is, $Q_{(n)}\,(veh/h) = 218.18\overline{v}_{s_{(n)}} - 5.95\overline{v}^2{}_{s_{(n)}}$ (unit of $\overline{v}_{s_{(n)}}$: m/s). Finally, the calculated speed-volume relationship is compared with the detected real speed and flow relationship, and the results are basically consistent, which shows the feasibility of the simulation model.

Meanwhile, due to the low penetration rate of CVs and traffic volume, it may not be able to collect data at some moments. To analyze the missing data under different penetration rates and different flows, we show the distribution of missing data in Figure 3. The white line indicates vacant data, and the black line indicates data. Clearly, the severity of vacant data would decrease with the increase of traffic density and penetration rate. In the case of low penetration rate, even if there is a large traffic volume, there would still be many vacant data. This situation could be alleviated when the penetration rate reaches 0.03. At the same time, when the penetration rate reaches 0.3, the loss basically disappears, even at low traffic volume.

We use ACC-KNN data filling model to fill the original data. There is not a standard for selecting the best 'k' in KNN [42], so we determine it from the characteristics of the data set and the experimental results. From the perspective of the data set, firstly, there should be data in the k neighborhood under low penetration rate, and secondly, the influence interval of speed should be considered. Therefore, we take 'k' as 3–6. From the perspective of the experimental results, we set the value of 'k' to 3, 4, 5, and 6, respectively, and the mean difference between the filling speed and the real speed is 6.07 m/s, 5.73 m/s, 5.81 m/s, and 6.40 m/s, respectively. Therefore, we take 'k' as 4.

Figure 4 shows the real speed (red line), the original data (blue line) collected when the penetration rate is 0.05, and the filling speed (orange line and green line) when KNN filling and ACC-KNN filling are used. The results show that, in the case of 0.05 penetration rate, error with acceleration is reduced by 5.8% compared with that without acceleration. In three cases of dynamic penetration rate, the errors are reduced by 4.0%, 12.7%, and 7.9%, respectively. Clearly, our filling model is more effective than a single KNN filling model.

4. Experiment and Comparison

4.1. Speed Estimation. In this method, the LSTM model needs to learn multiple speed characteristics and penetration rate to estimate the whole speed, which is a multiple-input single-output model. Firstly, according to the spatial average speed and MFD, CVs penetration rate is estimated. Then, six speed data and the penetration rate are input into the LSTM model for estimation.

In our experiment, the first 115 groups of data are used for training, and the last 115 groups are used for testing. According to the results of many experiments, we set the parameters of the LSTM model. There are 3 hidden layers and 2 dense layers in LSTM, and the number of neurons in the hidden layer is 256, 256, and 128. Meanwhile, hyperbolic tangent function (tanh) is used for activation function, mean absolute error (MAE) is used for the loss function, adaptive moment estimation (Adam) optimizer is used for optimization, and the learning rate is set to 0.005.

We conducted traffic state estimation experiments using ACC-P-LSTM that LSTM model uses KNN filling with fusion acceleration and penetration rate estimation, P-LSTM that LSTM model uses penetration rate estimation, ACC-LSTM that LSTM model uses KNN filling with fusion acceleration, LSTM, ACC-KF that Kalman Filtering uses KNN filling with fusion acceleration, and Kalman

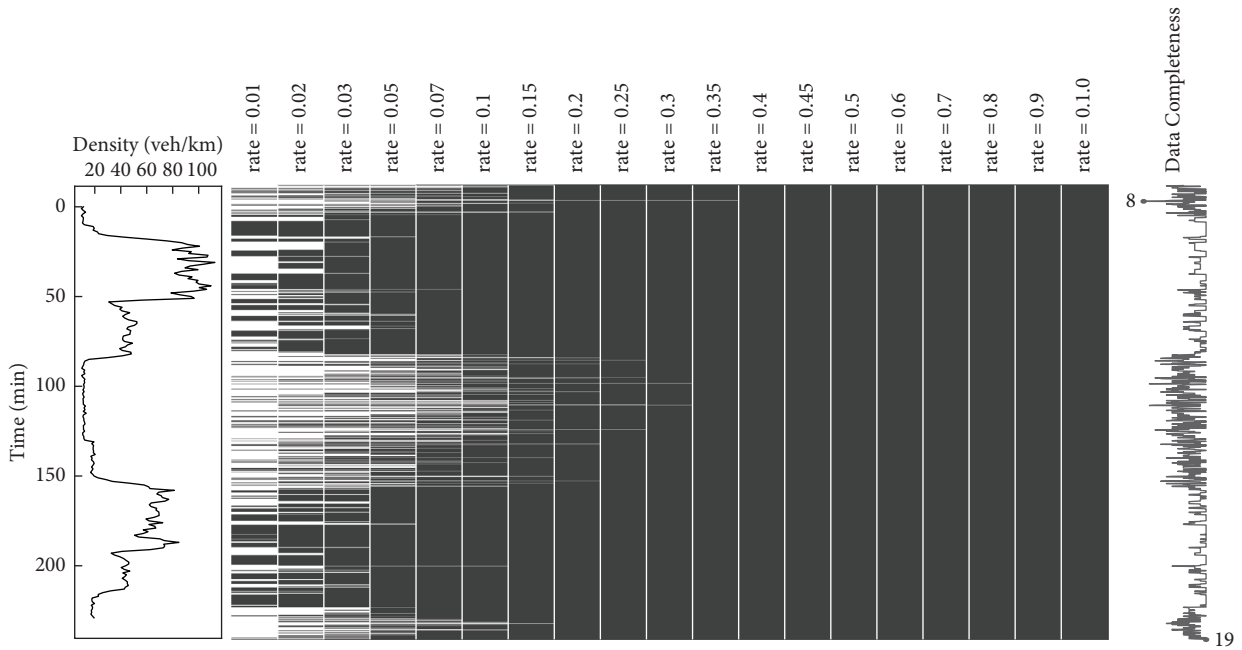


Figure 3: Vacant data of different penetration rates and density distribution map.

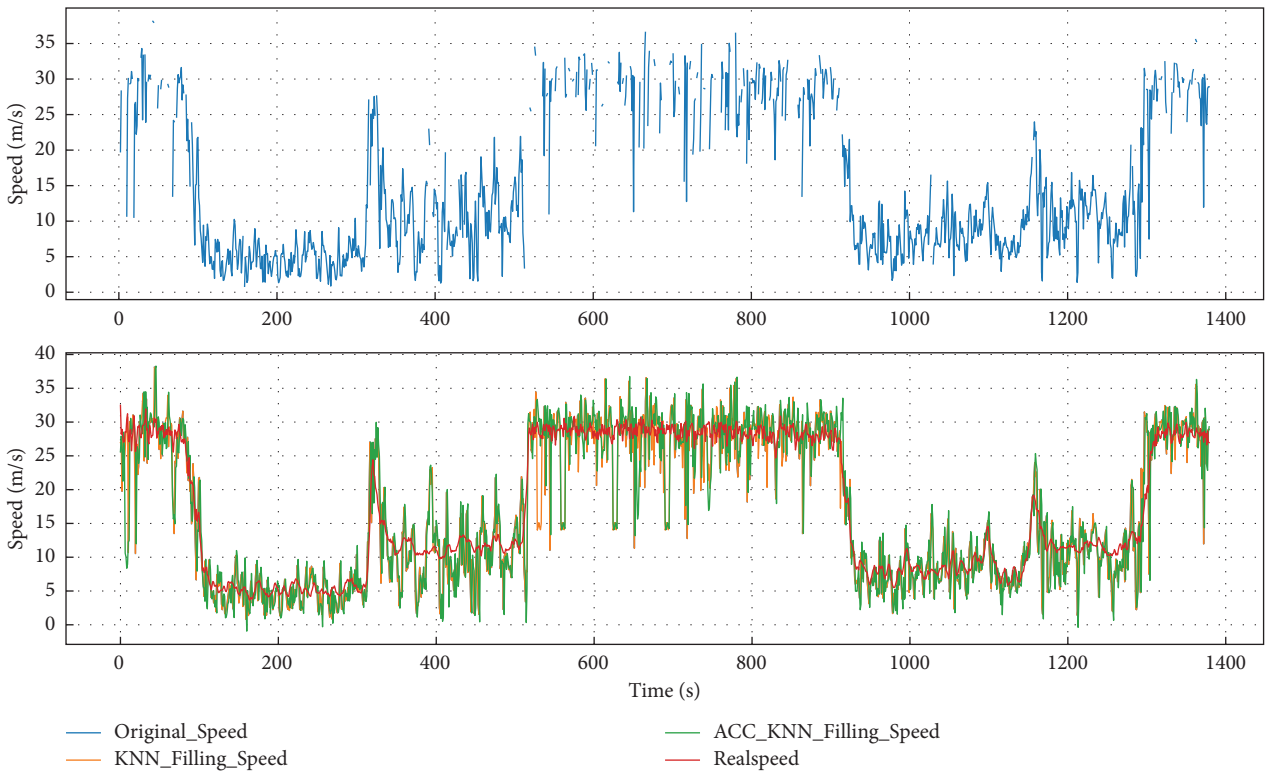


Figure 4: Speed before and after filling using different models. Note: KNN_Filling_speed = speed filled by K-Nearest Neighbors; and ACC_KNN_Filling_speed = speed filled by K-Nearest Neighbors combined with acceleration.

Filtering (KF), respectively. On the one hand, we set up cases of the penetration rate from 0.01 to 1.0 to compare these models' performance. On the other hand, we set up three groups of unknown and dynamic penetration rate cases to compare their performance. In these experiments, the performance of the traffic state estimation model is measured by three indicators: Root Mean Square Error (RMSE), Mean Relative Error (MRE), and Mean Absolute Error (MAE). Their calculation formulas are as follows:

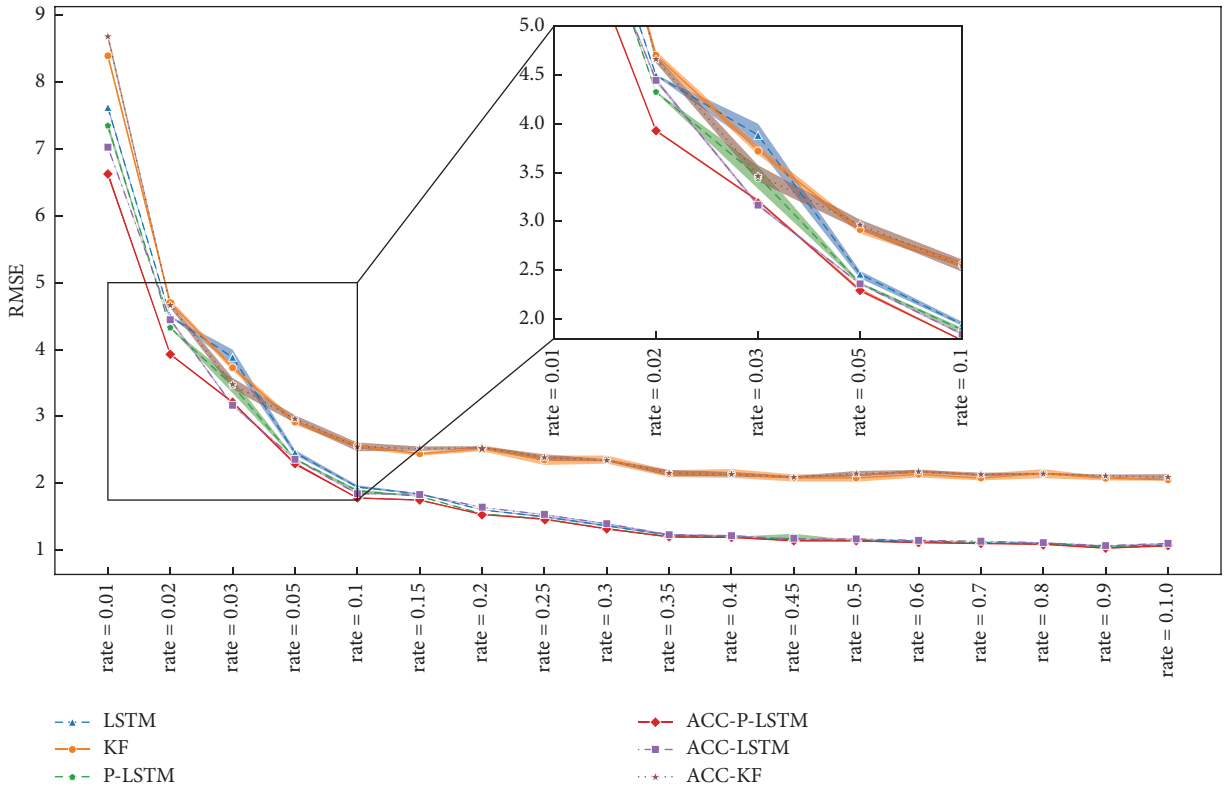


Figure 5: RMSE comparison of estimation results of various models under different penetration rates. Note: LSTM = Long Short-Term Memory; ACC-P-LSTM = LSTM model using KNN filling with fusion acceleration and penetration rate estimation; P-LSTM = LSTM model using penetration rate estimation; ACC-LSTM = LSTM model using KNN filling with fusion acceleration; ACC-KF = Kalman Filtering using KNN filling with fusion acceleration; KF = Kalman Filtering; and RMSE = Root Mean Square Error.

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(v_{s_{(n)}} - v_{(n)}\right)^2}, \\ \text{MRE} &= \frac{1}{N}\sum_{n-1}^{N}\frac{\left|v_{s_{(n)}} - v_{(n)}\right|}{v_{(n)}}, \\ \text{MAE} &= \frac{1}{N}\sum_{n=1}^{N}\left|v_{s_{(n)}} - v_{(n)}\right|. \end{aligned} \tag{11}$$

Figure 5 shows the RMSE of these models running results for five times and their trend in the 95% confidence range from rate = 0.01 to rate = 1.0. It can be found that the LSTM algorithm is better than the Kalman filter algorithm, especially in the cases of the high penetration rate. ACC-P-LSTM has the best prediction effect, especially in the cases of the low penetration rate. Specifically, when the penetration rate is 0.01, 0.02, 0.03, 0.05, 0.10, 0.15, and 0.20, the error of the ACC-P-LSTM is reduced by 13.5%, 12.6%, 18.3%, 7.8%, 7.6%, 5.6%, and 4.5%, respectively, compared with the single LSTM. When the penetration rate exceeds 0.35, all LSTM models show almost the same effect, but they are still half the error of the KF model. In addition, the effect of using KNN filling with fusion acceleration is greater than that of using penetration rate. This may be caused by the fixed penetration rate in these cases we set. The fixed penetration rate does not give full play to the advantages of the P-LSTM model, but it still improves the accuracy to a limited extent. In short, we believe that this data filling method and penetration rate estimation method proposed can improve the estimation effect under the condition of low CVs penetration rate.

To verify the estimation effect under the condition of dynamic penetration rate, we set up three groups of experiments that their penetration rate could change with time. In these three cases, our penetration rate is set to change in a reasonable interval, and the average penetration rate is kept below 30%, because the accuracy of models above 30% is not different. We also use ACC-P-LSTM, P-LSTM, ACC-LSTM, LSTM, ACC- KF, and KF conducted experiments and counted their RMSE, MRE, and MAE to compare, as shown in Table 1. It can be seen that, in three cases, ACC-P-LSTM shows good results, while other effects are that ACC-LSTM is similar to P-LSTM and better than the Kalman filter algorithm. It can be seen that the accuracy of the model after using acceleration to correct the data and the model with penetration rate as input can achieve similar improvements. This result is different from the first experiment of fixed penetration rates, which reflects that the P-LSTM model can well perceive changes in penetration rates. At the same time, these two improvement measures are complementary to each other, and the model that uses them both achieves the best results. In short, ACC-P-LSTM can be considered that it achieves significant performance in cases of dynamic penetration rate.

Table 1: Error comparison of different models in three cases (the results with the best performance are highlighted).

Case	Algorithm	RMSE	MRE	MAE
Case 1	KF	2.32031	0.15474	1.81752
	ACC-KF	2.27986	0.13954	1.74709
	LSTM	1.80295	0.09115	1.29157
	ACC-LSTM	1.62693	0.06925	1.17177
	P-LSTM	1.71519	0.07896	1.16926
	ACC-P-LSTM	**1.54142**	**0.06452**	**1.05299**
Case 2	KF	2.20117	0.10711	1.59424
	ACC-KF	2.17710	0.10660	1.58011
	LSTM	1.26577	0.06295	0.99065
	ACC-LSTM	1.25447	0.06122	0.98373
	P-LSTM	1.24782	**0.06019**	**0.95079**
	ACC-P-LSTM	**1.24741**	0.06022	0.96165
Case 3	KF	2.64087	0.16417	1.92558
	ACC-KF	2.52561	0.15863	1.78602
	LSTM	1.93180	0.09619	1.36391
	ACC-LSTM	1.92861	0.09618	1.34253
	P-LSTM	**1.91543**	**0.09422**	1.35214
	ACC-P-LSTM	1.91870	0.09661	**1.33502**

Note: KF = Kalman Filtering; ACC-KF = Kalman Filtering using KNN filling with fusion acceleration; LSTM = Long Short-Term Memory; ACC-LSTM = LSTM model using KNN filling with fusion acceleration; P-LSTM = LSTM model using penetration rate estimation; ACC-P-LSTM = LSTM model using KNN filling with fusion acceleration and penetration rate estimation; RMSE = Root Mean Square Error; MRE = Mean Relative Error; and MAE = Mean Absolute Error.

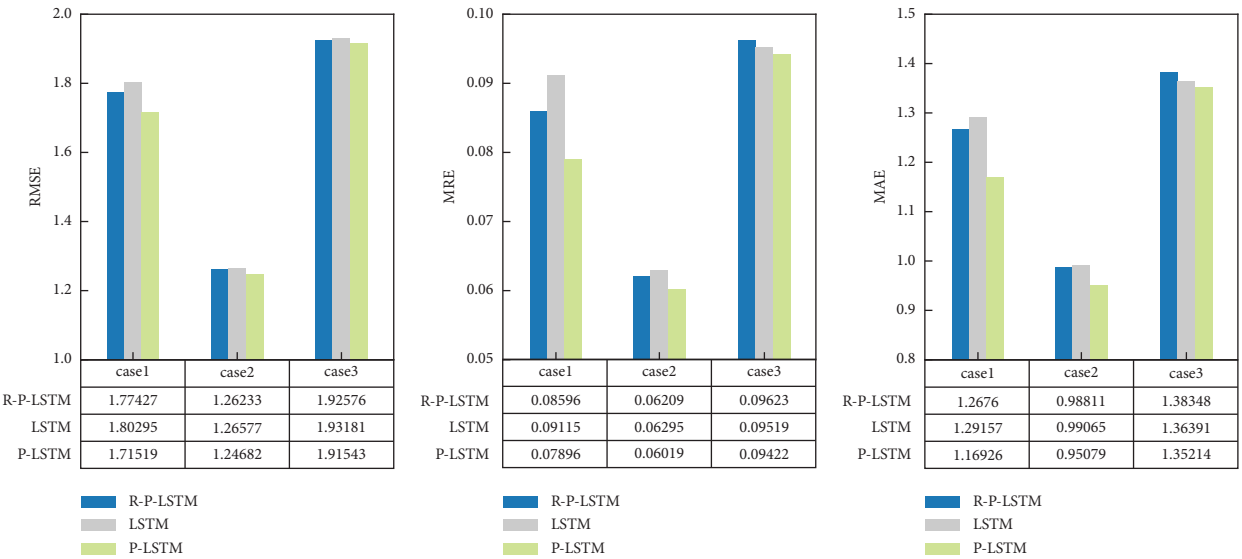


Figure 6: Error bar graph of error of R-P-LSTM, LSTM, and P-LSTM in three cases. Note: LSTM = Long Short-Term Memory; P-LSTM = LSTM model using penetration rate estimation; and R-P-LSTM = LSTM model using random penetration rate; RMSE = Root Mean Square Error; MRE = Mean Relative Error; and MAE = Mean Absolute Error.

In our model, estimation of the penetration rate is difficult and inaccurate. It is necessary to exclude that the increased accuracy of adding penetration rate is caused by the addition of noise. Therefore, we have added a set of experiments with random penetration rate, that is, replacing the penetration rate with random noise from 0 to 1. We use LSTM, P-LSTM, and LSTM with random penetration rate as an input (R-P-LSTM) for comparison experiments. Three cases are also set up, and the estimation error is shown in Figure 6. In case 1 and case 2, the error of the R-P-LSTM is lower than that of the LSTM, but the error of the P-LSTM decreases more. Meanwhile, the random penetration rate in case 3 is counterproductive. Therefore, we can eliminate the interference of random noise and believe that the addition of the estimated penetration rate could help improve the estimation accuracy. This model can learn the impact of different penetration rates on the accuracy of estimation and can autonomously adjust and weaken this impact.

4.2. Traffic State Judgment. The final task is to judge the state of traffic flow. The speed data can be directly obtained only

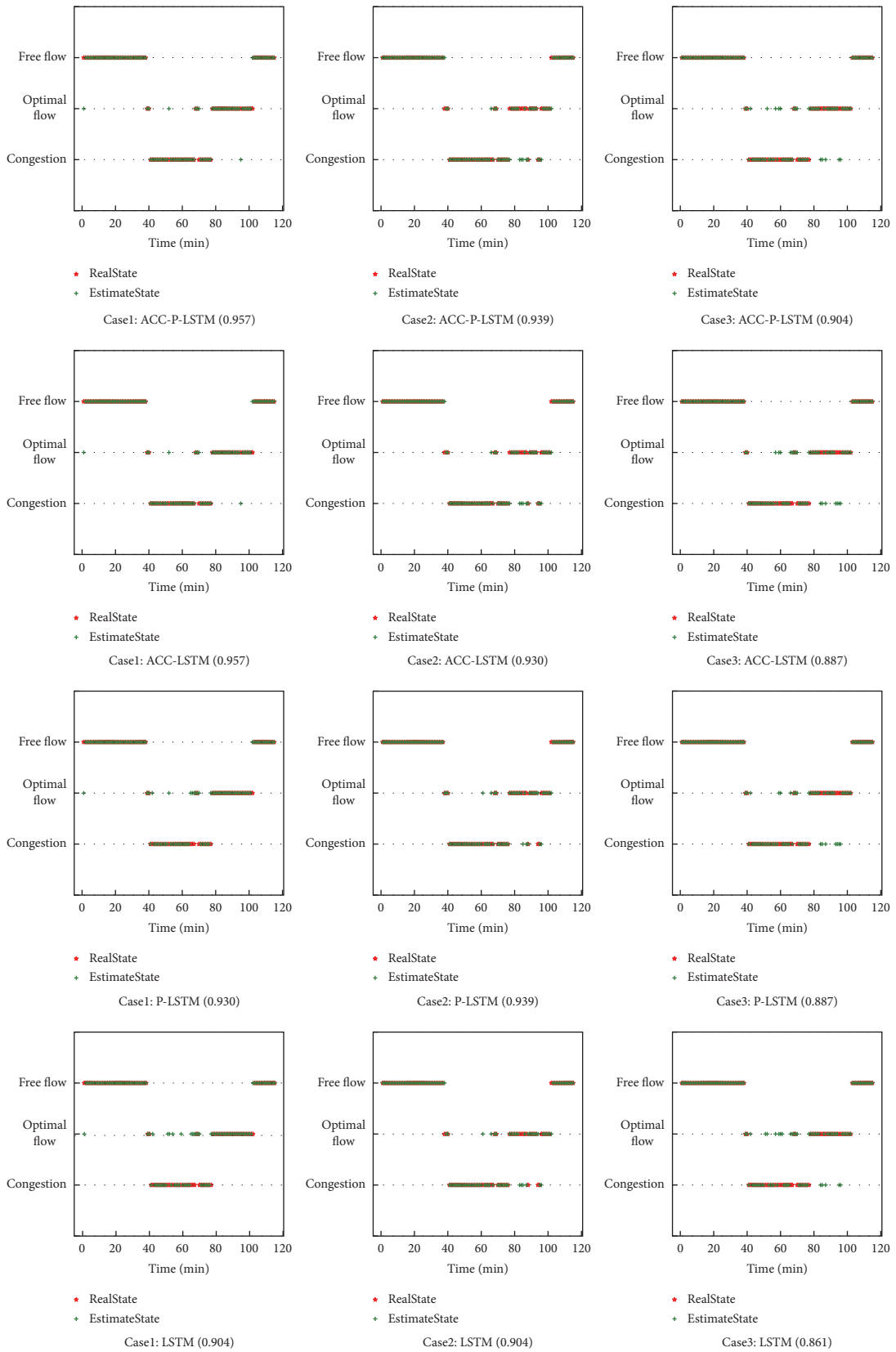


Figure 7: Comparison of traffic state estimation of various models under different penetration rates (the accuracy is recorded in the brackets of the subgraphs). Note: LSTM = Long Short-Term Memory; KF = Kalman Filtering; ACC-P-LSTM = LSTM model using KNN filling with fusion acceleration and penetration rate estimation; P-LSTM = LSTM model using penetration rate estimation; ACC-LSTM = LSTM model using KNN filling with fusion acceleration; and ACC-KF = Kalman Filtering using KNN filling with fusion acceleration.

through CVs and can be used to judge the traffic state. We refer to the concept of operation efficiency proposed by Xu et al. [6] and believe that the operation efficiency of the road is related to its volume and speed, as shown in formula (12). Operation efficiency suggests that the best traffic state means that the number of vehicles is large, and speed is also fast at the same time. Therefore, the formula of operation efficiency $E_{(n)}$ can be expressed as

$$E_{(n)} = Q_{(n)} \cdot \overline{v}_{s_{(n)}} = K_j \cdot \overline{v}^2_{s_{(n)}} - \frac{K_j \cdot \overline{v}^3_{s_{(n)}}}{V_f}. \tag{12}$$

Upper two-side $\overline{v}_{s_{(n)}}$ derivation:

$$\dot{E}_{(n)} = 2K_j \cdot \overline{v}_{s_{(n)}} - \frac{3K_j \cdot \overline{v}^2_{s_{(n)}}}{V_f}. \tag{13}$$

Order $\dot{E}_{(n)} = 0$; you can get $\overline{v}_{s_{(n)}} = 2/3V_f$. Then, find the derivative, and let $\ddot{E}_{(n)} = 0$; you can get $\overline{v}_{s_{(n)}} = 1/3V_f$.

Taking $2/3V_f$ and $1/3V_f$ as the dividing line, we divide the traffic state into free flow, optimal flow, and congestion. Therefore, in this experiment, free flow state is speed (22.22 m/s, 33.33 m/s), optimal flow state is speed (11.11 m/s, 22.22 m/s), and congestion state is speed (0, 11.11 m/s).

Three dynamic penetration rate cases are used as validation experiments. The results are shown in Figure 7. In three cases, ACC-P-LSTM has achieved better results than other models. Specifically, the average accuracy of ACC-P-LSTM, ACC-LSTM, P-LSTM, and LSTM in the three cases is 93.91%, 93.33%, 93.04%, and 91.88%. In addition, estimation errors are mostly in optimal flow and congestion. These two states have large random fluctuations and fuzzy boundaries, which are the main reasons for errors. Moreover, due to the randomness and dynamics of CVs, vehicle state performance is difficult to show the whole state, especially at low penetration rates. For example, in the case of heavy traffic, it takes a certain time for aggregation wave to reach the rear of the vehicle platoon. Further, if there are only CVs in the rear, the result would be inaccurate. This is also a reason why the penetration rate of CVs is introduced in this article. In general, our model has achieved good performance under a dynamic penetration rate.

5. Conclusion

CVs environment provides a new way to obtain data, but it is still difficult to estimate the traffic state under low or dynamic penetration rates of CVs. Due to the small amount of acquisition and unstable representativeness, the data transmitted by CVs need to be filled effectively and evaluated. In this article, a KNN data filling model integrating acceleration and an LSTM speed estimation model introducing penetration rate estimation are proposed. Specifically:

(1) Acceleration of CVs is used for data filling to fuse the time feature of speed, while KNN is used to mine distribution features. The results show that adding acceleration to the speed filling reduces the error by about 10%.

(2) The volume is estimated according to spatial speed based on MFD, inferring the penetration rate of CVs. The penetration rate that is directly related to estimation accuracy is taken as an input factor of the speed estimation model. Further, we construct a multifactor LSTM model that extracts the features of multisource data to estimate space speed. The speed estimation results show that the accuracy of using these strategies is better than that of not using these strategies. Especially when the penetration rate is lower than 30%, the ACC-P-LSTM model has the lowest speed estimation error and maintains the best stability.

(3) The speed is used to divide the traffic state, which can directly show traffic characteristics without additional sensor information. The results show that the accuracy of our model (average accuracy 0.933) is better than benchmark models (average accuracy 0.890) in various cases.

These can illustrate that the acceleration is effective for filling vacant data, and the penetration rate can improve the accuracy of estimation. But this article does not further study the correlation between them. The influence of penetration rate can be discussed in detail. At the same time, the performance of this estimation model is different under different traffic states. Especially under the optimal flow and congestion state, the probability of model error is high. The estimation accuracy in heavy traffic flow can be strengthened in the future. Besides, dynamic characteristics of CVs different from conventional vehicles are not considered in our model, which may affect speed estimation. Future research can consider the influence of the dynamic performance of CVs.

Data Availability

The data used to support the findings of this study were generated from SUMO simulation software and could be 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 work was supported by the National Key Research and Development Program of China (2019YFB1600100).

References

[1] S. M. Khan, K. C. Dey, and M. Chowdhury, "Real-time traffic state estimation with connected vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1687–1699, 2017.

[2] F.-Y. Wang, "Parallel control and management for intelligent transportation systems: concepts, architectures, and

applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 630–638, 2010.

[3] H. Shi, Y. Zhou, K. Wu, X. Wang, Y. Lin, and B. Ran, "Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment," *Transportation Research Part C: Emerging Technologies*, vol. 133, Article ID 103421, 2021.

[4] Y. Zhou, S. Ahn, M. Wang, and S. Hoogendoorn, "Stabilizing mixed vehicular platoons with connected automated vehicles: an H-infinity approach," *Transportation Research Part B: Methodological*, vol. 132, pp. 152–170, 2020.

[5] A. Sunderrajan, V. Viswanathan, W. Cai, and A. Knoll, "Traffic state estimation using floating car data," *Procedia Computer Science*, vol. 80, pp. 2008–2018, 2016.

[6] F. Xu, Z. He, Z. Sha, W. Sun, and L. Zhuang, "Traffic state evaluation based on macroscopic fundamental diagram of urban road network," *Procedia - Social and Behavioral Sciences*, vol. 96, pp. 480–489, 2013.

[7] T. Seo and T. Kusakabe, "Probe vehicle-based traffic state estimation method with spacing information and conservation law," *Transportation Research Part C: Emerging Technologies*, vol. 59, no. OCT, pp. 391–403, 2015.

[8] C. Nanthawichit, T. Nakatsuji, and H. Suzuki, "Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1855, no. 1, pp. 49–59, 2003.

[9] F. Ding, Z. Zhang, Y. Zhou, X. Chen, and B. Ran, "Large-Scale full-coverage traffic speed estimation under extreme traffic conditions using a big data and deep learning approach: case study in China," *Journal of Transportation Engineering, Part A: Systems*, vol. 145, no. 5, Article ID 05019001, 2019.

[10] S. He, F. Ding, Y. Zhou, Y. Cheng, and B. Ran, "Investigating and modelling the relationship between traffic volume and extracts from cellphone activity data," *Intelligent Transport Systems*, IET, vol. 13, 2019.

[11] S. Papadopoulou, C. Roncoli, N. Bekiaris-Liberis, I. Papamichail, and M. Papageorgiou, "Microscopic simulation-based validation of a per-lane traffic state estimation scheme for highways with connected vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 441–452, 2018.

[12] Z. M. Fadlullah, F. Tang, B. Mao et al., "State-of-the-Art deep learning: evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2432–2455, 2017.

[13] J. Barrachina, "V2X-d: a vehicular density estimation system that combines V2V and V2I communications," in *Proceedings of the 2013 IFIP Wireless Days (WD)*, pp. 1–6, IEEE, Valencia, Spain, November, 2013.

[14] Y. Lin, Y. Zhou, S. Yao, F. Ding, and P. Wang, "Real-time fine-grained freeway traffic state estimation under sparse observation," in *Proceedings of the ECML/PKDD*, no. 1, pp. 561–577, Ghent, Belgium, November, 2020.

[15] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, "Highway traffic state estimation with mixed connected and conventional vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3484–3497, 2016.

[16] N. Geroliminis and J. Sun, "Properties of a well-defined macroscopic fundamental diagram for urban traffic," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 605–617, 2011.

[17] S. He, X. Guo, F. Ding, Y. Qi, and T. Chen, "Freeway traffic speed estimation of mixed traffic using data from connected and autonomous vehicles with a low penetration rate," *Journal of Advanced Transportation*, vol. 2020, no. 1178, pp. 1–13, 2020.

[18] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, "Highway traffic state estimation per lane in the presence of connected vehicles," *Transportation Research Part B: Methodological*, vol. 106, pp. 1–28, 2017.

[19] F. Vuolo, W.-T. Ng, and C. Atzberger, "Smoothing and gap-filling of high resolution multi-spectral time series: example of Landsat data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 57, pp. 202–213, 2017.

[20] C. Zhang, J. Kai, H. C. Feng, and T. Yang, "The nearest neighbor algorithm of filling missing data based on cluster Analysis," *Applied Mechanics and Materials*, vol. 347, no. 2, pp. 2324–2328, 2013.

[21] X. Qi, H. Guo, and W. Wang, "A reliable KNN filling approach for incomplete interval-valued data," *Engineering Applications of Artificial Intelligence*, vol. 100, no. 315, p. 104175, 2021.

[22] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research Part C: Emerging Technologies*, vol. 28, pp. 15–27, 2013, mar..

[23] C. N. Babu, P. Sure, and C. M. Bhuma, "Sparse Bayesian learning assisted approaches for road network traffic state estimation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1733–1741, 2020.

[24] Y. Wang, Z. Yao, Y. Cheng, Y. Jiang, and B. Ran, "Kalman filtering method for real-time queue length estimation in a connected vehicle environment," *Transportation Research Record*, vol. 2675, Article ID 03611981211011996, 2021.

[25] T. Seo, A. M. Bayen, T. Kusakabe, and Y. Asakura, "Traffic state estimation on highway: a comprehensive survey," *Annual Reviews in Control*, vol. 43, pp. 128–151, 2017.

[26] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach," *Transportation Research Part B*, vol. 39, 2007.

[27] M. Fountoulakis, N. Bekiaris-Liberis, C. Roncoli, I. Papamichail, and M. Papageorgiou, "Highway traffic state estimation with mixed connected and conventional vehicles: microscopic simulation-based testing," *Transportation Research Part C: Emerging Technologies*, vol. 78, pp. 13–33, 2017.

[28] C. F. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[29] C. M. J. Tampere and L. H. Immers, "An extended kalman filter application for traffic state estimation using CTM with implicit mode switching and dynamic parameters," in *Proceedings of the Intelligent Transportation Systems Conference*, September, 2007.

[30] C. Antoniou, H. N. Koutsopoulos, and G. Yannis, "Dynamic data-driven local traffic state estimation and prediction," *Transportation Research Part C: Emerging Technologies*, vol. 34, pp. 89–107, 2013.

[31] X. Zhan, R. Li, and S. V. Ukkusuri, "Link-based traffic state estimation and prediction for arterial networks using license-plate recognition data," *Transportation Research Part C: Emerging Technologies*, vol. 117, Article ID 102660, 2020.

[32] I. Laña, J. L. Lobo, E. Capecci, J. Del Ser, and N. Kasabov, "Adaptive long-term traffic state estimation with evolving spiking neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 126–144, 2019.

[33] W. Wei, H. Wu, and H. Ma, "An AutoEncoder and LSTM-based traffic flow prediction method," *Sensors*, vol. 19, no. 13, p. 2946, 2019.

[34] R. Vinayakumar, K. P. Soman, and P. Poornachandran, "Applying deep learning approaches for network traffic prediction," in *Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics*, September, 2017.

[35] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values," *Transportation Research Part C: Emerging Technologies*, vol. 118, 2020.

[36] X. Du, H. Zhang, H. V. Nguyen, and H. Zhu, "Stacked LSTM deep learning model for traffic prediction in vehicle-to-vehicle communication," in *Proceedings of the 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Toronto, ON, Canada, September, 2017.

[37] Q.-J. Qing-Jie Kong, Z. Zhipeng Li, Y. Yikai Chen, and Y. Yuncai Liu, "An approach to urban traffic state estimation by fusing multisource information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 3, pp. 499–511, 2009.

[38] Y. Yuan, H. V. Lint, F. V. Wageningen-Kessels, and S. Hoogendoorn, "Network-Wide traffic state estimation using loop detector and floating car data," *ITS Journal*, vol. 18, no. 1-4, pp. 41–50, 2014.

[39] H. Chen, H. A. Rakha, and S. Sadek, "Real-time freeway traffic state prediction: a particle filter approach," in *Proceedings of the International IEEE Conference on Intelligent Transportation Systems*, Washington, DC, USA, October, 2011.

[40] V. Knoop, S. P. Hoogendoorn, and H. van Zuylen, "Empirical differences between time mean speed and space mean speed," in *Traffic and Granular Flow'07*, pp. 351–356, Springer, New York, NY, USA, 2009.

[41] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," *Computer Science*, vol. 1402, p. 1128, 2014.

[42] U. Pujianto, A. P. Wibawa, and M. I. Akbar, "K-nearest neighbor (K-nn) based missing data imputation," in *Proceedings of the 2019 5th International Conference on Science in Information Technology (ICSITech)*, pp. 83–88, IEEE, Yogyakarta, Indonesia, October, 2019.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 2259113, 12 pages
https://doi.org/10.1155/2022/2259113

Research Article

Improved Real-Time Traffic Obstacle Detection and Classification Method Applied in Intelligent and Connected Vehicles in Mixed Traffic Environment

Luyao Du,[1] **Xiongjie Chen**,[2] **Zhonghui Pei**,[3] **Donghua Zhang**,[4] **Bo Liu**,[4] **and Wei Chen**[1]

[1]*School of Automation, Wuhan University of Technology, Wuhan 430070, China*
[2]*Department of Computer Science, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU2 7XH, UK*
[3]*School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China*
[4]*Wuhan Zhongyuan Electronics Group Co., Ltd., Wuhan 430205, China*

Correspondence should be addressed to Wei Chen; greatchen@whut.edu.cn

Received 22 January 2022; Revised 20 March 2022; Accepted 24 March 2022; Published 7 April 2022

Academic Editor: Gen Li

Mixed traffic is a common phenomenon in urban environment. For the mixed traffic situation, the detection of traffic obstacles, including motor vehicle, non-motor vehicle, and pedestrian, is an essential task for intelligent and connected vehicles (ICVs). In this paper, an improved YOLO model is proposed for traffic obstacle detection and classification. The YOLO network is used to accurately detect the traffic obstacles, while the Wasserstein distance-based loss is used to improve the misclassification in the detection that may cause serious consequences. A new established dataset containing four types of traffic obstacles including vehicles, bikes, riders, and pedestrians is collected under different time periods and different weather conditions in urban environment in Wuhan, China. Experiments are performed on the established dataset on Windows PC and NVIDIA TX2, respectively. From the experimental results, the improved YOLO model has higher mean average precision than the original YOLO model and can effectively reduce intolerable misclassifications. In addition, the improved YOLOv4-tiny model has a detection speed of 22.5928 fps on NVIDIA TX2, which can basically realize the real-time detection of traffic obstacles.

1. Introduction

Traffic conditions in urban societies can be highly complex, since vehicles, pedestrians, and riders may be on the same road, especially in developing countries. In recent years, the rapid rise of bike-sharing and take-away industries has aggravated this phenomenon to a certain extent. The co-existence of vehicles, bikes, riders, and pedestrians has brought great challenges to driving safety in urban areas. The detection and classification of vehicles, bikes, riders, and pedestrians is essential for ICVs [1].

Vision-based object detection and classification is an important method to achieve traffic obstacle detection and classification. In traditional object detection methods, traditional machine learning methods such as scale-invariant feature transform (SIFT) [2] and histogram of oriented gradients (HOG) [3] extract the object features and input the extracted features into the classifiers like support vector machine (SVM) [4] and AdaBoost [5]. The design of these features can be very complicated; in particular, these features are handcrafted features, and their performances are task-dependent, which is not scalable to large-scale applications and can hardly be generalized. At this stage, traditional machine learning object detection methods can barely meet the requirements in practical applications; therefore, new target detection methods are needed. With the development of deep learning, many deep learning techniques have been

applied to the field of object detection, among which the deep convolutional neural network (CNN) [6] is the most prominent one. Unlike traditional feature extraction algorithm relying on domain knowledge, CNN has shown to be invariant to geometric transformation, deformation, and illumination, thus effectively overcoming the difficulties caused by the variability of non-motorized vehicle appearance. It also can adaptively capture complex feature patterns by learning from data, leading to its high flexibility and generalization ability. Many deep learning-based object detection methods were proposed in recent years, including one-stage and two-stage detection methods, as shown in Figure 1 [7]. One-stage detection algorithms, such as YOLO [8], SSD [9], and Retina-Net [10], do not need to predict region proposals. In particular, they directly generate the label and the location of objects. After a single test, the final detection result can be obtained in an end-to-end manner, so that the detection speed is faster. In contrast, the two-stage detection algorithm divides the detection problem into two stages. Firstly, the region proposals are generated, and then the regional proposals are classified. In most cases, the predicted positions need to be refined. One typical example of the two-stage algorithms is the family of R-CNN algorithm, which is based on the region proposal, including R-CNN [11], SPPNet [12], Fast R-CNN [13], Faster R-CNN [14], and FPN [15].

Among all the aforementioned state-of-the-art object detection algorithms, YOLOv3 [16] and YOLOv4 [17] are arguably the most promising approaches. Proposed by Redmon et al. in 2018 and by Bochkovskiy et al. in 2020, YOLOv3 and YOLOv4 have both high detection speed and accuracy and can be used for the detection and classification of traffic obstacles. Researchers have conducted many research studies on the detection of traffic obstacles based on YOLO [18–22]. Wang et al. [18] used YOLOv3 to detect vehicles, pedestrians, and non-motor vehicles, which improved the detection accuracy. Narayanan et al. [19] proposed a model using HOG and YOLO algorithm for pedestrian detection in thermal images. Hung et al. [20] performed real-time obstacle detection with the YOLO model on an embedded system. Wang [21] proposed a real-time vehicle detection algorithm that integrates vision and lidar point cloud information, which achieved high detection accuracy and good real-time performance. Arvind et al. [22] developed a near-range obstacle sensing system based on vision sensor, which can ensure early detection and tracking of the obstacle. Zhang et al. [23] proposed a classification method for four classes of moving object using 3D point cloud, which recognized the moving objects effectively. Feng et al. [24] presented a 32-layer multibranch method for object detection in traffic scenes, which achieved the state-of-the-art performance. Li et al. [25] proposed an improved multivehicle detection method considering traffic flow, which achieved good performance and robustness. Wang et al. [26] presented a vision-based crash detection framework in mixed traffic flow environment, which achieved a high detection rate with relatively low false alarm rate. Cai et al. [27] presented an improved framework for object detection based on YOLOv4. Hnewa et al. [28] outlined the state-of-the-art frameworks for object detection under rainy conditions. Liu et al. [29] proposed a radar and camera information fusion method for object recognition. Bell et al. [30] presented a real-time system for night time vehicle detection. Satyanarayana et al. [31] proposed a vehicle method for heterogeneous and lane-less traffic. However, the above research seldom carried out on-vehicle real-time detection and classification of traffic obstacles based on the target characteristics of real hybrid traffic scenes, and the detection accuracy and real-time performance can be further improved.

In the task of traffic obstacle identification and classification, each misclassification is considered to be the same in terms of the potential costs it may bring. However, in actual applications, different misclassifications can result in significant different consequences for ICVs, and some may only lead to minor mistakes, while the others can bring disastrous consequences. To improve the safety of ICVs and avoid disastrous consequences caused by wrong predictions, one may need to assign different weights to different mislabelled results. Recently, the application of Wasserstein distance in object detection system has attracted much attention from the machine learning community [32]. Wasserstein distance [33] is a measure of distance between probability distributions, combining with which the loss function of YOLO could effectively reduce the probability of producing intolerable misclassification in ICVs, thereby reducing the security risk caused by misclassification.

In this paper, an improved Wasserstein distance loss is proposed based on the YOLO model. The main contributions of this paper can be summarized as follows:

(i) A new dataset, containing traffic obstacles including vehicles, bikes, riders, and pedestrians under different time periods and different weather conditions in urban environment in Wuhan, China, is collected and established for detection.

(ii) Based on YOLO network, the improved model is designed for traffic obstacle detection. The Wasserstein distance-based loss, which assigns different weights for one sample classified to different classes with different values, so that the misclassified objects are classified to similar classes with a higher probability, is combined with the loss function of YOLO to enhance the performance of traffic obstacle detection.

(iii) The improved model is deployed on NVIDIA TX2 for real-time detection and then compared with the original model. Empirical experiments show that the improved model presents more accurate and robust results than the original model, and its real-time performance can basically meet the requirements of real-time detection applications.

The remainder of this paper is organized as follows. In Section 2, the dataset collected in real scenes is described. Section 3 presents the Wasserstein loss-based YOLO model, including the network architecture of the designed model and the loss function for training it. The experimental results

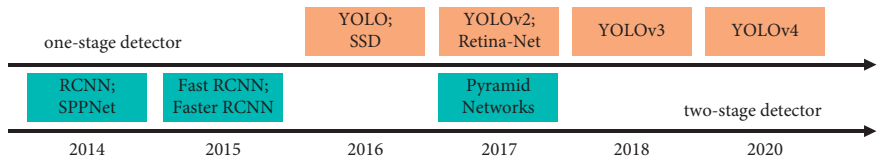


Figure 1: A road map of deep learning-based object detection methods.

are reported in Section 4. Finally, the conclusions are presented in Section 5.

2. Dataset

2.1. Data Acquisition. In order to achieve accurate and efficient traffic obstacle detection, image data specifically for traffic obstacles including vehicle, bike, rider, and pedestrian were collected by a camera at a 1920 ∗ 1080 pixel resolution in Wuhan, China. The collection was conducted during different time periods including daytime and nightfall, and the weather conditions included sunny and cloudy. 496 image data were selected as the original image data to establish the dataset.

2.2. Data Classification. In the urban hybrid traffic scenario, vehicle, bike, rider, and pedestrian are the main traffic obstacles that affect the driving safety of intelligent and connected vehicles. Therefore, as shown in Figure 2, the detection objects in the collected data are divided into these four categories.

2.3. Data Augmentation. As shown in Figure 3, in order to enrich the dataset and enhance the robustness, data augmentation operations including rotation and brightness transformation were performed on the image data. The dataset after data augmentation contains a total of 2976 image data in hybrid traffic scenes.

2.4. Data Annotation. After the above processing, the dataset was manually labelled. In the images, objects with contours less than 50% and small targets that cannot be seen clearly were not labelled. The detailed sample size of each category that has been labelled is shown in Table 1.

3. Methodology

3.1. YOLO Model. In this paper, the YOLO-based detection models, including YOLOv3, YOLOv4, and YOLOv4-tiny, are established. In the YOLOv3 model [16], the image is divided into S ∗ S grid cells, and the grid cell at the center of the object is responsible for completing the prediction of the object. In view of the large number of vehicles, bikes, riders, and pedestrians in the urban hybrid traffic environment and the large difference in scale, the model uses the multiple scale fusion method to make predictions. The features of the three detection scales with sizes of 13 ∗ 13, 26 ∗ 26 and 52 ∗ 52 are fused, so as to be compatible with large and small objects.

The network is mainly composed of a series of 1 x 1 and 3 x 3 convolutional layers (each convolutional layer is

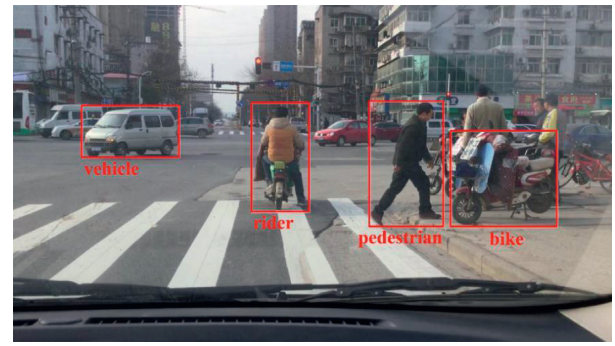


Figure 2: Object classification of traffic obstacles in urban hybrid traffic scenarios.

followed by a BN layer and a LeakyReLU layer). Three detections were performed in the network, which were performed during 32 times downsampling (25), 16 times downsampling (24), and 8 times downsampling (23). After the 79th layer of the convolutional network, it passes through several convolutional layers to obtain a scale of detection results. Compared with the input image, the feature map used for detection here has 32 times downsampling. Due to the high downsampling factor, the receptive field of the feature map here is relatively large, so it is suitable for detecting objects of relatively large size in the image data. In order to achieve fine-grained detection, start sampling from the feature map of the 79th layer and then fuse it with the feature map of the 61st layer (concatenation) to obtain a fine-grained feature map of the 91st layer, which also passes through several convolutional layers, and then get a 16 times downsampled feature map relative to the input image, which has a medium-scale receptive field and is suitable for detecting medium-scale objects. Finally, the 91st layer feature map is upsampled again and fused with the 36th layer feature map to obtain a feature map that is downsampled 8 times relative to the input image. It has the smallest receptive field and is suitable for detecting small-sized objects.

YOLOv4 [17] has made a series of improvements on the basis of YOLOv3, mainly including the following: the backbone feature extraction network is changed from DarkNet53 to CSPDarkNet53 [34], the feature pyramid is changed to SPP [35] and PAN [36], the classification regression layer is unchanged for YOLOv3, etc.

The YOLOv4-tiny network structure is a simplified version of YOLOv4, which is a lightweight model with only 6 million parameters equivalent to one-tenth of the original. As shown in Figure 4, the overall network structure has 38 layers, using three residual units, the activation function uses LeakyReLU, the classification and regression of the target are

Figure 3: Examples of data augmentation. (a) Original image data. (b) Rotated 180°. (c) Dimmed brightness. (d) Dimmed brightness and rotated. (e) Brightened brightness. (f) Brightened brightness and rotated.

Table 1: Detailed sample size of each category.

Category	Vehicle	Bike	Rider	Pedestrian
Sample size	10668	1440	5814	13182

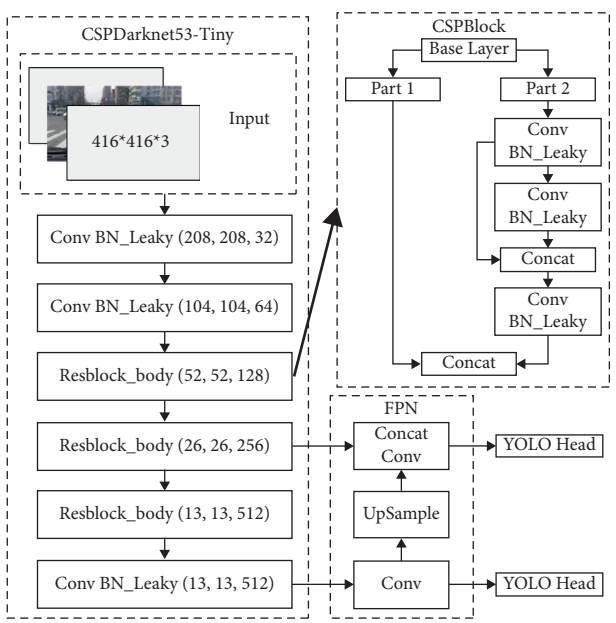


Figure 4: The network structure of YOLOv4-tiny model.

changed to use two feature layers, and the feature pyramid network (FPN) is used when merging the effective feature layers. It also uses the CSPnet structure, performs channel segmentation on the feature extraction network, divides the feature layer channel output after 3x3 convolution into two parts, and takes the second part. The detection speed of the YOLOv4-tiny model has been greatly improved, which makes it possible to be deployed on mobile embedded terminals such as NVIDIA TX2 for real-time detection.

3.2. Wasserstein Distance-Based Loss. To alleviate the undesirable consequences caused by misclassification, we propose to incorporate the Wasserstein distance into the framework of YOLO and apply it to ICVs. The Wasserstein distance is a metric for measuring the discrepancy or dissimilarity between probability measures, and it calculates the cost of moving one distribution to another one [37]. For discrete distributions $s = \{s_m\}_{m=1}^{N}$ and $t = \{t_n\}_{n=1}^{N}$, the Wasserstein distance between s and t can be formulated as follows:

$$W(s,t) = \inf_{\gamma \in \Gamma(s,t)} \sum_{m=1}^{N} \sum_{n=1}^{N} D_{m,n}\gamma_{m,n}, \tag{1}$$

where $\Gamma(s,t)$ is the set of all possible optimal transport plan between s and t and $D_{N\times N}$ is the distance matrix, whose element $D_{m,n}$ measures the distance between s_m and t_n. In particular, for arbitrary optimal transport plan γ, it has to satisfy

$$\begin{aligned} &\gamma(m,n) \geq 0, \\ &\sum_{n=1}^{N} \gamma(m,n) = s_m, \\ &\sum_{m=1}^{N} \gamma(m,n) = t_n, \end{aligned} \tag{2}$$

which implies that the optimal transport plan can also be interpreted as a joint distribution of s_m and t_n. To say it in another way, the Wasserstein distance tries to find the optimal joint distribution of s_m and t_n, which can produce the minimal cost of transporting s_m to t_n. Compared to other

distance metrics for probability measures such as Kullback–Leibler divergence, Hellinger distance, and Jensen–Shannon divergence, the Wasserstein distance has some favourable geometry properties. Firstly, it is a valid distance metric, i.e., it is symmetric and non-negative, and also satisfies the triangular inequality and identity of indiscernible. Secondly, it can capture the geometry in the underlying space [38].

In object detection, we consider the source distribution $s = \{s_m\}_{m=1}^{N}$ as the prediction of the probability distribution of the label of objects and $t = \{t_n\}_{n=1}^{N}$ as the ground truth of the label of objects. More specifically, in this paper, the discrepancy between the predictions produced by classifier and the ground-truth labels will be measured by the Wasserstein distance. In [39], it is proved that if either the source distribution or the target distribution is a one-hot histogram, there is only one possible transport plan, and the Wasserstein distance between the source distribution and the target distribution can be calculated by

$$W(s,t) = \sum_{i=1}^{N} \langle s_i - t_i, D_{c_i,1:C} \rangle, \tag{3}$$

where c_i is the index of the one-hot element in t_i, C is the number of object classes, and $D_{c_i,1:C}$ represents the c_i-th row of the distance matrix $D_{C\times C}$. The distance matrix, which specifies the distance between categories, needs to be predefined. In this paper, there are four categories in the dataset, namely, vehicle, bike, rider, and pedestrian. As discussed in the Introduction, different misclassifications may result in different consequences, and if the classifier is able to discriminate two different misclassifications, then disastrous consequences can be avoided. For example, classifying "bike" as "rider" may not change the decision made by autonomous driving system, as bike and rider share the same behaviour pattern in a large degree. However, classifying "bike" as "vehicle" is very likely to have significant influence on the decision-making process of a self-driving vehicle, not only because bike and vehicle are different objects but also because they are expected to have distinct trajectories. To prevent the above undesirable problem, in the proposed method, the distance matrix is defined as in Figure 5.

Denote by (P, P^*) the predicted location and the ground-truth location, (T, T^*) the predicted confidence and ground-truth confidence, and (F, F^*) the predicted class and ground-truth class; in the original Yolov3, the loss function is composed of three parts, the location loss $L_{\text{loc}}(P, P^*)$, confidence loss $L_{\text{conf}}(T, T^*)$, and classification loss $L_{\text{class}}(F, F^*)$. In this paper, we propose to use an additional loss, the Wasserstein loss, and thus the modified YOLO loss function becomes

$$\begin{aligned} L_{\text{YOLO}} = {} & L_{\text{loc}}(P, P^*) + L_{\text{conf}}(T, T^*) + L_{\text{class}}(F, F^*) \\ & + \lambda W(F, F^*, D), \end{aligned} \tag{4}$$

where λ is a hyperparameter that controls the weight of the Wasserstein distance.

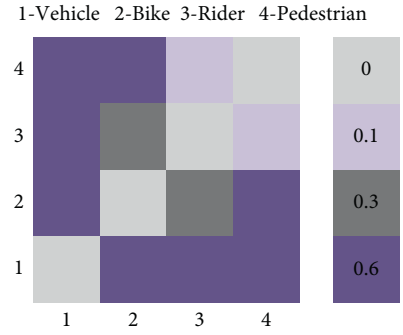


Figure 5: The distance matrix.

4. Experimental Results

4.1. Experimental Environment. The experiments were trained and tested on a Windows PC with two Intel Xeon processors, a CPU at 3.5 GHz, 128 G DDR4, and an NVIDIA GeForce RTX 2080 with 8 GB memory. The established dataset is divided into training set and test set at a ratio of 9 : 1. During training, all but three output layers were first frozen to get a stable loss and then unfrozen, and training was continued to fine-tune. To avoid overfitting, when the loss cannot be reduced within ten epochs, training is terminated. In addition, the original and improved YOLO models were performed on NVIDIA TX2 for real-time detection.

4.2. Evaluation Metric. In this study, the precision-recall curve (P-R curve), F1 score, and mean average precision (mAP) were used to evaluate the performance of the model.

The P-R curve is a curve composed of the value of precision (P) as the ordinate and the value of recall (R) as the abscissa, where P can be defined as

$$P = \frac{TP}{TP + FP}. \tag{5}$$

R can be defined as

$$R = \frac{TP}{TP + FN}, \tag{6}$$

where the definitions of TP, FN, and FP are shown in Table 2.

F1 score, an index that comprehensively considers the values of P and R to reflect the performance of the detection model, can be defined as

$$\text{F1} = \frac{2 \times P \times R}{P + R}. \tag{7}$$

The area under the P-R curve is the value of the average precision (AP), and the AP value over four categories of the obstacle objects in the hybrid traffic scene is defined as mAP. The AP and mAP value can be defined as

$$\begin{aligned} AP &= \int_0^1 P(R)\mathrm{d}R, \\ mAP &= \frac{1}{4} \sum_{c \in \text{categories}} AP(c). \end{aligned} \tag{8}$$

Table 2: Definition of different detection results.

Labelled	Detected	Definition
Positive	Positive	TP
Positive	Negative	FN
Negative	Positive	FP
Negative	Negative	TN

4.3. Result of Designed Models on Established Dataset. In order to verify the detection effect of the designed models, the models including YOLOv3, YOLOv4, and YOLOv4-tiny were performed on the four categories of obstacle objects. The loss curves of the designed models are shown in Figures 6, 7, and 8, respectively.

It can be seen from the loss curves that the loss value of the improved model is higher than that of the original model at the beginning of training, and the loss value of the improved model and original model is basically the same when the loss value stabilizes. This is because of the addition of Wasserstein distance-based loss to the improved model. The final loss values of the YOLOv3, YOLOv4, and YOLOv4-tiny models are about 24.5, 10, and 11.5, respectively.

The experimental results of designed models are shown in Table 3, and the P-R curves are shown in Figure 9. It can be seen from the experimental results that the mAP of the improved YOLOv3, YOLOv4, and YOLOv4-tiny models is 98.57%, 98.19%, and 80.39%, respectively, slightly higher than that of each original model, and the F1 value of the improved models is basically the same as each original model.

4.4. Result of Designed Models on BDD Dataset. BDD is one of the latest published autonomous driving datasets with dense traffic scenes, on which the detection effect of the designed models is also verified. In the BDD dataset, there are few objects in the bike and rider categories, so we selected the data containing these two categories of objects for testing to maintain the relative balance between the various categories. The experimental results of designed models are shown in Table 4, and the P-R curves are shown in Figure 10.

It can be seen from the experimental results that the mAP of the improved YOLOv3, YOLOv4, and YOLOv4-tiny models is 92.97%, 91.23%, and 77.97%, respectively, higher than that of each original model, and the F1 value of the improved models is basically the same as each original model. The detection mAP value of the designed model on the BDD dataset is slightly lower than that on the established dataset. This is because the training of the model is carried out on the training set in the established dataset, which is similar to the testing set scene but different from the BDD dataset scene. However, the detection results on both datasets could meet the basic application requirements.

4.5. The Application-Oriented Performance on NVIDIA TX2. NVIDIA TX2 is a mobile terminal that can be deployed directly on the vehicle. The vehicle application scenarios on NVIDIA TX2 are shown in Figure 11. The trained improved and original models are deployed on NVIDIA TX2, respectively, and then tested on the established dataset. In addition, the NVIDIA TX2 with a camera is installed on the vehicle for real-time detection to verify the detection effect and real-time performance of the proposed model.

The detection speed of different models is shown in Table 5. As can be seen from the table, the detection speed

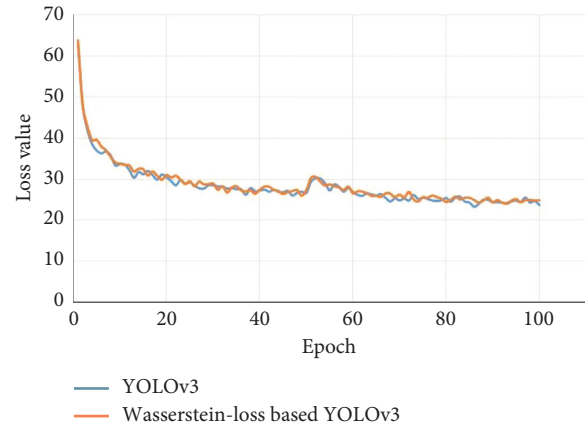


Figure 6: The loss curve of YOLOv3 model.

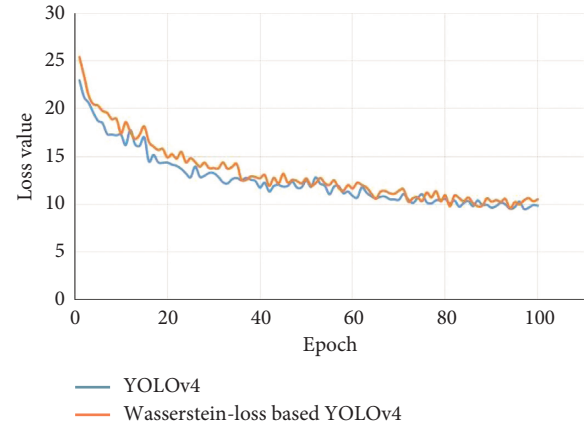


Figure 7: The loss curve of YOLOv4 model.

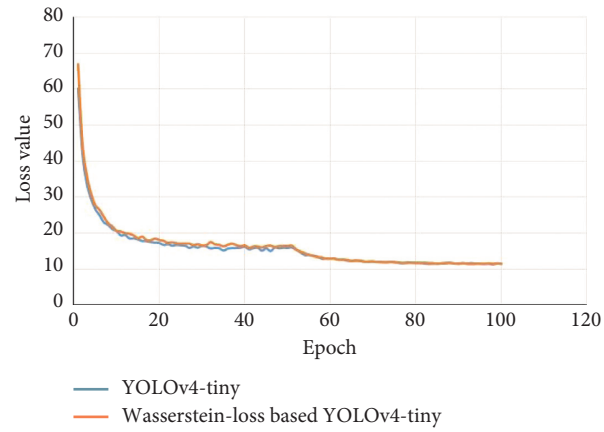


Figure 8: The loss curve of YOLOv4-tiny model.

Table 3: The experimental results.

	YOLOv3		Improved YOLOv3	
	AP	F1	AP	F1
Vehicle	99.30%	0.97	99.43%	0.97
Bike	99.36%	0.99	99.35%	0.99
Rider	98.16%	0.95	98.39%	0.95
Pedestrian	96.66%	0.93	97.09%	0.94
mAP	98.37%		98.57%	
	YOLOv4		Improved YOLOv4	
	AP	F1	AP	F1
Vehicle	98.66%	0.96	99.00%	0.96
Bike	99.49%	0.97	99.33%	0.96
Rider	97.44%	0.95	97.52%	0.95
Pedestrian	97.09%	0.93	96.91%	0.93
mAP	98.17%		98.19%	
	YOLOv4-tiny		Improved YOLOv4-tiny	
	AP	F1	AP	F1
Vehicle	86.98%	0.83	86.77%	0.83
Bike	83.05%	0.85	83.47%	0.85
Rider	80.19%	0.79	80.44%	0.78
Pedestrian	70.55%	0.72	70.89%	0.72
mAP	80.19%		80.39%	

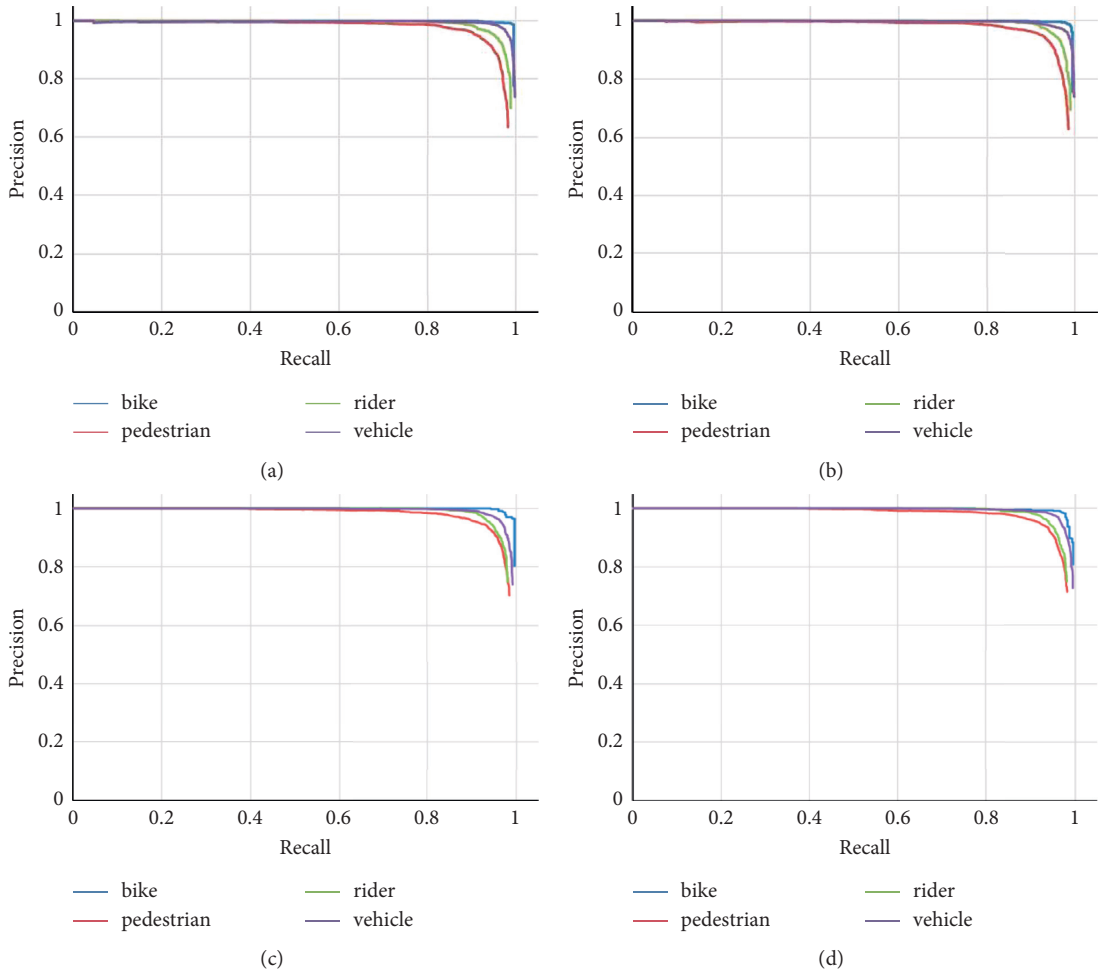


Figure 9: Continued.

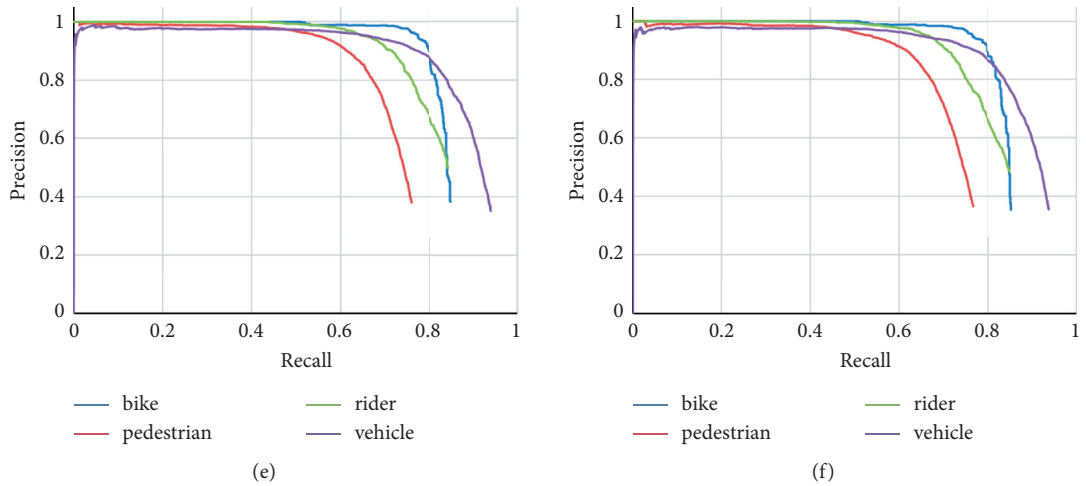


Figure 9: P-R curves of different models. (a) Original YOLOv3. (b) Improved YOLOv3. (c) Original YOLOv4. (d) Improved YOLOv4. (e) Original YOLOv4-tiny. (f) Improved YOLOv4-tiny.

Table 4: The experimental results.

	YOLOv3		Improved YOLOv3	
	AP	F1	AP	F1
Vehicle	97.70%	0.95	97.80%	0.95
Bike	81.27%	0.86	81.55%	0.87
Rider	97.10%	0.92	96.97%	0.92
Pedestrian	95.37%	0.92	95.54%	0.92
mAP	92.86%		92.97%	
	YOLOv4		Improved YOLOv4	
	AP	F1	AP	F1
Vehicle	97.13%	0.94	97.07%	0.94
Bike	77.37%	0.84	77.82%	0.84
Rider	95.60%	0.91	95.46%	0.92
Pedestrian	94.42%	0.91	94.57%	0.91
mAP	91.13%		91.23%	
	YOLOv4-tiny		Improved YOLOv4-tiny	
	AP	F1	AP	F1
Vehicle	86.56%	0.82	86.57%	0.83
Bike	74.60%	0.78	75.44%	0.78
Rider	79.52%	0.78	79.51%	0.78
Pedestrian	70.37%	0.72	70.34%	0.72
mAP	77.76%		77.97%	

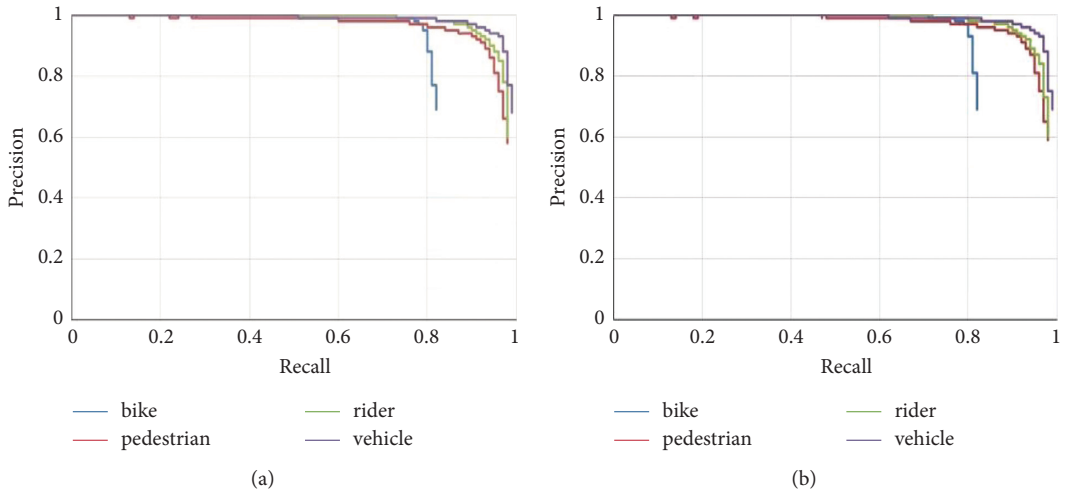


Figure 10: Continued.

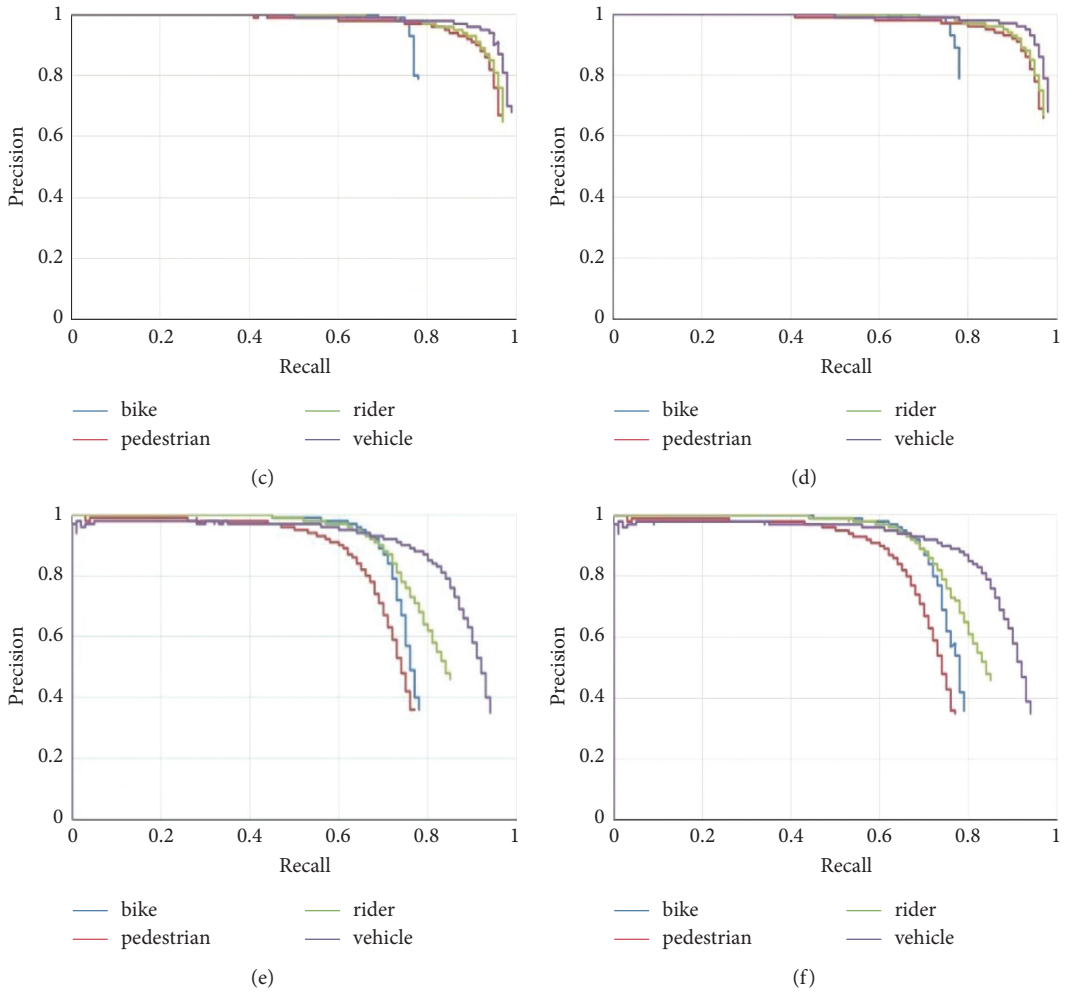


Figure 10: P-R curves of different models. (a) Original YOLOv3. (b) Improved YOLOv3. (c) Original YOLOv4. (d) Improved YOLOv4. (e) Original YOLOv4-tiny. (f) Improved YOLOv4-tiny.

Figure 11: Vehicle application scenarios on NVIDIA TX2.

of the improved YOLOv3 and YOLOv4 models on NVIDIA TX2 is between 3 fps and 4 fps, while on Windows PC, it is between 8 fps and 9 fps, which is a little poor in real-time performance. The detection speed of the improved YOLOv4-tiny model on NVIDIA TX2 is above 22 fps, while on Windows PC, it is above 27 fps, which can basically realize the real-time detection of traffic obstacles.

Table 5: Detection speed of different models.

Detection model		Windows PC	
Speed (fps)	YOLOv3	YOLOv4	YOLOv4-tiny
		Model	
Original model	8.6786	8.0926	27.7137
Improved model	8.8412	8.2137	27.7925
Detection model		NVIDIA TX2	
Speed (fps)	YOLOv3	YOLOv4	YOLOv4-tiny
		Model	
Original model	3.8385	3.3939	22.4285
Improved model	3.8586	3.4049	22.5928

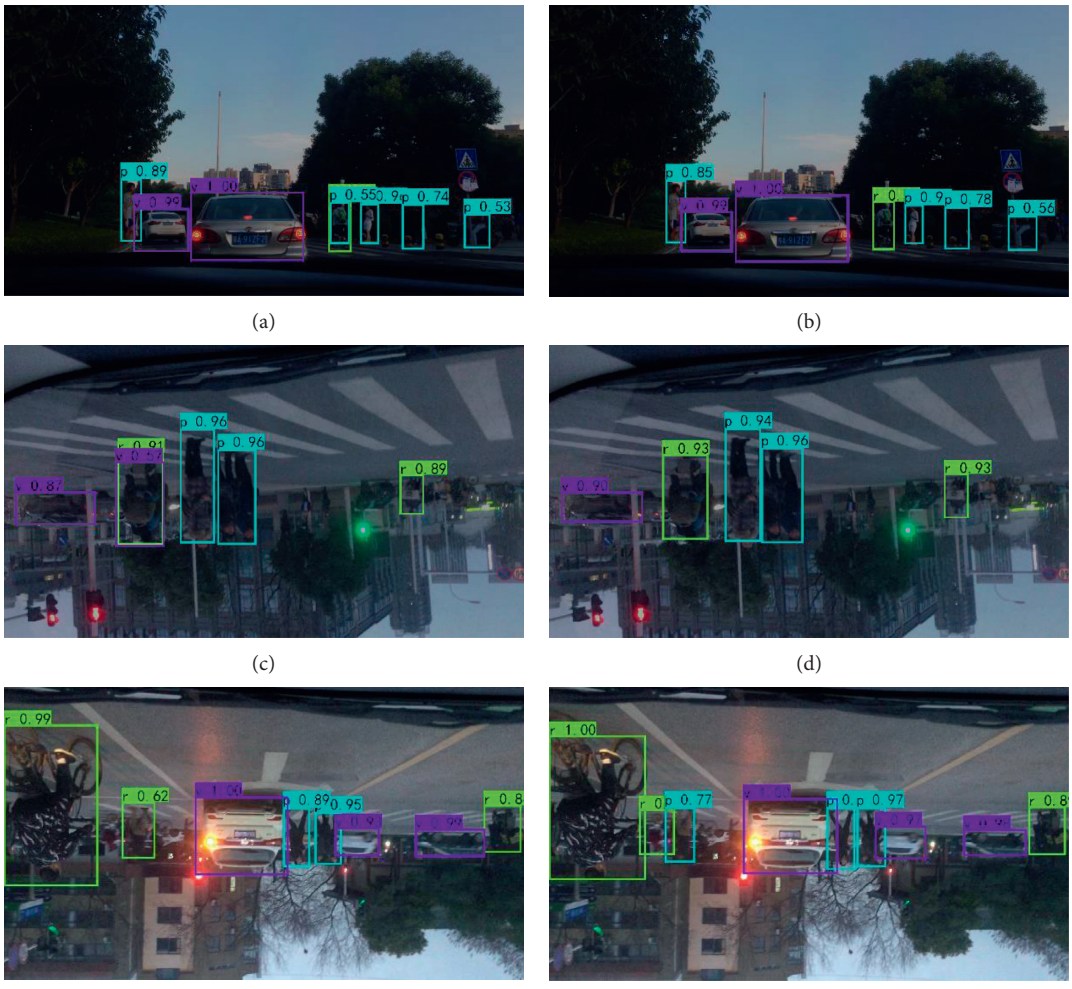


Figure 12: Examples of detection results. (a) Original YOLOv4-tiny. (b) Improved YOLOv4-tiny. (c) Original YOLOv4-tiny. (d) Improved YOLOv4-tiny. (e) Original YOLOv4-tiny. (f) Improved YOLOv4-tiny.

The real-time detection effect of the improved YOLOv4-tiny model was verified on the NVIDIA TX2 and compared with the original YOLOv4-tiny model. As shown in Figure 12, some misclassifications detected by the original model can be effectively and correctly classified by the improved model, proving that the improved model can effectively reduce intolerable misclassifications between different categories.

5. Conclusions

In this paper, an improved YOLO model for traffic obstacle detection and classification applied in ICVs is presented. A new dataset containing traffic obstacles collected under different time periods and different weather conditions in urban environment was established. The improved models, which reduce the intolerable misclassification and enhance

the performance of traffic obstacle detection by combining the Wasserstein distance-based loss with the YOLO models, were designed and implemented. The improved model was trained and then tested on established dataset and selected BDD dataset and deployed on NVIDIA TX2 for real-time detection.

Experimental results showed that the mAP values of the improved YOLOv3, YOLOv4, and YOLOv4-tiny models are 98.57%, 98.19%, and 80.39%, respectively, higher than those of each original model. From the application-oriented performance on NVIDIA TX2, the detection speed of the improved YOLOv4-tiny model is 22.5928 fps, which is much better than that of the YOLOv3 and YOLOv4 models and basically meets the real-time detection requirements of traffic obstacles. In addition, in the real-time vehicle verification, the improved YOLOv4-tiny model can reduce the intolerable misclassifications between different categories more effectively than the original model. In practical applications, the improved model could effectively improve the accuracy of decision making for ICVs, thereby improving the driving safety. In the future study, the dataset could be enriched and the detection model could be further optimised.

Data Availability

The established dataset used to support the findings of this study is 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 supported in part by the National Key R&D Program of China under grant no. 2018YFB0105205 and in part by the Hubei Province Technological Innovation Major Project under grant no. 2019AAA025.

References

[1] K. Q. Li, Y. F. Dai, S. B. Li, and M. Y. Bian, "State-of-the-art and technical trends of intelligent and connected vehicles," *Journal of Automotive Safety and Energy*, vol. 8, no. 1, pp. 1–14, 2017.

[2] X. Y. Ma and W. E. L. Grimson, "Edge-based rich representation for vehicle classification," in *Proceedings of the 10th IEEE International Conference on Computer Vision*, pp. 1185–1192, Beijing, China, October 2005.

[3] Y. Taigman, M. Yang, and M. Ranzato, "Deepface: closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708, Columbus, OH, USA, 2014.

[4] F. M. Kazemi, S. Samadi, and H. R. Poorreza, "Vehicle recognition using curvelet transform and svm," in *Proceedings of the 4th IEEE International Conference on Information Technology*, pp. 516–521, Las Vegas, NV, USA, April 2007.

[5] Y. K. L. Lai, Y. H. C. Chou, and T. Schumann, "Vehicle detection for forward collision warning system based on a cascade classifier using adaboost algorithm," in *Proceedings of the 7th IEEE International Conference on Consumer Electronics*, pp. 47-48, Berlin, Germany, 3–6 Sept. 2017.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, p. 640, 2017.

[7] Z. X. Zou, Z. W. Shi, Y. H. Guo, and J. P. Ye, "Object detection in 20 years: a survey," pp. 1–40, 2019, https://arxiv.org/abs/1905.05055.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, Las Vegas, NV, USA, 27–30 June 2016.

[9] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: single shot multibox detector," in *Proceedings of the European conference on computer vision*, pp. 21–37, Springer, Amsterdam, The Netherlands, December 2016.

[10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020.

[11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, Columbus, OH, USA, 23–28 June 2014.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Proceedings of the European conference on computer vision*, pp. 346–361, Springer, Cham, April 2014.

[13] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, Santiago, Chile, 7–13 Dec. 2015.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, pp. 91–99, 2015.

[15] T. Y. Lin, P. Doll´ar, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. 1, no. 2, p. 4, Honolulu, HI, USA, 21–26 July 2017.

[16] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2016, https://arxiv.org/abs/1804.02767.

[17] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: optimal speed and accuracy of object detection," 2020, https://arxiv.org/abs/2004.10934.

[18] S. Y. Wang and T. Ahmad, "A real-time detection method of traffic targets based on YOLO," *Computer & Digital Engineering*, vol. 48, no. 9, pp. 2162–2167, 2020.

[19] A. Narayanan, R. Darshan Kumar, R. RoselinKiruba, and T. Sree Sharmila, "Study and analysis of pedestrian detection in thermal images using YOLO and SVM," in *Proceedings of the 2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 431–434, Chennai, India, March 2021.

[20] S. H. Hung, K. W. Chen, C. H. Chen, H. T. Chou, and C. Y. Yao, "Real-time obstacle detection on embedded system," in *Proceedings of the 2018 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1-2, Taichung, Taiwan, China, December 2018.

[21] H. Wang, X. Lou, Y. Cai, Y. Li, and L. Chen, "Real-time vehicle detection algorithm based on vision and lidar point cloud fusion," *Journal of Sensors*, vol. 2019, Article ID 8473980, 9 pages, 2019.

[22] C. S. Arvind, R. Jyothi, K. Mahalakshmi, C. K. Vaishnavi, and U. Apoorva, "Vision based driver assistance for near range obstacle sensing under unstructured traffic environment," in *Proceedings of the Proceedings of 2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1163–1170, Xiamen, China, December 2019.

[23] M. Zhang, R. Fu, Y. Guo, and L. Wang, "Moving object classification using 3D point cloud in urban traffic environment," *Journal of Advanced Transportation*, vol. 2020, Article ID 1583129, 12 pages, 2020.

[24] J. Feng, F. Wang, S. Feng, and Y. Peng, "A multibranch object detection method for traffic scenes," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 3679203, 16 pages, 2019.

[25] X. Li, Y. Liu, Z. Zhao, Y. Zhang, and L. He, "A deep learning approach of vehicle multitarget detection from traffic video," *Journal of Advanced Transportation*, vol. 2018, Article ID 7075814, 11 pages, 2018.

[26] C. Wang, Y. Dai, W. Zhou, and Y. Geng, "A vision-based video crash detection framework for mixed traffic flow environment considering low-visibility condition," *Journal of Advanced Transportation*, vol. 2020, Article ID 9194028, 11 pages, 2020.

[27] Y. Cai, T. Luan, H. Gao et al., "YOLOv4-5D: an effective and efficient object detector for autonomous driving," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2021.

[28] M. Hnewa and H. Radha, "Object detection under rainy conditions for autonomous vehicles: a review of state-of-the-art and emerging techniques," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 53–67, 2021.

[29] Z. Liu, Y. Cai, H. Wang et al., "Robust target recognition and tracking of self-driving cars with radar and camera information fusion under severe weather conditions," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–14, 2021.

[30] A. Bell, T. Mantecon, C. Diaz, C. R. del-Blanco, F. Jaureguizar, and N. Garcia, "A novel system for nighttime vehicle detection based on foveal classifiers with real-time performance," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2021.

[31] G. S. R. Satyanarayana, S. Majhi, and S. K. Das, "A vehicle detection technique using binary images for heterogeneous and lane-less traffic," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.

[32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 214–223, Xiamen, China, October 2019.

[33] C. Villani, *Optimal Transport: Old and New*, Vol. 338, Springer Science & Business Media, , Berlin, Germany, 2008.

[34] C. Y. Wang, H. Y. M. Liao, Y. H. Wu, P. Y. Chen, J. W. Hsieh, and I. H. Yeh, "CSPNet: a new backbone that can enhance learning capability of cnn," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPR Workshop)*, Seattle, WA, USA, June 2020.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.

[36] S. Liu, L. Qi, H. F. Qin, J. P. Shi, and J. Y. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8759–8768, Salt Lake City, UT, USA, June 2018.

[37] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy, "Optimal transport for domain adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 9, pp. 1853–1865, 2016.

[38] G. Peyré and M. Cuturi, "Computational optimal transport," *Foundations and Trends in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.

[39] N. Bonnotte, *Unidimensional and Evolution Methods for Optimal Transportation*, vol. 11, Ph. D. thesis , Paris, France, 2013.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 8389229, 10 pages
https://doi.org/10.1155/2022/8389229

Research Article

A Multiscale Symbolic Dynamic Entropy Analysis of Traffic Flow

Zhanyou Cui,[1] **Gaoli Chen**,[1] **Bing Liu**,[2] **and Deguang Li**[3]

[1]*College of Mechanical and Electrical Engineering, Zhengzhou University of Industrial Technology, Zhengzhou 451150, China*
[2]*Henan Forestry Vocational College, Luoyang 471002, China*
[3]*School of Information Technology, Luoyang Normal University, Luoyang 471934, China*

Correspondence should be addressed to Zhanyou Cui; zhanyou_cui@126.com

Received 24 September 2021; Revised 15 November 2021; Accepted 14 March 2022; Published 30 March 2022

Academic Editor: Gen Li

The complexity analysis of traffic flow is important for understanding the property of traffic system. Being good at analyzing the regularity and complexity, multiscale SamEn has attracted much attention and many methods have been proposed for complexity analysis of traffic flow. However, there may exist discontinuity of the calculated entropy value which makes the regularity of the traffic system difficult to understand. The phenomenon occurs due to an inappropriate selection of the parameter r in the multiscale SamEn. Moreover, it is difficult to select an appropriate r for the accurate evaluation of the complexity, which limits the application of multiscale entropy for traffic flow analysis. To solve this problem, a new entropy-based method, multiscale symbolic dynamic entropy, for evaluating the traffic system is proposed here. To verify the effectiveness of the proposed method, traffic data collected from stations in different cities are preprocessed by the proposed method. Both results of two cases show that the weekend patterns and weekday patterns are effectively distinguished using the proposed method, respectively. Specifically, compared with the traditional methods including multiscale SamEn and the multiscale modified SamEn, the complexity of the corresponding traffic system can be better evaluated without considering the selection of r, which demonstrates the effectiveness of the proposed method.

1. Introduction

The research of the traffic system plays great role in reducing the traffic congestion, which increases traveling delays and causes huge economic losses in cities [1]. In recent years, the number of vehicles has increased rapidly and the traffic congestion has become more serious, so the research of the traffic system has become more important and has attracted much attention. Specifically, there are a lot of studies about the research of traffic including the merging behavior [2–4], traffic flow forecasting [5, 6], model of traffic flow [7, 8], traffic theory [9], and traffic resource allocation [10]. For example, Wang et al. [11] proposed a short-term traffic flow prediction approach based on multiple traffic flow basic parameters, in which the chaos theory and support vector regression are utilized. Li et al. [12] proposed a novel hybrid forecasting model by combining three predictors, namely, the autoregressive integrated moving average (ARIMA), backpropagation neural network (BPNN), and support vector regression (SVR). Yao et al. [13] presented a two-level optimization method of scheduling and trajectory planning for connected automated vehicles.

Furthermore, the traffic dynamics and regularity of traffic flow is still a hot topic and a big challenge to transportation engineering [14]. Choi and Lee [15] found that the traffic equation exhibits a variety of behaviors including homogeneous flow, turbulent behavior, and density waves with fluctuations in the appropriate regime, and they investigated the possibility of $1/f$ fluctuations. Li et al. [16] conducted a comprehensive analysis of the motivations, gap acceptance, duration, and speed adjustment of heavy vehicle lane changes, which gives a better understanding of the lane-changing behaviors of heavy vehicles. Castillo et al. [17] found that the emergent state shows a stochastic resonance-like behavior that the average traffic velocity increases with respect to the system without noise for different initial jammed densities.

With help of these above findings, the dynamics of traffic jams in cities can be understood better. Then, an approach based on algebraic topological methods [18] was used to accurately characterize jamming in dynamical systems with queues. Yin and Shang [19] proposed a multiscale multifractal distended cross-correlation analysis method to describe the cross-correlation properties depending on the timescale in which the multifractality is computed. They found that the main distinction between weekday and weekend patterns is the different periodic patterns hidden in them. In addition, these different periodic patterns play an important role in the Hurst surface of cross-correlation investigation. The chaos characteristics of the traffic system were also studied by a nonlinear time series modeling technique [20]. To better evaluate the complexity of the traffic system, the multiscale SamEn (MSE) method is the most used method and has obtained many valuable results. The MSE method was first proposed in 2002 for evaluating the complexity of physiologic time series [21]. Wang et al. [22] introduced the multiscale permutation entropy analysis to investigate the complexities of different traffic time series and found that the complexity of weekend traffic time series is different from that of weekday time series. This finding is helpful for classifying the series when making prediction. Afterwards, to study the correlation degree and complexity between multiple variables, the multivariate multiscale sample entropy [23, 24] was also proposed and the more accurate and helpful knowledge about the complexity of traffic time series was obtained.

From these studies, we can find that the MSE and its variants are very effective for complexity evaluation of the traffic system. However, there are still shortcomings of multiscale entropy which limit its application for traffic flow. As pointed out in [25], the similarity definition of vectors based on Heaviside function may lead to discontinuous and hard boundary and cause some problems in the validity and accuracy of SampEn. Specifically, there are no vectors satisfying the evaluating condition, and the value of Heaviside function is zero. As a result, the denominator of SamEn is zero, which makes the value of SamEn be NaN [26]. This meaningless result makes the complexity of traffic flow difficult to understand. Furthermore, it is difficult to select an optimal parameter r for the calculation of MSE or multiscale modified SamEn (MMSE) reported in [25]. Therefore, it is difficult to evaluate the complexity of some traffic systems. To solve this problem, inspired the symbolic sequences which are able to reflect the change of dynamic characteristics of the system state [27], a multiscale symbolic dynamic entropy (MSSDE) method is proposed for evaluating the complexity of the traffic system in this paper. In the proposed method, first, the original time series defined is transformed into the coarse-grained time series for representing the system dynamics on different time scales. Second, the coarse-grained time series is transformed into the symbolic time series. Third, we construct the template vector with the length of m and calculate the frequency of occurrence. Naturally, based on the frequency of occurrence, the average of the frequency can be obtained. Similarly, the average of the frequency with the dimensionality of m+1 can be obtained. Finally, the MSSDE can be constructed based on the obtained average of the frequency with m+1 and m. The MSSDE holds both merits of multiscale entropy and symbolic dynamic entropy. The complexity and similarity can be well analyzed with help of the proposed MSSDE. Two cases of traffic data collected from different stations are used to verify the effectiveness of the proposed method. The results show that the proposed method is useful for the complexity analysis of traffic flow.

1.1. Multiscale Symbolic Dynamic Entropy. The MSSDE is proposed based on SamEn for better analysis of the complexity of traffic flow. The MSSDE is described in detail as follows.

(1) First, original time series defined is transformed into the coarse-grained time series for representing the system dynamics on different time scales. The process is achieved by averaging original time series inside consecutive but nonoverlapping windows of length τ. τ is selected from 1 to 6 in the proposed method. The corresponding equation is as follows:

$$u_j^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j_\tau} x_i, \frac{1 \le j \le N}{\tau}. \tag{1}$$

(2) Then, $\left\{u_j^\tau, j = 1, 2, \ldots, N/\tau\right\}$ of the scale τ is transformed into the symbolic time series $s^\tau\{s(k), k = 1, 2, \cdots, N\}$ by

$$s^\tau(i) = \begin{cases} \sigma_1, u_1^\tau \in C_1 \\ \sigma_2, u_2^\tau \in C_2 \\ . \\ . \\ . \\ \sigma_M, u_{N/\tau}^\tau \in C_M \end{cases}, \tag{2}$$

where N is the length of the time series and M $(M < N)$ is the number of symbolic. $\sigma_i (i = 1, 2, \cdots, N)$ represents different symbolic which is further replaced by Arabic numerals in this paper; specifically, $\sigma_i = i$. $C_i (i = 1, 2, \cdots, N)$ are N disjoint sets obtained by dividing the value space of u. Specifically, C_i satisfy both equations (3) and (4).

$$\bigcup_{i=1}^{N} C_i = \bigcup_{i=1}^{N} u_i^\tau, \tag{3}$$

$$C_i \cap C_j = \varnothing, \forall i \ne j; i, j \in [1, \cdots, N]. \tag{4}$$

The space of C is divided into M partitions based on the normal distribution [28]. Through this step, the time series u^τ with length N/τ at various τ are synchronized into the symbolic time series with length M.

(3) Next, the template vector is constructed by

$$S_i^m = [s(i), s(i+1), \cdots, s(i+m-1)], i = 1, 2, \cdots, \frac{N}{\tau - m + 1}, \quad (5)$$

where m is the length of sequences to be compared. We can find that the number of possible values of S_i^m is M_m. In addition, $N/\tau - m + 1$ template vectors can be obtained.

(4) Afterwards, for a given s_i^m, calculate the number of vectors Q_i^m satisfying $S_i^m = S_j^m$, where $1 \le j \le N/\tau - m + 1$.

(5) Calculate the frequency of occurrence $C_i^m = 1/N - m + 1Q_i^m, j = 1, 2, \cdots, N/\tau - m + 1$.

(6) The average of the frequency is defined by $C^m = 1/N - m + 1 \sum_{i=1}^{N/\tau - m+1} C_i^m$.

(7) Extend the dimensionality from m to m+1, and then S_i^{m+1} are calculated as follows:

$$S_i^{m+1} = [s(i), s(i+1), \cdots, s(i+m)], \left(i = 1, 2, \cdots, \frac{N}{\tau - m}\right). \quad (6)$$

(8) Similarly, the frequency of occurrence is calculated according to

$$C_i^{m+1} = \frac{1}{N-m} Q_i^{m+1}, \left(j = 1, 2, \cdots, \frac{N}{\tau - m}\right), \quad (7)$$

where Q_i^{m+1} represent the number of vectors satisfying $S_i^{m+1} = S_j^{m+1}$. The corresponding average of C_i^{m+1} is

$$C^{m+1} = \frac{1}{N-m} \sum_{i=1}^{N/\tau - m+1} C_i^{m+1}. \quad (8)$$

(9) Finally, the MSSDE can be obtained according to

$$\mathrm{MSSDE}(m, \tau, M) = -ln\left(\frac{C^{m+1}}{C^m}\right). \quad (9)$$

The computational complexity of the proposed method is more expensive than the traditional methods including symbolic dynamic entropy and sample entropy but is almost the same as the other traditional multiscale entropy-based methods. In other words, the computational complexity depends greatly on the scales of entropy. However, the computation complexity is still small and the proposed method can be quickly computed with help of a personal computer.

2. Results and Analysis

2.1. Case 1: Complexity Analysis of Traffic Flow from PeMS. To verify the effectiveness of the proposed method, the test dataset from PeMS open-access traffic flow database is used. The database was recorded every 5 min in Sacramento County. Therefore, 288 data elements were collected for a daily flow series of one loop detector, and the data of 7 days from March 31 to April 6, 2014, as shown in Figure 1, were analyzed by the proposed MSSDE to explore the complexities of the traffic time series [29, 30].

We first apply the traditional MSE to process the data, and the results are shown below. The parameter r is set to be 0.2 $*$ std (std stands for the standard deviation of the data to be analyzed), empirically. The corresponding MSE of various scales from Monday to Sunday is shown in Figure 2(a). MSE of many days at some scales has no value, and the existing MSE values on the traffic data of Saturday and Sunday show that the traffic data of the two days belong to the same pattern with the weekday pattern. The MSE with no value refers to the meaningless value, which is obtained because some values are divided by zero. This condition is commonly seen because the template vectors are coarsely segmented just based on the original time series without preprocessing. As a result, the different patterns of traffic data cannot be well extracted and the difference between different patterns is so small that the frequency of occurrence is zero. Therefore, the complexity of the traffic data cannot be evaluated by the MSE when r = 0.2 $*$ std. To investigate the influence of r on MSE, r varying from 0.01 to 0.6 at a step of 0.05 is used to calculate the MSE at different scales from 1 to 6 for the traffic data of March 31 as shown in Figure 2(b). It can be seen that r has a great influence on the MSE and the MSE gave no entropy values when r is very small at some scales. Moreover, the MSE at the same r of various scale changes randomly, which is inconsistent with the fact of the traffic flow. Because these analysis results are conducted under one pattern of traffic flow, they should have the same pattern even under different r. Undoubtedly, it is a hard work to select a suitable r for the traditional MSE.

As shown in Figure 2(c), the MSE of scale 1 versus different r is plotted. It can be seen that the MSE gave no values when r is smaller than 0.1 and the MSE of various r is greatly different. Therefore, it is difficult to select an appropriate r for the effective evaluation of traffic data using MSE. A MMSE method is proposed in [25] to solve the problem of discontinuousness boundary caused by the Heaviside function, and the MMSE is constructed by replacing the nonlinear function, sigmoid function, in equation (10) with the Heaviside function.

$$f(d, c) = \frac{1}{1 + \exp[(d - 0.5)]/c}, \quad (10)$$

where c and d are parameters for the slope or the steepness of the function. Although the discontinuity problem is well solved in MMSE, it is doubtful that the MMSE is effective to the evaluate the complexity of traffic data. In addition, it is also difficult to select an optimal r in equation (11) of Castillo et al. [17] for MMSE to accurately evaluate the nonlinearity of the traffic system.

$$D_{ij}^m = f\left(d_{ij}^m, r\right) = \frac{1}{1 + \exp\left[\left(d_{ij}^m - 0.5\right)/r\right]}, \quad (11)$$

where d_{ij} denotes the similarity degree between two different vectors in MMSE. To illustrate this point, the MMSE is also used to analyze the traffic flow for PeMS. The corresponding result when r is 0.1 in MMSE is shown in Figure 3(a). As

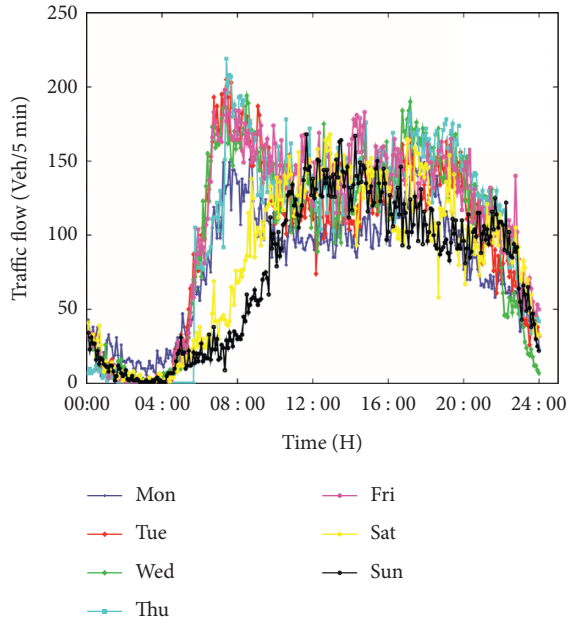


Figure 1: The traffic data of week model from Monday to Sunday.

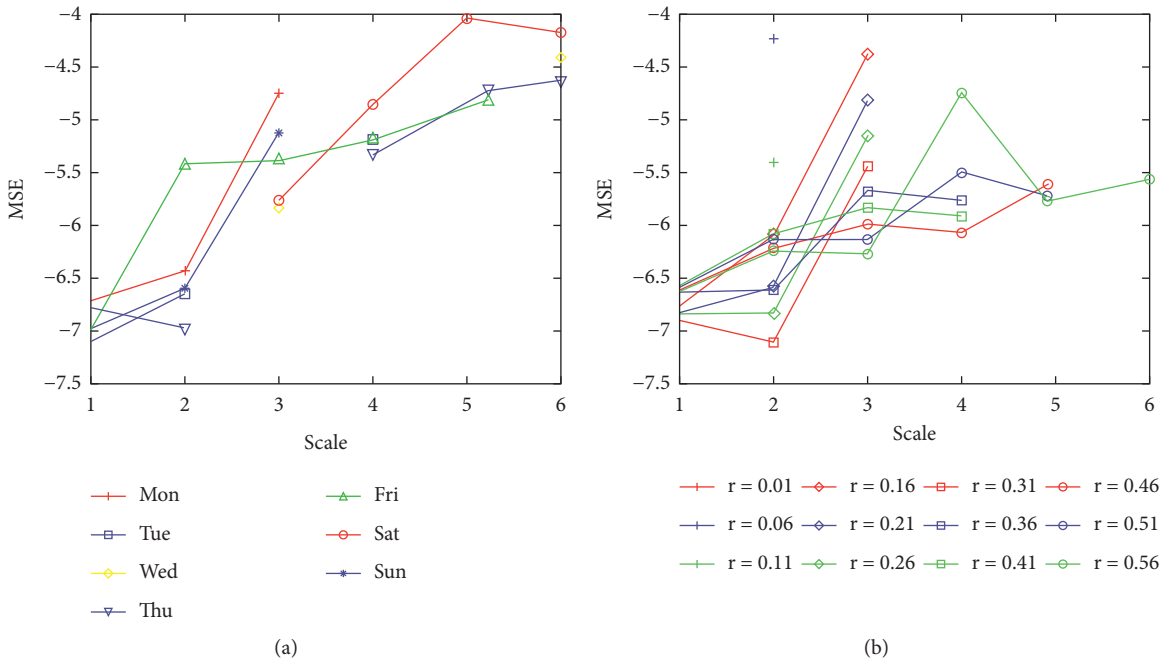
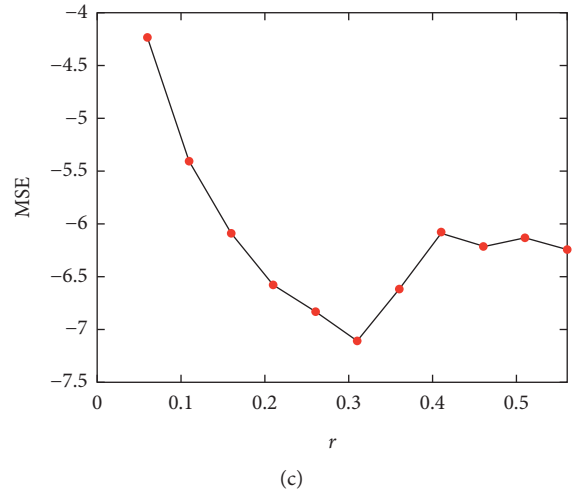
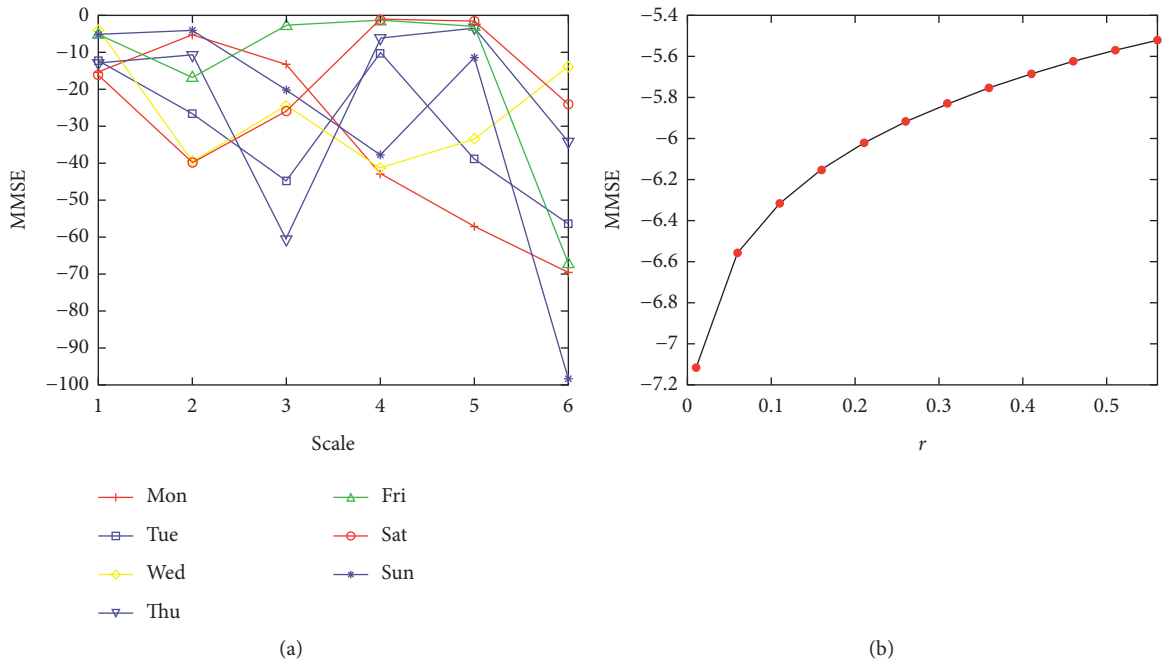


Figure 2: Continued.



(c)

Figure 2: The analysis result of the traffic data using traditional MSE. (a) The analysis of the traffic data using MSE. (b) The MSE of Monday versus various scales under different r. (c) The MSE of Monday of scale two under different r.



(a) (b)

Figure 3: The analysis result of the traffic data using MMSE. (a) The analysis of the traffic data using MMSE. (b) The MMSE of Monday versus various scales under different r.

shown in Figure 3(a), different day patterns including weekday pattern and weekend data pattern cannot be found, which indicates that the MMSE method fails to analyze the complexity of these traffic data. To further explain this, the MMSE values with various r from 0.01 to 0.6 with a step of o.05 are also calculated for the traffic data on Monday. The result is shown in Figure 3(b), and we can find that the MMSE values are different for various r, which shows that r influences MMSE greatly. Therefore, the MMSE may not be a proper method for evaluating the complexity of the traffic system. The problem of meaningless values is well solved by the MMSE method because the method replaces the division with a similarity distance. Consequently, even when two patterns are very similar, the MMSE can also be successfully computed with help of the similarity degree. But we can find that this method also suffered from the problem of coarse evaluation, leading to the weakness of the proposed method.

To evaluate the complexity of these traffic data, the proposed method MSSDE is used to process these traffic data. Because the MSSDE is calculated by counting the number of the same patterns of symbolic vectors rather than counting the number of two vectors smaller than r, we can complete the calculation of MSSDE without considering the selection of r. Therefore, the problem of traditional MSE about how to select an optimal r is well solved. The traffic data of 7 days are analyzed by the MSSDE, and the result is shown in Figure 4. A

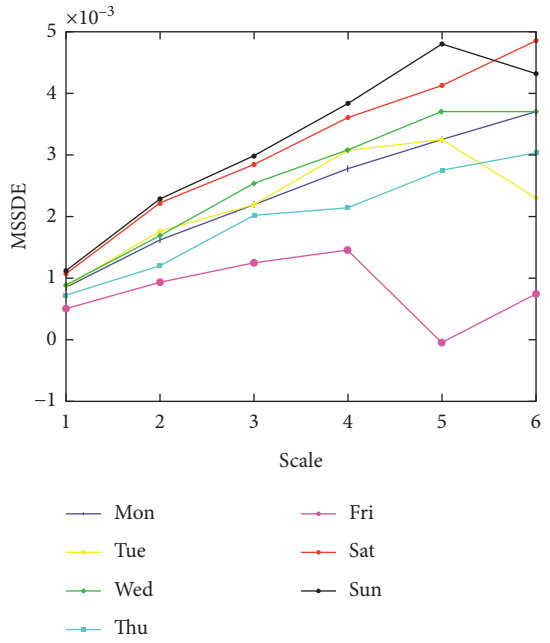


Figure 4: The result of analysis on the traffic day using the proposed MSSDE.

monotonic increase in the MSSDE values versus the scale is found, and the gap between the MSSDE value of Saturday and Sunday is small, while the gap between weekdays and the weekend days is large. This result indicates that Saturday and Sunday fall into the same pattern named weekend pattern and have different patterns with the weekdays, which is consistent with the fact. In particular, the difference between the weekday pattern and the weekend patter becomes larger with the increase of the scale. As we all know, at workday, people have to go to work or school, while a lot of people prefer to stay at home or nearby areas regularly at weekends. As a result, the traffic data of weekdays are less complex than the traffic data of weekends. Obviously, the traffic data of Saturday and Sunday belong to the different pattern from the other days. It can be inferred that the weekend pattern can be effectively distinguished with the help of MSSDE.

2.2. Case 2: Complexity Analysis of Traffic Flow from the WisTransPortal System. To further demonstrate the effectiveness of the proposed method, traffic flow data obtained from the WisTransPortal system [31] are used. The traffic data collection station is shown in Figure 5, which is marked by a blue triangle. Moreover, it can be found that there exist many companies and hospitals around the station, which indicates that the complexity of these traffic data is similar to the traffic data of Case 1. Specifically, these traffic data probably contain two patterns including the weekday pattern and the weekend pattern. Detector 5561 is used to collect the traffic flow of entrance ramp, while there also exist detectors 5562, 5566, and 5570 to collect the traffic flow of other lanes.

The traffic flow data collected by detector 5561 versus various times are shown in Figure 6(a). To evaluate and learn the complexity of this traffic day, we first apply the proposed

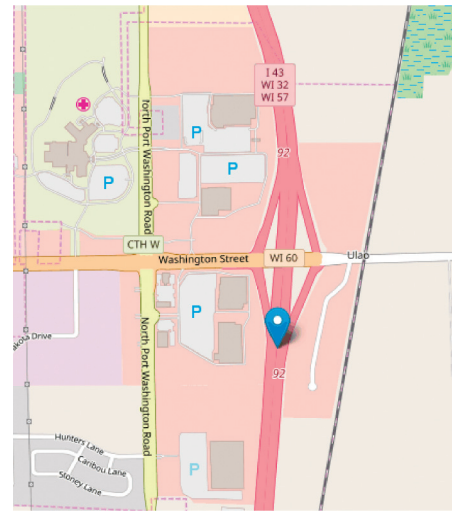


Figure 5: The traffic data collection station in a city of Wisconsin.

MSSDE method to deal with this day, and the corresponding result is shown in Figure 6(b). It can be seen that the difference between the weekday traffic flow and weekend traffic flow is large, which indicates that there also exist a weekday pattern and weekend pattern in the traffic data. This finding is in agreement with the condition of the data collection station, which verifies the performance of our proposed method. The tractional MSE and the MMSE methods are also used, and the results are shown in Figures 6(c) and 6(d), respectively. Figure 6(c) shows that discontinuities occur in the MSE result, and the complexity of the traffic data cannot be illustrated from this result. Moreover, as shown in Figure 6(d), though the discontinuity problem is solved, the weekday and weekend patterns cannot be found from the MMSE result. Compared with our proposed method, both

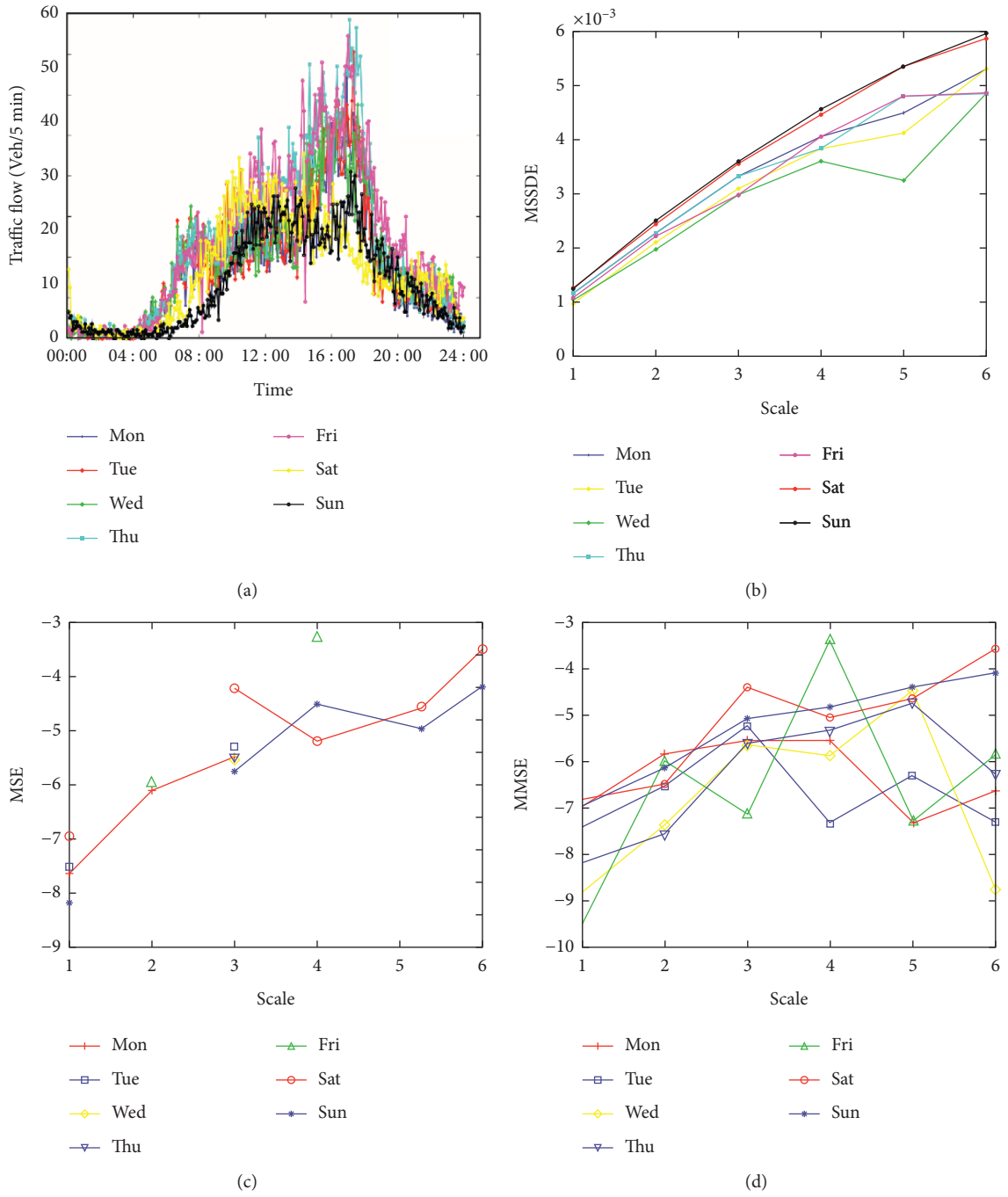


Figure 6: The result of traffic data collected by detector 5561. (a) The traffic data of week model from Monday to Sunday. (b) The analysis result of the traffic data using the proposed method. (c) The analysis result of the traffic data using the traditional MSE. (d) The analysis result of the traffic data using the MMSE.

MSE and MMSE methods fail to analyze the traffic flow. In addition, more traffic data collected by 5562, 5566, and 5570 are also used for verifying the proposed MSSDE. The traffic data versus different times of detectors 5562, 5566, and 5570 are shown in Figures 7(a), 8(a), and 9(a), respectively. The proposed method is applied to these data, and the corresponding results are shown in Figures 7(b), 8(b), and 9(b). It can be found that the MSSDE of Saturday and Sunday is larger than that of other days. The MSSDE of Saturday approximately equals MSSDE of Sunday, which indicates that Saturday and Sunday have the same pattern, weekend pattern. Specifically, two patterns, including weekday pattern and weekend pattern, occur in the traffic system, which is similar to the result of the analysis of traffic data obtained from PeMS. In other words, different patterns of both weekday and weekend can also be clearly distinguished with the help of the proposed method. Therefore, the proposed MSSDE yields excellent performance and can be used for evaluating the complexity of traffic flow.

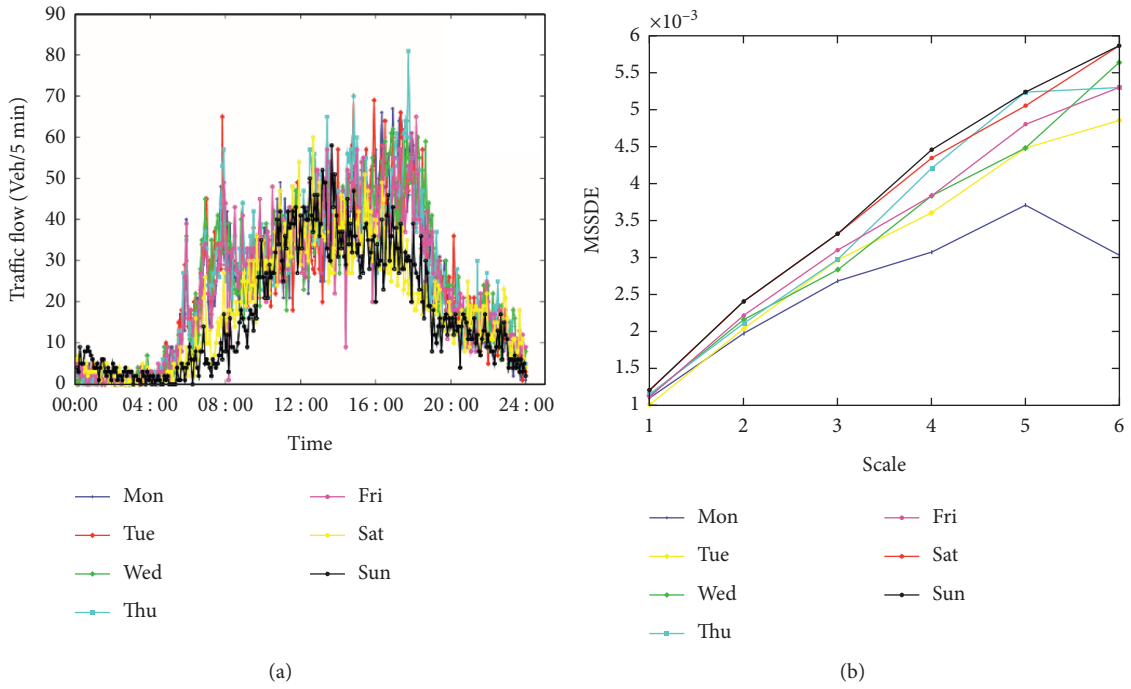


Figure 7: The result of traffic data collected by detector 5562. (a) The traffic data collected of week model from Monday to Sunday. (b) The result of analysis on the traffic day using the proposed MSSDE.

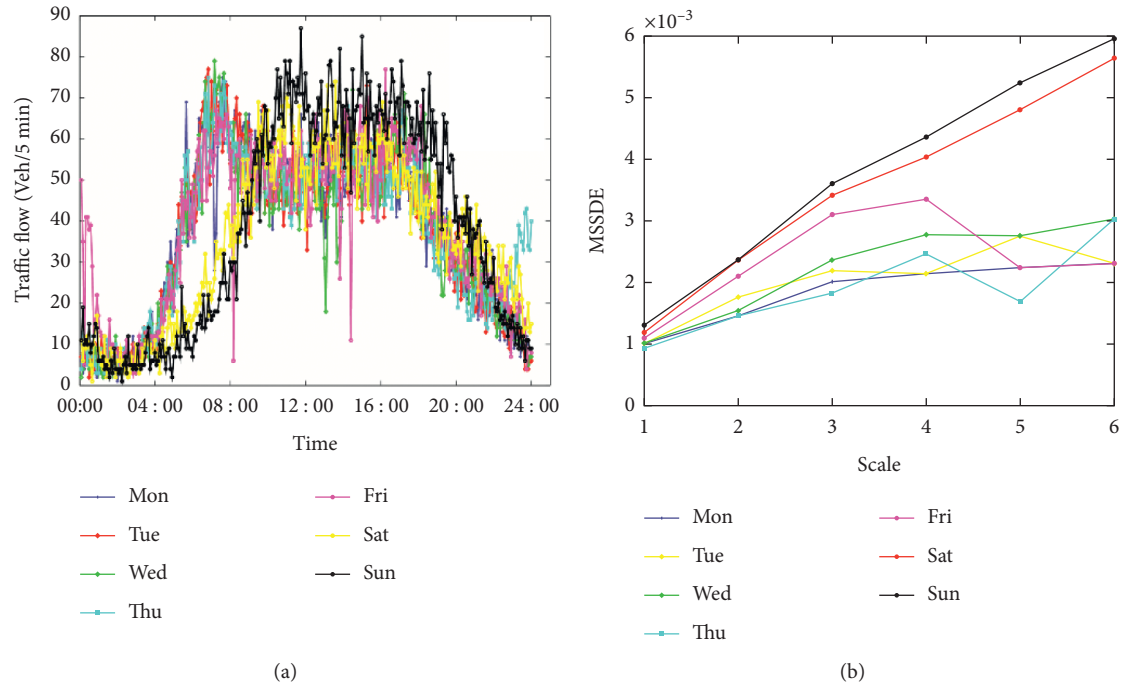


Figure 8: The result of traffic data collected by detector 5566. (a) The traffic data collected of week model from Monday to Sunday. (b) The result of analysis on the traffic day using the proposed MSSDE.

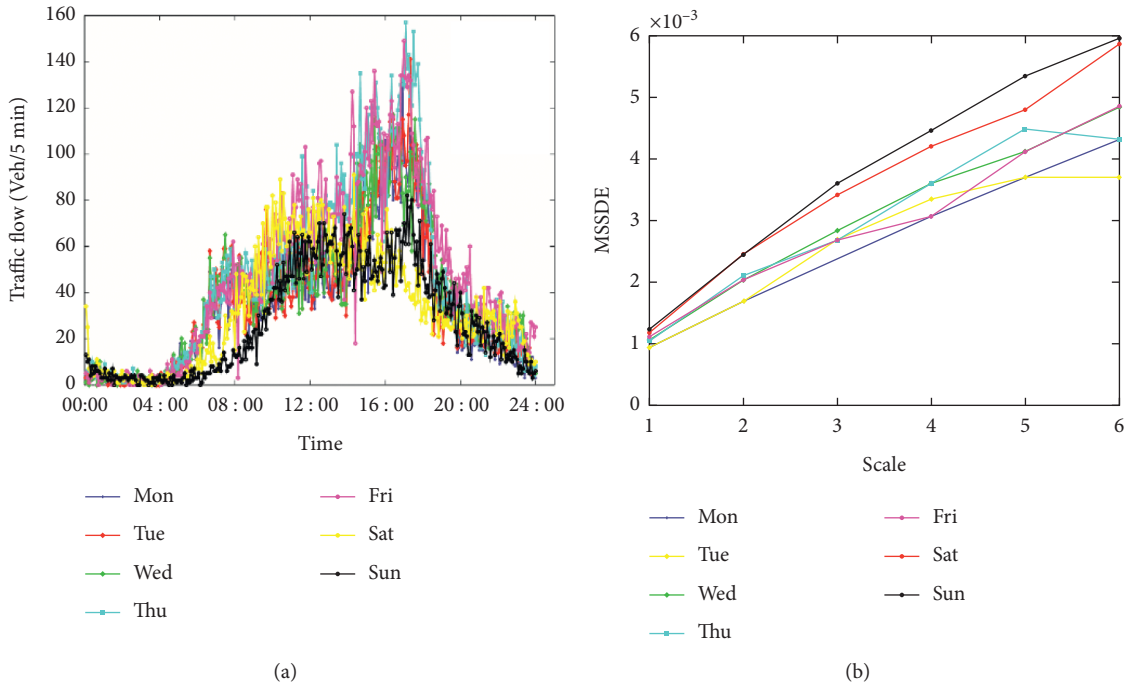


Figure 9: The result of traffic data collected by detector 5570. (a) The traffic data collected of week model from Monday to Sunday. (b) The result of analysis on the traffic day using the proposed MSSDE.

3. Discussion

In this paper, the MSSDE is proposed for accurately analyzing the complexity of traffic data. With help of the proposed method, the problem of discontinuity phenomenon caused by the Heaviside function in traditional multiscale entropy is well solved by the proposed method. Furthermore, the traffic flow can be effectively evaluated without considering the selection of r. In addition, the result of the analysis on the traffic data from PeMS and WisTransPortal systems shows that weekday and weekend patterns of the traffic system can be distinguished using the MSSDE, which demonstrates the effectiveness of the proposed method. Therefore, the proposed MSSDE method is able to evaluate the regularity of time series effectively and thus can be more convenient and powerful for traffic system analysis. In the future, we will consider the application of the MSSDE method to the multivariate traffic data analysis. Furthermore, some inspired methods [32–34] will also be considered and introduced for traffic analysis.

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.

Authors' Contributions

ZC and DL were responsible for investigation and original draft preparation. GC and BL were responsible for methodology and review and editing. ZC was responsible for project administration.

Acknowledgments

This study was supported in part by the Science and Technology Key Project of Henan Province under grant no. 202102210370.

References

[1] D. Helbing, "Traffic and related self-driven many-particle systems," *Reviews of Modern Physics*, vol. 73, 2001.

[2] G. Li, Z. Yang, Q. Yu, J. Ma, and S. Fang, "Characterizing heterogeneity among merging positions: a comparison study between random parameter and latent class Accelerated hazard mode," *Journal of Transportation Engineering, Part A: Systems*, vol. 147, no. 3, Article ID 04021029, 2021.

[3] G. Li, J. Ma, and Q. Shen, "Modeling of merging decision during execution period based on random forest," *Journal of Advanced Transportation*, vol. 2021, Article ID 6654096, 11 pages, 2021.

[4] G. Li, S. Fang, J. Ma, and J. Cheng, "Modeling merging acceleration and deceleration behavior based on gradient-boosting decision tree," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 7, Article ID 05020005, 2020.

[5] J. Wang, W. Deng, and Y. Guo, "New Bayesian combination method for short-term traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 79–94, 2014.

[6] Y. Liu and J. Zhang, "Predicting traffic flow in local area networks by the largest le," *Entropy*, vol. 18, no. 1, p. 32, 2016.

[7] S. Benzoni-Gavage and R. M. Colombo, "An n -populations model for traffic flow," *European Journal of Applied Mathematics*, vol. 14, no. 5, pp. 587–612, 2003.

[8] I. Gasser, G. Sirito, and B. Werner, "Bifurcation analysis of a class of 'car following' traffic models," *Physica D: Nonlinear Phenomena*, vol. 197, no. 3-4, pp. 222–241, 2004.

[9] Y. Yan, S. Zhang, J. Tang, and X. Wang, "Understanding characteristics in multivariate traffic flow time series from complex network structure," *Physica A: Statistical Mechanics and Its Applications*, vol. 477, pp. 149–160, 2017.

[10] W. Deng, J. Xu, and H. Zhao, "A novel gate resource allocation method using improved PSO-based QEA," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, 2020.

[11] X. Wang, W. Li, and C. Yin, "A multiple-parameter approach for short-term traffic flow prediction," *Modern Physics Letters B*, vol. 35, no. 14, Article ID 2150245, 2021.

[12] W. Li, S. Chen, and X. Wang, "A hybrid approach for short-term traffic flow forecasting based on similarity identification," *Modern Physics Letters B*, vol. 35, no. 13, Article ID 2150212, 2021.

[13] Z. Yao, H. Jiang, and Y. Cheng, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, 2020.

[14] Y. Yokoya, "Dynamics of traffic flow with real-time traffic information, Physical review," *E, Statistical, nonlinear, and soft matter physics*, vol. 69, Article ID 016121, 2004.

[15] M. Y. Choi and H. Y. Lee, "Traffic flow and 1/f fluctuations," *Physical Review A*, vol. 52, pp. 5979–5984, 1995.

[16] G. Li, J. Ma, and Z. Yang, "Characteristics of Heavy Vehicle Discretionary Lane Change Based on Trajectory Data," *Transportation Research Record*, 2021.

[17] F. Castillo, B. A. Toledo, V. Muoz et al., "City traffic jam relief by stochastic resonance," *Physica A: Statistical Mechanics and Its AppliCations*, vol. 403, pp. 65–70, 2014.

[18] M. Andjelkovic, N. Gupte, and B. Tadic, "Hidden geometry of traffic jamming, Physical review," *E, Statistical, nonlinear, and soft matter physics*, vol. 91, Article ID 052817, 2015.

[19] Y. Yin and P. Shang, "Multiscale multifractal detrended cross-correlation analysis of traffic flow," *Nonlinear Dynamics*, vol. 81, pp. 1329–1347, 2015.

[20] P. Shang, X. Li, and S. Kamae, "Chaotic analysis of traffic time series, Chaos," *Solitons & Fractals*, vol. 25, pp. 121–128, 2005.

[21] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, Article ID 068102, 2002.

[22] P.S. J. Wang, X. J. Zhao, and J. Xia, "Multiscale detrended fluctuation analysis of traffic index series," *International Journal of Modern Physics C*, vol. 24, 2014.

[23] Y. Yin and P. Shang, "Multivariate multiscale sample entropy of traffic time series," *Nonlinear Dynamics*, vol. 86, pp. 479–488, 2016.

[24] W. C. Juan Bao, "Coupling complexity analysis of traffic time series based on multivariate multiscale sample entropy," *Boletín Técnico*, vol. 55, pp. 1–9, 2017.

[25] H. B. Xie, W. X. He, and H. Liu, "Measuring time series regularity using nonlinear similarity-based sample entropy," *Physics Letters A*, vol. 372, pp. 7140–7146, 2008.

[26] S. W, C. W, and S. L, "Analysis of complex time series using refined composite multiscale entropy - ScienceDirect," *Physics Letters A*, vol. 378, no. 20, pp. 1369–1374, 2014.

[27] T. L. J. Wang, R. Xie, X. M. Wang, and Y. Y. Cao, "Fault Feature Extraction for Multiple Electrical Faults of Aviation Electro-Mechanical Actuator Based on Symbolic Dynamics Entropy," in *Proceedings of the 2015 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, IEEE, Ningbo, China, September 2015.

[28] J. Lin and K. Eammon, "Experiencing SAX a novel symbolic representation of time series," *Data Mining and Knowledge Discovery*, vol. 5, no. 2, pp. 107–144, 2007.

[29] H. Tan, G. Feng, J. Feng, W. Wang, Y. J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research Part C: Emerging Technologies*, vol. 28, pp. 15–27, 2013.

[30] PeMS, "California Performance Measurement System," http://pems.eecs.berkeley.edu.

[31] "The WisTransPortal System," http://transportal.cee.wisc.edu.

[32] W. Deng, S. Shang, and X. Cai, "An improved differential evolution algorithm and its application in optimization problem," *Soft Computing*, vol. 25, no. 7, pp. 5277–5298, 2021.

[33] X. Cai, H. Zhao, and S. Shang, "An improved quantum-inspired cooperative co-evolution algorithm with muli-strategy and its application," *Expert Systems with Applications*, vol. 171, Article ID 114629, 2021.

[34] W. Deng, S. Shang, and X. Cai, "Quantum differential evolution with cooperative coevolution framework and hybrid mutation strategy for large scale optimization[J]," *Knowledge-Based Systems*, vol. 224, Article ID 107080, 2021.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 1245639, 12 pages
https://doi.org/10.1155/2022/1245639

Research Article

Evaluation of Public Welfare Level of Urban Rail Transit considering Operation Management

Ran Meng,[1] **Baohua Mao**,[1] **Qi Xu**,[1] **and Yang Yang**[2]

[1]*Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China*
[2]*School of Transportation Science and Engineering, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Baohua Mao; bhmao@bjtu.edu.cn

Received 15 February 2022; Accepted 4 March 2022; Published 27 March 2022

Academic Editor: Zhihong Yao

Evaluating the public welfare level of the urban rail transit systems has not only great significance for the government to provide fair and reasonable subsidies but for the better operation and management of urban rail transit enterprises. An evaluation index system composed of 3 criterion-level indicators and 12 subindexes has been established in this paper. The 3 criterion-level indicators conclude service level, social benefit, and policy loss which are all affected by operation management. Besides, the subjective and objective comprehensive weighting method combined with the analytic hierarchy process method and the entropy weight method is proposed to calculate the index-level weights. Furtherly, the grey correlation-TOPSIS comprehensive evaluation method was designed to calculate the comprehensive evaluation value of the public welfare level of each city. To verify the effectiveness of the proposed method, urban rail transit systems in 16 Chinese cities are studied as a case study. The results show that (1) the three indicators of passenger travel cost (25.69%), the increase in housing prices around urban rail transit(10.74%), and operating cost ratio(9.95%) are more likely to affect the evaluation of public welfare level of urban rail transit. (2) the level of public welfare in different cities is not balanced. The cities with a relatively high level of public welfare relative closeness exceeding 0.5 include Shanghai, Beijing, Shenzhen, Guangzhou, Suzhou, Wuhan, Nanjing, Wuxi, and Dalian. (3) Both GDP and urban population are positively correlated with the relative closeness of social benefit and service level. (4) The level of public welfare can be improved by reducing the fare price and improving the service level, such as increasing the network density, reducing the departure interval, and increasing the average speed.

1. Introduction

Urban rail transit companies often suffer from financial losses due to large operating costs and government regulations on service level and fare. Although urban rail transit has great social benefits, enterprises have not received corresponding benefits. The government has to subsidize the policy losses of operating companies. Recently, the deficit of China's urban rail transit enterprises is becoming more and more serious, and it accounts for a large proportion of government fiscal expenditures, which has become a considerable financial burden to the local government. Therefore, evaluating the public welfare level of urban rail transit and distinguishing public welfare and profitability has great significance for urban rail transit operation management and subsidy optimization.

Some scholars have studied the definition of public welfare. Qin [1] introduced the term "public welfare" from Japan at the end of the nineteenth century. Mike [2] proposed that public welfare is synonymous with general welfare, or refers to a broader professional social obligation that exceeds the interests of customers. Greve [3] recommended that public welfare refers to "welfare determined by the public sector and mainly funded by the public sector. Public welfare can pass the tax system or support voluntary sectors, enterprises and families". Zi [4] studied that "public welfare" is closer to philanthropy (caring for human well-being). Wu [5] found that the connotation of "public welfare" includes "nonprofit" and "to promote public welfare."

As far as we know, in this recent decade, there were very few studies focused on the public welfare of urban rail transit, although some scholars have done some research about the welfare of public transit. The welfare of public transit is mainly studied from the perspective of the contribution of public transport services to social welfare and the profit and loss of public transit enterprises. From the perspective of increasing social welfare, Masatoshi [6] proposed that the conventional rules of public facilities and services should aim to increase the economic welfare of the society by realizing the effective distribution of resources and the fair distribution of income. Taylor et al. [7] proposed that the public policy goal of public transit is to solve the often difficult problems such as traffic congestion, car dependence, and suburban expansion while promoting community economic development, employment opportunities, the revitalization of poor communities, urban aesthetics, livability, and mobility of people who are unable or unwilling to drive. Stjernborg et al. [8] recommended that one of the most important welfare effects of a well-functioning public transit system is to establish a fair and open barrier-free environment. Wachs and Taylor [9] studied that public transit can help move welfare recipients out of home and engage in wage jobs. Holmgren [10] analyzed that expanding the demand for public transit requires lower fares from the perspective of maximizing social welfare. From the perspective of public welfare leading to corporate losses, Guerra [11] estimated the passenger welfare and operating deficit of the public transit system and found that from the perspective of economic cost-effectiveness, urban rail transit systems may not be optimal, but they obviously create value for consumers and society. Arcier [12] believes that the financial loss of public transit companies is driven by public policy goals (the entry of vulnerable people into cities, reducing car use and carbon dioxide emissions), rather than demand.

Based on the abovementioned literature, the current research on the public welfare of urban rail transit is mainly through qualitative research, and there are very few researches integrated quantitative methods in the welfare estimation of urban rail transit. The major contributions of this paper can be summarized as follows: (1) firstly, we select 3 criterion-level indicators and 12 index-level indicators through the analysis of the factors that influence the public welfare of urban rail transit. (2) The subjective and objective combination method of the Analytic Hierarchy Process (AHP) and entropy weight method was adopted to determine the weight of public welfare indicators. (3) The grey correlation-Technique for Order Preference by Similarity to an Ideal Solution (GC-TOPSIS) comprehensive evaluation method was used to comprehensively evaluate the public welfare level of the urban rail transit systems.

The remainder of this paper is organized as follows. In Section 2, firstly, the main characteristics of the public welfare of urban rail transit are presented. Then, we introduce the evaluation method of the public welfare level of the urban rail transit system. The performance and application of our proposed models are evaluated through numerical experiments in Section 3. Finally, conclusions with major findings are provided in Section 4.

2. Evaluation Method of Public Welfare Level of the Urban Rail Transit System

2.1. Main Characteristics of the Public Welfare of Urban Rail Transit. The main characteristics of the public welfare of urban rail transit are "universal service" [13] and "not for profit" [14]. Firstly, urban rail transit provides universal service regardless of region, space, time, and population, and takes the public's interests as the initial point to provide equal services. Secondly, urban rail transit operators do not aim for profit. Urban rail transit operators need to provide high-quality service levels that meet the needs of passengers in accordance with government regulations, which will bring large costs and losses to urban rail transit companies. What's more, urban rail transit has positive externalities, which bring large external benefits to the region, such as relieving road traffic congestion, saving energy and reducing emissions, and promoting regional economic development, however, they have not received corresponding returns [15].

2.2. Selection of Public Welfare Level Indicators of the Urban Rail Transit System. This article draws on public welfare level evaluation indicators in some other adjacent fields [16–18], combines the literature of urban rail transit public welfare research and the characteristics of the urban rail transit system, and is based on the principles of scientific, comprehensive, comparable, accessible, absolute and relative number, and finally constructs an urban rail transit public welfare level evaluation index system covering 12 indicators from three aspects: service level, social benefits, and policy loss. The public welfare level evaluation index system of urban rail transit is listed in Table 1.

(1) Service level indicators. The operation and management of urban rail transit enterprises will affect the service level, which in turn affects the benefits of passengers. This indicator reflects the quality of urban rail transit service level. The purpose of the urban rail transit system is to satisfy passengers demand with a service level regulated by the government. Providing operation services is a part of urban rail transit operation management. The service level reflects the degree of effort of urban rail transit companies to serve passengers. The selected indicators are average operating time, maximum load factor during peak time, departure interval, and average speed.

(2) Social benefit indicators. The energy of urban rail transit is electrical energy, so urban rail transit can save energy and reduce emissions. Moreover, the operation of urban rail transit enhances the land value and housing prices along the line [19]. The diversion of urban rail transit also reduces the cost of conventional public transit facilities. The

Table 1: Public welfare level evaluation index system of urban rail transit.

Criterion layer	Index layer	Index explanation	Index unit	Indicator attributes
Service level (A1)	Average operating service time (B1)	Length of service hours in a day	h	+
	Maximum load factor during peak time (B2)	Sum of the number of people staying in high one-way and high section during peak hours divided by sum of the hourly capacity	%	-
	Departure interval (B3)	The time interval between two consecutive transport vehicles passing through a fixed position along the same direction on an operating route	min	-
	Average speed (B4)	Line length divided by run time	km/h	+
	Network density (B5)	The sum of the lengths of all lines of the urban rail transit network divided by the area covered by the urban rail transit network	km/km^2	+
Social benefit (A2)	GDP growth rate (B6)	(Annual regional GDP after operation-annual regional GDP when the project is not constructed)/Annual regional GDP when the project is not constructed × 100%, the annual regional GDP data when the project is not under construction use the data of the year before the subway opened	%	+
	Energy-saving and emission reduction benefits (B7)	Energy-saving and emission reduction benefits are the product of energy-saving and emission reduction benefits per passenger volume and passenger turnover [21]. The benefits of energy-saving and emission reduction per passenger volume are regarded as a fixed value, so energy-saving and emission reduction benefits can be represented by network energy-saving and emission reduction benefits.	Ten thousand people	+
	The increase in housing prices along urban rail transit lines (B8)	(Annual regional average house price after operation−annual regional average house price when the project is not constructed)/annual regional average house price when the project is not under construction × 100%, the annual regional average housing price data when the project is not under construction is selected from the data of the year before the subway opened	%	+
	Replacement of conventional public transport facilities investment benefits (B9)	The investment benefit of replacing conventional public transit facilities is proportional to the benefit of replacing public transit [21] and can be replaced by public transit benefit index	Ten thousand people	+
Policy loss(A3)	Passenger travel cost indicators (B10)	Per capita ticket expenditure/per capita GDP=(average fare ∗ 2 ∗ 250)/per capita GDP	%	-
	Welfare ticket with or without (B11)	If there is a welfare ticket, take 1; if there is no welfare ticket, take 0	-	+
	Operating cost ratio (B12)	Operating ticket revenue/total operating cost	%	-

improvement of the urban rail transit network has strengthened intracity communication and further promoted the economic growth of related industries [20]. The selected indicators include GDP growth rate, energy-saving and emission reduction benefits, the increase in housing prices along urban rail transit lines, and the benefits of replacing conventional public transport facilities.

(3) Policy loss indicators. The government's regulations on fare caps and service levels reflect the characteristics of urban rail transit "serving people." To reduce the burden on passengers, it is necessary to consider the affordability of commuter passengers for fares, which cannot be too high despite the high operating costs. In addition, the departure interval, all-day operating time, speed, etc. must meet the service level stipulated by the government to meet the needs of passengers, which leads to huge operating costs. As a result, fares are generally lower than operating costs, causing operating companies to lose money. The level of fare also reflects the level of welfare. Therefore, the ratio of the average per capita ticket expenditure to the city's per capita GDP is selected as an index for evaluating public welfare. In addition, urban rail transit has implemented preferential policies for vulnerable groups such as the elderly, children, soldiers, and the disabled, which fully shows the welfare nature of urban rail transit. Therefore, the availability of welfare tickets is selected as an index for evaluating the public welfare.

2.3. Evaluation Method of Public Welfare Level of the Urban Rail Transit System. Modern comprehensive evaluation methods include AHP method [22, 23], fuzzy comprehensive evaluation method, TOPSIS evaluation method [24], grey comprehensive evaluation method [25], machine learning method [26, 27] and so on. The basic idea of the entropy weight method is to determine the objective weight according to the variability of the index, which can objectively and truly reflect the index information. AHP method and entropy weight method [28–30] were combined to calculate the index weights. This method balances the subjectivity of expert decision-making and makes the results more scientific and accurate. TOPSIS method analyzes and makes decisions based on the degree to which the data is close to the ideal point and far from the critical point, and the geometric meaning is intuitive. However, this method mainly considers the static distance between the corresponding standard values of the data, but does not consider the consistency of the dynamic trend of the series. The basic principle of the grey correlation method is to define the trend correlation degree to comprehensively characterize the similarity of the dynamic changes of the system indicators and the closeness of the index value change rate. The closer the dynamic change trend of the decision indicators is, the greater the correlation degree. Aiming at the shortcomings of the TOPSIS method, this paper uses the grey correlation method to improve it, so that the evaluation results are more reasonable, and the combination of static distance and dynamic trend can be realized. The basic steps of the AHP-entropy weight-GC-TOPSIS method model are as follows.

(1) Construct the initial decision matrix $X = (x_{ij})_{m\times n}$ by formula

$$X = \left(x_{ij}\right)_{m\times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}, \tag{1}$$

where X is the initial decision matrix; x_{ij} is the j-th index of the i-th city, $i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n$; m and n are the number of cities and indexes, respectively.

(2) Standardize the initial decision matrix using the range method [31], when the index is positive, the standardized calculation $R = (r_{ij})_{m\times n}$ can be expressed as formula (2); when the index is negative, the standardized calculation $R = (r_{ij})_{m\times n}$ can be expressed as formula (3).

$$r_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}}, \tag{2}$$

$$r_{ij} = 1 - \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}}, \tag{3}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the j-th index, respectively.

(4) Use the AHP method [22, 23] to obtain the weight matrix H of each index by formula .

$$H = [h_1 h_2 \ldots h_n], \tag{4}$$

where h_j is the weight of the j-th index; H is the weight matrix.

(5) Use the entropy method [28–30] to obtain the weight matrix W. Calculate the proportion p_{ij} of the i-th item in the j-th index, the entropy value e_j of the j-th item, and the information entropy redundancy t_j. Calculated by formula (5)–(7) [32].

$$p_{ij} = \frac{r_{ij}}{\sum_{i=1}^{m} r_{ij}}, \tag{5}$$

$$e_j = \begin{cases} -\dfrac{1}{\ln m} \displaystyle\sum_{i=1}^{m} p_{ij} \ln p_{ij}, & p_{ij} \neq 0, \\ 0, & p_{ij} = 0, \end{cases} \tag{6}$$

$$t_j = 1 - e_j. \tag{7}$$

(6) Calculate the indicator weight W_j by formula :

$$W_j = \frac{t_j}{\sum_{j=1}^{n} t_j}, \tag{8}$$

where w_j is the weight of the j-th index; W is the weight matrix.

(8) The comprehensive weight is the weighted average of H and W by formula .

$$G = \alpha H + (1 - \alpha) W \, (0 \leq \alpha \leq 1), \tag{9}$$

Where G is the comprehensive weight; α is a parameter, usually 0.5.

(9) Multiply each row element in the normalized matrix R by the corresponding weight G to get the weighted normalization matrix V by formula .

$$V = \left(v_{ij}\right)_{m\times n} = \left(g_j r_{ij}\right)_{m\times n}, \tag{10}$$

where V is the weighted normalization matrix.

(10) Determine the positive ideal solution and the negative ideal solution. The optimal and inferior vectors composed of the maximum and minimum values of each column of the matrix are the positive ideal solution and the negative ideal solution in formula (11) and formula (12).

$$V^+ = \left(v_1^+, v_2^+, \ldots, v_m^+\right), \quad j = 1, 2, \ldots, n, \tag{11}$$

$$V^- = \left(v_1^-, v_2^-, \ldots, v_m^-\right), \quad j = 1, 2, \ldots, n. \tag{12}$$

(11) Calculate the Euclidean distance between each index value and the positive ideal solution and the negative ideal solution by formula (13) and formula (14).

$$D_i^+ = \left\{ \sum_{j=1}^{n} \left(V_{ij} - V_j^+\right)^2 \right\}^{1/2} \quad (i = 1, 2, \ldots, m), \tag{13}$$

$$D_i^- = \left\{ \sum_{j=1}^{n} \left(V_{ij} - V_j^-\right)^2 \right\}^{1/2} \quad (i = 1, 2, \ldots, m), \tag{14}$$

where D_i^+ and D_i^- are the Euclidean distance of the j-th index of the i-th item with concerning the positive ideal solution V^+ and the negative ideal solution V^-.

(12) Calculate the grey correlation degree between the sample and the positive ideal solution and the negative ideal solution, as shown in formula (15) and formula (16).

$$l_{ij}^+ = \frac{\min \min \left|v_j^+ - v_{ij}\right| + \rho \max \max \left|v_j^+ - v_{ij}\right|}{\left|v_j^+ - v_{ij}\right| + \rho \max \max \left|v_j^+ - v_{ij}\right|} \quad (i = 1, 2, \ldots, m), \tag{15}$$

$$l_{ij}^- = \frac{\min \min \left|v_j^- - v_{ij}\right| + \rho \max \max \left|v_j^- - v_{ij}\right|}{\left|v_j^- - v_{ij}\right| + \rho \max \max \left|v_j^- - v_{ij}\right|} \quad (i = 1, 2, \ldots, m). \tag{16}$$

Where l_{ij}^+ and l_{ij}^- are the grey correlation coefficients from the sample to the positive and negative ideal solutions, respectively; $\rho \in (0,1)$ is the resolution coefficient, generally 0.5.

(13) The grey correlation degrees between the i-th sample and the positive and negative ideal solutions are shown in formula (17) and formula (18), respectively.

$$O_i^+ = \frac{1}{n} \sum_{j=1}^{n} l_{ij}^+, \tag{17}$$

$$O_i^- = \frac{1}{n} \sum_{j=1}^{n} l_{ij}^-, \tag{18}$$

where O_i^+ and O_i^- are the grey correlation degrees between the i-th sample and the positive and negative ideal solutions, respectively.

(14) Perform dimensionless processing on the relative closeness degree and grey relational degree, respectively, by formula (19) and formula (20):

$$d_i^+ = \frac{D_i^+}{\max D_i^+}, \quad d_i^- = \frac{D_i^-}{\max D_i^-}, \tag{19}$$

$$o_i^+ = \frac{O_i^+}{\max O_i^+}, \quad o_i^- = \frac{O_i^-}{\max O_i^-}, \tag{20}$$

where d_i^+ and d_i^- represent the result of dimensionless processing of Euclidean distance, and o_i^+ and o_i^- represent the result of dimensionless processing of grey relational degree.

(15) Combine Euclidean distance and grey relational degree by formula (21) and formula (22):

$$S_i^+ = \beta o_i^+ + (1 - \beta) d_i^-, \tag{21}$$

$$S_i^- = \beta o_i^- + (1 - \beta) d_i^+, \tag{22}$$

where S_i^+, S_i^- are the distances from the measurement point to the best and worst points, respectively; β is the preference coefficient, which reflects the preference degree of decision makers for location and trend, and the value range is [0, 1].

(16) Calculate the relative closeness to the ideal solution by formula (23):

$$A_i = \frac{S_i^+}{S_i^+ + S_i^-}, \tag{23}$$

where A_i is the relative closeness degree.

In this study, A_i values calculated from the AHP-entropy weight-GC-TOPSIS model are in the range between 0 and 1; the solution is classified as a higher performance as the values approach closer to 1.

3. Case Analysis

Taking the urban rail transit systems of 16 representative cities in China as an example, this paper uses the above method and MATLAB software to evaluate their levels of public welfare. The data relating to urban rail transit operation adopts the operation data collected from the urban rail transit operating companies of each city in 2016. The urban GDP data was derived from the National Bureau of Statistics website, the population data from the official website of the Ministry of Housing and Urban-Rural Development of the People's Republic of China, and the house price data from the Anjuke website. The original data are shown in Table 2.

3.1. Calculation of the Weight of the Evaluation Index of the Public Welfare Level. The subjective and objective combination method and grey relation-TOPSIS evaluation method established above are used to evaluate the public welfare level of urban rail transit.

The weight of the AHP method comes from the scores of 10 experts in the field. α, ρ, and β all take the value 0.5. According to formulas (1)–(9), the weights of each indicator can be calculated, as listed in Table 3.

As can be seen from Table 3, the AHP method shows that the weights of the criterion layer from large to small are policy loss (62.67%), service level (27.97%), and social benefit (9.36%). The top five weights of the indicator layer are the passenger travel cost indicator (46.56%), operating cost ratio (12.15%), network density (9.67%), departure interval (8.12%), and average speed (5.43%). The subjective weight method indicates that policy loss and service level are more important for public welfare. Since policy loss and service level reflect the benefits of passengers, this shows that the subjective weight method pays more attention to the benefits of passengers for public welfare.

The result of the entropy weight method is that the weight of the criterion layer from large to small is the social benefit (63.17%), service level (21.31%), and policy loss (15.52%). The top five weights of the indicator layer are the increase in housing prices around rail transit (21.07%), the benefit of energy conservation and emission reduction (15.26%), the investment benefit of replacing conventional public transport facilities (15.22%), the GDP growth rate (11.62%), the network density (8.19%). It can be seen that the results of the entropy weight method show that social benefit index and service level index are more important for evaluating the level of public welfare. This is mainly because the principle of the entropy weight method is based on the degree of dispersion of the indicators, and the degree of dispersion of these indicators in different cities is relatively large, so the weights of these indicators are relatively large.

The result calculated by the comprehensive weight method is that the weight of the criterion layer from large to small is the policy loss (39.09%), the social benefit (36.27%), and the service level (24.64%). The top three weights of the indicator layer are the passenger travel cost index (25.69%), the increase in housing prices around rail transit (10.74%), and operating cost ratio (9.95%). For the service level indicator, the network density (8.93%), departure interval (5.87%), and average speed (4.47%) are the more important indicators; For the policy loss indicator, the passenger travel cost index (25.69%) and operating cost ratio (9.95%) are the more important indicators.

The weighted standardized decision matrix is calculated according to formula (10). To further reflect the public welfare level of different types of cities, cities are classified according to region and scale, and the weighted standardized results of different types of cities are averaged. Figure 1 shows the weight normalization results under different classifications.

Note: the northeast region includes Shenyang, Dalian, and Harbin; the eastern region includes Beijing, Shanghai, Suzhou, Guangzhou, Shenzhen, Nanjing, Wuxi, and Dongguan; the western region includes Chongqing, Xi'an, and Nanning; the central region includes Wuhan and Nanchang. Megacities (>10 million people) include Beijing, Shanghai, Chongqing, and Shenzhen; supercities (5–10 million people) include Guangzhou, Nanjing; large cities (1–5 million people) include Shenyang, Xi'an, Suzhou, Harbin, Wuhan, Dalian, Wuxi, Nanchang, Dongguan and Nanning.

As shown in Figure 1(a), it can be found from the perspective of regional geographical location as follows. (1) From east to west, the public welfare, service level, social benefits, and policy loss are all gradually decrease. This may be due to from east to west, the economic level and population gradually decrease, and the passenger flow also decreases relatively, so the social benefit and service level are relatively low. (2) The northeast region ranks first in terms of policy-related losses, but its social welfare and service level indicators are smaller than those in the central region, and its social benefit index ranks last.

As shown in Figure 1(b), it can be found from the perspective of city scale as follows. (1) It can be seen that with the increase of city size, the weighted average results of service level indicator and social benefit indicator both increase. The service level results of large cities, supercities, and megacities are 0.423, 0.549, and 0.572, respectively, and the

Table 2: Original data.

City	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	Urban population(10 thousand people)	GDP (billion)
Beijing	18.6	118	120	37.9	0.39	0.16	4332	584	500	3.6	0	0.57	1880	27041
Shanghai	17.9	114	135	35.8	0.59	1.01	8305	3376	928	4.4	0	0.94	2420	29887
Chongqing	17.0	106	150	43.6	0.16	0.4	1782	194	190	5.7	1	0.56	1103	18023
Guangzhou	17.7	115	118	35.4	0.22	0.63	4860	616	679	2.8	1	0.92	627	19611
Shenzhen	16.6	101	150	37.3	0.29	0.44	2538	682	299	2.5	1	1.02	1191	20686
Wuhan	16.8	79	216	31.8	0.31	0.47	1485	563	196	2.5	1	0.95	473	11531
Nanjing	17.0	78	140	43.6	0.29	0.37	1864	504	227	2.4	1	0.71	591	10819
Shenyang	17.7	128	225	24.3	0.09	0.04	611	76	82	4.1	1	0.87	471	5290
Dalian	15.4	84	210	39.4	0.33	0.30	436	393	35	2.8	1	0.36	327	5650
Xi'an	17.0	92	173	33.5	0.10	0.16	881	46	106	4.2	1	0.96	436	6396
Harbin	15.6	82	330	30.5	0.04	0.09	121	38	19	3.7	1	0.52	422	4370
Suzhou	16.3	94	220	32.0	0.15	0.09	292	39	41	2.1	1	0.39	263	15400
Wuxi	17.0	56	410	34.5	0.17	0.05	161	0	23	2.2	1	0.31	216	9340
Nanchang	16.4	94	390	32.4	0.09	0.09	161	9	22	3.4	1	1.04	244	4140
Dongguan	16.0	76	410	51.8	0.04	0.09	135	39	10	5.8	1	0.29	201	7260
Nanning	13.3	97	480	31.2	0.10	0.09	20	8	3	5.6	1	0.08	225	3730

Table 3: Weights and rankings for different indicators.

Criterion layer	AHP method	Entropy method	Comprehensive weighting method	Index layer	AHP method	Sort	Entropy method	Sort	Comprehensive weighting method	Sort
A1	0.2797	0.2131	0.2464	B1	0.0201	10	0.0208	12	0.0204	12
				B2	0.0275	9	0.0391	8	0.0333	11
				B3	0.0812	4	0.0362	9	0.0587	8
				B4	0.0543	5	0.0351	10	0.0447	9
				B5	0.0967	3	0.0819	5	0.0893	6
A2	0.0936	0.6317	0.3627	B6	0.0144	11	0.1162	4	0.0653	7
				B7	0.0389	7	0.1526	2	0.0958	4
				B8	0.0042	12	0.2107	1	0.1074	2
				B9	0.0361	8	0.1522	3	0.0941	5
A3	0.6267	0.1552	0.3909	B10	0.4656	1	0.0483	7	0.2569	1
				B11	0.0396	6	0.0295	11	0.0346	10
				B12	0.1215	2	0.0774	6	0.0995	3

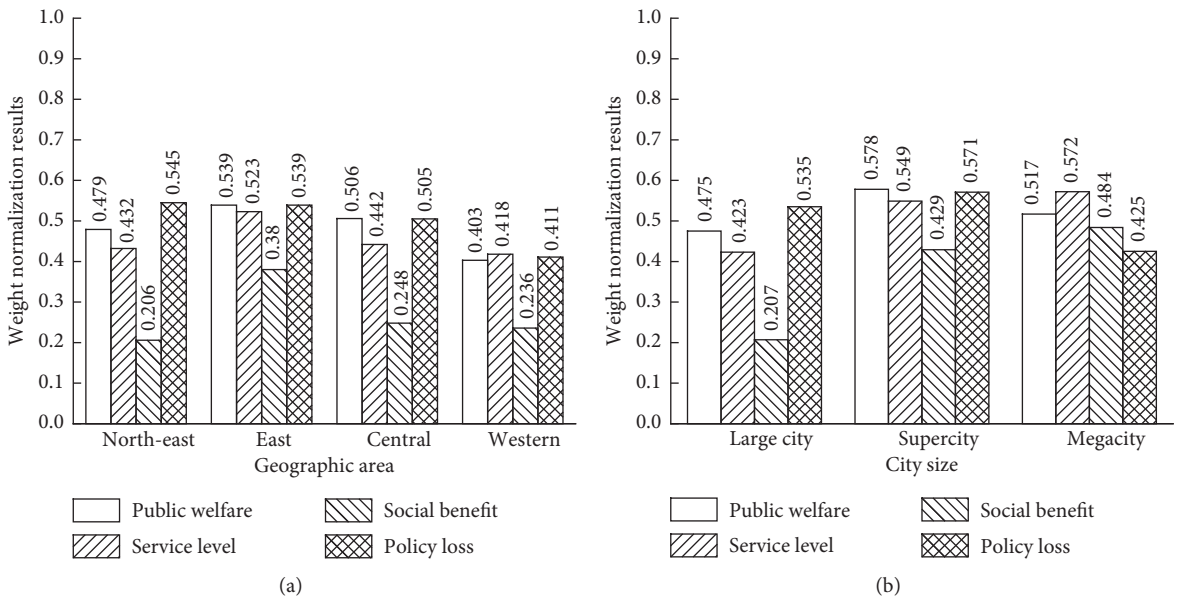


Figure 1: Weight normalization results under different classifications (a) Geographic area (b) City size.

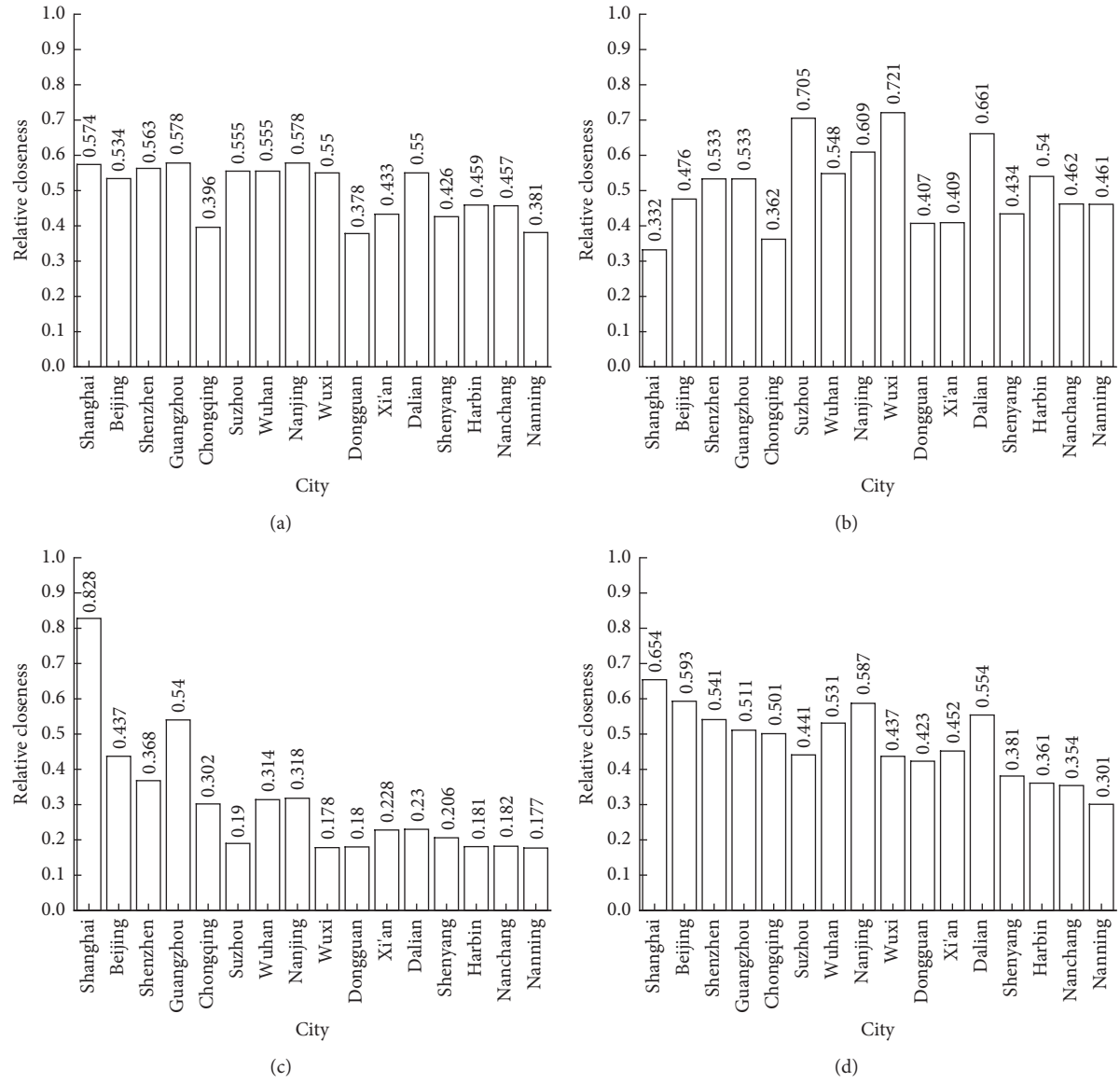


Figure 2: The relative closeness of different indicators in different cities: (a) public welfare, (b) policy loss, (c) social benefit, and (d) service level.

social benefit indicators are 0.207, 0.429, and 0.484, respectively. This may be because larger cities tend to have higher passenger flow and economic levels, the service level will increase accordingly, and the social benefits will also increase. (2) The result of policy-related losses is the smallest in megacity, only 0.425, this is mainly because the passenger travel cost indicators in these cities are relatively high, and the operating cost ratio of Shanghai and Shenzhen is also relatively high.

3.2. Measurement and Calculation of Comprehensive Evaluation Indicators. According to equations (11)–(23), the relative closeness of the public welfare of the rail transit systems of 16 cities is shown in Figure 2. Cities are listed in descending order of GDP.

Figure 2(a) shows that the level (relative closeness) of public welfare in different cities is not balanced. From Shanghai to Wuxi, these cities with relatively high GDP, except for Chongqing, have a relatively low degree of closeness, and the relative closeness of other cities is around 0.55. This is mainly because the passenger travel cost indicator in Chongqing is relatively high, reaching 5.7. In the latter seven cities, except for Dalian, where the relative closeness suddenly increased, the relative closeness of several other cities was less than 0.5, which was at a relatively low level. This is mainly due to the relatively low passenger travel cost index which is 2.8, the relatively low operating cost ratio which is 0.36, and the relatively high network density which is 0.33 in Dalian.

Figure 2(b) shows that from Shanghai to Wuxi, the policy loss relative closeness of cities with relatively high GDP, except for Shanghai and Chongqing, is above 0.47.

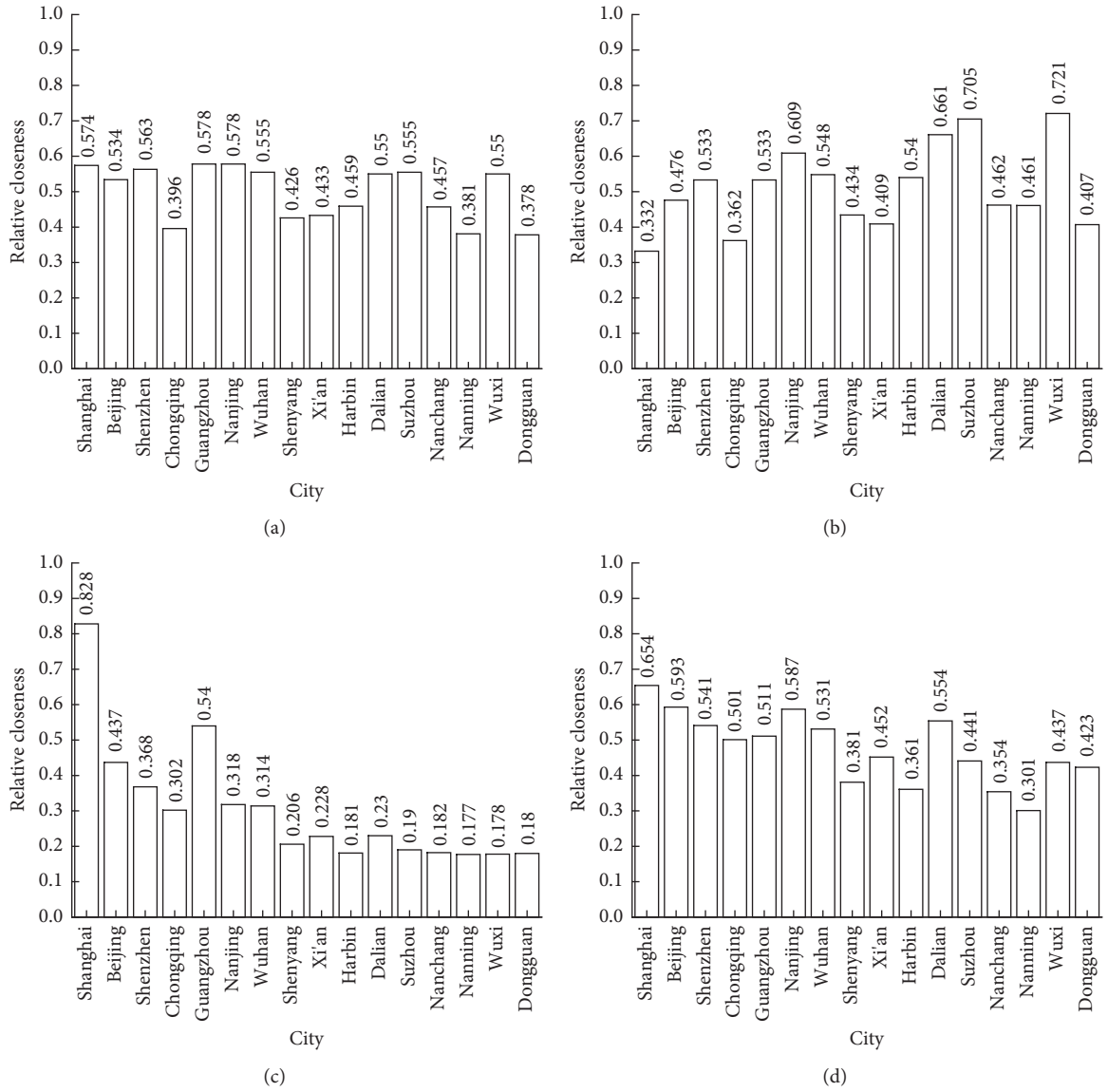


Figure 3: The relative closeness of different indicators in different cities: (a) public welfare, (b) policy loss, (c) social benefit, and (d) service level.

This is mainly because the passenger travel cost index in Chongqing is relatively high, reaching 5.7, and the passenger travel cost index and operating cost ratio index in Shanghai are also very high, reaching 4.4 and 0.94 respectively. In the latter seven cities, except for Dalian and Harbin, where the policy loss relative closeness of policy loss suddenly increased, the policy-loss relative closeness of several other cities was less than 0.5, which was at a relatively low level. This is mainly due to the relatively low passenger travel cost indicators and operating cost ratios in Dalian and Harbin, 2.8 and 0.36 in Dalian and 3.7 and 0.52 in Harbin respectively.

Figure 2(c) shows that the relative closeness of the social benefit indicator shows an obvious trend of decreasing gradually with the decline of the economic level, which indicates that the urban rail transit in the more developed cities tends to have higher benefits.

Figure 2(d) shows that with the decline of GDP, the relative closeness of the service level of each city shows an obvious downward trend. This may be because the cities with lower economic levels tend to have smaller populations, so the service level will be relatively low.

When cities are listed in descending order of urban population, the relative closeness of public welfare, policy loss, social benefit, and service level are shown in Figure 3.

Figure 3(a) shows that from Shanghai to Wuhan, cities with a large population except Chongqing have a relative closeness to public welfare of more than 0.5, the relative closeness of public welfare in Chongqing is only 0.396. This may be due to the relatively low indicators of social benefits and policy losses in Chongqing. Among the other cities, the relative closeness of public welfare in Dalian, Suzhou, and Wuxi is more than 0.5, which is mainly due to their relatively

Table 4: Correlation analysis results.

		The relative closeness of public welfare	The relative closeness of service level	The relative closeness of social benefit	The relative closeness of policy loss	Urban population	GDP
Urban population	Pearson correlation	0.308	0.688**	0.866**	−0.453	1	0.885**
	Sig. (double tail)	0.246	0.003	0.000	0.078		0.000
	Number of cases	16	16	16	16	16	16
GDP	Pearson correlation	0.359	0.616*	0.776**	-0.300	0.885**	1
	Sig. (double tail)	0.172	0.011	0.000	0.258	0.000	
	Number of cases	16	16	16	16	16	16

**indicates a significant correlation at the 0.01 level (two-tailed). * indicates a significant correlation at the 0.05 level (two-tailed).

high policy-related loss indicators and service level indicators, while the relative closeness of public welfare in other cities is lower than 0.5.

Figure 3(b) shows that the relative closeness of policy losses in some cities is significantly lower than that of other cities, including Shanghai, Chongqing, Shenyang, Xi'an, Nanchang, Nanning and Dongguan, all of which are lower than 0.5. This is mainly due to the relatively high passenger travel cost index and operating cost ratio index in Shanghai, Shenyang, and Xi'an, which are 4.4 and 0.94, respectively; while although Chongqing, Nanning, and Dongguan have relatively low operating cost ratios, but their passenger travel cost index is relatively high, were 5.7, 5.6 and 5.8; Nanchang's operating cost ratio was very high, reaching 1.04.

Figure 3(c) and Figure 3(d) show that with the reduction of the urban population, the relative closeness of social benefit and service level both show a downward trend. This is because cities with smaller populations tend to have less passenger flow. Therefore, various social benefits such as replacement of public transport benefits and benefits of reducing pollution will also decrease. At the same time, urban rail transit construction in cities with fewer populations usually starts later, and the density of the network are always being smaller, and due to less passenger flow, the departure interval and average speed will also decrease to save costs.

This paper uses SPSS software to carry out a correlation analysis on GDP, urban population, and the relative closeness of public welfare, social benefit, policy loss, and service level. The results are shown in Table 4.

It can be seen from Table 4 that there is a significant positive correlation between GDP and urban population; on the other hand, there is a positive correlation between GDP and the relative closeness of social benefit and service level; there is a positive correlation between urban population and the relative closeness of social benefit and service level. However, there is no obvious correlation between GDP, population, and the relative closeness of public welfare and policy losses. This verifies the correctness of the above results in this paper.

4. Conclusions

The evaluation index system for the public welfare level of urban rail transit is established, which adopts the combination of a subjective and objective method to determine the weight of public welfare indicators, and the grey relation-TOPSIS method is adopted to evaluate the levels of public welfare of different urban rail transit systems in different cities. The conclusions are as follows:

(1) The public welfare level evaluation indicators account for a large proportion of the passenger travel cost index(25.69%), the increase in housing prices around rail transit(10.74%), and operating cost ratio(9.95%); for the service level indicator, network density(8.93%), departure interval(5.87%) and average speed(4.47%) are the more important indicators, so governments and enterprises can increase public welfare by reducing the fare price and improving the service level, such as increasing the network density, reducing the departure interval, and increasing the average speed.

(2) The levels(relative closeness) of public welfare of each city are not balanced. Cities with a large population except Chongqing have a public welfare relative closeness of more than 0.5, the relative closeness of public welfare in Chongqing is only 0.396. This may be due to the relatively low indicators of social benefits and policy losses in Chongqing, such as its high passenger travel cost index of 5.7. Among the other cities, the relative closeness of public welfare in Dalian, Suzhou, and Wuxi is more than 0.5, which is mainly due to their relatively high policy-related loss indicators and service level indicators, while the relative closeness of public welfare in other cities is lower than 0.5.

(3) There is a positive correlation between GDP and urban population; there is a positive correlation between GDP and the relative closeness of social benefit and service level; Furthermore, there is also a positive correlation between urban population and

the relative closeness of social benefit and service level. However, there is no obvious correlation between GDP, population, and the relative closeness of public welfare and policy losses.

The evaluation of the level of public welfare of urban rail transit in this paper has a certain guiding significance for the operation and management strategies of the government and urban rail transit operating enterprises to improve the level of public welfare. Future studies could fruitfully explore this issue further by conducting research on subsidy and fare optimization on the basis of current public welfare research.

Data Availability

Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (71971021) and Beijing Natural Science Foundation (J210001).

References

[1] H. Qin, *Modernization beyond Government and Enterprises: A Comparative Study of the History of Chinese and Western Public Welfare*, Zhejiang People's Publishing House, Hangzhou, China, 1999, (in Chinese).

[2] S. Mike, "Analyzing the professions: the case for the neo-Weberian approach," *Comparative Sociology*, vol. 9, no. 6, pp. 887–915, 2010.

[3] B. Greve, "What is welfare?," *Central European Journal of Public Policy*, vol. 2, no. 1, 2008.

[4] Z. Zi, *The Fate of Wealth: Review of American Modern Public Welfare Foundation*, Shanghai People's Publishing House, Shanghai, China, 2003, (in Chinese).

[5] J. Wu, "A study of public welfare in public hospitals," *Comparative Economic & Social Systems*, vol. 4, pp. 13–20, 2012, (in Chinese).

[6] M. A. Abe, *Pricing and Welfare in Urban Transportation*, Economics Faculty Research and Publications, New Jersey, NJ, USA, 1973.

[7] B. D. Taylor and E. A. Morris, "Public transportation objectives and rider demographics: are transit's priorities poor public policy?" *Transportation*, vol. 42, no. 2, pp. 347–367, 2015.

[8] V. Stjernborg and O. Mattisson, "The role of public transport in society-A case study of general policy documents in Sweden," *Sustainability*, vol. 8, no. 11, p. 1120, 2016.

[9] M. Wachs and B. D. Taylor, "Can transportation strategies help meet the welfare challenge?" *Journal of the American Planning Association*, vol. 64, no. 1, pp. 15–19, 1998.

[10] J. Holmgren, "A strategy for increased public transport usage - the effects of implementing a welfare maximizing policy," *Research in Transportation Economics*, vol. 48, pp. 221–226, 2014.

[11] E. Guerra, "Valuing rail transit," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2219, no. 1, pp. 50–58, 2011.

[12] B. Faivre d'Arcier, "Measuring the performance of urban public transport in relation to public policy objectives," *Research in Transportation Economics*, vol. 48, pp. 67–76, 2014.

[13] S. Kim and W. Vandenabeele, "A strategy for building public service motivation research internationally," *Public Administration Review*, vol. 70, no. 5, pp. 701–709, 2010.

[14] B. H. Mao, Z. Zhang, Z. J. Chen, J. W. zheng, and H. T. kin, "A review on operational technologies of urban rail transit networks," *Journal of Transportation Systems Engineering and Information Technology*, vol. 17, no. 6, pp. 155–163, 2017, (in Chinese).

[15] Y.-K. Qiao, F.-L. Peng, and Y. Wang, "Valuing external benefits of underground rail transit in monetary terms: a practical method applied to Changzhou City," *Tunnelling and Underground Space Technology*, vol. 83, pp. 91–98, 2019.

[16] Y. L. Jiao, L. H. Fan, and F. Yu, "Research on the construction of public welfare evaluation system from the perspective of party building inpublic hospitals," *Chinese Hospital Management*, vol. 40, no. 5, pp. 47–50, 2020, (in Chinese).

[17] J. Li, F. Liu, C. L. Tan, X. Lingzhong, Z. Liang, and F. Zhanchun, "Research on fuzzy comprehensive evaluation of public welfare in public hospitals in sichuan province from the patients' perspective," *Medicine in Society*, vol. 38, no. 8, p. 10, 2019, (in Chinese).

[18] W. X. Lv, T. Y. Wang, and D. Zhu, "Researches on the assessment and maintenance mechanism of university canteen's public welfare under socialization reform," *Education Economics*, vol. 2016, no. 5, pp. 66–72, 2016, (in Chinese).

[19] D. Zhang and J. Jiao, "How does urban rail transit influence residential property values? Evidence from an emerging Chinese megacity," *Sustainability*, vol. 11, no. 2, p. 534, 2019.

[20] Y. Sun and Y. Cui, "Evaluating the coordinated development of economic, social and environmental benefits of urban public transportation infrastructure: case study of four Chinese autonomous municipalities," *Transport Policy*, vol. 66, pp. 116–126, 2018.

[21] S. X. Chen and X. M. Tao, "Estimative analysis of the social benefit of shanghai UMT system," *Urban Mass Transit*, vol. 7, no. 1, pp. 1–5, 2004, (in Chinese).

[22] M. Nassereddine and H. Eskandari, "An integrated MCDM approach to evaluate public transportation systems in Tehran," *Transportation Research Part A: Policy and Practice*, vol. 106, pp. 427–439, 2017.

[23] E. Broniewicz and K. Ogrodnik, "Multi-criteria analysis of transport infrastructure projects," *Transportation Research Part D: Transport and Environment*, vol. 83, Article ID 102351, 2020.

[24] Z. Deng, Z. F. Li, Y. T. Zhou, X. Chen, and S. S. Liang, "Measurement and spatial spillover effects of port comprehensive strength: empirical evidence from China," *Transport Policy*, vol. 99, pp. 288–298, 2020.

[25] S. Chen, Y. Leng, B. Mao, and S. Liu, "Integrated weight-based multi-criteria evaluation on transfer in large transport terminals: a case study of the beijing south railway station," *Transportation Research Part A: Policy and Practice*, vol. 66, pp. 13–26, 2014.

[26] Y. Yang, K. He, Y. P. Wang, Z. Z. Yuan, Y. H. Yin, and M. Z. Guo, "Identification of dynamic traffic crash risk for cross-area freeways based on statistical and machine learning methods," *Physica A: Statistical Mechanics and Its Applications*, vol. 595, Article ID 127083, 2022.

[27] X. Chen, H. Chen, Y. Yang et al., "Traffic flow prediction by an ensemble framework with data denoising and deep learning model," *Physica A: Statistical Mechanics and Its Applications*, vol. 565, Article ID 125574, 2021.

[28] Y. Yang, Z. Yuan, J. Chen, and M. Guo, "Assessment of osculating value method based on entropy weight to transportation energy conservation and emission reduction," *Environmental Engineering and Management Journal*, vol. 16, no. 10, pp. 2413–2423, 2017.

[29] Z. Z. Yuan, K. He, and Y. Yang, "A roadway safety sustainable approach: modeling for real-time traffic crash with limited data and its reliability verification," *Journal of Advanced Transportation*, vol. 2022, Article ID 1570521, 14 pages, 2022.

[30] L. J. Li, Y. Yang, Z. Z. Yuan, and Z. Chen, "A spatial-temporal approach for traffic status analysis and prediction based on Bi-LSTM structure," *Modern Physics Letters B*, vol. 35, no. 31, Article ID 2150481, 2021.

[31] F. W. Gao, Z. B. Zhang, and M. X. Shang, "Risk evaluation study of urban rail transit network based on entropy-TOPSIS-coupling coordination model," *Discrete Dynamics in Nature and Society*, vol. 2021, Article ID 5124951, 8 pages, 2021.

[32] H. Liu, Y. Dong, and F. Wang, "Gas outburst prediction model using improved entropy weight grey correlation analysis and IPSO-LSSVM," *Mathematical Problems in Engineering*, vol. 2020, no. 7, 10 pages, Article ID 8863425, 2020.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 2205292, 17 pages
https://doi.org/10.1155/2022/2205292

Research Article

A Person-Based Adaptive Traffic Signal Control Method with Cooperative Transit Signal Priority

Wei-Hsun Lee [1,2] **and Hsuan-Chih Wang** [1]

[1]*Department of Transportation and Communication Management Science, National Cheng Kung University, Tainan, Taiwan*
[2]*Center for Innovative FinTech Business Models, National Cheng Kung University, Tainan, Taiwan*

Correspondence should be addressed to Wei-Hsun Lee; leews@mail.ncku.edu.tw

Received 18 December 2021; Revised 21 January 2022; Accepted 8 February 2022; Published 27 March 2022

Academic Editor: Gen Li

Real-time traffic signal control has long been a critical way to improve traffic congestion. Transit Signal Priority (TSP) is seen as a cost-effective way to reduce travel time variability. Most of the previous studies develop real-time signal control systems on a vehicle basis, which is unable to efficiently provide preferential treatment on transit vehicles. Person-based signal control systems, which transform traffic delay computation units from vehicle to passenger, have been proposed to try to address this limitation. However, their models, optimizing signal plan cycle-by-cycle, cannot rapidly respond to traffic variations. This study proposes a Person-based Adaptive traffic signal control method with Cooperative Transit signal priority (PACT). In PACT, not only do Road-Side Units (RSUs) perform signal optimization, but also On-Board Units (OBUs) provide in-vehicle speed advisory to reduce delays. The interaction between RSU and OBU is conducted second-by-second, which has high adaptability to traffic variations. Experiments are performed based on real traffic data via traffic simulation platform SUMO. The results indicate that PACT can efficiently reduce delays of both bus passengers and auto passengers at a signalized intersection. Compared to preoptimized signal plans, the results show that each passenger on transit vehicles experiences 33%–70% decreases in delays, and each auto passenger experiences 3%–29% decreases in delays. PACT can reduce 80%–98% in delays when the occupancy weight factor is relatively large, showing the potential of extending PACT on performing signal preemption.

1. Introduction

With the growth of the city, traffic congestion has become a serious problem since it causes high air pollution and fuel consumption. One of the solutions to such a problem is to enhance the mobility of the transportation system. In urban transportation networks, traffic signal control plays a significant role in the efficiency of urban mobility. Adaptive traffic signal control (ATSC) systems can rapidly adjust signal settings in response to dynamic changes in traffic flow. In the past several decades, abundant studies of ATSC, such as SCAT [1], SCOOT [2], OPAC [3], PRODYN [4], and RHODES [5], have been proposed. In order to optimize signal settings, ATSC strategies need to predict traffic conditions based on traffic data collected from vehicle detection techniques. The quality and dimensions of traffic data significantly affect the performance of ATSC systems [6]. Most of the ATSC systems [1–5] are developed based on fixed-point detectors (e.g., loop detectors), which provide limited dimensions of traffic information. The high cost of such detectors on deployment, maintenance, and operation when collecting data on a large scale weakens the signal control accuracy [7].

More recently, the emergence of Connected Vehicle (CV) technology brings new sights of the approaches on traffic signal control. CV technology has shown its great potential for improving signalized intersection efficiency and safety [8–10]. With Vehicle-to-Infrastructure (V2I) communication, real-time states of vehicles such as vehicle type, speed, position, and trajectories can be used to enhance signal control modelling. Conventional loop-detector-based ATSC strategies are unable to deal with such abundant real-time traffic information. Abundant studies have contributed to CV-based real-time signal control [11–14]. However, it is

expected that the penetration rate of On-Board Unit (OBU) will remain low in the near future due to technological and economic challenges [15]. The requirement of a high penetration rate of OBU limits the applicability of CV-based systems in real life. In recent years, Vision-based Traffic flow Detection (VTD) systems powered by deep learning techniques, such as YOLO [16], can capture all movement (i.e., speed and position) of vehicles with high-resolution cameras [17]. The systems can provide detailed real-time traffic states without requiring a high penetration rate of OBU.

Transit Signal Priority (TSP) is a collection of techniques that provide transit vehicles riding through a signalized intersection smoothly. TSP decreases the travel time variability of buses and enhances operation reliability by adjusting signal settings that favour bus arrivals [18]. However, serving signal priority to buses may deteriorate nonprioritized traffic. To lessen negative impacts, several studies have developed optimization models of TSP under various constraints [19–21]. Though both ATSC and TSP aim to mitigate traffic congestion, there is a conflict between the objectives of ATSC and TSP. The objective of TSP is to minimize bus delays at intersections; however, traditionally, the objective of ATSC is to minimize total vehicle delays. A study has shown that this is a trade-off relationship [22]. In contrast to vehicle-based signal control, which ignores person mobility in the transportation network, a series of studies have proposed a person-based traffic signal control framework [23–28]. The objective of person-based signal control is to minimize total person delays at an isolated intersection or multiple intersections. Person-based systems transform the number of vehicles to persons via a weighted function. With such a system, signal control can decrease delays of vehicles and make transit vehicles naturally get "signal priority" due to high passenger occupancy. Christofa et al. [23] firstly proposed a responsive traffic signal control system that incorporates Transit Signal Priority on the basis of passenger occupancy. Then the person-based signal control model is revised to reduce computation complexity [24]. In addition to adaptive signal control, the person-based approach has also been applied on optimizing lane allocation and passive TSP timing settings [26], signal optimization for center transit lanes with various types of travelers [27], and dynamic exclusive bus lane optimization [28]. Zeng et al. [29] formulated the person-based signal control model under the CV environment, and Yu et al. [30] relaxed the constrained on fixed cycle length and considered uncertain bus arrival time. Li et al. [31] considered the effects of passenger and pedestrian delays at downstream intersections under coordinated green wave control.

In addition to CV-based signal control, another emerging CV application is in-vehicle Driving Advisory Systems (DASs). DASs provide driver speed advisories for various objectives (e.g., eco-driving and passengers' comfort) based on Signal Phase and Timing (SPaT) information. A well-known example of DASs is GLOSA [32], followed by a series of studies on DASs for eco-driving [33, 34]. The Cooperative Intelligent Transportation System (C-ITS) established a two-way communication scheme between signals and vehicles, which allows cooptimizing traffic signal control and vehicle traveling efficiency. Several studies have evaluated C-ITS frameworks, and the results show significant improvement in either fuel consumption or travel time. However, these works either assume a 100% penetration rate of OBUs on all vehicle types [35] or postulating vehicles are homogenous, which does not take different vehicle types (e.g., transit vehicles) and passenger numbers into consideration [36, 37]. Hu et al. [38] and Wu et al. [39] enabled DASs on TSP to coordinate traffic signal plans among multi-intersections. Seredynski et al. [40–43] presented a complete analysis on implementing DASs on TSP. The results demonstrate that DASs can enhance TSP control efficiency without dramatically interfering in nonprioritized traffic. Nonetheless, little attention has been paid to integrate DASs into person-based real-time signal control.

In summary, several studies have contributed to the person-based signal control and cooperation method on a signalized intersection. All the literature on person-based signal control systems optimizes signal settings cycle-by-cycle, which may not be able to respond to traffic variations in real time. Although transit vehicles can get preferential treatment in person-based systems due to higher occupancy, buses may still experience delays at intersections since the interference of other traffic factors, such as the volume of competitive traffic. Inspired by the above points, this study proposed a Person-based Adaptive traffic signal control with Cooperative Transit signal priority (PACT) system. The contribution of the proposed PACT method can be summarized as follows:

(1) PACT is an integrated system featuring a signal-vehicle cooperate-control structure. PACT optimizes signal settings in a short period (per second) with the objective of minimizing total person delay. The real-time optimal signal settings are transmitted to OBUs via I2V, and then OBUs can calculate optimal speed advisory with the objective of maximizing passing probability of the intersection.

(2) PACT performs both signal optimization and speed advisory calculation for buses. The rolling horizon procedure provides additional reduction in bus delays without dramatically deteriorating the performance of nonprioritized traffic. Therefore, PACT can minimize total person delays and enhance the traveling efficiency of transit vehicles.

The remainder of this paper is organized as follows: Section 2 depicts the structure of the proposed PACT system and model formulation of signal control and speed advisory algorithm. In Section 3, a case study of an isolated intersection is introduced, following the evaluation of the effectiveness and performance of the PACT system. Finally, the findings of this study and future works are discussed in Section 4.

2. Methodology

2.1. System Framework. The concept of PACT is illustrated in Figure 1. At each second, PACT sequentially conducts the

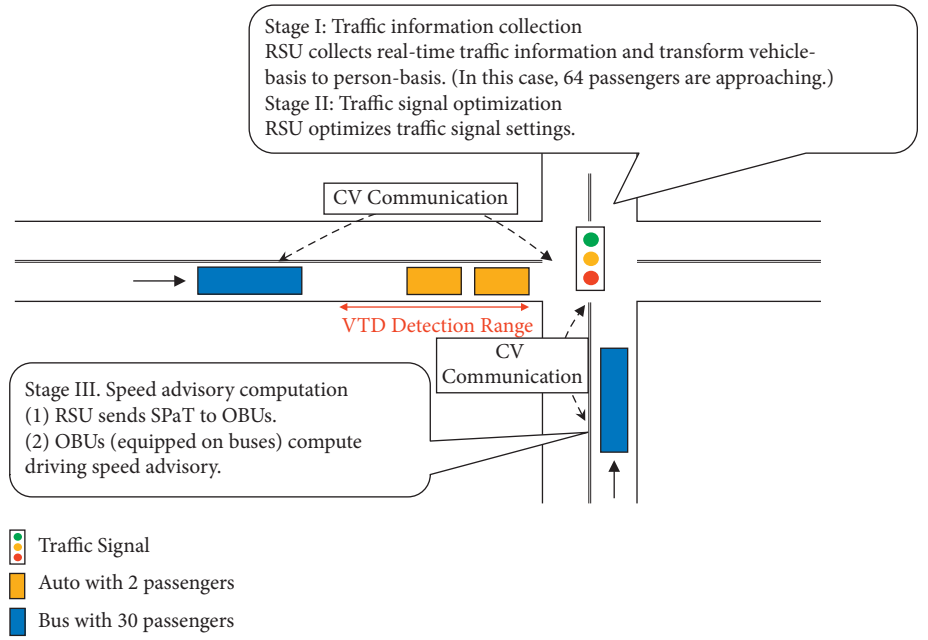


Figure 1: The process of PACT.

following three stages. In stage I, the RSU retrieved real-time traffic information (i.e., speed and position) using VTD techniques and transformed the vehicle unit into person unit. In stage II, the RSU optimizes signal timing settings with the objective of minimizing total person delay. A Binary Mixed Integer Linear Program (BMILP) model revised by Zeng et al. [29] is constructed and utilized by the signal timing optimization. In stage III, the RSU sends SPaT information to OBUs on transit vehicles via CV communication. OBUs perform driving speed advisory computation with the objective of maximizing intersection passing probability. In the next second, three stages are repeated.

2.2. Model Formulation. For model simplicity, we have the following assumptions in the proposed PACT:

(1) All the transit vehicles are equipped with OBUs. OBUs send vehicle information (i.e., speed and position) and receive SPaT via V2I communication such as IEEE 802.11p or C-V2X.

(2) Bus stops are far side located. The impacts of bus dwell time are neglected.

(3) Bus drivers fully comply with the optimal driving speed advisory as long as safety spacing is maintained.

(4) Since the standard CV domain roughly ranges from 150 to 300 m [44], both CV communication range and VTD detection range are set as 300 m.

2.2.1. Objective of Signal Optimization. The signal optimization model of PACT is developed by Zeng et al. [29], in which delays of each vehicle are calculated individually under the CV environment assumption. PACT divides vehicles in the VTD detection range into two groups: queueing vehicles and approaching vehicles. Vehicle's speed <1 km/hr is defined as queueing; the others are defined as approaching. In each group, every vehicle is labelled an order x according to relative position to the stop bar. J is the number of total phases, I_j^Q refers to the number of queueing vehicles, and I_j is the total number of vehicles in the VTD detection range. $W_{x,j}$ represents the person-based weighted parameter of vehicle x of phase j. The objective of signal optimization is to minimize both delays of queueing vehicles $d_{x,j}^Q$ and approaching vehicles $d_{x,j}$. The objective function is formulated as follows:

$$\text{Min} \sum_{j=1}^{J} \sum_{x=1}^{I_j^Q} W_{x,j} d_{x,j}^Q + \sum_{j=1}^{J} \sum_{x=I_j^Q+1}^{I_j} W_{x,j} d_{x,j}. \quad (1)$$

Person-based weighted parameter ($W_{x,j}$) is obtained through equation (2). O_A is auto occupancy and O_B is transit vehicle occupancy. The parameter α is introduced to increase the priority level on transit vehicles. Real-time auto passenger occupancies data are currently not available. However, occupancies of autos vary slightly from day to day for a specific time of a day. Historical data can provide estimates of average occupancy per auto. Real-time information about bus passenger occupancies can be obtained by smart card automated fare collection systems or auto passenger counter systems. Such systems are widely used in transit systems and can be connected to OBUs to send passenger number data.

$$\begin{aligned} W_{x,j} &= O_A \text{ if vehicle } x \text{ is Auto,} \\ W_{x,j} &= (1+\alpha) O_B \text{ if vehicle } x \text{ is Bus.} \end{aligned} \quad (2)$$

2.2.2. Modelling Ring-and-Barrier Signal Control. As shown in Figure 2, an effective model of eight-phase, dual-ring

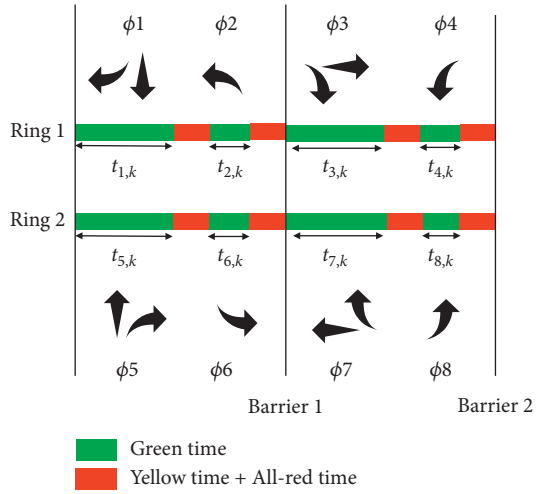


Figure 2: Standard ring-and-barrier signal logic.

signal developed by L. Head et al. [44] is adopted. The decision variable is the actual green time $t_{j,k}$ (phase j and cycle k) at the intersection.

Zeng et al. [29] assume that signal optimization is conducted per cycle. Once the optimization completes, signal parameters are implemented instantly and remain the same until the next optimization process (i.e., next cycle) begins. However, optimization is conducted per second in PACT. The impacts of elapsed time in a cycle should be considered. Two necessary revisions are summarized as follows.

(1) Minimum green constraint relaxation: Figure 3 shows the idea that the minimum green constraint $g_j^{\min}$ of phase j gradually relaxes along with time elapses. To clearly describe the phase elapsed time $\tau_{j,k}$ at a particular time, the time order is shown in the bracket. At time t_1, one second has elapsed, and the phase elapsed time $\tau_{j,k}$ should be subtracted from minimum green $g_j^{\min}$. At time t_9, since the accumulated time length at time t_9 has exceeded the minimum green, the model allows assigning zero to feasible green time variable $g_{j,k}$. The equations are listed as follows:

$$\begin{aligned} &g_{j,k} \geq g_j^{\min} - \tau_{j,k} \text{ if } \tau_{j,k} < g_j^{\min}, \\ &g_{j,k} \geq 0 \text{ if } \tau_{j,k} \geq g_j^{\min}. \end{aligned} \tag{3}$$

The if-else statement in equation (3) can be transformed into mathematical formulations in

$$\tau_{j,k} - g_j^{\min} < M \cdot \rho_{j,k}, \tag{4}$$

$$\tau_{j,k} - g_j^{\min} \geq -M \cdot \left(1 - \rho_{j,k}\right), \tag{5}$$

$$g_{j,k} \geq g_j^{\min} - \tau_{j,k} - M \cdot \rho_{j,k}, \tag{6}$$

$$g_{j,k} \geq -M \cdot \left(1 - \rho_{j,k}\right). \tag{7}$$

A binary variable, $\rho_{j,k}$, is introduced to indicate that the elapsed time is less or larger than the minimum green. When the elapsed time is less than minimum green, $\rho_{j,k}$ is constrained to be 0. Otherwise, $\rho_{j,k}$ is constrained to be 1.

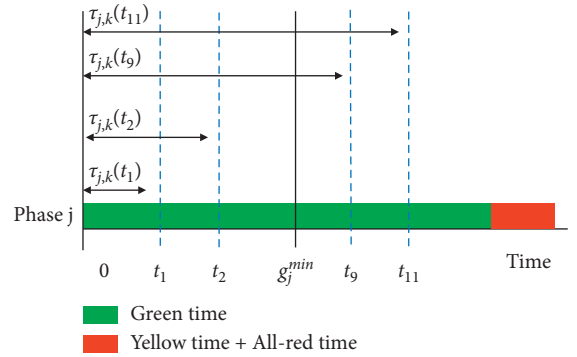


Figure 3: Minimum green time relaxation.

Therefore, the constraints of minimum green time (equations (6) and (7)) can be chosen correctly. M is the big number constraint.

(2) Equal constraint of green time: in PACT, once phase j ends, the green time of phase j ($g_{j,k}$) becomes a constant that equals elapsed time $\tau_{j,k}$ (shown in Figure 4).

For the example shown in Figure 4, at time t_1, all actual green time $t_{j,k}$ is free to be optimized. At time t_{26}, the optimization assigns $g_{1,k}$ and $g_{5,k}$ to be zero, which means terminating phases 1 and 5. At time t_{27}, since phase 1 and phase 5 end, $t_{1,k}$ and $t_{5,k}$ are no longer adjustable and should be circumscribed to equal phase elapsed time (i.e., 26 seconds). Moreover, $t_{j,k}$ cannot be larger than maxim green time $g_j^{\max}$. These conditions are mathematically expressed as follows:

$$\begin{aligned} &t_{j,k} = g_{j,k} \cdot \left(1 - P_{j,k}\right) + \tau_{j,k} \cdot P_{j,k}. \\ &g_j^{\max} \geq t_{j,k}. \end{aligned} \tag{8}$$

$P_{j,k}$ represents the activation status of phase j. If phase j ends, $P_{j,k}$ equals 1. Otherwise, $P_{j,k}$ equals 0. If the phase j ends, $g_{j,k}$ becomes a constant and is replaced with the elapsed time $\tau_{j,k}$. Finally, the ring-and-barrier model can be completed through equations (9) to (14). $T_{j,k}$ denotes start

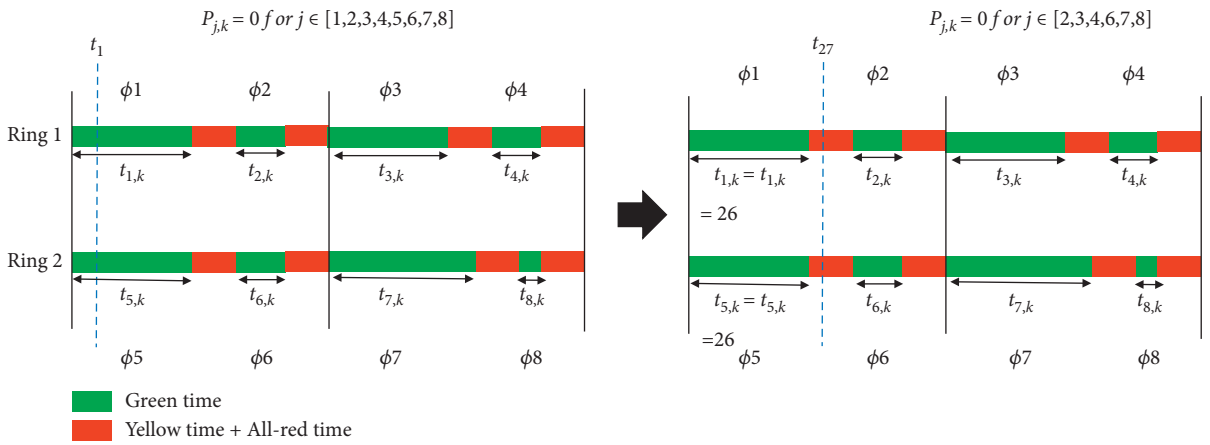


Figure 4: Example of green time equal constraint.

time of green of phase j in cycle k; yellow interval and all-red interval are Y and AR, respectively. The duration of cycle k is C_k. Note that PACT allows cycle length to be variable within a user-defined range, and the maximum and minimum cycle length are denoted as $C_{\min}$ and $C_{\max}$, respectively.

$$T_{1,k} = T_{5,k} = 0, \tag{9}$$

$$T_{3,k} = T_{7,k}, \tag{10}$$

$$T_{j,k} = T_{j-1,k} + t_{j-1,k} + Y + AR + \tau_{j-1,k} \cdot \left(1 - P_{j-1,k}\right) \cdot \forall j \in [2,3,4,6,7,8], \tag{11}$$

$$T_{4,k} + t_{4,k} + Y + AR + \tau_{4,k} \cdot \left(1 - P_{4,k}\right) - T_{1,k} = C_k, \tag{12}$$

$$T_{8,k} + t_{8,k} + Y + AR + \tau_{8,k} \cdot \left(1 - P_{8,k}\right) - T_{5,k} = C_k, \tag{13}$$

$$C_{\max} \geq C_k \geq C_{\min}. \tag{14}$$

2.2.3. Modelling Vehicle Delays. Since Zeng et al. [29] optimized signal settings per cycle to take into consideration future traffic flows, the planning horizon of Zeng's model is two-cycle lengths, and the communication range of CV technology is set as 2 km. In this study, one cycle is adopted as PACT's planning horizon. Delays that are unable to be captured in the planning horizon are estimated with the background signal plan.

(1) Queueing vehicles: equation (7) shows the delay calculation on queueing vehicles, as illustrated in Figure 5. For vehicle x in the queue, its delay depends on the saturation time of the vehicles in front of it. The number of lanes of phase j is N_j and start time of phase j in cycle k at the intersection is $T_{j,k}$. Delays of queueing vehicles $d^Q_{x,j}$ can be derived through (15) with a saturation flow rate s.

$$d^Q_{x,j} \geq T_{j,k} + \frac{\lfloor (x-1)/N_j \rfloor}{S}. \tag{15}$$

To avoid oversaturation, the green time of each phase should not be less than the dissipating time of queueing vehicle $I^Q_j/N_j/s$. The situation is formulated as follows:

$$g_{j,k} \geq \left\lfloor \frac{I^Q_j}{N_j} \right\rfloor / s - M \cdot P_{j,k}. \tag{16}$$

Note that the effectiveness of this constraint depends on the status of the phase. The constraint becomes ineffective if the phase j ends ($P_{j,k} = 1$).

(2) Approaching vehicles: the arrival time of each vehicle at the stop bar equals the free-flow travel time to the stop bar $T^P_{x,j}$. Delay of approaching vehicles can be separated into two categories:

Category 1. Arrival before the end of green time in the current cycle (cycle $k = 0$).

Category 2. Arrival after the end of green time in the current cycle (cycle $k = 0$).

$$T^P_{x,j} \leq T_{j,k} + \tau_{j,k} \cdot \left(1 - P_{j,k}\right) + t_{j,k} - y^k_{x,j} \cdot M, \tag{17}$$

$$T^P_{x,j} \leq T_{j,k} + \tau_{j,k} \cdot \left(1 - P_{j,k}\right) + t_{j,k} + \left(1 - y^k_{x,j}\right) \cdot M, \tag{18}$$

$$\theta^k_{x,j} \leq y^k_{x,j}. \tag{19}$$

The binary variable $y^k_{x,j}$ is used to identify the category, and the binary variable $\theta^k_{x,j}$ is introduced to identify whether

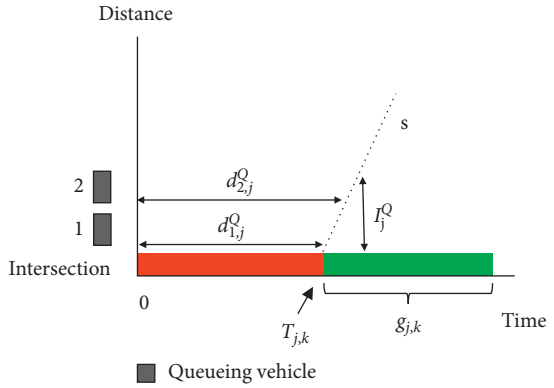


Figure 5: Delay of queueing vehicle.

a vehicle can leave the intersection in cycle k or not. Equation (19) indicates that the value of $\theta^k_{x,j}$ is constrained by the category of vehicle x. $y^k_{x,j}$ equals 0 means that vehicle x arrives after the end of green time in cycle k. Hence, $\theta^k_{x,j}$ must be 0.

According to different queueing statuses, Category 1 includes three types (see Figure 6):

Type I. Vehicles that arrive at the back of the queue before green starts.

Type II. Vehicles that arrive at the back of the queue after green starts and the queue has not been cleared yet.

Type III. Vehicles that arrive at the stop bar after green starts without any queue remaining.

$T^A_{x,j}$ denotes actual arrival time at the stop bar of vehicle x for phase j. The arrival time includes waiting time in queue and travel time from queueing position to stop bar. The number of vehicles in the queue upon the x th vehicle's arrival is denoted as $q_{x,j}$. Given the average length of vehicles L_s and free-flow speed v_f, the time of x th vehicle that arrives at the back of the queue $T^r_{x,j}$ can be calculated by subtracting the average queue clearance time from $T^P_{x,j}$ (as shown in (20)). For type III, there is no queue when a vehicle arrives at the stop bar; thus, $T^r_{x,j}$ equals $T^P_{x,j}$.

$$T^r_{x,j} = T^P_{x,j} - \frac{q_{x,j} \cdot L_s}{v_f}. \tag{20}$$

The binary variable $\sigma_{x,j}$ is used to decide whether the arrivals of vehicles belong to Type I or II ($\sigma_{x,j} = 1$ for Type I and $\sigma_{x,j} = 0$ for Type II). The equations are listed as follows:

$$\begin{aligned} T_{x,j} &\le T_{j,k} + \left(1 - \sigma_{x,j}\right) \cdot M, \\ T^r_{x,j} &> T_{j,k} - \sigma_{x,j} \cdot M. \end{aligned} \tag{21}$$

The queue length upon vehicle x arriving behind the queue ($q_{x,j}$) can be estimated through equations (22)–(24), which refer to Types I, II, and III, respectively:

$$q_{x,j} \ge \lfloor \frac{x-1}{N_j} \rfloor - \left(1 - \sigma_{x,j}\right) \cdot M, \tag{22}$$

$$q_{x,j} \ge \lfloor \frac{x-1}{N_j} \rfloor - s \cdot \left(T^r_{x,j} - T_{j,k}\right) - \sigma_{x,j} \cdot M, \tag{23}$$

$$q_{x,j} \ge 0. \tag{24}$$

Equations (25) to (28) refer to the calculation of time to clear the queue. If the vehicle fails to leave the intersection in cycle k (i.e., saturation time of the queue is larger than green time), the vehicle needs to wait for the start of green in the next cycle.

$$\frac{q_{x,j}}{s} + \frac{q_{x,j} L_s}{v_f} \le t_{j,k} + \tau_{j,k} \cdot \left(1 - P_{j,k}\right) + \left(1 - \sigma_{x,j}\right) \cdot M + \left(1 - \theta^k_{x,j}\right) \cdot M, \tag{25}$$

$$\frac{q_{x,j}}{s} + \frac{q_{x,j} L_s}{v_f} > t_{j,k} + \tau_{j,k} \cdot \left(1 - P_{j,k}\right) - \left(1 - \sigma_{x,j}\right) \cdot M - \theta^k_{x,j} \cdot M, \tag{26}$$

$$\frac{q_{x,j}}{s} + T^r_{x,j} + \frac{\left(q_{x,j} + I^Q_j\right) L_s}{v_f} \le T_{j,k} + t_{j,k} + \tau_{j,k} \cdot \left(1 - P_{j,k}\right) + \sigma_{x,j} \cdot M + \left(1 - \theta^k_{x,j}\right) \cdot M, \tag{27}$$

$$\frac{q_{x,j}}{s} + T^r_{x,j} + \frac{\left(q_{x,j} + I^Q_j\right) L_s}{v_f} > T_{j,k} + t_{j,k} + \tau_{j,k} \cdot \left(1 - P_{j,k}\right) - \sigma_{x,j} \cdot M - \theta^k_{x,j} \cdot M. \tag{28}$$

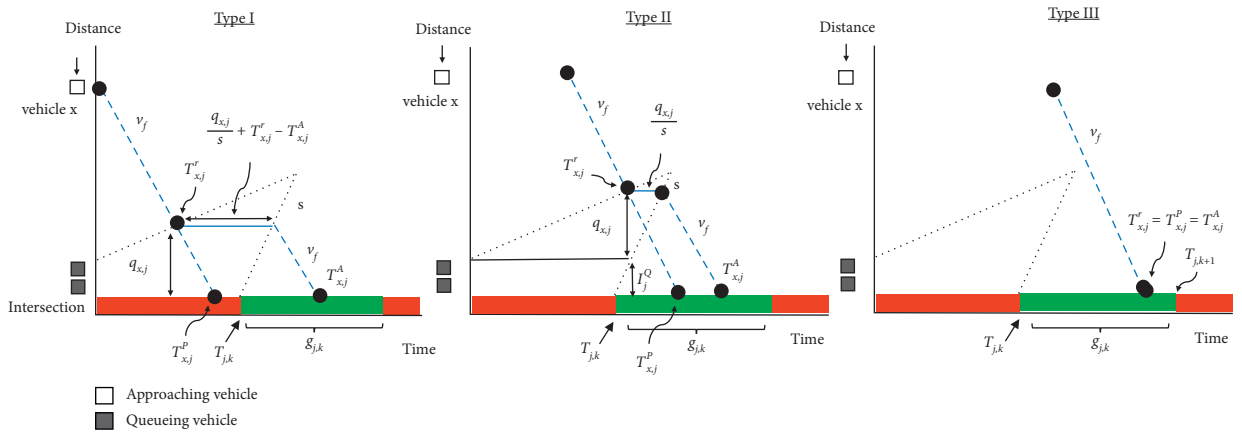


Figure 6: Illustration of three types of arrivals.

As shown in (29), if the vehicles cannot leave the intersection in the cycle k, that is, $(\theta_{x,j}^{k} = 0)$, extra waiting time $d_{x,j}^{E}$ estimated from the background signal timing plan is calculated in equation (29). $T_{x,j}^{A}$ is actual arrival time.

$$d_{x,j}^{E} = T_{j,k+1} - T_{x,j}^{A}. \tag{29}$$

According to the arrival type of each vehicle, the calculation of $T_{x,j}^{A}$ is formulated as follows:

$$T_{x,j}^{A} \le \frac{q_{x,j}}{s} + T_{j,k} + \frac{q_{x,j} \cdot L_s}{v_f} + \theta_{x,j}^{k} \cdot M + \left(1 - \sigma_{x,j}\right) \cdot M,$$

$$T_{x,j}^{A} \le T_{x,j}^{r} + \frac{q_{x,j}}{s} + \frac{\left(q_{x,j} + I_j^{Q}\right) \cdot L_s}{v_f} + \theta_{x,j}^{k} \cdot M + \sigma_{x,j} \cdot M. \tag{30}$$

Finally, delays for an approaching vehicle of each scenario are formulated as

$$d_{x,j} \ge \frac{q_{x,j}}{s} + T_{j,k} - T_{x,j}^{r} - \left(1 - \sigma_{x,j}\right) \cdot M - \left(1 - \theta_{x,j}^{k}\right) \cdot M, \tag{31}$$

$$d_{x,j} \ge \frac{q_{x,j}}{s} + T_{j,k} - T_{x,j}^{r} + d_{x,j}^{E} - \left(1 - \sigma_{x,j}\right) \cdot M - \theta_{x,j}^{k} \cdot M, \tag{32}$$

$$d_{x,j} \ge \frac{q_{x,j}}{s} - \sigma_{x,j} \cdot M - \left(1 - \theta_{x,j}^{k}\right) \cdot M, \tag{33}$$

$$d_{x,j} \ge \frac{q_{x,j}}{s} + d_{x,j}^{E} - \sigma_{x,j} \cdot M - \theta_{x,j}^{k} \cdot M. \tag{34}$$

2.2.4. Driving Speed Advisory Computation. The flowchart of calculating speed advisory is shown in Figure 7. At first, OBU requests SPaT information of the intersection. Next, OBU computes passing probability based on SPaT. A threshold of passing probability is used to determine the activation of speed advisory calculation. If the passing probability is larger than the threshold, there is no need to provide speed advisory since current speed can almost ensure that transit vehicles pass intersections without delays (Case 3). If the passing probability is less than the threshold, computation of speed advisory will be conducted. The OBU may suggest an advisory speed (Case 1) or advise maintaining the vehicles' current speed (Case 2).

(1) Computation of passing probability at the intersection: in this study, transit vehicles operate under mixed traffic circumstances without exclusive lanes. Results of bus travel time prediction may be biased if a simple calculation (i.e., distance to intersection divided by vehicle's current speed) is adopted. To consider the uncertain nature of arrival time, we assume that the arrival time of bus b at the intersection is a random variable T_b^B, which follows a normal distribution with a standard error e.

The detail of the passing probability calculation is shown in Figure 8. At time t_0, there is a bus b with speed v_b riding towards the intersection. The expected arrival time $E(T_b^B)$ is calculated from the distance to stop bar (L) divided by bus current speed (v_b). The standard error (e) is a function related to the bus distance to the stop bar. Therefore, we can calculate the passing probability by the cumulative density function of the normal distribution. The integration range is the intersection of green duration $(g_{j,k})$ and arrival time (T_b^B). Since bus expected arrival time may cross several cycles, preoptimized signal plans are applied on cycles that are not involved in the planning horizon of ATSC.

The detail of calculating the probability of passing the intersection is listed in Algorithm 1. The input parameters include bus order b, bus speed v_b, signal settings $(T_{j,k})$ and $(g_{j,k})$, and distance to the intersection. The order, speed, and distance to the intersection are retrieved via CV communication. The result of passing probability, PassProb, is initialized as 0 at the beginning of the algorithm. Lines (2) to (3) calculate the mean and standard error. The standard error is a function related to distance to intersection L, and the equation is shown as follows:

$$e = \frac{1}{3} \cdot \left(2 + \frac{L}{50}\right). \tag{35}$$

Through lines (4) to (5), three times standard error extended from mean is used to approximate the earliest

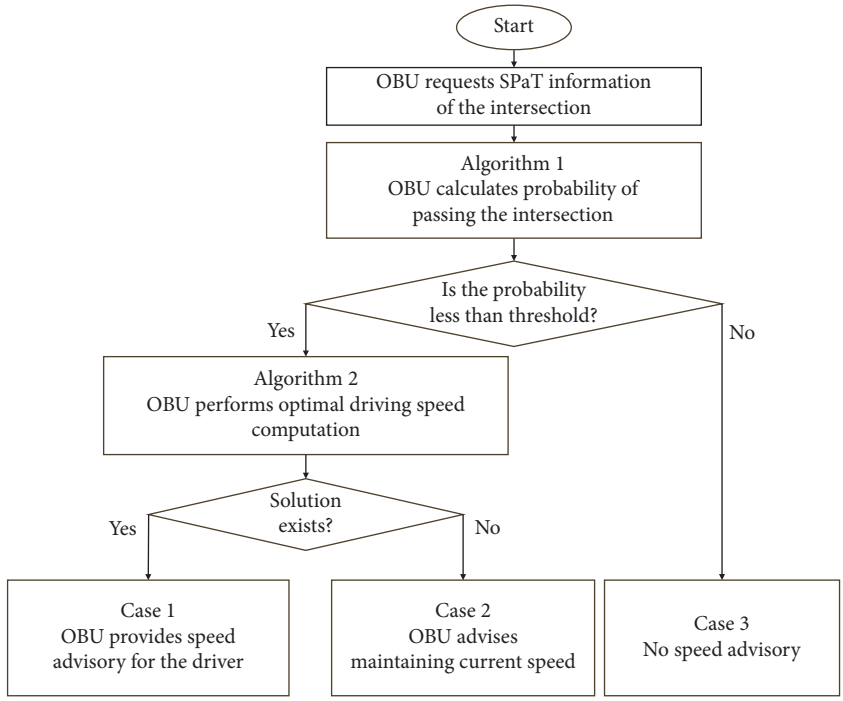


Figure 7: The flowchart of optimal driving speed calculation.

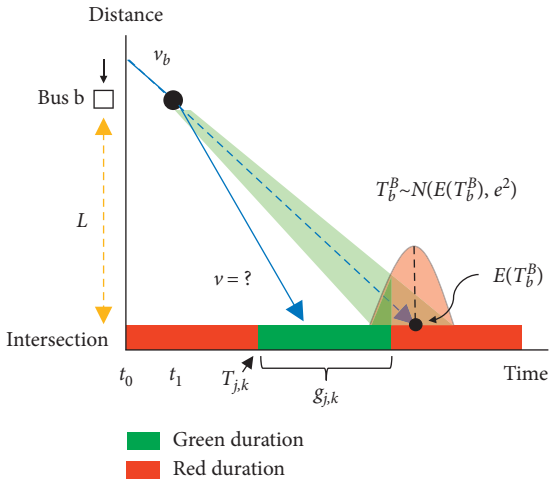


Figure 8: Illustration of passing probability calculation.

arrival time EarliestArrival and latest arrival time LatestArrival. Line (6) shows the intersection of green duration and arrival time (i.e., F to H), which is computed by intersecting the arrival time range and green duration. Finally, we can obtain the passing probability by integrating F to H with the cumulative density function of normal distribution.

(2) Finding an optimal driving speed: the objective of speed advisory is to find a speed v that can maximize the probability of passing intersection. Due to the limited space of the feasible solutions, it is possible to get a global optimal solution by the enumeration method. Algorithm 2 shows the process of finding optimal driving speed.

Line (1) is the initialization of variables. Line (2) indicates the algorithm start enumerating a speed v from 0.7 v_{lim} to v_{lim}. v_{lim} is speed limit of the road. In line (3), the algorithm calculates the probability of passing the intersection; the result is PassProb. In lines (4) to (6), a comparison between PassProb and MaxPassProb is performed to choose a speed v that can maximize passing probability. Lines (7) to (9) are a stabilization mechanism designed to prevent OBU from frequently providing different speed advisory. PreStepPassProb is the passing probability of the previous time step. An optimal speed is updated only if the difference between MaxPassProb and PreStepPassProb is larger than β. For example, at time step t, the optimal speed of bus b is 13 m/s, and the bus has an 85% probability of passing the intersections. However, at time step $t-1$, the speed of bus b is 11 m/s, and the probability of passing the intersection is 82%. Here, we assume β as 5%. Although the optimal speed at time step t has a higher passing probability (85% > 82%), the difference is less than β (85%–82% = 3% < 5%). Thus, the optimal speed at time t is 11 m/s, not 13 m/s. If the difference between MaxPassProb and PreStepPassProb is less than β, the algorithm advises maintaining current speed v_b (i.e., Case 2).

3. Experimental Evaluation

Traffic simulation is adopted to evaluate the performance of the PACT system. The simulation platform and process are introduced, and then the intersection layout and related traffic settings are depicted.

3.1. Simulation Settings. The simulation tool, Simulation of Urban Mobility (SUMO), is selected as the simulation

platform [45]. A well-known optimizer GUROBI [46] is employed for optimizing the signal control model. As shown in Figure 9, the Python programs can control and retrieve results from SUMO via Traffic Control Interface (TraCI).

For signal parameters, $C_{\min}$ is set as 50 seconds and $C_{\max}$ is $(1 + \gamma) \cdot$ Original Cycle Length. The parameter γ stands for an adjustable range of cycle length, a number in [0, 1]. For example, the cycle length is 100 seconds, γ is 0.2, and PACT allows generating an optimal signal plan that has a maximum cycle $C_{\max} = 100 \cdot (1 + 0.2) = 120$. A fixed-time signal plan optimized by PASSER-V [47] serves as the background plan. For OBUs, the threshold of passing probability is set at 50%. Parameter β is set as 5% to avoid OBUs providing unstable speed advisory. The occupancy of each auto is set as 1.5 passengers, and bus is 30 passengers. The speed limit of each approach is 60 km/hr. V_f is set as 15 m/s (i.e., about 54 km/hr) for both autos and buses.

A real four-leg with a left-turn bay intersection, Xiao-Dong Road and Chunghwa Road, in Tainan, Taiwan, was established in a simulation environment. As shown in Figure 10. The length of each leg is 500 m. The detection range of VTD and CV communication ranges are set as 300 m. Seven bus lines pass through this intersection: Bus 20, Bus 15, Bus 2, Bus 19, Bus G17, Bus O12, and Bus H62. The headway of each bus line is 30, 20, 15, 20, 30, 60, and 25 minutes, respectively.

Detailed signal settings and historical traffic volume are shown in Figure 11. Historical traffic volumes are surveyed in morning peak hours (8:00 to 9:00 am), and the corresponding V/C ratio is 0.9. The preoptimized signal plan is adopted as the baseline in the following experiments.

Simulation environment parameters are listed in Table 1. The saturation flow rate is 0.5 veh/sec, and the maximum acceleration and deceleration rate is 4 m/s^2 and 5 m/s^2, respectively. The safety spacing between two vehicles is 2 m. The average vehicle length L_S is set as 4 m.

3.2. Results Evaluation. Three types of evaluation modes are evaluated: PASSER-V optimization plan (PASSER-V), PACT without Optimal Speed advisory (PACT-No OS), and PACT with Optimal Speed advisory (PACT-OS). The simulation time of each experiment scenario is 2 hours of 5 runs with 8 random seeds, excluding 5-minute warm-up time. The CPU is Intel i5-9600K, and the average computation time in the most complex case (i.e., $\gamma = 0.9$ and V/C = 0.9 case) is 0.15 seconds. All experimental results are statistically significant at a p-value < 0.05.

3.2.1. Various V/C Ratios (without Buses). This experiment aims to evaluate the performance of PACT under four levels of V/C ratios where transit vehicles are not included. As listed in Table 2, four V/C ratios are tested: 0.3, 0.5, 0.7, and 0.9. The results show that V/C = 0.9 case has the greatest significant delay reduction (i.e., −29.69% in-vehicle-based and −29.67% person-based) compared to other V/C ratio cases. As the V/C ratio becomes lower, the amount of delay reduction decreases as well. There is only about a 2.7% delay reduction in V/C = 0.3 case. The minimum cycle length setting causes the result of insignificant delay reduction in lower V/C ratio cases. The minimum cycle length is 50 s in simulation settings, limiting the improvement of bus delay in a lower V/C ratio.

3.2.2. Various V/C Ratios (with Buses). Figures 12 and 13, respectively, illustrate the person-based delay and vehicle-based delay changes among various control types. In Figure 12, bus passenger delay reduces by 68% in PACT-No OS and 81% in PACT-OS compared with the PASSER-V plan in 0.9 V/C ratio case. It is evident that PACT-OS outperforms PACT-No OS in reducing bus passenger delays. The additional benefit is 17% (V/C = 0.5 case) and 13% (V/C = 0.7 case). For light traffic volume (V/C ratio <0.5), the amount of delay reduction is minor since traffic variations are lighter. A significant reduction in auto passenger delay is presented in high traffic volume (V/C ratio >0.5). Auto passenger delay reduces by 10%–15% in 0.7 V/C ratio case and 25%–30% in 0.9 V/C ratio case, showing that PACT can favour transit vehicles without dramatically interfering with non-prioritized traffic. For vehicle-based results (Figure 13), bus delay is 23.12 s, and auto delay is 23.72 s in V/C = 0.9 case. PACT-No OS decreases bus delay to 7.47 s, and PACT-OS further decreases to 4.58 s. For auto delay, PACT-No OS is 17.97 s, and PACT-OS is 17.66 s. Since speed advisory is only applied on transit vehicles, the results of PACT-No OS are nearly the same as PACT-OS. The results show that PACT can reduce auto delays even if they are not equipped with OBUs (OBUs are only equipped on transit vehicles).

3.2.3. Various Signal Adjustable Parameters γ. In comparison with the PASSER-V plan, both PACT-No OS and PACT-OS decrease bus passenger delays for all γ cases. As shown in Figure 14, for PACT-OS, bus passenger delay is 0.58 s in $\gamma = 0.1$ case and 0.19 s in $\gamma = 0.9$ case; total passenger delay is 12.56 s in $\gamma = 0.1$ case and 10.04 s in $\gamma = 0.9$ case. PACT generates high benefits in delay reduction with large γ in either bus or system perspective. However, the benefit difference of bus passenger delays between PACT-No OS and PACT-OS becomes smaller when γ increases. The difference is 10% in $\gamma = 0.9$ case but increases to 21% in $\gamma = 0.1$ case. The difference manifests that speed advisory can considerably improve delays when a strong limitation on the signal adjustable range (γ) was applied. There is an interesting finding in the results of bus passenger delays. The amount of delay reduction dramatically increases as γ gets higher, but the pattern remains near the same when γ is above 0.5. The maximum delay improvement of PACT occurs when $\gamma = 0.5$, which is −67% in bus passenger delays and −24% in auto. The result is coincidental with the traditional traffic view of the relationship between cycle length and intersection capacity: as cycle length increases, the capacity initially increases but eventually remains constant. Due to the limitation of intersection capacity, the maximum improvement of the PACT system occurs when $\gamma = 0.5$.

Input: bus order (b), bus speed (v_b), distance to intersection (L), $T_{i,j}$, and $g_{j,k}$.
Output: probability of passing intersection (*passProb*).
(1) Initialization: set $passProb = 0$.
(2) $E(T_b^B) = L/v_b$.
(3) $e = f(L)$.
(4) Latest_Arrival = $E(T_b^B) + 3e$.
(5) Earliest_Arrival = $E(T_b^B) - 3e$.
(6) $[F, H]$ = [Earliest_Arrival, Latest_Arrival] $\cap$ $[T_{j,k}, T_{j,k} + g_{j,k}]$.
(7) passProb = $\int_p^H P(F \le T_b^B \le H) dT_{i,b}^B = \Phi(F + H - E(T_b^B)/e) - \Phi(F - E(T_b^B)/e)$.
(8) **return** *passProb*.

Algorithm 1: Calculation of passing probability at an intersection.

Input: bus order (b), bus speed (v_b), distance to intersection (L), $T_{i,j}$, and $g_{j,k}$ passing probability of pervious time step (*PreStepPassProb*)
Output: speed advisory (OptimalSpeed or v_b)
(1) Initialization: $passProb = 0$ $MaxPassProb = 0$ $PreStepPassProb = 0$
(2) **For** $v \in [0.7v_{\text{lim}}, v_{\text{lim}}]$**do**:
(3) $PassProb$ = **Algorithm 1** (input: b, v, L, $T_{j,k}$, $g_{j,k}$)
(4) **If** PassProb > MaxPassProb **then**:
(5) $Max_P assProb$ = PassProb
(6) OptimalSpeed = v
(7) **If** (MaxPassProb-PreStepPassProb) > β **then:**
(8) PreStepPassProb = Max_PassProb
(9) **Return** OptimalSpeed
(10) **Else:**
(11) **Return** v_b

Algorithm 2: Finding an optimal speed.

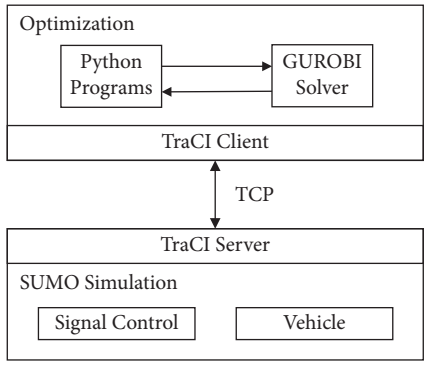


Figure 9: Integration of signal optimization and traffic simulation by TraCI.

3.2.4. Sensitivity Test of Bus Occupancy. The results of the bus passenger sensitivity test are shown in Figure 15 (person-based delay changes) and Figure 16 (vehicle-based delay changes). Based on various passenger occupancy, the experiment suggests the extent of preferential treatment that PACT can generate for buses. In Figure 15, it is evident that PACT provides favourable bus delays in all passenger occupancy cases, even if passenger occupancy is low. In the 2-passenger case, PACT-No OS decreases bus passenger delays by 3.69 s, and PACT-OS decreases bus passenger delays by 5.42 s. As the bus passenger occupancy gets higher, the delay benefits get higher as well. However, if bus passenger occupancy is greater than 20, the improvement on bus delays gets smaller. The benefits maintain about 65% and 77% reduction in PACT-No OS and PACT-OS cases, respectively. In Figure 16, since PASSER-V optimized signal plans on a vehicle basis, the delay results are unaffected by bus passenger occupancy. In the 2-passenger case, the average bus delay under the PASSER-V plan is 23.72 s. PACT-No OS decreases bus delay to 15.75 s, and PACT-OS further decreases bus delay to 12.27 s.

3.2.5. Traffic Signal Preemption. Priority level factor α is used to increase the priority level of a vehicle. This experiment aims to demonstrate the maximum preferential treatment that PACT can provide for vehicles with extreme large α. In Figure 17, the results manifest that PACT can provide above 80% benefits on vehicle delays as long as $\alpha \ge 100$ and $\gamma \ge 0.5$. In $\alpha = 1000$ and $\gamma = 0.7$ scenario, the amount adds up to 98% when optimal speed guidance is implemented. Since traffic signals provide near unconditional "signal preemption" for special vehicles, the advantage of speed guidance becomes minor when $\alpha \ge 100$. The benefit of delays is near 95% to 97% in both PACT-No OS and PACT-OS. Even signal adjustable range (γ) is small, PACT can reduce considerable delays in special vehicles as long as α is large enough. For $\gamma = 0.3$ case, the benefit of the delay is

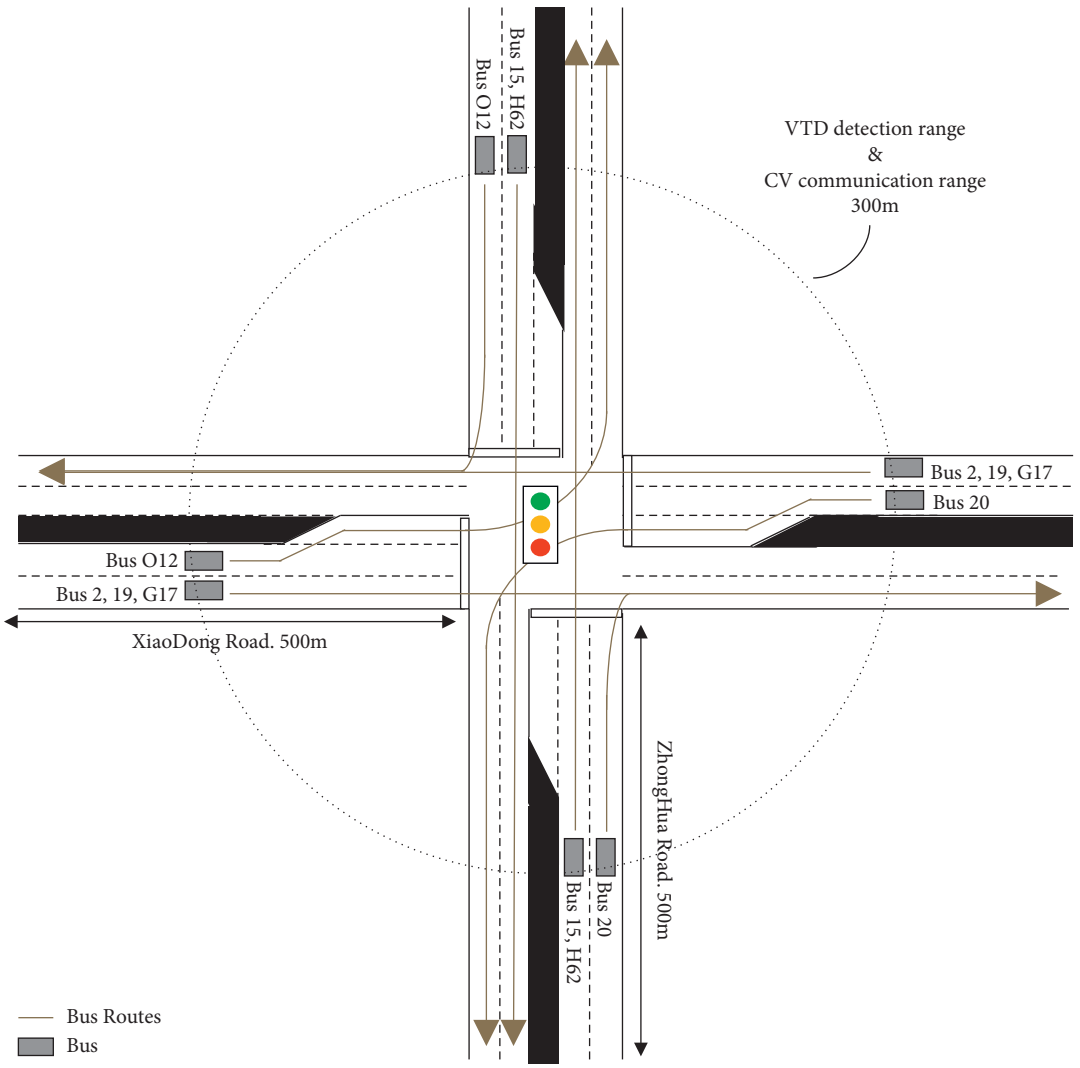


Figure 10: Layout and bus routes for the intersection of XiaoDong and ZhongHua Road.

Signal Parameters and Traffic Volume

Phase order		$\phi 1$	$\phi 2$	$\phi 3$	$\phi 4$
Number of lanes		2	1	2	1
Volume (veh/hr)		617	149	738	235
Signal Parameters (sec)	Green	24	8	20	13
	Yellow	3			
	All-red	2			
	Cycle	85			

Phase order		$\phi 5$	$\phi 6$	$\phi 7$	$\phi 8$
Number of Lanes		2	1	2	1
Volume (veh/hr)		872	116	685	251
Signal Parameters (sec)	Green	24	8	20	13
	Yellow	3			
	All-red	2			
	Cycle	85			

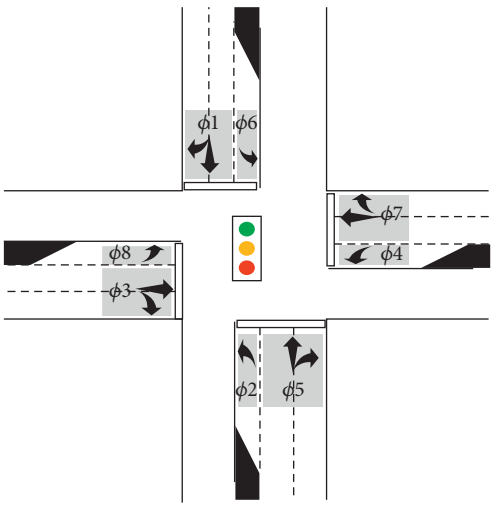


Figure 11: Detailed signal parameters and traffic volume settings.

Table 1: Details of simulation parameters.

Parameters	Value
Saturation flow rate (s)	0.5 veh/s
Maximum acceleration rate	4 m/s^2
Maximum deceleration rate	5 m/s^2
Safety spacing	2 m
Vehicle length (L_S)	4 m

Table 2: Person-based and vehicle-based delay results without buses under various V/C ratios.

V/C	Type	Measures	PASSER-V	PACT-No OS	Percentage changes
0.9	Auto	Vehicle delay (D-AV)	23.61 (23.44–23.78)	16.6 (16.48–16.72)	−29.69%
		Person delay (D-AP)	15.74 (15.62–15.86)	11.07 (10.99–11.15)	−29.67%
0.7	Auto	Vehicle delay (D-AV)	17.3 (17.24–17.36)	14.13 (14.04–14.22)	−18.32%
		Person delay (D-AP)	11.53 (11.49–11.57)	9.42 (9.36–9.48)	−18.30%
0.5	Auto	Vehicle delay (D-AV)	14.37 (14.22–14.52)	13.28 (13.22–13.34)	−7.59%
		Person delay (D-AP)	9.58 (9.48–9.68)	8.85 (8.81–8.89)	−8.14%
0.3	Auto	Vehicle delay (D-AV)	13.16 (13.07–13.25)	12.81 (12.72–12.9)	−2.66%
		Person delay (D-AP)	8.78 (8.72–8.84)	8.53 (8.47–8.59)	−2.85%

Note. The numbers in the bracket show 95% confidence intervals.

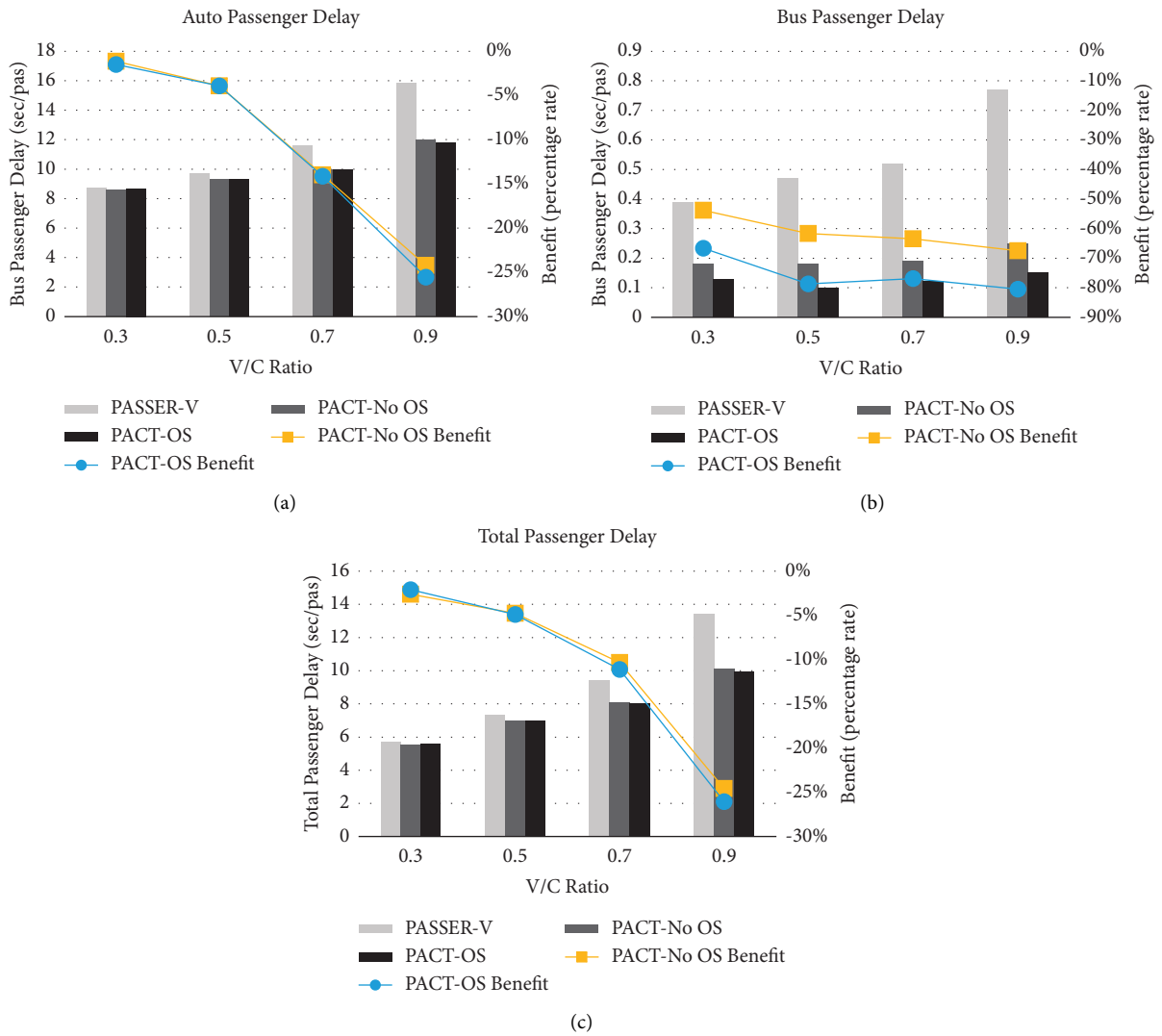


Figure 12: Delay changes in (a) auto passenger, (b) bus passenger, and (c) total passenger under various V/C ratios.

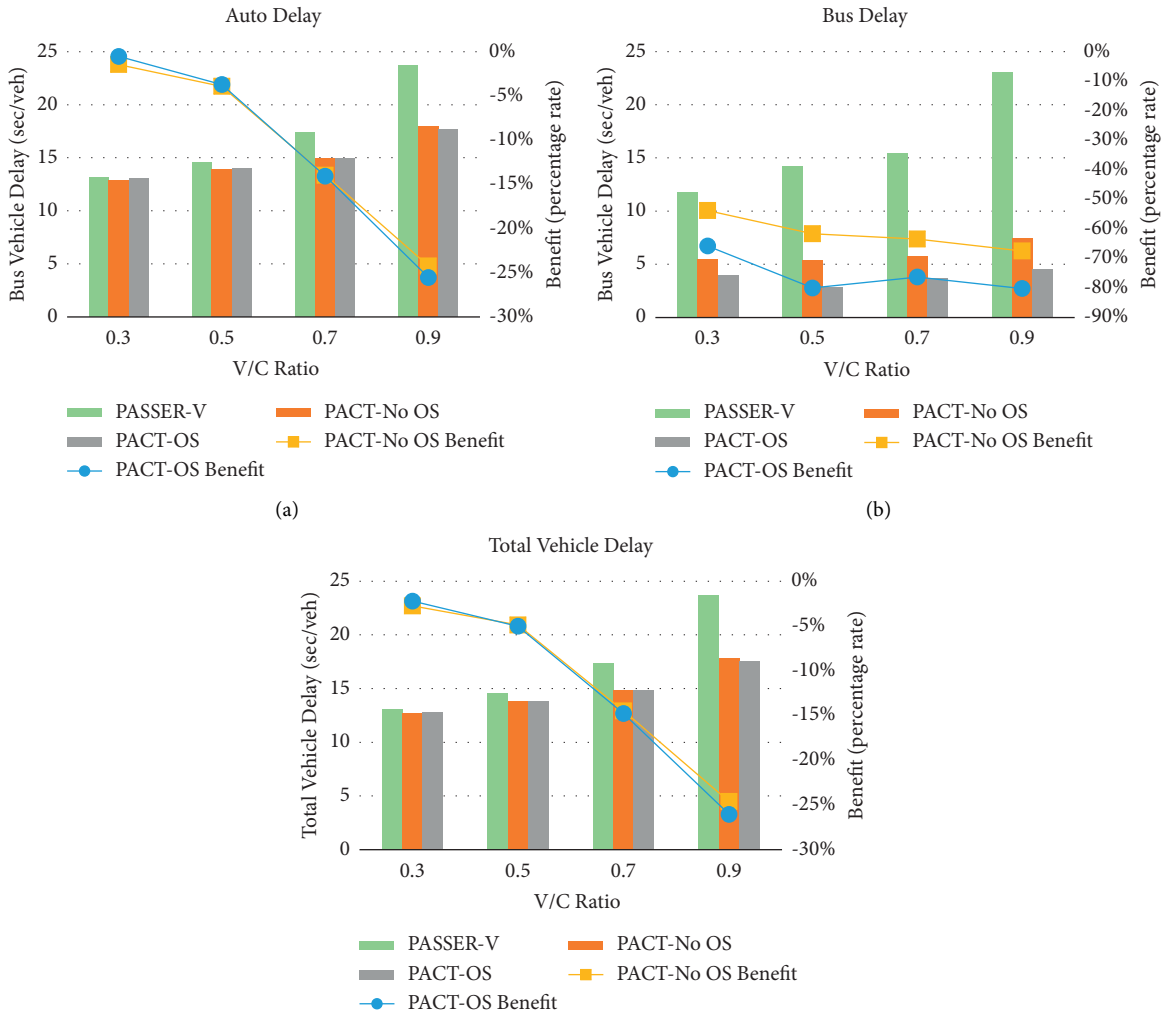


Figure 13: Delay changes in (a) auto, (b) bus, and (c) total vehicle under various V/C ratios.

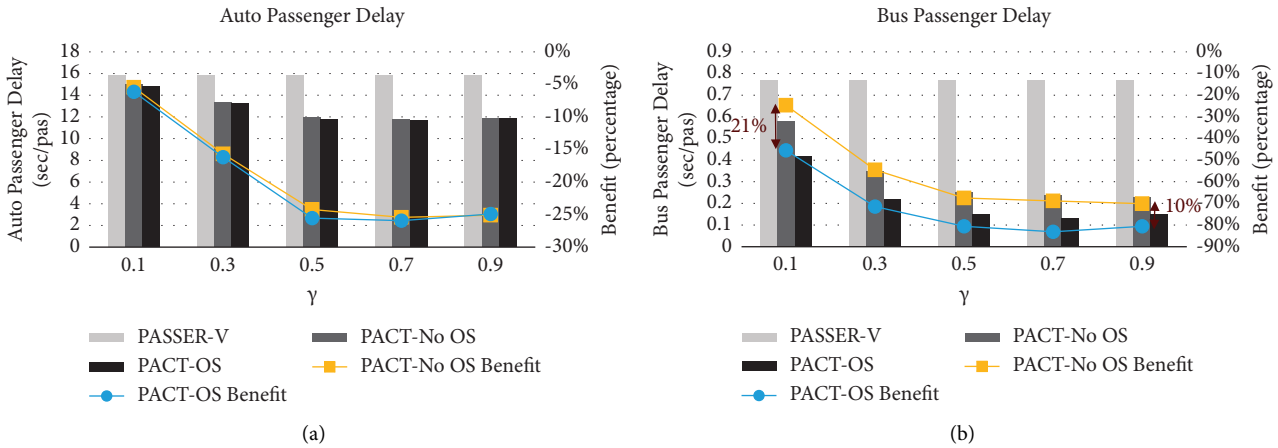
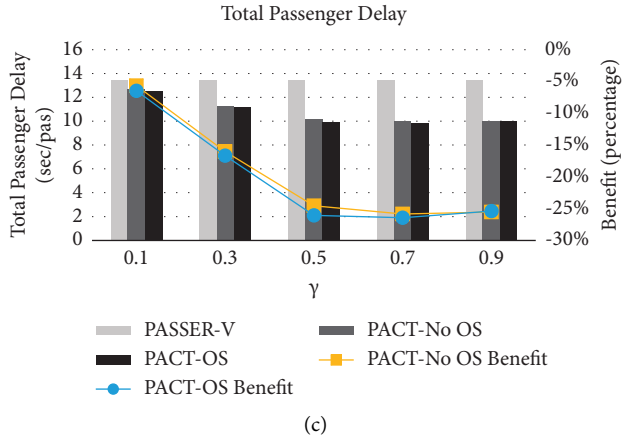


Figure 14: Continued.



(c)

Figure 14: Delay changes in (a) auto passenger, (b) bus passenger, and (c) total passenger under various signal adjustable parameters.

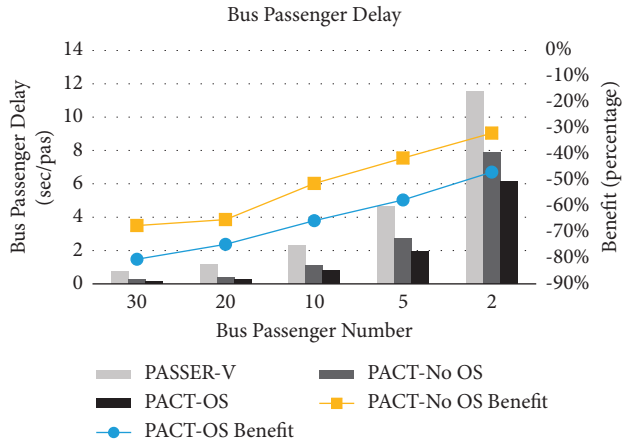


Figure 15: Delay changes in bus passenger under various bus occupancy.

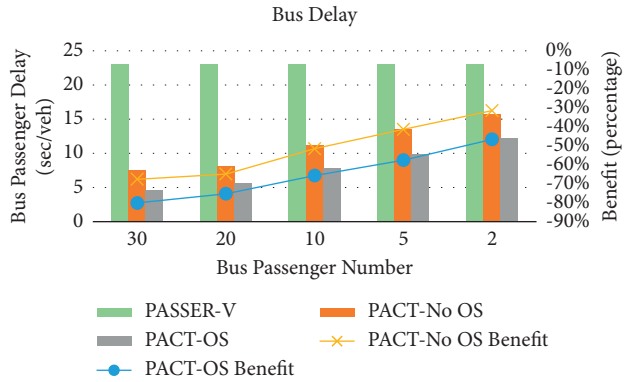


Figure 16: Delay changes in bus under various bus occupancy.

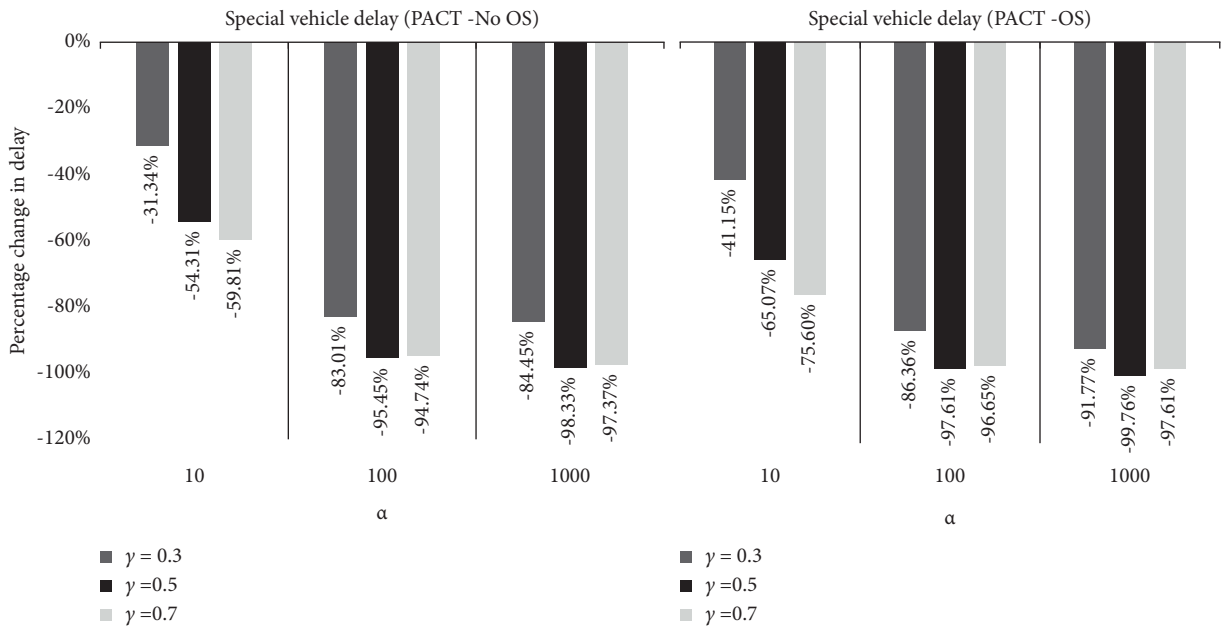


Figure 17: Percentage changes in delays compared to various α and γ.

83% in PACT-No OS and 86% in PACT-OS when $\alpha = 100$. In summary, one can make PACT provide near signal preemption by enlarging weight factor α and enabling speed guidance.

4. Conclusion

This study proposed a person-based cooperative adaptive traffic signal control system named PACT. PACT operates in a rolling horizon procedure, which includes two parts: second-basis adaptive traffic signal control (performed by RSU) and second-basis optimal driving speed guidance (performed by OBU). The signal control model optimizes signal parameters to minimize total person delays. Once optimization completes, the SPaT information is sent to OBUs to calculate optimal speed advisory. Optimal speed advisory algorithms provide an advisory speed that can maximize intersection passing probability. The algorithms consider the stochastic nature of bus arrival time with the normal distribution assumption. After transit vehicles apply recommended speed, RSUs optimize signal parameters based on the latest traffic status, and then the process repeats. PACT's signal-vehicle cooperate-control structure can enhance the travel efficiency of both transit vehicles and autos.

The experiment results show that PACT has a strong ability to handle real-time traffic variation, reducing total person delays up to about 28% and bus person delays to 60%. Moreover, PACT-OS can generate additional benefits up to 20% compared to PACT-No OS, which implies that optimal speed advisory can further strengthen the benefits of delay reduction. In the sensitivity test of bus occupancy, each bus passenger experiences about a 65% reduction in delay when average bus occupancy is above 20. Furthermore, the maximum benefit of delay reduction of high priority level (i.e., high α value) can be 99%, showing that PACT has high adaptability to manage various priority level demands and can be further extended to emergency vehicle signal preemption.

A promising research direction is to extend the signal control model of PACT to multiple intersections. The effect of arterial traffic flow (e.g., platoon dispersion) must be considered in the multi-intersections model. Future studies can design an experiment to explore the value of α on maintaining the stability of bus operation, such as bus headway and bus bunching problems. This study assumes that bus drivers fully comply with the speed guidance, which is hard to implement in the real world. More experiments on the degree of compliance with OBU's guidance should be conducted. In the near future, Connected Vehicle technology will advance to virtual-signal (or nonsignal) intersection control. The authority of driving through intersections depends on the direct cooperation between vehicles and virtual signals via V2I communication. The cooperative concept proposed in PACT shows a possible scheme of interaction between vehicles and traffic signals. Future research can apply this concept to virtual-signal control.

Data Availability

The data are generated by a simulation program, and part of the program files included in this study are available by sending a request to the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported in part by the Ministry of Science and Technology under contract "MOST 110-2124-M-006-002" and "Center for Innovative FinTech Business Models" of National Cheng Kung University (NCKU), Taiwan.

References

[1] A. G. Sims and K. W. Dobinson, "The Sydney coordinated adaptive traffic (SCAT) system philosophy and benefits," *IEEE Transactions on Vehicular Technology*, vol. 29, no. 2, pp. 130–137, 1980.

[2] P. Hunt, D. Robertson, R. Bretherton, and R. Winton, *SCOOT-a Traffic Responsive Method of Coordinating Signals*, pp. 0266–7045, TRRL Urban Networks Division, Great Britain, UK, 1981.

[3] N. H. Gartner, *OPAC: A Demand-Responsive Strategy for Traffic Signal Control*, Transportation Research Record Journal of the Transportation Research Board (no. 906), Washington, D.C, 1983.

[4] J. J. Henry, J. L. Farges, and J. Tuffal, "The PRODYN real time traffic algorithm," in *Control in Transportation Systems*, pp. 305–310, Elsevier, Amsterdam, Netherlands, 1984.

[5] P. Mirchandani and L. Head, "A real-time traffic signal control system: architecture, algorithms, and analysis," *Transportation Research Part C: Emerging Technologies*, vol. 9, no. 6, pp. 415–432, 2001.

[6] Q. Guo and L. Li, "Urban traffic signal control with connected and automated vehicles: a survey," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 313–334, 2019.

[7] K. Tiaprasert, Y. Zhang, X. B. Wang, and X. Zeng, "Queue length estimation using connected vehicle technology for adaptive signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2129–2140, 2015.

[8] M. Steadman and B. Huntsman, *Connected vehicle infrastructure: deployment and funding overview*, vol. 17-77, F. TexasA&M Transp. Inst., Texas, TX, USA, 2018, Tech. Rep. PRC.

[9] K. F. Turnbull, L. Cherrington, Z. Elgart et al., *Automated and connected vehicle (AV/CV) test bed to improve transit, bicycle, and pedestrian safety*, Tech. Rep. FHWA/TX-17/0-6875-1, Texas, Dept. Transp. Res. Technol. Implement. Office, Texas, TX, USA, 2017.

[10] X. Zeng, Y. Zhang, K. N. Balke, and K. Yin, "A real-time transit signal priority control model considering stochastic bus arrival time," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1657–1666, Aug. 2014.

[11] T. Wang, J. Cao, and A. Hussain, "Adaptive traffic signal control for large-scale scenario with cooperative group-based multi-agent reinforcement learning," *Transportation Research Part C: Emerging Technologies*, vol. 125, Article ID 103046, 2021.

[12] C. B. Rafter, B. Anvari, S. Box, and T. Cherrett, "Augmenting traffic signal control systems for urban road networks with connected vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1728–1740, 2020.

[13] Y. Feng, K. L. Head, S. Khoshmagham, and M. Zamanipour, "A real-time adaptive signal control in a connected vehicle environment," *Transportation Research Part C: Emerging Technologies*, vol. 55, pp. 460–473, 2015.

[14] X. Liang, S. I. Guler, and V. V. Gayah, "An equitable traffic signal control scheme at isolated signalized intersections using Connected Vehicle technology," *Transportation Research Part C: Emerging Technologies*, vol. 110, pp. 81–97, 2020.

[15] J. A. Volpe, *Vehicle-infrastructure Integration (VII) Initiative Benefit-Cost Analysis Version 2.3 (Draft)*, National Transportation Systems Center, FHWA, USA, 2008.

[16] H. Liang, H. Song, H. Li, and Z. Dai, "Vehicle counting system using deep learning and multi-object tracking methods," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2674, no. 4, pp. 114–128, 2020.

[17] J. I. Engel, J. Martin, and R. Barco, "A low-complexity vision-based system for real-time traffic monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1279–1288, 2017.

[18] H. R. Smith, B. Hemily, and M. Ivanovic, *Transit Signal Priority (TSP): A Planning and Implementation Handbook*, TRID, Washington, DC, USA, 2005.

[19] J. Lee, A. Shalaby, J. Greenough, M. Bowie, and S. Hung, "Advanced transit signal priority control with online microsimulation-based transit prediction model," *Transportation Research Record*, vol. 1925, no. 1, pp. 185–194, 2005.

[20] X. Zeng, Y. Zhang, J. Jiao, and K. Yin, "Route-based transit signal priority using connected vehicle technology to promote bus schedule adherence," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 1174–1184, 2020.

[21] L. T. Truong, G. Currie, M. Wallace, C. De Gruyter, and K. An, "Coordinated transit signal priority model considering stochastic bus arrival time," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1269–1277, 2019.

[22] G. Liu and T. Z. Qiu, "Trade-offs between bus and private vehicle delays at signalized intersections: case study of a multiobjective model," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2539, no. 1, pp. 72–83, 2016.

[23] E. Christofa and A. Skabardonis, "Traffic signal optimization with application of transit signal priority to an isolated intersection," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2259, no. 1, pp. 192–201, 2011.

[24] E. Christofa, I. Papamichail, and A. Skabardonis, "Person-based traffic responsive signal control optimization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1278–1289, 2013.

[25] E. Christofa, K. Ampountolas, and A. Skabardonis, "Arterial traffic signal optimization: a person-based approach," *Transportation Research Part C: Emerging Technologies*, vol. 66, pp. 27–47, 2016.

[26] W. Ma, K. L. Head, and Y. Feng, "Integrated optimization of transit priority operation at isolated intersections: a person-capacity-based approach," *Transportation Research Part C: Emerging Technologies*, vol. 40, pp. 49–62, 2014.

[27] J. Zhao and W. Ma, "Optimizing vehicle and pedestrian trade-off using signal timing in intersections with center transit lanes," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 6, Article ID 04018023, 2018.

[28] J. Zhao and X. Zhou, "Improving the operational efficiency of buses with dynamic use of exclusive bus lane at isolated intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 642–653, 2018.

[29] X. Zeng, X. Sun, Y. Zhang, and L. Quadrifoglio, "Person-based adaptive priority signal control with connected-vehicle information," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2487, no. 1, pp. 78–87, 2015.

[30] Z. Yu, V. V. Gayah, and E. Christofa, "Person-based optimization of signal timing," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2620, no. 1, pp. 31–42, 2017.

[31] J. Li, Y. Liu, H. Yang, and B. Chen, "Bus priority signal control considering delays of passengers and pedestrians of adjacent intersections," *Journal of Advanced Transportation*, vol. 2020, Article ID 3935795, 12 pages, 2020.

[32] K. Katsaros, R. Kernchen, M. Dianati, and D. Rieck, "Performance study of a Green Light Optimized Speed Advisory (GLOSA) application using an integrated cooperative ITS simulation platform," in *Proceedings of the 2011 7th International Wireless Communications and Mobile Computing Conference*, pp. 918–923, Istanbul, Turkey, 2011.

[33] W.-H. Lee, Y.-C. Lai, and P.-Y. Chen, "A study on energy saving and $\hbox{CO}_{2} $ emission reduction on signal countdown extension by vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 3, pp. 890–900, 2015.

[34] W.-H. Lee and J.-Y. Li, "An eco-driving advisory system for continuous signalized intersections by vehicular ad hoc network," *Journal of Advanced Transportation*, vol. 2018, Article ID 5060481, 2018.

[35] B. Xu, X. J. Ban, Y. Bian et al., "Cooperative method of traffic signal optimization and speed control of connected vehicles at isolated intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1390–1403, 2019.

[36] Y. Du, W. ShangGuan, and L. Chai, "A coupled vehicle-signal control method at signalized intersections in mixed traffic environment," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 3, pp. 2089–2100, 2021.

[37] Y. Feng, C. Yu, and H. X. Liu, "Spatiotemporal intersection control in a connected and automated vehicle environment," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 364–383, 2018.

[38] J. Hu, B. B. Park, and Y.-J. Lee, "Coordinated transit signal priority supporting transit progression under Connected Vehicle Technology," *Transportation Research Part C: Emerging Technologies*, vol. 55, pp. 393–408, 2015.

[39] W. Wu, W. Ma, K. Long, and Y. Wang, "Integrated optimization of bus priority operations in connected vehicle environment," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 1853–1869, 2016.

[40] M. Seredynski, G. Laskaris, and F. Viti, "Analysis of cooperative bus priority at traffic signals," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 5, 2019.

[41] G. Laskaris, M. Seredynski, and F. Viti, "A real time hybrid controller for regulating bus operations and reducing stops at signals," in *Proceedings of the 2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 1–7, Kraków, Poland, 2019.

[42] G. Laskaris, M. Seredynski, and F. Viti, "Sensitivity analysis on regularity based driver advisory systems," in *Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, Rhodes, Greece, September 2020.

[43] J. Barbaresso, G. Cordahi, D. Garcia, C. Hill, A. Jendzejec, and U.S. Department of Transportation, "ITS Research 2015-2019: Connected Vehicle (Connected Vehicles: Benefits, Roles, Outcomes)," 2015, https://www.its.dot.gov/research_areas/WhitePaper_connected_vehicle.htm.

[44] L. Head, D. Gettman, and Z. Wei, "Decision model for priority control of traffic signals," *Transportation Research Record*, vol. 1978, no. 1, pp. 169–177, 2006.

[45] P. A. Lopez, M. Behrisch, and L. Bieker-Walz, "Microscopic traffic simulation using sumo," in *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2575–2582, Maui, HI, USA, November 2018.

[46] L. Gurobi Optimization, *Gurobi Optimizer Reference Manual*, Gurobi Optimization, Inc, Houston, Texas, 2021.

[47] N. A. Chaudhary and C.-L. Chu, *New PASSER Program for Timing Signalized Arterials*, Texas Transportation Institute, Texas A & M University System, Texas, 2003.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 1829104, 14 pages
https://doi.org/10.1155/2022/1829104

Research Article

Differential Congestion Pricing Strategies for Heterogeneous Users in the Mixed Traffic Condition

Yifan Chen,[1] **Yuliang Zhang**,[2] **and Ziyuan Gu**[3]

[1]*Institute of Transport Studies, Department of Civil Engineering, Monash UniversityClayton, VIC 3800, Australia*
[2]*Hangzhou Institute of Advanced Technology, Hangzhou 310018, China*
[3]*Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing 210096, China*

Correspondence should be addressed to Yifan Chen; yifan.chen@monash.edu

Received 15 October 2021; Accepted 23 December 2021; Published 19 January 2022

Academic Editor: Zhihong Yao

Congestion pricing is one effective demand management strategy to alleviate traffic congestion. This work investigates pricing schemes for mixed traffic flow systems where the human-driven vehicles (HVs) and autonomous vehicles (AVs) coexist. The emerging and integration of autonomous vehicles can help improve the overall transportation efficiency and safety. Given the coexistence of HVs and AVs in the near future, there is need to adjust the existing traffic management strategies to adapt to the mixed traffic conditions. In this study, congestion pricing is imposed on the HVs and the AVs differently, that is, a distance-based toll to the HVs while a delay-based toll to the AVs. We consider six user groups based on the value of time (VOT) and the vehicle types. Compared with the unified distance-based toll, the advantage of delay-based toll is demonstrated first. Then, a surrogate-based optimization framework, namely the regressing Kriging (RK) model, is formulated. Three pricing schemes are investigated and compared: equity-oriented (EQ), environment friendliness-oriented (EN), and revenue-oriented (RE) schemes. Results show that the RE scheme collects the highest revenues; however, its cost-efficiency is weakened. The EQ scheme reduces the variance in the average travel costs among user groups, thus solving the equity issue.

1. Introduction

With the rapidly increasing population, traffic congestion has become a major problem, especially in urban areas. Due to the limited land resources and high costs, it is unsustainable to address this issue by merely building new infrastructures. Given this situation, congestion pricing, one of the promising demand-oriented strategies, is widely advocated to mitigate congestion. As an economic lever, it originates from the Pigou's theory where extra traveler entering the network should be charged an additional fee for his negative impact imposed on other road users [1]. This type of pricing model is known as the first-best pricing. However, the first-best pricing is difficult to be implemented because it is unrealistic to charge all links in a network strictly by the marginal cost surplus of traffic [2]. Instead, the second-best pricing model was proposed wherein just a subset of links are charged based on the traffic conditions [3]. A branch of congestion pricing practices emerged since 1975, when the area licensing scheme (ALS) was successfully operated in Singapore. However, quite a few schemes were aborted at the stage of trial or referendum. Low public acceptance was the main reason which led to the failure of these cases [4].

Equity is a vital factor influencing public acceptance towards the congestion pricing policy [5]. As a result, vast literature discussed the equity issue on the design of congestion pricing schemes from different perspectives. Lucas et al. pointed out that drivers' places of residence will affect their willingness to pay. Under a cordon-based pricing scheme, people living outside the cordon and commuting towards it may become losers while the winners are those living and conducting activities inside the cordon [6]. Gaunt et al. investigated residents' attitudes towards congestion

pricing in Edinburgh (Scotland). He found that people with high car dependency were less supportive of congestion tolls [7]. Chen at al. conceived a novel pricing scheme considering drivers' perceived level of service (LOS). People who enjoyed higher LOS on their trips would pay more [8].

Value of time (VOT) is an important direction when we deal with the equity issue. According to Arnott's research, the heterogeneity in VOT affected the calculation of welfare effects. The congestion toll favored people with high VOT relative to the schedule delay [9]. Later, more studies were dedicated to examine the impact of congestion pricing on users with different VOTs [10–12]. However, few researches designed the equitable toll rates directly based on the difference in VOTs. Zheng et al. carried out a seminal study where the group-based tolls were determined according to the distribution of VOTs. They argued that when the total time savings after the implementation of congestion pricing were monetized, drivers with higher VOT would gain more. Hence, it is unfair to ask all road users to pay the same amount of toll [13]. However, the VOT categorization in their research is only related to income levels. Obviously, the vehicle type also influences VOT. The advent of autonomous vehicles (AV) has the potential to impact travel behaviour as well as the transportation system. The AV technologies enable drivers to do other activities on their trips, thus making travel time more productive [14–16]. Zhong et al. applied a mixed logit model to quantify the changes in VOT if taking AVs. The results showed that the VOT would be reduced by 30% at most [17]. On the contrary, the majority of actualized congestion pricing schemes are cordon-based or distance-based [5]. The drawback of the cordon-based scheme is that drivers are charged equally regardless of the actual distance travelled inside the cordon. As for the distance-based case, it may induce people to choose the shortest paths, thus making the distribution of congestion more uneven. Gu et al. once put forward the joint distance and delay toll (JDDT) scheme to remedy the limitation of the distance-based congestion pricing policy [2]. But the novel scheme is difficult to be put into practice in the era of human-driven cars. With the embedded vehicle-to-vehicle communication technology, AVs can collect rich information for the real-time traffic condition. As a result, it provides opportunities to apply advanced congestion pricing approaches.

The traditional way to solve the congestion pricing problem is formulating a bi-level mathematical program with equilibrium constraints (MPECs) [18–20]. Nonetheless, it remains challenging to deal with MPEC in a dynamic large-scale transportation network. As the dynamic congestion pricing (DCP) problem often involves high-dimension decision variables and a nonconvex, nonlinear, and nonclosed form objective function [21–23]. In recent years, with the help of mature dynamic traffic assignment (DTA) simulators, simulation-based optimization (SBO) is recognized as an alternative way to handle this problem. It does not require an explicit mathematical formulation of the objective function which is usually unavailable under stochastic traffic dynamics. In general, SBO methods can be classified into four types: (1) direct search; (2) gradient-based approach; (3) feedback control; (4) surrogate-based method. The direct search partitions the search space into hyper rectangles. By comparing the objective function values, the potential optimal rectangle is determined for further partition [24]. Given its exhaustive partitioning characteristics, it is time-consuming to improve the current best solution. The gradient-based method is derivative-based, which tries to approximate the gradient by finite difference. However, it is not suitable for objective functions where the gradient does not exist everywhere [25]. The feedback control is an efficient optimization method. It adjusts the input variables iteratively to let the system output approach the set point [26]. But this method is not applicable when it comes to high-dimension problem with complex constraints [27]. In this context, the surrogate-based methods gain popularity for their ability in approximating expensive-to-evaluate functions. With limited number of evaluations, they can approximate simulation input-output mapping [28]. The regressing Kriging (RK) model outperforms other surrogate models for its excellent performance in prediction [29]. By constructing a probability model, it generates not only an interpolated spatial correlation but also an estimate of the uncertainty. Meanwhile, the feature of regressing sample points overcomes the influence of simulation noise [30].

In order to monitor the effects of the congestion pricing schemes, it is essential to understand the traffic dynamics at the network level. Geroliminis and Levinson proposed the macroscopic fundamental diagram (MFD), which related the network average density to the average flow [31]. The congestion can be identified when the flow decreases with the density, and it indicates the oversaturated state of the network. The critical density is the one at which the maximal production (flow) is achieved [32]. After the implementation of the pricing scheme, we hope the network density would be around the critical value.

Although plenty of studies attempted to design appropriate congestion pricing schemes, there are still some research gaps: (1) few works considered the mixed traffic of human-driven vehicle (HV) and AVs. The inserted communication technologies of AVs enable us to apply the advanced pricing strategy, which can be different from that of HVs; (2) reduction in VOT for AV users was not well considered; and (3) existing studies rarely compared the pricing schemes with different objectives. In this paper, we propose a surrogate-based optimization to solve the dynamic pricing problem for mixed traffic with drivers of heterogeneous VOTs. The HVs are charged by distance while AV users pay according to the average link delay. The VOT-based tolls are adopted among different groups. The results of three pricing schemes, namely equity-oriented (EQ), environment friendliness-oriented (EN), and revenue-oriented (RE) schemes are compared. To the best of the authors' knowledge, the congestion pricing optimization under mix traffic with heterogeneous users has never been proposed before and hence offers significant contributions to the literature.

2. Methodology

2.1. Simulation-Based Dynamic Assignment. The stochastic route choice (SRC) model is a kind of nonequilibrium-seeking DTA model. It has been extensively used for transportation operations owing to the lower computational cost compared to the dynamic user equilibrium (DUE) model [33]. Table 1 summarizes the notations used in SRC model. Let $G = (N, L)$ denote the object network. The time-varying distance-based toll rates for the HV user group i can be represented as a decision vector $\mathbf{v}_i = [v_{1i}, v_{2i}, \ldots, v_{ni}, \ldots, v_{Ni}]$. And the delay-based toll rates for AV user group j is expressed as $\mathbf{w}_j = [w_{1j}, w_{2j}, \ldots, w_{nj}, \ldots, w_{Nj}]$.

The generalized travel cost for HV user group i to choose link a in the nth interval is as follows:

$$C_{ai}(n) = \overline{t}_a(n-1) + \frac{l_a v_{ni}}{\mathrm{VOT}_i}. \tag{1}$$

The generalized travel cost for AV user group j to choose link a in the nth interval is as follows:

$$C_{aj}(n) = \overline{t}_a(n-1) + \frac{\overline{d}_a(n-1) w_{nj}}{\mathrm{VOT}_j}. \tag{2}$$

Note that the communication is only among AVs. Hence $\overline{d}_a(n-1)$ is the average value acquired from AVs which crossed link a during the $(n-1)$th time interval. Given the mixed traffic, it can approximately represent the average delay for all groups. The C-logit model is used to simulate the route choice behavior. The probability of choosing path r among the path set R^{OD} for a given OD pair in the nth interval is as follows:

$$P_r(n) = \frac{\exp\left(-C_r(n) - CF_r\right)}{\sum_{m \in R^{OD}} \exp\left(-C_m(n) - CF_m\right)}, \tag{3}$$

where the term CF_r (CF_m) is the commonality factor of path r (m) which describes the degree of overlapping with other alternative paths [34]. $C_r(n)$ is the sum of the travel costs of all links in path r.

2.2. Proposed Framework. In this paper, travelers are categorized according to the income level and the vehicle type (HV or AV). Given the same vehicle type, high-income travelers have higher VOT due to high productivity [35]. Given the same income level, AV users will have a smaller VOT because they can spend in-vehicle time engaged in other activities. The differentiated VOTs influence the route choice behavior. When paying the same amount of toll, travelers of high VOT will perceive fewer costs (according to equation (1) and (2)), thus resulting in inequity. As a result, let $v_{n,\mathrm{low}}$ denote the distance-based toll rate for the HV user group of the lowest VOT (HV reference group) in the nth time interval. Then, the toll rates for other HV user groups are set based on the VOT difference from the HV reference group. Similarly, denote the delay-based toll rate for the AV user group of the lowest VOT (AV reference group) in the nth time interval as $w_{n,\mathrm{low}}$. Then, the toll rates for other HV user groups are set based on the VOT difference from the AV reference group. We focus on the optimization of congestion pricing for the congested central area of the city. Three different congestion pricing schemes are investigated: (1) EQ scheme; (2) EN scheme; and (3) RE scheme.

As we adopt the distance-based and delay-based tolls for HV and AV user groups separately, the average travel costs (AUD/trip) for different groups would be disperse after the congestion pricing. Hence, the objective of the EQ scheme is to minimize the variance in average travel costs:

$$\min Z_1 = \sigma^2\left(\mathrm{ATC}_g\right), \tag{4}$$

where Z_1 is the objective function for the EQ scheme and ATC_g is the average travel cost for the user group g after the implementation of congestion pricing. Note that ATC_g includes the trip travel time (in monetary unit) and toll paid.

As for the EN scheme, we aim to minimize the total emission in the pricing zone. The QUARTET pollution emission model is utilized [36]. In the mesoscopic simulation, we can only acquire the average speed for a link during each time interval. The impact of the acceleration and deceleration processes has been considered via the average speed:

$$\min Z_2 = \sum_{g=1}^{G} EM_g, \tag{5}$$

where Z_2 is the objective function of the EN scheme; EM_g is the total emission of the user group g during the toll period; and G is the number of user groups.

The RE scheme is devoted to maximizing the total toll paid. It is equivalent to minimize the negative version of this term. The collected revenues can be redistributed to road users or used to maintain the transport infrastructures:

$$\min Z_3 = -\sum_{g=1}^{G} \mathrm{REV}_g, \tag{6}$$

where Z_3 is the objective function of the RE scheme; REV_g is the revenues collected from the user group g; and G is the number of user groups.

There are several constraints considered for all three schemes:

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left|\overline{K}_n - K_{cr}\right| < e,\\
&\left|v_{n,\mathrm{low}} - v_{n+1,\mathrm{low}}\right| \le \alpha, \quad n = 1, 2, \ldots, N-1,\\
&\left|w_{n,\mathrm{low}} - w_{n+1,\mathrm{low}}\right| \le \beta, \quad n = 1, 2, \ldots, N-1,\\
&v_{\min} < v_{n,\mathrm{low}} \le v_{\max}, \quad n = 1, 2, \ldots, N,\\
&w_{\min} < w_{n,\mathrm{low}} \le w_{\max}, \quad n = 1, 2, \ldots, N.
\end{aligned} \tag{7}$$

Constraint 7 controls the network performance. $\overline{K}_n$ is the network average density in the nth time interval; K_{cr} is the critical density of the network; and e is the preset threshold.

Table 1: Notations used in the SRC model.

Notation	Description
N	Node set for the object network
L	Directed link set for the object network
v_{ni}	Distance-based toll rate for HV user group i in the nth time interval
w_{nj}	Delay-based toll rate for AV user group j in the nth time interval
N	Number of time intervals
VOT_i	Value of time for HV user group i
VOT_j	Value of time for AV user group j
$\bar{d}_a(n-1)$	Average delay in link a in the $(n-1)$th interval
$\bar{t}_a(n-1)$	Average travel time for link a in the $(n-1)$th interval
CF_r	Commonality factor of path r

This constraint makes sure that the network density is close to the critical value over time. Constraint 8 and 9 ensure that the toll rates between adjacent time intervals do not change sharply. $v_{n,\mathrm{low}}$ ($v_{n+1,\mathrm{low}}$) is the distance-based toll rate for the HV reference group in the nth ($(n+1)$th) time interval; $w_{n,\mathrm{low}}$ ($w_{n+1,\mathrm{low}}$) is the delay-based toll rate for the AV reference group in the nth ($(n+1)$th) time interval; and α and β are the smoothing parameters. Constraint 10 and 11 set the upper and lower bound for the toll rates. $v_{\min}$ ($v_{\max}$) is the lower and upper bound for the distance-based toll rate of the HV reference group. $w_{\min}$ ($w_{\max}$) is the lower and upper bound for the delay-based toll rate of the AV reference group. Note that the toll rates of other groups are set proportionally. Once the toll rate ranges for reference groups are confirmed, the ranges for other groups are also determined.

2.3. Surrogate-Based Optimization. The RK model is employed to do the constrained optimization. Figure 1 shows the procedure of the surrogate optimization. Some initial points (or toll plans) are generated by a certain sampling method. In our work, Latin hypercube sampling (LHS) is adopted. It is a space-filling approach which stratifies each dimension of the decision variables into an equal number of partitions [37]. At least $(m+1)$ initial points are needed for m-dimension problems [38]. Considering the high dimension of our optimization problems, totally $2(m+1)$ points are generated. The we run the network simulation to get the objective function values for these initial inputs. The simulation input-output mapping is used to build a preliminary surrogate model. In order to enhance the model, additional infill points are required. The additional samples can provide more information in the potential regions where the good solution may exists. The infill process will not terminate until the stop criteria are met. Finally, the accuracy of the model is checked. When the accuracy is validated, we can declare that the optimum is achieved. Otherwise, the surrogate model should be redesigned.

2.3.1. Model Construction. For the RK model, the prediction for a nonsampled point $\mathbf{x}_{n+1}$ is written as follows:

$$\hat{y}(\mathbf{x}_{n+1}) = \hat{\mu} + \varepsilon(\mathbf{x}_{n+1}), \tag{8}$$

where $\hat{\mu}$ is the constant mean; $\varepsilon(\mathbf{x}_{n+1})$ is a normally distributed and independent estimation error, and it can be expanded as follows:

$$\varepsilon(\mathbf{x}_{n+1}) = \boldsymbol{\psi}^{\mathrm{T}}(\mathbf{R} + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{9}$$

where $\mathbf{y}$ is the output vector for n sampled points ($\mathbf{y} = [y_1, y_2, \ldots, y_n]^{\mathrm{T}}$); $\mathbf{I}$ is the identity matrix; $\mathbf{R}$ is the correlation matrix between the sampled points; and λ ($\lambda \in (0, 10]$) is the regression constant. With the positive λ, the RK model regresses the data to reduce the influence from the simulation noise. Ψ is the correlation vector between n sampled points and the new point. The correlation between point $\mathbf{x}_i$ and $\mathbf{x}_j$ is given as follows:

$$\mathrm{Corr}[\mathbf{x}_i, \mathbf{x}_j] = \exp\left[-\sum_{h=1}^{k} \hat{\theta}_h \left| x_{i,h} - x_{j,h} \right|^{\hat{p}_h}\right], \tag{10}$$

where k is the dimension of variables; $\hat{\theta}_h$ and $\hat{p}_h$ is the scaling and smoothness coefficient for the hth dimension, separately; and $\hat{\theta}_h$ ($\hat{\theta}_h \in [0.001, 100]$) denotes the sensitivity to the objective function value. A larger $\hat{\theta}_h$ means the certain dimension is more sensitive than others. The use of $\hat{p} \in [0, 1]$ is usually advocated. However, $\hat{p} = 2$ proves to be suitable for the engineering-based problem [39].

Then, the vector of scaling coefficients θ, the variance σ^2 and the mean μ are estimated by maximizing the likelihood probability [40]:

$$\frac{1}{2\pi^{1/2}\left(\sigma^2\right)^{1/2}|\mathbf{R} + \lambda\mathbf{I}|^{1/2}} \exp\left(\frac{-(\mathbf{y} - \mathbf{1}\mu)^T\mathbf{R} + \lambda\mathbf{I}^{-1}(\mathbf{y} - \mathbf{1}\mu)}{2\sigma^2}\right). \tag{11}$$

The limited memory Broyden–Fletcher–Goldfarb–Shanna (LBFGS) algorithm is employed to solve the maximization problem. This iterative method is appropriate for problems with large numbers of variables [41].

2.3.2. Infill Strategy. In order to improve the accuracy of the surrogate model by the augmented data set, the acquisition function is used to guide the search direction. The expected improvement (EI) function is a widely-used acquisition function, which estimates the magnitude of improvement at a new point [42]. It can explore the unvisited region and exploit the domain of interest at the same time. Let $y_{\min}$

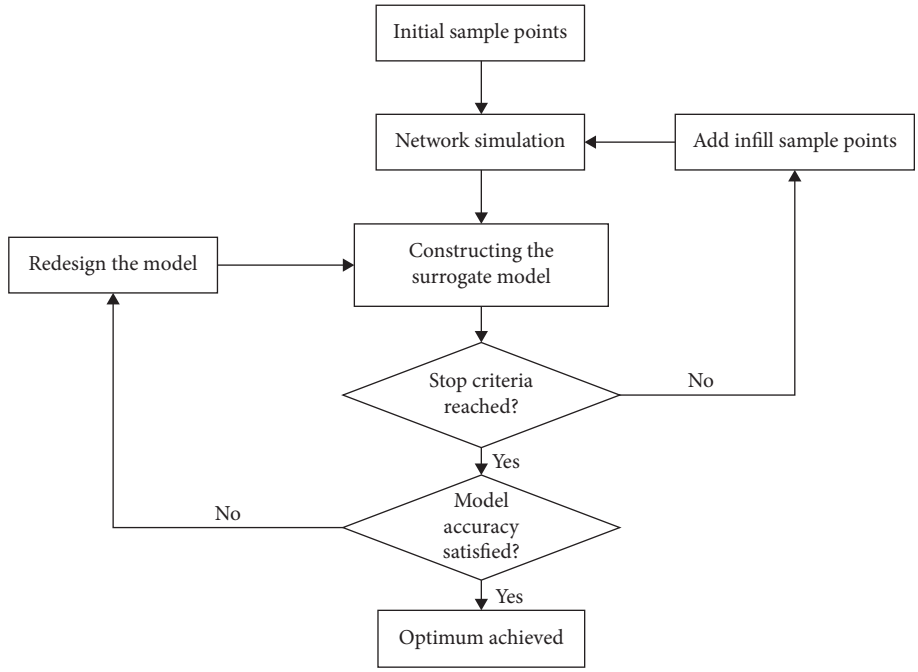


Figure 1: Flowchart of the surrogate-based optimization.

denote the current best solution. The improvement of a new point $\mathbf{x}$ can be expressed as $I(\mathbf{x}) = \max(0, y_{\min} - \hat{y}(\mathbf{x}))$. Then, the formulation of EI is demonstrated as [43]:

$$E[I(\mathbf{x})] = \begin{cases} [y_{\min} - \hat{y}(\mathbf{x})]\Phi\left(\dfrac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}_{ri}(\mathbf{x})}\right) + \hat{s}_{ri}(\mathbf{x})\phi\left(\dfrac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{s}_{ri}(\mathbf{x})}\right), & \text{if } \hat{s}_{ri}(\mathbf{x}) > 0, \\ 0, & \text{if } \hat{s}_{ri}(\mathbf{x}) = 0, \end{cases} \tag{12}$$

where $\Phi(.)$ is the normal cumulative distribution function; $\phi(.)$ is the probability density function; and $\hat{s}_{ri}(\mathbf{x})$ is the re-interpolation prediction error for the objective:

$$\hat{s}_{ri}{}^2(\mathbf{x}) = \hat{\sigma}_{ri}{}^2\left[1 - \boldsymbol{\psi}^T(\mathbf{R} + \lambda\mathbf{I})^{-1}\boldsymbol{\psi}\right], \tag{13}$$

where $\hat{\sigma}^2{}_{ri}$ is the estimate of σ^2 for the re-interpolation:

$$\hat{\sigma}^2_{ri} = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})^T(\mathbf{R} + \lambda\mathbf{I})^{-1}\mathbf{R}(\mathbf{R} + \lambda\mathbf{I})^{-1}(y - \mathbf{1}\hat{\mu})}{n}. \tag{14}$$

Note that by replacing $\hat{\sigma}^2$ with $\hat{\sigma}^2_{ri}$, $E[I(\mathbf{x})]$ stays at zero for all the existing points. Hence, the EI sampling will not be trapped at existing sample points. The new points are obtained with the aim of maximizing the $E[I(\mathbf{x})]$.

The EI function is sufficient for unconstrained optimization. However, when it comes to the constrained optimization. A new point that has a good objective value may not meet the expensive-to-evaluation constraint at the same time. As such, another acquisition, namely the probability of improvement (PI) function is utilized. The mechanism of PI is to research for a new point which has the highest probability to meet the complex constraint. We need to construct a new surrogate model for the constraint separately. Denote $\hat{c}(\mathbf{x})$ as the estimated constraint value. Then, we can formulate the PI function as follows:

$$\mathrm{P}\left[\hat{c}(\mathbf{x}) > c_{\min}\right] = \begin{cases} \Phi\left(\dfrac{c_{\max} - \hat{c}(\mathbf{x})}{\hat{s}_{c,ri}(\mathbf{x})}\right), & \hat{s}_{c,ri}(\mathbf{x}) > 0, \\ 0, & \hat{s}_{c,ri}(\mathbf{x}) = 0, \end{cases} \tag{15}$$

where $c_{\max}$ is the preset threshold as in Constraint 7 and $\hat{s}_{c,ri}(\mathbf{x})$ is the re-interpolation prediction error for the constraint. The constrained EI function $E[CI(\mathbf{x})]$ can be written as follows:

$$E[CI(\mathbf{x})] = \begin{cases} E[I(\mathbf{x})] \cdot P\left[\hat{c}(\mathbf{x}) \le c_{\min}\right], & \hat{c}(\mathbf{x}) \le c_{\min}, \\ 0, & \hat{c}(\mathbf{x}) > c_{\min}. \end{cases} \tag{16}$$

2.3.3. Model Validation. The leave-one-out cross-validation is a method frequently used to measure the accuracy of surrogate models [44]. This approach leaves out one observation and reconstructs the surrogate model with the

remaining points. The estimated output of the removed point from the re-fitted model will be compared with the actual output. The key metric for the leave-one-out cross-validation is the standardized cross-validated residual (SCVR). It is calculated as follows:

$$\frac{y(\mathbf{x}_i) - \hat{y}_{-i}(\mathbf{x}_i)}{\hat{s}_{-i}(\mathbf{x}_i)}, \tag{17}$$

where $y(\mathbf{x}_i)$ is the actual output for $\mathbf{x}_i$; $\hat{y}_{-i}(\mathbf{x}_i)$ is the estimated output of $\mathbf{x}_i$ without $\mathbf{x}_i$; and $\hat{s}_{-i}(\mathbf{x}_i)$ is the prediction error of $\mathbf{x}_i$ from the re-fitted model without $\mathbf{x}_i$. The model is validated when the SCVRs for all points roughly fall in the interval [−3,3]. It means the model proves to be 99.7% confident that $y(\mathbf{x}_i)$ lies in the interval $[\hat{y}_{-i}(\mathbf{x}_i) - 3\hat{s}_{-i}(\mathbf{x}_i), \hat{y}_{-i}(\mathbf{x}_i) - 3\hat{s}_{-i}(\mathbf{x}_i)]$ [41].

3. Results and Discussion

3.1. Experiment Setup. In this paper, we perform a mesoscopic simulation-based DTA in the Melbourne city, Australia in the traffic simulation software AIMSUN. The travel demand is collected via the loop detector and changes every 15 min between 6 : 00 AM and 10 : 00 AM. Figure 2(a) shows the congested central business district (CBD) (inside the red box). We choose it as the pricing zone. There are 75 nodes and 218 links in the object network. The information of daily income of residents is extracted from the Victorian Integrated Survey of Travel and Activity (VISTA). We assume that people work 8 hours a day. As a result, the VOTs equate to the daily income divided by eight. Travelers are classified into three groups based on the distribution of VOT initially. The VOT of each group is the median of the corresponding population. Since the mass market penetration of AVs is challenging in the near future, 30% AV penetration rate is used in our work. Meanwhile, we assume that more high-income travelers can afford AVs. Finally, there are totally six user groups shown in Table 2. Note that AV users will have a reduction in their VOTs by 30% [14].

The agents of different user groups are uniformly distributed in the network. In order to avoid simulation randomness, we run three replications with different random seeds. The MFDs for the three replications (R1, R2, and R3) and the average result (AVE) are presented in Figure 2(b). According to the MFDs, the flow starts to drop at 25 veh/km/lane. Hence, we set the critical density as 25 veh/km/lane. The network density over time in the no-toll scenario is demonstrated in Figure 2(c). Observe that the density exceeds the critical value around 8:30. So the toll period is from 8:30 to 10:00.

In our study, $v_{\min}$ is set as 0 AUD/km, and $v_{\max}$ is set as 1 AUD/km. $w_{\min}$ is set as 0 AUD/h, and $w_{\max}$ is set as 15 AUD/h. The smoothing parameter α is set as $1/3 \times (1 - 0) = 0.3$, while β is set as $1/3 \times (15 - 0) = 5$. One can refer to the study by Gu et al. [34], wherein the analysis on the smoothing parameters is conducted. The threshold e is set as 2.45 veh/km/lane which is nearly one-fourth of the value (9.8 veh/km/lane) in the no-toll scenario. The toll period is divided into six 15-minute intervals. Hence, the decision variable vector has 12 dimensions, six of which are distance-based toll rates for the HV reference group, while the rest are delay-based toll rates for the AV reference group.

3.2. Comparison between Unified Delay-Based Toll and Distance-Based Toll. In order to unveil the advantage of the delay-based toll, we compare the network performance under the unified delay-based toll and distance-based toll. First of all, let us introduce a term, namely the spatial spread of density [45]. It measures the heterogeneity of congestion distribution and is calculated as follows:

$$\gamma = \sqrt{\frac{\sum_i l_i n_i (k_i - K)^2}{\sum_i l_i n_i}}, \tag{18}$$

where l_i is the length of link i; n_i is the number of lanes of link i; k_i is the density of link i; and K is the network density.

The relationship between the spatial spread of density and the network density can be expressed as the third-polynomial function [34]:

$$\gamma(K) = aK^3 + bK^2 + cK. \tag{19}$$

We collect 1000 data points to plot Figure 3(a). The lower envelop of Figure 3(a) is fitted to Equation (19). Parameters are estimated as: $a = -0.0004$, $b = 0.0278$, and $c = 1.2345$. According to Figure 3(a), the spatial spread of density will naturally increase with the network density. Simoni et al. further proposed an advanced metric: the deviation from spread [46]:

$$\Delta = \gamma - \gamma(K). \tag{20}$$

Smaller Δ indicates the lower heterogeneity of congestion distribution, thus leading to the higher network flow. We select a delay-based toll scheme with the unified delay toll rate of 13 AUD/h and a distance-based toll scheme with the unified distance toll rate of 0.5 AUD/km. According to Figure 3(b), the network density under the distance-based toll scheme (DI_density) is smaller than that of the delay-based toll (DE_density) in most of the time. However, the result of the network flow is the opposite. When we look into the average deviation from spread during the toll period, we find that it is much smaller under the delay-based toll scheme (1.93 veh/km/lane vs. 4.93 veh/km/lane). This conclusion is in line with Gu's research [27]. It indicates that the delay-based toll can reduce the heterogeneity of congestion distribution to a great extent. This is also one motivation of our research to implement the delay-based toll on AVs which are equipped with the perfect technology to realize it. The inserted advanced communication technology of autonomous vehicles enables us to collect data such as the vehicle position and the average delay of queues.

3.3. Results of Optimal Toll Designs with Different Objectives. The stop criterion in this paper is that 30 infill points are collected or $EI = 0$ for consecutive 10 points. Figures 4(a)–4(f) show the results for the EQ scheme. When we track the convergence history of EI, the value decreases gradually

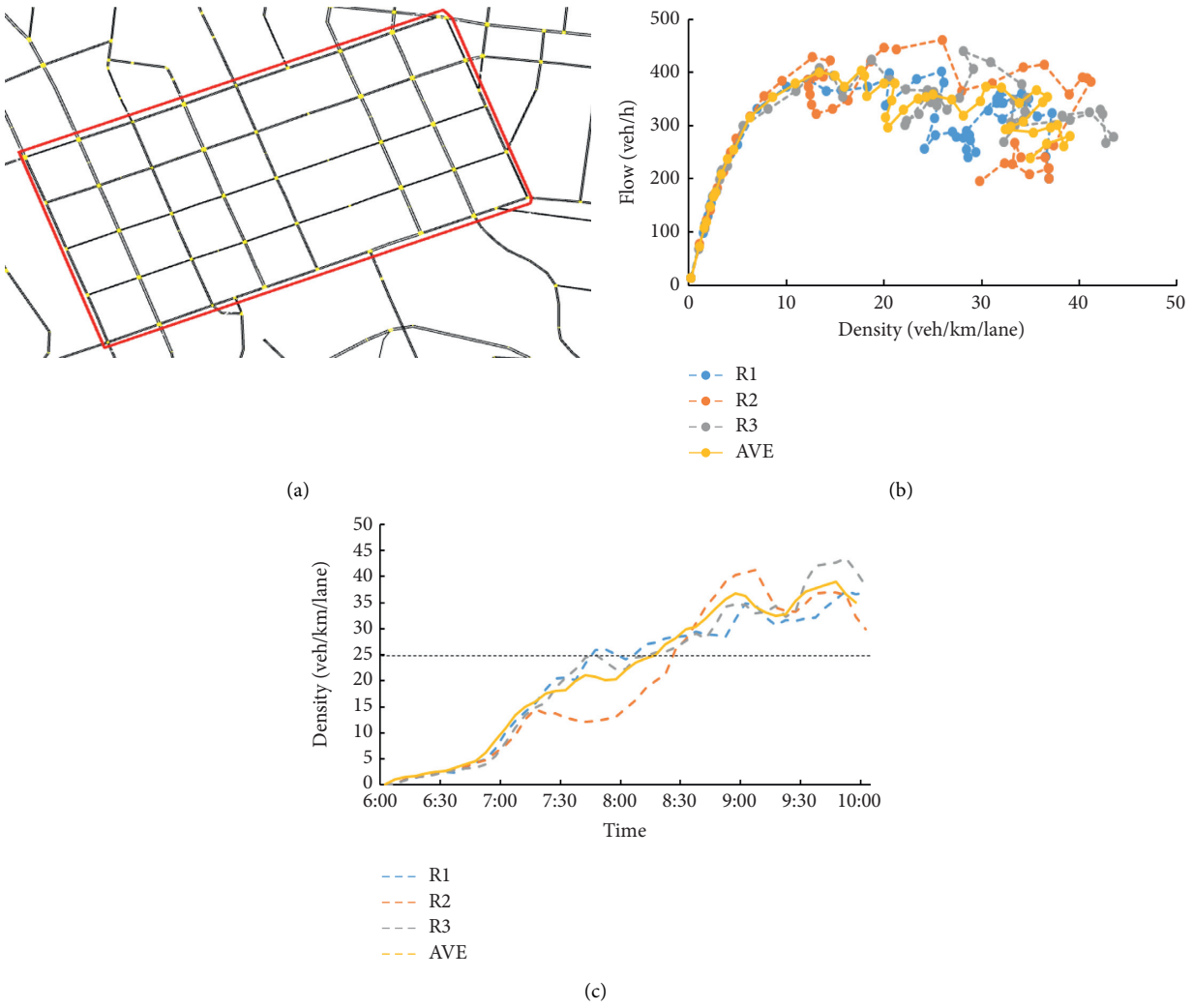
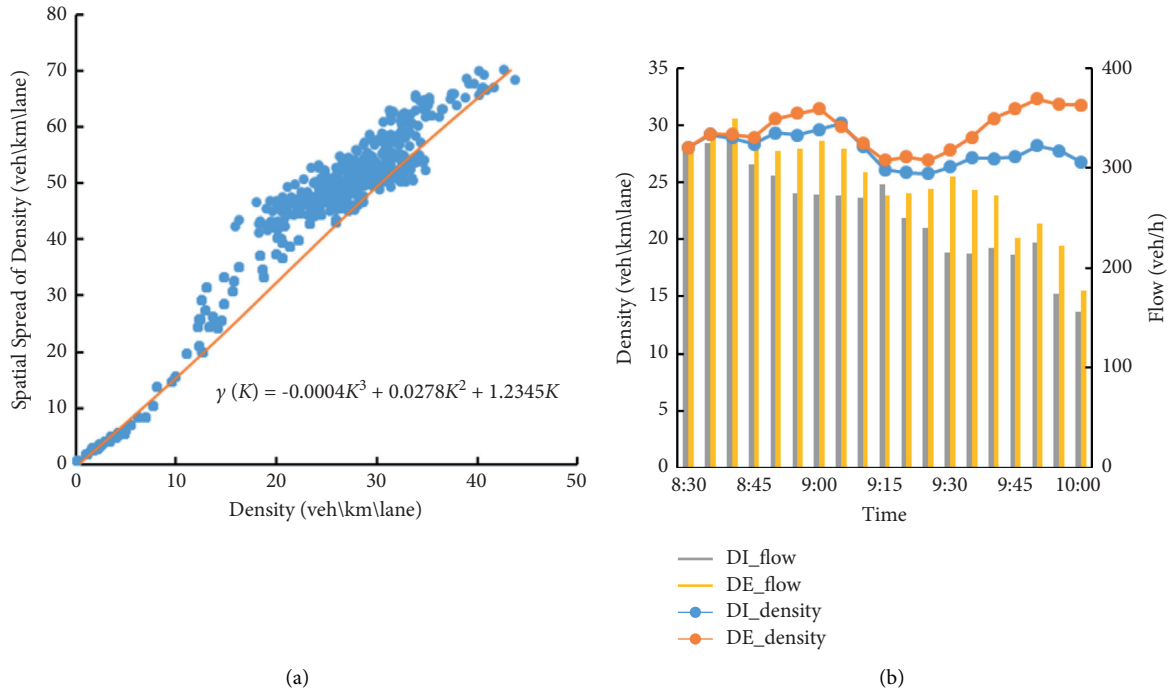


Figure 2: (a) Pricing zone; (b) MFD for the pricing zone in no-toll scenario; (c) density over time in no-toll scenario.

Table 2: The VOTs and percentages of different user groups.

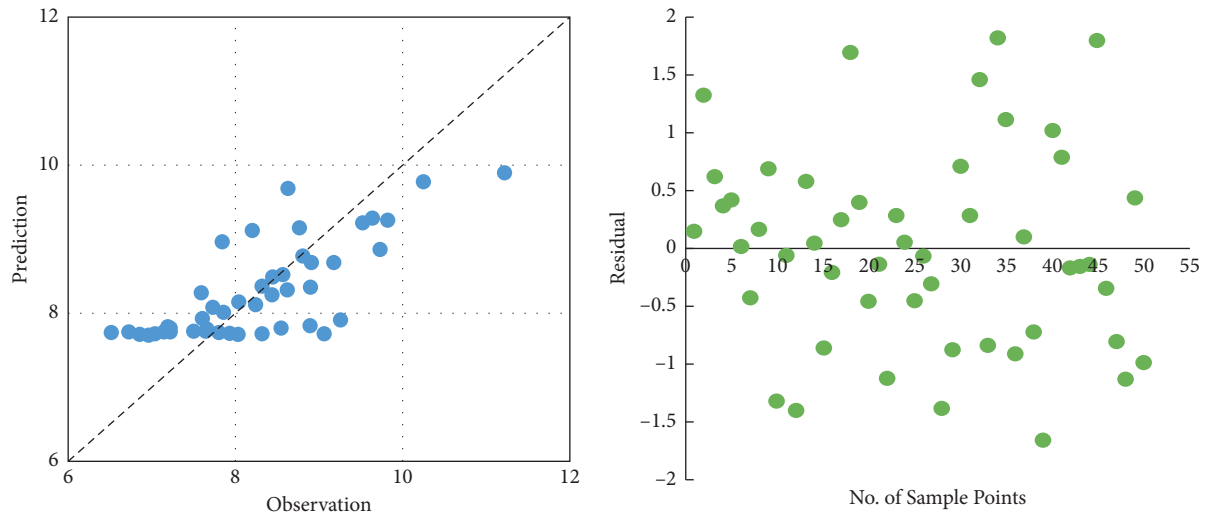
Group name	VOT (AUD/h)	Percentage	AV penetration (%)
*AV*_low	8	15	20
*AV*_medium	19	10	50
*AV*_high	28	5	85
*CV*_low	12	59	0
*CV*_medium	28	10	0
*CV*_high	42	1	0

(shown in Figure 4(e)). The infill process terminates after the 24th infill point because EI equates to zero for 10 function evaluations successively. It means, it is unlikely to find a new point which improves the current best solution. According to Figures 4(a) and 44(b), the accuracy of the objective and constraint models are both validated, seeing all the sample points lie in the interval [−3,3]. The distribution of sample points is demonstrated in Figure 4(f). Note that only the infill points which have $EI = 0$ are plotted. As the limited memory LBFGS algorithm will randomly select a point when it cannot find a new point with $EI > 0$. As a result, the infill points with $EI = 0$ are not representative in terms of the objective and constraint values. The X-axis is the constraint value while Y-axis is the objective value. The dashed red line is the constraint limit. Most of the initial points violate the constraint while only one infill point is beyond the constraint limit, which proves that the infill strategy works well. The point inside the red circle is the optimal solution for the EQ schemes. We solve the EN (Figures 5(a)–55(f)) and RE scheme (Figures 6(a)–66(f)) in the same way.



(a) (b)

Figure 3: (a) Relationship between the spatial spread of density and density and (b) comparison of flow and density between unified delay toll scheme and distance toll scheme.



(a) (b)

Figure 4: Continued.

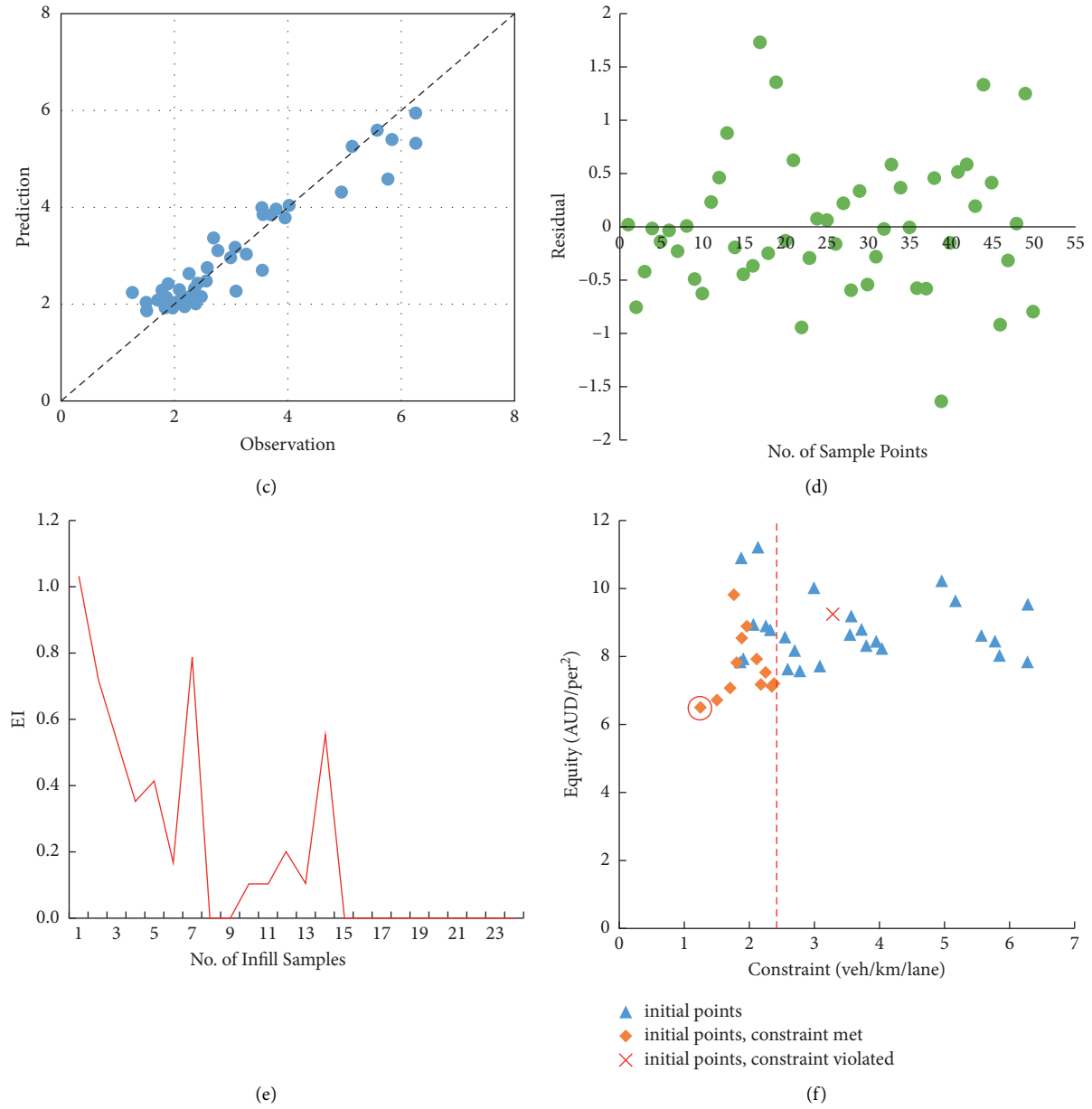


Figure 4: Results for EQ scheme: (a) comparison between predictions and observations for the objective model; (b) residuals of sample points for the objective model; (c) comparison between predictions and observations for the constraint model; (d) residuals of sample points for the constraint model; (e) convergence history of EI; (f) distribution of sample points.

3.4. Comparison among Three Schemes. In this section, we compare the results of three schemes. The optimal toll plans for three schemes are illustrated in Figures 7(a)–7(c). The primary Y-axis is the distance toll rate for HV groups, while the secondary Y-axis is the delay toll rates for AV groups. According to Figure 7(d), all schemes control the network density over time around the critical value (25 veh/km/lane) well. The constraint values under all three schemes are within 2.45 veh/km/lane (shown in Table 3). The RE scheme lets road users to pay 15.79 (10 × 3 AUD) which is 40% higher than the revenues collected in the EQ or EN scheme. However, when we look into the total travel time reduction and the constraint value, the network performance under RE scheme is similar to those under other two schemes. It indicates that RE scheme is less cost-effective. The extremely high toll may induce the low marginal effect.

The EN scheme produces least emissions (7.02 kg). According to the QUARTET pollution emission model, the emission rate goes up with the increase of the cruising speed. Moreover, the optimal toll plan for the EN scheme reduce the emissions by controlling the total time travelling inside the pricing zone, seeing the least total travel time (7726 min.veh) among three schemes. It suggests that the EN scheme can encourage individuals to travel less time in the pricing zone. Nonetheless, since the HVs and AVs are charged differently in our work, in order to increase the

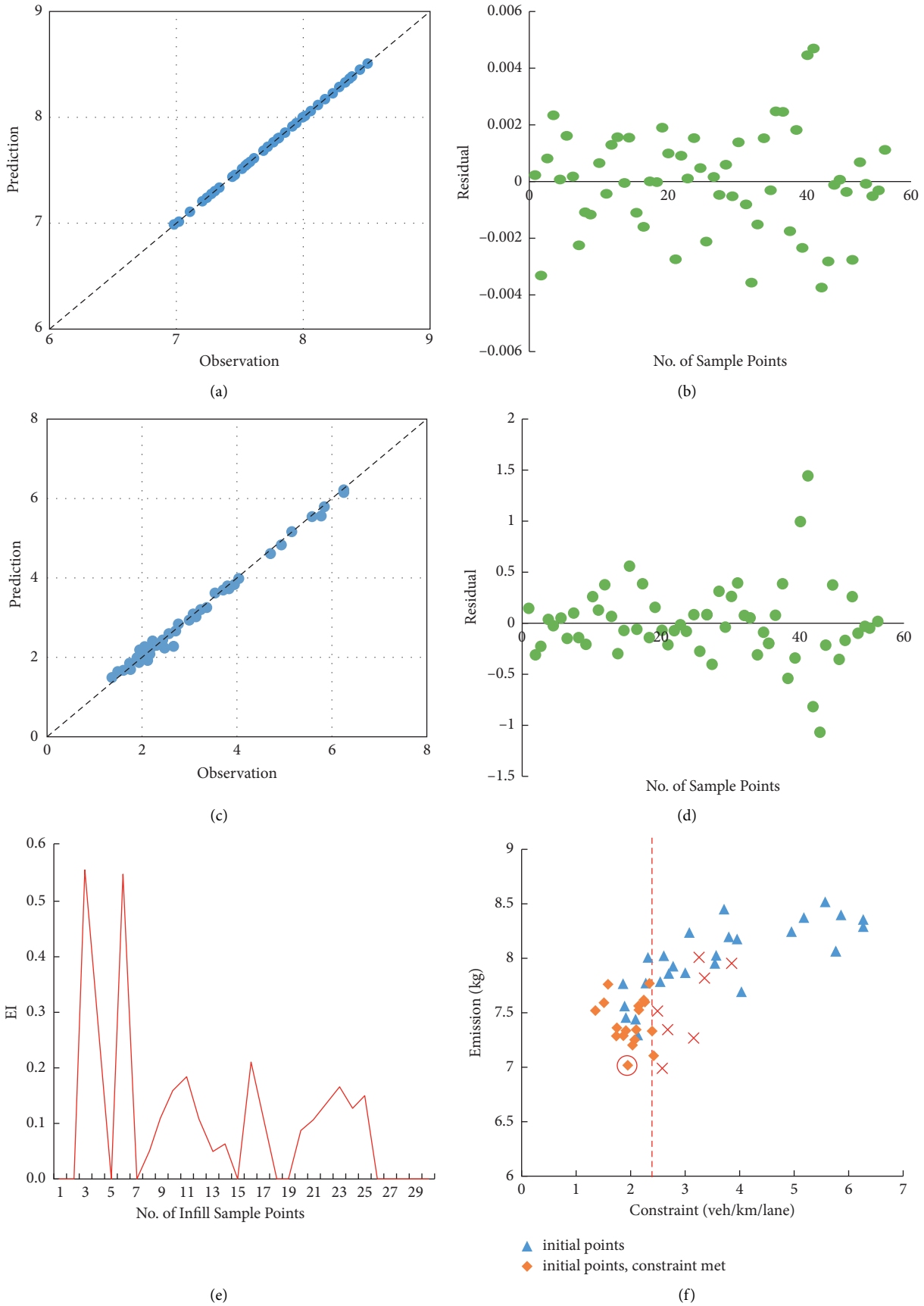


Figure 5: Results for EN scheme: (a) comparison between predictions and observations for the objective model; (b) residuals of sample points for the objective model; (c) comparison between predictions and observations for the constraint model; (d) residuals of sample points for the constraint model; (e) convergence history of EI; (f) distribution of sample points.

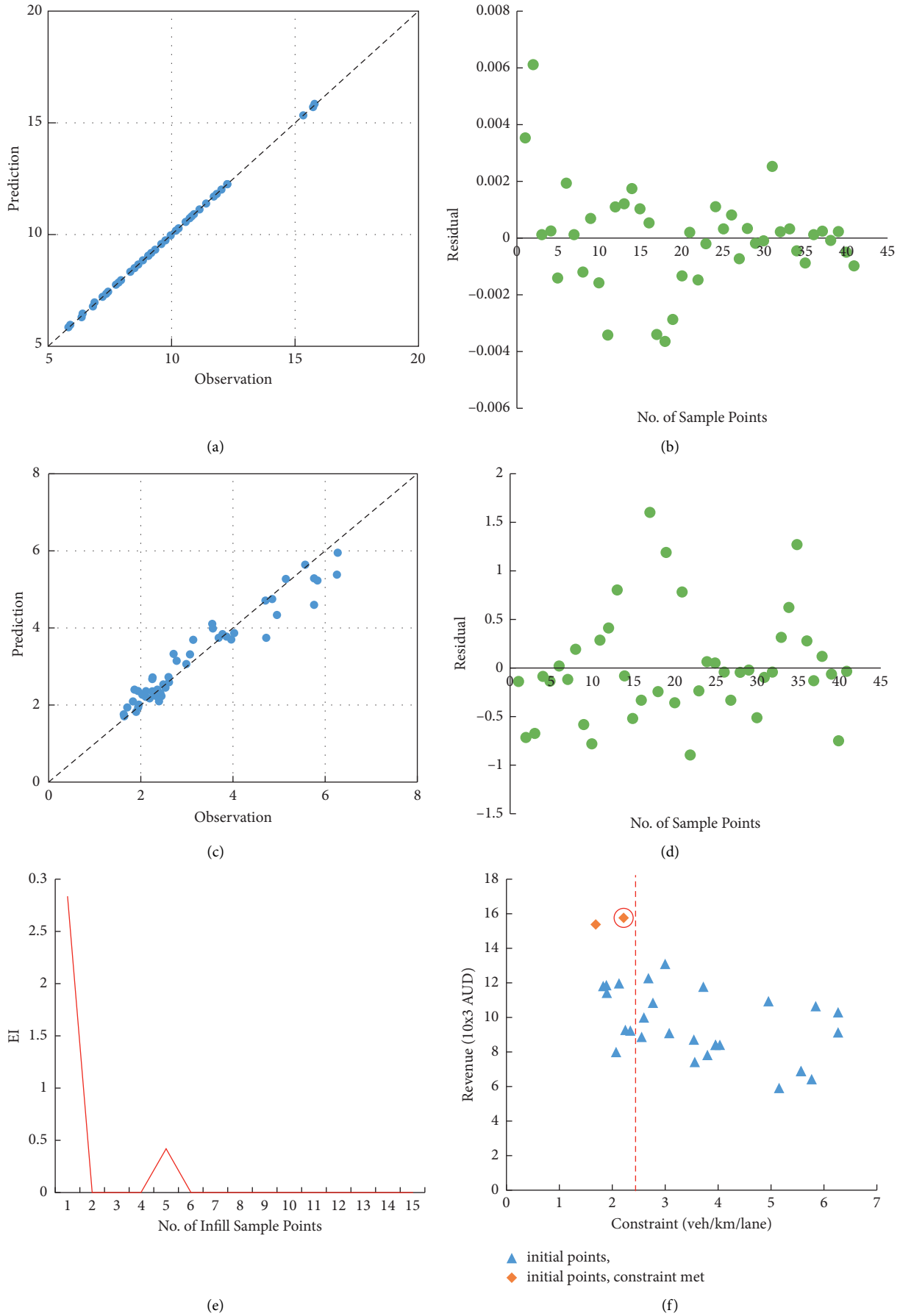


Figure 6: Results for RE scheme: (a) comparison between predictions and observations for the objective model; (b) residuals of sample points for the objective model; (c) comparison between predictions and observations for the constraint model; (d) residuals of sample points for the constraint model; (e) convergence history of EI; (f) distribution of sample points.

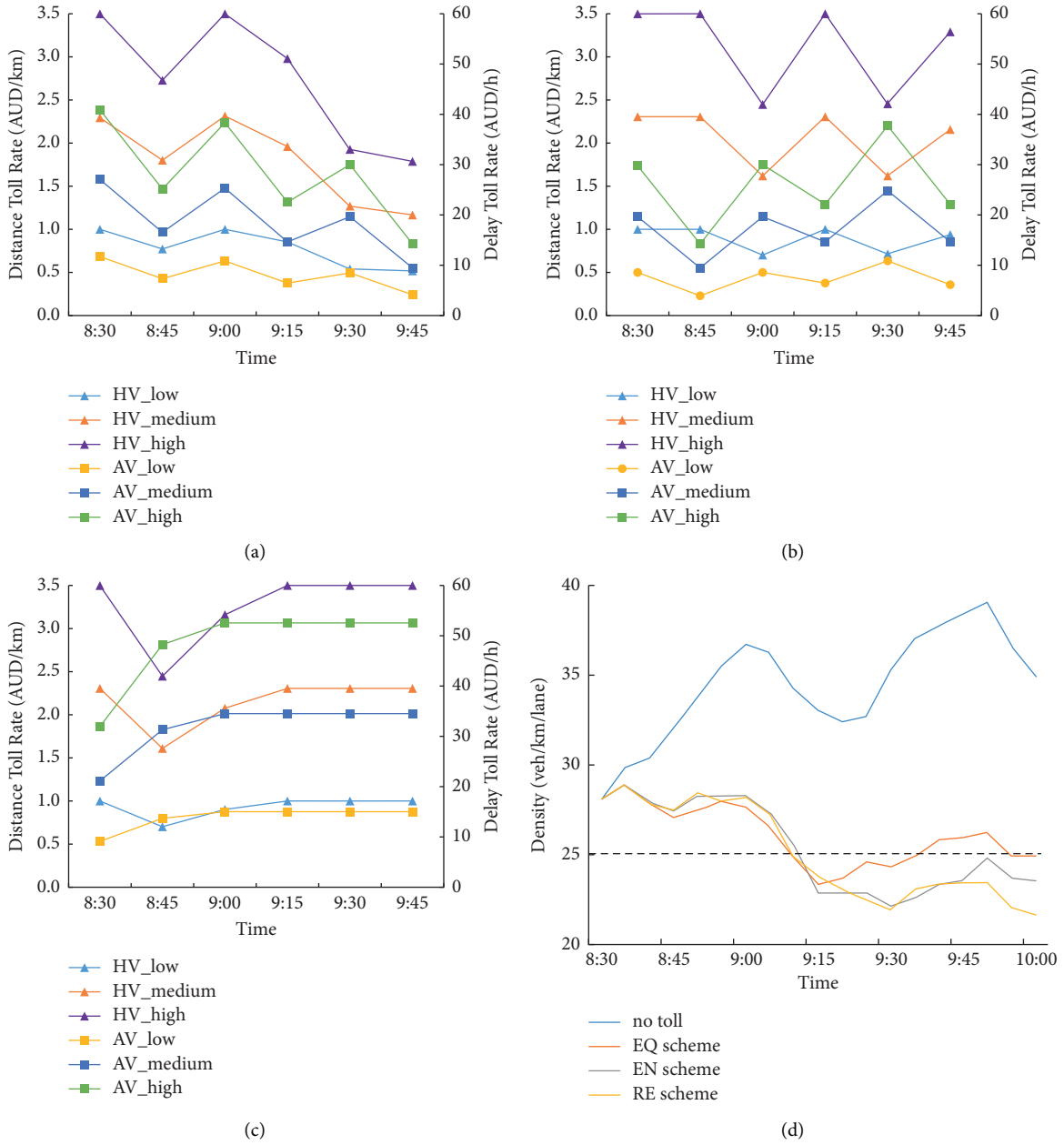


Figure 7: (a) Optimal toll plan for EQ scheme; (b) optimal toll plan for EN scheme; (c) optimal toll plan for RE scheme; (d) density over time under different schemes.

Table 3: Metric comparison among different schemes.

Scheme name	No-toll	EQ scheme	EN scheme	RE scheme
Equity (AUD/per^2)	9.14	**6.53**	8.00	9.50
Emission (kg)	9.35	7.64	**7.02**	7.63
Revenue (10 × 3 AUD)	0	11.92	11.20	**15.79**
Constraint value (veh/km/lane)	9.8	1.25	1.94	2.23
Total travel time (min.veh)	10,665	8589	**7726**	7829
Average speed (km/h)	27.69	30.30	30.47	30.62

The bold ones are the most important values.

acceptance of the policy, $\sigma^2(\mathrm{ATC}_g)$ should be as small as possible. From this perspective, the EQ scheme outperforms other schemes with $\sigma^2(\mathrm{ATC}_g) = 6.53$ AUD/per^2. In addition, compared with the no-toll scenario, the EQ scheme reduces the NOx emissions by 18.3% and collects 11.92 (10×3 AUD), which is appreciable.

4. Conclusion

The formulation of DCP optimization involves high-dimension decision variables as well as the expensive-to-evaluate objective function. Meanwhile, with the emerging of autonomous technology, HVs and AVs will coexist in the transportation network in the near future. How to design appropriate congestion pricing schemes under mixed traffic is challenging. In this paper, we divide the road users into six groups according to the vehicle type and VOTs. Three congestion pricing schemes with different objectives are investigated. The network performance is considered as a complex constraint to keep the network density close to the critical value. The RK model is adopted to estimate the input-output mapping. Initial and infill sampling strategies are combined to search for optimal toll plans.

A large-scale simulation-based DTA model of Melbourne, Australia, is used to demonstrate the results of proposed pricing schemes. The RE scheme is not advocated because it charges road users too much money without achieving much better network performance. The phenomenon illustrates that demand originates from the pricing zone is evitable. High tolls may obtain low marginal effect. The EN scheme produces least emissions and maintains the relatively high average speed. In this study, AVs are charged by delay while HVs are charged by distance. Meanwhile, people with higher VOTs will have high toll rate. As a result, we should guarantee that the proposed pricing scheme can be accepted by all groups. Hence, the EQ scheme is superior to EN and RE schemes for its smallest $\sigma^2(\mathrm{ATC}_g)$. To ensure the efficiency of our pricing schemes.

We should also guarantee that travelers report their VOTs honestly. Nan et al. [15] conducted an in-depth discussion on how to stimulate people to tell truth. For instance, the mechanism design which does not require travelers to provide confidential information. In addition, punishments such as fines or banning the liars from the pricing zone for a certain period are considerable. As long as the government takes a lead role to boost the group-based discrimination, the proposed pricing schemes can be put into practice.

Data Availability

The simulation-based traffic model is available from GitHub through https://github.com/meeadsaberi/dynamel.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

The first author was supported by the China Scholarship Council grant scheme.

References

[1] A. Pigou, *The Economics of Welfare*, MacMillan and Co., Ltd, London, UK, 1st edition, 1932.

[2] Z. Y. Gu, S. Shafiei, Z. Y. Liu, and M. Saberi, "Optimal distance- and time-dependent area-based pricing with the network fundamental Diagram," *Transportation Research Part C*, vol. 95, pp. 1–28, 2018.

[3] M. Merchand, "A note on optimal tolls in an imperfect environment," *Econometrica*, vol. 36, pp. 575–581, 1968.

[4] Z. Y. Gu, Z. Y. Liu, Q. X. Cheng, and M. Saberi, "Congestion pricing practices and public acceptance: a review of evidence," *Case Studies on Transport Policy*, vol. 6, no. 1, pp. 94–101, 2018.

[5] D. A. Hensher and Z. Li, "Referendum voting in road pricing reform: a review of the evidence," *Transport Policy*, vol. 25, pp. 186–197, 2013.

[6] L. M. D. Freitasm, "Equity effects of congestion charges: an exploratory analysis with MATSim," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2670, pp. 75–82, 2017.

[7] M. Gaunt, T. Rye, and S. Allen, "Public acceptability of road user charging: the case of Edinburgh and the 2005 referendum," *Transport Reviews*, vol. 27, no. 1, pp. 85–102, 2007.

[8] Y. F. Chen, N. Zheng, and H. Vu, "A novel urban congestion pricing scheme considering travel cost perception and level of service," *Transportation Research Part C*, vol. 125, 2021.

[9] R. Arnott, A. D. Palma, and R. Lindsey, "Schedule delay and departure time decisions with heterogeneous commuters," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1197, pp. 56–67, 1988.

[10] H. Yang and X. Zhang, "Multi-class network toll design problem with social and spatial equity constraints," *Journal of Transportation Engineering*, vol. 128, no. 5, pp. 420–428, 2002.

[11] F. Xiao, Z. Qian, and M. Zhang, "The morning commute problem with coarse toll and non-identical commuters," *Networks and Spatial Economics*, vol. 11, no. 2, pp. 343–369, 2011.

[12] V. V. D. Berg and E. Verhoef, "Congestion tolling in the bottleneck model with heterogeneous values of time," *Transportation Research Part B*, vol. 45, pp. 60–78, 2011.

[13] N. Zheng and N. Geroliminis, "Area-based equitable pricing strategies for multimodal urban networks with heterogeneous users," *Transportation Research Part A*, vol. 136, pp. 357–374, 2020.

[14] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transportation Research Part A*, vol. 77, pp. 167–181, 2015.

[15] Z. H. Yao, H. R. Jiang, Y. Cheng, Y. Jiang, and B. Ran, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2020.

[16] Z. H. Yao, T. R. Xu, Y. Jiang, and R. Hu, "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, 2021.

[17] H. T. Zhong, W. Li, M. Burris, A. Talebpour, and K. Sinha, "Will autonomous vehicles change auto commuters' value of travel time?" *Transportation Research Part D*, vol. 83, 2020.

[18] L. N. Liu and J. F. McDonald, "Economic eficiency of second-best congestion pricing schemes in urban highway systems," *Transportation Research Part A*, vol. 33, no. 3, pp. 157–188, 1999.

[19] Z. Liu, Q. Meng, and S. Wang, "Speed-based toll design for cordon-based congestion pricing scheme," *Transportation Research Part C*, vol. 31, pp. 83–98, 2013.

[20] Z. Liu, S. Wang, and Q. Meng, "Optimal joint distance and time toll for cordon-based congestion pricing," *Transportation Research Part C*, vol. 69, pp. 81–97, 2014.

[21] X. Chen, C. Xiong, H. Xiang, Z. Zheng, and Z. Lei, "Time-of-day vehicle mileage fees for congestion mitigation and revenue generation: a simulation-based optimization method and its real-world application," *Transportation Research Part C*, vol. 63, pp. 71–95, 2016.

[22] Z. Tan, H. Yang, and R. Y. Guo, "Dynamic congestion pricing with day-to-day flow evolution and user heterogeneity," *Transportation Research Part C*, vol. 61, pp. 87–105, 2015.

[23] S. V. Ukkusuri, L. Han, and K. Doan, "Dynamic user equilibrium with a path based cell transmission model for general traffic networks," *Transportation Research Part B*, vol. 46, no. 10, pp. 1657–1684, 2012.

[24] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the lipschitz constant," *Journal of Optimization Theory and Applications*, vol. 79, no. 1, pp. 157–181, 1993.

[25] S. Amaran, N. V. Sahinidis, B. Sharda, and S. J. Bury, "Simulation optimization: a review of algorithms and applications," *Annals of Operations Research*, vol. 240, pp. 351–380, 2016.

[26] N. Zheng, G. R´erat, and N. Geroliminis, "Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment," *Transportation Research Part C*, vol. 62, pp. 133–148, 2016.

[27] Z. Y. Gu and M. Saberi, "Simulation-based optimization of toll pricing in large-scale urban networks using the network fundamental Diagram: a cross-comparison of methods," *Transportation Research Part C*, vol. 122, 2021.

[28] X. Chen, L. Zhang, X. He, C. Xiong, and Z. Li, "Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, pp. 359–381, 2013.

[29] X. He, X. Q. Chen, C. F. Xiong, Z. Zhu, and L. Zhang, "Optimal time-varying pricing for toll roads under multiple objectives: a simulation-based optimization approach," *Transportation Science*, vol. 51, pp. 412–426, 2017.

[30] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science*, vol. 4, pp. 409–423, 1989.

[31] N. Geroliminis and D. Levinson, "Cordon pricing consistent with the physics of overcrowding," in *Proceedings of the 18th International Symposium on Transportation and Traffic Theory*, Springer, Berlin, Germany, July 2009.

[32] N. Zheng, R. A. Waraich, K. Axhausen, and N. Geroliminis, "A dynamic cordon pricing scheme combining the macroscopic fundamental Diagram and an agent-based traffic model," *Transportation Research Part C*, vol. 46, pp. 1291–1303, 2012.

[33] M. Ashfaq, Z. Y. Gu, S. T. Waller, and M. Saberi, "Comparing dynamic user equilibrium and noniterative stochastic route choice in a simulation-based dynamic traffic assignment model: practical considerations for large-scale networks," *Journal of Advanced Transportation*, vol. 2021, Article ID 6667335, 16 pages, 2021.

[34] Z. Y. Gu, S. T. Waller, and M. Saberi, "Surrogate-based toll optimization in a large-scale heterogeneously congested network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, pp. 638–653, 2019.

[35] M. Börjesson, J. Eliasson, M. Hugosson, and K. Brundell-Freij, "The stockholm congestion charges - 5 Years on effects, acceptability and lessons learnt," *Transport Policy*, vol. 20, pp. 1–12, 2012.

[36] Quartet Deliverable, *Assessment of Current Tools for Environmental Assessment in QUARTET, DRIVE II Project V2018*, QUARTET, London, UK, 1992.

[37] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, pp. 239–245, 1979.

[38] J. Ekstrom, I. Kristoffersson, and N. H. Quttineh, "Surrogate-based optimization of cordon toll levels in congested traffic networks," *Journal of Advanced Transportation*, vol. 50, pp. 1008–1033, 2016.

[39] A. I. J. Forrester, A. J. Keane, and N. W. Bressloff, "Design and analysis of "noisy" computer experiments," *AIAA Journal*, vol. 44, pp. 2331–2339, 2006.

[40] H. M. Gutmann, "A radial basis function method for global optimization," *Journal of Global Optimization*, vol. 19, no. 1, pp. 201–227, 2001.

[41] Y. H. Xiao, Z. X. Wei, and Z. G. Wang, "A limited memory BFGS-type method for large- scale unconstrained optimization," *Computers and Mathematics with Applications," Computers & Mathematics with Applications*, vol. 56, pp. 1001–1009, 2008.

[42] E. Brochu, V. M. Cora, and N. D. Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," 2014, https://arxiv.org/abs/1012.2599.

[43] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[44] A. Molinaro, R. Simon, and R. Pfeiffer, "Prediction error estimation: a comparison of resampling methods," *Bioinformatics*, vol. 21, no. 15, pp. 3301–3307, 2005.

[45] V. L. Knoop and S. P. Hoogendoorn, "Empirics of a generalized macroscopic fundamental Diagram for urban freeways," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2391, no. 1, pp. 133–141, 2013.

[46] M. D. Simoni, A. J. Pei, R. A. Waraich, and S. P. Hoogendoorn, "Marginal cost congestion pricing based on the network fundamental Diagram," *Transportation Research Part C*, vol. 56, pp. 221–238, 2015.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 8150617, 26 pages
https://doi.org/10.1155/2022/8150617

Research Article

Influence of Lane-Changing Behavior on Traffic Flow Velocity in Mixed Traffic Environment

Han Xie, Qinghua Ren, and Zheng Lei

Xihua University, School of Management, Jinniu, Chengdu, China

Correspondence should be addressed to Qinghua Ren; ren-qinghua@163.com

Received 4 November 2021; Revised 7 December 2021; Accepted 11 December 2021; Published 18 January 2022

Academic Editor: Zhihong Yao

In mixed traffic with autonomous vehicles, the relationship between speed and lane-changing behavior is an important basis for mixed traffic control. In this study, we use empirical, simulation, and data-driven methods to study the relationship between speed and lane change rates in mixed traffic under different autonomous vehicle penetration rates. We use the empirical data to establish the corresponding road simulation models. Based on the simulation model, the traffic flow simulation experiments under the conditions of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% penetration rate of autonomous vehicles were carried out. The analysis of the simulation results found that: (1) the penetration of autonomous vehicles into the road has a positive impact on the lanes far away from the entrance and exit, while the impact on the lanes closer to the entrance and exit is not obvious. (2) Lane-changing behavior has effectively decreased with the penetration of autonomous vehicles, but it is not obvious when the penetration rate exceeds 10%, and there is no significant drop in the lane connecting the entrance and exit. (3) There is a linear relationship between speed and lane-changing rate. Under different penetration rates, the data-driven analysis is used to perform multiple linear regressions, and the regression formula fits are all above 0.7. Based on the above findings, the linear formula of the fitting is proposed, and the value interval of the parameters in different states is given as well. Due to the small changes in the parameter values under different permeability conditions, the model has a certain degree of stability. The speed-lane change rate model proposed in this study can better describe the relationship between the speed of the ring-shaped urban expressway and the lane-changing behavior in the mixed traffic environment with the larger traffic flow.

1. Introduction

In recent years, with the development of autonomous driving technology, autonomous vehicles will appear on urban roads in the future. Related researchers predict that by 2040, as many as 75% of vehicles may be self-driving (IEEE Press Release, 2012). Mixed driving of autonomous vehicles and manually driven vehicles will be the normal state of traffic for a long time in the future. Due to the mixing of autonomous vehicles, the traffic state is not the same as that under manual driving. It is important to the management and control of mixed traffic by understanding the operation mechanism of mixed traffic in the future.

Speed, as one of the important parameters that describes the characteristics of traffic flow, has great significance to the description of the state of traffic flow. Lane-changing behavior, as one of the basic behaviors of vehicles running on the road, has a great disturbance to traffic flow and is also one of the main reasons for traffic congestion. In the case of small traffic flow, the behavior of changing lanes has little interference with the speed, but in the case of large traffic flow, this interference is more obvious.

A self-driving vehicle is a vehicle that can automatically drive on the road in a controlled environment. This vehicle will automatically choose to follow the car when the traffic does not allow it to change lanes. But, under special circumstances (such as obstacles ahead, the target lane can only be reached after changing lanes, and a more suitable lane needs to be selected), the lane change behavior will be performed. In mixed traffic, autonomous vehicles will affect speed and will also effectively reduce the occurrence of lane-changing behavior [1–5]. It is important to understand the

impact of autonomous vehicles on the speed and lane-changing behavior in a 4-lane closed road in mixed traffic. The affection is very important to understand the operating mechanism of traffic flow in mixed traffic. The current research studies on the lane-changing behavior of autonomous vehicles are mainly focused on the lane-changing behavior of a single autonomous vehicle. Research methods mainly used data-driven model [6–8] and simulation model. There are few studies on the impact of lane-changing behavior on traffic flow, especially in each lane. For traffic flow, lane-changing behavior has a greater impact on each lane than on a road section. Therefore, it has great significance to study the relationship between speed and lane-changing behavior in each lane. The relationship is important to understand the operation mechanism of traffic flow in mixed traffic in the urban road with closed loops, multiple lanes, and multiple entrances and has practical significance for future traffic control under mixed traffic conditions.

Through the current studies, we found that speed is a very important parameter to describe the traffic flow state. Most research studies on speed are based on basic graph theory. Basic graph theory is the basis of traffic flow under manual driving. For mixed traffic, some scholars had proposed a basic graph. The basic graph of mixed traffic mainly studied the capability, that is, the relationship between flow and density within nonfull density [9–13]. The current studies are mainly from two aspects. On the one hand, the data of flow and density are obtained through simulation, and the relationship graph of flow and density is fitted [14, 15]. On the other hand, the relationship graph of flow, density, and speed is obtained from the average head-to-head distance of mixed traffic in a balanced state by theoretical analysis methods [16–21]. In the research studies of the basic graph of mixed traffic, the main studies are focused on the impact of car-following behavior on the three basic parameters of traffic flow. The important perspective is focus on car-following behavior. However, there are not only car-following behavior and lane-changing behavior of autonomous vehicles in mixed traffic but also lane-changing behavior of human-driven vehicles. But the study of lane-changing behavior is not involved in the basic graph of mixed traffic. At the same time, the research studies on speed are only from the three basic parameters of traffic flow and the car-following model of auto-driving vehicles, while the relationship between speed and lane-changing behavior in mixed traffic has not been studied.

The main research studies on the impact of lane-changing behavior on traffic flow under human driving are considered from the traffic flow model. Based on the classic LWR model, Laval and Daganzo [22] proposed a modified model [23, 24] and further proposed a model that used four parameters of speed, speed, capacity, and lane change rate [24]. In the model, lane-changing vehicles are described as particles that completely blocked the traffic flow in each lane. Jin [25] clearly proposed that lane-changing behavior had an impact on traffic capacity, but it still used the LWR model and introduced lane-changing factor to modify the three parameters of traffic flow, density, and speed. In HCM2010 [26], specific calculation methods for interwoven and non-interwoven lanes are given for the interweaving area of expressways. The lane-changing rate is introduced for interwoven lanes, but the relationship between lane-changing behavior and speed is not analyzed. The above studies are all within the basic graph of manual driving vehicles, and none of them analyzes the relationship between speed and lane-changing behavior. Park and Rictchie [27] studied the relationship between lane-changing behavior and speed. The study proposed a regression model on the measured data of lane-changing behavior, vehicle types, the difference between upstream and downstream traffic flow, and speed, respectively, to study the influence of lane-changing behavior, vehicle types, and the difference between upstream and downstream traffic flow on speed in the detection area. The lane-changing behavior is defined as the number of lane-changing behaviors performed by a single vehicle in the detection interval, and the different lane-changing times are counted to establish a multiple regression model. The model shows that lane-changing behavior can affect speed changes more than factors such as vehicle type and the differences between upstream and downstream traffic flow. Park and Rictchie gave a multiple linear regression model, as shown in [27]

$$v = 6.169 - 4.410L0 + 5.154L2 + 11.090L3 + 44.150L4. \quad (1)$$

In the equation, v is the average speed in the measurement section; $L0$ is the total number of vehicles, in which lane changes are 0 time in the measurement section per unit time/the total number of vehicles in the measurement section per unit time; $L2$ is the total number of vehicles, in which lane changes are 2 times in the measurement section per unit time/the total number of vehicles in the measurement section per unit time; $L3$ is the total number of vehicles, in which lane changes are 3 times in the measurement section per unit time/the total number of vehicles in the measurement section per unit time; and L4 is the total number of vehicles, in which lane changes are greater than or equal to 4 times in the measurement section per unit time/the total number of vehicles in the measurement section per unit time.

A simulation model was used to study the relationship between speed and lane-changing behavior through each vehicle trajectory in the measurement section from the microscopic view. However, this study has not been conducted on the affect of lane-changing behavior on each lane. The behavior of changing lanes has a greater impact on each lane, but the impact on the entire road is not as great as that in each lane. Moreover, in low traffic flow, the lane-changing behavior of a single vehicle has a minor impact on the road. But in large traffic flow, the impact of lane-changing behavior is more prominent. Therefore, it is more important to study the influence of lane-changing behavior in each lane under large traffic flow, especially speed. The above studies are all carried out in a human-driving environment.

In the autonomous driving environment, the study of lane-changing behavior mainly focuses on behavior prediction [28–30]and lane-changing behavior [31, 32]. Liu et al. [33] used a dynamic control model to describe the lane-changing motivation at high speed. The motive of changing

lanes was considered from both horizontal and vertical aspects, using speed as the motive for safety guarantee, and predicting the behavior of lane changing. Zheng et al. [34] also adopted lane-changing motivation models from horizontal and vertical aspects in a dynamic driving environment. The model used speed, acceleration, and ending time as inputs in the longitudinal direction and used trajectories as inputs in the vertical direction to make decisions about lane-changing behavior. Of course, it is not enough to make lane-changing decision from the intention. Decision also need to consider collisions [35], front and rear cars [36, 37], current road condition [38, 39], and other aspects. In addition, Jongsang et al. [35] used a deep auto-encoding network to establish a multivariate model for lane-changing decisions. Most studies are based on human behavior, and most models are similar to human behavior models. Compared with human behavior models, autonomous vehicle models used more and more accurate information when making decisions. When an autonomous vehicle enters mixed traffic, it will be naturally affected by human-driving vehicles, and there are more factors to be considered when changing lanes. However, the current research studies are mostly based on the auto-driving vehicle, and the research studies in the mixed traffic are mostly from the road section and rarely considered from each lane. For multilane road, the impact of lane-changing behavior on each lane is greater than the impact on the entire road or the section. In this study, the influence of lane-changing behavior of each lane on speed is studied. At the same time, speed is an important parameter to measure the traffic state. It is important to understand the relationship between speed and lane-changing behavior in mixed traffic in each lane. The relationship is important for understanding the operating mechanism of mixed traffic flow.

In summary, the current research studies are less involved in the impact of lane-changing behavior on traffic flow in mixed traffic. The research studies on the relationship between lane-changing behavior and speed are mainly from the human-driving environment and mainly involved the trajectory of lane-changing vehicles and from the road section. The research on the relationship between lane-changing behavior and speed from each lane is rarely involved. The behavior of changing lanes has a great impact on speed in each lane, especially in heavy traffic. Therefore, it is necessary to study the relationship between speed and lane-changing behavior in each lane on closed city roads with circular multilane under heavy traffic flow in mixed traffic. This study uses empirical, simulation, and data-driven research methods to study the relationship between lane-changing behavior and speed in each lane based on the different penetration rates of autonomous vehicles. The research conducted an empirical study on the third ring expressway in Chengdu, China, which has the characteristics of a typical ring-shaped urban expressway [40]. We collected relevant data and established a corresponding simulation model. The relationships between speed and lane-changing behavior, under the autonomous vehicle penetration rates of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%, are studied through simulation models. A data-driven modeling method was used to establish a linear regression model. For the first time, we proposed a single-lane speed and lane change rate model for the closed-loop multilane urban road under mixed traffic environment and have given the corresponding value interval of the parameters. The model can better describe the relationship between the speed and the lane-changing behavior of a circular urban expressway in mixed traffic environment under the large traffic flow.

This study introduces the research background in the first part. The second part introduces the method we used. In this part, we introduce the verification of empirical research, the verification of simulation models, the simulation experiments, and the results obtained from the experiments. The third part is the results and discussion. In this part, the results of the analysis were obtained from the previous experiments, and the speed and lane change models were established. The last part is the conclusion.

2. Method

This study uses empirical, simulation, and data-driven methods to study the relationship between speed and lane-changing behavior. First, empirical methods were used to obtain data for simulation modeling. Then, we verified the accuracy and validity of the data. Finally, simulation and data-driven methods were used to analyze the data characteristics. The validity of the model is verified.

2.1. Empirical Verification. The data used in this study are video data, which were taken in 2011, 2015, and 2018. The videos came from on-site shooting and Chengdu Transportation Administration. We used manual method and self-made traffic flow recognition software to directly obtain the data from the videos. Collection points are shown in Figure 1.

In Figure 1, the points represent each selected collection point. The size of the points represents the number of data, which were collected in each collection point. The collection point satisfies the conditions of the section, which has entrances or exits, pedestrian overpasses, commercial and residential land usage on both sides of road, and the different road types. According to the statistics, the eligible sections on the third ring road can be classified into 4 types, namely, entrances or exits connected to the straight section and connected to overpasses, entrances or exits connected to curved section and connected to overpasses, entrances or exits connected to the straight section and not connected to overpasses, and entrances or exits connected to bends and not connected to overpasses, as shown in Figure 2.

There are four types, and at least one point of each type should be selected as the collection point. The nine points are used as the collection points in this study. The three parameters of the speed, the number of vehicles that changes into lane, and the number of vehicles that changes out of lane are collected from the videos. The collection time is the time period in the morning and afternoon of the working day with a large traffic flow. The collection situations are shown in Table 1.

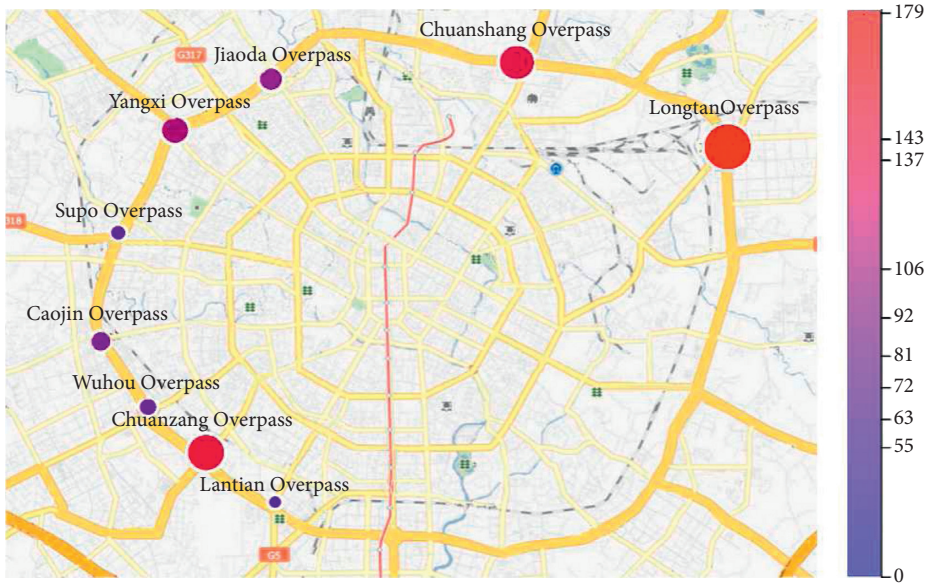


Figure 1: Collection point situation.

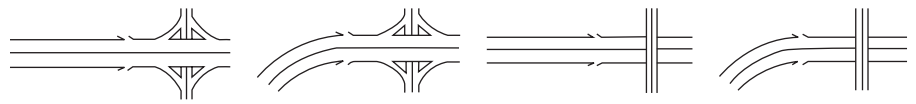


Figure 2: The sketches of connection and linear section. (a) The sketch of the connecting overpass and straight linear section. (b) The sketch of connecting overpass and bend linear section. (c) The sketch of connecting straight linear section. (d) The sketch of connecting bend linear section.

Table 1: Introduction of each collection location.

Collection location	Location	Overpass mode	Connecting road		Date of data acquisition	Time (minutes)	Detection distance (meters)	Dataset
			Inner ring	Outer ring				
Location 1 (Sichuan—Shaanxi overpass)	Northeast	Compound	Zhaojue temple south road	Rongdu Avenue	2011-1-11 2011-1-12	740	100	137
Location 2 (Jiaoda overpass)	Northwest	Compound	Jiao Tong University road	Xihua Avenue	2010-5-27 2015-7-21 2011-1-5 2011-1-11	455	200	92
Location 3 (Yangxi overpass)	Northwest	Compound	Shuhan road	Shuhan West road	2011-1-12 2011-1-11 2011-1-12	600	100	106
Location 4 (Supo overpass)	West	Compound	Su Po West road	Su Po West road	2015-7-19 2015-7-20 2015-7-27	525	100	63
Location 5 (Caojin overpass)	West	Compound	Jinyang road	Yongkang road	2011-1-12 2011-1-13	420	100	81
Location 6 (Wuhou overpass)	South west	Interchange	Wuhou Avenue	Wuhou Avenue	2011-1-12 2011-1-13	610	100	72
Location 7(Sichuan–Tibet overpass)	Southwest	Compound	Jialing road	Twin Avenue	2011-1-13 2011-1-18	710	200	143
Location 8 (Lantian overpass)	Southwest	Compound	Airport expressway	Airport expressway	2015-7-19 2015-7-20 2015-7-27	525	100	55
Location 9 (Fenghuang overpass)	Northeast	Compound	Shili shop road	Chenghua Avenue	2015-7-10 2018-7-11 2018-7-12	1170	100	179

In order to describe the road characteristics, the minimum sample size of statistics is adopted. Equation (2) [41] was used.

$$n \geq \left(\frac{\sigma \bullet K}{E}\right)^2. \tag{2}$$

In the equation, n is the minimum sample size for observation; σ is the standard deviation of the sample of the observed speed; K is the constant of confidence to meet expectations; and E is the allowable error of speed. It depends on the accuracy of the mean speed, generally 1.5~2 km/h.

We used $\sigma = 7$ km/h (according to the two-way eight-lane road), $K = 1.96$ (95%), and $E = 2$ km/h. Then, we got the minimum sample size as 48. That is to say, if the data groups collected are greater than 48 groups, the data can be used to describe the road section characteristics.

Comparing the videos, the unreasonable data were excluded, and the groups of available data were from 55 groups to 179 groups in each collection point. In this study, we studied three parameters: speed, the number of vehicles that changes into lanes, and the number of vehicles that changes out of lanes.

The speed v describes the distance the vehicle passes in a unit of time. In traffic flow theory, speed is classified into 4 types: point speed, travel speed, design speed, and running speed. The average speed of the collected section is used in this study. At first, we used the travel speed to describe the speed of a single vehicle. The travel speed is calculated by [41]

$$v = \frac{L \times 3600}{t \times 1000}. \tag{3}$$

In the equation, v is the travel speed in each lane (km/h); L is the distance of the detection section (m); and t is the travel time (h).

Then, the average speed of the collected section in each lane in the collection time interval is calculated by [41]

$$\overline{V_s} = \frac{1}{1/n \sum_{i=0}^{n} 1/v_i}. \tag{4}$$

In the equation, $(\overline{V_s})$ is the average speed of the detection lane in the detection section (km/h); n is the number of vehicles passing through the detection lane; and v_i is the speed of the i-th vehicle passing through the detection lane (km/h).

The speed unit is km/h. In this study, the collection time interval was 5 minutes, and the data from each lane in the selected section were collected. In our software, if over half part of the vehicle is in the tested lane, then the data of this vehicle are calculated in the tested lane.

Lane change in or lane change out behavior refers to the behavior of vehicles entering the tested lane from other lanes or entering other lanes from the tested lane while running on the road, as shown in Figure 3.

In Figure 3, the blue vehicle is the vehicle to be changed in lanes, and it is planned to change to the target lane from the original driving lane. When the lane change happens, there are two states of the vehicle to the target lane, such as partial entry state and complete entry state. In this study, the behavior of changing into lanes in the target lane is calculated in >50% of the part of the vehicle in the target lane. The behavior of changing out of lanes is similar to the behavior of changing into lanes, except that the calculation is for vehicles leaving the tested lane.

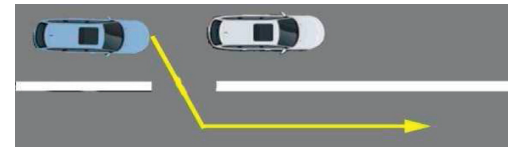


Figure 3: Schematic diagram of vehicle changing lanes (picture from the internet).

The number of vehicles changing into the tested lane is also called the rate of changing into lanes, which refers to the number of vehicles entering the tested lane from other lanes within a unit time and unit length. The number of vehicles changing out of the tested lane is also called the rate of changing out of lanes, which refers to the number of vehicles entering other lanes from the tested lane in a unit time and unit length. Here, the unit time is 1 h, and the unit length is 1 km. It is consistent with the definition of lane change in and change out in HCM2010 [26]. However, HCM2010 [26] did not study lane-changing behavior of vehicles in the section between the entrance ramp and exit ramp. On the highway, the distance between the entrance ramp and exit ramp is generally long, and the impact area of the vehicles, which came from the entrance ramp or exit ramp, is relatively minor. On the urban expressway, the distance between the on-ramp and off-ramp is relatively short, and almost the entire road is in the weaving area, and the impact of changing lanes in the road is greater. If we only considered the impact of the lane-changing behavior on the entrance ramp and exit ramp without the section between the on-ramp and off-ramp, we cannot accurately describe the impact of lane-changing behavior on the urban expressway.

In the data collection, the collection time interval is 5 min, and the distance of each section is 100 m or 200 m. Before data analysis, the data were standardized and converted, the final time interval was 1 h, and the length was 1 km. The unit of the rate of changing into lanes and the rate of changing out of lanes are times that are as same as the number of lane changes in HCM2010 [26]. The data are represented by LCi and LCo, respectively.

In the data arrangement, we found that some data were missing. This situation happens because of the discontinuity of the collection locations and time. We used the time difference method to complement the data, as shown in Equation (5). [41]. After completion, the data are 923 groups. The histograms of the data are shown in Figure 4.

$$Z^t\left(t_j, x_i\right) = \begin{cases} \left(t_a, x_i\right), & t_j = t_1, \\ z\left(t_{j-1}, x_i\right) + \dfrac{\Delta t}{t_{j-1} - t_a} z\left(t_a, x_i\right), & t_1 < t_j < t_p, \\ z\left(t_j, x_{i-1}\right), & t_j = t_p. \end{cases} \tag{5}$$

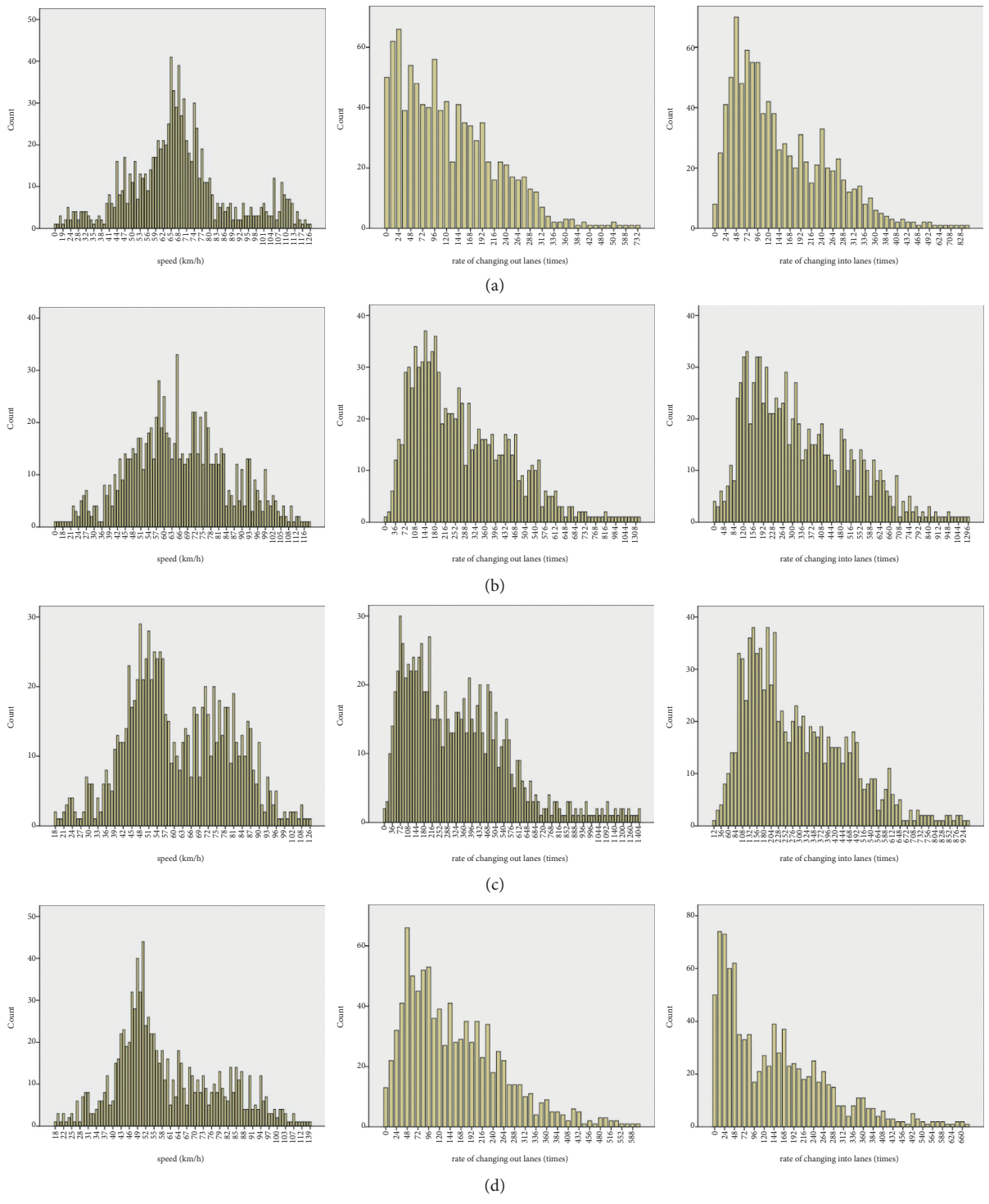


Figure 4: Data distribution histogram. (a) The histogram of Lane 1 speed, the rate of changing into lanes, and the rate of changing out of lanes. (b) The histogram of Lane 2 speed, the rate of changing into lanes, and the rate of changing out of lanes. (c) The histogram of Lane 3 speed, the rate of changing into lanes, and the rate of changing out of lanes. (d) The histogram of Lane 4 speed, the rate of changing into lanes, and the rate of changing out of lanes.

In the formula, Z is the interpolation in time t_j, location x_i; t_a is the first valid data in x_i location, t_1 is the initial moment; t_p is end time; and x_i is the place where interpolation needs to be performed.

In Figure 4, the data distribution is relatively discrete. We use histograms to describe the data characteristics. The speed is concentrated within 100 km/h. This is mainly due to the maximum speed limit (which is 100 km/h) on

the third ring expressway in Chengdu. The data distribution of lane change-in rate and lane change-out rate is slightly different in each lane. From the shape of the plot, the speed in each lane is similar, but the rate of changing out of lanes and the rate of changing into lanes in each lane are different. The shapes in Lane 2 and Lane 3 are similar, while the shapes in Lane 1 and Lane 4 are different. For speed, the speed limit is used in all lanes. For lane-changing behavior, Lane 1 is the lane farthest from the entrance ramp and exit ramp, Lane 4 is the lane connecting the entrance ramp and exit ramp, and Lane 2 and Lane 3 are the connecting lanes of Lane 1 and Lane 4. Therefore, the shapes of Lane 2 and Lane 3 are similar, while the shapes of Lane 1 and Lane 4 are different. In addition, Lane 1 is the lane less affected by the entrance ramp and exit ramp, and Lane 4 is the lane affected by the entrance ramp and exit ramp. These characteristics are shown in Figure 4. The discrete of the lane-changing rate in Lane 4 is more than that in Lane 1. The speed, the rate of changing out of lanes, and the rate of changing into lanes in each lane show partial normal distributions. This distribution is very common on the road. The data distribution shows that the data collected from videos are really effective. The collected data can be used for the establishment of the simulation model.

2.2. Simulation Model. In order to study the relationship between the lane-changing behavior and speed under different penetration rates of autonomous vehicles on urban expressways, we used the third ring road in Chengdu, China, as a benchmark and established a simulation model of the urban expressway. The simulation model takes the ring-shaped urban expressway as the object. In order to make the simulation model close to the real road, we used 9 road sections according to the previously collected data. By simplifying the road network, only the exits and entrances are retained. Finally, a simulation model is established.

Using sumo as the platform, the road section is set as a one-way 4 lane. In previous studies, we found that the lane-changing behavior has an impact on the speed, but there are a slight effect in the free flow and more effect in large traffic flow. In order to increase enough flow, each section has 4 entrances and 1 exit. The road section setting is shown in Figure 5.

In Figure 5, different types of vehicles are used in the simulation model. In order to better simulate the real traffic, the vehicle types are set to four types of large trucks, medium trucks, cars, and autonomous vehicles. The autonomous vehicle types are uniform in the model. We used the CACC model to simulate the autonomous vehicle and the Krauss model to simulate the human-driving vehicle. In the road network, each section contains 4 entrances and one exit. The distance between the ramps is 1 km, the simulation distance of each section is 5 km, and the simulation distance of the whole road is 45 km, which is close to the length of the third ring expressway of Chengdu (the length is 51 km). The simulation time is 86,400 s (that is 24 h). The simulation time period is 4 hours in the morning and 4 hours in the afternoon of the working day in large traffic flow. The simulation step is 0.1 s. The tranCI is used to add different traffic flow to different intersections of the road. The collection time interval is 5 minutes, and the collected data in each lane include the speed, the rate of changing out lanes, and the rate of changing into lanes. The groups of collected data are 2,591 groups.

In order to verify the applicability of the simulation model, the first simulation involved three vehicle types: large trucks, medium trucks, and cars. The distribution of simulation data is shown in Figure 6.

In Figure 6, the data distribution is relatively discrete and similar to the empirical data. The rate of changing into lanes and the rate of changing out of lanes are relatively similar. In addition, the shapes of the simulation data and the empirical data (Figure 4) have a slight difference. The main reason is that the empirical data contain data in small traffic flow. In the simulation model, we only simulated the large traffic flow. There is a certain gap between the simulated data and the empirical data. But the data distributions are partial normal distributions.

In order to verify the similarity between the simulation model and the real traffic state, we drew diagrams of the relationship between the speed and the rate of changing out of lanes, and the relationship between the speed and the rate of changing into lanes of simulation data and empirical data, as shown in Figure 7.

In Figure 7, the data of the plots in the left part are collected from empirical research, and the data of the plots in the right part are collected by the simulation model. From the form of data, the shape and aggregation state of the empirical data are similar to that of the simulation data, except for group number, which is collected in simulations is more than that in empirical research. That is, from the view of data relationships, the data representation has a certain degree of similarity in empirical and simulation methods. Combining the histogram graphs and the plots of the relationships, we believe that the established simulation model can be used to study the relationship between lane-changing behavior and speed under different autonomous vehicle penetration rates.

2.3. Experiment

2.3.1. Experimental Design. Through the established simulation model, the experiment is designed to study the relationship between lane-changing behavior and speed in a circular traffic network under different autonomous vehicle penetration rates. Due to the addition of autonomous vehicles, the lane-changing behavior of vehicles is slightly different from that of human-driving vehicles. We would like to know whether different penetration rates of autonomous vehicles have an impact on the speed, the rate of changing out of lanes, and the rate of changing into lanes in the traffic. Through the data characteristics, we would like to know the relationship between speed and lane-changing behavior in mixed traffic. In the experiments, car-following behavior is mostly adopted, unless there is an obstacle in front of the lane-changing behavior must be performed.

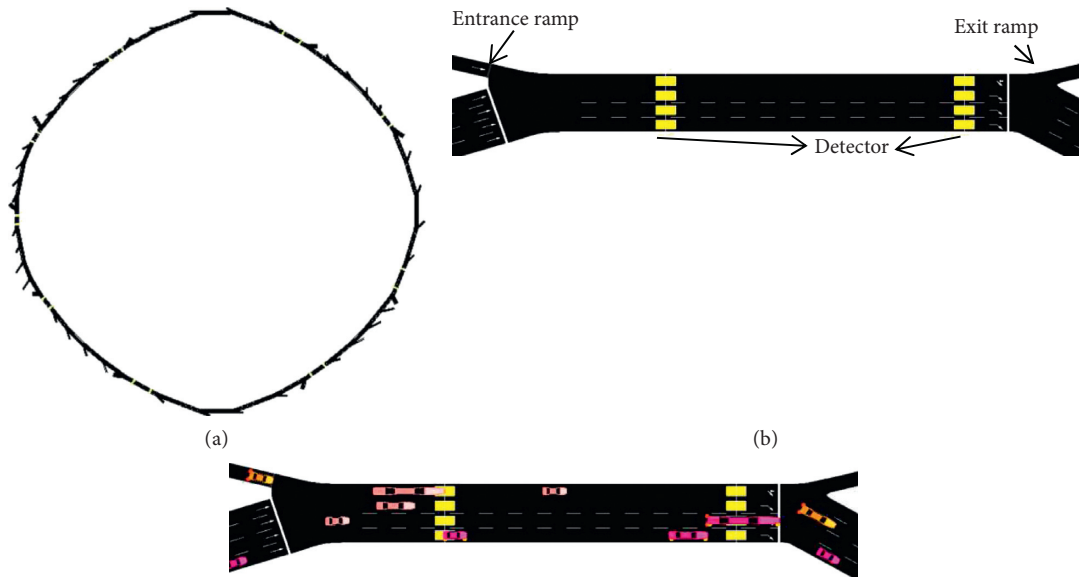


Figure 5: Schematic diagram of simulation modeling. (a) Road network model of 9 sections established by simulation. (b) Schematic diagram of the road section including the detector. (c) A road running diagram with one exit, one entrance, and two detectors.

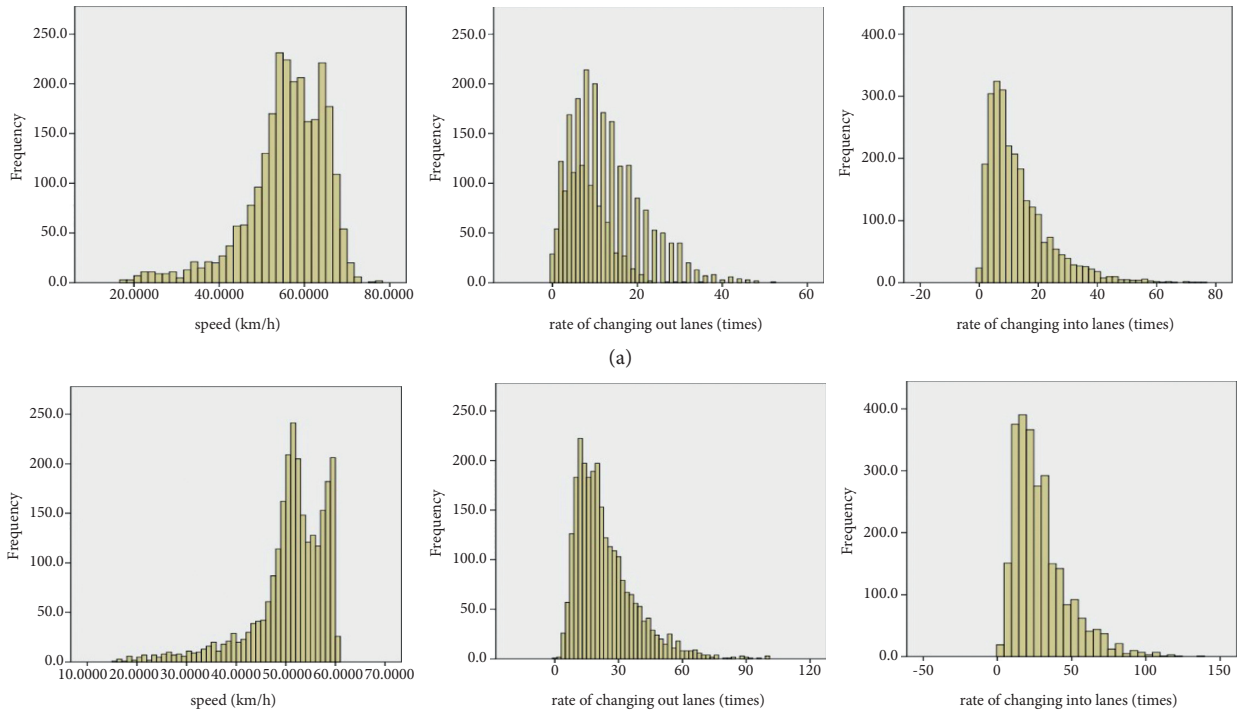


Figure 6: Continued.

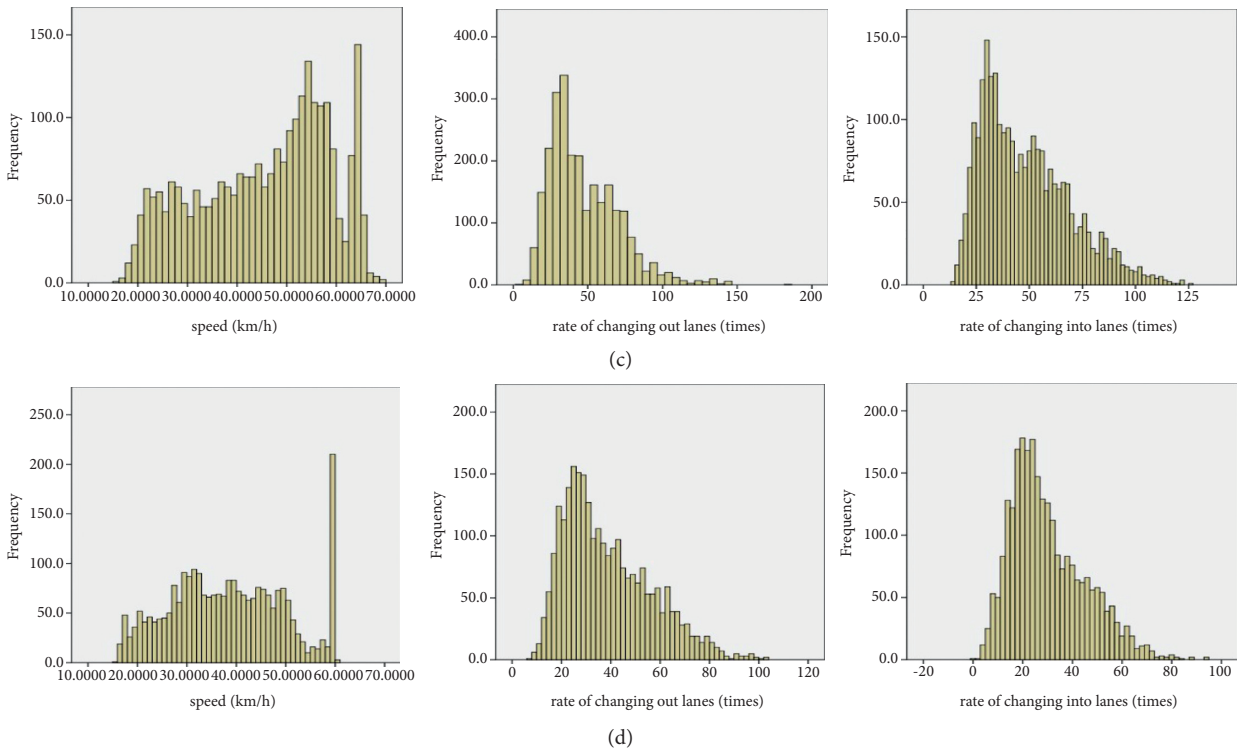


Figure 6: Simulation data distribution histogram. (a) The histogram of Lane 1 speed, the rate of changing into lanes, and the rate of changing out of lanes. (b) The histogram of Lane 2 speed, the rate of changing into lanes, and the rate of changing out of lanes. (c) The histogram of Lane 3 speed, the rate of changing into lanes, and the rate of changing out of lanes. (d) The histogram of Lane 4 speed, the rate of changing into lanes, and the rate of changing out of lanes.

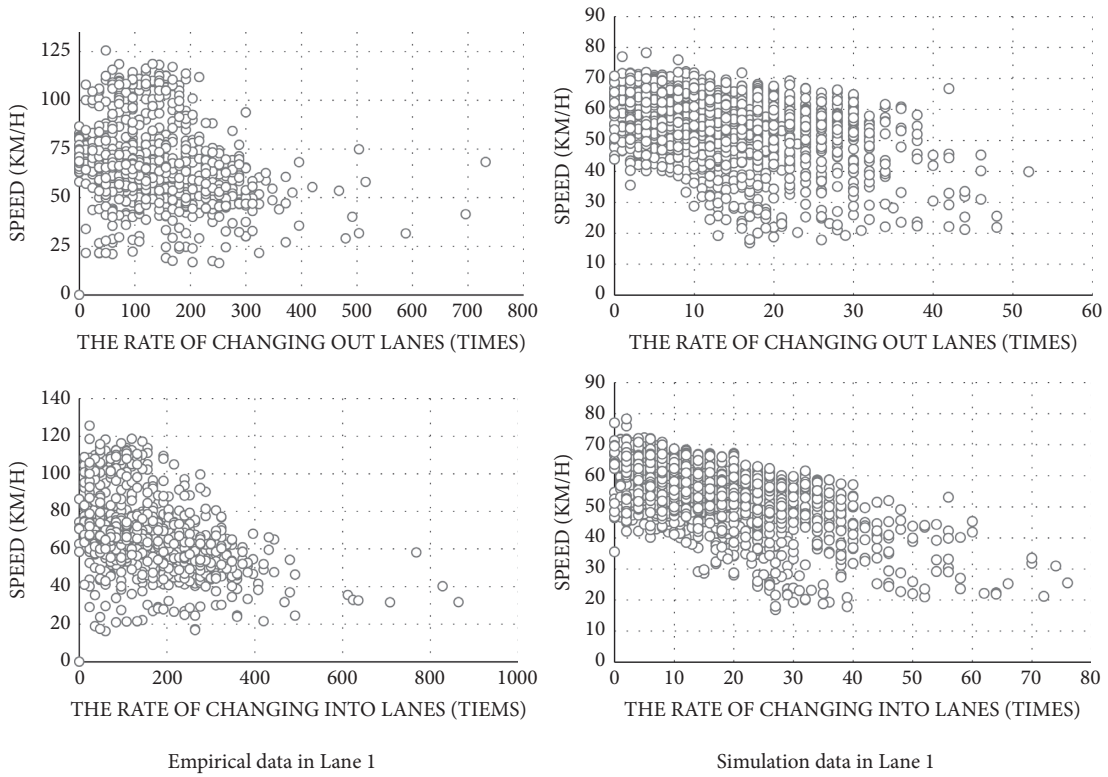
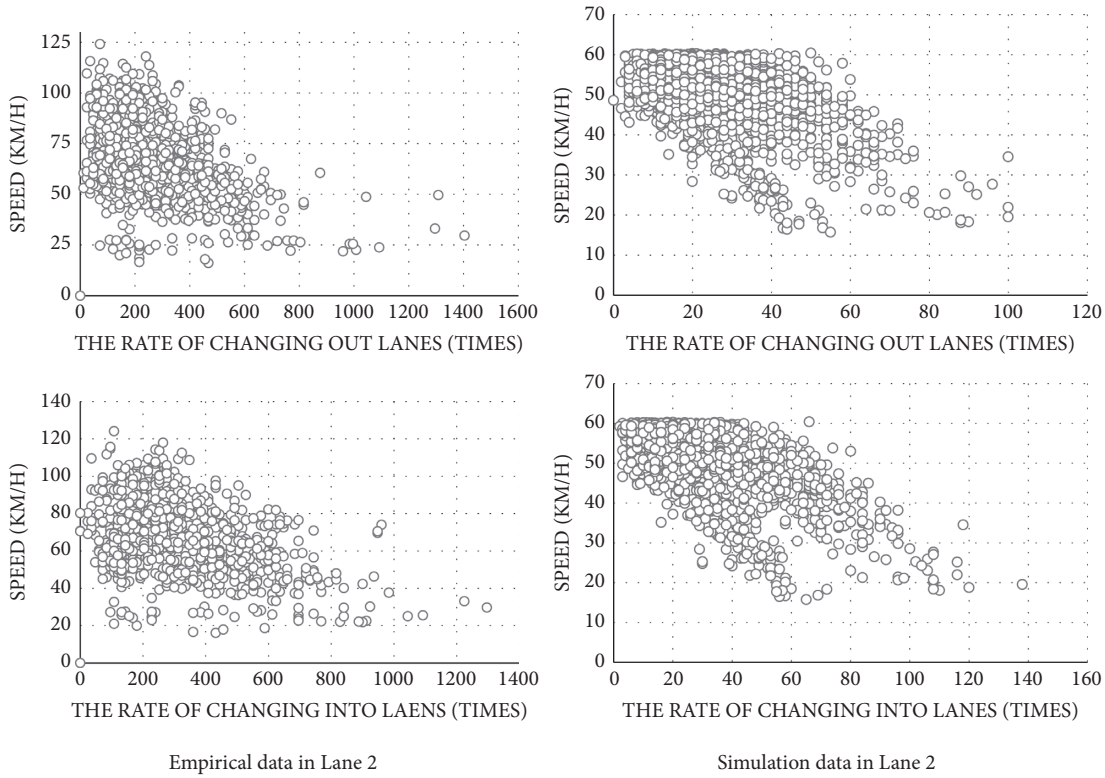
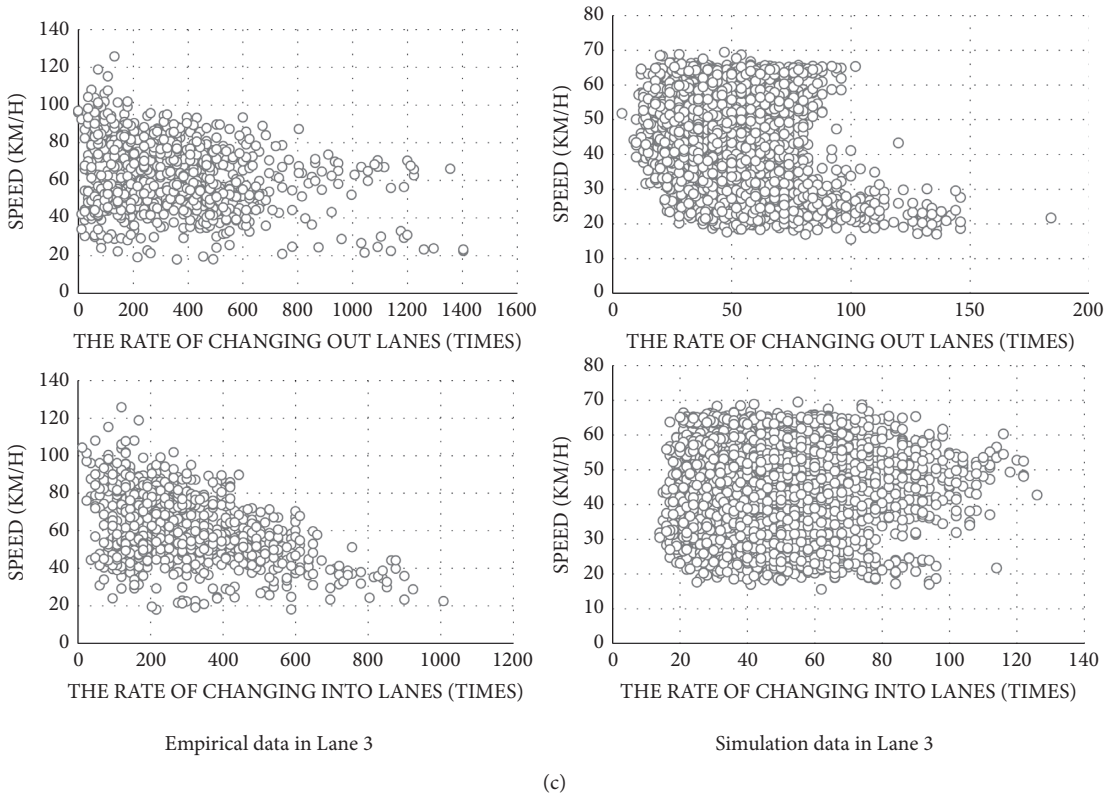


Figure 7: Continued.



Empirical data in Lane 2 Simulation data in Lane 2

(b)



Empirical data in Lane 3 Simulation data in Lane 3

(c)

Figure 7: Continued.

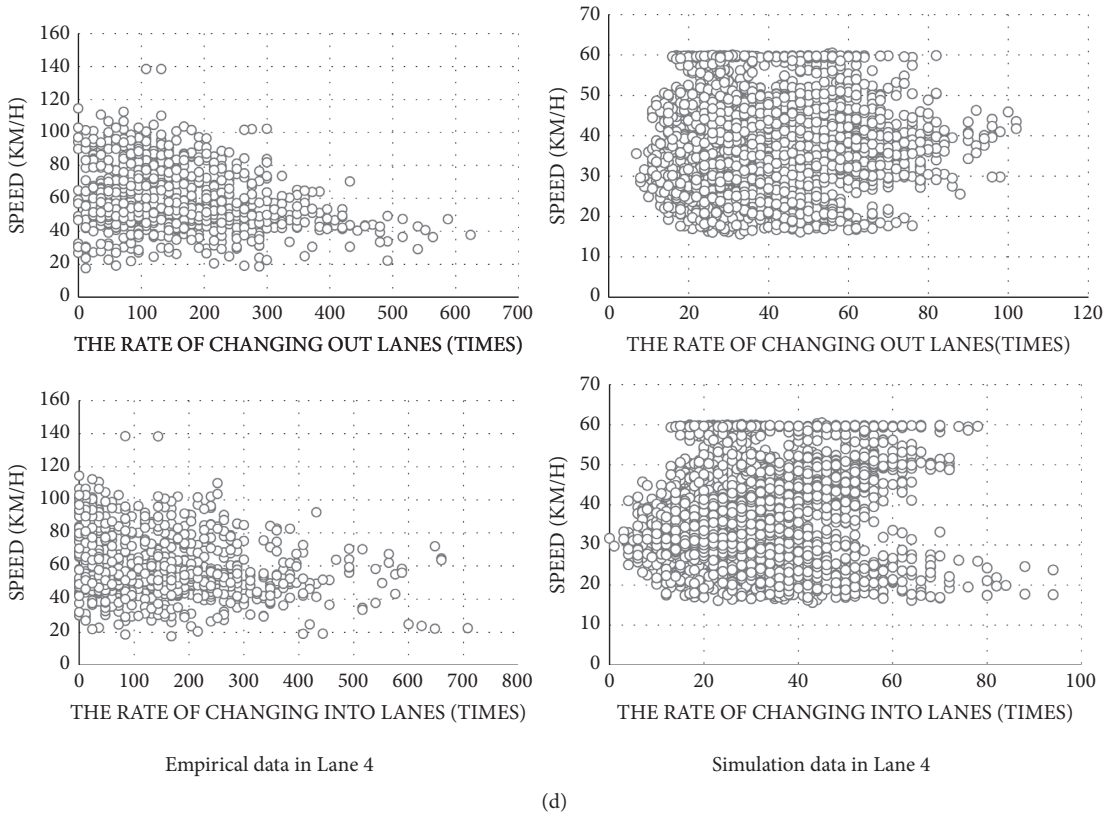


Figure 7: A scatter diagram of the relationship between the speed and the rate of changing out of lanes, and the relationship between the speed and rate of changing into lanes in each lane in the empirical data and simulation data. (a) The scatter plot of Lane 1. (b) The scatter plot of Lane 2. (c) The scatter plot of Lane 3. (d) The scatter plot of Lane 4.

Therefore, in the simulation process, CACC vehicles are used to simulate autonomous vehicles. The human-driving vehicles used the DK2008 model to change lanes. The simulation environment setting is consistent with the first simulation.

The experimental steps are as follows:

Step1. Different proportions of CACC vehicles (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%, respectively) are added to the ring road for simulation experiments, using 86,400 s as the simulation time and step length of 0.1 s. The collection time interval is 5 min. The data of the speed, the rate of changing out of lanes, and the rate of changing into lanes are collected.

Step2. According to statistical data, the relationship between the speed and the rate of changing out of lanes, and between the speed and the rate of changing into lanes under different penetration rates of autonomous vehicles are analyzed.

Step3. The final conclusion is drawn based on the previous analysis.

2.3.2. Experimental Results. Through the design of the experiment, we collected the data of speed, the rate of changing into lanes, and the rate of changing out of lanes by the smooth operation of the experiment. In order to better analyze the relationship between speed and lane change behavior, the characteristics of the collected data are analyzed to verify the accuracy and effectiveness of the data. The data analysis can be shown in the statistical description data in Tables 2~4.

In Table 2, the data representations of Lane 1 and Lane 2 are relatively close. The average speed slightly decreases with the increase in the penetration rate of autonomous vehicles. When the penetration rates are 10% and 30%, the average speed drops by 0.1%, respectively. When the penetration rate increases to 40%, the average speed is increased by 5%, which is compared with the penetration rate of 30%. Compared with no autonomous vehicle penetration, the average speed of Lane 1 is increased by 2%, and of Lane 2 is increased by 4%. When the penetration rate increases to 90%, the average speed increases by 4% compared to the penetration rate of 40%, by 6% in Lane1, and by 8% in Lane2 compared to no autonomous vehicle penetration. However, when the penetration rates increase, the median speed slightly increases, and the speed mode does not change much. About 75% of speed has decreased, and 25% of speed has increased. Those representations show that with the continuous penetration of autonomous vehicles, the speed tends to converge toward the speed mode, and the median speed gradually moves closer to the speed mode. But it is still different from the average speed, and it is still non-normally distributed. Lane 1

Table 2: The statistical description table of the speed of each lane under different autonomous vehicle penetration rates.

		p=0%	p=10%	p=20%	p=30%	p=40%	p=50%	p=60%	p=70%	p=80%	p=90%
Lane1	Mean value	55.85	55.45	55.08	54.046	54.47	54.630	57.20	58.745	59.72	60.54
	Standard error	0.182	0.190	0.191	0.203	0.193	0.185	0.130	0.104	0.083	0.075
	Median	56.98	57.18	57.37	56.735	57.29	57.235	58.6	59.46	60.04	60.44
	Mode	64.95	58.64	64.83	64.405	61.33	61.73	60.43	61.615	61.96	60.99
	Variance	85.90	9.706	95.13	106.96	96.56	89.340	44.19	28.32	17.99	14.92
	75%	70.26	94.20	68.89	68.645	68.09	67.605	67.91	68.5	68.59	69.56
	25%	23.49	69.68	24.26	23.135	23.90	25.825	34.99	40.35	45.65	47.58
Lane2	Mean value	50.85	23.69	50.52	49.80	50.30	50.44	52.98	54.27	55.14	55.90
	Standard error	0.151	50.63	0.164	0.176	0.166	0.159	0.104	0.079	0.059	0.048
	Median	51.94	0.164	52.64	52.36	52.595	52.6	54.03	54.93	55.62	56.25
	Mode	52.47	52.4	54.44	52.05	49.985	56.495	54.79	55.71	56.89	56.74
	Variance	59.08	51.92	70.37	80.65	71.64	65.79	28.37	16.43	9.139	6.029
	75%	60.00	8.194	59.86	59.805	59.705	59.73	59.75	59.845	59.77	59.93
	25%	21.96	67.19	21.90	21.585	22.985	24.58	33.29	38.84	45.60	48.53
Lane3	Mean value	45.59	59.94	39.96	36.84	35.66	35.22	35.69	38.31	40.25	42.80
	Standard error	0.258	21.54	0.239	0.221	0.203	0.204	0.181	0.1726	0.154	0.123
	Median	48.1	42.91	39.68	36.08	35.1	34.065	35.54	39.055	41.07	43.29
	Mode	54.55	0.256	30.47	29.855	36.135	38.845	36.315	42.57	42.80	45.11
	Variance	173.4	44.09	148.7	127.13	107.05	108.04	85.46	76.67	61.55	39.78
	75%	65.43	24.91	64.62	63.25	59.82	59.765	58.005	57.65	57.37	56.30
	25%	19.26	13.03	19.10	18.39	18.5	18.52	18.825	20.145	21.65	25.77
Lane4	Mean value	38.46	169.8	33.81	31.24	30.14	29.618	29.05	31.01	32.97	36.65
	Standard error	0.228	65.65	0.211	0.198	0.186	0.189	0.184	0.191	0.186	0.164
	Median	38.05	18.67	32.98	30.235	29.29	28.14	28.125	30.975	33.88	38.82
	Mode	59.60	36.30	44.85	22.36	20.56	31.455	35.215	18.945	40.19	42.60
	Variance	135.0	0.226	116.2	101.62	90.056	92.974	88.243	94.86	89.75	70.30
	75%	59.79	35.85	59.5	57.385	53.73	54.195	51.11	50.155	50.76	51.16
	25%	17.16	31.83	16.35	15.215	15.2	14.545	13.41	12.995	13.22	16.26

Table 3: The statistical description table of the rate of changing out of lanes of each lane under different autonomous vehicle penetration rates.

		p=0%	p=10%	p=20%	p=30%	p=40%	p=50%	p=60%	p=70%	p=80%	p=90%
Lane1	Mean value	11.76	8.07	8.151	8.31	8.176	8.08	6.852	5.92	5.5	4.81
	Standard error	0.158	0.087	0.083	0.082	0.076	0.076	0.066	0.059	0.056	0.052
	Median	10	8	8	8	8	8	6	6	5	4
	Mode	8	6	6	6	6	6	5	5	4	4
	Variance	65.10	4.445	17.88	17.52	15.04	15.03	11.60	9.02	8.242	7.08
	75%	38	21	20	22	20	20	16	15	14	13
	25%	0	1	1	2	1	2	1	1	1	0
Lane2	Mean value	23.35	16.11	16.73	17.64	17.65	17.98	17.25	16.06	15.28	14.55
	Standard error	0.280	0.160	0.162	0.181	0.166	0.174	0.166	0.157	0.143	0.132
	Median	20	14	15	16	16	16	15	15	14	13
	Mode	18	14	13	12	14	13	12	11	11	12
	Variance	203.5	8.169	68.341	84.92	71.40	78.79	71.99	63.98	53.22	45.29
	75%	70	44	44	49	46	47	44	43	39	36
	25%	4	4	4	5	5	4	4	4	3	3
Lane3	Mean value	47.30	34.71	37.64	41.20	42.63	44.01	47.12	46.88	45.84	44.11
	Standard error	0.458	0.241	0.260	0.298	0.291	0.309	0.334	0.339	0.3165	0.280
	Median	42	32	35	38	40	41	44	43	43	42
	Mode	36	30	33	32	32	37	42	36	41	40
	Variance	543.7	12.27	175.27	230.9	218.80	247.8	290.7	297.9	259.27	203.44
	75%	126	77	83	89	91	93	98	99	94	87
	25%	14	16	17	18	19	19	20	20	20	19
Lane4	Mean value	38.01	25.35	25.28	26.01	26.57	26.88	27.16	25.51	23.64	20.22
	Standard error	0.342	0.148	0.148	0.152	0.163	0.164	0.173	0.180	0.173	0.156
	Median	34	24	24	25	26	26	26	25	23	19
	Mode	24	22	22	22	22	24	25	20	19	18
	Variance	304.2	7.575	56.978	59.91	69.539	70.16	78.22	83.96	77.67	63.738
	75%	86	46	45	46	49	49	51	51	46	41
	25%	12	11	11	11	11	11	11	9	8	6

Table 4: The statistical description table of the rate of changing into lanes of each lane under different autonomous vehicle penetration rates.

		p=0%	p=10%	p=20%	p=30%	p=40%	p=50%	p=60%	p=70%	p=80%	p=90%
Lane1	Mean value	13.05	9.231	9.814	10.50	10.543	10.651	9.592	8.370	7.538	6.872
	Standard error	0.212	0.127	0.125	0.130	0.123	0.121	0.104	0.091	0.076	0.065
	Median	10	7	8	9	9	9	9	8	7	6
	Mode	8	5	7	6	7	6	8	6	5	6
	Variance	116.7	6.480	41.001	44.08	39.59	38.07	28.26	21.67	15.17	11.08
	75%	52	31	31	32	31	30	26	24	20	16
	25%	1	1	1	1	1	1	1	1	1	1
Lane2	Mean value	29.15	20.61	21.63	23.15	23.64	24.01	23.55	22.05	20.72	19.00
	Standard error	0.363	0.207	0.202	0.221	0.210	0.215	0.214	0.205	0.186	0.162
	Median	24	18	20	21	21	22	21	20	19	18
	Mode	28	15	15	16	17	16	20	15	18	15
	Variance	342.3	10.53	106.62	126.7	114.95	120.43	118.8	109.86	89.82	67.95
	75%	94	55	54	60	56	57	56	55	49	43
	25%	5	5	6	6	6	7	6	5	5	4
Lane3	Mean value	48.31	32.23	32.21	33.18	33.68	34.22	34.82	33.20	31.38	27.90
	Standard error	0.414	0.171	0.173	0.186	0.194	0.201	0.216	0.228	0.218	0.198
	Median	44	31	31	32	32	33	34	32	30	27
	Mode	52	29	28	25	28	32	34	32	27	23
	Variance	445.6	8.72	77.707	90.13	97.904	104.14	121.5	135.56	123.7	101.79
	75%	108	55	55	59	62	62	65	66	62	55
	25%	17	16	16	16	16	15	15	12	12	9
Lane4	Mean value	29.90	22.185	24.15	26.33	27.17	28.080	30.42	30.762	30.62	29.928
	Standard error	0.298	0.142	0.152	0.167	0.163	0.177	0.195	0.202	0.199	0.194
	Median	26	21	23	25	26	26	29	29	29	28
	Mode	24	20	21	23	24	25	25	23	25	24
	Variance	231.6	7.25	60.01	72.33	69.06	81.36	99.01	105.68	102.8	97.76
	75%	72	46	51	53	53	56	61	62	61	61
	25%	6	9	11	12	13	13	14	13	13	14

and Lane 2 belong to the faster lanes. When autonomous vehicles enter the road, there will have a certain impact on the road speed. This is a slight downward trend. With the increase in penetration of autonomous vehicles, the average speed of the lanes will have a significant increase when the penetration rate is over 40%, compared with no autonomous vehicles on the road. Lane 3 and Lane 4 are close to the ramps and have the influence of entrances and exits, so there will have a certain impact on the speed after the autonomous vehicle running on the road. In Table 2, the average speed of Lane 3 drops by 12% under the 10% penetration rate of autonomous vehicles. When the penetration rates of autonomous vehicles continue to increase, the average speed still shows a downward trend. When the penetration rate reaches 40%, the average speed has rebounded. When the penetration rate reaches 90%, the average speed has risen by 12.7% compared to 40% of the autonomous vehicle penetration rate, but it is still 11% lower than the penetration rate of 0%. The median speed slightly increased, the speed mode did not change much, the 75% speed did not change much, and the 25% speed increased. It shows that with the penetration of autonomous vehicles in Lane3, the speed is concentrated in the speed mode, and the median speed and the speed mode have a similar trend with the increase in penetration rates, but there is still a gap with the average speed. So the speed is still non-normally distributed in Lane 3. However, compared with 0% of the penetration rate of autonomous vehicles, the average speed of Lane 3 has decreased. This is mainly because Lane 3 is responsible for a large number of lane change behaviors of vehicles from Lane 1 and Lane 2 that need to exit the road. So the lane changing behavior effect the average speed of Lane 3. Lane 4 is the lane directly connected to the entrances and exits. The average speed is most affected by the behavior of changing lanes. With the increased penetration rates of autonomous vehicles, the average speed decreases by 12% when the penetration rate of autonomous vehicles is 10%. Then, the speed of decline increases with the increase in autonomous vehicle penetration rates. However, the average speed of Lane 4 will not increase until the penetration rate of autonomous vehicles is increased to 40%. Although the average speed has increased when the penetration rate is increased to 90%, it is still 14% lower than the penetration rate of 0%. The median speed and speed mode slightly changed with the average speed but did not change much. The median speed and speed mode did not show a close trend due to the increase in the penetration rates of autonomous vehicles. The 75% speed and 25% speed slightly decrease with the increase in penetration rates, but the speed still converges toward the speed mode. Taking a comprehensive of the speeds in Table 2 from Lane 1 to Lane 4, with the increase in penetration rates of autonomous vehicles, the average speed shows a trend of decline at the first and then rise up. The changes in the median speed and speed mode are more obvious in Lane 1 and Lane 2 and are not obvious in Lane 3 and Lane 4. But the median speed and speed mode have a trend of approaching

as the penetration rates of autonomous vehicles increase, except for data in Lane 4. The 75% speed has decreased, and the 25% speed has increased except for the decrease in Lane 4. However, it shows that the speed is converging toward the speed mode as the penetration rates of autonomous vehicles increase.

The impact of autonomous vehicles on speed can be seen in Table 2. Different lanes play different roles in traffic. The lane-changing behavior is most affected in Lane 3 and Lane 4, especially Lane 4, which directly affects speed. It can be clearly seen in Figure 8 that the addition of autonomous vehicles has a greater impact on lane-changing behavior. It can also be proved from the statistical data in Tables 3 and 4.

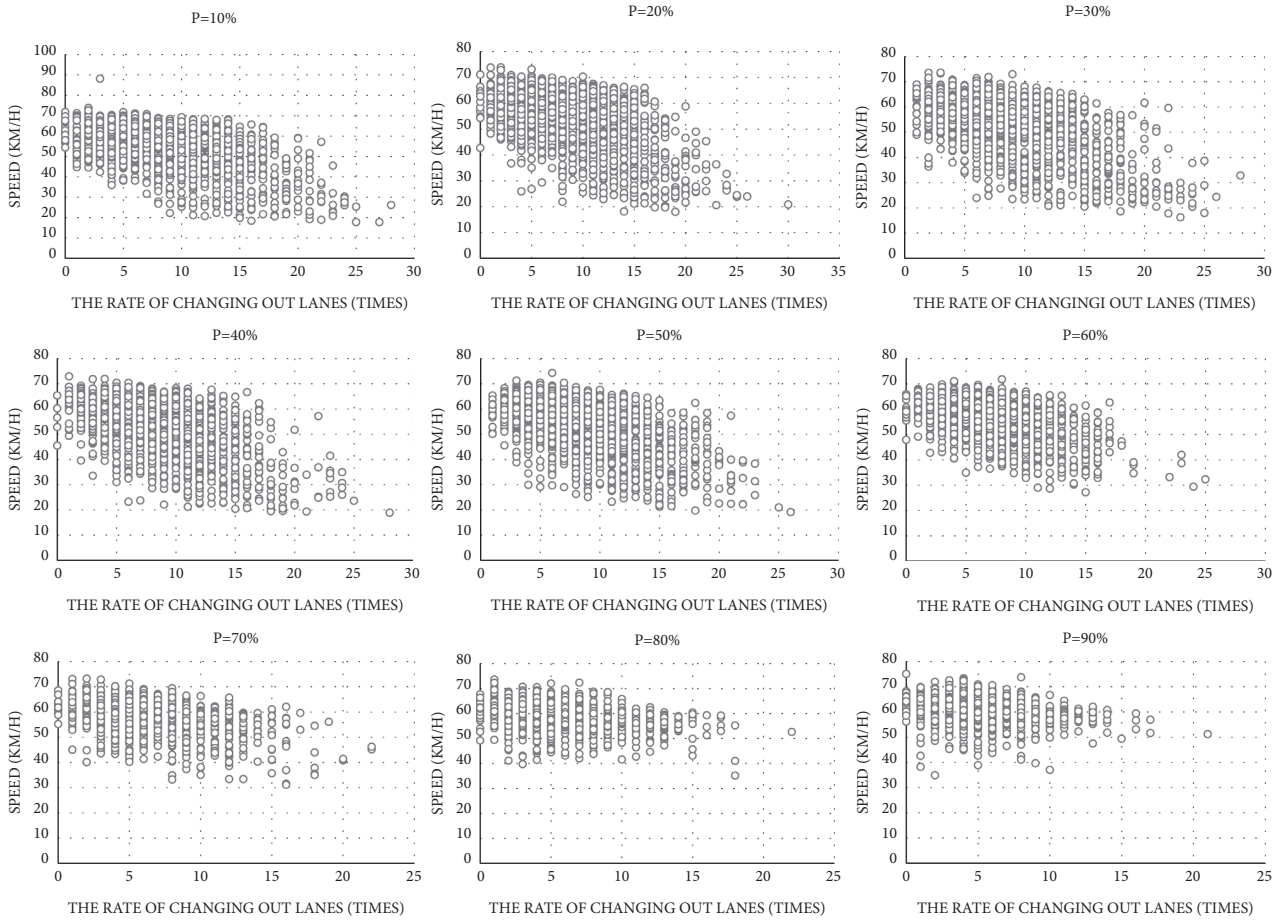
In Table 3, as the penetration rates of autonomous vehicles continue to increase, the average rate of changing out of lanes does not significantly change and has declined compared with 0% penetration rate. Except for Lane 4, the median and mode of the rate of changing out of lanes tend to approach with the increase in penetration rates of autonomous vehicles, but there are still differences from the average rate of changing out of lanes, and the distribution is still non-normal. The rate of changing out of lanes in the 75% has decreased with the increase in penetration rates of autonomous vehicles, except for the significant decrease under the penetration rate of 10%. But in the 25%, the increases with the increase in penetration rates are not significant changes. It can be seen from the data changes of the rate of changing out of lanes that the increases in autonomous vehicles have a significant impact on multilane traffic, but as the penetration rates increase, the impact is not significant. In Table 4, this situation still occurs at the rate of changing into lanes. When the penetration rate of autonomous vehicles is 10%, the changes in the average rate of changing into lanes are not obvious in Lane 4 and are obvious in the other three lanes. The decline is above 10% in the other three lanes. However, as the penetration rates of autonomous vehicles increase, the trend of changes is not obvious. Except for Lane 4, the median and mode of the rate of changing into lanes tend to approach with the increase in penetration rates of autonomous vehicles, but there are differences from the average rate of changing into lanes. The change in the median and mode of the rate of changing into lanes in Lane 4 is not obvious. The 75% of the rate of changing into lanes has no significant change with the increase in the penetration rates of autonomous vehicles, except for a significant drop in the penetration rate of 10%. However, the changes in the 25% of the rate of changing into lanes in all lanes are not obvious. On the whole, the rate of changing into lanes and the rate of changing out of lanes tend to increase in Lane 1~Lane 3, while the data in Lane 4 slightly decrease. Lane 3, as the busiest lane on the road, is mainly responsible for vehicles, which enter or exit the road. With the increase in the penetration rates of autonomous vehicles, the rate of changing into lanes and the rate of changing out of lanes have declined, but most lane change behaviors still happened in Lane 3. At the same time, the lane change-in rate and lane change-out rate of Lane 2, Lane 3, and Lane 4 have a similar trend. The influence of lane-changing behavior in Lane 1 is relatively small because of the lane far away from the entrance and exit. With the increase in the penetration rates of autonomous vehicles, the rate of changing into lanes and the rate of changing out of lanes in Lane 1 decreased, but the decline is not very obvious.

Considering data in Tables 2–4, we found that on a circular closed 4-lane road with a large traffic flow in mixed traffic, the speed presents an obvious trend of first decline and then rise up. In different lanes, the different penetration rates of autonomous vehicles will have different effects on speed. The speed of the lane far away from the entrance and exit has increased when the penetration rate of autonomous vehicles reaches 40%, compared with the penetration rate of 0%. The speed of the lane closer to the entrance and exit is not significantly increased or even slightly decreased. At the same time, the rate of changing out of lanes and the rate of changing into lanes significantly decreased as the penetration rate of autonomous vehicles reached 10% and then did not significantly decrease as the penetration rates increased. Lane 3 is still the busiest lane among the four lanes, and the lane-changing behavior of Lane 4 does not significantly respond to the addition of autonomous vehicles. This phenomenon is consistent with the speed changes. At the same time, it shows that after the penetration rate of autonomous vehicles reaches 10%, the penetration rate has little effect on the lane-changing behavior of each lane. With the increase in the penetration rates of autonomous vehicles, the speed of the lane far away from the entrance and exit has decreased before reaching 30% penetration rate and has increased after the penetration rate exceeds 40%, compared with 0% penetration rate. For the lane close to the entrance and exit, lane-changing behavior is still the main factor that affects the speed in each lane. A greater impact on the rate of changing into lanes and the rate of changing out of lanes only occurs when the penetration rate is 10%. This situation shows that autonomous vehicles running on the road take a limited impact on lane-changing behavior. With the increase in the penetration rates of autonomous vehicles, the average speed of the lanes has decreased compared with 0% penetration rate. It shows that for the lane close to the entrance and exit, the autonomous vehicles actually affect the average speed of the lane. However, all data show that as the penetration rates of autonomous vehicles increase, there is a certain degree of convergence in the speed, the rate of changing into lanes, and the rate of changing out of lanes in each lane.

Through the above analysis, we believe that the collected data of the speed, the rate of changing out of lanes, and the rate of changing into lanes have certain changes and stability. The simulation model is valid. So the data can be used to study the relationship between speed and lane change rate. But the relationship requires further research and analysis.

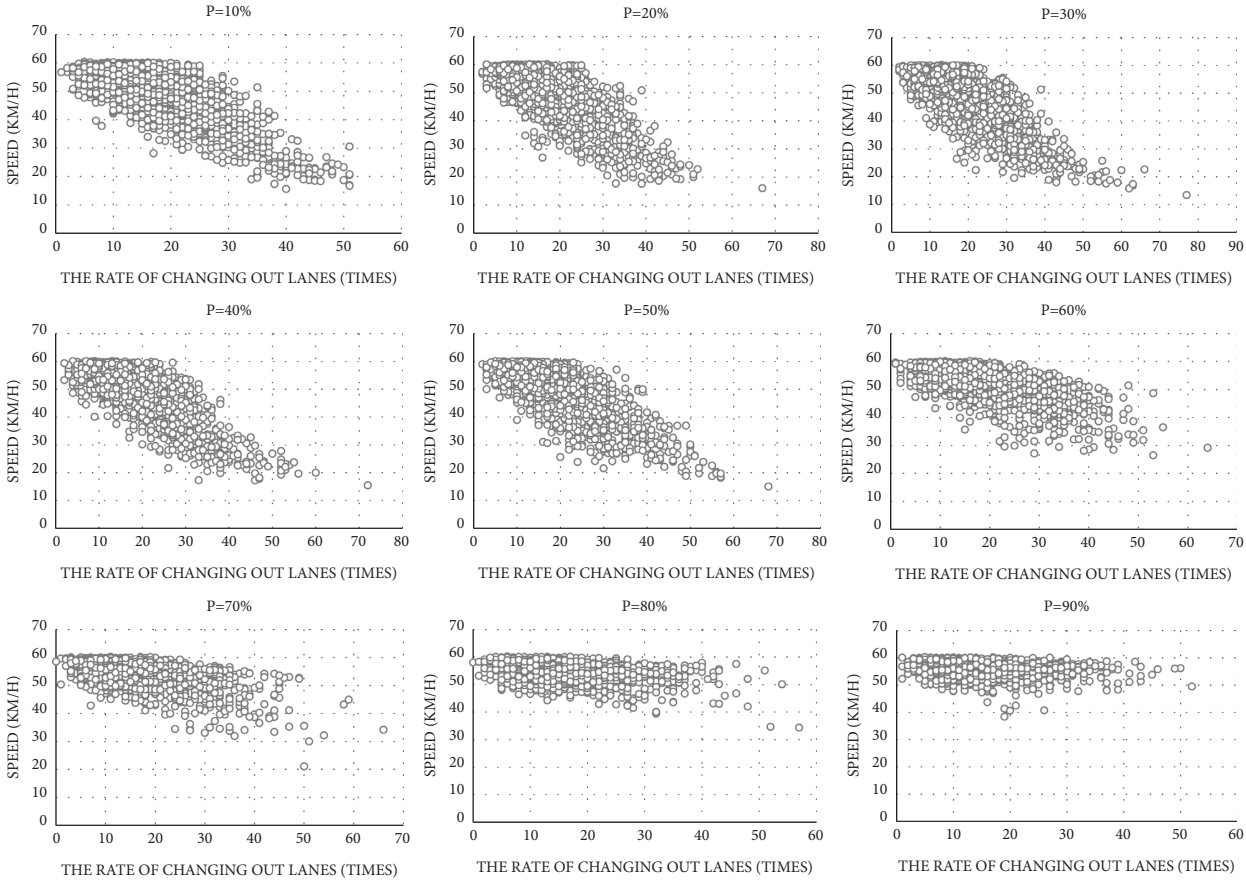
3. Results and Conclusions

3.1. Analysis of the Relationship between Speed and Lane Change Rate in Mixed Traffic. After the previous analysis, we found that we can study the relationship between speed and the rate of changing out of lanes, and the relationship between speed and the rate of changing into lanes to research



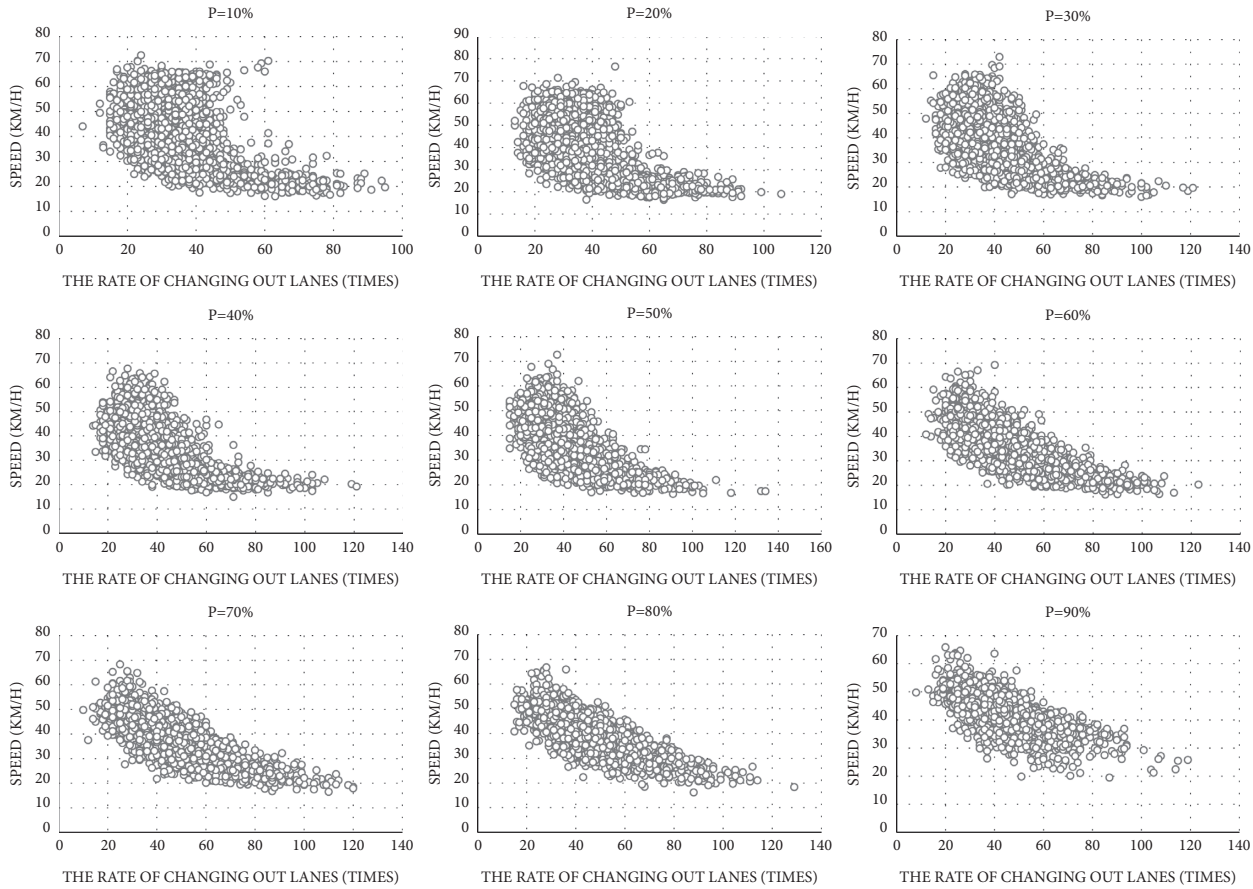
(a)

Figure 8: Continued.



(b)

Figure 8: Continued.



(c)

Figure 8: Continued.

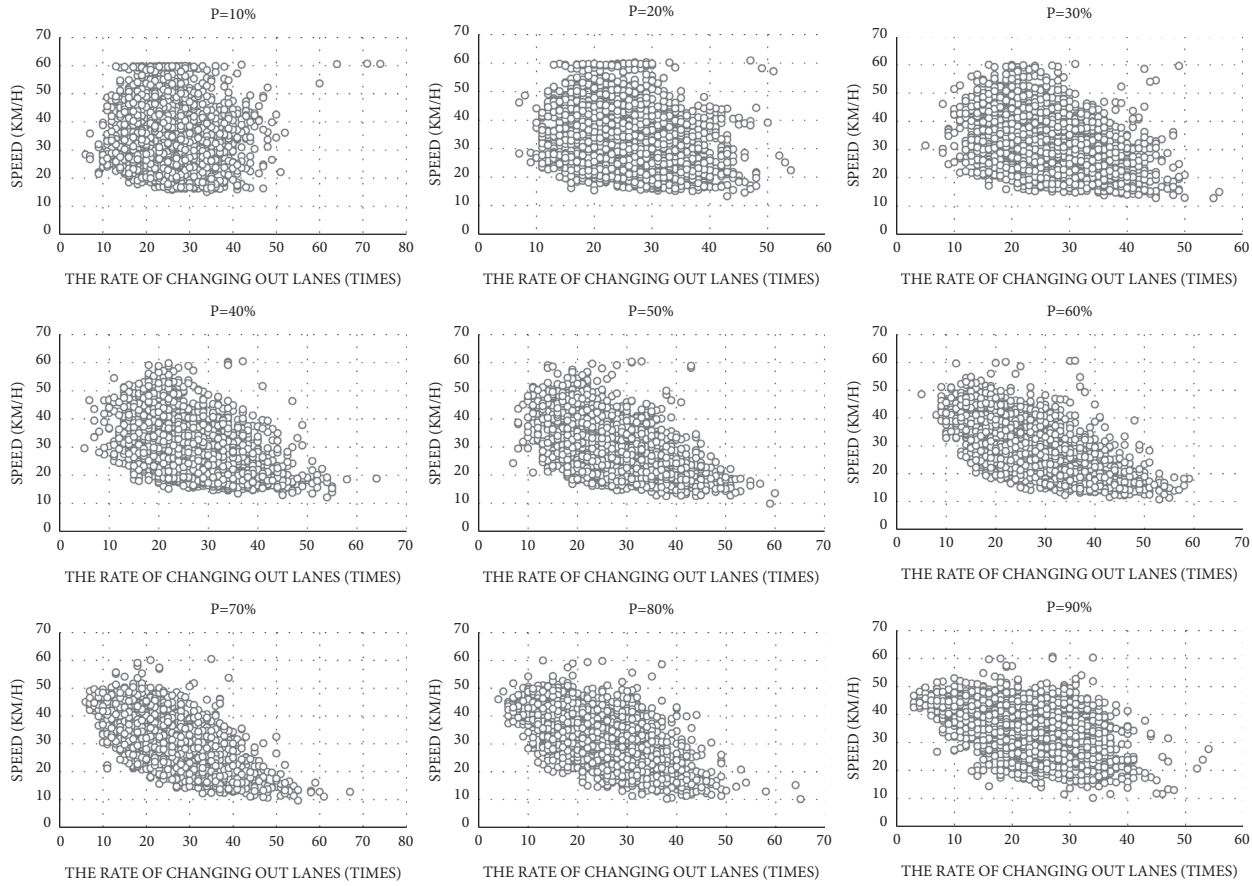


Figure 8: Scatter plot of the relationship between speed and the rate of changing out of lanes. (a) A scatter plot of Lane 1. (b) A scatter plot of Lane 2. (c) A scatter plot of Lane 3. (d) A scatter plot of Lane 4.

the relationship between speed and lane change rate. Since autonomous vehicles have different effects on the speed in each lane, the relationships between the speed and the lane change rate under different autonomous vehicles penetration rates need further analysis.

To verify there is a linear correlation between the speed and the lane change rate, the correlation analysis between the speed and the lane change rate in each lane under the different penetration rates of autonomous vehicles is shown in Table 5.

In Table 5, with the increase in the penetration rates of autonomous vehicles, the correlation in the speed and the rate of changing out of lanes, and the correlation in the speed and the rate of changing into lanes in Lane 1 and Lane 2 decrease, and the correlations in Lane 3 and Lane 4 enhanced. The relationship between speed and the rate of changing into lanes, and the relationship between speed and the rate of changing out of lanes can be observed from the scatter plot of the relationships, as shown in Figures 8 and 9.

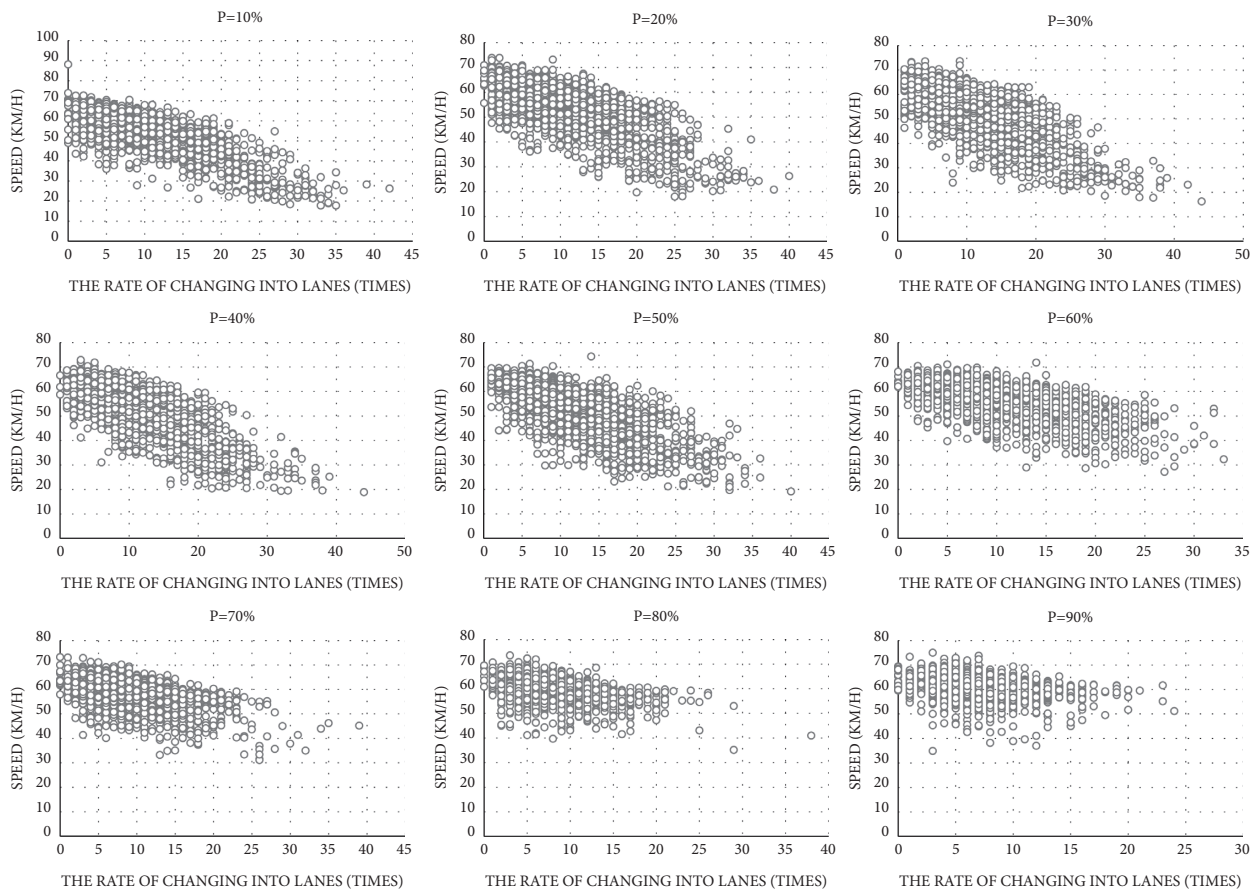
In Figures 8 and9, with the increase in the penetration rates of autonomous vehicles, the relationship between speed and the rate of changing out of lanes, and the relationship between speed and the rate of changing into lanes are closer in the same lane, but the relationships in different lanes are shown slightly different in the shape. Combined with Table 5, the relationships in Lane 1 and Lane 2 show a higher level of linearity when the penetration rate of autonomous vehicles is 10%, and the relationships in Lane 3 and Lane 4 show a higher level of linearity when the penetration rate of autonomous vehicles is 80%. For Lane 3 and Lane 4, with the increase in the penetration rates of autonomous vehicles, the speed eventually decreases. According to the analysis of data distribution in Part 2 of this study, the speed, the rate of changing out of lanes, and the rate of changing into lanes present a partial normal distribution. Considering the analysis of the relationship between the speed and the rate of changing out of lanes, and the relationship between the speed and the rate of changing into lanes in Part 3 of this study, there are certain linear correlations between the speed and lane change rate. So regression analysis can be used to analyze the relationship between speed and lane change behavior.

3.2. Speed-Lane Change Rate Regression Model. In Figures 8, 9, and Table 5, there is a certain linear relationship between the speed V and the lane change rate. Regression analysis can be used to describe the relationship.

Table 5: The linear correlation table between V, LCi, and LCo.

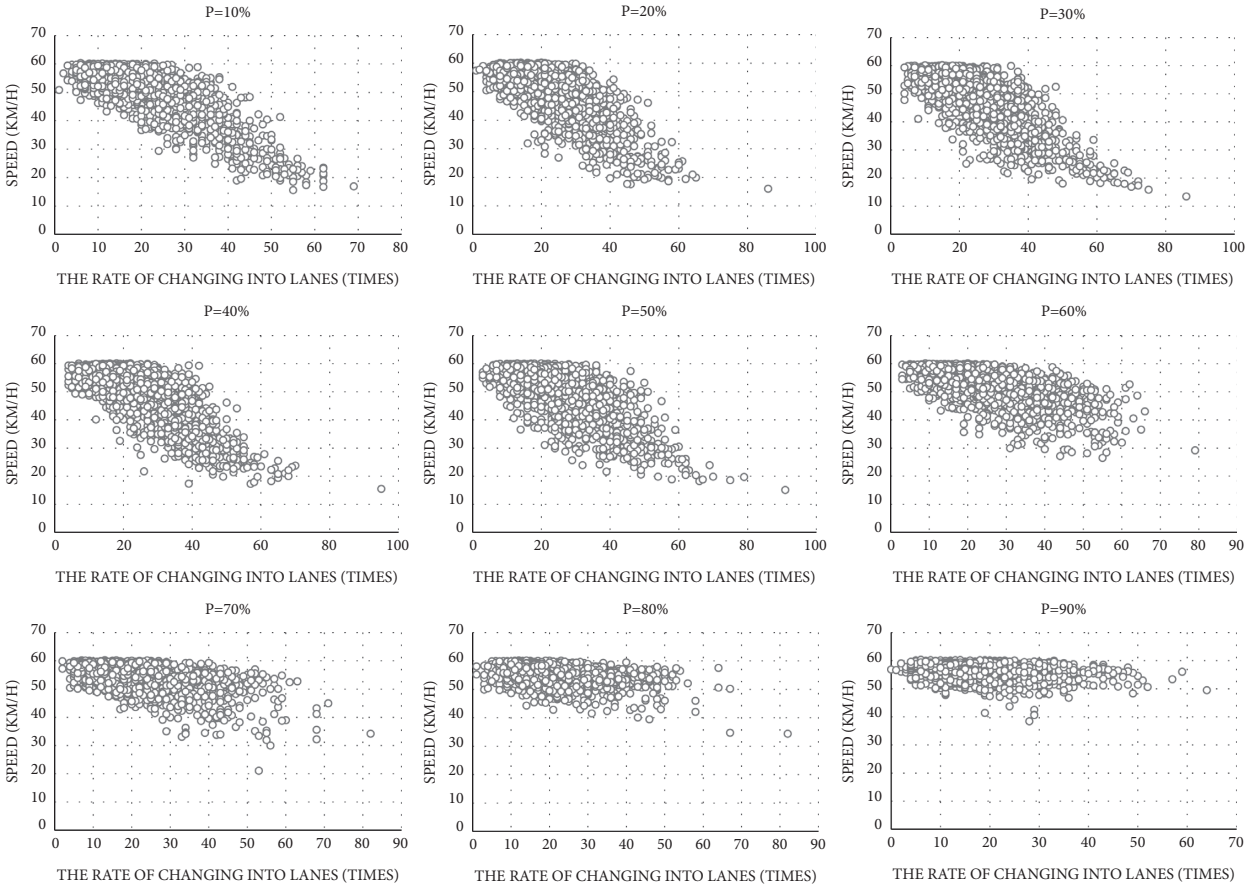
		Lane1		Lane2		Lane3		Lane4	
		LCo	LCi	LCo	LCi	LCo	LCi	LCo	LCi
p=10%	Pearson's Correlation	−0.508**	−0.767**	−0.761**	−0.788**	−0.512**	−0.135**	−0.093**	−0.042*
	Sig. (2-tailed)	0	0	0	0	0	0	0	0.033
p=20%	Pearson's Correlation	−0.505**	−0.735**	−0.759**	−0.762**	−0.590**	−0.295**	−0.266**	−0.309**
	Sig. (2-tailed)	0	0	0	0	0	0	0	0
p=30%	Pearson's Correlation	−0.562**	−0.760**	−0.789**	−0.769**	−0.637**	−0.431**	−0.397**	−0.478**
	Sig. (2-tailed)	0	0	0	0	0	0	0	0
p=40%	Pearson's Correlation	−0.528**	−0.750**	−0.776**	−0.759**	−0.665**	−0.470**	−0.467**	−0.518**
	Sig. (2-tailed)	0	0	0	0	0	0	$9.16E-141$	$3.80E-178$
p=50%	Pearson's Correlation	−0.549**	−0.736**	−0.770**	−0.732**	−0.692**	−0.576**	−0.542**	−0.576**
	Sig. (2-tailed)	0	0	0	0	0	0	0	0
p=60%	Pearson's Correlation	−0.500**	−0.617**	−0.640**	−0.593**	−0.757**	−0.665**	−0.632**	−0.707**
	Sig. (2-tailed)	0	0	0	0	0	0	$2.72E-28$	0
p=70%	Pearson's Correlation	−0.423**	−0.531**	−0.502**	−0.482**	−.788**	−.694**	−0.673**	−0.762**
	Sig. (2-tailed)	0	0	0	0	0	0	0	0
p=80%	Pearson's Correlation	−0.328**	−0.401**	−0.324**	−0.341**	−0.792**	−0.682**	−0.655**	−0.776**
	Sig. (2-tailed)	0	0	0	0	0	0	0	0
p=90%	Pearson's Correlation	−0.277**	−0.319**	−0.169**	−0.194**	−0.712**	−0.540**	−0.540**	−0.719**
	Sig. (2-tailed)	0	0	0	0	0	0	0	0

Note: (a)*when 0.05, significant correlation. (b)**When 0.01, significant correlation.



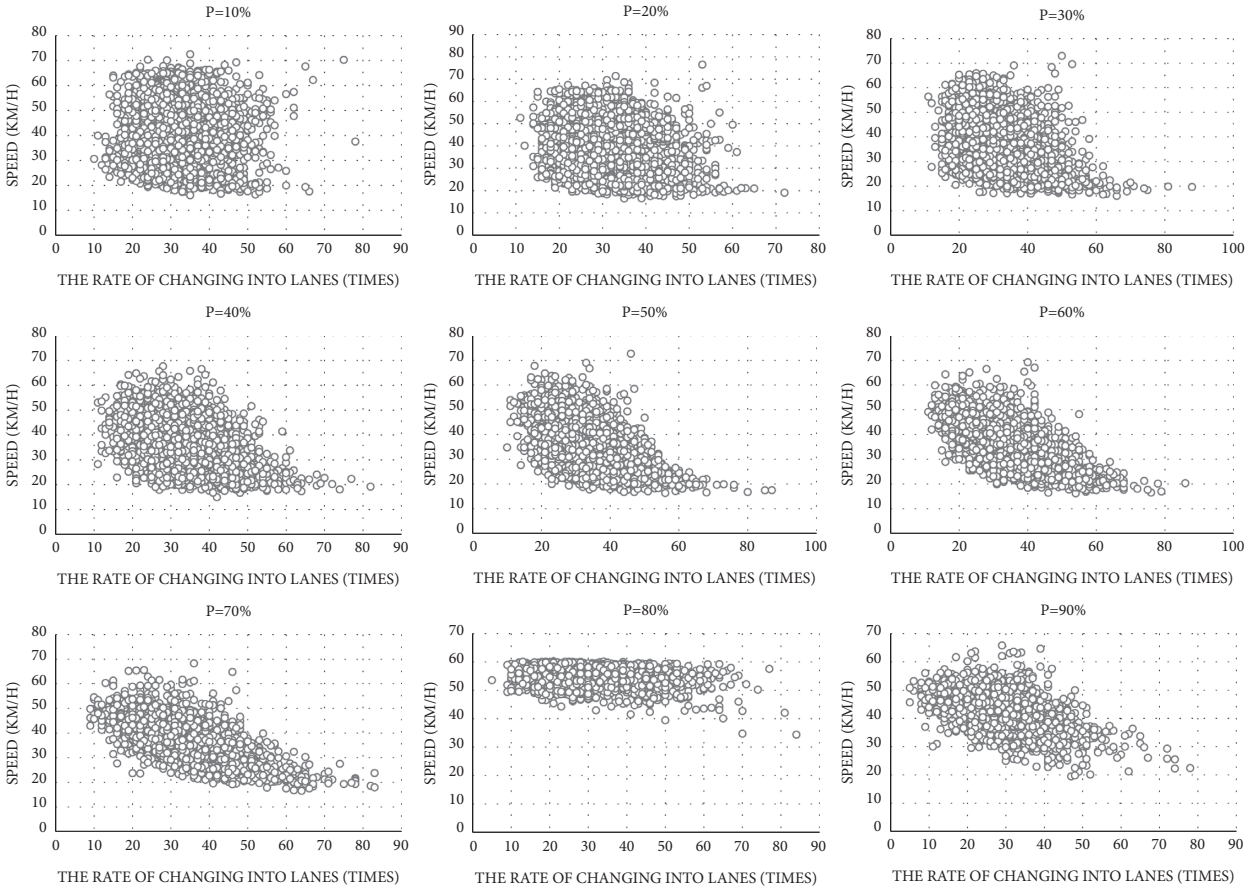
(a)

Figure 9: Continued.



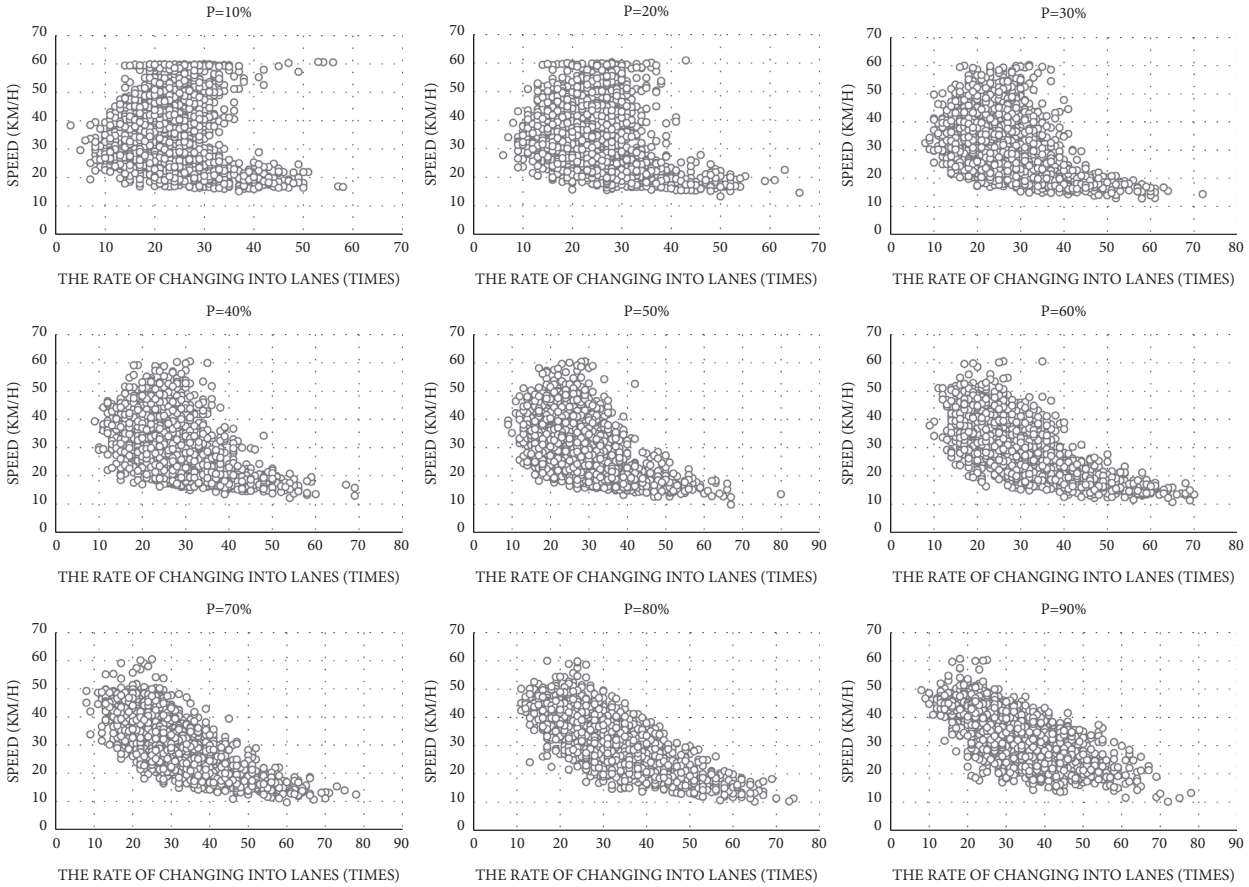
(b)

Figure 9: Continued.



(c)

Figure 9: Continued.



(d)

Figure 9: Scatter plot of the relationship between speed and the rate of changing into lanes. (a) A scatter plot of Lane 1. (b) A scatter plot of Lane 2. (c) A scatter plot of Lane 3. (d) A scatter plot of Lane 4.

Table 6: The degree of the fit table of the relationship between the speed and the lane change rate under different penetration rates of autonomous vehicles in each lane.

		p=10%	p=20%	p=30%	p=40%	p=50%	p=60%	p=70%	p=80%	p=90%
Lane1	Multiple R	0.777	0.837	0.832	0.852	0.849	0.865	0.535	0.889	0.343
	R square	0.604	0.702	0.693	0.726	0.721	0.749	0.286	0.791	0.118
	Adjusted R square	0.604	0.701	0.692	0.725	0.720	0.749	0.286	0.791	0.117
	Standard error	6.10	30.54	30.49	28.96	29.29	28.81	4.49	27.33	3.62
Lane2	Multiple R	0.796	0.837	0.798	0.846	0.774	0.878	0.511	0.904	0.202
	R square	0.633	0.702	0.636	0.716	0.599	0.771	0.261	0.817	0.041
	Adjusted R square	0.633	0.701	0.636	0.715	0.599	0.779	0.260	0.817	0.040
	Standard error	4.96	27.96	5.41	27.18	5.13	25.46	3.48	23.58	2.40
Lane3	Multiple R	0.512	0.899	0.645	0.884	0.712	0.872	0.800	0.885	0.716
	R square	0.262	0.809	0.416	0.782	0.508	0.761	0.640	0.783	0.513
	Adjusted R square	0.261	0.809	0.415	0.782	0.507	0.761	0.640	0.783	0.512
	Standard error	11.19	18.22	8.61	17.31	7.29	17.98	5.25	19.07	4.40
Lane4	Multiple R	0.097	0.899	0.538	0.878	0.663	0.850	0.805	0.851	0.753
	R square	0.009	0.809	0.290	0.772	0.439	0.723	0.648	0.725	0.568
	Adjusted R square	0.008	0.809	0.289	0.772	0.439	0.723	0.648	0.724	0.568
	Standard error	11.47	15.47	8.49	15.06	7.21	16.04	5.77	17.97	5.51

Table 7: The parameter value interval table of the regression equation under different penetration rates of autonomous vehicles.

		Lane1		Lane2		Lane3		Lane4	
		Lower 95%	Upper 95%	Lower 95%	Upper 95%	Lower 95%	Upper 95%	Lower 95%	Upper 95%
	c	0	0	0	0	0	0	0	0
p=10%	a	[0.334	0.4987]	[−0.31	−0.20]	[−0.57	−0.50]	[−0.19	−0.07]
	b	[−1.42	−1.312]	[−0.47	−0.38]	[−0.09	0.009]	[−0.107	0.015]
	c	0	0	0	0	0	0	0	0
p=20%	a	[6.08	6.83]	[0.32	0.92]	[−0.08	0.02]	[0.67	0.79]
	b	[−1.44	−0.85]	[1.08	1.55]	[1.09	1.22]	[0.46	0.59]
	c	0	0	0	0	0	0	0	0
p=30%	a	[5.991	6.759]	[−0.55	−0.44]	[−0.45	−0.39]	[−0.38	−0.29]
	b	[−1.45	−0.88]	[−0.28	−0.19]	[−0.18	−0.09]	[−0.50]	[−0.41]
	c	0	0	0	0	0	0	0	0
p=40%	a	[6.10	6.84]	[−0.05	0.51]	[−0.05	0.05]	[0.46	0.59]
	b	[−1.27	−0.72]	[1.28	1.71	[0.87	1.00]	[0.42	0.54]
	c	0	0	[63.14	64.10]	[57.95	59.97]	[52.457	54.61]
p=50%	a	[5.812	6.622]	[−0.60	−0.49]	[−0.38	−0.33]	[−0.454	−0.38]
	b	[−1.04	−0.45]	[−0.17	−0.09]	[−0.26	−0.19]	[−0.48	−0.41]
	c	0	0	0	0	0	0	0	0
p=60%	a	[4.81	5.69]	[0.04	0.55]	[−0.09	0.03]	[0.48	0.62]
	b	[0.62	1.23]	[1.40	1.77]	[0.84	1.00]	[0.48	0.62]
	c	[63.79	64.590]	[58.32	58.95]	[57.90	59.17]	[55.538	57.04]
p=70%	a	[−0.25	−0.091]	[−0.20	−0.13]	[−0.33	−0.29]	[−0.38	−0.31]
	b	[−0.58	−0.47]	[−0.09	−0.04]	[−0.19	−0.13]	[−0.55	−0.50]
	c	0	0	0	0	0	0	0	0
p=80%	a	[3.95	4.86]	[0.89	1.34]	[0.22	0.35]	[0.38	0.56]
	b	[2.84	3.51]	[1.21	1.54]	[0.58	0.77]	[0.47	0.62]
	c	0	0	0	0	0	0	0	0
p=90%	a	[−0.27	−0.15]	[−0.04	0]	[−0.29	−0.26]	[−0.24	−0.27]
	b	[−0.32	−0.22]	[−0.06	−0.03]	[−0.09	−0.04]	[−0.48]	[−0.32]

Assuming that there is a relationship between the speed V, the rate of changing out of lanes LCo, and the rate of changing into lanes LCi, it can be expressed by the following formula:

$$V = f(LCo, LCi) \tag{6}$$

It is assumed that the relationship between the speed V, the rate of changing out of lanes LCo, and the rate of changing into lanes LCi is expressed as a multivariate linear relationship, as shown in

$$V = aLCo + bLCi + c. \tag{7}$$

In the formula, a, b, and c all represent constants.

Regression analysis is performed on the relationship between speed and lane change rate under different penetration rates of autonomous vehicles, and the degree of fit is shown in Table 6.

In Table 6, the degree of fit R is above 0.7 for each penetration rate in each lane, indicating that the relationship between speed and lane change rate is a linear relationship in mixed traffic. It means that the relationship between speed and lane change rate can be described by a linear relationship in mixed traffic, when the penetration rates of autonomous vehicles are greater than 10%. The parameter value interval of Equation (7) under different penetration rates is shown in Table 7.

In Table 7, parameter c obtained in all lanes with different penetration rates is all 0. Parameter a shows an upward trend in Lane 1, and it also shows rise and fall trends in the other lanes. Parameter b shows a downward trend in Lane 1, and it has the same trend in the other lanes, which is like parameter a. However, there is a slight change in trends of the parameters. All of those mean that formula (7) shows certain stability on the road in mixed traffic.

4. Conclusions

In this study, empirical and simulation methods are used to study the 4-lane urban expressway with typical ring road characteristics. When autonomous vehicles running on the road in a certain proportion, the speed of each lane on the road section is related to the lane-changing behavior in large traffic flow. In the empirical stage, video data, which are collected from 9 road sections on the third ring road in Chengdu, China, were used as the basis for establishing the simulation model. In the simulation model, the third ring road in Chengdu, China, was used as the benchmark object, and a simulation model of a 45 kilometer ring road with 9 exits and 36 entrances was established. In the model, the speed limit is also set according to the maximum speed limit of the third ring expressway of Chengdu. In order to further study the relationship between speed and lane change rate, only the large traffic flow was simulated. We simulated the situations under the penetration rates of autonomous vehicles of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and

90%, respectively. We analyzed the data of the speed, the rate of changing out of lanes, and the rate of changing into lanes in each lane. Then, we found that

(1) Speed in each lane is affected by autonomous vehicles running on the road to a certain extent. The speed of the two lanes, which are far away from the entrance and exit, shows a downward trend when the penetration rates of autonomous vehicles are 10% and 30%. But with the increase in the penetration rates of autonomous vehicles, the speed is effectively increased. The speed of the two lanes, which are close to the entrance and exit, shows a downward trend with the increase in autonomous vehicle penetration rates, especially the lane that is directly connected with the entrance or exit.

(2) The rate of changing out of lanes and the rate of changing into lanes are closer to the changes in each lane with different penetration rates of autonomous vehicles. When the penetration rate of autonomous vehicles is 10%, the rate of changing into lanes and the rate of changing out of lanes have relatively obvious downward trends from Lane 1 to Lane3 and have no obvious trend in Lane 4. Then, with the increase in the penetration rates of autonomous vehicles, the rate of changing into lanes and the rate of changing out of lanes did not significantly change.

(3) With the penetration of autonomous vehicles, the speed, the rate of changing out of lanes, and the rate of changing into lanes in all lanes show tendencies to converge toward the mode position. Compared with 0% penetration rate of autonomous vehicles, the autonomous vehicles take the behavior of changing lanes more balanced in the whole time.

(4) With the penetration of autonomous vehicles, there is a linear relationship between speed, the rate of changing out of lanes, and the rate of changing into lanes. Although the data representations of the three factors are still non-normally distributed, the fit degree obtained after regression analysis remains above 0.7 under different penetration rates of autonomous vehicles in different lanes. Those mean that a linear relationship between speed and lane change rate exists in the mixed traffic.

In conclusion, we believe that in the 4-lane urban expressway with large traffic flow, autonomous vehicles have a certain effect on improving traffic flow. With the penetration of autonomous vehicles, the speed is significantly increased in the lane far away from the entrance and exit, but it is not obvious or even shows a slight downward trend for the lane closer to the entrance and exit. However, the lane-changing behavior only is decreased to a certain extent when the penetration rate is 10%. As the penetration rate increases, the decreasing trend is not obvious. The speed and lane change rate show a linear relationship. Under different penetration rates, its parameters do not change much, and the formula has a certain stability.

In addition, this study also has some shortcomings. The speed-lane change rate model proposed in this study has not been tested for its applicability in mixed traffic on other roads. These are the directions for further research studies in the future.

Data Availability

The data used in this research can be found via the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the Applied Basic Research Programs and Technology Commission Foundation of Sichuan Province of China (Grant Nos. 2019JDR0093, 2020JDR0253, 2021YJ0066, 2017JY0246, and 2017JY0269), the Chengdu Science and Technology Project (Grant Nos. 2017RK0000306ZF and 2017RK0000357ZF), and the Key Scientific Research Fund of Xihua University (Grant No. Z17131). The authors would like to thank Dr. Qinhua Ren, Dr. Juanxiu Zhu, Dr. Huawei Duan, Dr. Kehong Li, and Mr. Zheng Lei for their contribution to data collection and analysis and preparation of the manuscript. The authors would also like to thank their school brothers and school sisters for their support and help in the process of research and writing.

References

[1] G. Orosz, R. E. Wilson, and G. Stépán, "Traffic jams: dynamics and control," *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, vol. 368, no. 1928, pp. 4455–4479, 2010.

[2] D. Ngoduy, "Analytical studies on the instabilities of heterogeneous intelligent traffic flow," *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 10, pp. 2699–2706, 2013.

[3] J. Sun, Z. Zheng, and J. Sun, "Stability analysis methods and their applicability to car-following models in conventional and connected environments," *Transportation Research Part B: Methodological*, vol. 109, pp. 212–237, 2018.

[4] C. Wang, Q. Sun, Z. Li, H. Zhang, and K. Ruan, "Cognitive competence improvement for autonomous vehicles: a lane change identification model for distant preceding vehicles," *IEEE Access*, vol. 7, pp. 83229–83242, 2019.

[5] Y. S. Jiang, R. Hu, Z. H. Yao, P. C. Wu, and X. L. Luo, "Stability and safety analysis for heterogeneous traffic flow composed of intelligent and connected vehicles," *Journal of Veijing Jiaotong University*, vol. 44, no. 01, pp. 27–33, 2020.

[6] L. Huang, H. Guo, R. Zhang, H. Wang, and J. Wu, "Capturing drivers' lane changing behaviors on operational level by data driven methods," *IEEE Access*, vol. 6, pp. 57497–57506, 2018.

[7] H. Wang, Y. Qin, W. Wang, and J. Chen, "Stability of CACC-manual heterogeneous vehicular flow with partial CACC performance degrading," *Transportation Business: Transport Dynamics*.vol. 7, no. 1, pp. 1–26, 2019.

[8] Y. Dou, Y. Fang, C. Hu, R. Zheng, and F. Yan, "Gated branch neural network for mandatory lane changing suggestion at the on-ramps of highway," *IET Intelligent Transport Systems*, vol. 13, no. 1, pp. 48–54, 2019.

[9] A. Kesting, M. Treiber, and D. Helbing, "Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity," *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, vol. 368, no. 1928, pp. 4585–4605, 2010.

[10] Y. Liu, J. Guo, J. Taplin, and Y. Wang, "Characteristic analysis of mixed traffic flow of regular and autonomous vehicles using cellular automata," *Journal of Advanced Transportation*, vol. 2017, pp. 1–10, Article ID 2854895, 2017.

[11] L. Ye and T. Yamamoto, "Impact of dedicated lanes for connected and autonomous vehicle on traffic flow throughput," *Physica A: Statistical Mechanics and Its Applications*, vol. 512, pp. 588–597, 2018.

[12] W. Li, S. Chen, X. Wang, C. Yin, and Z. Huang, "A hybrid approach for short-term traffic flow forecasting based on similarity identification," *Modern Physics Letters B*, vol. 35, no. 13, p. 2150212, 2021.

[13] S. Wang, J. Zhao, C. Shao, C. Dong, and C. Yin, "Truck traffic flow prediction based on LSTM and GRU methods with sampled GPS data," *IEEE Access*, vol. 8, pp. 208158–208169, 2020.

[14] A. Talebrour and H. S. Mahmassani, "Influence of connected and autonomous vehicles on traffic flow stability and throughput," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 143–163, 2016.

[15] L. Ye and T. Yamamoto, "Modeling connected and autonomous vehicles in heterogeneous traffic flow," *Physica A: Statistical Mechanics and Its Applications*, vol. 490, pp. 269–277, 2018.

[16] M. W. Levin and S. D. Boyles, "A multiclass cell transmission model for shared human and autonomous vehicle roads," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 103–116, 2016.

[17] G. Li, Z. Yang, Q. Yu, J. Ma, and S. Fang, "Characterizing heterogeneity among merging positions: comparison study between random parameter and latent class Accelerated hazard model," *Journal of Transportation Engineering, Part A: Systems*, vol. 147, no. 6, p. 4021029, 2021.

[18] X. Wang, W. Li, C. Yin, S. Zeng, and P. Liu, "A multiple-parameter approach for short-term traffic flow prediction," *Modern Physics Letters B*, vol. 35, no. 14, p. 2150245, 2021.

[19] Z. Yao, H. Jiang, Y. Cheng, Y. Jiang, and B. Ran, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2020.

[20] Y. Y. Qin Yan-Yan, H. Wang Hao, W. Wang Wei, and Q. Wan Qian, "Stability analysis and fundamental diagram of heterogeneous traffic flow mixed with cooperative adaptive cruise control vehicles," *Acta Physica Sinica*, vol. 66, no. 9, p. 094502, 2017.

[21] Z. Yao, R. Hu, Y. Wang, Y. Jiang, B. Ran, and Y. Chen, "Stability analysis and the fundamental diagram for mixed connected automated and human-driven vehicles," *Physica A: Statistical Mechanics and Its Applications*, vol. 533, p. 121931, 2019.

[22] J. A. Laval and C. F. Daganzo, "A hybrid model of traffic flow: impacts of roadway geometry on capacity," in *Proceedings of the TRB 2003 Annual Meeting CD-ROM*, The Georgia Institute of Technology, Atlanta, Georgia, 2003.

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

[24] J. A. Laval and C. F. Daganzo, "Lane-changing in traffic streams," *Transportation Research Part B: Methodological*, vol. 40, no. 3, pp. 251–264, 2006.

[25] W.-L. Jin, "A kinematic wave theory of lane-changing traffic flow," *Transportation Research Part B: Methodological*, vol. 44, no. 8-9, pp. 1001–1021, 2010.

[26] Transportation Research Board, *Highway Capacity Manual (HCM2010)*, National Academy of Sciences, USA, 2010.

[27] S. Park and S. G. Ritchie, "Exploring the relationship between freeway speed variance, lane changing and vehicle heterogeneity[R]," in *Proceedings of the 83rd Annual Meeting of the Transportation Research Board*, Washington D. C.UCI-ITS-TS-WP-04-4, Washington D. C., 2004.

[28] F. Gmez -Bravo, F. Cuesta, A. Oliero, and A. Viguria, "Continuous curvature path generation based on β -spline curves for parking manoeuvres," *Robotics and Autonomous Systems*, vol. 56, no. 4, pp. 360–372, 2008.

[29] H. Sun, W. W. Deng, and S. M. Zhang, "Micro vehicle dynamic trajectory plan with global optimality," *Journal of the University (Engineering and Technology Edition)*, vol. 44, no. 4, pp. 918–924, 2014.

[30] F. C. Lan, S. C. Li, J. Q. Chen, and Z. L. Liu, "Comprehensive approach for trajectory optimization of autopilot vehicles considering handling stability," *Journal of Hunan University*, vol. 46, no. 10, pp. 36–45, 2019.

[31] A. Kesting, M. Treiber, and D. Helbing, "General lane-changing model MOBIL for car-following models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1999, no. 1, pp. 86–94, 2007.

[32] C. Vallon, Z. Ercan, A. Carvalho, and F. Borrelli, "A machine learning approach for personalized autonomous lane change initiation and control," in *Proceedings of the 2017 IEEE Intelligent Vehicles Symposium(IV)*, pp. 1590–1595, IEEE, Los Angeles, CA, USA, June 2017.

[33] K. Liu, J. Gong, A. Kurt, H. Chen, and U. Ozguner, "Dynamic modeling and control of high-speed automated vehicles for lane change maneuver," *IEEE TRANSACTIONS ON INTELLIGENT VEHICLES*, vol. 3, no. 3, pp. 329–339, 2018.

[34] H. Zheng, J. Zhou, Q. Shao, and Y. Wang, "Investigation of a longitudinal and lateral lane-changing motion planning model for intelligent vehicles in dynamical driving environments," *IEEE Access*, vol. 7, pp. 44783–44802, 2019.

[35] S. Jongsang, C. Heungseok, and Y. Kyongsu, "Stochastic model-predictive control for lane change decision of automated driving vehicles," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 4771–4782, 2018.

[36] Y. Ekim, Y. Suguru, M. Chiyomi et al., "Integrating driving behavior and traffic context through signal symbolization for data reduction and risky lane change detection," *IEEE TRANSACTIONS ON INTELLIGENT VEHICLES*, vol. 3, no. 3, pp. 242–253, 2018.

[37] Z. Wang, X. Zhao, Z. Xu, X. Li, and X. Qu, "Modeling and field experiments on autonomous vehicle lane changing with surrounding human-driven vehicles," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 7, pp. 877–889, 2020.

[38] L. Seolyoung, O. Cheol, and H. Sungmin, "Exploring lane change safety issues for manually driven vehicles in vehicle platooning environments," *IET Intelligent Transport Systems*, vol. 12, no. 9, pp. 1142–1147, 2018.

[39] P. Cao, Z. Xu, Q. Fan, and X. Liu, "Analysing driving efficiency of mandatory lane change decision for autonomous vehicles," *IET Intelligent Transport Systems*, vol. 13, no. 3, pp. 506–514, 2019.

[40] *Ministry of Housing and Urban-Rural Development of the People's Republic of China.Design Discipline of Urban Expressway*, pp. 1–5, China Architecture & Building Press, Beijing, China, 2009.

[41] Q. S. Zhang and Y. P. Zhang, *Analysis of Road Traffic Capacity*, China communication press, Beijing, 2002.

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 4499477, 40 pages
https://doi.org/10.1155/2022/4499477

Research Article

Analysis of the Relationship between the Density and Lane-Changing Behavior of Circular Multilane Urban Expressway in Mixed Traffic

Han Xie, Juanxiu Zhu, and Huawei Duan

Xihua University, School of Management, Jinniu Disitrict Jinzhou Road 999, Chengdu, China

Correspondence should be addressed to Juanxiu Zhu; zhujuanxiu@163.com

Received 14 October 2021; Accepted 15 November 2021; Published 4 January 2022

Academic Editor: Gen Li

The behavior of changing lanes has a great impact on road traffic with heavy traffic. Traffic flow density is one of the important parameters that characterize the characteristics of traffic flow, and it will also be affected by the behavior of changing lanes, especially in the case of each lane. The penetration of autonomous vehicles can effectively reduce lane-changing behavior. Studying the relationship between traffic flow density and lane-changing behavior under different autonomous vehicle penetration rates is of great significance for describing the operation mechanism of mixed traffic flow and the control of mixed traffic. In this article, we use empirical, simulation, and data-driven methods to analyze the urban expressway of autonomous vehicles with penetration rates of 10%, 20%, 30%, 40%, 50%, 60%, 70%, and 80%, respectively. A simulation experiment was carried out on the road, and data related to density, the rate of changing into the lanes, and the rate of changing out lanes were collected. The analysis of the experimental results found the following: (1) The increase in penetration of autonomous vehicles leads to a certain degree of downward trend in density, the rate of changing into the lanes, and the rate of changing out lanes. (2) Different lanes have different effects on the penetration of autonomous vehicles. In a 4-lane road, the two lanes farther from the entrance and exit are closer in appearance, while the two lanes closer to the entrance and exit are similar. (3) The relationship between density and the rate of changing into the lanes and the rate of changing out lanes shows a linear relationship with the penetration of autonomous vehicles. Although the performance of each lane is slightly different, in general, it can be carried out by a multiple regression model. The given parameter value range is relatively close under different permeability. In summary, autonomous vehicles effectively reduce the traffic density and lane-changing behavior of each lane. There is a linear relationship between traffic flow density and lane-changing behavior with the penetration of autonomous vehicles. The density-lane-changing behavior model proposed in this paper can better describe the relationship between the density of the circular multilane urban expressway and the lane-changing behavior in the case of a large traffic flow in mixed traffic.

1. Introduction

The relationship between traffic flow density and lane-changing behavior is very important for understanding the mechanism of traffic flow. Traffic flow density is one of the important parameters of traffic flow, which can directly reflect traffic demand and is also the basis for the classification of road service levels. The behavior of changing lanes can quickly change the density of traffic flow, especially on a single lane. Therefore, studying the relationship between traffic flow density and lane-changing behavior is an important part of understanding the operation mechanism of urban expressway traffic flow.

With the development of artificial intelligence, autonomous vehicles have gradually appeared in people's sight in recent years. Vehicles in the controlled environment will adopt car-following behavior in most cases and will adopt lane-changing behaviors when there are obstacles in front or need to change lanes to reach the target lane. Compared with humans, the lane-changing behavior of autonomous vehicle is more concerned with safety and standardization. When autonomous vehicles are driving on roads in mixed traffic,

they will affect the road traffic flow [1–4]. At present, there are many studies on lane-changing behavior of a single autonomous vehicle [5–8], and most of the research studies used data-driven and simulation methods. There are fewer studies on the influence of lane-changing behavior on density and fewer studies on the relationship between individual lanes. As an important traffic flow parameter, it is important to understand the impact of autonomous vehicles on traffic flow density and lane-changing behavior in mixed traffic.

This paper uses empirical, simulation, and data-driven methods to study the relationship between density and lane-changing behavior under mixed traffic conditions. Taking a typical ring-shaped multilane urban expressway [9], the third ring expressway in Chengdu, China, as the simulation model benchmark object, this study first adopts empirical research, through the traffic flow density, the rate of changing into lanes, and the rate of changing out lanes of the third ring expressway, and data collection and analysis were carried out to verify the validity of the data. Then, the method of benchmarking the three-loop modeling was adopted to establish the corresponding simulation model and compare the empirical data to verify the usability of the simulation model. Finally, 8 simulation experiments with the penetration rate of 10%, 20%, 30%, 40%, 50%, 60%, 70%, and 80% of autonomous vehicles are designed to collect the data of traffic flow density, the rate of changing into lanes, and the rate of changing out lanes. A multiple linear regression model of density and lane-changing behavior was proposed for the first time through a data-driven method. And the parameter value range was given. The model better describes the relationship between the density and the lane-changing behavior of a typical circular multilane urban expressway in a mixed traffic environment when the traffic volume is large.

The first part of this article is an introduction to the research background. The second part is the method used in the article, specifically for empirical research and simulation model establishment. The third part introduces simulation experiments and experimental results. The fourth part analyzes the experimental results and proposes the established relationship model between density and lane-changing behavior. Finally, a summary is made.

1.1. Literature Review. Traffic flow density is an important parameter that characterizes the state of traffic flow. At present, most of the research on density is based on basic graph theory. The basic graph is the basic theory of traffic flow. The theory describes the nonlinear relationship between traffic flow, speed, and density, and a pairwise relationship model of three parameters is given. The basic graph theory is based on a human driving environment. At present, scholars have proposed the basic graph of mixed traffic flow. The research mainly adopts two kinds of simulation research and theoretical research. In simulation research, the current focus is on the flow-density relationship [10–14] in the non-full-density state. In terms of theoretical analysis, the average headway distance is often used in the research studies, and the relationship between flow-speed-density is derived [14–18]. The above research studies are based on car-following behavior. As the main behavior of autonomous vehicles running on road, in addition to car-following, there are also lane-changing behaviors. There are few studies on the impact of lane-changing behavior on traffic flow in mixed traffic. The relationship between traffic flow density and lane-changing behavior under mixed traffic conditions has not been studied.

At present, the research studies on the impact of lane-changing behavior on traffic flow are still focused on the traffic flow model. In the human driving, lane-changing behavior is basically based on the LWR model. The lane-changing behavior is used as parameters [19–21] or factors [22]. In HCM2010 [23], the lane-changing rate is introduced for the interweaving lanes of expressways, but the relationship between lane-changing behavior and density is not analyzed. In the mixed traffic, there are few studies on the influence of lane-changing behavior on traffic flow, and more researches focus on the study of lane-changing behavior of autonomous vehicles. The research studies on lane-changing behavior of autonomous vehicles mainly include lane-changing intention [24, 25], lane-changing decision [26, 27], cooperative lane-changing between front and rear vehicles [28–31], and lane-changing position [32]. Research studies use human behavior as the research basis to construct human-like behavior models. Although there is a gap with human behavior, compared with human behavior, self-driving vehicles can reach a partial or full information state when making decisions, and the accuracy of decision-making has also been improved. However, the above research is only conducted from autonomous vehicles, and the research on the impact of lane-changing behavior on traffic flow in mixed traffic is not involved. With the further development of artificial intelligence, it can be predicted that, in the future urban traffic, there will be a state where autonomous vehicles and human-driven vehicles will be mixed. It is very important to understand the operation mechanism of traffic flow in the mixed state. Studying the relationship between density, which is one of the important parameters of traffic flow, and lane-changing behavior, which is one of the main behaviors of vehicles in traffic, is of great significance for understanding the operation mechanism of traffic flow in mixed traffic.

In summary, there are few studies on the influence of lane-changing behavior on density in mixed traffic. Lane-changing behavior is only used as a factor or parameter to participate in the study of traffic flow in the human driving situation. And the relationship between lane-changing behavior and traffic flow parameters has not been studied. The behavior of changing lanes has the greatest impact on each lane on the road, especially in heavy traffic. The urban expressway fits this situation. In the case of frequent entrances and exits, more lane-changing behaviors, greater traffic flow, and more interweaving areas, the relationship between lane-changing behavior and density under non-full traffic flow conditions can be better studied. Therefore, it is feasible to use a typical circular multilane urban expressway as the

benchmark object for empirical research and simulation modeling to conduct simulation experiments to study the relationship between traffic flow density and lane-changing behavior in a mixed state. A clear relationship between density and lane-changing behavior in mixed traffic is of great significance for understanding the traffic flow mechanism of typical circular multilane urban expressways in mixed traffic, and it can also provide a theoretical basis for future mixed traffic management and control.

2. Research Protocols

Empirical, simulation, and data-driven methods are used in this paper to verify the effectiveness and accuracy of the empirical data. Furthermore, a simulation model is established, corresponding data through simulation experiments are collected, and a data-driven method is used to establish a model of the relationship between density and lane-changing behavior. In this part, empirical data testing and simulation modeling methods are used.

2.1. Empirical Research and Data Testing. In this paper, the data used in this study is to extract data from videos, and the video data were collected from 2011, 2015, and 2018. The videos come from on-site shooting and Chengdu Transportation Administration. The data is extracted from the videos using manual collection and self-developed traffic flow data collection software. The distribution of the collection points is shown in Figure 1.

In Figure 1, there are a total of 9 collection points. The collection points consider the surrounding land use, the connection mode of the road section and the overpass, and whether it is convenient for data collection and other conditions. Through on-site inspections, map selection, and inspection of relevant specifications, the final selection is consistent with connecting entrances and exits, pedestrian bridges, and roads for commercial and residential use, and the connection with the overpass is connecting straight and connecting with the overpass and connecting the curved road. Data collection is performed on 9 road sections connected with an overpass, connected to a straight road but not connected to an overpass, and connected to a curve but not connected to an overpass in one of four ways. In this study, data of traffic flow density, the rate of changing into lanes, and the rate of changing out lanes are collected. The collection is carried out in the morning and afternoon of the working day with heavy traffic.

In order to make the data representative of roads, the minimum sample size is adopted as formula (1) [33], and the minimum sample size is calculated to be 48. That is, as long as the data is larger than 48 groups, it can represent the road characteristics.

$$n \geq \left(\frac{\sigma \bullet K}{E}\right)^2, \tag{1}$$

where n is the minimum sample size for observation; σ is the standard deviation of the sample size of the observed vehicle speed; K is the constant level of confidence to meet expectations; and E is the allowable error of vehicle speed depending on the accuracy of the average vehicle speed, generally 1.5~2 km/h.

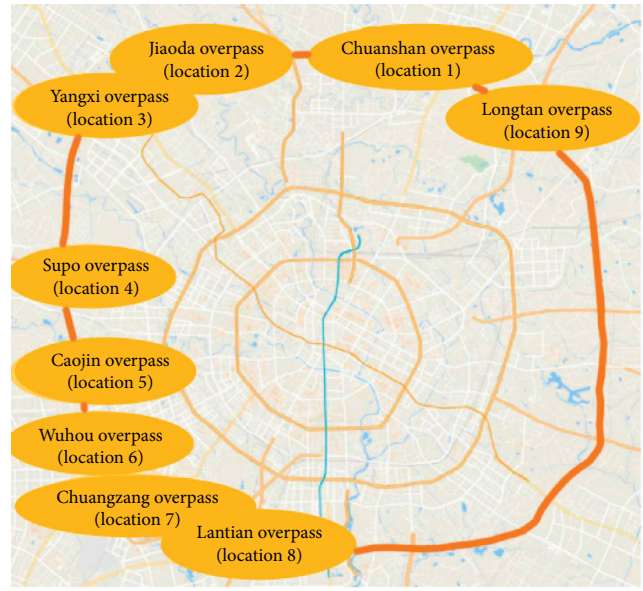


Figure 1: Collection point situation.

In the formula, the parameter value is $\sigma = 7$ km/h (according to two-way eight lanes), $K = 1.96$ (95%), and $E = 2$ km/h, and after calculation, the minimum sample size is 48.

The collected data at each point is compared with the video data to eliminate the collection of unreasonable data. Finally, the number of available data groups is at least 55 groups and at most 179 groups.

The density, the rate of changing into lanes, and the rate of changing out lanes studied in this paper are defined as follows.

The density k refers to the number of vehicles present on the road per unit length at a certain moment [33], as shown in the following formula:

$$k = \frac{N}{L}, \tag{2}$$

where k is the traffic flow density at a certain moment (vehicles/km), N is the number of vehicles, and L is the length of road section.

The photography method used in the density collection method [34] is shown as follows:

$$k = \frac{\sum_{i=0}^{n} k_i}{n} \times \frac{1}{L}, \tag{3}$$

where n is the total time, i is the number of screens when reading the number of vehicles, k_i is the number of vehicles in the interval measured on the i-th screen, and L is the length of observation interval (km).

The rate of changing into lanes and the rate of changing out lanes are used to describe the behavior of a vehicle driving on the road changing into the intended lane and changing out of the current lane, as shown in Figure 2.

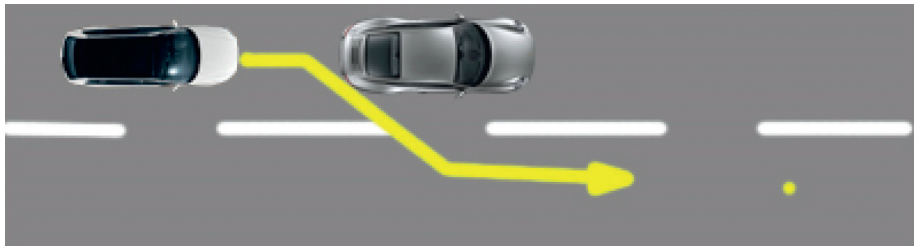


Figure 2: Schematic diagram of vehicle changing lanes (picture from the Internet).

In Figure 2, the white vehicle is a vehicle that intends to change lanes. For some reason, the white vehicle is not satisfied with the current driving lane, so it intends to change lanes from the current driving lane and continue driving. When the vehicle changes lanes, there are two states: partially entering the lane and fully entering the lane. This paper uses more than 50% of the vehicle's location to determine whether the vehicle belongs to the state of changing into the lane or changing out of the lane, so as to include the rate of changing into lanes and the rate of changing out lanes.

The rate of changing into lanes is the number of vehicles changing into the lane, which refers to the number of vehicles entering the tested lane from other lanes within a unit time and unit length. Similarly, the rate of changing out lanes is also the number of vehicles that change lanes, that is, the number of vehicles entering other lanes from the tested lane in a unit time and unit length. Here the unit time is 1 h, and the unit length is 1 km. The definition of the rate of changing into lanes and the rate of changing out lanes is consistent with HCM2010 [23]. However, HCM2010 [23] did not conduct research on the sections of the vehicle on and off the ramp. The main reason is that the main research object of HCM2010 [23] is the expressway. The distance between the on-ramps on the expressway is relatively long, and the ramp has a short influence on the road section, while the distance between the on-ramps on the urban expressway is almost in the interlaced area. The change of lanes of incoming and outgoing vehicles has a greater impact. If only the vehicles on and off the ramp are considered, and the lane-changing behavior in the road section is not considered, the impact of lane-changing behavior on expressways cannot be better described.

In the empirical research, the data collection time interval is 5 min, and the distance length is 100 m or 200 m. Before conducting the research, the data was standardized and converted, and the final data time interval was 1 h, and the distance length was 1 km. The unit of the rate of changing into lanes and the rate of changing out lanes is the number of lane changes in HCM2010 [23], which are represented by LCi and LCo, respectively.

Due to the discontinuity of the collection location and time, some of the collected data are missing. The time difference method formula (4) [33] was used to complete the data. A total of 923 data were obtained after completion. The data distribution box diagram of the density, the rate of changing into lanes, and the rate of changing out lanes of a single lane is shown in Figure 3.

$$Z^t\left(t_j, x_i\right) = \begin{cases} z\left(t_a, x_i\right), & t_j = t_1 \\ z\left(t_{j-1}, x_i\right) + \dfrac{\Delta t}{t_{j-1} - t_a} z\left(t_a, x_i\right), & t_1 < t_j < t_p, \\ z\left(t_j, x_{i-1}\right) & t_j = t_p \end{cases} \tag{4}$$

where Z means time interpolation in time t_j and x_i location; t_a is the first valid data of x_i location; t_1 is initial moment; t_p is end time; x_i is the place where interpolation needs to be performed.

In Figure 3, it can be seen from Figures 3(a) to 3(c) that the distribution of data is relatively discrete, and there are many outliers in each lane in density, the rate of changing into lanes, and the rate of changing out lanes. The 50% quantile data are all distributed low, which is biased towards the 25% quantile data. In Figure 3(a), the density distribution in Lane 1 to Lane 3 is relatively similar, and the distribution in Lane 4 is less. The graphs in Figures 3(b) and 3(c) are similar. Lane 1 and Lane 4 have a lower lane change rate, and Lane 2 and Lane 3 have a higher lane change rate. The graph shows that the lane change-out rate of Lane 4 is lower than the lane change-in rate of Lane 4. This is mainly because Lane 4 is directly connected to the entrance and exit. For vehicles, except for those that must leave the road, they are basically reluctant to drive in this lane, the density of the lane is the lowest, and fewer vehicles are swapped out. Vehicles that have to get out of the road must change into the lane, so the rate of changing into lanes is slightly higher than that of changing out lanes. The situation shown in Figure 3 is basically the same as what actually happened on the road. The statistical analysis of the three parameters can also be further proved. The statistical analysis of the data description is shown in Table 1.

It can be seen from Table 1 that the average density of Lane 1 to Lane 3 is relatively close, and the density of Lane 4 is lower. At the same time, there is a gap between the density mean, the density median, and the density mode. This is reflected in the rate of changing into lanes and the rate of changing out lanes. The data of the three parameters on each lane show that the data is nonnormally distributed. In addition, the rate of changing into lanes and the rate of changing out lanes have lower mean values in Lane 1 and Lane 3, while the mean values of Lane 2 and Lane 3 are both

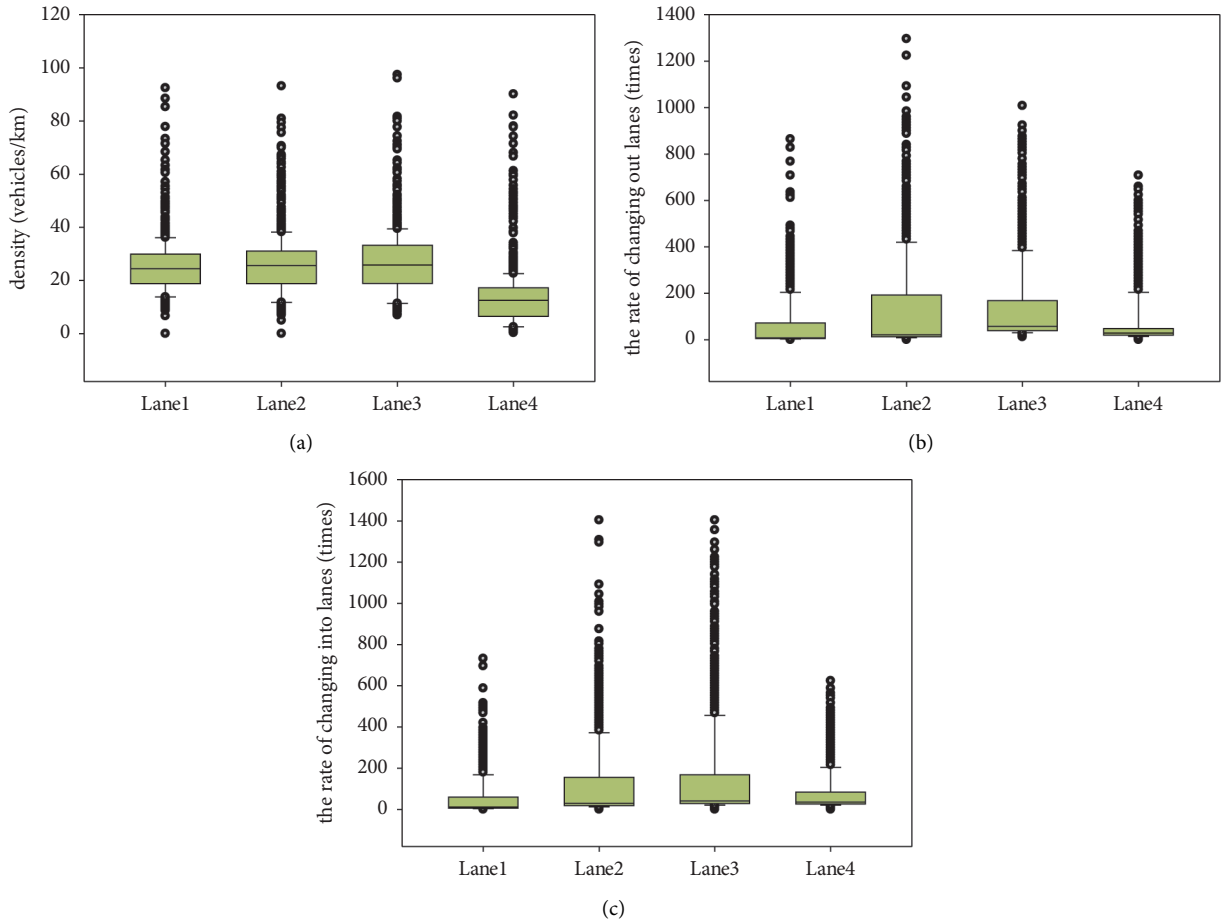


Figure 3: Box plot of empirical data distribution. (a) Box diagram of density distribution of a single lane. (b) A box chart of the rate of changing out lanes of a single lane. (c) A single lane box chart of the rate of changing into lanes.

Table 1: Empirical data by the statistical description.

		Lane 1	Lane 2	Lane 3	Lane 4
Density (vehicle/km)	Mean value	25.33149	26.14947	26.79935	13.66627
	Standard error	0.334813	0.375786	0.40465	0.372006
	Median	24.44	25.6438	25.79167	12.54167
	Mode	22.66667	25.08333	25.79167	2.583333
	Variance	103.4679	130.3416	151.1339	127.7324
	Kurtosis	6.95363	3.987783	4.232871	10.95704
	Skewness	1.730317	1.305238	1.327613	2.656881
Rate of changing out lanes (times)	Mean value	123.7573	277.5991	339.7183	154.7129
	Standard error	3.271707	6.147944	8.187349	3.616784
	Median	108	240	300	132
	Mode	24	144	72	48
	Variance	9879.854	34886.83	61871.17	12073.88
	Kurtosis	3.218006	3.836383	2.183544	1.075749
	Skewness	1.279292	1.438276	1.303985	1.04637
Rate of changing into lanes (times)	Mean value	148.6024	326.8732	293.2264	143.0639
	Standard error	3.734585	6.509294	5.754521	4.287722
	Median	120	276	252	120
	Mode	48	132	192	12
	Variance	12873.2	39108.35	30564.69	16968.94
	Kurtosis	4.616732	1.158177	0.726926	1.667461
	Skewness	1.563167	0.993322	0.956815	1.258795

higher. The performance of the above data is consistent with the box plot display and further confirms that the data presents a nonnormal distribution, and the data distribution conforms to the actual road conditions. The data is true and valid and can be further used for simulation.

2.2. Simulation Model Establishment and Verification. In order to study the relationship between lane-changing behavior and density of urban expressways under different penetration rates of autonomous vehicles, the actual urban expressway was compared with the road network simulation model. The model is established based on the 9 road sections in the empirical study. By simplifying the road network, only exits and entrances are retained.

Using SUMO as the simulation platform, the road section is set as a one-way 4-lane. In order to reflect the actual situation of the road more realistically, in the simulation model, each road section has 4 entrances and 1 exit. The road section setting is shown in Figure 4.

In this article, the simulation model is shown in Figure 4. According to the actual situation in the empirical study, four types of vehicles are set as large trucks, medium trucks, cars, and autonomous vehicles. Among them, the autonomous vehicle is set as car. Each section contains 4 entrances and one exit. The distance between the entrance and the exit is 1 km, the simulation distance of each road section is 5 km, and the simulation distance of the whole road is 45 km, which is close to the total length of 51 km of Chengdu's third ring road. The simulation time is 86400 s, that is, 24 h. The simulation time period is 4 hours in the morning and 4 hours in the afternoon of the working day. The simulation step is 0.1 s. The tranCI interface is used to add different traffic flows to different intersections of the road. The data collection time interval is 5 minutes, and the collected data includes the density, the lane change-in rate, and the lane change-out rate. Data collection was performed on 2591 groups.

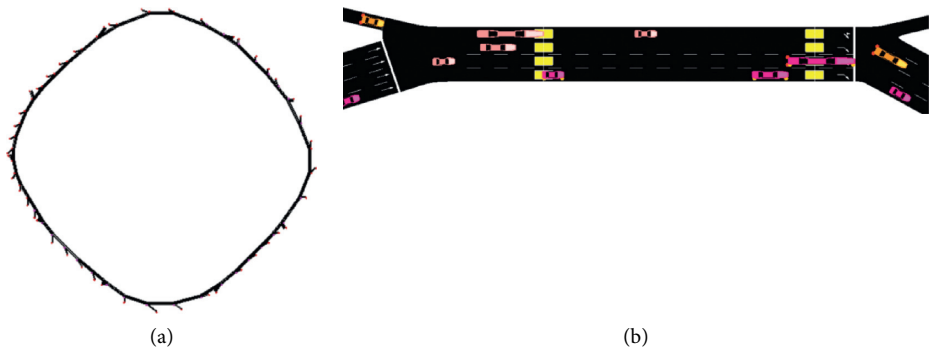
Before conducting autonomous vehicles' penetration rate experiment, first the availability of the simulation model is verified. The traffic flow was run which composed of three types of large trucks, medium trucks, and cars in the simulation model, and traffic flow density, the rate of changing into lanes, and the rate of changing out of lanes were collected to compare with the data in the empirical study. Model verification was carried out. The simulation data distribution box diagram is shown in Figure 5.

In Figure 5, from Figures 5(a) to 5(c), the data distribution of density, the rate of changing into lanes, and the rate of changing out lanes are, respectively, shown. From the figure, the density distribution in Figure 5(a) is Lane 4, which is the largest, Lane 2 is the second, and Lane 1 is the smallest distribution. The distributions of the rate of changing into lanes and the rate of changing out lanes in Figures 5(b) and 5(c) are relatively similar. Lane 1 is the smallest, lanes 2 and 3 are larger, and Lane 3 is the busiest. The 50% quantile of the data shown in Figure 5 is close to the 25% quantile, indicating that the data distribution is nonnormal. It can also be further confirmed from the simulation data statistics, as shown in Table 2.

It can be seen from Table 2 that the average density of Lane 1 and Lane 3 is lower, and the average density of Lane 4 is the largest. The rate of changing out lanes and the rate of changing into lanes are the smallest in Lane 1, and the remaining three lanes are more balanced, of which Lane 3 is the largest. This is mainly due to the fact that Lane 3 accepts a large number of traffic flow from Lane 1 and Lane 2 that need to leave the road. At the same time, it also accepts the traffic flow from Lane 4, so lane change behavior in Lane 3 occurs more frequently. In addition, the data in the table shows that the average is inconsistent with the median and mode, and the data has a skewness higher than 0.7. Therefore, the data presents a nonnormal distribution. The data of simulation and empirical research differ in distribution. However, we found that the data from the simulation experiment and the empirical research both present a nonnormal distribution. At the same time, the relationship between density and the rate of changing out lanes and the relationship between density and the rate of changing into lanes is relatively close, as shown in Figure 6.

It can be seen from Figure 6 that Figures 6(a), 6(c), 6(e), and 6(g) are empirical data and Figures 6(b), 6(d), 6(f), and 6(h) are the simulation data. The data is concentrated in the lower left corner, and the discrete form is beyond the upper right corner. There are differences between the empirical data and the simulation data, but the relationship characteristics are relatively similar, so we believe that the simulation model can be used to show realistic road conditions.

2.3. Data Sources. In our research, the data was used to analyze the relationship between density and the rate of changing into lanes and the relationship between density and the rate of changing out lanes. The data from videos and simulation model were used in this paper. Most of the videos were captured from live shooting, and part of them were from the Chengdu Traffic Management Bureau. Data was collected data from video via the traffic data collection software and manual work. And then, the data was cleaned and complemented before analyzed. In this research, the data of density, the rate of changing out lanes, and the rate of changing into lanes in different lane was used. The data can be accessed from our team, only be allowed in research. Because of some data came from Bureau, only desensitized data can be proved. The part of data we used is shown in Table 3. In Table 3, K1 means density of Lane 1, K2 means density of Lane 2, LCo1 means the rate of changing out lanes of Lane 1, LCo2 means the rate of changing out lanes of Lane 2, LCi1 means the rate of changing into lanes of Lane 1, and LCi2 means the rate of changing into lanes of Lane 2.



(a) (b)

Figure 4: Schematic diagram of simulation modeling. (a) Road network model established by simulation. (b) A road running diagram with one exit, one entrance, and two detectors.

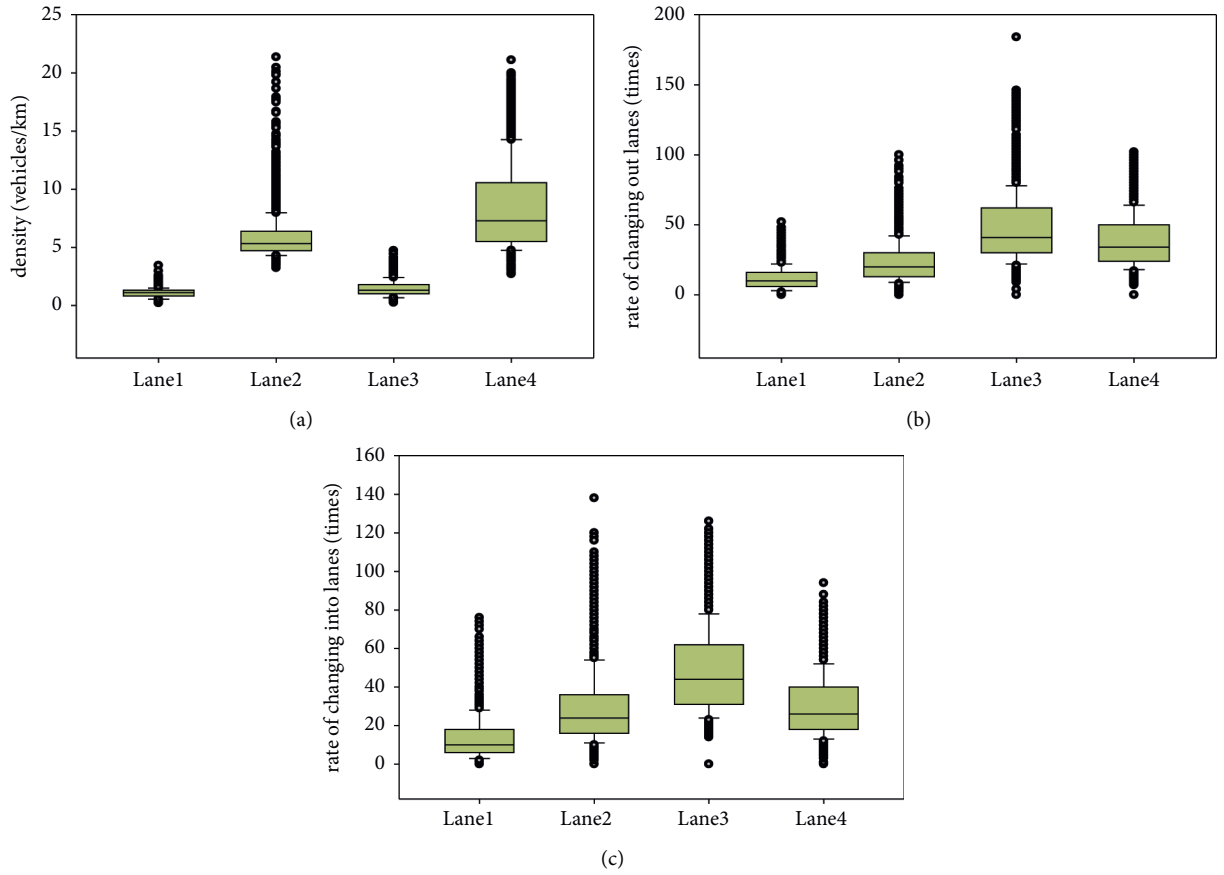


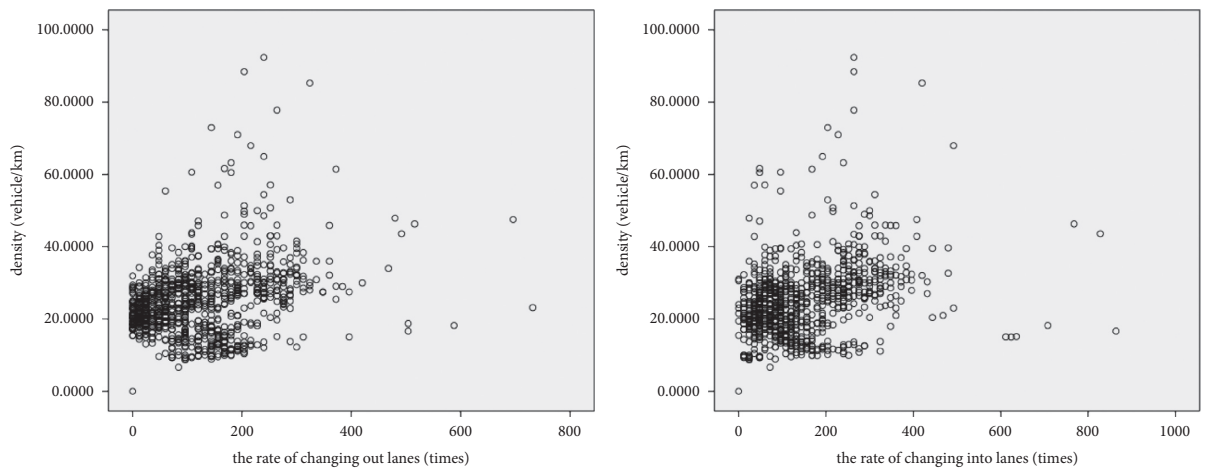
Figure 5: Box diagram of simulation data distribution. (a) A box diagram of the density distribution of a single lane. (b) A box chart of the rate of changing out lanes of a single lane. (c) A box chart of the rate of changing into lanes of a single lane.

3. Simulation Experiment and Results

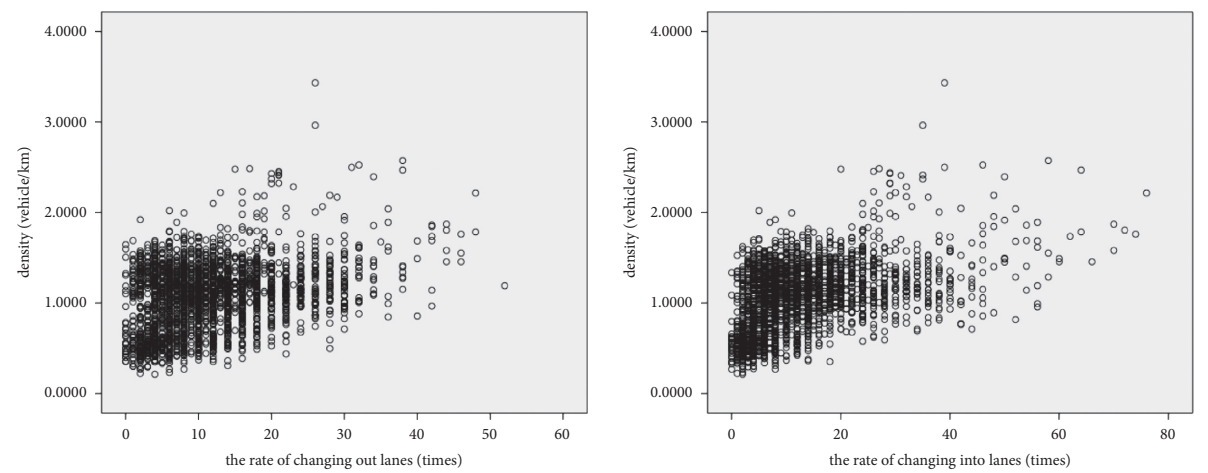
3.1. Design of Simulation Experiments for the Penetration Rate of Different Autonomous Vehicles. Through the previous verification, the simulation model and the research object in the empirical study have a certain similarity in the relationship between density and lane-changing behavior. Therefore, the simulation model can be used to study the relationship between density and lane-changing behavior of autonomous vehicles under different permeability. The addition of autonomous vehicles will have a certain impact on density distribution and lane-changing behavior. For example, autonomous vehicles will automatically choose to follow the car when the speed is currently maintained at the optimal situation, instead of actively changing lanes, and will be in front of the lane. In the case of impossibility or only changing lanes to get out of the road, the lane will definitely be changed actively, and the behavior of changing lanes will have a certain impact on the density of each lane. Therefore, studying the relationship between density and lane-changing

Table 2: Simulation data by the statistical description.

		Lane 1	Lane 2	Lane 3	Lane 4
Density (vehicle/km)	Mean value	1.075899	5.900002	1.457771	8.466858
	Standard error	0.007279	0.039991	0.013194	0.072797
	Median	1.11	5.33	1.32	7.28
	Mode	1.17	5.35	1.295	4.8
	Variance	0.137273	4.14377	0.451078	13.73066
	Kurtosis	1.165419	12.17008	1.090342	0.184097
	Skewness	0.325542	2.920906	0.981555	1.00062
Rate of changing out lanes (times)	Mean value	11.76071	23.35006	47.30066	38.01775
	Standard error	0.158519	0.280317	0.458097	0.342657
	Median	10	20	42	34
	Mode	8	18	36	24
	Variance	65.1072	203.5944	543.7285	304.219
	Kurtosis	1.640203	2.948731	1.678656	0.115027
	Skewness	1.179921	1.461843	1.128802	0.797563
Rate of changing into lanes (times)	Mean value	13.05558	29.15515	48.31223	29.90621
	Standard error	0.21225	0.363476	0.414709	0.298989
	Median	10	24	44	26
	Mode	8	28	52	24
	Variance	116.7251	342.3103	445.6102	231.6209
	Kurtosis	3.971726	2.844836	0.08808	0.28378
	Skewness	1.729006	1.485479	0.784475	0.80799

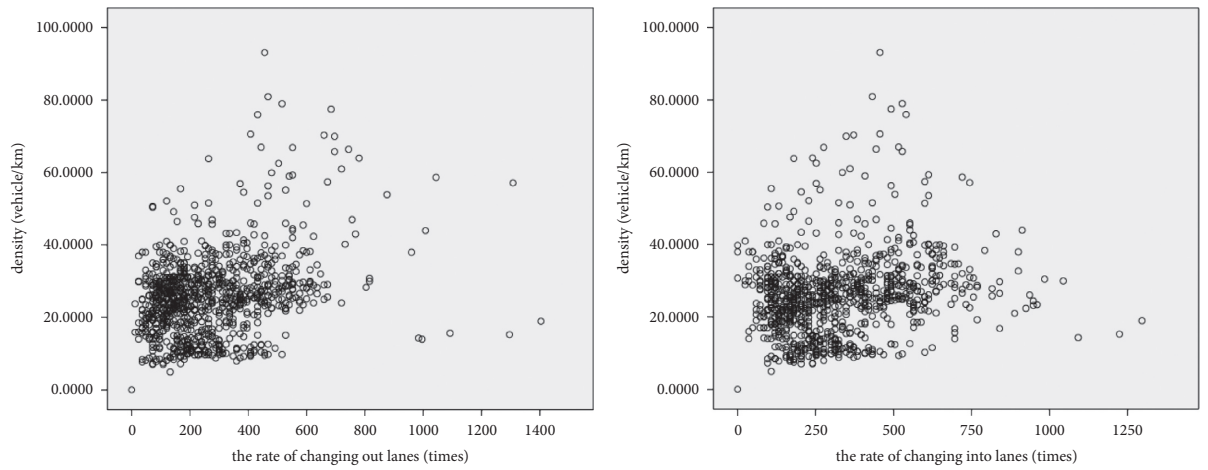


(a)

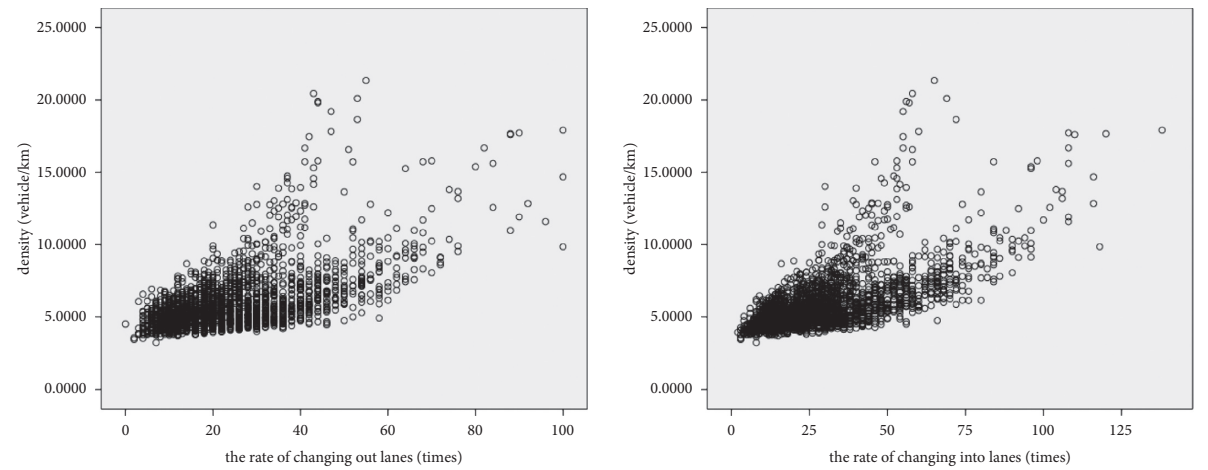


(b)

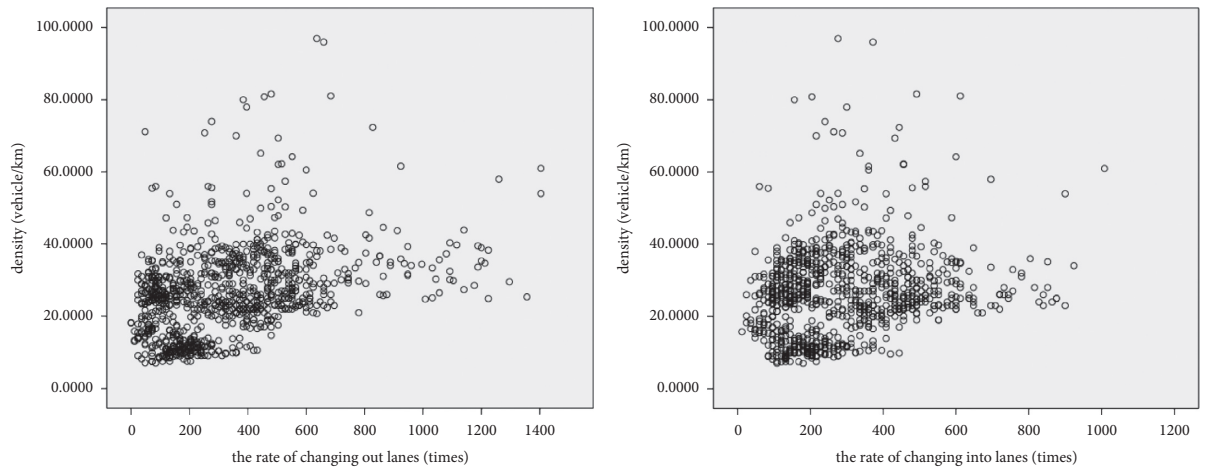
Figure 6: Continued.



(c)



(d)



(e)

Figure 6: Continued.

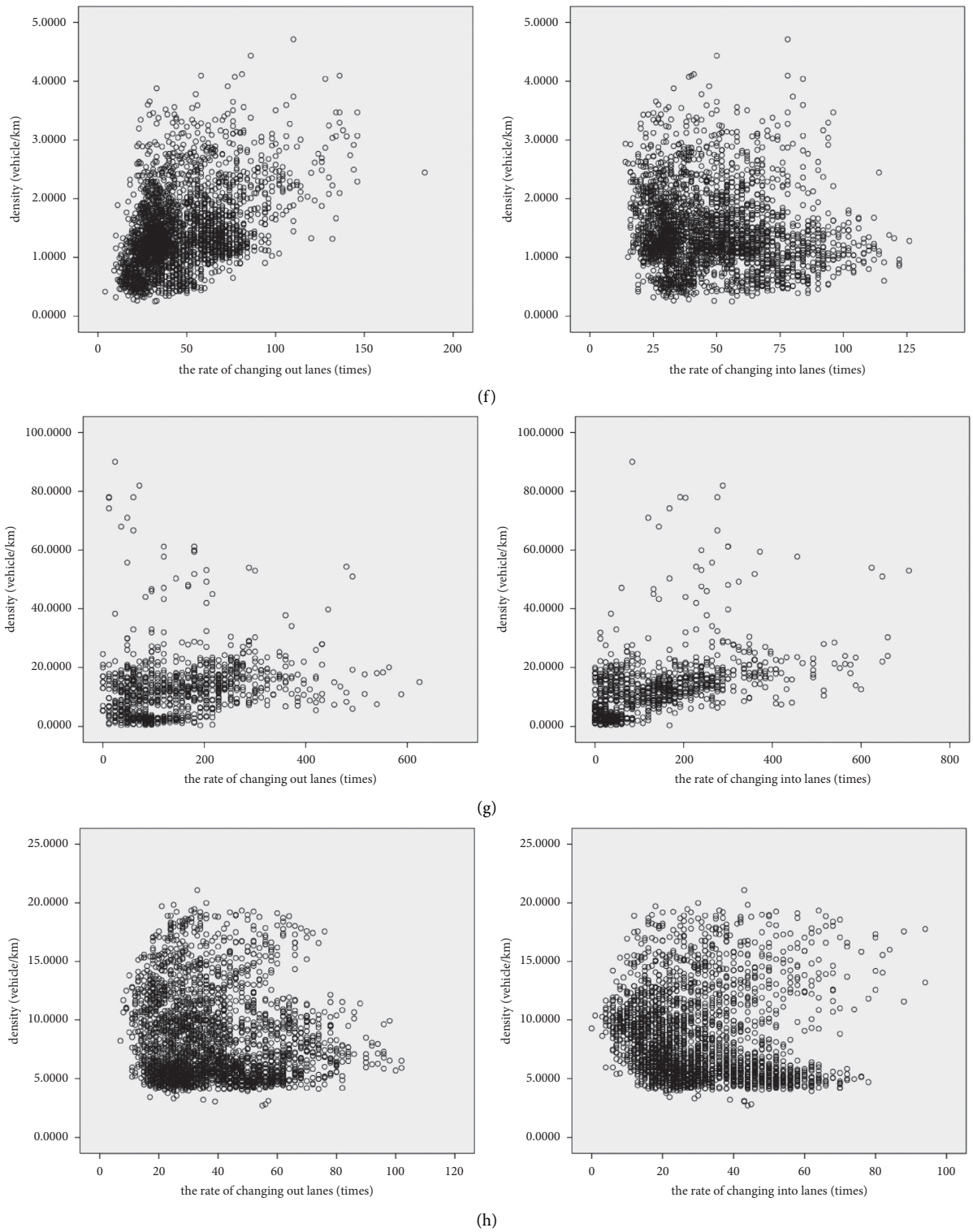


Figure 6: Scatter diagram of empirical and simulation models for the relationship between density and the rate of changing out lanes and between density and the rate of changing into lanes of a single lane. (a) The empirical data of Lane 1. (b) The simulation data of Lane 1. (c) The empirical data of Lane 2. (d) The simulation data of Lane 2. (e) The empirical data of Lane 3. (f) The simulation data of Lane 3. (g) The empirical data of Lane 4. (h) The simulation data of Lane 4.

Table 3: The data of density and the number of lane-changing times of Lane 1 and Lane 2.

k1 (vehicle/km)	k2 (vehicle/km)	LCo1 (times)	LCo2 (times)	LCi1 (times)	LCi2 (times)
41.3534	49.397	276	732	420	588
45.8286	57.419	216	672	276	600
45.9102	51.4407	360	600	348	600
51.3625	65.8116	204	696	264	528
46.316	58.6692	516	1044	768	720
43.6069	57.1446	492	1308	828	744
54.4434	66.4226	240	744	312	444
47.5244	59.9603	696	480	408	336
32.8996	46.9474	216	276	180	444
32.6967	40.62	108	396	156	564
39.4808	42.3747	156	624	384	396
25.3695	26.1902	144	384	192	300
29.7359	36.3863	108	492	300	360
31.3534	38.9231	252	396	228	504
23.1025	44.1669	156	1008	492	912
21.307	43.1977	192	768	360	828
22.7805	38.1319	204	960	360	900
23.1025	36.3308	84	252	144	408
23.7316	43.2011	96	336	192	252
26.0077	45.7935	48	420	240	408
22.6332	35.1054	12	168	60	132
20.5373	25.0703	48	120	120	276
30.3718	39.5103	12	96	24	192
23.2872	42.0075	60	552	288	516
20.4341	31.7524	60	540	300	528
25.9159	33.3484	12	168	84	156
20.6745	30.2142	24	24	12	108
23.1988	35.4795	12	72	36	192
23.3655	37.1458	60	384	180	276
22.4643	38.4357	60	180	96	324
50.7629	56.3251	252	468	216	492
61.4895	59.3006	372	552	168	612
77.8395	70.6239	264	408	264	456
85.3344	77.5126	324	684	420	492
65.3061	67.3181	240	444	192	516
68.3836	69.9785	216	696	492	348
71.4634	75.5906	192	432	228	540
73.3314	79.4491	144	516	204	528
92.4337	80.9599	240	468	264	432
88.489	93.1523	204	456	264	456
23.5978	31.4044	24	108	48	96
24.5965	34.4922	36	132	48	96
21.9508	31.8681	36	72	36	72
18.5071	30.3185	0	60	24	24
22.5266	34.0184	0	84	0	36
25.2426	40.8139	0	120	48	24
23.99	38.4165	24	36	24	48
23.5253	38.1263	0	108	0	48
28.4573	38.2759	0	96	24	0
22.8712	36.2333	12	108	48	60
21.2639	46.638	192	756	468	384
40.0034	64.448	168	780	276	240
33.9265	60.908	300	720	312	360
31.8102	59.3811	300	540	324	408
21.7264	38.3303	12	48	36	36
19.7166	29.0354	48	264	84	324
17.793	28.6409	12	336	84	288
17.6955	30.2176	0	216	24	420
16.6629	26.9723	24	204	96	408
19.8912	29.3868	12	264	108	336

Table 3: Continued.

k1 (vehicle/km)	k2 (vehicle/km)	LCo1 (times)	LCo2 (times)	LCi1 (times)	LCi2 (times)
19.0909	29.4854	48	336	156	372
19.6112	32.2682	0	360	144	420
18.6976	33.3303	12	444	192	360
18.4969	30.6064	48	288	60	480
20.8796	34.329	12	276	132	528
21.9531	34.1578	36	300	84	576
17.6105	30.9374	0	360	108	324
20.6801	35.3208	12	348	120	384
22.1571	37.4518	12	324	120	372
21.2866	34.4638	48	384	132	480
22.6139	36.7649	24	396	48	456
21.5212	36.7411	36	444	132	348
24.584	41.1188	36	240	84	516
21.808	38.5786	48	276	84	636
20.0102	35.5554	36	420	192	408
18.599	33.0537	24	348	192	396
19.9377	35.204	24	420	156	504
20.2188	31.3829	0	264	96	576
20.2131	36.3557	24	432	96	480
22.0177	31.7434	60	420	156	456
23.9129	36.3455	48	420	108	648
25.5588	23.1977	144	456	288	948
26.5212	23.3961	96	468	228	960
25.8422	24.4366	96	456	252	948
25.0057	22.5448	132	528	312	744
19.4831	31.5144	12	228	84	648
20.781	35.4636	60	276	60	444
23.0027	39.3199	12	324	48	696
22.1072	34.7121	48	336	84	684
19.6293	34.1419	48	240	36	660
19.6803	31.2469	24	192	72	696
21.1879	34.7733	48	252	60	408
22.8225	37.6026	0	336	60	624
22.6615	34.0784	72	360	108	624
19.5182	33.7168	60	276	48	588
17.0642	37.3883	24	252	12	468
18.7214	39.8209	504	252	12	0
19.5749	41.3942	24	384	96	480
18.3122	35.9703	12	288	120	444

behavior under different self-driving vehicle penetration rates is very important for understanding the traffic operation mechanism of mixed traffic. In the simulation process, we use the CACC vehicle to simulate the self-driving vehicle, and the remaining settings of the simulation environment remain unchanged. The simulation experiment adopts 8 states that the CACC vehicle penetration rate is designed to be 10%, 20%, 30%, 40%, 50%, 60%, 70%, and 80%. Among them, the permeability is represented by P. The simulation collects the traffic density, the rate of changing into lanes, and the rate of changing out lanes data of each lane at a time interval of 5 minutes and analyzes them. Finally, the relationship between density and lane-changing behavior in the mixed traffic is obtained.

3.2. Permeability Simulation Experiment Results of Different Autonomous Vehicles. After 8 rounds of experiments, we obtained the data of density, the rate of changing out lanes, and the rate of changing into lanes under the penetration rate of autonomous vehicles is 10%, 20%, 30%, 40%, 50%, 60%, 70%, and 80%. For a better comparison, the 8-round test data is compared with the data without autonomous vehicles. The data distribution is shown in Figures 7–9.

In Figure 7, with the penetration of autonomous vehicles, the density still shows more outliers, but the number decreases with the addition of autonomous vehicles. In 25% and 75% quantile of density, as the penetration rate of autonomous vehicles increases, Lane 1 to Lane 3 are gradually approaching, but Lane 4 increases with the increase in penetration rate. The 50% quantile of density is still close to the 25% quantile. The average density of Lane 1 is still the smallest, and the average density of Lane 4 is still the largest. From the state shown in Figure 7, the addition of autonomous vehicles has a certain improvement in the density of lanes from Lane 1 to Lane 3, while Lane 4 is not obvious and even has an increasing trend.

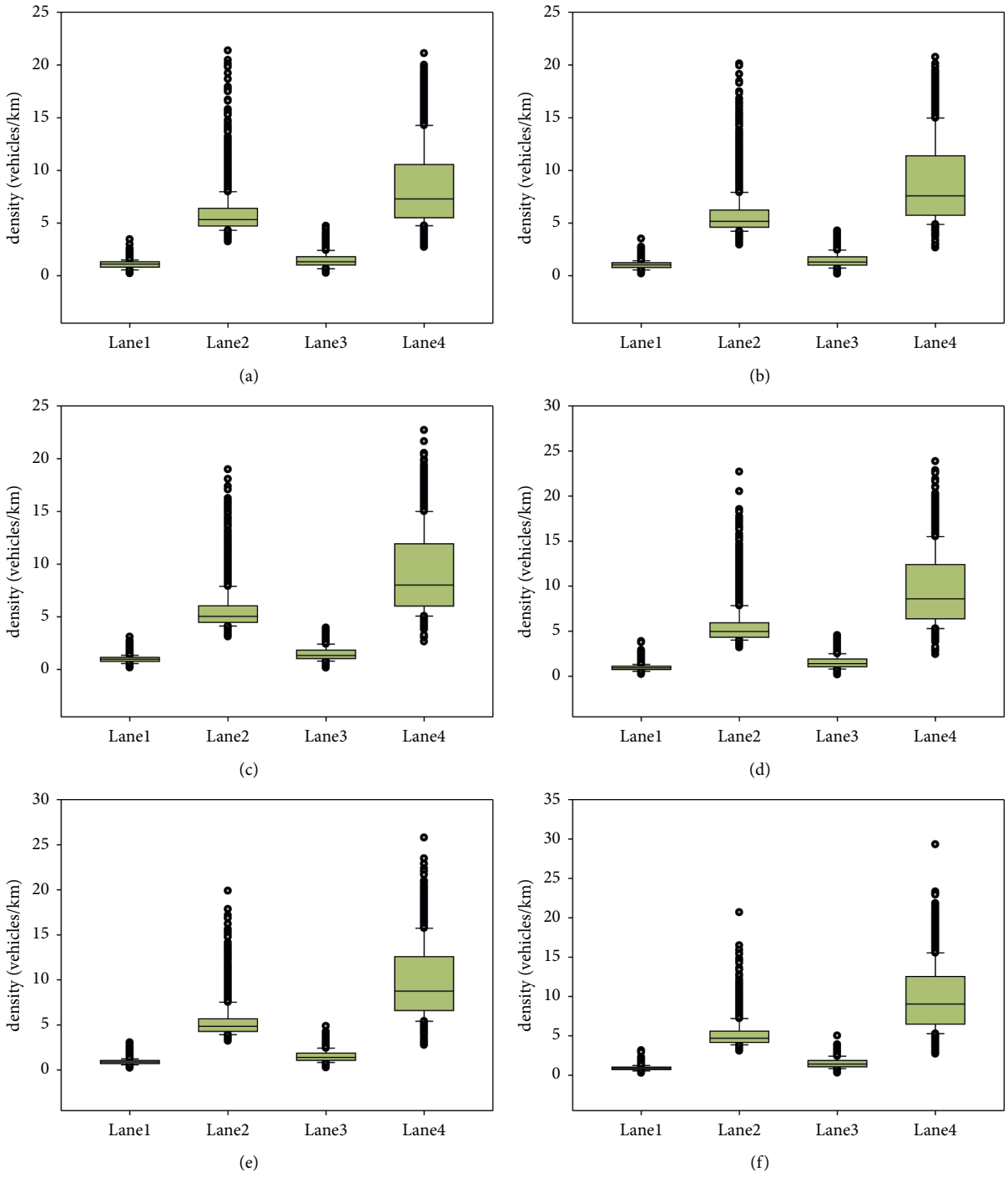


Figure 7: Continued.

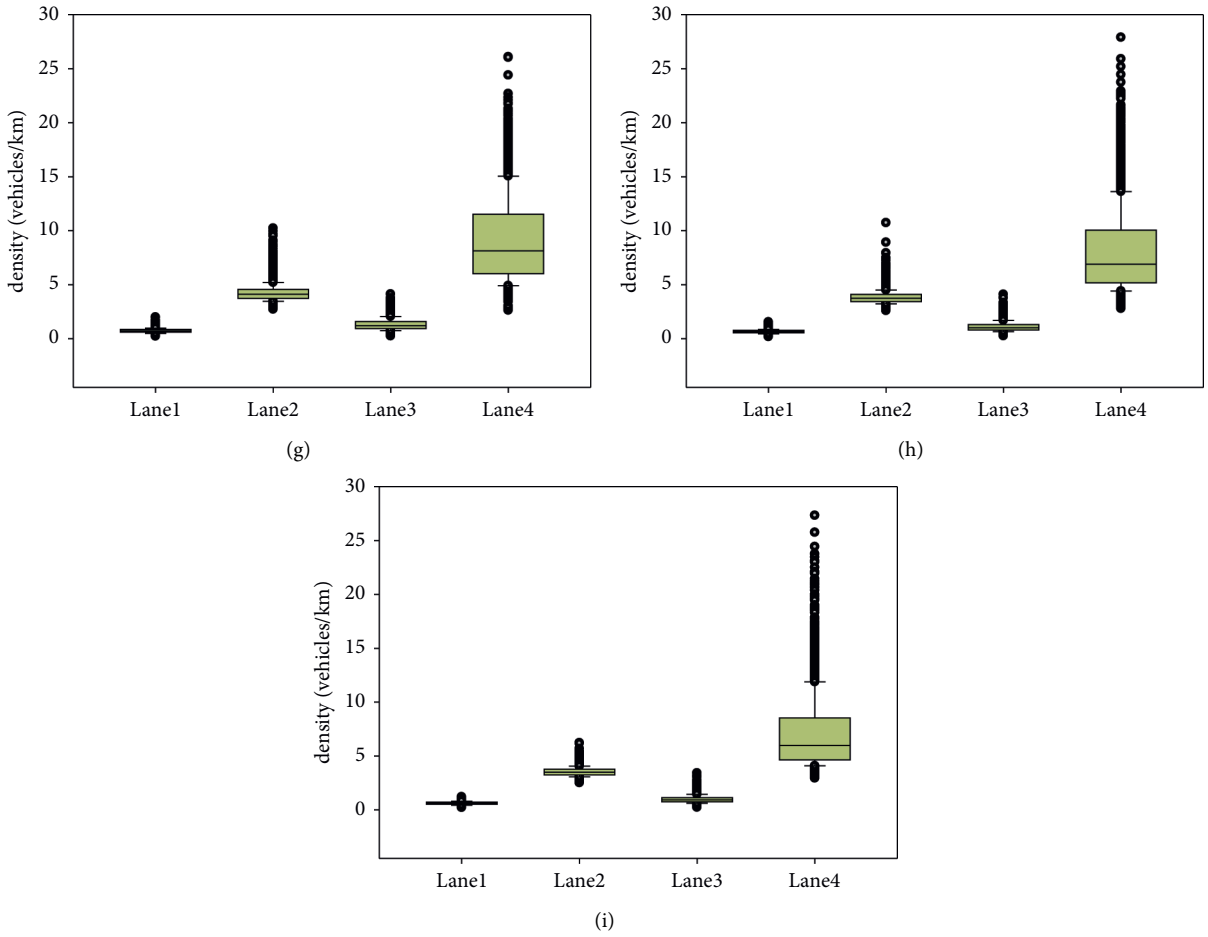


FIGURE 7: The density distribution of a single lane under different autonomous vehicle penetration rates. The lane density distribution map (a) for the 0% joining rate of autonomous vehicles, (b) for the 10% joining rate of autonomous vehicles, (c) for the 20% joining rate of autonomous vehicles, (d) for the 30% joining rate of autonomous vehicles, (e) for the 40% joining rate of autonomous vehicles, (f) for the 50% joining rate of autonomous vehicles, (g) for the 60% joining rate of autonomous vehicles, (h) for the 70% joining rate of autonomous vehicles, and (i) for the 80% joining rate of autonomous vehicles.

Figures 8 and 9 are generally close in performance. With the penetration of autonomous vehicles, there are still many outliers in the data. In Figure 8, the data distribution in 25% and 75% is relatively close. There is no significant difference. At the same time, the 50% quantile is still close to the 25% quantile. In Figure 8, compared to the other three lanes, the rate of changing into lanes of Lane 1 has a large gap in both the mean value and interval. In Figure 9, the gap between Lane 1 and the other three lanes is not as obvious as in Figure 8. This is mainly because, compared to Lane 1, most vehicles will be less affected by the entrance and exit because Lane 1 is farther away. At the same time, Lane 1 is less dense and less traffic pressure, so more choose to change into this lane, unless necessary unwilling to leave the lane. The above analysis can be further verified from Tables 4 to 6 of the statistical description data.

In Table 4, the data representations of Lane 1 and Lane 2 are relatively close. The average density decreases slightly with the increase in the penetration rate of autonomous vehicles. Before the penetration rate is 60%, as the penetration rate increases. The density drops from 1% to 5%, but at 60%, the density drops by more than 17%, at 70%, it drops slightly, and at 80%, it is slightly lower than 70%. As the penetration rate in Lane 3 increases, the density increases and then decreases before 60% and suddenly decreases when it reaches 60%, and the degree of 70% decrease increases, and 80% decreases slightly from 70%. Lane 4 has an increase in density with the penetration of autonomous vehicles. The increase reaches the highest penetration rate at 30% and then begins to decrease at 40%, but the density increases compared to 30%, until 60%. It begins to decline, and the rate of decline increases with the increase of permeability, but the rate of decrease of density at 80% is less than that at 70%. However, changes in the average density data show that when the penetration rate of autonomous vehicles reaches 70%, the density of all lanes decreases compared with the situation without the autonomous vehicles. The density median and density mode decrease slightly in Lane 1 and Lane 2 and have a similar trend, while those in Lane 3 and Lane 4 increase and decrease and then decrease after the 60% quantile. The 75% quantile density decreases with the penetration of autonomous vehicles from

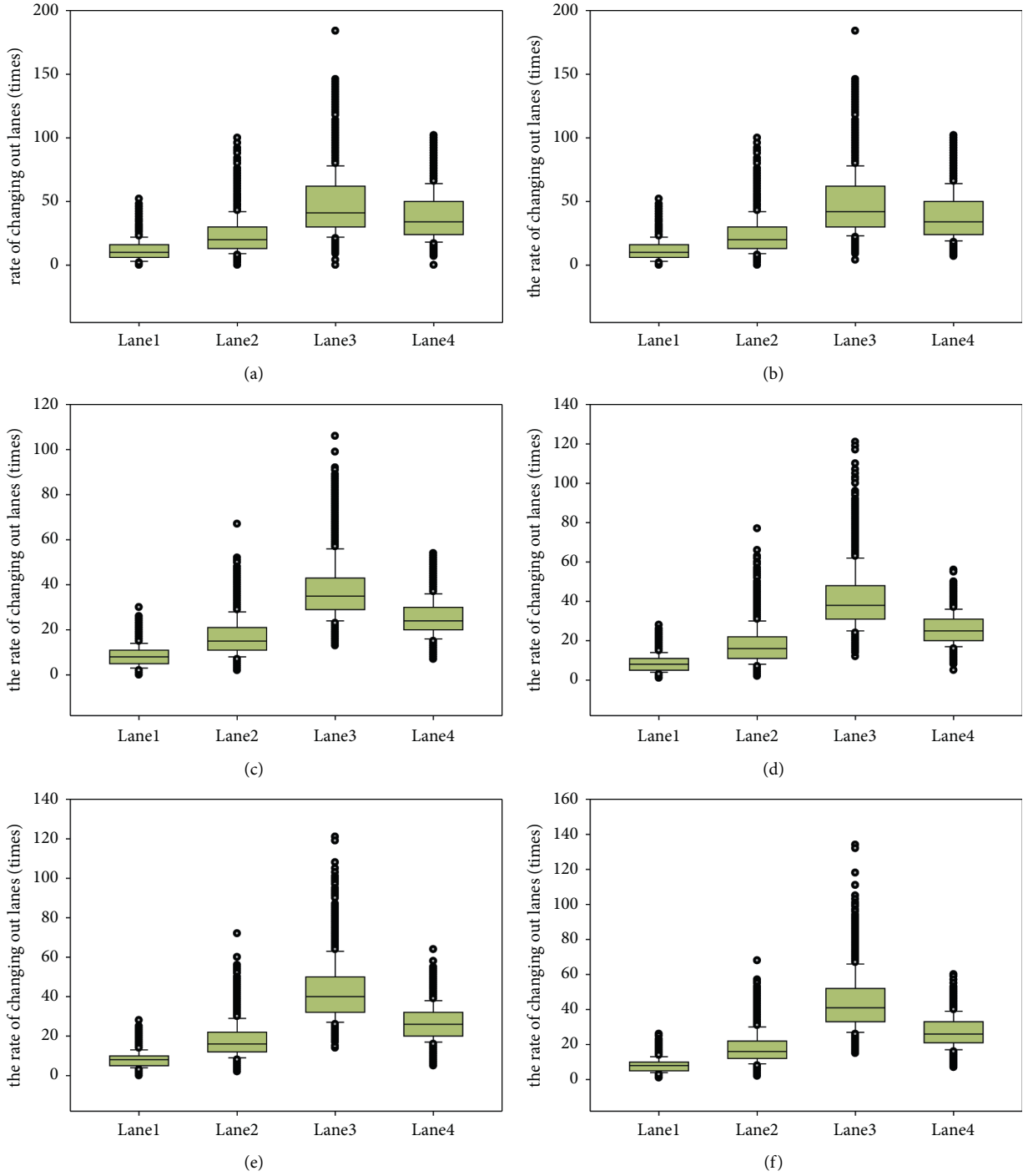


Figure 8: Continued.

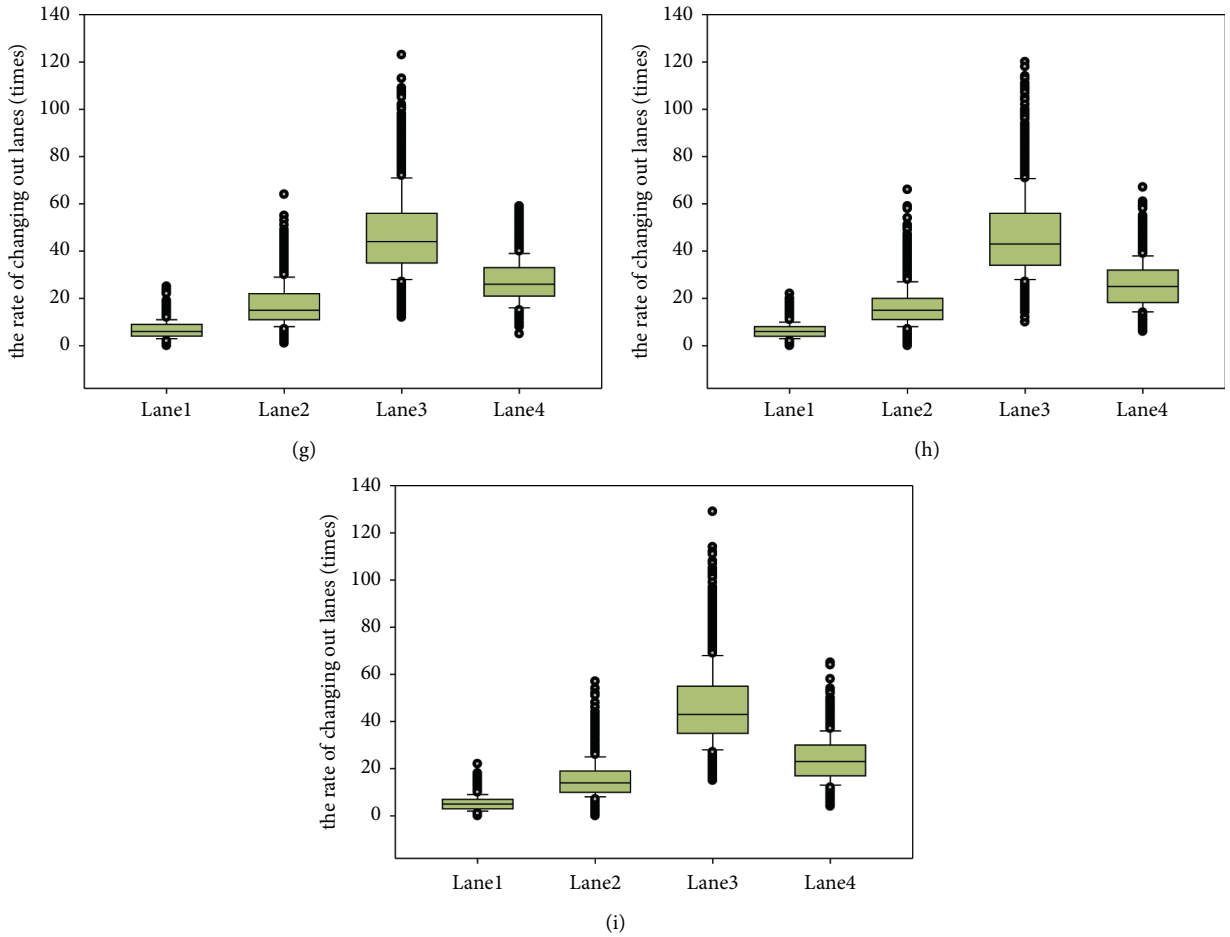


Figure 8: Distribution of the rate of changing out lanes for a single lane under different penetration rates of autonomous vehicles. Distribution map of the rate of changing out lanes (a) for the 0% joining rate of autonomous vehicles, (b) for the 10% joining rate of autonomous vehicles, (c) for the 20% joining rate of autonomous vehicles, (d) for the 30% joining rate of autonomous vehicles, (e) for the 40% joining rate of autonomous vehicles, (f) for the 50% joining rate of autonomous vehicles, (g) for the 60% joining rate of autonomous vehicles, (h) for the 70% joining rate of autonomous vehicles, and (i) for the 80% joining rate of autonomous vehicles.

Lane 1, while Lane 4 has increased or decreased with little change. The 25% quantile density increased slightly in Lane 1 and Lane 2 with little change, while that in Lane 3 and Lane 4 increased and decreased. However, both the 25% and 75% quantile densities converge towards the mode. Taking a comprehensive look at the density in Table 4, the average density from Lane 1 and Lane 2 decreases with the penetration of autonomous vehicles, while that from Lane 3 and Lane 4 only begins to decrease significantly when the penetration rate reaches 60%. But when the penetration rate reaches 80%, the density of all lanes is lower than that without autonomous vehicles. And the 75% and 25% quantile densities both tend to converge towards the mode density. It shows that the penetration of autonomous vehicles affects for the road traffic density. As the penetration rate exceeds 60%, the density decreases significantly, and the density is further converged. In combination with Figure 7, it can be seen that as the penetration rate of autonomous vehicles increases, the density presents a nonnormal distribution, and the data is concentrated in the direction of the mode.

The impact of autonomous vehicles on density can be seen in Table 4 and Figure 7. After the autonomous vehicles have penetrated, the density of each lane begins to decrease when the penetration rate reaches more than 60%. Lane 3 and Lane 4 are close to the entrance and exit, especially Lane 4, which becomes the lane with the largest distance between the 25% and 75% quantile in the density distribution. It can be seen that the addition of autonomous vehicles can improve the road density, but this kind of improvement of the whole road section will be highlighted until the penetration rate reaches more than 60%. Otherwise, for multilane roads, the improvement of lanes farther from the entrance and exit will be more obvious. However, the lanes that are close to the entrance or directly connected to the entrance and exit are not significantly improved at the beginning. Autonomous vehicles have a direct impact on lane-changing behavior, as shown in Tables 5 and 6.

In Table 5, when the penetration rate of self-driving vehicles reaches 10%, the average rate of changing out lanes decreases significantly, and this data increases as the penetration rate increases. It starts to decrease after 40% of

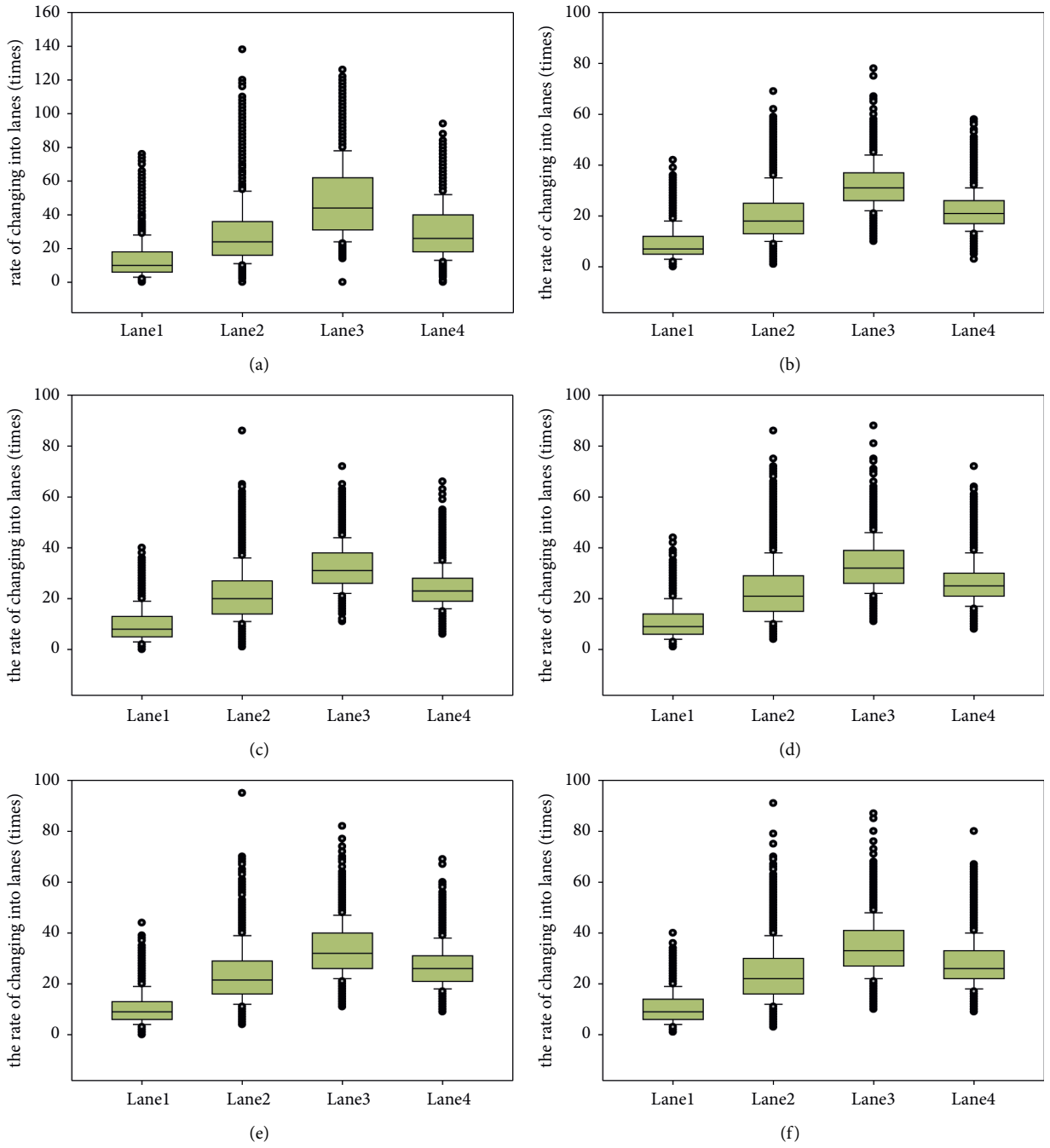


Figure 9: Continued.

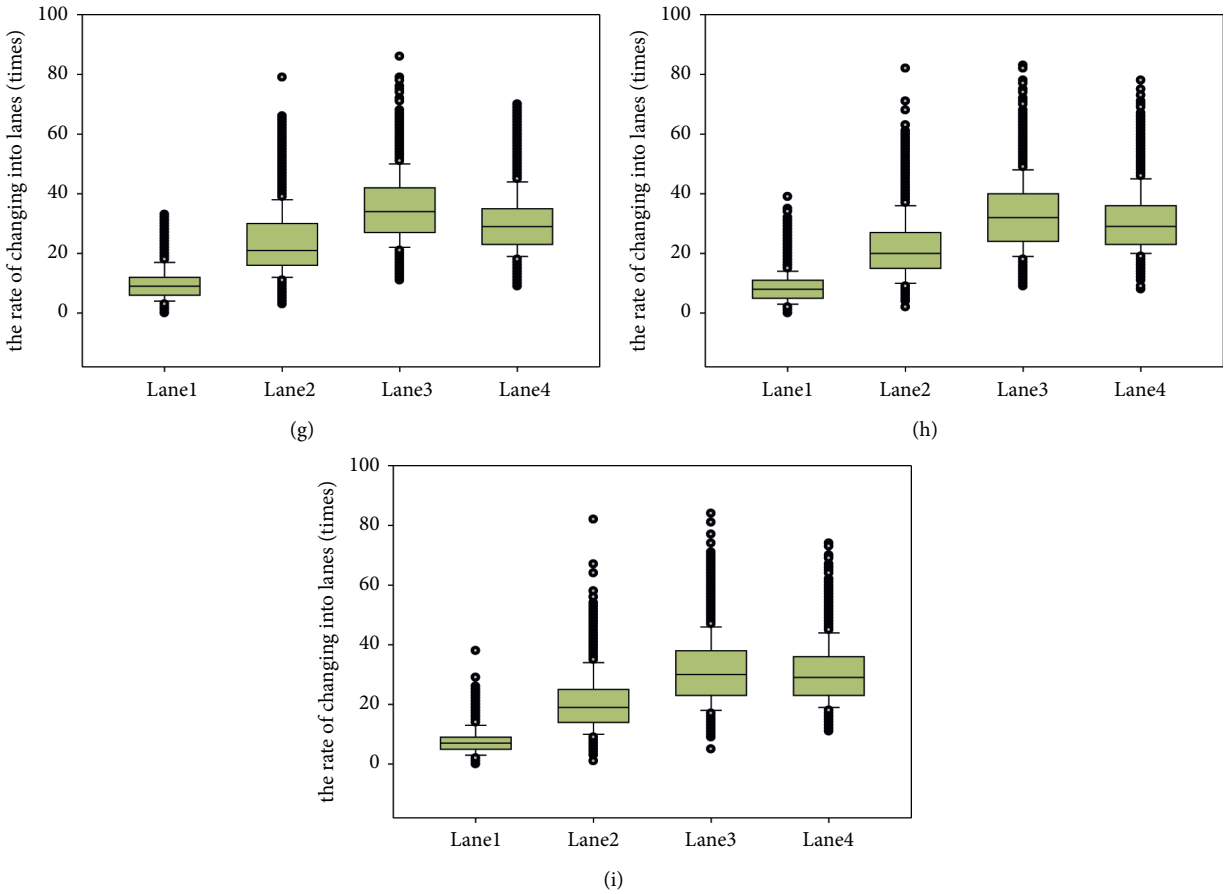


Figure 9: The distribution map of the rate of changing into lanes of a single lane under different autonomous vehicle penetration rates. Distribution map of the rate of changing into lanes (a) for the 0% joining rate of autonomous vehicles, (b) for the 10% joining rate of autonomous vehicles, (c) for the 20% joining rate of autonomous vehicles, (d) for the 30% joining rate of autonomous vehicles, (e) for the 40% joining rate of autonomous vehicles, (f) for the 50% joining rate of autonomous vehicles, (g) for the 60% joining rate of autonomous vehicles, (h) for the 70% joining rate of autonomous vehicles, and (i) for the 80% joining rate of autonomous vehicles.

Table 4: The density of a single lane under different autonomous vehicle penetration rates by the statistical description (unit: vehicle/km).

		$P = 0\%$	$P = 10\%$	$P = 20\%$	$P = 30\%$	$P = 40\%$	$P = 50\%$	$P = 60\%$	$P = 70\%$	$P = 80\%$
Lane 1	Mean value	1.076	1.015	0.971	0.956	0.913	0.889	0.741	0.666	0.626
	Standard error	0.007	0.007	0.006	0.007	0.006	0.006	0.004	0.003	0.003
	Median	1.110	1.030	0.970	0.930	0.895	0.865	0.730	0.660	0.625
	Mode	1.170	1.100	1.070	0.995	0.985	0.830	0.790	0.610	0.630
	Variance	0.137	0.122	0.105	0.120	0.090	0.082	0.039	0.027	0.021
	Kurtosis	1.165	1.854	2.864	6.652	4.397	4.144	1.574	0.883	0.020
	Skewness	0.326	0.527	0.838	1.526	1.256	1.191	0.545	0.391	0.169
	75%	2.170	2.005	1.975	2.250	2.020	1.825	1.275	1.115	0.985
	25%	0.350	0.335	0.350	0.345	0.355	0.375	0.330	0.310	0.310
Lane 2	Mean value	5.900	5.810	5.640	5.581	5.373	5.183	4.269	3.826	3.535
	Standard error	0.040	0.041	0.038	0.040	0.036	0.032	0.016	0.012	0.008
	Median	5.330	5.160	5.040	4.965	4.840	4.695	4.125	3.738	3.490
	Mode	5.350	5.065	4.530	4.160	4.825	4.115	3.870	3.255	3.295
	Variance	4.144	4.416	3.777	4.224	3.388	2.698	0.693	0.354	0.169
	Kurtosis	12.170	9.929	8.160	9.814	8.766	9.959	7.289	13.361	2.463
	Skewness	2.921	2.771	2.514	2.677	2.572	2.538	2.063	2.264	0.941
	75%	15.260	15.235	13.845	13.935	12.765	11.610	7.620	6.015	4.875
	25%	3.840	3.690	3.650	3.560	3.490	3.395	3.105	2.900	2.785

Table 4: Continued.

		$P = 0\%$	$P = 10\%$	$P = 20\%$	$P = 30\%$	$P = 40\%$	$P = 50\%$	$P = 60\%$	$P = 70\%$	$P = 80\%$
Lane 3	Mean value	1.458	1.453	1.474	1.538	1.529	1.524	1.322	1.112	0.983
	Standard error	0.013	0.013	0.013	0.013	0.012	0.012	0.011	0.009	0.007
	Median	1.320	1.295	1.318	1.395	1.400	1.400	1.215	1.015	0.920
	Mode	1.295	1.000	1.225	1.240	1.240	1.055	1.200	0.875	0.945
	Variance	0.451	0.452	0.410	0.451	0.398	0.392	0.292	0.202	0.133
	Kurtosis	1.090	0.780	0.598	0.654	0.939	0.708	1.793	3.493	4.681
	Skewness	0.982	0.969	0.905	0.886	0.934	0.861	1.125	1.477	1.632
	75%	3.470	3.365	3.330	3.505	3.400	3.340	3.035	2.575	2.330
	25%	0.400	0.350	0.395	0.415	0.500	0.500	0.435	0.425	0.420
Lane 4	Mean value	8.467	8.806	9.155	9.627	9.772	9.791	9.147	8.092	7.101
	Standard error	0.073	0.075	0.075	0.077	0.077	0.078	0.078	0.075	0.066
	Median	7.280	7.570	8.008	8.595	8.760	9.030	8.130	6.908	5.980
	Mode	4.800	5.590	7.060	5.445	7.175	6.290	5.155	4.725	4.205
	Variance	13.731	14.607	14.404	15.424	15.219	15.569	15.728	14.718	11.444
	Kurtosis	0.184	−0.226	−0.406	−0.343	−0.342	−0.218	0.417	1.662	3.503
	Skewness	1.001	0.861	0.742	0.711	0.700	0.658	0.957	1.324	1.693
	75%	18.715	18.470	18.480	19.345	19.300	19.745	20.110	20.375	18.875
	25%	4.130	4.055	4.080	4.125	4.220	4.110	3.780	3.545	3.465

Table 5: The rate of changing out lanes of a single lane under different autonomous vehicle penetration rates by statistical description (unit: times).

		$P = 0\%$	$P = 10\%$	$P = 20\%$	$P = 30\%$	$P = 40\%$	$P = 50\%$	$P = 60\%$	$P = 70\%$	$P = 80\%$
Lane 1	Mean value	1.076	1.015	0.971	0.956	0.913	0.889	0.741	0.666	0.626
	Standard error	0.007	0.007	0.006	0.007	0.006	0.006	0.004	0.003	0.003
	Median	1.110	1.030	0.970	0.930	0.895	0.865	0.730	0.660	0.625
	Mode	1.170	1.100	1.070	0.995	0.985	0.830	0.790	0.610	0.630
	Variance	0.137	0.122	0.105	0.120	0.090	0.082	0.039	0.027	0.021
	Kurtosis	1.165	1.854	2.864	6.652	4.397	4.144	1.574	0.883	0.020
	Skewness	0.326	0.527	0.838	1.526	1.256	1.191	0.545	0.391	0.169
	75%	2.170	2.005	1.975	2.250	2.020	1.825	1.275	1.115	0.985
	25%	0.350	0.335	0.350	0.345	0.355	0.375	0.330	0.310	0.310
Lane 2	Mean value	5.900	5.810	5.640	5.581	5.373	5.183	4.269	3.826	3.535
	Standard error	0.040	0.041	0.038	0.040	0.036	0.032	0.016	0.012	0.008
	Median	5.330	5.160	5.040	4.965	4.840	4.695	4.125	3.738	3.490
	Mode	5.350	5.065	4.530	4.160	4.825	4.115	3.870	3.255	3.295
	Variance	4.144	4.416	3.777	4.224	3.388	2.698	0.693	0.354	0.169
	Kurtosis	12.170	9.929	8.160	9.814	8.766	9.959	7.289	13.361	2.463
	Skewness	2.921	2.771	2.514	2.677	2.572	2.538	2.063	2.264	0.941
	75%	15.260	15.235	13.845	13.935	12.765	11.610	7.620	6.015	4.875
	25%	3.840	3.690	3.650	3.560	3.490	3.395	3.105	2.900	2.785
Lane 3	Mean value	1.458	1.453	1.474	1.538	1.529	1.524	1.322	1.112	0.983
	Standard error	0.013	0.013	0.013	0.013	0.012	0.012	0.011	0.009	0.007
	Median	1.320	1.295	1.318	1.395	1.400	1.400	1.215	1.015	0.920
	Mode	1.295	1.000	1.225	1.240	1.240	1.055	1.200	0.875	0.945
	Variance	0.451	0.452	0.410	0.451	0.398	0.392	0.292	0.202	0.133
	Kurtosis	1.090	0.780	0.598	0.654	0.939	0.708	1.793	3.493	4.681
	Skewness	0.982	0.969	0.905	0.886	0.934	0.861	1.125	1.477	1.632
	75%	3.470	3.365	3.330	3.505	3.400	3.340	3.035	2.575	2.330
	25%	0.400	0.350	0.395	0.415	0.500	0.500	0.435	0.425	0.420
Lane 4	Mean value	8.467	8.806	9.155	9.627	9.772	9.791	9.147	8.092	7.101
	Standard error	0.073	0.075	0.075	0.077	0.077	0.078	0.078	0.075	0.066
	Median	7.280	7.570	8.008	8.595	8.760	9.030	8.130	6.908	5.980
	Mode	4.800	5.590	7.060	5.445	7.175	6.290	5.155	4.725	4.205
	Variance	13.731	14.607	14.404	15.424	15.219	15.569	15.728	14.718	11.444
	Kurtosis	0.184	−0.226	−0.406	−0.343	−0.342	−0.218	0.417	1.662	3.503
	Skewness	1.001	0.861	0.742	0.711	0.700	0.658	0.957	1.324	1.693
	75%	18.715	18.470	18.480	19.345	19.300	19.745	20.110	20.375	18.875
	25%	4.130	4.055	4.080	4.125	4.220	4.110	3.780	3.545	3.465

Table 6: The rate of changing into lanes of a single lane under different autonomous vehicle penetration rates by statistical description (unit: times).

		$P = 0\%$	$P = 10\%$	$P = 20\%$	$P = 30\%$	$P = 40\%$	$P = 50\%$	$P = 60\%$	$P = 70\%$	$P = 80\%$
Lane 1	Mean value	11.761	8.075	8.151	8.318	8.174	8.084	6.857	5.926	5.500
	Standard error	0.159	0.087	0.083	0.082	0.076	0.076	0.067	0.059	0.056
	Median	10.000	8.000	8.000	8.000	8.000	8.000	6.000	6.000	5.000
	Mode	8.000	6.000	6.000	6.000	6.000	6.000	5.000	5.000	4.000
	Variance	65.107	19.75	17.88	17.50	15.04	15.028	11.612	9.024	8.242
	Kurtosis	1.640	0.829	0.842	1.221	1.084	0.953	1.072	1.732	1.382
	Skewness	1.180	0.845	0.801	0.963	0.839	0.889	0.805	0.966	0.899
	75%	38.000	21.00	20.00	22.00	20.00	20.000	16.000	15.000	14.000
	25%	0.000	1.000	1.000	2.000	1.000	2.000	1.000	1.000	1.000
Lane 2	Mean value	23.350	16.11	16.73	17.64	17.65	17.983	17.254	16.065	15.286
	Standard error	0.280	0.161	0.162	0.181	0.166	0.174	0.167	0.157	0.143
	Median	20.000	14.00	15.00	16.00	16.00	16.000	15.000	15.000	14.000
	Mode	18.000	14.00	13.00	12.00	14.00	13.000	12.000	11.000	11.000
	Variance	203.59	66.69	68.34	84.86	71.45	78.740	71.974	63.963	53.225
	Kurtosis	2.949	2.163	1.877	3.032	2.510	2.126	1.594	2.658	2.387
	Skewness	1.462	1.335	1.239	1.445	1.311	1.275	1.145	1.323	1.243
	75%	70.000	44.00	44.00	49.00	46.00	47.000	44.000	43.000	39.000
	25%	4.000	4.000	4.000	5.000	5.000	4.000	4.000	4.000	3.000
Lane 3	Mean value	47.301	34.71	37.64	41.20	42.63	44.017	47.123	46.893	45.844
	Standard error	0.458	0.241	0.260	0.298	0.291	0.309	0.335	0.339	0.316
	Median	42.000	32.00	35.00	38.00	40.00	41.000	44.000	43.000	43.000
	Mode	36.000	30.00	33.00	32.00	32.00	37.000	42.000	36.000	41.000
	Variance	543.72	150.6	175.2	230.7	218.8	247.65	290.65	297.918	259.274
	Kurtosis	1.679	2.861	2.250	2.248	1.934	1.732	0.814	1.004	1.160
	Skewness	1.129	1.481	1.344	1.310	1.210	1.108	0.946	0.986	0.975
	75%	126.000	77.000	83.000	89.000	91.000	93.000	98.000	99.000	94.000
	25%	14.000	16.000	17.000	18.000	19.000	19.000	20.000	20.000	20.000
Lane 4	Mean value	38.018	25.357	25.288	26.016	26.571	26.881	27.165	25.524	23.642
	Standard error	0.343	0.149	0.148	0.152	0.164	0.165	0.174	0.180	0.173
	Median	34.000	24.000	24.000	25.000	26.000	26.000	26.000	25.000	23.000
	Mode	17.442	7.574	7.548	7.738	8.341	8.377	8.843	9.171	8.813
	Variance	304.219	57.369	56.978	59.879	69.569	70.169	78.200	84.110	77.671
	Kurtosis	0.115	1.349	−0.052	−0.118	0.273	−0.053	0.010	0.173	−0.001
	Skewness	0.798	0.683	0.497	0.449	0.581	0.507	0.528	0.562	0.534
	75%	86.000	46.000	45.000	46.000	49.000	49.000	51.000	51.000	46.000
	25%	12.000	11.000	11.000	11.000	11.000	11.000	11.000	9.000	8.000

Lane 1. Lane 2 did not begin to decline until 60%, and the data for Lane 3 and Lane 4 did not show a downward trend until the penetration rate reached 70%. Although when the penetration rate reaches 80%, the average rate of changing out lanes decreases compared with the case of no autonomous vehicle, but in 10% case of the penetration rate of autonomous vehicles, lane-changing behavior of all lanes has an obvious decrease, as the penetration rate increases, the lane-changing behavior will increase, and the changes in each lane are also different. The median and mode of lane swapping rates show a downward trend in Lane 1, Lane 2, and Lane 4, while Lane 3 is not obvious. The 75% quantile of the lane swap rate showed a downward trend in Lane 1 to Lane 4, while the 25% quantile showed an upward trend in Lane 3, Lane 4 showed a downward trend, and Lane 1 and Lane 2 were not obvious. The final data converges towards the mode. The situation in Table 6 is similar to that in Table 5, except that the average rate of changing into lanes in Lane 4 decreases sharply with the addition of autonomous vehicles and then increases and is finally close to the situation without autonomous vehicles. The data all show a tendency to converge in the direction of the mode. The performance of the remaining data is relatively close. Combining Figures 8, 9, Tables 5, and 6, it can be seen that, with the addition of autonomous vehicles, it has an impact on lane-changing behavior. However, the rate of changing into lanes and the rate of changing out lanes are both declined or remained with no autonomous vehicle when the penetration rate reaches 80%, but as the penetration increased, the lane change rate showed a law of first decreasing and then slightly increasing, reaching a certain percentage and then decreasing. Due to the inconsistency distance between the ramps and different lanes, different lanes have different rate of changing into lanes and rate of changing out lanes. However, Lane 3, as the busiest lane on the entire road, burden on a large number of lane changes.

Table 7: The linear correlation of k and LCi, and LCo.

		Lane 1		Lane 2		Lane 3		Lane 4	
		LCo	LCi	LCo	LCi	LCo	LCi	LCo	LCi
$P = 10\%$	Pearson correlation	0.469**	0.566**	0.819**	0.838**	0.602**	0.071**	0.052**	0.218**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000
$P = 20\%$	Pearson correlation	0.508**	0.559**	0.808**	0.807**	0.606**	0.078**	0.233**	0.450**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$P = 30\%$	Pearson correlation	0.579**	0.634**	0.836**	0.814**	0.644**	0.286**	0.384**	0.600**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$P = 40\%$	Pearson correlation	0.546**	0.588**	0.823**	0.807**	0.623**	0.265**	0.453**	0.618**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$P = 50\%$	Pearson correlation	0.555**	0.579**	0.817**	0.785**	0.667**	0.412**	0.537**	0.675**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$P = 60\%$	Pearson correlation	0.475**	0.412**	0.733**	0.700**	0.714**	0.567**	0.629**	0.783**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$P = 70\%$	Pearson correlation	0.412**	0.359**	0.653**	0.653**	0.694**	0.571**	0.660**	0.807**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$P = 80\%$	Pearson correlation	0.393**	0.312**	0.560**	0.595**	0.673**	0.539**	0.630**	0.795**
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: (a) *when 0.05, significant correlation. (b) **when 0.01, significant correlation.

The lane-changing behavior of Lane 4 does not respond significantly with the addition of autonomous vehicles. However, all data show that as the penetration rate of autonomous vehicles increases, there is a certain degree of convergence in lane density, the rate of changing into lanes, and the rate of changing out lanes. Through the above analysis, we believe that the relationship between density and lane-changing behavior requires further research and analysis.

4. Analysis and Comparison

4.1. Analysis of the Relationship between Density and Lane-Changing Behavior in Mixed Traffic. From the previous analysis, we can see that there is a certain relationship between density and lane-changing behavior. In terms of the form of data, with the penetration of autonomous vehicles, the density and lane-changing behavior have improved. For the relationship between density and lane-changing behavior under different permeability, different lanes need to be analyzed accordingly.

In order to verify whether there is a linear correlation between the density and the lane change rate, the correlation analysis between the density and the lane change rate in each lane under the different penetration rates of autonomous vehicles is shown in Table 7.

It can be seen from Table 7 that, with the increase in the penetration rate of autonomous vehicles, the correlation between the density and the lane change-out rate and lane change-in rate in Lane 1 increased before the penetration rate reached 30% and then continued to decrease. Lane 2 decreased as the penetration rate increased after 30%. Lane 3 and Lane 4 have increased and decreased, and the relationship change is not obvious. The relationship between density and the rate of changing into lanes, and density and the rate of changing out lanes can be observed more accurately from the scatter plot of the relationship between them, as shown in Figures 10 and 11.

It can be seen from Figures 10 and 11 that, with the penetration of autonomous vehicles, the relationship between density and the rate of changing into lanes and the relationship between density and the rate of changing out lanes appears basically the same in the same lane, but displays in different lanes, the shapes are slightly different. However, they gather in the direction of lower density and slightly fewer lanes. In Figure 10, as the penetration rate of autonomous vehicles increases, the rate of lane swapping out increases, and the degree of dispersion decreases. Especially when the permeability increases to more than 50%, this situation is more obvious. In Figure 11, although the aggregation also shows an increasing trend, the performance is not obvious in Lane 2. In Figure 10, the density in Lane 2 and Lane 3 and the lane change rate show a relatively obvious linear relationship, while the relationship between Lane 1 and Lane 4 is not obvious. The same linear relationship is shown in the graph of the relationship in Lane 2 and Lane 4 after the penetration rate over 50% in Figure 11. In Lane 1 and Lane 3, the appearance of this relationship is not obvious. However, from Lane 1 and Lane 4 in Figure 10, and Lane 1 and Lane 3 in Figure 11, it can also be seen that the data has a tendency to converge and develop linearly. Therefore, the possibility of a linear relationship between density and lane-changing behavior can be considered. Multiple regression is used to describe the relationship between them.

4.2. Density-Lane-Changing Behavior Regression Model. It can be seen from Figures 10, 11, and Table 7 that there is a significant linear relationship between the density and the rate of changing into lanes, and the density and the rate of changing out lanes. Multiple regression is used to describe it.

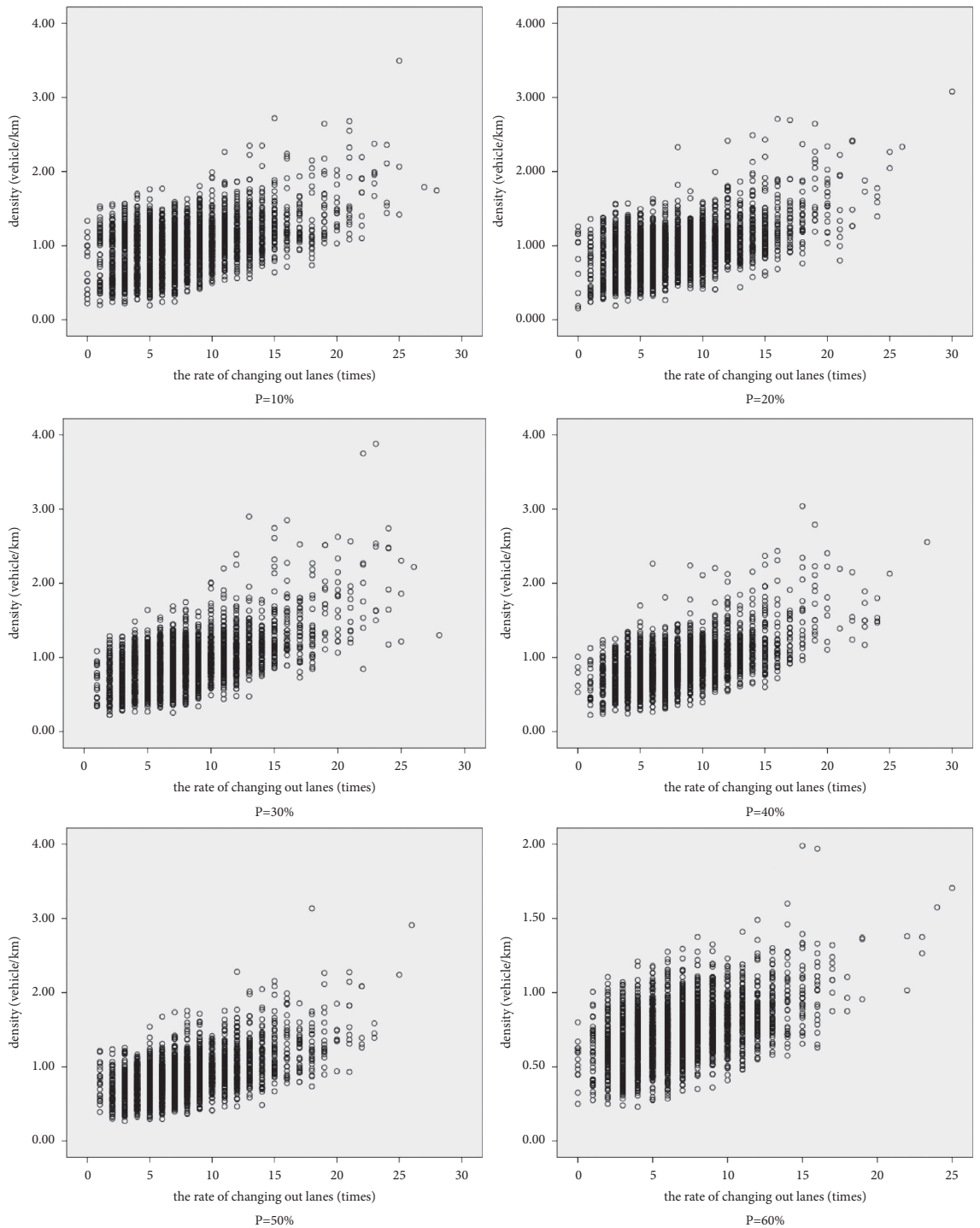
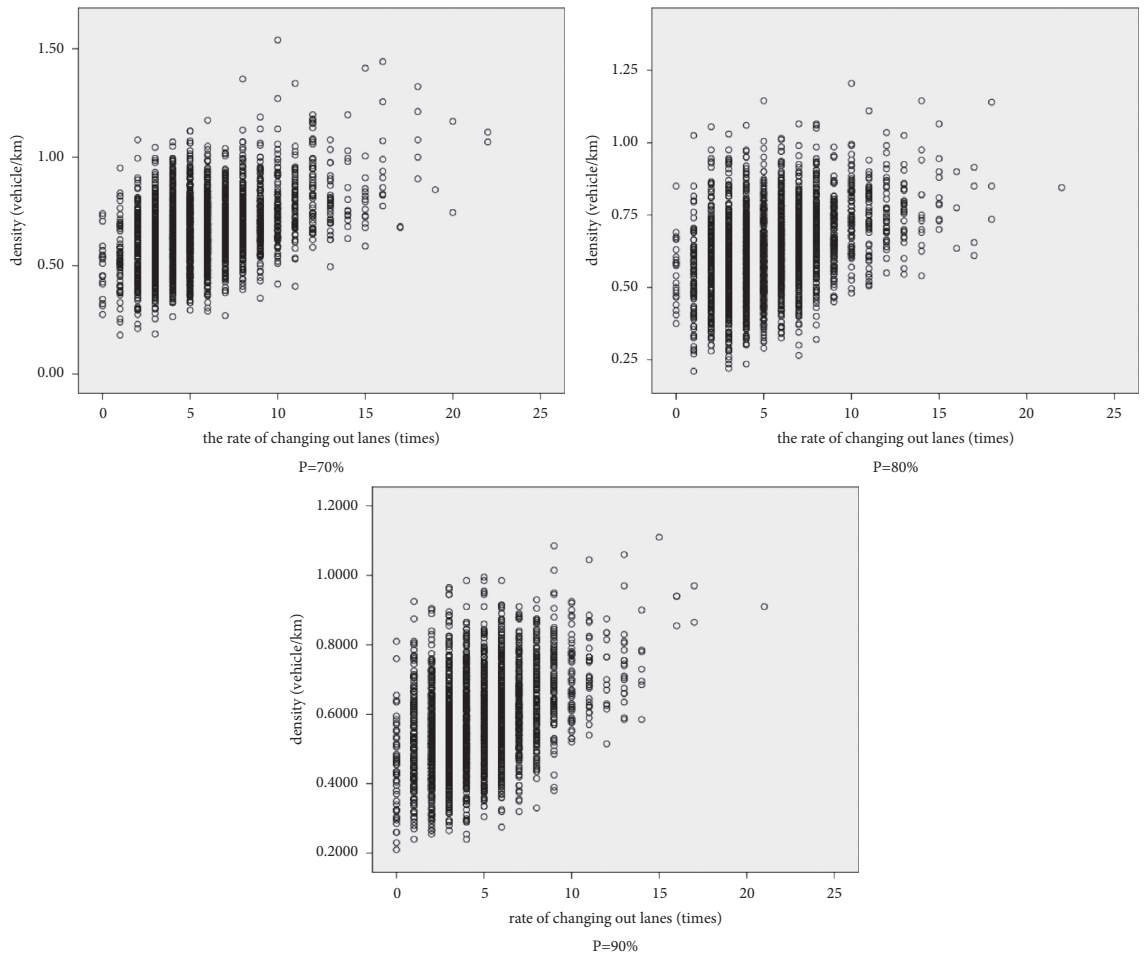


Figure 10 (a): Continued.



(a)

Figure 10: Continued.

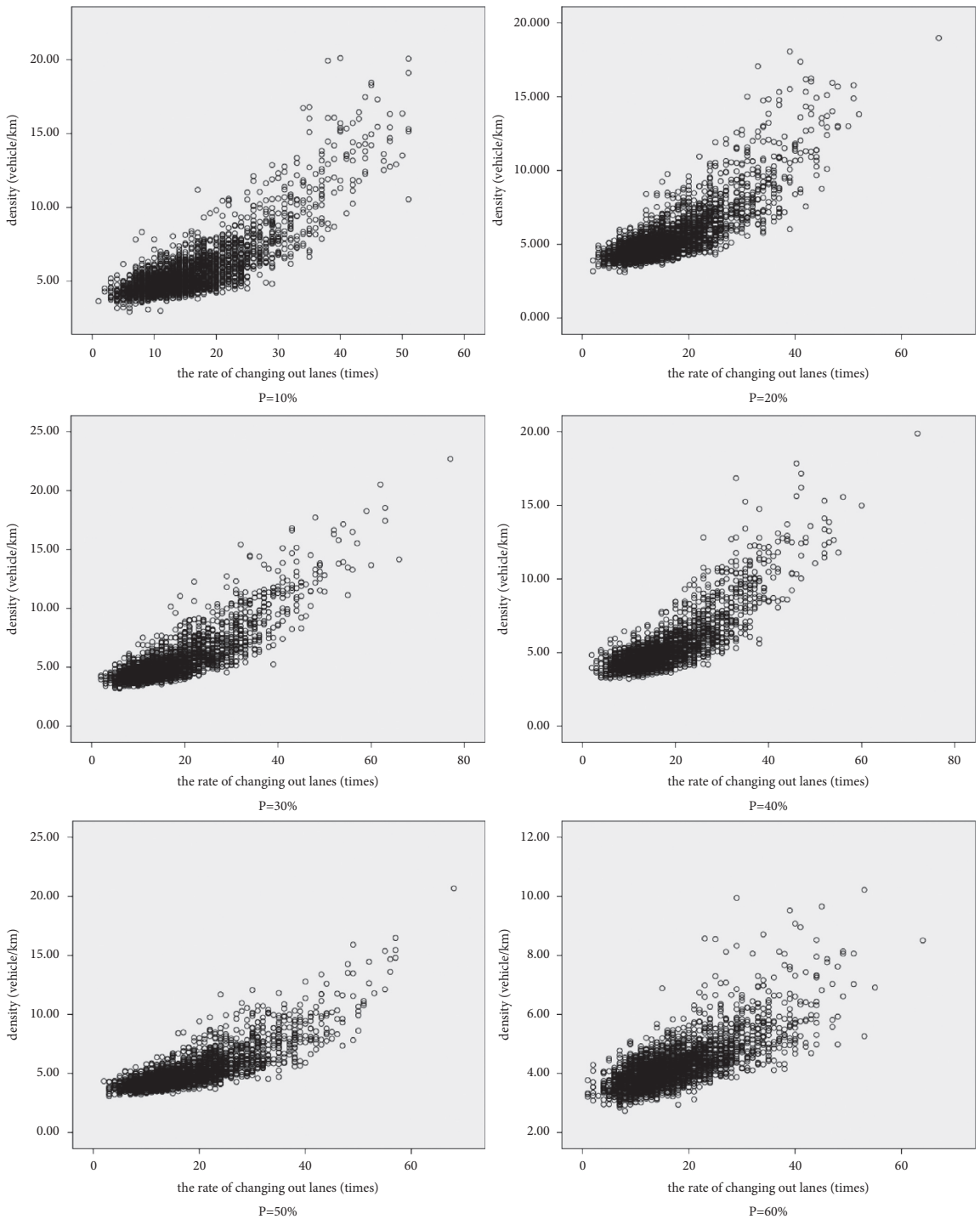
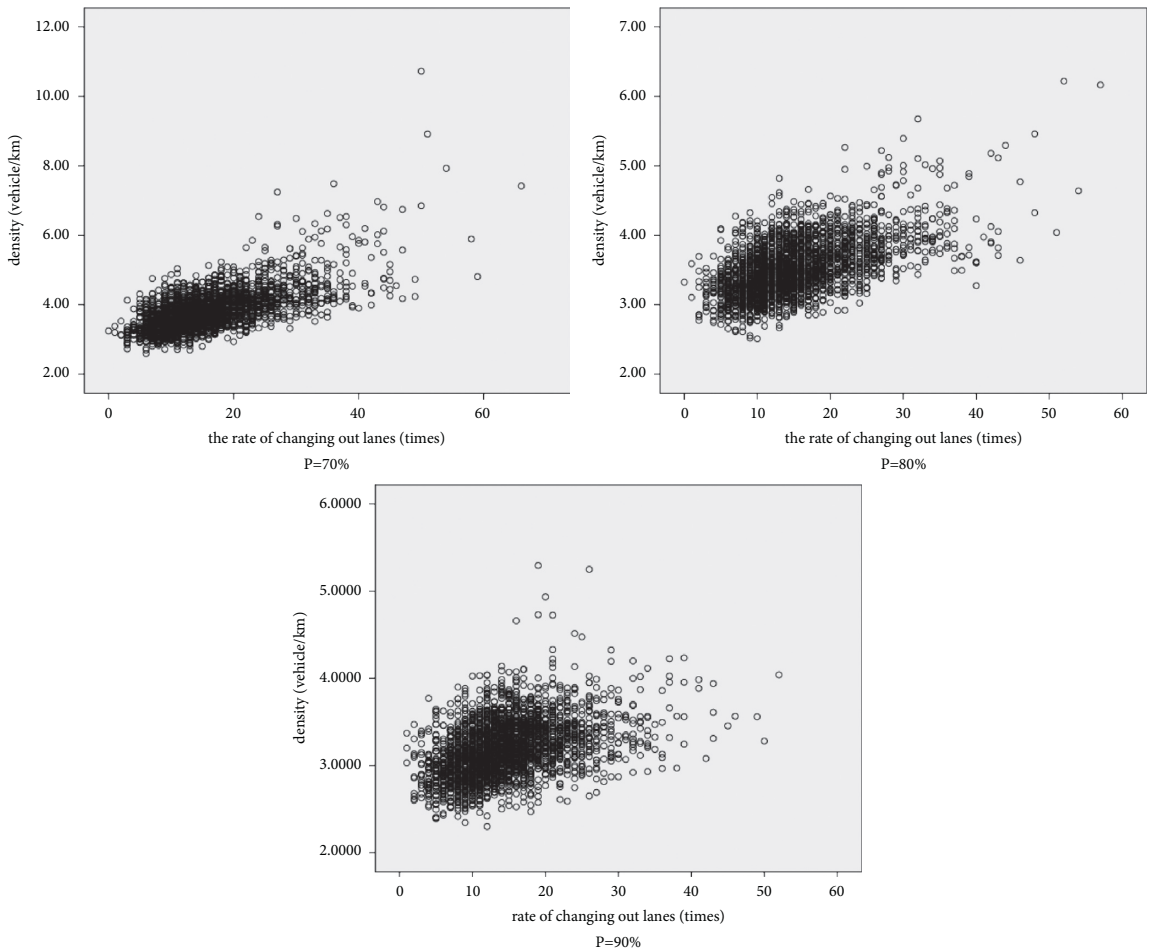


Figure 10 (b): Continued.



(b)

Figure 10: Continued.

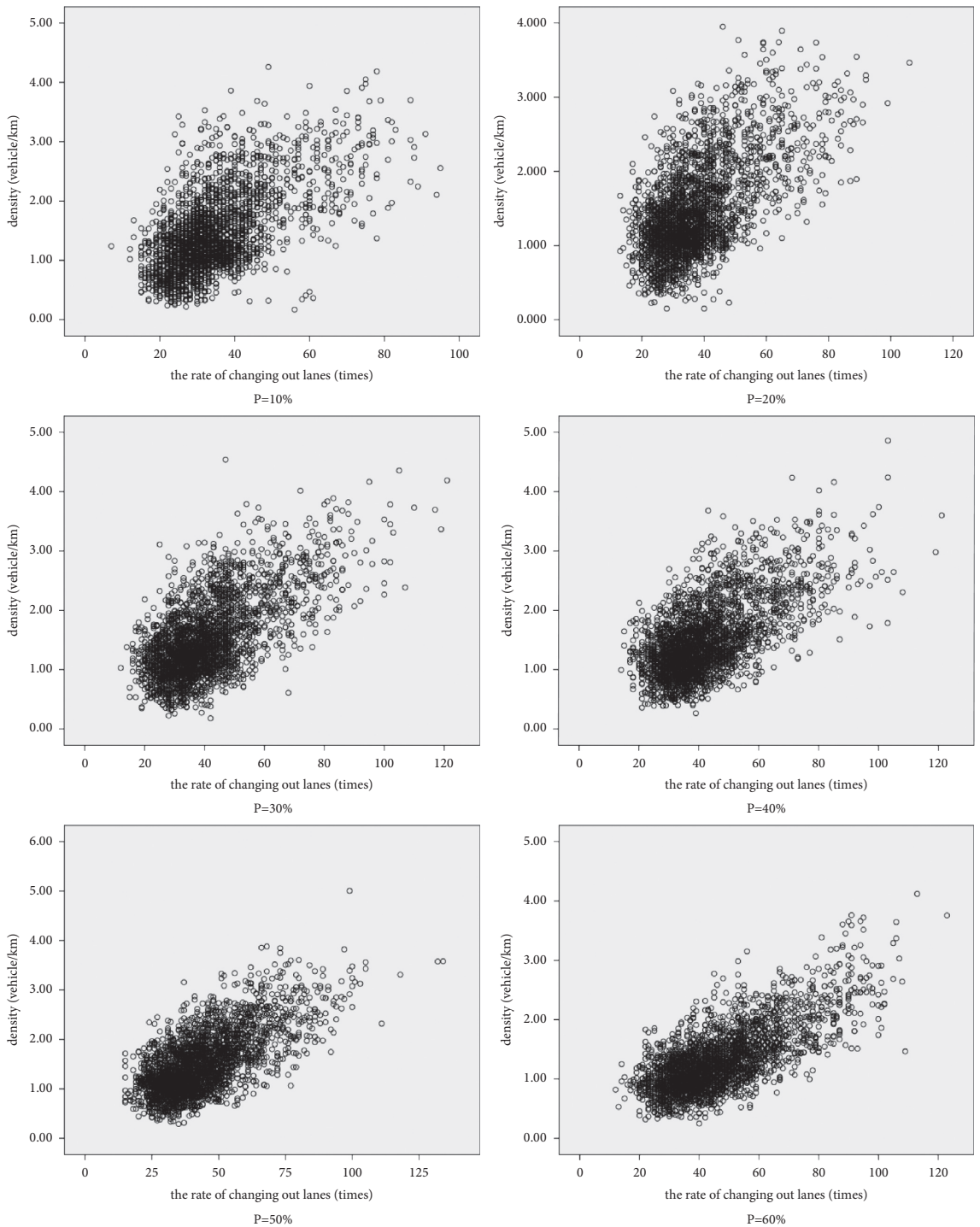
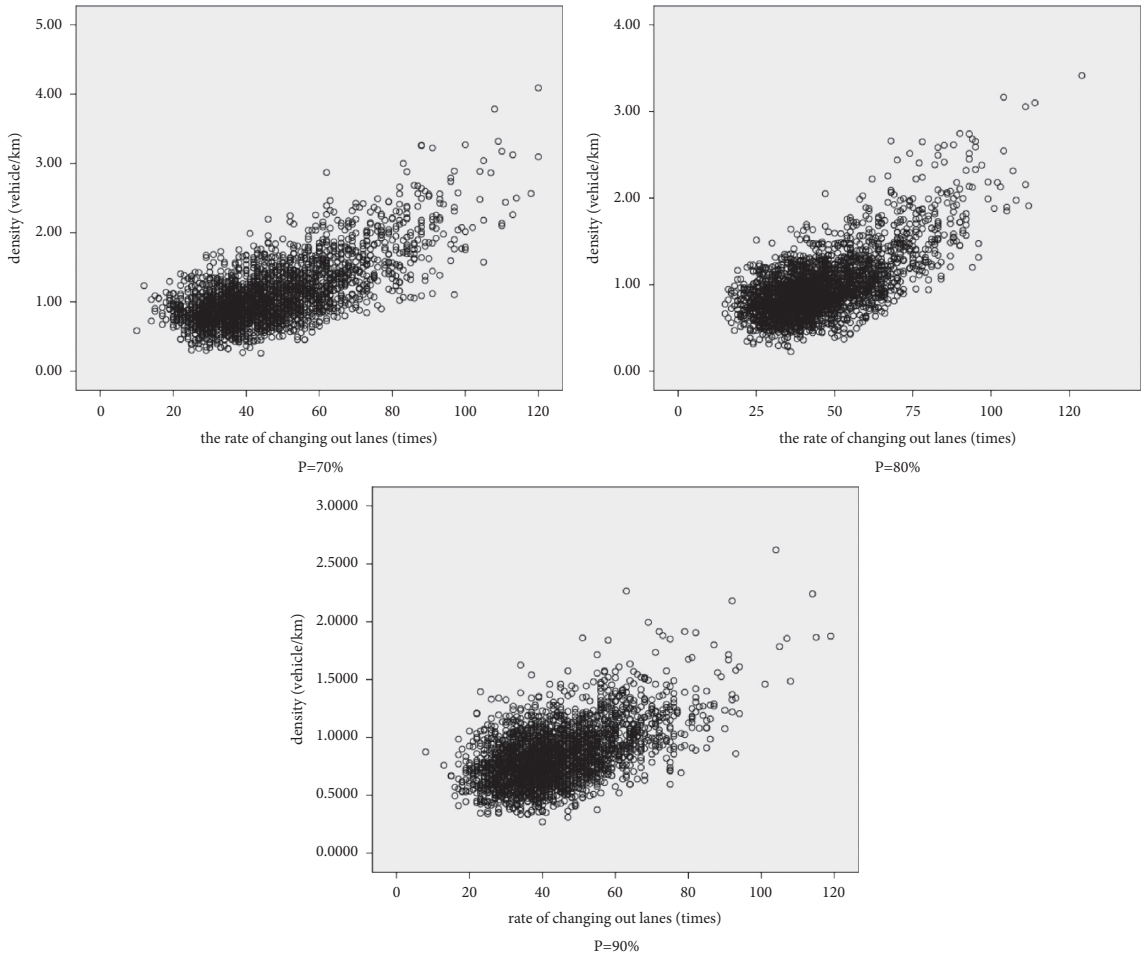


Figure 10(c): Continued.



(c)

Figure 10: Continued.

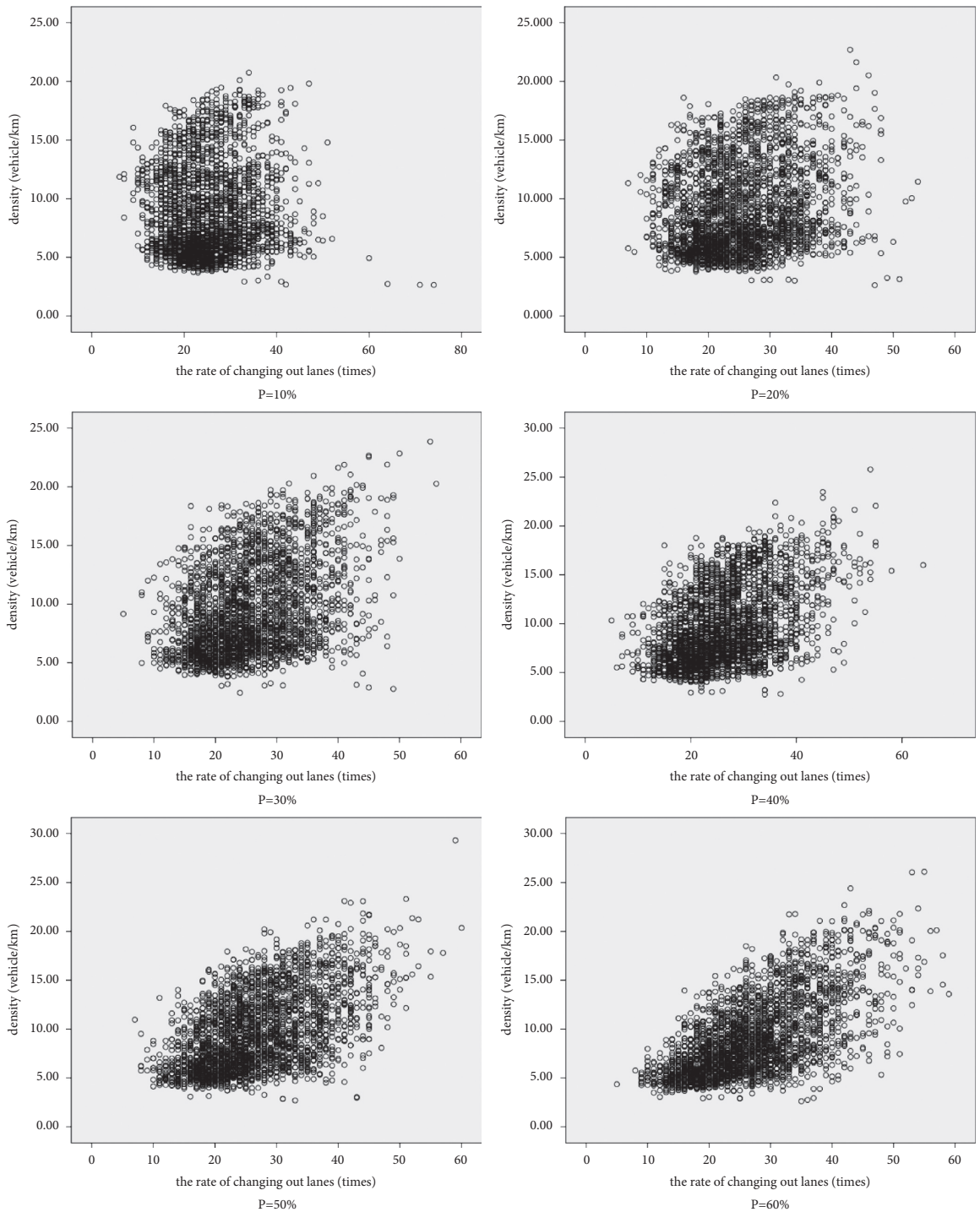


Figure 10 (d): Continued.

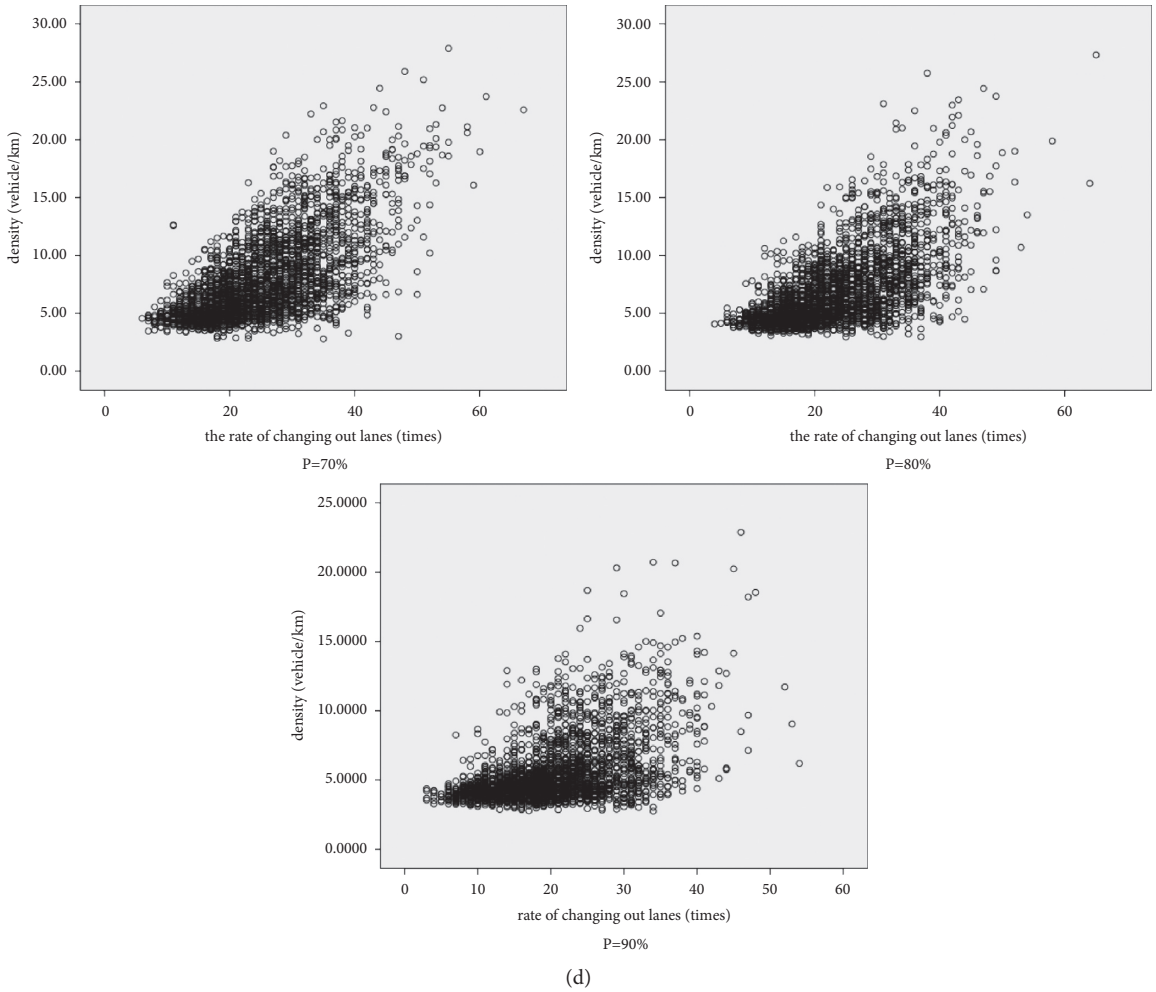


Figure 10: The relationship between density and the rate of changing out lanes by scatter plot. (a) The scatter plot of Lane 1. (b) The scatter plot of Lane 2. (c) The scatter plot of Lane 3. (d) The scatter plot of Lane 4.

Assume that the relationship between the density k and the rate of changing out lanes LCo and the rate of changing into lanes LCi is expressed by

$$k = f(LCo, LCi). \tag{5}$$

Assume that the relationship between the density k and the rate of changing out lanes LCo and the rate of changing into lanes LCi is shown in

$$k = aLCo + bLCi + c, \tag{6}$$

where a, b, and c all represent constants.

The relationship between the density and lane-changing behavior under different autonomous vehicle penetration rates was analyzed by multiple regression analysis, and the fit degree is shown in Table 8.

It can be seen from Table 8 that the degree of fit R in Lane 1 increases as the penetration rate increases to 50% and then shows a downward trend. The same trend is observed in Lane 2, but the penetration rate in Lane 2 shows a downward trend after 30%, but it is not obvious, and it does not decrease significantly until after 60%. Lane 3 does not have this trend. Lane 4 does not increase significantly with the penetration of autonomous vehicles. This trend slows down when the penetration rate reaches 30%, and there is a more obvious increase after 50%, and after 60%, it slows down. However, from the perspective of the degree of fit, with the penetration of autonomous vehicles, there is a certain linear relationship between the density and the rate of changing into lanes and the rate of changing out lanes, although the penetration rate on Lane 1 reaches 80%. The correlation is low, and Lane 4 has a low correlation when the penetration rate is 10%. It shows that when the autonomous vehicles begin to penetrate, the relationship between lane-changing behavior and density is not obvious because Lane 4 is directly connected to the entrance and exit. Because there are fewer autonomous vehicles, this relationship does not appear. When the number of autonomous vehicles continues to increase, the influence on the relationship between density and lane-changing behavior presents a more obvious linear

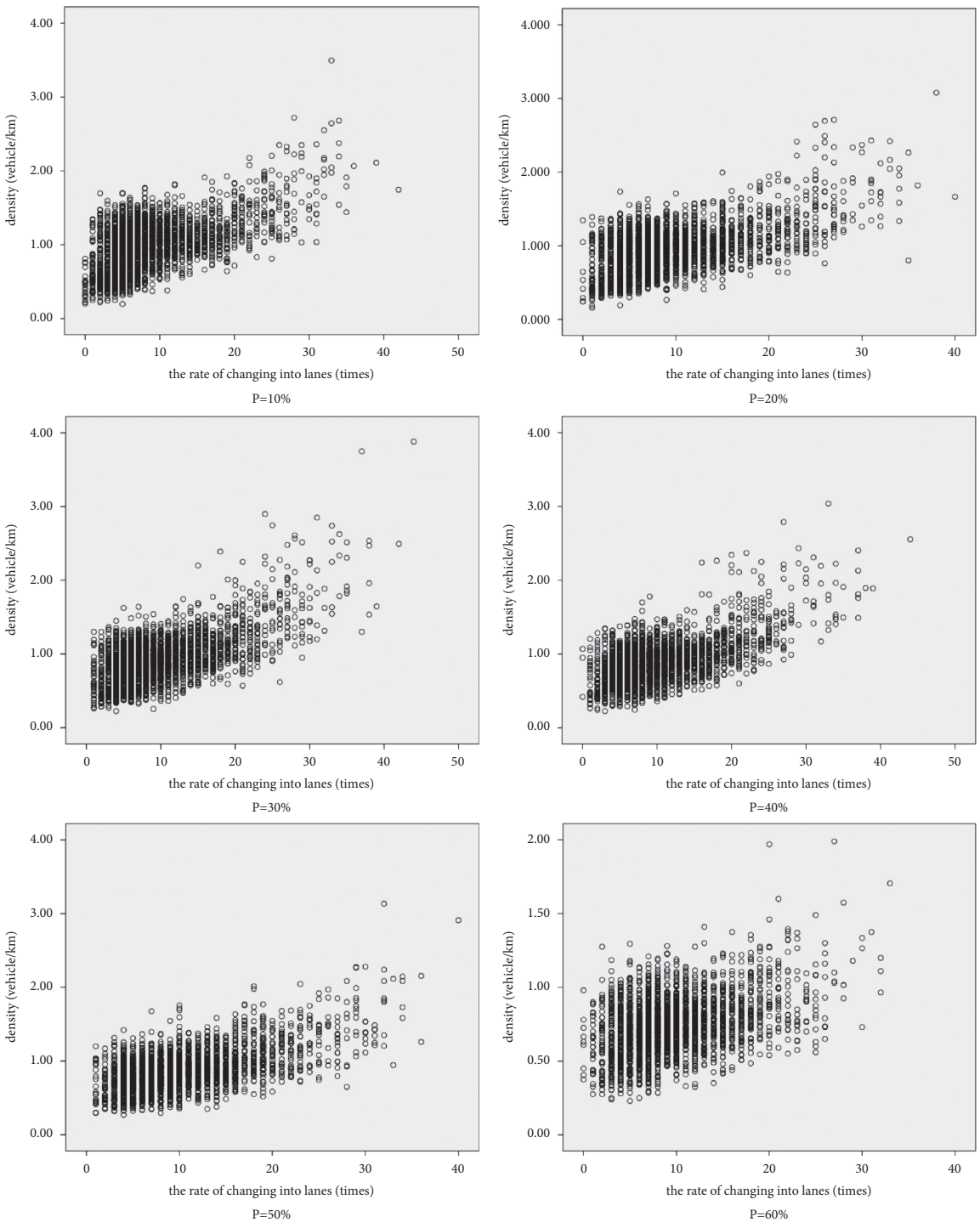
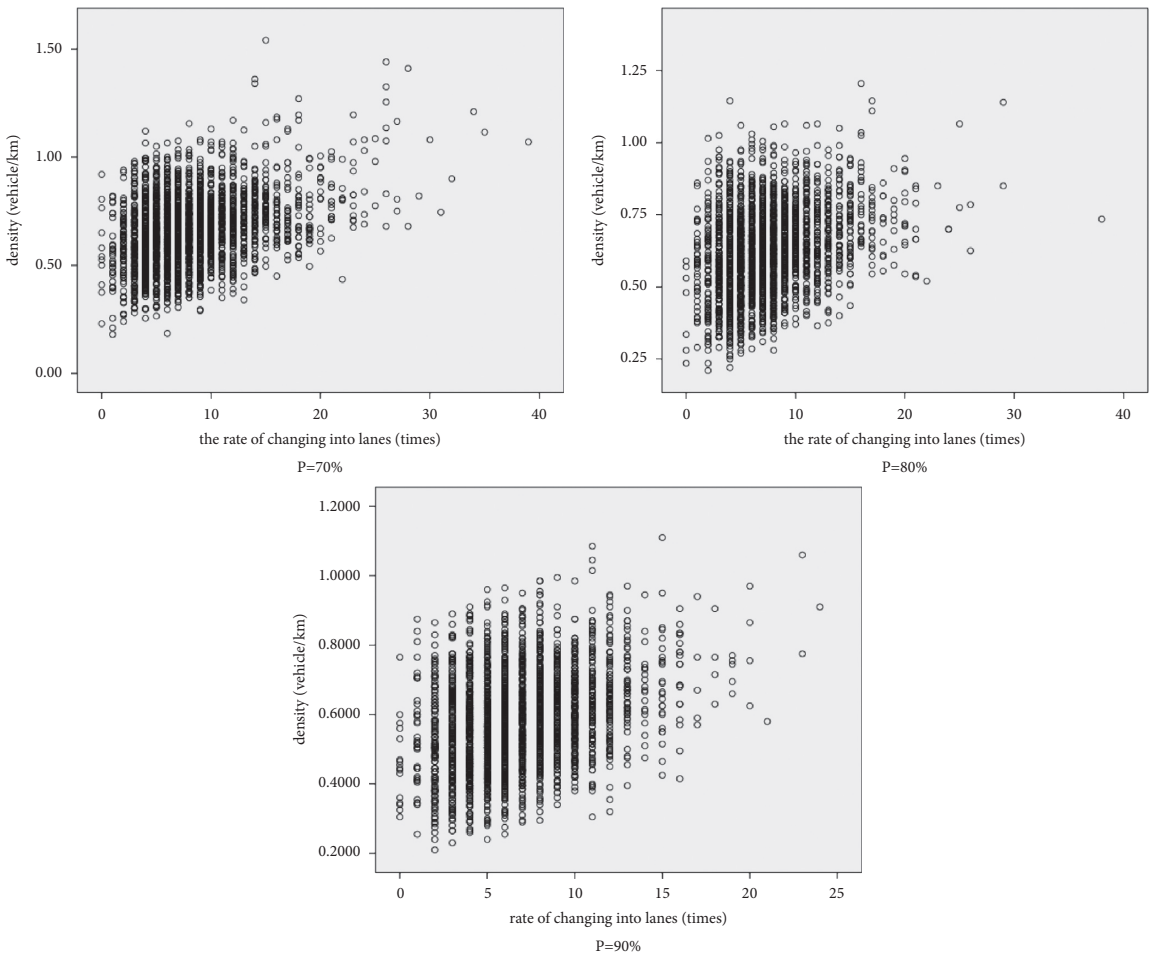


Figure 11(a): Continued.



(a)

Figure 11: Continued.

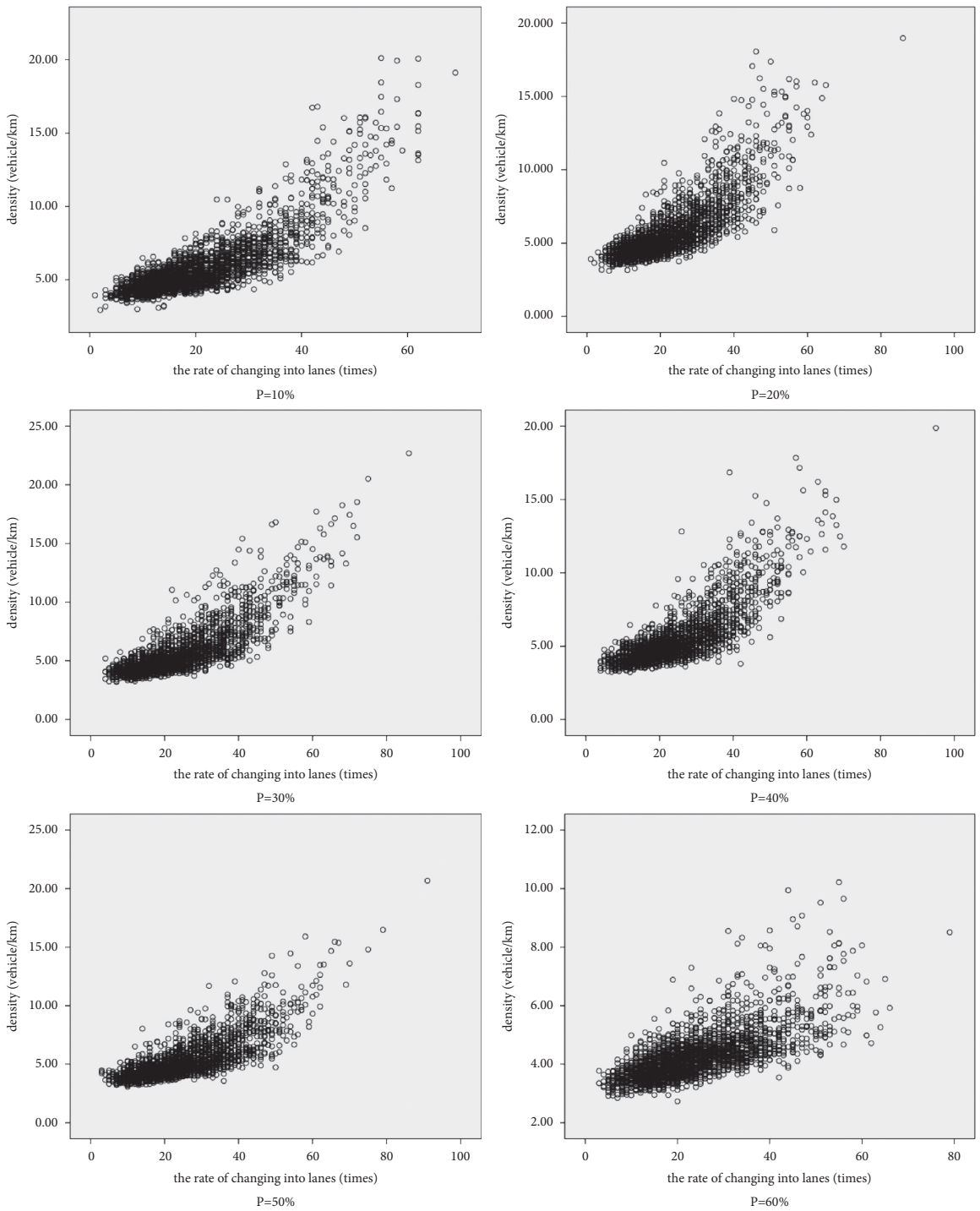
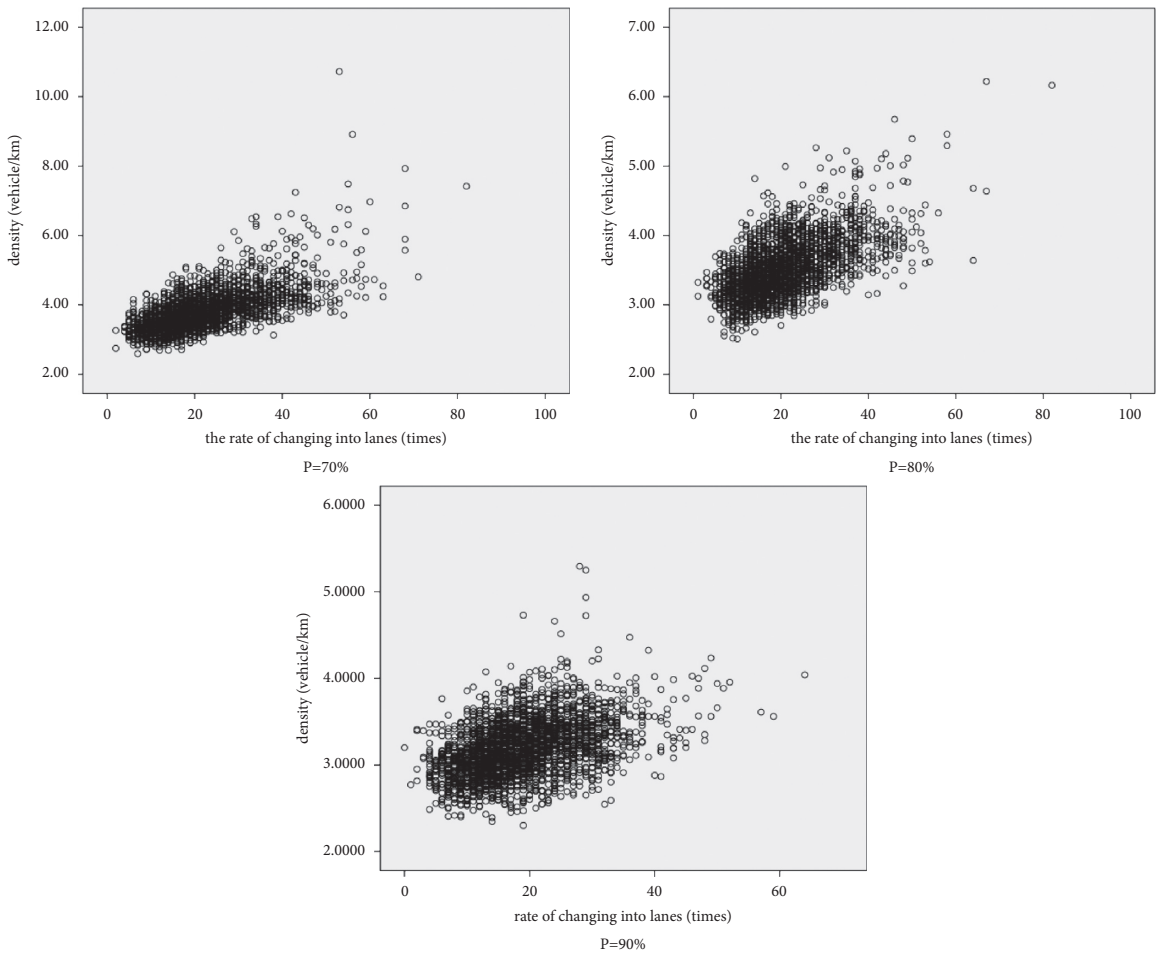


Figure 11(b): Continued.



(b)

Figure 11: Continued.

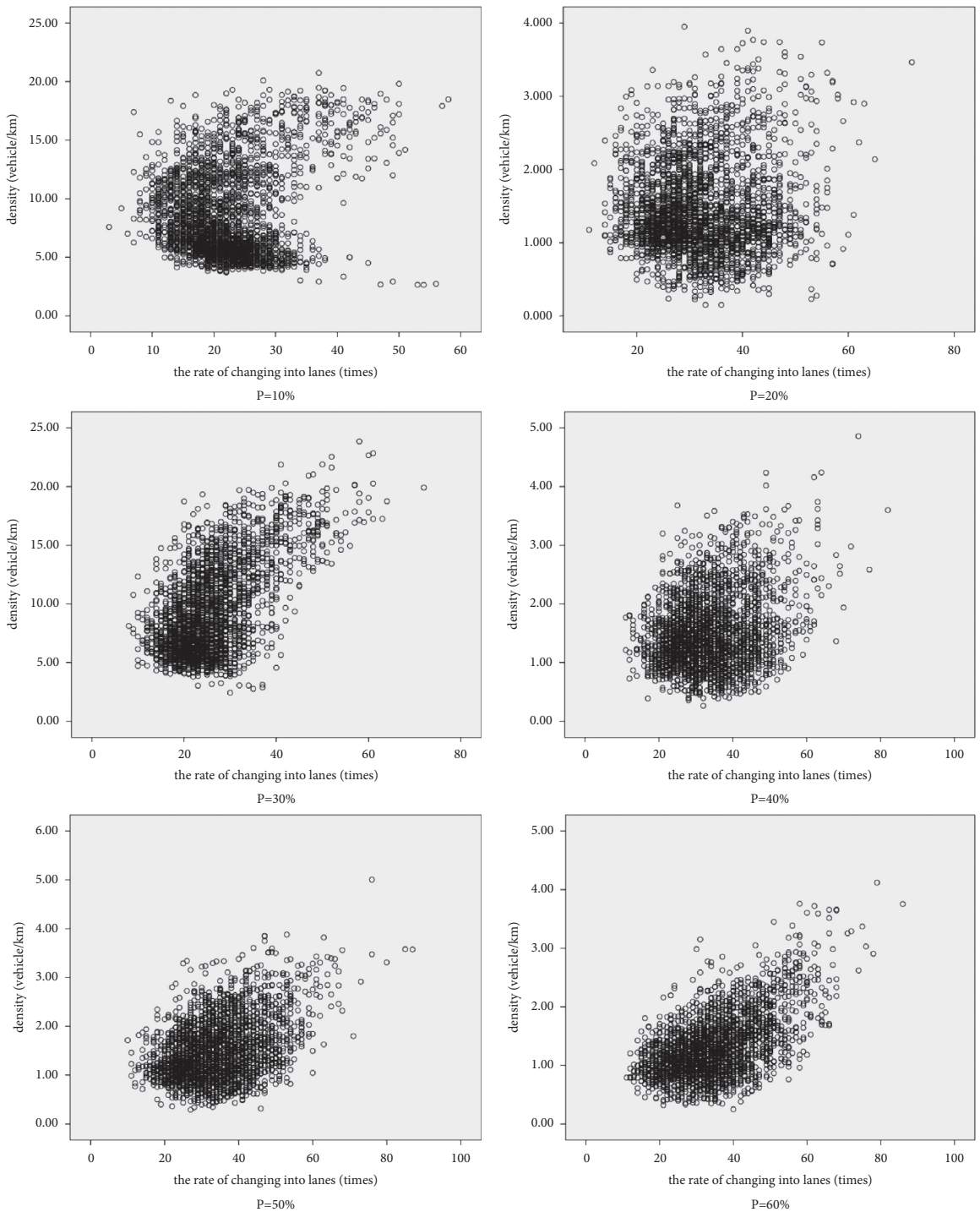
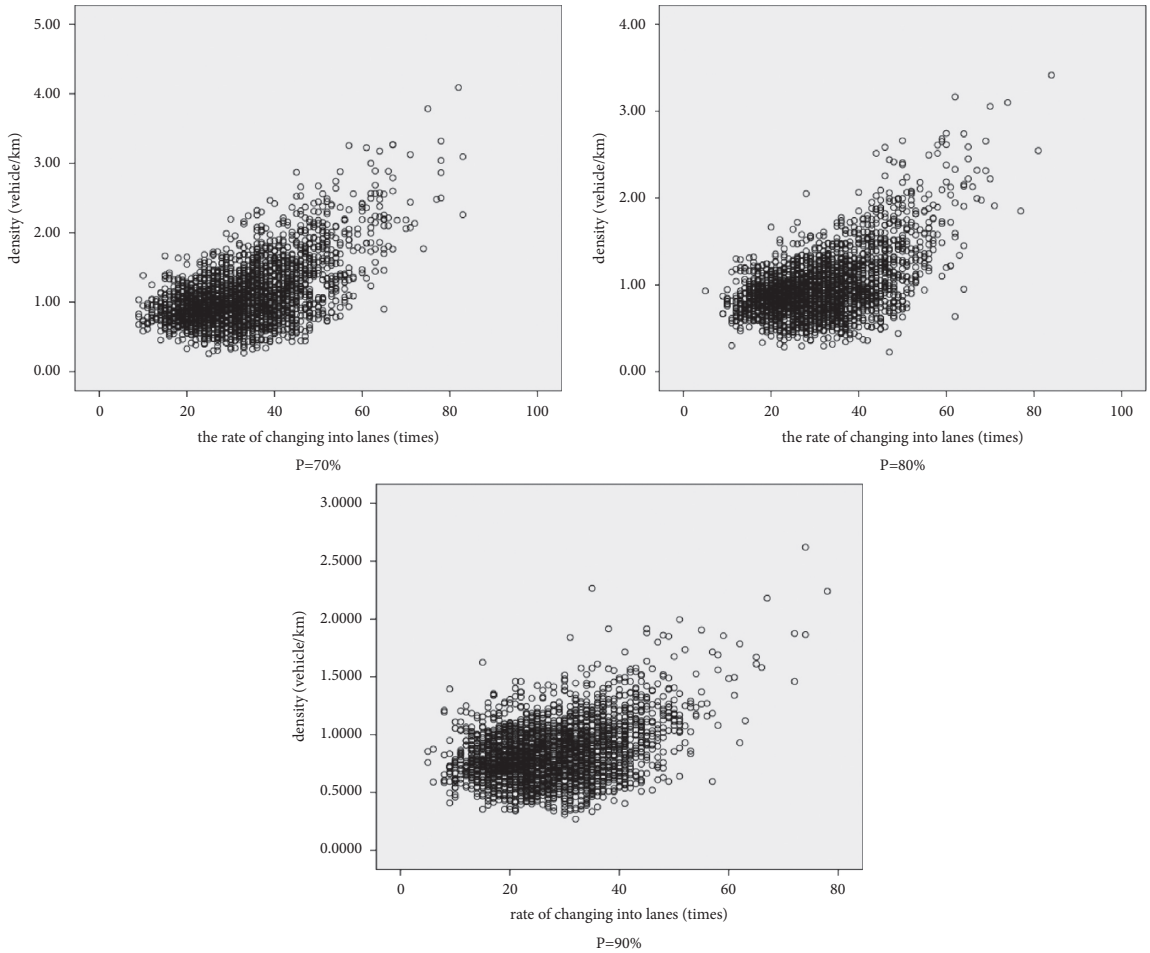


Figure 11 (c): Continued.



(c)

Figure 11: Continued.

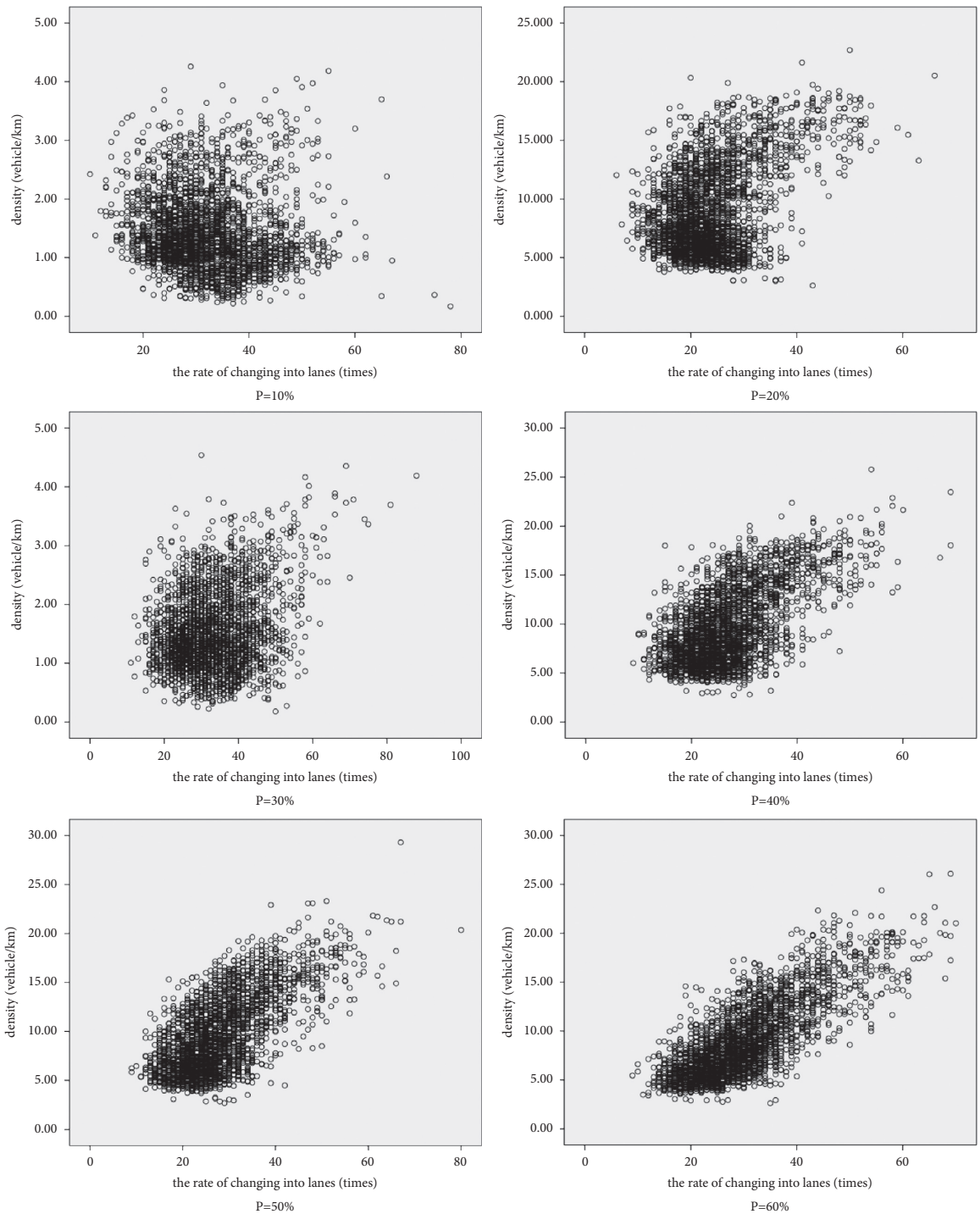
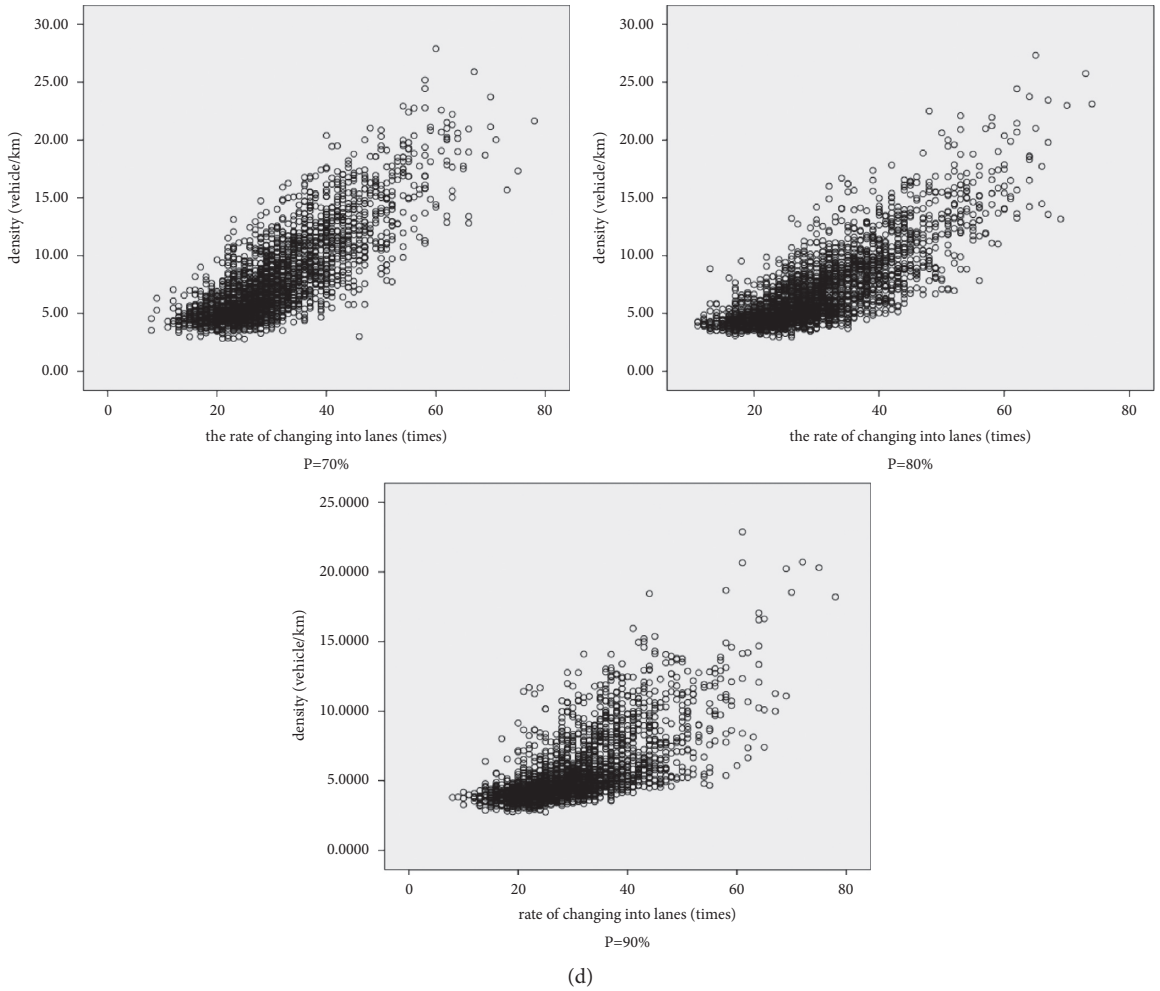


Figure 11 (d): Continued.



(d)

Figure 11: The relationship between density and the rate of changing into lanes by scatter plot. (a) The scatter plot of Lane 1. (b) The scatter plot of Lane 2. (c) The scatter plot of Lane 3. (d) The scatter plot of Lane 4.

relationship. Lane 1 has an impact on the relationship between traffic flow density and lane-changing behavior when it first with autonomous vehicles, and the linear relationship has gradually emerged. However, as the penetration rate of autonomous vehicles exceeds 50%, most autonomous vehicles form queues. Vehicles that need to change lanes also follow a certain order, so they do not necessarily drive in Lane 1, which effectively reduces lane-changing behavior, so the relationship between density and lane-changing behavior gradually becomes discrete. This can also be seen in Figures 10 and 11. But in general, the relationship between the density and the behavior of changing lanes has developed towards a linear relationship with the penetration of autonomous vehicles. In particular, the degree of fit of Lane 2 is higher, as shown in Figures 10 and 11. It can also be clearly observed. Therefore, multiple regression can be used to describe the relationship between density and lane-changing behavior. Table 9 shows the parameter value interval under different permeability.

It can also be seen from Table 9 that under the same lane, the value interval of the same parameter under different permeability is not big. It further proves that the relationship between density and lane-changing behavior can be described by multiple regression. That is, with the continuous penetration of autonomous vehicles, there is an obvious multiple linear relationship between the density and the lane change rate. This relationship model can be used to describe the density and lane change in mixed traffic.

5. Summary

In this paper, empirical, simulation, and data-driven methods are used to study the relationship between single-lane traffic density and lane-changing behavior in mixed traffic on a typical 4-lane ring-shaped urban expressway under different autonomous vehicle penetration rates. Through empirical research, the traffic density, the rate of changing into lanes, and the rate of changing out lanes data of each lane of a ring-shaped urban expressway are obtained, and the usability of the data is verified through a data-driven method. Based on empirical research, a corresponding simulation model was established. Through simulation and

Table 8: The relationship between the speed of a single lane and the lane change rate under different permeability by different fitting degrees.

		$P = 10\%$	$P = 20\%$	$P = 30\%$	$P = 40\%$	$P = 50\%$	$P = 60\%$	$P = 70\%$	$P = 80\%$
Lane 1	Multiple R	0.569	0.571	0.647	0.607	0.602	0.484	0.423	0.399
	R square	0.324	0.326	0.418	0.368	0.362	0.234	0.179	0.159
	Adjusted R square	0.323	0.325	0.418	0.368	0.362	0.233	0.179	0.159
	Standard error	0.287	0.266	0.264	0.239	0.229	0.174	0.149	0.134
Lane 2	Multiple R	0.850	0.828	0.845	0.837	0.824	0.740	0.675	0.606
	R square	0.722	0.685	0.715	0.700	0.679	0.547	0.456	0.367
	Adjusted R square	0.722	0.685	0.714	0.700	0.678	0.547	0.455	0.367
	Standard error	1.109	1.091	1.099	1.008	0.932	0.561	0.439	0.327
Lane 3	Multiple R	0.636	0.627	0.648	0.630	0.668	0.716	0.697	0.674
	R square	0.404	0.393	0.420	0.397	0.446	0.513	0.485	0.455
	Adjusted R square	0.404	0.393	0.420	0.396	0.446	0.513	0.485	0.454
	Standard error	0.519	0.499	0.512	0.490	0.466	0.378	0.323	0.269
Lane 4	Multiple R	0.219	0.468	0.629	0.657	0.730	0.821	0.834	0.822
	R square	0.048	0.219	0.395	0.432	0.533	0.674	0.695	0.675
	Adjusted R square	0.047	0.219	0.395	0.432	0.533	0.673	0.695	0.675
	Standard error	3.731	3.354	3.055	2.941	2.696	2.266	2.119	1.929

Table 9: Regression formula parameter value interval table under different autonomous vehicle penetration rates.

		Lane 1		Lane 2		Lane 3		Lane 4	
		Lower 95%	Upper 95%	Lower 95%	Upper 95%	Lower 95%	Upper 95%	Lower 95%	Upper 95%
$P = 10\%$	c	[0.6883	0.7343]	[2.1282	2.3202]	[0.6529	0.8275]	[5.403	6.6543]
	a	[0.0029	0.0107]	[0.0731	0.0979]	[0.0337	0.0371]	[−0.009	0.0292]
	b	[0.0243	0.0296]	[0.0975	0.1167]	[−0.018	−0.013]	[0.0935	0.1337]
$P = 20\%$	c	[0.6275	0.6720]	[2.1444	2.3413]	[0.5771	0.7352]	[1.9489	3.0449]
	a	[0.0103	0.0181]	[0.0890	0.1127]	[0.0309	0.0340]	[0.0499	0.0852]
	b	[0.0184	0.0235]	[0.0695	0.0884]	[−0.014	−0.010]	[0.1878	0.2221]
$P = 30\%$	c	[0.5334	0.5784]	[1.9980	2.1917]	[0.4177	0.5660]	[0.0361	0.9946]
	a	[0.0137	0.0216]	[0.1133	0.1357]	[0.0292	0.0323]	[0.0861	0.1185]
	b	[0.0217	0.0267]	[0.0465	0.0648]	[−0.009	−0.004]	[0.2302	0.2597]
$P = 40\%$	c	[0.5386	0.5815]	[1.9051	2.0941]	[0.4526	0.5943]	[−0.355	0.5477]
	a	[0.0144	0.0217]	[0.0998	0.1209]	[0.0277	0.0307]	[0.0996	0.1292]
	b	[0.0173	0.0218]	[0.0520	0.0686]	[−0.009	−0.004]	[0.2294	0.2591]
$P = 50\%$	c	[0.5284	0.5692]	[2.2359	2.4087]	[0.3185	0.4478]	[−1.202	−0.396]
	a	[0.0156	0.0230]	[0.1008	0.1200]	[0.0258	0.0288]	[0.1310	0.1583]
	b	[0.0150	0.0196]	[0.0287	0.0443]	[−0.004	0.0005]	[0.2260	0.2514]
$P = 60\%$	c	[0.5282	0.5589]	[2.9123	3.0157]	[0.1515	0.2497]	[−2.287	−1.654]
	a	[0.0190	0.0248]	[0.0465	0.0578]	[0.0193	0.0218]	[0.1177	0.1412]
	b	[0.0031	0.0068]	[0.0129	0.0216]	[0.0025	0.0064]	[0.2394	0.2603]
$P = 70\%$	c	[0.5103	0.5366]	[2.9312	3.0110]	[0.1923	0.2703]	[−2.417	−1.865]
	a	[0.0146	0.0200]	[0.0217	0.0303]	[0.0151	0.0173]	[0.1002	0.1227]
	b	[0.0031	0.0066]	[0.0165	0.0231]	[0.0020	0.0053]	[0.2301	0.2502]
$P = 80\%$	c	[0.4953	0.5194]	[2.9462	3.0080]	[0.2339	0.3000]	[−2.071	−1.569]
	a	[0.0144	0.0192]	[0.0088	0.0150]	[0.0129	0.0149]	[0.0876	0.1083]
	b	[0.0017	0.0053]	[0.0157	0.0206]	[0.0011	0.0039]	[0.2067	0.2248]

empirical data comparison, the relationship between density and the rate of changing into lanes, and the relationship between density and the rate of changing out lanes was found that under two methods the graphics are close. So, it is considered that the simulation model is close to the actual road conditions, and the density and lane-changing behavior under the penetration rate of autonomous vehicles can be studied. Based on the simulation model, simulation experiments were carried out when the penetration rate of autonomous vehicles was 10%, 20%, 30%, 40%, 50%, 60%, 70%, and 80%. The experiment obtained data on density, the rate of changing into lanes, and the rate of changing out lanes. Research using a data-driven approach found the following:

(1) The density of a single lane decreases to a certain extent with the penetration of autonomous vehicles. The downward trend of the two lanes farther from the entrance and exit is more obvious. Lane 3 and Lane 4, which are closer to the entrance and exit, first

increase and then decrease. Although it reaches 80%, the situation compared with no autonomous vehicle is still reduced, but there has been an upward trend in the front. The phenomenon indicates indicating that the penetration of autonomous vehicles has caused vehicles that can temporarily not drive in Lane 3 and Lane 4 to appear in those lanes. After the permeability reaches 50%, the density decreases.

(2) The rate of changing into lanes and the rate of changing out lanes are closer with the addition of autonomous vehicles. When the penetration rate of autonomous vehicles is 10%, all lanes show a significant decrease in lane-changing behavior. Then, with the penetration of autonomous vehicles, Lane 1 to Lane 2, the rate of changing into lanes and the rate of changing out of lanes has a relatively obvious downward trend, while Lane 3 and Lane 4 have no obvious trend. Especially, in Lane 3, as the penetration rate of autonomous vehicles increases, lane-changing behavior has increased. Lane 3 is still the main place where the behavior of changing in and out of lanes occurs.

(3) With the penetration of autonomous vehicles, the density, the rate of changing into lanes, and the rate of changing out lanes all tend to converge towards the mode.

(4) With the penetration of autonomous vehicles, the linear relationship between density and the rate of changing into lanes and the rate of changing out lanes appears. Although the data distribution of the three is still nonnormal, the degree of fit of the relationship obtained after multiple regression analysis is mostly above 0.6 under the conditions of different permeability and different lanes. It shows that the penetration of autonomous vehicles leads to a multiple linear relationship between density and lane-changing behavior.

In summary, we believe that autonomous vehicles have a certain effect on improving traffic on multilane circular urban expressways with large traffic volumes. When the penetration rate of autonomous vehicles exceeds 10%, the density, the rate of changing into lanes, and the rate of changing out lanes will decrease to a certain extent. Although there is a slight upward trend with the increase in penetration rate, it will be lower for most lanes. In other words, after the permeability exceeds 50%, the three parameters all show a downward trend. Finally, when the permeability reaches 80%, compared with the case of no autonomous vehicle, the three parameters all have a certain degree of decline. At the same time, the relationship between density and the rate of changing into lanes and the rate of changing out lanes shows a certain degree of linearity with the penetration of autonomous vehicles, and the relationship can be described by a multiple regression model. The parameter value interval given by the model also has little difference, which can prove the stability of the model.

In addition, this paper does not test the applicability of the density-lane change rate model in mixed traffic on other roads. This will be the direction of the next research.

Data Availability

The data used in this research can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This research was supported by Applied Basic Research Programs and Technology Commission Foundation of Sichuan Province of China (Grand nos. 2019JDR0093, 2020JDR0253, 2021YJ0066, 2017JY0246, and 2017JY0269), Chengdu Science and Technology Project (Grand nos. 2017RK0000306ZF and 2017RK0000357ZF), and Key Scientific Research Fund of Xihua University (Grand no. Z17131). The authors would like to express their special thanks of gratitude to their colleagues for their support and help in the process of research and writing.

References

[1] A. Talebpour and H. S. Mahmassani, "Influence of connected and autonomous vehicles on traffic flow stability and throughput," *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 143–163, 2016.

[2] Y.-Y. Qin, H. Wang, W. Wang, and W. Qian, "Stability analysis and fundamental diagram of heterogeneous traffic flow mixed with cooperative adaptive cruise control vehicles," *Acta Physica Sinica*, vol. 66, no. 9, Article ID 094502, 2017.

[3] X. Chang, H. J. Li, J. Rong, Q. Lingqiao, and Y. Yanfang, "Analysis on fundamental diagram model for mixed traffic flow with connected vehicle platoons," *Journal of Southeast University (Natural Science Edition)*, vol. 50, no. 4, pp. 782–788, 2020.

[4] L. Gen and C. Juan, "Exploring the effects of traffic density on merging behavior," *IEEE Access*, vol. 7, pp. 51608–51619, 2019.

[5] L. Huang, H. Guo, R. Zhang, H. Wang, and J. Wu, "Capturing drivers' lane changing behaviors on operational level by data driven methods," *IEEE Access*, vol. 6, pp. 57497–57506, 2018.

[6] C. Wang, Q. Sun, Z. Li, H. Zhang, and K. Ruan, "Cognitive competence improvement for autonomous vehicles: a lane change identification model for distant preceding vehicles," *IEEE Access*, vol. 7, pp. 83229–83242, 2019.

[7] Y. Dou, Y. Fang, C. Hu, R. Zheng, and F. Yan, "Gated branch neural network for mandatory lane changing suggestion at the on-ramps of highway," *IET Intelligent Transport Systems*, vol. 13, no. 1, pp. 48–54, 2019.

[8] L. Gen, F. Song, M. Jianxiao, and C. Juan, "Modeling merging acceleration and deceleration behavior based on gradient boosting decision tree," *JournalofTransportation Engineering ,Part A:Systems*, vol. 146, no. 7, Article ID 05020005, 2020.

[9] Ministry of Housing and Urban-Rural, *Development of the People's Republic of China.Design Discipline of Urban Expressway[S]*, China Architecture & Building Press, Beijing, China, 2009.

[10] W.-X. Zhu and H. M. Zhang, "Analysis of mixed traffic flow with human-driving and autonomous cars based on car-following model," *Physica A: Statistical Mechanics and Its Applications*, vol. 496, pp. 274–285, 2018.

[11] L. Gen, M. Jianxiao, and Y. Zhen, "Characteristics of heavy vehicle discretionary lane change based on trajectory data," *Transportation Research Record*, Article ID 03611981211051337, 2021.

[12] Z. Yao, H. Jiang, and Y. Cheng, "Integrated schedule and trajectory optimization for connected automated vehicles in a conflict zone," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2020.

[13] X. Wang, W. Li, C. Yin, S. Zeng, and P. Liu, "A multiple-parameter approach for short-term traffic flow prediction," *Modern Physics Letters B*, vol. 35, no. 14, Article ID 2150245, 2021.

[14] W. Li, S. Chen, X. Wang, and Z. Huang, "A hybrid approach for short-term traffic flow forecasting based on similarity identification," *Modern Physics Letters B*, vol. 35, no. 13, Article ID 2150212, 2021.

[15] C. L. Davis, "Effect of adaptive cruise control systems on traffic flow," *Physical Review E - Statistical, Nonlinear and Soft Matter Physics*, vol. 69, no. 6, Article ID 066110, 2004.

[16] M. W. Levin and S. D. Boyles, "A multiclass cell transmission model for shared human and autonomous vehicle roads," *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 103–116, 2016.

[17] D. Xie, X. Zhao, and Z. He, "Heterogeneous traffic mixing regular and connected vehicles: modeling and stabilization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2060–2071, 2019.

[18] S. Wang, J. Zhao, C. Shao, and C. Yin, "Truck traffic flow prediction based on LSTM and GRU methods with sampled GPS data," *IEEE Access*, vol. 8, pp. 208158–208169, 2020.

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

[20] J. A. Laval and C. F. Daganzo, "A hybrid model of traffic flow: impacts of roadway geometry on capacity," in *Proceedings of the TRB 2003 Annual Meeting CD-ROM 2003*, North Avenue, Atlanta, GA, USA, 2003.

[21] J. A. Laval and C. F. Daganzo, "Lane-changing in traffic streams," *Transportation Research Part B*, vol. 40, pp. 251–264, 2006.

[22] W.-L. Jin, "Kinematic wave-theory of lane-changing traffic flow," *Transportation Research Part B*, vol. 44, no. 8-9, pp. 1001–1021, 2009.

[23] Transportation Research Board, *Highway Capacity Manual*, HCM2010), USA, 2010.

[24] K. Liu, J. W. Gong, K. Arda, H. Y. Chen, and O. Umit, "Dynamic modeling and control of high-speed automated vehicles for lane change maneuver," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 3, pp. 329–339, 2018.

[25] Y. Xing, C. Lv, H. J. Wang et al., "Driver lane change intention inference for intelligent vehicles: framework, survey, and challenges," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4377–4390, 2019.

[26] X. P. Gu, Y. P. Han, and J. F. Yu, "A novel lane-changing decision model for autonomous vehicles based on deep autoencoder network and XGBoost," *IEEE Access*, vol. 8, pp. 9846–9863, 2020.

[27] L. Gen, P. Yiyong, Y. Zhen, and M. Jianxiao, "Modelling vehicle merging position selection behaviors based on a finite mixture of linear regression models," *IEEE Access*, vol. 7, 2019.

[28] N. Ding and X. H. Meng, W. G. Xia, D. Wu, L. Xu, and B. C. Chen, Multi-vehicle coordinated lane change strategy in the roundabout under Internet of vehicles based on game theory and cognitive computing," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 1–8, 2015.

[29] B. Li, Y. M. Zhang, Y. H. Feng, Y. Zhang, Y. M. Ge, and Z. J. Shao, "Balancing computation speed and quality: a decentralized motion planning method for cooperative lane changes of connected and automated vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 3, pp. 340–350, 2018.

[30] T. T. Li, J. P. Wu, C. Y. Chan et al., "A cooperative lane change model for connected and automated vehicles," *IEEE Access*, vol. 8, pp. 1–7, 2019.

[31] Y. Zheng, B. Ran, X. Qu, J. Zhang, and Y. Lin, "Cooperative lane changing strategies to improve traffic operation and safety nearby freeway off-ramps in a connected and automated vehicles environment," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2020.

[32] X. Liu, J. Liang, and B. Xu, "A deep learning method for lane changing situation assessment and decision making," *IEEE Access*, vol. 7, pp. 133749–133759, 2019.

[33] Q. S. Zhang and Y. P. Zhang, *Analysis of Road Traffic Capacity*, China Communication Press, Beijing, China, 2002.

[34] J. Li, *Traffic Engineer*, China Communication Press, Beijing, China, 2007.