Journal of Advanced Transportation

# Traffic Efficiency and Safety in Mixed Traffic Flow Environments 2020

Lead Guest Editor: Chunjiao Dong
Guest Editors: Baoshan Huang, Xiaoming Chen, Chengxiang Zhuge, and Meng Meng

# Traffic Efficiency and Safety in Mixed Traffic Flow Environments 2020

Journal of Advanced Transportation

# Traffic Efficiency and Safety in Mixed Traffic Flow Environments 2020

Lead Guest Editor: Chunjiao Dong
Guest Editors: Baoshan Huang, Xiaoming Chen, Chengxiang Zhuge, and Meng Meng

# Contents

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 6624452, 14 pages
https://doi.org/10.1155/2021/6624452

*Research Article*

# A Multifeatures Spatial-Temporal-Based Neural Network Model for Truck Flow Prediction

**Shengyou Wang**[1], **Chunfu Shao**[1], **Yajiao Zhai**[2], **Song Xue**[1] and **Yan Zheng**[1]

[1]*Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, School of Traffic and Transportation, Beijing Jiaotong University, No. 3 Shangyuan Cun, Haidian District, Beijing 100044, China*
[2]*Road Construction Project Management Center of Beijing Road Administration Bureau, No. 36, Dongwei Street, Xicheng District, Beijing 100031, China*

Correspondence should be addressed to Chunfu Shao; cfshao@bjtu.edu.cn

Received 18 October 2020; Revised 22 December 2020; Accepted 23 April 2021; Published 4 May 2021

Academic Editor: Zhixiang Fang



The majority of studies on road traffic flow prediction have focused on the flow of passenger cars or the flow of traffic as a whole, which ignore the significant impact of trucks with different sizes and operational characteristics on traffic flow efficiency. Therefore, in this paper, we focus on truck traffic flow and propose a Multifeatures Spatial-Temporal-Based Neural Network model (M-BiCNNGRU) to improve its prediction. The proposed model not only comprises conventional temporal characteristics and spatial relationships but also includes a range of multifeatures. These multifeatures include policy limit, optimal time delay, road resistance, and traffic congestion state. The impacts of upstream and downstream road sections are considered on the spatial relationship by using a Convolutional Neural Network (CNN). A Bidirectional Gated Recurrent Unit (Bi-GRU) is employed to account for the temporal characteristics. To evaluate the proposed model, traffic flow data were collected from a major expressway in Beijing and the results were compared with those derived from existing models. The results show that the prediction accuracy of the BiCNNGRU model, with spatial-temporal characteristics, and the M-BiGRU model, with multifeatures and temporal, is, respectively, 4.13% and 2.15% greater than that of the Bi-GRU model, with temporal characteristics. The prediction accuracy of the proposed M-BiCNNGRU model is 92.86%, which is 7.12% greater than that of the Bi-GRU model and 13.83% greater than that of the Support Vector Regression (SVR) model. In general, therefore, the proposed M-BiCNNGRU model, which combines multifeatures, temporal characteristics, and spatial relationships, can significantly improve accuracy in predicting truck traffic flow.

## 1. Introduction

It is generally acknowledged that the efficient operation of modern supply chains relies heavily on the use of trucks. This is particularly the case in China which has seen a growth of over 40% in truck road traffic during the last decade [1]. This increased transportation dependency has essentially reinforced the importance of trucks but given their poor dynamic performance and large sizes, the net effect of increased truck traffic is the problem of potentially more congestion and reduced traffic capacity overall [2]. Consequently, this problem is now attracting the attention of academia, especially concerning the development of an accurate prediction model for truck traffic flow, aiming to find solutions that can reduce such traffic congestion that is caused primarily by the increasing numbers of trucks on roads.

Despite the increasing congestion problems caused by trucks, currently, most research related to congestion is associated with cars [3] and mainly relies on historical traffic data for traffic prediction and seeking rules from time characteristics [4–7]. Li et al. [8] proposed a model based on ensemble empirical mode decomposition and a random vector functional link network to predict travel time in mixed traffic flow. There are relatively few studies on the influence of truck flow, even though it is a significant component of overall road traffic. Several researches have

predicted traffic flow from the perspective of temporal characteristics and spatial relationships and obtained better prediction results, indicating that increasing the spatial relationship will benefit prediction accuracy [9–11]. Compared with regular passenger cars, however, trucks have a range of operational features or constraints. For example, there are policies that restrict truck operation on specific road sections and time periods. Such policies, therefore, suggest that it is necessary to analyze such features that affect truck traffic flow and explore whether these features can effectively improve the efficiency of truck traffic flow prediction.

In terms of traffic flow prediction, the models currently employed can be classified as either parametric models or nonparametric models [12]. The parametric models mainly include ARIMA, subset ARIMA, and Kalman filtering [13–18]. However, a major drawback of the parametric models is their inability to reflect nonlinear traffic flow data with strong randomness, leading to traffic flow prediction inaccuracies. Addressing this problem in order to improve prediction accuracy, nonparametric models have been proposed which are more adept in the capture of nonlinear characteristics of traffic flow. The nonparametric models include K-Nearest Neighbor (KNN), Support Vector Regression (SVR), Back Propagation Neural Network (BPN), and Fuzzy Neural Networks [16, 19–22].

In recent years, not least with improvements in computer performance, the field of deep learning has developed rapidly. Of particular relevance, deep learning methods can capture the characteristics of nonlinear traffic data and achieve excellent results. However, each deep learning model has its fields and characteristics. For example, the Convolutional Neural Network (CNN) has an excellent performance in processing spatial data, and the Recurrent Neural Network (RNN) has advantages in capturing temporal features [23]. It also has been found that both historical time data and the road network spatial relationship have an impact on traffic flow, and such a finding led to the joint deep learning model combining spatial-temporal advantages being applied to traffic flow prediction [24]. Asadi and Regan [25] proposed a CNN-LSTM deep learning framework for spatiotemporal forecasting problems. LSTM (Long Short Term Memory) is a variant of RNN [26], which solves the problem of gradient explosion and disappearance of RNN by adding memory modules, and has achieved positive performance in time series prediction [27, 28]. Based on temporal-spatial features, Li et al. [29] combined an advanced multiobjective particle swarm optimization algorithm and deep belief networks to forecast traffic flow for the next day. Yang et al. [28] proposed a hybrid deep learning model to predict traffic flow, in which the LSTM and a CNN component were applied to learn temporal and spatial features, separately. A CNN-LSTM model was proposed by Cheng et al. [30] to capture traffic flow evolution in a transportation network. A CNN layer was applied to both downstream and upstream road sections for spatial characteristic mining, followed by the incorporation of an LSTM layer. It is worth noting that the prediction performance of the LSTM method depends on a large number of parameters and training data and that its computing ability is limited by computer bandwidth and memory. The Gated Recurrent Unit (GRU) was proposed to solve the deficiencies of LSTM, and its parameters are relatively small in number and it is easier to obtain a converged solution [31]. Fu et al. [3] applied the GRU method to predict car traffic flow for the first time, which performed better than the ARIMA model. Considering the advantages of GRU and CNN in capturing temporal and spatial features, they can be combined to capture the temporal and spatial characteristics of truck traffic flow to improve the prediction accuracy.

Compared with passenger cars, trucks are required to adhere to special operational policies, mainly in terms of driving time and space. For example, the transportation department stipulates that trucks are prohibited from passing on certain urban roads within 24 hours or trucks are prohibited from passing through certain areas from 7:00–8:30 and 16:30–18:30 during passenger car rush hours. Also, traffic condition, the impedance of the road, and the mixed flow rate of trucks all can affect traffic efficiency but there are relatively few researches into the influence of these factors on truck traffic flow prediction; most rely instead on the use of single-source data from historical traffic flow records.

In general, to improve the prediction accuracy of truck traffic flow, additional factors that affect such flow will be analyzed from the perspective of policy, traffic status, and road impedance. Within this approach, multiple factors are defined as variables to form multiple features, and the input variables are, therefore, changed from one-dimensional traffic flow data to multidimensional feature data. Finally, the spatial relationship between upstream and downstream road sections is added to form a multifeatures spatial-temporal-based joint deep learning model to predict truck traffic flow and evaluate the prediction results. The overall aim and main contributions of this study are summarized as follows:

(1) The primary objective of the study is to improve the prediction accuracy of truck traffic flow. Although passenger car traffic flow forecasting has been widely studied, the characteristics of trucks and passenger cars are different, it is not feasible to apply the forecasting method developed for passenger cars to trucks. For this reason, it is necessary to incorporate truck characteristics into the method for predicting truck traffic flow.

(2) A Multifeatures Spatial-Temporal-Based Neural Network model (M-BiCNNGRU) is proposed in this study to improve the prediction accuracy of truck traffic flow. In the proposed model, three types of factors are considered, namely, "spatial relationship," "temporal series," and "road operation." Moreover, the impacts of upstream and downstream road sections are considered on the spatial relationship using a CNN. A Bidirectional Gated Recurrent Unit (Bi-GRU) is employed to account for the temporal characteristics. The road operation factor includes a range of multifeatures. These multifeatures include "policy limit," "optimal time

delay," "road resistance," and "traffic congestion state." These multifeatures are also through the CNN model to explore the relationship between the truck flow and multifeatures.

(3) The prediction performance of the M-BiCNNGRU model is explored. Several baseline models are also trained and predicted, including (i) the BiCNNGRU model, which considers spatial relationships and temporal characteristics; (ii) the M-BiGRU model, which considers multifeatures and temporal characteristics; (iii) the Bi-GRU model, which considers only temporal characteristics; (iv) the SVR model, which again, only considers temporal characteristics; (v) finally, the ARIMA model, which likewise only considers temporal characteristics."

This paper is organized as follows. Section 2 introduces the methodology of the multifeatures spatial-temporal-based truck traffic flow prediction method. In Section 3, the properties of road sections are introduced, and the spatial relationship and congestion coefficient associated with road sections are analyzed. The results of the proposed model are analyzed in Section 4. Section 5 discusses the prediction performance during peak and low peak periods. Section 6 summarizes the conclusion.

## 2. Methodologies

In previous models, traffic flow predictions associated with passenger cars were only inferred from historical traffic flow data. However, other than historical traffic flow, there are multiple other features that affect truck traffic flow, such as current road conditions, road resistance, traffic congestion index, and traffic restriction policies. To consider these additional features, in this paper, we propose a Multifeatures Spatial-Temporal-Based Neural Network prediction method for truck traffic flow, which is termed M-BiCNNGRU. The CNN model is used for multifeatures and spatial feature learning and Bi-GRU model is employed for temporal characteristics learning. The proposed M-BiCNNGRU model structure is shown in Figure 1.

*2.1. Multifeatures.* Five features have been analyzed and incorporated into the model to improve the prediction performance of truck traffic flow. The interpretation and definition of the considered features are as follows:

(1) Truck traffic flow: Truck traffic flow has a self-correlation, which means that historical truck traffic flow will have an impact on future truck traffic. The truck traffic flow $q$ refers to the number of trucks passing through a road section in a certain time interval and is also defined as a continuous variable. The definition of truck is that mainly used for carrying goods in terms of design and technical characteristics, including special vehicles whose main purpose is to carry goods (GA 802-2014). In this paper, the authors regard all truck types as the research object of truck traffic flow prediction. To unify the different truck types, the authors converted the traffic flow of each truck type to the Passenger Car Equivalent (PCE). The classification of truck type and the PCE conversion factor is shown in Table 1.

(2) Policy limit: For different road sections and different periods, truck operation will be restricted by policy, which will cause changes in truck traffic flow. The policy limit variable is defined as a discrete variable. If trucks are allowed to operate during a period, it is defined as 1; if not, it is defined as 0.

(3) Optimal time delay: In this paper, the road section where traffic flow is to be predicted is called the target section. For situations when truck flow "upstream" of the target section changes, a period of time is incurred before the traffic information can be disseminated to the target section. Therefore, it is necessary to find a time delay difference $r$, which represents the time taken for the upstream and downstream sections to transmit truck flow characteristics. The optimal time delay $r_b$ refers to the delay which provides the best match between two road segments. The time delay diagram of upstream and downstream sections is shown in Figure 2. $n$ is defined as the number of truck flow time series.

In this paper, the Pearson correlation coefficient (PCC) is used to calculate the correlation of truck flow series under a different time delay $r$, where the time delay under the maximum PCC is the optimal time delay $r_b$. Moreover, $r_b$ is a discrete variable whose value is equal to the optimal time delay value.

(4) Road resistance: In previous models, the BPR function was often used to calculate the road resistance of passenger cars. However, for truck traffic, the traditional BPR function is inadequate and led to the development of an improved truck road resistance function (1).

$$t_a = \frac{l_0^t}{v_0^t}\left[1 + \alpha\left(\frac{\sum q_a^t \cdot e^t}{c_a \cdot m_a^t}\right)^{\beta}\right], \tag{1}$$

where $l_a^t$ is the length of the road section $a$ at time $t$. $v_0^t$ represents the speed of the truck when the traffic flow of the road section is zero at time $t$. $q_a^t$ represents the truck traffic flow on road section $a$. $e^t$ represents the coefficient for the conversion of trucks into equivalent passenger cars; its respective values are 1.5, 2, and 3 for small, medium, and heavy trucks. $m_a^t$ represents the proportion of trucks in the overall traffic flow. $c_a$ represents the traffic capacity of road section $a$. For the improved truck road resistance function, the length of the road section, the current truck traffic flow, the truck traffic flow rate in the mixed traffic flow, and the truck equivalent are taken into account to express the road operation more accurately.

(5) Congestion coefficient: At present, the standards for measuring congestion indicators are not uniform. Due to the complex and diverse traffic operating

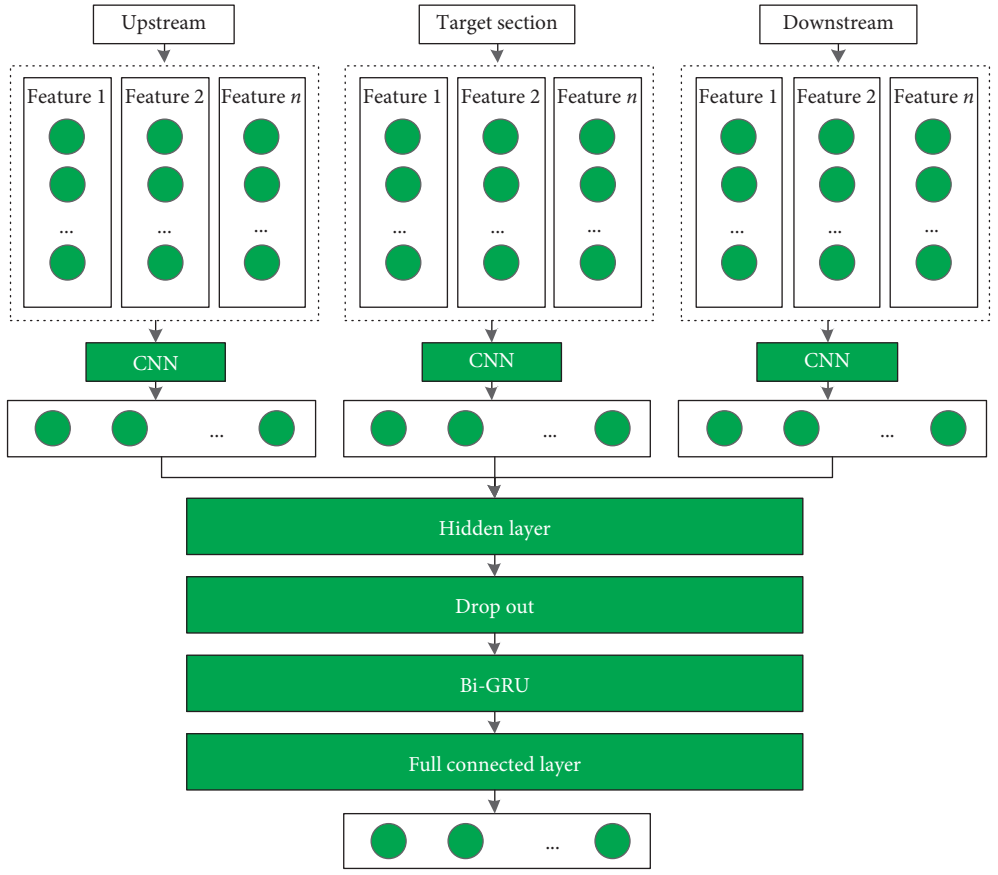



Figure 1: The proposed M-BiCNNGRU model structure.

Table 1: The classification of truck type and the Passenger Car Equivalent (PCE) conversion factor.

| Truck type | Conversion factor | Load and passenger capacities |
|---|---|---|
| Small truck | 1 | Load capacity ≤ 2.0 tons |
| Medium truck | 1.5 | 2.0 tons < load capacity ≤ 7.0 tons |
| Large truck | 2 | 7.0 tons < load capacity ≤ 14.0 tons |
| Extra large truck | 3 | Load capacity > 14.0 ton |

environment of urban roads, the classification of traffic operating status is often not accurate enough and has a certain degree of ambiguity [32, 33]. For this reason, the fuzzy evaluation method is used to judge the traffic state. This paper applies the truck traffic flow fuzzy comprehensive evaluation method to calculate the congestion coefficient on the road section. The function of the congestion coefficient (CC) is

$$f_a^q = \frac{z_a \cdot q_a^t}{c_a \cdot m_a}. \tag{2}$$

When $(q_a^t/m_a)$ reaches the design flow limit, the road section is in a serious congestion state, where the CC is equal to 10; however, when the $(q_a^t/m_a)$ is equal to 0, the road section is in a clear state where the CC is close to 0.

For the fuzzy comprehensive evaluation method, the threshold range according to the congestion degree needs to be defined first. For the CC, the larger its value means the more congested the road section is. The CC of a road network is ranked in a descending order and divided into five levels according to a certain proportion. In this paper, the definition of congestion is obtained with reference to Beijing's "Evaluation Index System for Urban Road Traffic Operation" (EISU) published on April 28, 2011. EISU proposed the recommended conversion relationship between the Road Network Operation Level (RNOL) and Traffic Performance Index (TPI). When the TPI is between the values [0, 2], [2, 4], [4, 6], [6, 8], and [8, 10], the RNOL is defined as "unblocked," "basically smooth," "light congestion," "moderate congestion," and "severe congestion," respectively. Moreover, the respective variables associated with these RNOL definitions are 1, 2, 3, 4, and 5. Further details of the above variable definitions can be found in Table 2. To unify the types of discrete variables and continuous variables, the author performs one-hot encoding of discrete variables.

*2.2. Construct Feature Vector.* Studies have shown that the traffic conditions of the upstream and downstream sections have an impact on the target section flow. However, the

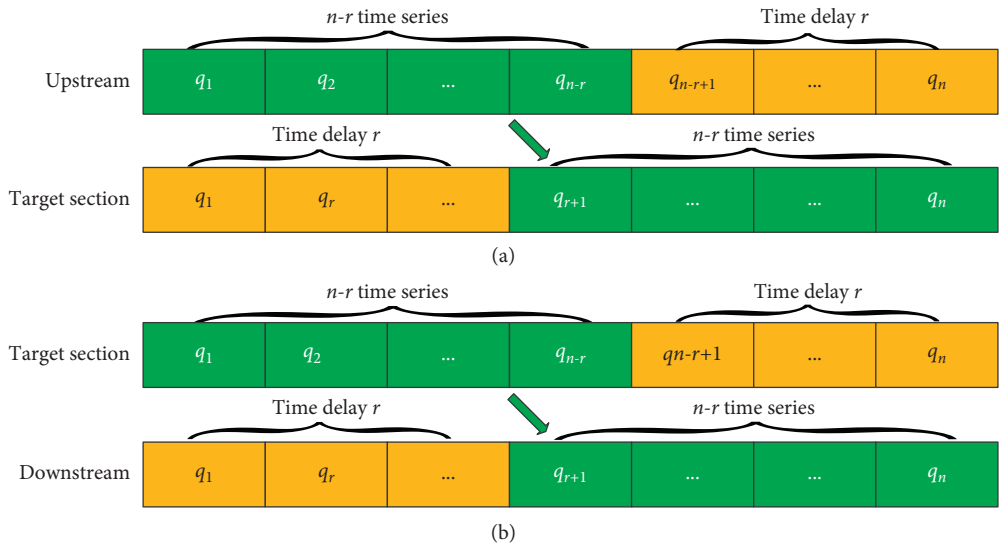



Figure 2: Schematic diagram of time delay.

Table 2: The details of the variable definition.

| Symbol | Feature | Type | Description |
|---|---|---|---|
| $r_p$ | Policy limit | Discrete variable | 1: open to traffic; 0: closed to traffic |
| $r_b$ | Optimal time delay | Discrete variable | 1: optimal time delay is 1; 2: optimal time delay is 2; 3: optimal time delay is 3 |
| $r_s$ | Road resistance | Continue variable | Value of truck road resistance function |
| $q_a$ | Truck traffic flow | Continue variable | Value of truck traffic flow |
| $r_c$ | Congestion coefficient | Discrete variable | 1: unblocked; 2: basically smooth; 3: light congestion; 4: moderate congestion; 5: the severe congestion |

target road section may have multiple upstream road sections as well as multiple downstream road sections, as shown in Figure 3.

The feature variables of the upstream section, the downstream section, and the target section are defined as $X_u$, $X_d$, and $X_o$, respectively, where $X_u = [r_p^u, r_b^u, r_s^u, q_a^u, r_c^u]$, $X_d = [r_p^d, r_b^d, r_s^d, q_a^d, r_c^d]$, and $X_o = [r_p^o, r_s^o, q_a^o, r_c^o]$. The input feature vector $X$ of the proposed model is, therefore, constructed, as shown in equation (3).

$$X = [X_u, X_t, X_d]. \tag{3}$$

If the target road section has only one upstream and one downstream road section, as shown in Figure 2(a), it is relatively easy to compose the input feature vector. However, a target road section often has multiple upstream road sections and multiple downstream road sections, as shown in Figure 3(b). Moreover, if the feature variables of multiple road sections are all simply applied to the input feature vector, there will be inconsistencies in the dimensions of the input feature vectors. However, if one of the multiple upstream sections is simply selected as being representative of the upstream section and because the upstream representative section may not accurately display the truck traffic flow status of all upstream sections, errors may also be introduced. It should be noted that discussions on the multiple scenarios linking the target road section with the upstream and downstream road sections are often ignored in previous papers.

Therefore, when there are multiple upstream road sections, this paper proposes aggregating the feature variables of multiple upstream sections into one upstream section. The aggregated rule is as follows:

(1) For the truck traffic flow variable, it is formed by summing up the multiple upstream sections.

(2) For the policy limit, it is set as the policy for the road section with the highest levels of truck traffic flow.

(3) For the optimal time delay, CC, and road resistance, they are recalculated according to the sum of the truck traffic flow in the multiple upstream sections.

The rule for the downstream road section is the same as the above upstream section. Through the above processing, multiple scenarios linking the target section with the upstream and downstream section are aggregated into a unified dimension of the input feature vector.

*2.3. Spatial Relationships.* To mine the spatial relationship between the upstream and downstream sections and the target section, the CNN was chosen. This is because CNN

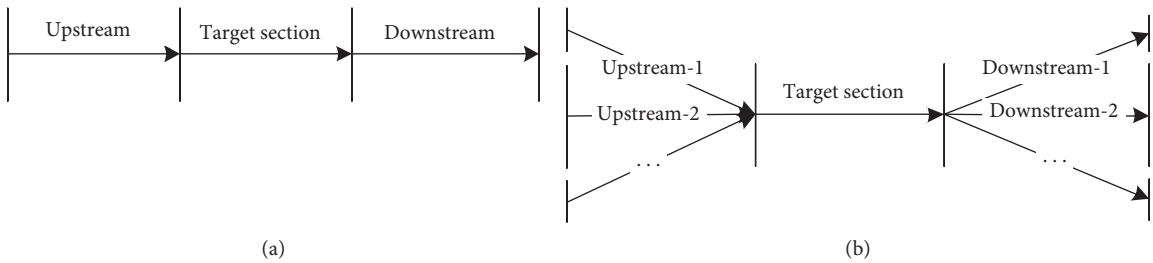



Figure 3: Multiple scenarios linking the target section with the upstream and downstream. (a) One upstream and one downstream section. (b) Multiple upstream and multiple downstream sections.

has advantages in mining spatial local related information and has achieved significant results in other fields such as natural language processing. The conventional 1D CNN is applied to learn the spatial relationship with the locality structure of key features. The CNN contains a convolutional layer, activation layer, and a pooling layer. By performing 1D convolution on the input feature vector $X$, the m-th feature map of the 1D CNN is calculated:

$$h_i = \text{pool}\left(k_i\left(w_i * X_i + b_i\right)\right), \tag{4}$$

where $h_i$ is the weight vector of the $i$-th feature, $b_i$ is the bias, $k_i$ is the nonlinear activation, $*$ represents convolution, and pool denotes the pool function. To avoid overfitting, the depth of the 1D CNN is set to 3. The ReLU is applied for nonlinear activation.

*2.4. Temporal Characteristics.* Truck traffic flow has a long-term temporal dependency. In this paper, a Bi-GRU is introduced to learn the temporal characteristics associated with truck traffic flow. This is because the Bi-GRU has advantages in capturing time characteristics. Recently, an LSTM method has been applied in numerous problems as it, in general, performs well in terms of prediction accuracy, but compared with LSTM, the Bi-GRU contains fewer parameters and has achieved a better performance in the field of natural languages. However, for truck traffic flow, few scholars have studied its potential benefits. Specifically, in the Bi-GRU model, the reset gate and update gate were applied to control the update status of the time series. The reset gate $R_i$ and the update gate $U_i$ are calculated using

$$R_i = \sigma\left(W_r S_i + U_r H_{i-1} + b_r\right), \tag{5}$$

$$U_i = \sigma\left(W_u S_i + U_z H_{i-1} + b_u\right), \tag{6}$$

where $S_i$ is the input from the Bi-BRU, which is also the output of the 1D CNN, $\sigma$ is the activation, $W_r$ and $W_u$ are the weight vectors, $b_r$ and $b_z$ are biases, and $H_{i-1}$ is the hidden layer output value. Also, the structure of the Bi-GRU is composed of the forward and reverse sequences. The forward sequence is from start to end, and the reverse sequence is from end to start.

In summary, the Multifeatures Spatial-Temporal-Based Neural Network prediction method, namely, the M-BiCNNGRU model, is a new joint neural network model. Compared with the CNN model, the proposed M-BiCNNGRU model adds the structure of mining temporal characteristics from the Bi-GRU neural network. As far as the authors are aware, this study is the first to combine the Bi-GRU and CNN neural networks for truck traffic flow tasks. The proposed model has three advantages; firstly, the impact of the truck traffic flow in the upstream and downstream sections on the target road section can be explored using the spatial relationship. Therefore, the road conditions of the upstream and downstream sections can be well represented in the proposed model. Secondly, the Bi-GRU neural network is well capable of mining the temporal characteristics. Through the full connectivity layer, the output result of the spatial relationship is placed in the Bi-GRU model which can then continue to explore the time relationship. Thirdly, the proposed model is more capable of expressing the characteristics of truck flow from the traffic perspective.

## 3. Data

*3.1. Properties of the Road Section.* The data for the period June 1st to June 5th, 2019, were provided by the traffic survey in Beijing, China, and contains truck traffic flow as well as passenger car traffic flow information. The data were collected by manual counting methods. The data were provided by Beijing Transport Institute, which invited professional survey companies to conduct data collection work, and a total of 30 investigators participated in this work. The time interval of truck traffic flow is one hour, and the size of the data in the CSV format is 3.48 GB. As an important part of Beijing's northwest freight corridor, the Sixth Loop Expressway is selected as the research object and is divided into 1,293 road sections and numbered accordingly. To consider different scenarios, six road sections (S1–S6) were randomly selected as target road sections, as shown in Figure 4.

From Figure 4, it can be seen that S3 and S6 have three upstream sections and one downstream section, whereas S1, S2, S4, and S5 each have one upstream section and one downstream section. The average daily truck traffic flow of the selected six target section is shown in Figure 5. It is known that the hourly distribution of daily truck traffic flow is different from the distribution of morning and evening peak flow of passenger cars. The peak hour of the truck traffic flow is during 11:00–23:00 and the low peak period is during 0:00–11:00.

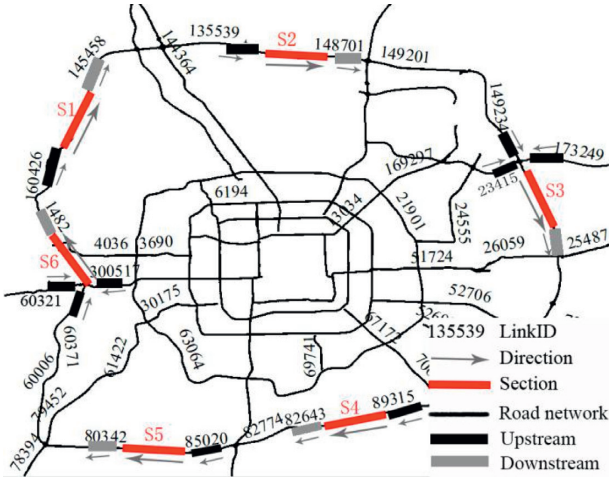



Figure 4: Truck traffic flow locations.

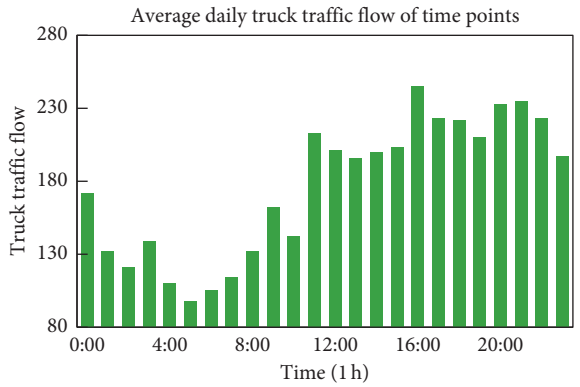



Figure 5: The average daily truck traffic flow.

The properties of the six selected target sections are shown in Table 3. In this context, mean refers to the mean truck traffic flow in 1 hour. Std refers to the mean standard deviation. The mixed ratio refers to the rate of trucks to the overall traffic flow. Peak mean and peak std refer to the mean truck traffic flow and standard deviation in peak hours (11:00–23:00). Low peak mean and low peak std refer to the mean truck traffic flow and standard deviation in low peak hours (0:00–10:00).

*3.2. Spatial Relationships.* To show that the upstream and downstream sections have a spatial relationship, PCC is usually applied to study the interaction between sections. The PCC is defined as the quotient of the covariance and standard deviation between the truck traffic flows of two road sections. The value of PCC is between −1 and 1. When it is closer to 1, it indicates that there is a positive correlation between road sections. The PCC with the upstream and downstream sections of the six selected target sections is shown in Figure 6.

As can be seen from Figure 6, the PCC value is close to 1, indicating that there is a strong spatial correlation between the upstream and downstream sections and the target section. Therefore, there is a certain basis for improving the accuracy of predicting truck traffic by using the spatial relationship between upstream and downstream sections.

*3.3. CC.* Using the fuzzy evaluation method, the TPI of six road sections is obtained as shown in Figure 7.

The congestion situation in the upstream and downstream will affect the flow of the target road section. However, by calculating the CC, the operating status of the road section can become well understood, which is helpful for the prediction of truck flow. Similarly, the CCs of the upstream and downstream sections of the six sections are also calculated.

*3.4. Evaluation.* To analyze the performance of the proposed model, model training and evaluation of the truck traffic in the six selected road sections were carried out. The first four days of data were used for training purposes, and data from the last day were used for verification, within which the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) were applied as parameters for evaluation.

## 4. Results

To explore the prediction performance of the M-BiCNN-GRU model, several baseline models were also trained and used in predicting truck traffic flows. These model include the following:

(i) BiCNNGRU, which considers spatial relationships and temporal characteristics.

(ii) M-BiGRU, which considers multifeatures and temporal characteristics.

(iii) Bi-GRU, which considers temporal characteristics.

(iv) SVR, which considers temporal characteristics.

(v) ARIMA, which considers temporal characteristics.

For the deep learning algorithm, several superparameters that affect the performance of the proposed model need to be discussed; these superparameters include batch size, hidden unit size, hidden layer number, time steps, and epochs. When the values of the evaluation parameter of MAPE, MAE, and RMSE reach their minimum, the superparameters are selected as the optimal parameters. The learning rate in the study is set to 0.001; the batch size is set to 20. Epochs are 100, 230, 130, 240, 160, and 210, respectively, for the six sections of S1, S2, S3, S4, S5, and S6, as determined from multiple iterations of training to find the optimal parameters.

(1) For the training environment, Keras and TensorFlow provided the deep learning packages, and Python 3.6 was used as a general-purpose programming language. These were hosted on a desktop computer, comprising an Intel Xeon (*R*) E5-2640 2.5 GHz CPU and 32 GB memory.

Table 3: The properties of the six selected sections.

| Section | Mean | Std | Mixed ratio | Peak mean | Rush std | Low peak mean | Low peak std |
|---|---|---|---|---|---|---|---|
| S1 | 400.84 | 90.22 | 0.49 | 454.79 | 43.93 | 251.25 | 13.88 |
| S2 | 216.11 | 38.02 | 0.42 | 232.64 | 25.95 | 150.00 | 12.04 |
| S3 | 124.84 | 37.91 | 0.42 | 143.21 | 29.17 | 77.50 | 11.19 |
| S4 | 46.21 | 14.90 | 0.21 | 52.21 | 9.31 | 20.75 | 4.02 |
| S5 | 277.89 | 75.98 | 0.43 | 327.64 | 22.08 | 163.25 | 26.02 |
| S6 | 182.47 | 46.67 | 0.32 | 210.21 | 24.04 | 113.00 | 15.60 |

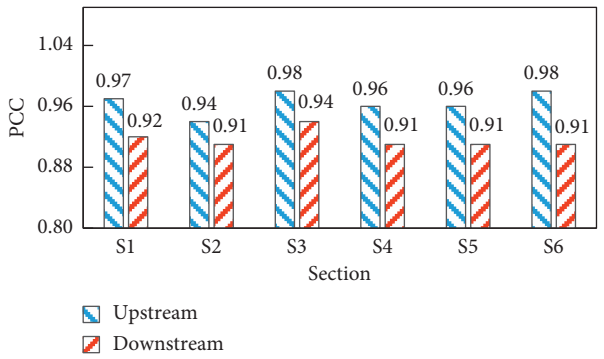



Figure 6: The Pearson correction coefficient with the upstream and downstream sections.

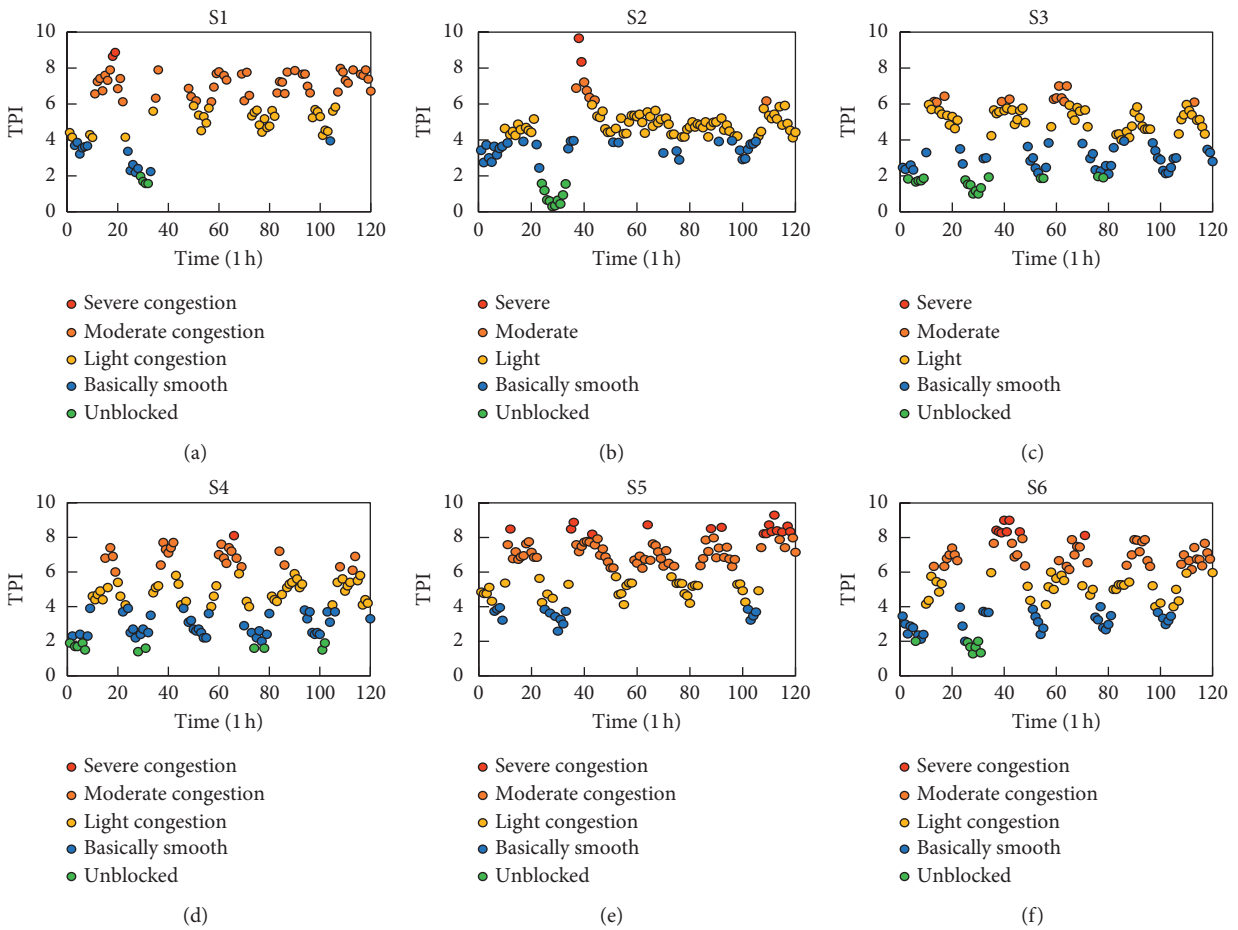



Figure 7: The TPI for the six road sections. (a) S1 (b) S2. (c) S3. (d) S4. (e) S5. (f) S6.

(2) For choosing the optimal hidden units, the evaluation value with different hidden units was compared. Firstly, the hidden units from [5, 10, 15, 30, 50, 100] were tested, respectively. The results indicated that the best parameter occurs for a unit of 5. Then the hidden units from [1, 2, 3, 4, 5] were used in further testing, where it was found that the respective MAPE value for each hidden unit for the road section S1 is 8.32%, 9.31%, 8.59%, 7.62%, and 9.65%. The value of hidden unit 4 represented by the lowest value of MAPE (7.62%) is, therefore, the optimal parameter of the hidden unit.

(3) For different time steps, the proposed M-BiCNN-GRU model has different evaluation performances. The evaluation values of MAPE, MAE, and RMSE with different time steps [4, 8, 12, 16, 20, 24] were tested, and the corresponding evaluation distributions are shown in Figure 8.

From Figure 8, it can be seen that when the time step is gradually increased from 4 to 24, the distributions of MAPE, MAE, and RMSE for road sections S1, S2, S3, S4, S5, and S6 all show a gradually increased trend. When the time step is 4, the error values of MAPE, MAE, and RMSE are at their lowest, suggesting that the proposed model has the best performance and that time step 4 is the optimal parameter. That is to say, the best model fit is obtained by learning the flow of the first 4 hours to predict the flow of the fifth hour. The evaluation results of the above six models are shown in Table 4, and Figure 9.

It can be seen in Table 4 and Figure 9 that the M-BiCNNGRU model, which combines multifeatures, spatial relationships, and temporal characteristics performs better than the baseline models. The average MAPE of the six sections is 7.14, and the prediction accuracy is 92.86%, which is 2.99%, 4.97%, 7.12%, 13.83%, and 13.14% lower than that of BiCNNGRU, M-BiGRU, Bi-GRU, SVR, and ARIMA, respectively. The improved performance in prediction accuracy of the proposed M-BiCNNGRU model indicates that the spatial relationships, temporal characteristics, and operational factors with multifeatures are significantly helpful in predicting truck traffic flow accurately. To further analyze multifeatures, spatial relationships, and temporal characteristics that contribute to the accuracy improvement, the authors designed several models with and without single factors to evaluate the prediction performance.

The BiCNNGRU model was designed by considering the factors of spatial relationships and temporal characteristics; its purpose is to explore whether the road operation factors of multifeatures play a positive role in improving the prediction accuracy of truck traffic flow. The average MAPE value of the BiCNNGRU truck traffic flow prediction model considering spatial and temporal factors is 10.13, which is 2.99 higher than the proposed M-BiCNNGRU model. However, without the road operation factors of multifeatures, the prediction accuracy reduces by 2.99%; hence, the results show that the operation factors of multifeatures is effective in improving the prediction accuracy of truck flow.

Concerning the M-BiGRU model, it was designed by considering the factors of temporal characteristics and the road operation factors of multifeatures; its purpose is to explore whether spatial relationships of upstream and downstream sections play a positive role in improving prediction accuracy. The average MAPE value of the BiCNNGRU model is 12.11, which is 4.97 higher than the proposed M-BiCNNGRU model. The result shows that, without the factor of spatial relationships, the prediction accuracy is reduced by 4.97%. The reduced accuracy of 4.97% by not considering spatial relationships is higher than the reduced accuracy of 2.99% by not considering multifeatures. This result indicates that, compared with multifeatures, the consideration of spatial relationships has a higher accuracy contribution.

The Bi-GRU model was designed by only considering the factor of temporal characteristics. It is a deep learning model which has been widely studied and applied as a result of its good performance in the mining of temporal characteristics. However, does this good performance also apply in truck traffic flow prediction? How much of an increase in truck flow prediction accuracy can be achieved using the Bi-GRU model to mine the temporal characteristics compared to the parameter-type model ARIMA and machine learning model SVR?

The average MAPE of the Bi-GRU model is 14.26, which is 6.71 and 6.02 lower than that of SVR model and ARIMA model, respectively, showing that the Bi-GRU model selected in this paper to capture time characteristics has a good prediction performance and perform better than the parameter-type models and machine learning models. It also indicates that the Bi-GRU model can play a fundamental role in obtaining a high prediction accuracy in the constructed M-BiCNNGRU model.

Furthermore, the average MAPE of the M-BiGRU model considering multifeatures and temporal characteristics is 12.11, which is 2.14 lower than that of Bi-GRU model only considering temporal characteristics. It also indicates that the operation factor of multifeatures is helpful in improving truck flow prediction accuracy, and the improved prediction accuracy value is 2.14%. The average MAPE of the BiCNNGRU model considering the spatial relationship and temporal characteristics is 4.13 lower than the Bi-GRU model only considering temporal characteristics. It also indicates that the factor of spatial relationship is conducive to improving the truck flow prediction accuracy, and the improved prediction accuracy value is 4.13%, which is higher than the improved prediction accuracy value 2.14% of multifeatures. The result shows that the contribution to improving the prediction accuracy of spatial relationship factor is greater than the operation factor of multifeatures.

In summary, the proposed M-BiCNNGRU model, considering three type factors of spatial relationship, temporal characteristics, and road operation multifeatures, outperforms other models in predicting truck traffic flow. Besides, for the contribution to improving the prediction accuracy, the temporal characteristics are the biggest, followed by the spatial relationship and then the road operation factor of multifeatures.

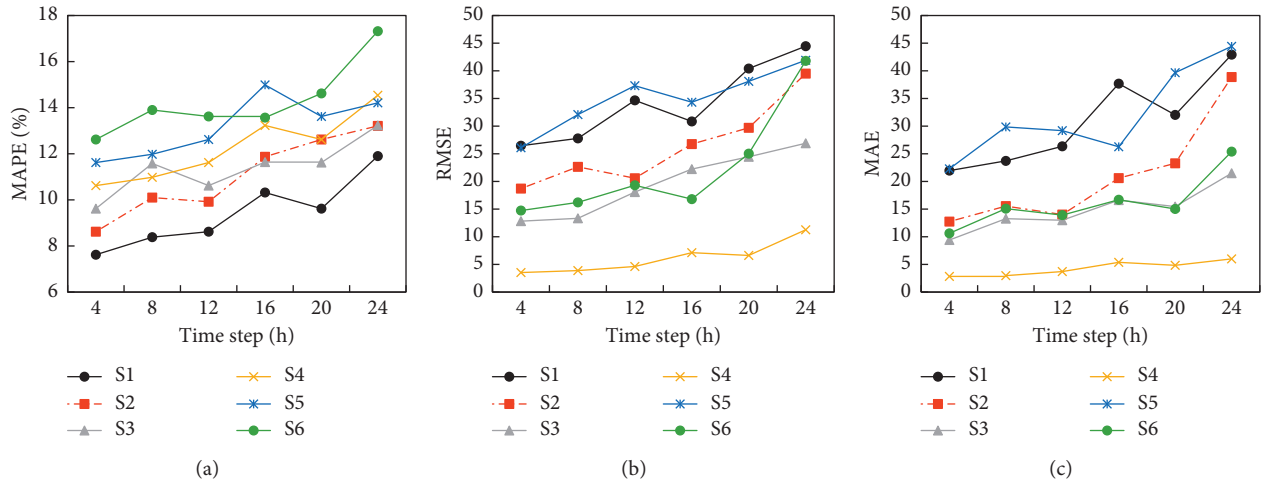



Figure 8: The evaluation values of MAPE, MAE, and RMSE with different time steps.

Table 4: Forecasting effect of peak and low peak.

| Period | Section | S1 | S2 | S3 | S4 | S5 | S6 | Mean |
|---|---|---|---|---|---|---|---|---|
| Peak | RMSE | 28.88 | 13.68 | 12.84 | 3.30 | 37.40 | 16.88 | 18.83 |
| | MAE | 24.73 | 11.82 | 10.14 | 2.63 | 32.77 | 11.62 | 15.62 |
| | MAPE | 5.35 | 5.10 | 6.73 | 4.78 | 9.37 | 5.26 | 6.10 |
| Low peak | RMSE | 39.23 | 6.69 | 12.79 | 3.90 | 22.10 | 10.08 | 15.80 |
| | MAE | 32.12 | 5.20 | 8.16 | 3.17 | 19.43 | 8.89 | 12.83 |
| | MAPE | 11.53 | 2.79 | 8.14 | 10.37 | 10.31 | 7.07 | 8.37 |
| Difference* | RMSE | −10.35 | 6.99 | 0.05 | -0.60 | 15.29 | 6.80 | 3.03 |
| | MAE | −7.39 | 6.63 | 1.98 | −0.54 | 13.34 | 2.73 | 2.79 |
| | MAPE | −6.18 | 2.31 | −1.42 | −5.59 | −0.94 | −1.81 | −2.27 |

*Peak value minus low peak value.

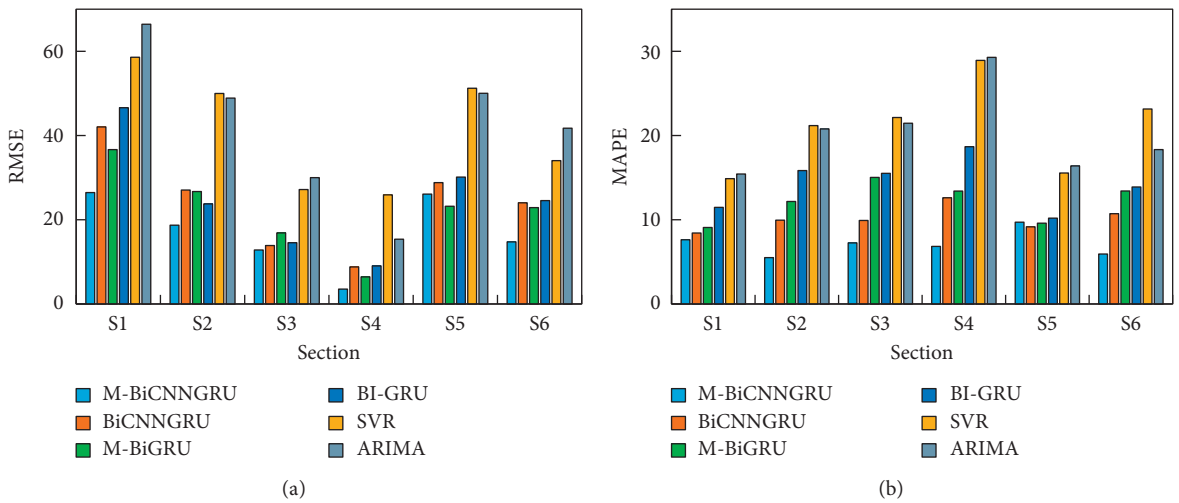



Figure 9: Continued.

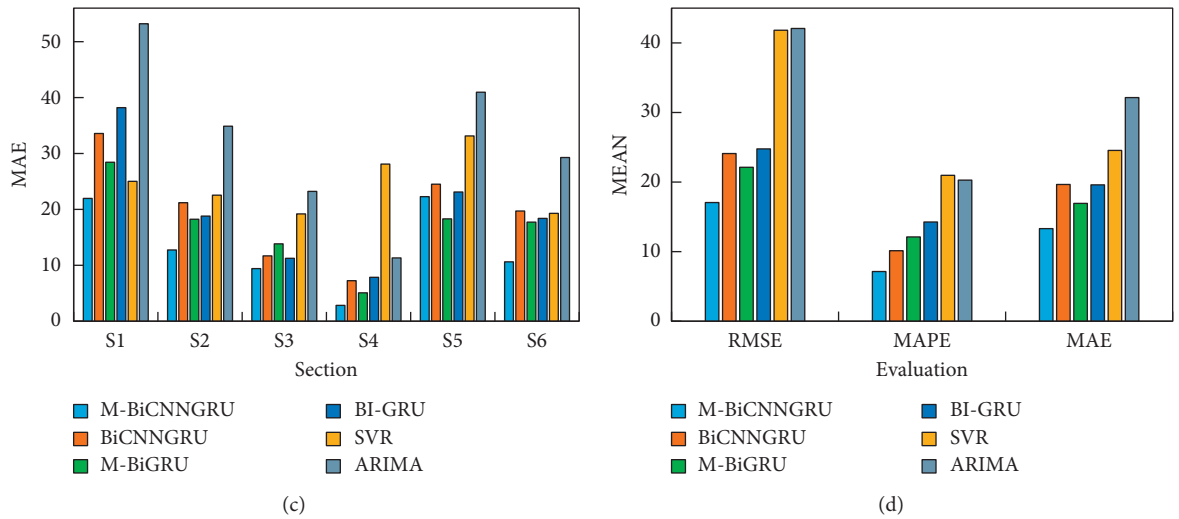



Figure 9: Comparison of prediction performance of various models.

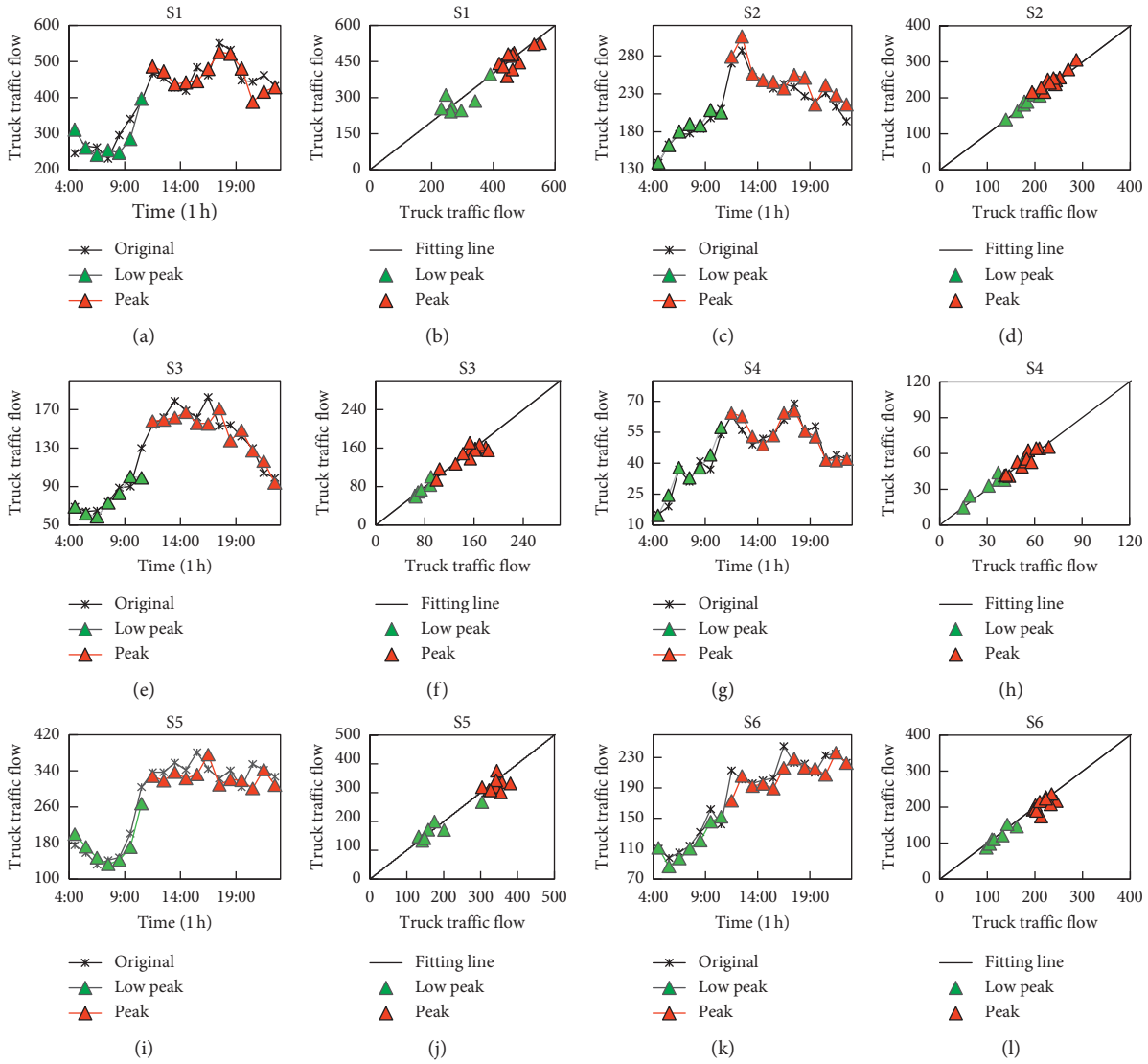



Figure 10: Comparison of the predicted truck flow of peak and low peak.

Table 5: The RMSEs, MAPEs, and MAEs of various models.

| Model | | Evaluation | S1 | S2 | S3 | S4 | S5 | S6 | Mean |
|---|---|---|---|---|---|---|---|---|---|
| M-BiCNNGRU | Multifeatures/spatial/temporal | RMSE | 26.46 | 18.71 | 12.82 | 3.53 | 26.09 | 14.74 | 17.06 |
| | | MAPE | 7.62 | 5.50 | 7.25 | 6.84 | 9.71 | 5.93 | 7.14 |
| | | MAE | 21.96 | 12.74 | 9.41 | 2.83 | 22.28 | 10.62 | 13.31 |
| BiCNNGRU | Spatial/temporal | RMSE | 42.04 | 27.05 | 13.88 | 8.79 | 28.82 | 24.02 | 24.10 |
| | | MAPE | 8.42 | 9.95 | 9.92 | 12.61 | 9.16 | 10.72 | 10.13 |
| | | MAE | 33.58 | 21.19 | 11.69 | 7.25 | 24.51 | 19.71 | 19.65 |
| M-BiGRU | Multifeatures/temporal | RMSE | 36.66 | 26.70 | 16.87 | 6.43 | 23.20 | 22.88 | 22.12 |
| | | MAPE | 9.08 | 12.17 | 15.03 | 13.40 | 9.59 | 13.41 | 12.11 |
| | | MAE | 28.43 | 18.24 | 13.84 | 5.08 | 18.30 | 17.72 | 16.94 |
| Bi-GRU | Temporal | RMSE | 46.60 | 23.76 | 14.54 | 9.03 | 30.13 | 24.55 | 24.77 |
| | | MAPE | 11.47 | 15.84 | 15.51 | 18.67 | 10.19 | 13.89 | 14.26 |
| | | MAE | 38.20 | 18.80 | 11.25 | 7.85 | 23.11 | 18.39 | 19.60 |
| SVR | Temporal | RMSE | 58.59 | 49.99 | 27.18 | 25.91 | 51.23 | 34.04 | 41.16 |
| | | MAPE | 14.88 | 21.18 | 22.14 | 28.92 | 15.55 | 23.15 | 20.97 |
| | | MAE | 25.02 | 22.54 | 19.19 | 28.10 | 33.14 | 19.29 | 24.55 |
| ARIMA | Temporal | RMSE | 66.43 | 48.89 | 30.00 | 15.38 | 50.03 | 41.74 | 42.08 |
| | | MAPE | 15.43 | 20.78 | 21.46 | 29.28 | 16.39 | 18.33 | 20.28 |
| | | MAE | 53.20 | 34.88 | 23.22 | 11.32 | 40.96 | 29.28 | 32.14 |

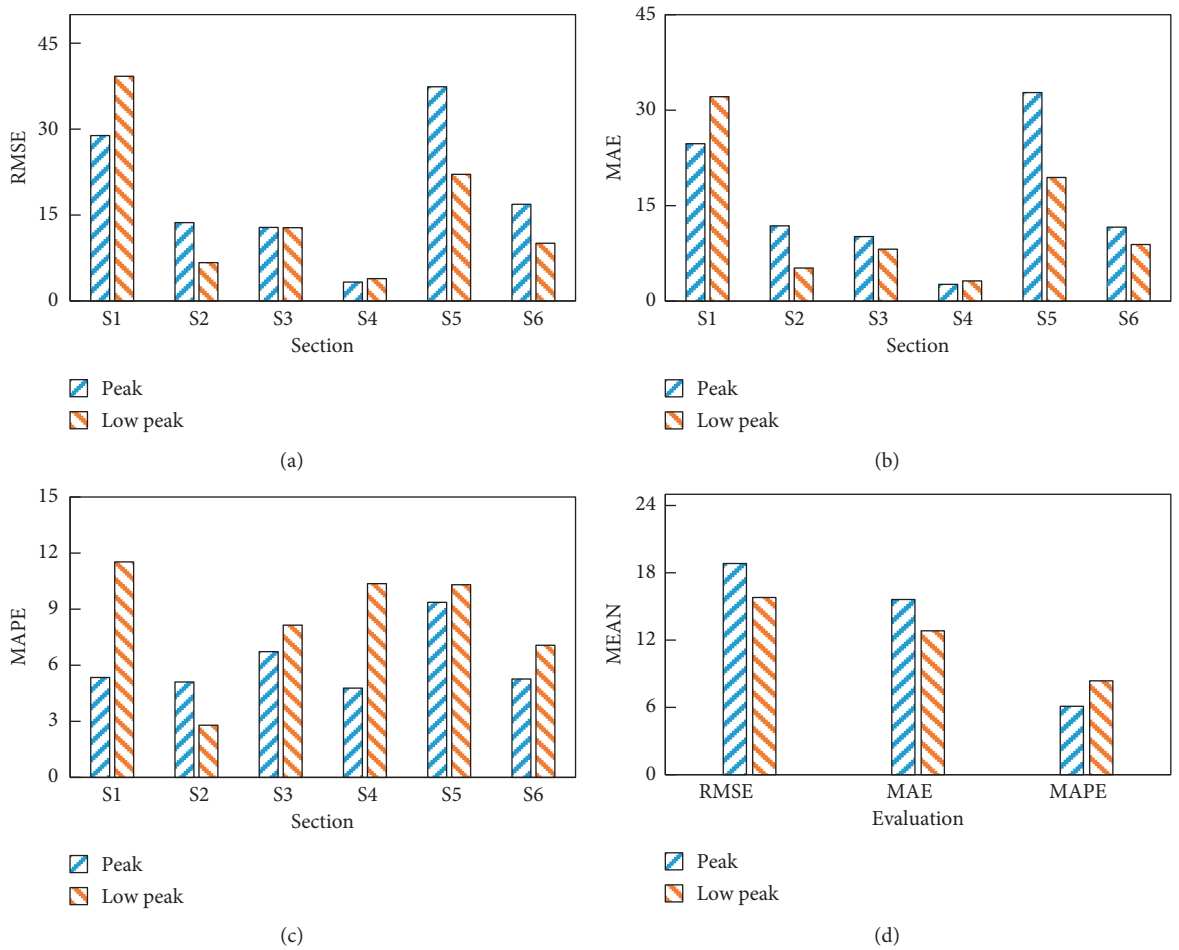



Figure 11: Prediction performance of peak and low peak.

## 5. Discussion

To further discuss the performance of the M-BiCNNGRU model, the original truck traffic flow is compared with the predicted truck traffic flow values. At the same time, this paper divides the day into two periods, low peak period (4: 00–10:00) and peak period (11:00–22:00), and compares the prediction results of truck traffic flow in the two periods. The comparison results are shown in Figure 10.

For an improved quantitative analysis of the truck traffic flow prediction model in both peak and low peak periods, MAE, MAPE, and RMSE indexes were also used for evaluation. The results are shown in Table 5.

It can be seen from Figure 10 and Table 4 that the overall average MAPE of truck flow in the peak period is 2.27 lower than that in a low peak period. The MAPE value of the S1, S3, S4, S5, and S6 road sections in the peak period is 6.18, 1.42, 5.59, and 0.94, respectively. This means that the prediction results are more accurate during peak periods than they are in low peak periods. The comparison of the evaluation parameters of the six sections in peak and low peak periods is shown in Figure 11.

## 6. Conclusions

The paper proposes a Multifeatures Spatial-Temporal-Based model (M-BiCNNGRU) for the accurate forecasting of truck traffic flow. The multifeatures include factors such as truck operational policy, road resistance, truck traffic flow rate, traffic congestion state, and optimal time delay which were all incorporated into the model. The novelty of the model is that it includes the impact factors of multifeatures, as well as temporal characteristics and spatial relationships. The model was trained using data from six road sections from a major expressway network in China collected over a five-day period, where the final day's data were excluded from the training but were instead employed in verifying the performance of the model. To gain further insight into the prediction results of the M-BiCNNGRU model, five baseline models, namely, BiCNNGRU, M-BiGRU, Bi-GRU, SVR, and ARIMA were also applied. The results show that the prediction accuracy of the BiCNNGRU model is 89.87%, which is 4.13%, 10.84%, and 10.15% higher than that of the Bi-GRU, SVR, and ARIMA models. This indicates that the prediction performance of truck traffic flow can be improved by increasing the spatial relationship between upstream and downstream sections. The prediction accuracy of the M-BiGRU model is 87.89%, which is 2.15%, 8.86%, and 8.17% higher than that of the Bi-GRU, SVR, and ARIMA models. This indicates that the incorporation of multifeatures can help to improve prediction accuracy. The prediction accuracy of the M-BiCNNGRU is 92.86%, which is 13.83% and 13.14% higher than that of the SVR and ARIMA models with single temporal characteristics. In summary, the proposed model based on multifeatures, temporal characteristics, and spatial relationships outperforms baseline algorithms as demonstrated in this paper. However, in this paper, the contributions of each individual feature to the overall prediction effectiveness were not specifically analyzed but it is suggested that this could form part of a future study to determine the influence of each feature in terms of weighting factors.

## Data Availability

Some data and code used during the study are available in a repository in accordance with funder data retention policies (https://github.com/uubest/-LSTM-and-GRU).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

The authors confirm contribution to the paper as follows: Chunfu Shao and Shengyou Wang conceptualized and designed the study; Yan Zheng performed data collection; Shengyou Wang, Song Xue, and Yajiao Zhai performed analysis and interpretation of results; Shengyou Wang and Yajiao Zhai prepared the original manuscript. All authors reviewed the results and approved the final version of the manuscript.

## Acknowledgments

The research was supported by funding provided by the Fundamental Research Funds for the Central Universities (Grant no. 2019RC027) and the National Key R&D Program of China (Grant no. 2019YFF0301403). The authors would like to thank Professor Chunjiaodong from Beijing Jiaotong University and Dr. Chaoying Yin from Nanjing Forestry University for their assistance in editing the paper and providing data.

## References

[1] National Bureau of Statistics of China (NBSC), 2020, https://data.stats.gov.cn/easyquery.htm?cn=C016.

[2] K. Gingerich, H. Maoh, and W. Anderson., "Classifying the purpose of stopped truck events: an application of entropy to GPS data," *Transportation Research Part C: Emerging Technologies*, vol. 64, pp. 17–27, 2014.

[3] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proceedings of the 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pp. 324–328, Wuhan, China, November 2016.

[4] G. Miguel, M. Carlos, P. Jared, and M. Eric, "Measurement of travel time reliability of road transportation using GPS data: a freight fluidity approach," *Transportation Research Part A: Policy and Practice*, vol. 130, pp. 240–288, 2019.

[5] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, no. 1, pp. 3–19, 2014.

[6] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: deep belief networks with multitask learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.

[7] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 1–17, 2017.

[8] L. Li, X. Qu, J. Zhang, H. Li, and B. Ran, "Travel time prediction for highway network based on the ensemble empirical mode decomposition and random vector functional link network," *Applied Soft Computing*, vol. 73, pp. 921–932, 2018.

[9] J. Guo, Z. Xie, Y. Qin, L. Jia, and Y. Wang, "Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM," *IEEE Access*, vol. 7, pp. 42946–42955, 2019.

[10] Z. Zhang, M. Li, X. Lin, Y. Wang, and F. He, "Multistep speed prediction on traffic networks: a deep learning approach considering spatio-temporal dependencies," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 297–322, 2019.

[11] X. Luo, D. Li, Y. Yang, and Y. Yang, "Spatiotemporal traffic flow prediction with KNN and LSTM," *Journal of Advanced Transportation*, vol. 1, no. 1, pp. 1–10, 2019.

[12] L. Zhao, Y. Song, C. Zhang, Y. Liu et al., "T-GCN: A temporal graph convolutional network for traffic prediction," 2019, https://arxiv.org/abs/1811.05320.

[13] H. Wang, L. Liu, S. Dong, Z. Qian, and H. Wei, "A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD-ARIMA framework," *Transportmetrica B: Transport Dynamics*, vol. 4, no. 3, pp. 159–186, 2015.

[14] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[15] Y. Xie, Y. Zhang, and Z. Ye, "Short-term traffic volume forecasting using Kalman Filter with discrete wavelet decomposition," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 326–334, 2007.

[16] C. Dong, H. R. Stephen, Q. Yang, and C. Shao, "Combining the statistical model and heuristic model to predict flow rate," *Journal of Transportation Engineering*, vol. 140, no. 10, Article ID 6014001, 2014.

[17] J. Guo, W. Huang, and M. Billy, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 50–64, 2014.

[18] T. Zhou, D. Jiang, Z. Lin, G. Han, X. Xu, and J. Qin, "Hybrid dual Kalman Filtering model for short-term traffic flow forecasting," *IET Intelligent Transport Systems*, vol. 13, no. 6, pp. 1023–1032, 2019.

[19] S. Guo, Y. Lin, S. Li et al., "Deep spatial-temporal 3D convolutional neural networks for traffic data forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, pp. 1–14, 2019.

[20] X. Wang, C. Shao, C. Yin, C. Zhuge, and W. Li, "Application of bayesian multilevel models using small and medium size city in China: the case of changchun," *Sustainability*, vol. 10, no. 2, 2018.

[21] W. Zhang, Y. Yu, Y. Qi, F. Shu, and Y. Wang, "Short-term traffic flow prediction based on spatiotemporal analysis and CNN deep learning," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1688–1711, 2019.

[22] X. Wang, L. Xu, and K. Chen, "Data-driven short-term forecasting for urban road network traffic based on data processing and LSTM-RNN," *Arabian Journal for Science and Engineering*, vol. 4, no. 44, pp. 3043–3060, 2018.

[23] S. Bai, J. Z. Kolter, and V. T. C. N. Koltun, "An empirical evaluation of Generic Convolutional and Recurrent Networks for sequence modeling," in *Proceedings of the International Conference on Learning Representations (ICLR)*, Vancouver, Canada, April 2018.

[24] T. Bogaerts, A. D. Masegosa, J. S. Angarita-Zapata et al., "A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data," *Transportation Research Part C: Emerging Technologies*, vol. 112, no. 3, pp. 62–77, 2020.

[25] R. Asadi and A. Regan, "A spatial-temporal decomposition based deep neural network for time series forecasting," *Applied Soft Computing*, vol. 87, 2020.

[26] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, "LSTM-based traffic flow prediction with missing data," *Neurocomputing*, vol. 27, pp. 297–305, 2018.

[27] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[28] D. Yang, K. R. Chen, M. N. Yang, and X. C. Zhao, "Urban rail transit passenger flow forecast based on LSTM with enhanced long-term features," *IET Intelligent Transport Systems*, vol. 13, no. 10, pp. 1475–1482, 2019.

[29] L. Li, L. Qin, X. Qua, J. Zhang, Y. Wang, and B. Ran, "Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm," *Knowledge-Based Systems*, vol. 172, pp. 1–14, 2019.

[30] X. Cheng, R. Zhang, J. Zhou, and W. Xu, "Deep Transport: learning spatial-temporal dependency for traffic condition forecasting," in *Proceedings of the International Joint Conference on Neural Networks*, pp. 1–8, Rio de Janeiro, Brazil, August 2018.

[31] D. Zhang and M. R. Kabuka, "Combining weather condition data to predict traffic flow: a GRU-based deep learning approach," *IET Intelligent Transport Systems*, vol. 12, no. 7, p. 578, 2018.

[32] Z. Cheng, W. Wang, J. Lu et al., "Classifying the traffic state of urban expressways: a machine-learning approach," *Transportation Research Part A Policy & Practice*, vol. 137, pp. 411–428, 2018.

[33] D. Xu, C. Wei, P. Peng, Q. Xuan, and H. G. E.-G. A. N. Guo, "A novel deep learning framework for road traffic state estimation," *Transportation Research Part C: Emerging Technologies*, vol. 117, 2020.

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 8849234, 11 pages
https://doi.org/10.1155/2021/8849234

*Research Article*

# The Impact of Pedestrian and Nonmotorized Vehicle Violations on Vehicle Emissions at Signalized Intersections in the Real World: A Case Study in Beijing

**Jianchang Huang**[1], **Guohua Song**[1], **Jianbo Zhang**[2], **Zufen Li**[1], **Yizheng Wu**[1], **and Lei Yu**[1,3,4]

[1]*Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China*
[2]*Department of Intelligent Transportation System, Beijing Transport Institute, Beijing 100073, China*
[3]*Department of Transportation Studies, Xuchang University, Xuchang 461000, China*
[4]*Department of Transportation Studies, Texas Southern University, 3100 Cleburne Avenue, Houston, TX 77004, USA*

Correspondence should be addressed to Guohua Song; ghsong@bjtu.edu.cn

Received 29 May 2020; Revised 8 July 2020; Accepted 6 February 2021; Published 2 March 2021

Academic Editor: Keping Li



Emission around intersections has become an issue in the urban traffic network. This paper aims to investigate the impact of pedestrian and nonmotorized vehicle violations on emissions at mixed-traffic flow intersection based on the volumes of vehicles, nonmotor vehicles, and pedestrians. Also, it focuses on the arterial and collector intersections with high vehicle volume and limited space. Running red light and crossing intersection diagonally are two critical violations, accounting for 91.75% of effective violations (interference with vehicles' operation). In this context, a violation blocking model is developed to estimate the blocking probability for each vehicle based on the volumes of pedestrians and nonmotor vehicles. The model includes two scenarios. (1) Through phase: the violation blocking model of running red light is developed based on the survival curve (the relationship between waiting time and running red light probability). (2) Left-turn phase: the violation blocking model at this phase includes two parts: (i) crossing the intersection diagonally model is developed for the first vehicle and (ii) running red light model is developed for subsequent vehicles. The existing emission model can estimate the emissions based on the blocking positions. In the case study, emissions increase with the vehicle volume approaching the saturated flow rate and the volumes of nonmotor vehicles and pedestrians increasing. Results show that the maximum emission increase of CO (carbon monoxide) for through phase and left-turn phase can reach 16.7% and 36.4%.

## 1. Introduction

Emission around intersections has become an issue in the urban traffic network. [1], especially in densely populated metropolitans. Previous studies illustrate that high emissions mainly resulted from the stop-and-go vehicle activities at the intersection [2–4], which further lead to high pollutant exposure pedestrians around the intersections [5].

Violation behaviors frequently exist in developing countries such as China [6]. Frequent violations can not only merely intensify the traffic risk but also increase the stop-and-go activities (see Figure 1). Few studies have investigated the impact of violations of pedestrians and nonmotorized vehicles on the vehicle emissions at real world intersections. A violation blocking model based on real world data is recently developed instead of hypothetical violations [7]. In this paper, the main problem is that how many emissions were caused by pedestrian and nonmotorized vehicle violations based on the volumes of vehicles, nonvehicles, and pedestrians in the real world.

A lot of research focused on intersection emissions, which could be summarized into six categories: the influence

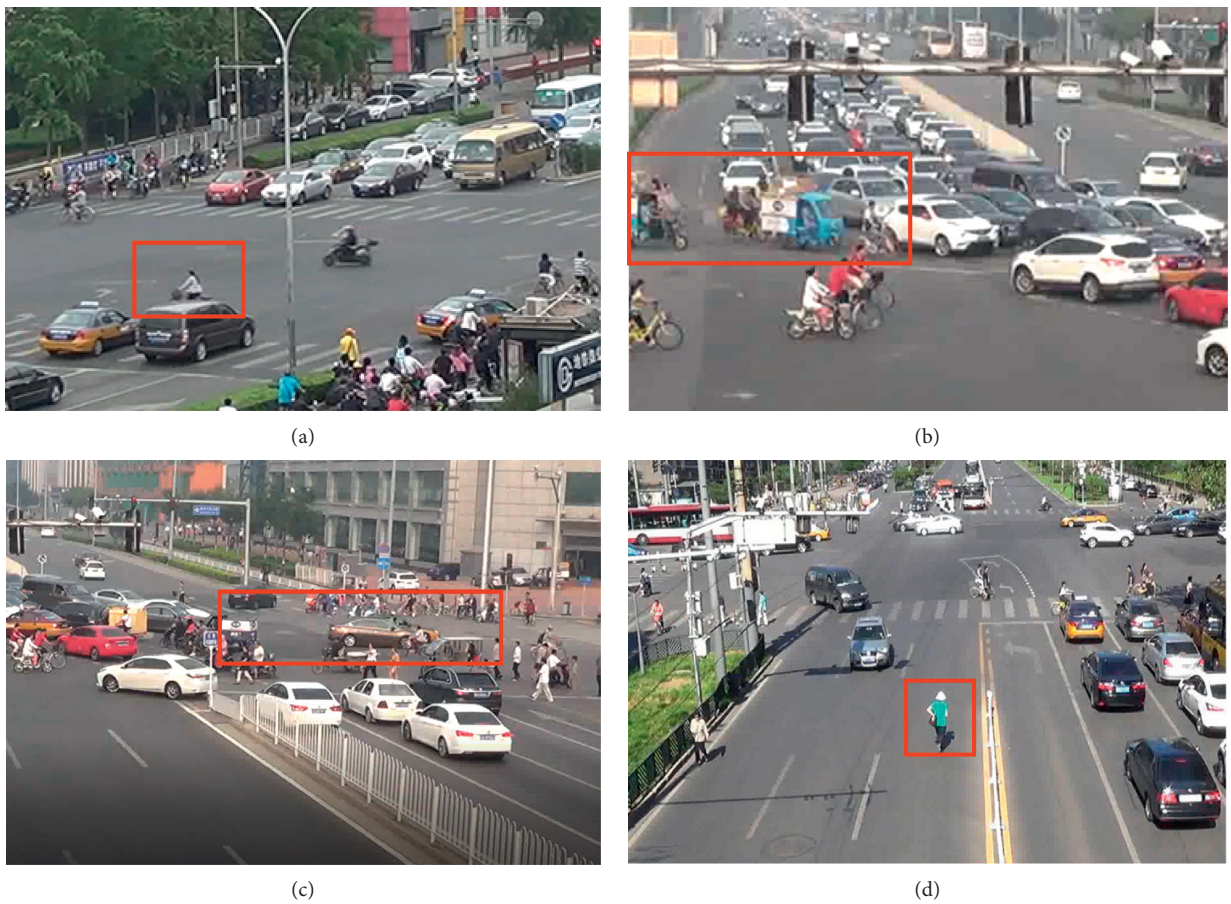


Figure 1: Violations at signalized intersections under mixed-traffic in Beijing. (a) Running red light at the through phase. (b) Running red light at the left-turn phase. (c) Crossing the intersection diagonally. (d) Intruding into the lane.

of emissions caused by traffic congestion, single point signal timing [8–10], signal coordination between intersections [2, 11, 12], intersection shape [13], road characteristics [14], and traffic behaviour. And existing research related to intersection emissions can be divided into two categories: (1) intersection emissions estimation and (2) traffic congestion and traffic behaviour at the intersection.

In the existing studies, Rakha et al. [15] studied the impact of acceleration and deceleration behavior at intersections on emissions and found that emission was more sensitive to cruise under the same speed levels. Papson et al. [16] estimate emissions in a combination with the average emission rate at the four driving modes at signalized intersections. However, the aggregation of four driving modes cannot reflect trajectory characteristics. Zhang et al. [17] developed SIDRA model for the emission estimation at the intersection. Gokhale et al. [18] developed a CO concentration model based on the traffic flow pattern at intersections. Braven et al. [19] estimated emissions based on fuel consumption and emission data obtained from an existing emission inventory dataset.

In terms of congestion, Stevanovic et al. [20] suggested that the moderate speed, shortest delays, and fewest stops are the best vehicles' operation of traffic on arterial roads for emissions. Frequent stops and accelerating at intersections can result in high fuel consumption. The accelerations lead to higher fuel consumption rates compared to idling or deceleration [20]. Many researchers studied determining the factors affecting the emission levels at intersections. These studies show that nonsmooth operations and stop-and-go activities are the most important factors for high emissions at urban intersections [21, 22]. More time was spent in acceleration at intersections because of stop-and-go activities. The vehicle's engine power operates at a higher level during acceleration, and it causes excessive emissions [23, 24].

In terms of traffic behaviour, Sun et al. [25] quantified the effects of Advanced Traffic Signal Status Warning Systems (ATSSWS) on driving behaviour. The systems can reduce traffic emissions at intersections by reducing unnecessary brakes and accelerations. Przybyla et al. [26] studied the changes of vehicle trajectory and follow-up laws under the distracted driving behaviour. The following model of the front and subsequent vehicles was developed under the distracted behaviour. Mudgal et al. [27] investigated the impact of variability of driving behaviours on vehicle emissions at the roundabout intersection.

However, the quantification of the impact of real-world violations (based on the volumes of vehicle, nonmotor vehicle, and pedestrian) on the operation and emissions was

not investigated in the existing studies. Therefore, this paper develops a violation blocking model to estimate vehicles' operation and emissions at the protected intersections of arterials and collectors.

## 2. Materials and Methods

This paper includes three steps. (1) The first is determining the arterial and collector intersections as research object by comparing the violations' frequency of pedestrians and nonmotorized vehicles of different types of intersections and signal control. (2) Second is developing the violation blocking model according to the characteristics of the violations. The blocking probability for each vehicle can be estimated based on the volumes of pedestrians and nonmotor vehicles. (3) Third is estimating emissions based on blocking positions and average lost time (see Figure 2) [7].

*2.1. Data Source.* The paper includes two data sources. (1) Field survey data is used for analysing vehicle behaviors and developing violation mode under the violation at intersections. (2) Emission data is used for quantifying emission.

*2.1.1. Field Survey Data.* The video data of 5 signalized intersections under mixed-traffic flow were collected in Beijing, China, 2017. The data includes intersection attributes and operation data listed as follows:

(1) Intersection attributes
   (a) Channelization information
   (b) Signal information
(2) Operation data
   (a) Fundamental data: this includes two parts: (1) the volume of vehicle, nonmotor vehicle, and pedestrian and (2) the time headway of all vehicles with a precision of 0.02
   (b) Trajectories data: ten vehicles' trajectories, which are interfered by violations, are collected
   (c) Time headway: 388 groups of time headway were collected with a precision of 0.02 s under the influence of violations

*2.1.2. Emission Data.* The vehicle emissions data are derived from local emission rates model for light-duty gasoline vehicles [28, 29]. The emission standard of China III is selected to provide the emission rates for LDVs. The VSP (Vehicle Specific Power) is estimated after the data quality control [30, 31].

The emission factors are estimated with the following procedures based on idling time and the number of stops [7]:

(1) Eighty-five VSP distributions are developed according to the number of stops, idling times, and divisions upstream and downstream (see Table 1);
(2) The average emission factors in each VSP bin are estimated and the emission factors in 85 intervals are estimated, which are corresponding to each operation state of intersection vehicles [7]. The emissions can be quantified based on the idling time and number of stops of vehicles before and after violations.

*2.2. Determination of Study Object.* In Beijing, the arterial roads have abundant traffic volume. Therefore, violation pyramids are developed based on three types of intersections (arterial and arterial, arterial and collector, and arterial and branch) (see Figure 3). Taking pedestrian as an example, the pyramid of violations is divided into three levels: (1) pedestrian volume, (2) violation volume, and (iii) effective violation (interference with vehicles' operation). Effective violation is defined as the violation which can interfere with the vehicles' operation. In the real world, the arterial and collector intersections have a higher proportion of the effective violations, due to the high vehicle volume and limited space (see Figure 3). There are more conflicts and violations between the three components (vehicle, nonmotor vehicle, and pedestrian) in limited space. Therefore, this paper aims to investigate the impact of pedestrian and nonmotorized vehicle violations on emissions at arterial and collector intersections.

At mixed-traffic flow intersection, the category of violations is diverse and their occurrence probability is different. Running red light and crossing the intersection diagonally account for 91.75% of effective violations. Therefore, the object of this research is running red light and crossing the intersection diagonally at arterial and collector intersections.

*2.3. Impact of Pedestrian and Nonmotorized Vehicle Violations on Vehicle Trajectories.* Figure 4 shows the vehicle trajectories which are affected by pedestrian and nonmotorized vehicle violations in which $x$-axis is time, and $y$-axis is distance. The positive ordinate is the downstream the intersection, and the negative ordinate is the position of queuing vehicles. The red point represents the position and time of the violation. $\tau$ is the sum of driver's reaction time and braking time, and $S_\tau$ is the corresponding distance. The dotted lines and solid lines with the same color indicate the same vehicle before and after violations, respectively (see Figure 4).

The vehicle fleet affected by pedestrian and nonmotorized vehicle violations includes two types. (1) The first is idling vehicles (see Figure 4(a)). The trajectory is similar regardless of violations. When idling vehicles delay a start-up by violations, the fleet's idling time and the number of stops will increase. (2) Second is running vehicles (see Figure 4(b)). The impact of violations on the trajectories of running vehicles is different. The first affected vehicle has three processes: deceleration, idling, andacceleration. The fluctuation of the fleet is gradually transmitting to the subsequent vehicles and becoming smaller. The following vehicles will have longer idling time such as the fifth car if the fluctuation is small enough. Then, the subsequent vehicles will cross the intersection with the saturated headway.

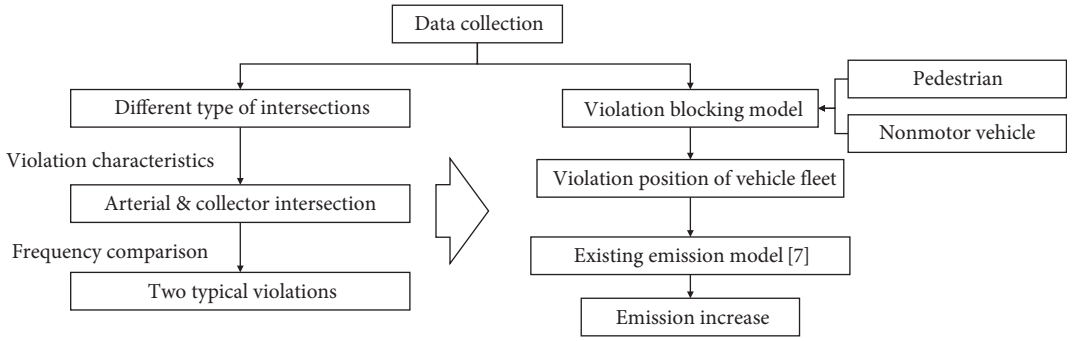



Figure 2: Overview of the methodology.

Table 1: Information of the eighty-five VSP distributions [7].

| The serial number | The spatial position | Number of stops | Idling time(s) |
|---|---|---|---|
| 1 | Downstream intersection | 0 | 0 |
| 2 | Upstream intersection | 0 | 0–2 |
| 3 | Upstream intersection | 1 | 2–5 |
| 4 | Upstream intersection | 1 | 5–10 |
| 5 | Upstream intersection | 1 | 10–15 |
| ⋮ | | | |
| 30 | Upstream intersection | 1 | 135–140 |
| 31 | Upstream intersection | 1 | 140–145 |
| 32 | Upstream intersection | 1 | 145–150 |
| 33 | Upstream intersection | 2 | 10–15 |
| 34 | Upstream intersection | 2 | 15–20 |
| 35 | Upstream intersection | 2 | 20–25 |
| ⋮ | | | |
| 83 | Upstream intersection | 2 | 260–265 |
| 84 | Upstream intersection | 2 | 265–270 |
| 85 | Upstream intersection | 2 | 270–275 |

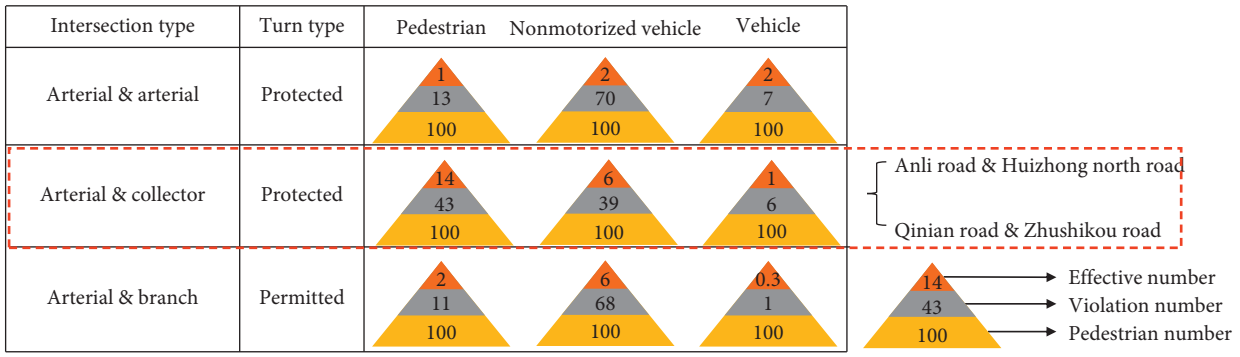



Figure 3: Violation pyramids. Note: protected: go straight and turn left at the same time; permitted: go straight and turn left at the different time.

*2.4. Violation Blocking Model.* In order to estimate the blocking probability for each vehicle, a violation blocking model is developed based on the volumes of vehicles, nonmotor vehicles, and pedestrians. It is a basis for the estimation of the emission increase of violations.

At the intersection of arterial and collector protected, straight and left-turn vehicles cross the intersection at different time and have different violation characteristics. Therefore, this section includes two scenarios: (1) through phase and (2) left-turn phase. All the violation blocking models are developed based on the 55-cycle data of arterial and collector intersections (Anli and Huizhong North have 25 cycles, and Qinian and Zhushikou have 30 cycles).

*2.4.1. Through Phase.* Running red light is the main violation of through phase, because the diagonally crossing nonmotor vehicles can cross the intersection twice legally. The 95.6% of red light runners will lead the head vehicle to

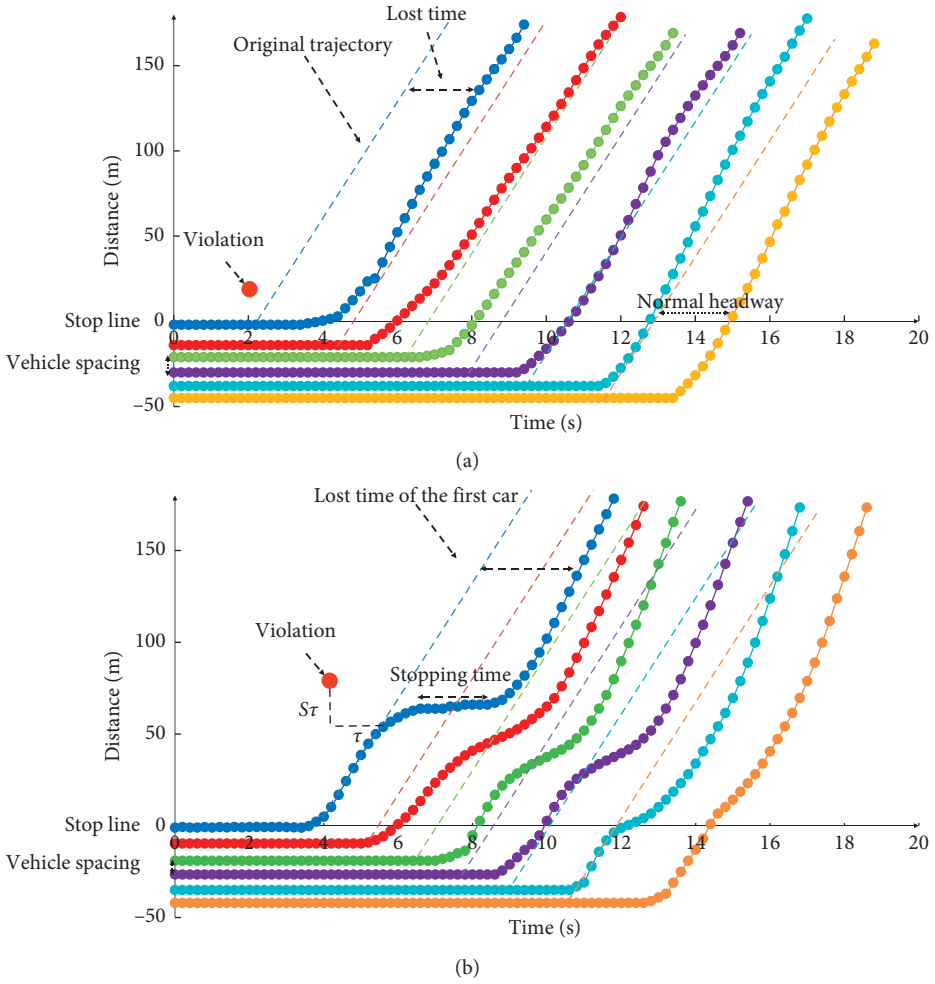



Figure 4: Vehicle trajectories affected by the violation. (a) Idling vehicles. (b) Running vehicles.

delay a start-up or restart; thus, the blocking position can be regarded as the head vehicle at the through phase.

Huang et al. [32] have developed a survival curve to describe the relationship between the probability of running red light and the waiting time.

The blocking probability at the through phase can be estimated based on the probability of running red light and the probability of the fleet being blocked under the running red light condition.

The violation blocking model at the through phase includes three parts: (1) correction coefficient is defined as the probability of vehicle fleet being affected when pedestrian and nonmotorized vehicle violations occur; $k$ equals 0.402 according to the collected data; (2) the volume of pedestrians and nonmotor vehicles is used to estimate how many pedestrians arrive at the waiting area during the specific time; (3) pedestrians and nonmotor vehicles' survival curve is used to estimate how many pedestrians will run red light, and the violation blocking model at the through phase is developed as follows:

$$
\begin{aligned}
P_{\text{through}} &= 1 - \overline{P_{\text{ped}}} \cdot \overline{P_{\text{non}}}, \\
P_{\text{ped}} &= \sum_{i=1}^{\theta} k \cdot C_{\theta}^{i} \cdot a_{\text{ped}}^{i} \cdot \left(1 - a_{\text{ped}}^{i}\right)^{\theta - i} \cdot \left[1 - \left(1 - \beta_{\text{ped}}\right)^{i}\right], \\
P_{\text{non}} &= \sum_{i=1}^{\theta} k \cdot C_{\theta}^{i} \cdot a_{\text{non}}^{i} \cdot \left(1 - a_{\text{non}}^{i}\right)^{\theta - i} \cdot \left[1 - \left(1 - \beta_{\text{non}}\right)^{i}\right], \\
a_{\text{ped}} &= \frac{V_{\text{ped}}}{C}, \\
a_{\text{non}} &= \frac{V_{\text{non}}}{C},
\end{aligned}
\tag{1}
$$

*where* $P_{\text{through}}$ is the probability that the vehicle will be affected in the through phase. $P_{\text{ped}}$ and $P_{\text{non}}$ are the violation probability of pedestrian and nonmotor vehicles. $K$ is correction coefficient, defined as the affected probability of

vehicle fleet when violations occur; $k = 0.402$. $\theta$ ($s$) is the effective interval of nonmotor vehicle and pedestrian; effective interval is the crossing time of pedestrian and nonmotor vehicle. $I$ is used to describe that there are $i$ pedestrians running red light within the time $\theta$. $a_{\text{ped}}$ and $a_{\text{non}}$ are the probability of reaching pedestrians per second. $\beta_{\text{ped}}$ and $\beta_{\text{non}}$ are the probability of running red light when pedestrians and nonmotor vehicles arrive. $V_{\text{ped}}$ and $V_{\text{non}}$ (veh) are the volumes of pedestrian and nonmotor vehicle in every signal period. $C$ ($s$) is signal cycle time.

The two intersections' relative errors are 4.1% and 8.4%, respectively, after inputting the actual data into the model. The result is acceptable due to the randomness of violation behaviors (see Table 2).

*2.4.2. Left-Turn Phase.* At the left-turn phase, the vehicle fleet can be blocked anywhere because the runners can insert the fleet easily due to the lower speed and fewer lanes. The violation blocking model at the left-turn phase includes two scenarios: (1) the first vehicle affected by crossing the intersection diagonally and (2) the subsequent vehicles affected by running red light.

Both pedestrians and nonmotor vehicles have violation behaviors. However, few pedestrians will have effective violations which can interfere with vehicle fleet at the left-turn phase, due to the poor mobility of pedestrians and the long distance between pedestrians and vehicles. Thus, the nonmotorized vehicle is the main research object at the left-turn phase.

*(1) Violation Blocking Model for Crossing the Intersection Diagonally*. The violation blocking model for crossing the intersection diagonally has remarkable scale effect on vehicle fleet. The head vehicle will delay a start-up or restart when several nonmotor vehicles cross the intersection diagonally together.

Based on the above analysis, the model for crossing intersection diagonally includes two steps: (1) the estimation of violation probability based on the volumes of nonmotor vehicles and (2) the estimation of probability that the volume of nonmotor vehicle crossing intersection diagonally exceeds specific scale.

The violation blocking model for crossing the intersection diagonally is fitted by the logarithm model based on the volumes of nonmotor vehicles, and the $R^2$ equals 0.826.

$$\begin{aligned} &P(N_{\text{non}}) = 0.1641 \cdot \ln(N_{\text{non}}) - 0.0978 \quad N_{\text{non}} \geq 0, \\ &P(N_{\text{non}}) \leq 1, \end{aligned} \tag{2}$$

where $P(N_{\text{non}})$ is the probability of crossing intersection diagonally. $N_{\text{non}}$ is the number of nonmotor vehicles that want to arrive the diagonally opposite of intersections when the turn-left light is on.

The nonmotor vehicles will affect the start of the left-turning fleet, when the number of nonmotor vehicles crossing the intersection diagonally reaches a specific scale. The probability model of the number of violations exceeds a specific scale that can be described as follows:

$$P_{\text{left,diagonally}}(N_{\text{non}}) = 1 - \sum_{i=0}^{N_{\text{scale}}-1} C_{N_{\text{non}}}^{i} P(N_{\text{non}})^{i} \cdot [1 - P(N_{\text{non}})]^{N_{\text{non}}-i}, \tag{3}$$

where $P_{\text{left,diagonally}}$ is the probability of the head vehicle affected. $N_{\text{scale}}$ is the critical scale of violations, $N_{\text{scale}} = 6$.

The two intersections' relative errors are 0.32% and 19.56%, respectively, after inputting the actual data into the model (see Table 3).

*(2) Violation Blocking Model for Running Red Light*. Nonmotor vehicles, which run the red light during the all-red time, are always besides the left-turn fleet when the left-turn light turns green. And red light runners will only affect the subsequent vehicles due to the existence of the left-turn waiting zone. The existing study indicates that the blocking probability is increasing with time headway [33] (see Table 4). The violation blocking model for running red light at the left-turn phase includes two steps: (1) developing the Gauss distribution of the time headway and (2) estimating the corresponding blocking probability under the specific time headway.

The probability model of time headway is fitted by Gauss model, and $R^2$ equals 0.782. Then, the corresponding blocking probability can be estimated under the specific time headway.

$$\begin{aligned} &P(t) = 0.0077 + 0.0521 \cdot e^{-2.3222 \cdot (t-2.2864)^2}, \\ &P_{\text{left,run}}(t) = t \longrightarrow G(t), \end{aligned} \tag{4}$$

where $P(t)$ means that the probability of the time headway is $t$. $T$ ($s$) is time headway. $P_{\text{left, run}}(t)$ is the blocking probability of running red light. $G(t)$ is the correspondence between the time headway and the blocking probability.

*2.4.3. Summary of Violation Blocking Model.* The model's purpose is to estimate the blocking probability for each vehicle based on the volumes of nonmotor vehicles and pedestrians. The violation blocking model includes two scenarios: through phase and left-turn phase.

At the through phase, the main pedestrian and nonmotorized vehicle violations are running red light, which will lead the head vehicle to delay a start-up or restart. The violation blocking model at through phase includes two input parameters: (1) volumes of pedestrian and nonmotor vehicle and (2) the survival curve which can quantify the relationship between the probability of running red light and the waiting time.

Table 2: Running red light at the through phase: model error.

| Intersection name | Blocking probability | | Absolute error (%) |
|---|---|---|---|
| | Actual blocking (%) | Violation blocking model (%) | |
| Anli and Huizhong North | 53.3 | 57.4 | 4.1 |
| Qinian and Zhushikou | 60.6 | 69.0 | 8.4 |

Table 3: Cross intersection diagonally at the left-turn phase: model error.

| Intersection name | Blocking probability | | Absolute error (%) |
|---|---|---|---|
| | Actual blocking (%) | Violation blocking model (%) | |
| Anli and Huizhong North | 12.00 | 12.32 | 0.32 |
| Qinian and Zhushikou | 46.67 | 65.23 | 19.56 |

Table 4: Gap accepted by nonmotor vehicles [33].

| Time headway(s) | Refuse block | Accept block | Accept proportion (%) |
|---|---|---|---|
| 1.5–2.0 | 41 | 0 | 0 |
| 2.0–2.5 | 163 | 16 | 9 |
| 2.5–3.0 | 111 | 37 | 25 |
| 3.0–3.5 | 92 | 48 | 34 |
| 3.5–4.0 | 57 | 82 | 59 |
| 4.0–4.5 | 33 | 90 | 73 |
| 4.5–5.0 | 10 | 56 | 85 |
| 5.0–5.5 | 8 | 96 | 92 |
| 5.5–6.0 | 0 | 61 | 100 |

At the left-turn phase, crossing the intersection diagonally affects the head vehicle's operation, and running the red light affects the subsequent vehicles' blockage. The violation blocking model at left-turn phase includes three input parameters: (1) volumes of pedestrian and nonmotor vehicle, (2) the probability model of running diagonally, and (3) the correspondence between the time headway and the blocking probability.

*2.5. Emission Model for Violation Blocking.* The existing emission model can estimate the emissions based on the blocking positions. It includes two parts [7]. (1) The linear emission model is developed considering the number of stops and idling times, which is used for estimating emissions under nonviolation and violation conditions. (2) Violation emission model is developed clearly at two levels: trajectory level and traffic flow level (see Figure 5).

At the trajectory level, the study focuses on the first four vehicles affected by pedestrian and nonmotorized vehicle violations. First, the trajectory model of the head car is developed. Then, the Gipps' car-following model is also applied for other three vehicles. Figure 5(a) shows the trajectories under the impact of pedestrian and nonmotorized vehicle violations in which $x$-axis is time, and $y$-axis is distance. The red point represents the location and time of the violation. $\tau$ is equal to the sum of driver's reaction time and braking time, and $S_\tau$ is the corresponding distance. The first affected vehicle has three processes: deceleration, idling, and acceleration. The fluctuation of the fleet is gradually transmitting to the subsequent vehicles and becomes smaller. When the fluctuation is small enough, the following vehicles will have longer idling time such as the fifth car. Then, the subsequent vehicles will cross the intersection with the saturated headway. The emissions can be estimated based on the speed and acceleration at 1 second interval (see Figure 5(a)).

At the traffic flow level, the study focuses on the subsequent vehicles after the first four vehicles, and the study is divided into unsaturated and saturated scenarios. It is assumed that the total lost time of the first four effected vehicles is 4 s. In the unsaturated scenario, the subsequent vehicles' idling time increases by 4 s, and the number of stops is still one. In the saturated scenario, two vehicles will transform from one stop to two stops due to the increase of idling time. And the two-stop vehicles' idling time increases by 4 s. The emissions can be estimated according to the developed linear emission model, whose input parameters are the number of stops and idling times (see Figure 5(b)).

The sum of the emissions on these two levels is the total emissions at intersections affected by pedestrian and nonmotorized vehicle violations, which can be estimated as in the following equation [7]:

$$\mathrm{AE} = \frac{\left(\sum_{x=e}^{e+3} \mathrm{EF}_{\mathrm{increased,trajectory},x} + \sum_{y=e+4}^{n} \mathrm{EF}_{\mathrm{increased,flow},y}\right) \cdot D}{\sum_{i=1}^{n} \left(\mathrm{EF}_{\mathrm{upstream},i} + \mathrm{EF}_{\mathrm{downstream}}\right) \cdot D} - 1 \tag{5}$$

where AE (%) is the increase of emissions. $E$ is the $e^{\mathrm{th}}$ vehicle, which is the location of the first affected vehicle. $N$ is the vehicle number of the cycles. $\mathrm{EF}_{\mathrm{increased,trajectory},x}$ (g/km) is the increased emission factors of the first four affected vehicles. $\mathrm{EF}_{\mathrm{increased,flow},y}$ (g/km) is the increased emission factors of the subsequent vehicles after the first four affected vehicles. $\mathrm{EF}_i$ (g/km) is the normal emission factors of vehicles. $\mathrm{EF}_{\mathrm{upstream}}$ and $\mathrm{EF}_{\mathrm{downstream}}$ (g/km) are the emission factors at upstream and downstream the intersection. $D$ (km) is the distance of the study range, which is 0.2 km.

## 3. Case Study

Numerical simulations are designed in the case. Based on the violation blocking model and the emission model, the impact of pedestrian and nonmotorized vehicle violations on intersection emissions can be quantified based on the

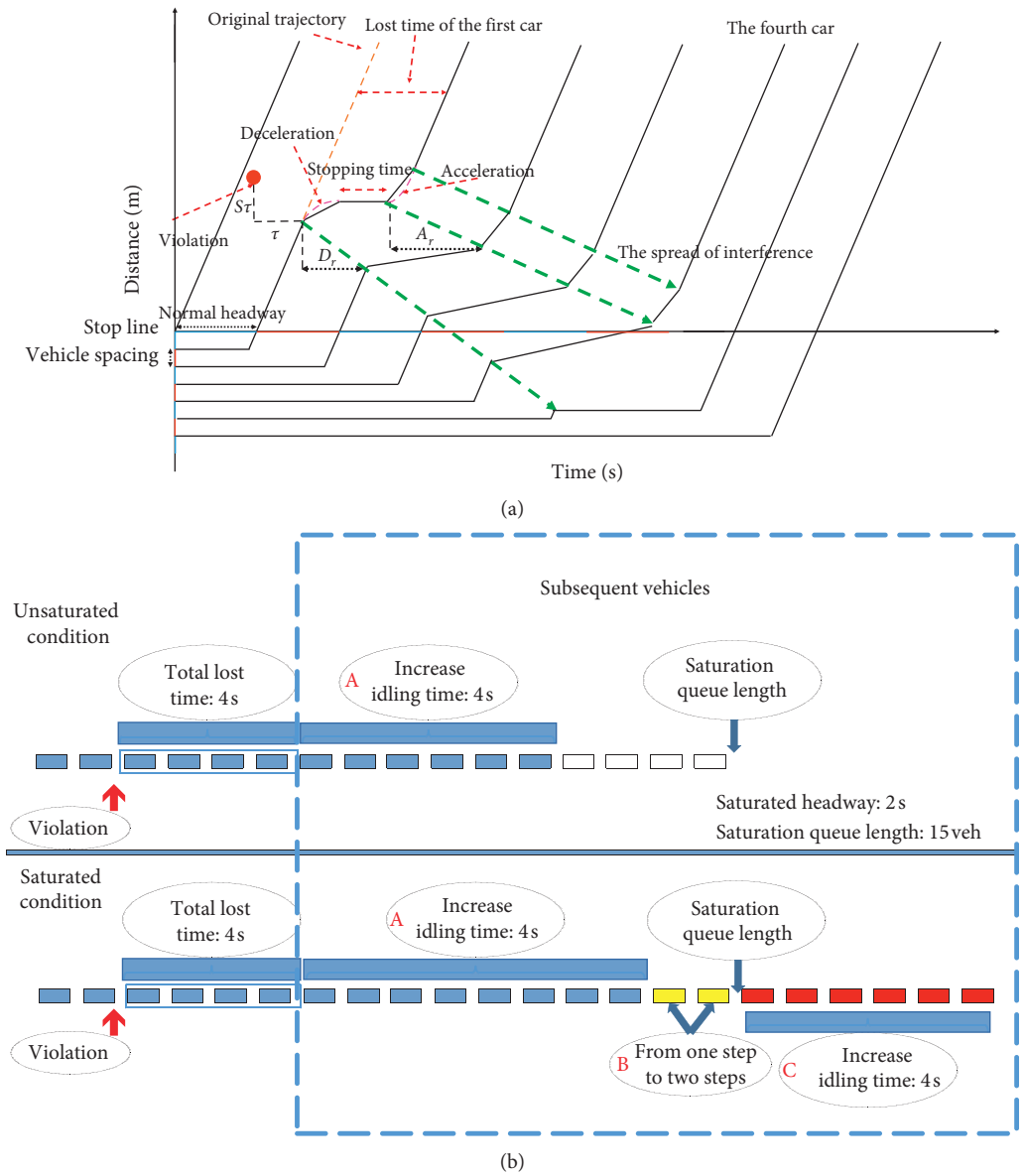



Figure 5: Vehicle characteristics affected by pedestrian and nonmotorized vehicle violations. (a) Trajectory level. (b) Traffic flow level. [7].

volumes of vehicles, nonmotorized vehicles, and pedestrians. The numerical simulation object of the case is the arterial (north-south) direction of the intersection (see Table 5). The five simulation conditions are listed as follows:

(1) Violation average lost time is 5.52 s

(2) The research range is 200 meters around the intersection

(3) The subject of the simulation is all the vehicles arriving at the research scope

(4) The number of pedestrians and nonmotor vehicles is assumed to be the same

(5) The max number of stops is two times

Repetitive simulations are really necessary due to the randomness of violations. The average increase of 50 groups simulations can improve the simulation accuracy.

## 4. Results and Discussion

The emission increase of pedestrian and nonmotorized vehicle violations has $x$-axis indicating the pedestrian/nonmotor vehicles volume and $y$-axis indicating the vehicle volume (see Figure 6).

Emissions increase with the vehicle volume approaching the saturated flow rate and the volumes of nonmotor vehicles and pedestrians increasing. The maximum emission increase of CO for through phase and left-turn phase can reach 16.7% and 36.4%, as shown in Table 6.

The emission increase of left-turn phase is higher than that of the through phase, because the blockings have a greater impact on the left-turn phase due to the shorter green time.

As the volume of pedestrians and nonmotor vehicles' increasing, the emission increase has two processes with the

Table 5: Signal and channelization information of case intersection [7].

| Signal information | | | | Channelization information | | | |
|---|---|---|---|---|---|---|---|
| Phase | Green | Yellow | All red | Direction | Lane number | | |
| North-south straight | 74 | 3 | 2 | | Straight | Left | Right |
| North-south left-turn | 42 | 3 | 2 | North | 3 | 1 | 1 |
| East phase | 34 | 3 | 2 | South | 3 | 1 | Straight-right |
| West phase | 28 | 3 | 2 | East | 0 | 1 | Straight-right |
| Cycle time | | 198 | | West | 1 | 1 | Straight-right |

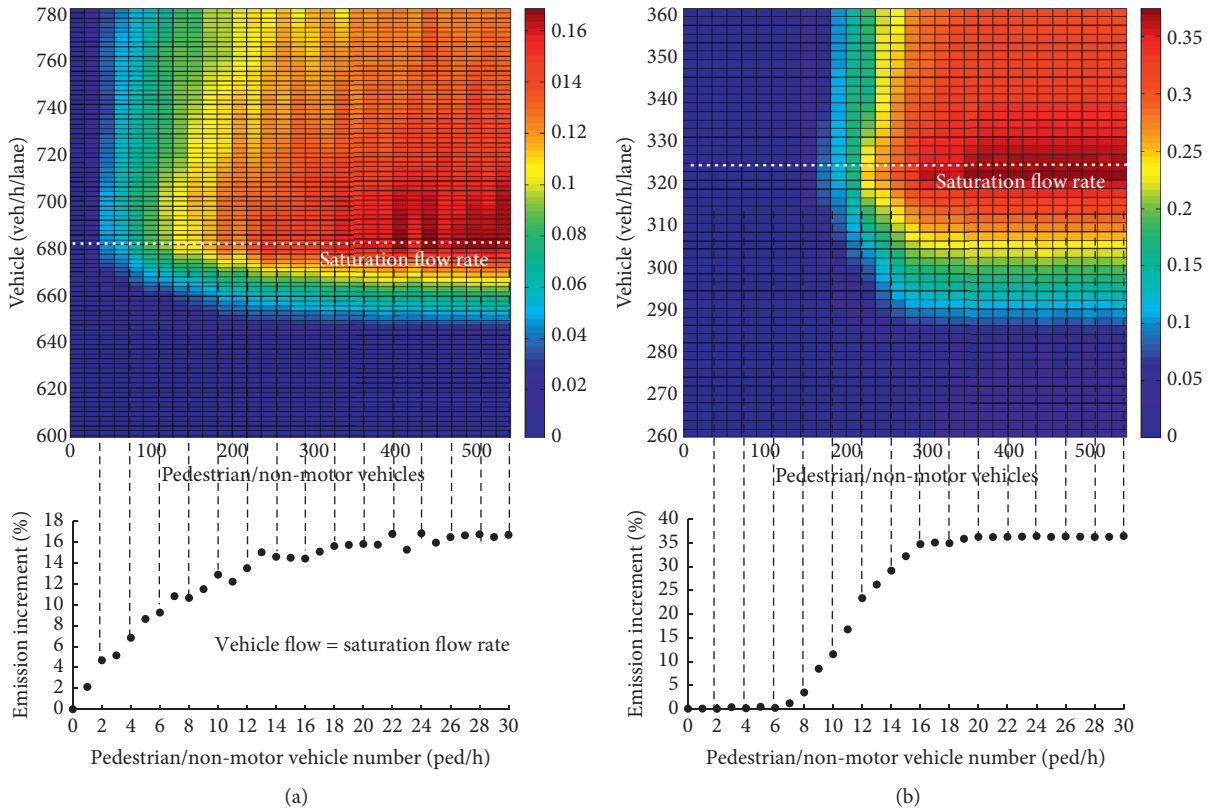



Figure 6: Emission increase of CO after pedestrian and nonmotorized vehicle violations based on the volumes of vehicles, nonmotor vehicles, and pedestrians. (a) Through phase. (b) Left-turn phase.

Table 6: Maximum emission increase of pedestrian and nonmotorized vehicle violations.

| Pollutant | Through phase (%) | Left-turn phase (%) |
|---|---|---|
| $CO_2$ | 32.6 | 66.6 |
| CO | 16.7 | 36.4 |
| $NO_X$ | 20.1 | 47.5 |
| HC | 28.1 | 59.2 |

specific number of vehicles: (1) rising and (2) stable. The violation probability is relatively stable when pedestrians and nonmotor vehicles' volume are at a high level.

As the volume of motor vehicles' increasing, there are three periods of the emission increase. (1) The first is rising slowly. Before the traffic flow reaches the saturated flow rate, there is only one stop regardless of the violation; thus, pedestrian and nonmotorized vehicle violations will only increase the idling time. (2) Second is rising rapidly. When the vehicle approaches the saturated flow rate, vehicles transform from one stop to two stops due to pedestrian and nonmotorized vehicle violations. (3) Third is declining steadily. The emission increase will decline steadily when the volume of vehicles exceeds the saturation flow rate. The proportion of the vehicles, which transform from one stop to two stops because of pedestrian and nonmotorized vehicle violations, reduces. The proportion of the vehicles, which always stops two times regardless of violation, is increasing.

## 5. Conclusions

This paper studies the impact of pedestrian and nonmotorized vehicle violations on emissions at arterial and collector protected intersections in the real world. First, the

characteristics of violations are compared under different types of intersections and violations. Second, violation blocking model is developed to evaluate the blocking position based on the volumes of vehicles, nonmotor vehicles, and pedestrians. Finally, a numerical simulation is conducted to evaluate the impact of pedestrian and nonmotorized vehicle violations on the emissions based on the existing study. Main conclusions can be summarized as follows:

(i) The arterial and collector intersections have a higher proportion of effective violations due to the high vehicle volume and intersection's limited space. Running red light and crossing intersection diagonally are two typical violations at the arterial and collector intersection, which occupy 91.75% of effective violations.

(ii) The proposed violation blocking model based on the volumes of vehicles, nonmotor vehicles, and pedestrians can be used for estimating the blocking probability for each vehicle at the through phase and left-turn phase.

(iii) The variation of emissions is significant based on the volumes of vehicles, nonmotor vehicles, and pedestrians:

  (a) Emissions increase with the vehicle volume approaching the saturated flow rate and the volumes of nonmotor vehicles and pedestrians increasing. The maximum emission increase of CO for through phase and left-turn phase can reach 16.7% and 36.4%.

  (b) As the volume of pedestrians and nonmotor vehicles' increasing, the emission increment has two processes under the specific volume of vehicles: (1) rising and (2) stable. The blocking probability is relatively stable when pedestrians and nonmotor vehicles' volume are at a high level.

  (c) Emission increase is increasing as the vehicle volume is approaching the saturated flow rate. Therefore, as the increasing of the vehicle volume, the emission increase has three processes: (1) rising slowly, (2) rising rapidly, and (3) declining steadily.

This paper develops a method to quantify the impact of pedestrian and nonmotorized vehicle violations on emissions at mixed-traffic flow signalized intersections. Further studies would be conducted on the general model, distribution of lost time of different violations, and red light pedestrians at left-turn phase.

## Data Availability

The emission data come from cited literature in [7, 28–31], and violation data come from manual collection of the project.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was supported by the National Key R&D Program of China (no. 2018YFB1600700) and the Natural Science Foundation of China (NSFC) (nos. 71871015, 51678045, and 71901018).

## References

[1] Federal Highway Administration, *Carbon Monoxide Categorical Hot-Spot Finding*, Federal Highway Administration, New York, NY, USA, 2017.

[2] Y. Zhang, X. Chen, X. Zhang, G. Song, Y. Hao, and L. Yu, "Assessing effect of traffic signal control strategies on vehicle emissions," *Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 1, pp. 150–155, 2009.

[3] L. Zhang, Y. Yin, and S. Chen, "Robust signal timing optimization with environmental concerns," *Transportation Research Part C: Emerging Technologies*, vol. 29, no. 1, pp. 55–71, 2013.

[4] S. Yagar, B. Han, and J. Greenough, "Real-time signal control for mixed traffic and transit based on priority rules," in *Proceedings of the Presented at 71th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, March 1992.

[5] M. Franklin, X. Yin, R. Urman, R. Lee, S. Fruin, and R. Mcconnell, "Environmental factors affecting stress in children: interrelationships between traffic-related noise, air pollution, and the built environment," in *Proceedings of the HEI Annual Conference*, Washington, DC, USA, May 2019.

[6] J. Xing, J. Hua, and P. Hao, "Study on pedestrian crossing rate at signal-controlled intersections," *Journal of Technology & Economy in Areas of Communications*, vol. 16, no. 5, pp. 14–19, 2014, Chinese article.

[7] J. C. Huang, G. Song, J. Zhang, C. Li, Q. Liu, and L. Yu, "The impact of violations on vehicles on vehicle emissions at signalized intersections," *Journal of Advanced Transportation*, vol. 202014 pages, 2020.

[8] B. D. Coensel, A. Can, B. Degraeuwe, I. D. Vlieger, and D. Botteldooren, "Effects of traffic signal coordination on noise and air pollutant emissions," *Journal of Environmental Modelling & Software*, vol. 35, no. 4, pp. 74–83, 2012.

[9] J. Lai, L. Yu, G. Song, and X. Chen, "Emission characteristics of heavy-duty diesel transit buses at intersections of urban area: case study in beijing," *Journal of Aids Care*, vol. 26, no. 10, pp. 1242–1248, 2014.

[10] S. Midenet, F. Boillot, and J.-C. Pierrelée, "Signalized intersection with real-time adaptive control: on-field assessment of CO2 and pollutant emission reduction," *Transportation Research Part D: Transport and Environment*, vol. 9, no. 1, pp. 29–47, 2004.

[11] F. Tao, Q. Shi, and L. Yu, "Evaluation of effectiveness of coordinated signal control in reducing vehicle emissions during peak hours versus nonpeak hours," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2233, no. 2233, pp. 45–52, 2011.

[12] J. Lv and Y. Zhang, "Effect of signal coordination on traffic emission," *Transportation Research Part D: Transport and Environment*, vol. 17, no. 2, pp. 149–153, 2012.

[13] J. C. Krogscheepers and M. Watters, "Roundabouts along rural arterials in South Africa," in *Proceedings of the Preseented at 93th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2014.

[14] S. Pandian, S. Gokhale, and A. K. Ghoshal, "Evaluating effects of traffic and vehicle characteristics on vehicular emissions near traffic intersections," *Transportation Research Part D: Transport and Environment*, vol. 14, no. 3, pp. 180–196, 2009.

[15] H. Rakha and Y. Ding, "Impact of stops on vehicle fuel consumption and emissions," *Journal of Transportation Engineering*, vol. 129, no. 1, pp. 23–32, 2003.

[16] A. Papson, S. Hartley, and K. Kuo, "Analysis of emissions at congested and uncongested intersections with motor vehicle emission simulation," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2012pp. 124–131, Washington, DC, USA, 2010.

[17] Y. Zhang, J. Lv, and W. Wang, "Evaluation of vehicle acceleration models for emission estimation at an intersection," *Transportation Research Part D: Transport and Environment*, vol. 18, pp. 46–50, 2013.

[18] S. Gokhale and S. Pandian, "A semi-empirical box modeling approach for predicting the carbon monoxide concentrations at an urban traffic intersection," *Atmospheric Environment*, vol. 41, no. 36, pp. 7940–7950, 2007.

[19] K. Braven, A. Abdel-Rahim, K. Henrickson et al., *Modelling Vehicle Fuel Consumption and Emissions at Signalized Intersection*, Publication KLK721. National Institute for Advanced Transportation Technology, University of Idaho, Moscow, ID, USA, 2012.

[20] A. Stevanovic, J. Stevanovic, K. Zhang et al., "Optimizing traffic control to reduce fuel consumption and vehicular emissions: integrated approach with VISSIM, CMEM, and VISGAOST. Transportation research record," *Journal of the Transportation Research Board*, vol. 2128, p. 105, 2009.

[21] K. Chen and L. Yu, "Microscopic traffic-emission simulation and case study for evaluation of traffic control strategies," *Journal of Transportation Systems Engineering and Information Technology*, vol. 7, no. 1, pp. 93–99, 2007.

[22] H. Rakha and Y. Ding, "Impact of stops on vehicle fuel consumption and emissions," *Journal of Transportation Engineering*, vol. 129, no. 1, pp. 23–32, 2002.

[23] H. C. Frey, N. M. Rouphail, and H. Zhai, "Speed- and facility-specific emission estimates for on-road light-duty vehicles on the basis of real-world speed profiles," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1987, no. 1, pp. 128–137, 2006.

[24] H. Zhai, H. C. Frey, and N. M. Rouphail, "A vehicle-specific power approach to speed- and facility-specific emissions estimates for diesel transit buses," *Environmental Science & Technology*, vol. 42, no. 21, p. 7985, 2008.

[25] X. Sun, X. Chen, Y. Qi et al., "Analyzing the effects of different advanced traffic signal Status warning systems on vehicle emission reductions at signalized intersections," in *Proceedings of the Presented at 95th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, December 2016.

[26] J. Przybyla, J. Taylor, J. Jupe, and X. Zhou, "Estimating risk effects of driving distraction: a dynamic errorable car-following model," *Transportation Research Part C: Emerging Technologies*, vol. 50, pp. 117–129, 2015.

[27] A. Mudgal, S. Hallmark, A. Carriquiry, and K. Gkritza, "Driving behavior at a roundabout: a hierarchical bayesian regression analysis," *Transportation Research Part D: Transport and Environment*, vol. 26, no. 1, pp. 20–26, 2014.

[28] D. Xie, G. Song, J. Guo, J. Sun, and L. Yu, "Development and application of an online dynamic emission model for traffic networks: a case study of Beijing," in *Proceedings of the Presented at97th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, June 2018.

[29] C. Li, L. Yu, W. He, Y. Cheng, and G. Song, "Development of local emissions rate model for light-duty gasoline vehicles: beijing field data and patterns of emissions rates in EPA simulator," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2627, no. 1, pp. 67–76, 2017.

[30] Report of overall architecture design and technical difficult of the distribution regular pattern of traffic emissions, *Technical Report for Center for Transportation Sector Energy Reduction and Emissions Mitigation of Beijing*, Beijing Jiaotong University, Beijing, China, 2014.

[31] J. Zhang, L. Yu, J. Guo, Y. Cheng, W. He, and G. Song, "Optimized adjustment of speed resolution and time alignment data for improving emissions estimations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2570, no. 1, pp. 77–86, 2016.

[32] Y. L. Huang and S. Ni, "Analysis on violation behavior of non-motor vehicles at plane intersection," *Journal of Transportation Engineering and Information*, vol. 15, no. 2, pp. 64–70, 2017, Chinese article.

[33] L. J. Xu and W. Wang, "Analysis of influence of left-turn non-motors in signalized intersection," *China Journal of Highway and Transport*, vol. 19, no. 1, pp. 89–92, 2006, Chinese article.

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 6652278, 14 pages
https://doi.org/10.1155/2021/6652278

*Research Article*

# Determining the Operator for the Public Toll Road

**Bin Shang**,[1] **Qiang Sun**,[2] **Hao Feng**,[3,4] **and Jiancong Chang**[1]

[1]*Shanghai Jian Qiao University, Shanghai 201306, China*
[2]*China Northeast Municipal Engineering Design and Research Institute, Changchun 130021, China*
[3]*The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China*
[4]*Academy of Forensic Science, Shanghai 200063, China*

Correspondence should be addressed to Hao Feng; fengh@ssfjd.cn

Received 13 October 2020; Revised 24 December 2020; Accepted 31 December 2020; Published 18 January 2021

Academic Editor: Chunjiao Dong



After the BOT road operation contract expires, generally, the road will be transferred to the government, and then the government operates the road independently without charging costs from its users. Facing the huge amount of the operation cost, Chinese government tends to continue to charge the road users to guarantee the high quality of road operation. Then, the government will have to decide whether a private firm or government itself would be suitable to operate the road. A model is presented for decision-making through balancing interests between the government and the private firm with an introduction of an intermediate variable, i.e., bidding price. Three scenarios are investigated in the model, including the optimization of government operation, the optimization of private firm operation, and government operation with an improper decision of the intermediate variable. Improper intermediate variable will result in a higher toll charged by the government than by a private firm. The method of avoiding an improper decision is investigated. The result shows that the intermediate variable should be determined to be the government operation cost, based on which the private operator could be chosen, if available. With consideration of the private operator's profit to be guaranteed by the government, the maximum subsidy should be equal to the minimum private operator's profit to be disclosed when the contract is signed.

## 1. Introduction

Due to the tight government budget, a lot of transportation facilities, especially highways, were built in the build-operate-transfer (BOT) mode in the past forty years. At present, more and more BOT contracts for the operation of roads have expired or will be going to expire.

When the roads previously operated in the BOT mode are transferred to the government, it is usually assumed that they will become public roads, and the government will take over the responsibility of operating and maintaining these roads with its own budget. In fact, the operation cost (OC) including but not limited to the maintenance cost is much less than the initial construction cost of the roads. In general, no toll will be charged for the public roads. In fact, in order to guarantee the standard of the service relating to the road operations, the government needs to invest heavily to cover the OC. Due to the tight fiscal budget and for decreasing the government financial burden, continuing tolling after the transfer of the roads becomes a logical decision and, to some extent, a universal operation mode. In fact, Chinese government intends to charge the fees for these roads [1]. Chinese government could decide either to charge in its own name or entrust a private firm with this task.

Based on the above description, the road outside the BOT concession period is referred to as the public toll road (PTR) hereinafter, and the toll will be charged continuously to make up the OC. Tan et al. [2] and Niu and Zhang [3] pointed out that normally, a private firm would define the whole road life cycle as the BOT concession period in the BOT contract. However, it is not the case in China. All the BOT roads will have to be transferred to the government when the contracts expire. In this paper, we assume that PTR begins a new period after the road has been transferred to the government. The road life cycle, including the BOT concession period and PTR period, is illustrated in Figure 1.

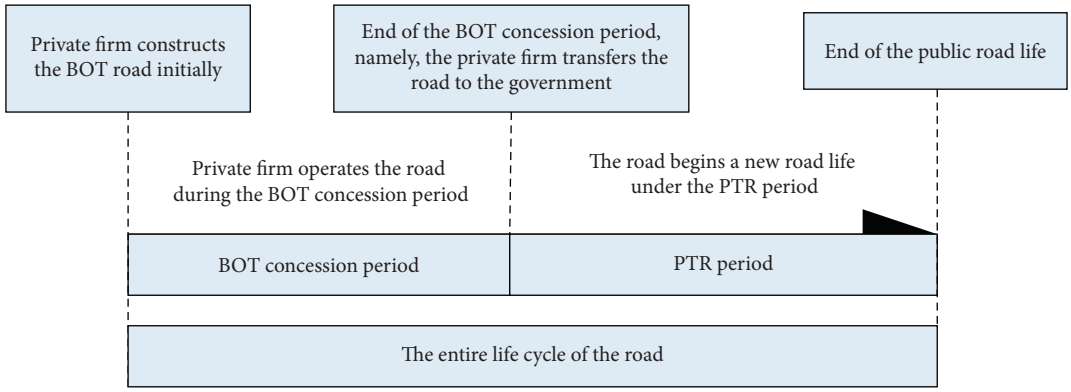



Figure 1: Relationship of the BOT concession period and PTC period.

There are numerous studies on the BOT road construction and operation [2–12]; however, few researchers consider the case after the roads are transferred to the government. To our best knowledge, the influence of OC during the PTR period has not been intensively studied yet. This paper aims to investigate whether the government or a private firm should operate the road under the charge of road toll if the government decides to charge to compensate the OC. This paper comes up with a new concept compared with previous studies concerning the BOT road. Previous research studies generally focus on the selection of the road operator among private firms in the BOT mode. This paper will explicitly examine the selection mechanism to determine whether the government or a private firm will operate the road under the charge of road toll during the PTR period. Note that the revenue collected from road users will be used to compensate the OC, which is different from traditional pricing scenarios such as tolling for the return of the loan (BOT and public-private-partnership).

Referring to Shi et al.'s study [13], we regard that OC is composed of two parts: the demand-related OC and capacity-related OC. The demand-related OC denotes the cost relating to traffic demand; the capacity-related OC denotes the cost relating to the capacity of the road, including but not limited to the maintenance cost. Then, we introduced an intermediate variable presented by the government in the bidding document, which serves as perfect information to both government and private firms. The government and private firms are fully aware of their respective OCs, while the government is not clear about the potential profit to be earned by private firms. We consider the intermediate variable as crucial, based on which we studied the case by analyzing the relationship of the intermediate variable, OC of the private firm, and that of the government. The government may determine an improper intermediate variable based on the limited information. In contrast, we provide an optimal solution that the government sets its anticipated operation cost as the intermediate variable.

The remainder of the paper is organized as follows: Section 2 reviews the previous studies relating to toll road operation. Section 3 introduces the basic notations and assumptions necessary to our analysis model. Section 4 looks into the issue whether the government or a private firm is proper to operate the PTR under the condition that the toll is charged if the government decides to collect the toll to cover the OC. Section 5 investigates the optimal solution to be adopted by the government. Finally, Section 6 concludes this paper.

## 2. Literature Review

Literature on auction mechanisms of the BOT project is closely related to our research. These studies can let the government select the operator from several private firms efficiently. The least-present-value-of-revenue (LPRV) auction mechanism is used on highway franchise. According to auction theory [14], LPRV mechanism equaled to a type of auction: the first-price sealed-bid where the winner was the firm who offered the lowest bid for total revenue. Engel et al. [5] presented a framework to evaluate different awarding mechanisms for the franchise contract, and based on the LPRV auction mechanism, they presented a simple auction. Furthermore, Engel et al. [15] showed that the commonly used fixed-term contract of franchise highway did not allocate the demand risk optimally and then proposed a new auction mechanism for the franchise highway based on the LPRV mechanism.

On the variable-term-based concept [5], Rus and Nombela [16] pointed out that if operation cost was not negligible, the LPVR mechanism would not guarantee the best concessionaire; then, they proposed the least-present-value-of-net-revenue (LPRNV) mechanism which required private firm making offers on total revenue, maintenance costs, annual operation and routine maintenance cost, and so on; and this mechanism eliminated the demand risk effectively. Based on a flexible-term contract and bidimensional bids for maintenance costs and total net revenue, Nombela and de Rus [17] proposed a new franchising mechanism, comparing with fixed-term concessions, which could promote the selection of efficient concessionaires through eliminating the traffic risk. Verhoef [18] investigated welfare impacts of franchise regimes for congestible toll roads and found that patronage-maximizing auction was not optimal when considering the second-best network

aspects. Ubbels and Verhoef [19] looked into how to get the desired welfare through analyzing the auction rules which were set by the government in the process of bidding and found that a bid with minimal price on the toll road would lead to satisfying results for unsubsidy auction; however, considering subsidies, the best bid would appear to involve the minimum generalized travel costs. van den Berg [20] investigated private supply of two congestible infrastructures under four market structures including a monopoly and three duopolies that differed in how private firms interact with each other and found that a duopoly could lead to a different rule of capacity than the first-best one which differed from the duopoly with the facilities.

However, all the above studies focus on the BOT road operation as well as how to choose a private road operator. Any existing selection mechanisms decide a private road operator without considering the government as one of the potential road operators. Therefore, compared to the previous studies, we investigate a mechanism to select the best party to operate the road during the PTR period among the government and private firms.

## 3. Notation and Assumptions

As mentioned above, we are aware that Chinese government will have to decide whether to collect the toll in its own name or to entrust a private firm with this task. Furthermore, we assume that the PTR is the unique one connecting two neighboring cities, and the private firms are able to operate the road in a more efficient and profitable way than the government. Besides, we further assume that the social welfare for the public and the private firm's profit derived from operating the PTR are functions of the toll pricing and the road capacity.

Let $v \geq 0$ and $c \geq c_1$ be the travel demand and road capacity, respectively, which are measured by the number of vehicles per unit time; $c_1$ is the road capacity when it is transferred to the government. Let $B(v)$ be the inverse demand function (or a function of the marginal benefit) and $t(v,c)$ be the travel time function. Moreover, the demand-supply equilibrium condition holds, i.e., $B(v) = p + \beta t(v,c)$, where $p$ is the toll charged by the operator and $\beta$ is the value of time that converts time into an equivalent monetary cost. The above condition means that the travel demand of a new road is determined by the total cost (including travel time and toll collected from road users) of a trip. In this paper, we only consider homogeneous users. Subsequently, we get a price function:

$$p(v,c) = B(v) - \beta \cdot t(v,c). \tag{1}$$

Equation (1) implies that the toll $p$ is determined by travel demand $v$ uniquely for a given road capacity $c$. Hereinafter, $v$ and $c$ are equivalent to $p$ and $c$.

Let $M_s(v)$ and $M_g(v)$ be the demand-related OC of a private firm and the government, respectively. Let $M_d(v)$ be an intermediate demand-related OC which is a function of the intermediate variable. Then, we have

$$\begin{aligned} M_s(v) &= m_s v, \\ M_g(v) &= m_g v, \\ M_d(v) &= m_d v, \end{aligned} \tag{2}$$

where $m_s$ and $m_g$ are the unit demand-related OC of the private firm and the government, respectively, and $m_d$ is the intermediate variable.

Let $J_s(v,c)$ and $J_g(v,c)$ be the OC of a private firm and the government, respectively. Let $J_d(v,c)$ be an intermediate OC. Then, we have

$$\begin{aligned} J_s(v,c) &= I(c) + M_s(v) = I(c) + m_s v, \\ J_g(v,c) &= I(c) + M_g(v) = I(c) + m_g v, \\ J_d(v,c) &= I(c) + M_d(v) = I(c) + m_d v, \end{aligned} \tag{3}$$

where $I(c)$ is the capacity-related OC of the toll road. The construction material information of a road is known in the market, so we assume that $I(c)$ is the perfect information and identical for a private firm and the government in the paper. Given the same $I(c)$, we can regard $m_d$ as the intermediate variable that is set by the government in the bidding document. In practice, based on $m_d$, the government will evaluate the road OC including the management cost and construction cost before determining the road operator. Throughout this paper, we make the following assumptions on $B(v), t(v,c), M_s(v), M_g(v)$, and $I(c)$.

*Assumption 1*

(a) The inverse demand function, $B(v)$, is a strictly continuously decreasing and differentiable function of $v$ for any $v \geq 0$.

(b) The travel time function $t(v,c)$ is a continuously differentiable and strictly convex function of $v$ and $c$ for $v \geq 0$ and $c \geq 0$. $t(v,c)$ strictly decreases with $c$ for $v > 0$ and strictly increases with $v$ for $c \geq c_1$.

(c) The demand-related OC function of the government, a private firm, and the intermediate variable, $M_g(v), M_s(v)$, and $M_d(v)$, are continuously increasing and differentiable with respect to $v$ for $v \geq 0$; the road construction cost function $I(c)$ is continuously increasing and differentiable with respect to $c$ for $c \geq c_1$.

(d) For any given $c \geq c_1$, the revenue collected from road users is given by $R(v,c) = v \cdot p(v,c)$, which is a strictly concave function of $v$ for $v \geq 0$, i.e., $\partial R^2 / \partial v^2 < 0$.

It is assumed that the government knows the OC of a private firm, but is not clear of its profit; so, the government will select a private firm through a determined intermediate variable $m_d$. With the given intermediate variable $m_d$, the problem in the following cases can be analyzed. If private firms take part in the bidding process and obtain the right to operate the PTR with $m_d$, the selection procedure will terminate. If the government sets the value of $m_d$ too high, there will be no private firms participating in the bidding. The government has to take the right approaches to avoid

this situation. In order to offer proposals for the government to make a proper decision, two solution approaches are proposed: set the right value of the intermediate variable; provide guarantee for the private firms to ensure their profit. If not selecting a private firm to operate the PTR, the government will operate the road by itself.

We first consider the profit of both roles. Let $P_g(v,c)$, $P_s(v,c)$, and $P_{sc}(v,c)$ be the governmental profit, private firm profit without governmental guarantee, and private profit with governmental guarantee, respectively.

$$\begin{aligned} P_g(v,c) &= R(v,c) - J_g(v,c) = v \cdot B(v) - v\beta t(v,c) \\ &\quad - \left[I(c) + m_g v\right], \\ P_s(v,c) &= R(v,c) - J_s(v,c) = v \cdot B(v) - v\beta t(v,c) \\ &\quad - \left[I(c) + m_s v\right], \end{aligned} \tag{4}$$

$$P_{sc}(v,c,n) = R(v,c) - J_s(v,c) + n = P_s(v,c) + n, \tag{5}$$

where $R(v,c)$ is the total revenue obtained from the road users; the last term of equation (5) is the government guarantee that the government will offer subsidy for the private firm operating the PTR. Equation (5) shows that private firm profit includes two parts: one is profit from the project $P_s(v,c)$, and the other is the government subsidy $n$. Then, the revenue is

$$R(v,c) = v \cdot p(v,c) = v \cdot B(v) - v\beta t(v,c). \tag{6}$$

Next, we define the function of social welfare. When the government operates the road, the social welfare function is formulated as

$$W_g(v,c) = U_s + P_g(v,c) = \int_0^v B(w)\mathrm{d}w - \beta v t(v,c) - J_g(v,c), \tag{7}$$

where $U_s$ is the total social surplus.

With equation (1), the total social surplus is defined by

$$U_s = \int_0^v B(w)\mathrm{d}w - vB(v) = \int_0^v B(w)\mathrm{d}w - v[p(v,c) + \beta t(v,c)]. \tag{8}$$

When the private firm operates the toll road without the government guarantee, the social welfare is given by

$$W_s(v,c) = U_s + P_s(v,c) = \int_0^v B(w)\mathrm{d}w - \beta v t(v,c) - J_s(v,c). \tag{9}$$

When the private firm operates the toll road with the government guarantee, the social welfare can be computed as

$$\begin{aligned} W_{sc}(v,c) &= U_s + \delta \cdot P_{sc}(v,c) - n = \int_0^v B(w)\mathrm{d}w \\ &\quad - (1-\delta)vB(v) - \beta v t(v,c) - \delta J_s(v,c) - (1-\delta)n, \end{aligned} \tag{10}$$

where $\delta$ is a weighting factor, $0 < \delta < 1$. The first term on the right side of equations (7)–(10) is the total social surplus of road users; the second term of equation (7) is the government profit; the second term of equation (9) is the private profit; the second term of equation (10) is the weighted provider's surplus [21]; and the last term of equation (10) is the government guarantee.

*Assumption 2*

(a) $t(v,c)$ is homogeneous of degree zero in $v$ and $c$, i.e., $t(\alpha v, \alpha c) = t(v,c)$ for any $\alpha > 0$

(b) Constant return to scale in the construction and capacity-related cost, namely, $I(c) = kc$, where $k$ denotes the constant cost per unit capacity

(c) $vB(v)$ is a concave function of $v$ for any $v \geq 0$

Part (a) of Assumption 2 shows that vehicle travel time on the road is only related with the volume-capacity ratio, $r = v/c$. For example, the BPR (Bureau of Public Roads) travel time function satisfies this assumption. In equation (1), the toll $p$ can be uniquely determined by the travel demand $v$ for a given road capacity $c$. In other words, social welfare and profit, created by both the government and a private firm, are determined by the travel demand $v$, road capacity $c$, and the government subsidy $n$.

## 4. The First Selection of the PTR Operator by the Government

The government has the right to set an intermediate variable $m_d$ as the threshold value of the public toll road for private firms. The government will not select $m_d > m_g$; otherwise, it will operate the road by itself. So, when the government aims to choose a private firm to build/rebuild and operate the road, obviously, it will set the intermediate variable $m_d \leq m_g$.

*4.1. Under the Condition of $m_s > m_g$ with the Government Operation.* The government and private firm know the information of OC for each other. When the unit demand-related OC of the private firm is greater than the government's, namely, $m_s > m_g$, the government will operate the PTR by itself to maximize social welfare. In order to distinguish from the analysis of Section 4.3, the superscript "$s > g$" is used to represent the case $m_s > m_g$. Then, if $m_s > m_g$, the unconstraint social welfare maximization problem is formulated as

$$\max_{(v,c)} W_g^{s>g}(v,c) = \int_0^v B(w)\mathrm{d}w - \beta v t(v,c) - J_g^{s>g}(v,c), \tag{11}$$

where the first-order optimal conditions of equation (11) can be found in Tan et al. [2].

There are many highways operated by the government in China. The government needs the revenue collected from the road users at least to cover the road investment including operation and construction cost. So, when analyzing the operating PTR, it is necessary to analyze the social welfare maximization under the condition that the total investment including but not limited to operation costs is equal to the total revenue $R(v,c)$, namely, the road operator obtains

positive profit or zero profit from the project. Then, if the government operates the road on its own, the constrained problem is defined as

$$\max_{(v,c)} W_g^{s>g}(v,c) = \int_0^v B(w)\mathrm{d}w - \beta v t(v,c) - J_g^{s>g}(v,c), \tag{12}$$

subject to

$$P_g^{s>g}(v,c) = R_g^{s>g}(v,c) - J_g^{s>g}(v,c) \geq 0, \tag{13}$$

$$\begin{aligned} v &\geq 0, \\ c &\geq c_1, \end{aligned} \tag{14}$$

where equation (13) implies nonnegative profit; $P_g^{s>g}(v,c)$ is the profit of the government; and $R_g^{s>g}(v,c)$ is the governmental revenue.

*4.2. Under the Condition of $m_s \leq m_g$ with the Private Firm Operation.* In practice, the OC of a private firm, with stronger execution and more flexibility, is usually lower than that of the government. If the government makes a correct decision on $m_d$, the private firm will be willing to operate the PTR with proper profit. In this section, we first propose the condition that the private firm operates the road under the condition of $m_s \leq m_g$; then, based on the unconstrained and constrained maximization private profit models, the definitions of the first-best and second-best problems for the private firm are provided, respectively. For simplicity, the subscript "$s$" was used to denote $m_s \leq m_g$.

First, the following proposition provides the condition for the private firm to operate the road. We assume the acceptable profit to be $P_a$.

**Proposition 1.** *If the sum of the unit demand-related OC and unit benefit of the private firm is less than $m_d$, namely, $m_s + P_a/v \leq m_d$, and the government sets $m_d$ in the interval $[m_s + P_a/v, m_g)$, then the contract between the government and a private firm will be signed.*

*Proof.* If allowed to operate the PTR, the private firm will aim to make maximal profit from the project; and the government will certainly not let the private firm get so much profit. In other words, the private profit will be less than the difference between the intermediate OC and the OC of the private firm. Then, we have

$$\begin{aligned} v \cdot p(v,c) - J_s(v,c) &\geq P_a, \\ v \cdot p(v,c) - J_s(v,c) &\leq J_d(v,c) - J_s(v,c) \end{aligned} \tag{15}$$

such that

$$J_d(v,c) \geq P_a + J_s(v,c). \tag{16}$$

Since $I(c)$ is the perfect information,

$$M_d(v) \geq P_a + M_s(v). \tag{17}$$

The proof is completed.

Although Proposition 1 includes the OC of the private firm and the intermediate OC, it does not mean that the government knows the amount of the private profit. In reality, only when the government sets the intermediate variable $m_d$, the private firm will decide whether or not to operate the road according to its OC and profit. However, the government usually evaluates the trade secret $P_a$ of private firms to determine $m_d$.

Getting the road franchising, the private firm will maximize its profit by selecting the road capacity $c$ and toll $p(c)$. In the extreme case of the private profit maximization without any constraints, the unconstraint optimal problem can be formulated as

$$\max_{(v,c)} P_s(v,c) = vp(v,c) - J_s(v,c). \tag{18}$$

Then, we investigate the profit maximization under the constraint that the private profit is not larger than the difference between the intermediate OC and the private firm OC. The constrained optimization problem can be formulated as

$$\max_{(v,c)} P_s(v,c) = vp(v,c) - J_s(c), \tag{19}$$

subject to

$$vp(v,c) - J_s(v,c) \geq P_a, \tag{20}$$

$$vp(v,c) - J_s(v,c) \leq J_d(v,c) - J_s(v,c), \tag{21}$$

$$\begin{aligned} m_s < m_d &\leq m_g, \\ v &\geq 0, \\ c &\geq c_1, \\ P_a &\geq 0, \end{aligned} \tag{22}$$

where $P_a \geq 0$ is the minimum profit of the private firm. Even if $m_d > m_s$, the government is still willing to let a private firm operate the road, which meets constraint (20). Alternatively, only when the private profit is larger than the profit margin $P_a$, the private firm will be willing to operate the road. Constraint (21) means that the private firm profit will not be larger than the difference between the intermediate OC and the private firm OC, and the government will permit the private firm to operate the PTR.

Based on the aforementioned problems, the definition of the first-best and second-best solutions of the private firm is introduced. □

*Definition 1.* (first-best and second-best profit solutions of the private firm). Let $(v_s, c_s)$ be an optimal solution to optimal problems (19)–(22); it will be the first-best solution if it maximizes problem (18); otherwise, it will be the second-best solution.

Definition 1 is based on the profit maximization that could classify the private profit problems. It is different from the definition of the first-best and second-best contracts based on the social welfare [22]. In fact, the private firm will easily reach the second-best solution with constraints (20)

and (21). By comparing first-best problem (18) and second-best problems (19)–(22), the following observation can be obtained based on Definition 1.

*Observation 1.* The first-best solution of the private firm exists if there only exists a solution $(v_s^*, c_s^*)$, which solves the first-best problem and satisfies

$$v \cdot p(v_s^*, c_s^*) - J_s(v_s^*, c_s^*) \geq P_a, \tag{23}$$

$$v_s^* \cdot p(v_s^*, c) - J_d(v_s^*, c_s^*) \leq 0. \tag{24}$$

The above observation is intuitive. If the total revenue generated by the first-best profit solution can cover the OC and create an acceptable profit threshold for the private firm, then the first-best profit solution can be archived.

From Definition 1 and Observation 1, it is clear that if the government permits the private firm to determine the optimal profit by itself, the private firm will choose either the first-best or second-best solution, but not both. The first-best solution exists only if conditions (23) and (24) are satisfied; otherwise, the private firm will obtain the second-best solution. Therefore, to solve private profit problems (19)–(22), it is necessary to solve problem (18) and check whether its solution satisfies constraints (23) and (24). If these conditions hold, it is unnecessary to consider problems (19)–(22).

The Kuhn–Tucker conditions can be applied to obtain the first-order conditions (Appendix A). Note that let $P_s^{sb}$ be the optimal value of the objective function of problems (19)–(22), and it distinguishes from the value of $\overline{P}_s$ that is a monopoly optimal solution of problem (18).The relationship between $P_s^{\mathrm{sb}}$ and $\overline{P}_s$ is that $P_s^{\mathrm{sb}} \leq \overline{P}_s$, and $P_s^{\mathrm{sb}}$ is on the left side of $\overline{P}_s$ in Figure 2. The following observation about $P_s^{sb}$ can be obtained through analyzing the constraints.

*Observation 2*

(a) If constraints (20) and (21) are effective, the acceptable private firm profit is equal to the maximal profit $J_d - J_s$, namely, $P_s^{\mathrm{sb}} = P_a = J_d - J_s = M_d - M_s$

(b) If constraint (20) is effective but constraint (21) is inactive, the private firm profit is $P_s^{\mathrm{sb}} = P_a$

(c) If constraint (21) is effective but constraint (20) is inactive, the private firm profit is $P_s^{\mathrm{sb}} = J_d(v, c) - J_s(v, c)$

(d) If both constraints (20) and (21) are ineffective and the second-best problem becomes the first-best problem, then the private firm profit is $P_s^{\mathrm{sb}} = \overline{P}_s$

From the aforementioned observations, we investigate the case of the first-best and second-best private profit problem, which is presented on the left side of the monopoly optimum in Figure 2. However, when will the monopoly optimum case happen? As shown in Figure 2, we can see that, responding to the same private profit, social welfare on the right side of the monopoly optimum is higher than that on the left side. Additionally, the government can make some rules in the contract to make the private firm set the social welfare maximization as the goal [23]. So, it is necessary to analyze the following two problems with private firm management: one is the unconstrained social welfare maximization problem; the other is the constrained social welfare maximization problem.

Base on equation (7), the unconstrained social welfare maximization problem can be formulated by

$$\max_{(v,c)\in\Omega} W_s(v, c) = \int_0^v B(w)\mathrm{d}w - \beta v t(v, c) - J_s(v, c). \tag{25}$$

The constrained social problem is given by

$$\max_{(v,c)\in\Omega} W_s(v, c) = \int_0^v B(w)\mathrm{d}w - \beta v t(v, c) - J_s(v, c), \tag{26}$$

subject to

$$vp(v, c) - J_s(v, c) \geq P_a, \tag{27}$$

$$vp(v, c) - J_s(v, c) \leq J_d(v, c) - J_s(v, c), \tag{28}$$

$$\begin{aligned} &m_s < m_d \leq m_g, \\ &v \geq 0, \\ &c \geq c_1, \\ &P_a \geq 0, \end{aligned} \tag{29}$$

where $\Omega = \{(v, c): v \geq 0, c \geq 0\}$.The constrained conditions are the same as those of the second-best problem of the private firm, i.e., (20)–(22), which guarantee the private firm to operate the road.

Since this kind of constrained social welfare problem has more constraints than the traditional Pareto optimal problem, the win-win optimal problem is defined as follows.

*Definition 2.* (win-win optimal problem). A pair $(v_s^{\mathrm{ww}}, c_s^{\mathrm{ww}}) \in \Omega$ is called the win-win solution if there are no other feasible solutions $(v, c) \in \Omega$ such that $W(v, c) \geq W(v_s^{\mathrm{ww}}, c_s^{\mathrm{ww}})$ and $P(v, c) \geq P(v_s^{\mathrm{ww}}, c_s^{\mathrm{ww}})$ with at least one strict inequality.

The definition of the win-win optimal problem is a meaningful and important concept which meets all the conditions of the Pareto problem, but with more constraints, and its solution is a part of the Pareto optimal solution.

In general, more than one constrained condition in problems (27)–(29) will be effective. In response to the case on the right side of monopoly optimum in Figure 2, similar to Observation 2, the following observation is provided through analyzing the constrained conditions (Appendix B).

*Observation 3*

(a) If constraints (27) and (28) are effective, the acceptable private firm profit is the same as its maximal profit, which is equal to $J_d - J_s$, namely, $P_a = M_d - M_s$

(b) If constraint (27) is effective and constraint (28) is inactive, the private firm profit is equal to $P_a$

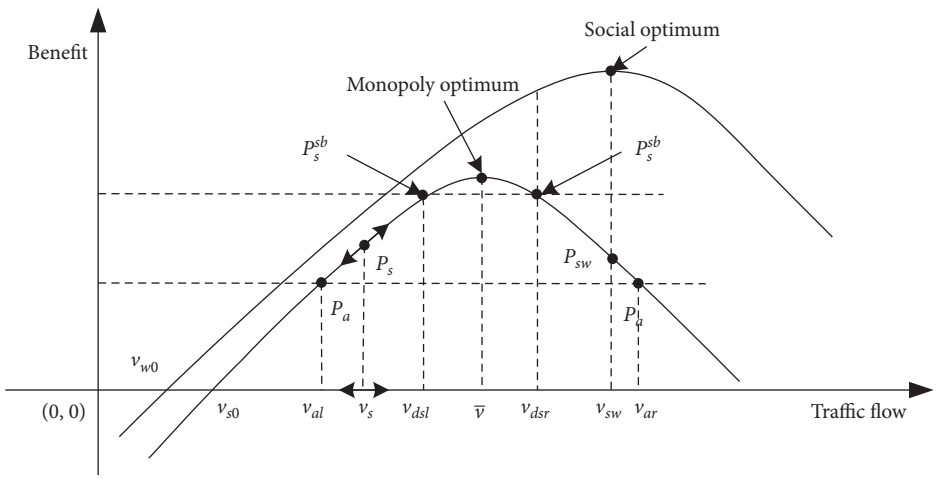



Figure 2: A geometric illustration of $P_a, \overline{P}_s, P_s$, and $P_s^{\mathrm{sb}}$.

(c) If constraint (28) is effective and constraint (27) is inactive, the private firm profit is equal to $M_d - M_s$

Furthermore, let $\widetilde{v}_s$ and $\overline{v}_s$ be the social welfare optimality solution of problem (18) and the monopoly optimal solution of problem (25), respectively. The following proposition is obtained in terms of the above second-best solution of the private problem and win-win optimal problem (see more details in Appendix C).

**Proposition 2.** *It holds that* $v_s^{\mathrm{sb}} < \overline{v}_s \le v_s^{\mathrm{ww}} \le \widetilde{v}_s$ *for the win-win optimal solution* $(v_s^{\mathrm{ww}}, c_s^{\mathrm{ww}})$ *and the second-best solution of the private firm* $(v_s^{\mathrm{sb}}, c_s^{\mathrm{sb}})$ *when the private firm operates the road.*

Proposition 2 means that travel demand of the win-win solution should not be strictly less than the level of the profit maximization and less than the level of the social welfare maximization. Travel demand of the second-best optimal solution is smaller than that of any other optimal solutions. Figure 2 implies that the government must set the regulation to gain the win-win optimal solution, namely, make $M_d - M_s$ on the right side of the monopoly optimal solution. Otherwise, the private firm will optimize the travel demand $v_s^{\mathrm{sb}}$.

Based on the second-best solution and the Win-Win optimal solution, Proposition 3 can be derived:

**Proposition 3.** *For the second-best solution of the private firm, the bigger* $M_d - M_s$ *is, the more possible it is that the government and a private firm have the same objective; For the win-win solution, the smaller* $M_d - M_s$ *is, the more possible it is that the government and private firm have the same objective.*

Refer to the proof in Appendix D. Proposition 3 tells us why the private firm wants to know the value of $M_d$ and the government also wants to evaluate the demand-related OC of the private firm to set a better $M_d$.The result is consistent with Figure 2, which implies the objective of the government and private firm will be realized at the same time on the left side of monopoly optimality, namely, the private firm obtains the maximum solution, while the government achieves the maximal social welfare. If the difference $M_d - M_s$ is bigger, the private firm will obtain more profit. On the right side of monopoly optimality in Figure 2, the government and private firm will have different objectives, namely, when the private firm chooses the profit maximization, the government will reach the minimum social welfare. If the difference $M_d - M_s$ is smaller, the private firm will obtain more profit, and the government can also gain more social welfare. However, welfare on the right side that belongs to the win-win optimal solution is bigger than that on the left side.

Let $P_{sw}$ be the private profit corresponding to the social optimal welfare. Based on Proposition 3, an extreme case is analyzed: if private profit satisfies $P_a \le P_{sw}$, when the intermediate variable set by the government meets $(M_d - M_s) \ge P_{sw}$, namely, the variable relationship meets $P_a \le P_{sw} \le (M_d - M_s)$; under the win-win optimal regulation in the contract, not only that the government may gain the optimal social welfare but also the private firm meets its profit. So, the ideal condition is $P_a \le P_{sw} = (M_d - M_s)$, which is also the best case for the win-win optimal solution.

Some private firms usually earn huge profit from the project, which happened in China such as Beijing Airport Expressway and Guangzhou-Shenzhen Highway. Although these roads were built with the largest road capacity (allowed by the design code), there is oversaturated traffic flow every day. In contrast with the traditional contract, we can regard constraint (21) as a regulation to allow the private firm to make the decision. Then,

$$vp(v,c) - J_s(v,c) > J_d(v,c) - J_s(v,c). \tag{30}$$

The above condition implies that if the private firm profit is bigger than the difference between the intermediate OC and private firm OC, the government will operate the road by itself. Condition (30) can be simplified by

$$vp(v,c) - [I(c) + m_d v] > 0. \tag{31}$$

If there is heavy traffic pressure, then the private firm will get profit from the project quickly. Meanwhile, condition (31) implies that if the road is operated, all terms of (31) are the symmetric information for the government and a private

firm, namely, the road capacity has been determined when the road is operating; the government can get the information of road capacity, toll, and travel demand from real data. Then, the government knows the profit of the private firm based on the intermediate variable even if it knows nothing about the private firm operation. When condition (31) is satisfied, the government can operate the road by itself, namely, condition (31) points out that the difference between profit $vp(v,c)$ and intermediate variable $m_d$ determines the road operation right of the private firm.

In the last part of this section, the private firm zero profit will be analyzed. If private firms get the road franchise, they will have the right to determine locations of on-ramps and off-ramps that have an impact on the land development, and so on. With the above description, the private firm will be willing to operate the toll road projects with zero profit or negative profit to earn more profit from other aspects. So, it is necessary to analyze the zero or negative profit case. However, in the light of knowing nothing about how much loss the private firm can bear, the zero profit is analyzed as follows.

**Proposition 4.** *Under the condition of zero profit, the unit demand-related OC of the private firm $m_s$ can approach the biggest value, namely, $m_s$ is equal to $m_g$.*

*Proof.* From constraints (20) and (21),

$$J_d(v,c) - J_s(v,c) \geq P_a. \tag{32}$$

If the private profit is equal to zero, namely, $P_a = 0$, then

$$M_d(v,c) - M_s(v,c) = 0. \tag{33}$$

If $M_d(v,c) = M_g(v,c)$, then

$$M_s(v,c) = M_g(v,c). \tag{34}$$

The proof is completed. □

*4.3. Under the Condition of $m_s \leq m_g$ with the Government Operation.* Different from Section 4.1, it is assumed that the government knows nothing about the private profit, which may make the government set an improper decision on the intermediate variable $m_d$. The private firm will not operate the road if it cannot gain the satisfied profit from the project. The following two cases are related with an improper decision of $m_d$.

In contrast with Section 4.2, we provide the condition that the private firm cannot gain enough profit. Using constraints (20) through (22),

$$J_d(v,c) - J_s(v,c) \geq P_a, \tag{35}$$

which implies that only when the difference between the intermediate OC and the private OC is larger than the private firm profit, the private firm will be willing to operate the toll road.

On the contrary,

$$J_d(v,c) - J_s(v,c) < P_a, \tag{36}$$

which implies that the government will be reluctant to operate the road by itself.

If $m_d$ given by the government is less than the private OC, namely, $m_d \leq m_s < m_g$, the private firm will not operate the road due to the negative profit.

In Section 4.1, we analyze the case that the government is willing to operate the road. In this section, there will be no contract between the government and a private firm. The superscript "$s<g$" is used to represent $m_s \leq m_g$. Then, the problem can be formulated as

$$\max_{(v,c)} W_g^{s<g}(v,c) = \int_0^v B(w)\mathrm{d}w - \beta v t(v,c) - J_g^{s<g}(v,c). \tag{37}$$

In the above analysis, two kinds of government operation scenarios were considered: the first one was analyzed in Section 4.1 given $m_s > m_g$; the second one was analyzed in this section under the condition of $m_s \leq m_g$. If the government does not make an improper decision on $m_d$, we assume that the government will choose to maximize the social welfare without loss, so it is not necessary to analyze this case. Edelson [24] showed that a private firm may charge a lower toll than the government for a given and fixed capacity. Let $c_d^*$ be the given road capacity, and all analyses of this section are based on this capacity. Then, we compare the toll charged by the government and the private firm based on the assumption that the government makes an improper decision on the variable $m_d$ under the condition of $m_s \leq m_g$.Note that the toll charged by the government is the solution at the zero profit, namely, the governmental revenue just covers its capital cost.

Then, we define the average social cost (ASC) of the government and private firm, respectively:

$$\mathrm{ASC}_g^{s<g} = \frac{\beta v t(v,c_d^*) + J_g^{s<g}(v,c_d^*)}{v} = \beta t(v,c_d^*) + \frac{J_g^{s<g}(v,c_d^*)}{v}, \tag{38}$$

$$\mathrm{ASC}_s = \frac{\beta v t(v,c_d^*) + J_s(v,c_d^*)}{v} = \beta t(v,c_d^*) + \frac{J_s(v,c_d^*)}{v}. \tag{39}$$

The first term of equations (38) and (39) on the right side is the average travel time in monetary units; the second term is the cost of operation per trip. Intuitively, with Assumptions 1 and 2 and the condition of $m_s \leq m_g$, the average social cost of the government will be bigger than the private firm OC for given road capacity and travel demand.

If considering the average social profit in the cost of the private firm, equation (39) can be rewritten as

$$\mathrm{ASC}_{sp} = \beta t(v,c_d^*) + \frac{J_s(v,c_d^*)}{v} + \frac{P_s}{v}. \tag{40}$$

Then, combining equations (38) and (40) for the given travel demand and road capacity $c_s^*$,

$$\mathrm{ASC}_g^{s<g} - \mathrm{ASC}_{sp} = \frac{M_g^{s<g}(v) - [M_s(v) + P_s]}{v}, \tag{41}$$

where $P_s$ is in the interval $[P_a, M_d - M_s]$.

From equation (41), though $P_s = M_d - M_s$, the average social cost of the government will be bigger than that of

private firm operation under $M_d \le M_g$. We know that the government may charge a higher toll than the private firm under the condition that the government operates the road without loss.

Let $\overline{p}_s$, $\widetilde{p}_s$, and $\widetilde{p}_g^{s<g}$ be the monopoly price of the private firm, the price of social optimal welfare of the private firm, and the price of social optimal welfare of the government, respectively. Then, the following proposition can be obtained (Appendix E).

**Proposition 5.** *With a given road capacity $c_d^*$, the toll $\overline{v}_s, \widetilde{v}_s, \widetilde{v}_g^{s<g}$ has the following relationship: $\widetilde{p}_s \le \widetilde{p}_g^{s<g} < \overline{p}_s$.*

Proposition 5 shows that if the social welfare maximum is reached, the toll charged by a private firm is less than the government, which is derived from the improper decision of the government on the intermediate variable. In order to avoid the circumstance that the government operates the road with higher charge than the private firm, the best option for the government is investigated to get a satisfactory result.

## 5. The Optimal Intermediate Variable

The case discussed in Section 4.3 occurs because of the improper intermediate variable $m_d$ given by the government. With the governmental operation, road users will pay higher toll that is analyzed in Section 4.3 for $m_s \le m_g$. So, in order to find a proper private firm to operate the road and not let the road users undertake higher toll, the government will make a change on the selection tactics. In the following, two modes will be studied to avoid the improper decision on $m_d$: one is without guarantee, and the other is considering guarantee provided by the government.

*5.1. The Selection Mode without the Government Guarantee.* At first, we investigate the mode that the government will not provide guarantee for the private firm. In Section 4, we discussed the case that the private firm and the government were willing to operate the road; we also discussed the case that the government was reluctant to operate the road under the improper decision. Under the condition of knowing nothing about private profit $P_a$, the government will try its best to make a correct decision. In Section 4, we pointed out that the case of $m_d > m_g$ was impossible in that the government would operate the road by itself. So, in this section, we only investigate the value of $m_d$ which can result in a satisfactory result for the government under $m_d \le m_g$. We firstly provide the satisfactory result and unsatisfactory result under $m_d \le m_g$ and $P_a \ge 0$.

The definition of satisfactory results:

(a)If $m_d = m_g$ and $m_s > m_d$, the government will operate the road by itself

(b) If $m_s \le m_d = m_g$ and $m_s + P_a/v > m_d$, the government will operate the road by itself

(c) If $m_s \le m_d \le m_g$ and $m_s + P_a/v > m_d$, a private firm will operate the road

(d) If $m_d < m_s \le m_g$ and $m_s + P_a/v > m_d$, the government will operate the road by itself

The definition of unsatisfactory results:

(e) If $m_s \le m_d < m_g$ and $m_s + P_a/v > m_g$, the government will be reluctant to operate the road by itself

(f) If $m_d < m_s \le m_g$ and $m_s + P_a/v > m_g$, the government will be reluctant to operate the road by itself

With the definition of satisfactory results and unsatisfactory results of the government, we have the following proposition. We know that there are two constraints: $m_s + P_s/v > m_d$ and $P_s \ge P_a$. Then, we make another assumption that the government must obtain the satisfactory result.

**Proposition 6.** *If the government wants to obtain a satisfied result, the intermediate variable $m_d$ should be equal to the demand-related OC of the government $m_g$, namely, $m_d = m_g$.*

At the beginning of this section, we pointed out that the government is not clear about profit of the private firm. The governmental objective is to obtain the satisfied result with the private firm. If the government determines the value of the intermediate variable so that the private firms, after their independent calculation, believe that they will be able to make necessary profit from running the project, the private firms may decide to take part in the bid. The government does not need to know the OC of the private firm. So, when the government sets $m_d = m_g$, the government can easily choose a proper road operator under the condition of satisfactory results.

When $m_d = m_g$, the problem can be transformed to the following:

$$\max_{(v,c)} W_s(v,c) = \int_0^v B(w)\mathrm{d}w - \beta v t(v,c) - J_s(v,c), \tag{42}$$

subject to

$$\begin{aligned} v \cdot p(v,c) - J_s(v,c) &\ge P_a, \\ v \cdot p(v,c) - J_s(v,c) &\le J_g(v,c) - J_s(v,c), \\ m_s &\le m_g, \\ v &\ge 0, \\ c &\ge c_1, \\ P_a &\ge 0. \end{aligned} \tag{43}$$

Then, the government can make the decision to obtain satisfactory results through comparing formulas (39)–(42) with the social optimal welfare problem in Section 4.3. Private firm can obtain its profit through setting the toll and road capacity based on the solution of problems (39)–(42).

*5.2. The Selection Mode with the Government Guarantee.* In this section, the government guarantee will be considered to make the government reach a satisfactory result. In order to obtain satisfactory results, like in Section 4.3, the government can make proper decisions through providing subsidy for the private firm.

We firstly give the condition that the government may provide the guarantee for a private firm:

$$v \cdot p(v,c) - J_s(v,c) = P_s(v,c) < P_a. \tag{44}$$

From equation (44), we know that the private firm will not operate the road in that it will not get satisfactory profit from the project or get negative profit. If the government wants the private firm to operate the road, it must provide guarantee, namely, additional profit for the private firm, which makes sure the private profit is not less than a threshold value $P_a$ [25]. Then, we have

$$P_{sc}(v,c) = P_s(v,c) + n = v \cdot p(v,c) - J_s(v,c) + n \geq P_a, \tag{45}$$

where the first term on the left side of the inequality is the private profit and $n$ is the government subsidy. Therefore, private profit is composed of a private firm profit and the government subsidy.

Even if providing subsidy for the private firm, the government will be reluctant that the total amount of profits, the OC of the private firm, and subsidies are bigger than the governmental OC. According to Proposition 6, we point out that $m_d = m_g$ is a better option for the government. Then, the second constraint becomes

$$v \cdot p(v,c) - J_s(v,c) + n \leq J_g(v,c) - J_s(v,c). \tag{46}$$

Equation (46) can be simply written as $v \cdot p(v,c) + n \leq J_g(v,c)$. According to equation (46), the government can control the biggest subsidy which is $n = J_g(v,c) - v \cdot p(v,c)$. With this constraint, even if nothing is known about private profit, the government can determine the minimal subsidy. An extreme case is that the minimal private profit is equal to the maximum of guarantee, namely, $P_a = J_g(c) - J_s(c)$. From Proposition 6, as long as the subsidies do not exceed the maximum subsidies, the government will regard it as a satisfactory result.

In the bidding process, many private companies have put forward different subsidy requirements, which can further determine subsidies. If the government provides subsidies for the private firm, it must require a private firm to select the win-win optimal solution. With constraints (45) and (46) and objective function (10), the optimal problem with the government guarantee can be formulated by

$$\begin{aligned} \max_{(v,c)} W_{sc}(v,c,n) = {} & \int_0^v B(w)\mathrm{d}w - vB(v) \\ & + \delta\left[v \cdot p(v,c) - J_s(c)\right] - (1-\delta)n, \end{aligned} \tag{47}$$

subject to equations (45) and (46), and

$$\begin{aligned} & m_s \leq m_g, \\ & v \geq 0, \\ & c > c_1, \\ & 0 < \delta < 1, \\ & P_a \geq 0. \end{aligned} \tag{48}$$

Based on the above problem, we have the following proposition (Appendix F).

**Proposition 7.** *If the government provides guarantee for the private firm, the minimum private profit is equal to the maximum subsidies.*

*The above proposition shows that if the government wants to provide guarantee for the private firm, the private firm will choose the maximum subsides. And the government will not permit the total amount of toll sum charged by the private firm and guarantee larger than the governmental OC on the PTR.*

## 6. Conclusions

The operation of the PTR becomes a significant issue of the government. This study offers valuable insights into the selection mechanism of PTR operators by comparing the OC of the government and the private firms. Previous studies conducted few research on the PTR operation. Considering a combination of the toll charge, road capacity, and guarantee, this paper analyses the operation problem through an intermediate variable. If the government sets the right intermediate variable in the bidding document, private firms will be willing to operate the road actively with the profit maximization. The first-best and second-best optimal solutions for the private firm and the win-win solution of the private operation were investigated. The average social cost and the toll charged by operators were compared; in some cases, the toll charged by the government is higher than private firms. The case that toll collected by the private firm from the road users is lower than that of the government will happen. Analytical results show that the social optimal pricing set by the private firm is less than that of the government, and then two approaches are proposed to solve the problem given a biased decision of the intermediate variable. The first approach sets the intermediate variable equal to the governmental OC without guarantee; the second one is that the government provides profit guarantee for the private firm.

To optimize the succeeding operation of different kinds of facilities developed in the BOT mode after the concession period expires, future research can be carried out under information asymmetry to investigate whether and how to charge the toll and to set regulations to be followed by the private firms when operating these facilities. Even if the government is in a position to build/rebuild and operate its transportation facilities, it is still necessary for it to seriously consider how to operate them in a sustainable way to maximize social welfare.

## Appendix

## A. Proof of Observation 2

*Proof.* It is assumed that $(v_s^{sb}, c_s^{sb})$ is any second-best optimal solution of private problem (19). From constrained

optimization problems (19)–(22), the following Lagrange-like function can be defined:

$$L = P_s(v,c) + \lambda_{sb1}^{s<g}\left[P_s(v,c) - P_a\right] + \lambda_{sb2}^{s<g}\left[-v \cdot p(v,c) + J_d(v)\right], \tag{A.1}$$

where $\lambda_{sb1}^{s<g} \geq 0$ and $\lambda_{sb2}^{s<g} \geq 0$ are the Lagrangian multipliers which are associated with constraints (20) and (21), respectively. Using the Kuhn–Tucker condition, the first-order conditions can be determined:

$$\begin{aligned}
&\frac{\partial L}{\partial v}\Big|_{(v_s^{sb},c_s^{sb})} = \left[1 + \lambda_{sb1}^{s<g} - \lambda_{sb2}^{s<g}\right] \cdot \left[p\left(v_s^{sb},c_s^{sb}\right) + v_s^{sb} \cdot p_q'\left(v_s^{sb},c_s^{sb}\right) - m_s\right] + \lambda_{sb2}^{s<g}\left[m_d - m_s\right] = 0,\\
&\frac{\partial L}{\partial c}\Big|_{(v_s^{sb},c_s^{sb})} = \left[1 + \lambda_{sb1}^{s<g} - \lambda_{sb2}^{s<g}\right] \cdot \left[v_s^{sb} \cdot p_y' - k\right] + \lambda_{sb2}^{s<g} \cdot k = 0,\\
&\frac{\partial L}{\partial \lambda_{sb1}^{s<g}}\Big|_{(v_s^{sb},c_s^{sb})} = P_s\left(v_s^{sb},c_s^{sb}\right) - P_a \geq 0,\\
&\frac{\partial L}{\partial \lambda_{sb2}^{s<g}}\Big|_{(v_s^{sb},c_s^{sb})} = -\overline{v}_{sb} \cdot p\left(v_s^{sb},c_s^{sb}\right) + J_d\left(v_s^{sb},c_s^{sb}\right) \geq 0,
\end{aligned} \tag{A.2}$$

$$\lambda_{sb2}^{s<g} \geq 0, \tag{A.3}$$

$$\lambda_{sb1}^{s<g} \geq 0, \tag{A.4}$$

$$\lambda_{sb2}^{s<g} \cdot \frac{\partial L}{\partial \lambda_{sb2}^{s<g}}\Big|_{(v_s^{sb},c_s^{sb})} = \lambda_{sb2}^{s<g} \cdot \left[-v_s^{sb} p\left(v_s^{sb},c_s^{sb}\right) + J_d\left(v_s^{sb},c_s^{sb}\right)\right] = 0, \tag{A.5}$$

$$\lambda_{sb1}^{s<g} \cdot \frac{\partial L}{\partial \lambda_{sb1}^{s<g}}\Big|_{(v_s^{sb},c_s^{sb})} = \lambda_{sb1}^{s<g} \cdot \left[P_s\left(v_s^{sb},c_s^{sb}\right) - P_a\right] = 0. \tag{A.6}$$

If constrained conditions (20) and (21) are effective and the Lagrangian multipliers $\lambda_{sb1}^{s<g} > 0$ and $\lambda_{sb2}^{s<g} > 0$, then

$$P_s(q,y) - P_a = 0, \tag{A.7}$$

$$P_s(v,c) = vp(v,c) - J_s(v,c) = J_d(v,c) - J_s(v,c). \tag{A.8}$$

From equations (A.7) and (A.8),

$$P_a = P_s^{sb} = J_d(v,c) - J_s(v,c) = M_d - M_s. \tag{A.9}$$

Note that if profit $M_d(v) - M_s(v)$ is on the right side of the monopoly optimum (see Figure 2), the private firm will choose travel demand $\overline{v}$ to obtain the monopoly optimum solution. So, the private firm will choose the optimal profit point, i.e., $M_d(v) - M_s(v)$, on the left side of the monopoly optimum, the corresponding travel demand of which is $v_{dsl}$. Thus, part (a) of Observation 2 is proved.

If constraint (20) is effective and constraint (21) is inactive, the Lagrangian multipliers $\lambda_{sb1}^{s<g} > 0$ and $\lambda_{sb2}^{s<g} = 0$. If $\lambda_{sb2}^{s<g} = 0$, constraint (21) can be ignored. From equations (A.3) and (A.5), we can determine that the private firm will choose $P_s(v,c) = P_a$ corresponding to travel demand $q_{dal}$ (Figure 2). Thus, part (b) of Observation 2 is proved.

If constraint (21) is effective and constraint (20) is inactive, the Lagrangian multipliers $\lambda_{sb1}^{s<g} = 0$ and $\lambda_{sb2}^{s<g} > 0$. If $\lambda_{sb1}^{s<g} = 0$, we can ignore constraint (20). Then, from equations (A.4) and (A.6), we can determine that the private firm will choose $P_s(v,c) = M_d - M_s$ corresponding to travel demand $v_{dsl}$ (Figure 2). Thus, part (c) of Observation 2 is proved.

If constraints (20) and (21) are inactive, the Lagrangian multipliers $\lambda_{sb1}^{s<g} = 0$ and $\lambda_{sb2}^{s<g} = 0$. We can ignore all of the constraints. The problem becomes a monopoly optimal problem (or the first-best problem of the private firm), and the private firm can obtain the maximum profit $\overline{P}_s$. Thus, part (d) of Observation 2 is proved. This completes the proof. □

## B. Proof of Observation 3

*Proof.* It is assumed that $(v_s^{ww}, c_s^{ww})$ is any win-win optimal solution of private problem (26). From constrained optimization problems (26)–(29), the following Lagrange-like function can be defined:

$$L = W_s(v,c) + \lambda_{sw1}^{s<g}\left[P_s(v,c) - P_a\right] + \lambda_{sw2}^{s<g}\left[-vp(v,c) + J_d(c)\right], \tag{B.1}$$

where $\lambda_{sw1}^{s<g} \geq 0$ and $\lambda_{sw2}^{s<g} \geq 0$ are Lagrangian multipliers, which are associated with constraints (27) and (28), respectively. Using the Kuhn–Tucker condition, the first-order conditions can be determined by

$$\begin{aligned}
&\frac{\partial L}{\partial v}\Big|_{(v_s^{ww},c_s^{ww})} = \left[1 + \lambda_{sw1}^{s<g} - \lambda_{sw2}^{s<g}\right] \cdot \left[B\left(v_s^{ww}\right) - \beta t\left(v_s^{ww},c_s^{ww}\right) - \beta r t'(r) - m_s\right] + \lambda_{sw1}^{s<g} v B_v'(v) + \lambda_{sw2}^{s<g}\left[m_d - m_s - v B_v'(v)\right] = 0,\\
&\frac{\partial L}{\partial c}\Big|_{(v_s^{ww},c_s^{ww})} = \left[1 + \lambda_{sw1}^{s<g} - \lambda_{sw2}^{s<g}\right] \cdot \left[\beta\left(r_s^{ww}\right)^2 \cdot t'\left(r_s^{ww}\right) - k\right] = 0,\\
&\frac{\partial L}{\partial \lambda_{sw1}^{s<g}}\Big|_{(v_s^{ww},c_s^{ww})} = P_s\left(v_s^{ww},c_s^{ww}\right) - P_a \geq 0,\\
&\frac{\partial L}{\partial \lambda_{sw2}^{s<g}}\Big|_{(v_s^{ww},c_s^{ww})} = -v_s^{ww} \cdot p\left(v_s^{ww},c_s^{ww}\right) + J_d\left(v_s^{ww},c_s^{ww}\right) \geq 0,
\end{aligned} \tag{B.2}$$

$$\lambda_{sw1}^{s<g} \cdot \frac{\partial L}{\partial \lambda_{sw1}^{s<g}}\Big|_{(v_s^{ww},c_s^{ww})} = \lambda_{sw1}^{s<g} \cdot \left[P_s\left(v_s^{ww},c_s^{ww}\right) - P_a\right] = 0, \tag{B.3}$$

$$\lambda_{sw2}^{s<g} \cdot \frac{\partial L}{\partial \lambda_{sw2}^{s<g}}\Big|_{(v_s^{ww},c_s^{ww})} = \lambda_{sw2}^{s<g} \cdot \left[-v_s^{ww} p\left(v_s^{ww},c_s^{ww}\right) + J_d\left(c_s^{ww}\right)\right] = 0, \tag{B.4}$$

$$\lambda_{sw1}^{s<g} \geq 0, \tag{B.5}$$

$$\lambda_{sw2}^{s<g} \geq 0. \tag{B.6}$$

Note that profit $M_d(v) - M_s(v)$ is on the right side of the monopoly optimum (see Figure 2), and the private firm will choose the optimal profit point, $M_d(v) - M_s(v)$, which is close to the monopoly optimum (Figure 2), the corresponding travel demand of which is $v_{dsr}$. Thus, part (a) of Observation 3 is proved.

If constraint (26) is effective and constraint (27) is inactive, the Lagrangian multipliers $\lambda_{sw1}^{s<g} > 0$ and $\lambda_{sw2}^{s<g} = 0$. If $\lambda_{sw2}^{s<g} = 0$, we can ignore constraint (21). From equations (B.3) and (B.5), we can determine that the private firm will choose $P_s(v, c) = P_a$ corresponding to travel demand $v_{ar}$ (Figure 2) for the same reason of part (a).Thus, part (b) of Observation 3 is proved.

If constraint (27) is effective and constraint (26) is inactive, the Lagrangian multipliers $\lambda_{sw1}^{s<g} = 0$ and $\lambda_{sw2}^{s<g} > 0$. If $\lambda_{sw1}^{s<g} = 0$, we can ignore constraint (26). Then, from equations (B.4) and (B.6), we can determine that the private firm will choose $P_s(v, c) = M_d - M_s$ corresponding to travel demand $v_{dsr}$ (Figure 2) for the same reason of part (a). Thus, part (c) of Observation 3 is proved. This completes the proof. □

## C. Proof of Proposition 2

*Proof.* We first take the partial derivatives of social welfare $W_s(v, c)$ and private profit $P_s(v, c)$ with respect to demand $v$ and capacity $c$. Based on social welfare function $W_s(v, c)$ given by (25), we have

$$\frac{\partial W_s(v,c)}{\partial v} = B(v) - \beta t(v,c) - \beta v \frac{\partial t(v,c)}{\partial v} - m_s, \tag{C.1}$$

$$\frac{\partial W_s(v,c)}{\partial c} = -\left(\beta v \frac{\partial t(v,c)}{\partial c} + k\right), \tag{C.2}$$

and according to the private profit function given by (18), we have

$$\frac{\partial P_s(v,c)}{\partial v} = R'(v) = B(v) - \beta t(v,c) + vB'(v) - \beta v \frac{\partial t(v,c)}{\partial v} - m_s, \tag{C.3}$$

$$\frac{\partial P_s(v,c)}{\partial c} = -\left[\beta v \frac{\partial t(v,c)}{\partial c} + k\right]. \tag{C.4}$$

For exposition simplicity, we define private profit $P_s(v)$ and social welfare $W_s(v)$ for given $c_s^*$. Let $\widetilde{v}_s$ be the unique $v$ that could maximize $W_s(v, c)$; then, with equation (C.1),

$$B(\widetilde{v}_s) - \beta \cdot t(\widetilde{v}_s, c_s^*) - \beta \cdot \widetilde{v}_s \cdot \frac{\partial t(\widetilde{v}_s, c_s^*)}{\partial v} - m_s = 0. \tag{C.5}$$

Let $\overline{q}_s$ be unique $q$ that could maximize $P_s(q, y)$; then, with equation (C.3)

$$B(\overline{v}_s, c_s^*) + \overline{v}_s \cdot B'(\overline{v}_s) - \beta \cdot t(\overline{v}_s, c_s^*) - \beta \cdot \overline{v}_s \cdot \frac{\partial t(\overline{v}_s, c_s^*)}{\partial v} - m_s = 0. \tag{C.6}$$

With Assumption 1, there exist unique $\overline{v}_s$ and $\widetilde{v}_s$ satisfying (C.5) and (C.6), respectively. By comparing equations (C.5) and (C.6), we can observe that $\overline{v}_s < \widetilde{v}_s$.

Let $(v_s^{ww}, c_s^{ww})$ be the solution of win-win optimality. Comparing (C.1) and (C.3), we have $\partial W/\partial v > \partial P/\partial v$ for any feasible solution $(v, c)$ with the condition $B'(v) < 0$. If $(v_s^{ww}, c_s^{ww})$ is a win-win solution, $\partial W(v_s^{ww}, c_s^{ww})/\partial v$ and $\partial P(v_s^{ww}, c_s^{ww})/\partial v$ will not be positive or negative at the same time; otherwise, the change of $v$ will increase or decrease private profit and social welfare simultaneously, which contradicts with the fact that $(v_s^{ww}, c_s^{ww})$ is a win-win optimal solution. Thus, we have $\partial W(v_s^{ww}, c_s^{ww})/\partial v \geq 0$ and $\partial P(v_s^{ww}, c_s^{ww})/\partial v \leq 0$. Then, $\overline{v}_s \leq v_s^{ww} \leq \widetilde{v}_s$.

From Definition 1, we know that $P_s(v_s^{sb}, c_s^*) \leq P_s(\overline{v}_s, c_s^*)$ for given road capacity. If $P_s(v, c)$ is the strictly concave function in $v$ and $P_s(v, c)$ is the strictly increasing function in $v$ for $v < \overline{v}_s$, we have that $v_s^{sb} < \overline{v}_s$. This completes the proof. □

## D. Proof of Proposition 3

*Proof.* For constraints (20) and (21), we have $M_d - M_s \geq P_a$, which implies that the objective value of equation (19), $P_s$, is in the interval $[P_a, M_d - M_s]$. Based on Assumption 1, we know that $W_s(v, c)$ and $P_s(v, c)$ are the concave functions of $v$. According to Proposition 2, we know that $\overline{v}_s < \widetilde{v}_s$.

Since $W_s(v, c)$ is an increasing function of $v$ for $v \leq \widetilde{v}_s$ and $P_s(v, c)$ is an increasing function of $v$ for $v \leq \overline{v}_s$, $W_s(v, c)$ and $P_s(v, c)$ are increasing functions of $v$ for $v \leq \overline{v}_s$. Then, under the condition of the second-best problem of the private firm, social welfare will increase with private profit. For the profit approaching maximization profit, $M_d - M_s$, social welfare approaches a maximizing value under the case of the second-best problem of the private firm. So, the larger $M_d - M_s$ is, the larger the social welfare is.

Since $W_s(v, c)$ is an increasing function of $q$ for $v \leq \widetilde{v}_s$ and $P_s(v, c)$ is a decreasing function of $v$ for $v \geq \overline{v}_s$, $W_s(v, c)$ is an increasing function of $v$ for $\overline{v}_s \leq v \leq \widetilde{v}_s$ and $P_s(v, c)$ is an increasing function of $q$ for $v \geq \overline{v}_s$. Then, under the condition of the win-win optimal problem, social welfare will decrease with private profit. And with the profit approaching maximization $M_d - M_s$, social welfare approaches the minimum value under the case of the win-win optimal problem of the private firm. So, the smaller $M_d - M_s$ is, the larger the social welfare is. This completes the proof. □

## E. Proof of Proposition 5

*Proof.* Assume $c_d$ is a given and fixed capacity. Let $\overline{p}_s$, $\widetilde{p}_s$, and $\widetilde{p}_g^{s<g}$ be the monopoly price of the private firm, social optimal price of the private firm, and social optimal price of the government, respectively.

According to equation (1), regard $v$ as the function of toll charge $p$ and road capacity $c$. Using the implicit function theorem, we take derivative $v$ with respect to $p$:

$$v_p' = \frac{1}{\left[B_v'(v) - \beta t_v'(v,c)\right]}. \tag{E.1}$$

For the first-best problem of private firm (18), we take partial derivatives of (19) with respect to $p$, and we have

$$\frac{\partial P_s}{\partial p} = \frac{\partial\left[vp(v,c) - J_s(v,c)\right]}{\partial p} = v + p \cdot v_p' - m_s \cdot v_p' = 0. \tag{E.2}$$

Recall equation (E.1). We have

$$\overline{p}_s = m_s - v\left[B_v'(v) - \beta t_v'(v, c_d)\right]. \tag{E.3}$$

For the social optimal problem of private firm (25), we take partial derivatives of (25) with respect to $p$, and we have

$$\begin{aligned}\frac{\partial W_s}{\partial p} &= \frac{\partial\left[\int_0^v B(w)dw - \beta v t(v,c) - J_s(v,c)\right]}{\partial p} \\ &= B(v) \cdot v_p' - \left[\beta \cdot t(v, c_d) \cdot v_p' + \beta v \cdot t_v'(v, c_d) \cdot v_p'\right] \\ &\quad - m_s \cdot v_p' = 0.\end{aligned} \tag{E.4}$$

Thus,

$$\widetilde{p}_s = m_s + \beta v \cdot t_v'(v, c_d). \tag{E.5}$$

For the social optimal problem of the government (37), we take partial derivatives of (37) with respect to $p$, and we have

$$\begin{aligned}\frac{\partial W_g^{s<g}}{\partial p} &= \frac{\partial\left[\int_0^v B(w)dw - \beta v t(v,c) - J_g^{s<g}(v,c)\right]}{\partial p} \\ &= B(v) \cdot v_p' - \left[\beta \cdot t(v, c_d) \cdot v_p' + \beta v \cdot t_v'(v, c_d) \cdot v_p'\right] \\ &\quad - m_g^{s<g} \cdot v_p' = 0.\end{aligned} \tag{E.6}$$

Thus,

$$\widetilde{p}_g^{s<g} = m_g^{s<g} + \beta v \cdot t_v'(v, c_d). \tag{E.7}$$

From Assumption 1 that $B(v)$ is a decreasing function of $v$, $B_v'(v) < 0$; recall equations (E.3), (E.5), and (E.7) and $m_s \le m_g^{s<g}$; we have $\widetilde{p}_s \le \widetilde{p}_g^{s<g} < \overline{p}_s$. This completes the proof. □

## F. Proof of Proposition 7

*Proof.* It is assumed that $(v_s^{sc}, c_s^{sc}, n^*)$ is any optimal solution of problem equations (45)–(48). From constrained optimization problems (45)–(48), the following Lagrange-like function can be defined:

$$\begin{aligned}L &= \int_0^v B(w)\mathrm{d}w - vB(v) + \delta\left[v \cdot p(v,c) - J_s(v,c)\right] - (1-\delta)n \\ &\quad + \lambda_{sc1}^{s<g} \cdot \left[v \cdot p(v,c) - J_s(v,c) + n - P_a\right] + \lambda_{sc2}^{s<g} \\ &\quad \cdot \left[-v \cdot p(v,c) + J_g(v,c) - n\right],\end{aligned} \tag{F.1}$$

where $\lambda_{sc1}^{s<g} \ge 0$ and $\lambda_{sc2}^{s<g} \ge 0$ are Lagrangian multipliers which are associated with constraints (45) and (46), respectively. Using the Kuhn–Tucker condition, the first-order conditions of the problem can be determined by

$$\begin{aligned}&\frac{\partial L}{\partial v}\Big|_{(v_s^{sc}, c_s^{sc}, n^*)} = -v_s^{sc} \cdot B_v'(v_s^{sc}) + \lambda_{sc2}^{s<g}\left(m_g - m_s\right) + \\ &\left[\delta + \lambda_{sc1}^{s<g} - \lambda_{sc2}^{s<g}\right] \cdot \left[B(v_s^{sc}) - \beta t(v_s^{sc}, c_s^{sc}) + v_s^{sc} \cdot B_v'(v_s^{sc})\right. \\ &\qquad \left. - \beta \cdot v_s^{sc} \cdot t_v'(v_s^{sc}, c_s^{sc}) - m_s\right] = 0, \\ &\frac{\partial L}{\partial c}\Big|_{(v_s^{sc}, c_s^{sc}, n^*)} = \left[\delta + \lambda_{sc1}^{s<g} - \lambda_{sc2}^{s<g}\right] \cdot \left[v_s^{sc} \cdot p_c' - k\right] = 0, \\ &\frac{\partial L}{\partial \lambda_{sc1}^{s<g}}\Big|_{(v_s^{sc}, c_s^{sc}, n^*)} = P_s(v_s^{sc}, c_s^{sc}) + n^* - P_a = v_s^{sc} \cdot p(v_s^{sc}, c_s^{sc}) \\ &\qquad - J_s(v_s^{sc}, c_s^{sc}) + n^* - P_a \ge 0, \\ &\frac{\partial L}{\partial \lambda_{sc2}^{s<g}}\Big|_{(v_s^{sc}, c_s^{sc}, n^*)} = -v_s^{sc} \cdot p(v_s^{sc}, c_s^{sc}) + J_g(v_s^{sc}, c_s^{sc}) + n^* \ge 0,\end{aligned} \tag{F.2}$$

$$\begin{aligned}\lambda_{sc1}^{s<g} \cdot \frac{\partial L}{\partial \lambda_{sc1}^{s<g}}\Big|_{(v_s^{sc}, c_s^{sc}, n^*)} &= \lambda_{sc1}^{s<g} \cdot \left[v_s^{sc} \cdot p(v_s^{sc}, c_s^{sc}) - J_s(v_s^{sc}, c_s^{sc})\right. \\ &\quad \left. + n^* - P_a\right] = 0,\end{aligned} \tag{F.3}$$

$$\begin{aligned}\lambda_{sc2}^{s<g} \cdot \frac{\partial L}{\partial \lambda_{sc2}^{s<g}}\Big|_{(v_s^{sc}, c_s^{sc}, n^*)} &= \lambda_{sc2}^{s<g} \cdot \left[-v_s^{sc} \cdot p(v_s^{sc}, c_s^{sc}) + J_g(v_s^{sc}, c_s^{sc})\right. \\ &\quad \left. - n^*\right] = 0,\end{aligned} \tag{F.4}$$

$$\frac{\partial L}{\partial c} = -(1-\delta) + \lambda_{sc1} - \lambda_{sc2} = 0, \tag{F.5}$$

$$\lambda_{sc1} \ge 0, \tag{F.6}$$

$$\lambda_{sc2} \ge 0. \tag{F.7}$$

Form equation (F.5) and $0 < \delta < 1$, we have $\lambda_{sc1}^{s<g} - \lambda_{sc2}^{s<g} = 1 - \delta > 0$ and $\lambda_{sc1}^{s<g} > \lambda_{sc2}^{s<g}$. With equations (F.6) and (F.7), we know that $\lambda_{sc1}^{s<g} = 0$ and $\lambda_{sc2}^{s<g} = 0$ will not happen simultaneously. From the above analysis, we know that the government must let equation (46) hold, which results in $\lambda_{sc2}^{s<g} > 0$. If the private firm takes part in the bid actively, constraint (45) will be effective, and then $\lambda_{sc1} > 0$. In general, if the government provides guarantee, it will satisfy the

requirement of the private firm, which also implies the effectiveness of constraint (45).

Then, there is only one case to investigate, namely, $\lambda_{sc1}^{s<g} > \lambda_{sc2}^{s<g} > 0$. With equations (F.3) and (F.4), we have

$$v_s^{sc} \cdot p\left(v_s^{sc}, c_s^{sc}\right) - J_s\left(v_s^{sc}, c_s^{sc}\right) + n^* - P_a = 0, \tag{F.8}$$

$$-v_s^{sc} \cdot p\left(v_s^{sc}, c_s^{sc}\right) + J_g\left(v_s^{sc}, c_s^{sc}\right) - n^* = 0. \tag{F.9}$$

Recall equations (F.8) and (F.9), and we have

$$n^* = P_a - v_s^{sc} \cdot p\left(v_s^{sc}, c_s^{sc}\right) + J_s\left(c_s^{sc}\right) = J_g\left(c_s^{sc}\right) - v_s^{sc} \cdot p\left(v_s^{sc}, c_s^{sc}\right). \tag{F.10}$$

This completes the proof. □

## 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 no conflicts of interest.

## Acknowledgments

This work was supported by the Application-Oriented Undergraduate Pilot Major Construction (Engineering Management Major) Project (Z30011-17-02), the School-Level Project of Shanghai Jianqiao University (SJQ17011), the Central-Level Research Institutes Public Welfare Project (GY2020G-7 and GY2020Z-1), the Science and Technology Committee of Shanghai Municipality (16DZ2290900), and the Ministry of Education Humanities and Social Sciences Fund's Project of the year 2017 (17YJA790007). To Advocate for Legislation for Mandatory Helmets Wearing by E-bike Riders in China (GRSP CHNXX-RD14-1145).

## References

[1] Ministry of Transport of the People's Republic of China, *Toll Road Management Regulations (Revised Draft)*, Ministry of Transport of the People's Republic of China, Beijing, China, 2018, http://www.gov.cn/xinwen/2018-12/21/content_5350738.htm.

[2] Z. Tan, H. Yang, and X. Guo, "Properties of Pareto-efficient contracts and regulations for road franchising," *Transportation Research Part B: Methodological*, vol. 44, no. 4, pp. 415–433, 2010.

[3] B. Niu and J. Zhang, "Price, capacity and concession period decisions of Pareto-efficient BOT contracts with demand uncertainty," *Transportation Research Part E: Logistics and Transportation Review*, vol. 53, pp. 1–14, 2013.

[4] G. Roth, *Roads in a Market Economy*, Avebury Technical, Ashgate Publishing Limited, Aldershot, UK, 1996.

[5] E. Engel, R. D. Fischer, and A. Galetovic, "Highway franchising: pitfalls and opportunities," *American Economic Review*, vol. 87, no. 2, pp. 68–72, 1997.

[6] H. Mohring and M. Harwitz, *Highway Benefits: An Analytical Framework*, Northwestern University Press, Evanston, IL, USA, 1962.

[7] H. Yang and Q. Meng, "Highway pricing and capacity choice in a road network under a build-operate-transfer scheme," *Transportation Research Part A: Policy and Practice*, vol. 34, no. 3, pp. 207–222, 2000.

[8] H. Yang and Q. Meng, "A note on "highway pricing and capacity choice in a road network under a build-operate-transfer scheme," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 7, pp. 659–663, 2002.

[9] Z. Feng, S.-B. Zhang, and Y. Gao, "Modeling the impact of government guarantees on toll charge, road quality and capacity for build-operate-transfer (BOT) road projects," *Transportation Research Part A: Policy and Practice*, vol. 78, pp. 54–67, 2015.

[10] Z. Feng, Y. Zhang, S. Zhang, and J. Song, "Contracting and renegotiating with a loss-averse private firm in BOT road projects," *Transportation Research Part B: Methodological*, vol. 112, pp. 40–72, 2018.

[11] F. Wang, M. Xiong, B. Niu, and X. Zhuo, "Impact of government subsidy on BOT contract design: price, demand, and concession period," *Transportation Research Part B: Methodological*, vol. 110, pp. 137–159, 2018.

[12] S. Shi, Q. An, and K. Chen, "Optimal choice of capacity, toll, and subsidy for build-operate-transfer roads with a paid minimum traffic guarantee," *Transportation Research Part A: Policy and Practice*, vol. 139, pp. 228–254, 2020.

[13] S. Shi, Y. Yin, and X. Guo, "Optimal choice of capacity, toll and government guarantee for build-operate-transfer roads under asymmetric cost information," *Transportation Research Part B: Methodological*, vol. 85, pp. 56–69, 2016.

[14] P. Milgrom, "Auctions and bidding: a primer," *Journal of Economic Perspectives*, vol. 3, no. 3, pp. 3–22, 1989.

[15] E. M. R. A. Engel, R. D. Fischer, and A. Galetovic, "Least-present-value-of-revenue auctions and highway franchising," *Journal of Political Economy*, vol. 109, no. 5, pp. 993–1020, 2001.

[16] G. D. Rus and G. Nombela, "Least present value of net revenue: a new auction-mechanism for highway concessions," University Library of Munich, Munich, Germany, MPRA Paper, 2000.

[17] G. Nombela and G. de Rus, "Flexible-term contracts for road franchising," *Transportation Research Part A: Policy and Practice*, vol. 38, no. 3, pp. 163–179, 2004.

[18] E. T. Verhoef, "Second-best road pricing through highway franchising," *Journal of Urban Economics*, vol. 62, no. 2, pp. 337–361, 2007.

[19] B. Ubbels and E. T. Verhoef, "Auctioning concessions for private roads," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 1, pp. 155–172, 2008.

[20] V. A. C. van den Berg, "Serial private infrastructures," *Transportation Research Part B: Methodological*, vol. 56, pp. 186–202, 2013.

[21] R. S. Bower, "Impact of regulation on economic behavior: discussion," *The Journal of Finance*, vol. 36, no. 2, pp. 397–399, 1981.

[22] X. Guo and H. Yang, "Analysis of a build–operate–transfer scheme for road franchising," *International Journal of Sustainable Transportation*, vol. 3, no. 5-6, pp. 312–338, 2009.

[23] A. M. Spence, "Monopoly, quality, and regulation," *The Bell Journal of Economics*, vol. 6, no. 2, pp. 417–429, 1975.

[24] N. M. Edelson, "Congestion tolls under monopoly," *The American Economic Review*, vol. 61, no. 5, pp. 873–882, 1971.

[25] H. R. Varian, *Microeconomic Analysis*, W. W. Norton & Company, Inc., New York, NY, USA, 1992.

Hindawi
Journal of Advanced Transportation
Volume 2021, Article ID 8816540, 19 pages
https://doi.org/10.1155/2021/8816540

*Research Article*

# Traffic Flow Characteristics and Lane Use Strategies for Connected and Automated Vehicles in Mixed Traffic Conditions

**Zijia Zhong**[1], **Joyoung Lee**[1] **and Liuhui Zhao**[2]

[1]*John A. Reif, Jr. Department of Civil and Environmental Engineering, New Jersey Institute of Technology, NJ, USA*
[2]*College of Engineering, Georgia Institute of Technology, GA 30332, Atlanta, USA*

Correspondence should be addressed to Zijia Zhong; zijia.zhong@njit.edu

Received 22 April 2020; Revised 20 August 2020; Accepted 22 November 2020; Published 13 January 2021

Academic Editor: Meng Meng



Managed lanes, such as a dedicated lane for connected and automated vehicles (CAVs), can provide not only technological accommodation but also desired market incentives for road users to adopt CAVs in the near future. In this paper, we investigate traffic flow characteristics with two configurations of the managed lane across different market penetration rates and quantify the benefits from the perspectives of lane-level headway distribution, fuel consumption, communication density, and overall network performance. The results highlight the benefits of implementing managed lane strategies for CAVs: (1) A dedicated CAV lane significantly extends the stable region of the speed-flow diagram and yields a greater road capacity. As the result shows, the highest flow rate is 3400 vehicles per hour per lane at 90% market penetration rate with one CAV lane. (2) The concentration of CAVs in one lane results in a narrower headway distribution (with smaller standard deviation) even with partial market penetration. (3) A dedicated CAV lane is also able to eliminate duel-bell-shape distribution that is caused by the heterogeneous traffic flow. (4) A dedicated CAV lane creates a more consistent CAV density, which facilitates communication activity and decreases the probability of packet dropping.

## 1. Introduction

The mobility landscape is experiencing a paradigm shift due to rapid advancements of the information and vehicular technologies. Among them, the connected and automated vehicle (CAV) technologies have been contributing to the adoption of next-generation vehicles that are equipped with connectivity (i.e., connected vehicles) and/or automation (i.e., automated vehicles). In spite of CAV's immense benefits and potentials in reshaping the mobility landscape, the adoption of CAVs by consumers is still uncertain [1], although some lower-level vehicle automation in the form of driver-assistance system has been commercially available.

The near-term deployment of CAVs is characterized by mixed traffic conditions, where human-driven vehicles (HVs) and CAVs constantly interact with each other. The potential benefits from CAVs may be offset by the interactions among different types of vehicles. For example, the short following time gap (e.g., 0.6 s) is only feasible when a CAV follows another CAV. To overcome such shortcoming in near-term CAV deployment, managed lane strategies, such as CAV dedicated lane, are one of the promising solutions in order to facilitate the formation of the CAV strings. Practically, managed lane strategies are freeway lanes that are set aside and operated under various fixed and/or real-time strategies in response to certain objectives, such as improving traffic operation [2]. It is also anticipated that managed lane strategies incentivize the adoption of CAV, just as they did for encouraging car-pooling or low emission vehicles.

The goal of this study is to investigate the impact of different lane use strategies under mixed traffic conditions at vehicle trajectory, as well as lane, level. For clarity, we refer mixed traffic condition to the condition that CAVs and HVs operate on the same roadway network in the following discussions. The contributions of the paper include the following:

(1) The analysis of CAV-enhanced traffic flow characteristics at the lane level and vehicle level

(2) The investigation of traffic performance with gradual introduction of CAV platoons under difference managed lane strategies

(3) The implications of managed lane strategies from a dedicated short-range communication (DSRC) perspective

The remainder of the paper is organized as follows. Related work regarding the research of CAVs in mixed traffic and managed lanes is reviewed in Section 2, followed by the evaluation methodology, including customized CAV module and defined scenarios, in Section 3. The simulation results are presented and discussed in Section 4. Lastly, findings and recommendations are discussed in Section 5.

## 2. Literature Review

There have been numerous studies on the implementation and evaluation of CAVs in various traffic settings. Aligning with our research topic, we focused our literature search on two key aspects of CAV studies: (1) CAV evaluation in mixed traffic conditions at network level and (2) managed lane strategies for CAV. A list of abbrevations used can be found in Table 1 in the Appendix.

*2.1. CAV Evaluation in Mixed Traffic Conditions.* Three main approaches have been used to assess the benefits of CAVs: (1) analytical study, (2) simulation evaluation, and (3) field test with equipped vehicles.

On-road testing provides the utmost degree of realism with equipped automated driving systems (ADS) and real-world traffic environment. However, the safety and efficiency issues for testing CAV on public roads have been the major concern, especially after several severe CAV-involved accidents in recent years. Due to safety, technological, and budgetary limitations, the scale of a CAV field test at current stage tends to be small (e.g., with a handful of CAVs). As a result, the conclusions from these small-scale field tests may not be reliably generalized to a traffic flow level. Furthermore, it was estimated by Kalra and Paddock that billions of kilometers of road test would be required to achieve the desired level of confidence in terms of safety of an ADS [3]. Thus, analytical and simulation approaches serve as two primary methods for evaluating traffic flow impact of CAVs.

The majority of the analytical models is based on macroscopic traffic flow models and may experience difficulty in faithfully capturing the complex phenomena in transportation networks, such as lane drop. Smith et al. proposed an analytical framework for assessing the benefits of CAV operations [4]. The results indicated that CAVs improved network mobility performance, even with low MRP and no managed lane policies. Throughput, without managed lanes increased by 4%, 10%, and 16% at the MPR of 10%, 20%, and 30%, respectively. It was also discovered that the managed lane policy facilitated homogeneous CAV traffic flow, leading to more consistent and stable network outputs. An analytical model for determining the optimal managed lane strategy was proposed in [5], where the maximum system throughput in a mixed traffic condition could be calculated under the assumption of random vehicle distribution on a freeway facility. Three types of headways (i.e., conservative, neutral, and aggressive) were used in the model. Wang et al. proposed a second-order traffic flow model for mixed traffic streams with HVs and AVs. The authors found that the second-order model consistently outperformed the first-order one in terms of the accuracy of traffic density when the variability of the penetration rate increased [6].

At the corridor level, a capacity of 4250 vph/pl (vehicle per hour per lane) was observed in [7] on a 6 km highway segment with uniformly distributed ramps under full market penetration of CAVs. The study by Shladover et al. observed a pipeline capacity of 3600 at 90% MPR of CAVs [8], where the pipeline capacity refers to the throughput observed on a single-lane roadway without any interference of lateral movements [9].

Arnaout and Arnaout evaluated CAVs under moderate, saturated, and oversaturated demand levels on a hypothetical 4-lane highway under different market penetrations. They found that 9400 vehicles could be served within an hour when the CAV MPR reached 40% [10]. Songchitruksa et al. assessed the network

Table 1: List of abbreviations.

| Abbreviation | Definition |
|---|---|
| ADAS | Advanced driver-assistance systems |
| ADS | Automated driving systems |
| ACC | Adaptive cruise control |
| AV | Automated vehicles |
| API | Application programming interface |
| BSM | Basic safety message |
| CV | Connected vehicles |
| CAV | Connected and automated vehicles |
| CACC | Cooperative adaptive cruise control |
| CAH | Constant-acceleration heuristic |
| CDF | Cumulative probability function |
| CHEM | Comprehensive modal emission model |
| DSRC | Dedicated short-range communication |
| DLL | Dynamic-link library |
| DTG | Desired time gap |
| E-IDM | Enhanced intelligent driver model |
| GPL | General purpose lane |
| HV | Human-driven vehicle |
| HOV | High-occupancy vehicles |
| IEEE | Institute of Electrical and Electronics Engineers |
| MPR | Market penetration rate |
| MOVES | Motor vehicle emission simulator |
| PET | Postencroachment time |
| SSAM | Surrogate safety assessment model |
| SAE | Society of Automotive Engineers International |
| SUMO | Simulation of urban mobility |
| TTC | Time to collision |
| VAD | Vehicle awareness device |
| WAVE | Wireless access in vehicular environment |

performance with CAVs on the 26-mile I-30 freeway in Dallas, TX, and found the highest throughput being 4400 vph at 50% MPR [11] among four MPR scenarios (i.e., 10%, 30%, 50%, and 70%). Another study [12] also revealed that the mobility benefits of CAV emerged at 30% MPR.

Liu et al. investigated the benefits of alleviating freeway merge bottleneck and compared the performance of CACC with ACC under full market penetration. The results showed that CACC yielded a 50% reduction in fuel consumption (as estimated with the EPA MOVES model) while increasing corridor capacity by 49%, compared to the ACC scenario [13]. With a subsequent test on an 18 km segment of SR-99, the research team found that deploying vehicle awareness device- (VAD-) equipped vehicles along with managed lane strategies was helpful in improving corridor-level traffic flow under low and medium CAV market penetrations [14]. Besides MOVES, comprehensive modal emission model (CMEM) [15], VT-Micro [16], and the Future Automotive Systems Technology Simulator (FASTSim) [17] are among the commonly used vehicle emission models in quantifying potential environmental impact of deploying CAVs.

The potential impact of the short following time headway of CAVs on HVs has also been studied in previous studies. Among them, the KONVOI project found the carry-over effect for CACC drivers in manual driving after the disengagement of the CACC system [18]. In recent years, driving simulator has been employed to study the behavioral adaptation of human drivers operating in the vicinity of CAVs. Nowakowski et al. found that test participants are likely to drive under a shorter following distance in the presence of CACC platoons in the adjacent lane [19]. A similar study was conducted by Gouy et al. to investigate the behavioral adaptation of human drivers along a CACC platoon, in which two CACC platoon configurations were tested: (1) a 10-truck platoon with 0.3 s intraplatoon headway and (2) a 3-truck platoon with 1.4 s intraplatoon headway. It was found that a smaller average HV headway was observed in the first scenario, under which participants spent more time under a 1 s headway, although such short headway was generally deemed unsafe in previous studies (e.g., [20]).

*2.2. CAVs and Managed Lanes.* Managed lanes have been in practice over the years to improve target operation objectives, such as (1) promoting the adaptation of environment-friendly vehicles by offering priority usage to specific travel lanes (e.g., the California Clean Air Vehicle Decal [21]), (2) encouraging car-pooling by adopting high-occupancy vehicle (HOV) lanes [22], and (3) performing active traffic management with the aid of high-occupancy toll (HOT) lanes [23]. A CAV lane is one variant of managed lane strategies which provides exclusive lane use privilege to CAVs. Although managed lane strategies have been widely applied to highway operation with successful cases, due to distinctive operational characteristics of CAVs, knowledge learned from a conventional managed lane may not be directly transferable to the implementation of a CAV lane.

The provision of a CAV-managed lane has two primary reasons. First, CAV-managed lanes can incentivize the adaptation of CAVs by offering priority usage to managed lanes, which typically provides better and more reliable travel because of active traffic management. More importantly and unique to CAVs, CAV-managed lanes can provide accommodations for the underlying operational characteristics of CAVs. A CAV is able to operate at a much closer headway than a human driver with the assistance of V2V wireless communication and the automated driving system (ADS) [24, 25]. Hence, the necessary condition for realizing such a short following headway is the availability of the vehicle driving information of at least one of the predecessors on the same lane, that is, through a CAV following another CAV. Otherwise, the string stability of CAVs cannot be guaranteed [26], and the lack of thereof is termed as CAV degradation [27], which could potentially be a major hurdle for CAVs operating in mixed traffic. A numerical example by Wang et al. has showed that the current technological maturity of CACC contributed negatively to the stability of heterogeneous flow [27].

To mitigate CAV degradation, ad hoc coordination, local coordination, and global coordination are the three major strategies that outline the organization of CAV platoons [28]. Ad hoc coordination assumes that CAVs arrive in random sequence and do not actively seek clustering opportunities in a traffic stream. By extension, the probability of driving around other CAVs is highly correlated to MPR. On the contrary, CAVs actively identify and approach an existing CAV cluster (or other free-agent CAVs) to form a new cluster through local coordination, regardless of CAV MPRs. Finally, global coordination (a.k.a. end-to-end platooning) requires a high-level route planning and extensive communication to coordinate vehicles traveling with the same origin-destination pair even before the CAVs enter highway sections [29].

To successfully form and maintain platoons, accurate and cost-effective localization of CAVs in a dynamic traffic environment remains one of the biggest challenges, especially for local coordination [28, 30]. In the presence of a CAV-managed lane, a higher concentration of CAVs facilitates local coordination with much less stringent requirements on the accuracy of vehicle localization. In addition, the CAV-managed lane strategy aligns well with the three-stage deployment roadmap considering market diffusion and technological maturation for CAVs [31]. In the first stage, the adoption of CAVs is incentivized by allowing the use of the managed lane free of charge. At this stage, the following headway of CAVs on the managed lane may be comparable to that which has been observed from HVs for safety reasons in a mixed traffic condition. In the second stage, a shorter following headway for CAVs could be implemented to further increase the carrying capacity of the managed lane when the demand of CAVs along with the familiarity of road users to CAVs increases.

In the third stage, when the CAVs reach a critical level of MPR, high-performance driving enabled by the CAVs can be achieved due to homogeneous CAV traffic flow on the managed lane.

To assess the impact of CAV-managed lane strategies, Zhang et al. compared the performance of a managed lane and general propose lanes (GPL) based on average speed, throughput, and travel time [32]. The results indicated that the speed improvement in the managed lane was significant compared to that of GPLs. With 20% MPR, the latent demand (the demand that cannot enter the simulation network due to congestion) decreased to zero. Inspired by the fluid approximation of traffic, Wright et al. proposed an algorithm for simulating the weaving activity at the interface of a managed lane and the adjacent GPL at a macroscopic scale [33]. Chen et al. proposed a time-dependent deployment framework that was formulated with a network equilibrium model and a diffusion model. With the constraint of a given set of candidate lanes which corresponds to the field condition, the social cost was minimized with the consideration of different MPR levels [34]. Zhong and Lee studied four managed lane strategies and compared the benefits for GPL and managed lane users in terms of mobility, safety, emission, and equity [35]. In freeway settings, the authors recommended a 30% minimal MRP for deploying a CAV-managed lane to avoid lane use imbalance that could degrade the performance [31, 36].

Qom et al. proposed a multiresolution framework to study the mobility impact of CAV lanes. Traffic flow-based static traffic assignment and the mesoscopic simulation-based dynamic traffic assignment were adapted in the bilevel framework. The former yielded the MPR-based trends, whereas the latter refined the trend based on traffic congestion. The results indicated that it was not beneficial to provide toll incentive for CAVs at lower MPR due to the marginal increase in highway capacity [37]. Ghiasi et al. proposed an analytical capacity model for mixed traffic [38]. The model relied on the Markov chain representation of the spatial distribution of heterogeneous and stochastic headway. With the sufficient and necessary condition of capacity increase proven, the authors emphasized the importance of quantitative analysis of the actual headway setting.

The introduction of a CAV lane to a signalized corridor was reported in [39]. Two configurations of a CAV lane, along with other managed lanes, were evaluated. To accommendate for the turning movements, buffer zones were implemented, where HVs are allowed to temporarily use the CAV lane. Papadoulis et al. evaluated the safety impact of CAVs using the Surrogate Safety Assessment Model (SSAM) [40]. The time to collision (TTC) and the postencroachment time (PET) were adapted with safety thresholds of 1.5 s and 5 s, respectively. They observed substantial safety benefits in terms of reduction in traffic conflicts: 12–47% at 25% MPR to 90–94 % at 100% MPR. In [41], TTC was also used to assess the safety conditions for HVs when CAV local clustering strategy was employed. Ali et al. found that drivers with advanced traffic information enabled by connectivity tend to wait longer and maintained a larger space on mandatory lane change (the communication delay for lane merging assistance was unnoticeable when it was less than 1.5 s). Postencroachment time (PET) analysis also indicated improved travel safety from CAV implementation [42].

*2.3. Summary.* The vast majority of previous studies evaluated the benefits of CAVs at an aggregated level with the emphasis of overall traffic improvement. Analytic models are in macroscopic nature under overly ideal conditions, and they have difficulty in factoring the stochastic nature of human drivers in a mixed traffic environment. CAV-managed lane strategy could be instrumental in the near-term deployment of CAVs, but it is still an underexplored area, despite its increasing recognition.

## 3. Evaluation Framework and Experiment Design

This study focuses on analyzing mixed traffic flow characteristics at a corridor level considering different CAV MRPs and managed lane strategies. In this section, the integrated simulation test bed, transportation network, and simulation scenarios are discussed in detail.

*3.1. CAV Behavior Model.* The PTV Vissim [43], a commercial-off-the-shelf microscopic simulation package, is chosen for the evaluation. Vissim has been widely adapted by transportation practitioners and researchers, owing to its high-fidelity simulation mechanism and flexible modules. Compared to other open-source traffic simulators (e.g., SUMO), one reservation for Vissim being a commercial software is its close-sourced nature. As shown in Table 2, a calibrated Wiedemann car-following model and the enhanced intelligent driver model (E-IDM) [44] were used to model HVs and CAVs, respectively. The intelligent driver model (IDM) and its variants have been used to design the ACC/CACC controller that resembles human-like car-following behaviors [45–49]. As an improved iteration of the collision-free IDM [50], the E-IDM deals with CAV longitudinal maneuver. The behavior model of the E-IDM is expressed in equations (1)–(3): where $a$ is the maximum acceleration; $b$ is the desired deceleration; $c$ is the coolness factor; $\delta$ is the free acceleration exponent; $\dot{x}$ is the current speed of the subject vehicle; $\dot{x}_{\text{des}}$ is the desired speed, $\dot{x}_{\text{lead}}$ is the speed of the lead vehicle; $s_0$ is the minimal distance; $\ddot{x}$ is the acceleration of the subject vehicle; $\ddot{x}_{\text{lead}}$ is the acceleration of the lead vehicle; $\ddot{x}_{\text{IDM}}$ is the acceleration calculated by the original IDM model [50]. The minimal distance can be calculated as $s^*(\dot{x}, \dot{x}_{\text{lead}})$,

Table 2: Differences between HVs and CAVs in the simulation models.

| Vehicle type | Longitudinal control | DTG | Stochasticity |
|---|---|---|---|
| HV | Wiedemann 99 | 1.4 s | Y |
| CAV | E-IDM | 0.6, 1.2 s | N |

$$\ddot{x} = \begin{cases} a\left[1 - \left(\frac{\dot{x}}{\dot{x}_{\text{des}}}\right)^{\delta} - \left(\frac{s^*\left(\dot{x}, \dot{x}_{\text{lead}}\right)}{s_0}\right)\right], & \text{if } \ddot{x}_{\text{IDM}} \geq \ddot{x}_{\text{CAH}}, \\ (1-c)\ddot{x}_{\text{IDM}} + c\left[\ddot{x}_{\text{CAH}} + b \cdot \tanh\left(\frac{\ddot{x}_{\text{IDM}} - \ddot{x}_{\text{CAH}}}{b}\right)\right], & \text{otherwise,} \end{cases} \tag{1}$$

$$s^*\left(\dot{x}, t\dot{x}_{\text{lead}}\right) = s_0 + \dot{x}T + \frac{\dot{x}\left(\dot{x} - t\dot{x}_{\text{lead}}\right)}{2\sqrt{ab}}, \tag{2}$$

$$\ddot{x}_{\text{CAH}} = \begin{cases} \frac{\dot{x}^2 \cdot \min\{\ddot{x}_{\text{lead}}, \ddot{x}\}}{\dot{x}_{\text{lead}}^2 - 2x \cdot \min\{\ddot{x}_{\text{lead}}, \ddot{x}\}}, & \text{if } \dot{x}_{\text{lead}}\left(\dot{x} - \dot{x}_{\text{lead}}\right) \leq -2x \cdot \min\{\ddot{x}_{\text{lead}}, \ddot{x}\}, \\ \min\{\ddot{x}_{\text{lead}}, \ddot{x}\} - \frac{\left(\dot{x} - \dot{x}_{\text{lead}}\right)^2 \Theta\left(\dot{x} - \dot{x}_{\text{lead}}\right)}{2x}, & \text{otherwise.} \end{cases} \tag{3}$$

where $T$ is the desired time gap; and $\ddot{x}_{\text{CAH}}$ is the acceleration calculated by the constant-acceleration heuristic (CAH) component, where $\Theta$ is the Heaviside step function that is used to eliminate the negative approaching rate of subject vehicle [44].

In this study, the E-IDM model is selected as the longitudinal control for the CAVs. Although without built-in multianticipative car-following function, as the literature shows, E-IDM is still a good simple car-following model for CAVs, as the stochastic nature of human driving is removed (i.e., automation property), and the acceleration of the preceding vehicle is taken into account in the driving model (i.e., connectivity property). As shown in Table 3, all the parameters remain the same as those originally specified in [44], with the exception of the desired time gap (DTG), which is defined with two values: 0.6 s and 1.2 s. The former DTG is used when the communication between a preceding CAV and the subject CAV is successful, whereas the latter one is in effect when the communication failure occurs. The updating frequency for the E-IDM model in Vissim is 10 Hz. The density of CAVs which is used to calculate the communication activity is updated at a 2 Hz frequency to reflect the traffic dynamic. Each transmission is assumed to have up to five attempts (four retransmissions). At least one successful attempt is required for a transmission to be considered successful, upon which the DTG is determined.

To implement these two car-following models in Vissim, the subset of the human driving behavior is realized by adjusting car-following parameters of the Wiedemann car-following model, which is relatively straightforward. The E-IDM, on the other hand, is implemented via the external driver model application programming interface (API) and connected with Vissim through a dynamic link library (DLL). The DLL is invoked in each simulation time step such that the default car-following behavior will be overwritten for a specified vehicle type. The DSRC wireless communication module, discussed later in Section 3.2, is also implemented in the API to achieve a dynamic response based on prevailing traffic conditions.

One of the most prominent features in CAV behavior modeling is the short time headway during car-following, which is manifested by several key differences between a CAV and a HV. First, the stochasticity of the CAVs is significantly lower than that of human drivers. This is enabled by the on-board sensors that are able to continuously and accurately perceive the surrounding environment. However, the stochasticity cannot be completely eliminated due to sensor noise and communication delay/error. Second, a CAV has minimal reaction time due to its algorithmic decision-making process and computational power. Past studies have already identified the impact of the reaction time of human drivers in various traffic phenomena, including capacity drop [51] and flow stability [47], whereas driving simulation tests revealed that the information augmented by connectivity could decrease the reaction time for drivers [52].

In addition, human factor plays a crucial role in the resumption of control of a CAV when an ADS exits its operational domain (e.g., high risk of collision, sensor failure, and communication interference). Quantitative evidence regarding the transition of control from traffic psychology or human-machine interactions is still limited [53], though few frameworks have been proposed to simulate human behavior endogenously [54, 55]. For example, the prospect theory was used to model the risk and human perception [56, 57]. The Risk Allostasis Theory [58] was adopted for modeling relationship between cognitive processing of information and physical performance. The Task Capacity Interface [59] was employed by Saifuzzaman et al. for quantifying situational awareness of a driver.

Table 3: E-IDM vehicle control parameters.

| Parameter | $T_{intra}$ | $T_{inter}$ | $s_0$ | $a$ | $b$ | $c$ | $\theta$ | $\dot{x}_{des}$ |
|---|---|---|---|---|---|---|---|---|
| Value | 0.6 s | 1.2 s | 1 m | 2 m/s$^2$ | 2 m/s$^2$ | 0.99 | 4 | 105 km/h |

Calvert and van Arem developed a framework that encompasses the driving task demand and driver task saturation [53]. The framework's main goal is to assess the performance impact during the transition of control for AVs. The total task demand, situational awareness, and reaction time during the transition of control from AVs were explored. The framework showed promising capability in capturing the interactive effects of humans with lower-level AVs. However, empirical evidence is still needed to relax the assumptions used in the framework not only from the cognitive point of view but also from vehicle dynamics and intervehicle interactions.

Another human factor is driver compliance to the ADS. Since, in lower or medium level of automation, the driver is ultimately responsible for his or her vehicle, which means overwriting, when deemed necessary, is possible by the human driver, this control authority, in extreme cases, could cancel out the benefits promised by the CAV technologies. In a recent study [57], Sharma et al. employed the prospect theory to model driver decision-making mechanisms including irrational ones and captured the negative relationship between headway and compliance decision by a driver.

In this study, we represent the differences of a CAV and a HV with different desired time headways through separate car following models, with the following assumptions made for CAVs: (1) no error for the on-board sensors and the vehicle controller, that is, perfect perception; (2) no human factor modeling pertaining to the transition of authority; and (3) no behavior adaptation for CAVs for non-CAV drivers.

*3.2. Wireless Communication Model.* In an early study, we implemented a packet-level communication module through Vissim API [36]. Similar adaptations for the model were also found in previous studies [11, 36, 60]. The analytical model [61] was developed from ns-2, an empirical packet-level network simulator that returns the probability of one-hop broadcast reception of basic safety message (BSM) under IEEE 802.11p, an approved amendment tailored to wireless access in vehicular environment (WAVE) in the 802.11 family protocol. The model uses the concept of communication density level, a metric representing channel load in vehicular communication in the form of the sensible transmission per unit of time and per unit of the road [62]. The data reception rate is determined jointly by communication density level and transmission power. An illustration for the reception probability is shown in Appendix B. Note that this communication model only pertains to the physical layer of the DSRC communication (e.g., no MAC layer delay):

$$P_r(x, \delta, \varphi, f) = e^{-3(x/\varphi)^2}\left(1 + \sum_{i=1}^{4} h_i(\xi, \varphi)\left(\frac{x}{\varphi}\right)^i\right), \quad \xi = \delta \cdot \varphi \cdot f, \tag{4}$$

where $h_i(\xi, \varphi)$ is the two-dimensional polynomial of fourth-degree for all curving fitting parameters [63], which is also shown in Appendix B; $\xi$ is communication density, events/s/km; and $\varphi$ is the transmission power, m; $\delta$ is vehicle per kilometer that periodically broadcast messages, veh/km; and $f$ is transmission rate, Hz.

*3.3. Transportation Network.* A 9.3 km 4-lane hypothetical network was constructed in Vissim with two interchanges located at mile markers 2 (km) and 6 (km), respectively. An abstract geometry of the network along with vehicle demand of the origins and destination is shown in Figure 1. The primary reason for using a simply synthetic network is to limit variables for the simulation. Note that the driving behavior parameters for the Wiedemann car-following model (for HVs) are the same as those in previous studies [31, 41, 64, 65], representing a subset of the calibrated driving behavior in the I-66 segment in northern Virginia. The demand originated on the mainline is deliberately set higher than usual to create a congested network. The speed limit for the mainline of the network is set as 120 km/h. Three data collectors are placed at "C1," "C2," and "C3" locations.

*3.4. Managed Lane Scenarios.* Three cases of CAV lanes, as shown in Table 4, are implemented in the network:

(1) No managed lane (NML): This scenario serves as the base condition of the study. There is no priority lane use for CAVs, and they are mixed with HVs throughout the network;

(2) One CAV lane (CAV-1): In this strategy, one CAV lane is implemented in the left-most lane (the fourth lane from the right);

(3) Two CAV lanes (CAV-2): An additional CAV lane is added to the CAV-1 case, making two CAV lanes available at the left-most lane and the second-left-most lane in the roadway segment. It aims to investigate the duel managed lane configuration.

As revealed in previous studies [31, 32, 66, 67], a managed lane may have a detrimental effect on traffic performance if implemented prematurely, that is, usually with an MPR less than 30%. Therefore, in this study, we set CAV MPRs for "CAV-1" to start from 30%. With the same logic, the "CAV-2" cases start with 40% to cover certain transition MPR, since the linear extrapolation may not hold.

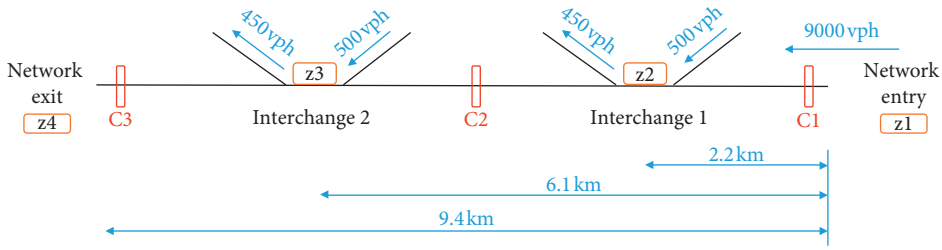



Figure 1: Network geometry and demand.

Table 4: Managed lane evaluation plan.

| Policy<br>ID | No managed lane<br>NML | Managed lane #1<br>CAV-1 | Managed lane #2<br>CAV-2 |
|---|---|---|---|
| 1st lane | HV + CAV | HV + CAV | HV + CAV |
| 2nd lane | HV + CAV | HV + CAV | HV + CAV |
| 3rd lane | HV + CAV | HV + CAV | CAV |
| 4th lane | HV + CAV | CAV | CAV |
| MPR | 0%–100% | 30%–100% | 40%–100% |

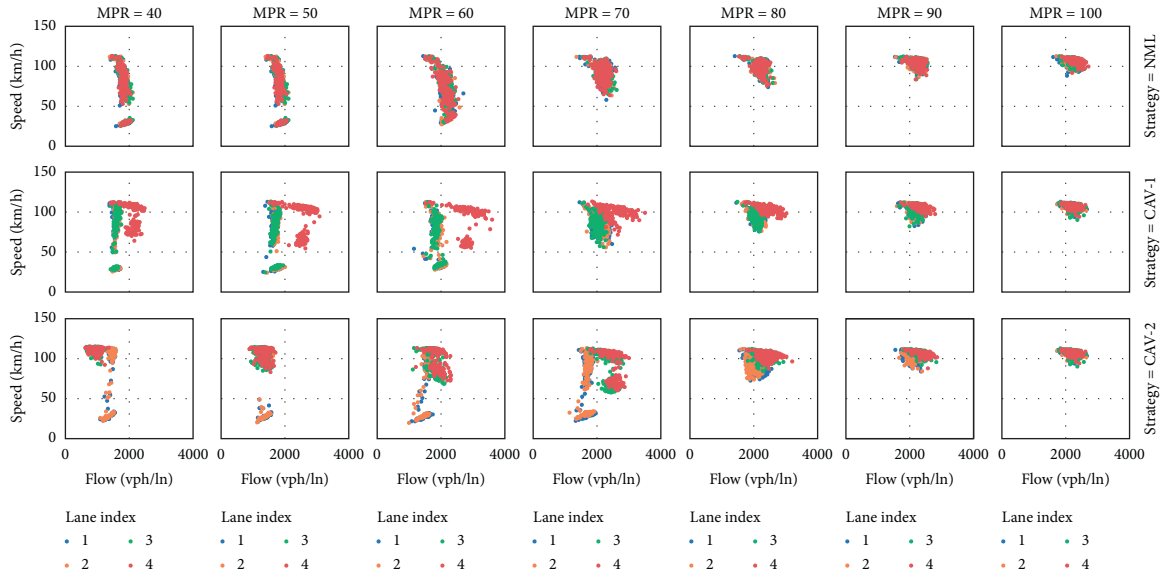



Figure 2: Speed-flow curves.

## 4. Results and Analysis

Five replications are run for each combination of managed lane policies and MPRs. Aggregated data are collected at 5-minute intervals, and the raw data are collected at each simulation time step. The analysis is performed on five aspects: (1) traffic flow characteristics, (2) headway distribution, (3) fuel consumption, (4) wireless communication, and (5) overall network performance.

*4.1. Traffic Flow Characteristic.* Figure 2 exhibits the speed-flow characteristics of the simulation scenarios having 40% MPR and above. The plot is color-coded by travel lanes with index "1" representing the right-most lane and "4" the left-most lane. The speed-flow diagram is comprised of a stable region and an unstable (congested) region, separated by the optimum (maximum) flow. Several distinctive patterns can be observed. First, regardless of the managed lane strategy, the sample points become more concentrated as the MPR increases, with the disappearance of the congested region typically found in the lower speed region. Second, the CAV lane has a distinct pattern compared to the GPLs. Such pattern is most apparent in CAV-1, where the traffic samples on the left-most lane (CAV lane) shift to the right along the flow axis. The congested region disappears when MPR reaches 70% in the CAV-1 case for all of the lanes. The improvement for the GPs is due to a higher carrying capacity of the CAV lane, which results in less traffic on the GPLs. The homogeneity of the CAV traffic is the primary factor in realizing the mobility benefit of CAVs: in NML cases, the sample points from different lanes are evenly distributed, in

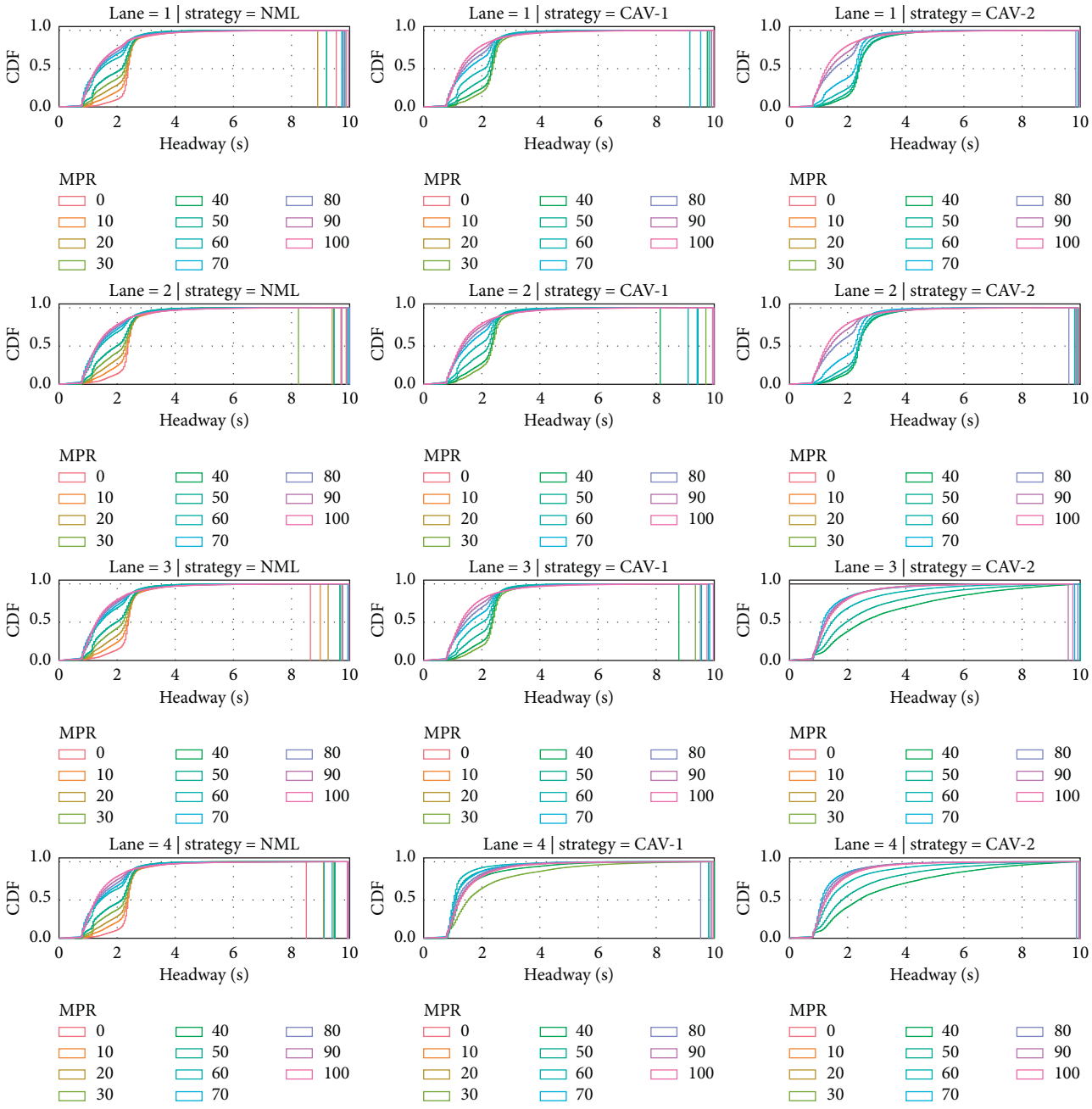



Figure 3: Cumulative distribution for headway distribution among travel lanes.

contrast to managed lane cases. For the CAV-2 case, the separation of the CAV lanes (the left-most and the second left-most) started to appear at 70% MPR. At full penetration (100%), the traffic patterns are very similar, as the managed lane becomes irrelevant.

*4.2. Headway Distribution.* The simulation collects raw data from the data collector, an equivalent of real-world detectors (e.g., loop detectors, video cameras, and microwave sensors). By analyzing the high-resolution raw data (collected every 0.1 s), the headway distribution in CAV lanes can be obtained. Recall that the collectors are placed in three sections of the roadway segment, as shown in Figure 1.

The cumulative probability function (CDF) curves are displayed in Figure 3. The vertical lines in the figure are the headways when 100% cumulative probability is reached. The slope of the CDF indicates the level of concentration of the samples within a distribution. In NML cases, two types of tipping points exist: the one at lower headway resulting from a high MPR and the one with higher headway observed at a low MPR (below 40%). For CAV-1, the pattern for CDF at 30% and 40% is transformed to the pattern observed at high MPRs. With 2 CAV lanes, the CDF increases gradually in the

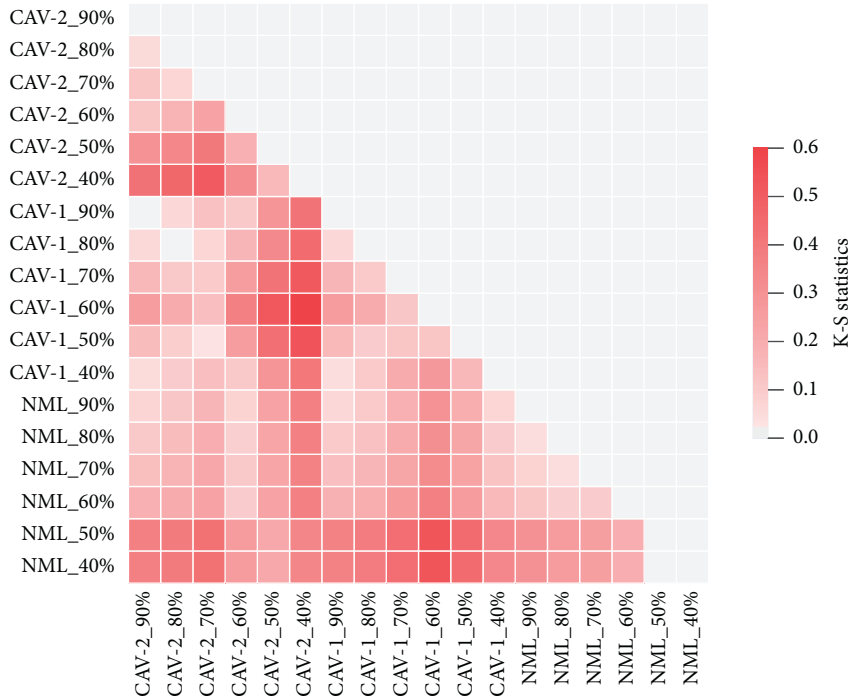



Figure 4: K-S statistics for CDF comparison.

mid-range MPR (40% to 60%) because of undersaturation on the CAV lanes, as illustrated in the CDF on the 3rd and 4th lanes. Such undersaturation situation is alleviated when the MPR reaches 70%. A similar pattern in CDFs is observed at a high MPR range (i.e., 80% to 90%) regardless of the managed lane strategies, indicating a high concentration of samples with headway above 1 s.

Two-sample Kolmogorov-Smirnov (K-S) test is adopted to analyze the CDFs to check whether two random samples are from the same population [68]. It is a nonparametric test where no assumption is made regarding the distribution of the variables [69]. The null hypothesis ($H_0$) of the two-sample K-S test is that the two sample sets are from the same continuous distribution. Nearly all the CDFs in the pairwise comparison reject the null hypothesis with a low $p$ value at the 0.05 significance level, with the exception of the comparison of 40% and 50% in NML. Figure 4 is a heatmap that shows the pairwise K-S statistics that represent the supremum of the two tested empirical CDFs. The denser the color, the higher the difference in cumulative probability between two comparing scenarios.

The average headway for HVs and CAVs in every travel lane is shown in Figure 5. The row represents the vehicle types, whereas the column represents the travel lane. Recall that the 4th lane is the left-most lane. For HVs, their averaged headway decreases as the MPR increases in CAV-1 and CAV-2 cases. While the headway also decreases in the NML case, it is at a lesser rate. When it comes to CAVs, the decreasing rate in CAV-2 is greater than that in CAV-1 or NML. The mean headway is around 4 s in CAV-2 case when the MPR is low or in middle range due to low lane utilization in the CAV lanes. The average headway in CAV-2 case reaches a comparable level to its counterparts at 70% MPR, which is the deflection point. The lowest mean headway achieved among all scenarios is observed at 70% MPR in CAV-1 case for CAVs, which corresponds to the maximum capacity with all other factors being equal. Lastly, the headway trend for CAVs remains a similar pattern across four travel lanes in the NML case, since CAVs are uniformly distributed across all lanes.

Figure 6 shows the comparison of headway distributions in the left-most lane among three managed lane scenarios under different MPRs. In the 40% to 70% MPR range, it is shown that implementing a managed lane for CAVs clearly shifts the distribution to the left-hand side, which represents smaller headways. The distributions of headway collected for either CAV-1 or CAV-2 become "narrower" (with less standard deviation), as the MPR increases from 40% to 70%. The highest bin of the histogram for both CAV-1 and CAV-2 cases is 1 s–1.2 s when the MPR is below 50%. When the MPR is higher than 50%, the highest bin of the histogram shifts to 0.8 s–1 s. In comparison, the NML case does not exhibit such a concentration pattern as the MPR increases. The result indicates that a homogeneous traffic flow comprised of only CAVs is able to realize the short headway benefits from deploying CAVs.

*4.3. Fuel Consumption.* The VT-Micro model [16], an individual vehicle- and operation-level emission model, is adopted to calculate the instantaneous fuel consumption rate. The inputs for the VT-Micro model are instantaneous vehicle speed and acceleration, and the output is the second-by-second fuel consumption rate, as shown in equation (5), where $\dot{x}$ is the instantaneous speed, $\ddot{x}$ is instantaneous acceleration, and $L^e_{i,j}$ and $M^e_{i,j}$ are regression coefficients for emission and fuel consumption at speed power $i$ and acceleration power $j$, respectively:

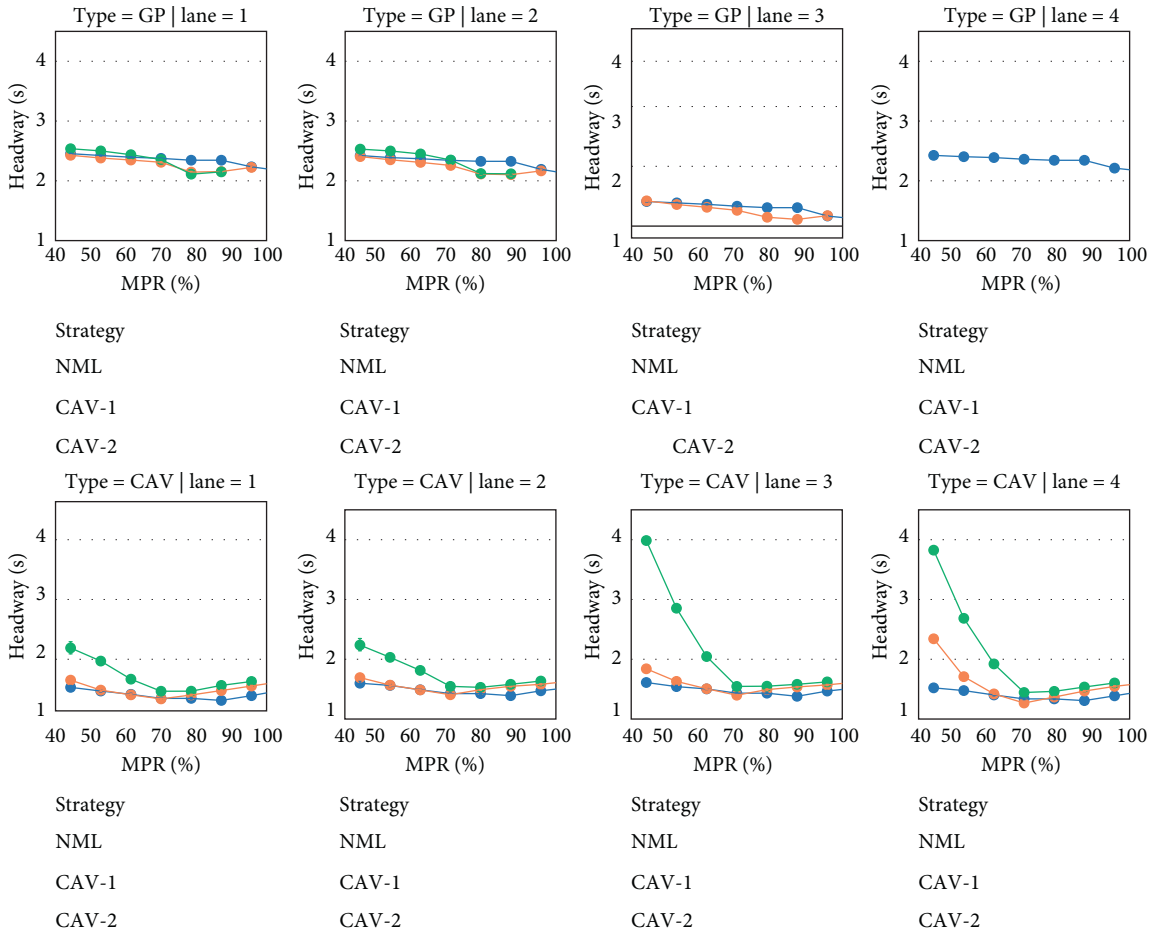



Figure 5: Average headway.

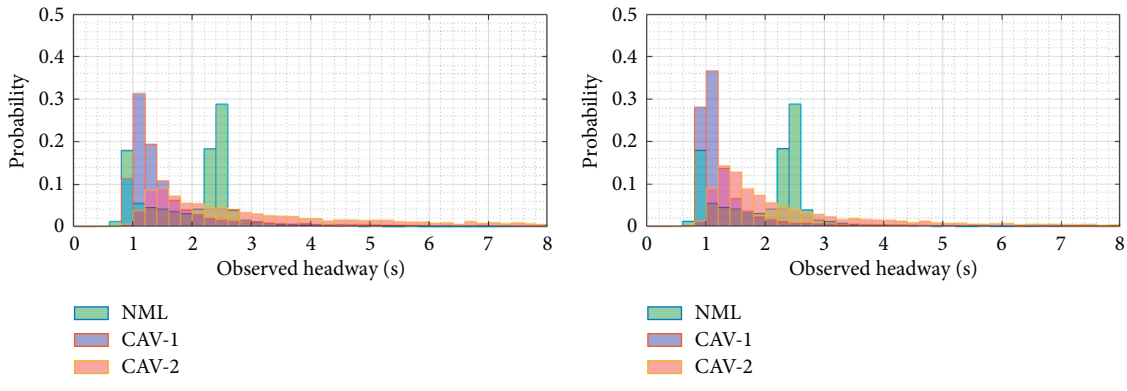



Figure 6: Continued.

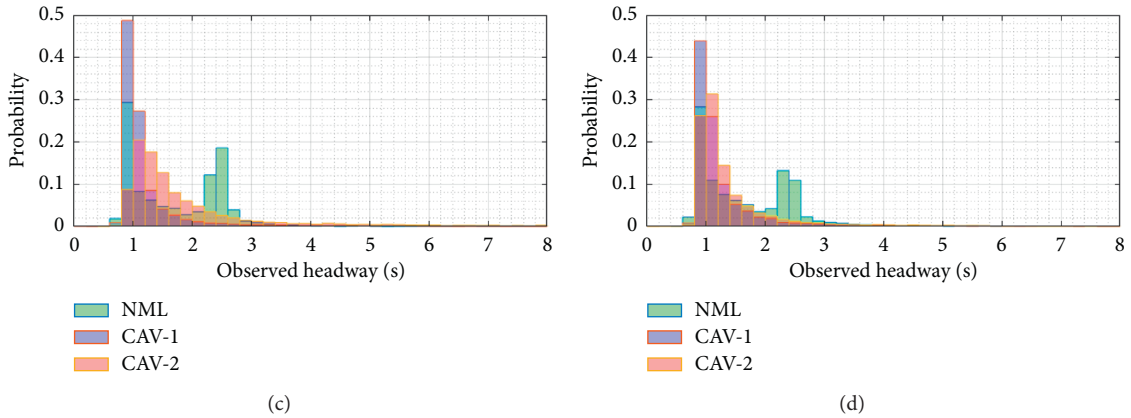



Figure 6: Headway distributions in the left-most lane. (a) 40% MPR. (b) 50% MPR. (c) 60% MPR. (d) 70% MPR.

$$f(\dot{x},\ddot{x}) = \begin{cases} \exp\left(\sum_{i=0}^{3}\sum_{j=0}^{3}\left(L_{i,j}^{e}\cdot\dot{x}^{i}\cdot\ddot{x}^{j}\right)\right), & \text{for } \ddot{x}\geq 0, \\ \exp\left(\sum_{i=0}^{3}\sum_{j=0}^{3}\left(M_{i,j}^{e}\cdot\dot{x}^{i}\cdot\ddot{x}^{j}\right)\right), & \text{for } \ddot{x} < 0. \end{cases} \tag{5}$$

The vehicle data was derived from the raw data from the detectors in three locations marked in Figure 1. The result for the fuel consumption is plotted in Figure 7, which shows two distinctive patterns for the GPLs and the managed lane. The concentration of fuel consumption is within 5 ml/s to 10 ml/s for lanes that allow HV operation (i.e., mixed traffic), when the MPR for CAVs is equal to or less than 60%. When the MPR rises to above 60%, the instantaneous fuel consumption shifts to lower values with a "narrower" slope: higher concentration between 5 ml/s and 7 ml/s.

We then isolate the CDF curve for both CAVs and HVs, when they operate on the left-most lane under homogeneous flow condition. More specifically, the separated CDF curves represent the observations of HVs from the 0% MPR in NML case and the observations for CAVs from the 100% MPR for CAV-1 case. The CDF curves in Figure 7 exhibit two different patterns for CAVs and HVs. The former with 60% of the observations fall below 4 ml/s, whereas the latter with 60% of the observations fall below 12 ml/s with a wider spread. The wider spread for HVs is probably caused by the stochastic nature of human drivers (which is simulated by the Wiedemann model). Hence, the mixed traffic condition is comprised of two competing flows that excrete their influence.

In the GPLs, the MPR plays a role as an indicator for the dominance of each traffic flow. The higher the MPR is, the closer the CDF curves approach the pattern of managed lane that is used by CAV exclusively. In the managed lane, the CAV traffic is the sole dominating traffic. Therefore, the fuel consumption curve exhibits only CAV traffic characteristics, regardless of the MPR. We include the fuel consumption rate CDF curves for HVs and CAVs in Appendix C, Figures 8 and 9: both figures demonstrate the shift towards CAV fuel consumption CDF pattern as the MPR grows. The difference in fuel consumption between two types of homogeneous flow can be found in Figure 10 in Appendix.

*4.4. Wireless Communication.* Figure 11(a) shows the maximum and the average density for instances of V2V communication among three managed lane policies. Recall that the DSRC communication model only deals with the physical layer. While the transmission density increases as the MPR increases, the maximum density in NML is higher than CAV-1 and CAV-2, because the CAV platoons were broken down by certain HVs, which are susceptible to shockwaves. As such, the traffic flow is compressed, producing a higher traffic density and thus higher transmission density. With the aid of CAV lane, the communication density is thus maintained at a lower level. In a CAV lane, the CAVs distribute longitudinally on the managed lane. The NML, in comparison with two managed lane cases, is more likely to generate pockets of traffic with CAVs across multiple lanes, which could result in localized higher transmission activity.

The probability of successful reception of BSM from a leading vehicle to a subject vehicle is shown in Figure 11(b). The probability curves under CAV-1 and CAV-2 scenarios are in close proximity to each other and they are showing the same trend. The maximum difference between these two curves is 0.04 at 90% MPR. The probability of successful communication of NML at high MPR range (60% to 90%) is consistently lower than those of CAV-1 and CAV-2. This is caused by the compression of traffic flow by localized shockwaves. There is an overall decreasing trend of the probability as the MPR increases but there still remains a successful rate of 94% and above.

*4.5. Network Performance.* The measures used in this section gauge the overall performance of the simulation network at an aggregated level. The throughput represents the total number of vehicles that have arrived at their destinations,

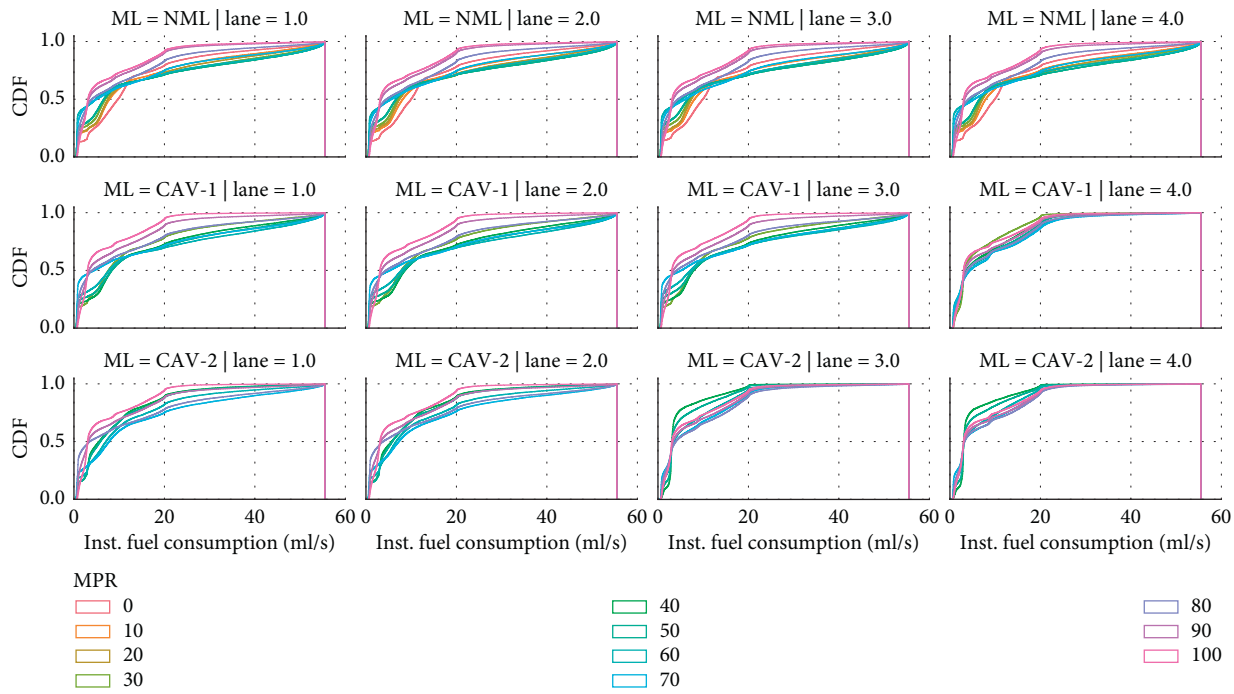



Figure 7: Instantaneous fuel consumption for all vehicles.

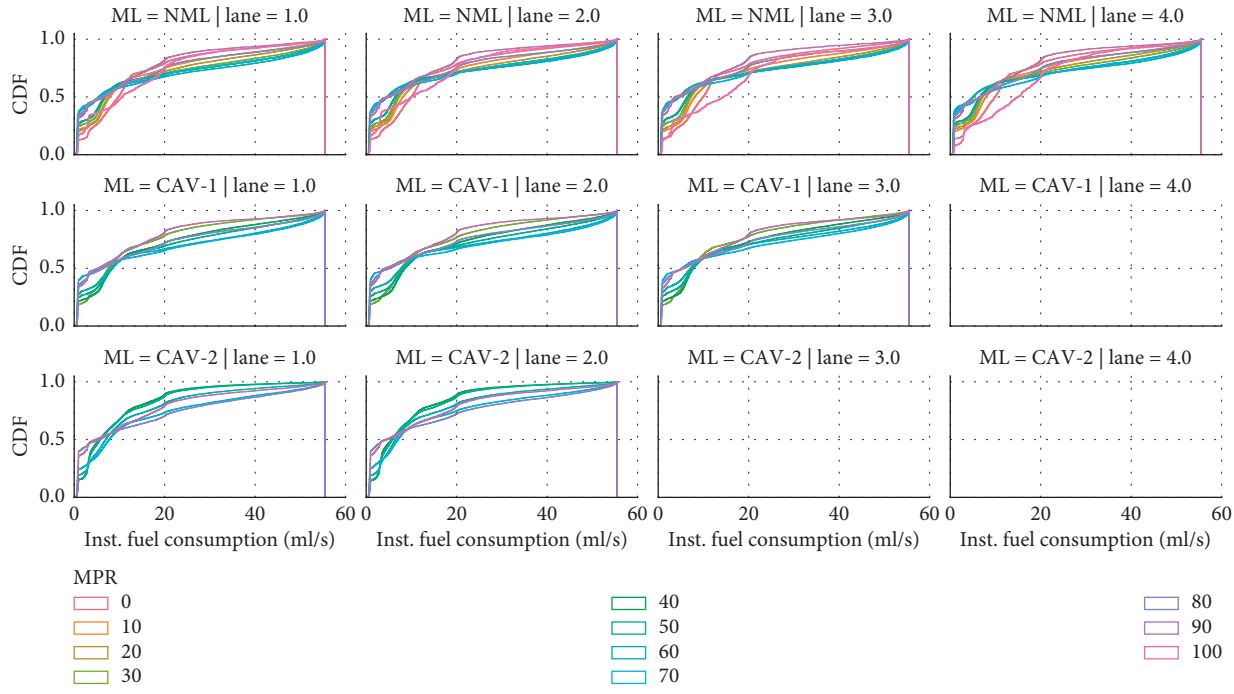



Figure 8: Instantaneous fuel consumption for HVs.

shown in Figure 12. As mentioned before, the network was configured with a higher than usual demand. With a 10,000 vph demand for a four-lane highway, the network was only able to process 6500 vph in the absence of CAVs. Under the NML scenario, as the MPR of CAVs increases, so does the network throughput. The throughput reaches approximately 8000 vph with 40% and 50% MPRs. However, at 60% MPR, the network throughput is boosted again and remains at the same level at 9600 vph when the MPR is above 70%. The throughput in CAV-1 case begins to outperform the NML case at MPR 50% and keeps increasing to 9700 vph at 70% MPR, where the throughput starts leveling in spite of

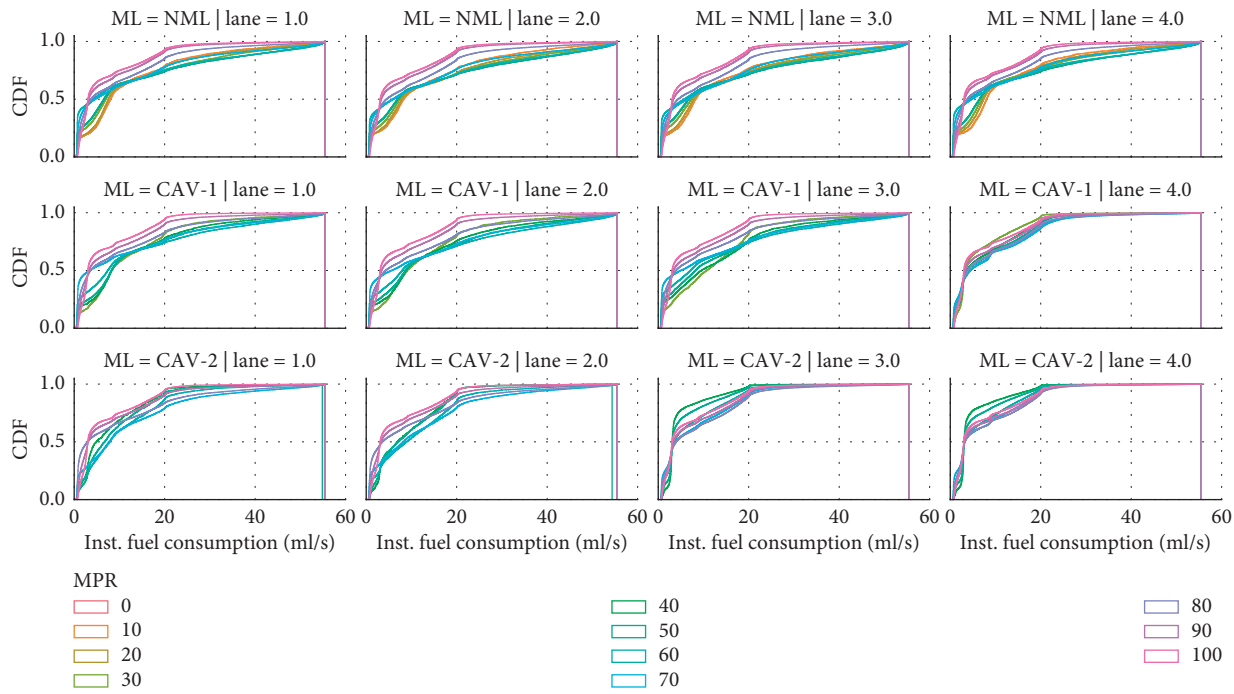



Figure 9: Instantaneous fuel consumption for CAVs.

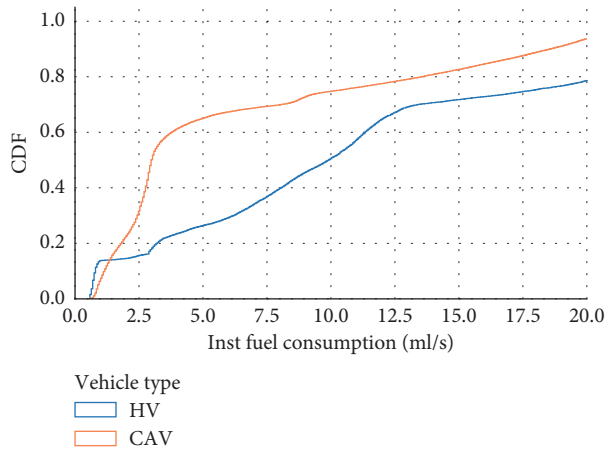



Figure 10: Instantaneous fuel consumption curve for homogeneous flow.

the increase in MPR. For the CAV-2 policy, the system throughput only reaches the same level of the two counterparts at 70% MPR due to underutilization of CAV lanes with low MRPs.

The average delay experienced by vehicles (plotted in Figure 13(a)) within the network is calculated by dividing the total delay by the sum of the vehicles within the network and the vehicles that have exited the network. For three strategies, the average delay starts to decrease as the throughput levels off: at 60% for NML and CAV-1 and at 70 % for CAV-2. Such seemingly counter-intuitive phenomena could be explained by taking into account the average speed, which is shown in Figure 13(b): when the throughput is in a graduate increase as the MPR goes until 60%, the average speed exhibits a decreasing trend, which is in an inverse relationship with vehicle delay. This trend is in agreement with the speed-flow fundamental diagram.

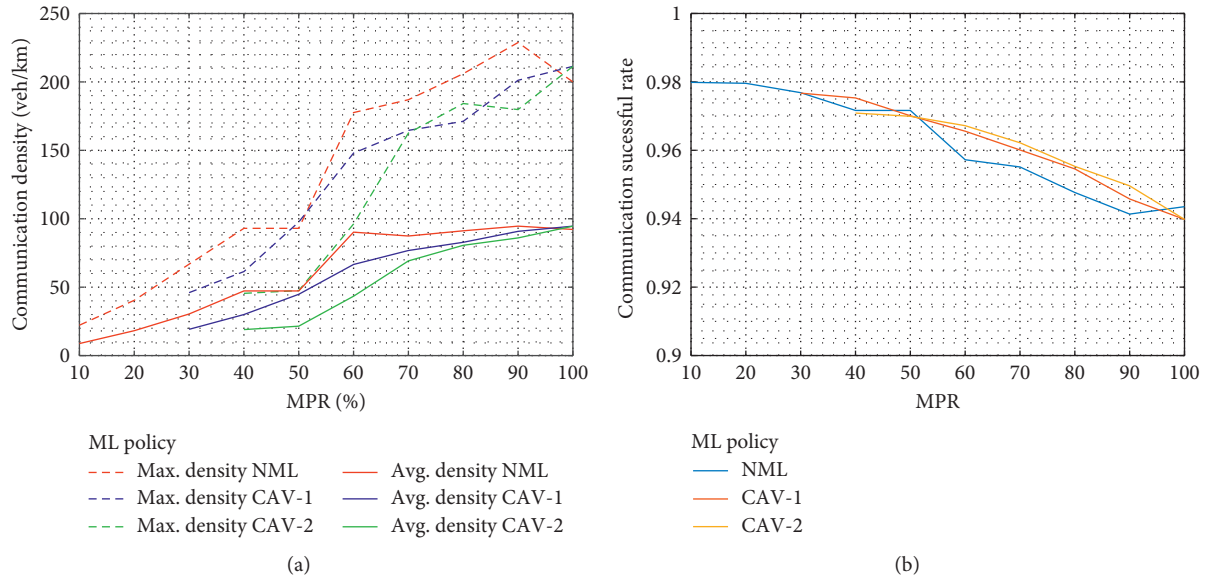



Figure 11: V2V communication performance measure. (a) Vehicle density. (b) Packet perception rate.

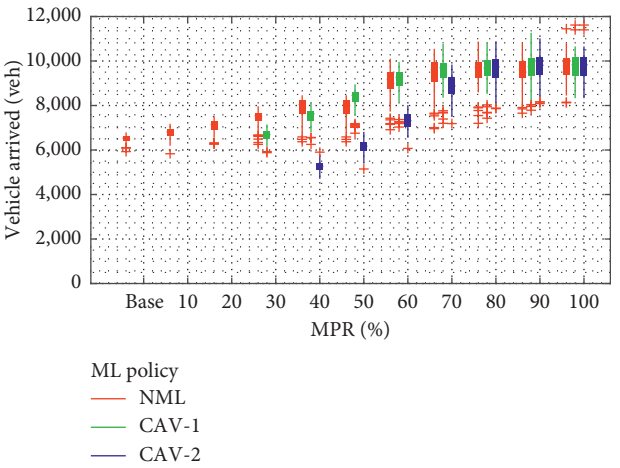



Figure 12: Network throughput.

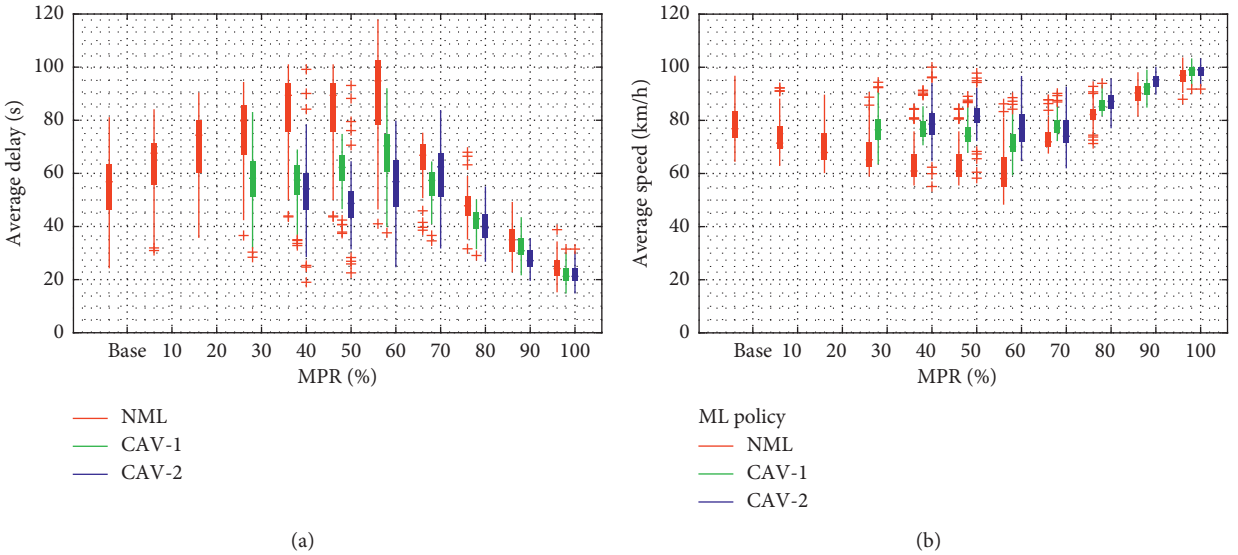
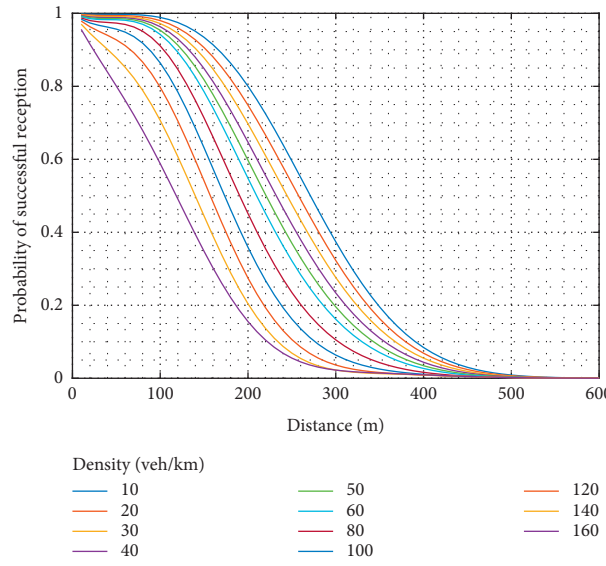



Figure 13: Average speed and delay. (a) Average delay. (b) Average speed.

| | (j,k) | | | | |
|---|---|---|---|---|---|
| | (0,0) | (1,0) | (2,0) | (3,0) | (4,0) |
| $h_1^{(j,k)}$ | 0.0209865 | $-9.66304e-07$ | $-1.72786e-11$ | $5.09506e-17$ | $-7.91921e-23$ |
| $h_2^{(j,k)}$ | 2.24743 | $7.84884e-07$ | $2.28533e-10$ | $-5.89802e-16$ | $3.55262e-22$ |
| $h_3^{(j,k)}$ | 2.56426 | $2.82287e-05$ | $-7.09939e-10$ | $1.34371e-15$ | $-3.01956e-22$ |
| $h_4^{(j,k)}$ | 2.41146 | $-9.32859e05$ | $6.77403e-10$ | $-9.64188e-16$ | $3.69652e-23$ |
| | (3,1) | (2,1) | (2,2) | (1,1) | (1,2) |
| $h_1^{(j,k)}$ | $3.16577e-20$ | $2.13587e-14$ | $-5.05716e-17$ | $4.00928e-09$ | $-1.88707e-11$ |
| $h_2^{(j,k)}$ | $4.07120e-19$ | $-2.66510e-13$ | $8.64273e-17$ | $-7.31274e-08$ | $2.98549e-10$ |
| $h_3^{(j,k)}$ | $-1.85451e-18$ | $1.02847e-12$ | $1.80250e-16$ | $1.56259e-07$ | $-8.50944e-10$ |
| $h_4^{(j,k)}$ | $1.85043e-18$ | $-1.13894e-16$ | $-4.05333e-16$ | $-2.56738e-08$ | $6.24415e-10$ |
| | (1,3) | (0,1) | (0,2) | (0,3) | (0,4) |
| $h_1^{(j,k)}$ | $3.25406e-14$ | 0.000418109 | $-4.30875e-06$ | $1.00775e-08$ | $-7.32254e-12$ |
| $h_2^{(j,k)}$ | $-3.24982e-13$ | 0.00498750 | $-7.22232e-06$ | $1.69755e-08$ | $-2.94381e-11$ |
| $h_3^{(j,k)}$ | $7.59094e-13$ | −0.0227008 | $7.50391e-05$ | $-1.81469e-07$ | $2.02182e-10$ |
| $h_4^{(j,k)}$ | $-3.57571e-13$ | 0.0191490 | $-6.92678e-07$ | $1.79917e-07$ | $-2.07263e-10$ |

(a)




(b)

Figure 14: DSRC model coefficients and PDF. (a) Coefficient $h_i^{(j,k)}$ in equation (4). (b) PDF of successful reception (300 m power range).

## 5. Discussion and Conclusions

In this section, we highlight the findings from the previous section and discuss the study in a boarder context.

*5.1. Summary.* The analysis results indicate that the introduction of CAV could increase the throughput of the overall system, even when no managed lane policy is in place. The congestion region in the speed-flow diagram disappears as the MPR of the CAVs increases. This is an indication of the improvement of roadway capacity owing to CAVs, which is consistent with the findings of previous studies. More importantly, the congestion region first disappears in the CAV lane in CAV-1 case, illustrating that the homogeneity of CAV traffic results in a more stable traffic flow with a high throughput. A CAV lane, with an MPR as low as 40%, is able to accommodate more traffic compared to a GP lane and it helps to alleviate the overall congestion of the network. The average vehicle delay exhibits a decreasing trend, even after the network throughput levels after 70% MPR. This is an indicator that the network is able to carry more traffic than the high demand specified in Figure 1.

The individual headways among consecutive vehicles are measured for each lane. From the headway distribution, one can measure not only the compactness of the traffic but also the stability of the traffic flow. Both HVs and CAVs have a predominate headway as shown in Table 4. In a heterogeneous traffic flow, two spikes with different tipping points can be observed in the headway CDF curve. Each segment indicates a high concentration for the headway samples. One is for the following headway samples observed on HVs, and the other is for the headway samples for CAVs. With traffic homogeneity on the CAV lane, there is only one spike on the CDF curves. The magnitude of the spike also depends on the lane occupancy, as evidenced by the comparison of CAV-1 and CAV-2 at the same MPR. The two-spike pattern remains even at high-range MPR (i.e., 60–80%) in the absence of CAV lane (the NML case).

The VT-Micro model, which produces instantaneous fuel consumption for individual vehicles, was employed to estimate the environmental impact of the CAV lane. The vehicle speed and acceleration were collected as inputs and the relative fuel consumption, instead of the absolute one, was examined. Again, distinct patterns for a GPL and a CAV lane were observed. The average instantaneous fuel consumption for CAV lane has a narrower distribution.

Lastly, the DSRC communication was measured using an analytical communication model that is derived from a package-level network simulator. It simulates the physical layer of the DSRC communication that is an integral element of CAVs. We found a lower communication density in CAV lane, as the CAVs were more evenly distributed longitudinally. A lower communication density indicates a less congested communication channel, which increases the performance of the V2V communication. Compared to CAV-1 and CAV-2 scenarios, it is more likely under NML scenario to generate pockets of traffic with CAVs across multiple lanes, which could introduce higher localized transmission activity and increase the loss of BSM packets.

The overall results show that a single CAV lane in a four-lane highway network is able to provide the necessary technical accommodation efficiently in the mixed traffic conditions with a wide range of MPR. A CAV dedicated lane is helpful to guarantee the benefits of CAVs, as it creates a homogeneous

CAV flow. Implementing two CAV lanes, however, may adversely affect the overall traffic, especially when the MPR of CAV does not warrant an additional CAV lane.

*5.2. Limitations.* While the paper demonstrates the benefits of managed lane for CAV at lane level and vehicle level, we should note that there are limitations in this study and the benefits are realized in a controlled environment under certain assumptions. First of all, although the Wiedemann model is behaviorally sound and has been adopted by numerous researchers for simulating human drivers, the complexity of a human driver under dynamic traffic conditions is difficult, if possible at all, to be captured by simulation models. In addition, the behavioral adaptation for human drivers in the presence of CAV is not known yet, due to the lack of empirical evidence in the public domain. Preliminary results revealed that a smaller time headway was adopted by a HV when driving along side closed platooned CAVs [70]. Note that the Wiedemann driver behavior parameters were calibrated using field data where CAVs have not been deployed on the roadway yet. The calibrated parameters represent a subset of the driving population, and they may not be directly transferable to other driving conditions or demographics. The E-IDM, while being widely adopted, does not contain the multianticipative car-following feature, which has been promoted as one of the crucial features enabled by V2V communication. Therefore, the performance of the CAVs is expected to be more conservative. Like many existing CAV car-following models, the E-IDM does not factor the aspects of human factor that is anticipated to be more pronounced in the lower levels of automation.

In addition, there are several salient issues regarding the low-level automation and its modeling as well. For a CAV, the drivers' acceptance of short following headway (e.g., 0.6 s) is still an open question [24], given that the short following headway is technologically attainable. It is reasonable to expect that the acceptance of extremely short headway would be low initially, although it will gradually increase as CAV penetration increases. The pace of adaption, though, is largely depending on the level of confidence to the ADS from human drivers. The level of compliance from drivers (in the absence of automation) is also an important factor in harnessing rich information brought by the connectivity. The layer of driver stochasticity in reacting to traffic information remains. In the extreme case, a complete disregard of useful information could negate the benefits of connectivity.

Another crucial issue is the transition of control from the ADS back to the human driver. As per the definition of vehicle automation by the SAE, the level 3 automation (and below) requires a fallback receptive driver when the ADS exits its designed operational domain. As studies have shown, such fallback process is way more complicated than merely retaking the steering wheel. First, a driver needs to regain situational awareness of the traffic environment from the disengagement of driving. The surge in cognitive demand during the initial period of reengaging in driving tasks could result in deterioration in driver's performance (e.g., increased reaction time and inadequate situational awareness). This aspect rarely exists in current CAV models, and much likely it will require an endogenous cognitive model that is able to take into account the driving task demand and the cognitive capacity of human drivers [53]. Therefore, the human-machine interfacing is seldom captured in current simulation model, including the one used in this study.

*5.3. Future Research.* The future research would focus on relaxing the assumptions in this study. The first direction is the CAV behavior modeling. Researchers have recently started the incorporation of human factor aspect, such as an extension module in IDM to model driver's responses to advanced traffic information [57], and an explicit human cognitive model for the transition of control [53]. Such developments offer a great opportunity to introduce human factor in a mixed traffic flow in the future. Second, the innermost lane is generally assigned as the managed lane in current practices, which requires eligible users to merge to access the managed lane and induces additional demand of lane changing. The access plan (e.g., ingress and egress points of the managed lane, eligibility) requires further study to minimize the negative impacts caused by induced weaving activity. A cost-benefit analysis may also be warranted for comparing managed lane strategies with other emerging technologies, such as vehicle awareness device (VAD), for the near-term deployment of CAV. Some researchers have started the exploration of right-most managed lane in USA [65]. Lastly, the characteristics of mixed traffic flow that is anticipated in the near-term deployment of CAV need further exploration. In particular, the impact of CAVs at individual trajectory level by analyzing high-resolution vehicle trajectory data requires further exploration (Appendix A).

# Appendix

## A. List of Abbreviations

The list of abbreviation used is provided in Table 1.

## B. Coefficients for the Wireless Communication Model

The coefficients obtained from the polynomial function $h_i(\xi, \varphi)$ are shown in Figure 14(a). It is worth stressing that even seemingly negligible values, if omitted, could result in deviation in the probability of reception from 8% to 100% [63]. The probability distribution curves for transmission scenarios are shown in 14(a).

## C. Instantaneous Fuel Consumption for HV and CAV

Instantaneous fuel consumption for HV and CAV is shown in Figures 8 and 9. The instantaneous fuel consumption curve for homogeneous flow is shown in Figure 10.

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

## Acknowledgments

This work was supported in part by the National Science Foundation under Grant No. CMMI-1844238.

## References

[1] C.-Y. Chan, "Advancements, prospects, and impacts of automated driving systems," *International Journal of Transportation Science and Technology*, vol. 6, no. 3, pp. 208–216, 2017.

[2] Managed Lanes: A Primer, Technical Report, 2008.

[3] N. Kalra and S. M. Paddock, "Driving to safety: how many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 182–193, 2016.

[4] S. Smith, J. Bellone, S. Bransfield et al., "Benefits estimation framework for automated vehicle operations," United States Department of Transportation, Intelligent Transportation, Technical Report, 2015.

[5] O. Hussain, A. Ghiasi, and X. Li, "Freeway lane management approach in mixed traffic environment with connected autonomous vehicles," 2016, http://arxiv.org/abs/1609.02946.

[6] R. Wang, Y. Li, and D. B. Work, "Comparing traffic state estimators for mixed human and automated traffic flows," *Transportation Research Part C: Emerging Technologies*, vol. 78, pp. 95–110, 2017.

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

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

[9] R. W. Hall and H. . J. Tsao, "Capacity of automated highway systems: merging efficiency," in *Proceedings of the 1997 American Control Conference (Cat. No. 97CH36041)*, pp. 2046–2050, Albuquerque, New Mexico, 1997.

[10] G. Arnaout and J.-P. Arnaout, "Exploring the effects of cooperative adaptive cruise control on highway traffic flow using microscopic traffic simulation," *Transportation Planning and Technology*, 2014.

[11] P. Songchitruksa, A. Bibeka, L. I. Lin, Y. Zhang et al., "Incorporating driver behaviors into connected and automated vehicle simulation," Center for Advancing Transportation Leadership and Safety (ATLAS Center), Technical Report, 2016.

[12] J. Lee, J. Bared, and B. Park, "Mobility impacts of Cooperative Adaptive Cruise Control (CACC) under mixed traffic conditions," in *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2014.

[13] H. Liu, S. E. Shladover, X.-Y. Lu, and X. Kan, "Freeway vehicle fuel efficiency improvement via cooperative adaptive cruise control," *Journal of Intelligent Transportation Systems*, pp. 1–13, 2020.

[14] H. Liu, X. Kan, S. E. Shladover, X.-Y. Lu, and R. E. Ferlis, "Modeling impacts of cooperative adaptive cruise control on mixed traffic flow in multi-lane freeway facilities," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 261–279, 2018.

[15] F. An, M. Barth, J. Norbeck, and M. Ross, "Development of comprehensive modal emissions model: operating under hot-stabilized conditions," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1587, no. 1, pp. 52–62, 1997.

[16] H. Rakha, K. Ahn, and A. Trani, "Development of VT-micro model for estimating hot stabilized light duty vehicle and truck emissions," *Transportation Research Part D: Transport and Environment*, vol. 9, pp. 49–74, 2004.

[17] A. Brooker, J. Gonder, L. Wang, E. Wood, S. Lopp, and L. Ramroth, "FASTSim: a model to estimate vehicle efficiency, cost and performance," SAE Technical Paper, Technical Report, 2015.

[18] S. M. Casey and A. K. Lund, "Changes in speed and speed adaptation following increase in national maximum speed limit," *Journal of Safety Research*, vol. 23, no. 3, pp. 135–146, 1992.

[19] C. Nowakowski, S. E. Shladover, D. Cody et al., "Cooperative adaptive cruise control: testing drivers' choices of following distances," Technical Report, 2011.

[20] S. H. Fairclough, A. J. May, and C. Carter, "The effect of time headway feedback on following behaviour," *Accident Analysis & Prevention*, vol. 29, no. 3, pp. 387–397, 1997.

[21] S. Shewmake and L. Jarvis, "Hybrid cars and HOV lanes," *Transportation Research Part A: Policy and Practice*, vol. 67, pp. 304–319, 2014.

[22] M. Chang, J. Wiegmann, A. Smith, and C. Bilotto, "A review of HOV lane performance and policy options in the United States," Federal Highway Administration, Technical Report, United States, 2008.

[23] J. A. Gomez-Ibanez, C. B. Casady, M. Fagan, J. Foote, and E. Marsh, "Toll-managed lanes: benefit-cost analyses of seven projects," Technical Report, 2018.

[24] C. Nowakowski, S. E. Shladover, and D. Cody, *Cooperative Adaptive Cruise Control: Testing Drivers' Choices of Following Distances*, California Partners for Advanced Transit and HIghways (PATH),Richmond, CA, USA, Technical Report, 2011.

[25] S. E. Shladover, C. Nowakowski, and X.-Y. Lu, *Using Cooperative Adaptive Cruise Control (CACC) to Form High-Performance Vehicle Streams Definitions, Literature Review and Operational Concept Alternatives*, Technical Report, University of California Berkeley, California Partners for Advanced Transportation Technology, 2018, https://escholarship.org/uc/item/3w6920wz.

[26] D. Swaroop, J. K. Hedrick, C. C. Chien, and P. Ioannou, "A comparision of spacing and headway control laws for automatically controlled vehicles1," *Vehicle System Dynamics*, vol. 23, no. 1, pp. 597–625, 1994.

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

[28] S. E. Shladover, C. Nowakowski, X.-Y. Lu, and R. Ferlis, "dCooperative Adaptive Cruise Control," *Transportation*

*Research Record: Journal of the Transportation Research Board*, vol. 2489, no. 1, pp. 145–152, 2015.

[29] J. Larson, C. Kammer, K.-Y. Liang, and K. H. Johansson, "Coordinated route optimization for heavy-duty vehicle platoons," in *Proceedings of the 2013 16th International IEEE Conference on Intelligent Transportation Systems-(ITSC)*, pp. 1196–1202, The Hague, Netherlands, 2013.

[30] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 829–846, 2018.

[31] Z. Zhong, "Assessing the effectiveness of managed lane strategies for the rapid deployment of cooperative adaptive cruise control technology" Ph.D. thesis, Newark, NJ, USA, 2018.

[32] X. Zhang, J. Ma, B. Smith, and J. Liu, "Operational performance evaluation of the managed lane strategy for early deployment of cooperative adaptive cruise control," Technical Report, 2018.

[33] M. A. Wright, R. Horowitz, and A. A. Kurzhanskiy, "A dynamic-system-based approach to modeling driver movements across general-purpose/managed lane interfaces," in *Proceedings of the ASME 2018 Dynamic Systems and Control Conference*, Atlanta, GA, USA, October 2018.

[34] 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*, pp. 143–156, 2016, Washington, DC, USA.

[35] Z. Zhong and J. Lee, "The effectiveness of managed lane strategies for the near-term deployment of cooperative adaptive cruise control," *Transportation Research Part A: Policy and Practice*, vol. 129, pp. 257–270, 2019.

[36] Z. Zhong and J. Lee, "Simulation framework for cooperative adaptive cruise control with empirical DSRC module," in *Proceedings of the 44th Annual Conference of the IEEE Industrial Electronics Society*, Washington, DC, USA, 2018, http://arxiv.org/abs//1810.06510.

[37] S. F. Qom, Y. Xiao, and M. Hadi, "Evaluation of Cooperative Adaptive Cruise Control (CACC) vehicles on managed lanes utilizing macroscopic and mesoscopic simulation," in *Proceedings of the Transportation Research Board 95th Annual Meeting*, Washington, DC, USA, 2016.

[38] 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.

[39] Z. Zhong, L. Joyoung, and L. Zhao, "Evaluations of managed lane strategies for arterial deployment of cooperative adaptive cruise control," in *Proceedings of the 96th Transportation Research Board Annual Meeting*, Washington, DC, USA, January 2017.

[40] A. Papadoulis, M. Quddus, and M. Imprialou, "Evaluating the safety impact of connected and autonomous vehicles on motorways," *Accident Analysis & Prevention*, vol. 124, pp. 12–22, 2019.

[41] 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.

[42] Y. Ali, Z. Zheng, and M. M. Haque, "eConnectivity's impact on mandatory lane-changing behaviour: evidences from a driving simulator study," *Transportation Research Part C: Emerging Technologies*, vol. 93, pp. 292–309, 2018.

[43] PTV Vissim, Retrieved from PTV Group, http://vision-traffic.ptvgroup.com/enus/products/ptv-vissim/use-cases/junction-geometry, 2018.

[44] 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 and Engineering Sciences*, vol. 368, no. 1928, pp. 4585–4605, 2010.

[45] M. Wang, S. van Maarseveen, R. Happee, O. Tool, and B. van Arem, "Benefits and risks of truck platooning on freeway operations near entrance ramp," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 8, pp. 588–602, 2019.

[46] 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, pp. 668–683, 2008.

[47] A. Talebpour, H. S. Mahmassani, and F. E. Bustamante, "Modeling driver behavior in a connected environment: integrated microscopic simulation of traffic and mobile wireless telecommunication systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2560, no. 1, pp. 75–86, 2016.

[48] A. Spiliopoulou, G. Perraki, M. Papageorgiou, and C. Roncoli, "Exploitation of acc systems towards improved traffic flow efficiency on motorways," in *Proceedings of the 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 37–43, Naples, Italy, June 2017.

[49] M. Guériau, R. Billot, N.-E. E. Faouzi, J. Monteil, F. Armetta, and S. Hassas, "How to assess the benefits of connected vehicles? a simulation framework for the design of cooperative traffic management strategies," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 266–279, 2016.

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

[51] S. C. Calvert, F. L. M. van Wageningen-Kessels, and S. P. Hoogendoorn, "Capacity drop through reaction times in heterogeneous traffic," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 5, no. 2, pp. 96–104, 2018.

[52] A. Sharma, Z. Zheng, J. Kim, A. Bhaskar, and M. M. Haque, "Estimating and comparing response times in traditional and connected environments," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 4, pp. 674–684, 2019.

[53] S. C. Calvert and B. van Arem, "A generic multi-level framework for microscopic traffic simulation with automated vehicles in mixed traffic," *Transportation Research Part C: Emerging Technologies*, vol. 110, pp. 291–311, 2020.

[54] J. W. C. van Lint and S. C. Calvert, "A generic multi-level framework for microscopic traffic simulation-Theory and an example case in modelling driver distraction," *Transportation Research Part B: Methodological*, vol. 117, pp. 63–86, 2018.

[55] M. Saifuzzaman, Z. Zheng, M. M. Haque, and S. Washington, "Understanding the mechanism of traffic hysteresis and traffic oscillations through the change in task difficulty level," *Transportation Research Part B: Methodological*, vol. 105, pp. 523–538, 2017.

[56] S. H. Hamdar, H. S. Mahmassani, and M. Treiber, "cFrom behavioral psychology to acceleration modeling: Calibration, validation, and exploration of drivers' cognitive and safety

parameters in a risk-taking environment," *Transportation Research Part B: Methodological*, vol. 78, pp. 32–53, 2015.

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

[58] R. Fuller, "Driver control theory: from task difficulty homeostasis to risk allostasis," in *Handbook of Traffic Psychology*Elsevier, Amsterdam, Netherlands, 2011.

[59] M. Saifuzzaman, Z. Zheng, M. Mazharul Haque, and S. Washington, "Revisiting the Task-Capability Interface model for incorporating human factors into car-following models," *Transportation Research Part B: Methodological*, vol. 82, pp. 1–19, 2015.

[60] A. Bibeka, P. Songchitruksa, and Y. Zhang, "Assessing environmental impacts of ad-hoc truck platooning on multilane freeways," *Journal of Intelligent Transportation Systems*, vol. 0, pp. 1–12, 2019.

[61] M. Killat, F. Schmidt-Eisenlohr, H. Hartenstein et al., "Enabling efficient and accurate large-scale simulations of vanets for vehicular traffic management," in *Proceedings of the Fourth ACM International Workshop on Vehicular Ad Hoc Networks*, pp. 29–38, Santa Barbara, CA, USA, 2007.

[62] D. Jiang, Q. Chen, and L. Delgrossi, "Communication density: a channel load metric for vehicular communications research," in *Proceedings of the 2007 IEEE International Conference on Mobile Adhoc and Sensor Systems*, pp. 1–8, Pisa, Italy, October 2007.

[63] M. Killat and H. Hartenstein, "An empirical model for probability of packet reception in vehicular ad hoc networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 721301, 2009.

[64] Leidos, *Simulation of Evolutionary Introduction of Cooperative Adaptive Cruise Control Equipped Vehicles into Traffic*, Saxton Transportation Operations Laboratory, Technical Report, 2016.

[65] T. Li, J. Ma, and D. K. Hale, "High-occupancy vehicle lanes on the right: an alternative design for congestion reduction at freeway merge, diverge, and weaving areas," *Transportation Letters*, vol. 12, no. 4, pp. 233–245, 2019.

[66] L. Xiao, M. Wang, and B. van Arem, "Traffic flow impacts of converting an HOV lane into a dedicated CACC lane on a freeway corridor," *IEEE Intelligent Transportation Systems Magazine*, vol. 12, pp. 60–73, 2019.

[67] Transportation Research Board, *Dedicating Lanes for Priority or Exclusive Use by Connected and Automated Vehicles*, The National Academies Press, Washington, DC, USA, 2018.

[68] L. A. Goodman, "Kolmogorov-Smirnov tests for psychological research," *Psychological Bulletin*, vol. 51, no. 2, p. 160, 1954.

[69] I. T. Young, "Proof without prejudice: use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources," *Journal of Histochemistry & Cytochemistry*, vol. 25, no. 7, pp. 935–941, 1977.

[70] M. Gouy, K. Wiedemann, A. Stevens, G. Brunett, and N. Reed, "Driving next to automated vehicle platoons: how do short time headways influence non-platoon drivers' longitudinal control?" *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 27, pp. 264–273, 2014.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8897325, 18 pages
https://doi.org/10.1155/2020/8897325

*Research Article*

# Traffic Speed Forecast in Adjacent Region between Highway and Urban Expressway: Based on MFD and GRU Model

**Yuan Gao**,[1] **Jiandong Zhao**,[2,3] **Ziyan Qin**,[2] **Yingzi Feng**,[2] **Zhenzhen Yang**,[2] **and Bin Jia**[2,3]

[1]*School of Mechanical and Electronic Control Engineering, Beijing Jiaotong University, Beijing 100044, China*
[2]*School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*
[3]*Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing, 100044, China*

Correspondence should be addressed to Jiandong Zhao; zhaojd@bjtu.edu.cn

Received 26 September 2020; Revised 15 November 2020; Accepted 20 November 2020; Published 3 December 2020

Academic Editor: Xiaoming Chen



Traffic congestion in the adjacent region between the highway and urban expressway is becoming more and more serious. This paper proposes a traffic speed forecast method based on the Macroscopic Fundamental Diagram (MFD) and Gated Recurrent Unit (GRU) model to provide the necessary traffic guidance information for travelers in this region. Firstly, considering that the road traffic speed is affected by the macroscopic traffic state, the adjacent region between the highway and expressway is divided into subareas based on the MFD. Secondly, the spatial-temporal correlation coefficient is proposed to measure the correlation between subareas. Then, the matrix of regional traffic speed data is constructed. Thirdly, the matrix is input into the GRU prediction model to get the predicted traffic speed. The proposed algorithm's prediction performance is verified based on the GPS data collected from the adjacent region between Beijing Highways and Expressway.

## 1. Introduction

With the continuous growth of the scale of China's highway network and traffic volume, the traffic load of the intercity highway in some developed cities is increasing. Many adjacent regions (network of in- and out-of-town roads) between highway and expressway have become part of the urban commuting road. Besides, due to the restrictions of traffic management measures, trucks are not allowed to enter the urban area in fixed hours. They can only drive on adjacent regions between the highway and urban expressway, which caused the trucks to accumulate, resulting in low road capacity and service level during peak hours, severe traffic congestion, and frequent traffic accidents. What is more, the abnormal weather also often leads to local traffic congestion, which gradually evolves into spread congestion in the regional network. To sum up, it is essential to conduct research on traffic states predicting at the intersection of highways and urban expressways and publish accurate traffic guidance information to travelers to alleviate traffic congestion.

The evolution of road traffic flow has complex nonlinear characteristics [1], which makes it challenging to realize accurate traffic flow predictions. Many machine learning algorithms have been used to research transportation [2], especially traffic flow prediction. Liu et al. proposed a hybrid road network traffic speed prediction model based on the state-space neural network and extended Kalman filter [3]. Zhang et al. predicted the traffic speed considering the heterogeneity of different roads [4]. Dhivyabharathi et al. proposed a method for predicting river traffic time using the particle filtering method [5]. Zhao et al. integrated the charging data and microwave detection data to predict traffic speed [6]. Zhao et al. proposed a prediction algorithm combining equal spacing interpolation and Sage–Husa adaptive Kalman filtering [7]. Wang et al. improved the reliability method of driving time prediction based on GPS point velocity distribution by calculating the

variable velocity distribution coefficient [8]. Zhou et al. proposed a recurrent neural network based microscopic car-following model on predicting traffic oscillation [9]. Wang and Goodchild developed a logit model to determine the truck route's influencing factors and estimate the driving time [10]. Jula et al. developed a hybrid method composed of dynamic programming and genetic algorithm to find trucks' shortest path [11]. Dong et al. proposed a traffic crash prediction method based on the support vector regression (SVR) model [12]. Yu et al. proposed the random forests based on the near neighbor (RFNN) method to predict bus travel time [13]. Xie and Wei proposed Elman neural network to predict truck speed [14]. Wang and Xu constructed the short-term traffic flow prediction model of urban expressway based on the Long Short-Term Memory (LSTM) network under deep learning [15]. Luo et al. proposed a short-term traffic flow prediction model based on deep learning combining the features of convolutional neural network and support vector regression classifier [16]. Yao et al. discussed the application of Support Vector Machine theory to predict road travel time [17]. Wu et al. [18] proposed a traffic flow prediction model based on Deep Neural Networks (DNN) by utilizing the weekly/daily periodicity and space-time characteristics of traffic flow [18]. Jia et al. proposed a deep learning method for short-term traffic speed information prediction-deep belief network (DBN) model [19]. Due to the problem that the commonly used weight optimization algorithm could not adjust the learning rate adaptably, Zhao et al. adopted Adam, Adadelta, and Rmsprop to optimize the weight in the GRU model of the deep learning algorithm [20]. Wang et al. established a travel time prediction model based on the LSTM (Long Short-Term Memory) considering the precipitation data [21]. Zang et al. proposed an all-day traffic speed prediction method for elevated highways based on deep learning [22]. Peng et al. proposed a 3D Convolutional Neural Network-Deep Neural Network method to recognize and predict traffic status from aerial videos [23]. The author has also conducted many studies on how to improve prediction accuracy [24–26]. The existing machine learning algorithm cannot fully dig out the essence of traffic flow characteristics. Deep learning model, such as GRU, can help us learn and seize the inherent complex features effectively and predict traffic flow without prior knowledge [27]. Using the deep learning algorithm to mine traffic flow rules becomes the direction of traffic state prediction.

The prediction of road segment traffic speed belongs to the study of microscopic traffic flow characteristics. It is also affected by the traffic states at the macrolevel, such as road network and neighboring region. At present, it is rare to combine macro- and microcharacteristics to predict traffic speed. The Macroscopic Fundamental Diagram (MFD) is a model reflecting the road network's macroscopic traffic state. According to specific indicators, MFD divides a complex and large road network into several independent subregions and implements appropriate control optimization strategies according to the subareas' characteristics. Based on the subregion division results, several road sections that are most similar to the predicted target road can be selected. Meanwhile, the spatial and temporal correlation between traffic flows in each subregion can be analyzed. The subregions with strong correlation can be selected to construct the traffic flow sequence dataset and input the prediction model, which is conducive to improving the prediction accuracy. Many scholars have studied the subregional division method of MFD. Ji and Geroliminis proposed a static normalized cut (Ncut) based subregions division method based on traffic congestion's spatial characteristics and minimizing the subarea's vehicle density [28]. Ji et al. also proposed a dynamic subarea delineation method based on GPS data targeting the maximum connected element [29]. Haddad and Geroliminis proposed a division method based on the operational stability of the subregions [30]. Ma et al. used a spectral approach to divide traffic zones based on neighboring intersections [31]. Ncut is a graph theory-based partitioning algorithm derived from the image partitioning domain. This algorithm not only considers the similarities within regions but also normalizes the similarities within regions using the similarities between regions. And then, the cutting scheme that minimizes the similarities between regions after normalization is found. In this paper, the Ncut algorithm is used to divide the traffic subregions. Then, the stability of the MFD for each subdivided subregion is calculated and analyzed to justify the division results.

GPS data has a wide coverage area and can better reflect the characteristics of urban road traffic flow [32]. What is more, in recent years, China has strengthened the supervision of freight vehicles. It is required that the GPS devices be installed on the large heavy-haul trucks to monitor the trucks' running status. This produces a large amount of trajectory data, especially in the adjacent region between highway and expressway. In this paper, based on the average roadway speed and flow data extracted from the truck GPS data, a short-time traffic flow prediction method combining MFD and GRU is proposed. Using the characteristics of MFD, the road network area is divided into subregions, and the microtraffic flow characteristics and macrotraffic conditions are combined to develop a traffic forecasting method. The test results of real traffic flow data show that the method proposed in this paper has lower prediction errors and higher accuracy than the existing prediction models. It is a reasonable and effective method to predict short-time traffic flow. The technical framework of this paper is shown in Figure 1.

## 2. Subdivision Method of Road Network Based on MFD

*2.1. Construction of Road Network Weighted Graph Based on Traffic Operation Similarity.* A stable MFD exists in a network of roads with operational homogeneity. A large area can be divided into subregions based on the operational homogeneity of traffic. The starting point of the MFD theory is to study the relationship between traffic demand and traffic supply in the road network, the maximum traffic volume can directly reflect the traffic supply of each road section and the overall road network, and the traffic volume data can be easily obtained through traffic flow detection. Thus, the traffic volume is taken as the fundamental traffic characteristic of the road section in this paper. The road

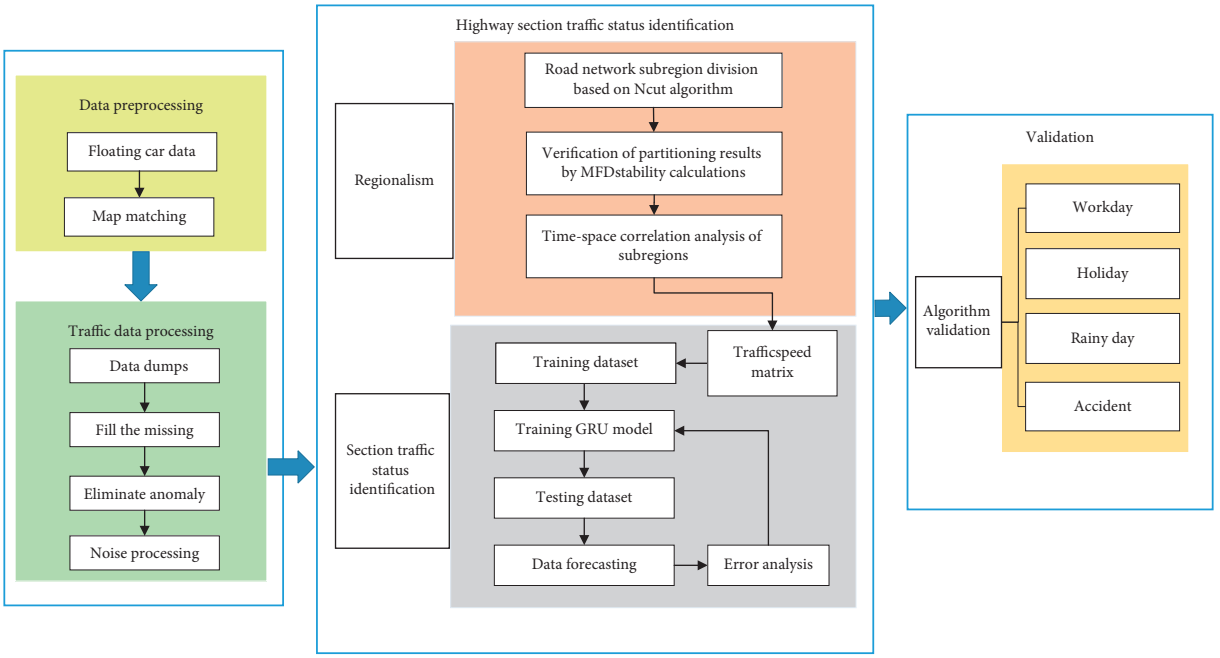



Figure 1: The technical framework of the paper.

section's maximum traffic volume is used to define the traffic operation similarity between adjacent connected sections.

Let the similarity degree of traffic operation between adjacent connected sections $i$ and $j$ in road network $G$ be:

$$w(i,j) = \exp\left[-\left(q_i^m - q_j^m\right)^2\right], \tag{1}$$

where $q_i^m$ is the maximum traffic volume of section $i$ in road network $G$ and $q_j^m$ is the maximum traffic volume of section $j$ in road network $G$. Using the natural constant transformation, the difference value of the maximum traffic volume between adjacent connected sections is squared mapped to the interval of 0-1. If the similarity is 0, the traffic operation similarity between the road sections is the least. If the similarity is 1, the traffic operation similarity between the road sections is the biggest.

Based on graph theory, the road network is first "node-arc transformation," so that the similarity degree of traffic operation between adjacent road sections is expressed as the weight of the arc section in the graph. The Laplace matrix is constructed based on the similarity degree of traffic operation, which is the basis for the subgraph division in graph theory. The details are as follows.

When road segment $i$ in road network $G$ is connected with road segment $j$, $a_{ij} = 1$. When segment $i$ is disconnected from segment $j$, $a_{ij} = 0$. When $i$ is equal to $j$, $a_{ij} = 0$. The weighted adjacency matrix of road network $G$ is $W$, and the element in $W$ is

$$w_{ij} = \begin{cases} w'(i,j), & a_{ij} = 1, \\ 0, & a_{ij} = 0. \end{cases} \tag{2}$$

This paper adopts road sections as node $V$ in the undirected graph $G$, so that $w'(i,j) = -w(i,j)$. The matrix $D$ is the diagonal matrix:

$$D = \operatorname{diag}\{d_i\}, \quad d_i = \sum_j w'(i,j). \tag{3}$$

$D$-$W$ is the Laplace matrix of road network $G$, in which the sum of all the rows and columns is zero. Based on the transformations and calculations, it is possible to obtain a road network weighted by traffic operations similarity between adjacent connecting sections.

*2.2. Road Network Subarea Division Method Based on Normalized Cut.* Ncut is one of the neutron graph partition methods in graph theory, which is a subgraph division method at the macrolevel. The focus is not on the graph's details, but on the overall characteristics of the graph. The optimal normalized cut problem of the graph can be expressed as

$$\min_x \operatorname{Ncut}(x) \quad \text{s.t.} \quad x \in V. \tag{4}$$

It is an NP-hard problem to solve the minimum value of the normalized cut. The spectral clustering method is a widely used method that can solve NP-hard problem approximately by solving eigenvalue and eigenvalue vector. Therefore, the Fiedler method is used to calculate the

eigenvalues and eigenvectors of the matrix and divide the subareas of the road network. The point set $V$ of figure $G$ is divided into two subsets, and the transformation can express the optimal normalized cut problem of $A$ and $B$:

$$\begin{aligned} &\min_x \mathrm{Ncut}(x) = \min_y \frac{y^T(D-W)y}{y^T D y} \\ &x_i = 1, \quad \text{if } x_i \in A \\ &x_i = -1, \quad \text{if } x_i \in B \\ &\text{s.t.} \\ &y_i \in \{-1, b\} \\ &y^T D 1 = 0, \end{aligned} \tag{5}$$

where $y = (1+x) - b(1-x)$, $b = ((\sum_{x_i>0} d_i)/(\sum_{x_i<0} d_i))$, $d_i = \sum_j w_{ij}$, and $x$ is a vector of columns consisting of 1 and $-1$. When the $i$-th node is subdivided into subregion $A$, $x = 1$; when it is subdivided into subregion $B$, $x = -1$.

Since all rows and columns in a Laplace matrix have a sum of 0, the matrix always has an eigenvalue of 0. If graph $G$ is connected, then the second small eigenvalue is positive. The corresponding eigenvector is called the Fiedler vector, which contains important information about the graph; that is, the numerical size of the elements in the Fiedler vector reflects the correlation of their corresponding vertices. When the road network is divided according to the Fiedler vector, the vertices corresponding to the Fiedler vector can be divided according to different critical value $S$. There are many methods to select the $S$ value, among which the 0-point method is practical and straightforward.

## 3. Stability Calculation of MFD

By calculating the stability of MFD, the rationality of the subregion division can be proved. MFD stability depends on stability in the critical state. In the critical state, if the average traffic volume fluctuates less under the same road network density, the road network traffic operation will be more controllable. This article refers to the method (Fuzzy c-means algorithm) in our research [33] to divide the test data sets into three categories: unblocked, critical, and congested. Two indicators, the regional traffic volume and regional density, are used to classify the three traffic states. Firstly, FCM (Fuzzy c-means algorithm) is used to divide the data points of the spatial distribution of multidimensional data into specific classes. Each data point belongs to a certain class to some extent, and the membership degree is used to indicate the degree to which each data point belongs to a certain clustering. FCM divides $n$ vectors into $c$ fuzzy groups and calculates each group's clustering center to minimize the objective function of nonsimilarity indexes. Then, the traffic state is divided into three stages: unblocked, critical, and congested.

The dispersion of road network traffic operation, that is, the dispersion of weighted average traffic volume of road network in the critical state, represents MFD's stability in the critical state. The lower the dispersion of road network traffic operation is, the more stable the road network operation is, and the higher the MFD stability is. The higher the dispersion of road network traffic operation, the more unstable the road network operation will be and the lower the MFD stability. The dispersion of road network traffic operation is

$$\begin{aligned} s &= \frac{\sqrt{(1/n)\sum_{e=1}^{n}\left(q_e^w - q_c^w\right)}}{q_c^w}, \\ q_c^w &= \alpha q_m^w, \end{aligned} \tag{6}$$

where $q_e^w$ is the average traffic volume in the test data. $q_c^w$ is the critical average traffic volume of road network, $q_m^w$ is the maximum average traffic volume of road network, and $\alpha$ is undetermined parameters, $0 < \alpha < 1$.

The whole road network's dispersion degree is calculated by the weighted average of each subarea's dispersion degree. The calculation results can be used as the judgment index of the subarea division to characterize the whole network's MFD stability. If the entire road network is divided into $N$ subareas, and the dispersion of road network traffic operation in subarea $i$ is $s_i$, then the dispersion of the whole road network traffic operation is $S_N$

$$S_N = \frac{\sum_{i=1}^{N} q_{\mathrm{ci}}^w s_i}{\sum_{i=1}^{N} q_{\mathrm{ci}}^w}, \tag{7}$$

where $q_{\mathrm{ci}}^w$ is the critical average traffic volume in subarea $i$.

## 4. Traffic Speed Prediction Method Based on the Spatial-Temporal Correlation of Subareas

The evolution of traffic speed on a road section in a specific subarea is affected not only by the temporal evolution law of the traffic flow on the road sections but also by the spatial influence of the road sections in other subareas. This paper proposes a traffic speed prediction algorithm that considers the spatial-temporal correlation of subareas.

### 4.1. Correlation Analysis of Subareas

*4.1.1. Spatial Correlation Analysis.* Firstly, the spatial correlation between each subarea is analyzed.

In spatial correlation analysis, it is necessary to measure the adjacency relationship of the neighboring subregions. This requires quantitatively describing the adjacency relationship of adjacent regions to perform the calculation of spatial correlation statistics.

In this paper, the spatial adjacency matrix is used to express the spatial relationship between subregions.

Suppose that there are $m$ subregions in the study area, and the spatial weight matrix $W_{\mathrm{sp}} = [w_{ij}^{\mathrm{sp}}]_{m\times m}$ is used to express the spatial correlation between the subregions:

$$w_{ij}^{\mathrm{sp}} = \begin{cases} 1, & \text{when subregions } i \text{ and } j \text{ are adjacent,} \\ 0, & \text{when subregions } i \text{ and } j \text{ are not adjacent,} \end{cases} \tag{8}$$

where $W_{\mathrm{sp}}$ is $m \times m$ dimensional spatial weight matrix and $w_{ij}^{\mathrm{sp}}$ is the spatial weight between the regional units $i$ and $j$.

Besides, to ensure that the subregions cannot be adjacent to themselves specifies that when $j = i$, $w_{ij}^{sp} = 0$. When two subregions are sharing one or more nodes, it is adjacent.

The standardized formula of the spatial weights is

$$\omega_{ij}^{sp} = \frac{w_{ij}^{sp}}{\sum_j w_{ij}^{sp}}. \tag{9}$$

*4.1.2. Temporal Correlation Analysis.* The Pearson correlation coefficient formula is improved to measure the time correlation of the two regions. If two subregions in the study area have spatial adjacencies, the time correlation can be calculated by the following formula over a certain period:

$$\gamma_{ij} = \frac{E\left(q_i(t) - \mu_i\right)\left(q_j(t) - \mu_j\right)}{\sigma_i \sigma_j}, \tag{10}$$

where $q_i(t)$ and $q_j(t)$ are the traffic volume of subregions $i$ and $j$ at time $t$; $u_i$ and $u_j$ are the mean traffic volume for $i$ and $j$; and $\sigma_i$ and $\sigma_j$ are the variances.

According to the road network area studied in this paper, the regional correlation is calculated as $\omega_{ij}^{sp}\gamma_{ij}(\mathrm{ac})$, where $\omega_{ij}$is the spatial correlation of areas $i$ and $j$ and $\gamma_{ij}(\mathrm{ac})$ is the correlation of two areas in the study period.

In this paper, the data of the $K$ most relevant regions to the region where the predicted target segment is located are selected for constructing the input matrix of the prediction model. In this paper, the $K$ values are determined as follows: $K$ ($K = 0, 1, ..., 4$) are used to input the data from the most relevant regions into the GRU model, and the $K$-value with the smallest prediction error is taken.

One input sample of the deep learning algorithm is

$$\begin{bmatrix} V_{11}(t-1) & \cdots & V_{1(o-1)}(t-1) & V_{21}(t-1) & \cdots & V_{2o}(t-1) & \cdots & V_{ko}(t-1) & V_a(t-1) \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ V_{11}(t-L) & \cdots & V_{1(o-1)}(t-L) & V_{21}(t-L) & \cdots & V_{2o}(t-L) & \cdots & V_{ko}(t-L) & V_a(t-L) \end{bmatrix}, \tag{11}$$

where $L$ is timestep, $o$ is the number of sections of road included in Area $k$, and $a$ represents the target section.

*4.2. GRU-Based Traffic Speed Prediction Algorithm.* RNN (recurrent neural network) is a kind of deep neural network designed to process sequence data, which plays an important role in the field of sequence mining.

The GRU model is an improvement of recurrent neural network, which is one of the hot technologies of deep learning in recent years. Different from the traditional recurrent neural network, the internal structure of the GRU's hidden layer nodes does not use a single activation function.

The specific calculation steps of GRU are as follows: firstly, the current state input $z_t$ and the previous time output $h_{t-1}$ are input into the update gate, and then a value between 0 and 1 can be output, where 0 represents the complete discarding information and 1 represents the complete reserving information, and the calculation formula is as shown in formula (12). Secondly, $z_t$ and $h_{t-1}$ entering the reset gate of the sigmoid layer output the value between 0 and 1. Meanwhile, tanh layer will create a new candidate value vector $\widetilde{h}_t$, and the calculation formulas are shown in equations (13) and (14). Thirdly, the update gate is used as the weight vector, and the candidate vector and the output vector at the last moment are weighted averages to obtain the output $h_t$ of GRU cells. The calculation formula is shown in equation (15):

$$\eta_t = \sigma\left(p_\eta + U_\eta z_t + Q_\eta h_{t-1}\right), \tag{12}$$

$$r_t = \sigma\left(p_r + U_r z_t + Q_r h_{t-1}\right), \tag{13}$$

$$\widetilde{h}_t = \tanh\left(p_s + U_s z_t + Q_s r_t h_{t-1}\right), \tag{14}$$

$$h_t = \left(1 - \eta_t\right)h_{t-1} + \eta_t \widetilde{h}_t, \tag{15}$$

where $\eta$ represents the update gate vector; $r$ represents the reset gate vector; $p$ represents the bias vector; $U$ represents the input weight; $Q$ represents the cyclic weight; $z_t$ represents the input vector at $t$ time; and $h_t$ represents the output vector at $t$ time.

Regularization is generally defined as the modification of the learning algorithm, and the goal is to reduce generalization error rather than training error. Common regularization methods include *L1* and *L2* parameter paradigm penalty, Dropout, multitask learning, and early termination, which are common, where the penalty terms *L2* and *L1* of parameter normal form can be expressed as

$$L2(\theta) = \frac{1}{2}\partial \sum_i |\theta_i^2|, L1(\theta) = \partial \sum_i |\theta_i|, \quad (16)$$

where $\theta_i$ can be expressed as the reciprocal of weight $Q$ of each layer, indicating that, for the layer with too high weight learned, its updating degree should be reduced. On the contrary, for the node with too low weight learned in the layer, its updating degree should be improved to achieve the goal of amortizing the ownership value in the layer.

To sum up, the flow of the stroke speed prediction algorithm proposed in this paper is shown in Figure 2. The input is a three-dimensional vector composed of features, time step, and samples. This 3D tensor is input to the GRU model with a dropout layer and fully connected layer to get the travel time's predicted result. One column of the matrix in equation (11) corresponds to the input of one time step of the GRU model.

## 5. Algorithm Validation

*5.1. Road Network Subarea Division and Stability Calculation of MFD.* This paper uses truck GPS data as the basis for algorithm validation. As shown in Figure 3, the experimental area selected in this paper is located in Beijing's southeast. Beijing's expressway and the main road, including the 5$^{th}$ and 6$^{th}$ ring roads, Jingtai, Jinghu, and Jingha highway, are selected to verify algorithm accuracy. The area is approximately 110 square kilometers. The total length of roads in the road network is about 131 km. According to the analysis, the selected area has more accidents and more GPS data of truck.

The time range of the data used for validation is May 1, 2018, to July 31, 2018.

The methods of map matching, anomaly data processing, and traffic speed time series extraction of truck GPS data in this paper are from literature [20]. In this paper, the collected truck GPS data is organized into the form of a time series of the traffic speed of the road section with a period of 5 minutes. Then, according to the chosen $K$-value, it is organized into equation (11). $L$ takes 12; that is, the prediction is made using the previous hour's data.

Sample size is a critical concern when using probe vehicles to collect real-time traffic information, and it is necessary to determine the number of probe vehicles needed for traffic state estimation. In this paper, the required sample size for different combinations of confidence levels of the study area is determined with reference to the method in [34].

May and June's average speed data are used as the training data set for GRU model training. The rest of the data serves as a test set for the algorithm. In this paper, the study area's road network is abstracted into the road network diagram, as shown in Figure 4, and there are 32 road sections and 21 nodes.

The regional division of the cases of the parties is shown in Figure 5. The dispersion of traffic operation in subareas and the whole road network is shown in Table 1. When the network is divided into 5 subareas, the whole road network's traffic discrete degree is the smallest, 0.05673. The network has been divided into smaller regions. If continuing, the change of discrete degree of the whole road network is not too big. However, for speed prediction, the dimensionality of the data input to the model will increase. Thus, the prediction difficulty will increase. So, take five as the optimal scheme of areas division.

The MFD of each subarea is shown in Figure 6. The traffic state classification results based on the FCM algorithm for subarea 1 are shown in Figure 7. The clustering centers are shown in Table 2.

*5.2. Traffic Speed Prediction.* Two indexes, MAPE and RMSE, are selected to evaluate the prediction accuracy of the model. The calculation method of MAPE and RMSE is shown in the following formulas:

$$\text{MAPE} = \frac{1}{L}\sum_L \frac{|V_Y(t) - V(t)|}{V(t)}, \quad (17)$$

$$\text{RMSE} = \sqrt{\frac{1}{L}\sum_L (V_Y(t) - V(t))^2}, \quad (18)$$

where $V_Y(t)$ is the predicted traffic speed at time $t$, $V(t)$ is the actual traffic speed at time $t$, and $L$ is the total number of predicted cycles.

This paper selects nodes 17 to 18 (Section 1) and 9 to 10 (Section 2) belonging to different regions as the experimental verification sections. The accuracy of the algorithm was verified in four different scenarios. The prediction results were compared with the GRU prediction algorithm based on a single time series of the road segment. This GRU model has the same parameter settings as the model presented in this paper.

The first step is to determine the number of regions $K$ that are input to the GRU model, so the relationship between the number of inputs and the prediction accuracy of the model is analyzed. Table 3 shows the prediction accuracy of Section 1 for different $K$ values. Table 4 shows the prediction accuracy of Section 2 for different $K$ values. From Table 3, it can be seen that the $K$ of road Section 1 take 1. From Table 4, it can be seen that the K of road Section 2 take 1.

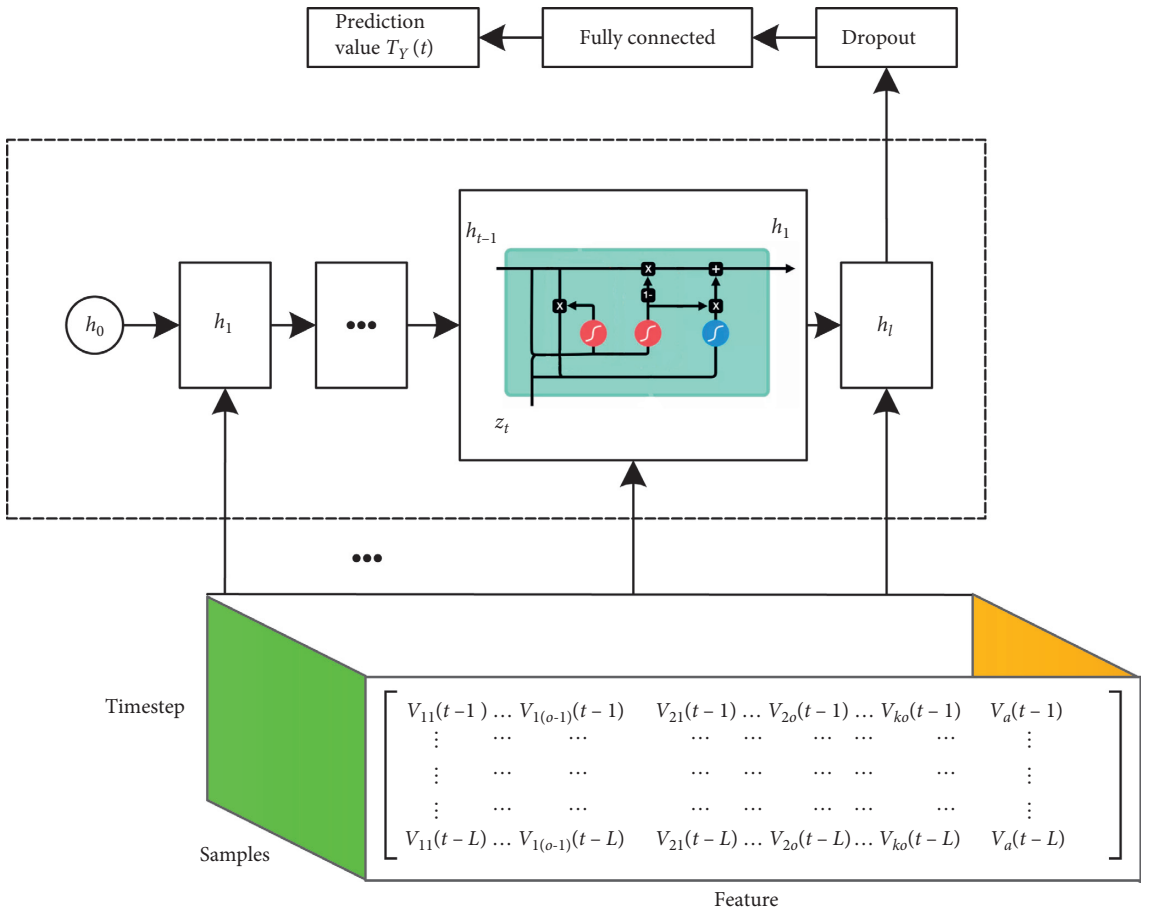



Figure 2: GRU algorithm framework.

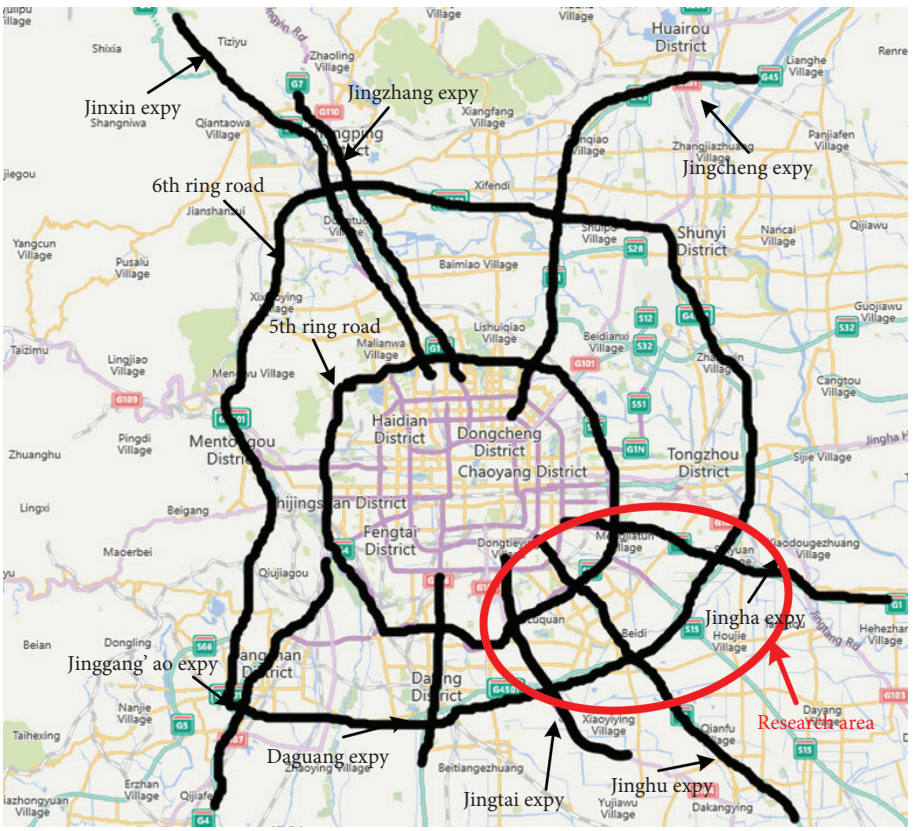



Figure 3: Adjacent region between the highway and urban expressway.

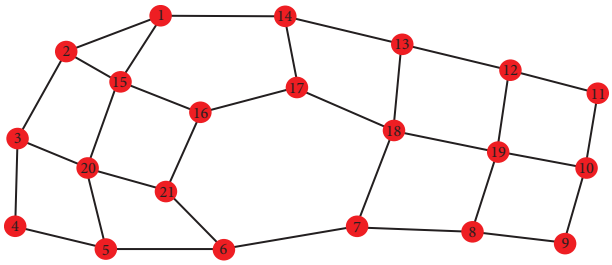


Figure 4: Node number of the road network.

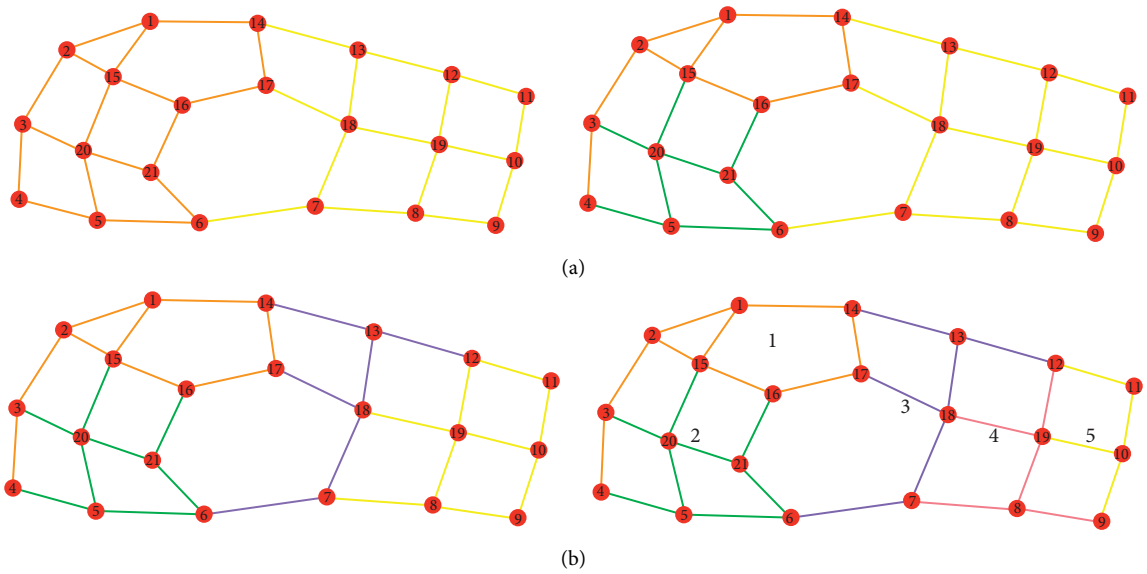



Figure 5: Subregional division of each scheme.

Table 1: Dispersion of traffic flow of each scheme.

| Subarea number | Transportation dispersion in subareas | Dispersion of traffic flow in the whole road network |
|---|---|---|
| 2 | 0.08350<br>0.07194 | 0.06334 |
| 3 | 0.05627<br>0.05957<br>0.06334 | 0.06130 |
| 4 | 0.05627<br>0.05957<br>0.03872<br>0.11177 | 0.06499 |
| 5 | 0.05627<br>0.05957<br>0.03872<br>0.07729<br>0.06224 | 0.05673 |

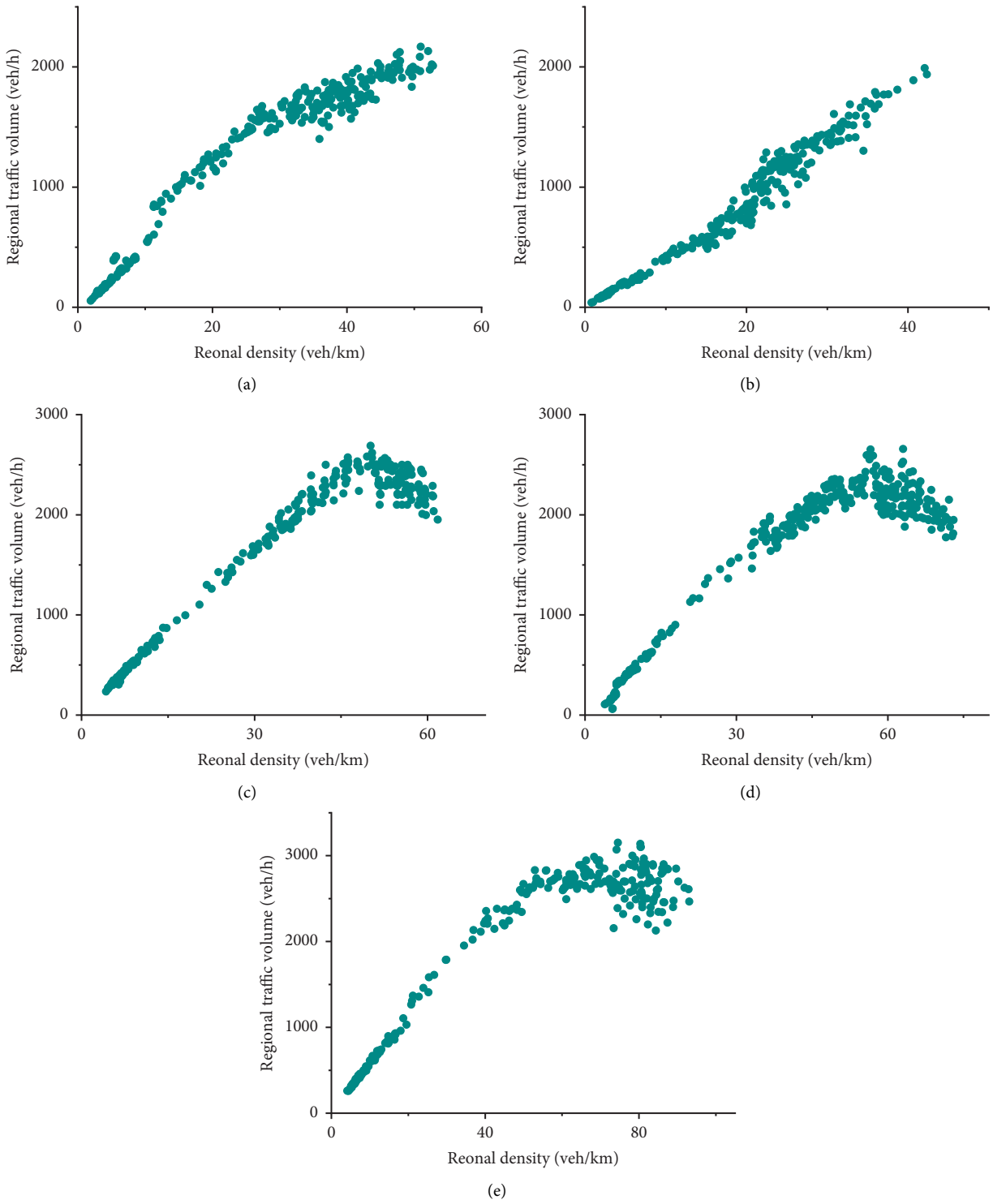



Figure 6: MFD of each subregional. (a) Subregional one. (b) Subregional two. (c) Subregional three. (d) Subregional four. (e) Subregional five.

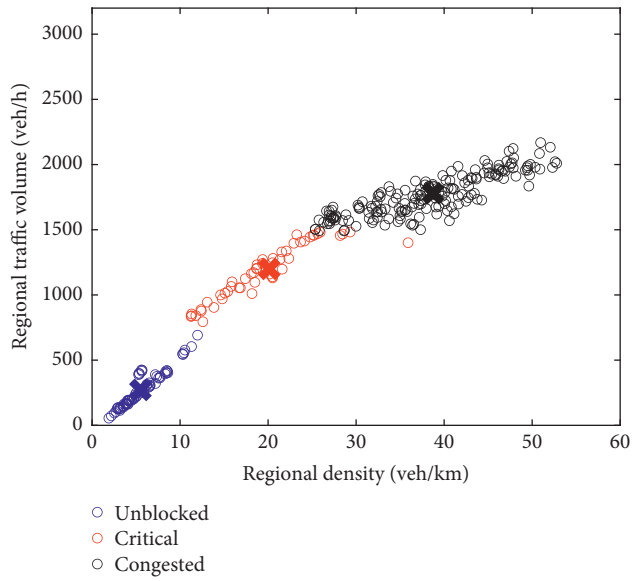



Figure 7: The traffic state classification results for subarea 1.

Table 2: The clustering centers.

| Traffic state | Regional density | Regional traffic volume |
|---|---|---|
| Unblocked | 5.49 | 272 |
| Critical | 20.13 | 1200 |
| Congested | 38.74 | 1780 |

Table 3: Prediction accuracy of road Section 1 for different K values.

| | GRU | $K=0$ | $K=1$ | $K=2$ | $K=3$ | $K=4$ |
|---|---|---|---|---|---|---|
| MAPE (%) | 4.15 | 4.275 | 4.018 | 4.120 | 4.49 | 4.667 |
| RMSE | 2.331 | 2.251 | 2.217 | 2.294 | 2.402 | 2.42 |

Table 4: Prediction accuracy of road Section 2 for different K values.

| | GRU | $K=0$ | $K=1$ | $K=2$ | $K=3$ | $K=4$ |
|---|---|---|---|---|---|---|
| MAPE (%) | 4.003 | 4.016 | 3.903 | 4.009 | 4.16% | 3.955 |
| RMSE | 3.136 | 3.122 | 2.97 | 3.051 | 3.142 | 3.145 |

*5.2.1. Working Days*

*(1) Section 1.* The predicted results of Section 1 on July 2 (working day) are shown in Figure 8. The errors are shown in Figure 9. It can be seen that on July 2, the average speed of the morning rush hour and noon rush hour sections is low, in a congested state, and the road section was in a state of smooth flow at night. The algorithm proposed in this paper can achieve acceptable prediction results. The results of error evaluation indicators MAPE and RMSE are shown in Table 5. MAPE was 2.30%, 3.05%, and the RMSE was 1.34 and 1.68, respectively.

*(2) Section 2.* The predicted results of Section 2 on July 2 (working day) are shown in Figure 10. The errors are shown

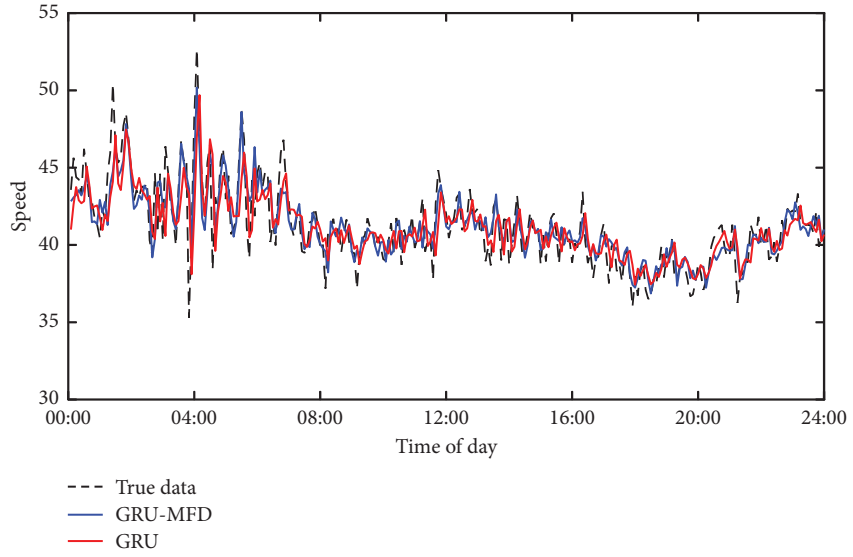



Figure 8: Prediction results on July 2, 2018, Section 1.

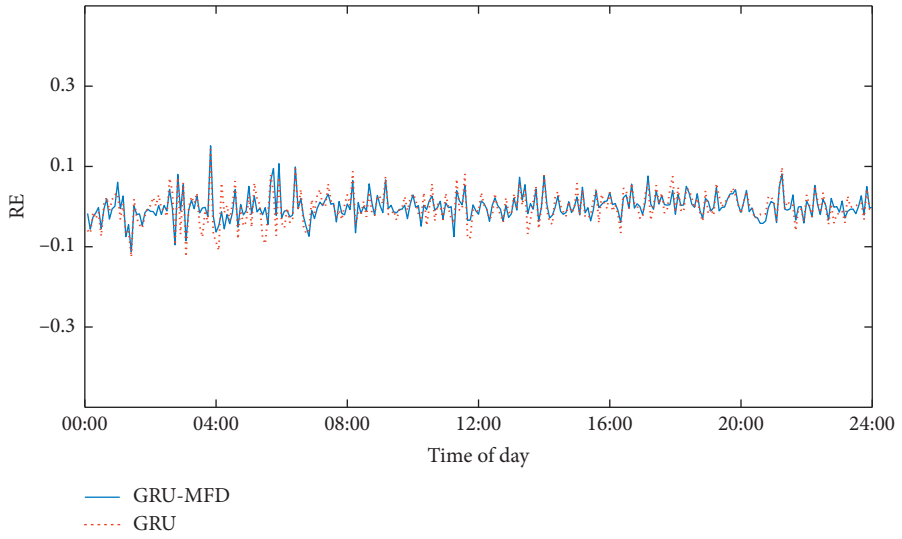



Figure 9: Errors on July 2, 2018, Section 1.

Table 5: Section 1 prediction error.

| | Working day | | Weekend | | Rainy day | |
|---|---|---|---|---|---|---|
| | 7.2 | | 7.1 | | 7.5 | |
| Evaluation index | MAPE | RMSE | MAPE | RMSE | MAPE | RMSE |
| GRU-MFD | 2.30 | 1.34 | 3.93 | 2.17 | 2.86 | 1.65 |
| GRU | 3.05 | 1.68 | 4.25 | 2.31 | 3.23 | 1.86 |

in Figure 11. The algorithm proposed in this paper can achieve acceptable prediction results. The results of error evaluation indicators MAPE and RMSE are shown in Table 6. MAPE was 1.42%, 2.53%, and the RMSE was 1.05 and 1.79, respectively.

*5.2.2. Weekend. (1) Section 1.* The predicted results of Section 1 on July 1 (weekend) are shown in Figure 12. The errors are shown in Figure 13. The algorithm proposed in this paper can achieve acceptable prediction results. The results of error evaluation indicators MAPE and RMSE are

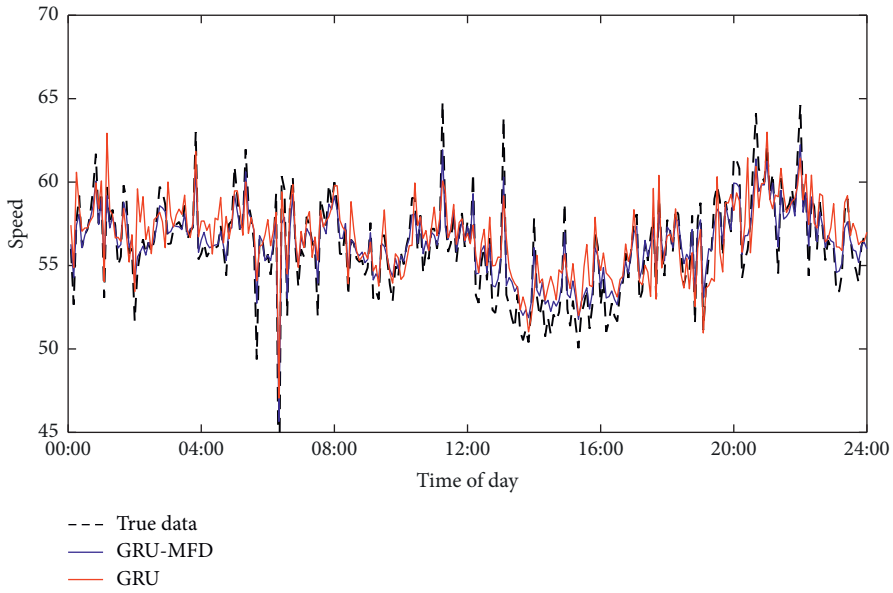



Figure 10: Prediction results on July 2, 2018, Section 2.

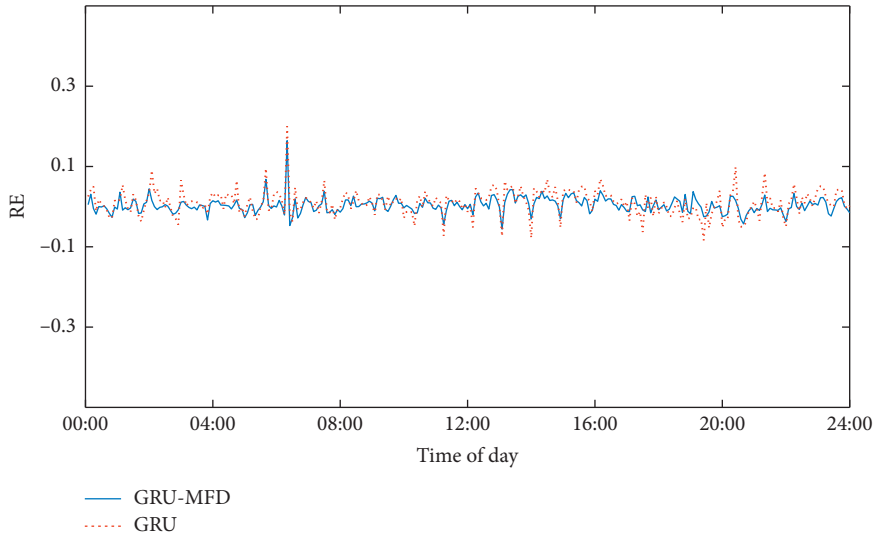



Figure 11: Errors on July 2, 2018, Section 2.

shown in Table 5. MAPE was 3.93%, 4.25%, and the RMSE was 2.17 and 2.31, respectively.

*(2) Section 2.* The predicted results of Section 2 on July 1 (weekend) are shown in Figure 14. The errors are shown in Figure 15. The algorithm proposed in this paper can achieve acceptable prediction results. The results of error evaluation indicators MAPE and RMSE are shown in Table 6. MAPE was 2.46%, 3.93%, and the RMSE was 1.89 and 2.18, respectively.

*5.2.3. Rainy Day. (1) Section 1.* The predicted results of Section 1 on July 5 (rainy day) are shown in Figure 16. The errors are shown in Figure 17. The algorithm proposed in this paper can achieve acceptable prediction results. The results of error evaluation indicators MAPE and RMSE are shown in Table 5. MAPE was 2.86%, 3.23%, and RMSE were 1.65 and 1.86, respectively.

*(2) Section 2.* The predicted results of Section 2 on July 5 (rainy day) are shown in Figure 18. The errors are shown in

Table 6: Section 2 prediction error.

| | Working day | | Holiday | | Rainy day | | Accident | |
|---|---|---|---|---|---|---|---|---|
| | 7.2 | | 7.1 | | 7.5 | | 7.11 | |
| Evaluation index | MAPE | RMSE | MAPE | RMSE | MAPE | RMSE | MAPE (%) | RMSE |
| GRU-MFD | 1.42 | 1.05 | 2.46 | 1.89 | 1.49 | 1.03 | 7.35 | 5.05 |
| GRU | 2.53 | 1.79 | 3.93 | 2.18 | 2.55 | 1.82 | 8.94 | 5.83 |

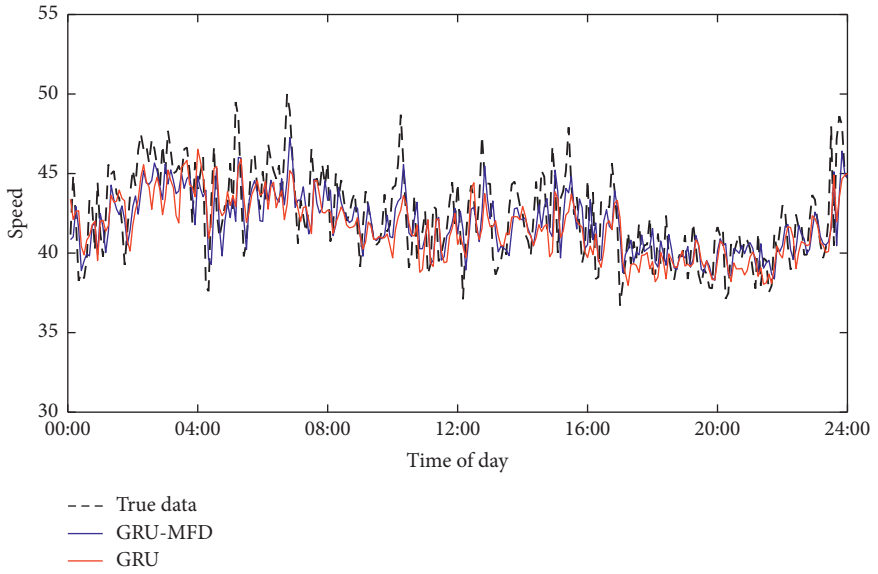



Figure 12: Prediction results on July 1, 2018, Section 1.

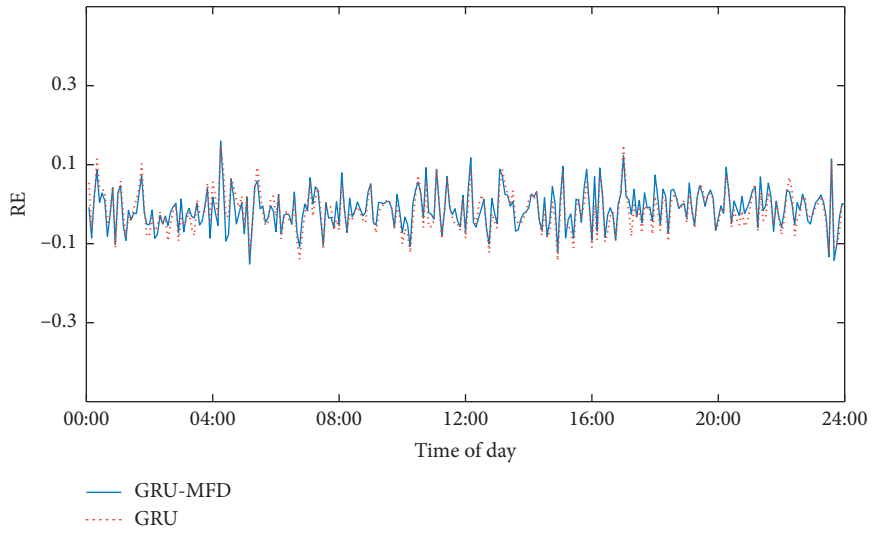



Figure 13: Errors on July 1, 2018, Section 1.

Figure 19. The algorithm proposed in this paper can achieve acceptable prediction results. The results of error evaluation indicators MAPE and RMSE are shown in Table 6. MAPE was 1.49% and 2.55%, and the RMSE was 1.03 and 1.82, respectively.

*5.2.4. Accident.* The predicted results of Section 2 on July 11 are shown in Figure 20. The errors are shown in Figure 21. There was a traffic accident in the early hours of the morning. The results of MAPE and RMSE are shown in Table 6. MAPE was 7.35% and 8.94%, and the RMSE was 5.05 and 5.83, respectively.

It can be seen from the above figures and tables that, compared with the traditional prediction method based on single segment time series, the prediction accuracy of the proposed algorithm is improved. When no accident happened,

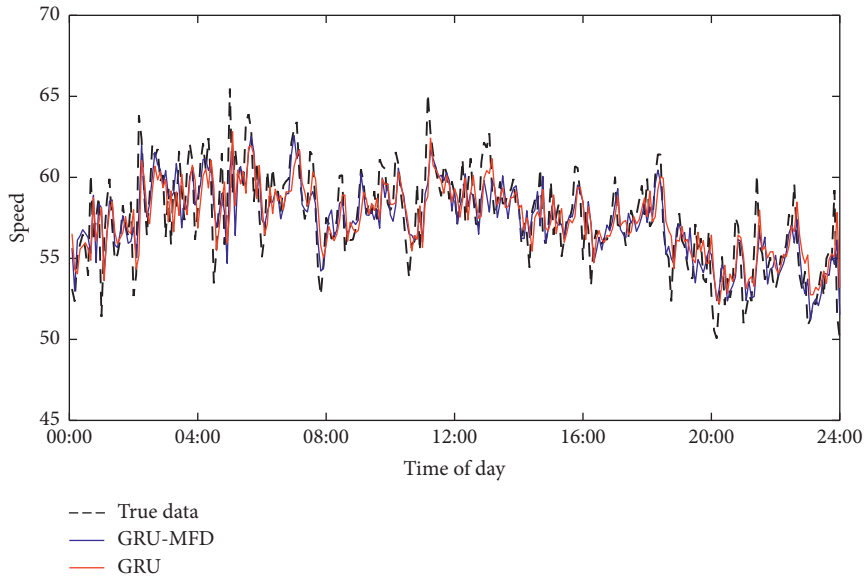



Figure 14: Prediction results on July 1, 2018, Section 2.

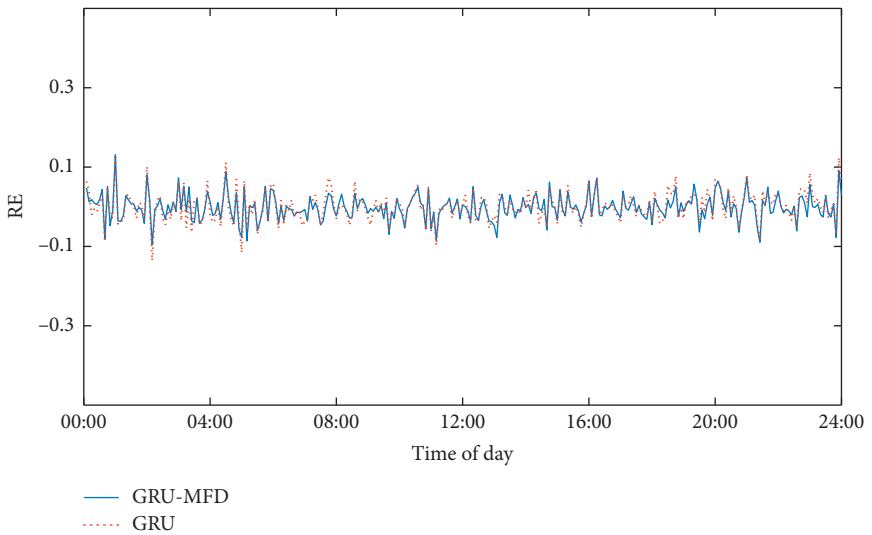



Figure 15: Errors on July 1, 2018, Section 2.

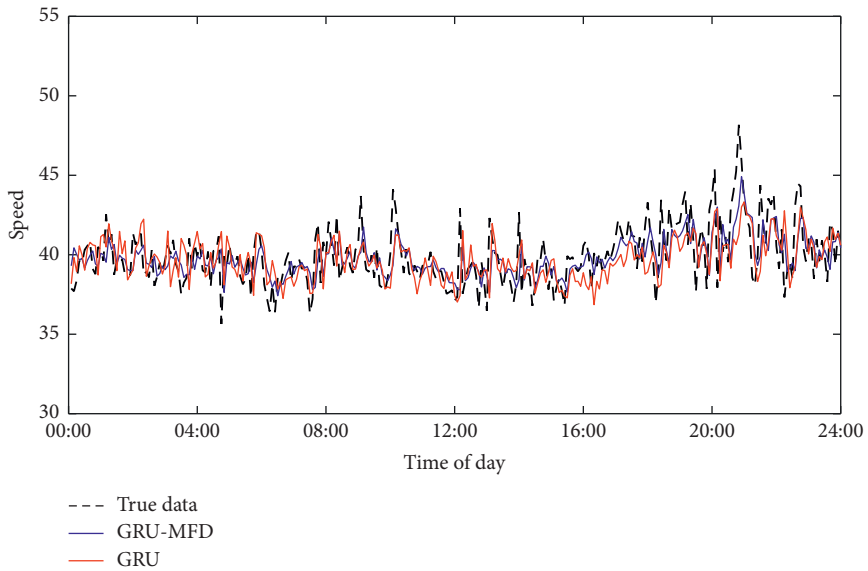



Figure 16: Prediction results on July 5, 2018, Section 1.

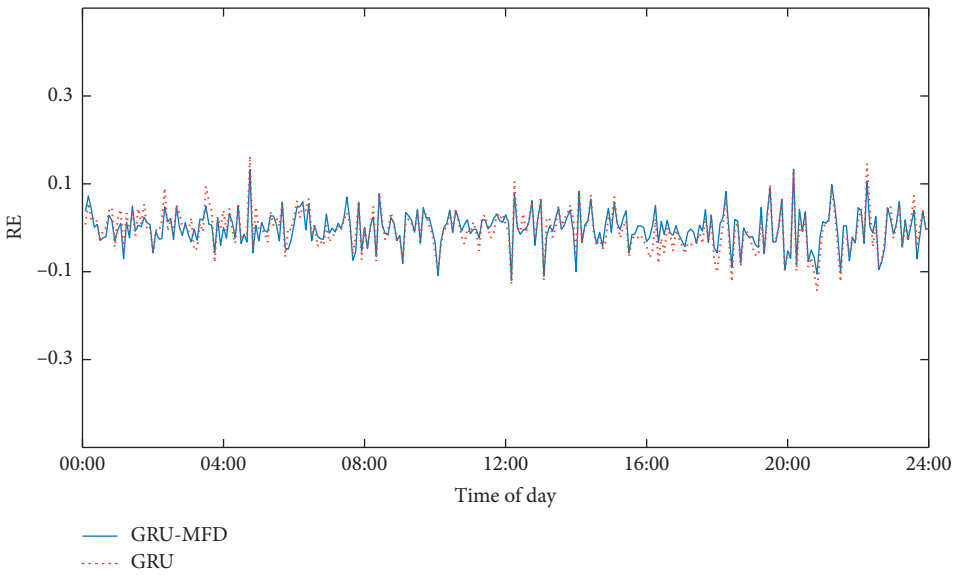



Figure 17: Errors on July 5, 2018, Section 1.

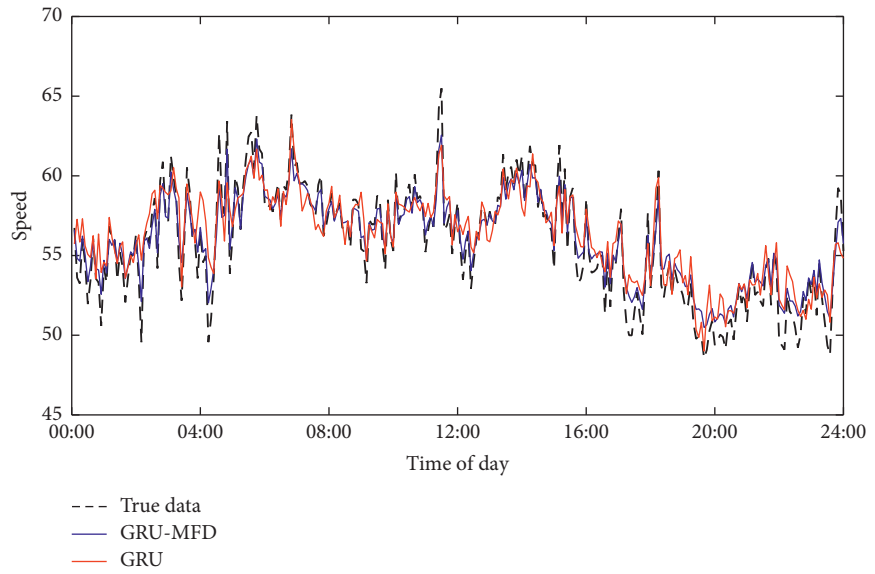



Figure 18: Prediction results on July 5, 2018, Section 2.

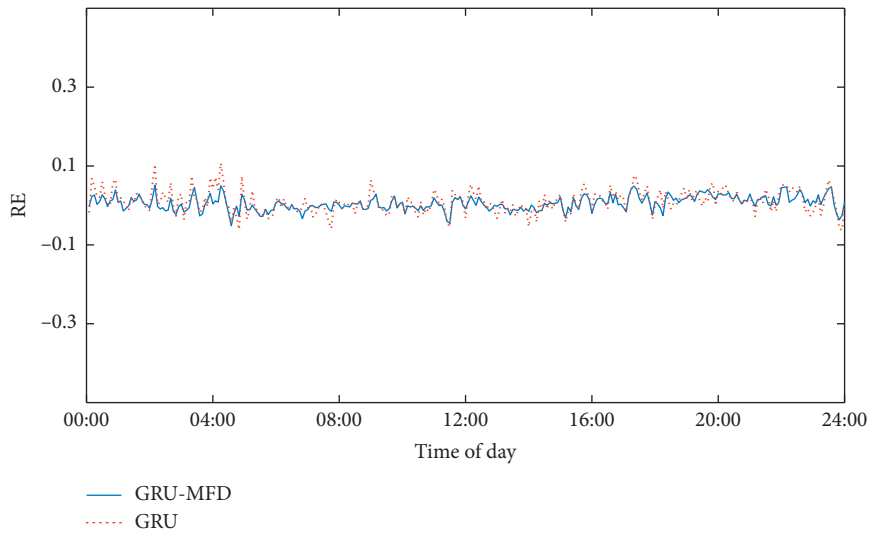



Figure 19: Errors on July 5, 2018, Section 2.

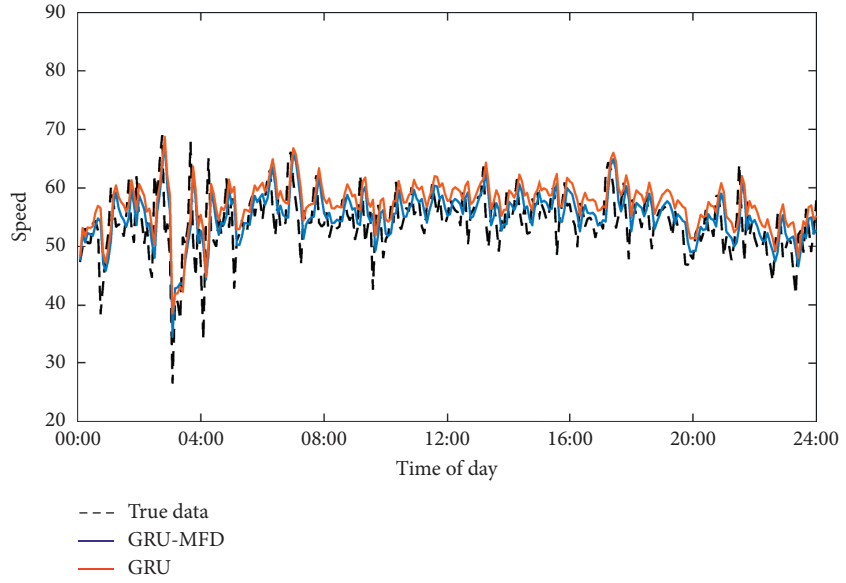



Figure 20: Prediction results on July 11, 2018, Section 2.

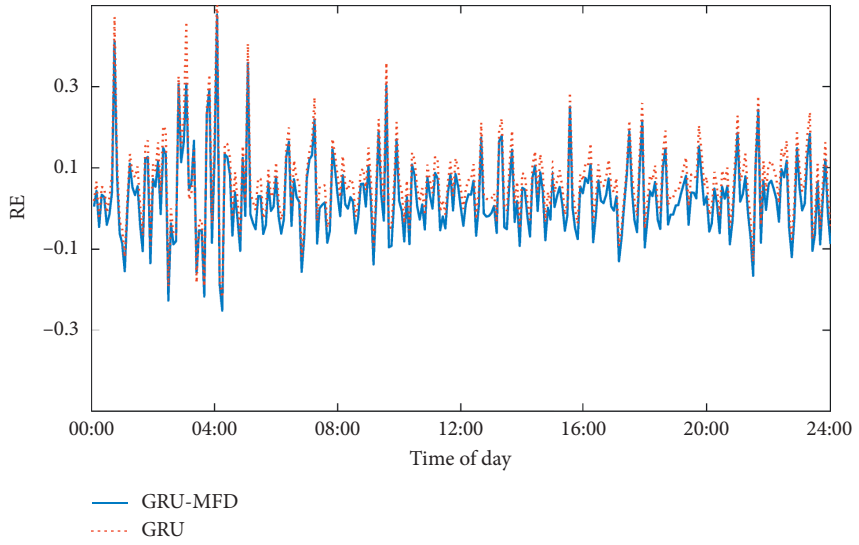



Figure 21: Errors on July 11, 2018, Section 2.

MAPE increased by about 0.5%, and on the day of the accident, MAPE increased by about 1.5%. The RMSE also improved more on accident day, compared with the day of no accident.

## 6. Conclusions

The traffic speed in adjacent regions between highway and expressway has gradually become important information concerned by highway managers and travelers. This paper proposes a prediction method of road traffic speed that considers microscopic traffic flow characteristics and macroscopic traffic status based on the road section average speed and flow data extracted from the GPS data.

Based on MFD, road network subareas are divided and evaluated. Firstly, the Ncut algorithm is used for the division of the road network. Secondly, to ensure the stability of the divided subarea's MFD, the definition of the road network's discrete degree is proposed. The traffic state is divided combined with FCM to get the best scheme for dividing the subregions after the calculation of the discrete degree of the whole network.

The spatial-temporal correlation coefficient is proposed to measure the correlation between subareas. Then, the traffic speed time sequence of the study subarea and the related area is used to build a matrix of traffic speed. The regional matrix of traffic speed data is input into the GRU model, and the output result is the predicted traffic speed of the studied region.

This paper takes the adjacent region between the highway and expressway of Beijing as an example to verify the algorithm. The southeast corner of the Beijing road network is selected as the research area. The area consists of two ring expressways and

three highways with a total area of approximately 110 square kilometers. Truck GPS data from this region is the basis of this study. The proposed algorithm's accuracy is verified under the working days, weekend, rainy days, and accident scenarios. The result shows that, compared with the traditional prediction method based on single segment time series, the prediction accuracy of the proposed algorithm is improved. This will enhance the level of traffic information services in the adjacent region between the highway and urban expressway and ease traffic congestion.

## Data Availability

The data used to support the findings of this study have not been made available because the authors have signed the confidentiality agreement with the data providers.

## 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 (no. 2018YFB1600900) and National Natural Science Foundation of China (71871011, 71890972/71890970, and 71621001).

## References

[1] L. Peng, Z. Li, C. Wang, and T. Sarkodie-Gyan, "Evaluation of roadway spatial-temporal travel speed estimation using mapped low-frequency AVL probe data," *Measurement*, vol. 165, Article ID 108150, 2020.

[2] X. Qu, Y. Yu, M. Zhou, C.-T. Lin, and X. Wang, "Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: a reinforcement learning based approach," *Applied Energy*, vol. 257, Article ID 114030, 2020.

[3] H. Liu, K. Zhang, and J. Z. Henk, "Short-term travel time prediction for urban networks," *Journal of Transportation Systems Engineering & Information Technology*, vol. 7, no. 3, pp. 118–124, 2007.

[4] D. Zhang, C. OuYang, and W. Wu, "N-order neighbor road network based vehicle travel time estimation model," *Chinese Journal of Electronics*, vol. 43, no. 12, pp. 2491–2496, 2015.

[5] B. Dhivyabharathi, E. S. Hima, and L. Vanajakshi, "Stream travel time prediction using particle filtering approach," *Transportation Letters*, vol. 10, no. 9, pp. 1–8, 2016.

[6] J. Zhao, F. Xu, K. Zhang, and J. Bai, "Highway travel time prediction based on multi-source data fusion," *Journal of Transportation Systems Engineering & Information Technology*, vol. 16, no. 1, pp. 52–57, 2016.

[7] J. Zhao, H. Wang, and W. Liu, "Prediction of highway travel time by adaptive kalman filtering," *Journal of South China University of Technology (Natural Science Edition)*, vol. 2014, no. 2, pp. 109–115, 2014.

[8] Z. Wang, A. Goodchild, and E. Mccormack, "Measuring truck travel time reliability using truck probe GPS data," *Journal of Intelligent Transportation Systems*, vol. 20, no. 2, pp. 103–112, 2016.

[9] M. Zhou, X. Qu, and X. Li, "A recurrent neural network based microscopic car following model to predict traffic oscillation," *Transportation Research Part C: Emerging Technologies*, vol. 84, pp. 245–264, 2017.

[10] Z. Wang and A. V. Goodchild, "GPS data analysis of the impact of tolling on truck speed and routing," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2411, no. 1, pp. 112–119, 2014.

[11] H. Jula, M. Dessouky, P. Ioannouc, and A. Chassiakosd, "Container movement by trucks in metropolitan networks: modeling and optimization," *Transportation Research Part E*, vol. 41, no. 3, pp. 235–259, 2011.

[12] C. Dong, K. Xie, X. Sun, M. Lyu, and H. Yue, "Roadway traffic crash prediction using a state-space model based support vector regression approach," *PLoS One*, vol. 14, no. 4, Article ID e0214866, 2019.

[13] B. Yu, H. Wang, W. Shan, and B. Yao, "Prediction of bus travel time using random forests based on near neighbors," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 4, pp. 333–350, 2018.

[14] S. Xie and L. Wei, "Application of elman neural networks to predict truck's operating speed," *Advanced Materials Research*, vol. 2013, no. 4, pp. 694–697, 2013.

[15] X. Wang and L. Xu, "Study on short-term traffic flow prediction based on deep learning," *Journal of Transportation Systems Engineering & Information Technology*, vol. 18, no. 1, pp. 81–88, 2018.

[16] W. Luo, B. Dong, and Z. Wang, "Short-term traffic flow prediction based on CNN-SVR hybrid deep learning model," *Journal of Transportation Systems Engineering & Information Technology*, vol. 2017, no. 5, pp. 68–74, 2017.

[17] Z. Yao, C. Shao, and Z. Xiong, "Research on use of support vector machine for forecasting link travel time," *Journal of Highway and Transportation Research and Development*, vol. 24, no. 9, pp. 96–99, 2007.

[18] Y. Wu, H. Tan, L. Qin, B. Ran, and Z. Jiang, "A hybrid deep learning based traffic flow prediction method and its understanding," *Transportation Research Part C: Emerging Technologies*, vol. 90, no. 90, pp. 166–180, 2018.

[19] Y. Jia, J. Wu, and Y. Du, "Traffic speed prediction using deep learning method," *IEEE International Conference on Intelligent Transportation Systems*, vol. 2016, no. 11, pp. 1–4, 2016.

[20] J. Zhao, Y. Gao, Z. Yang et al., "Truck traffic speed prediction under non-recurrent congestion: based on optimized deep learning algorithms and GPS data," *IEEE Access*, vol. 7, no. 7, pp. 9116–9127, 2019.

[21] Z. Wang, D. Li, and X. Cui, "Travel time prediction based on LSTM neural network in precipitation," *Journal of Transportation Systems Engineering & Information Technology*, vol. 20, no. 1, pp. 137–144, 2020.

[22] D. Zang, J. Ling, J. Cheng, K. Tang, and X. Li, "Using convolutional neural network with asymmetrical kernels to predict speed of elevated highway," in *Proceedings of the International Conference on Ion Sources*, pp. 212–221, Shanghai, China, October 2017.

[23] B. Peng, Ju. Tang, X. Cai, J. Xie, Y. Zhang, and Y. Wang, "3DCNN-DNN based traffic status prediction from aerial videos," *Journal of Transportation Systems Engineering & Information Technology*, vol. 20, no. 3, pp. 39–46, 2020.

[24] J. Zhao, Y. Gao, J. Tang, L. Zhu, and J. Ma, "Highway travel time prediction using sparse tensor completion tactics and K-nearest neighbor pattern matching method," *Journal of Advanced Transportation*, vol. 2018, Article ID 5721058, 2018.

[25] J. Zhao, Y. Gao, Y. Guo, and Z. Bai, "Travel time prediction of expressway based on multi-dimensional data and the particle swarm optimization–autoregressive moving average with exogenous input model," *Advances in Mechanical Engineering*, vol. 10, no. 2, pp. 1–16, 2018.

[26] J. Zhao, Y. Gao, Y. Qu, H. Yin, Y. Liu, and H. Sun, "Travel time prediction: based on gated recurrent unit method and data fusion," *IEEE Access*, vol. 6, no. 6, pp. 70463–70472, 2018.

[27] M. Zhou, Y. Yu, and X. Qu, "Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: a reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 1, pp. 433–443, 2020.

[28] Y. Ji and N. Geroliminis, "On the spatial partitioning of urban transportation networks," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1639–1656, 2012.

[29] Y. Ji, J. Luo, and N. Geroliminis, "Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2422, no. 2422, pp. 1–11, 2014.

[30] J. Haddad and N. Geroliminis, "On the stability of traffic perimeter control in two-region urban cities," *Transportation Research Part B: Methodological*, vol. 46, no. 9, pp. 1159–1176, 2012.

[31] Y. Ma, X. Yang, and Y. Zeng, "Urban traffic signal control network partitioning using spectral method," *Systems Engineering-Theory & Practice*, vol. 30, no. 12, pp. 2290–2296, 2010.

[32] F. Yang, L. Peng, A. Teng, and C. Wang, "Measuring temporal and spatial travel efficiency for transit route system using low-frequency bus automatic vehicle location data," *Advances in Mechanical Engineering*, vol. 10, no. 10, pp. 1–11, 2018.

[33] H. Su, Z. Bai, L. Deng, S. She, Y. Qi, and J. Zhao, "Research on expressway state recognition based on Beidou positioning and navigation data," in *Proceedings of International Conference on Smart Grid and Electrical Automation*, pp. 354–357, Changsha, China, June 2018.

[34] H. Han, L. Peng, A. Teng, C. Wang, and T. Z. Qiu, "Evaluation of freeway travel speed estimation using anonymous cellphones as probes: a field study in," *Canadian Journal of Civil Engineering*, 2020.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8818496, 11 pages
https://doi.org/10.1155/2020/8818496

*Research Article*

# Modeling Drivers' Stopping Behaviors during Yellow Intervals at Intersections considering Group Heterogeneity

**Juan Li**,[1,2] **Hui Zhang**,[1,2] **Yanru Zhang**,[3] **and Xuan Zhang**[4]

[1]*School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China*
[2]*Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing, China*
[3]*JP Morgan Chase, New York, NY, USA*
[4]*Sam's Club Technology, Austin, TX, USA*

Correspondence should be addressed to Juan Li; juanli@bjtu.edu.cn

Received 25 May 2020; Revised 12 June 2020; Accepted 31 October 2020; Published 26 November 2020

Academic Editor: Chunjiao Dong



Stopping behavior during yellow intervals is one of the critical driver behaviors correlated with intersection safety. As the main index of stopping behavior, stopping time is typically described by Accelerated Failure Time (AFT) model. In this study, the comparison of survival curves of stopping time confirms the existence of group specific effects on drivers. However, the AFT model is developed based on the homogeneity assumption. To overcome this drawback, shared frailty survival models are developed for stopping time analysis, which consider the group heterogeneity of drivers. The results show that log-logistic based frailty model with age as a grouping variable has the best goodness of fit and prediction accuracy. Analysis of the models' parameters indicates that phone status, maximum deceleration, vehicles' speed, and the distance to stopping line at the onset of the yellow signal have significant impacts on stopping time. Additionally, heterogeneity analysis illustrates that young, middle-aged, and female drivers are more likely to brake harshly and stop past the stop line, which may block the intersection. Furthermore, drivers, who are more familiar with traffic environments, are more possible to make reasonable stopping decisions approaching intersections. The results can be utilized by traffic authorities to implement road safety strategies, which will help reduce traffic incidents caused by improper stopping behavior at intersections.

## 1. Introduction

Signalized intersections are crucial components in road networks, where traffic accidents occur frequently [1]. 15,188 vehicles are involved in fatal crashes at intersections and more than 50 percent of all crashes occurred in intersections according to the statistics on the National Highway Traffic Safety Administration (NHTSA) in 2017 and the Federal Highway Administration (FHWA) in 2018 accordingly [2, 3]. One of the reasons causing high accident rates at intersections is drivers' incorrect decisions of going or stopping at the onset of a yellow signal. Improper stopping behavior may lead to stopping at intersections illegally, or rear-end collision. Recent research has shown that stopping behavior is a crucial influencing factor for designing traffic facilities, such as red-light camera and countdown timers [4, 5], so it is urgent to understand the principle of stopping behavior if we want to reduce the accident rate at intersections.

Most driver behavior studies focus on drivers' decision-making processes at the onset of yellow signals, especially in the dilemma zone. Since driving decisions can be seen as a binary classification problem, methods such as binary logistic model [6, 7], binary probit model [8, 9], and ordered probit model [10] are widely adopted in this issue. In addition, influencing factors on driver behaviors are widely detected, such as vehicle's factors (e.g., speed, deceleration rates, and distance to stop line) [11, 12], driver characteristics (e.g., age and gender) [13, 14], distracting factors (e.g., cell phone use and listening to music) [15–17]. Results have

indicated that higher approach speeds are always associated with higher deceleration rates and aggressive driving behavior [12, 18]. Older drivers usually employ greater deceleration levels compared with younger and middle-aged drivers [11], while female drivers generally have a greater level of variability in their driving style [19], and cell phone use has detrimental effects on both novice and experienced drivers [20]. However, the major concern of the above studies is drivers' stop/go decisions, and few studies investigate drivers' stopping behavior approaching intersections.

Driver stopping behavior can be divided into four steps: perception, judgement, manipulation, and stop. Previous research pays more attention to the driver's preparation reaction time (PRT) of braking behavior [21, 22]. Methods commonly focus on the analysis of variance (ANOVA) and a series of linear mixed models [23–25]. However, the stopping process is a continuous-time state, while PRT can only reflect the initial state of the period. Therefore, stopping time is selected to describe the entire stopping behavior at intersections, and the survival analysis model is an appropriate and effective way for timing analysis.

In recent years, the survival analysis model has attracted more attention from researchers as it is convenient and suitable to analyze time-related data [26, 27]. This model has also been widely used in the transportation field, including traffic accident analysis, travel behavior, pedestrian crossing behavior, and so forth [28–32]. Among survival analysis models, the nonparametric methods and semiparametric models are desirable methods if only the duration times are available or distributional assumptions for the duration data are unexplicit. For example, Tiwari et al. the select nonparametric methods to produce the Kaplan–Meier survival curves analyzing pedestrian risk exposure at signalized intersections [33]. Hao and Cheng conduct nonmotor vehicles' waiting time survival curve to evaluate nonmotor vehicles' street-crossing behavior at signalized crosswalk [34]. However, the above methods cannot build a relationship between event time and key affecting factors, respectively. The parametric AFT duration model is approved by further studies, which can embody specific assumptions about the distribution of failure times. Bella et al. select the AFT model with a Weibull distribution to analyze reaction time and speed reduction time for the evaluation of the effects on driving performance [35]. Li and Silvestri develop an AFT duration model to evaluate the intersection crossing completion time of drivers [36]. Though some key factors were analyzed in these studies, unobserved heterogeneity was neglected for the limitation of homogeneity assumption among objectives in AFT model.

The reasons causing heterogeneity are various, including unobserved covariates, traffic conditions (e.g., traffic volume, vehicles' type, road geometry), and driver characters (e.g., gender, age, and driving experience) [37–39]. Studies have shown that there are clear differences in driving sensitivity, crash severity, and decision-making among different age and gender driver groups when approaching intersections [40–42]. Therefore, it is necessary to consider behavioral differences among driver groups. Methods like random parameters model and latent class model have been combined to develop a mixed logit model and probit model to reflect these differences in driver decision-making behavior [42–44], but these studies mainly focus on unobserved heterogeneity, few studies explore group heterogeneity in stopping behavior at intersections.

Considering the group heterogeneity among driver group, the shared frailty survival model is developed to predict stopping time approaching intersections. The primary contribution of the study is that it makes an initial attempt to avoid the effect of group heterogeneity by incorporating a fragility to describe the difference among driver groups. In this way, the frailty is assumed to be a random distribution across driver groups. Furthermore, the relationship between stopping time and influencing factors has been amply explored. More importantly, the results of the model can be used to make effective traffic management and operational strategies, which will reduce the accidents caused by improper stopping behavior at intersections.

## 2. Methodology

Because the survival analysis model is well-suitable in analyzing the law between events and time, it is employed as a basic model to explore various factors that influence vehicles' stopping time approaching intersections.

*2.1. Survival Analysis Model.* The survival analysis model has three basic types, nonparametric methods, semiparametric models, and full-parametric models [26]. If there are only the duration times available, then life tables and survival curves can be definitely a good choice to analyze the survival trend for a dataset, so nonparametric methods usually applied for the first step in survival analysis. Semiparametric models, like Cox's proportional hazards model, are suitable for modeling duration data with one or more covariates observed and only minimal assumptions about the underlying distribution. This kind of model is more specific than the nonparametric model undoubtedly, but due to its limited flexibility, it cannot deal with heterogeneity among objectives. Therefore, for further study, more specific and flexible models, parametric models are developed for analyzing survival data.

*2.2. AFT Model.* AFT model, as a kind of parametric survival analysis model, can mine the impact directly of different influencing factors on survival time and hazard function. In this study, to analyze the different factors that affect the driver's stopping behavior, the timing variable $T$ is defined as the stopping time, which has a probability function $f(t)$, survival function $S(t)$, and hazard function $h(t)$. The survival function is defined as the probability that an individual survives longer than a certain time point $t$ and the hazard function of survival time $T$ gives the conditional failure rate, which is defined as the probability of failure during a small time interval. The specific functions and their relationships are as follows:

$$S(t) = P(T \geq t),$$
$$h(t) = \lim_{\Delta t \longrightarrow 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t}, \quad (1)$$
$$h(t) = \frac{f(t)}{S(t)}.$$

According to the meaning of stopping time, the survival function in this study can reflect the probability of a driver who has not stopped at a certain time point and the hazard function represents the probability of stop during a very small time interval. The greater the hazard rate, the more likely drivers stop at that time point.

The regression form between the log of stopping time $T$ and various variables in the AFT model is as follows:

$$\log T_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \sigma\varepsilon_i, \quad (2)$$

where $\varepsilon_i$ is a random disturbance term, $x_{ij}\,(j = 1, 2, \ldots, k)$ represents the $j$ variable of individual$i$, $\beta_0, \ldots, \beta_k$ and $\sigma$ are parameters to be estimated, $\beta_0$is the model's constant term, $\beta_1, \ldots, \beta_k$is the corresponding coefficient of$x_{ij}$, and $T_i$ represents the survival time of the individual $i$.

One of the advantages of AFT model is that it can provide explicit distributions for stopping time, such as exponential, Weibull, log-normal, log-logistic, and generalized gamma distributions. In this study, the above distributions were all considered in conducting models like equation (2), and the maximum likelihood estimation (MLE) method was used to select the best model.

*2.3. Shared Frailty Survival Model.* The shared frailty survival model is a type of developed mixture AFT model, which can solve the group heterogeneity among different groups by setting an unobservable multiplicative effect, called frailty, $\alpha$, which is assumed to be random in each cluster [45, 46].

For individuals $q\,(l = 1, \ldots, n, q = 1, \ldots, n_p)$ in the $l$ group, the hazard function is set as

$$h_{lq}(t|\alpha_l) = \alpha_l h_{lq}(t) = \alpha_l h_0(t)\exp\left(\beta^T X_{lq}\right), \quad (3)$$

where $X_{lq}$ is the covariate vector and $\beta$ is the fixed coefficient vector to be estimated. $h_0(t)$ refers to the baseline hazard function and $h_{lq}(t)$ represents the individual hazard when the covariate is $X_{lq}$. The frailty, $\alpha$, is a random positive quantity and, for identifiability of the model, supposed to have mean 1 and variance $\theta$. The larger the variance value, the greater the difference between different driver groups. For mathematical tractability, we limit the choice of frailty distribution to either the gamma $((1/\theta), \theta)$ distribution or the inverse-Gaussian $(1, (1/\theta))$ distribution, and the corresponding survival functions are

$$\text{gamma}: \cdot S_\theta(t) = \exp\left[1 - \theta\log\{S(t)\}\right]^{-(1/\theta)},$$
$$\text{inverse} - \text{Gaussian}: \cdot S_\theta(t) = \exp\left\{\frac{1}{\theta}\left(1 - \left[2\theta\log\{S(t)\}\right]^{(1/2)}\right)\right\}, \quad (4)$$

where $S_\theta(t)$ is the survival function under the corresponding frailty distribution and $S(t)$ is the basic survival function.

It can be seen from the hazard function of the shared frailty survival model that the model assumes that all stopping times in a driver group are independent given the frailty, $\alpha$. In other words, the model is a conditionally independent model and mainly deals with differences between different driver groups.

## 3. Application to Stopping Time Prediction

*3.1. Data Description and Preparation.* The dataset comes from the Driver Behavior Analysis Competition organized by the Transportation Research Board (TRB) in 2014 [47, 48]. The data study was conducted at the National Advanced Driving Simulator (NADS) at the University of Iowa, which is a high-fidelity driving simulator, aiming to detect the influence of wireless phone use on driver behaviors. During the experiment, each participant had three drives. Each drive consisted of 3 equivalent segments and each segment had one urban field and one rural field. Besides, five signal intersections were set in each segment, and two of the intersections' traffic signals were randomly triggered to change states to yellow when the driver approaches. In addition, secondary tasks were randomly applied, which exposed participants to 3 cell phone interfaces, including baseline (no phone call), outgoing call (calling out), and incoming call (answering a call). For each drive, the participant experienced a different order of segments/interfaces that are otherwise equivalent. The incoming and outgoing calls were started before the arrival at each segment. Therefore, there were 6 times yellow light decisions for each drive and each participant needed to go through 18 times of that.

The original dataset was collected from 49 participants, with a total of 1157 observations. After deleting the missing and abnormal data, there are still 838 observations left, including 306 "go" observations and 532 "stop" observations. This study selects 532 "stop" observations as the research object.

In this study, seven factors related to driver's stopping behavior were selected as independent variables, including driver's age (Age), gender (Gender), phone status (PS), maximum deceleration (MD), maximum acceleration (MA), vehicle's approaching speed (GTYV), and the distance to the stopping line (GTYD) at the onset of yellow light. The detailed descriptions and values for each variable are shown in Table 1. There are more young drivers in the subject groups and male drivers are a bit more than female, while the numbers of three kinds of phone status' drivers are relatively equal, noticing that maximum decelerations are negative, and the mean of their absolute value is 6.11 $\text{m} \cdot \text{s}^{-2}$. Besides, the mean of the vehicle's speed at the onset of yellow light is 18.95 $\text{m} \cdot \text{s}^{-1}$, and the mean distance is 62.36 meters.

The stopping time (FST) was selected as the duration variable, which was the time from the onset of yellow light to the vehicle stopping completely. Statistics on overall stopping time are shown in Table 2. Vehicles begin to stop at 3.63 seconds after the onset of yellow light and all vehicles stop

Table 1: Statistical description of independent variables.

| Variables | Description | Value | Count | Mean | Std. Dev. |
|---|---|---|---|---|---|
| Age | Old (50~60 years) | 0 | 131 | | |
| | Middle (30~45 years) | 1 | 197 | | |
| | Young (18~25 years) | 2 | 204 | | |
| Gender | Female | 0 | 242 | | |
| | Male | 1 | 290 | | |
| PS | Outcoming call ($O$) | 0 | 178 | | |
| | Incoming call ($I$) | 1 | 181 | | |
| | Baseline ($B$) | 2 | 173 | | |
| MD ($m \cdot s^{-2}$) | Maximum deceleration | Continuous | 532 | −6.11 | 1.53 |
| MA ($m \cdot s^{-2}$) | Maximum acceleration | Continuous | 532 | 3.47 | 0.43 |
| GTYV ($m \cdot s^{-1}$) | Vehicle's speed at the onset of yellow light | Continuous | 532 | 18.95 | 2.27 |
| GTYD (m) | Distance between the vehicle and the stopping line at the onset of yellow light | Continuous | 532 | 62.36 | 10.56 |

Table 2: Quartile statistics of stopping time.

| Quantile | 90% | 75% | 50% | 25% | Min. | Max. | Mean | Std. Dev. |
|---|---|---|---|---|---|---|---|---|
| Stopping time (s) | 8.92 | 7.68 | 5.84 | 4.69 | 3.63 | 12.97 | 6.24 | 0.078 |

under 12.97 seconds. The mean stopping time is 6.24 seconds and 90%, 75%, 50%, and 25% of drivers stop within 8.92 seconds, 7.68 seconds, 5.84 seconds, and 4.69 seconds accordingly.

To understand the probability that vehicles haven't stopped at a certain time better, this study conducts an overall survival curve for 532 "stop" data. Figure 1 presents that vehicles begin to stop at 3.63 seconds and most vehicles can stop completely under 10 seconds.

*3.2. Heterogeneity Analysis.* Drivers' stopping behaviors may be different under different driver groups; therefore survival curves under different grouping variables were developed for preliminary verification.

The survival curves of stopping time are varied under different driving times, age, gender, and phone states, as shown in Figure 2, which presents evidence for the existence of group heterogeneity. So, in this study, we employ the shared frailty survival model to analyze stopping time, which can avoid the homogeneity assumption of the AFT model.

Behavioral characters of different driver groups are assessed by the dangerous driving index. According to early research, driver behavior is assumed to be statistically consistent within a certain population when facing certain levels of conflicts [49, 50]. Therefore, this study employs the possibility of dangerous stopping in the original dataset to evaluate the dangerous stopping behavior in different driver groups. The dangerous driving index can be described as the probability of dangerous stopping behaviors and can be calculated:

$$R = \frac{1}{m} \sum_{r=1}^{m} b_r, \tag{5}$$

where $R$ represents the dangerous driving index of each driver group and $m$represents the number of stopping observations in each driver group, when the $r$ observation's record is over-the-line stopping and the value of $b_r$ is 1; otherwise, $b_r$ is set to 0.

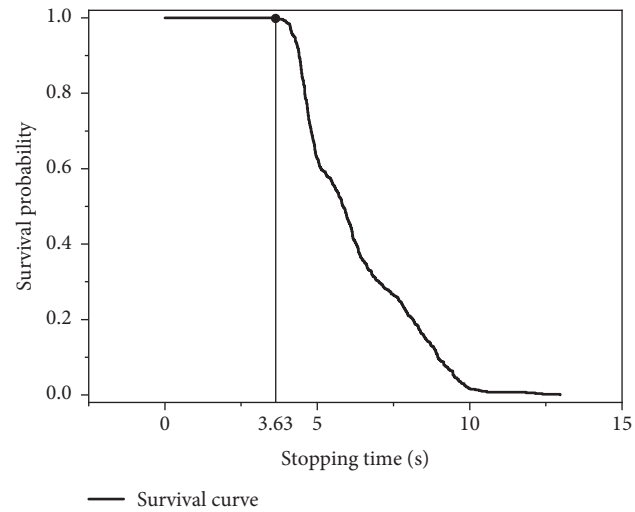



Figure 1: Survival curve of overall stopping time.

According to the definition of the dangerous driving index, the greater it is, the greater the tendency of this driver group to stop over the line. Table 3 presents the dangerous driving index for different driver groups. It can be found that the dangerous driving index of middle-aged drivers is the highest among age groups, and young drivers in it are also very high while old drivers' are the lowest. Besides, female drivers' index is higher than male, and the index is a little higher when there are phone interferences than no phone call. The results also show that dangerous driving indexes are changeable under different driving times; especially the index is the highest under the second drive.

## 4. Results and Discussion

*4.1. Model Results.* Based on 532 "stop" observations approaching intersections, the shared frailty survival models with grouping variables, including driving times, age, gender, and phone status, are constructed, and parameters can

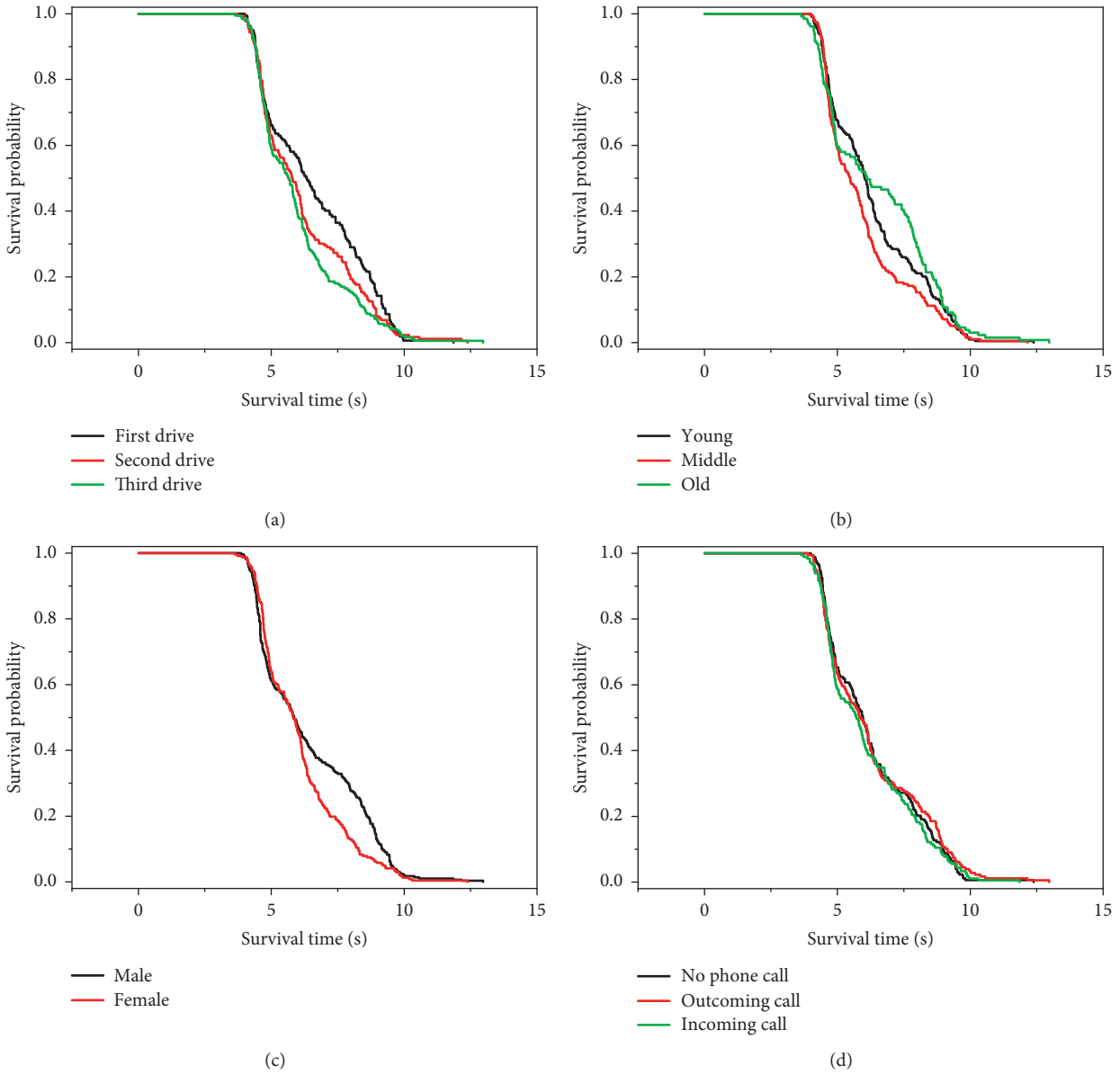



Figure 2: Survival curves under different grouping variables. (a) Survival curves of different driving times. (b) Survival curves of different age groups. (c) Survival curves of different gender groups. (d) Survival curves of different phone status.

Table 3: Dangerous driving index of different driver groups.

| Group | Description | Number of drivers | Number of unsafe stops | Dangerous driving index |
|---|---|---|---|---|
| Age | Old | 131 | 8 | 0.061 |
| | Middle | 197 | 24 | 0.122 |
| | Young | 204 | 24 | 0.112 |
| Gender | Female | 242 | 29 | 0.120 |
| | Male | 290 | 27 | 0.093 |
| PS | Outcoming call | 178 | 19 | 0.107 |
| | Incoming call | 181 | 19 | 0.105 |
| | Baseline | 173 | 18 | 0.104 |
| Drive | Drive 1 | 162 | 13 | 0.080 |
| | Drive 2 | 176 | 22 | 0.125 |
| | Drive 3 | 194 | 21 | 0.108 |

be estimated by using the maximum likelihood estimation method at a significance level of 0.1.

In order to select the best distribution type and grouping variable for modeling the stopping time, two model selection statistics, including the Akaike information criterion (AIC) statistic and the Bayesian information criterion (BIC) statistic are adopted. In general, the model has better goodness of fit with smaller values of the AIC and BIC. Criterion statistics in Table 4 illustrates that generalized gamma distribution is the best-fitted distribution among AFT models, while the log-logistic distribution is proved to be the best distribution among shared frailty survival models, especially when age is selected as grouping variables.

In addition, better distributions are also selected by comparing Cox-Snell residuals [51]. Figures 3(a)–3(d) illustrate the Cox–Snell residuals for different distributions of AFT models; it can be seen that the plotted points in generalized gamma distribution of AFT model fall closer to the reference line with a slope of 1, which confirms the goodness of fit of the model. At the same time, Figures 3(e)–3(h) indicate that log-logistic distribution is the best fit distribution among shared frailty survival models.

Table 5 presents the estimated results of the parameters of the above good fitted models. Among the shared frailty survival models, except for the model that grouping variable is gender, in which the frailty $\alpha$ follows gamma distribution, other shared frailty survival models' $\alpha$ follow the inverse Gaussian distribution. Variances $\theta$ of frailty $\alpha$ are all far greater than 0, which also indicate the existence of significant differences between different driver groups. Besides, referring to Table 4, it is easy to find that all shared frailty survival models have smaller values of AIC and BIC than that of the AFT model, which indicates the improved goodness of fit for the shared frailty survival model. Besides, the model with the grouping variable of Age has the smallest value of AIC, BIC. Therefore, the shared frailty survival model with the age grouping variable is selected to analyze the key factors that affect the stopping time.

The results of the optimal model in Table 5 show that phone status, maximum deceleration, vehicle's speed, and the distance to the stopping line at the onset of yellow light are significant variables that affect the stopping time. Different variables' effects on stopping time can be calculated referring to equation (2). Table 6 illustrates the influence of significant variables on the stopping time. It can be seen when incoming calls (PS = 1) that the stopping time is relatively reduced by 1.0% ($e^{-0.001} - 1$) compared to outcoming calls (PS = 0) states where the stopping time is increased by 1.3% ($e^{0.013} - 1$) compared to no phone calls condition (PS = 2). Besides, the stopping time under incoming calls is relatively shorter than that of outcoming calls, which means that the driver may stop harshly when receiving a call. The reason for this phenomenon may be that the driver does not have sufficient psychological preparation when receiving a call rather than calling out, so greater distraction is caused, which agrees with the previous research [20, 52]. At the same time, referring to the dangerous driving index in Table 3, it is found that the dangerous driving index is higher than no phone call situation when there are phoning tasks, indicating that the driver is more likely to stop over the line when distracted by phone use. Therefore, when approaching intersections, drivers are suggested to avoid distractions such as making and receiving calls. At the same time, because the value of the maximum deceleration (MD) variable in this study is negative, according to the sign of variable coefficients, it is found that the maximum deceleration (MD) and vehicle's speed at the onset of yellow light (GTYV) have a negative impact on the stopping time. When there is a unit increase in the absolute value of maximum deceleration and the value of vehicle speed at the onset of yellow light, the stopping time decreases by 5.8% and 1.3% accordingly. Therefore, there is a close relationship between the maximum deceleration, vehicle's speed at the onset of yellow light, and the emergency stop behaviors, so these two factors should be reasonably controlled. Additionally, the distance to the stopping line at the onset of yellow light (GTYD) has a positive effect on the stopping time, and for every meter increase of GTYD, the stopping time increases by 1.1%. Therefore, it is necessary to reasonably control the braking distance to ensure a smooth stopping of the vehicle.

Additionally, other model results in Table 5 also show that Age and Gender variables have an impact on stopping time. The coefficients of young (Age = 2), middle (Age = 1), and male (Gender = 1) drivers are all negative at a significance level; this means young, middle, and female drivers have shorter stopping time and stop more urgently than older and male drivers accordingly, which agree with the previous research finding [19]. Referring to the dangerous driving index in Table 3, the dangerous driving index of young (0.112) and middle (0.122) drivers is higher than that of old (0.061) drivers, and the index of female drivers (0.120) is higher than that of male drivers (0.093), indicating that young, middle, and female drivers are also more likely to stop over the line, which may cause intersection jams.

*4.2. Model Accuracy.* The mean absolute percentage error (MAPE) is used to compare the prediction accuracy of the model [53]. The smaller the value, the better the prediction accuracy. The median survival time is selected to calculate the predicted value in this article:

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{n}\left|\frac{y_i - y_i^*}{y_i}\right|, \tag{6}$$

where $N$ is the number of observed individuals, $y_i$ is the actual observed value of the individual $i$, and $y_i^*$ is the predicted value.

Table 7 presents values of MAPE for the AFT model (15.04%) and the shared frailty survival model with the age grouping variable (13.92%). It can be seen that both models have a good prediction accuracy [54]. However, compared with the AFT model from a global perspective, there is a negligible improvement in MAPE, 1.12%, in the shared frailty survival model. Therefore, this study adopts a categorical analysis in prediction accuracy. Note that 75% of drivers will stop under 7.7 seconds in Table 2, so this study

Table 4: Statistics of criterion based on different distribution types.

| Criterion | AFT model distribution type | | | | Shared frailty survival model distribution type | | | |
|---|---|---|---|---|---|---|---|---|
| | Weibull | Log-normal | Log-logistic | Generalized gamma | Loglogistic (d-rive) | Log-logistic (age) | Log-logistic (gender) | Log-logistic (PS) |
| AIC | 90 | −88 | −101 | **−508** | −517 | **−526** | −529 | −500 |
| BIC | 146 | −33 | −46 | **−448** | −474 | **−491** | −482 | −457 |

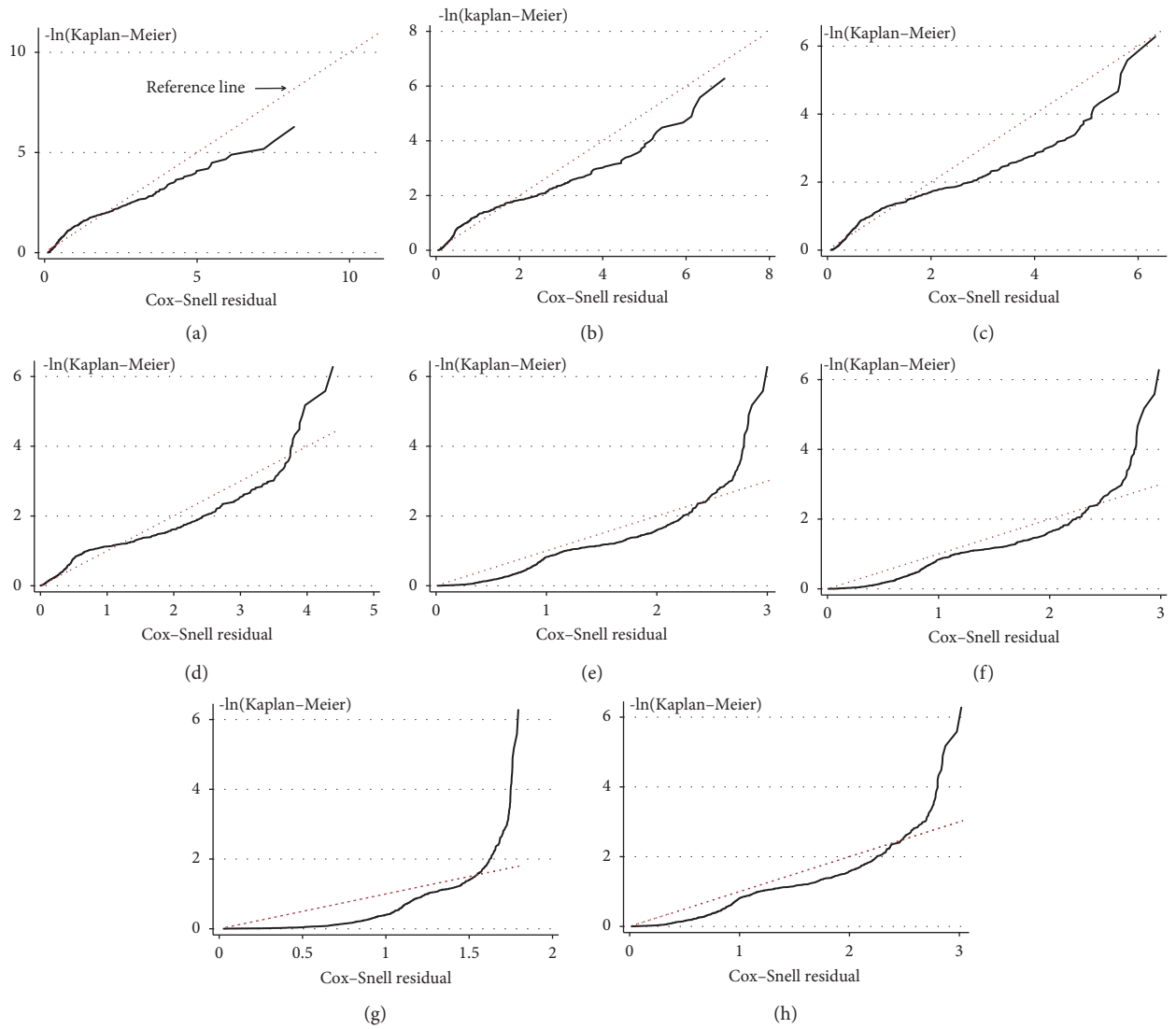



Figure 3: Cox–Snell residuals for different distribution types.(a) Weibull. (b) Log-normal. (c) Log-logistic. (d) General gamma. (e) Log-logistic (Drive). (f) Log-logistic (Age). (g) Log-logistic (Gender). (h) Log-logistic (PS).

divides the dataset into two categories at this time point and analyzes the prediction accuracy for each data category. Table 7 presents the accuracy results.

When the stopping time is less than 7.7 seconds, the improved shared model can have a very high accurate prediction (MAPE is 6%) and a small mean prediction error (0.36 sec); besides, it improves prediction accuracy by 3% compared with AFT model. Though both the above models do not have good prediction accuracy when stopping time is larger than 7.7 seconds, the prediction accuracy is also acceptable [54]. Therefore, the predictions can be used to analyze driver stopping behaviors reasonably.

*4.3. Effects of Group Heterogeneity.* As mentioned above, the evidence of heterogeneity is seen in the core results of the models presented in Figure 2 and Table 5. It is found that, compared with the AFT model, the shared frailty survival model's advantages are reflected not only in improving the model's goodness of fitting and prediction accuracy but also

Table 5: Model estimation results.

| Variables | AFT model | Shared frailty survival model | | | |
|---|---|---|---|---|---|
| | | Group variable (drive) | Group variable (age) | Group variable (gender) | Group variable (PS) |
| Distribution of FST | Generalized gamma | Log-logistic | Log-logistic | Log-logistic | Log-logistic |
| Distribution of $\alpha$ | — | Inverse Gaussian | Inverse Gaussian | Gamma | Inverse Gaussian |
| Variable | Coefficient | Coefficient | Coefficient | Coefficient | Coefficient |
| Age = 1 | −0.005 | −0.007 | — | −0.007 | −0.006 |
| Age = 2 | −0.017*** | −0.013** | — | −0.011* | −0.014** |
| Gender = 1 | — | — | — | −0.014*** | — |
| PS = 1 | 0.004 | 0.001 | −0.001 | −0.002 | 0.004 |
| PS = 2 | 0.021*** | 0.016*** | 0.013** | 0.012** | 0.019*** |
| MD | 0.056*** | 0.056*** | 0.056*** | 0.055*** | 0.056*** |
| MA | — | — | — | — | — |
| GTYV | −0.014*** | −0.013*** | −0.013*** | −0.013*** | −0.013*** |
| GTYD | 0.011*** | 0.011*** | 0.011*** | 0.011*** | 0.011*** |
| Constant | 1.511*** | 1.483*** | 1.472*** | 1.477*** | 1.481*** |
| $\theta$ | — | 31.273 | 27.512 | 3.380 | 32.601 |

$\theta$ is the variance of frailty $\alpha$. ***, **, * mean statistical significance at $\alpha = 0.01, 0.05, 0.1$, respectively.

Table 6: Driver's stopping risk percentage change.

| Variable | Coefficient | EXP($\beta$) | Percentage change (%) |
|---|---|---|---|
| PS = 1 | −0.001 | 0.999 | −1.0 |
| PS = 2 | 0.013 | 1.013 | 1.3 |
| MD | 0.056 | 1.058 | 5.8 |
| GTYV | −0.013 | 0.987 | −1.3 |
| GTYD | 0.011 | 1.011 | 1.1 |

Table 7: Statistics on model accuracy.

| Stopping time category (sec) | Count | Model | MAPE (%) | Mean prediction error (sec) |
|---|---|---|---|---|
| Overall | 532 | AFT | 15.04 | 1.11 |
| | | Shared (Age) | 13.92 | 1.13 |
| FST ≤ 7.7 | 400 (75%) | AFT | **9** | 0.51 |
| | | Shared (Age) | **6** | **0.36** |
| FST > 7.7 | 132 (25%) | AFT | 32 | 2.94 |
| | | Shared (Age) | 38 | 3.46 |

in the mining impacts of other potential variables on drivers' stopping behaviors.

Comparing with the AFT model, the value and sign of significant variables' coefficients have changed in the shared frailty survival model. For example, in the shared frailty survival model with grouping variable *age*, the sign of incoming calls' (PS = 1) coefficient changes to negative compared to the AFT model, which is more in line with the fact that the incoming distraction situation has a greater impact on the driver's behavior. What is more, the reduced absolute value of coefficients for phoning states and vehicle's speed at the onset of yellow light variables, reduced constant terms of all shared frailty survival models compared to the AFT model, also illustrates that some of the significant variables' effect on stopping time is shared after considering the heterogeneity of varied driver groups. Besides, other shared frailty survival models also highlight the significant effects caused by age, gender, and different driving times on drivers' stopping behavior. Table 3 shows that, in the first, second, and third drives, the dangerous driving indexes show a trend of increasing first and then decreasing. The dangerous driving index in the second drive is the highest with a value of 0.125. The reason for this phenomenon may be drivers in the second drive have been already familiar with the driving environment and are not concentrating on driving enough. While in the third drive, drivers accumulate the previous two driving experiences and make the stopping decision more accurate and reduced the probability of stopping over the line, which reflects that familiarity with the traffic environment directly affects the safety of driver's stopping behavior, drivers who are more familiar with the traffic environment can make more reasonable stopping decisions, and these effects are difficult to mine by the AFT model.

## 5. Conclusions

To further understand drivers' stopping behaviors at intersections during the yellow interval, a survival analysis model was proposed. However, the parameter AFT model has a homogeneity assumption among drivers. To overcome this limitation, the developed shared frailty survival models

with grouping variables, including driving times, age, gender, and phone status, were conducted in this study.

Results show that the shared frailty survival models can improve the fitness and prediction accuracy compared with the AFT model, and the model with the grouping variable of age fits best. In this best-fitting model, variables, including phone status, maximum deceleration, vehicle's speed, and the distance to the stopping line at the onset of yellow light, have significant effects on stopping time. Drivers stop relatively in emergency and are easier to stop over the line when there are phoning tasks' distraction. The greater the vehicle's speed at the onset of yellow light and the absolute value of the maximum deceleration, the more likely the vehicle to have an emergency stop, while the greater the value of the distance to the stopping line at the onset of yellow light, the smoother the vehicle stop.

By heterogeneity analysis, other potential influencing factors on stopping time are also explored in this study. Except for the best-fitting shared frailty survival model, models with other grouping variables also highlight the driving behaviors' significant differences in varied age, gender, and driving times groups. Results show that young, middle, and female drivers are more likely to stop emergency and stop over the line, which may cause intersection jams. Besides, familiarity with the traffic environment has a direct impact on the safety of the stopping behavior; drivers who more familiar with the traffic condition could make a more reasonable stopping decision.

Based on the above results, drivers are recommended to reasonably predict the braking distance when approaching the signalized intersections, control the vehicle's speed and deceleration, and avoid distracting behaviors such as having phone calls. Besides, young, middle-aged, and female drivers are more likely to stop over the stop line and block the intersections. These groups of people may need more attention from traffic managers such as intensive observation. Familiarity with the traffic environment may help drivers make proper decisions, so intersection safety will be improved if drivers are provided useful road and traffic information such as speed limits and signal ahead. However, the dataset in this study only contains three driving times, so related data collection is needed for further studying the explicit effects of the familiarity with traffic environment on drivers' stopping behaviors. This study can help researchers better understand driver stopping behavior at the onset of yellow light and also can be applied to driver assistance systems and intersection design studies.

## Data Availability

The data used for this study were taken from experiments conducted at the University of Iowa-National Advanced Driving Simulator (NADS). The raw data are attached, which can be accessed from the website: http://depts.washington.edu/hfsm/upload.php.

## 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 Fundamental Research Funds for the Central Universities (no. 2019JBM036).

## Supplementary Materials

Our processed stopping data are attached. (*Supplementary Materials*)

## References

[1] H. Huang, D. Wang, L. Zheng, and X. Li, "Evaluating time-reminder strategies before amber: common signal, green flashing and green countdown," *Accident Analysis & Prevention*, vol. 71, pp. 248–260, 2014.

[2] National Highway Traffic Safety Administration, "Traffic safety facts 2017 a compilation of motor vehicle crash data (annual report)," 2017, https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812806.

[3] The Federal Highway Administration (FHWA), "Intersection safety: background and objectives," 2018, https://highways.dot.gov/research-programs/safety/intersection-safety.

[4] F. Baratian-Ghorghi, H. Zhou, and A. Franco-Watkins, "Effects of red light cameras on driver's stop/go decision: assessing the green extension hypothesis," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 46, pp. 87–95, 2017.

[5] M. Huang, M. Fujita, and W. Wisetjindawat, "Countdown timers, video surveillance and drivers' stop/go behavior: winter versus summer," *Accident Analysis & Prevention*, vol. 98, pp. 185–197, 2017.

[6] H. Köll, M. Bader, and K. W. Axhausen, "Driver behaviour during flashing green before amber: a comparative study," *Accident Analysis & Prevention*, vol. 36, no. 2, pp. 273–280, 2004.

[7] T. J. Gates, D. A. Noyce, L. Laracuente, and E. V. Nordheim, "Analysis of driver behavior in dilemma zones at signalized intersections," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2030, no. 1, pp. 29–39, 2007.

[8] Y. Sheffi and H. Mahmassani, "A model of driver behavior at high speed signalized intersections," *Transportation Science*, vol. 15, no. 1, pp. 50–61, 1981.

[9] P. Li and M. Abbas, "Stochastic dilemma hazard model at high-speed signalized intersections," *Journal of Transportation Engineering*, vol. 136, no. 5, pp. 448–456, 2010.

[10] Y. Liu, G.-L. Chang, and J. Yu, "Empirical study of driver responses during the yellow signal phase at six Maryland intersections," *Journal of Transportation Engineering*, vol. 138, no. 1, pp. 31–42, 2012.

[11] I. El-Shawarby, H. Rakha, A. Amer, and C. McGhee, "Impact of driver and surrounding traffic on vehicle deceleration behavior at onset of yellow indication," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2248, no. 1, pp. 10–20, 2011.

[12] J. Wang, K. K. Dixon, H. Li, and J. Ogle, "Normal deceleration behavior of passenger vehicles at stop sign-controlled intersections evaluated with in-vehicle global positioning system data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1937, no. 1, pp. 120–127, 2005.

[13] C. Y. D. Yang and W. G. Najm, "Examining driver behavior using data gathered from red light photo enforcement

cameras," *Journal of Safety Research*, vol. 38, no. 3, pp. 311–321, 2007.

[14] J. J. Lu and J. C. Pernía, "The differences of driving behavior among different driver age groups at signalized intersections," *IATSS Research*, vol. 24, no. 2, pp. 75–84, 2000.

[15] R. M. Brumfield and S. S. Pulugurtha, "Effect of driver cell phone use on queue discharge patterns at signalized intersections," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2257, no. 1, pp. 71–79, 2011.

[16] M. E. Rakauskas, L. J. Gugerty, and N. J. Ward, "Effects of naturalistic cell phone conversations on driving performance," *Journal of Safety Research*, vol. 35, no. 4, pp. 453–464, 2004.

[17] A. K. Huemer, M. Schumacher, M. Mennecke, and M. Vollrath, "Systematic review of observational studies on secondary task engagement while driving," *Accident Analysis & Prevention*, vol. 119, pp. 225–236, 2018.

[18] H. Rakha, A. Amer, and I. El-Shawarby, "Modeling driver behavior within a signalized intersection approach decision-dilemma zone," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2069, no. 1, pp. 16–25, 2008.

[19] S. M. Lavrenz, V. Dimitra Pyrialakou, and K. Gkritza, "Modeling driver behavior in dilemma zones: a discrete/continuous formulation with selectivity bias corrections," *Analytic Methods in Accident Research*, vol. 3-4, pp. 44–55, 2014.

[20] J. Cooper, C. Yager, and S. T. Chrysler, "An investigation of the effects of reading and writing text-based messages while driving," Southwest Region University Transportation Center, College Station, TX, USA, No. SWUTC/11/476660-00024-1, 2011.

[21] Y. D. Wong and P. K. Goh, "Driver perception-response time for braking action during signal change interval," *Road Transport Research*, vol. 9, pp. 17–26, 2000.

[22] P.-K. Goh and Y.-D. Wong, "Driver perception response time during the signal change interval," *Applied Health Economics and Health Policy*, vol. 3, no. 1, pp. 9–15, 2004.

[23] C. Ariën, E. M. M. Jongen, K. Brijs, T. Brijs, S. Daniels, and G. Wets, "A simulator study on the impact of traffic calming measures in urban areas on driving behavior and workload," *Accident Analysis & Prevention*, vol. 61, pp. 43–53, 2013.

[24] Q. Hussain, A. Pirdavani, C. Arien, T. Brijs, and W. K. M. Alhajyaseen, "The impact of perceptual countermeasures on driving behavior in rural-urban transition road segments: a driving simulator study," *Advanced Transportation Stud.: International Journal*, vol. 46, pp. 83–96, 2018.

[25] Q. Hussain, W. K. M. Alhajyaseen, K. Brijs, A. Pirdavani, and T. Brijs, "Innovative countermeasures for red light running prevention at signalized intersections: a driving simulator study," *Accident Analysis & Prevention*, vol. 134, Article ID 105349, 2020.

[26] S. P. Washington, M. G. Karlaftis, and F. L. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall/CRC, Boca Raton, FL, USA, 2nd ed. edition, 2011.

[27] D. A. Hensher and F. L. Mannering, "Hazard-based duration models and their application to transport analysis," *Transport Reviews*, vol. 14, no. 1, pp. 63–82, 1994.

[28] J. Weng, Y. Zheng, X. Yan, and Q. Meng, "Development of a subway operation incident delay model using accelerated failure time approaches," *Accident Analysis & Prevention*, vol. 73, pp. 12–19, 2014.

[29] X. Yang, Z. Gao, H. Guo, and M. Huan, "Survival analysis of car travel time near a bus stop in developing countries," *Science China Technological Sciences*, vol. 55, no. 8, pp. 2355–2361, 2012.

[30] A. Mohammadian and S. T. Doherty, "Modeling activity scheduling time horizon: duration of time between planning and execution of pre-planned activities," *Transportation Research Part A: Policy and Practice*, vol. 40, no. 6, pp. 475–490, 2006.

[31] P. Van den Berg, T. Arentze, and H. Timmermans, "A latent class accelerated hazard model of social activity duration," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 1, pp. 12–21, 2012.

[32] X. Yang, M. Abdel-Aty, M. Huan, Y. Peng, and Z. Gao, "An accelerated failure time model for investigating pedestrian crossing behavior and waiting times at signalized intersections," *Accident Analysis & Prevention*, vol. 82, pp. 154–162, 2015.

[33] G. Tiwari, S. Bangdiwala, A. Saraswat, and S. Gaurav, "Survival analysis: pedestrian risk exposure at signalized intersections," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 10, no. 2, pp. 77–89, 2007.

[34] P. Hao and W. Cheng, "Nonparametric survival analysis of non-motor vehicles violation in signal intersection," *Applied Mechanics and Materials*, vol. 713-715, pp. 2115–2118, 2015.

[35] F. Bella and M. Silvestri, "Effects of directional auditory and visual warnings at intersections on reaction times and speed reduction times," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 51, pp. 88–102, 2017.

[36] X. Li, O. Oviedo-Trespalacios, and A. Rakotonirainy, "Drivers' gap acceptance behaviours at intersections: a driving simulator study to understand the impact of mobile phone visual-manual interactions," *Accident Analysis & Prevention*, vol. 138, Article ID 105486, 2020.

[37] A. B. Ellison, S. Greaves, and M. Bliemer, "Examining heterogeneity of driver behavior with temporal and spatial factors," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2386, no. 1, pp. 158–167, 2013.

[38] P. Maheshwary, K. Bhattacharyya, B. Maitra, and M. Boltze, "A methodology for calibration of traffic micro-simulator for urban heterogeneous traffic operations," *Journal of Traffic and Transportation Engineering*, vol. 7, no. 4, 2019.

[39] B. K. Pathivada and V. Perumal, "Modeling driver behavior in dilemma zone under mixed traffic conditions," *Transportation Research Procedia*, vol. 27, pp. 961–968, 2017.

[40] S. Hong, B. Min, S. Doi, and K. Suzuki, "Approaching and stopping behaviors to the intersections of aged drivers compared with young drivers," *International Journal of Industrial Ergonomics*, vol. 54, pp. 32–41, 2016.

[41] A. Rakotonirainy, D. Steinhardt, P. Delhomme, M. Darvell, and A. Schramm, "Older drivers' crashes in Queensland, Australia," *Accident Analysis & Prevention*, vol. 48, pp. 423–429, 2012.

[42] P. T. Savolainen, "Examining driver behavior at the onset of yellow in a traffic simulator environment: comparisons between random parameters and latent class logit models," *Accident Analysis & Prevention*, vol. 96, pp. 300–307, 2016.

[43] P. Seraneeprakarn, S. Huang, V. Shankar, F. Mannering, N. Venkataraman, and J. Milton, "Occupant injury severities in hybrid-vehicle involved crashes: a random parameters approach with heterogeneity in means and variances," *Analytic Methods in Accident Research*, vol. 15, pp. 41–55, 2017.

[44] P. T. Savolainen, A. Sharma, and T. J. Gates, "Driver decision-making in the dilemma zone-examining the influences of clearance intervals, enforcement cameras and the provision of advance warning through a panel data random parameters probit model," *Accident Analysis & Prevention*, vol. 96, pp. 351–360, 2016.

[45] P. K. Andersen, J. P. Klein, K. M. Knudsen, and R. T. Palacios, "Estimation of variance in cox's regression model with shared gamma frailties," *Biometrics*, vol. 53, no. 4, pp. 1475–1484, 1997.

[46] H. T. V. Vu, "Estimation in semiparametric conditional shared frailty models with events before study entry," *Computational Statistics & Data Analysis*, vol. 45, no. 3, pp. 621–637, 2004.

[47] Transportation Research Board, "2014 TRB data competition," 2014, http://trbstats.weebly.com/2014-trb-data-competition.html.

[48] The National Advanced Driving Simulator, 2014, http://www.nads-sc.uiowa.edu/sim_nads1.php.

[49] X. Zou and D. M. Levinson, "Modeling pipeline driving behaviors," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1980, no. 1, pp. 16–23, 2006.

[50] J. Li, Q. He, H. Zhou, Y. Guan, and W. Dai, "Modeling driver behavior near intersections in hidden markov model," *International Journal of Environmental Research and Public Health*, vol. 13, no. 12, p. 1265, 2016.

[51] J. Weng, W. Qiao, X. Qu, and X. Yan, "Cluster-based log-normal distribution model for accident duration," *Transportmetrica A: Transport Science*, vol. 11, no. 4, pp. 345–363, 2015.

[52] M. M. Haque, A. D. Ohlhauser, S. Washington, and L. N. Boyle, "Decisions and actions of distracted drivers at the onset of yellow lights," *Accident Analysis & Prevention*, vol. 96, pp. 290–299, 2016.

[53] Q. Meng and X. Qu, "Estimation of rear-end vehicle crash frequencies in urban road tunnels," *Accident Analysis & Prevention*, vol. 48, pp. 254–263, 2012.

[54] Y. Chung, "Development of an accident duration prediction model on the Korean freeway systems," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 282–289, 2010.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8878346, 15 pages
https://doi.org/10.1155/2020/8878346

*Research Article*

# Expansion of the Fundamental Diagram from a Microscopic Multilane Modeling Framework of Mixed Traffic

**Mudasser Seraj, Jiangchen Li, and Tony Z. Qiu**

*Department of Civil and Environmental Engineering, University of Alberta, Edmonton T6G 2R3, Canada*

Correspondence should be addressed to Mudasser Seraj; seraj@ualberta.ca

Received 25 June 2020; Revised 28 October 2020; Accepted 4 November 2020; Published 24 November 2020

Academic Editor: Chengxiang Zhuge



Microscopic modeling of mixed traffic (i.e., automaton-driven vehicles and human-driven vehicles) dynamics, particularly car-following, lane-changing, and gap-acceptance, provides the opportunity to gain a more accurate estimation of flow-density relationships for both traditional traffic with human-driven vehicles and different mixed traffic scenarios. Our paper proposes a microscopic framework to model multilane traffic for both vehicle types on shared roadways which sets the stage to explore the capability of macroscopic car-following models in general to explain the fundamental flow-density relationship. Since prior models inadequately represent the fundamental diagram realistically, we propose a rectified macroscopic flow model that can account for the impact of both lane-changing and gap-acceptance. Differentiability, boundary conditions, and flexibility of the proposed model are tested to validate its applicability. Finally, the capability to interpret the flow-density relationship by the proposed model is verified for different mixed traffic scenarios. Although few model parameter values were obtained directly from the simulation input, the rest of the parameters have been calibrated by flow and density outputs from the simulations. The analysis results show a distinct correlation between the proposed model parameters with automation-driven vehicle shares and lane-changing rates of traffic. The findings from this study emphasize the importance of taking complete motion dynamics into account, rather than partial motion dynamics (i.e., car-following) as has been the case in the previous studies, to explain macroscopic traffic flow characteristics, irrespective of the vehicle type.

## 1. Introduction

The rapid development of Connected and Automated Vehicle (CAV) technology has motivated researchers and practitioners to consider collaborative approaches, such as AutoDrive Challenge by SAE and General Motors [1], to solve traditional transportation problems. Consequently, researchers have been exploring traditional, fundamental concepts of traffic flow theories from new perspectives and questioning the understanding of the concepts in resulting scenarios. With an improved grasp on fundamental concepts of traffic flow and informed decision making, sustainable growth of CAV technology is possible, as is the identification of the optimal combinations of new technology with traditional transportation infrastructure. The resultant paradigm shifts on traffic operational technologies have prompted researchers to determine the limitations of partial motion dynamics (i.e., car-following) in explaining macroscopic features of multilane traffic flow and ask the question of how to complete vehicle motion dynamics could be considered in equilibrium flow models of traditional and mixed traffic. As such, this study attempts to determine the effectiveness of the macroscopic adaptation of a traditional microscopic car-following model (i.e., the Intelligent Driver Model (IDM) [2] in explaining fundamental features of traffic through detailed microscopic multilane modeling. More significantly, this study proposes a newly conceived, realistic traffic flow model that can (i) account for complete motion dynamics (i.e., car-following, lane-changing, and gap-acceptance) and (ii) interpret the transformation of the fundamental diagram in mixed traffic (i.e., Human-Driven (HuD) and Automaton-Driven (AuD) vehicles) scenarios.

A considerable issue in the research findings up to this point is that, while car-following is the dominant maneuver

in traffic flow, the capability of this microscopic feature to explain the relationship of flow with density is debatable. More promisingly, the IDM is a microscopic car-following model that aims to represent the car-following behavior through stimulus-response characteristics of human drivers. Since car-following models such as the IDM can describe steady-state and homogenous conditions, the macroscopic form can be attained from this model through local aggregation. Since IDM-prescribed acceleration is generated from both speed and headway, one can redevelop the fundamental diagram from a speed-spacing diagram of car-following [3]. The initial goal of this research, then, is to assess the proficiency of the derived fundamental diagram in reproducing an actual flow-density relationship. We, then, put forward an inclusive traffic flow model that reflects the complete motion dynamics of traffic, thereby addressing significant shortcomings in the previous efforts. We also explore the suitability of the proposed model to assess the remaining microscopic features. Finally, we analyze varying mixed traffic flow scenarios developed through the multilane microscopic modeling framework for adaptability attributes of the proposed model.

The rest of the paper is organized as follows: the literature review summarizes the findings from leading studies on macroscopic car-following models and their ability to explain the fundamental features of traffic. Additionally, articles exploring mixed traffic scenarios and their implications on both microscopic and macroscopic perspectives are reviewed. The following section provides a detailed description of the microscopic modeling framework that establishes the foundation of this comprehensive study. An in-depth discussion of the limitations of the macroscopic adaptation of the car-following model and a macroscopic model that accounts for complete motion dynamics of vehicles are, then, described in the next section. The functionality of the proposed model when identifying flow-density patterns is evaluated in the penultimate section, followed by the final section summarizing the research findings and contributions to the literature.

## 2. Literature Review

To review scholarly contributions made by previous studies that correlate with the present one, we chose to divide the review discussion into two segments and address each segment separately. The two segments of review discussion are (a) macroscopic fundamental relations from car-following models and their capability to explain the fundamental diagram and (b) microscopic and macroscopic aspects of mixed traffic. Some remarkable research efforts and studies that coincide with these two topics have been summarized in this section of the paper to provide insight into current research and existing groundwork available to support the research presented here.

Inspired by Greenshield et al.'s [4] single-regime continuity model, many models have been proposed to formulate speed-spacing (microscopic) or flow-density (macroscopic) relationships [5–17]. These models share two commonalities: (1) the variations of traffic states are explained from one specific perspective, either microscopic or macroscopic, and (2) the traffic is assumed to be homogenous. Under the assumption of homogenous traffic, microscopic models can be upscaled to macroscopic traffic flow models. Ni [18] showed this type of expansion by converting the IDM and longitudinal control model, which are microscopic car-following models, to macroscopic fundamental equations through local aggregation. The macroscopic IDM model is capable of generating a realistic flow-density diagram by employing four parameters to obtain a desirable shape with good fitting quality. While Ni did not explore the generated macroscopic models' competence in explaining traffic flow features, our study will utilize the macroscopic adaptation of the IDM for this direction of research. A reversed approach was taken by Duret et al. [19] where the authors simplified the fundamental diagram in an attempt to quantify individual vehicles' speed-spacing relationship. The results indicated that taking individual variability into account could improve the accuracy of vehicle trajectories. Chiabaut et al. [20] proposed a method to calibrate Newell's car-following model parameter [6], eventually assisting in calibrating the macroscopic model from individual observation. The influence of driver behavioral heterogeneity on the capacity drop was studied by Chen et al. [21]. The findings from the study concluded that both the lane-changing maneuver and variations in driver characteristics could reduce the bottleneck discharge rate by 8–23%. Treiber and Kesting [22] proposed a calibration and validation method for car-following models in the microscopic scale. However, the implications of calibrated parameters and its effect on best fit on the fundamental diagram remain unexplored. A similar approach to calibration and validation of car-following models was taken by Zheng et al. [23]. This study provided an intriguing insight into the validation and accuracy of car-following models by suggesting that the inclusion of a higher number of parameters in a car-following model could provide greater accuracy but might overfit in some scenarios. To address this issue, the authors suggested using different parameters and car-following models in different traffic conditions.

Due to the rapid expansion of CAV vehicle technology use, the coexistence of HuD with AuD on shared roadways have been studied extensively over the last few years. Liu et al. [24] proposed comprehensive instruction for modeling mixed traffic on multilane freeway facilities. This study specifically looked into mobility improvements brought upon by the introduction of AuD, warranting further study on fundamental features of mixed traffic. The authors were influenced by the modeling structure proposed in this study but pivoted the study focus towards macroscopic correlations of flow-density. An analytical capacity model was proposed by Gjaisi et al. [25] for mixed traffic. This study examined the different mixed traffic scenarios (i.e., varying demand, market shares, platooning, and technology use) to determine optimal AuD vehicle lanes for maximizing mixed traffic throughput. Based on a similar goal, Chen et al. [26] developed a general formulation to estimate the capacity of mixed traffic based on vehicle spacing, platoon configurations, and AuD shares. Gong and Du [27] developed a

cooperative platoon control based on a model predictive control strategy for mixed traffic that focused on reducing oscillation propagation and stabilizing traffic flow. A simulation-based traffic state estimation study was executed by Fountoulakis et al. [28] that demonstrated satisfactory estimation results for varying traffic state and penetration rates. Talebpour and Mahmassani [29] developed a simulation framework to analyze overall stability and throughput at different mixed traffic conditions. Additionally, the coexistence of HuD and AuD were found to be effective in preventing shockwave formation and propagation. Deng [30] proposed a simulation framework of mixed traffic incorporating human driving behavior from VISSIM and Adaptive Cruise Control and Cooperative Adaptive Cruise Control models from Van Arem et al. [31]. Ye and Yamamoto [32] proposed a modeling strategy to study the impact of dedicated AuD lanes on traffic flow throughput.

What is evident from the literature review is that a need exists to connect microscopic models with macroscopic models for both traditional and mixed traffic scenarios. While microscopic models could be accurate, to some extent, in estimating an individual vehicle's movement, the fit of these models to macroscopic scale requires significant extrapolation. Furthermore, although mixed traffic dynamics modeling has initiated considerable research efforts recently, major theoretical aspects that understand the true potential of CAV technology are yet to be explored. Critical reservation of mixed traffic-related studies would be a complete dependence on car-following models of hypothetical single-lane roadways. Hence, developing a traffic flow model that can incorporate the implications of complete motion dynamics of traffic movements will be the initial focus of this study and a significant effort to address the aforementioned shortcomings. At the same time, this study will present realistic, multilane, complex motion dynamics of both vehicle types and explore the resulting implications on the flow-density diagram, an important step in the study. Altogether, this study will attempt to specify the limitations of partial motion dynamics on a macroscopic scale and provide a reasonable approach to overcome those limitations.

## 3. Microscopic Modeling Framework of Mixed Multilane Traffic

To meet the developed objectives, we performed numerical simulations of 22 vehicles in a hypothetical homogenous two-lane roadway segment in MATLAB. At the beginning of the simulation, the number of vehicles varied between 8–12 vehicles in each lane. Significantly for our study, the simulated environment developed complete motion dynamics that included car-following, lane-changing, and gap-acceptance rules for individual vehicles corresponding to the vehicle type. All vehicles in the simulation revised their acceleration, lane-changing, and gap-acceptance decisions at each time-step with response to the surrounding vehicles' relative positions and velocities. Each simulation time-step duration was taken as 0.1 second (sec) of real-world time, and the simulation length was 9001 time-steps (0 to 9000) totaling 15 minutes of duration. First, two control-vehicles, one in each of the two lanes, followed a predefined acceleration profile. At the beginning of the simulation (0 time-step), all vehicles were placed on the roads by uniformly spacing them according to a simulated flow rate (from 1000 vehicles per hour to 2600 vehicles per hour) at 90 kilometers per hour (kph). The acceleration profiles of each lead vehicle were formed to generate bottlenecks for the following vehicles in each lane. At the 3001 time-step, the lead vehicles of each lane decelerated at the rate of $-2$ meter/sec$^2$ (m/s$^2$) to drop the speed of vehicles from 90 kph to 30 kph. This speed drop created an intentional bottleneck on the roadway segment where the following vehicles in each lane were unable to pass the lead vehicles. This provided an important opportunity to explore congested traffic states. Then, at the 4917 time-step, both lead vehicles accelerated at the rate of 2 m/s$^2$ to regain the speed from 30 kph to 90 kph.

*3.1. Car-following Principles.* Car-following, being the predominant maneuver during driving, has a greater influence on traffic states. In the proposed modeling framework, the car-following principles of vehicles primarily relied on paired vehicle types (e.g., HuD, AuD) of a subject vehicle and corresponding lead vehicle in the same lane of that subject vehicle. The HuD subject vehicles maintained the car-following IDM [2] irrespective of the lead vehicles type (i.e., HuD or AuD) for car-following maneuvers. The AuD vehicles adhered to car-following rules (i.e., Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC)) in response to the lead vehicle's type. If the lead vehicle was a HuD vehicle, then the AuD vehicles would follow ACC with relatively higher safety headway. Otherwise, the HuD vehicle would choose CACC and try to form a CACC platoon of AuD vehicles with the lead vehicle. The ideal ACC/CACC car-following model and parameter values were adopted from the work of Hu et al. [33]; an AuD vehicle would prefer to form CACC platoon with other AuD vehicles in the traffic stream. Platoon formation depended on the platoon configuration (i.e., intraplatoon headway, interplatoon headway, and maximum platoon length) (detailed descriptions of different platoon configurations and principles are available in the work of Seraj et al. [34]). Influenced by the findings from Seraj et al. [34], the present study employed a fixed platoon configuration (i.e., intraplatoon headway = 0.50 sec, interplatoon headway = 2.0 sec, and max. platoon length = 5 vehicles) to attain optimal mobility and safety benefits in mixed traffic environments.

In addition to usual car-following models, our modeling framework adopted a few subjective car-following models acquired from the IDM and ACC. The purpose of introducing these additional car-following models was to ensure a smooth adjustment to accommodate other two maneuvers (i.e., lane-changing and gap-acceptance) of vehicle motion dynamics. The preceding car-following (PCF) model was triggered for a subject vehicle prior to lane-changing when the vehicle was to change lanes but could not execute the maneuver due to lack of an adequate gap in the target lane. By contrast, the succeeding car-following (SCF) model was

activated to facilitate a subject vehicle's transition from the PCF principle right after the lane-changing maneuver. The accommodative car-following (ACF) model was used to enable a subject vehicle to increase spacing with its lead vehicle and accept a lane-changing vehicle from the nearby lane. The maximum duration a vehicle could remain in PCF, SCF, or ACF was chosen to be 5 sec (50 time-steps). Each vehicle had an acceleration equation specific to its' car-following state for each time-step. Relevant formulae to simulate vehicle movements are provided in Figure 1(a).

*3.2. Lane-Changing Principles.* The complex elements of motion dynamics (i.e., lane-changing and gap-acceptance) for vehicles had restricted application on the developed simulation structure. For HuD vehicles, only discretionary lane-changing that was motivated to gain free-flow speed was considered for modeling in this simulation. If a HuD vehicle were to drive below free-flow speed for 5 consecutive seconds, the vehicle driver would indicate the intention to change lane and start checking for an acceptable safety gap in the target lane. As for AuD vehicles, lane-changing intentions were only triggered by the possibility to form a CACC platoon with the lead vehicle in the target lane. However, an AuD vehicle would not intend to change lanes if it were already a part of the platoon. Both lane-changing and gap-acceptance rules developed for simulation were influenced by strategies proposed by Liu et al. [35] for mixed traffic movements. While the intentions for lane changes were conveyed as soon as the requirements were fulfilled, execution of lane-changing operations only occurred once the available gap in the target lane was governed by the gap-acceptance principles for respective vehicle types. In mixed traffic scenarios, vehicles were assigned with a Platoon Position ID ($P^2ID$) that indicated whether the vehicle was a part of CACC platoon, as well as identifying their physical position in that platoon. The $P^2ID$ for all HuD vehicles was assigned as zero (0). AuD vehicles' $P^2ID$ derived from the lead vehicle's type and $P^2ID$. For instance, if the lead vehicle in the same lane was a HuD vehicle, then the AuD vehicle's $P^2ID$ would be one "1," and it would follow ACC car-following rules. However, if the lead vehicle was an AuD vehicle, then the subject AuD vehicle would form a CACC platoon with the lead vehicle, and a $P^2ID$ would be assigned to the subject AuD vehicle according to the CACC platoon configuration (i.e., maximum platoon length). In another scenario, if the lead vehicle in the lane of an AuD vehicle were a HuD vehicle and the lead vehicle in the target lane was an AuD vehicle, the subject vehicle would attempt lane-changing from the current lane to the target lane, and the following vehicle in the target lane would try to safely receive the subject vehicle.

Since the motivation for discretionary lane-changing is not as strong as mandatory lane-changing, a conservative approach was taken to estimate required clearance for the maneuver by discounting extreme braking scenarios. A gap-searching subject vehicle would execute a lane-changing maneuver by implementing the following acceleration equation based on the vehicle type and start SCF from the next time-step.

$$a_{LC}(t) = \begin{cases} a_{\max}\left[1 - \left(\dfrac{v(t-1)}{v_f}\right)^4 - \left(\dfrac{s_0 + \max\left[0,\ v(t-1) \times T' + \left((v(t-1) \times \Delta v(t-1))/2\sqrt{a_{\max} b}\right)\right]}{s(t-1)}\right)^2\right], & \text{(for HuD)}, \\ k_1\left[\Delta p(t-1) - v(t-1) \times T' - s_0\right] + k_2 \Delta v(t-1), & \text{(for AuD)}, \end{cases} \tag{1}$$

and within 5-seconds of a lane-changing operation, the subject vehicle would reclaim its actual safety headway ($T$) and transfer to the car-following state according to the vehicle type and lead vehicle type. Meanwhile, the following vehicle would start adjusting the spacing with a new lead vehicle right after the lane-changing by the subject vehicle. A complete motion dynamic modeling scheme for both HuD and AuD vehicles are presented in Figures 1(b) and 1(c).

*3.3. Gap-Acceptance Principles.* The gap-acceptance of discretionary lane-changing for both vehicle types was based on the minimum accepted gap by the driver if the driver chose to merge into an available gap in the target lane. This gap must meet the following conditions to be considered a possibility for lane-changing by the subject vehicle:

$$\begin{aligned} G_{s,\min} &\geq 2g_0 + L, \\ G_{l,\min}, G_{f,\min} &\geq g_0. \end{aligned} \tag{2}$$

Here, $(G_{s,\min})$ = minimum total clearance in the target lane, $(G_{l,\min})$ = minimum lead clearance, and $(G_{f,\min})$ = minimum following clearance. Upon satisfying the minimum boundary conditions, the total available clearance in the target lane, between the lead and following vehicles, would be compared with the acceptable gap of the subject vehicle for the corresponding time-step. If the subject vehicle had a total gap in the target lane smaller than an acceptable lane-changing gap, the subject vehicle would move to PCF (provided that the vehicle was not in PCF for last 5-second time-step). The subject vehicle would stay on PCF for the maximum 5-second time-step and actively examine the available gap in the target lane to execute a lane-changing

- Intelligent driver model (IDM)

$$a(t) = a_{\max}\left[1 - \left(\frac{v(t-1)}{v_f}\right)^4 - \left(\frac{s_0 + \max\left[0, v(t-1)\times T + \frac{v(t-1)\times \Delta v(t-1)}{2\sqrt{a_{max} b}}\right]}{s(t-1)}\right)^2\right]$$

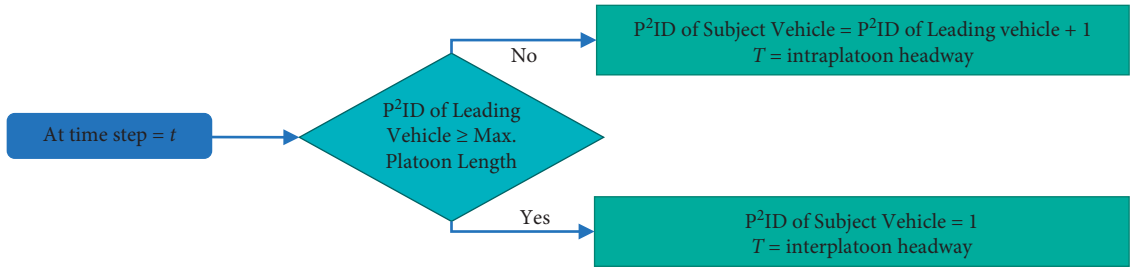
$a(t)$ = acceleration at time step $t$
$a_{\max}$ = maximum acceleration, 2.5 m/s$^2$
$v(t-1)$ = speed at time step $(t-1)$
$v_f$ = free flow speed
$s_0$ = minimum spacing (7.5 m) = $g_0 + L$
$g_0$ = jam clearance, 2 m
L = average vehicle length, 5.5 m
$\Delta v(t-1)$ = speed difference with lead vehicle at $(t-1)$
b = desirable deceleration, 2.5 m/s$^2$
$s(t-1)$ = clearance with lead vehicle at $(t-1)$
$T$ = safety head way, for a HuD vehicle T is fixed throughout the simulation period and varied randomly from 2 to 3 sec.

- Adaptive cruise control (ACC)

$a(t) = k_1[\Delta p(t-1) - v(t-1)\times T - s_0] + k\Delta v(t-1)$
$k_1, k_2$ = control constants for relative distance and relative velocity respectively ($k_1 = 0.3$, $k_2 = 0.5$)
$T$ = safety headway = 1.5 sec for vehicles at ACC

- Cooperative adaptive cruise control (CACC)

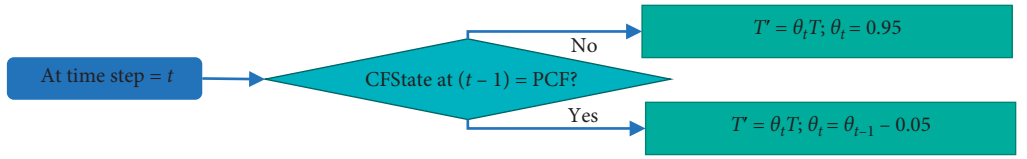
$a(t)$ = same as ACC with $T$ (safety headway) according to Intra/Inter platoon headway




- Preceding car-following (PCF)

✓ For HuD
$a(t)$ = same as IDM

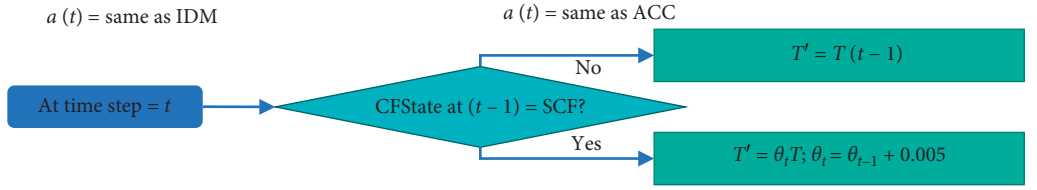
✓ For AuD
$a(t)$ = same as ACC




- Succeeding car-following (SCF)

✓ For HuD
$a(t)$ = same as IDM

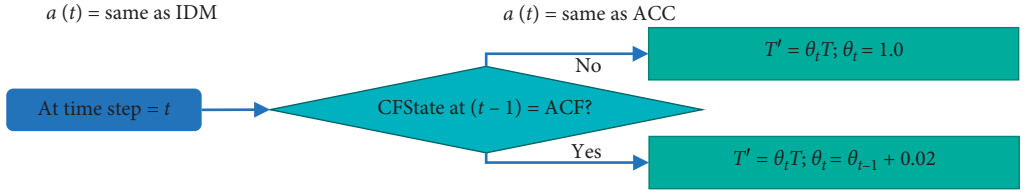
✓ For AuD
$a(t)$ = same as ACC




- Accommodating car-following (SCF)

✓ For HuD
$a(t)$ = same as IDM

✓ For AuD
$a(t)$ = same as ACC




- Lane-changing (LC)

✓ For HuD
$a(t)$ = same as IDM

✓ For AuD
$a(t)$ = same as ACC

(a)

Figure 1: Continued.

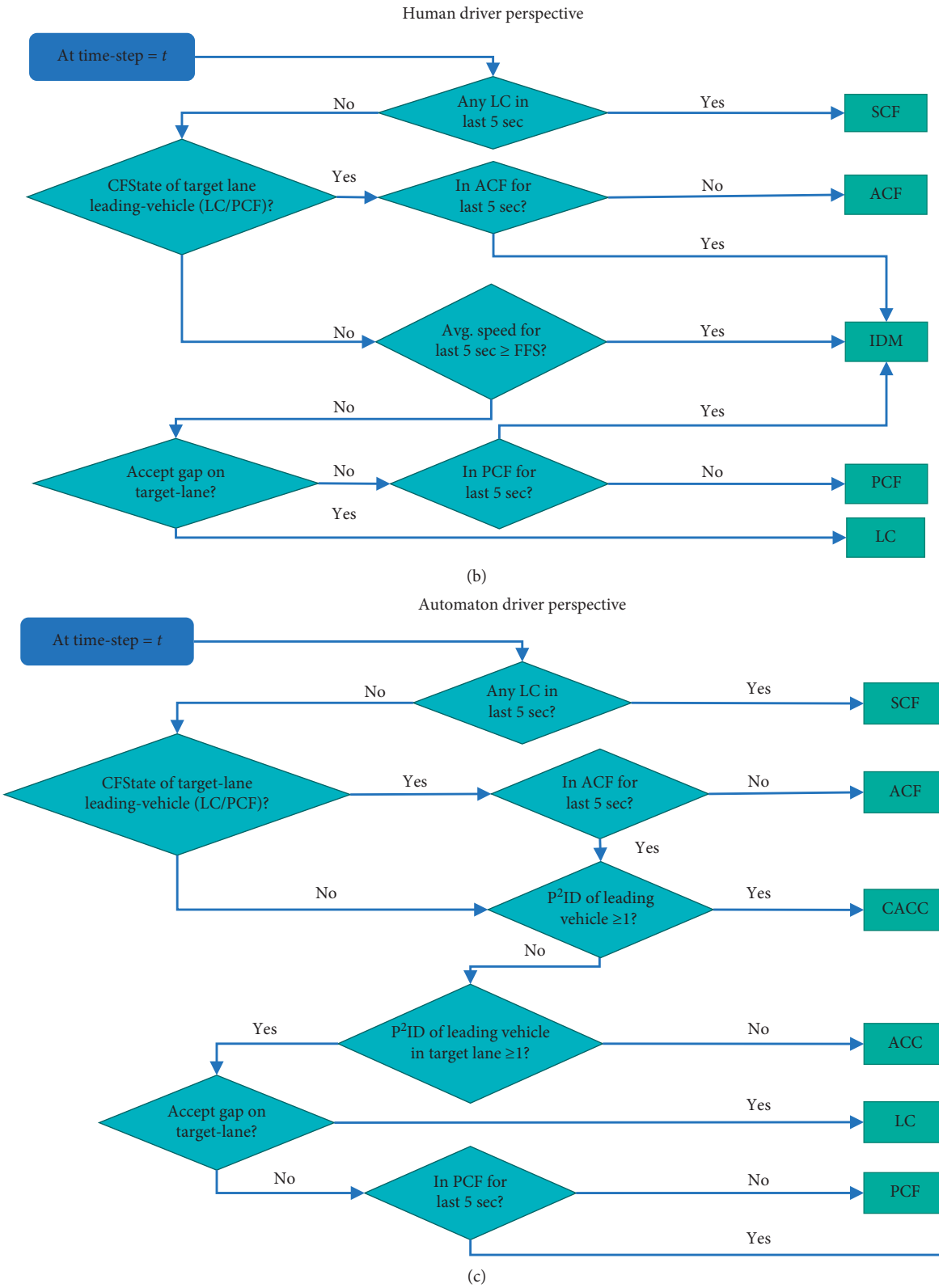



Figure 1: (a) Implemented formulae for the modeling framework, (b) principles for a human driver, and (c) principles for an automaton driver to model multilane traffic movements.

operation. If the vehicle could not find an acceptable gap within this 5-second time-step in PCF, it would move back to the primary car-following mode (i.e., IDM, ACC). The subject vehicle would, then, seek the following conditions to identify an acceptable gap ($G_{s,\text{acc}}$) for lane-changing at any time-step:

$$\begin{aligned} G_{s,\text{acc}}(t') &= \max\left\{G_{s,\min}, G_{s,\text{req}}(t')\right\}, \\ G_{s,\text{req}}(t') &= \left[v_s(t') + \Delta v_f(t')\right]T_s' + \left[v_s(t') + \Delta v_l(t')\right]T_s' + L, \\ \Delta v_l &= \left[v_s(t') - v_l(t')\right], \\ \Delta v_f &= \left[v_s(t') - v_f(t')\right], \end{aligned} \tag{3}$$

where $G_{s,\text{acc}}$ = acceptable clearance, $t'$ = time-step in PCF, $s$ = subject vehicle, $l$ = lead vehicle in the target lane, $f$ = following vehicle in the target lane, $G_{s,\text{req}}$ = required clearance, $\Delta v$ = speed difference, and $T_s'$ = modified safety headway of the subject vehicle.

Similarly, if a subject vehicle observed that a lead vehicle in a neighbour lane had activated PCF (assumed to be conveyed by indicator light), the subject vehicle would move to the ACF mode for a maximum 5-second time-step to receive that vehicle as a new lead vehicle. Upon successful lane-changing by the lead vehicle in the target lane within the 5-second time-step, the subject vehicle would shift to the PCF mode specific to the vehicle type. Otherwise, the subject vehicle would automatically move back to the primary car-following mode (i.e., IDM, ACC) after the 5-second time-step in ACF. Within this 5-second window, the safety headway ($T'$) for the subject vehicle would change gradually to allow for a smaller acceptable gap for lane-changing (in PCF) to broaden the gap between the lead vehicle and itself in the same lane to accommodate the lane-changing of the lead vehicle from a neighbor lane (in ACF) and to increase the gap between a new lead vehicle (in SCF) and itself. The safety headway would meet the conditions mentioned in Figure 1(a), specific to subjective car-following modes. If a vehicle can successfully execute a lane-changing maneuver, the vehicle would shift to SCF in the following time-step and stay on that mode until it reclaimed its initial safety headway ($T$).

## 4. Flow-Density Relationships for a Human-Driven Vehicle

*4.1. Limitations of the Macroscopic Adaptation of the Car-following Model.* In a real-world scenario, car-following action is the most frequent task performed by drivers in vehicle motion. Therefore, the amplification of a microscopic car-following model for homogenous traffic in a state of steady equilibrium should provide an ideal representation of the fundamental diagram. In this regard, the IDM was expanded to obtain traffic flow models between speed and density. The following forms of equation were obtained [18]:

$$\begin{aligned} k &= \frac{1}{(s_0 + vT)\left[1 - \left(v/v_f\right)^{\delta}\right]^{-(1/2)}}, \\ v &= v_f\left[1 - \left(\frac{k}{k^*}\right)^2\right], \\ q &= \frac{v}{(s_0 + vT)\left[1 - \left(v/v_f\right)^{\delta}\right]^{-(1/2)}}, \end{aligned} \tag{4}$$

where $k$ = mean density, $s_0$ = minimum spacing, $v_f$ = free-flow speed, $k^*$ = desired mean density, and $q$ = mean traffic flow. The macroscopic form of the IDM involved four parameters (i.e., $v_f, s_0, T, \delta$) that determine the shapes and features of fundamental relations (i.e., flow, speed, and density). To check the capability of the derived macroscopic model to explain these relations, a wide range of traffic flow scenarios (i.e., flow rates = 1000 vph–2600 vph) was simulated under the proposed modeling framework containing only HuD vehicles in the traffic stream. Measured outputs from the simulations were computed and calibrated to obtain flow-density values and model parameters, respectively. Macroscopic parameter values were measured according to the techniques proposed in the Highway Capacity Manual [36]. Flow and respective density values for different simulated traffic flow scenarios are plotted in Figure 2. The macroscopic form of the IDM is also plotted in the same figure with the premeditated parameter values (i.e., $(s_0 = 7.5\,\text{m},\ v_f = 90\,\text{kph},\ T = 1.98\,\text{sec},\ \delta = 4)$ used for the simulation.

Since the parameters required to generate the Fundamental Diagram (FD), derived from the IDM, were obtained from the parameters provided as inputs of the car-following IDM in the simulation, calibration of the parameters was exempt. Therefore, the FD obtained from the given parameter set should represent the real-world implications of an expanded IDM on the flow-density relation. However, as observed in Figure 2(a), the plotted FD, derived from IDM, did not capture the extremity and variations observed from the simulated flow-density results. Analysis of the simulated traffic's microscopic features revealed that, for 90.73% of the total duration, the vehicles maintained the car-following state (i.e., IDM). The remaining share of vehicle movements was spent on lane-changing-related motion dynamics (i.e., LC, PCF, SCF, and ACF). The comparison between the flow-density points from the simulation with the generated FD revealed that the lane-changing gap-acceptance control decisions had an influential impact from a macroscopic perspective. While car-following was a predominant maneuver on the microscopic scale, the impact of lane-changing gap-acceptance was far more significant from a macroscopic viewpoint.

To further investigate the capability of the FD derived from the IDM, we explored empirical data obtained from a recurrent bottleneck segment of an urban freeway corridor: Whitemud Drive, Edmonton, Canada. Whitemud Drive is a multilane freeway with a posted speed limit of 80 kph and

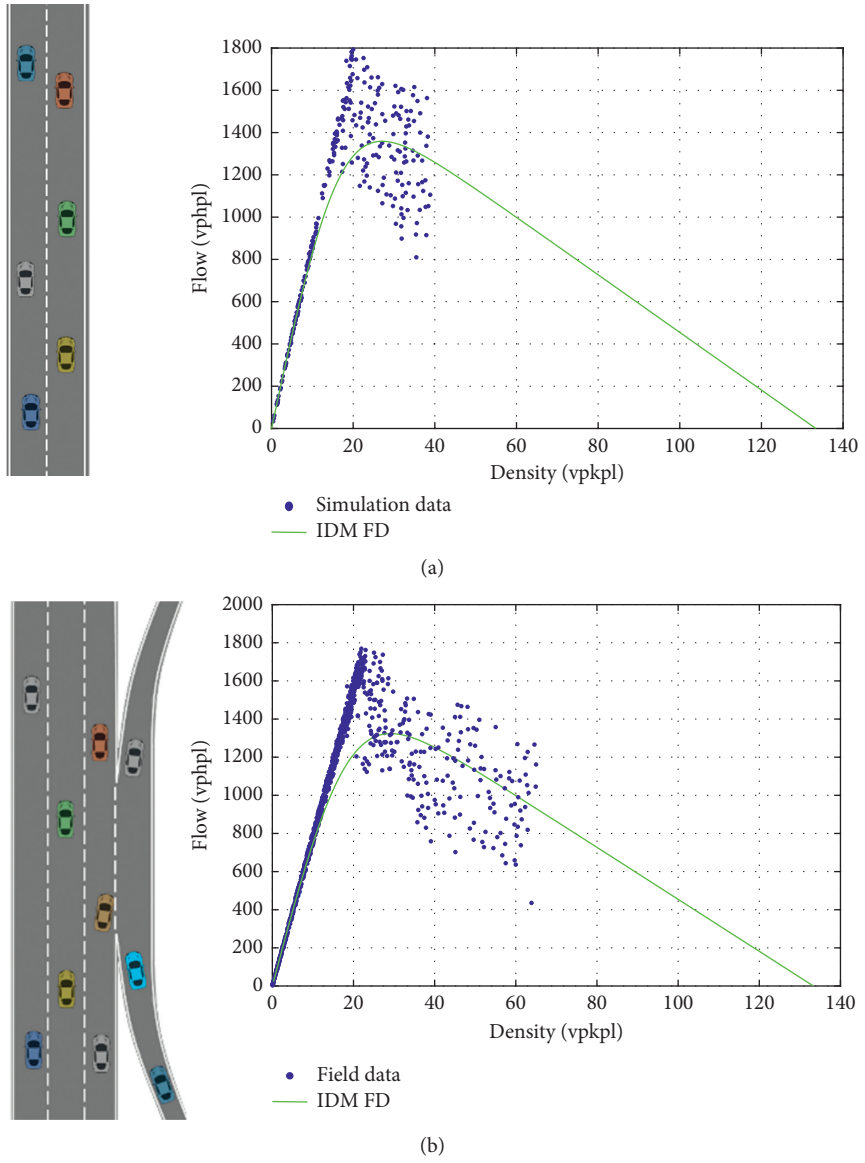



Figure 2: Resulting fundamental diagrams by the macroscopic IDM from (a) simulation data and (b) empirical data.

serves as a part of Edmonton's inner ring road. The annual, average daily traffic (AADT) of its westbound section was greater than 100,000 vehicles in 2017 [37]. For the study purpose, we have used 24 hr data from 30 workdays (from August 7, 2017, to September 18, 2017). Based on field observations and bottleneck information, the weaving segment after the 122 Street on-ramp and the segment around Fox Drive were selected as critical segments for this study. The selected segments were, then, equipped with loop detectors on each lane to collect traffic flow data (i.e., speed, flow, and occupancy) at 20-second intervals. The studied roadway segment had four (4) lanes with a one-lane on-ramp at the beginning and a one-lane off-ramp at the end of the segment. The free-flow speed for the FD was taken to be 80 kph since this was the maximum speed limit of the freeway. The minimum spacing was taken as 7.5 m (5.5 m (average vehicle length) + 2 m (safety clearance)), and $\delta = 4$ was considered. The safety headway ($T = 1.98$ sec) for traffic was calibrated for the FD using empirical data points. Analogous to earlier observations, the FD failed to capture the entirety of the fundamental relation between flow and density. The FD provided a lower capacity value and failed to explain the capacity drop phenomenon. Therefore, traffic flow models, by expanding the car-following dynamic, do not adequately explain most of the traffic flow states and events. While the microscopic car-following model considered vehicle-to-vehicle interactions in the same lane, this model failed to apprehend interactions of vehicles in multiple lanes. As such, FDs derived from car-following models cannot be expected to comprehensively interpret traffic flow characteristics. In that context, a macroscopic model can only consider safety spacing during car-following and accommodate speed, as well as space sensitivity, from drivers relevant to lane-changing and gap-acceptance.

*4.2. Proposed Traffic Flow Model.* Premised on these findings, a new, flexible, and integrated traffic flow model can now be formulated. In general, the proposed model must consist of a safety spacing principle, predominant in car-following dynamics, while integrating broad-scale speed and space sensitivity, which are prevailing aspects on lane-changing and gap-acceptance. These considerations are incorporated when reviewing numerous forms of traffic flow and result in the following formulation:

$$k = \frac{1}{\left(\left(s_0 + vT + \lambda v^2\right)\left[1 - \ln\left(1 - \left(v/v_f\right)\right)\right]^{(1/\eta)}\right)},$$

$$v = v_f\left[1 - e^{\{1-(k^*/k)\}^{\eta}}\right], \tag{5}$$

$$q = \frac{v}{\left(\left(s_0 + vT + \lambda v^2\right)\left[1 - \ln\left(1 - \left(v/v_f\right)\right)\right]^{(1/\eta)}\right)},$$

where $v$ = mean speed of traffic, $v_f$ = free-flow speed, $k$ = mean density, and $k^*$ = the desired mean density. The latter, considering traffic safety rules, then takes the following form:

$$k^* = \frac{1}{s_0 + vT + \lambda v^2}. \tag{6}$$

Here, $\lambda$ = represents the awareness of the mean speed of traffic; $T$ = the mean safety headway of the driving population; $s_0$ = minimum spacing; and $\eta$ = represents sensitivity to the average spacing of traffic.

Since simulated, discretionary lane-changing was assumed to be motivated by speed gain and successful lane-changing maneuvers depend on the available spacing between vehicles, and the new parameters (i.e., $\lambda$ and $\eta$) should account for and relate to the lane-changing of vehicles in traffic.

*4.3. Differentiability of the Proposed Traffic Flow Model.* The ability of the proposed model to explain the flow-density relationship can be further analyzed to generate sensible information. For instance, one can take the first derivatives of flow ($q$) with respect to velocity ($v$) to find the capacity and set the result to zero.

$$\frac{dq}{dv} = k + v\frac{dk}{dv} = v\left[-\frac{(T + 2\lambda v)\left[1 - \ln\left(1 - \left(v/v_f\right)\right)\right]^{-(1/\eta)}}{\left(s_0 + vT + \lambda v^2\right)^2} - \frac{\left[1 - \ln\left(1 - \left(v/v_f\right)\right)\right]^{-1-(1/\eta)}}{\eta v_f\left(1 - \left(v/v_f\right)\right)\left(s_0 + vT + \lambda v^2\right)}\right] + \frac{\left[1 - \ln\left(1 - \left(v/v_f\right)\right)\right]^{-(1/\eta)}}{\left(s_0 + vT + \lambda v^2\right)}. \tag{7}$$

Additionally, the traffic flow ($q$) can be differentiated with respect to density ($dq/dk$) to obtain the slope of a tangent at any point on a flow-density curve. Hence,

$$\frac{dq}{dk} = v + k\frac{dv}{dk} = v + \frac{k}{(dk/dv)}$$

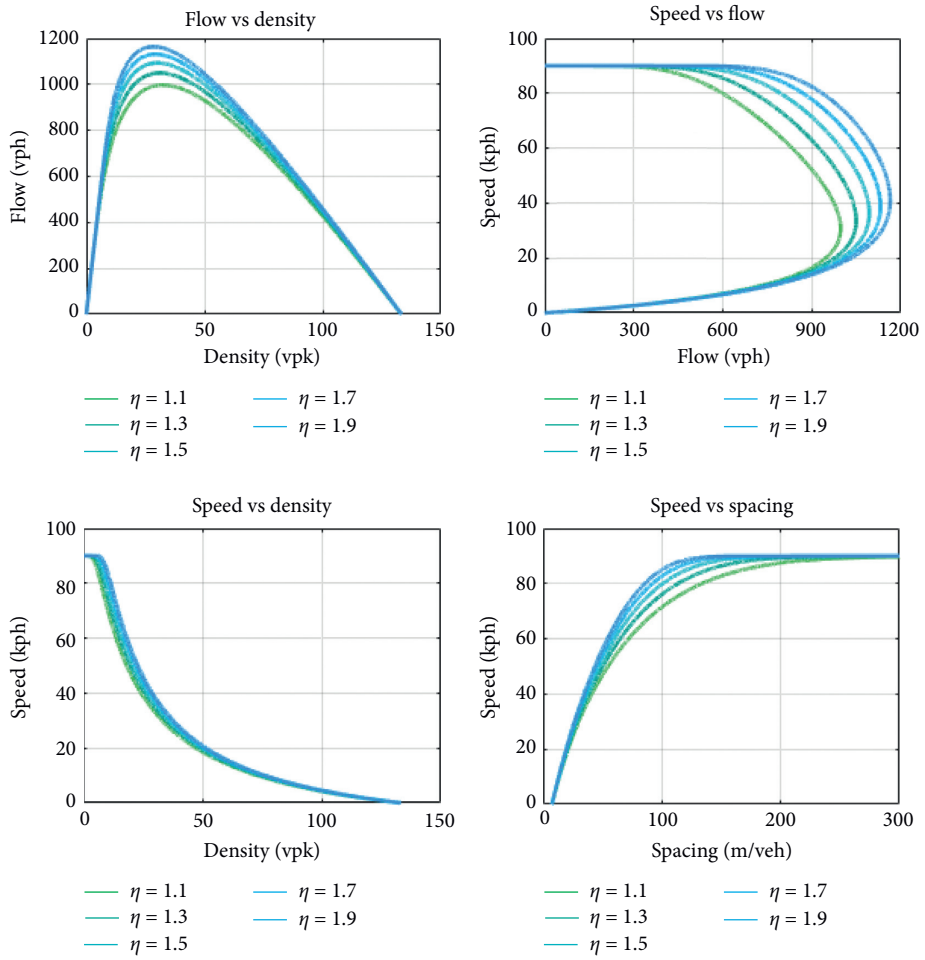
$$= \frac{\left[1 - \ln\left(1 - \left(v/v_f\right)\right)\right]^{-(1/\eta)}}{\left(s_0 + vT + \lambda v^2\right)\left[-\left(\left((T + 2\lambda v)\left[1 - \ln\left(1 - \left(v/v_f\right)\right)\right]^{-(1/\eta)}\right)/\left(s_0 + vT + \lambda v^2\right)^2 - \left(\left[1 - \ln\left(1 - \left(v/v_f\right)\right)\right]^{-1-(1/\eta)}/\eta v_f\left(1 - \left(v/v_f\right)\right)\left(s_0 + vT + \lambda v^2\right)\right)\right]} + v. \tag{8}$$

Moreover, as a boundary condition of the flow-density relationship, shockwave speed ($w_j$) at density can be determined from the abovementioned equation:

$$w_j = \frac{dq}{dk}|k = k_j, v = 0 = -\frac{s_0}{T + \left(s_0/\eta v_f\right)}. \tag{9}$$

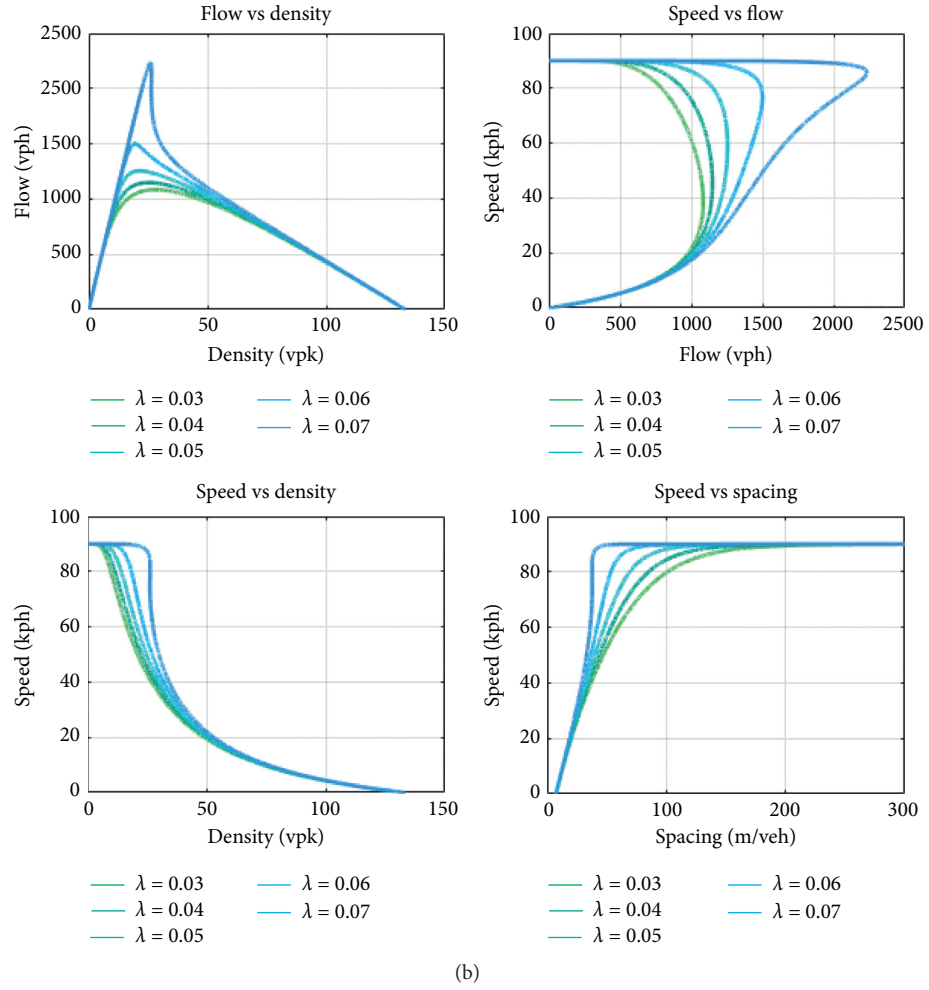
*4.4. Sensitivity of the Proposed Traffic Flow Model Parameters.* An expected feature of the traffic flow model is the flexibility to explain different shapes of fundamental diagrams from a single model. Two additional parameters introduced in the proposed model (i.e., $\lambda$ and $\eta$) provide greater flexibility and precision in replicating wide variations of flow-density relationships. Figure 3 illustrates a group of fundamental diagrams generated from the proposed model with varying speed and spacing sensitivity parameters. Figure 3 establishes that both speed and spacing sensitivity can effectively influence the shape of the flow-density relationship and produce diverse shapes based on empirical data. However, the interaction of speed and spacing sensitivity with other parameters in the model requires further analysis.

*4.5. Applicability on Simulated and Empirical Observations.* With the potential of the proposed model now established, its proficiency in portraying fundamental features of traffic,




(a)

Figure 3: Continued.




(b)

Figure 3: Variability of the fundamental diagram resulting from changes in the (a) spacing sensitivity parameter and (b) speed sensitivity parameter.

for both simulated and empirical data, was tested and compared with the macroscopic IDM. Common parameters for both models (i.e., $s_0 = 7.5$ m, $v_f = 90$ kph, and $T = 1.98$ sec) were maintained for fundamental diagram calibration. As shown in Figure 4, the fundamental diagrams generated through the proposed model were a better fit than the macroscopic extrapolation of the IDM. The Root Mean Square Error (RMSE) was reduced by 42.67% for the simulation data and by 37.41% for the field data in comparison to the IDM. Additionally, the congestion element to the fundamental diagram of the proposed model had curvilinear shapes that were instrumental in explaining the capacity drop phenomenon. A similar output was experienced for empirical observations, which established the absolute superiority of the proposed model over macroscopic adaptation of the IDM with respect to the competence of explaining traffic flow features.

*4.6. Flow-Density Relationship for Mixed Traffic.* The capability of the proposed model to apprehend the macroscopic implications of complete motion dynamics has been established in the previous section of this paper. We will now explore the model's adeptness to explain flow-density relationships for different mixed traffic scenarios and correlations of model parameters with AuD vehicle shares and lane-changing rates in this section. Our mixed traffic scenarios, for this study, were developed considering different AuD vehicle shares in simulated traffic while adhering to the aforementioned traffic dynamic principles (i.e., car-following, lane-changing, and gap-acceptance) according to vehicle types (i.e., AuD and HuD vehicle). The obtained results from the simulation were utilized to calibrate the model parameters and plot the flow-density diagrams for different mixed traffic conditions; Table 1 and Figure 5. The flexibility of the proposed model allowed for more precise replication of a wide variety of fundamental relations. Additionally, both parameters introduced in the proposed model demonstrated their sensitivity to different traffic conditions.

As described in the section on the microscopic modeling framework of mixed multilane traffic, two types of vehicle were simulated by following the developed car-following, lane-changing, and gap-acceptance principles for varying flow rates (i.e., 1000–2600 vphpl) and specific AuD vehicle shares. Microscopic parameters of individual vehicles were

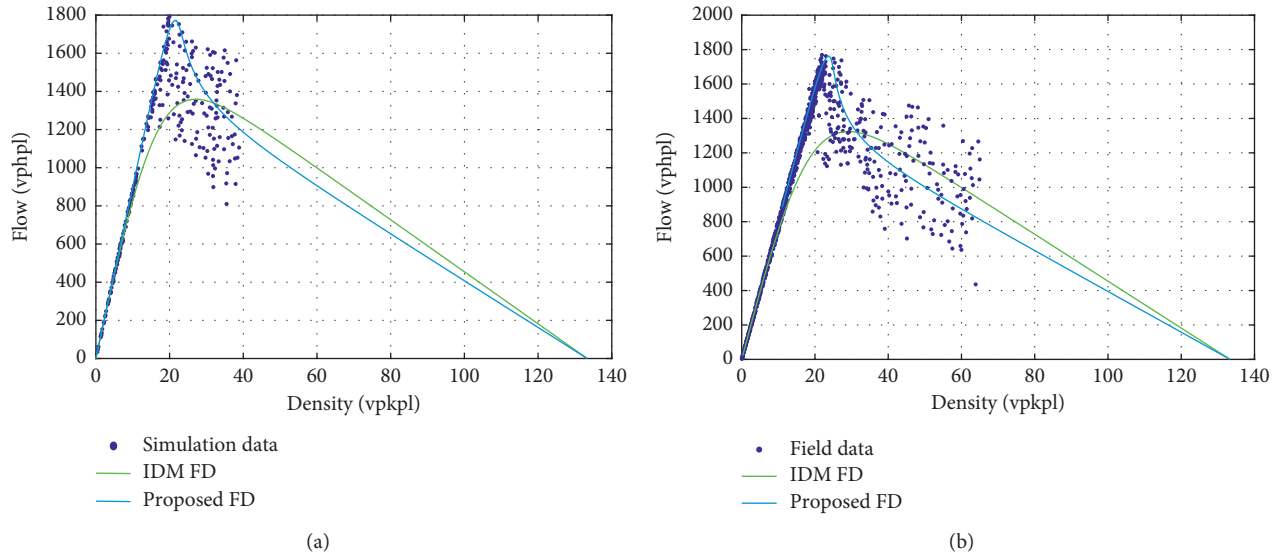



Figure 4: Comparison of the calibrated fundamental diagrams resulting from the macroscopic IDM and proposed traffic flow model on (a) simulation data and (b) empirical data.

Table 1: Model parameters of the proposed model for distinct mixed traffic scenarios.

| AuD share (%) | $s_0$ (m) | $v_f$ (kph) | $T$ (sec) | $\lambda$ (s$^2$/m) | $\eta$ |
|---|---|---|---|---|---|
| 0 | 7.5 | 89.86 | 1.98 | −0.0668 | 1.349 |
| 25 | 7.5 | 87.89 | 1.76 | −0.0617 | 1.401 |
| 50 | 7.5 | 88.16 | 1.46 | −0.0523 | 1.470 |
| 75 | 7.5 | 88.50 | 1.31 | −0.0447 | 1.690 |
| 100 | 7.5 | 90.00 | 1.18 | −0.0394 | 1.849 |

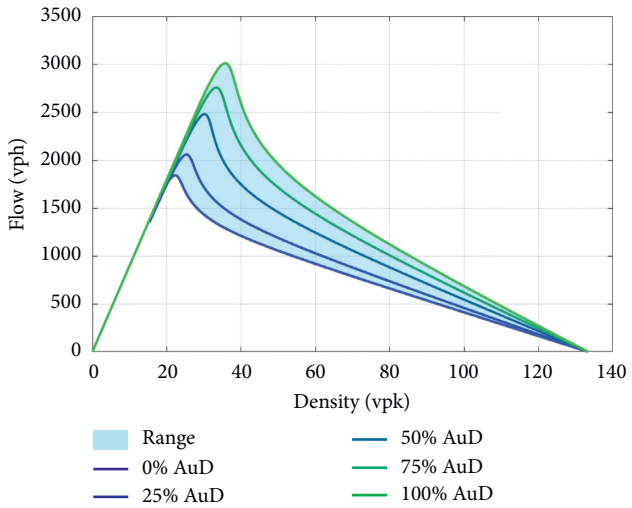



Figure 5: Resulting fundamental diagrams for distinct mixed traffic conditions generated from numerical simulations.

recorded throughout the simulation period for specific simulation scenarios (i.e., flow rate and AuD share). These parameters included position, velocity, acceleration, and current lane. The average number of lane changes by the vehicles in traffic was measured from assessing these features. This average was presumed to have a notable influence on traffic flow and was, therefore, measured for each simulation scenario. The recorded microscopic features were, then, translated into macroscopic traffic flow parameters for the simulated roadway segment. The flow, density, and average speed of traffic were calculated from microscopic data for specific AuD market shares and plotted to develop a traffic condition-specific flow-density diagram Figure 5. The calibrated flow-density curves demonstrated the combinatorial impact of introduced parameters (i.e., $\lambda$ and $\eta$) in effectively portraying the correlation between traffic flow parameters.

To demonstrate the implications of lane-changing on mixed traffic scenarios, the interactions of the lane-changing rate with increasing traffic flow rates were assessed. The lane-changing rate (number of lane changes/minute) of traffic was measured by counting the total number lane changes by simulated traffic (22 vehicles) over the simulation period in minutes (15 minutes). The simulated environment provided the opportunity to precisely measure every instance of lane-changing by the vehicles and otherwise a difficult task when working with real-world traffic. Results showed a strong linear correlation of the lane-changing rate with increasing flow rates Figure 6. The observations outlined in Figure 6 show the regression line for 0% displaying an upward trend with increasing flow rates. However, the higher residuals indicated an unstable correlation between the variables at 0% AuD traffic. Further analysis demonstrated a reduction in residual values with an increased number of AuD vehicles in traffic. Moreover, a gradual decrease in the lane-changing rate with increased AuD vehicle shares was observed. Although the extent of reduction in lane-changing rates varied for different flow rates, the trend of diminishing lane-changing was notable. The result of this particular analysis warrants future examination of lane-changing demand and supply equilibrium in different mixed traffic scenarios.

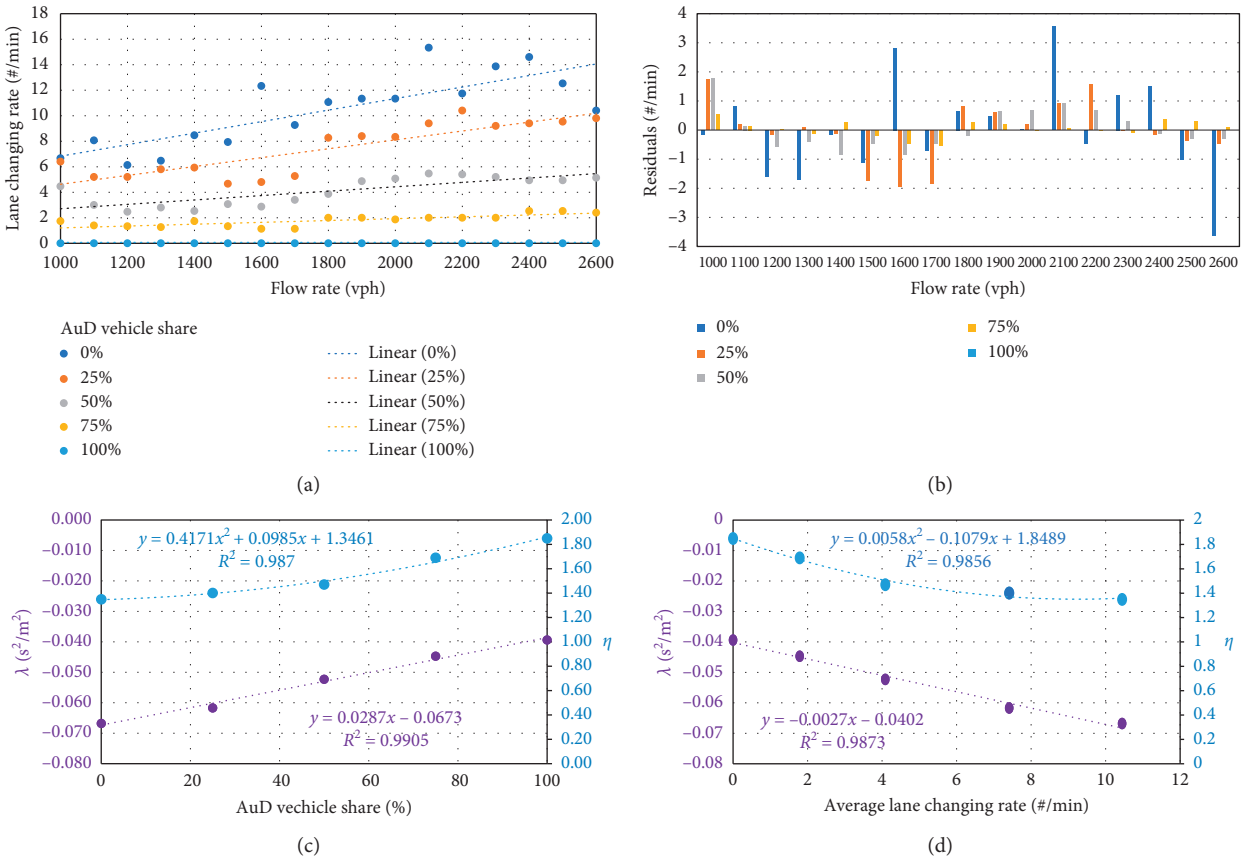



FIGURE 6: (a) Patterns of lane-changing rates, (b) regression residuals with increasing traffic flow rates at increasing flow rates for different mixed traffic conditions, (c) interactions of the introduced model parameters with AuD vehicle shares in traffic, and (d) average lane-changing rates.

To justify the inclusion of speed and space sensitivity in the traffic flow model, the correlation of both parameters with changing traffic conditions was analyzed. Analysis of the flow-density relationship in mixed traffic scenarios demonstrated the evolution of the fundamental diagram, plotted by the proposed macroscopic model, with an increasing share of AuD vehicles. While minimum spacing ($s_0$) was predefined for the simulation, free-flow speed ($v_f$) and safety headway of traffic ($T$) were computed from the simulation outcomes. Afterwards, the speed and spacing sensitivity parameter ($\lambda$ and $\eta$) were calibrated to best fit the obtained flow-density points of different mixed traffic scenarios. This study used neutral regression by normalizing the RSME of two fundamental parameters (flow and density) to minimize overall errors. Analysis of the best fit of $\lambda$ and $\eta$ values uncovered their correlation with AuD vehicle shares; Figure 6(c). While $\lambda$ was linearly correlated with the simulated mixed traffic scenarios (i.e., 0%, 25%, 50%, 75%, and 100% AuD vehicle shares), $\eta$ showed a quadratic correlation with AuD vehicle shares. High $R^2$ values for both cases established a strong correlation between $\lambda$ and $\eta$ with AuD vehicles' presence in traffic. Further analysis of $\lambda$ and $\eta$ values at different AuD vehicle shares proved their association with lane-changing maneuvers; Figure 6(d).

## 5. Concluding Remarks

This paper has considered the fundamental drawbacks to the conversion of a microscopic model into a macroscopic model. In place of this, we have proposed an efficient, flexible traffic flow model that is capable of capturing limits and variations of fundamental diagrams. The proposed model is formulated by taking the impact of vehicles' complete motion dynamics into account. While the proposed model was portrayed in the macroscopic scale, it also balanced the microscopic vehicle dynamics through speed and spacing sensitivity parameters that gave the model better adaptability to recognize the fundamental features with greater precision. Proficiency of the proposed model was compared with the macroscopic car-following model by fitting both simulation and empirical data. The flexibility of the proposed model was also evaluated by calibrating the model parameters to shape the fundamental diagrams for different mixed traffic conditions.

Considering the magnitude of research warranted by the identified issues from previous research, this study makes notable contributions to the transportation literature by (i) revealing the limitations of the macroscopic car-following model to explain the fundamental features of multilane traffic flow, (ii) establishing the prominence and capability of

a single-regime equilibrium traffic flow model to account for the complete motion dynamics of traffic, (iii) capturing the variations of the fundamental diagram with different mixed traffic scenarios through single traffic flow models, and (iv) explaining the correlation of model parameters with vehicle motion dynamics and AuD shares on traffic. Given these key contributions, the most substantial theoretical significance of this study is in addressing the highly contested explanation of the macroscopic fundamental diagram by exploring the implications of the lane-changing maneuver. Although lane-changing has long been proposed as an elementary source of disturbance with broader consequences on traffic, surprisingly little empirical evidence that meets scientific standards was available to support this claim. Hence, this study carefully examined the macroscopic implications of complete motion dynamics to address the considerable gap in the existing body of knowledge. What we found was that incorporating the influences of bidirectional movements of traffic through speed and space sensitivity parameters would promote more accurate estimation of the fundamental diagram while providing adequate flexibility to adapt with varying mixed traffic scenarios.

The unique properties characterizing the proposed traffic flow model facilitate the replication of numerous traditional and mixed traffic scenarios. This substantially improved model can successfully explain the roadway capacity and capacity drop phenomenon, thus recognizing both opportunities and limitations of the roadway. Furthermore, leveraging the proposed model can facilitate more advanced and accurate highway performance functions to assist dynamic route choice on a traffic network. A future extension of this study will be to examine the applicability of the proposed model in signalized, arterial traffic and modifications required when considering the implications of a traffic signal on traffic flow. Besides, the lane-changing implications on traffic flow at varying mixed traffic scenarios will be further explored.

## Data Availability

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

## Conflicts of Interest

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

## Acknowledgments

The authors want to acknowledge the support from the Natural Sciences and Engineering Research Council (NSERC) of Canada, City of Edmonton, and Transport Canada for funding this research. The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of NSERC, City of Edmonton, and Transport Canada. The authors also acknowledge the contributions by Dr. Sharon Harper for her assistance in editing and proofreading the paper. This research work was jointly supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, City of Edmonton, and Transport Canada.

## References

[1] SAE International, *AutoDrive Challenge*, SAE International, Warrendale, PA, USA, 2016, https://autodrivechallenge.com/.

[2] 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.

[3] V. L. Knoop, *Introduction to Traffic Flow Theory: An Introduction with Exercises*, Delft University of Technology, Delft, Netherlands, 2017.

[4] B. D. Greenshields, J. R. Bibbins, W. S. Channing, and H. H. Miller, "A study of traffic capacity," *Highway Research Board Proceedings*, vol. 14, 1935.

[5] H. Greenberg, "An analysis of traffic flow," *Operational Research*, vol. 7, no. 1, pp. 79–85, 1959.

[6] G. F. Newell, "Nonlinear effects in the dynamics of car following," *Operational Research*, vol. 9, no. 2, pp. 209–229, 1961.

[7] L. C. Edie, "Car-following and steady-state theory for noncongested traffic," *Operational Research*, vol. 9, no. 1, pp. 66–76, 1961.

[8] B. S. Kerner and P. Konhäuser, "Structure and parameters of clusters in traffic flow," *Physical Review E*, vol. 50, no. 1, pp. 54–83, 1994.

[9] J. M. D. Castillo and F. G. Benítez, ""On the functional form of the speed-density relationship-I: general theory," *Transportation Research Part B: Methodological*, vol. 29, no. 5, pp. 373–389, 1995.

[10] N. Wu, "A new approach for modeling of fundamental diagrams," *Transportation Research Part A*, vol. 36, no. 10, pp. 867–884, 2002.

[11] Y. Ji, W. Daamen, S. Hoogendoorn, S. Hoogendoorn-Lanser, and X. Qian, "Investigating the shape of the macroscopic fundamental diagram using simulation data," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2161, no. 1, pp. 40–48, 2010.

[12] H. Wang, J. Li, Q. Chen, and D. Ni, ""Logistic modeling of the equilibrium speed-density relationship," *Transportation Research Part A*, vol. 45, no. 6, pp. 554–566, 2011.

[13] J. Li and H. M. Zhang, "Fundamental diagram of traffic flow," *Transportation Research Record*, vol. 2260, no. 1, pp. 50–59, 2011.

[14] X. Wu, H. X. Liu, and N. Geroliminis, "An empirical analysis on the arterial fundamental diagram," *Transportation Research Part B*, vol. 45, no. 1, pp. 255–266, 2011.

[15] M. Keyvan-Ekbatani, A. Kouvelas, I. Papamichail, and M. Papageorgiou, "Exploiting the fundamental diagram of urban networks for feedback-based gating," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1393–1403, 2012.

[16] M. Keyvan-Ekbatani, M. Papageorgiou, and I. Papamichail, "Urban congestion gating control based on reduced operational network fundamental diagrams," *Transportation Research Part C*, vol. 33, pp. 74–87, 2013.

[17] B. Coifman and S. Kim, "Extended bottlenecks, the fundamental relationship, and capacity drop on freeways," *Procedia Social and Behavioral Sciences*, vol. 17, pp. 44–57, 2011.

[18] D. Ni, *Traffic Flow Theory*, Elsevier, Amsterdam, Netherlands, 2016.

[19] A. Duret, C. Buisson, and N. Chiabaut, "Estimating individual speed-spacing relationship and assessing ability of newell's car-following model to reproduce trajectories," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2088, pp. 188–197, 2008.

[20] N. Chiabaut, L. Leclercq, and C. Buisson, "From heterogeneous drivers to macroscopic patterns in congestion," *Transportation Research Part B*, vol. 44, no. 2, pp. 299–308, 2010.

[21] D. Chen, S. Ahn, J. Laval, and Z. Zheng, "On the periodicity of traffic oscillations and capacity drop : the role of driver characteristics," *Transportation Research Part B*, vol. 59, pp. 117–136, 2014.

[22] M. Treiber and A. Kesting, "Microscopic calibration and validation of car-following models—a systematic approach," *Procedia Social and Behavioral Sciences*, vol. 80, pp. 922–939, 2014.

[23] J. Zheng, K. Suzuki, and M. Fujita, "Evaluation of car-following models using trajectory data from real traffic," *Procedia Social and Behavioral Sciences*, vol. 43, pp. 356–366, 2012.

[24] H. Liu, X. David Kan, S. E. Shladover, X. Y. Lu, and R. E. Ferlis, "Modeling impacts of cooperative adaptive cruise control on mixed traffic flow in multi-lane freeway facilities," *Transportation Research Part C*, vol. 95, pp. 261–279, 2018.

[25] A. Ghiasi, O. Hussain, Z. Sean 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*, vol. 106, pp. 266–292, 2017.

[26] 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*, vol. 100, pp. 196–221, 2017.

[27] S. Gong and L. Du, "Cooperative platoon control for a mixed traffic flow including human drive vehicles and connected and autonomous vehicles," *Transportation Research Part B*, vol. 116, pp. 25–61, 2018.

[28] 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, 2017.

[29] 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.

[30] Q. Deng, "A general simulation framework for modeling and analysis of heavy-duty vehicle platooning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 11, pp. 3252–3262, 2016.

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

[32] 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.

[33] S. G. Hu, H. Y. Wen, L. Xu, and H. Fu, "Stability of platoon of adaptive cruise control vehicles with time delay," *Transportation Letters*, vol. 7867, pp. 1–10, 2017.

[34] M. Seraj, J. Li, and T. Z. Qiu, "Modeling microscopic car-following strategy of mixed traffic to identify optimal platoon configurations for multiobjective decision-making," *Journal of Advanced Transportation*, vol. 201815 pages, Article ID 7835010, 2018.

[35] H. Liu, S. E. Shladover, X. Lu, and R. A. Ferlis, "Impact of cooperative adaptive cruise control (CACC) on multilane freeway merge capacity," in *Proceedings of the Transportation Research Board 97th Annual Meeting*, Washington, DC, USA, January 2018.

[36] Transportation Research Board, *HCM 2010 : Highway Capacity Manual*, Transportation Research Board, Washington, DC, USA, Fifth edition, 2010.

[37] E. O. F. Bear and H. Rd, *Alberta Highways 1 TO 986 Traffic Volume History 2006–2017*, Government of Alberta publications, Edmonton, Canada, 2018.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8892693, 16 pages
https://doi.org/10.1155/2020/8892693

*Research Article*

# A Discretionary Lane-Changing Decision-Making Mechanism Incorporating Drivers' Heterogeneity: A Signalling Game-Based Approach

**Haipeng Shao**[1], **Miaoran Zhang**[1], **Tao Feng**[2] **and Yifan Dong**[1]

[1]*College of Transportation Engineering, Chang'an University, Xi'an 710064, China*
[2]*Urban Planning Group, Department of the Built Environment, Eindhoven University of Technology, 5600 MB, Eindhoven, Netherlands*

Correspondence should be addressed to Miaoran Zhang; zhangmr@chd.edu.cn

Received 20 August 2020; Revised 11 September 2020; Accepted 17 September 2020; Published 7 October 2020

Academic Editor: Meng Meng



This paper attempts to propose a discretionary lane-changing decision-making model based on signalling game in the context of mixed traffic flow of autonomous and regular vehicles. The effects of the heterogeneity among different drivers and the endogeneity of same drivers in lane-changing behaviours, e.g., aggressive or conservative, are incorporated through the specification of different payoff functions under different scenarios. The model is calibrated and validated using the NGSIM dataset with a bilevel calibration framework, including two kinds of methods, genetic algorithm and perfect Bayesian equilibrium. Comparative results based on simulation show that the signalling game-based model outperforms the traditional space-based lane-changing model in the sense that the proposed model yields relatively stable reciprocal of time to collision and higher success rate of lane-changing under different traffic densities. Finally, a sensitivity analysis is performed to test the robustness of the proposed model, which indicates that the signalling game-based model is stable to the varying ratios of driver type.

## 1. Introduction

Lane-changing behaviour is vital in its effects on traffic flow. The behaviour that drivers execute lane-changing to obtain a better driving condition (higher speed, lager space, etc.) or get to the correct lane is a complicated decision-making process which is affected by many observable and unobservable factors of the surrounding environment and drivers. A bad lane-changing decision can lead to serious traffic problems [1], such as traffic breakdowns, bottleneck discharge rate reduction, stop-and-go oscillations, and safety hazards [2], while cooperative lane-changing may potentially improve traffic situation.

Compared to traditional cars, connected and autonomous vehicles enabled in vehicle-to-vehicle and vehicle-to-infrastructure communication technology are promising in improving traffic efficiency and safety [3]. However, one can foresee that within a certain period of time, regular vehicles and autonomous vehicles coexist on the road [4]. This means a certain level of cooperation is needed between human drivers and autonomous vehicles. Autonomous vehicles need to respond correctly and accurately to the environment and with respect to the specific drivers.

Existing studies on lane-changing of autonomous vehicles are mostly focused on the development of lane-changing trajectory, design of lane-changing controller, and prediction of lane-changing intention, while the lane-changing decision-making of autonomous vehicles has not been addressed sufficiently, especially in the interactions and heterogeneity between different drivers, which is an important issue in lane-changing process. To fill this research gap, this paper proposes a lane-changing decision-making mechanism for autonomous vehicles in the context of mixed traffic flow with conventional vehicles. Drivers' heterogeneous lane-changing behaviour is integrated based on a signalling game approach, whereas different payoff functions

are established according to the type of drivers, aggressive or conservative. The model is calibrated and validated using NGSIM data. Furthermore, sensitivity analysis is performed through microsimulations to examine the model robustness in comparison with the traditional space-based model.

The remainder of this paper is organized as follows: Section 2 reviews existing lane-changing approaches. Section 3 introduces the methodology that is proposed in this paper. Data processing and model calibration and validation are presented in Sections 4 and 5, respectively. Sections 6 and 7 present the details of simulation process and main findings. The paper is summarized in Section 8 with a discussion of future possible research aspects.

## 2. Literature Review

In the literature of lane-changing research, three types of models dominate the modelling of lane-changing decisions: rule-based models, utility-based models, and game theory-based models [2].

The pioneer rule-based model proposed by Gipps [5] contains a couple of factors influencing on lane-changing behaviour, such as collision avoiding, position of barrier, type of vehicles, special lane, and the possibility of advanced speed. When there are multiple lanes to change, a set of priority rules is used to decide the target lane. Although the involved affecting factors are reasonable and rather exhaustive, the interactions and communication between vehicles, e.g., considering other vehicles as moving obstacles, were ignored. Another limit is that this model fails to take heterogeneous drivers into account and assumes the characteristics of the same driver is constant over time (e.g., sometimes a conservative driver can do some aggressive actions for certain reasons). Based on Gipps's work, Yang and Koutsopoulos [6] made some improvement by splitting this lane-changing process into four steps: considering a lane-changing, determining the target lane, searching for the acceptable gap, and executing the change. However, no calibration and validation of this model had been done. Hidas [7] classified three types of lane-changing behaviour: free lane-changing, cooperative lane-changing, and forced lane-changing and proposed a new model which is capable to describe the interactions between vehicles and overcome one of flaws in Gipps's model.

Differing from rule-based models, utility-based models assume that the different levels of lane-changing behaviour depend on driver characteristics and status which vary across individual drivers [8]. Ahmed [9] expanded the model by incorporating mandatory lane-changing decisions. Furthermore, to address the possible dependency between discretionary lane-changing and mandatory lane-changing decisions, Toledo et al. [10] proposed an integrated model and introduced a random term to capture unobserved variables, which may explain the heterogeneity between different drivers and change in styles of the same driver over time.

The third approach in modelling lane-changing behaviour is game theory which provides a way to model the interactions between drivers. The first work focusing on the merging behaviour of regular vehicles was conducted by Kita [11]. After that, Meng et al. [12] used vehicles' velocity and acceleration to construct payoff functions without considering drivers' heterogeneity. Yu et al. [13] represented a Stackelberg game model to describe the lane-changing decision-making process of autonomous vehicles. The payoff function is defined as a linear combination of vehicle spacing and safety factor. In addition, the coefficients were considered as a function of driver's aggressiveness. The model has great advantages in comparison with the traditional rule-based model as it involves aggressiveness revealed by vehicle's lateral movement. Nevertheless, other psychological factors such as regret and conformity were considered. Ali et al. [2] proposed a game model under mandatory lane-changing scenarios in which the strategy set was screened by empirical data. Based on the NGSIM US101 data [14] and driving simulator data, the model was validated through predefined microscopic indicators, such as confusion matrix, time, and location error.

In addition to the models presented above, researchers have also applied artificial intelligence to model the lane-changing decision process, such as BP neural network [15] and fuzzy inference system [16]. It is no doubt that these models can generally achieve a better prediction, and the model performance largely depends on the quality and quantity of the mixed traffic flow data which are still difficult to obtain. Moreover, artificial intelligent models are often inexplainable models in the sense it could not help much to understand the mechanism of lane-changing behaviour.

Relative to the situation of human drivers only, future traffic condition will be mixed with autonomous vehicles, indicating a variety of interactions between human drivers and autonomous vehicles. Several issues that differ from the traditional traffic environment may arise. First, autonomous vehicles receive massive information about surroundings (i.e., real-time traffic information and vehicle-to-vehicle communication) through various sensors or cameras, which indicates that an optimal lane-changing decision can be made relatively easier than human drivers if the information can be correctly and sufficiently recognized and used. This means when modelling the lane-changing behaviour of autonomous vehicles, additional assumptions may be needed. Moreover, the lane-changing decision of autonomous vehicles plays a vital part in the coordination of traffic system. Human drivers use their experience to decide whether to change lanes, while autonomous vehicles need to learn to interact with human drivers instead of following fixed rules or being constantly courteous. Autonomous vehicles need a higher degree of accuracy and effectiveness in responses than human drivers to cope with different styles of driving and lane-changing behaviour.

Recent studies have attempted to differentiated drivers from the perspective of discretionary lane-changing behaviour and specified the payoff functions based on safety and efficiency. Depending on a group of interviewers' subjective opinion or using clustering algorithms, different driving styles such as aggressive or not aggressive can be identified [17–19]. The effects of the heterogeneity among different drivers in their lane-changing behaviour on other

vehicles are critical, especially in a connected environment and with autonomous vehicles. However, this heterogeneity may be insufficient to reflect the variation of decisions of a same driver. Research has shown that drivers typically adjust their driving styles in different traffic situations [1] such that the same driver may have different lane-changing styles in different temporal and spatial situations [20,21].

This signifies the importance and necessity that, when building a reasonable lane-changing decision-making mechanism for autonomous vehicles in a mixed traffic flow, one should systematically incorporate the interactions and heterogeneity between different drivers. First, discretionary lane-changing is not a one-way decision-making process, which needs interactions with others. In a lane-changing scenario, decisions of two participants are affected by each other. However, such an interaction has been ignored by all existing approaches except the game theory. Second, in addition to the heterogeneity between different drivers, the changes in the characteristics of the same driver have not been paid attention. Decisions should be made by considering the real-time reaction of drivers rather than a fixed driving style of classification since drivers can act in an opposite way to normal in certain situations.

Therefore, in this paper, we attempt to incorporate the heterogeneous behaviour among different drivers and of the same driver into the lane-changing decision-making process. To the best of our knowledge, this is the first attempt to take into account the effects of both the heterogeneity between different drivers and the endogeneity of same drivers in lane-changing decisions. In order to mimic the dynamic feature of such decisions, a signalling game approach is proposed. In the following sections, we will present the methods and our simulation results.

## 3. Materials and Methods

Here, we regard a lane-changing decision-making process as a game considering that game theories are potentially useful to the modelling of lane-changing behaviour. In the following sections, we first present the concept of game theory and the signalling game that is adopted in this study. Then we present the specification of games in lane-changing behaviour in the setting of mixed traffic flow of autonomous and human driver cars.

*3.1. Game Theory and Signalling Games.* Game theory is a powerful and well-developed mathematic tool. It is usually used to model the interaction process between two or more players in economic fields. With a number of disciplines, game theory has been widely used in logical decision-making in humans. It has six main concepts: game, player, strategy, payoff, information set, and equilibrium. Game is defined as any circumstances that are affected by players' decision, and a player is seen as a decision-maker in this game. Strategy is a decision that a player may take in that circumstance. Payoff is used to measure the cost due to a pair of decisions. Information set consists all available information. Equilibrium is the optimal state of a game, yielding every player's decision and their payoffs.

Solutions under equilibrium in game theory can be classified in four types by the nature of timing and information. In the case of simultaneous move games with complete information, the appropriate solution concept is Nash equilibrium (NE), while for games of sequential timing with complete information, the best solution is subgame perfect equilibrium (SPE). When there is incomplete information, Bayesian Nash equilibrium (BNE) and perfect Bayesian equilibrium (PBE) are the solutions to solve simultaneous move and sequential timing games, respectively. It is a set of strategies and beliefs such that the strategies are sequentially rational given the players' beliefs and players update beliefs via Bayes rule wherever possible [22].

Among the various games, signalling game is a simple type of dynamic Bayesian game in game theory [23]. It includes two players, called sender and receiver, and a party called Nature who randomly decides the type of a sender for a receiver. The receiver does not know the type of senders for sure but knows about the probability of the sender type. For example, assuming there are two potential types, strong and weak. Even if the Nature decides that the sender is a weak type, the receiver only infers the type of the sender by observing its actions, e.g., the receiver has information that the probability of being weak is 0.4 and 0.6 of being strong.

Since lane-changing behaviour is sequential (one takes actions according to the other's actions) and the information may be incomplete (the type of vehicles on the target lane is unknown to the vehicle who wants to change lanes), we apply signalling games based on PBE in a lane-changing process.

*3.2. Signalling Game of Lane-Changing.* In the process of lane-changing decision with mixed traffic flow, autonomous (CAR-E) and regular vehicles (CAR-TF) are the two players. When a CAR-E has a demand of changing lanes, it indicates its intention to the other vehicle, CAR-TF. CAR-TF will respond to the intention by sending a signal as soon as the intention is received. Note that sending a signal could be generally regarded as a warning behaviour of drivers, such as a significant acceleration or horning. However, an aggressive driver may have a higher probability of sending a signal than other types of drivers. Observing the reaction of CAR-TF, CAR-E can then analyse the behaviour of the CAR-TF and calculate the payoff values in order to make an optimal decision.

Due to the fact that the type of drivers in CAR-TF, e.g., aggressive or conservative, which largely affects lane-changing decision of CAR-E is unknown to CAR-E, we assume the probability that CAR-E knows CAR-TF as an aggressive driver is $p$. CAR-TF knows that this is the belief of CAR-E, and CAR-E knows that CAR-TF knows that this is the belief of CAR-E, which means $p$ is the common knowledge between both players [24].

Here, we defined that the set of CAR-TF types is $\Theta = \{\theta_1, \theta_2\}$, where $\theta_1$ and $\theta_2$ represent the aggressive type and conservative type, respectively. The prior probability of the one type is $p$, then the prior probability of being the other type is $1 - p\,(0 \le p \le 1)$ Let the signal set of CAR-TF be

$A_1 = \{S, NS\}$, where S and NS denote signalling and not signalling, respectively. The action set of CAR-E is defined as $A_2 = \{C, W\}$, where $C$ denotes lane-changing behaviour and $W$ represents waiting decision. For each signal that CAR-E receives and the possible action that each CAR-TF takes, there is a payoff for CAR-TF ($P_{ij}$) and CAR-E ($Q_{ij}$). All sets of possible strategies of this game are presented in Table 1.

In the game sequence, "Nature" selects the type of CAR-TF, and CAR-E knows the prior probability of type $\theta_i$, $p(\theta_i)$. Note that "Nature" is virtual and not participating in the game. When the CAR-TF receives the intention of CAR-E, it chooses an action $a_1 \in A_1$ (usually is related to the driver's types). Then, CAR-E observes the action of CAR-TF, $a_1$, then infers the posterior probability $\tilde{p}(\theta \mid a_1)$ using Bayes' rules. Then CAR-E takes an action, $a_2 \in A_2$. Considering the multiple possibilities at each stage of the sequence, a game tree can be built to describe the decision mechanism, as shown in Figure 1.

In Figure 1, Ct is the cost of CAR-TF for sending a signal, such as decreased attention and acceleration fluctuation. $F_{\mathrm{TF}}^A$ and $F_E^A$ are losses of CAR-TF and CAR-E, respectively, in the situation that CAR-E changes lanes when the driver of CAR-TF is aggressive. $G_{\mathrm{TF}}^C$ and $G_E^C$ are losses of CAR-TF and the benefit of CAR-E, respectively, in the situation that CAR-E changes lane when the driver of CAR-TF is conservative. $R_{\mathrm{TF}}^A$ and $R_E^A$ are regrets of aggressive and conservative CAR-TFs in the situation that CAR-E changes lanes when the driver of CAR-TF did not send a signal.

Figure 2 shows an example of the lane-changing process where CAR-E is assumed to change lanes. $E$ represents CAR-E, and TF represents CAR-TF. TL and L1 are the leading vehicles in the target lane and current lane, respectively, called CAR-TL and CAR-L1.

Considering that speed and driving space are the two main factors influencing lane-changing behaviour, we include the space and speed into the payoff function. In addition, an indicator of regret is also taken into account to mimic individuals' decision-making.

Let $t_0$ be the starting time of lane-changing and $\Delta t$ be the time to finish this process. Then, the distance between vehicles can be calculated as below:

$$\begin{aligned} D_1 &= y_{L1}^0 - y_E^0 - L, \\ D_1' &= y_{TL}^0 - y_E^0 - L, \\ D_2 &= y_E^0 - y_{TF}^0 - L, \end{aligned} \tag{1}$$

where, $D_1, D_1'$, and $D_2$ denote the distance between CAR-E and CAR-L1, CAR-E and CAR-TL, and CAR-E and CAR-TF at $t_0$, respectively; $L$ denotes the length of vehicles. $y_{L1}^0$, $y_E^0$, $y_{\mathrm{TL}}^0$, and $y_{\mathrm{TF}}^0$ denote the longitudinal position of the centre of CAR-L1, CAR-E, CAR-TL, and CAR-TF at $t_0$ respectively. Thus, the distance between CAR-E and CAR-TF at $t_0 + \Delta t$, $D_2'$, is

$$D_2' = \left(v_{\mathrm{TF}}^0 - v_E^0\right) \times \Delta t + \left(\frac{\Delta t^2}{2}\right) \times \left(a_{\mathrm{TF}}^0 - a_E^0\right) - L, \tag{2}$$

where $v_{\mathrm{TF}}^0$ and $v_E^0$ are the instantaneous speed of CAR-TF and CAR-E at $t_0$ and $a_{\mathrm{TF}}^0$ and $a_E^0$ are the acceleration rates of CAR-TF and CAR-E at $t_0$.

Here, the regret of aggressive and conservative CAR-TFs are defined as functions of the speed difference between CAR-TF and CAR-E as below:

$$\begin{aligned} R_{\mathrm{TF}}^A &= \begin{cases} \gamma_1 \times \left(v_E' - v_{\mathrm{TF}}'\right), & v_E' - v_{\mathrm{TF}}' > 0, \\ 0, & v_E' - v_{\mathrm{TF}}' \le 0, \end{cases} \\ R_{\mathrm{TF}}^C &= \begin{cases} \gamma_2 \times \left(v_E' - v_{\mathrm{TF}}'\right), & v_E' - v_{\mathrm{TF}}' > 0, \\ 0, & v_E' - v_{\mathrm{TF}}' \le 0, \end{cases} \end{aligned} \tag{3}$$

where $v_{\mathrm{TF}}'$ and $v_E'$ are the speed of CAR-TF and CAR-E at $t_0 + \Delta t$, respectively. $\gamma_1$ and $\gamma_2$ are the parameters to be estimated.

The cost of CAR-TF to send a signal is defined based on the difference in acceleration at different time, which indicates distraction caused by sending a signal, given by the following equation:

$$\mathrm{Ct} = \varepsilon \times \left(a_{\mathrm{TF}}' - a_{\mathrm{TF}}^0\right), \tag{4}$$

where $a_{\mathrm{TF}}'$ is the acceleration of CAR-TF at $t_0 + \Delta t$. $\varepsilon$ is a parameter to be estimated.

For aggressive CAR-TFs, the loss is dependent on the difference in distance between the two vehicles at time $t_0$ and $t_0 + \Delta t$, which can be calculated by using the following equation:

$$F_{\mathrm{TF}}^A = f_{\mathrm{TF}}^1 \times \left(D_2 - D_2'\right), \tag{5}$$

where $f_{\mathrm{TF}}^1$ is the parameter to be estimated.

Let $v_{\mathrm{TF}}'$ and $v_E'$ be the speed of CAR-TL and CAR-E at $t_0 + \Delta t$, respectively. $G_{\mathrm{TF}}^C$ is quantified by loss of space and speed; thus, the loss of conservative CAR-TFs can be calculated by using the following equation:

$$G_{\mathrm{TF}}^C = g_{\mathrm{TF}}^1 \times \left(D_1' + D_2 + L - D_2'\right) + g_{\mathrm{TF}}^2 \times \left(v_{\mathrm{TL}}{}' - v_E'\right), \tag{6}$$

where $g_{\mathrm{TF}}^1$ and $g_{\mathrm{TF}}^2$ are the parameters to be estimated.

Similarly, the payoff functions of aggressive and conservative CAR-Es can be calculated as below:

$$\begin{aligned} F_E^A &= f_E^1 \times (1 - \eta_1) \times \left(D_2 - D_2'\right), \\ G_E^C &= g_E^1 \times (1 - \eta_1) \times \left(D_1' - D_1\right) + g_E^2 \times \eta_1 \times \left(v_{\mathrm{TL}}^0 - v_E^0\right), \end{aligned} \tag{7}$$

where $\eta_1$ denotes the weight coefficient of speed benefit in the payoff of CAR-E, which can be set in the controller manually. The higher $\eta_1$ is, the more the attention paid on speed is. Lower $\eta_1$ can be set in some risky road segments to create a more cautious driving style for autonomous vehicles. $g_E^1$, $g_E^2$, and $f_E^1$ are the parameters to be estimated.

*3.3. Equilibrium Setting and Model Solution.* According to the previous studies on Game Theory, signalling game can be solved by the following method [23]. Let $\theta (\theta \in \Theta)$ be the type of CAR-TF, which is selected by the Nature. $a_1 (a_1 \in A_1)$ and $a_2 (a_2 \in A_2)$ denote the actions selected by CAR-TF and CAR-E. The payoffs of the CAR-TF and CAR-E are represented as $u_1(a_1, a_2, \theta)$ and $u_2(a_1, a_2, \theta)$. $\sigma_1(\cdot \mid \theta)$ is a probability distribution to each type of CAR-TF over action $a_1$, and $\sigma_2(\cdot \mid a_1)$ is a probability distribution to each

Table 1: Possible strategies in the game.

| TF ($P$) | E ($Q$) | | | |
|---|---|---|---|---|
| | Aggressive TF | | Conservative TF | |
| | Change lane (C) | Wait (W) | Change lane (C) | Wait (W) |
| Signalling (S) | $P_{11}$, $Q_{11}$ | $P_{12}$, $Q_{12}$ | $P_{13}$, $Q_{13}$ | $P_{14}$, $Q_{14}$ |
| Not signalling (NS) | $P_{21}$, $Q_{21}$ | $P_{22}$, $Q_{22}$ | $P_{23}$, $Q_{23}$ | $P_{24}$, $Q_{24}$ |

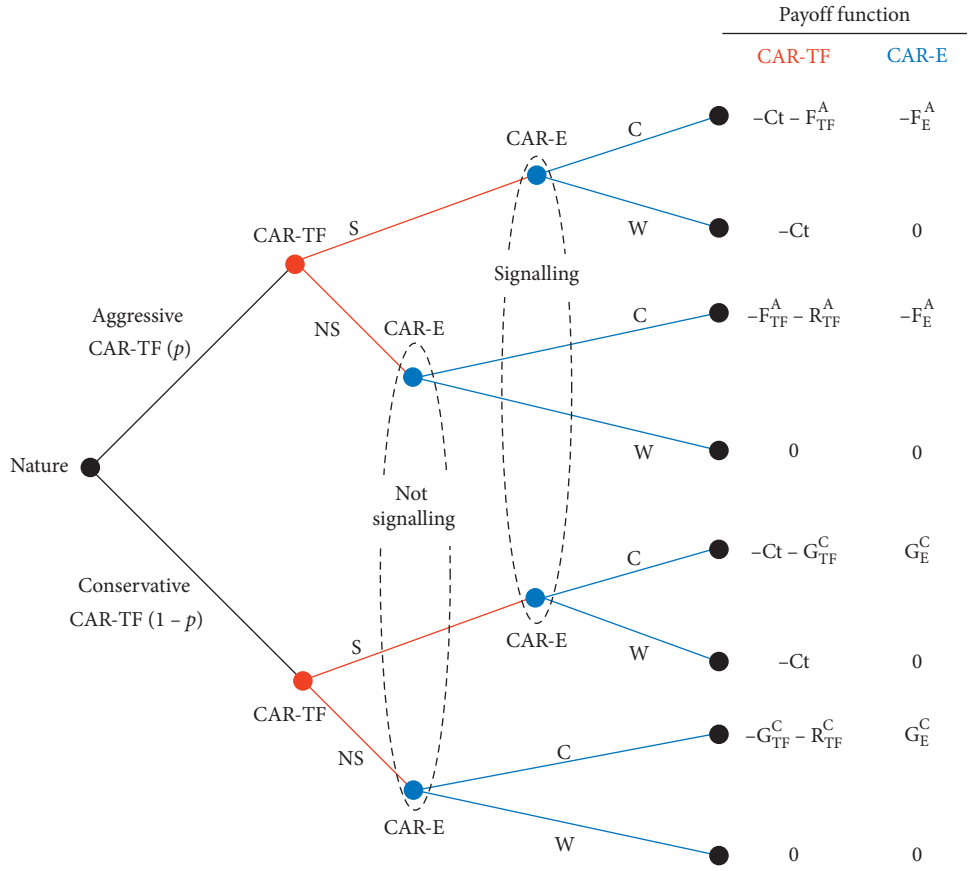



Figure 1: Extended signalling game of lane-changing behaviour.

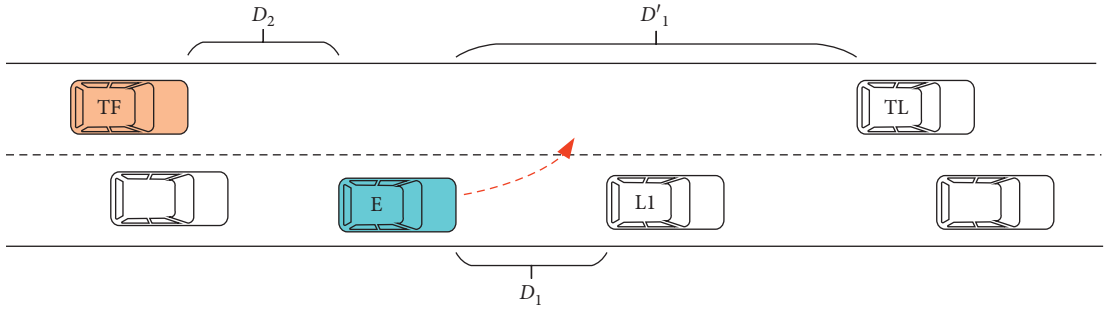



Figure 2: Lane-changing representation in a two-lane scenario.

action $a_1$ over action $a_2$. In addition, when CAR-E updates the probability of CAR-TF's type by Bayesian rules, the strategies of CAR-TF and CAR-E become $\sigma_1^*(\cdot \mid \theta)$ and $\sigma_2(\cdot \mid a_1)$, respectively. Consequently, a perfect Bayesian equilibrium (PBE) for a signalling game is a strategy profile $\sigma^*$ and a posterior belief $\mu(\cdot \mid a_1)$, which satisfy the following equation:

$$
\begin{aligned}
&(P_1)\forall\theta, \sigma_1^*(\cdot|\theta) \in \arg\max_{a_1} u_1(a_1, \sigma_2^*, \theta),\\
&(P_2)\forall a_1, \sigma_2^*(\cdot|a_1) \in \arg\max_{a_2} \sum_{\theta} \mu(\theta|a_1) u_2(a_1, \sigma_2^*, \theta),\\
&(B)\mu(\theta|a_1) = \frac{p(\theta)\sigma_1^*(a_1|\theta)}{\sum_{\theta'\in\Theta} p(\theta')\sigma_1^*(a_1 \mid \theta')},\\
&\text{if } \sum_{\theta'\in\Theta} p(\theta')\sigma_1^*(a_1 \mid \theta') > 0,\\
&\mu(\cdot|a_1) \text{ is any probability distribution over } \Theta.
\end{aligned}
\tag{8}
$$

Note that condition $(P_1)$ means that CAR-TF takes into account the influence of $a_1$ on CAR-E and condition $(P_2)$ means that CAR-E makes the best response to the action of CAR-TF by knowing the posterior beliefs. Moreover, condition $(B)$ is the application of the Bayesian rule.

According to different actions of players, the PBE of this game can be divided into three subcategories, namely, separating equilibrium, pooling equilibrium, and semi-separating equilibrium. As discussed below, these equilibriums are used to derive the full set of game solutions.

A separating equilibrium represents that different types of CAR-TFs select different actions with the probability of 1, i.e., actions reveal types definitively. For example, aggressive drivers always select to send a signal, while conservative drivers always choose to be silent (i.e., doing nothing). In this situation, once CAR-TFs take an action, CAR-E can distinguish its type correctly and make the optimal action.

There are two different possible situations: (1) all aggressive drivers send a signal, and all conservative drivers do nothing; (2) all aggressive drivers do nothing, and all conservative drivers send a signal.

In the first case, we assume all aggressive drivers send a signal, and all conservative ones do nothing. Firstly, let us consider payoffs from CAR-E's perspective. If CAR-E observes a signal, it believes the CAR-TF is an aggressive one based on the assumption. According to the payoff matrix in the game tree, it costs $F_E^A$ (i.e., the payoff is $-F_E^A$) to change lane, while that would be 0 in case of waiting decision. Apparently, 0 is larger than $-F_E^A$; therefore, it tends to wait instead of changing a lane. On the contrary, if CAR-E does not observe any actions (i.e., CAR-TF does nothing), it will reckon that the driver is more conservative. If CAR-E changes lane, the payoff will be $G_E^C$, which is higher than the payoff of waiting (0). Thus, CAR-E is more likely to change lanes. Consequently, CAR-E's strategy can be concluded as below:

$$
\begin{cases}
C, & \text{if CAR} - \text{TF NS},\\
W, & \text{if CAR} - \text{TF S}.
\end{cases}
\tag{9}
$$

To reach the equilibrium, payoffs of both sides are necessary to be considered. Therefore, further analysis is conducted in terms of CAR-TF's interest. Several situations can be concluded in Table 2 according to corresponding branches of the game tree.

In order to meet the assumption made before, a set of inequality can be derived:

$$
\begin{cases}
\text{payoff}(S \mid \text{aggressive}) > \text{payoff}(\text{NS} \mid \text{aggressive}),\\
\text{payoff}(S \mid \text{conservative}) < \text{payoff}(\text{NS} \mid \text{conservative}).
\end{cases}
\tag{10}
$$

Therefore, the condition of this equilibrium is

$$
G_{\text{TF}}^C + R_{\text{TF}}^C < \text{Ct} < F_{\text{TF}}^A + R_{\text{TF}}^A. \tag{11}
$$

The same procedure can be performed in the second case.

A pooling equilibrium denotes that different types of CAR-TFs choose the same action, which makes CAR-Es unable to update the prior probability. In other words, no further information is included in the sender's choice. The only thing CAR-E can do is to predict the type according to $p$ (the original probability).

There are two situations: all drivers go for signalling or do nothing. Under the first circumstance, CAR-E's payoffs can be derived as below:

$$
\begin{aligned}
\text{payoff}_C &= -pF_E^A + (1-p)G_E^C,\\
\text{payoff}_W &= -p \times 0 + (1-p) \times 0 = 0.
\end{aligned}
\tag{12}
$$

Here, if $\text{payoff}_C > \text{payoff}_W$, then CAR-E will choose to change a lane. On the contrary, if $\text{payoff}_C < \text{payoff}_W$, then CAR-E will choose to wait for another chance.

In case that the first condition is matched, the payoffs of CAR-TF can be summarized in Table 3 based on its interest.

According to the discussion above, a set of inequality conditions can be derived:

$$
\begin{cases}
\text{payoff}_C > \text{payoff}_W,\\
\text{payoff}(S \mid \text{aggressive}) > \text{payoff}(\text{NS} \mid \text{aggressive}),\\
\text{payoff}(S \mid \text{conservative}) > \text{payoff}(\text{NS} \mid \text{conservative}),
\end{cases}
\tag{13}
$$

here comes the first strategy.

When $\begin{cases} -pF_E^A + (1-p)G_E^C > 0,\\ R_{\text{TF}}^A > \text{Ct},\\ R_{\text{TF}}^C > \text{Ct}, \end{cases}$ all CAR-TFs will send a signal and CAR-E will change lanes. The remaining branches can be derived by the same way.

Semiseparating equilibrium is the most complicated situation. This means that some types of CAR-TFs randomly choose actions, and other types of CAR-TFs choose specific actions. A conservative driver is more likely to yield when a vehicle in the other lane shows lane-changing intention. However, he/she may also pretend to be an aggressive driver in order to protect his/her own interest. Consequently, in this study, semiseparation equilibrium can be divided into two categories. One is that a part of aggressive CAR-TFs choose to send a signal and the rest not, while the conservative ones do not take any actions. On the contrary, the other case is that a part of conservative CAR-TFs chooses to send a signal and the rest not, while all aggressive drivers choose to send a signal.

Table 2: Payoffs for CAR-TF in the first case of separating equilibrium.

| Type | Strategy (CAR-TF, CAR-E) | Payoff (CAR-TF) |
|---|---|---|
| Aggressive | S, W | $-\mathrm{Ct}$ |
| | NS, C | $-F_{\mathrm{TF}}^{A} - R_{\mathrm{TF}}^{A}$ |
| Conservative | S, W | $-\mathrm{Ct}$ |
| | NS, C | $-G_{\mathrm{TF}}^{C} - T_{\mathrm{TF}}^{C}$ |

Table 3: Payoffs for CAR-TF in the first condition of pooling equilibrium.

| Type | Strategy (CAR-TF, CAR-E) | Payoff (CAR-TF) |
|---|---|---|
| Aggressive | S, C | $-\mathrm{Ct} - F_{\mathrm{TF}}^{A}$ |
| | NS, C | $-F_{\mathrm{TF}}^{A} - R_{\mathrm{TF}}^{A}$ |
| Conservative | S, C | $-\mathrm{Ct} - G_{\mathrm{TF}}^{C}$ |
| | NS, C | $-G_{\mathrm{TF}}^{C} - R_{\mathrm{TF}}^{C}$ |

In the first case, when CAR-E receives a signal, it will determine CAR-TF is an aggressive type and choose to wait in order to get the optimal payoff. The other scenario is more complicated, when there is no signal, CAR-E will have to determine the best strategy according to the probability. Let $x$ $(x < 1)$ be the probability of the signalling action of aggressive drivers, i.e., $P(\mathrm{S} \mid \text{aggressive}) = x$. Let $y$ $(y < 1)$ be the probability of the lane-changing action of autonomous vehicles when observing a signal, i.e., $P(\mathrm{C} \mid \mathrm{S}) = y$. Then, the probability of CAR-E observing nothing is

$$\begin{aligned} P(\mathrm{NS}) &= p \times P(\mathrm{NS} \mid \text{aggressive}) + (1-p) \\ &\quad \times P(\mathrm{NS} \mid \text{conservative}) \\ &= p(1-x) + (1-p) \\ &= 1 - px. \end{aligned} \tag{14}$$

Then,

$$\begin{aligned} P(\text{aggressive} \mid \mathrm{NS}) &= \frac{p(1-x)}{1-px}, \\ P(\text{conservative} \mid \mathrm{NS}) &= \frac{1-p}{1-px}. \end{aligned} \tag{15}$$

Therefore, the payoffs of CAR-E under different strategies can be concluded below:

$$\begin{aligned} \text{payoff}_C &= P(\text{aggressive} \mid \mathrm{NS}) \times \left(-F_E^A\right) \\ &\quad + P(\text{conservative} \mid \mathrm{NS}) \times G_E^C \\ &= \frac{-p(1-x) \times F_E^A + (1-p) \times G_E^C}{1-px}, \\ \text{payoff}_W &= P(\text{aggressive} \mid \mathrm{NS}) \times 0 \\ &\quad + P(\text{conservative} \mid \mathrm{NS}) \times 0 = 0. \end{aligned} \tag{16}$$

$x$ can be determined by equation $\text{payoff}_C = \text{payoff}_W$; thus,

$$x = 1 - (1-p) \times \frac{G_E^C}{pF_E^A} < 1. \tag{17}$$

According to CAR-TF's strategies, its payoffs can be summarized in Table 4.

Accordingly, $y$ can be determined by aggressive drivers' payoff equation, $\text{payoff}_{\mathrm{S}} = \text{payoff}_{\mathrm{NS}}$; thus,

$$y = \frac{\mathrm{Ct}}{\left(R_{\mathrm{TF}}^{A} + F_{\mathrm{TF}}^{A}\right)} < 1. \tag{18}$$

The assumption is that all the conservative drivers do not take any actions; therefore, from their perspective, $\text{payoff}_{\mathrm{S}} < \text{payoff}_{\mathrm{NS}}$, i.e.,

$$-\mathrm{Ct} < -y\left(G_{\mathrm{TF}}^{C} + R_{\mathrm{TF}}^{C}\right). \tag{19}$$

In conclusion, when $\begin{cases} (1-p) \times G_E^C < pF_E^A, \\ \mathrm{Ct} < R_{\mathrm{TF}}^{A} + F_{\mathrm{TF}}^{A} < G_{\mathrm{TF}}^{C} + R_{\mathrm{TF}}^{C}, \end{cases}$ aggressive CAR-TFs choose to send a signal or not, while the conservative ones do nothing.

According to the analysis presented above, the whole model solution can be summarized in Table 5.

## 4. Data

The NGSIM US-101 dataset is used to calibrate the proposed model. This dataset, which was collected on Hollywood Freeway, Los Angeles, includes the detailed vehicle trajectories at every 0.1 second. Figure 3 shows the study area of the NGSIM US-101 dataset.

In order to implement the proposed model, the data were first cleaned according to the following principles:

(1) Since the model focuses on car's lane-changing behaviour, the records of other types of vehicles (i.e., motorcycles and heavy vehicles) were removed.

(2) Hollywood Freeway is a five-lane highway with an auxiliary lane and two ramps (i.e., Lane 6, Lane 7, and Lane 8). Therefore, records related to these three lanes, which could be seen as mandatory lane-changing behaviour, were removed.

(3) Considering vehicles' lane-changing behaviour is quite different when the traffic flow is overly congested or in a free-flow situation, records with spacing headway longer than 30 meters or shorter than 6 meters were filtered out [26].

In addition, a driver's decision time window was empirically considered as 2 seconds [2]; thus, we extract the dataset based on a 2-second interval.

Moreover, we split the process of lane-changing into three phases, namely, lane-keeping (LK) decision horizon, lane-changing (LC) decision horizon, and LC duration. LC point is determined when the centre of the vehicle passes the lane boundary. Details can be seen in Figure 4.

In order to simulate lane-changing decision of autonomous vehicles, the records of lane-changing and waiting were selected randomly from the data. Additionally, there are more lane-keeping cases than lane-changing cases in the dataset. Too many lane-keeping cases will have an enormous effect on model calibration and cause a low predictive rate in

Table 4: Payoffs for CAR-TF in the first case of semiseparating equilibrium.

| Type | Strategy (CAR-TF, CAR-E) | Payoff (CAR-TF) |
|---|---|---|
| Aggressive | S, W | −Ct |
| | NS, $P(\mathrm{C}) = y$ | $-y(F_{\mathrm{TF}}^{A} + R_{\mathrm{TF}}^{A})$ |
| Conservative | S, $W$ | −Ct |
| | NS, $P(\mathrm{C}) = y$ | $-y(G_{\mathrm{TF}}^{C} + R_{\mathrm{TF}}^{C})$ |

Table 5: Solutions in three equilibriums.

| Type | Strategy: The sender (CAR-TF) Aggressive | Strategy: The sender (CAR-TF) Conservative | Strategy: The receiver (CAR-E) | Condition |
|---|---|---|---|---|
| Separating equilibrium | S | NS | If S, then W; if NS, then C | $G_{\mathrm{TF}}^{C} + R_{\mathrm{TF}}^{C} < \mathrm{Ct} < F_{\mathrm{TF}}^{A} + R_{\mathrm{TF}}^{A}$ |
| | NS | S | If S, then C; if NS, then W | Not available |
| Pooling equilibrium | NS | NS | If NS, then C/W | C, when $pF_{E}^{A} < (1-p)G_{E}^{C}$, $R_{\mathrm{TF}}^{A} < \mathrm{Ct}$, and $R_{\mathrm{TF}}^{C} < \mathrm{Ct}$<br>W, when $pF_{E}^{A} > (1-p)G_{E}^{C}$ |
| | S | S | If S, then C | $pF_{E}^{A} < (1-p)G_{E}^{C}$, $R_{\mathrm{TF}}^{A} > \mathrm{Ct}$, and $R_{\mathrm{TF}}^{C} > \mathrm{Ct}$ |
| Semiseparating equilibrium | S | $P(\mathrm{S}) = x$ | Update information using $p$ and $x$.<br>If S, then $P(\mathrm{C} \mid \mathrm{NS}) = y$; if NS, then C | $pF_{E}^{A} < (1-p)G_{E}^{C}$, $yF_{\mathrm{TF}}^{A} < F_{\mathrm{TF}}^{A} + R_{\mathrm{TF}}^{A}$, and $R_{\mathrm{TF}}^{C} > \mathrm{Ct}$,<br>where $x = pF_{E}^{A}/(1-p)G_{E}^{C}$ and $y = (G_{\mathrm{TF}}^{C} + R_{\mathrm{TF}}^{C} - \mathrm{Ct})/G_{\mathrm{TF}}^{C}$. |
| | $P(\mathrm{S}) = x$ | NS | Update information using $p$ and $x$.<br>If NS, then $P(\mathrm{C} \mid \mathrm{NS}) = y$<br>Update information; if S, then W | $pF_{E}^{A} > (1-p)G_{E}^{C}$, $y(G_{\mathrm{TF}}^{C} + R_{\mathrm{TF}}^{C}) > \mathrm{Ct}$, and $R_{\mathrm{TF}}^{A} + F_{\mathrm{TF}}^{A} > \mathrm{Ct}$,<br>where $x = 1 - ((1-p)G_{E}^{C}/pF_{E}^{A})$ and $y = \mathrm{Ct}/(R_{\mathrm{TF}}^{A} + F_{\mathrm{TF}}^{A})$. |

S = signalling; NS = not signalling; C = changing lane; W = waiting.

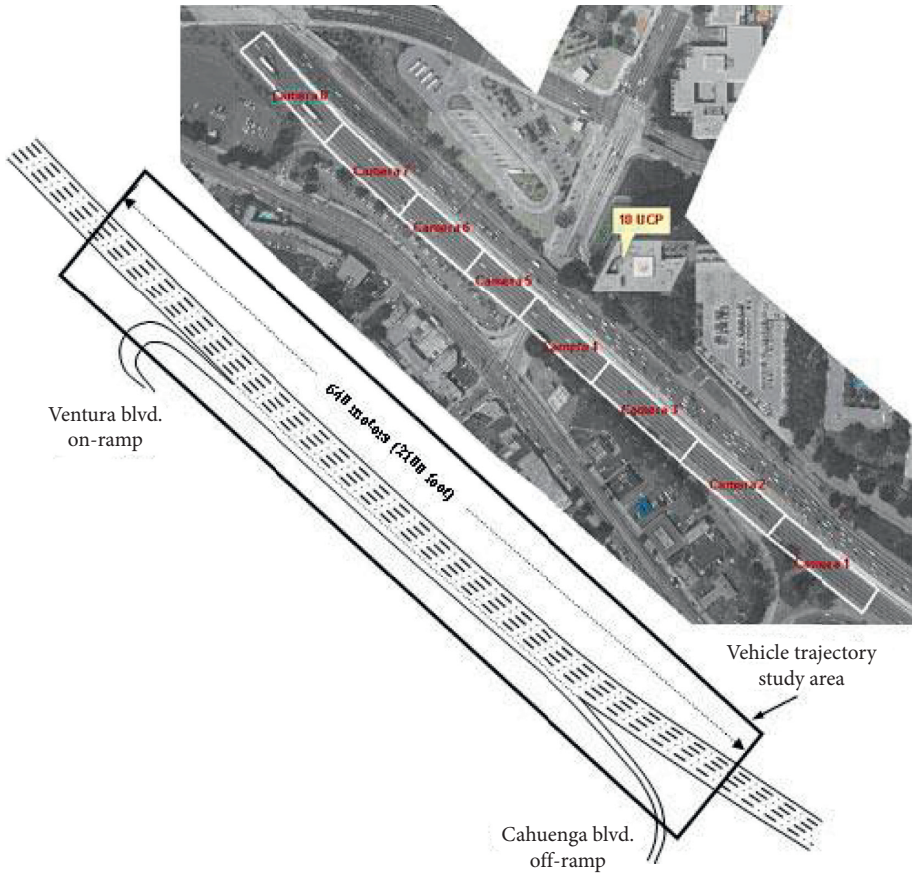



Figure 3: The study area of the NGSIM US-101 dataset [25].

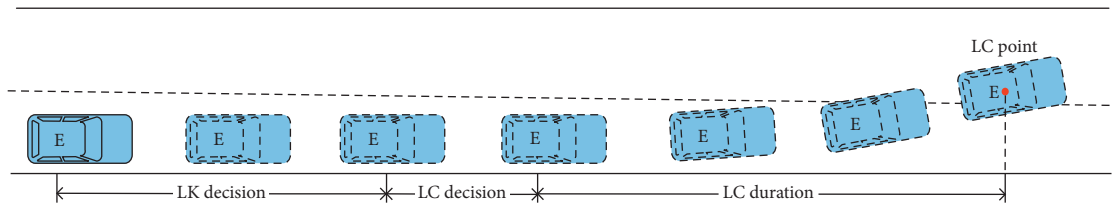



Figure 4: Lane-changing process.

lane-changing behaviour. Therefore, it is important to determine the ratio between the records of lane-changing and lane-keeping. The longer the LK decision horizon is, the more the LK cases are. A sensitivity analysis therefore is performed to determine the LK decision horizon. As a result, this study takes 2 seconds as the LK decision horizon, i.e., the ratio between the two cases are equal.

The strategies of CAR-E (i.e., waiting or changing) can be determined by the existence of lane-changing point. However, it is a challenge to extract the strategies of CAR-TF from vehicle trajectories, i.e., signalling or not. Therefore, the steady-state regime is adopted to identify CAR-TF's action based on vehicles' acceleration status in the sense that a steady-state regime is achieved when the acceleration or deceleration rate is lower than 0.05 g ("$g$" is the gravitational acceleration) [27]. When the acceleration or deceleration is smaller than 0.05 g, CAR-TF's strategy can be seen as "not signalling"; otherwise it is treated as "signalling."

Due to the limitation of trajectory data, it is impossible to find out the exact time and position when a LC decision is made. However, it can be derived from LC duration since the LC point is easy to be determined. LC duration is normally defined as the time taken by completing LC manoeuvre. Many researches have been done to figure out the LC duration (Table 6). In this study, 4s is adopted, which is used to find out the endpoint of LC decision. Then, the start point for LC decision can be determined according to the 2-second decision time window.

Finally, a total of 741 observations were selected, including 391 lane-changing records and 350 lane-keeping records. Table 7 shows the distribution of extracted strategies.

## 5. Model Calibration

The calibration of the proposed model is implemented in the way that we estimate the parameters in order to minimize the difference between observed decision in the NGSIM dataset and the decision predicted by the proposed model. This paper adopts the calibration framework established by Liu et al. [35], which was also applied by Ali et al. [2] and Kang and Rakha [36].

As shown in Figure 5, the whole problem is divided into two levels: upper and lower level. The upper level is used to narrow the difference between predicted decisions and observed decision, while the lower level aims to find a full set of equilibrium for this game. Before programming, decisions need to be extracted from the original observed data. Let $k$ be the iteration index. At the first stage, a set of parameters are initialized. By applying these parameters into payoff functions defined before, strategies can be found for each vehicle using the PBE method. Then, we calculate the difference between predicted strategies and the actual decisions in upper level. These steps are iterated and do not stop until the convergence is reached.

The function of the upper problem of this programming is defined as the difference between the observed data and predicted data as below:

$$\min \sum_{i=1}^{n} \left(y_i^* - y_i\right)^2, \tag{20}$$

where $n$ is the number of observations; $i$ is the index of records; and $y_i^*$ and $y_i$ are the predicted action and observed action, respectively.

The genetic algorithm is used in the upper level problem to minimize the objective function since it is not limited by the form of functions [37].

On the contrary, the lower level problem is solved according to the perfect Bayesian equilibrium. In order to seek the entire solution set for the Nash equilibrium, tools like Gambit [38] and the Nashpy [39] package in Python [40] have been developed to find the equilibrium of a simple game. However, signalling game is more complicated because of its dynamic feature with incomplete information. Therefore, instead of adopting these packages directly, we explore three different equilibriums (i.e., separating equilibrium, pooling equilibrium, and semiseparating equilibrium) to seek the optimal solution in the level problem.

Two indicators were calculated to measure model predictability, namely, recall (also called sensitivity) and $F_1$ score [41]. They are widely used in the classification model in machine learning and can be derived from equations (21) and (22):

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{21}$$

$$F_{1\_\text{Score}} = 2 \times \frac{\text{TP}^2}{(\text{TP} + \text{FP}) \times (\text{TP} + \text{FN}) \times ((\text{TP}/(\text{TP} + \text{FP})) + (\text{TP}/(\text{TP} + \text{FN})))}, \tag{22}$$

Table 6: Summary of lane-changing duration.

| Source | Duration | |
|---|---|---|
| | Mean (s) | Range (s) |
| Olsen et al. [28] | 6.28 | — |
| Toledo and Zohar [29] | 4.6 | 1.0 to 13.3 |
| Gurupackiam and Jones [30] | 4.19 | — |
| Cao et al. [31] | 2.54 | 1.0 to 6.8 |
| Sajjad et al. [32] | 4.3 | — |
| Yang et al. [33] | 3.75/4.22 | — |
| Ali et al. [34] | 5.11 | — |

Table 7: Distribution of extracted strategies.

| No. of cases | | CAR-E | |
|---|---|---|---|
| | | Changing lane (C) | Waiting (W) |
| CAR-TF | Signalling (S) | 194 | 171 |
| | Not signalling (NS) | 197 | 179 |

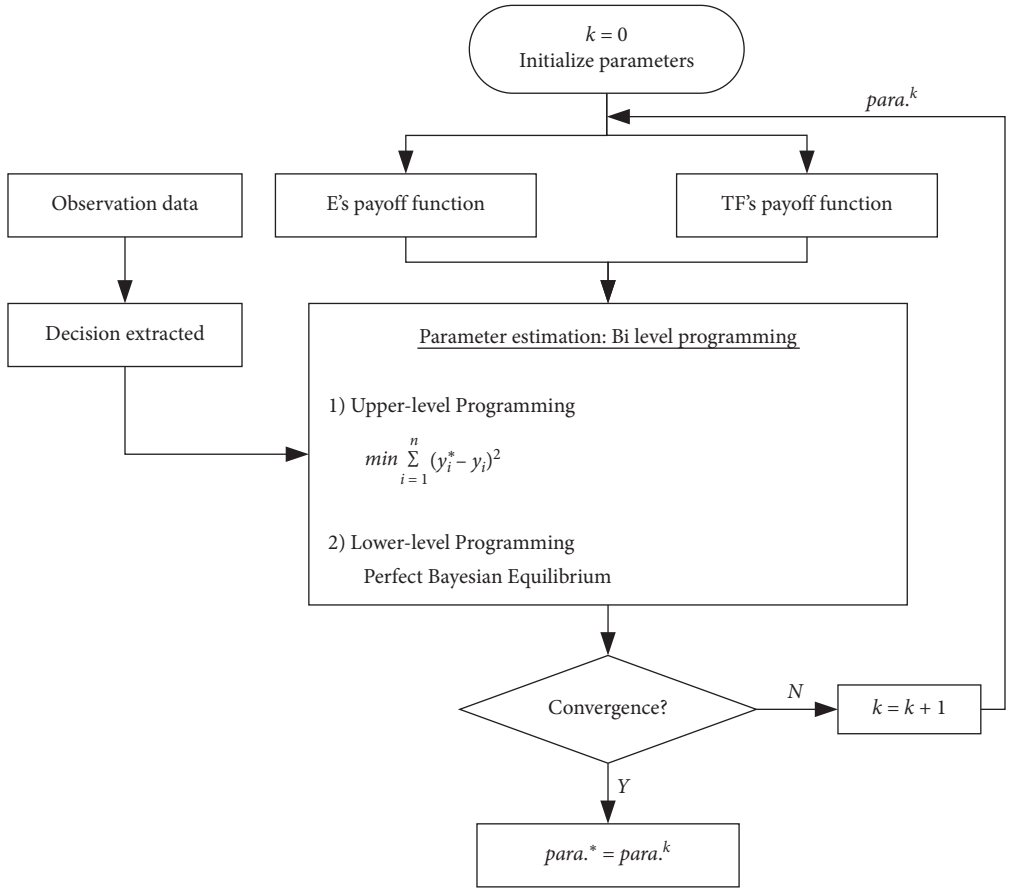



Figure 5: Calibraion framework.

where TP denotes the true positive cases, i.e., the observation and the predicted action are both positive; FP denotes the false positive cases, i.e., the observation is negative, while the predicted action is positive; FN denotes the false negative cases, i.e., the observation is positive, while the predicted action is negative.

Using 70% of the data to calibrate the model and the remaining in validation, Table 8 summarizes the results of parameters. We can see from the table that $\gamma_1$ is lower than $\gamma_2$, which means that conservative drivers are more likely to regret than aggressive drivers. The negative $\varepsilon$ indicates that distraction exists when CAR-TFs send a signal. Also, it can

Table 8: Results of calibration and validation.

| | Parameter | Value | Parameter | Value |
|---|---|---|---|---|
| Calibration | $\gamma_1$ | 0.170 | $g^1_{\mathrm{TF}}$ | 1.165 |
| | $\gamma_2$ | 0.363 | $g^2_{\mathrm{TF}}$ | −1.584 |
| | $\varepsilon$ | −1.198 | $g^1_E$ | 0.103 |
| | $f^1_{\mathrm{TF}}$ | 0.406 | $g^2_E$ | 1.789 |
| | $f^1_E$ | 0.144 | $p$ | 0.127 |
| Validation | Recall | 0.590 | | |
| | $F_1$_Score | 0.632 | | |

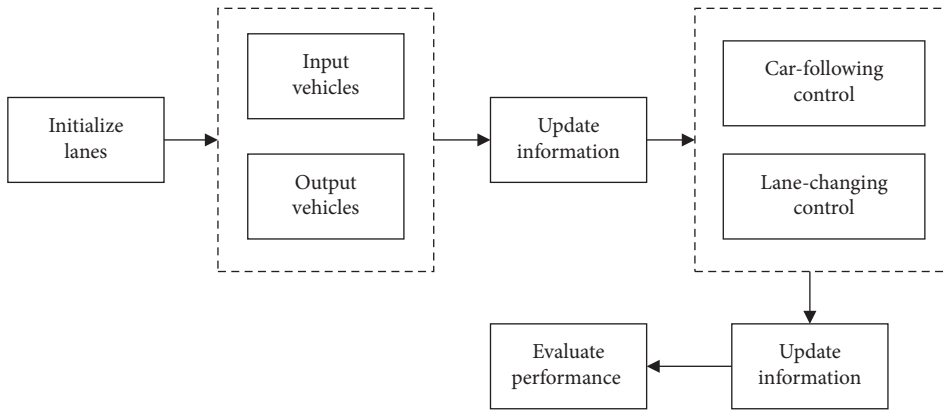



Figure 6: Framework of simulation.

be concluded that conservative drivers value distance more than aggressive drivers since $f^1_{\mathrm{TF}}$ is lower than $g^1_{\mathrm{TF}}$.

## 6. Simulation

Based on the calibrated parameters presented above, simulation experiments are performed. For the purpose of compassion, a traditional space-based LC model which only considers gaps between vehicles is also applied. The simulation framework is presented in Figure 6. It can be divided into five parts, lanes initializing, vehicles input and output, manoeuvre controlling, information updating, and performance evaluation.

The length of the lane is set to 10 kilometres. It is assumed that a vehicle is randomly distributed in two lanes at the initial time (initializing lanes). And there are three pairs of ramps on roads, where vehicles enter or exit the road randomly every second (vehicle input and output). In addition, the information of vehicles, such as position, speed, and acceleration, is updated every 0.1 second in the simulation (information updating). The car-following part is designed according to equation (23) [42]:

$$v_i = \begin{cases} v_i, D \geq D_s, \\ \min\left(v_{i+i}, \mathrm{sp}\right), D < D_s, \end{cases} \tag{23}$$

where sp is calculated by $\mathrm{sp} = (-\varphi + \sqrt{\varphi^2 + (4\gamma D_s)})/2\gamma$; $v_i$ and $v_{i+1}$ are defined as the instantaneous speed of the following vehicle and the leading vehicle, respectively; $D$ is the distance between two vehicles; $D_s$ denotes the safety distance; $\varphi = 0.75s$ is the reaction time; and $\gamma = 0.0070104\mathrm{s}^2/\mathrm{m}$ is the reciprocal of twice the maximum average deceleration of the vehicle. A detailed parameter setting is presented in Table 9. And an example of the simulation scenario is presented in Figure 7 (modified from Anushagj [43]).

## 7. Results and Discussion

This section evaluates the model performance using efficiency and safety indicators, which can be presented by the successful rate of LC for the first time (LC rate) and the reciprocal of time to collision (TTC) [44].

*7.1. Results of LC Rate.* LC rate means that drivers can finish LC successfully after the first attempt when there is a need to change lanes. Figure 8 presents the LC rates of the proposed model and the space-based model under different traffic flow densities. One can see that, with the increasing density, the space headway between vehicles decreases. Therefore, LC rates of the two models are both in the trend of declining. However, the LC rates of signalling game-based model are higher than those of the spaced-based model over the period, which means signalling game-based model is more efficient. A possible scenario is when the space headway between autonomous vehicles and regular vehicles is not big enough, a space-based controller chooses to wait. On the contrary, the signalling game-based controller analyses and interacts with the other, taking the best action, i.e., changing lanes. This helps autonomous vehicles understand the traffic information and make a more reasonable strategy instead of being overly cautious, i.e., waiting for bigger gaps.

Note that the rates represented are the results of the first attempt of a vehicle to change lanes at the very first time

Table 9: Simulation parameters.

| Parameter | Value |
|---|---|
| Road map | Two lanes, one direction, three pairs of ramps |
| Road length | 10 kilometres |
| Safety distance | 10 meters |
| Speed | [22.35, 31.29] m/s |
| Acceleration | [0, ±5] m/s$^2$ |
| Arrival/departure rate | 0.611 |
| Traffic density | 20–55 veh/ln/km |
| $p$ | 0.2 |

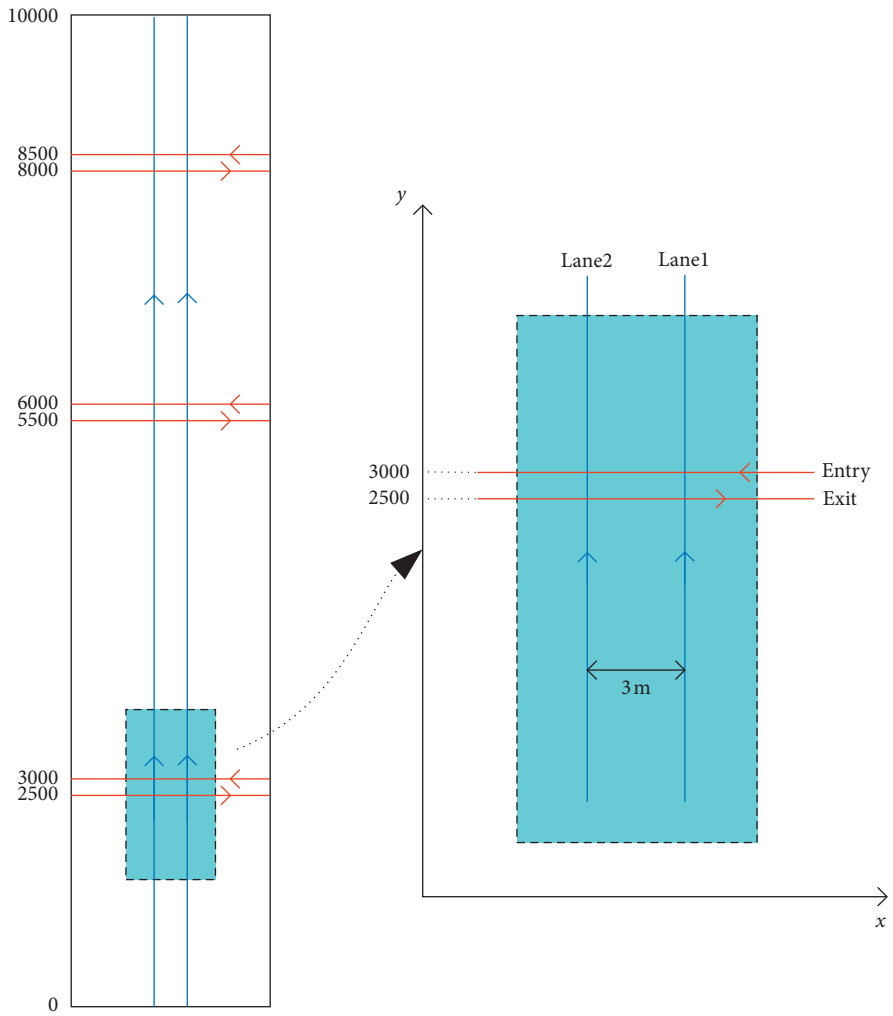



Figure 7: Simulation scenario.

when the lane-changing desire emerges. In reality, it is impossible that every driver can always change lane successfully as soon as he or she wants to because there still needs time to find a proper position to cut in. Therefore, it is comprehensible that the rates of two models are only 20% or so. In conclusion, the model established is superior to the space-based model in terms of efficiency.

*7.2. Results of $TTC^{-1}$.* TTC refers to the time that a target vehicle can use by adjusting its own speed to avoid collision with the preceding vehicle. Its reciprocal can be calculated by using the below equation:

$$\mathrm{TTC}^{-1} = \frac{v_F - v_L}{D}, \tag{24}$$

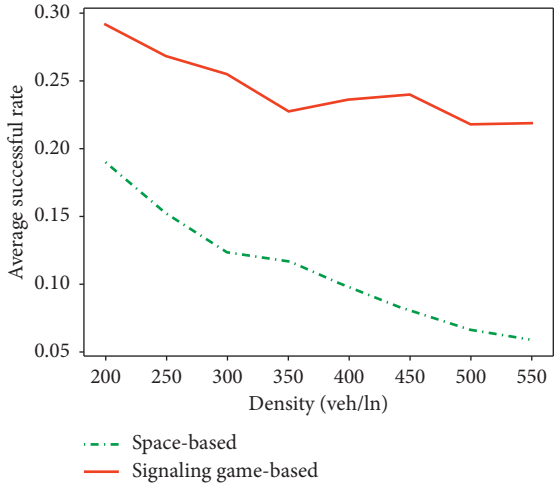



Figure 8: LC rate of the two models.

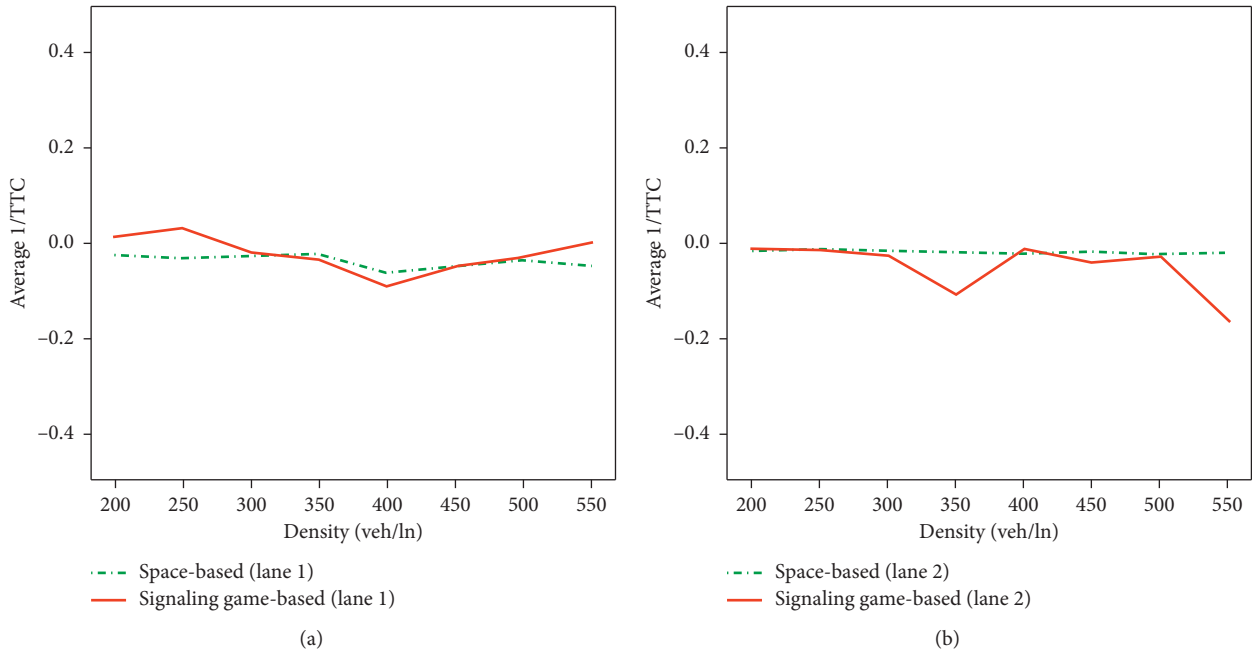



Figure 9: Average of $TTC^{-1}$. (a) $TTC^{-1}$ of the lane 1. (b) $TTC^{-1}$ of the lane 2.

where $v_F$ is the instantaneous speed of the following vehicle; $v_L$ is the instantaneous speed of the leading vehicle; and $D$ is the distance between the two vehicles.

Obviously, when the reciprocal of TTC is negative, the speed of the following vehicle is lower than that of the leading vehicle, so there is no danger of collision. Conversely, when the reciprocal of the collision time is positive (i.e., the speed of the following vehicle is greater than that of the leading vehicle), the risk of collision increases as the value increases. Figure 9 shows the results under different densities of traffic flow. The reciprocal of TTC of the two lanes stays relatively steady as density changes. Thus, the performance of signalling the game-based model is almost as good as the spaced-based model in terms of safety.

In order to examine the reliability of the above findings, we further compare the results of $TTC^{-1}$ of the two lanes under different scenarios of the decision-making controllers. The results are shown in Table 10. It can be seen that the 85$^{th}$

Table 10: $TTC^{-1}$ under different scenarios of the decision-making controllers.

| Density (veh/ln) | Indicator | $TTC^{-1}$ (1/s) | | | |
|---|---|---|---|---|---|
| | | Space-based | | Signalling game-based | |
| | | Lane 1 | Lane 2 | Lane 1 | Lane 2 |
| 200 | Mean | −0.023 | −0.034 | 0.013 | −0.024 |
| | 85th percentile | 0.134 | 0.125 | 0.136 | 0.129 |
| | 15th percentile | −0.170 | −0.179 | −0.156 | −0.170 |
| 250 | Mean | −0.031 | −0.022 | 0.033 | −0.030 |
| | 85th percentile | 0.131 | 0.138 | 0.130 | 0.138 |
| | 15th percentile | −0.174 | −0.171 | −0.175 | −0.191 |
| 300 | Mean | −0.026 | −0.031 | −0.019 | −0.052 |
| | 85th percentile | 0.144 | 0.155 | 0.129 | 0.132 |
| | 15th percentile | −0.185 | −0.187 | −0.200 | −0.187 |
| 350 | Mean | −0.024 | −0.043 | −0.034 | −0.216 |
| | 85th percentile | 0.153 | 0.140 | 0.127 | 0.124 |
| | 15th percentile | −0.186 | −0.201 | −0.189 | −0.204 |
| 400 | Mean | −0.062 | −0.044 | −0.092 | −0.022 |
| | 85th percentile | 0.138 | 0.138 | 0.114 | 0.126 |
| | 15th percentile | −0.196 | −0.186 | −0.200 | −0.193 |
| 450 | Mean | −0.048 | −0.035 | −0.049 | −0.079 |
| | 85th percentile | 0.125 | 0.126 | 0.131 | 0.122 |
| | 15th percentile | −0.188 | −0.180 | −0.211 | −0.201 |
| 500 | Mean | −0.036 | −0.047 | −0.028 | −0.054 |
| | 85th percentile | 0.122 | 0.118 | 0.088 | 0.117 |
| | 15th percentile | −0.186 | −0.189 | −0.188 | −0.195 |
| 550 | Mean | −0.047 | −0.043 | 0.002 | −0.328 |
| | 85th percentile | 0.118 | 0.118 | 0.108 | 0.105 |
| | 15th percentile | −0.173 | −0.188 | −0.205 | −0.196 |

percentile and the 15th percentile of $TTC^{-1}$ of the basic space-based model are slightly higher than those of the signalling game-based model, indicating that the proposed model results in a relatively low value of $TTC^{-1}$. In another word, the proposed model not only has higher efficiency (i.e., higher lane-changing rate) but also provides a relatively safer traffic condition.

To evaluate the performance of the proposed model in its stability, the model was run under different scenarios regarding the varied ratios of aggressive drivers ($p$) and traffic flow densities. Figure 10 and Table 11 show the results. It can be found that $TTC^{-1}$ values stay at a stable level under different scenarios, and the LC rate of the proposed model are in general higher than that of the space-based model. Therefore, we can conclude that the performance of the signalling game-based model is stable under different values of $p$.

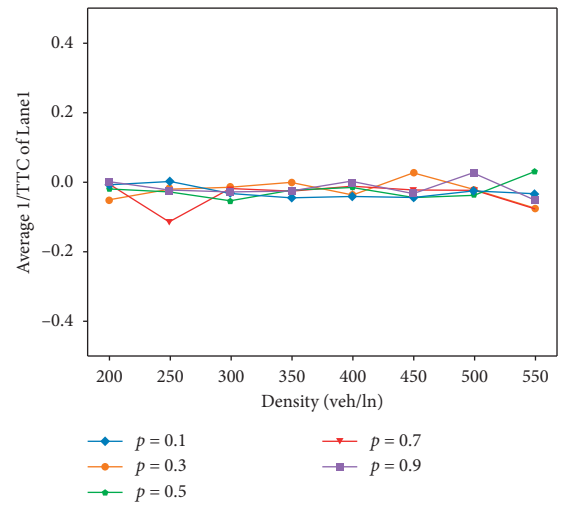



Figure 10: Average of $TTC^{-1}$ under different scenarios.

Table 11: Lane-changing rate under different scenarios (road traffic density and $p$).

| Density (veh/ln) | $p$ | LC rate (%) | |
|---|---|---|---|
| | | Signalling game-based | Space-based |
| 200 | 0.1 | 27.9 | 18.0 |
| | 0.3 | 28.4 | 17.2 |
| | 0.5 | 25.8 | 18.1 |
| | 0.7 | 22.9 | 16.0 |
| | 0.9 | 19.9 | 16.9 |
| 300 | 0.1 | 24.1 | 13.0 |
| | 0.3 | 26.3 | 12.3 |
| | 0.5 | 23.8 | 12.2 |
| | 0.7 | 21.8 | 11.9 |
| | 0.9 | 18.6 | 12.6 |
| 400 | 0.1 | 23.3 | 8.5 |
| | 0.3 | 24.7 | 9.5 |
| | 0.5 | 23.6 | 9.7 |
| | 0.7 | 19.6 | 8.6 |
| | 0.9 | 17.6 | 9.4 |
| 500 | 0.1 | 20.8 | 6.5 |
| | 0.3 | 23.4 | 6.4 |
| | 0.5 | 21.4 | 6.7 |
| | 0.7 | 20.7 | 6.5 |
| | 0.9 | 18.7 | 6.6 |

## 8. Conclusions

The main emphasis of this paper is to establish an integrated lane-changing model for autonomous vehicles under mixed traffic flow condition by applying the signalling game approach. Several simulations were performed to evaluate the performance of the proposed model. The results show that the proposed model has higher lane-changing rates than the space-based model and remains almost a similar level of $TTC^{-1}$ value at the same time under different densities. Therefore, we conclude that the proposed model improves the efficiency of lane-changing without decreasing the safety. Also, different ratios of the two types of drivers (conservative

or aggressive) are set to test the sensitivity of the model. One can see that the signal game-based model is stable to the varying ratios, which means that it can be used in different areas with different composition of drivers.

In spite that the proposal model outperforms the space-based model, there is still room to further improve the model. One possibility is to relax the assumption that only two types of drivers exist. This will unavoidably increase the complexity of the model. In addition, the data of mixed traffic flow involving autonomous vehicles do not exist in the NGSIM dataset, neither in many other available datasets. However, since autonomous vehicles are trained to imitate human behaviour, using traditional data to calibrate autonomous vehicles should not induce great difference. However, this should not rule out the possibility to recalibrate the proposed model when the data of connected vehicles become available.

Nevertheless, the proposed model shows its potential to be applied in complex situations, such as modelling the interactions and conflicts between connected vehicles with more driving styles and between vehicles with insufficient information of surroundings. When dealing with these cases, the scope of the game needs to be expanded further, e.g., add the behaviour of leading vehicles into the model, quantify the impacts of different driving behaviours on the process of lane-changing decision, and refine the payoff functions related to human psychology. It will also be interesting to consider the impact of the weather conditions and the appearance of road obstacles on the proposed model. We consider these aspects as our future work.

## Data Availability

Readers can access the data underlying the findings of the study from the corresponding author (zhangmr@chd.edu.cn).

## 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 Chang'an University Short-Term Study Abroad Program for Postgraduate Students (300203110004), the National Key R&D Program of China (2019YFB1600302), the Fundamental Research Funds for the Central Universities (CHD: 300102219210 and CHD: 300102210201), and the 111 Project of Sustainable Transportation for Urban Agglomeration in Western China (No. B20035).

## References

[1] Z. Zheng, "Recent developments and research needs in modeling lane changing," *Transportation Research Part B: Methodological*, vol. 60, pp. 16–32, 2014.

[2] Y. Ali, Z. Zheng, M. M. Haque, and M. Wang, "A game theory-based approach for modelling mandatory lane-changing behaviour in a connected environment," *Transportation Research Part C: Emerging Technologies*, vol. 106, pp. 220–242, 2019.

[3] D. Elliott, W. Keen, and L. Miao, "Recent advances in connected and automated vehicles," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 6, no. 2, pp. 109–131, 2019.

[4] A. Nagahama, D. Yanagisawa, and K. Nishinari, "Impact of next-nearest leading vehicles on followers' driving behaviours and traffic stability in mixed traffic," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 7, no. 1, pp. 42–51, 2020.

[5] P. G. Gipps, "A model for the structure of lane-changing decisions," *Transportation Research Part B: Methodological*, vol. 20, no. 5, pp. 403–414, 1986.

[6] Q. Yang and H. N. Koutsopoulos, "A Microscopic Traffic Simulator for evaluation of dynamic traffic management systems," *Transportation Research Part C: Emerging Technologies*, vol. 4, no. 3, pp. 113–129, 1996.

[7] P. Hidas, "Modelling vehicle interactions in microscopic simulation of merging and weaving," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 1, pp. 37–62, 2005.

[8] K. I. Ahmed, M. E. Ben-Akiva, H. N. Koutsopoulos, and R. G. Mishalani, "Models of freeway lane changing and gap acceptance behavior," in *Proceedings Of the 13th International Symposium On the Theory of Traffic Flow and Transportation*, pp. 501–515, Lyon, France, July 1996.

[9] K. I. Ahmed, *Modeling Drivers' Acceleration and Lane Changing Behavior*, Massachusetts Institute of Technology, Cambridge, MA, USA, 1999.

[10] T. Toledo, H. N. Koutsopoulos, and M. E. Ben-Akiva, "Modeling integrated lane-changing behavior," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1857, pp. 30–38, 2003.

[11] H. Kita, "A merging-giveway interaction model of cars in a merging section: a game theoretic analysis," *Transportation Research Part A: Policy and Practice*, vol. 33, no. 3-4, pp. 305–312, 1999.

[12] F. Meng, J. Su, C. Liu, and W. Chen, "Dynamic decision making in lane change: game theory with receding horizon," in *Proceedings of the 2016 UKACC 11th International Conference on Control (CONTROL)*, Belfast, UK, September 2016.

[13] H. Yu, H. E. Tseng, and R. Langari, "A human-like game theory-based controller for automatic lane changing," *Transportation Research Part C: Emerging Technologies*, vol. 88, pp. 140–158, 2018.

[14] FHWA, *US Highway 101 Dataset*, FHWA, Washington, DC, USA, 2007.

[15] Y. Liu, X. Wang, L. Li, S. Cheng, and Z. Chen, "A novel lane change decision-making model of autonomous vehicle based on support vector machine," *IEEE Access*, vol. 7, pp. 26543–26550, 2019.

[16] E. Balal, R. L. Cheu, and T. Sarkodie-Gyan, "A binary decision model for discretionary lane changing move based on fuzzy inference system," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 47–61, 2016.

[17] D. J. Sun and L. Elefteriadou, "Lane-changing behavior on urban streets: an "In-Vehicle" field experiment-based study," *Computer-Aided Civil and Infrastructure Engineering*, vol. 27, no. 7, pp. 525–542, 2012.

[18] C. Xue, G. Tan, and D. Nan, "Cooperative lane-changing model of human driving and unmanned driving based on game theory," *Computer Engineering*, vol. 43, pp. 261–266, 2017.

[19] M.-Y. Pang, B. Jia, D.-F. Xie, and X.-G. Li, "A probability lane-changing model considering memory effect and driver heterogeneity," *Transportmetrica B: Transport Dynamics*, vol. 8, no. 1, pp. 72–89, 2020.

[20] C. Zhang, J. Zhu, W. Wang, and J. Xi, *Spatiotemporal Learning of Multivehicle Interaction Patterns in Lane-Change Scenarios*, Cornell University, Ithaca, NY, USA, 2020.

[21] X. Li, W. Wang, and M. Roetting, "Estimating driver's lane-change intent considering driving style and contextual traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3258–3271, 2019.

[22] W. Spaniel, *Game Theory 101: The Complete Textbook*, CreateSpace Independent Publishing Platform, Scotts Valley, CA, USA, 2011.

[23] D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, Cambridge, MA, USA, 1991.

[24] A. Dixit and S. Skeath, *Games of Strategy*, W. W. Norton & Company, New York, NY, USA, 2004.

[25] FHWA, *US 101 Study Area*, FHWA, Washington, DC, USA, 2007.

[26] Q. Wang, Z. Li, and L. Li, "Investigation of discretionary lane-change characteristics using next-generation simulation data sets," *Journal of Intelligent Transportation Systems*, vol. 18, no. 3, pp. 246–253, 2014.

[27] H. Ozaki, "Reaction and anticipation in the car-following behavior," in *Proceedings of the 12th International Symposium on Theory of Traffic Flow and Transportation*, pp. 349–366, Berkeley, CA, USA, July 1993.

[28] E. C. B. Olsen, S. E. Lee, W. W. Wierwille, and M. J. Goodman, "Analysis of distribution, frequency, and duration of naturalistic lane changes," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 46, no. 22, pp. 1789–1793, 2002.

[29] T. Toledo and D. Zohar, "Modeling duration of lane changes," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1999, pp. 71–78, 2007.

[30] S. Gurupackiam and S. L. Jones, "Empirical study of accepted gap and lane change duration within arterial traffic under recurrent and non-recurrent congestion," *International Journal for Traffic and Transport Engineering*, vol. 2, no. 4, pp. 306–322, 2012.

[31] X. Cao, W. Young, and M. Sarvi, "Exploring duration of lane change execution," in *Proceedings of the Australasian Transport Research Forum*, Brisbane, Australia, October 2013.

[32] S. Sajjad, A. Shahram, K. Reza, and E. Arno, "Towards a decision-making algorithm for automatic lane change manoeuvre considering traffic dynamics," *PROMET-Traffic&-Transportation*, vol. 28, no. 2, 2016.

[33] L. Yang, X. Li, W. Guan, H. M. Zhang, and L. Fan, "Effect of traffic density on drivers' lane change and overtaking maneuvers in freeway situation-a driving simulator-based study," *Traffic Injury Prevention*, vol. 19, no. 6, pp. 594–600, 2018.

[34] Y. Ali, Z. Zheng, M. Haque, M. Yildirimoglu, and S. Washingtom, "Understanding the discretionary lane-changing behaviour in the connected environment," *Accident Analysis & Prevention*, vol. 137, 2020.

[35] H. Liu, W. Xin, Z. Adam, and X. Ban, "A game theoretical approach for modeling merging and yielding behavior at freeway on-ramp section," *Transportation and Traffic Theory*, pp. 197–211, Elsevier, Amsterdam, Netherlands, 2007.

[36] K. Kang and H. Rakha, "Game theoretical approach to model decision making for merging maneuvers at freeway on-ramps," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2623, pp. 19–28, 2017.

[37] A. Eiben and J. Smith, "Working with evolutionary algorithms," *Natural Computing Series, Introduction to Evolutionary Computing*, 2015.

[38] R. D. McKelvey, A. M. McLennan, and T. L. Turocy, Gambit: Software Tools for Game Theory," Version 16.0.1, 2016.

[39] V. Knight and J. Campbell, "Nashpy: a Python library for the computation of Nash equilibria," *Journal of Open Source Software*, vol. 3, no. 30, 2018.

[40] G. Rossum and F. Drake Jr., *Python Tutorial*, Centrum voor Wiskunde en Informatica Amsterdam, Amsterdam, Netherlands, 1995.

[41] A. Müller and S. Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*, Reilly-O'Reilly Media, Sebastopol, CA, USA, 2017.

[42] R. Rothery, *Car Following Models*, The University of Texas, Austin, TX, USA, 2002.

[43] Anushagj, "VANET-Mobiliy," 2015.

[44] J. Hayward, "Near-miss determination through use of a scale of danger," *Highway Research Record*, vol. 384, 1972.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8850591, 17 pages
https://doi.org/10.1155/2020/8850591

WILEY | Hindawi

*Research Article*

# An Approach for Handling Uncertainties Related to Behaviour and Vehicle Mixes in Traffic Simulation Experiments with Automated Vehicles

**Johan Olstam**,[1,2] **Fredrik Johansson**,[1] **Adriano Alessandrini**,[3] **Peter Sukennik**,[4] **Jochen Lohmiller**,[4] **and Markus Friedrich**[5]

[1]*Swedish National Road and Transport Research Institute (VTI), Linköping SE-581 95, Sweden*
[2]*Linköping University, Department of Science and Technology (ITN), Norrköping 60174, Sweden*
[3]*University of Florence, Dipartimento di Ingegneria Civile e Ambientale (DICEA), Firenze 50139, Italy*
[4]*PTV Group, Karlsruhe 76131, Germany*
[5]*University of Stuttgart, Institute for Road and Transport Science, Department for Transport Planning and Traffic Engineering, Stuttgart Pfaffenwaldring 7, 70569, Germany*

Correspondence should be addressed to Johan Olstam; johan.olstam@vti.se

Received 4 June 2020; Revised 25 August 2020; Accepted 8 September 2020; Published 18 September 2020

Academic Editor: Chengxiang Zhuge



The introduction of automated vehicles is expected to affect traffic performance. Microscopic traffic simulation offers good possibilities to investigate the potential effects of the introduction of automated vehicles. However, current microscopic traffic simulation models are designed for modelling human-driven vehicles. Thus, modelling the behaviour of automated vehicles requires further development. There are several possible ways to extend the models, but independent of approach a large problem is that the information available on how automated vehicles will behave is limited to today's partly automated vehicles. How future generations of automated vehicles will behave will be unknown for some time. There are also large uncertainties related to what automation functions are technically feasible, allowed, and actually activated by the users, for different road environments and at different stages of the transition from 0 to 100% of automated vehicles. This article presents an approach for handling several of these uncertainties by introducing conceptual descriptions of four different types of driving behaviour of automated vehicles (Rail-safe, Cautious, Normal, and All-knowing) and presents how these driving logics can be implemented in a commonly used traffic simulation program. The driving logics are also linked to assumptions on which logic that could operate in which environment at which part of the transition period. Simulation results for four different types of road facilities are also presented to illustrate potential effects on traffic performance of the driving logics. The simulation results show large variations in throughput, from large decreases to large increases, depending on driving logic and penetration rate.

## 1. Introduction

The introduction of automated vehicles (AVs) is expected to affect traffic performance. Both urban and national road authorities are interested in how the introduction should be handled and what measures they should or should not apply to avoid negative effects and boost positive effects of the introduction of AVs. Investigations on how road design and traffic control measures affect traffic performance are commonly based on results from traffic models.

Microscopic traffic simulation models are state-of-the-art tools in transport planning. By simulating the movements of every vehicle, the models provide indicators (travel time, queue length, vehicle throughput, etc.) describing the performance of road facilities. Traffic simulation models are typically applied for designing, testing, and analysing road network sections with their traffic control facilities. They extend traditional highway capacity manuals (HCM) by providing methods for capacity analysis with varying demand, demand-actuated traffic control facilities, and

coordinated signal control. Current microscopic traffic simulation models are designed for modelling vehicles with no automation. Hence, modelling the behaviour of AVs requires model extensions. These extended models also need to be calibrated and validated which is a problem since the systems to a large extent do not exist yet.

There are several possible approaches to incorporate the driving behaviour of AVs into traffic simulation models. However, there is limited information available on how AVs will behave. The information available so far is from tests with today's partly automated vehicles, mainly from test tracks. How future generations of AVs will behave will be unknown for some time. Thus, investigations of future traffic conditions with future versions of AVs and with higher penetration rates of AVs in general require consistent assumptions on how the behaviour of AVs will evolve.

Traffic simulation investigations of AVs commonly assume one type of automated vehicle and that all AVs behave the same. Furthermore, several traffic simulation investigations of AVs investigate vehicles with only one automation function, such as adaptive cruise control (ACC) or connected/cooperative adaptive cruise control (CACC). Investigations of one automation function at the time give valuable insights but limited knowledge on how the introduction of automated vehicles might affect the traffic system during the transition period. As discussed in Calvert et al. [1], the transition period from no to 100% AVs will most probably be long and it will likely include mixes of conventional (human-driven) vehicles and AVs with different levels of automation and different generations of automation functions.

Traffic simulation investigation of automated vehicles for estimation of effects during the transition period needs to apply a structured and systematic approach for handling the uncertainties related to how different generations of automated vehicles will behave and which mixes of different generations of automated vehicles that are likely to coexist at different stages of the transition period. The aim of this article is to present an approach for handling uncertainties related to the behaviour and composition of vehicle fleets with varying levels of automation in traffic simulation experiments for the transition period from mainly human-driven vehicles to 100% fully automated vehicles. This aim is approached by considering a discrete number of moments during the transition period, in addition to the commonly studied case of full automation and no automation.

This article is structured as follows. Section 2 gives an overview of traffic simulation models with respect to how the automated vehicles are modelled and how the transition period is handled in previous studies. The approach we suggest for the handling of different types of automated vehicles with different driving logics is presented in Section 3. Section 4 presents how these driving logics can be implemented in a microscopic traffic simulation model and Section 5 presents an example on how these driving logics affect traffic performance for four common road facility types. Section 6 ends the article with conclusions and needs for future research.

## 2. Approaches for Traffic Simulation including Automated Vehicles

Simulation of a traffic system that includes automated vehicles obviously requires modelling of the driving behaviour of the automated vehicles. There are several approaches for how the driving logics of automated vehicles can be incorporated in traffic simulation models (see Section 2.1). Furthermore, the transition from today's traffic situation with only a small fraction of partially automated vehicles to a future situation with a fleet consisting of exclusively fully automated vehicles will imply an extended time period with a range of mixes of vehicles with different levels of automation. Hence, traffic simulation experiments of the transition period need to consider a range of different automated vehicle behaviours. An overview of approaches for how the transition period has been handled in previously conducted simulation experiments is presented in Section 2.2.

*2.1. Approaches for Modelling of the Driving Behaviour of Automated Vehicles.* To the best of our knowledge, three approaches for how to include behaviour of driver assistance systems and automated vehicle functions are reported in the literature:

(i) Simulation of automated driving behaviour by adjustment of behavioural model parameters in the traffic simulation model

(ii) Replacement of behavioural models in the traffic simulation model with automated vehicle driving behaviour models

(iii) Extension of the driving behavioural models with "nanoscopic" modelling of automated vehicles, including simulation of sensors, vehicle dynamics, and driving behaviours

One of the simplest and most frequently used methods to include automated vehicles in traffic simulation models is to adjust parameters in an existing behaviour model for conventional vehicles to represent the driving behaviour of automated vehicles. This can, for example, be adjusting reaction time, gap-related parameters, acceleration parameters, and speed limit acceptance. There are traffic simulation investigations [2–13] that utilize this approach to simulate ACC/CACC equipped vehicles, or vehicles assumed to be highly automated in a specific road environment. Some investigations focus on one-lane roads without on/off-ramps and without overtaking possibilities (e.g., [14]), which, due to the fact that there is no lane-changing, is equivalent to simulation of just the ACC part of the automated vehicle functionality. The drawback of this approach is that the behavioural models in the traffic simulation model were developed to mimic human driver behaviour and it is unclear whether the driving behaviours of automated vehicles can be modelled by only adapting the parameters or if fundamental changes of the driving models are required. The advantage is that the unknown AV behaviour easily can be specified in terms of changes in relation to human driver

behaviour, e.g., shorter reaction times and changed desired speed distribution.

The second approach is to develop new submodels for the driving behaviour of the automated vehicles [15]. The most common approach is to replace the car-following model used for simulation of conventional vehicles with a new model describing the longitudinal control by some ACC or CACC logic, while the existing car-following model is still used for simulation of the conventional vehicles in the mixed flow.

The third approach extends the common vehicle-driver unit approach used in most microscopic traffic simulation models to a "nanoscopic" approach. Here, nanoscopic refers to more detailed modelling of the vehicle, which may include modelling of sensors, engine, gearbox, and vehicle dynamics. This approach can, for example, be used to analyse the dynamic characteristics and control algorithms of vehicles (see, e.g., Bahram et al. [15]). The practical solution for these kinds of simulations is commonly some kind of cosimulation in which an AV or AV vehicle simulation model is connected to a traffic simulation model using an application program interface. This approach has, for example, been used by automotive industry researches to support the vehicle simulation with a more realistic description of the traffic surroundings (see, e.g., Tapani et al. [6]). Other examples where traffic simulation has been combined with more detailed simulation models of vehicle dynamics, engine, and support systems, are in [16–19].

*2.2. Approaches for Simulation of a Mix of Automated and Manual Driven Vehicles.* Even though simulation investigations of platoons of ACC or CACC are relevant and provide important insights on platoon stability (see, e.g., Bose and Ioannou [20], Milanés and Shladover [21], or Xiao et al. [22]), these investigations give limited knowledge on how the introduction of automated vehicles might affect the traffic system during the transition period and for common road facilities.

Several studies aim to investigate how AVs affect traffic performance, but in principle the analysis is limited to how the longitudinal control (e.g., ACC or CACC) affects the traffic performance (see, e.g., Bierstedt et al. [4] and Van Arem et al. [23]). In some cases, also the speed limit acceptance is considered, as in Calvert et al. [1]. In addition to these AV/ACC investigations, there are several simulation investigations focusing purely on the effects of ACC [9, 21, 24–31], as well as investigations on effects of some kind of CACC [11, 16, 21, 32–34].

It is rather common that traffic simulation investigations of AVs focus on simulations of separate AV functions such as ACC, CACC, and lane keeping assistant (LKA). However, since AVs will be more than a vehicle with an ACC, several investigations focus on traffic effects of AVs with different assumptions of speed limit acceptance and gap acceptance in addition to the longitudinal control. Calvert et al. [1], for example, conduct simulations of vehicles with ACC and LKA, while Aria et al. [2] and Atkins [3] make adjustments to the car-following and lane-changing-related parameters in Vissim to reflect expected AV behaviour based on statements in the literature. Another example is Olia et al. [33] which includes modelling of both AVs and cooperative AVs by extending a car-following model, with detection range and desired spacing, and by introducing a cooperative merging model.

In addition to modelling of the driving behaviours of AVs, it may for some applications also be important to consider differences in perception between conventional vehicles and AVs. As mentioned in Section 2.1, nanoscopic modelling or cosimulation of vehicle and AV technology and traffic simulation are two ways to include sensor capabilities and the vehicles perception of the surroundings. However, it is also possible to capture some parts of the potential differences in perception by adjusting parameters, such as look-ahead distance and number of vehicles considered in car-following or lane-changing. Such approaches for taking the sensor range into account are, for example, used in Talebpour and Mahmassani [16], Aria et al. [2], and Atkins [3].

There are few investigations that address the issue that different types of automated vehicles will exist and that the capabilities of the AVs and their driving behaviour will evolve during the transition period. Calvert et al. [1] present estimations of how penetration rates of ACC, LKA, high automation, full automation, and vehicle cooperation might evolve during the transition period. The estimations are based on a wide range of various predictions from industry, academia, and authorities. The estimations are used to illustrate that the transition period will be quite long and include a mix of different types of automated vehicles. However, it does not seem like the simulations conducted and presented in Calvert et al. [1] investigate scenarios with different mixes of automated vehicles but only investigate different penetration levels of ACC + LKA equipped vehicles. Instead, the variation in AV behaviour is modelled by the distribution of the model parameters of the AVs, assuming the same distribution in parameters independently of the penetration rate. Variations in AV behaviour are also addressed in Atkins [3], which considers nine different AV capability levels ranging from cautious to assertive. However, no mix of capabilities is considered in the conducted simulation experiments. The experiments are investigating the traffic effects of penetration levels between 0 and 100% for each capability level separately. In the Analysis/Modelling/Simulation (AMS) Framework for Connected and Automated Vehicle Systems in the USA [35], the plan is to use four vehicle categories: manual nonconnected, manual connected, automated nonconnected, and automated connected. Similar simulation experimental designs of mixed traffic consisting of manual human-driven vehicles, automated nonconnected, and automated connected vehicles are presented in Mahmassani [36] and Mattas et al. [37]. However, both the planned investigations in Mahmassani et al. [35] and simulations presented in Mahmassani [36] and Mattas et al. [37] seem to assume that the driving behaviours of the two different versions of automated vehicles will stay the same independently of the penetration level. This indirect assumption that there is no correlation between

the technological development of AVs and the penetration rate seems to be made in most traffic simulation investigations described in the literature.

Another approach is to consider the effect of AV-human driver interaction in traffic flow simulation with the incorporation of more sophisticated human factors in mathematical models for driving behaviour. van Lint and Calvert [38] and Calvert and van Arem [39] proposed an improved simulation approach for the interactions between AVs and their drivers, and interactions with other human drivers. The framework especially considered the transition of control (ToC) as an important aspect of vehicle-driver interaction for the simulation of automated vehicles. More research is however needed in regard to human factors in driving and vehicle automation to refine the framework.

Do et al. [40] present a literature review of simulation-based investigations of connected and automated vehicles. Most of the studied articles simulate ACC or CACC systems, and for the ones simulating AVs or CAVs, the authors conclude (partly based on the conclusions in Milakis et al. [41]) that the model parameters are not calibrated based on real field data due to that AVs of level 3 or higher are still immature. Furthermore, Do et al. [40] conclude that the different studies use their own assumptions for the capabilities of the automated vehicles which can lead to inconsistencies in conclusions. Hence, "standardized driving characteristics of intelligent vehicles are necessary for future research studies as most studies use different assumptions on the key features of intelligent vehicles" and "the impact analysis of intelligent vehicles is still in a preliminary stage involving many uncertainties" [40].

## 3. Suggested Approach for How to Represent Different Levels of Automation

There are several uncertainties regarding the introduction of AVs. There are, for example, large uncertainties on how the mix of vehicles with different levels of automation will evolve during the transition period from 0 to 100% automated vehicles. One way of taking these uncertainties into account is to conduct scenario-based analysis.

To investigate the range of conditions that are likely to occur during the gradual introduction of AVs, we define three stages of coexistence: the introductory, established, and prevalent stages. These stages can be complemented by the stages of no automation and full automation. The use of three stages enables limitation of the number of scenarios, while still providing insight into the whole range of the transition period of introduction of AVs. The stages are not defined in terms of specific number of years in the future, but rather by the level of automation in a specific case study. The exact nature of the stages may vary significantly between different road facilities, and all three stages may not be relevant for all road facilities and case studies. Also, depending on many factors, including technological development and adoption rates, and changes to the legal framework, the stages may have vastly differing durations. Hence, defining the stages in terms of time is not only highly speculative but also problematic since the durations of the stages might vary. At a conceptual level, the stages are as follows:

(i) Introductory: automated driving has been introduced, but most vehicles are conventional cars. Automated driving is in general significantly constrained by limitations (real or perceived) in the technology.

(ii) Established: automated driving has been established as an important mode in some areas. Conventional driving still dominates in some road environments due to limitations (real or perceived) in the technology.

(iii) Prevalent: automated driving is the norm, but conventional driving is still present.

The descriptions of the stages above are not quantitative definitions, but rather qualitative descriptions. The quantitative definition of each stage is given by the penetration rates of the various AV classes defined below. When applying the suggested approach for analysing a specific case, relevant uncertainty factors, such as penetration rates and AV mixes, need to be quantified for each stage based on assumptions on the evolution and deployment of AVs for the area of study with respect to the road environment and driving context. The assumptions will be highly uncertain, especially for the later stages. Hence, several versions of each stage with different combinations of penetration rates and AV mixes need to be evaluated in order to obtain a range of simulation results for each stage to illustrate the uncertainty in traffic impacts during the different stages of the transition period. The result ranges may also be used to estimate at what stage one can expect significant benefits from the introduction of AVs.

In addition to handle the penetration rates of automated vehicles during the different stages of the transition period, there is a need to handle the variety of automated vehicles with different levels of automation. One possibility would be to use the SAE (SAE International 2018) levels, but they focus on to what extent the vehicle is driving by itself, where it can drive itself, and who is responsible for the driving. For example, SAE levels 1 and 2 relate to driver support systems that assist in the dynamic driving task, but the driver is responsible for the driving. At level 3, an automation driving system (ADS) performs the entire dynamic driving task but can only operate in some limited operational design domain (ODD) and the driver is responsible for the driving. Level 4 is an extension of level 3 in which the ADS is responsible for the driving. The focus of this article is automated vehicles with level 4 capabilities for some ODDs. However, the SAE levels do not distinguish how the driving behaviour varies between different levels or within a level. Therefore, we suggest that the level of automation is specified by the following two concepts:

(i) AV class (Basic AV, Intermediate AV, and Advanced AV)

(ii) Driving logic (Rail-safe, Cautious, Normal, and All-knowing) for different road environments

*3.1. AV Classes.* An AV class is a high-level description of the behaviour and capabilities of the vehicles. We assume that the main priority at each class is safety and that difference between the classes lies in the operational design domain (ODD) and how "offensively" the vehicles can handle different road environments and traffic contexts. We think that at least three classes are needed and suggest the following classes as a starting point:

(i) Basic AV: the first type of AVs with SAE level 4 capabilities only for one-directional traffic environments with physical separation with active modes. The behaviour is in general quite cautious and risk minimizing.

(ii) Intermediate AV: AVs with level 4 capabilities in some road environments and driving contexts. The behaviour at more complex road environments and driving contexts is still cautious and risk minimizing while the behaviour at less complex road environments and driving contexts can be less cautious and still be safe.

(iii) Advanced AV: AVs with level 4 capabilities in most road environments and driving contexts. The advanced AVs can drive more "offensively" but still safe in most road environments and driving contexts but still need to apply a more cautious behaviour in complex road environments and driving contexts.

*3.2. Driving Logics.* While an individual vehicle always belongs to the same AV class, its behaviour changes depending on the infrastructure and traffic conditions; structured environments with physical separation to other modes and directions of travel require less caution than complex environments with multidirectional interactions. That is, a given vehicle can drive more offensively—keep shorter following distance and accept smaller gaps—in highly structured environments like motorways, compared to less structured environments like urban streets or shared spaces. Thus, to specify the driving behaviour of a vehicle of a specific AV class, we need to specify how it behaves at each type of road environment it may encounter. This environment dependence of the behaviour could be specified taking into account details of the local conditions of the specific road link, but that would be prohibitively complex and would impede both model implementation and generalizations of the results. Thus, we propose using a small number of AV driving behaviours, called driving logics, paired with a small number of road types, to specify the environment-dependent behaviour of each AV class.

A basic assumption for all of the driving logics is that automation will lead to a shift in legal responsibility in case of accidents, from the driver to the manufacturer of the vehicle. Since this will lead to an accumulation of responsibility to a relatively small number of developers or corporate executives that will be legally responsible for all accidents of a given vehicle model, there will be strong incentives to minimize the number of accidents that can be seen as caused by the AV from a legal perspective. Thus, all automated vehicles are assumed to strictly follow the road code, and the user cannot, for example, set the desired speed above the legal limit.

The driving logics are given purely functional definitions, that is, the definitions specify the functionality of the vehicles without reference to what hardware or software that enables the functionality. This is an important simplification that allows specifying archetypal behaviours without differentiating based on speculations regarding the technology used to achieve the behaviour. We propose four driving logics: the first two correspond approximately to driving behaviour that have already been implemented in prototype AVs, while the remaining two correspond to possible future milestone capabilities of automated vehicles. We propose the following driving logics:

(i) Rail-safe: the vehicle follows a predefined path for the whole trip and emergency brakes if anything is on the collision course and slows down every time its sensors can have blind angles to avoid surprises. This driving logic is not dependent on communication or cooperation with other vehicles or the infrastructure.

(ii) Cautious: calculates gaps accurately and only merges when gaps are acceptable, and as the Rail-safe logic it slows down due to sensor blind angles. This driving logic is not dependent on communication or cooperation with other vehicles or the infrastructure.

(iii) Normal: uses the logic of an average driver but with the augmented (or diminished) capacities of the sensors for the perception of the surroundings. This type of driving logic may require devices for vehicle communication and cooperation.

(iv) All-knowing: perfect perception and prediction of the surroundings and the behaviour of the other road users. This automated driver is capable of forcing his way on other drivers whenever is needed without however ever being responsible for causing accidents. This type of driving logic requires devices for vehicle communication. If the devices fail, the logic may fall back to the Normal or Cautious driving logic depending on the driving context. With additional communication devices and control logics, the driving logic enables cooperation with other AVs with communication and cooperation functionality.

As indicated by the specifications of the driving logics above, communication is not explicitly modelled, but is for all driving logics except the All-knowing logic regarded as one of many possible technologies that could contribute to produce the specified functionality. For the All-knowing driving logic, however, communication is assumed to be required to achieve the functionality, but it is still not explicitly modelled.

The four driving logics are described in more detail in the following sections.

*3.2.1. The Rail-Safe Driving Logic.* Motivated by compliance with the machinery directive (Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC, ELI: http://data.europa.eu/eli/dir/2006/42/2016-04-20), the Rail-safe (RS) driving logic represents a fail-safe deterministic driving behaviour. The vehicle follows a predefined path from which it cannot deviate, and the speed is set by the guiding principle that it should be able to brake for anything that may come in its way. To guarantee the ability to emergency brake, the brakes are preactuated, as in rolling stock, so emergency braking will be executed even in the case of total power failure. This strategy is safe only up to certain speeds, depending on the restraining systems for the passengers, which significantly limits the top speeds of RS vehicles.

The maximum speed is determined by the distance to other road users and obstacles that blocks the field of view of the vehicle. The maximum speed is set such that the vehicle can brake with an acceptable deceleration for any road user that can reasonably enter its predefined path, including hypothetical road users currently out of view due to obstacles. In particular, this strategy implies the RS vehicle never approaches another road user closer than the "brick wall stop distance" (BWS), that is, the RS vehicle keeps a distance such that it can avoid collision by emergency braking even if the leading vehicle instantly stops.

The full focus on fail-safe operation leads to that the RS logic only can operate efficiently at higher speeds if it has a reserved lane with separation to both neighbouring lanes and sidewalks, in the form of either a buffer area or a physical barrier. Also, it can only cross conflicting streams if it has absolute right of way. Conceived to certify automated shuttle services on public roads together with some European Ministry of Transport, this driving logic is called Rail-safe logic because it follows for certification an approach derived from the rail technical standards as explained in Alessandrini [42]. The example in Figure 1 is an urban multilane arterial with three lanes per direction, unsignalised intersections at grade, and unprotected sidewalks. An infrastructure as such will never be certified for the "Rail-safe" logic but nevertheless is a good example to show what can and cannot be done with this logic and the others. More detailed explanations on how to insert automated road transport systems (ARTS) [43] in urban environments are given in Cignini et al. [44] and Tripodi et al. [45]. In Figure 1(a), the automated vehicle is all alone. It has a full view on the sidewalk, which is empty and can therefore ride at its maximum speed. As soon as a pedestrian is present on the sidewalk, the speed of the vehicle is limited by the time to collision with such "obstacle" that might arise if the pedestrian decides to suddenly cross the street. If the sidewalk is a metre away from the automated vehicle course and a pedestrian walk at a metre per second, the vehicle can only overtake the pedestrian at a speed which would allow it a full stop in one second. If the maximum allowed deceleration is 1.2 meters per square seconds (standing nonrestrained passengers on board), the maximum speed can be 1.2 meters per second (nearly 4.5 km/h), which grows to 18 km/h with sitting and belted passengers on board. To increase the vehicle's speed, the sidewalk needs either to be separated from the road or some safety boundary put between the sidewalk and the automated vehicle lane.

As shown in Figure 1(b), the vehicle cannot change lane and, as shown in Figure 1(c), it cannot turn left. To turn left, the vehicle must have a clearly marked lane (as shown in Figure 1(d)) and a dedicated traffic light without conflicting traffic stream allowed when the automated vehicle passes. Where necessary, this logic forces communication with the traffic lights in the infrastructures and with specially installed roadside sensors to guarantee operations as safe as in rail transport.

*3.2.2. The Cautious Driving Logic.* In contrast to the guiding principle of the RS logic, which is to do everything to not be involved in an accident, the guiding principle of the Cautious driving logic is to do everything not to *be responsible* for an accident. This implies that it always strictly follows the road code and always is on the safe side when accepting gaps.

Similar to the RS logic, the Cautious logic needs to assume that anything can turn up where its field of view is blocked and that following only can be done at BWS distance. However, in contrast to the RS logic, the Cautious logic assumes that other road users behave reasonably according to the road code; pedestrians are assumed to not suddenly jump out in the road, and vehicles in the neighbouring lanes are assumed not to suddenly cut in in front. Thus, the Cautious logic can handle nonsignalised conflicts, such as crossing or joining a traffic stream with right of way, changing lanes, or even overtaking slow vehicles. However, in all cases, it only acts when gaps are larger than BWS distance to guarantee that it does not cause an accident by conflicting with the prioritized flow.

BWS is not a criterion normally used on the roads; however, it is the only safety criterion which "can be proven safe". The reason for including BWS in the Cautious driving logic is that a vehicle keeping a BWS distance is never at fault in an accident. As discussed in Section 3.2, the legal responsibility in accidents will shift from drivers to makers concentrating responsibilities (especially the criminal one) for many accidents on few designers and company executives, many vehicle makers will want to be sure their vehicles will not be held responsible for causing accidents, and implementing BWS distance keeping would be one possible strategy to strongly reduce the number of accidents where the AV can be seen as causing the accident. Other criteria (for the Normal and the All-knowing driving logics) have been set to consider more normal and aggressive driving behaviours, but for the cautions it was necessary to set a reference which has a specific legal reasoning behind.

In Figure 2(c) the vehicle is allowed to turn left. However, it can do so only if each vehicle in the incoming lane is far enough away so that they do not have to decelerate. Assuming each vehicle will continue with its own speed and given the speed of the automated vehicle when the automated vehicle clears the path of the incoming vehicle, a BWS distance needs still to be in place. Should for any reason the automated vehicle stop in the middle of the crossing, the

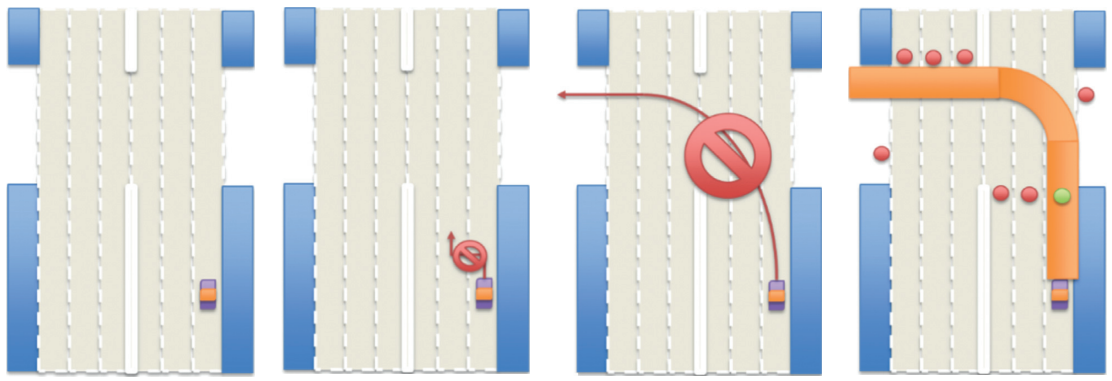



Figure 1: Illustration of the Rail-safe driving logic for road following (a), prohibited lane-changing (b), and left-turning (c). To allow left-turning for a Rail-safe logic, the road needs to be clearly marked and a traffic light inserted with a dedicated aspect for the automated vehicle left-turning which can even turn left from the right lane at this point (d).

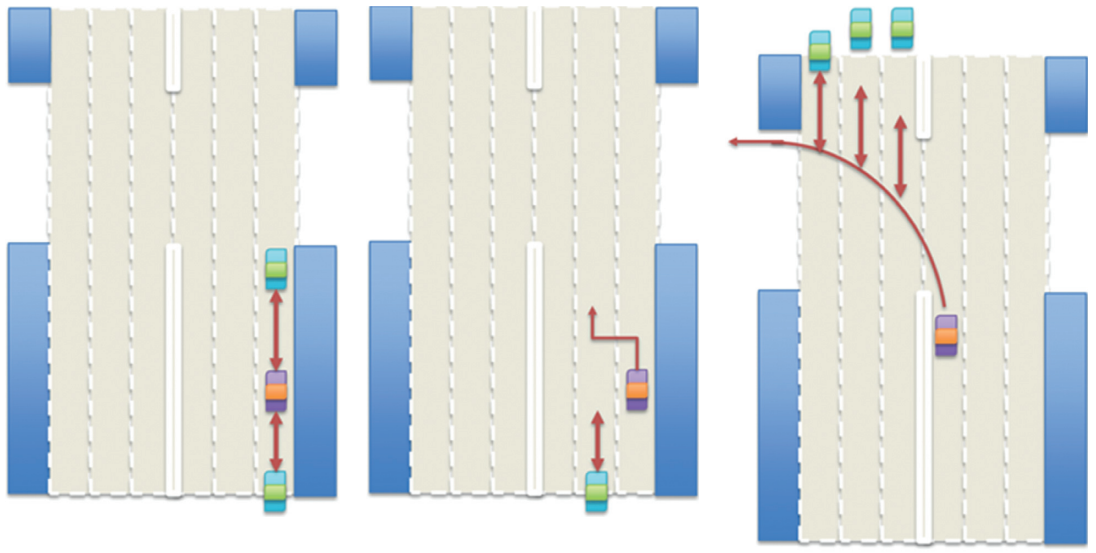



Figure 2: Gap acceptance for car-following (a), lane-changing (b), and left-turning (c) in the Cautious driving logic.

incoming vehicle can still come to a full stop without risking collision with the stopped automated vehicle. In case the automated vehicle passes through the intersection as planned, the incoming vehicle does not need to decelerate.

The case is similar in lane-changing (Figure 2(b)). The lane-changing manoeuvre is considered possible when the automated vehicle has reached the same speed of the flow of vehicles in the lane and the vehicle in the back (at the same speed) still has a BWS distance from the automated vehicles. This leads to the difficulty to conduct lane changes in heavy traffic.

*3.2.3. The Normal Driving Logic.* The Normal driving logic imitates the general human driver behaviour, but with the advantages and limitations of machines compared to humans, for example, shorter reaction time, more exact measurements of distances and time gaps, and precise execution of intended manoeuvres, but possibly more occlusion of the field of view by dynamical obstacles. However, the Normal driving logic does not imitate unwanted features of human drivers; it complies perfectly with the road code and there is no randomness in its behaviour.

The Normal driving logic is easy to implement in traffic simulation software: the driving logic is modelled using the regular model for conventional driving, but with stochasticity removed, reaction time improved, and occlusion of sensors approximately included. Accepted gaps and following gaps are similar in size to the ones of human drivers, due to that the Normal driving logic has shorter reaction time but lacks some of the predictive capabilities of human drivers.

*3.2.4. The All-Knowing Driving Logic.* For automation to lead to more efficient, in addition to safer, traffic, a vehicle

with the All-knowing driving logic predicts the behaviour of all road users detected by all connected detectors, on both vehicles and the infrastructure. This leads to very few blind spots and massive amounts of data to use as a basis for accurate predictions, which allows the All-knowing vehicle to keep short gaps and even force its way in conflicts when needed to achieve efficient flow, while keeping its manoeuvres safer than most human drivers.

This driving logic is of course much easier to simulate than implement in reality; in the simulation, we know the exact trajectory of all road users and their behaviour, so predicting it is not a problem. In reality, on the other hand, it is completely unknown how this driving logic would be implemented or how hard it would be. The All-knowing driving logic thus represents an extreme case that will not be realized in the foreseeable future.

Lane-changing as shown in Figure 3(b) happens with such small gaps that the following vehicle yields to let it merge as in a sort of cooperative manoeuvre. Left-turning as shown in Figure 3(c) can be done with small gaps and even when some of the vehicles in the furthest lanes are not in sight. Such behaviour can happen either because prediction models are perfect or because there is cooperation between vehicles; however, in simulation, it is not important which is the technology but only the resulting behaviour.

*3.3. Assumptions on the Relation between Driving Logics and AV Classes for Road Environments.* Simulations of automated vehicles do need to consider not only the driving logic of the vehicle but also the operational design domain (ODD) in which it can operate and according to which driving logic. As noticed in previous research [46], automated vehicles may have to adjust their driving behaviour depending on the road environment and the driving context. In this section, we present initial assumptions on which type of driving logic vehicles of different AV classes would be able to use for different types of vehicles and for different types of road environments. These assumptions were derived based on workshops and discussions between traffic model developers, traffic engineers, vehicle industry, and human factors researchers. It is important to note that these are general assumptions and that they might need to be adjusted depending on information or expectations for a specific case study. In this article, we used the following division and definitions of road environments:

(i) Motorway: multilane roads with physical barriers between directions and grade-separated intersections

(ii) Arterial: single- or multilane roads with at-grade intersections (mainly larger type of intersections as signalized intersections or roundabouts). Bicycle and pedestrian traffic are clearly separated from the vehicle traffic by either physical barriers or medians. Vehicles, bikes, and pedestrian may interact at intersections depending on if secondary conflicts between vehicles and active modes are separated or not.

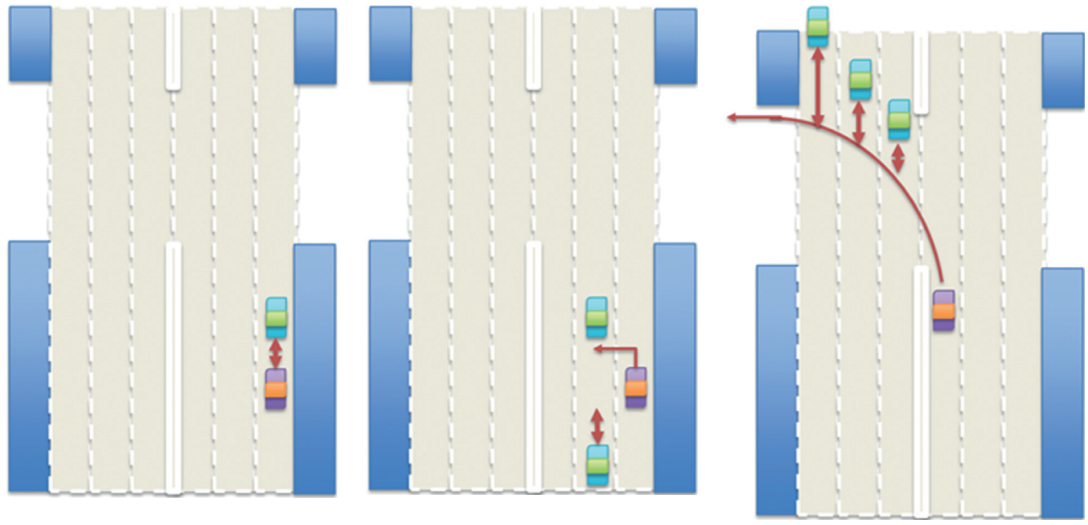
(iii) Urban street: single- or multilane roads with at-grade intersections (also stop or yield regulated intersection). No physical separation between vehicle traffic and pedestrian and bicycle traffic. Walkways and bikeways directly at the side of the vehicle lanes.

(iv) Shared space: vehicle, bicycles, and pedestrian share the same space, which can be unstructured or semistructured.

Table 1 presents the assumptions for which driving logic cars and trucks will use at the various road environments. Basic AVs are assumed to be able to drive in automated mode only on motorways and arterials. Furthermore, they are assumed to drive according to the Cautious driving logic on these road types. The Intermediate AVs are assumed to be able to use the Normal driving logic on motorways due to the development of sensor technology and anticipatory capabilities. The more complex arterials with at-grade intersections with interactions with active modes still constrain the AVs capabilities, and the behaviour is still according to the Cautious driving logic. Exceptions might be highly separated arterials with total conflict separation between vehicles and active modes, in which Intermediate AVs can be assumed to be able to drive according to the Normal driving logic. Furthermore, the Intermediate AV is assumed to be able to drive according to a Cautious driving logic on urban streets and according to the Rail-safe logic in shared spaces. However, it is questionable to what extent drivers would accept the Rail-safe logic driving in shared space due to the high "politeness" of the driving logic and the resulting low speed. Thus, it might from a traffic simulation point of view be more reasonable to assume manual driving. The Advanced AV class cars are assumed to be able to drive according to at least the Cautious logic in all the road environments, ranging from the All-knowing driving logic on motorways and arterials to the Normal driving logic on arterials and Cautious logic in shared spaces.

## 4. Implementation of AV Classes and Driving Logics in the Traffic Simulation Model VISSIM

The descriptions of the driving logics in the previous section are of conceptual nature. Further specifications are required to be able to implement the driving logics in a traffic simulation model. As described in Section 2.1, driving logics can be implemented either by replacing behavioural models in the traffic simulation model with automated vehicle driving behaviour models or by adjustment of behavioural model parameters in the traffic simulation model. The approach in this article has been to as far as possible try to adjust the behavioural parameters of the available behavioural models and when necessary extend the currently available models. This approach is chosen in favour of replacing the behavioural models. The idea behind this is that it, from a traffic performance point of view, is more important to capture the main characteristics of different



(a) (b) (c)

Figure 3: Gap acceptance for car-following (a), lane-changing (b), and left-turning (c) in the All-knowing driving logic.

Table 1: Specification of the driving logics envisioned for the three AV classes in the different road environments for cars.

| Road type | Basic AV | Intermediate AV | Advanced AV |
|---|---|---|---|
| Motorway | Cautious | Normal | All-knowing |
| Arterial | Cautious | Cautious/Normal[1] | All-knowing |
| Urban street | Manual | Cautious | Normal |
| Shared space | Manual | Rail-safe[2]/Manual | Cautious |

[1]The Normal driving logic will only be possible if all conflicts between vehicles and active modes are separated, e.g., by separate phases in a traffic signal. [2]It is questionable to what extent drivers would accept the Rail-safe logic driving in shared space due to the high "politeness" of the driving logic and the resulting low speed. Thus, it might from a traffic simulation point of view be more reasonable to assume manual driving.

types of automated vehicles instead of modelling all different types and combination of automation functions in detail.

Independently of approach, data for calibration of parameters are required. The Rail-safe and the Cautious driving logics correspond approximately to driving behaviour that have already been implemented in prototype AVs, while the Normal and All-knowing logics correspond to possible future milestone capabilities of automated vehicles. Hence, data from field test of current AV concepts can only be used to some extent to calibrate the suggested driving logics. We have therefore combined the data from field trials and cosimulations, which can give some information about currently available driving logics, e.g., with respect to following behaviour, with assumptions based on the conceptual descriptions of the driving logics presented in Section 3.2.

*4.1. Results from Test-Track Tests with Automated Vehicles with TNO's Driving Logic.* Within the CoEXist project, a test-track field test using CAV prototypes developed by TNO was conducted. Three vehicles were used in the field test, one with ACC, one with Cooperative ACC (CACC), and one with Degraded CACC (dCACC). Several different scenarios with different combinations of behaviour of the three vehicles in a platoon were tested (see example in Figure 4). In the example, the first vehicle is driven by a human, the second by an ACC-controlled vehicle, and the third by a CACC-controlled vehicle. Two different cases are shown: one in which the communication with the second vehicle regarding relative position and speed is on and then when it was turned off.

In this section, we summarize the key findings of this data collection, and the full details of the design, execution, and results are available in Sukennik et al. [47]. The key findings from the experiments are listed below. These findings might not be generalizable to all types of cooperative vehicle-following systems but may apply only to the specific control strategy that the test vehicles used in this research. However, it is always difficult to sustain that any conclusion is universally valid when most of the automated driving technology which needs to be implemented in the traffic simulation model is still to be developed.

(i) There is a linear deterministic relationship between headway and speed when an automated car is following another automated vehicle with car-to-car (C2C) communication. Human imperfection while driving is replaced by higher precision and deterministic nature of technical equipment and algorithms.

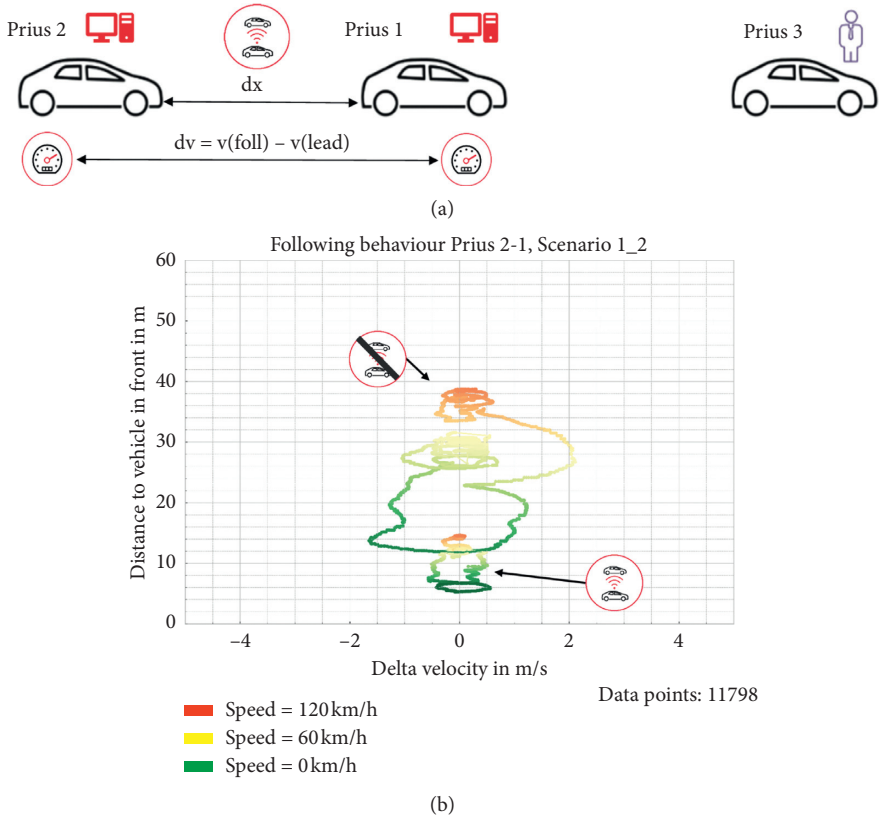



Figure 4: Relationship between speed difference and distance to preceding vehicle when following.

(ii) There is an almost linear relationship between headway and speed when an automated car is following a manually driven car or an automated car without C2C communication. The linear relationship is not as neat as with C2C communication but could be approximated.

(iii) Oscillations during the following process are small and without much variance in comparison with human drivers.

(iv) Safety distance without C2C communication is much higher than in the communication case: with C2C communication, the test vehicles were able to drive safely with a 0.6- or 0.3-second headway. After the disconnection of the C2C communication, the vehicle adapted larger following distance because of safety reasons.

(v) The safety distance in drive-away behaviour is larger when there is no communication. When following from standstill, the test vehicle kept a significantly larger safety distance in the case without C2C communication than with C2C.

(vi) No stochastic variation in drive-away behaviour.

(vii) When an automated vehicle followed another vehicle from a standstill (in front of a signal head), the following process did not show stochastic variations—the same behaviour applied each time.

*4.2. Parameter Changes in Vissim and Added Modelling Concepts.* The AV classes described in Section 3.1 are implemented in Vissim in a common way similarly as conventional vehicles. Each link representing a road segment is coupled with a link behaviour type. For a link behaviour type, the modeller can specify a driving behaviour for each vehicle class.

The driving logics are implemented in Vissim as new driving behaviours based on the Wiedemann 99 [48] driving behaviour model. The Wiedemann model operates at small reaction times (each time step) and takes oscillation in car-following into account (as observed by AVs in Figure 4). This is important for modelling AVs. But the model also accounts for psychological aspects as well as for physiological restrictions of drivers' perception of what we do not expect from AVs driver logics. Therefore, an extension to the model is required to model the behaviour of AVs as observed in Section 4.1. The contribution to the model results in these new features to allow for fundamental principles needed to model the AV driving logics:

(i) Reduction of implicit stochastic: the option "use implicit stochastic" can now be switched off for a specific driving behaviour. A vehicle using this driving behaviour does not use any internal stochastic variation, which is meant to model the imperfection of human drivers. For all distributions that cannot be explicitly set by the user, a median value is used instead of a random value.

(ii) Brick wall stop distance: the option "enforce absolute braking distance" can now be activated for a specific driving behaviour. Vehicles using this driving behaviour will always make sure that they can brake without a collision, even if the leading vehicle comes to an immediate stop (turns into a brick wall). Brick wall stop (BWS) distance is maintained by vehicles with the Rail-safe and Cautious driving logics.

(iii) Differentiable following parameters based on the leader vehicle class: this allows to set different headways for CAV following another CAV as for CAV following conventional vehicle or for any other combination of vehicle classes. This feature is applicable for driving logics such as the All-knowing logic which is based on the vehicle capability to recognise the leader vehicle class in the following process.

(iv) Sensors/equipment limitations: number of objects and vehicles that the vehicle can see and interact with can be defined separately for each driving logic. The number of interaction vehicles defines an upper limit for the observed leading vehicles; therefore, for example, this could be set to 1 for automated vehicles with a sensor equipment that cannot see through the leading vehicle. A red signal downstream of the leading vehicle could still be observed, but not the second real vehicle downstream.

The findings from the field trials in combinations with calibrations of the Wiedemann 99 model to the field trial vehicle trajectories were used to estimate parameters for the following behaviour in Vissim (see Table 2). To describe a more deterministic behaviour, the oscillation-related parameters are "turned off" expect that some minor acceleration oscillations are kept. The more advanced logics (e.g., Normal and All-knowing) possibilities to keep shorter gaps are implemented using the three following gap-related parameters (CC0, CC1, and CC3). Acceleration rate-related parameters are adjusted to reflect the more cautious behaviour of the Rail-safe and Cautious driving logics and the more offensive behaviour of the All-knowing logic.

Since no field data were available for adjustment of the lane-changing behaviour parameters, these are adjusted based on the conceptual descriptions of the driving logics according to Tables 3 and 4. Maximum and desired accepted deceleration rates follow the same pattern as for the car-following acceleration-related parameters. Hence, the Cautious driving logic does not accept as high deceleration rates as a consequence of a potential lane change. The parameter "$-1\,\mathrm{m/s^2}$ per distance" affects the distance at which a driver starts to accept higher deceleration rates than the desired acceleration rate for accepting a gap in the target lane. The Cautious driving logic is assumed to start accepting higher deceleration rates later while the Normal and All-knowing would behave similarly to a human driver in this matter.

Table 4 presents the parameter values related to cooperation and gap acceptance in connection with lane changes. The Normal logic is assumed to follow a human driver while the Cautious logic requires a larger minimum headway, less reduction of safety distance, and lower deceleration rates for cooperation.

In addition to the car-following and lane-changing parameters, AVs are assumed to drive more homogeneously compared to human-driven vehicles in terms of individual speeds and acceleration/deceleration behaviour. Desired speed distributions for AVs are assumed to be narrow and concentrated around the speed limit since the AVs are assumed to follow the road code. Individual settings of desired speeds may be possible for lower SAE levels (as driver support systems or conditional automation), like in today's automated cruise control vehicles, where the driver still is responsible for the driving. However, as discussed in Section 3.2, we assume that the producer of the vehicle will get increased legal responsibilities at SAE level 4 and we therefore assume complete compliance with the road code.

Furthermore, it is expected that automated vehicles will behave deterministically instead of stochastically like human drivers, which might have implications on the acceleration and deceleration behaviours. Therefore, all AVs are assumed to have the same maximum and desired acceleration and deceleration values. Different types of AVs or different vehicle brands might of course apply different acceleration rates in similar situations. However, when operated in automated mode, the utilized acceleration and deceleration rates will need to be constrained by comfort and safety requirements of the passengers, which lead to the assumption that the acceleration behaviour of automated vehicle will be very similar.

Moreover, the perception capabilities of the different driving logics are considered by limiting the number of vehicles ahead taken into account to one vehicle for the Rail-safe, Cautious, and Normal driving logics. Reactions on signals are assumed the same for all AV driving logics and the main differences compared to human drivers are that the AVs require full safety distance in interaction with other vehicles and that they only start passing the signal when it is green (red-amber is interpreted as stop). The AVs are assumed to have no reaction time while the reaction time of the human drivers depends on site-specific calibration. In the example simulations presented in Section 5, humans have zero reaction time.

## 5. Numerical Experiments

To illustrate the effects on traffic performance of the different driving logics, a set of simulations was conducted and results in form of maximal throughput and fundamental diagrams were produced. Here we present simulations of four different networks representing basic traffic situations:

(1) A simple one-lane link under ideal conditions without influence of intersections, parking manoeuvres, or other sources of disturbance

(2) A two-lane motorway without on- and off-ramps but with varying uphill gradient (from 0 to 3.4%)

Table 2: Recommended driving behaviour parameters for following in Vissim.

| Parameter (Wiedemann 99 following model) * * | Driving logic | | | | |
|---|---|---|---|---|---|
| | Rail-safe | Cautious | Normal | All-knowing | Def * * * |
| CC0—standstill distance (m) | 1.5 | 1.5 | 1.5 | 1 | 1.5 |
| CC1—spacing time (s) | 1.5 * | 1.5 * | 0.9 | 0.7 * * * * | 0.9 |
| CC2—following variation (m) | 0 | 0 | 0 | 0 | 4 |
| CC3—threshold for entering "following" (s) | −10 | −10 | −8 | −6 | −8 |
| CC4—negative "following" threshold (m/s) | −0.1 | −0.1 | −0.1 | −0.1 | −0.35 |
| CC5—positive "following" threshold (m/s) | 0.1 | 0.1 | 0.1 | 0.1 | 0.35 |
| CC6—speed dependency of oscillation ($10^{-4}$ rad/s) | 0 | 0 | 0 | 0 | 11.44 |
| CC7—oscillation acceleration ($m/s^2$) | 0.1 | 0.1 | 0.1 | 0.1 | 0.25 |
| CC8—standstill acceleration ($m/s^2$) | 2 | 3 | 3.5 | 4 | 3.5 |
| CC9—acceleration at 80 km/h ($m/s^2$) | 1.2 | 1.2 | 1.5 | 2 | 1.5 |

*If "enforce absolute braking distance" is on, brick wall stop distance is guaranteed. * *See Vissim manual [48] for detailed description. * * *Default values for Wiedemann99 following model in Vissim (conventional vehicles). * * * * If the followed vehicle is a conventional one, the follower maintains 0.9 s spacing time.

Table 3: Recommended driving behaviour parameters for necessary lane change in Vissim.

| | Driving logic—urban (motorway if the value differs) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Rail-safe | | Cautious * * | | Normal | | All-knowing | | Def | |
| Parameter for necessary lane change * | Own veh | Trailing veh | Own veh | Trailing veh | Own veh | Trailing veh | Own veh | Trailing veh | Own veh | Trailing veh |
| Maximum deceleration | n.a. | n.a. | −3.5 | −2.5 | −4 | −3 | −4 | −4 | −4 | −3 |
| $-1\ m/s^2$ per distance | n.a. | n.a. | 80 (160) | 80 (160) | 100 (200) | 100 (200) | 100 (200) | 100 (200) | 100 (200) | 100 (200) |
| Accepted deceleration | n.a. | n.a. | −1 | −1(−0,5) | −1 | −1 (−0,5) | −1 | −1.5 (−1) | −1 | −1 (−0,5) |

*Necessary lane change means a lane change which is necessary in order to follow a defined route (it is not overtaking because of higher own desired speed).
* * If "enforce absolute braking distance" is on, brick wall stop (BWS) distance is guaranteed. n.a. = not available.

Table 4: Recommended driving behaviour parameters for lane change in Vissim.

| Behavioural functionality | Driving logic | | | | |
|---|---|---|---|---|---|
| | Rail-safe | Cautious * * | Normal | All-knowing | Def |
| Advance merging * | n.a. | On * * */off | On * * * | On | On |
| Cooperative lane change * | n.a. | On * * */off | On * * * | On | Off |
| Safety distance reduction factor | n.a. | 1 + EABD | 0.6 | 0.75 | 0.6 |
| Min. headway (front/rear) | n.a. | 1 | 0.5 | 0.5 | 0.5 |
| Max. deceleration for cooperative braking | n.a. | −2.5 | −3 | −6 | −3 |

*Depends on technical equipment and implemented connectivity and cooperation functions. * *EABD (enforce absolute breaking distance) must be on.
* * *If the AV cannot detect that the other vehicle wants to change lanes, the value should be off/zero. n.a. = not applicable.

(3) A three-lane motorway section with one on-ramp and one off-ramp

(4) A one-lane approach to a traffic signal (considering only one approach)

The networks were populated with different shares of vehicles using one of the AV driving logics Cautious, Normal, and All-knowing at the time or a combination of those. The remaining shares of vehicles were simulated as human-driven vehicles using calibrated parameters (motorway networks with interchange and road with gradients) or the default parameters (models with simple link and link with signal) in Vissim. In case of combinations of different AV driving logics, the shares between the AV driving logics are equally distributed. For example, in case of 50% AV penetration rate for a combination of AV cautious and AV normal, 50% of the vehicles are human-driven vehicles, 25% use AV cautious, and 25% use AV normal. Simulations of the Rail-safe logic were not conducted since the logic mainly is applied to minibuses and shuttles which commonly do not reach high penetration rates of all vehicles on the road. To observe the impact of the driving logics compared to human-driven vehicles in terms of throughput, penetration rates (which equals the proportion of AVs compared to all vehicles in the simulation) of 10% steps from 0 to 100% are simulated. For each driving logic and penetration rate, several simulations with eight different demand configurations were conducted. For each combination of driving logic, penetration rate, and demand category, we run 10 simulations with different random seeds. Initially, the resulting relations between flow, density, and speed were then curve fitted to the functional relation by van Aerde [32]. However, the curve fitting was not appropriate at some locations of some networks because no breakdowns occur in

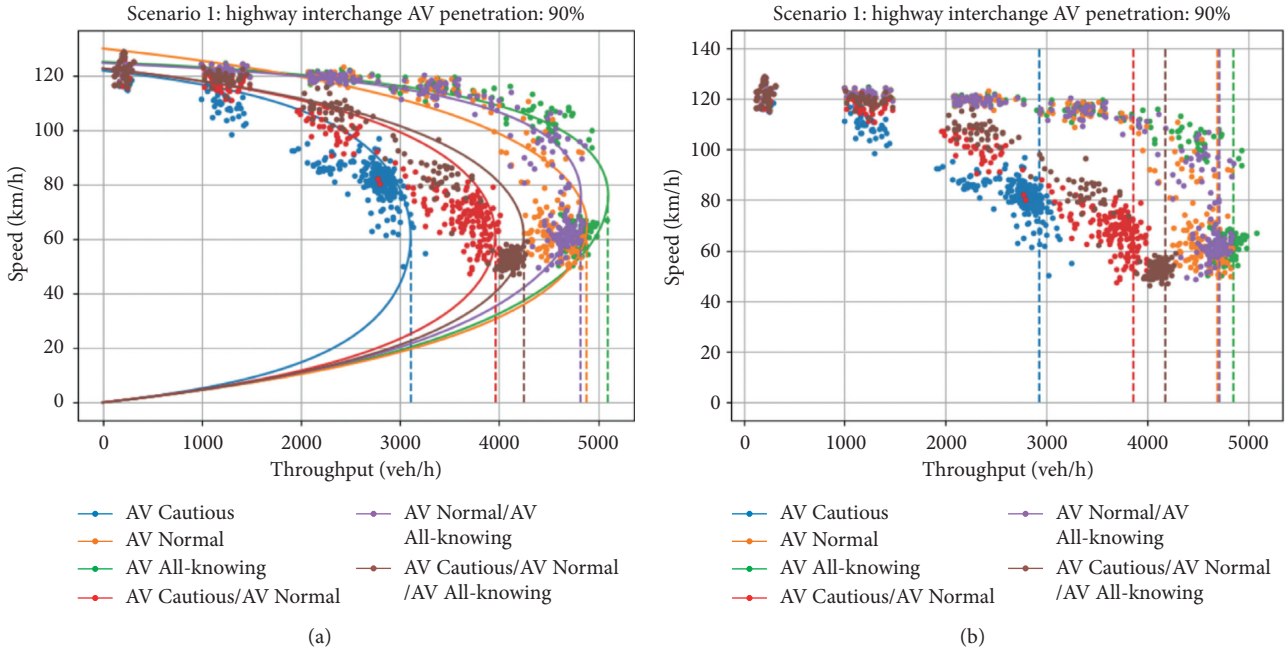



Figure 5: Example of van Aerde diagrams (a) and the 95% quantile (b) for the flow-speed relationships derived from simulations of a two-lane motorway with a one-lane off-ramp and on-ramp for an AV penetration rate of 90% of either the Cautious, Normal, All-knowing, and combined logics. For cases with mixes of driving logics, the shares of the AV logics are equal, e.g., 50/50 and 33/33/33.

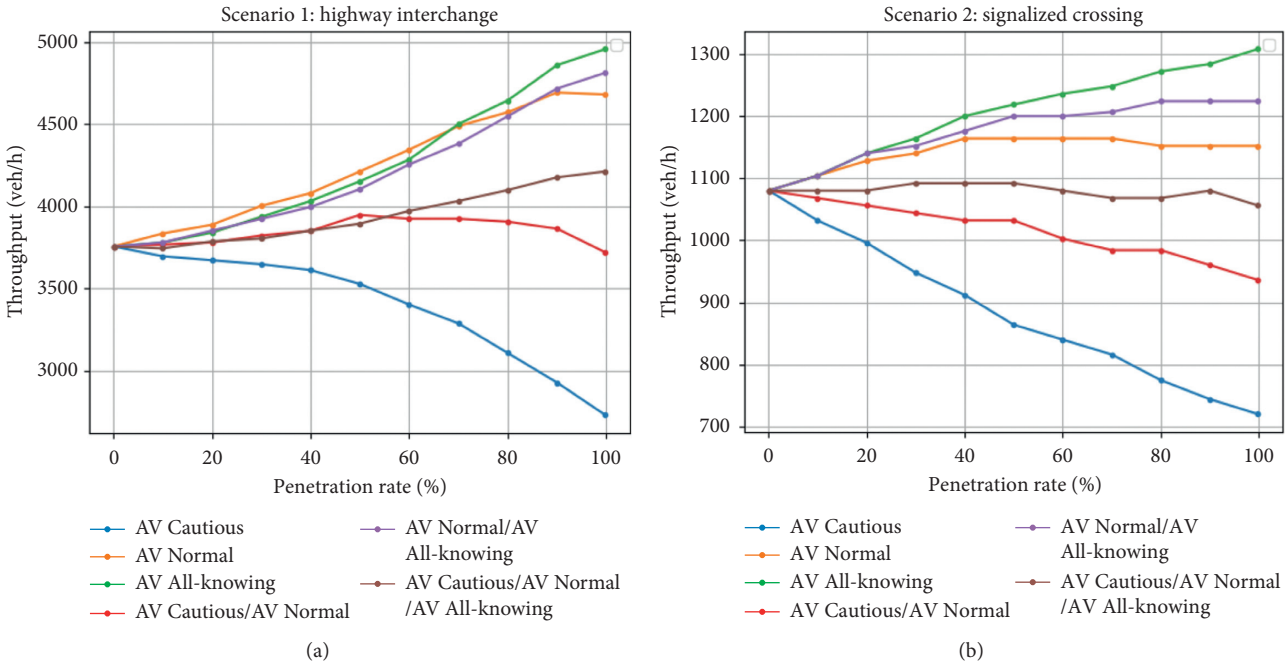



Figure 6: Continued.

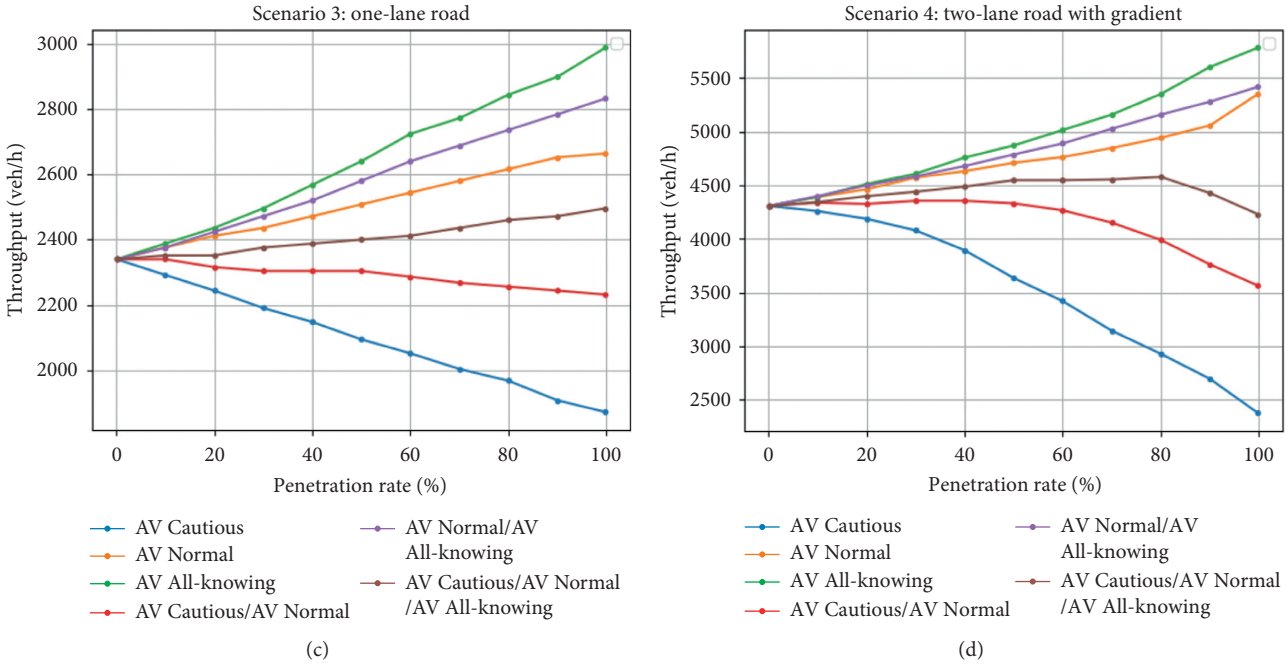



Figure 6: Results in maximum throughput depending on the AV penetration rate and driving logic mix for four different networks: gradient, simple link, on-off ramp and signal, and the AV driving logics: Cautious, Normal, and All-knowing. In the cases with a mix of driving logics, the shares of the AV logics are equal, e.g. 50/50, and 33/33/33.

the simulations with automated vehicles, e.g., for the one-lane road, the bottleneck is already at the inflow of the simulation network and traffic does not break down. Therefore, we derived the capacity not from van Aerde (Figure 5(a)) but from the 95% quantile of all demands (Figure 5(b)). The capacity derived from van Aerde is similar to the 95% quantile of all demands for the scenarios where we observed breakdowns as seen in Figure 5.

Figure 5 shows flow-speed for one of the networks used for numerical experiments (two-lane motorway with a one-lane off- and on-ramp) and one selected AV penetration rate (90%). There is a significant difference between Cautious logic and the other two (Normal and All-knowing). The Cautious driving logic shows lower maximum throughput and lower average speed at the same volume in comparison with the other two behaviour logics. That is because the Cautions driving logic follows the brick wall stop distance requirement for following and also for lane-changing. The All-knowing driving behaviour with settings corresponding to higher aggressiveness does not lead to high improvement because of causing disturbances in mixed flow with lot of lane change maneuverers. Higher gains with All-knowing behaviour logic would be theoretically possible with higher level of cooperation, which was not implemented in the model.

Figure 6 shows the results in maximum throughput depending on the penetration rate for the four different networks and for the Cautious, Normal, and All-knowing driving logics and for equal combinations (mixes) of these driving logics. Because of a larger headway compared to conventional vehicles, the throughput of the Cautious driver logic decreases with increasing penetration rate and on contrary because of the shorter headway, the All-knowing driver logic increases the throughput with increasing penetration rate. The Normal driving logic, where the headway is similar to conventional human-driven vehicle behaviour, increases the throughput with increasing penetration rate. This is because of a more consistent behaviour of AVs compared to conventional human-driven vehicles with less differences in individual behaviour. All-knowing behaviour logic brings an even higher increase in throughput in tested networks, except the highway scenario (motorway with off- and on-ramp), where the Normal and All-knowing driving logic and their combination lead to similar values, especially with penetration rates under 80%. As mentioned before, the number of lane-changing maneuverers and resulting disturbances in the traffic flow are the limiting factors. Further experiments with more varied mixes of conventional and automated vehicles with different driving logics are planned for the end phase of the CoEXist project.

## 6. Conclusions and Further Development Needs

There exist several uncertainties related to vehicle fleet composition during the transition period from 0 to 100% automated vehicles, e.g., regarding what automation functions are technically feasible, allowed, and actually activated by the users, for different road environments and at different stages of the introduction of AVs. In addition, we can also expect all these factors to be highly heterogeneous over the vehicle population, due to various functions being available for different brands and price levels, for the different times of production of the vehicles in the fleet. Hence, traffic simulation investigations of the transition period need to

consider that different types of automated vehicles with different capabilities and behaviour will exist and maybe coexist. Some possible heterogeneities, such as the possible heterogeneity in acceleration behaviour of AVs, have been simplified away in the presented model, which may impact the particular simulation results through effects on lane-changing behaviour. However, the proposed framework for handling uncertainties can in principle be extended to incorporate any and all uncertainties in the behaviour of the simulated road users, including any uncertainties in the heterogeneity of the acceleration behaviour, if this is deemed important for the considered application.

In this article, we present four conceptual descriptions of four different types of driving behaviour of automated vehicles and present how these driving logics can be implemented in a commonly used traffic simulation program. The idea behind the conceptual driving logic approach is that the information available so far on AV behaviour comes from tests with today's partly automated vehicles, mainly from test tracks. How future generations of AVs will behave will be unknown for some time. Thus, investigations of future traffic situations with future versions of AVs and with higher penetration rates of AVs require consistent assumptions on how the behaviour of AVs will evolve. The development and implementation of the four driving logics is one way of dealing with this issue. Furthermore, we believe that it is from a traffic performance point of view more important to capture the main characteristics of different types of automated vehicles instead of modelling all different types and combination of automation functions in detail, which are not available yet in real traffic.

To regard all uncertainties related to the introduction of automated vehicles as independent in a traffic simulation investigation would be infeasible in practice due to the curse of dimensionality; the number of simulation experiments required would become too large. We therefore suggest a simplified treatment of the uncertainties related to the vehicle fleet evolution by assuming that the penetration rate of AVs and the availability of advanced automation functions and driving logics covary and that they become available first for highly separated environments like motorways and later for more complex environments, such as urban streets. These assumptions allow us to constrain the space of possibilities in need of exploration to the vicinity of what we believe to be the most likely development. Furthermore, we suggest to divide the transition period into three stages: introductory, established, and prevalent. The stages are not defined in terms of specific number of years in the future, but rather by the level of automation in a specific case study. The reason is that depending on many factors, including technological development and adoption rates, and changes to the legal framework, the stages may have vastly differing durations. Hence, defining the stages in terms of time is not only highly speculative but also problematic since the durations of the stages might vary.

The developed approach for simulation of different driving logics taking into account their expected ODD and penetration rates for different stages of the transition period will in further work be applied to several different case studies in order to investigate the impact of the introduction of automated vehicles in, for example, signalized intersections, motorways, urban arterials, roundabouts, and shared spaces.

An important simplification in the approach presented here is that we do not consider the handover between automated and manual driving, even though we implicitly assume that such handovers are taking place. Hence, we assume that a specific simulated vehicle operates as either automated or human-driven. However, a vehicle might change driving logic when going from one type of road to another, and only then. Furthermore, this work does not consider potential behavioural adaptation of human drivers due to the introduction of automated vehicles, but this is an important uncertainty that needs to be addressed in future research. Since data availability on potential behavioural adaptation is limited, our suggestion is to conduct sensitivity analysis given different assumptions (e.g., regarding the adaptation of desired speed, following time gap, and overtaking willingness) [49].

## Data Availability

The data from the test-track tests with automated vehicles with TNO's driving logic (Section 4) is proprietary data of Siemens/TASS. The data from the numerical experiment (Section 5) can be reproduced using PTV Vissim (version 11 and higher) software with the defined driving logics (Section 4.2). Simulation results are available on request from Peter Sukennik or Jochen Lohmiller.

## Conflicts of Interest

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

## Acknowledgments

This research was part of the CoEXist project which was supported by funding from the European Union H2020-ART-2016-2017 under grant agreement no. 723201. Furthermore, the authors would like to thank all CoEXist partners and colleagues for their valuable input and contribution. Special thanks are due to Jaume Barcelo, Charlotte Fléchon, and Jörg Sonnleitner for valuable comments during the writing of the article.

## References

[1] S. C. Calvert, W. J. Schakel, and J. W. C. van Lint, "Will automated vehicles negatively impact traffic flow?" *Journal of Advanced Transportation*, vol. 2017, Article ID 3082781, 17 pages, 2017.

[2] E. Aria, J. Olstam, and C. Schwietering, "Investigation of automated vehicle effects on driver's behavior and traffic performance," *Transportation Research Procedia*, vol. 15, pp. 761–770, 2016.

[3] Atkins, *Research on the Impacts of Connected and Autonomous Vehicles (CAVs) on Traffic Flow-Stage 2: Traffic*

*Modelling and Analysis*, Atkins-Department for Transport, London, UK, 2016.

[4] J. Bierstedt, A. Gooze, C. Gray, J. Peterman, L. Raykin, and J. Walters, *Effects of Next-Generation Vehicles on Travel Demand and Highway Capacity*, F. P. Think, Ed., University of Oregon, Portland, OR, USA, 2014.

[5] F. Bohm and K. Häger, "Introduction of autonomous vehicles in the swedish traffic system: effects and changes due to the new self-driving car technology," Master thesis, Uppsala Universitet, Uppsala, Sweden, 2015.

[6] A. Tapani, F. Pereyron, and M. Viera Turnell, "Integrated vehicle and traffic simulation for emissions and energy efficiency estimation," in *Proceedings of the 17th International Conference of Hong Kong Society for Transportation Studies*, Hong Kong, China, 2012.

[7] A. Kesting, M. Treiber, M. Schönhof, F. Kranke, and D. Helbing, "Jam-avoiding adaptive cruise control (ACC) and its impact on traffic dynamics," *Traffic and Granular Flow'05*, Springer, Berlin, Germany, pp. 633–643, 2007.

[8] M. M. Minderhoud and P. H. L. Bovy, "Impact of intelligent cruise control on motorway capacity," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1679, no. 1, pp. 1–9, 1999.

[9] N. Motamedidehkordi, M. Margreiter, and T. Benz, "Effects of connected highly automated vehicles on the propagation of congested patterns on freeways," in *Proceedings of the Research Board 95th Annual Meeting*, Washington DC, USA, January 2016.

[10] A. Tapani, "Traffic simulation modelling of driver assistance systems," *Advances in Transportation Studies*, vol. 23, pp. 41–50, 2011.

[11] J. Vander Werf, S. E. Shladover, M. A. Miller, and N. Kourjanskaia, "Effects of adaptive cruise control systems on highway traffic flow capacity," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1800, no. 1, pp. 78–84, 2002.

[12] M. Makridis, K. Mattas, C. Mogno, B. Ciuffo, and G. Fontaras, "The impact of automation and connectivity on traffic flow and CO2 emissions. A detailed microsimulation study," *Atmospheric Environment*, vol. 226, Article ID 117399, 2020.

[13] A. D. Tibljaš, T. Giuffrè, S. Surdonja, and S. Trubia, "Introduction of autonomous vehicles: roundabouts design and safety performance evaluation," *Sustainability*, vol. 10, no. 4, 2018.

[14] P. Fernandes and U. Nunes, "Platooning of autonomous vehicles with intervehicle communications in SUMO traffic simulator," in *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, pp. 1313–1318, Madeira Island, Portugal, September 2010.

[15] M. Bahram, Z. Ghandeharioun, P. Zahn, M. Baur, W. Huber, and F. Busch, "Microscopic traffic simulation based evaluation of highly automated driving on highways," in *Proceedings of the IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, Quindao, China, October 2014.

[16] 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.

[17] J. Olstam and R. Elyasi-Pour, "Combining traffic and vehicle simulation for enhanced evaluations of powertrain related ADAS for trucks," in *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, pp. 851–856, The Hague, Netherlands, October 2013.

[18] J. L. F. Pereira and R. J. F. Rossetti, "An integrated architecture for autonomous vehicles simulation," in *Proceedings of the Trento 2012 27th Annual ACM Symposium on Applied Computing*, pp. 286–292, Trento, Italy, March 2012.

[19] M. C. Rahal, S. Pechberti, B. Heijke, and P. Sukennik, "D2.2: technical report on connecting CAV control logic and CAV simulator," 2017, https://www.h2020-coexist.eu/resources/.

[20] A. Bose and P. A. Ioannou, "Analysis of traffic flow with mixed manual and semiautomated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 4, no. 4, pp. 173–188, 2003.

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

[22] L. Xiao, M. Wang, and B. van Arem, "Realistic car-following models for microscopic simulation of adaptive and cooperative adaptive cruise control vehicles," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2623, no. 1, pp. 1–9, 2017.

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

[24] L. C. Davis, "Effect of adaptive cruise control systems on traffic flow," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 6 2, pp. 066111–066118, 2004.

[25] L. C. Davis, "Effect of adaptive cruise control systems on mixed traffic flow near an on-ramp," *Physica A: Statistical Mechanics and Its Applications*, vol. 379, no. 1, pp. 274–290, 2007.

[26] 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 and Engineering Sciences*, vol. 368, no. 1928, pp. 4585–4605, 2010.

[27] A. Kesting, M. Treiber, M. Schönhof, and D. Helbing, "Extending adaptive cruise control to adaptive driving strategies," *Transportation Research Record*, vol. 2000, no. 1, pp. 16–24, 2007.

[28] S. Kikuchi, N. Uno, and M. Tanaka, "Impacts of shorter perception-reaction time of adapted cruise controlled vehicles on traffic flow and safety," *Journal of Transportation Engineering*, vol. 129, no. 2, pp. 146–154, 2003.

[29] G. Klunder, M. Li, and M. Minderhoud, "Traffic flow impacts of adaptive cruise control deactivation and (Re)activation with cooperative driver behavior," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2129, no. 1, pp. 145–151, 2009.

[30] J. Ludmann, D. Neunzig, and M. Weilkes, "Traffic simulation with consideration of driver models, theory and examples," *Vehicle System Dynamics*, vol. 27, no. 5-6, pp. 491–516, 1997.

[31] R. Rajamani, D. Levinson, P. Michalopoulos, J. Wang, K. Santhanakrishnan, and X. Zou, "Adaptive cruise control system and its impact on traffic flow," Center for Transportation Studies, University of Minnesota, Minneapolis, MN, USA, CTS 05-01, 2005.

[32] M. Van Aerde, "A single regime speed-flow-density relationship for freeways and arterials," in *Proceedings of the 74th Annual Meeting of the Transportation Research Board*, Washington. DC, USA, 1995.

[33] 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.

[34] I.A. Ntousakis, I. K. Nikolos, and M. Papageorgiou, "On microscopic modelling of adaptive cruise control systems," *Transportation Research Procedia*, vol. 6, pp. 111–127, 2015.

[35] H. S. Mahmassani, A. Elfar, S. Shladover, and Z. Huang, *Development of an Analysis/Modeling/Simulation (AMS) Framework for Connected and Automated Vehicle Systems-Task 9: Draft V2I AMS Capability Development Case Study Plan* FHWA, Washington, DC, USA, 2018.

[36] 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.

[37] K. Mattas, M. Makridis, P. Hallac et al., "Simulating deployment of connectivity and automation on the Antwerp ring road," *IET Intelligent Transport Systems*, vol. 12, no. 9, pp. 1036–1044, 2018.

[38] J. W. C. van Lint and S. C. Calvert, "A generic multi-level framework for microscopic traffic simulation-Theory and an example case in modelling driver distraction," *Transportation Research Part B: Methodological*, vol. 117, pp. 63–86, 2018.

[39] S. C. Calvert and B. van Arem, "A generic multi-level framework for microscopic traffic simulation with automated vehicles in mixed traffic," *Transportation Research Part C: Emerging Technologies*, vol. 110, pp. 291–311, 2020.

[40] 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.

[41] 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.

[42] A. Alessandrini, "Existing legal barriers and the proposedCityMobil2 approach," *in Implementing Automated Road Transport Systems in Urban Settings,* A. Alessandrini, Ed.,pp. 273–278, Elsevier, Amsterdam, Netherlands, 2018.

[43] A. Alessandrini and D. Stam, "ARTS—automated roadtransport systems," *in Implementing Automated Road Transport Systems in Urban Settings,* A. Alessandrini, Ed.,pp. 9–16, Elsevier, Amsterdam, Netherlands, 2018.

[44] F. Cignini, C. Holguin, L. Domenichini, D. Stam, and A. Alessandrini, "Integrating ARTS in existing urban infrastructures:the general principles," *in Implementing Automated Road Transport Systems in Urban Settings,* A. Alessandrini, Ed., pp. 43–60, Elsevier, Amsterdam, Netherlands, 2018.

[45] A. Tripodi, F. Cignini, L. Domenichini, and A. Alessandrini, "Integrating ARTS on signalised and nonsignalised intersectionsfor safety maximisation and comparison with conventionalcar safety assessment," *in Implementing Automated Road Transport Systems in Urban Settings,* A. Alessandrini, Ed., pp. 61–80, Elsevier, Amsterdam, Netherlands, 2018.

[46] S. Lee, E. Jeong, M. Oh, and C. Oh, "Driving aggressiveness management policy to enhance the performance of mixed traffic conditions in automated driving environments," *Transportation Research Part A: Policy and Practice*, vol. 121, pp. 136–146, 2019.

[47] P. Sukennik, V. Zeidler, and J. van Etten, "Deliverable 2.6: technical report on data collection and validation," 2017, https://www.h2020-coexist.eu/resources/.

[48] PTV, "PTV VISSIM 11 User manual," A. G. Karlsruhe, 2018.

[49] SAE International, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," SAE International, Warrendale, PA, USA, J3016_201806, 2018.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8830731, 17 pages
https://doi.org/10.1155/2020/8830731

*Research Article*

# Data-Driven Analysis of the Chaotic Characteristics of Air Traffic Flow

**Zhaoyue Zhang**,[1] **An Zhang**,[1] **Cong Sun**,[1] **Shuaida Xiang**,[2] **and Shanmei Li**[2]

[1]*School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China*
[2]*College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China*

Correspondence should be addressed to Zhaoyue Zhang; zy_zhang@cauc.edu.cn

Received 14 June 2020; Revised 31 July 2020; Accepted 3 September 2020; Published 18 September 2020

Academic Editor: Chunjiao Dong



Understanding the chaos of air traffic flow is significant to the achievement of advanced air traffic management, and trajectory data are the basic material for studying the chaotic characteristics. However, at present, there are two main obstacles to this task, namely, large amounts of noise in the measured data and the tedium of existing data processing methods. This paper improves the incorrect trajectory processing method based on ADS-B trajectory data and proposes a method by which to quickly extract the traffic flow through a certain waypoint. Currently, the commonly used theoretical analysis tools for nonlinear complex systems include the classical nonlinear dynamics analysis method and the newly developed complex network-based analysis method. The latter is currently in an exploratory stage because it has just been introduced into the study of air traffic flow. From these two perspectives, the chaotic characteristics of air traffic flow are studied in the present work. From the perspective of nonlinear dynamics, the improved C-C method is used to calculate the reliability parameters, namely, the time delay $\tau$ and embedding dimension $m$, of phase-space reconstruction, and the maximum Lyapunov index is calculated by using the small data volume method to prove the existence of chaos in the system. From the perspective of complex networks, the construction of a visibility graph and horizontal visibility graph is used to prove the existence of chaos in the system, and the goodness-of-fit parameters of the degree distributions of two fitting methods under different time scales are evaluated, which provides support for the air traffic flow theory.

## 1. Introduction

With the growth of China's national economy, its civil aviation industry has also been greatly improved. Air traffic transportation is an important component of the modern transportation system, and the study of air traffic flow characteristics has gradually become a crucial task. Air traffic flow refers to the number of aircraft passing through a certain point or a certain route in a unit of time. China's air traffic flow is characterized by rapid growth and very uneven distribution. The dynamics of air traffic flow depend on the number and the length of air routes, the number of airports, the number of takeoffs and landings, etc. Air traffic flow is mainly concentrated in a few airports located in cities that are political, economic, and tourism centers, especially in the more developed areas in the eastern part of China. With the rapid increase of domestic flights, the problem of flight delay has become increasingly more obvious, and the study of the chaotic characteristics of air traffic flow is beneficial to its management and control.

In recent years, many scholars have studied the nonlinear mechanism of traffic flow, especially by applying chaos theory to understand the dynamics of transportation systems. Chaos refers to the seemingly random and irregular phenomenon in deterministic nonlinear systems, but this kind of quasi-random phenomenon contains some laws and order. Chaotic systems are very sensitive to initial values, and a small change in the initial value will lead to a great difference in the long-term evolution results of the system. There exist various models for traffic flow prediction,

including the chaos theory, fractal theory, wavelet analysis, and catastrophe theory models. The premise of the application of the chaos theory model is that the system has chaotic characteristics. Thus, when short-term traffic flow is predicted based on the chaos theory model, it must be determined in advance whether there are chaotic characteristics in the traffic flow.

The research on the chaotic dynamic characteristics of ground traffic flow has received substantial attention within the last decade. In 1984, Disbro and Frame first introduced chaos theory into the traffic system [1]. Low and Addision investigated the concept of chaotic behavior in a deterministic car-following model [2]. Tang et al. improved the chaos forecasting method to be effective in forecasting conflicting traffic flows [3]. Based on the maximum Lyapunov index, Liu and Zhang determined the chaotic characteristics of traffic flow and realized the prediction of short-term traffic flow [4].

With the development of ground traffic flow theory, scholars began to explore the nonlinear characteristics of air traffic systems. The time series of air traffic flow based on measured data is an effective method by which its nonlinear characteristics can be studied. Frank et al. found that the Lyapunov exponent calculation provides a clear indication of chaos [5]. Time series are closely related to time scales, and there are great differences between time series extracted from different time scales. If the time scale is too small, it is difficult to determine the fluctuation of traffic flow; in contrast, if the time scale is too large, the macrocharacteristics of traffic flow will be more reflected, and it is difficult to capture the detailed features. These are both mentioned in Zheng's research on the macro time series of air traffic flow under different time scales [6] and Wang's empirical analysis of time-dependent air traffic flow based on the Hurst index [7]. At present, the tool used most often for the analysis of the chaotic dynamic characteristics of air traffic flow is the classical nonlinear dynamic analysis method; however, there also exist new analysis methods based on complex networks.

*1.1. Nonlinear Dynamics Analysis Method.* Li et al. proposed an improved maximum Lyapunov exponential algorithm based on the small data volume method and wavelet noise reduction theory, identified the chaotic characteristics in flight conflict time series, and proved the feasibility of the application of chaos theory to flight conflict prediction [8]. Cong and Hu used time-series data of traffic flow in the airspace sector to study the chaotic characteristics of the air traffic system; they solved the correlation dimension and the maximum Lyapunov index of the traffic flow time series via the G-P algorithm and the small data volume method, thereby proving the existence of chaos and fractal characteristics in the air traffic flow time series [9]. Zheng used real trajectory data to study the chaotic characteristics of air traffic flow from both the macro- and microperspectives via the maximum Lyapunov index and recursive graphs [6]. Yang proposed a dynamic identification method for the chaotic characteristics of flow time series to solve the timeliness problem of judging the chaotic characteristics of flow time series; he proved that, with the increase of data, flow time series gradually possess chaotic characteristics [10].

*1.2. Complex Network-Based Analysis Method.* Complex networks are a new network research theory characterized by partial or total self-organization, self-similarity, and attractors and by being small-world and scale-free. In recent years, with the rise of network science, the study of time series from the perspective of complex networks has attracted widespread attention. Some scholars began to use complex networks to analyze nonlinear time series. The analysis method based on complex networks is often used in the research on the fluctuation of gold prices and vegetable species prices. Researchers have also explored the degree distribution characteristics of networks constructed from air traffic flow time series. Yang and Yang proposed a reliable procedure for the construction of complex networks from the correlation matrix of a time series [11]; they found that the degree distribution of the original series can be well fitted with a power law, while that of the return series can be well fitted with a Gaussian function. Gao and Jin proposed a reliable method for the construction of complex networks from a time series with each vector point of the reconstructed phase space represented by a single node and edge determined by the phase space distance [12]. Donner et al. investigated in detail the corresponding potentials and limitations of networks based on recurrences in the phase space [13]. Pan et al. introduced a novel method for the condition diagnosis of complex systems in the chemical process industry with complex network-based time series analysis [14]. Scholars also gradually introduced complex networks into the study of traffic flow. For example, Tang et al. converted a traffic flow time series into a complex network and then analyzed the characteristics of traffic flow states from the perspective of the complex network [15]. In 2018, Wang and Zhu first analyzed air traffic flow time series based on complex network theory [16]; he examined the mechanical characteristics of air traffic flow from a new perspective, laying a foundation for the future application of other research results of complex network theory in the air traffic field.

This remainder of this paper is structured as follows. Section 2 explains the processing of trajectory data, proposes four types of incorrect trajectory data, and introduces the process of automatic dependent surveillance-broadcast (ADS-B) trajectory data collection. Moreover, it presents an improvement of the efficient identification and processing method of incorrect trajectory data, and the results of clustering and nonclustering when extracting the air traffic flow passing through a waypoint are compared. Section 3 extracts time series of different scales. Section 4 analyzes the chaotic characteristics of air traffic flow based on the classical nonlinear dynamic analysis method, while Section 5 analyzes the chaotic characteristics of air traffic flow based on complex networks. Finally, the conclusions of this work are drawn in Section 6.

## 2. Processing of ADS-B Trajectory Data

In this paper, an analysis method based on the measured data of the operating environment is adopted, and the data source is ADS-B trajectory data. ADS-B mainly carries out air-to-air and air-to-ground surveillance. In general, only airborne electronic equipment is needed; the GPS (Global Positioning System) receiver, data link transceiver, and antenna, and CDTI (cockpit display of traffic information) can complete the relevant functions without any ground support equipment, and an aircraft equipped with ADS-B can broadcast its precise location and other data (such as its speed and height and whether the aircraft is turning, climbing, falling, etc.) through the data chain. The ADS-B receiver, in combination with air traffic control systems and airborne ADS-B from other aircraft, provides accurate, real-time conflict information both in the air and on the ground. ADS-B is a new technology that redefines the three key elements of modern air traffic control, namely, communication, navigation, and surveillance.

The ADS-B provides detailed data on individual flights, such as the aircraft serial number, registration number, ICAO, flight call sign, latitude, longitude, altitude, speed, heading, vertical speed, receiver IP address, and flight time. Because the research object of this paper is air traffic flow, it is necessary to extract the time series of traffic flow from the trajectory data.

In this work, it is found that there are some incorrect trajectory data in the dataset. Incorrect trajectories originate from system errors in recording the trajectory data, which leads to the recording of noise. Erroneous trajectory data are different from abnormal trajectory data; anomalous trajectory data refer to the trajectory deviating from the prevailing traffic flow, not an incorrect trajectory. The existence of an abnormal trajectory reflects the randomness of air traffic operation and the influence of control measures on air traffic flow. If the trajectory data are not processed, it will result in difficulties in the subsequent analysis, as presented in Figure 1.

Therefore, before traffic flow identification, it is necessary to identify and process the incorrect trajectory data. In this section, based on the types and characteristics of the incorrect trajectory data found in the research process, related identification and processing methods are studied. There were mainly four types of incorrect trajectory data found in the research process of the present work, namely, scattered, multiconnected, misplaced, and defective trajectories. The scatter diagram and trajectory diagram of incorrect trajectories are presented in Figure 2.

As shown in Figure 2(a), the scattered trajectory is composed of a series of scattered points that may be far away from each other, and the connecting lines may misrepresent the movement state of the aircraft. There are few track points, which are far less than the average of the number of normal track points. Such trajectory data can be either point or noise data consisting of several straight lines.

As shown in Figure 2(b), the multiconnected trajectory is the erroneous connection of two trajectories. The original data may be confused with the same track during transcription and encoding. Such trajectory data will create an extra line between two unrelated track points. Even if the points are filled by linear interpolation, the trajectory diagram will not change.

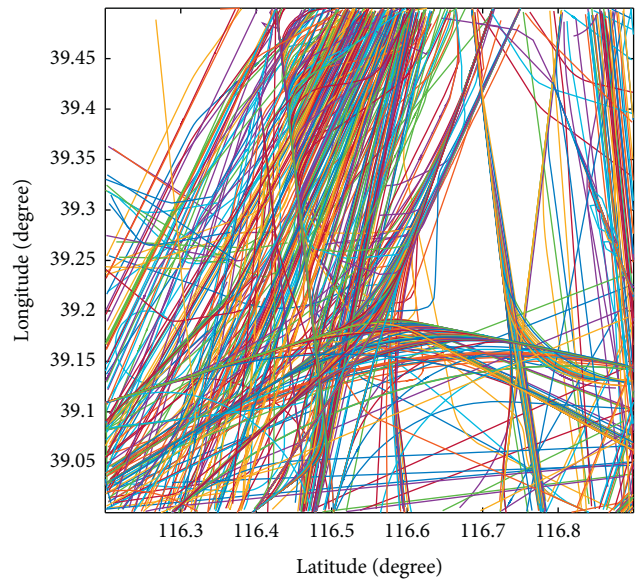



Figure 1: Original trajectory data.

As shown in Figure 2(c), the misplaced trajectory is represented as an acute turning point, and the aircraft will not turn at an acute angle when it turns. When the three adjacent track points are misaligned, it will appear as an acute angle, which is represented as an obvious sharp bulge in the trajectory diagram.

As shown in Figure 2(d), the defective trajectory shows that the distance between two adjacent track points is large, which is caused by data loss.

Incorrect trajectory data cannot represent the correct trajectory of the aircraft; this will increase the number of trajectories and affect the data processing efficiency, which is not conducive to the identification of prevailing traffic flow. According to the characteristics of the incorrect trajectory data, four processing methods for this data are proposed in this paper. Because each piece of data is information acquired by radar, it has the significance of existence. To retain the original trajectory characteristics, it is necessary to deal with the incorrect trajectory data according to the principle of "more dismantling and less deletion;" the scattered trajectory data must be deleted, the multiconnected trajectory data must be split, the misplaced trajectory data must be smoothed, and the defective trajectory data must be supplemented.

The process diagram is presented in Figure 3.

Figure 4 presents a comparison of the effects before and after the processing of incorrect trajectory data from the entire day of December 14, 2019, over LASAN (a waypoint) in Shanghai. By comparison, it can be found that there are many lines on the left side of the figure that do not pass through the position of the prevailing traffic flow. These lines are the lines that connect two unrelated trajectories in multiple pieces of data. When multiple pieces of data are

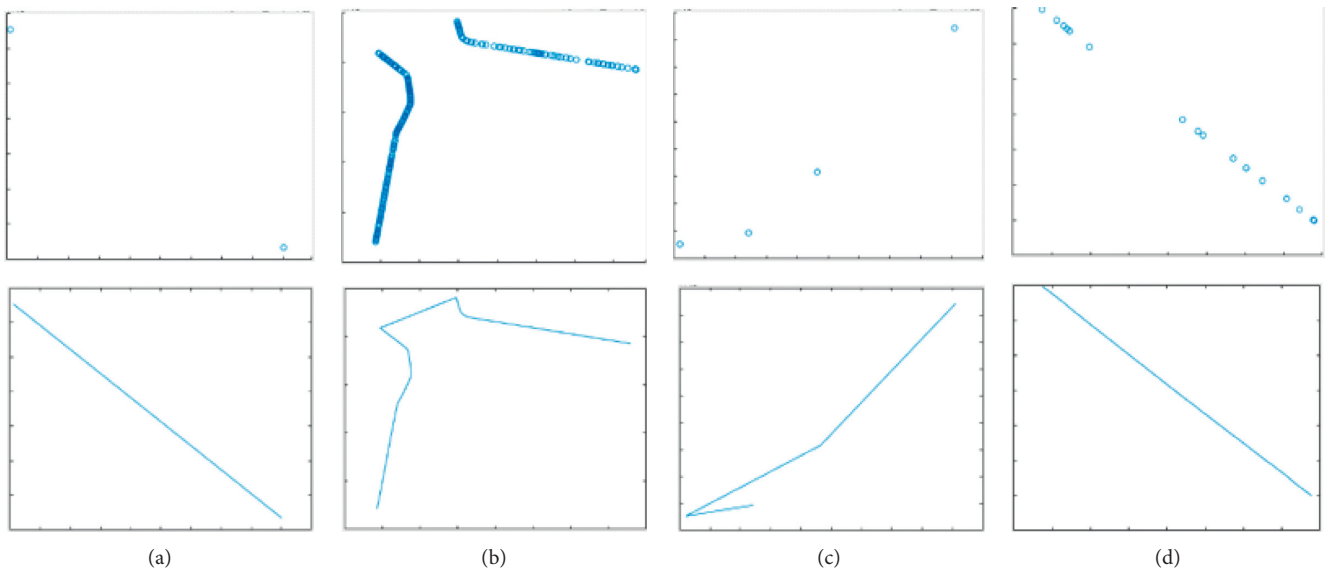



Figure 2: Scatter diagram and trajectory diagram corresponding to wrong trajectory. (a) Scattered trajectory. (b) Multiconnected trajectory. (c) Misplaced trajectory. (d) Defective trajectory.

split, the lines disappear. The comparison demonstrates that the traffic flow becomes clearer after the incorrect data are deleted, and the original features are still retained.

## 3. Extraction of Air Traffic Flow Time Series on Different Time Scales

Trajectory clustering is the main means by which traffic flow from chaotic trajectory data is identified; at present, the prevailing trajectory clustering method is adopted in most domestic and international research on air traffic flow identification. In this paper, the object of study for the traffic flow is on the air route, so the clustering of traffic flow in terminal areas need not be considered. Therefore, in this paper, a traffic flow identification method based on the air route network structure proposed by Zheng [6] is adopted; the air route network structure of air traffic flow is used for automatic identification, and the traffic flow is simultaneously determined via the waypoint sequence.

Taking the traffic flow over Laiyang City, Shandong Province, as an example, the trajectory data in the research area are processed and the traffic trajectory through P449 are extracted, as shown in Figure 5.

The complete clustering steps were found to be redundant for traffic flow extraction. For all the traffic flow data in Figure 5(a), if only extract the trajectory data whose closest distance to P449 is less than 10 km, it can be extracted as the traffic flow passing through the route point, as shown in Figure 6. The total traffic flow is 12% greater than the total traffic flow after clustering. Compared with the manual sorting of traffic flow sequences after pattern recognition, this not only realizes complete automation but is also more practical.

The air traffic flow time series $T_f = [t_{f1}, t_{f2}, \ldots, t_{fi}, \ldots, t_{fn'}]$ can be obtained by calculating the number of flights in a certain time scale, and the sequence length is $n'$. The traffic flow passing through P449 on December 14 is used as an example. Due to the low traffic flow during the period of 0:00 to 8:00 (PEK time), the traffic flow characteristics cannot be studied as a research sample. Therefore, the period of 8:00 to 24:00 is selected as the observation time, and the time series is constructed with four different time intervals of 2 min, 5 min, 10 min, and 15 min as the time scales, as shown in Figure 7.

By comparing the subgraphs in Figure 7, it can be determined that, although the time scales are different, similar peaks or troughs appear at the same time; however, the detailed characteristics of the time series are different. The reason for this is that larger time scales can smooth the data, but the resulting flow time series loses the internal details of the traffic flow. In this paper, a suitable time scale is selected to analyze the chaotic dynamics of air traffic flow.

## 4. Chaotic Characteristics of Air Traffic Flow Based on the Classical Nonlinear Analysis Method

*4.1. Judgment of Chaotic Characteristics.* In recent years, the chaotic time series analysis method has been widely used in many scientific studies and engineering fields. Phase-space reconstruction is the basis of chaotic time series analysis; according to this theory, the calculation of the time delay $\tau$ and embedding dimension $m$ [17] is the key problem of nonlinear time series phase-space reconstruction. For a given time series, such as the macroscopic traffic flow time series $T_f = [t_{f1}, t_{f2}, \ldots, t_{fi}, \ldots, t_{fn'}]$, it is assumed that the delay time is $\tau$ and the embedding dimension is $m$. The $m$-dimensional vector $X_n$ is the phase point produced by the phase space reconstruction as follows:

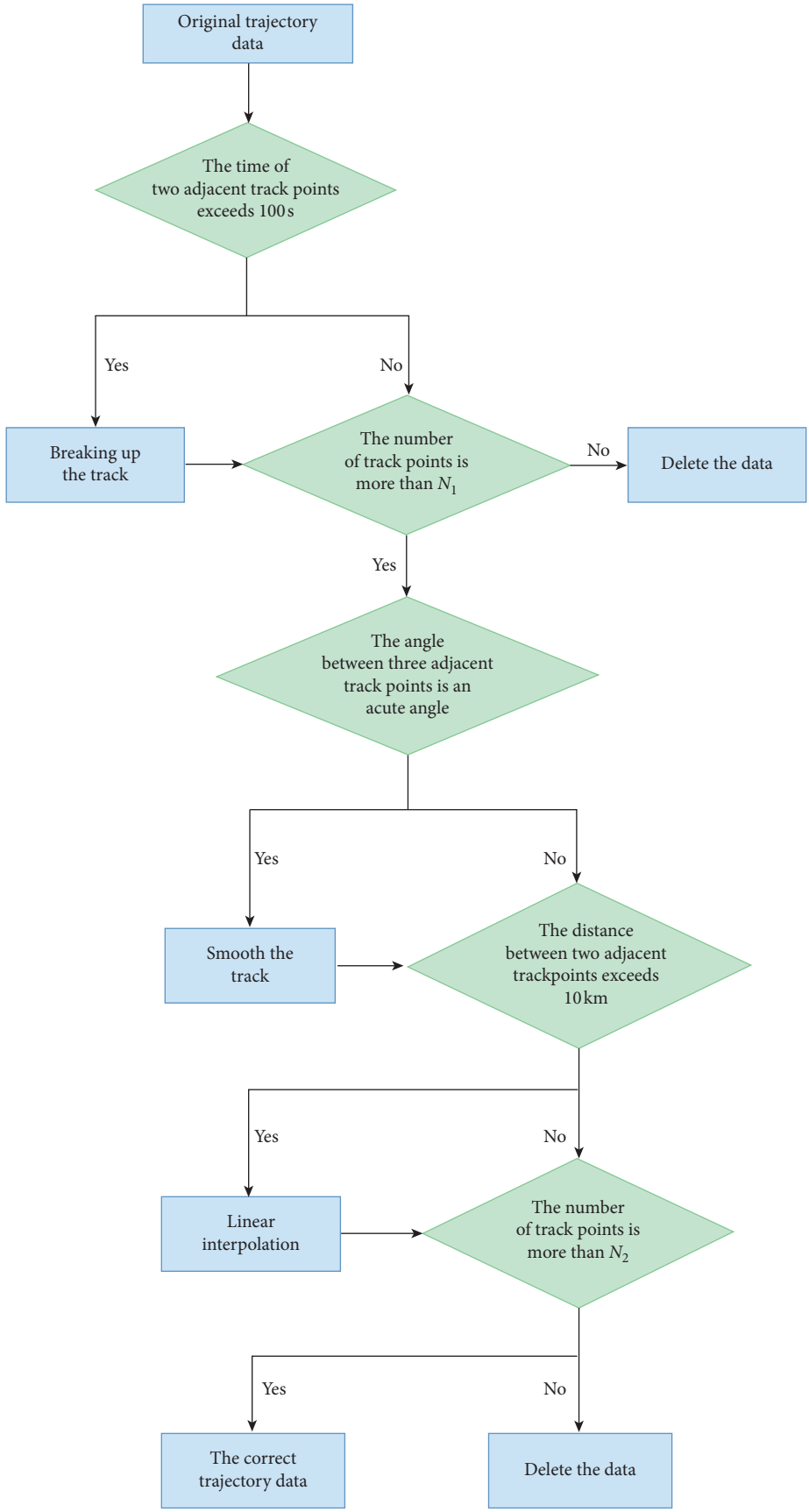



Figure 3: Wrong trajectory data processing algorithm flowchart.

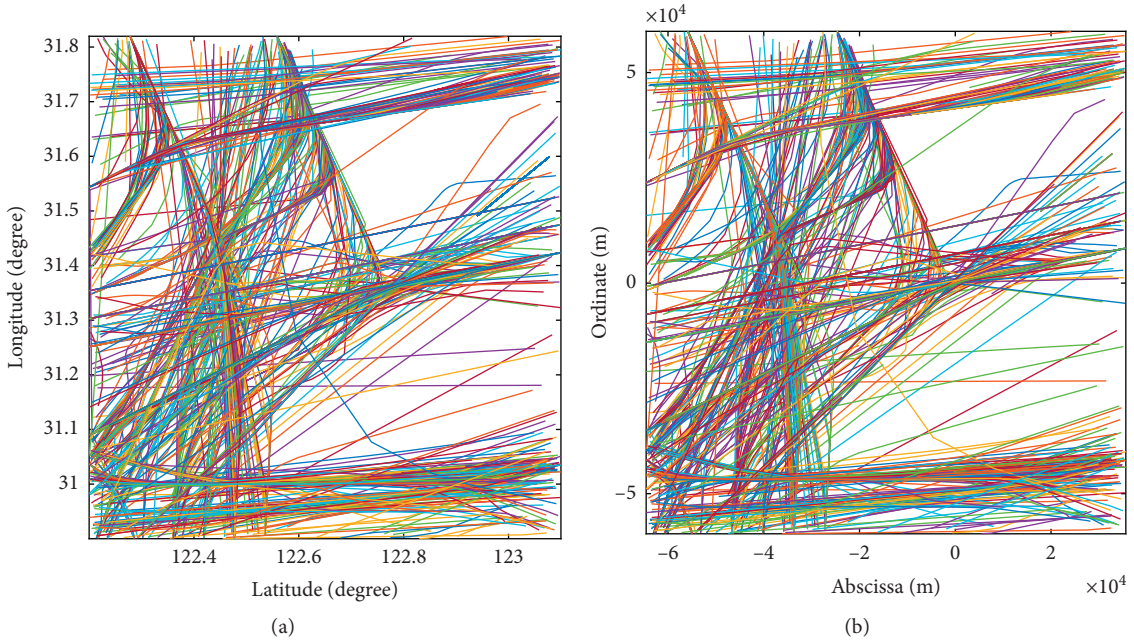



Figure 4: December 14, 2019, 122.2°–123.1° east longitude, 30.9°–31.82° north latitude trajectory data. (a) Before processing wrong. (b) After processing wrong.

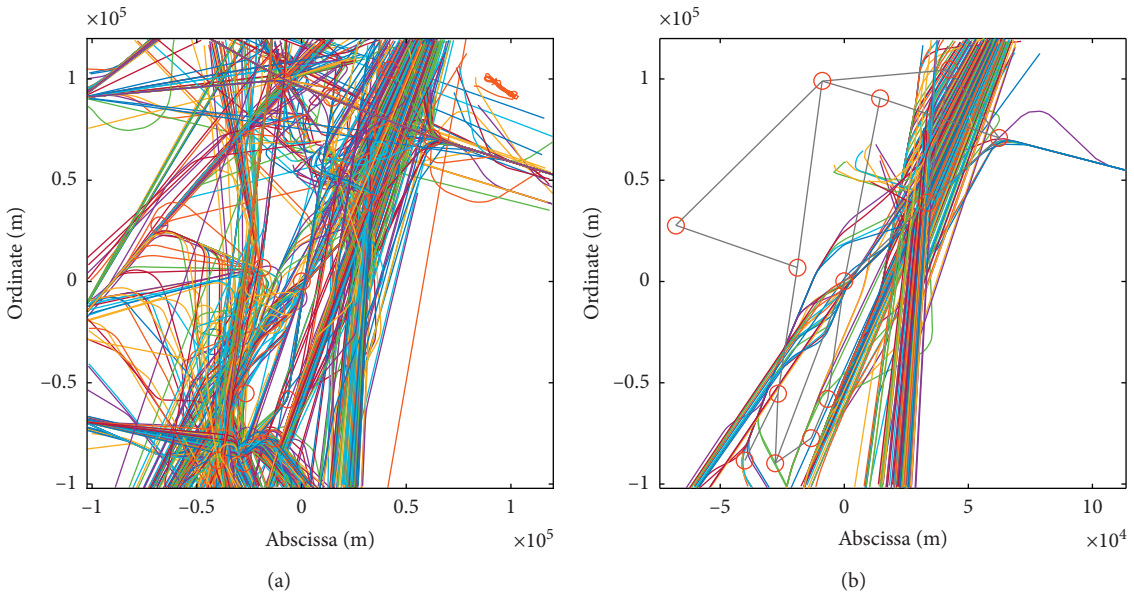



Figure 5: Traffic flow through P449 extracted on December 14, 2019. (a) All traffic flows over Laiyang on December 14, 2019. (b) Traffic flow through P449 extracted on December 14, 2019.

$$X_n = \left[t_{fn}, t_{f(n-t)}, t_{f(n-t)}, \ldots, t_{f[n-(m-1)t]}\right], \quad (1)$$

$$n \in [(m-1)\tau + 1, (m-1)\tau + 2, \ldots, N]. \quad (2)$$

Takens' theorem, which was put forward for infinitely long time series with the absence of noise, provided the foundation for the selection of the embedding dimension, and the time delay $\tau$ may take any value [18]. However, the system dynamics dimension $d$ is unknown, and actual time series have a finite length and noise interference; therefore, the selection of the appropriate time delay $\tau$ and embedding dimension $m$ are the key to calculation. There are currently two main views regarding the selection of these values. One view holds that the two are mutually independent, for instance, in the time delay autocorrelation method [19], the average mutual

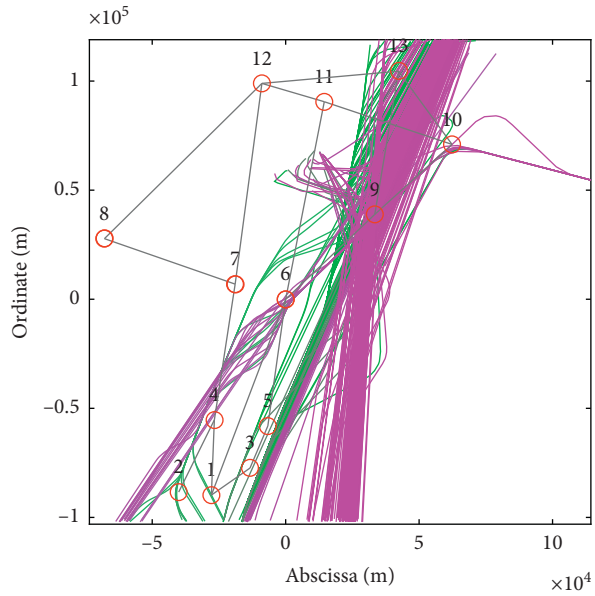



Figure 6: Traffic flow through P449 extracted on December 14, 2019.

information method [20], the embedded dimension G-P algorithm [21], the FNN method (false proximity method) [22], and the Cao method (the improvement of the false nearest proximity method) [23]. The other view holds that the two are related, for instance, in the embedded window method [24], C-C method [25], and improved C-C method [26].

In this paper, an improved C-C method is used to calculate these two parameters. The traditional methods have a certain advantage in the reconstructed phase space, while the improved method C-C method better conforms to the actual situation to achieve the estimates of the optimal time delay $\tau$ and the optimal delay time window $\tau_\omega$. The improved method uses the following two processes to modify the associated parameters.

(1) The time series standard deviation $\sigma$, which depends on the search radius $r$, is expanded to a certain extent according to the variation degree of the time series to reduce the interference caused by numerical oscillation. $\sigma = \text{std}(x)(1 + \text{cv}/3)$ is utilized, where cv represents the coefficient of variation for discrete sequences and $\text{cv} = \text{std}(x)/\text{mean}(x)$.

(2) The fixed value of $r$ is changed to a function related to the embedded dimension $m$: $r(m) = r\ln(m + 1)$. With the change of the refactoring dimension, the search scope is appropriately expanded to reduce interference, especially high-dimensional data oscillations.

The remaining statistics are calculated according to the rules defined by the traditional C-C method. When analyzing and calculating the results, the first local minimum peak $\tau_\omega$ of the cyclical nature of $S_{2_{\text{cor}}}(t)$ is searched as the optimal delay time window. For nonobvious cyclical qualitative results, the value of $S_{2_{\text{cor}}}(t)$ of the global minimum point is selected to determine the optimal delay time window $S_{2_{\text{cor}}}(t)$. Moreover, the value of $\Delta\overline{S}_2(t)$ of the first local minimum point is used as the optimal time delay $\tau$.

The Lyapunov index is an important index to measure the dynamic characteristics of a system. It represents the average exponential rate of convergence or divergence between adjacent orbits in the phase space. The basic characteristic of chaotic motion is that the motion state is highly sensitive to the initial conditions. The orbitals generated by two very close initial values are separated exponentially over time, and the Lyapunov index is a quantitative description of this phenomenon.

If the Lyapunov index is greater than 0, the system is in a chaotic state. This judgment method is simple and definite and has little error. There are many methods to calculate the maximum Lyapunov index, including the definition method, Wolf method, small data volume method, Jacobian method, p-norm algorithm, and singular value decomposition method. For the study of time series, the Wolf method and the small data volume method are more applicable. Both methods require the calculation of the average period $P$, which can be obtained by using the Fourier transform to draw the power spectrum curve. In the case of a limited data volume, the small data volume method can better meet the computing needs [27].

*4.2. Chaotic Characteristics at Different Time Scales.* Based on the previously mentioned methods, the time series $T_f$ of air traffic flow is reconstructed by the phase space $X_n = [t_{fn}, t_{f(n-\tau)}, t_{f(n-2\tau)}, \ldots, t_{f[n-(m-1)\tau]}]$. The maximum Lyapunov indices of the time sequence under different time scales presented in Section 3 are calculated based on the small data volume method, and the results are shown in Figure 8.

The maximum Lyapunov indices of the traffic flow time series in the sample at different time scales are all greater than zero ($\lambda_{\max} > 0$), indicating that there are chaotic characteristics in the air traffic flow. Via further comparison, it can be concluded that, with the increase of the time scale, the maximum Lyapunov index increases, indicating that the degree of chaos of the system increases. Consistent with previous research results [6], the length of the time scale has an important impact on the mining of traffic flow characteristics. The variation of the time scale will directly affect the morphological characteristics of traffic flow time series and will then affect the expression of system characteristics. If the time scale is too small, it is difficult to reflect the fluctuation of the traffic flow; if the time scale is too large, the overall rule of traffic flow variation throughout the day can be easily reflected, but it will be difficult to reveal the detailed characteristics of the traffic flow variation [6, 13].

## 5. Chaotic Characteristics of Air Traffic Flow Based on the Complex Network Theory

*5.1. Visibility Graph and Horizontal Visibility Graph of Time Series.* A complex network is a system composed of different individuals interacting with each other. The nodes correspond to units in the network, while an edge refers to the

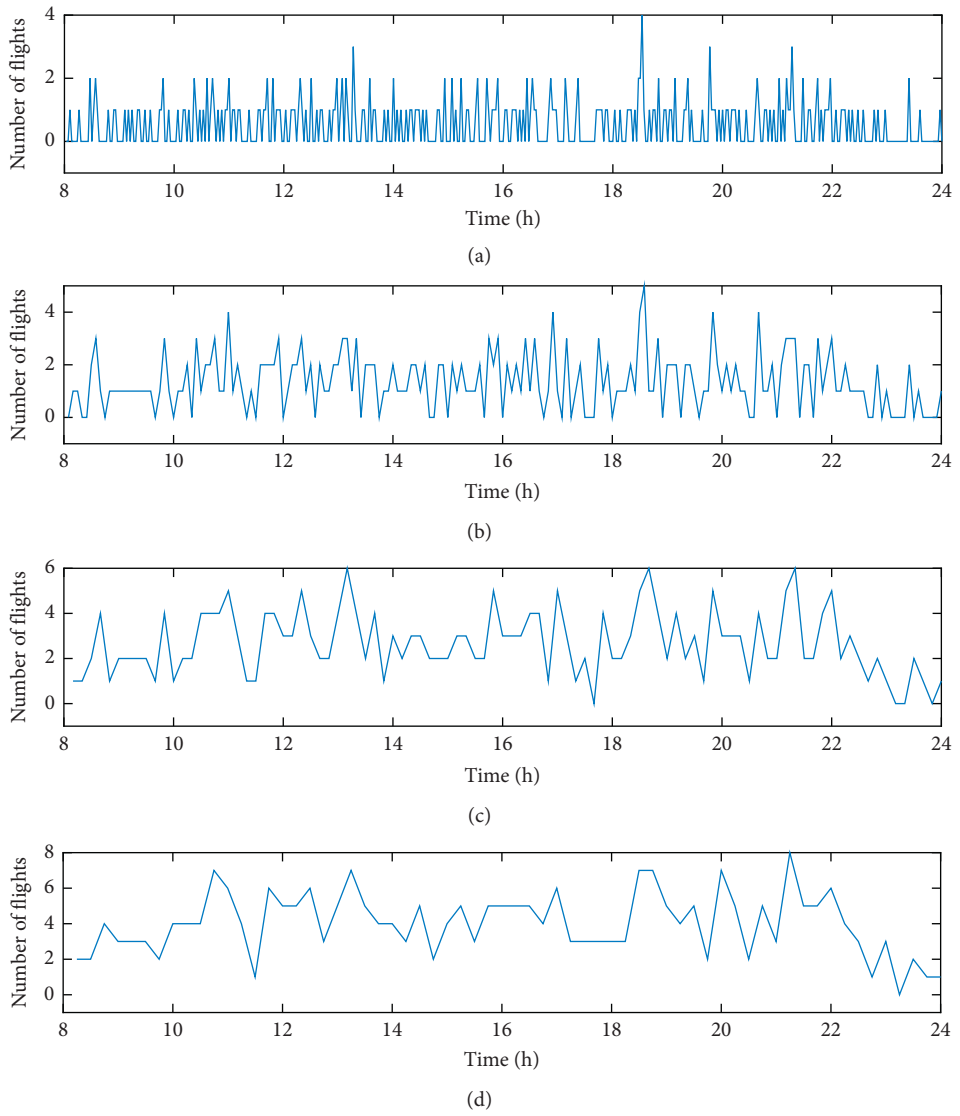



Figure 7: Different time scales of the time series. (a) $\Delta t = 2$ min. (b) $\Delta t = 5$ min. (c) $\Delta t = 10$ min. (d) $\Delta t = 15$ min.

connection between two nodes and represents the interaction between them. There are many statistical characteristics of complex networks, such as the degree distribution, accumulation coefficient, and community characteristics that are used to analyze the features contained in the original time series.

At present, complex network theory has become a new research direction for the use of complex networks as an analysis tool for the dynamic characteristics beneficial to the traffic flow theory, but also provide for time series analysis. To date, there has been little application of the complex network theory to the analysis of traffic flow time series. In this paper, complex networks are applied to the exploration of the characteristics of air traffic flow to expand support for actual traffic management and control.

In 2008, Lacasa et al. [28] proposed a new algorithm, namely, the visibility graph (VG) method, for the mapping of time series into complex networks. The principle of VG is that each node in the network corresponds to each time point in the discrete time series data. The edge of the network is the connection of two points in these nodes that meet the visibility rules specified in advance. The basic idea of the algorithm is that each node cannot be connected to itself, and each line of sight cannot pass through other straight bars.

The visibility criterion is as follows. If any two data points $(t_a, y_a)$ and $(t_b, y_b)$ in the time series are visible to each other, then for any point $(t_c, y_c)$, where $t_a < t_c < t_b$, the following is satisfied:

$$y_c > y_a + \frac{(y_a - y_b)(t_a - t_c)}{t_b - t_a}. \tag{3}$$

With the development of VG research, Luque et al. [29] proposed a geometrically simpler and less statistical algorithm based on the VG algorithm in 2009, namely, the horizontal visibility graph (HVG) method.

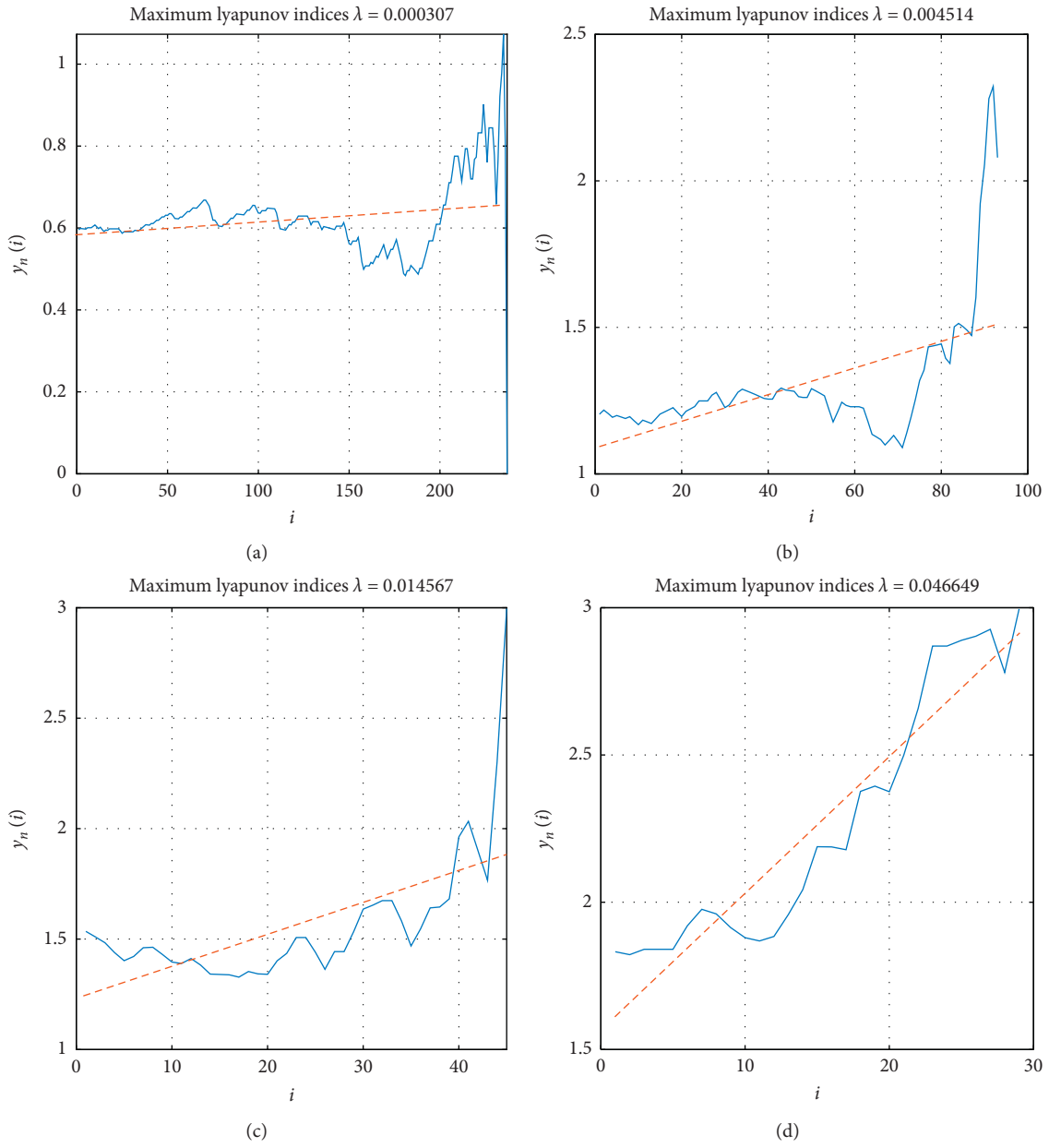



Figure 8: Analysis chart of the largest Lyapunov index of the sample. (a) $\Delta t = 2$ min. (b) $\Delta t = 5$ min. (c) $\Delta t = 10$ min. (d) $\Delta t = 15$ min.

The principle of HVG is expressed as follows. If a horizontal line of sight can be drawn between two straight bars without crossing other straight bars, the two points are considered to be connected in the network. In mathematical terms, the two data points are connected if the height of the corresponding bar is greater than that of the other bars located between the two points.

The visibility criterion is as follows. If any two data points in the time series $X_n$, such as $(t_a, y_a)$ and $(t_b, y_b)$, are visible to each other, then for any $(t_i, y_i)$, where $t_a < t_i < t_b$, the following is satisfied:

$$\begin{aligned} y_i &< y_a, \\ y_i &< y_b. \end{aligned} \tag{4}$$

Taking the time series corresponding to the four time scales presented in Section 3 as an example, the time series are, respectively, mapped into visibility and horizontal visibility graphs, and the results are presented in Figure 9.

*5.2. Analysis of the Degree Distributions of Complex Networks.* The degree distribution $p(k)$ is defined as a probability of the random selection of nodes with degree $k$. In reality, the degree distributions of many complex networks follow a power-law distribution: $p(k) = k^{-\lambda}$. Lacasa and Toral [30] demonstrated that random and chaotic time series of degree distributions of complex networks under the HVG algorithm present exponential distribution forms: $p(k) \sim e^{-\gamma k}$.

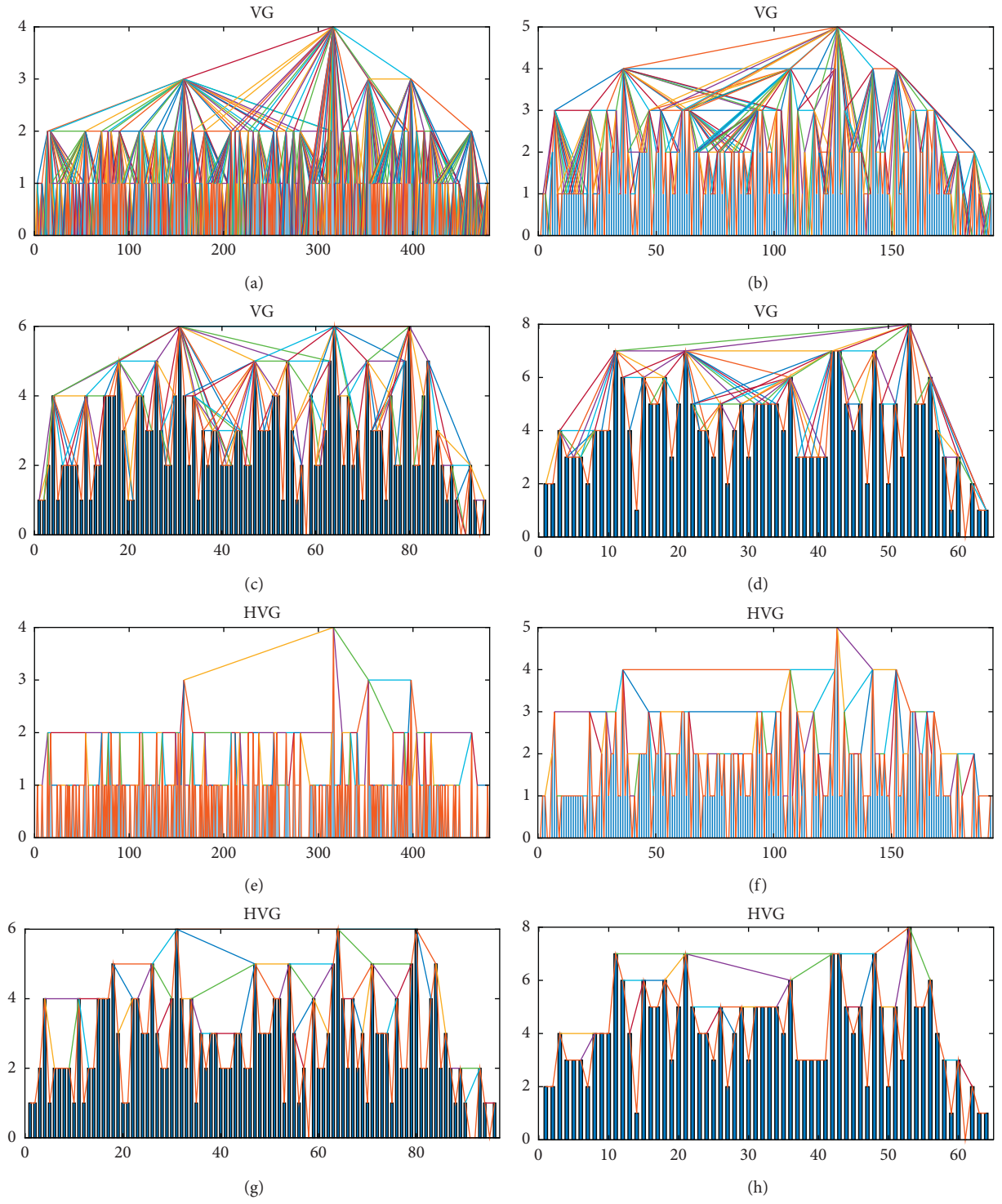



Figure 9: Visual graph and horizontal visual graph of the sample. (a) $\Delta t = 2$ min. (b) $\Delta t = 5$ min. (c) $\Delta t = 10$ min. (d) $\Delta t = 15$ min. (e) $\Delta t = 2$ min. (f) $\Delta t = 5$ min. (g) $\Delta t = 10$ min. (h) $\Delta t = 15$ min.

The VG and HVG in Figure 9 are used as examples to draw diagrams of the degree distributions, as presented in Figures 10 and 11.

In his research on the visual network analysis of nonlinear time series, Zhang [31] mentioned that the effect of the oblique line fitting of degree distributions can be described via the correlation coefficient $R^2$, the residual squared, and the sum of squares due to error (SSE). The smaller the value of SSE, the closer $R^2$ is to 1, and the higher the reference value of relevant coincidences, the better the fitting. The exponential distribution in a single logarithmic coordinate system is approximately a straight line. According to the

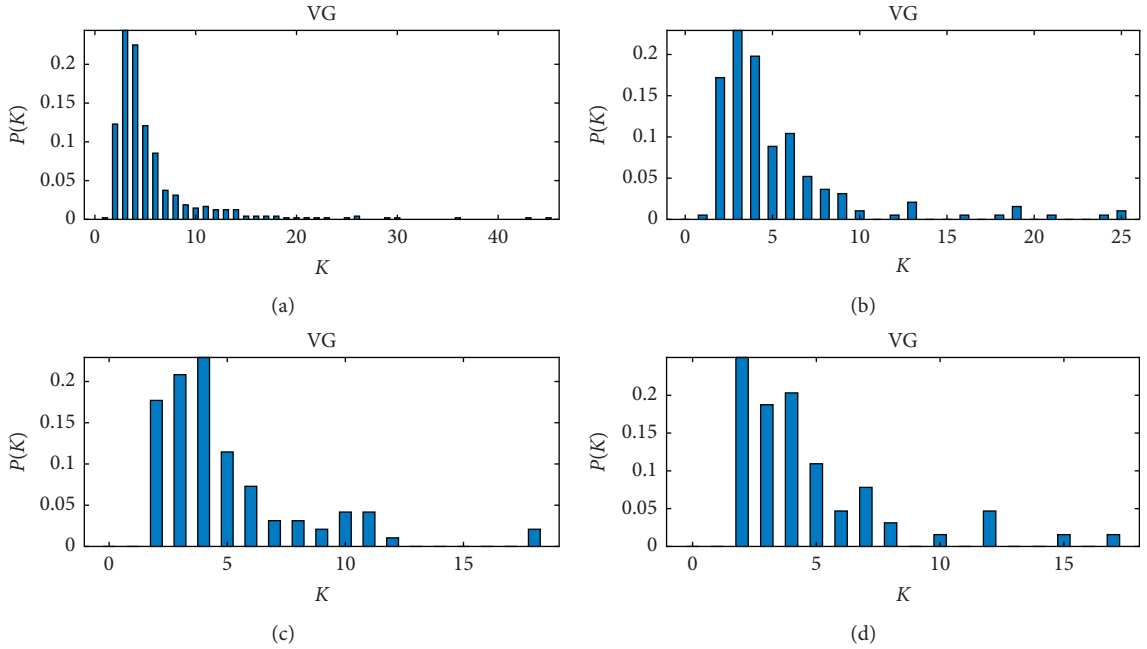



Figure 10: Degree distribution of visual graph. (a) $\Delta t = 2$ min. (b) $\Delta t = 5$ min. (c) $\Delta t = 10$ min. (d) $\Delta t = 15$ min.

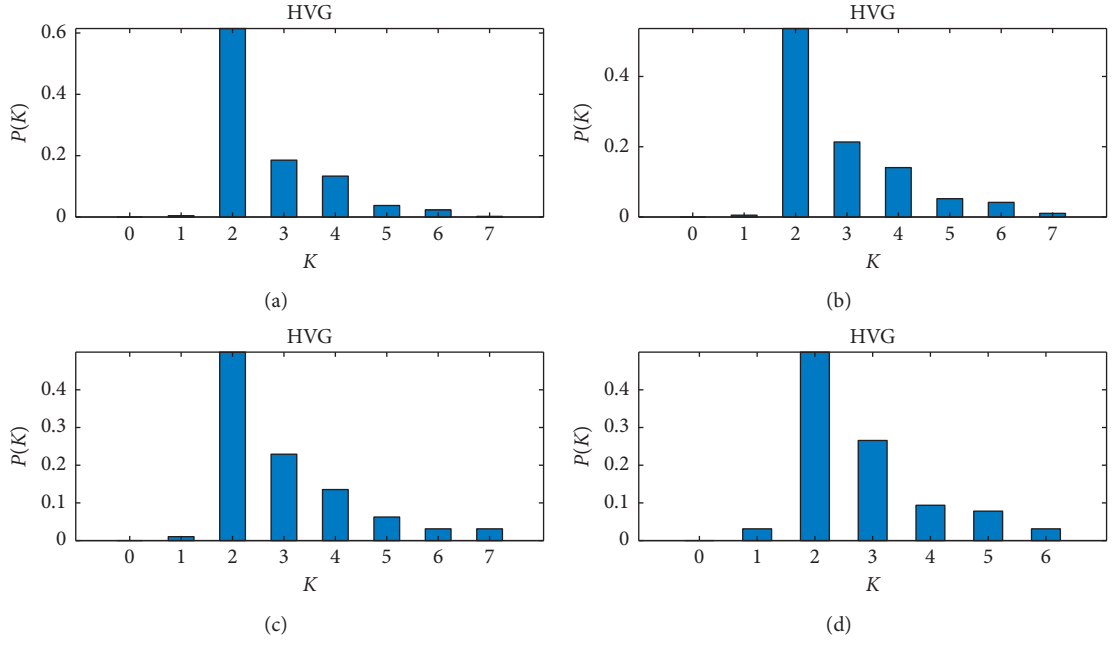



Figure 11: Degree distribution of horizontal visual graph. (a) $\Delta t = 2$ min. (b) $\Delta t = 5$ min. (c) $\Delta t = 10$ min. (d) $\Delta t = 15$ min.

steps outlined previously, the results of the four views in Figure 10 after fitting with the power-law distribution and exponential distribution are, respectively, calculated, as reported in Figure 12 and Table 1.

Similarly, according to the steps outlined previously, the results of the four horizontal visibility graphs in Figure 11 are calculated after fitting with the power-law distribution and exponential distribution, as reported in Figure 13 and Table 2.

In addition to $R^2$ and SSE, the mean squared error (MSE) is also used to evaluate the fitting effect. MSE is the mean sum of the squares of the corresponding point errors between the predicted data and the original data, i.e., SSE/$n$, and the root-mean-squared error (RMSE) is the square root of the MSE. When the original data increase, the representativeness of SSE will be affected, and the results of MSE and RMSE will be more convincing.

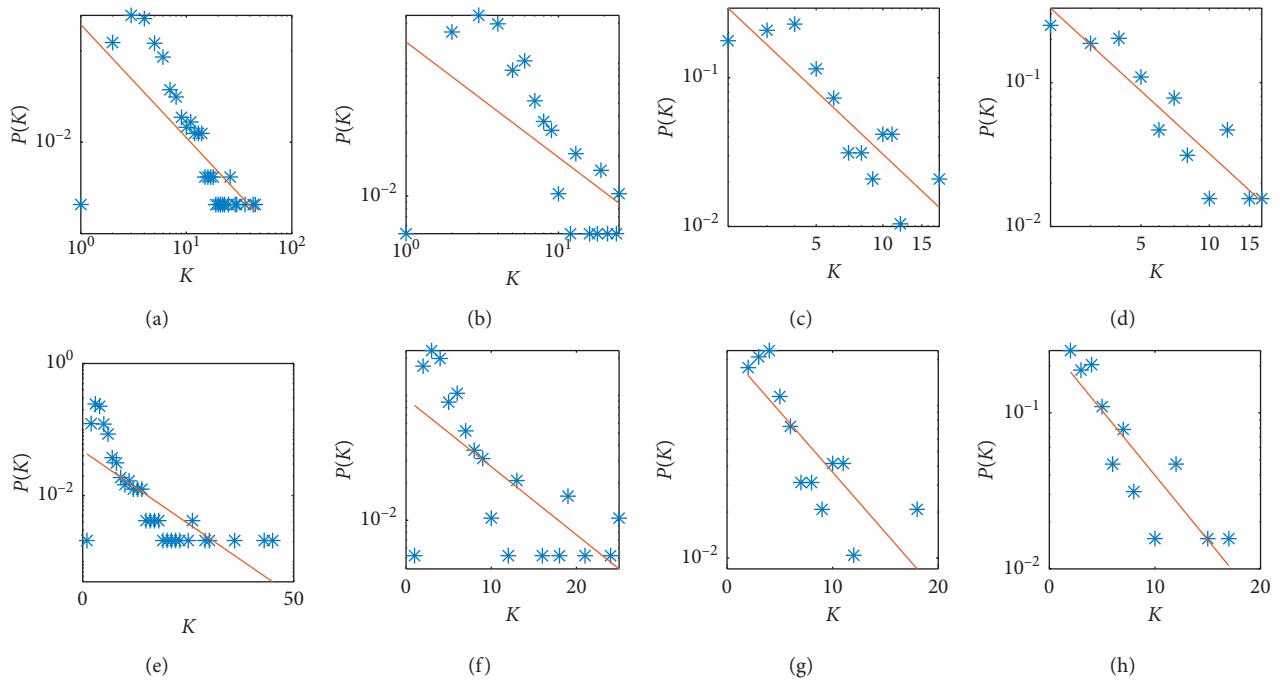



Figure 12: Results after power law distribution and exponential distribution fitting. (a) $\Delta t = 2$ min. (b) $\Delta t = 5$ min. (c) $\Delta t = 10$ min. (d) $\Delta t = 15$ min. (e) $\Delta t = 2$ min. (f) $\Delta t = 5$ min. (g) $\Delta t = 10$ min. (h) $\Delta t = 15$ min.

Table 1: Statistics of goodness of fit parameters.

| Figure | Fitting method | $\lambda/\gamma$ | SSE | $R^2$ | RMSE |
|---|---|---|---|---|---|
| (a) | Power law distribution | −1.236059 | 37.3785 | 0.4913 | 1.1554 |
| (b) | Power law distribution | −0.868055 | 23.3142 | 0.3079 | 1.2071 |
| (c) | Power law distribution | −1.403329 | 2.7372 | 0.7364 | 0.5232 |
| (d) | Power law distribution | −1.446997 | 1.7388 | 0.8448 | 0.4395 |
| (e) | Exponential distribution | −0.101365 | 33.7598 | 0.5406 | 1.098 |
| (f) | Exponential distribution | −0.125702 | 18.0168 | 0.4651 | 1.0612 |
| (g) | Exponential distribution | −0.182581 | 3.8681 | 0.6556 | 0.6219 |
| (h) | Exponential distribution | −0.190391 | 2.4722 | 0.7795 | 0.5241 |

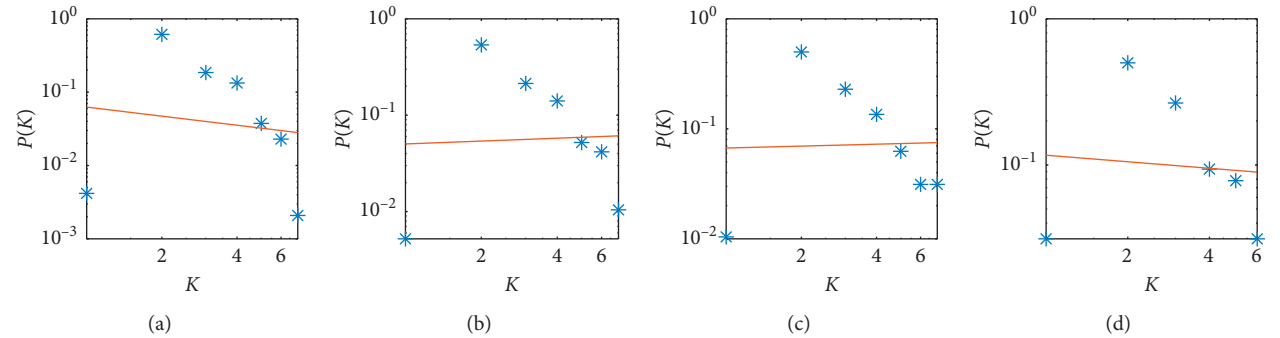



Figure 13: Continued.

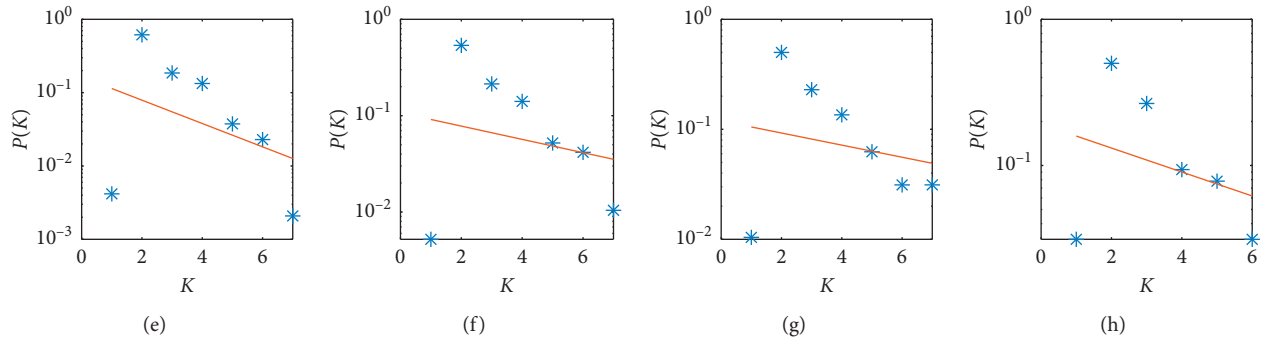



Figure 13: Results after power law distribution and exponential distribution fitting. (a) $\Delta t = 2$ min. (b) $\Delta t = 5$ min. (c) $\Delta t = 10$ min. (d) $\Delta t = 15$ min. (e) $\Delta t = 2$ min. (f) $\Delta t = 5$ min. (g) $\Delta t = 10$ min. (h) $\Delta t = 15$ min.

Table 2: Statistics of goodness of fit parameters.

| Figure | Fitting method | $\lambda/\gamma$ | SSE | $R^2$ | RMSE |
|---|---|---|---|---|---|
| (a) | Power law distribution | −0.410612 | 24.9337 | 0.0187 | 2.231 |
| (b) | Power law distribution | −0.099414 | 16.2789 | 0.0017 | 1.8044 |
| (c) | Power law distribution | −0.058225 | 10.6363 | 0.0009 | 1.4585 |
| (d) | Power law distribution | −0.148364 | 6.2668 | 0.0077 | 1.2517 |
| (e) | Exponential distribution | −0.366281 | 21.6516 | 0.1478 | 2.0809 |
| (f) | Exponential distribution | −0.158647 | 15.6019 | 0.0432 | 1.7665 |
| (g) | Exponential distribution | −0.126737 | 10.1961 | 0.0422 | 1.428 |
| (h) | Exponential distribution | −0.188867 | 5.6909 | 0.0988 | 1.1928 |

By comparing the results in Tables 1 and 2, it can be found that, for the VG algorithm, when the time scale is 2 min and 5 min, the degree distributions are closer to the exponential distribution; however, at this time, $R^2$ is too small, and the fitting results are not reliable. When the time scale is 10 min and 15 min, the distributions of the time degree are closer to the power-law distribution, and $R^2$ approaches 1; thus, the fitting results are reliable. However, for the HVG algorithm, the distribution of the uniform degree is closer to the exponential distribution. Additionally, with the increase of the time scale, the RMSE is smaller, but the results are unsatisfactory; moreover, $R^2$ is too close to 0, and the fitting results are very unreliable.

Inspired by the existing VG and HVG algorithms, this paper proposes a new visual graphic network construction method, namely, the equal-height horizontal visibility graph (EHHVG) method. The proposed EHHVG differs from the HVG in that a horizontal visual line can be drawn between two vertical bars and can pass through a vertical bar of the same height. The visibility criterion is as follows. For any two points, such as $(t_a, y_a)$ and $(t_b, y_b)$ that are visible in the time series $X_n$, for any $(t_k, y_k)$, where $t_a < t_k < t_b$, the following is satisfied,

$$\begin{aligned} y_k &\leq y_a, \\ y_k &\leq y_b. \end{aligned} \tag{5}$$

According to the new criterion, four EHHVG maps of the sample and their degree distributions are redrawn, and the results are presented in Figures 14 and 15.

The results of the fitting of the degree distributions are reported in Figure 16 and Table 3.

As is evident from Table 3, the fitting results are not ideal for the time scales of 2 min and 5 min, while they are good for the time scales of 10 min and 15 min. Compared with the degree distribution of the HVG presented in Table 2, the fitting results are more reliable. When the time scale is too small, the degree distribution points are very scattered, which is similar to the fitting results of the degree distribution of the VG reported in Table 1.

By comparing the results of the VG, HVG, and EHHVG network construction methods presented in Figures 9 and 14, it is clear that the connectivity between each node in the VG and EHHVG is relatively complex. By comparing Figures 10 and 15, the degree distribution results of the VG and EHHVG are also more dispersed, but there are some differences between the performances of various algorithms and the shape of air traffic flow after time-series feature recognition.

The conclusions of this section can be drawn as follows by comparing the goodness-of-fit parameters.

(1) In terms of the goodness of distribution fitting of the three complex network construction methods, the time series under the time scales of 10 min and 15 min are better than those under the time scales of 2 min and 5 min.

(2) The time series of air traffic flow are transformed into more concentrated and simple degree distributions of HVG.

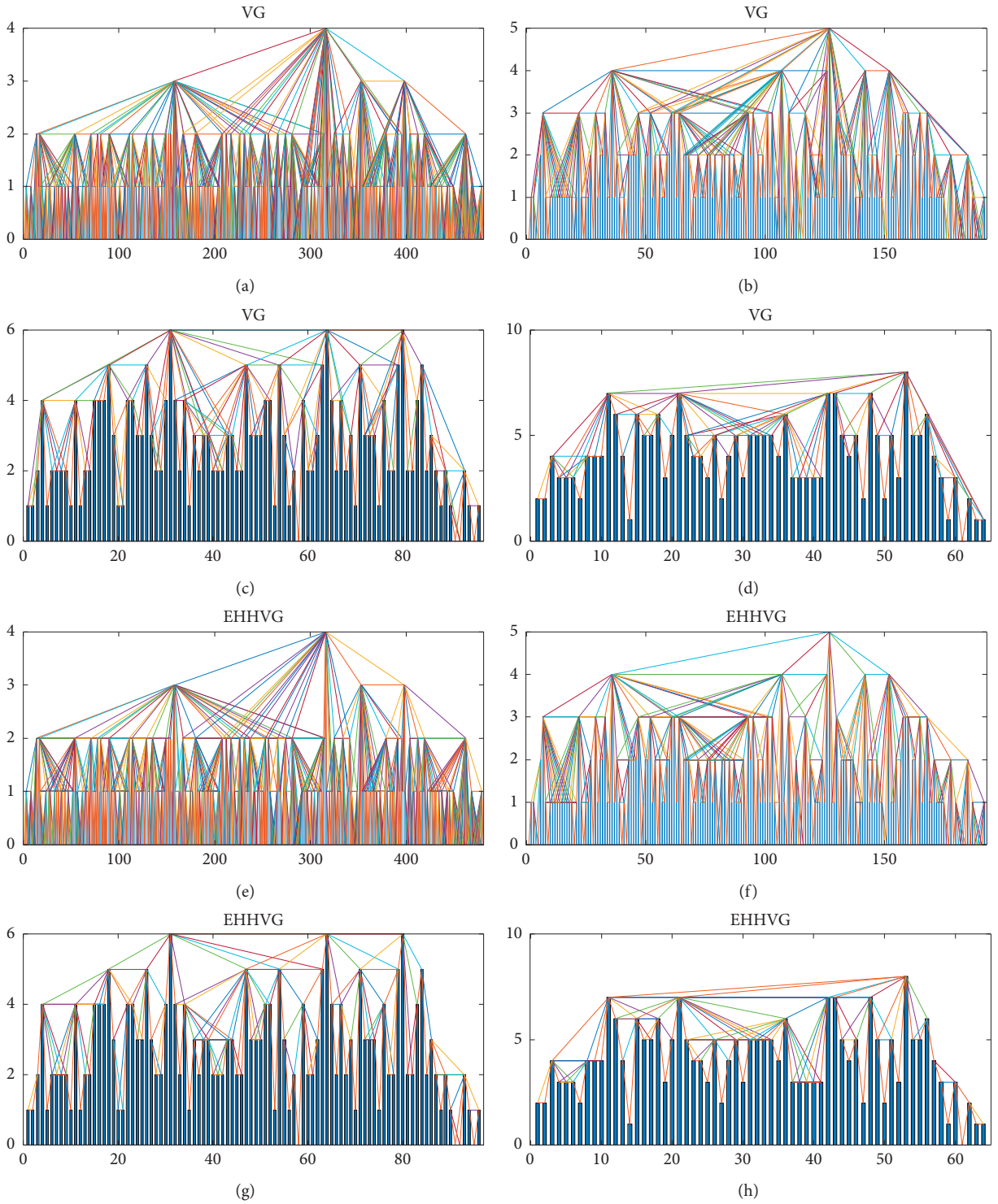



FIGURE 14: Equal height horizontal visibility graph of the sample. (a) $\Delta t = 2$ min. (b) $\Delta t = 5$ min. (c) $\Delta t = 10$ min. (d) $\Delta t = 15$ min. (e) $\Delta t = 2$ min. (f) $\Delta t = 5$ min. (g) $\Delta t = 10$ min. (h) $\Delta t = 15$ min.

(3) The fitting results of the power-law distribution and exponential distribution of the time series of the EHHVG algorithm are superior to those of the HVG method.

*5.3. Chaotic Characteristics at Different Time Scales.* Zhou et al. [32] found that the degree distribution of a complex network converted by the VG algorithm from chaotic time series presented an irregular multi-peak structure, but this

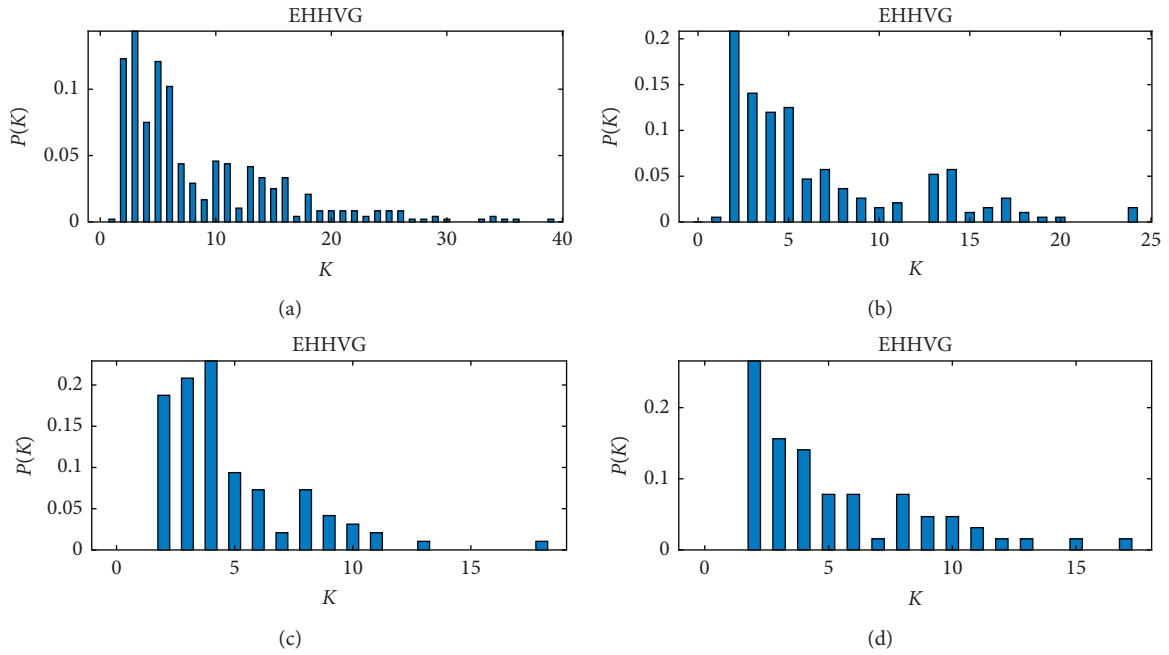



Figure 15: Degree distribution of visual graph and horizontal visual graph. (a) $\Delta t = 2$ min. (b) $\Delta t = 5$ min. (c) $\Delta t = 10$ min. (d) $\Delta t = 15$ min.

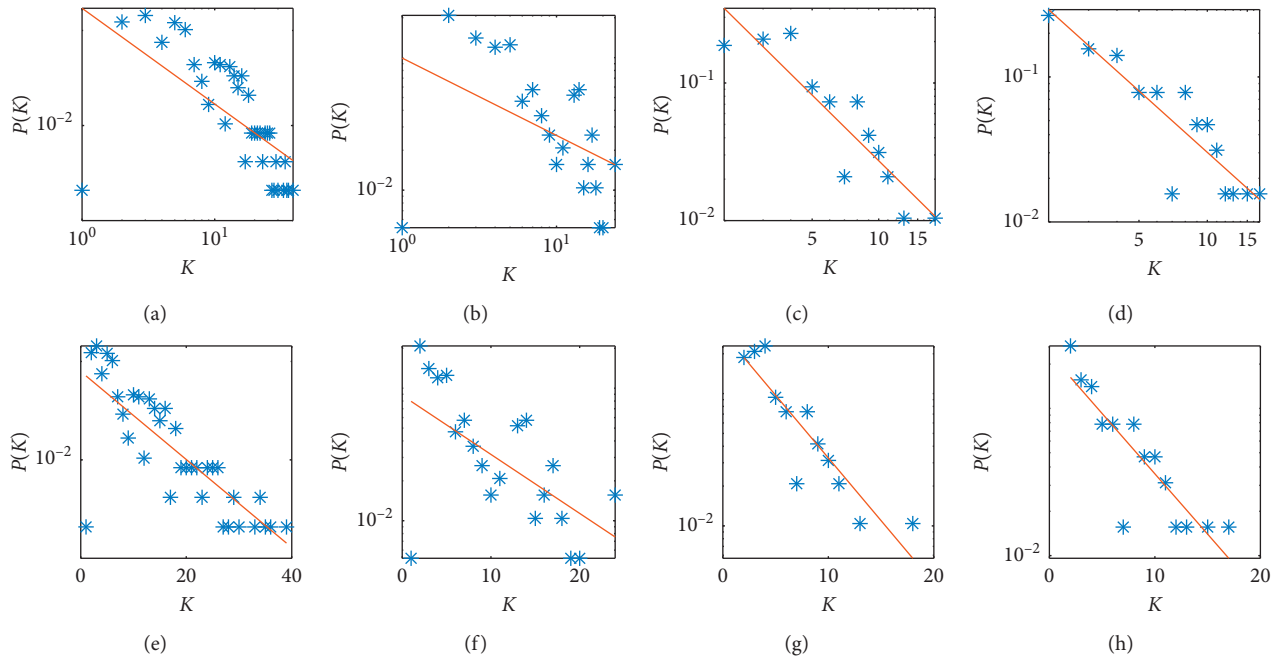



Figure 16: Results after power law distribution and exponential distribution fitting. (a) $\Delta t = 2$ min. (b) $\Delta t = 5$ min. (c) $\Delta t = 10$ min. (d) $\Delta t = 15$ min. (e) $\Delta t = 2$ min. (f) $\Delta t = 5$ min. (g) $\Delta t = 10$ min. (h) $\Delta t = 15$ min.

structure disappeared due to the weak antinoise capability of the HVG algorithm. As shown in Figure 10, the complex networks converted by the VG algorithm conform to the multipeak structure, proving that the four groups of air traffic flow data at different time scales all exhibit chaotic characteristics. A multipeak structure is not present in Figure 11, indicating that the noise-resistance capability of the HVG algorithm is weaker than that of the VG algorithm. By analyzing the degree distribution diagram of EHHVG presented in Figure 15, it can be found that a similar multipeak structure is also present. The EHHVG algorithm therefore has better noise resistance than the classical HVG algorithm.

In the research by Lacasa Toral [30], it was proven that the HVG algorithm could accurately distinguish random

Table 3: Statistics of goodness of fit parameters.

| Figure | Fitting method | $\lambda/\gamma$ | SSE | $R^2$ | RMSE |
|---|---|---|---|---|---|
| (a) | Power law distribution | −1.007224 | 38.9936 | 0.4031 | 1.0870 |
| (b) | Power law distribution | −0.583436 | 19.6875 | 0.1894 | 1.0458 |
| (c) | Power law distribution | −1.585173 | 2.4013 | 0.8220 | 0.4900 |
| (d) | Power law distribution | −1.408122 | 2.4476 | 0.8064 | 0.4516 |
| (e) | Exponential distribution | −0.102947 | 23.2955 | 0.6434 | 0.8402 |
| (f) | Exponential distribution | −0.102463 | 15.4925 | 0.3621 | 0.9277 |
| (g) | Exponential distribution | −0.218255 | 2.5323 | 0.8123 | 0.5032 |
| (h) | Exponential distribution | −0.188666 | 3.0696 | 0.7571 | 0.5058 |

time series from chaotic time series. After fitting the degree distribution into an exponential distribution form $p(k) \sim e^{-\gamma k}$, $\gamma = \ln(3/2)$ is the boundary between random and chaotic time series; when $\gamma > \ln(3/2)$, the time series is chaotic. As presented in Figure 13, the results under the four time scales all conform to $\gamma < \ln(3/2)$; thus, the time series have chaotic characteristics under the four time scales.

## 6. Conclusions

The innovations of the present work are as follows:

(1) This paper simplifies the steps of pattern recognition in air traffic flow extraction and proposes a method to quickly extract the traffic flow passing through a certain waypoint.

(2) This paper uses an advanced improved C-C method to calculate the reliability parameters. These methods are applied to the analysis of air traffic flow time series, thereby laying a solid foundation for the analysis of the chaotic characteristics of the system.

(3) This paper proposes a new visibility graph network construction method, analyzes its degree distribution characteristics, and compares it with two existing classic visibility algorithms, thereby enriching the research on complex networks for the determination of the nonlinear characteristics of time series.

(4) Three complex network analysis methods, namely, the VG algorithm, HVG algorithm, and proposed EHHVG algorithm, are compared in terms of the goodness-of-fit parameters of the degree distributions in different time scales. Regarding the goodness of distribution fitting of the three complex network construction methods, the time series under the time scales of 10 min and 15 min are better than those under the time scales of 2 min and 5 min.

There are many aspects of the characteristics of air traffic flow that are worthy of the in-depth study. Theoretical research on air traffic flow can better guide the management and control of local air traffic flow, which is of great importance for solving the problems of heavy traffic loads and frequent flight delays.

## Data Availability

The ADS-B data provide detailed data for individual flight, such as serial number, registration number, ICAO, flight call sign, latitude, longitude, altitude, speed, heading, vertical speed, receiver IP address, and flight time. The research object of this paper is air traffic flow, so it is necessary to extract time series of traffic flow from the track data. The data will be made 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 study was supported in part by the National Natural Science Foundation of China (No. 71801215) and in part by the Research Funds for Interdisciplinary Subject, Northwestern Polytechnical University.

## References

[1] J. E. Disbro and M. Frame, "Traffic flow theory and chaotic behavior," *Transportation Research Record*, vol. 1225, no. 1, pp. 109–115, 2018.

[2] D. Low and P. Addision, "Chaos in a car-following model with a desired headway time," *International Symposium on Automotive Technology & Automation Att/its Advances for Enhancing Passenger Freight & Intermodal Transportation Systems*, vol. 1225, no. 1, pp. 175–182, 1997.

[3] Y. S. Tang et al., "Chaos forecast for traffic conflict flow," *Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)*, vol. 35, no. 6, pp. 646–648, 2005.

[4] Y. Liu and J. Zhang, "Predicting traffic flow in local area networks by the largest Lyapunov exponent," *Entropy*, vol. 18, no. 1, p. 32, 2016.

[5] G. W. Frank, T. Lookman, M. A. H. Nerenberg, C. Essex, J. Lemieux, and W. Blume, "Chaotic time series analyses of epileptic seizures," *Physica D: Nonlinear Phenomena*, vol. 46, no. 3, pp. 427–438, 1990.

[6] X. Zheng, "Research on nonlinear characteristics of air traffic flow," Doctoral dissertation, Civil Aviation University of China, Tianjin, China, 2016.

[7] F. Wang, "Empirical analysis on air traffic flow long phase correlation based on Hurst exponent," *Journal of Civil Aviation University of China*, vol. 37, no. 2, pp. 1–4, 2019.

[8] S.M. Li, X.H. Xu, and L.H. Meng, "Flight conflict forecasting based on chaotic time series," *Transaction of Nanjing University of Aeronautics & Astronautics*, vol. 29, no. 4, pp. 388–394, 2012.

[9] W. Cong and M. Hu, "Chaotic characteristics analysis of air traffic system," *Transaction of Nanjing University of Aeronautics & Astronautics*, vol. 31, no. 6, pp. 636–642, 2014.
[10] Y. Yang, "Research on short term forecasting method of air traffic flow," Doctoral dissertation, Civil Aviation University of China, Tianjin, China, 2017.
[11] Y. Yang and H. Yang, "Complex network-based time series analysis," *Physica A: Statistical Mechanics and Its Applications*, vol. 387, no. 5-6, pp. 1381–1386, 2008.
[12] Z. Gao and N. Jin, "Complex network from time series based on phase space reconstruction," *Chaos An Interdiplinary Journal of Nonlinear Ence*, vol. 19, no. 3, p. 41, 2009.
[13] R. V. Donner, Y. Zou, and J. F. Donges, "Ambiguities in recurrence-based complex network representations of time series," *Physical Review E*, vol. 81, pp. 1–4, Article ID 015101, 2010.
[14] J. Pan, H. Jiang, J. Gao, and P. Yang, "Condition diagnosis with complex network-time series analysis," in *Proceedings of the Anuual Reliability & Maintainability Symposium*, January 2011.
[15] J. Tang, F. Liu, W. Zhang, S. Zhang, and Y. Wang, "Exploring dynamic property of traffic flow time series in multi-states based on complex networks: phase space reconstruction versus visibility graph," *Physica A: Statistical Mechanics and Its Applications*, vol. 450, pp. 635–648, 2016.
[16] C. Wang and M. Zhu, "Characterizing dynamic property of air traffic flow time series based on complex network," *Computer Engineering and Applications*, vol. 35, no. 6, pp. 81–85, 2018.
[17] K. Holger and S. Thomas, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, Second edition, 2003.
[18] F. Takens, "Determing strange attractors in turbulence," *Lecture Notes in Math*, vol. 898, pp. 361–381, 1981.
[19] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Vol. 127, Cambridge University Press, Cambridge, UK, 1997.
[20] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Physical Review A*, vol. 33, no. 2, pp. 1134–1140, 1986.
[21] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D: Nonlinear Phenomena*, vol. 9, no. 1-2, pp. 189–208, 1983.
[22] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Physical Review A*.vol. 45, no. 3403, pp. 34–45, 1992.
[23] J. Ma, Y. Zhang, and Z.-H. Cao, "Chaos characteristics of space object RCS based on Cao method," *Modern Radar*, vol. 31, no. 5, pp. 61–64, 2009.
[24] D. Kugiurmtzis, "State space reconstruction parameters in the analysis of chaotic times series-the role of the time window length," *Physica D*, vol. 95, pp. 13–28, 1996.
[25] H. S. Kim, R. Eykholt, and J. D. Salas, "Nonlinear dynamics, delay times, and embedding windows," *Physica D: Nonlinear Phenomena*, vol. 127, no. 1-2, pp. 48–60, 1999.
[26] N. Huang and L. Ma, "The prediction model of metro vault settlement based on developed C-C method phase space reconstruction and LS-SVM," *Mathematics in Practice & Theory*, vol. 44, no. 20, pp. 130–139, 2009.
[27] L. I. Xinjie, H. U. Tiesong, and D. Xiuming, "Application of Recurrence Plots to Nonlinear Analysis of Runoff Time Series," *Engineering Journal of Wuhan University*, vol. 46, no. 1, pp. 62–66, 2013.
[28] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, "From time series to complex networks: the visibility graph," *Proceedings of the National Academy of Sciences*, vol. 105, no. 13, pp. 4972–4975, 2008.
[29] B. Luque et al., "Horizontal visibility graphs: exact results for random time series," *Physical Review E*, vol. 80, Article ID 046103, 2009.
[30] L. Lacasa and R. Toral, "Description of stochastic and chaotic series using visibility graphs," *Physical Review E*, vol. 82, no. 3, pp. 1–26, 2010.
[31] R. Zhang, "Nonlinear time series analysis by means of visibility graphs," Doctoral dissertation, Civil Aviation University of China, Tianjin, China, 2016.
[32] T.T. Zhou, N.De Jin, Z.Ke Gao et al., "Limited penetrable visibility graph for establishing complex network from time series," *Acta Physica Sinica*, vol. 61, no. 3, Article ID 30506, 2012.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8814139, 16 pages
https://doi.org/10.1155/2020/8814139

*Research Article*

# Congestion Control for Mixed-Mode Traffic with Emission Cost

**Ya Li,[1,2] Renhuai Liu,[1,2] Yuanyang Zou,[3] Yingshuang Ma,[4] and Guoxin Wang[5]**

[1]*School of Management, Shanghai University, Shanghai 200444, China*
[2]*School of Management, Hangzhou Dianzi University, Hangzhou 310018, China*
[3]*School of Business Administration, Hubei University of Economics, Wuhan 430205, China*
[4]*Business College, Huanggang Normal University, Huanggang, Hubei 438000, China*
[5]*School of Mathematics and Physics, Nanyang Institute of Technology, Nanyang 473004, China*

Correspondence should be addressed to Yuanyang Zou; yuanyzou@gmail.com

Received 5 May 2020; Revised 14 July 2020; Accepted 27 July 2020; Published 28 August 2020

Academic Editor: Meng Meng



This paper presents two models to investigate the traffic assignment problem. In the two models, the emission cost for gasoline vehicles (GVs) is considered. The credit schemes are considered in the constraint of the models. The operation costs for battery electric vehicles (BEVs) and GVs are also studied. Particularly, the constraints related to the credit schemes can be utilized to adjust the number of GVs and to promote growth of the number of BEVs, which is a novel idea that was not studied. Preliminary numerical experiments demonstrate that the models are effective and the extended distance limit of BEVs can raise the volume of BEVs under the condition that the unit traffic cost of BEVs is lower than GVs. Therefore, it is feasible to control the quantity of GVs by adjusting the total credit schemes, and it is viable to reduce the emission by enlarging the number of BEVs' users.

## 1. Introduction

In recent years, in order to decrease petroleum consumption, various organizations are pushing to utilize various electric vehicles (e.g., [1–5]). It is predicted that electric vehicles will take up a significant market share in the near future as a result of the maturity of electric vehicle technologies and increasing public acceptance [6]. In light of engine technologies, the electric vehicles can be divided into two classes: plug-in hybrid electric vehicles (PHEVs) and battery electric vehicles (BEVs). The main difference between them is that the former is equipped with both gasoline engine and electric motor, while the latter is only equipped with electric motor. Bradley and Brank [7] stated that PHEVs do not fully mitigate environmental consequences since they still require gasoline. On the contrary, Lin and Greene [8] presented that, due to the characteristic of utilizing electricity entirely, BEVs may provide a definitive solution for electrification of personal transportation.

PHEVs and gasoline vehicles (GVs) can be substituted for BEVs to realize the goal of declining petroleum consumption, enhancing energy security, and improving environmental sustainability. However, similarly as stated by He et al. [3] and Nie and Ghamami [9], the users of BEVs may bump into a different kind of problems such as the cost caused by the limited driving range of BEVs, long charging time, scarce availability of charging stations, and limitation of battery technologies. These problems lead the BEVs not to be universally accepted by users who also have to bear the worry of being stranded when battery runs out of charge, which is normally referred to as range anxiety in the literature (e.g., [10]). It is still unrealistic to eliminate range anxiety in the near future, although more and more public charging stations have been deployed and many other strategies to deal with range anxiety have emerged (e.g., [3]).

PHEVs and gasoline vehicles (GVs) can be substituted for BEVs to realize the goal of declining petroleum consumption, enhancing energy security, and improving environmental sustainability. However, similarly as stated by He et al. [3] and Nie and Ghamami [9], the users of BEVs may bump into a different kind of problems such as the cost caused by the limited driving range of BEVs, long charging time, scarce availability of charging stations, and limitation of battery technologies. These problems lead the BEVs not to

be universally accepted by users who also have to bear the worry of being stranded when battery runs out of charge, which is normally referred to as range anxiety in the literature (e.g., [10, 11]). It is still unrealistic to eliminate range anxiety in the near future, although more and more public charging stations have been deployed and many other strategies to deal with range anxiety have emerged (e.g., [3]).

On the other hand, the fact that that the idea of tradable credit schemes was employed to decrease emissions and/or to reduce traffic congestion in urban was investigated in the latest decade. Yang and Wang [12] suggested a tradable credit plan to assuage jam which may eliminate unfairness of traditional jam pricing and analyze more reasonable pikes strategies. Three steps can be adopted as follows: initial credit allocations, credit charges, and credit transactions. Compared with traditional jam pikes plan, the financial transfer from users to governments is not contained. Afterward, transportation investigators have extensively concentrated on the tradable credit schemes. The effects of transaction costs on auction market and negotiated market for tradable credits were investigated by [13], whose investigation demonstrated that the initial allocation of credits may influence the equilibrium state. Under some proper conditions, the auction market can reach the equilibrium allocation of credits under appropriate conditions and, in the negotiated market, the transaction costs can divert the system from the desired equilibrium. The tradable credit plan for heterogeneous users with discrete value of time (VOT) was introduced by Wang et al. [14], who formulated adequate tradable credit plans and confirmed that these schemes could distribute optimal or Pareto-improving optimal traffic flow patterns to users. Moreover, the existence of optimal tradable credit scheme was proved by Xiao et al. [15], who eradicated the bottleneck queue. The price of tradable credits was studied by Shirmohammadi et al. [16], in which the aim of the study was to reduce congestion in urban by adjusting the price of tradable credits. A multiclass traffic network equilibrium issue under a given tradable credit plan with VOT decentralization was studied by Zhu et al. [17]. In addition, a novel jam decrease approach for optimizing tradable credit schemes on general traffic network was presented by Wang et al. [18]. On the other hand, a stochastic user equilibrium model containing tradable credit plan was studied by Han and Cheng [19], in which the maximization of volume of the traffic network capacity was the goal of the study. A public-private hybrid transportation network with tradable credits was scouted by Wang and Zhang [20] and Wang et al. [21], who solved the problem by equilibrium theory. For managing private financing and mobility, Bao et al. [22] proposed bilevel programming to study the problem, of which the tradable credits plan was considered. In view of the application of the tradable credits for operating the trip of car, a review paper was proposed by Dogterom et al. [23], which focused on the investigation of empiricism and the pertinent behavioral methods. In order to lessen the emission of vehicles, multiperiod tradable credit scheme methods were proposed by Miralinaghi and Peeta [24, 25]. In order to manage the queue length of vehicles at bottleneck, tradable credit plan was employed by Shirmohammadi and Yin [26]. Lahlou and Wynter [27] investigated a binary transport game containing tradable credit plans and introduced a Nash equilibrium model to solve the problem. In view of tradable credit scheme idea, tradable bottleneck permits were firstly introduced by Akamatsu and Wada [28], which was utilized to operate the transportation demand. In the design of the discrete traffic network, the noncontinuous credit pricing policy was scrutinized by Wang et al. [29]. Dogterom et al. [30] studied the adjustment actions at the trip level, in which the tradable credit plan was contained. Guo et al. [31] studied the tradable credit scheme based on the system optimum of the evolutionary traffic flows and evaluated the convergence of the system. The fact that the system optimum theory was utilized to investigate the tradable credit arrangement was introduced by Lv et al. [32]. Based on experimental economics method, Tian et al. [33] analyzed the tradable mobility credit plan with the influence factors of behavior. The experimental results demonstrate that the stated tradable mobility credit plan in their study was well-organized and economically maintainable. A novel model for tradable credit arrangement was stated by Lian et al. [34] and it was formulated based on driving-day under the congestion situation in urban. The existence of the unique equilibrium of dynamic jam noninternality with tradable credit policies was investigated by Bao et al. [35]. Gao et al. [36] studied the incremental-cost pricing of the tradable credits and utilized Cobweb model to explore the constancy of the credit price. The fact that the cyclic tradable credits were employed to improve the social justice was stated by [37]. In the same year, assuming that the nonhomogeneity of the demands and the commuters were conservative, Miralinaghi et al. [38] investigated the tradable credit scheme problem under the morning traffic congestion. The zero-emission vehicles were studied by Miralinaghi and Peeta [39]. The robust multiperiod tradable credit plan was employed to promote drivers to choose zero-emission vehicles and the target was to lessen the emission. For decreasing emission, the allocation efficiency of the method of tradable permit schemes was scrutinized by De Palma and Lindsey [40]. By their investigation, the good performance of the method was verified. The tradable credit scheme in a bimodal transportation network based on uncertain behaviors of traveler was reconnoitered by Han et al. [41], who formulated the issue to a variational equality model and proposed a heuristic approach to solve the model. Miralinaghi and Peeta [42] presented a bilevel model to optimize the multiperiod tradable credit scheme. The goal of the study was to diminish the emission and alleviate congestion.

However, how to distribute the preliminary credits is a significant and complicated procedure and it contains how to recognize the eligibility of travelers and allocate credits to competent travelers. Nie and Yin [43] suggested a bottleneck model that contains the allocation of credit without considering the initial distribution of credits to the avoidance of these difficulties. Liu and Huang [44] stated a model without initial credit distributions for general traffic network and gained a credit toll. They formulated the problem as mathematical programming with equilibrium constraints

and gained a credit charging mechanism under some conditions. Zhu et al. [45] presented a biobjective model without initial credit distribution to optimize jam and emissions and stated a system containing linear equalities and linear inequalities to solve the model.

Furthermore, the models mentioned above ignored the association between congestion and emission. Yin and Lawphongpanich [46] introduced a counterexample for social cost pricing and a nonnegative first-best emissions pricing strategy to the system optimum pricing issue for congestion and emissions. They provided methods to calculate the trade-off between conflicting objectives and alleviating congestion versus reducing traffic emissions. Chen and Yang [47] stated a biobjective optimization model to study the relationship between congestion and emissions on traffic networks. Abdul Aziz and Ukkusuri [48] also proposed a biobjective model to study the trade-off between travel time and emissions. In addition, Grant-Muller and Xu [49] scrutinized the tradable credit plan to operate a traffic jam and the behavior of the model choice, of which the goal was to decrease the total miles of vehicles. Furthermore, Li et al. [50] proposed a model for dynamic carbon credit plan to operate traffic network mobility and emissions. In the study, a bilevel model was constructed and a projection search approach was introduced to deal with the model.

In this paper, we consider a mixed traffic network in which travelers can select GVs or BEVs and the goal is to minimize the sum of traffic cost and emission cost of GVs with the constraints of credit schemes. In fact, some people may have GVs only, which makes the network optimization problem similar to the basic traffic assignment problem; some people may have BEVs only, which has already been studied by Jiang et al. [51]; others may own both GVs and BEVs, which is similar to the study conducted by Jiang and Xie [52]. In addition, a network user equilibrium problem for BEVs and GVs was studied by Hu et al. (2017), in which the BEVs and GVs were embraced in the study and the battery exchanging centers and road grade constraints also were included. Furthermore, two new impedance functions for the traffic time on the road of EVs and GVs were proposed by Lin et al. [53] and Zou et al. [54]. On the other hand, the problem that EVs and GVs were contained in ride-sourcing market was investigated by Ke et al. [55]. In the study, recharging schedules of EV drivers were considered in a time-expanded traffic network. Jensen et al. [56] stated an investigation of the real path problem, in which BEVs and internal combustion engine vehicles were contained in the study and the behavior of drivers was considered.

Based on the above discussions, the studies can be summarized as several types: some researchers investigated the traffic congestion with tradable credit scheme, some scholars studied the emissions with tradable credit plan of traditional fuel vehicles (GVs), and some investigators scrutinized the traffic assignment problem of EVs and GVs. Compared with the existing works, our main contributions can be stated as follows:

(i) Two novel models are presented in this paper, which contain BEVs and GVs. Particularly, the former model can be employed to expand the usage of BEVs and control the number of GVs in the traffic network.

(ii) In the models, the constraints of credit schemes have two cases: constraints containing both BEVs and GVs and constraints containing GVs only. The latter case can be regarded as a promotion of the travelers choosing BEVs with adjusting the total of credits displayed in our third experiment. That is, on the one hand, credit scheme can be utilized to adjust the traffic congestion. On the other hand, it can be employed to promote utilization of BEVs, which is an interesting perspective that was not investigated.

(iii) Particularly, the study of Miralinaghi and Peeta [39] must be noticed. The zero-emission vehicles were considered in the study. In the experiments of the investigation, the credits were distribution between internal combustion engine vehicles and zero-emission vehicles. Furthermore, in our study, it needs to be emphasized that the credits only for GVs are studied in the experiment and model (3.6)-(3.13) is investigated.

This paper is organized as follows: In Section 2, we introduce a new link travel time function and demonstrate its advantages. The models are constructed in Section 3, and numerical experiments are displayed in Section 4. The conclusions are drawn in Section 5.

## 2. Link Time Functions

As is known to us, the transportation optimization models are generally in the forms of

$$\begin{aligned} &\min \quad pt \sum_{a \in A} t_a\left(v_a(\rho)\right), \\ &\text{or } \min \quad pt \sum_{a \in A} t_a\left(x_a\right), \end{aligned} \tag{1}$$

$$\begin{aligned} &\min \quad pt \sum_{a \in A} pt_a\left(v_a(\rho)\right) + c\left(x_a\right), \\ &\text{or } \min \quad pt \sum_{a \in A} pt_a\left(x_a\right) + c\left(x_a\right), \end{aligned} \tag{2}$$

subject to link constraints, path constraints, and some other related constraints. Here, $A$ is the set of all links in the network, $\rho$ denotes traffic density per mile, $x_a$ denotes traffic flow on link $a \in A$, $v_a$ denotes a speed-density impedance function, $t_a$ denotes a time-speed or time-flow impedance function, $p$ is the price of unit time, and $c$ stands for other costs associated with vehicles such as energy cost.

The objective functions (1) and (2) require traffic network to minimize total travel costs and main difference between them is that (2) also minimizes other related costs. Model (2.1) was firstly applied to describe network equilibrium by Jiang et al. [51], while (2) was firstly considered by Jiang and Xie [52]. The GVs and BEVs were embraced in the two studies. Both are reasonable from their own perspectives. In this paper, we introduce two models different from the objective functions (1) and (2).

The impedance function $t_a$, which reveals the relationship between travel costs and traffic conditions on the road reflecting the crowed effect of traffic network, plays a significant role in (1) and (2). The most popular impedance function is the Bureau of Public Roads (BPR) function defined by

$$t_a^{\mathrm{BPR}}(x_a) = t_a^0\left(1 + \alpha\left(\frac{x_a}{C_a}\right)^{\beta}\right), \tag{3}$$

where $x_a$, $t_a^0$, and $C_a$ denote the traffic flow, free travel time, and capacity on link a, respectively, and $\alpha > 0$ and $\beta > 0$ are specified parameters. This function is gained through regression method based on investigating lots of road section by the Bureau of Public Roads in 1964.

In this paper, in view of the link travel time function given by [53], we define a new time impedance function:

$$t_a(x_a, b_a) = \begin{cases} t_a^0, & b_a \le x_a + b_a \le C_a, \\ t_a^0\left(1 + \alpha\left(\dfrac{x_a + b_a - C_a}{Z_a - C_a}\right)^{\beta}\right), & C_a < x_a + b_a \le Z_a, \\ t_a^0\left(1 + \alpha\left(\dfrac{\epsilon}{Z_a - x_a - b_a + \epsilon}\right)^{\beta}\right), & Z_a < x_a + b_a < Z_a + \epsilon. \end{cases} \tag{4}$$

Here, the meanings of $\{x_a, t_a^0, \alpha, \beta, C_a\}$ are the same as above, $Z_a$ is a critical value over which there will be a traffic jam on link $a$, $b_a \ge 0$ is the number of the buses on link $a$, and $\epsilon > 0$ is a small constant. This impedance function, which is depicted in Figure 1, has the following properties.

Comparing with the BPR function and the time function of [53], the number of the buses is contained on link $a$ in equation (4), which is more in line with the actual traffic conditions in urban. On the other hand, since the bus travel route is fixed, there is no need to assign the routing. Thus, we ignore the factor in the study.

**Theorem 1.** *Suppose that $\alpha > 0$, $\beta > 1$, $b_a \ge 0$, and $\epsilon \le Z_a - C_a - b_a$. Then, one has the following statements:*

*(1) The function $t_a(x_a, b_a)$ is continuous over $[0, Z_a + \epsilon)$*

*(2) The function $t_a(x_a, b_a)$ is differentiable everywhere over $[0, Z_a + \epsilon)$ except the point $Z_a$*

*(3) The functions $t_a(x_a, b_a)$ and $t_a(x_a, b_a) \cdot x_a$ are both convex over $[0, Z_a + \epsilon)$*

*Proof.* The proof of theorem (1) is analogous to [53].

Comparing with the BPR function and the time function of [53], the advantages of this link travel time function can be stated as follows:

(1) When the number of vehicles is no more than the capacity allowing to travel freely, travel time is stable in general.

(2) When the number of vehicles is more than the capacity, travel time begins to change sharply when approaching the threshold value.

(3) When congestion happens, few vehicles can travel into the network. In this case, we may think that the number of vehicles is a constant but travel time still increases. □

## 3. Models

The notations in the considered traffic network are as follows: $K$ denotes the set of O-D pairs; $k \in K$ means a single O-D; $k_g \in K$ and $k_e \in K$ mean feasible paths of GVs and BEVs of O-D pair $k$, respectively; $\omega_{k_g}$ and $\omega_{k_e}$ stand for the traffic flows of GVs and BEVs between O-D pairs $k$, respectively; $\omega_{k_g}^a$ and $\omega_{k_e}^a$ stand for the traffic flows of GVs and BEVs on link $a$ between O-D pairs $k$, respectively; $A$ is the set of all links in traffic network; $a \in A$ denotes a single link in the traffic network; $a_g \in A$ and $a_e \in A$, respectively, are feasible links of GVs and BEVs for $a \in A$; $x_{a_g}$ and $x_{a_e}$ denote the total traffic flows of GVs and BEVs on link $a$, respectively; $d_{k_g}$ and $d_{k_e}$ denote the original outflows or terminal inflows of GVs and BEVs between O-D pairs $k$, respectively.

*3.1. Constraints.* The set of feasible path flow patterns is defined by

$$\Gamma_\omega = \left\{ \left(\omega_{k_e}, \omega_{k_g}\right) \,\middle|\, \omega_{k_e} \ge 0, \omega_{k_g} \ge 0, \sum_k \omega_{k_e} = d_{k_e}, \sum_k \omega_{k_g} = d_{k_g}, k_e \in k, k_g \in k, k \in K \right\}, \tag{5}$$

and the set of feasible link flow patterns is defined by

$$\Gamma_x = \left\{ \left(x_{a_e}, x_{a_g}\right) \,\middle|\, x_{a_e} = \sum_{k_e \in K} \omega_{k_e}^a, x_{a_g} = \sum_{k_g \in K} \omega_{k_g}^a, x_{a_g} + x_{a_e} + b_a \le Z_a, a_g \in a, a_e \in a, a \in A \right\}. \tag{6}$$

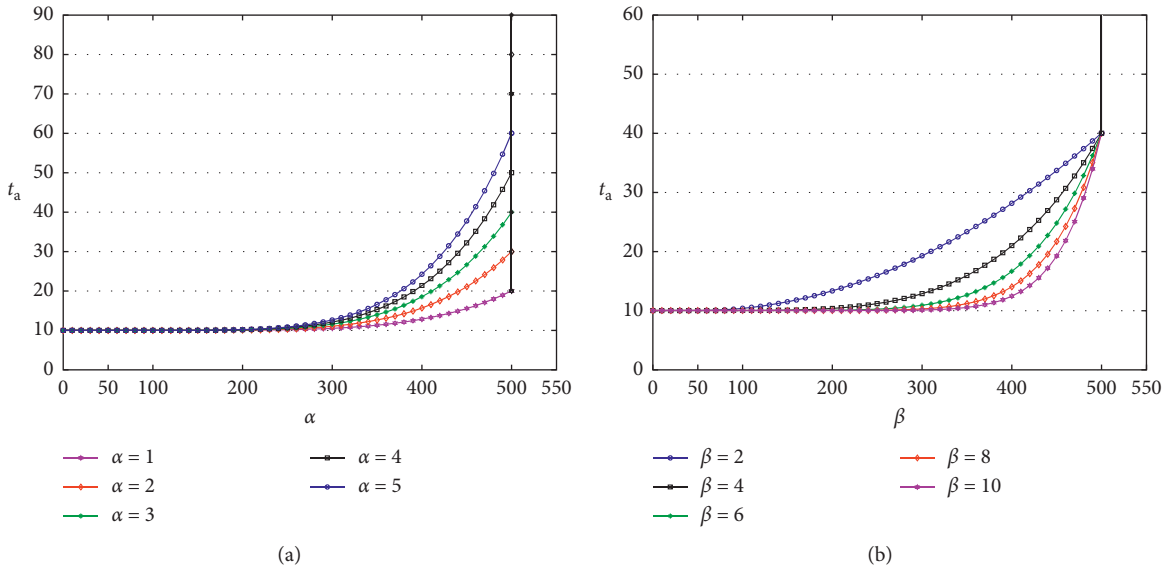



FIGURE 1: The impedance function $t_a(x_a)$}. (a) Different value of $\alpha$. (b) Different value of $\beta$.

In addition, let $Q$ be an upper bound of the total amount of credits, with which the traffic management department intends to control the total of the vehicles in the traffic network [43, 45]. In other words, in general, the total demand for the credits may be more than the supply. Thus, in order to adjust the traffic situation, the traffic management department can adjust the upper bound of the total amount of credits in the traffic network.

*3.2. Travel Cost Function.* In a traffic network, the traffic costs consist of several parts: the traffic time cost, operating cost, and depreciation charge of vehicles. Here, we consider a mixed traffic network with GVs and BEVs, which is distinguished by two factors: driving distance limit and travel cost. The operating cost contains the cost of consumption of energy and depreciation charge. For simplicity, since the routes of buses need not be assigned, thus, let $c_{a_e}(x_{a_e} + x_{a_g}, l_a)$ and $c_{a_g}(x_{a_g} + x_{a_e}, l_a)$ denote the total operating costs functions of BEVs and GVs, respectively; let $c_{a_e}$ and $c_{a_g}$ be the unit price of the BEVs and GVs on link $a$, respectively; let $p_a$ be the price of unit time on link $a$; and $l_a$ denotes the length of link $a$. Then, we can establish the system cost function for the mixed traffic network by

$$F = \sum_{a \in A} \left( p_a t_a \left( x_{a_e} + x_{a_g}, b_a \right) \left( x_{a_e} + x_{a_g} \right) + c_{a_e} \left( x_{a_e} + x_{a_g}, l_a \right) + c_{a_g} \left( x_{a_e} + x_{a_g}, l_a \right) \right). \tag{7}$$

Note that the objective function is similar to [51, 57]. However, the time function of equation (7) is different from that in papers [51, 57]. Thus, (7) is more reasonable.

*3.3. Emissions Function from Vehicles.* The emission engendered by gasoline vehicles on link $a$ is a detachable function $r_a(x_a)$ of link flow and speed. Yin and Lawphongpanich [46] investigated the first-best nonnegative emission tolls for minimization emission problem and proved that a first-best nonnegative emission toll exists when the emission function $r_a(x_a)$ is increasing with respect to link $a \in A$ in a transportation network, which is expressed as

$$R_a = r_a(x_a) x_a, \tag{8}$$

where $x_{a_g} = \sum_{k \in K} \omega_{k_g}^a$ is the sum of the traffic flow of GVs on link $a$ and $R(x_{a_g})$ denotes the sum of the emissions of the traffic network.

The relationship between emission and traffic flow on link $a$ has been investigated in recent years (e.g., [45–47, 58]), which is formulated as

$$r_a(x_a) = 0.2038 t_a(x_a) \exp\left( \frac{0.7962 l_a}{(t_a x_a)} \right), \quad \forall a \in A, \tag{9}$$

where $l_a$ denotes the length for link $a \in A$.

In this paper, we only consider the emissions of GVs. Then, the emissions function \$R_a\$ on link \$a\$ can be formulated as

$$R_a = r_a\left( x_{a_g} + x_{a_e} \right) x_{a_g} = 0.2038 t_a\left( x_{a_g} + x_{a_e} \right) \cdot \exp\left( \frac{0.7962 l_a}{t_a\left( x_{a_g} + x_{a_e} \right)} \right) x_{a_g}, \tag{10}$$

where $t_a(x_a)$ is defined as (4).

*3.4. Optimization Models.* In this subsection, we state two models to optimize the flow assignment problem and emissions of GVs in urban traffic network. The first model is to minimize the sum of travel time cost and operating cost of

BEVs and GVs, the anxiety cost of BEVs, and the emissions cost of GVs; that is,

$$\min_{\omega_{k_e},\omega_{k_g}} \quad \alpha_1 \sum_{a\in A}\left(p_a t_a\left(x_{a_e}+x_{a_g}\right)\left(x_{a_e}+x_{a_g}\right)+c_{a_e}\left(x_{a_e}+x_{a_g},l_a\right)+c_{a_g}\left(x_{a_e}+x_{a_g},l_a\right)\right)+\alpha_2\sum_{a\in A} r_a\left(x_{a_g}+x_{a_e}\right)x_{a_g}, \tag{11}$$

$$\text{s.t.} \quad E_{k_g}\omega_{k_g}=d_{k_g}i_{k_g}, \quad \forall k_g\in K, \tag{12}$$

$$E_{k_e}\omega_{k_e}=d_{k_e}i_{k_e}, \quad \forall k_e\in K, \tag{13}$$

$$x_{a_g}=\sum_{k\in K}\omega_{k_g}^{a}, \quad \forall a\in A, \tag{14}$$

$$x_{a_e}=\sum_{k\in K}\omega_{k_e}^{a}, \quad \forall a\in A, \tag{15}$$

$$x_{a_g}+x_{a_e}\le Z_a, \quad \forall a\in A, \tag{16}$$

$$\sum_a \kappa_a x_{a_g}\le Q, \quad \forall a\in A, \tag{17}$$

$$\begin{aligned}&\omega_{k_g}\ge 0,\\ &\omega_{k_e}\ge 0,\\ &\forall k_g\in K,\\ &k_e\in K.\end{aligned} \tag{18}$$

Here, $\alpha_1\ge 0$ and $\alpha_2\ge 0$ are weight factors, and $\lambda_a$ is the unit price of emissions. The first two equalities follow from the fact that each column of the matrix $E_{k_g}$ has only two nonzero elements $\{1,-1\}$ and so is the matrix $E_{k_e}$. The vectors $i_{k_g}=(1,0,\ldots,0,-1)^T$ and $i_{k_e}=(1,0,\ldots,0,-1)^T$ are with suitable dimensions, where the element 1 corresponds to the original of $k_g$ or $k_e$ and the element $-1$ corresponds to the terminal of $k_g$ or $k_e$. Conditions (12) and (13) are flow balance constraints of GVs and BEVs; that is, the flows at middle nodes are equal to zero and the flows at origins or terminals are equal to $d_k$. Conditions (14)–(16) mean that the total traffic flow on each feasible link is no more than its threshold value. Condition (17) is the sum of credits that is no more than the given number of the traffic managements.

In addition, constraint (17) is only for the GVs; that is, BEVs are not constrained by the credit schemes, which can be regarded as a new method to expand the usage of BEVs in the traffic network.

If the link flow of BEVs is not considered in constraint (17), we use

$$\sum_a \kappa_a\left(x_{a_e}+x_{a_g}\right)\le Q, \quad \forall a\in A, \tag{19}$$

to replace (17). That means that GVs and BEVs are constrained by the credit schemes and (17) can be utilized to adjust the volume of GVs and BEVs in the traffic network. Then, model (11)–(18) can be transferred as another model:

$$\min_{\omega_{k_e},\omega_{k_g}} \quad \alpha_1 \sum_{a\in A}\left(p_a t_a\left(x_{a_e}+x_{a_g}\right)\left(x_{a_e}+x_{a_g}\right)+c_{a_e}\left(x_{a_e}+x_{a_g},l_a\right)+c_{a_g}\left(x_{a_e}+x_{a_g},l_a\right)\right)+\alpha_2\sum_{a\in A} r_a\left(x_{a_g}+x_{a_e}\right)x_{a_g}, \tag{20}$$

$$\text{s.t.} \quad (12)-(16),\ (18),\ (19). \tag{21}$$

Models (11)–(18) and (20) and (21) are not studied in the former studies. Particularly, the constraint of the credits schemes is utilized to control the GVs number and to promote the usage of BEVs in the traffic network.

*Remark 1.* As the study of [57], the distance constraint of BEV is ignored in models (11)–(18) and (20) and (21). It is considered in algorithm in the next section.

**Proposition 1.** *In model (11)–(18), if $Q = 0$, there are no GVs in the traffic network and so the emission of GVs is 0; that is, all users choose BEVs.*

*Proof.* Since $Q = 0$, according to (17), (14), and (18), we have $x_{a_g} = 0$ and $R(\omega_{k_g}) = 0$. This completes the proof.

Based on model (11)–(18), we know that reducing the number of credits of GVs can promote the utilization of BEVs in the traffic network.

On the other hand, note that the Lagrangian function of model (11)–(18) can be expressed as

$$\begin{aligned} L\left(\omega_{a_e}, \omega_{a_g}, \gamma, \mu\right) = {} & \alpha_1 \sum_{a\in A}\left(p_a t_a\left(x_{a_e} + x_{a_g}\right)\left(x_{a_e} + x_{a_g}\right)\right. \\ & \left. + c_{a_e}\left(x_{a_e} + x_{a_g}, l_a\right) + c_{a_g}\left(x_{a_e} + x_{a_g}, l_a\right)\right) \\ & + \alpha_2 \sum_{a\in A} r_a\left(x_{a_e} + x_{a_g}\right)x_{a_g} \\ & + \sum_{k_g\in K} \gamma_{k_g}^T\left(E_{k_g}\omega_{k_g} - d_{k_g} i_{k_g}\right) \\ & + \sum_{k_e\in K} \gamma_{k_e}^T\left(E_{k_e}\omega_{k_g} - d_{k_e} i_{k_e}\right) \\ & + \sum_{a\in A} \mu_a\left(x_{a_g} + x_{a_e} - Z_a\right) \\ & + \mu_c\left(\sum_a \kappa_a x_{a_g} - Q\right) \\ & - \sum_{k_g\in K} \mu_{k_g}^T \omega_{k_g} - \sum_{k_e\in K} \mu_{k_e}^T \omega_{k_e}, \end{aligned} \tag{22}$$

where the Lagrangian multipliers $\mu_{k_g}^T$ and $\mu_{k_e}^T$ are the vectors of $\mu_{k_g}^a$ and $\mu_{k_e}^a$ for $a \in A$, respectively. One has

$$\begin{aligned} x_{a_g} &= \sum_{k_g\in K} \omega_{k_g}^a, \\ x_{a_e} &= \sum_{k_e\in K} \omega_{k_e}^a, \\ &\forall a \in A. \end{aligned} \tag{23}$$

Then, the optimality conditions for model (11)–(18) can be written as

$$\begin{aligned} \frac{\partial L}{\partial \omega_{k_g}^a} = {} & \alpha_1\left(pt_a'\left(x_{a_e} + x_{a_g}\right)\cdot\left(x_{a_e} + x_{a_g}\right)\right. \\ & + pt_a\left(x_{a_e} + x_{a_g}\right) + c_{a_e}'\left(x_{a_e} + x_{a_g}, l_a\right) \\ & \left. + c_{a_g}'\left(x_{a_e} + x_{a_g}, l_a\right)\right) \\ & + \alpha_2\left(r_a'\left(x_{a_e} + x_{a_g}\right)x_{a_g} + r_a\left(x_{a_e} + x_{a_g}\right)\right) \\ & + \left(\gamma_{k_g}^{a_i} - \mu_{k_g}^{a_j}\right) + \mu_a + \mu_c - \mu_{k_g}^a = 0, \\ \frac{\partial L}{\partial \omega_{k_e}^a} = {} & \alpha_1\left(pt_a'\left(x_{a_e} + x_{a_g}\right)\cdot\left(x_{a_e} + x_{a_g}\right)\right. \\ & \left. + pt_a\left(x_{a_e} + x_{a_g}\right) + c_{a_e}'\left(x_{a_e} + x_{a_g}, l_a\right)\right) \\ & + c_{a_g}'\left(x_{a_e} + x_{a_g}, l_a\right) + \alpha_2 r_a'\left(x_{a_e} + x_{a_g}\right)x_{a_g} \\ & + \left(\gamma_{k_e}^{a_i} - \mu_{k_e}^{a_j}\right) + \mu_a - \mu_{k_e}^a = 0, \\ x_{a_g} + x_{a_e} - Z_a \le 0, & \\ \mu_a \ge 0, & \\ \mu_a\left(x_{a_g} + x_{a_e} - Z_a\right) = 0, & \\ \sum_a \kappa_a x_{a_g} - Q \le 0, & \\ \mu_c \ge 0, & \\ \mu_c\left(\sum_a \kappa_a x_{a_g} - Q\right) = 0, & \\ x_{a_g} = \sum_{k_g\in K} \omega_{k_g}^a, & \\ x_{a_e} = \sum_{k_e\in K} \omega_{k_e}^a, & \\ \forall k_g \in K, k_e \in K, a = \left(a_i, a_j\right) \in A. & \end{aligned} \tag{24}$$

We then have the following results. □

**Proposition 2.** *The Lagrangian multipliers of* $\gamma_{k_g}^{a_i}, \mu_{k_g}^{a_j}, \mu_{k_g}^{a}, \gamma_{k_e}^{a_i}, \mu_{k_e}^{a_j}, \mu_a$, *and* $\mu_{k_e}^{a}$ *satisfy the equation*

$$\left(\mu_{k_g}^{a_j} + \gamma_{k_e}^{a_i}\right) - \left(\gamma_{k_g}^{a_i} + \mu_{k_e}^{a_j}\right) + \left(\mu_{k_g}^{a} - \mu_{k_e}^{a}\right) + \mu_a = \alpha_2 r_a\left(x_{a_e} + x_{a_g}\right). \tag{25}$$

*Proof.* Since $\partial L/\partial \omega_{k_g}^{a} = \partial L/\partial \omega_{k_e}^{a} = 0$, the conclusion can be proved immediately.

From Proposition 2, we have an interesting property that the Lagrangian multipliers only relate to the emission cost and the weight factor $\alpha_2$ for every $a$. In addition, in model (11)–(18), the definitions of the operating function do not impact the values of the Lagrangian multipliers. Furthermore, the emissions are only affected by the shadow prices, while the model reaches optimality. □

**Proposition 3.** *If there are no waiting users in the traffic network, then the Lagrangian multiplier* $\mu_a$ *is equal to zero for all* $a$ *in* $A$.

*Proof.* According to the definition of $Z_a$ and the assumption of this proposition, we have $x_{a_g} + x_{a_e} < Z_a$. It follows that $\mu_a = 0$ on all links in the traffic network.

It means that constraint equation (16) is noneffective. That is, the situation of traffic does not attain the critical value over which there will be a traffic jam. □

**Proposition 4.** *If the total credits are more than the demand in the traffic network, then the Lagrangian multiplier* $\mu_c$ *is equal to zero.*

It means that constraint equation (17) is noneffective. That is, decreasing emission by setting the total credits is invalid.

In addition, two properties can be obtained as follows:

(1) If $\alpha_1 > 0$ and $\alpha_2 = 0$, the objective functions of models (11)–(18) and (20) and (21) do not involve the emission factor. That is, in this situation, models (11)–(18) and (20) and (21) only study the credit schemes problem by minimizing the sum of the time cost and operating cost of BEVs and GVs. The model is only employed to reduce traffic congestion by choosing proper traffic paths. Of course, if the traffic is smooth, the emission also is declined.

(2) If $\alpha_1 = 0$ and $\alpha_2 > 0$, the models only consider the emission of the traffic network with the credit schemes. That is, the goals of models (11)–(18) and (20) and (21) are utilized to lessen the emission by assigning proper paths.

## 4. Algorithm and Numerical Experiments

In our numerical tests, we use a Windows 8.1-based PC equipped with a Core (TM) 2 CPU i5-4210M processor running at 2.60 GHz as the computing platform and algorithm is coded in Matlab2010b. In addition, the numerical experiments contained two main steps: find feasible path of GVs and BEVs; choose proper algorithm to solve the optimization models.

In particular, the emission function is chosen as (10) and the time function $t_a(x_{a_e} + x_{a_g})$ is defined as (4).

*4.1. Algorithm Frame.* In this subsection, we give an algorithm frame (see Table 1) to solve the models.

In view of Table 1, the algorithm can be summarized as two stages: find the feasible paths with distance limit \$D\$ of BEVs and feasible paths of GVs and solve model (11)–(18).

*4.2. Numerical Experiments.* In our tests, we had four aims: (1) checking the effect of changing distance limit of BEVs for choice of users in model (11)–(18); (2) testing the influence of changing unit operation cost of BEVs and GVs for pick of users in model (11)–(18); (3) verifying the impact of varying total credits for pick of users in model (11)–(18); (4) testing the influence of variation of $\lambda_a$ in $R_a = r_a(x_a)x_a = \lambda_a x_{a_g} l_a$ for selection of users in model (20) and (21).

Nguyen–Dupuis' network contains 13 nodes and 19 links (see Figure 2). More information about this network can be found in [59]. All feasible paths, O-D pairs, route compositions, and length of every path are shown in Table 2 [53, 54]. The travel demand of each O-D pair was set to be 400 and $\alpha_1 = \alpha_2 = 1$, $\alpha = 0.15, \beta = 100$, $\rho = 1$, and $p = 10$ in all experiments.

In the first experiment, for simplicity, in model (11)–(18), let the unit price $p$ of travel time be equal to 10. Let the operating cost functions of BEVs and GVs on link $a$ be expressed as $c_{a_e} = 0.3 l_a(x_{a_e} + x_{a_g})$ and $c_{a_g} = l_a(x_{a_g} + x_{a_g})$, respectively. Furthermore, let emissions cost $r_a(x_a)$ be equivalent to $x_{a_g} l_a$ and the upper bound of the credits $Q$ be set to 50000. In addition, the distance limit of BEVs was from 30 to 45.

Numerical results in Figures 3 and 4 reveal that the number of BEVs' users increases as the distance limit is getting longer. In more details, when the distance limit is 30, there is no BEV in the traffic network since no feasible path of BEVs exists. On the other hand, when the distance limit is 39, all users choose BEVs in the traffic network since the unit cost of BEVs is lower than the unit cost of GVs and there is no emission of BEVs in the network. When the distance limit is 31, on path 4, all users give up using GVs but other paths are not because only the O-D pair (1, 2) has a feasible path for BEVs; see Figure 4(a) and Table 2. Since the shortest path of the O-D (4, 2) is index 15, the users of the O-D pair (4, 2) select BEVs when the distance limit of BEVs is no less than 35. From this experiment, promoting the BEVs is utilized by expending the longest traveling distance of BEVs. This means that increasing the number of BEVs by technological progress is feasible.

In the second experiment, let the distance limit of BEVs be equal to 45. Let the operating cost functions of BEVs and GVs on link $a$ be expressed as $c_{a_e}(x_{a_e} + x_{a_g}) = 0.005 + \beta_e(x_{a_e} + x_{a_g})^2$ and $c_{a_g}(x_{a_e} + x_{a_g}) = 1$, respectively. Furthermore, let $\beta_e = 0.00001, 0.0001, 0.001, 0.01, 0.1, 1$ in model

Table 1: Algorithm frame.

*Step 1.* Initialization $x_0$, tolerances $\varepsilon_{\min}$, and maximum iterations $k_{\max}$.
*Step 2.* Find the feasible paths with distance limit $D$ of BEVs and feasible paths of GVs.
*Step 3.* Input $A$, $b$, $Aeq$, $beq$, $lb$, $ub$, $nonlcon$, and the constraints.
*Step 4.* Set $\omega_0 = \left\{\omega^a_{k_g 0}, \omega^a_{k_e 0} : k_g \in K, k_e \in K, a \in A\right\}$ such that the strict inequalities in (16)–(18) hold. Let $\mathbf{H}_0$ be the unit matrix.
*Step 5.* Solve the approximation quadratic programming problem

$$\min \quad 1/2 d^T \mathbf{H}_n d + \nabla g(\omega_n)^T d$$
$$\text{s.t.} \quad (16)\text{–}(18)$$

to get Lagrange multiplier $\lambda_n$ and the search direction $d_{(n)} = \omega_{(n+1)} - \omega_{(n)}$. Here, $g(\omega_n)$ is the objective function in model (11)–(18).
*Step 6.* Calculate the new iteration point $\omega_{n+1} = \omega_n + \alpha_n d_n$, where $\alpha_n$ is a stepsize calculated by one-dimensional search.
*Step 7.* If $\|\omega_{n+1} - \omega_n\| \le \varepsilon$, stop. Otherwise, go to step 8.
*Step 8.* Update the Hessian matrix $\mathbf{H}_n$ by BFGS algorithm, let $n = n + 1$, and go to Step 4.

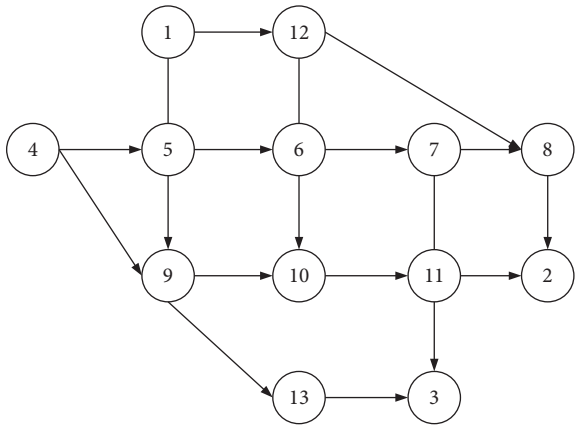



Figure 2: Nguyen–Dupuis' network.

Table 2: Analysis results for Nguyen–Dupuis' network.

| O-D | Route | Node sequence | Length |
|---|---|---|---|
| (1, 2) | 1 | 1-5-6-7-8-2 | 33 |
| | 2 | 1-5-6-7-11-2 | 38 |
| | 3 | 1-5-6-10-11-2 | 43 |
| | 4 | 1-12-8-2 | 31 |
| | 5 | 1-5-9-10-11-2 | 48 |
| | 6 | 1-12-6-7-8-2 | 46 |
| | 7 | 1-12-6-7-11-2 | 51 |
| | 8 | 1-12-6-10-11-2 | 56 |
| (1, 3) | 9 | 1-5-6-7-11-3 | 39 |
| | 10 | 1-5-6-10-11-3 | 44 |
| | 11 | 1-5-9-13-3 | 39 |
| | 12 | 1-5-9-10-11-3 | 49 |
| | 13 | 1-12-6-7-11-3 | 52 |
| | 14 | 1-12-6-10-11-3 | 57 |
| (4, 2) | 15 | 4-5-6-7-8-2 | 35 |
| | 16 | 4-5-6-7-11-2 | 40 |
| | 17 | 4-5-6-10-11-2 | 45 |
| | 18 | 4-9-10-11-2 | 45 |
| | 19 | 4-5-9-10-11-2 | 50 |
| | 20 | 4-5-6-7-11-3 | 41 |
| (4, 3) | 21 | 4-5-9-13-3 | 41 |
| | 22 | 4-9-13-3 | 36 |
| | 23 | 4-5-6-10-11-3 | 46 |
| | 24 | 4-5-9-10-11-3 | 51 |
| | 25 | 4-9-10-11-3 | 46 |

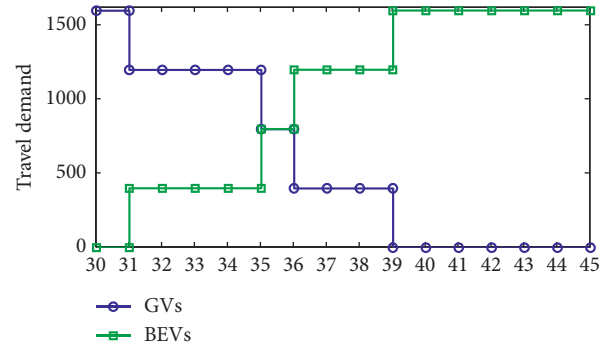



Figure 3: Changing distance limit of BEVs of all O-D pairs in Nguyen–Dupuis' network.

(11)–(18). The results displayed in Figures 5 and 6(a) indicate that the number of BEVs' users reduces with $\beta$ increasing. In detail, the first reduction is the O-D pair (4, 2) with $\beta = 0.001$. When $\beta_e = 0.1$, all users choose GVs; that is, there are no BEVs in the traffic network. This experiment demonstrates that we can increase the number of BEVs' users by reducing the operating cost such as cutting down the price of BEVs or increasing subsidies for BEVs.

In the third experiment, let the distance limit of BEVs be also equal to 45. Let the operating cost functions of BEVs and GVs on link $a$ be expressed as $c_{a_e} = 1$ and $c_{a_g}(x_{a_e} + x_{a_g}) = 0.005 + \beta_g (x_{a_e} + x_{a_g})^2$, respectively. Furthermore, let $\beta_g = 0.00001, 0.0001, 0.001, 0.01, 0.1, 1$ in model (11)–(18). The traffic flows of BEVs and GVs on each optimization feasible path for each O-D pair are displayed in Figure 7. The total traffic flow of all O-D pairs is exhibited in Figure 6(b). The results show that the amount of BEVs' users is increasing with $\beta_g$ increasing. In particular, when $\beta_g = 0.001$, the number of GVs' users reduces in all feasible paths of all O-D pairs. In the first experiment, the quantity of users is changing only between the O-D pairs (4, 2) when $\beta_e = 0.001$. The reason for this difference may be that the emission contains GVs only.

In the fourth experiment, model (11)–(18) was considered. In the model, the upper bound of total credit schemes $Q$ is set to be 0, 100, 500, 1000, 5000, 10000, 30000, 50000, 100000, and 500000 and there is no distance limit of BEVs.

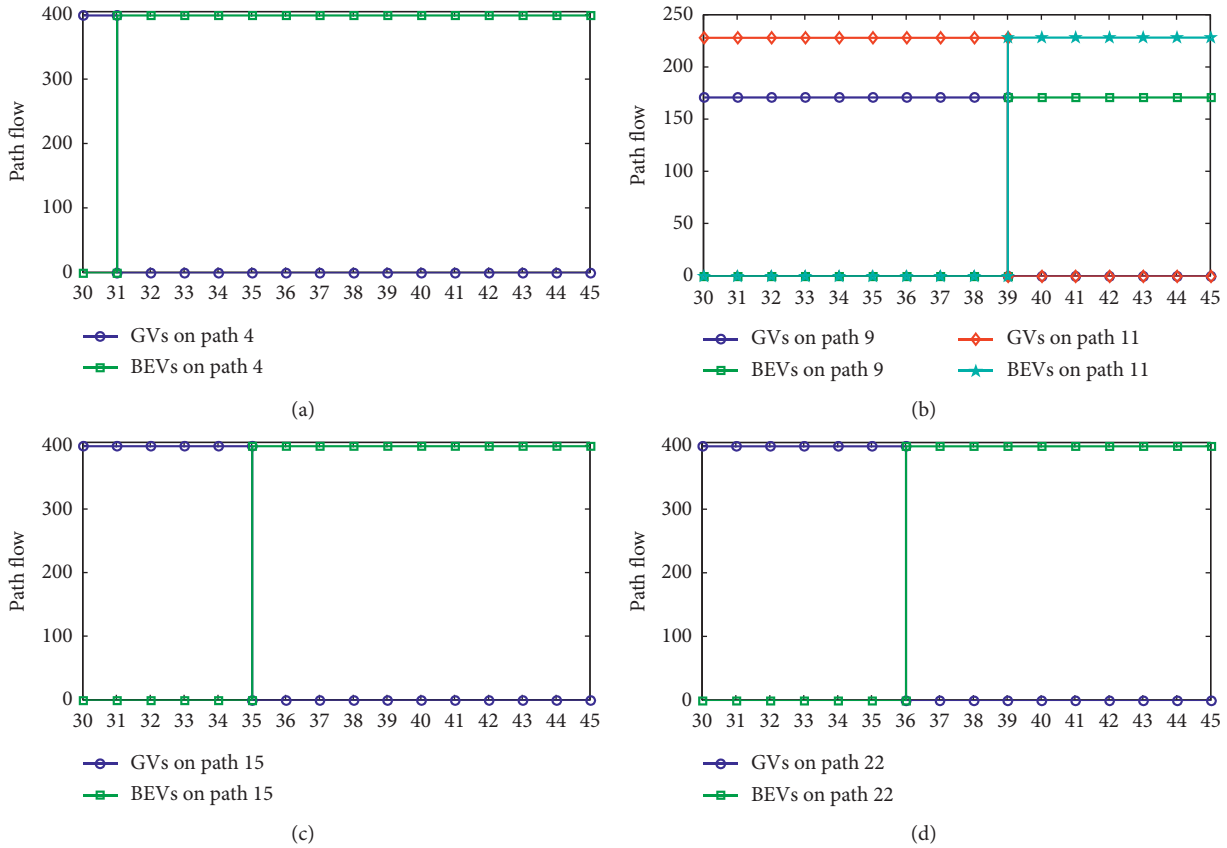



Figure 4: Changing distance limit of BEVs of each O-D pair in Nguyen–Dupuis' network. (a) O-D pair (1, 2). (b) O-D pair (1, 3). (c) O-D pair (4, 2). (d) O-D pair (4, 3).

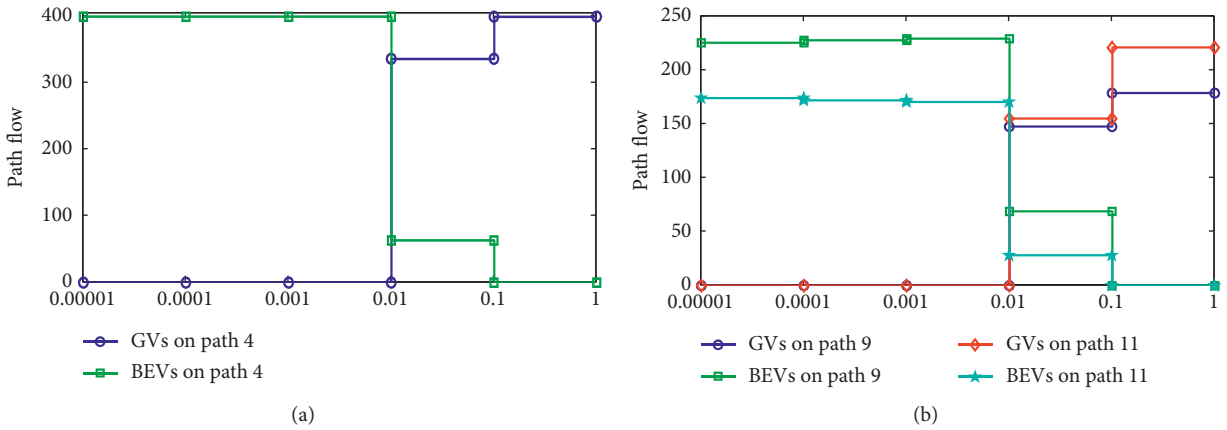



Figure 5: Continued.

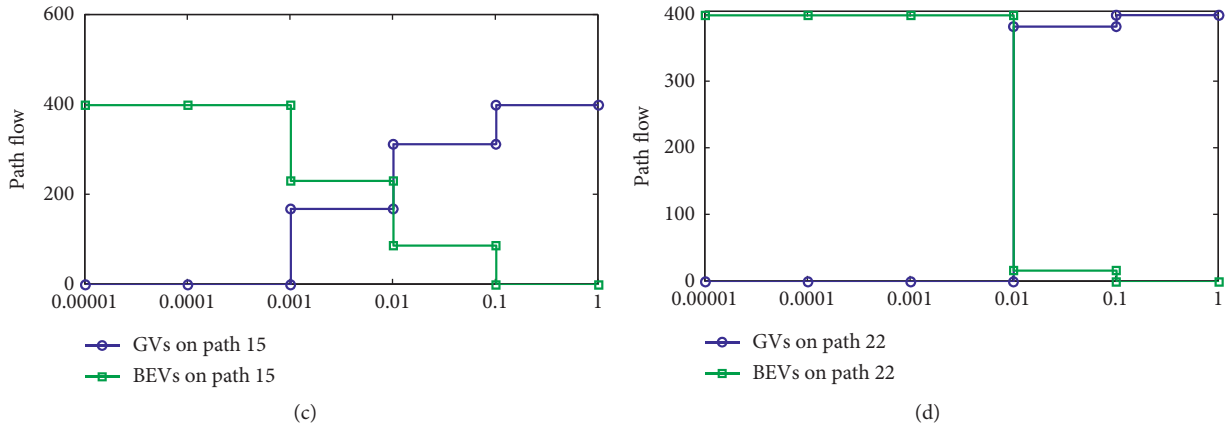



Figure 5: Changing unit operating cost of BEVs of each O-D pair in Nguyen–Dupuis' network. (a) O-D pair (1, 2). (b) O-D pair (1, 3). (c) O-D pair (4, 2). (d) O-D pair (4, 3).

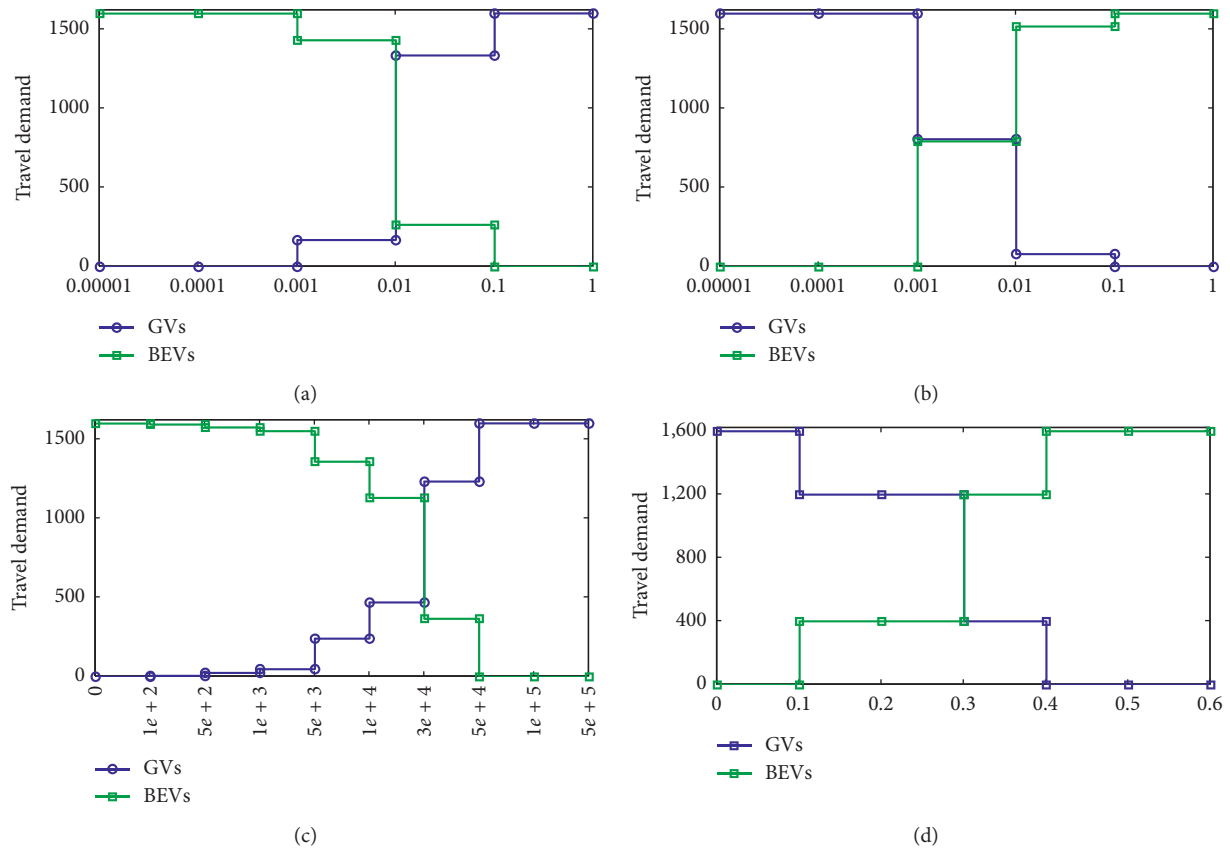



Figure 6: Total traffic flow. (a) Unit operating cost of EVs. (b) Unit operating cost of GVs. (c) Unit credit cost of GVs. (d) Unit price $\lambda_a$ of GVs.

The operation cost functions of BEVs and GVs were expressed as $c_{a_e} = 2l_a(x_{a_e} + x_{a_g})$ and $c_{a_g} = l_a(x_{a_e} + x_{a_g})$. The results stated in Figures 8 and 6(c) show that $Q = 0$, which means that the traffic network only includes BEVs of all O-D pairs. This experiment indicates that the volume of GVs' users increases with the upper bound of total credits enlarging in a proper range. In detail, when $Q$ is a large number such as $Q = 50000, 100000, 500000$, the traffic network only embodies GVs in Figures 5 and 6(b). In addition, from Figures 8 and 6(c), the number of GVs in the traffic network

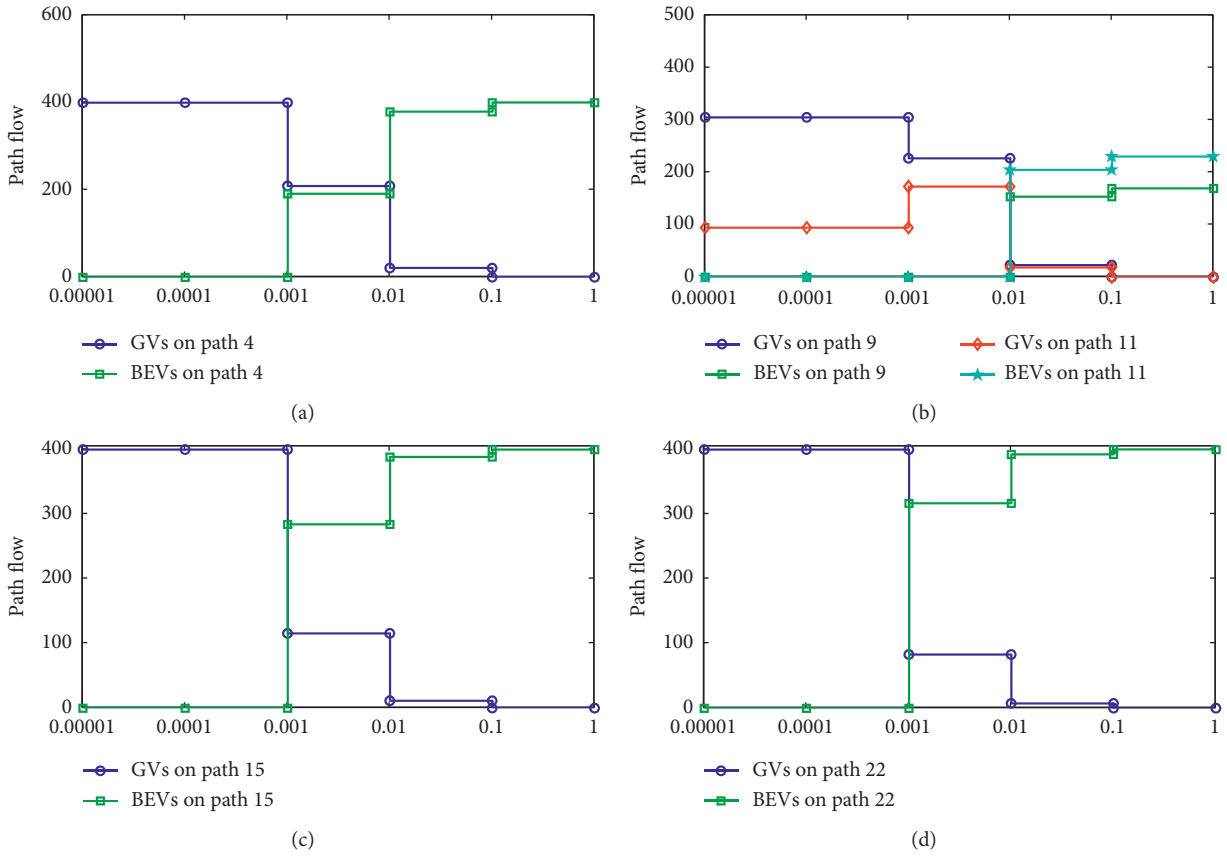



Figure 7: Changing unit operating cost of GVs in Nguyen–Dupuis' network. (a) O-D pair (1, 2). (b) O-D pair (1, 3). (c) O-D pair (4, 2). (d) O-D pair (4, 3).

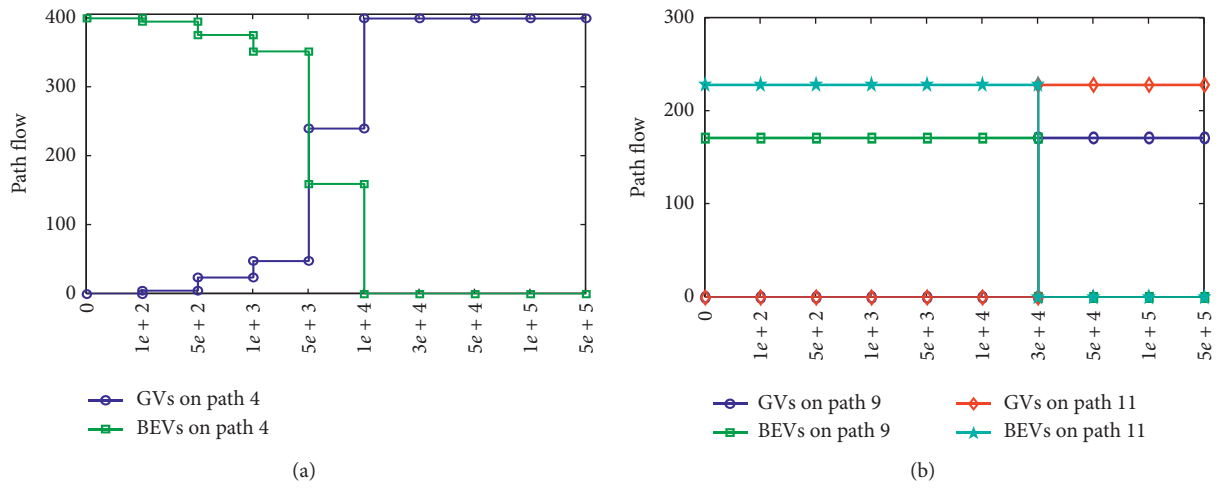



Figure 8: Continued.

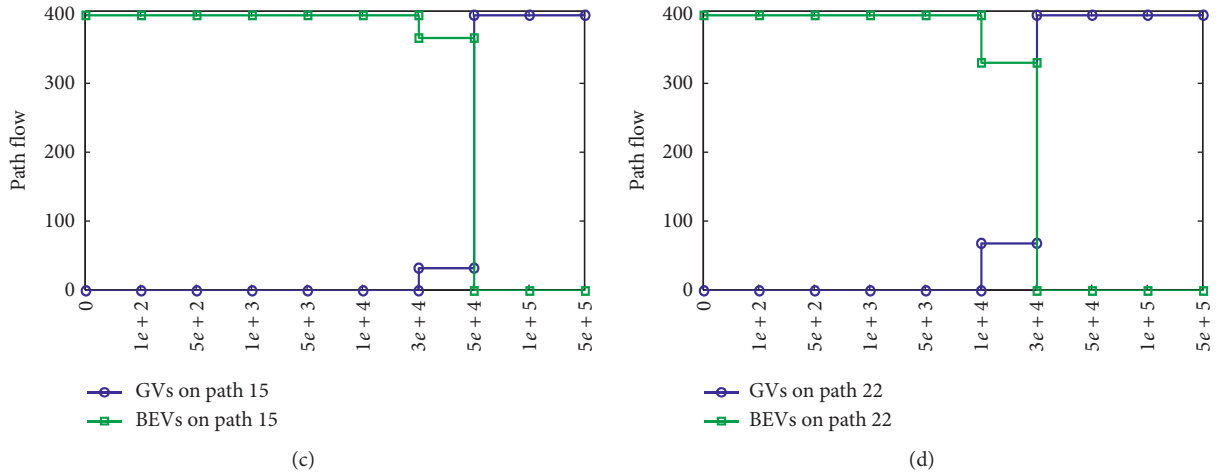



Figure 8: Changing total credits of GVs in Nguyen–Dupuis' network. (a) O-D pair (1, 2). (b) O-D pair (1, 3). (c) O-D pair (4, 2). (d) O-D pair (4, 3).

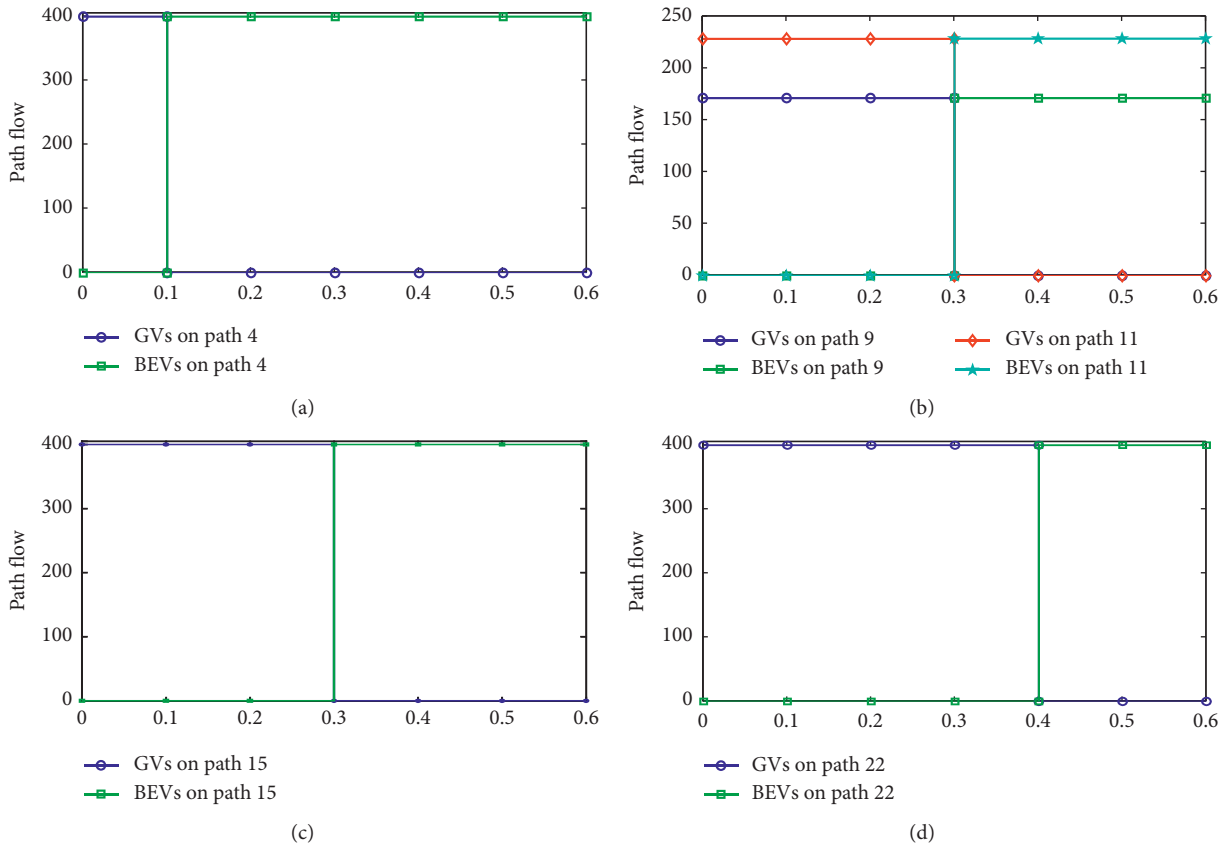



Figure 9: Changing $\lambda_a$ of GVs in Nguyen–Dupuis' network. (a) O-D pair (1, 2). (b) O-D pair (1, 3). (c) O-D pair (4, 2). (d) O-D pair (4, 3).

can be adjusted by modifying total credits for optimization model (11)–(18). This is an interesting approach to promote the number of BEVs.

In the last experiment, the unit price $\lambda_a$ on link $a$ of emissions of GVs was tested. The operation cost functions of BEVs and GVs were formulated as $c_{a_e} = 0.5 l_a (x_{a_e} + x_{a_g})$ and $c_{a_g} = 0.25 l_a (x_{a_e} + x_{a_g})$. The distance limit of BEVs was set to be 45. The upper bound of credits was equal to 500000. The unit price $\lambda_a$ of emissions of GVs was set to be 0, 0.1, 0.2, 0.3, 0.4, 0.5, and 0.6. The results are given in Figures 6(d) and 9. This experiment manifests that the volume of GVs' users decreases with the unit price of emissions rising. Thus, for accelerating the utilization of BEVs, raising the unit price of emissions may be employed.

## 5. Conclusions

We studied two types of vehicles (BEVs and GVs) in urban road network. We presented two models for the traffic flow assignment and emissions problems with the credits constraint. One of them is to minimize the sum of traffic cost and emissions cost by adjusting the total credits for BEVs and GVs in the traffic network. Another tries to adjust the volume of BEVs in the traffic network by changing the upper bound of total credits for GVs. Furthermore, we gave a simple algorithm to solve the models.

In our experiments, we first checked the influence of distance limits of BEVs and observed that expanding the distance limit of BEVs may promote the utilization of BEVs by technical progress. Then, we investigated the effect of changing unit operating cost of BEVs and GVs. It was observed that reducing the unit operating cost of BEVs may promote the utilization of BEVs and the total GVs in the traffic network may be controlled by changing the upper bound of total credits of GVs in model (11)–(18); that is, we may give a smaller upper bound of the volume of credits of GVs to promote the employment of BEVs in the traffic network. Finally, the unit price of emissions of GVs was also investigated and it was observed that the number of GVs is reducing with increasing the unit price of emission of GVs; in other words, diminishing the amount of GVs in the traffic network may be realized by raising the unit price of emissions of GVs.

As a future research direction, dynamic cases may be considered. The cases that some uncertainties (weather, traffic accidents, etc.) occur in the network may also be considered. Biobjective models or bilevel optimization models will be studied and the two-way street traffic network will also be considered. In addition, we will study more details about the anxiety cost function. The multiple charges of BEVs will also be considered in the future.

## Data Availability

The datasets are described in the experiments section of the paper.

## Conflicts of Interest

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

## Acknowledgments

The authors are grateful to Professor Guihua Lin who gave some help for the revision. This work was supported in part by the NSFC (no. 11901320).

## References

[1] W. Feng and M. Figliozzi, "An economic and technological analysis of the key factors affecting the competitiveness of electric commercial vehicles: a case study from the USA market," *Transportation Research Part C: Emerging Technologies*, vol. 26, pp. 135–145, 2013.

[2] L. M. Gardner, M. Duell, and S. T. Waller, "A framework for evaluating the role of electric vehicles in transportation network infrastructure under travel demand variability," *Transportation Research Part A: Policy and Practice*, vol. 49, pp. 76–90, 2013.

[3] F. He, D. Wu, Y. Yin, and Y. Guan, "Optimal deployment of public charging stations for plug-in hybrid electric vehicles," *Transportation Research Part B: Methodological*, vol. 47, pp. 87–101, 2013.

[4] M. A. Tamor, C. Gearhart, and C. Soto, "A statistical approach to estimating acceptance of electric vehicles and electrification of personal transportation," *Transportation Research Part C: Emerging Technologies*, vol. 26, pp. 125–134, 2013.

[5] M. Xu, Q. Meng, and K. Liu, "Network user equilibrium problems for the mixed battery electric vehicles and gasoline vehicles subject to battery swapping stations and road grade constraints," *Transportation Research Part B: Methodological*, vol. 99, pp. 138–166, 2017.

[6] J. Addison, "National academies predicts 13 to 40 million plug-ins by 2030," vol. 12, pp. 9–28, 2011, http://www.cleanfleetreport.com/plug-in-hybrids/national-academies-40-million-plug-ins.

[7] T. H. Bradley and A. A. Frank, "Design, demonstrations and sustainability impact assessments for plug-in hybrid electric vehicles," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 1, pp. 115–128, 2009.

[8] Z. Lin and D. L. Greene, "Promoting the market for plug-in hybrid and battery electric vehicles," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2252, no. 1, pp. 49–56, 2011.

[9] Y. Nie and M. Ghamami, "A corridor-centric approach to planning electric vehicle charging infrastructure," *Transportation Research Part B: Methodological*, vol. 57, pp. 172–190, 2013.

[10] N. S. Pearre, W. Kempton, R. L. Guensler, and V. V. Elango, "Electric vehicles: how much range is required for a day's driving?" *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1171–1184, 2011.

[11] K. S. Kurani, R. R. Heffner, and T. Turrentine, "Driving plug-in hybrid electric vehicles," Reports from US Drivers of HEVs Converted to PHEVs, Circa 2006-07, Institute of Transportation Studies, Berkeley, CA, USA, 2008.

[12] H. Yang and X. Wang, "Managing network mobility with tradable credits," *Transportation Research Part B: Methodological*, vol. 45, no. 3, pp. 580–594, 2011.

[13] Y. Nie, "Transaction costs and tradable mobility credits," *Transportation Research Part B: Methodological*, vol. 46, no. 1, pp. 189–203, 2012.

[14] X. Wang, H. Yang, D. Zhu, and C. Li, "Tradable travel credits for congestion management with heterogeneous users," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 2, pp. 426–437, 2012.

[15] F. Xiao, Z. Qian, and H. M. Zhang, "Managing bottleneck congestion with tradable credits," *Transportation Research Part B: Methodological*, vol. 56, pp. 1–14, 2013.

[16] N. Shirmohammadi, M. Zangui, Y. Yin, and Y. Nie, "Analysis and design of tradable credit schemes under uncertainty," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2333, no. 1, pp. 27–36, 2013.

[17] D. L. Zhu, H. Yang, C. M. Li, and X. L. Wang, "Properties of the multiclass traffic network equilibria under a tradable credit scheme," *Transportation Science*, vol. 49, no. 3, pp. 519–534, 2014.

[18] X. Wang, H. Yang, D. Han, and W. Liu, "Trial and error method for optimal tradable credit schemes: the network

case," *Journal of Advanced Transportation*, vol. 48, no. 6, pp. 685–700, 2014.

[19] F. Han and L. Cheng, "Stochastic user equilibrium model with a tradable credit scheme and application in maximizing network reserve capacity," *Engineering Optimization*, vol. 49, no. 4, pp. 549–564, 2016.

[20] H. Wang and X. Zhang, "Joint implementation of tradable credit and road pricing in public-private partnership networks considering mixed equilibrium behaviors," *Transportation Research Part E: Logistics and Transportation Review*, vol. 94, pp. 158–170, 2016.

[21] G. Wang, Y. Li, M. Xu, and Z. Gao, "Operating a public-private mixed road network via determining tradable credits and road tolls: an equilibrium problem with equilibrium constraints approach," *International Journal of Sustainable Transportation*, pp. 1–10, 2019.

[22] Y. Bao, Z. Gao, H. Yang, M. Xu, and G. Wang, "Private financing and mobility management of road network with tradable credits," *Transportation Research Part A: Policy and Practice*, vol. 97, pp. 158–176, 2017.

[23] N. Dogterom, D. Ettema, and M. Dijst, "Tradable credits for managing car travel: a review of empirical research and relevant behavioural approaches," *Transport Reviews*, vol. 37, no. 3, pp. 322–343, 2016.

[24] M. Miralinaghi and S. Peeta, "Multi-period equilibrium modeling planning framework for tradable credit schemes," *Transportation Research Part E: Logistics and Transportation Review*, vol. 93, pp. 177–198, 2016.

[25] M. Miralinaghi and S. Peeta, "A multi-period tradable credit scheme incorporating interest rate and traveler value-of-time heterogeneity to manage traffic system emissions," *Frontiers in Built Environment*, vol. 4, p. 33, 2018.

[26] N. Shirmohammadi and Y. Yin, "Tradable credit scheme to control bottleneck queue length," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2561, no. 1, pp. 53–63, 2016.

[27] S. Lahlou and L. Wynter, "A nash equilibrium formulation of a tradable credits scheme for incentivizing transport choices: from next-generation public transport mode choice to HOT lanes," *Transportation Research Part B: Methodological*, vol. 101, pp. 185–212, 2017.

[28] T. Akamatsu and K. Wada, "Tradable network permits: a new scheme for the most efficient use of network capacity," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 178–195, 2017.

[29] G. Wang, Z. Gao, and M. Xu, "Integrating link-based discrete credit charging scheme into discrete network design problem," *European Journal of Operational Research*, vol. 272, no. 1, pp. 176–187, 2018.

[30] N. Dogterom, "Willingness to change car use under a tradable driving credits scheme: a comparison between Beijing and the Netherlands," *Journal of Transport and Land Use*, vol. 11, no. 1, pp. 499–518, 2018.

[31] R.-Y. Guo, H.-J. Huang, and H. Yang, "Tradable credit scheme for control of evolutionary traffic flows to system optimum: model and its convergence," *Networks and Spatial Economics*, vol. 19, no. 3, pp. 833–868, 2018.

[32] J. Lv, F. Zong, M. Tang, and P. Yu, "Investigating a tradable credit scheme with the system optimum theory," *Advances in Mechanical Engineering*, vol. 11, no. 5, 2019.

[33] Y. Tian, Y.-C. Chiu, and J. Sun, "Understanding behavioral effects of tradable mobility credit scheme: an experimental economics approach," *Transport Policy*, vol. 81, pp. 1–11, 2019.

[34] Z. Lian, X. Liu, and W. Fan, "Does driving-day-based tradable credit scheme outperform license plate rationing? Examination considering transaction cost," *Journal of Modern Transportation*, vol. 27, no. 3, pp. 198–210, 2019.

[35] Y. Bao, E. T. Verhoef, and P. Koster, "Regulating dynamic congestion externalities with tradable credit schemes: does a unique equilibrium exist?" *Transportation Research Part B: Methodological*, vol. 127, pp. 225–236, 2019.

[36] G. Gao, X. Liu, H. Sun et al., "Marginal cost pricing analysis on tradable credits in traffic engineering," *Mathematical Problems in Engineering*, vol. 2019, Article ID 8461395, 10 pages, 2019.

[37] F. Xiao, J. Long, L. Li, G. Kou, and Y. Nie, "Promoting social equity with cyclic tradable credits," *Transportation Research Part B: Methodological*, vol. 121, pp. 56–73, 2019.

[38] M. Miralinaghi, S. Peeta, X. He, and S. V. Ukkusuri, "Managing morning commute congestion with a tradable credit scheme under commuter heterogeneity and market loss aversion behavior," *Transportmetrica B: Transport Dynamics*, vol. 7, no. 1, pp. 1780–1808, 2019.

[39] M. Miralinaghi and S. Peeta, "Promoting zero-emissions vehicles using robust multi-period tradable credit scheme," *Transportation Research Part D: Transport and Environment*, vol. 75, pp. 265–285, 2019.

[40] A. De Palma and R. Lindsey, "Tradable permit schemes for congestible facilities with uncertain supply and demand," *Economics of Transportation*, vol. 21, Article ID 100149, 2020.

[41] F. Han, X.-m. Zhao, and L. Cheng, "Traffic assignment problem under tradable credit scheme in a bi-modal stochastic transportation network: a cumulative prospect theory approach," *Journal of Central South University*, vol. 27, no. 1, pp. 180–197, 2020.

[42] M. Miralinaghi and S. Peeta, "Design of a multiperiod tradable credit scheme under vehicular emissions caps and traveler heterogeneity in future credit price perception," *Journal of Infrastructure Systems*, vol. 26, no. 3, Article ID 04020030, 2020.

[43] Y. Nie and Y. Yin, "Managing rush hour travel choices with tradable credit scheme," *Transportation Research Part B: Methodological*, vol. 50, pp. 1–19, 2013.

[44] B. Liu and C. Huang, "Alternative tradable credit scheme on general transportation network," *Wuhan University Journal of Natural Sciences*, vol. 19, no. 3, pp. 193–198, 2014.

[45] W. Zhu, S. Ma, and J. Tian, "Optimizing congestion and emissions via tradable credit charge and reward scheme without initial credit allocations," *Physica A: Statistical Mechanics and Its Applications*, vol. 465, pp. 438–448, 2017.

[46] Y. Yin and S. Lawphongpanich, "Internalizing emission externality on road networks," *Transportation Research Part D: Transport and Environment*, vol. 11, no. 4, pp. 292–301, 2006.

[47] L. Chen and H. Yang, "Managing congestion and emissions in road networks with tolls and rebates," *Transportation Research Part B: Methodological*, vol. 46, no. 8, pp. 933–948, 2012.

[48] H. M. Abdul Aziz and S. V. Ukkusuri, "Tradable emissions credits for personal travel: a market-based approach to achieve air quality standards," *International Journal of Advances in Engineering Sciences and Applied Mathematics*, vol. 5, no. 2-3, pp. 145–157, 2013.

[49] S. Grant-Muller and M. Xu, "The role of tradable credit schemes in road traffic congestion management," *Transport Reviews*, vol. 34, no. 2, pp. 128–149, 2014.

[50] Y. Li, S. V. Ukkusuri, and J. Fan, "Managing congestion and emissions in transportation networks with dynamic carbon credit charge scheme," *Computers & Operations Research*, vol. 99, pp. 90–108, 2018.

[51] N. Jiang, C. Xie, and S. T. Waller, "Path-constrained traffic assignment," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2283, no. 1, pp. 25–33, 2012.

[52] N. Jiang and C. Xie, "Computing and analyzing equilibrium network flows of gasoline and electric vehicles," in *Proceedings of the Transportation Research Board 92nd Annual Meeting*, 2013.

[53] G.-H. Lin, Y. Hu, and Y.-Y. Zou, "A mixed-mode traffic assignment model with new time-flow impedance function," *International Journal of Modern Physics B*, vol. 32, no. 3, p. 1850017, 2018.

[54] Y.-Y. Zou, X.-G. Xu, Y. Hu, and G.-H. Lin, "Traffic flow assignment model with modified impedance function," *KSCE Journal of Civil Engineering*, vol. 22, no. 10, pp. 4116–4126, 2018.

[55] J. Ke, X. Cen, H. Yang, X. Chen, and J. Ye, "Modelling drivers' working and recharging schedules in a ride-sourcing market with electric vehicles and gasoline vehicles," *Transportation Research Part E: Logistics and Transportation Review*, vol. 125, pp. 160–180, 2019.

[56] A. F. Jensen, T. K. Rasmussen, and C. G. Prato, "A route choice model for capturing driver preferences when driving electric and conventional vehicles," *Sustainability*, vol. 12, no. 3, p. 1149, 2020.

[57] N. Jiang and C. Xie, "Computing and analyzing mixed equilibrium network flows with gasoline and electric vehicles," *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, pp. 626–641, 2014.

[58] C. Mensink, I. De Vlieger, and J. Nys, "An urban transport emission model for the Antwerp area," *Atmospheric Environment*, vol. 34, no. 27, pp. 4595–4602, 2000.

[59] S. Nguyen and C. Dupuis, "An efficient method for computing traffic equilibria in networks with asymmetric transportation costs," *Transportation Science*, vol. 18, no. 2, pp. 185–202, 1984.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8842730, 21 pages
https://doi.org/10.1155/2020/8842730

WILEY | Hindawi

*Research Article*

# Fleet Management for HDVs and CAVs on Highway in Dense Fog Environment

**Bowen Gong,[1] Ruixin Wei,[1] Dayong Wu,[2] and Ciyun Lin [1,3]**

[1]*Department of Traffic Information and Control Engineering, Jilin University, Changchun 130022, China*
[2]*Texas A&M Transportation Institute, Texas A&M University, College Station, TX 77843, USA*
[3]*Jilin Engineering Research Center for ITS, Changchun 130022, China*

Correspondence should be addressed to Ciyun Lin; linciyun@jlu.edu.cn

Received 22 May 2020; Revised 5 July 2020; Accepted 29 July 2020; Published 14 August 2020

Academic Editor: Chunjiao Dong



Adverse weather conditions have a significant impairment on the safety, mobility, and efficiency of highway networks. Dense fog is considered the most dangerous within the adverse weather conditions. As to improve the traffic flow throughput and driving safety in dense fog weather condition on highway, this paper uses a mathematical modeling method to study and control the fleet mixed with human-driven vehicles (HDVs) and connected automatic vehicles (CAVs) in dense fog environment on highway based on distributed model predictive control algorithm (DMPC), along with considering the car-following behavior of HDVs driver based on cellular automatic (CA) model. It aims to provide a feasible solution for controlling the mixed flow of HDVs and CAVs more safely, accurately, and stably and then potentially to improve the mobility and efficiency of highway networks in adverse weather conditions, especially in dense fog environment. This paper explores the modeling framework of the fleet management for HDVs and CAVs, including the state space model of CAVs, the car-following model of HDVs, distributed model predictive control for the fleet, and the fleet stability analysis. The state space model is proposed to identify the status of the feet in the global state. The car-following model is proposed to simulate the driver behavior in the fleet in local. The DMPC-based model is proposed to optimize rolling of the fleet. Finally, this paper used the Lyapunov stability principle to analyze and prove the stability of the fleet in dense fog environment. Finally, numerical experiments were performed in MATLAB to verify the effectiveness of the proposed model. The results showed that the proposed fleet control model has the ability of local asymptotic stability and global nonstrict string stability.

## 1. Introduction

Adverse weather conditions have a significant impact on the safety, mobility, and efficiency of highway networks [1]. Based on the statistical data, weather contributed to about 20% of traffic accidents, 38.3% of traffic congestion, 23% of all nonreoccurring delays and caused billions dollars' loss by closed highways, vehicle delays, and traffic accidents [2]. The consequent adverse impact on the safety and mobility of highway networks makes it important to research and develop new and more efficient methods to address highway management and operation problems during adverse weather conditions [3]. Xu [4] explored crashes under different weather conditions on the highway and found that adverse weather conditions can lead to dangerous driving conditions and greatly increase the crash rate. Moreover, due to its limited visibility and accident susceptibility, dense fog is considered the most dangerous within the adverse weather conditions [4].

Besides visibility impairments in foggy weather conditions, the ability of risk perception is reduced as the driver inability to judge the driving state and safe distance accurately. Therefore, the driver tends to continue driving until the following distance is equal to or less than the safety distance. After realizing the crash-prone traffic condition, the driver will operate the vehicle with sequential rushed acceleration or nasty deceleration to keep out of traffic accidents [5]. Furthermore, the researchers found that most drivers prefer to drive at high speeds when they cannot see the vehicle in front. In this case, even if the driver finds the vehicle in front of the visibility boundary and braking timely, it is difficult to avoid a crash [4, 6].

In addition, the dense fog weather condition will greatly reduce the driver's ability of identification traffic conditions. For example, when the visibility is less than 50 m, the driver's visual search focus will increase by 23.6%, and the scope and efficiency visual selective attention will decrease [7]. In these situations, the combination of speed perception, speed feedback, driving performance feedback, and other capabilities that support the driver to make vehicle operation decisions is not always in the best safety interest [6]. Furthermore, the individual differences among drivers must also be considered in the driving conditions, such as differences in age, gender, and psychological quality.

The internal and external adverse impact on driver behavior in dense fog environment makes it becoming the most dangerous type within adverse weather conditions [8]. Wang [9] analyzed 1,513,792 traffic accidents and found that the number of fatal traffic accidents caused by dense fog is 35 times that of clear weather. Traffic safety concerns in dense fog on the highway were intense due to the fatal traffic accidents, as the individuals and groups are increasingly concerned with the adverse weather condition impact on highway travelling [10–12].

Based on the state-of-the-art in literatures and researches for driving safely in dense fog weather conditions, the proposed methods can be divided into macrolevel and microlevel. In respect of macrolevel mode, it mainly focuses on risk prediction under dense fog weather conditions based on historical data and then makes up corresponding risk management and control approaches such as setting warning flags and installing security infrastructure. Wu [13] analyzed the effects of real-time warning systems on driving under fog conditions by comparing the effectiveness of beacons and dynamic message signs in foggy areas. Zhai [14] used the historical traffic collision data, traffic flow data, and dense fog conditions to set up a collision risk prediction model to predict the collision risk under specific visibility. Ahmed [8] predicted the risk of traffic accidents on highways near the airport in dense fog weather conditions based on variables such as airport weather data, historical crash data, and road characteristics. Hassan [15] used detectors and radar sensors to receive real-time traffic flow data, which are used to predict the likelihood of traffic accidents in low-visibility situations, and then conduct proper traffic management 5–15 minutes before the highway may collide. In this study, the probability of correctly identifying a collision reached 69%. Winkle [16] proposed a systematic safety analysis framework that combines the spatial analysis function in ArcGIS with a clustering model to select areas on highways that are prone to fogging, thus providing guidance for highway safety strategies and active traffic management.

In respect of microlevel mode, the studies mainly focused on the vehicle itself to provide anticollision function. For example, a networked vehicle collision warning system (CWS) was designed on the basis of connected vehicles. This system used real-time data to predict the collision risk of vehicles and promptly alert drivers to improve the rationality and safety of driver's driving operations [17]. Li [18] developed a control strategy with a variable speed limit (VSL) to reduce the risk of secondary collision during adverse weather. In addition, self-driving vehicles were tested on-site in areas with poor visibility. It was expected that self-driving cars that adapt to various scenarios in the real world would be developed soon [16]. To the best of our knowledge, the current research is aimed at independent human-driven vehicles (HDVs) accident-related behavior analysis or connected automatic vehicles (CAVs) control strategy in dense fog weather conditions, respectively.

Although interest in CAVs has been growing exponentially in recent years, with increasing levels of automation being introduced to newer vehicles, many new technologies are being developed to intelligent infrastructure-based equipment on the smart highway construction [19]. Due to the wide range of potential applications, the objectives, and framework of control, CAVs will significantly enhance the safety, mobility, and efficiency of highway networks, especially in adverse weather conditions. However, the transition to CAVs is going to be a gradual process. It is expected to see a mixed flow of HDVs and CAVs for the next 50 years [20, 21].

A mixed flow of HDVs and CAVs in the traffic flow will lead to a highly heterogeneous traffic management and control environment. The dynamics, safety, and mobility of traffic flow will change due to the mutual interference caused by the performance differences of HDVs and CAVs, especially in adverse weather conditions. Therefore, how to manage and control the mixed traffic flow to improve the safety, mobility, and efficiency will be a consistent issue before the CAVs society fully realize. As the CAVs will be first practically applied in the highway environment, based on the research and development programs of governments and enterprises [22], this paper attempts to propose a fleet control model to manage the mixed flow of CAVs and HDVs passing through the dense fog environment safely on the highway. The main work of this paper is as follows:

(1) A cellular automatic- (CA-) based HDV following model is proposed to analyze the motion characteristics of HDVs in the mixed flow of HDVs and CAVs, along with considering the following behavior of drivers in dense fog weather conditions.

(2) A fleet control model based on the distributed model predictive control (DMPC) algorithm is proposed by using a mathematical modeling method to manage the HDVs and CAVs within the fleet. It attempts to provide a feasible solution for controlling the fleet to improve the mobility and efficiency of highway network in adverse weather conditions, especially in dense fog environment.

In the aspect of dense fog in the highway, the previous researches mostly focused on the association between the driver behavior and traffic accidents by considering the driver's physiology and psychology and environmental factors [12, 23, 24]. This paper focuses on modeling the driver behavior in the fleet mixed with HDVs and CAVs based on the Nagel–Schreckenberg (NaScH) cellular automatic (CA) model, which has advantages in describing the complex behavior and simulating the characteristics of traffic flow under various scenarios and traffic conditions [25–28].

In the aspect of traffic flow control in dense fog environment, the researches were aimed at independent HDVs or CAVs, respectively. This paper aims to manage the fleet mixed with HDVs and CAVs based on DMPC, which is widely used in control engineering and has advantages in dealing with the uncertainty during dynamical optimization and controlling operation, particularly in the area of vehicle dynamics and motion control [29–31].

The rest of the paper is organized as follows. Section 2 presents the assumptions and scenarios of the proposed model and formulates the fleet control model for HDVs and CAVs in dense fog weather conditions. Section 3 analyzes the feasibility and stability of the fleet control model in dense fog environment. The numerical experiments and discussions are presented in Section 4. Section 5 concludes this article with a summary of the contributions and the limitations of the proposed model, as well as the perspectives on future work.

## 2. Problem Description and Model Formulation

*2.1. Problem Description.* Given weather and traffic sensors installed along the side of the highway networks, the stability and time delay of communication among vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) are important factors that influence the safety and robustness in fleet management and control. The highway road alignment, gradient, and surface conditions also influence the mobility and acceleration performance of vehicles in adverse weather conditions. The driver's ability and intention play an important in car-following modeling and characteristics analysis in mixed traffic flow. This paper mainly focuses on the fleet management and control for HDVs and CAVs in dense fog weather conditions. As to restrict influence factors, we assume the following:

(1) The information interaction time delay between V2V or V2I is less than 20 ms. The time delay can be ignored in the CAVs control. The communication is uninterrupted and functional in dense fog weather conditions.

(2) All the drivers in HDVs have the same motivation that pass through dense fog environment safely. The drivers have no significant difference in driving ability. Nobody breaks out of the fleet during passing through the dense fog environment.

(3) The road alignment, gradient, and surface condition of highway in dense fog environment are consistent. The road gradient and surface condition do not affect the mobility and acceleration performance of vehicles during passing through the dense fog environment.

As to improve the traffic flow throughput and driving safety in dense fog weather condition on the highway, the fleet is composed of HDVs and CAVs. In the fleet, HDVs are driven by human completely and have not the capacity of V2I or V2V communication. CAVs are automatically driving and connect with each other. The HDVs are separated by CAVs in the mixed fleet as to reduce the collision risk caused by limited visibility in dense fog environment. The CAVs serve as an automated guided vehicle (AGV) to percept traffic conditions in dense fog environment. The fleet mixed with HDVs and CAVs is shown as Figure 1.

In Figure 1, the sequence numbers of CAVs and HDVs are labeled as the indices $i$ and $j$, respectively. The number sequence is from the head to tail of the fleet. $i = \{1, 2, 3, \ldots, n\}$. $j = \{1, 2, 3, \ldots, n\}$. $n$ is a natural number. The $i$th CAV and $j$th HDV in the fleet are defined as CAV$(i)$ and HDV$(j)$, respectively. If $j = i$, it means that the HDV$(j)$ is following the CAV$(i)$.

*2.2. Model Formulation*

*2.2.1. The State Space Model of CAVs.* In the fleet, the speed of CAV$(i)$ and HDV$(j)$ at time $t$ are defined as $v_i(t)$ and $v_j(t)$, respectively. The locations of CAV$(i)$ and HDV$(j)$ at time $t$ are defined as $d_i(t)$ and $d_j(t)$, respectively. The values of $v_j(t)$ and $d_j(t)$ can be detected and transferred by the following CAV. The minimum safety distance (MSD) between CAV$(i)$ and HDV$(j-1)(i=j)$ and between CAV$(i)$ and HDV$(j)(i=j)$ are defined as [32]

$$d_i^*(t) = -\frac{v_i^2(t)}{2 \times a_i(t)} + l_f, \tag{1}$$

$$d_f^*(t) = v_j(t).\gamma - \frac{v_j^2(t)}{2 \times a_j(t)} + l_f, \tag{2}$$

where $a_i(t)$ and $a_j(t)$ are minimum deceleration of the CAV$(i)$ and HDV$(j)$ at time $t$, respectively; $l_f$ is the critical headway; and $\gamma$ is the reaction time of drivers.

$s_i^*(t)$ is defined as the equilibrium spacing (ES) between CAV$(i)$ and HDV$(j-1)(i=j)$ at time $t$, and $s_i^*(t) > d_i^*(t)$ for CAV$(i)$ in the fleet. The ES of the CAVs in the fleet has the same value; that is, $s_i^*(t) = s^*(t)$, $\forall i = \{1, 2, \ldots .n\}$.

Within the fleet, we pursue the CAV following the HDV with the same velocity of HDV and keeping a safety distance in ES. Therefore, the state of CAV$(i)$ within the fleet is defined as

$$X_i := \left[\Delta d_i(\mathrm{t}), \Delta \mathrm{v}_i(t)\right]^T, \tag{3}$$

where $\Delta d_i(t)$ is the distance difference to ES at time t; $\Delta \mathrm{v}_i(t)$ is the speed difference to HDV$(j-1)$ at time $t$.

$$\Delta d_i(\mathrm{t}) = d_{j-1}(\mathrm{t}) - d_i(\mathrm{t}) - s^*(\mathrm{t}). \tag{4}$$

$$\Delta v_i(t) = v_{j-1}(t) - v_i(t). \tag{5}$$

Putting formulas (4) and (5) into formula (3), the state space model of CAVs within the fleet can be written as

$$X_i'(t) = AX_i(t) + Bu_i(t) + Da_{j-1}(t), \tag{6}$$

where $X_i'(t) = \begin{bmatrix} \Delta d_i'(t) \\ \Delta V_i'(t) \end{bmatrix}$, $\mathrm{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$, $D = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$; $\Delta d_i'(t)$ is the first derivative of $\Delta d_i(t)$, $\Delta v_t'(t)$ is

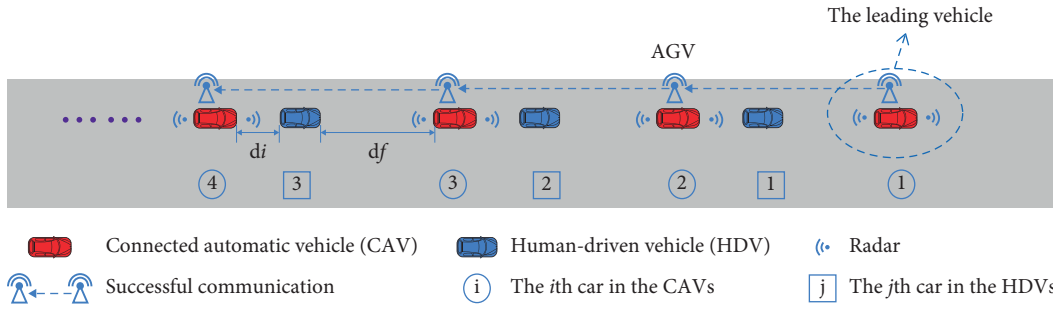



Figure 1: The layout of fleet in dense fog environment.

the first derivative of $\Delta v_i(t)$, $u_i(t)$ is the acceleration of CAV$(i)$ at time $t$; $a_{j-1}(t)$ is the acceleration of HDV$(j-1)$ at time $t$; and $j = i$ and $i = \{1, 2, 3, \ldots, n\}$.

As to describe the dynamic state space of CAVs in the fleet control, $\Delta t_s$ is defined as the discrete sampling time interval, and $\Delta t_s > 0$. K is the discrete sampling time, and $K\Delta t_s \le t \le (K+1)\Delta t_s$. If $\Delta t_s$ is small enough, the discretization equation can represent the dynamic moving process of CAVs [33, 34]. Therefore, the c2d function in MATLAB is used to disperse the state space of CAVs. The discretization equation is written as

$$X_i(K+1) = A' X_i(K) + B' u_i(K) + D' a_{j-1}(K), \tag{7}$$

where $X_i(K+1) = \begin{bmatrix} \Delta d_i(K+1) \\ \Delta V_i(K+1) \end{bmatrix}$; $X_i(K) = \begin{bmatrix} \Delta d_i(K) \\ \Delta v_i(K) \end{bmatrix}$; $A' = \begin{bmatrix} 2 & \Delta t_s \\ 0 & 1 \end{bmatrix}$; $B' = \begin{bmatrix} -(\Delta^2 t_s/2) \\ -\Delta t_s \end{bmatrix}$; $D' = \begin{bmatrix} \Delta^2 t_s/2 \\ \Delta t_s \end{bmatrix}$; $u_i(K)$ is the acceleration of CAV$(i)$ at time K; $a_{j-1}(K)$ is the acceleration of HDV$(j-1)$ at time K; $j = i$ and $i = \{1, 2, 3, \ldots, n\}$; $K = \{1, 2, 3, \ldots, m\}$; and m is a natural number, $m > 1$.

*2.2.2. The Car Following Model of HDVs in Dense Fog Environment on Highway.* Rosey [35] found that when the front vehicle is driving at a suitable speed in the dense fog environment, most of the following vehicles will drive and follow within the visibility range. Therefore, in the fleet management, it is important to guide the HDVs to keep safety driving in low-visibility environment by dynamically imposing restrictions on the velocity of AGV in the fleet, as to ensure the AGV always in the vision scope of HDVs and release the nervousness of driver in HDVs. In this situation, modeling the driver behavior of HDVs and the accuracy of car-following model of HDVs in dense fog environment will directly affect the performance of the fleet management, such as feasibility, reliability, stability, and robustness. The cellular automata (CA) model is widely used in traffic flow analysis and has advantages in describing the complex behavior and simulating the characteristics of traffic flow under various scenarios and traffic conditions [25–28]. However, as the CA models in the previous researches mainly focused on interactive between homogeneous vehicles, the current car-following model based on CA cannot be used directly between heterogeneous vehicles environment. Therefore, this paper attempts to model the car-following behavior with a Nagel–Schreckenberg (NaScH) cellular automatic (CA) model to describe the driver behavior in dense fog environment within the fleet mixed with HDVs and CAVs. The model is modeled as follows:

(1) If $d_i(K) - d_j(K) \ge L$ and $i = j$, then HDV driver will accelerate. It means that if the AGV is not in the scope of visibility, the driver of HDV will accelerate carefully to find the AGV [36]. The acceleration action is described as

$$u_j(K) = b \quad (b > 0), \tag{8}$$

where $d_i(K)$, $d_j(K)$ are the location of CAV$(i)$ and HDV$(j)$ at the time K, respectively. L is the distance of visibility in dense fog environment on the highway; $b$ is the random value within a reasonable range of acceleration. The uncertainty of $b$ reflects the randomness of the drivers under the interference of the actual situation.

(2) If $d_i(K) - d_j(K) < L$ and $i = j$, then the HDV driver will make an appropriate adjustment based on the speed and acceleration of HDV and the distance to AGV.

(i) If $d_i(K) - d_j(K) \ge d_f(K)$, then the HDV driver will try to keep pace with the velocity of AGV. That is, if $v_i(t) > v_j(t)$ and $i = j$, the HDV driver will accelerate to keep pace with the AGV; and if $v_i(t) < v_j(t)$ and $i = j$, the HDV driver will decelerate to keep pace with the AGV. The action is described as

$$u_j(K) = u_i(K) + c, \tag{9}$$

where $d_f(K)$ is the MSD between CAV$(i)$ and HDV$(j)$ $(i = j)$ at the time K. $c$ is the random value within a reasonable range of acceleration or deceleration.

(ii) If $d_i(K) - d_j(K) < d_f(K)$, then the HDV driver will decelerate. The deceleration action is described as

$$u_j(K) = f \quad (f < 0), \tag{10}$$

where $f$ is the random value within a reasonable range of deceleration.

Based on the driver behavior described above, the dynamic state of HDVs in the fleet in dense fog environment on the highway can be described as

$$\begin{aligned} v_j(\mathrm{K}+1) &= v_j(\mathrm{K}) + u_j(\mathrm{K}). \\ d_j(\mathrm{K}+1) &= d_j(\mathrm{K}) + v_j(\mathrm{K})\Delta t_s. \end{aligned} \tag{11}$$

We verified the effectiveness of the improved CA model in Section 4.1 and used it as the car-following model of HDVs in the fleet to further verify the fleet control model.

*2.2.3. Distributed Model Predictive Control (DMPC).* Model Predictive Control (MPC) can deal with the disturbance and uncertainty, and it is widely used in the dynamic vehicle control analysis and simulation [37]. The key difference between MPC and other control methods lies in the use of rolling optimization and rolling implementation of the control function, which makes the MPC more suitable for the complex traffic environments with low visibility. Therefore, CAVs using MPC can respond to the latest status of HDVs at every control moment. On the other hand, in the fleet which is mixed with HDVs and CAVs, it demands for the real-time and dynamical interactive among the CAVs to pass through the dense fog environment safely. As the central communication and control will increase time delay, computation complexity and reduce stability in the fleet control [38], the distributed control mode can take the advantage of CAVs computation, communication ability and reduce the time delay of communication and improve the stability of the system [39]. In addition, according to the layout of the fleet as shown in Figure 1, each CAV has the ability to perform the optimal control by itself. Therefore, this paper attempts to use the distributed model predictive control (DMPC) model to control the fleet in dense fog environment.

The interaction of CAVs within the fleet is described as in Figure 2. In the dense fog environment, the CAVs need to adjust their state according the state of the following HDVs in the fleet. As to adjust the state accurately, CAV ($i$) needs to obtain the driving state of HDV ($j-1$) from CAV ($i-1$) which can perceive the speed, acceleration, and location of HDV ($j-1$). At time K, CAV ($i-1$) processes the detected historical information, such as locations, speeds of CAV ($i-1$) and HDV ($j-1$). Then, CAV ($i-1$) will calculate the length $p$ and the acceleration of HDV ($j-1$) in the prediction horizon. Furthermore, CAV ($i-1$) will pass the information, such as $d_{j-1}(\mathrm{K})$, $v_{j-1}(\mathrm{K})$, $a_{j-1}(\mathrm{K})$, $a_{j-1}(\mathrm{K}+1)$, ..., $a_{j-1}(\mathrm{K}+\mathrm{p})$ to CAV ($i$). In synchronization, CAV ($i-1$) completes the optimization control according to the information transmitted by CAV ($i-2$). The optimization control process can be divided into two parts: (1) the fleet state prediction and (2) the fleet control based on DMPC.

*(1) The Fleet State Prediction.* In the processes of the fleet passing through the dense fog environment, CAV ($i-1$) perceives the speeds and locations of HDV ($j-1$) then evaluates the accelerations of HDV ($j-1$) at the time K. Then, the CAV ($i-1$) transfers these information of HDV ($j-1$) to CAV ($i$). In addition, CAV ($i-1$) processes the speeds and locations of HDV ($j-1$) before information interaction between CAVs ($j=i$ and $i=\{2, 3, \ldots, n\}$). The CA-based HDVs following model makes the trajectory of HDV ($j-1$) consistent with CAV ($j-1$) in a spatial-temporal delay $\tau_{j-1}(\mathrm{K})$ at time K, as shown in Figure 2. Thus, the locations $d_{j-1}(\mathrm{K}+\rho)$ and speeds $v_{j-1}(\mathrm{K}+\rho)$ of HDV ($j-1$) in prediction horizon can be obtained by dynamical equation. With the spatial-temporal delay $\tau_{j-1}(\mathrm{K})$, the accelerations or decelerations $a_{j-1}(\mathrm{K}+\rho)\,\rho = (0, 1 \ldots . p-1)$ of HDV ($j-1$) in the prediction horizon can be predicted by CAV ($i-1$) based on kinematic equation as

$$\begin{aligned} d_{j-1}(\mathrm{K}+\rho+1) = d_{j-1}(\mathrm{K}+\rho) + \mathrm{v}_{j-1}(\mathrm{K}+\rho)\times\Delta t_s \\ + 0.5\times a_{j-1}(\mathrm{K}+\rho)\times\Delta t_s^2. \end{aligned} \tag{12}$$

The spatial-temporal delay $\tau_{j-1}(\mathrm{K})$ can be obtained by curve matching based on Gong and Du's research [29]. The spatial-temporal delay matching curve is shown in Figure 3.

In Figure 3, it can be seen that there is a critical parameter $T_e$ (time range) in curve matching. And the simulation results showed that if the time range $T_e$ is longer, the more effective information and accurate can be obtained in the prediction. However, the longer time range $T_e$, the more computation resources are needed in CAVs. As to balance the calculation load and the accuracy of prediction, this paper sets the time range as $T_e = 100$s based on Chen's research [40].

$\{\mathrm{H}_{j,\mathrm{q1}} = (\mathrm{d}_{j,q_\mathrm{h}}, \mathrm{t}_{j,q_\mathrm{h}}), q_\mathrm{h} = 1, 2, 3, \ldots, |\mathrm{H}_j|\}$ and $\{\mathrm{C}_{i,q_\mathrm{c}} = (\mathrm{d}_{i,q_\mathrm{c}}, \mathrm{t}_{i,q_\mathrm{c}}), q_\mathrm{c} = 1, 2, 3, \ldots, |\mathrm{C}_i|\}$ $(i = j)$ are the historical trajectory and current trajectory of HDV ($j$) and CAV ($i$), respectively, in Figure 3. $\mathrm{d}_{j,q_\mathrm{h}}$, $\mathrm{t}_{j,q_\mathrm{h}}$ are the location and time of the $q_\mathrm{h}$ point in the historical trajectory of HDV ($j$), respectively. $\mathrm{d}_{i,q_\mathrm{c}}$, $\mathrm{t}_{i,q_\mathrm{c}}$ are the location and time of the $q_\mathrm{c}$ point in the current trajectory of CAV ($i$), respectively. The procedure of spatial-temporal delay matching is described as follows:

*Step 1.* Find the smallest distance point from $\mathrm{C}_i$ to $c$ based on the least square method as

$$\mathrm{D}\left(\mathrm{H}_j, \mathrm{C}_{i,q_\mathrm{c}}\right) = \min\left\|\mathrm{H}_j - \mathrm{C}^2_{i,q_\mathrm{c}}\right\|^2, \tag{13}$$

where $\mathrm{D}(\mathrm{H}_j, \mathrm{C}_{i,q_\mathrm{c}})$ is the minimum distance from the $q_\mathrm{c}$ point on $\mathrm{C}_i$ to $\mathrm{H}_j$.

*Step 2.* Calculate the minimum offset of spatial-temporal delay after finding the matching points pair of $q_\mathrm{h}$ and $q_\mathrm{c}$ in historical trajectory and current trajectory, respectively.

$$\mathrm{MinFt}_\mathrm{I} = \frac{1}{\mathrm{K}_\mathrm{I}}\sum_{\mathrm{K}=1}^{\mathrm{K}_\mathrm{I}}\left\|\mathrm{H}^\mathrm{I}_{j,\mathrm{K}} + \mathbf{t}_\mathbf{I} - \mathrm{C}^\mathrm{I}_{i,\mathrm{K}}\right\|^2, \tag{14}$$

where $Ft_I$ is the minimum offset of spatial-temporal delay between $\mathrm{C}_i$ and $\mathrm{H}_j$; $(\mathrm{H}^\mathrm{I}_{j,\mathrm{K}}, \mathrm{C}^\mathrm{I}_{i,\mathrm{K}})$ is the $\mathrm{K}^{\mathrm{th}}$ pairing point in the $\mathrm{I}^{\mathrm{th}}$ iteration; $\mathrm{K}_\mathrm{I}$ is the number of matching points in the $\mathrm{I}^{\mathrm{th}}$ iteration; and $\mathbf{t}_\mathbf{I}$ is the delay vector.

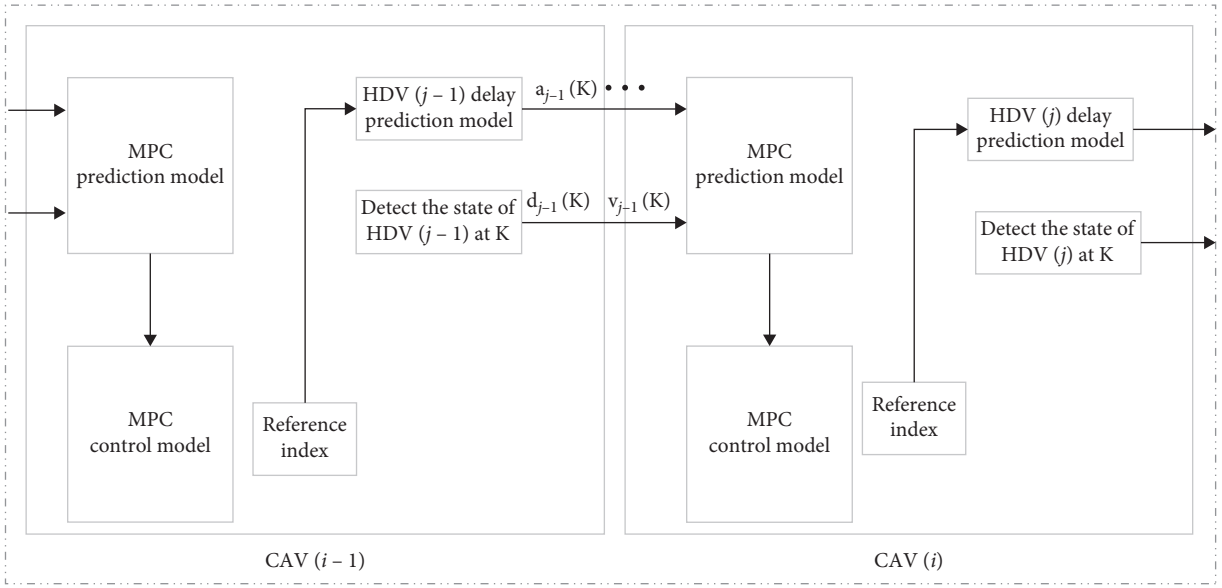



Figure 2: CAVs information interaction and DMPC control in the fleet.

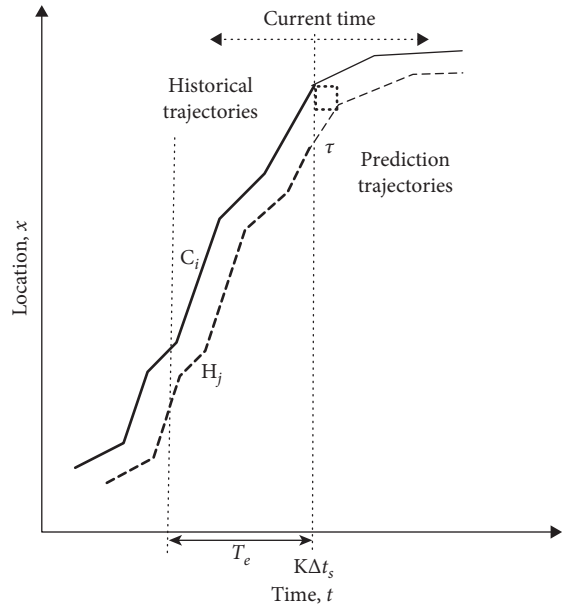



Figure 3: Spatial-temporal delay matching curve for HDV in prediction.

$$\mathbf{t}_{\mathbf{I}}^{*} = -\frac{\sum_{k=1}^{K_I}\left(H_{j,K}^{I} - C_{i,K}^{I}\right)}{K_I}, \tag{15}$$

where $t_I^*$ is the best solution of $t_I$.

*Step 3.* Repeat Step 1 and Step 2 until $t_I^*$ is small enough. Then, the HDV ($j$) trajectory can be predicted by the sum of $t_I^*$ as

$$\left(-\tau_j(K), ds_j(K)\right)^T = \sum t_I^*, \tag{16}$$

where $ds_j(K)$ is the spatial delay of HDV ($j$) at the time K; $-\tau_j(K)$ is the temporal offset of HDV ($j$) at the time K.

Then, the length of prediction horizon $p$ of the system at the time K can be calculated as

$$p = \left[\frac{\tau_{j-1}(K)}{\Delta t_s}\right] + 1. \tag{17}$$

*(2) The Fleet Control Based on DMPC. In the fleet.* CAV ($i$)*can obtain the following information by information interaction and prediction.*

(1) The location and speed of HDV ($j$ – 1) at time K

(2) The accelerations or decelerations of HDV ($j$ – 1) in prediction horizon and the length of prediction horizon

CAV $(i)$ obtains these pieces of information to update the state space of the system in horizon prediction and fleet control. In addition, the aim of the fleet control in dense fog environment is passing through the dense fog environment safely and steadily, keeping the vehicles in safety distance and within the scope of visibility. Thus, the object function of the fleet control in dense fog environment can be described as

$$\min J_i(\mathrm{K}) = \sum_{s=1}^{p} \left\{ \left(X_{i,\mathrm{K}+s}^{\mathrm{K}}\right)^T Q_i \left(X_{i,\mathrm{K}+s}^{\mathrm{K}}\right) + R_i \left(u_{i,\ \mathrm{K}+s-1}^{\mathrm{K}}\right)^2 \right\} + \left(X_{i,\mathrm{K}+p}^{\mathrm{K}}\right)^T W_i \left(X_{i,\mathrm{K}+p}^{\mathrm{K}}\right), \tag{18}$$

where $X_{i,\mathrm{K}+s}^{\mathrm{K}}$ is the $(\mathrm{K}+\mathrm{s})^{th}$ $(s = (1, 2, \ldots, p))$ state space of the system which is predicted at time K; $u_{i,\ \mathrm{K}+s-1}^{\mathrm{K}}$ is the $(\mathrm{K}+\mathrm{s})^{\mathrm{th}}$ control variable which is calculated at time K; $X_{i,\mathrm{K}+p}^{\mathrm{K}}$ is the terminal state space of system which is predicted at time K; and $\mathrm{Q}_i$, $\mathrm{R}_i$ are weight matric or weight value, respectively. $\mathrm{Q}_i$ is symmetric and positive definite matric, which is usually designed as $\mathrm{Q}_i = \mathrm{diag}(\alpha_{i1}, \alpha_{i2})(\alpha_{i1} > 0, \alpha_{i2} > 0)$; $\mathrm{R}_i$ is the coefficient that affects control variable and driver comfort, $\mathrm{R}_i > 0$. $\mathrm{Q}_i$ and $\mathrm{R}_i$ are the critical coefficient parameters in the fleet control and affect the stability of the fleet.

In formula (18), a comprehensive balanced control method is used to ensure the CAV $(i)$ performing the best operation at time K. It adopts two items to achieve this goal. The first item is $(X_{i,\mathrm{K}+s}^{\mathrm{K}})^T Q_i (X_{i,\mathrm{K}+s}^{\mathrm{K}})$, which attempts to reduce the position and speed errors of CAV $(i)$ and HDV $(j-1)$ during passing through the dense fog environment. And $R_i (u_{i,\ \mathrm{K}+s-1}^{\mathrm{K}})^2$ is used to reduce CAV $(i)$'s energy consumption and improve the comfort. The second item is $(X_{i,\mathrm{K}+m}^{\mathrm{K}})^T W_i (X_{i,\mathrm{K}+m}^{\mathrm{K}})$, which is used to reduce the errors between the state variables and the equilibrium state of the system at the end of the prediction horizon. In this study, $W_i$ is the solution of the discrete algebraic Riccati equation as formula (20), which is used to ensure the local asymptotic stability of the fleet [41].

$$W_i = \mathrm{Q}_i + \mathrm{A}' \left[ W_i - W_i B' \left( \mathrm{R}_i + B'^T W_i B' \right)^{-1} B'^T W_i \right] \mathrm{A}'. \tag{19}$$

The constraints of object function in the fleet control are set as follows:

(1) Control constraints:

$$u_{\min} \le u_{i,\mathrm{K}+s}^{\mathrm{K}} \le u_{\max} \quad \forall s \in \{0, 1, 2, \ldots p-1\}. \tag{20}$$

(2) State constraints:

$$v_{\min} \le v_{i,\mathrm{K}+s}^{\mathrm{K}} \le v_{\max} \ \ \forall s \in \{1, 2, 3, \ldots .. p\}, \tag{21}$$

where $v_{\min}$ and $v_{\max}$ are the minimum and maximum speed of the fleet during passing through the dense fog environment on highway, respectively.

$$\Delta d_{i,\mathrm{K}+s}^{\mathrm{K}} + s^*(t) \ge d_i^*(\mathrm{K}+s \mid \mathrm{K}), \tag{22}$$

where $\Delta d_{i,\mathrm{K}+s}^{\mathrm{K}}$ is the $(\mathrm{K}+s)^{\mathrm{th}}$ distance difference of system which is predicted at time K; $d_i^*(\mathrm{K}+s \mid \mathrm{K})$ is the $(\mathrm{K}+s)^{\mathrm{th}}$ MSD of CAV $(i)$ which is predicted at time K.

$$\Delta d_i^- \le \Delta d_{i,\mathrm{K}+s}^{\mathrm{K}} \le \Delta d_i^+ \quad \forall s \in \{1, 2, 3, \ldots p\}, \tag{23}$$

where

$$\begin{aligned} \Delta d_i^- &= -\max\left(\left|\Delta d_{i-1,\sigma}^r\right|\right) \text{for i} > 1 \text{ for} \quad \forall \sigma \in \{1, 2, 3, \ldots, \mathrm{K}\}, \\ \Delta d_i^+ &= \max\left(\left|\Delta d_{i-1,\sigma}^r\right|\right) \text{for i} > 1 \text{ for} \quad \forall \sigma \in \{1, 2, 3, \ldots, \mathrm{K}\}, \end{aligned} \tag{24}$$

where $\Delta d_{i-1,\sigma}^r$ is the actual state of CAV $(i-1)$ at time $\sigma$.

The constraint formulas (23) and (24) are used to guarantee the stability of the fleet.

## 3. Analysis Feasibility and Stability of Fleet Control in Dense Fog Environment

Based on the theory of Lyapunov stability [42, 43], if the fleet is in the status of feasibility, it means that it can find out the optimal control based on the state constraints. Within the feasibility analysis, the stability analysis includes analysis of the asymptotic stability and string stability. The asymptotic stability refers to the ability of the elements in the system to return to the initial stable states in a short period when subjected to slight disturbances. In this paper, it refers to the CAVs in the fleet. And the string stability reflects the ability of the system (the fleet) to resist slight disturbances. Therefore, the feasibility and stability analysis for the fleet is the premise of performing optimal control. Through the analysis, it can tell us whether the optimal control can be found out to keep the fleet stable.

*3.1. The Feasibility of Fleet Control.* In the fleet control, as to enforce the feasibility of the controller, the controller should have the capability to endure a certain range of disturbances. As to improve the feasibility of the fleet, the constraints of the fleet can be rewritten as

$$\begin{aligned} U = \Big\{ & \mathrm{u}_{i,\ \mathrm{K}+s}^{\mathrm{K}} \mid \mathrm{u}_{\min} \le \mathrm{u}_{i,\ \mathrm{K}+s}^{\mathrm{K}} \le \mathrm{u}_{\max}, \mathrm{v}_{\min} \le \mathrm{v}_{i,\ \mathrm{K}+s}^{\mathrm{K}} \\ & \le \mathrm{v}_{\max}, d_{j-1,\mathrm{K}+s}^{\mathrm{K}} - d_{i,\mathrm{K}+s}^{\mathrm{K}} + \frac{\left(\mathrm{v}_{i,K+s}^{K}\right)^2}{2\mathrm{u}_{\min}} \\ & - l_f \ge 0, \left| d_{j-2,\mathrm{K}+s}^{\mathrm{K}} - d_{i-1,\mathrm{K}+s}^{\mathrm{K}} \right| \ge \left| d_{j-1,\mathrm{K}+s}^{\mathrm{K}} - d_{i,\mathrm{K}+s}^{\mathrm{K}} \right| \Big\}, \end{aligned} \tag{25}$$

where $\mathrm{u}_{i,\ \mathrm{K}+s}^{\mathrm{K}}$ is the $s^{\mathrm{th}}$ control value of CAV $(i)$ in prediction horizon at time $K$; $d_{j-1,\mathrm{K}+s}^{\mathrm{K}}$ is the $s^{\mathrm{th}}$ location of HDV $(j-1)$ in prediction horizon at time $K$; $d_{i,\mathrm{K}+s}^{\mathrm{K}}$ is the $s^{th}$ location of CAV $(i)$ in prediction horizon at time $K$; and $\mathrm{v}_{i,\mathrm{K}+s}^{\mathrm{K}}$ is the $s^{\mathrm{th}}$ speed of CAV $(i)$ in prediction horizon at time K, $j = i$ and $i = \{2, 3, \ldots, n\}$ $(n \in R^+)$, $s = (1, 2, \ldots p)$.

Assume that the $(s-1)^{\mathrm{th}}$ optimal control is feasible; then the speed constraint in formula (25) can be obtained as

$$v_{\min} \le v^{\mathrm{K}}_{i,\mathrm{K}+s-1} + \Delta v^{\mathrm{K}}_{i,\mathrm{K}+s-1} \le v_{\max}. \tag{26}$$

Then, the range of $\mathrm{u}^{K}_{i,\,K+s-1}$ can be obtained by writing the formula 26 as

$$\frac{\mathrm{v}_{\min} - \mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s-1}}{\Delta t} \le \mathrm{u}^{\mathrm{K}}_{i,\,\mathrm{K}+s-1} \le \frac{\mathrm{v}_{\max} - \mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s-1}}{\Delta t}. \tag{27}$$

Define $\mathrm{u}_{v1} = (\mathrm{v}_{\min} - \mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s-1}/\Delta\mathrm{t})$ and $\mathrm{u}_{v2} = (\mathrm{v}_{\max} - \mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s-1}/\Delta\mathrm{t})$. As the $(s-1)^{\mathrm{th}}$ optimal control is feasible, and $\mathrm{v}_{\min} \le \mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s-1} \le \mathrm{v}_{\max}$, it can deduced that $u_{v1} < 0$ and $u_{v2} > 0$. Thus, we can get the control value that satisfies both the acceleration and speed value in the fleet can be restricted to satisfy the constraints and make sure the fleet is feasible.

In addition, as the $(s-1)^{\mathrm{th}}$ optimal control is feasible, the MSD constraint in formula (25) can be obtained as

$$\begin{aligned} & d^{\mathrm{K}}_{j-1,\mathrm{K}+s} - d^{\mathrm{K}}_{i,\mathrm{K}+s} + \frac{\left(\mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s}\right)^2}{2\mathrm{u}_{\min}} - l_f = \left( d^{\mathrm{K}}_{j-1,\mathrm{K}+s-1} + \mathrm{v}^{\mathrm{K}}_{j-1,\mathrm{K}+s-1}\Delta\mathrm{t} + \frac{\Delta\mathrm{t}^2}{2} a^{\mathrm{K}}_{j-1,\mathrm{K}+s-1} \right) \\ & - \left( d^{\mathrm{K}}_{i,\mathrm{K}+s-1} + \mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s-1}\Delta t + \frac{\Delta t^2}{2} u^{\mathrm{K}}_{i,\mathrm{K}+s-1} \right) + \frac{\left(\mathrm{v}^{K}_{i,K+s-1} + u^{\mathrm{K}}_{i,K+s-1}\right)^2}{2u_{\min}} - l_f = \frac{\Delta t^2}{2u_{\min}} \left(u^{\mathrm{K}}_{i,\mathrm{K}+s-1}\right)^2 \\ & + \left( \frac{\left(v^{\mathrm{K}}_{i,\mathrm{K}+s-1}\Delta t\right)}{u_{\min}} - \frac{\Delta t^2}{2} \right) u^{\mathrm{K}}_{i,\mathrm{K}+s-1} + \left[ \left( d^{\mathrm{K}}_{j-1,\mathrm{K}+s-1} - d^{\mathrm{K}}_{i,\mathrm{K}+s-1} \right) \right. \\ & \left. + \Delta\mathrm{t}\left(\mathrm{v}^{\mathrm{K}}_{j-1,\mathrm{K}+s-1} - \mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s-1}\right) + \frac{\Delta\mathrm{t}^2}{2} a^{\mathrm{K}}_{j-1,\mathrm{K}+s-1} + \frac{\left(\mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s}\right)^2}{2u_{\min}} - l_f \right] = 0, \end{aligned} \tag{28}$$

where $a^{\mathrm{K}}_{j-1,\mathrm{K}+s-1}$ is the $(s-1)^{th}$ control value of HDV $(j-1)$ in prediction horizon at time $K$; $\mathrm{v}^{K}_{j-1,K+s-1}$ is the $(s-1)^{th}$ speed of HDV $(j-1)$ in prediction horizon at time $K$.

Given $\mathrm{u}_{\min} < 0$ and $d^{\mathrm{K}}_{j-1,K+s} - d^{\mathrm{K}}_{i,K+s} + ((\mathrm{v}^{K}_{i,K+s})^2/2\mathrm{u}_{\min}) - l_f \ge 0$, then $u^{\mathrm{K}}_{i,K+s-1}$ is between the two roots of formula 30

$$u_{\mathrm{msd1}} = \frac{u_{\min}}{2} + \frac{u_{v1}}{\Delta t} + \frac{u_{\min}}{2\Delta t}\sqrt{C}, \tag{29}$$

$$u_{\mathrm{msd2}} = \frac{u_{\min}}{2} + \frac{u_{v1}}{\Delta t} - \frac{u_{\min}}{2\Delta t}\sqrt{C}, \tag{30}$$

$$C = \frac{\Delta\mathrm{t}^2 u_{\min} + 8\left(l_f - d^{\mathrm{K}}_{j-1,\mathrm{K}+s-1} + d^{\mathrm{K}}_{i,\mathrm{K}+s-1}\right) - 4\Delta\mathrm{t}\left(\mathrm{v}^{\mathrm{K}}_{j-1,\mathrm{K}+s-1} + \mathrm{v}^{\mathrm{K}}_{j-1,\mathrm{K}+s} - \mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s-1}\right)}{u_{\min}}. \tag{31}$$

As the $(s-1)^{\mathrm{th}}$ optimal control is feasible, then $l_f - d^{\mathrm{K}}_{j-1,\mathrm{K}+s-1} + d^{\mathrm{K}}_{i,\mathrm{K}+s-1} < 0$. If the control interval $\Delta\mathrm{t}$ is small enough, the vehicles in the fleet will have little difference in speed. Thus, $\mathrm{v}^{\mathrm{K}}_{j-1,\mathrm{K}+s-1} + \mathrm{v}^{\mathrm{K}}_{j-1,\mathrm{K}+s} - \mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s-1} > 0$, $A > 0$, and $\mathrm{u}_{\mathrm{msd1}} < \mathrm{u}_{\mathrm{msd2}}$.

$$\begin{aligned} \mathrm{u}_1 &= \max\left(\mathrm{u}_{\min,}\, \mathrm{u}_{v1},\, \mathrm{u}_{\mathrm{msd1}}\right), \\ \mathrm{u}_2 &= \min\left(\mathrm{u}_{\max}, \mathrm{u}_{v2},\, \mathrm{u}_{\mathrm{msd2}}\right), \end{aligned} \tag{32}$$

where $u_1$ is the lower boundary of the constraint; $u_2$ is the upper bound of the constraint.

Given $u_{v1} < 0 u_{\min} < 0$, and $0 < \Delta t < 1$, the change of $\mathrm{u}_{v1}$ can be obtained as

$$u_{\mathrm{msd1}} < \frac{u_{v1}}{\Delta t} < u_{v1}. \tag{33}$$

Then, $u_1 = \max(u_{\min,}\, u_{v1})$ according to formula (33). Based on formulas (1), (22), (27), (30), and (31), the range of $u_{\mathrm{msd2}}$ can be obtained as

$$\begin{aligned} u_{\mathrm{msd2}} &\ge \frac{u_{\min}\Delta t + 2u_{v1} + \sqrt{\Delta t^2 u^2_{\min} + 8u^2_{v1} + 4\Delta t^2 u_{\min} u_{v1}}}{2\Delta t} \\ &> \frac{u_{\min}\Delta t + 2u_{v1} + \left|u_{\min}\Delta t + 2\sqrt{2}\, u_{v1}\right|}{2\Delta t} > 0. \end{aligned} \tag{34}$$

Thus, $u_{\mathrm{msd2}} > u_{v1}$ and $u_{\mathrm{msd2}} > u_{\min}$. If $u_{v1} < 0$ and $u_{\min} < 0$, then $u_2 > u_1$. It can be concluded that the first three constraints in formula (25) are feasible.

Next, it can be found that the fourth constraint in formula (25) makes the constraint (25) loose. That is due to the fact that $|d^{\mathrm{K}}_{j-2,\mathrm{K}+s} - d^{K}_{i-1,\mathrm{K}+s}| > (-((v^{\mathrm{K}}_{i-1,\mathrm{K}+s})^2/2u_{\min}) + l_f)$, which $\left\{(-(v^{K}_{i-1,K+s})^2/2u_{\min})\right\}$ is approximately equal to $(-(v^{K}_{i,K+s})^2/2u_{\min})$ when the fleet is running normally. Thus, the fourth constraint is more relaxed than the third constraint in formula (25). In other words, the controller rolling

optimization at the control time is feasible under a certain range of disturbances.

*3.2. The Stability Analysis for Fleet Control.* Stability is an important feature for the fleet. In this study, we analyze the asymptotic stability of the single system and the string stability of the formation. According to the theorem developed by Mayne [44], Zhou [30], and the Lyapunov stability principle [42, 43], if the optimal control meets the following five conditions, the asymptotic stability of the system can be guaranteed:

(1) $X_{i,e} \in Z_f$
(2) $\mathrm{K}_f(X^{\mathrm{K}}_{i,\mathrm{K}+s}, a^{\mathrm{K}}_{j-1,\mathrm{K}+s}) \in \mathrm{U} \quad \forall X^{\mathrm{K}}_{i,\mathrm{K}+s} \in Z_f$
(3) $\mathrm{A}' X^{\mathrm{K}}_{i,\mathrm{K}+s} + B'\mathrm{K}_f(X^{\mathrm{K}}_{i,\mathrm{K}+s}, a^{\mathrm{K}}_{j-1,\mathrm{K}+s}) + D' a^{\mathrm{K}}_{j-1,\mathrm{K}+s} \in X_f, \quad \forall X^{\mathrm{K}}_{i,\mathrm{K}+s} \in Z_f$
(4) $\mathrm{F}(\mathrm{A}' X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}} + B'\mathrm{K}_f(X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}})) - \mathrm{F}(X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}}) + \mathrm{L}(\mathrm{A}' X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}} + B'\mathrm{K}_f(X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}})) \le 0 \quad X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}} \in Z_f$ and
(5) $\mathrm{CAV}(i)$ is robust recursive feasibility, $\forall i \in (2, 3, \ldots n)$.

Here, $X_{i,e}$ is the equilibrium state; $Z_f$ is the terminal state domain; $\mathrm{K}_f$ is the implicit control law; F, L are the cost functions, respectively. The other symbols have the same meaning as above.

$Z_f$ is set as a subset of the robust invariant set according to Zhou [30] as

$$i, \mathrm{RIS} = \left\{ X^{K}_{i,\mathrm{K}+s} = \left[\Delta \mathrm{d}^{\mathrm{K}}_{i,\mathrm{K}+s}, \Delta \mathrm{v}^{\mathrm{K}}_{i,\mathrm{K}+s}, a^{\mathrm{K}}_{j-1,\mathrm{K}+s}\right] \subseteq i, \mathrm{RIS} \,\middle|\, \mathrm{A}' X^{\mathrm{K}}_{i,\mathrm{K}+s} + B'\mathrm{K}_f\left(X^{\mathrm{K}}_{i,\mathrm{K}+s}, a^{\mathrm{K}}_{j-1,\mathrm{K}+s}\right) + D' a^{\mathrm{K}}_{j-1,\mathrm{K}+s} \in i, \mathrm{RIS}, \mathrm{K}_f\left(X^{\mathrm{K}}_{i,\mathrm{K}+s}, a^{\mathrm{K}}_{j-1,\mathrm{K}+s}\right) \in \mathrm{U} \text{ and } a^{\mathrm{K}}_{j-1,\mathrm{K}+s} \in \mathrm{U} \right\}. \tag{35}$$

Obviously, $X_{i,e} \in i, \mathrm{RIS}$ and let $Z_f \subseteq i, \mathrm{RIS}$; then, conditions (1), (2), and (3) as mentioned above are fulfilled. According to formula (7) and formula (18), condition (4) can be reorganized as

$$\begin{aligned} &\left(\mathrm{A}' X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}} + B'\mathrm{K}_f\left(X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}}\right)\right)^T W_i\left(\mathrm{A}' X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}} + B'\mathrm{K}_f\left(X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}}\right)\right) - \left(X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}}\right)^T W_i\left(X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}}\right) \\ &+\left(\mathrm{A}' X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}} + B'\mathrm{K}_f\left(X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}}\right)\right)^T Q_i\left(\mathrm{A}' X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}} + B'\mathrm{K}_f\left(X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}}\right)\right) + R_i\left(\mathrm{K}_f\left(X^{\mathrm{K}}_{i,\mathrm{K}+\mathrm{p}}, a^{\mathrm{K}}_{j-1,\mathrm{K}+\mathrm{p}}\right)\right)^2 \le 0. \end{aligned} \tag{36}$$

Assume the implicit control law $\mathrm{K}_f(X^{\mathrm{K}}_{1,\mathrm{K}+s}) = kX^{\mathrm{K}}_{1,\mathrm{K}+s}$ ($s = (1, 2, \ldots, p)$), formula (36) can be rewritten as

$$\begin{aligned} &\left(X^{\mathrm{K}}_{1,\mathrm{K}+\mathrm{p}}\right)^T \left\{ (\mathrm{A}' + B'k)^T W_i (\mathrm{A}' + B'k) - W_i \right. \\ &\left. + (\mathrm{A}' + B'k)^T Q_i (\mathrm{A}' + B'k) + R_i k \right\} \left(X^{K}_{1,\mathrm{K}+\mathrm{p}}\right) \le 0. \end{aligned} \tag{37}$$

Let $\vartheta = (\mathrm{A}' + B'k)^T W_i (\mathrm{A}' + B'k) - W_i + (\mathrm{A}' + B'k)^T Q_i (\mathrm{A}' + B'k) + R_i k$ and $k = 0$; the first three conditions are also satisfied and $\vartheta$ will be simplified as

$$\vartheta = (\mathrm{A}')^T W_i (\mathrm{A}') - W_i + (\mathrm{A}')^T Q_i (\mathrm{A}'), \tag{38}$$

where $A' = \begin{bmatrix} 2 & \Delta t_s \\ 0 & 1 \end{bmatrix}$; $Q_i = \mathrm{diag}(\alpha_{i1}, \alpha_{i2})$. Let $W_i = \mathrm{diag}(\beta_{i1}, \beta_{i2})$; then $\vartheta$ can be rewritten as

$$\vartheta = \begin{bmatrix} 3\beta_{i1} + 4\alpha_{i1}, & 2\Delta t_s \beta_{i1} + 2\Delta t_s \alpha_{i1}, \\ 2\Delta t_s \beta_{i1} + 2\Delta t_s \alpha_{i1}, & \Delta t_s^2 \beta_{i1} + \Delta t_s^2 \alpha_{i1} + \alpha_{i2}. \end{bmatrix}. \tag{39}$$

If $\vartheta$ is a negative semidefinite matrix, $\vartheta$ satisfies the conditions as follows:

$$3\beta_{i1} + 4\alpha_{i1} \le 0, \tag{40}$$

$$\begin{aligned} &(3\beta_{i1} + 4\alpha_{i1})\left(\Delta t_s^2 \beta_{i1} + \Delta t_s^2 \alpha_{i1} + \alpha_{i2}\right) \\ &- \left(2\Delta t_s \beta_{i1} + 2\Delta t_s \alpha_{i1}\right)^2 \ge 0. \end{aligned} \tag{41}$$

It can be deduced that formula (36) will be proved under formula (40). Furthermore, as the appropriate weight matrixes Q, w, and R are selected, formula (36) can be guaranteed in different $\mathrm{K}_f$ which satisfies conditions (1), (2), and (3) as mentioned above. Thus, the above four conditions have proved to be feasible; it can be concluded that the single system of fleet is asymptotically stable.

As to prove $\mathrm{CAV}(i)$ is robust recursive feasibility, it should be ensured that CAVs can predict HDVs accurately and interact successfully between CAVs. In the stability theory of dynamical systems, the string stability is well known as the higher level of formation control, which requires the control error gradually decreasing during the formation propagation. Based on the constraints of the object function of fleet control in dense fog environment, it can further set the constraint to meet the string stability as

$$\left|\Delta d^{\mathrm{K}}_{i,\mathrm{K}+s}\right| \le \max\left(\left|\Delta d^{r}_{i-1,\sigma}\right|\right), \tag{42}$$

where $\Delta d^{K}_{i,K+s}$ is the control error of state space at the time $K + s$ which is predicted by of CAV $(i)$ at the time K.

In the restriction in execution, DMPC-based fleet executes the first control variable in each time; $\Delta d^{K}_{i,K+s}$ can be defined as

$$\Delta d^{K}_{i,K+1} = \Delta d^{r}_{i,K+1}. \tag{43}$$

Then, formula (29) can be replaced as

$$\Delta d^{r}_{i,K+1} \leq \max\left(\left|\Delta d^{r}_{i-1,\sigma}\right|\right). \tag{44}$$

If the restrictions in formula (44) are adopted in the fleet control, the fleet will be consistent with the performance of CAVs formation. The control error can gradually decrease in the fleet during the formation propagation [11]. However, in the fleet, the uncertainty and randomness are generated by HDVs drivers. The CAVs following the HDVs need a greater control action to compensate the uncertainty and randomness made by the HDVs drivers. The string stability between CAVs can ensure the global asymptotic stability by reducing the control error during the formation propagation. Therefore, the fleet control in this paper can provide a nonstrict string stability during passing through dense fog environment on the highway.

## 4. Numerical Experiments

As to verify the performance of fleet in dense fog environment on highway, this paper uses numerical experiments to analyze the ability and capability of fleet with the proposed control model. Firstly, we verified the effectiveness of the improved HDV car-following model; that is, the model can reflect the driving characteristics of the driver in a dense fog environment condition. Secondly, we discussed the driving conditions of the fleet under different visibility to verify the feasibility and asymptotic stability of DMPC. Then, we changed the order of HDVs in the fleet and further analyzed the effective ability of the DMPC-based control model. Finally, we studied the relationship between string stability and control requirements for the fleet. The numerical experiments are performed by MATLAB R2018b on Windows 10 with Intel® Core™ i5-6200U CPU @ 2.30 GHz RAM 4.00 GB.

*4.1. The CA-Based NaScH Car following Model Verification.* Whether the HDV car-following model meets the driving characteristics of drivers in dense fog weather conditions is the first most important factor in managing and controlling the fleet mixed with HDVs and CAVs efficiently and precisely. Siebert and Wallis [45] had studied "how speed and visibility influence preferred headway distances in highly automated driving." Based on the study of Siebert and Wallis on the relationship between visibility and speed, the proposed car-following model is verified in three parameters groups. The parameters setting in the car-following model is shown in Table 1. $V_L$ is the limited speed of fleet in the visibility L. The initial headway between CAVs and HDVs in the fleet is set equal to the visibility L in the simulation. The random value of $b$, c, and $f$ is generated by the MATLAB random function within a defined range in Table 1.

It can be seen from Figures 4(c), 4(f), and 4(i) there is no significant different in acceleration characteristics in CAVs or HDVs under different visibility conditions. The control behavior of CAVs and HDVs seems to have the same characteristics. It means that the proposed car-following model can describe the characteristics of HDVs drivers trying to keep pace with AGV in the dense fog environment within the fleet. Based on this characteristic, the HDVs trajectories are following the CAVs trajectories in the fleet in different visibility conditions, as shown in Figures 4(a), 4(d), and 4(g). As to follow up the AGV within the scope of visibility, the HDV driver tries to keep pace with the speed of AGV by dynamically accelerate up or decelerate down the HDV, as shown in Figures 4(b), 4(e), and 4(h), corresponding to Figures 4(c), 4(f), and 4(i), respectively. It is worth noting that there is an acceleration behavior at the initial stage of HDV speed diagram in Figures 4(b), 4(e), and 4(h). It is due to the fact that the initial headway in the fleet is set equal to the distance of visibility. AGV is in the edge of visibility. Therefore, the HDV driver accelerates to keep pace with the AGV. The CA-based NaScH car-following model can reflect this phenomenon. As shown in Figure 4, the trajectory, speed, and acceleration of HDVs are keeping pace with the AGV in the dense fog environment within the fleet.

As to further verify the effectiveness of the proposed car-following model in the dense fog environment with the distance of visibility as 70 m, the initial headway is set as 30 m, 50 m, and 70 m in the verification simulation, respectively. In this situation, 30 m is less than the MSD. 50 m is a little more than the MSD. 70 m is equal to the distance of visibility. As shown in Figure 3, if the initial headway is 30 m, HDV will slow down to keep always from the AGV until the distance is in the safety distance. If the initial headway is 70 m, HDV will accelerate to keep pace with AGV and keep the AGV within the scope of visibility. If the initial headway is 50 m, the HDV will maintain the following state to keep the AGV within the scope of visibility and keep a safety distance from the AGV. The proposed car-following model also reflects these characteristics of HDVs drivers in the dense fog environment within the fleet.

From the phenomena shown in Figures 4 and 5, they are consistent with the assumptions of HDV driver characteristics discussed above and falls into the relationship of speed and visibility as studied by Siebert and Wallis [45]. Therefore, the proposed car-following model above can effectively describe the driver behavior within dense fog environment in the fleet control.

*4.2. The DMPC-Based Fleet Control Model Verification.* In this section, we focus on verifying the feasibility and stability of the mixed flow of HDVs and CAVs within the fleet. In the fleet, three CAVs and three HDVs are used to form the fleet in the numerical experiments. CAV alternates with HDV in the fleet. That is, the first, third, and fifth are CAV vehicles, and the second, fourth, and sixth are HDV vehicles. Three limited visibility ranges in dense fog environment are used as the fleet scenarios on the highway. The

Table 1: The parameters in the car-following model of HDVs.

| Parameter | $L$ (m) | $V_L$ (m/s) | $b$ | $c$ | $f$ | $d_f$ (m) | Initial headway (m) |
|---|---|---|---|---|---|---|---|
| Simulation 1 | 40 | 11 | (1, 2) | (−0.5, 0.5) | (−1, 0) | 21 | 40 |
| Simulation 2 | 60 | 17 | (1, 3) | (−0.5, 0.5) | (−1, 0) | 37 | 60 |
| Simulation 3 | 70 | 19 | (1, 3) | (−0.5, 0.5) | (−1, 0) | 44 | 70 |

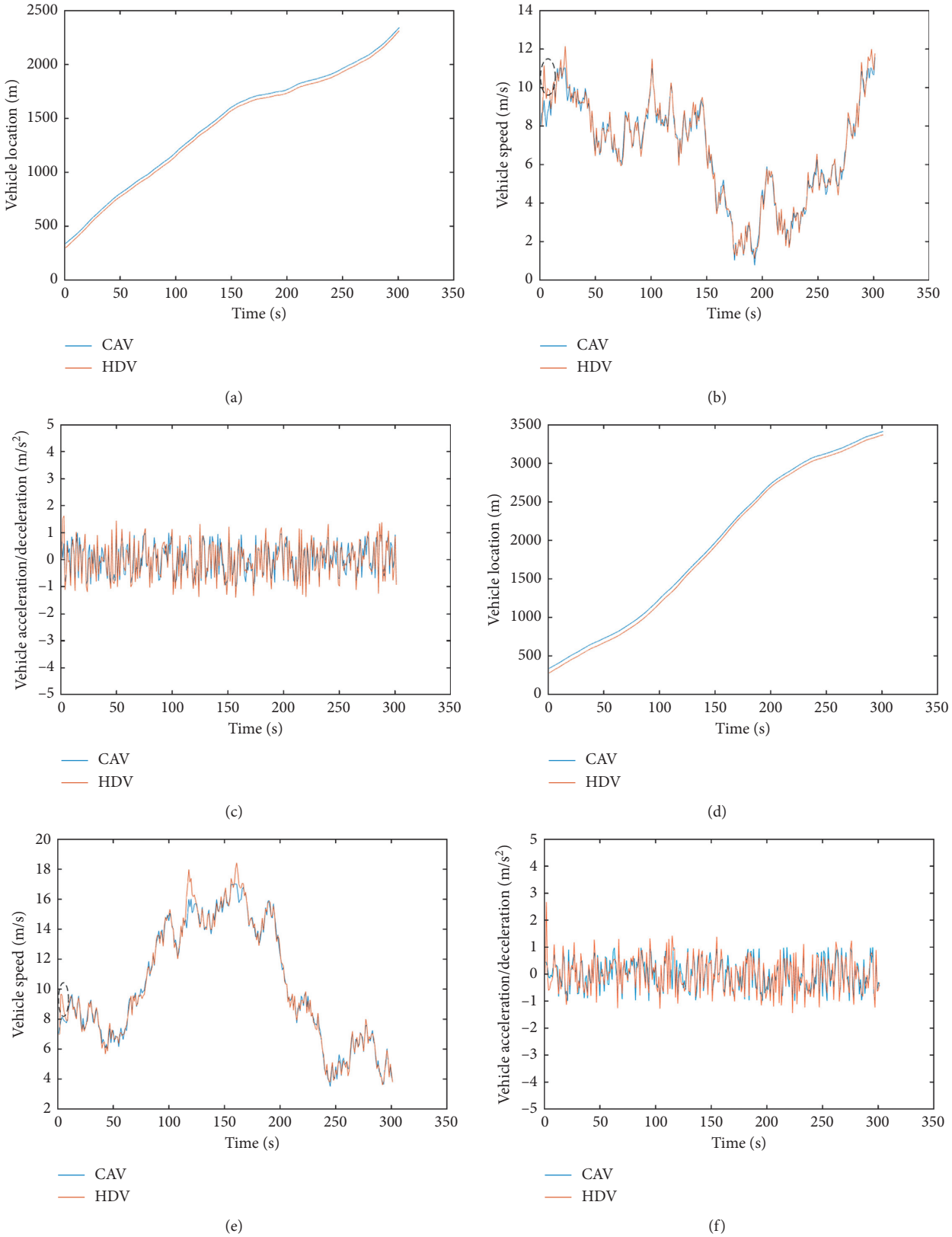



Figure 4: Continued.

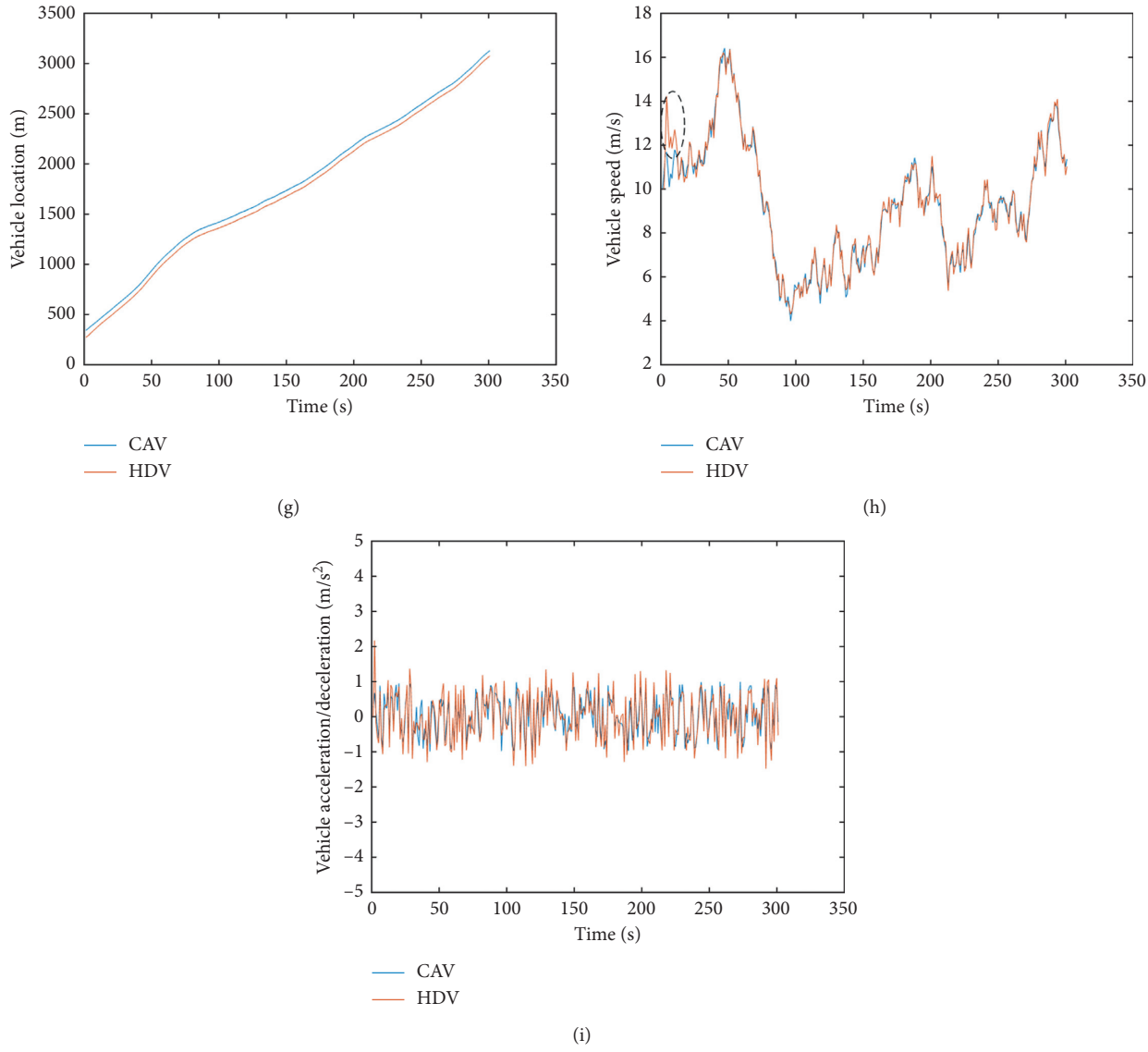



Figure 4: The car-following behavior of HDVs in the fleet within dense fog environment on highway. (a) The vehicle location of CAVs and HDVs ($L = 40$ m). (b) The vehicle speed of CAVs and HDVs ($L = 40$ m). (c) The vehicle acceleration of CAVs and HDVs ($L = 40$ m). (d) The vehicle location of CAVs and HDVs ($L = 60$ m). (e) The vehicle speed of CAVs and HDVs ($L = 60$ m). (f) The vehicle acceleration of CAVs and HDVs ($L = 60$ m). (g) The vehicle location of CAVs and HDVs ($L = 70$ m). (h) The vehicle speed of CAVs and HDVs ($L = 70$ m). (i) The vehicle acceleration of CAVs and HDVs ($L = 70$ m).

parameters setting in the numerical experiments are shown in Table 2.

In the control performance as shown in Figures 6(a), 6(b), 7(a), 7(b), 8(a), and 8(b), CAVs can be basically consistent with the HDVs in trajectory and speed under different driving states, which shows that the controller can guarantee the safe operation of the fleet. Furthermore, as shown in Figures 6(c), 7(c), and 8(c), the spacing in the fleet is controlled within the safety distance under different visibilities. It illustrates that the DMPC prediction model can accurately predict the states of HDVs and passes the information to the rear CAVs. On the other hand, the fleet can maintain the ideal distance between CAVs and HDVs under random disturbance. It shows that the system can ensure asymptotic stability. It should be noted that the continuous fluctuation of the states in the experiments is the result of the joint effect of the experimental inputs and the randomness of HDVs.

*4.3. Further Verification by Adjusting the Order of HDVs and CAVs in the Fleet.* As to further test the mixed flow of HDVs and CAVs and find out which layouts of the fleet have better performance by the DMPC-based control model, we changed the following number of HDVs between CAVs to two HDVs, three HDVs, and four HDVs in the fleet, respectively. We call these as regular alignment of the fleet, as shown in Figure 9. Then, we changed the queue of the fleet with one HDV between CAVs and three HDVs between

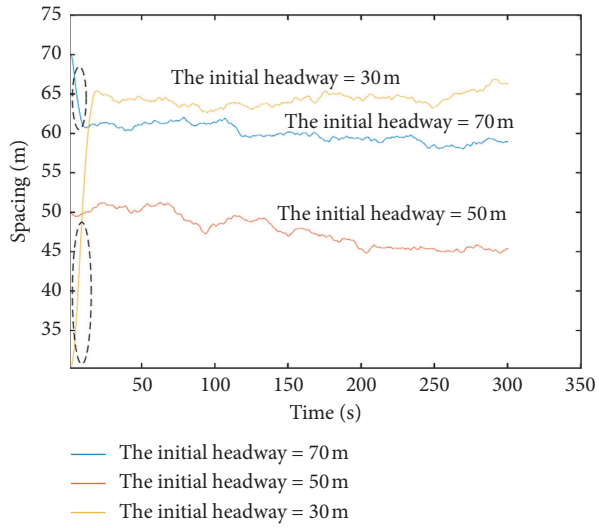



Figure 5: The distance between HDVs and CAVs in different initial headway under the distance of visibility = 70 m in dense fog environment on highway.

Table 2: The parameters in the numerical experiments.

| Visibility(m) | $V_L$(m/s) | $b$(m) | $c$(m) | $f$(m) | $d_j^*$ (m) | $d_i^*$ (m) |
|---|---|---|---|---|---|---|
| 50 | 14 | (1,2) | (−0.5,0.5) | (−0.5,0) | 28 | 14 |
| 70 | 19 | (1,3) | (−0.5,0.5) | (−1,0) | 44 | 20 |
| 90 | 25 | (1,4) | (−0.5,0.5) | (−2,0) | 66 | 41 |

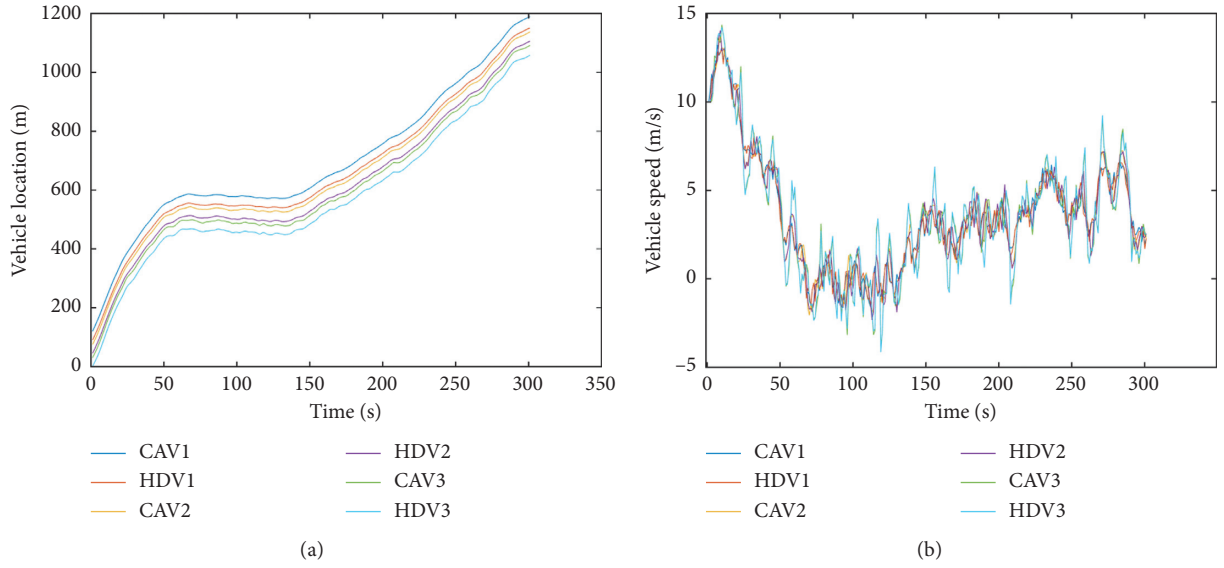
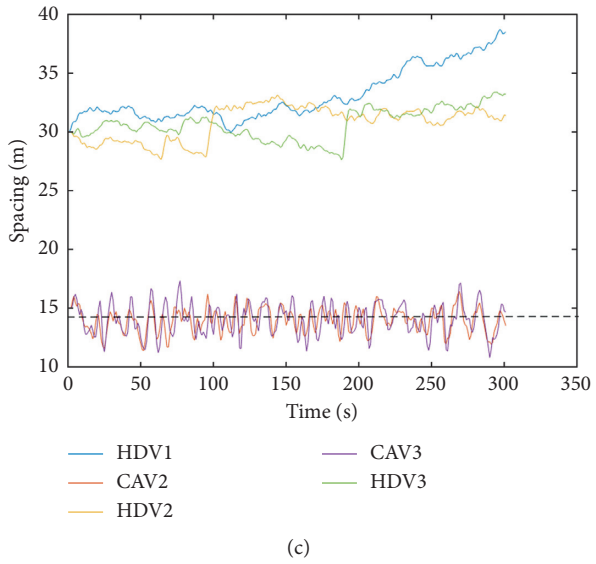
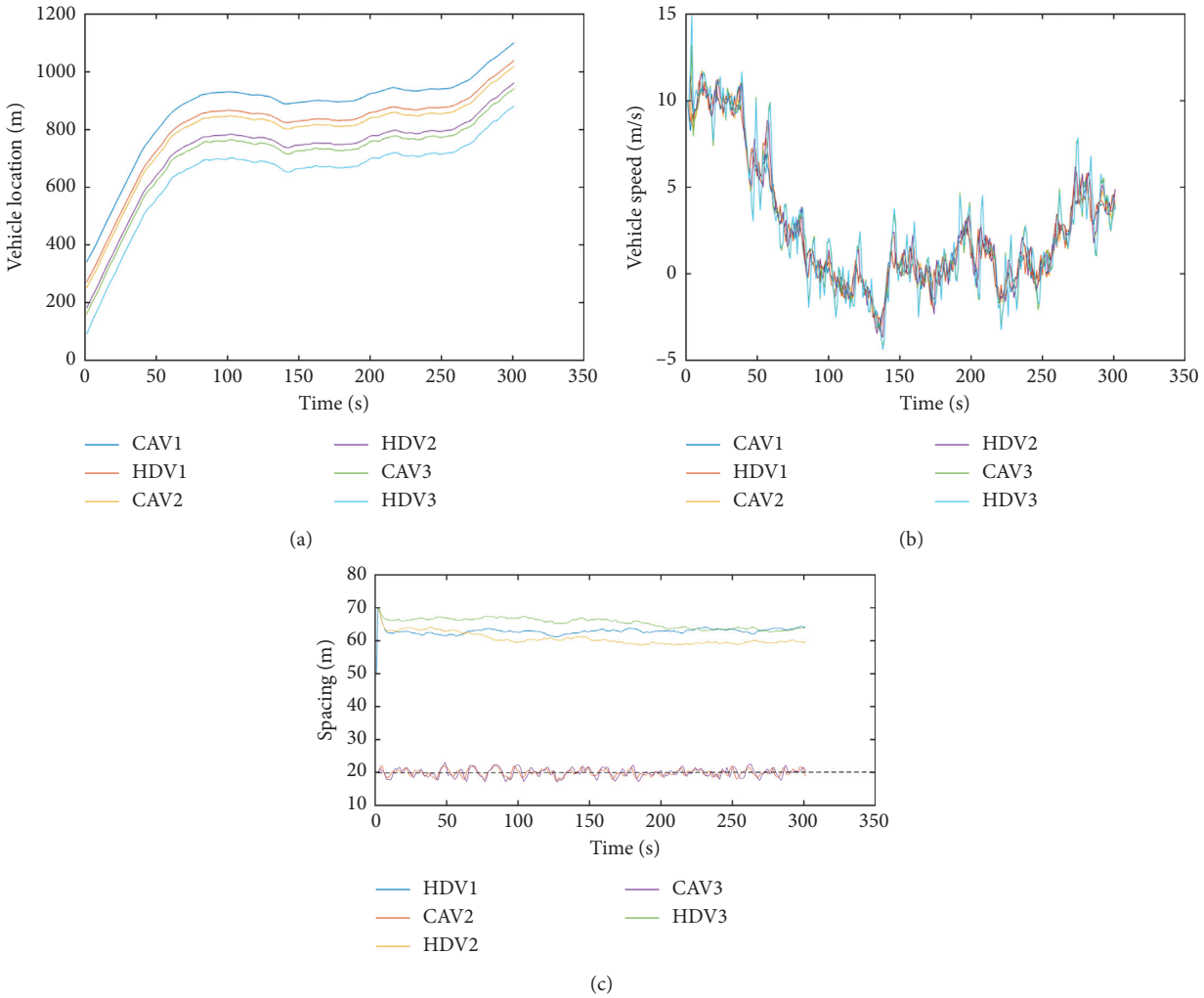



Figure 6: Continued.




(c)

Figure 6: The fleet in dense fog environment with the visibility being 50 m. (a) The location of formation. (b) The speed of formation. (c) The spacing of formation.




(a)

(b)

(c)

Figure 7: The fleet in dense fog environment with the visibility being 70 m. (a) The location of formation. (b) The speed of formation. (c) The spacing of formation.

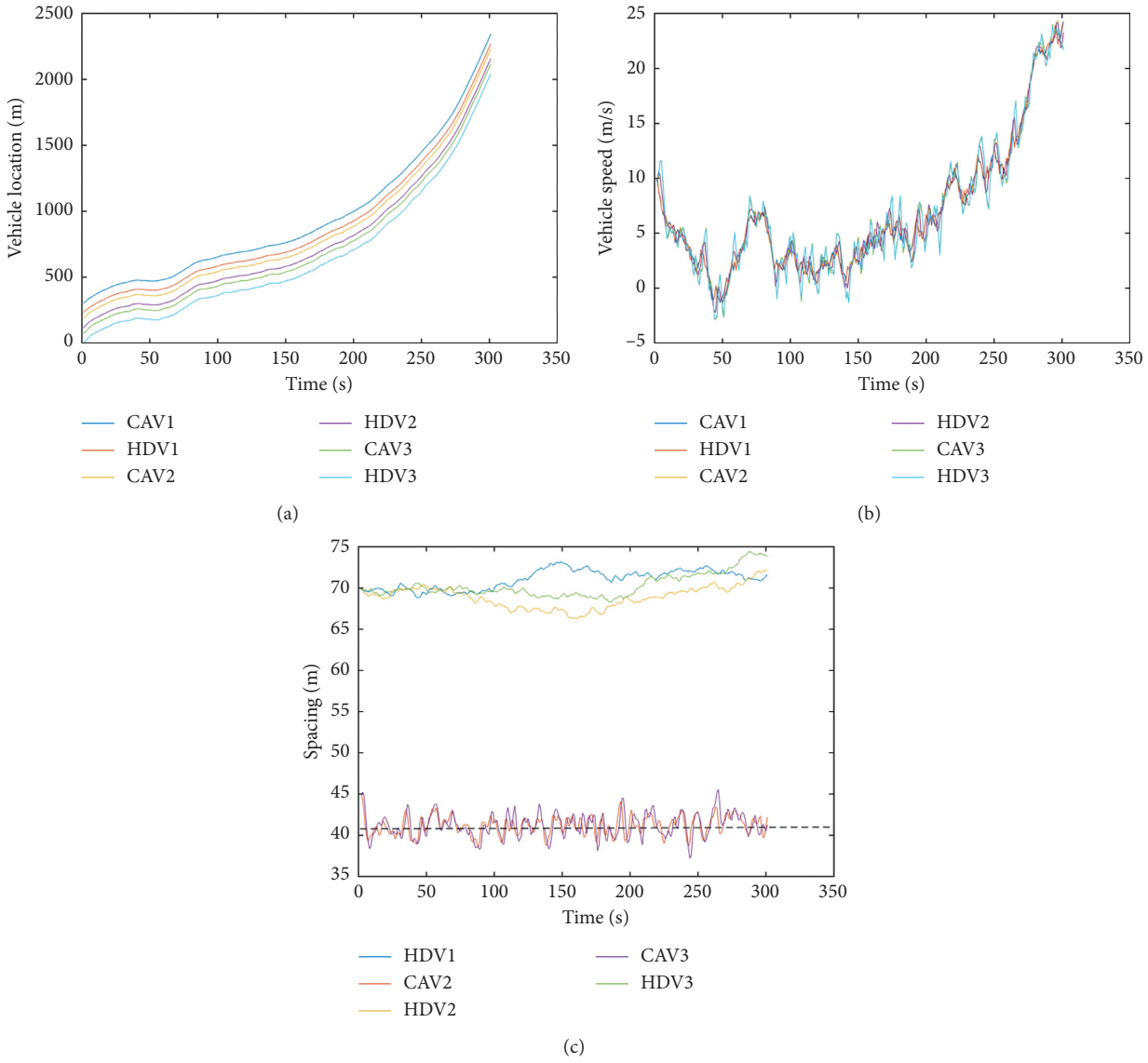



FIGURE 8: The fleet in dense fog environment with the visibility being 90 m. (a) The location of formation. (b) The speed of formation. (c) The spacing of formation.

CAVs. We call this as irregular alignment of the fleet, as shown in Figure 10. The car-following model between HDVs still used the NaScH CA-based car-following model in the experiments.

Based on the NaScH CA-based car-following model of HDV and the DMPC-based control model, the fleet in the experiments which have different regular alignment or irregular alignment were all can drive normally, as shown in Figures 11(a), 11(b), 11(d), 11(e), 11(g), and 11(h). However, as the number of HDVs between the CAVs within the fleet is increasing, the uncertainty of the fleet is also increasing. For example, by comparing among Figures 11(b), 11(e), and 11(h), it can be seen that as the number of HDVs between the CAVs within the fleet increases to two or three HDVs, the speed volatility of the fleet increases significantly.

In addition, the randomness of HDVs will be transmitted and accumulated between the adjacent HDVs. As shown in Figures 11(c), 11(f), 11(i), and (12), when the preceding vehicles in adjacent HDVs tend to travel at a higher speed and maintain an MSD from the preceding vehicles, the following vehicles in adjacent HDVs also have the same tendency. And the spacing fluctuation of the CAVs after the HDVs is clearly larger than that of the front CAVs. It is worth noting that the NaScH CA-based car-following model does not include aggressive drivers. If there are aggressive drivers in the fleet, it will be difficult to control and optimize the fleet. That is, the higher the penetration rates of CAV in the formation, the more effective the DMPC-based control model.

*4.4. The Stability of Fleet Control Verification.* Finally, the relationship between string stability and control quality is checked. The relationship checking is carried out under the dense fog environment with the visibility being 70m. The

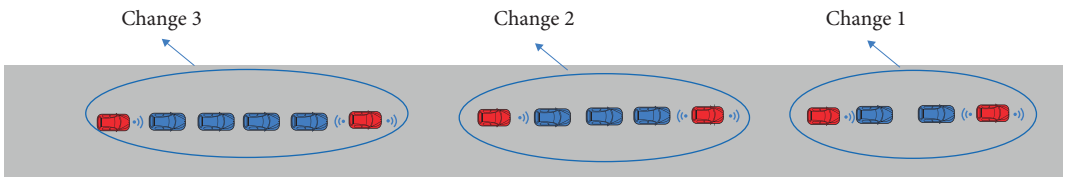



Figure 9: The regular alignment of the fleet.

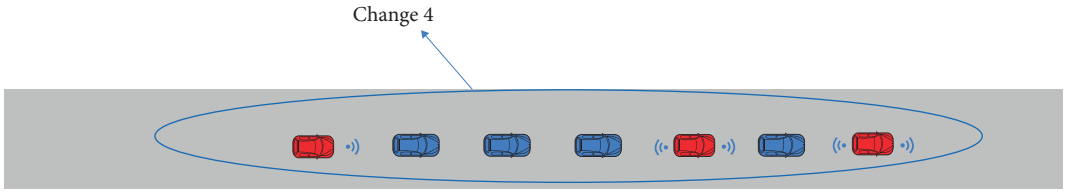



Figure 10: The irregular alignment of the fleet.

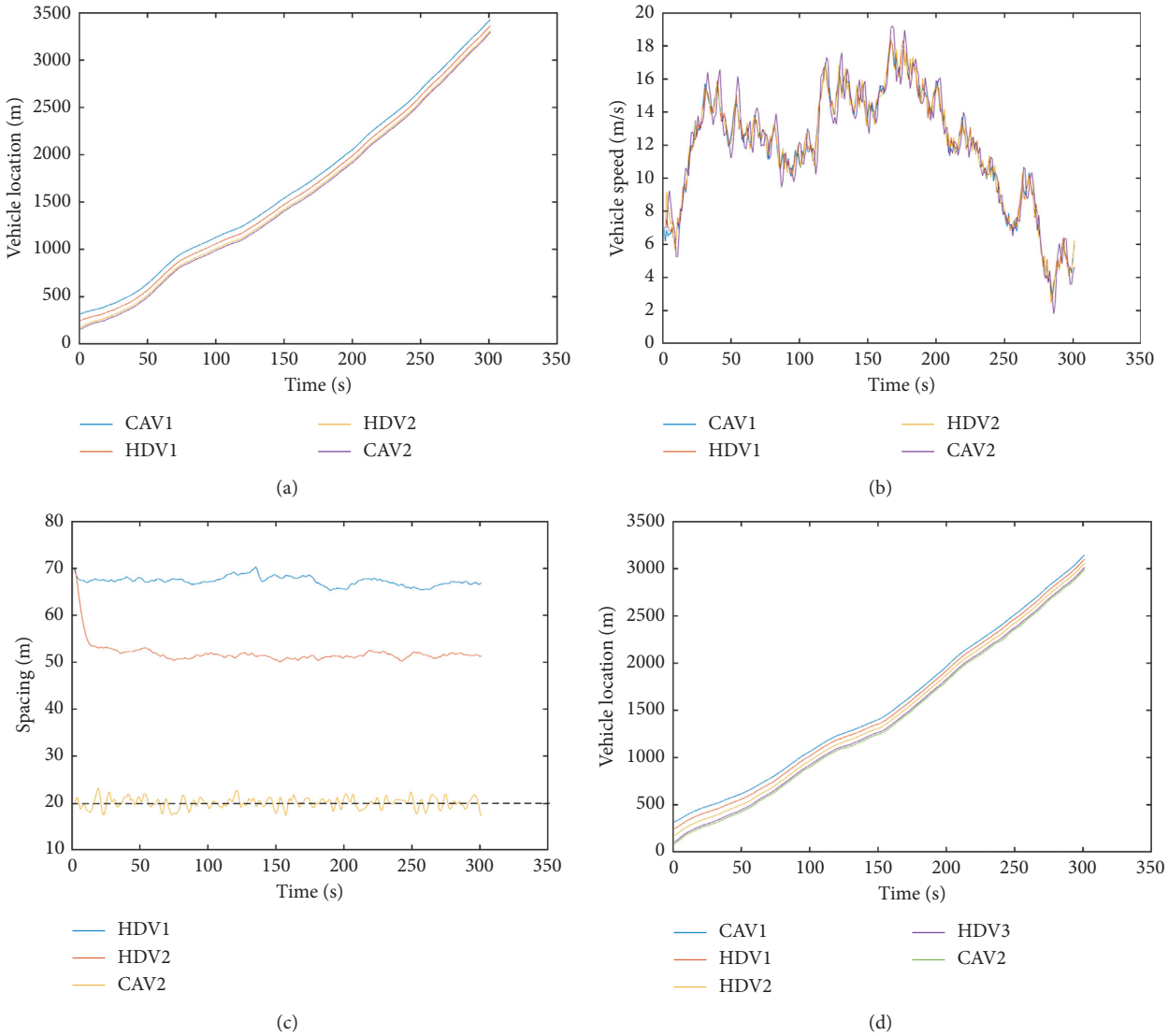



Figure 11: Continued.

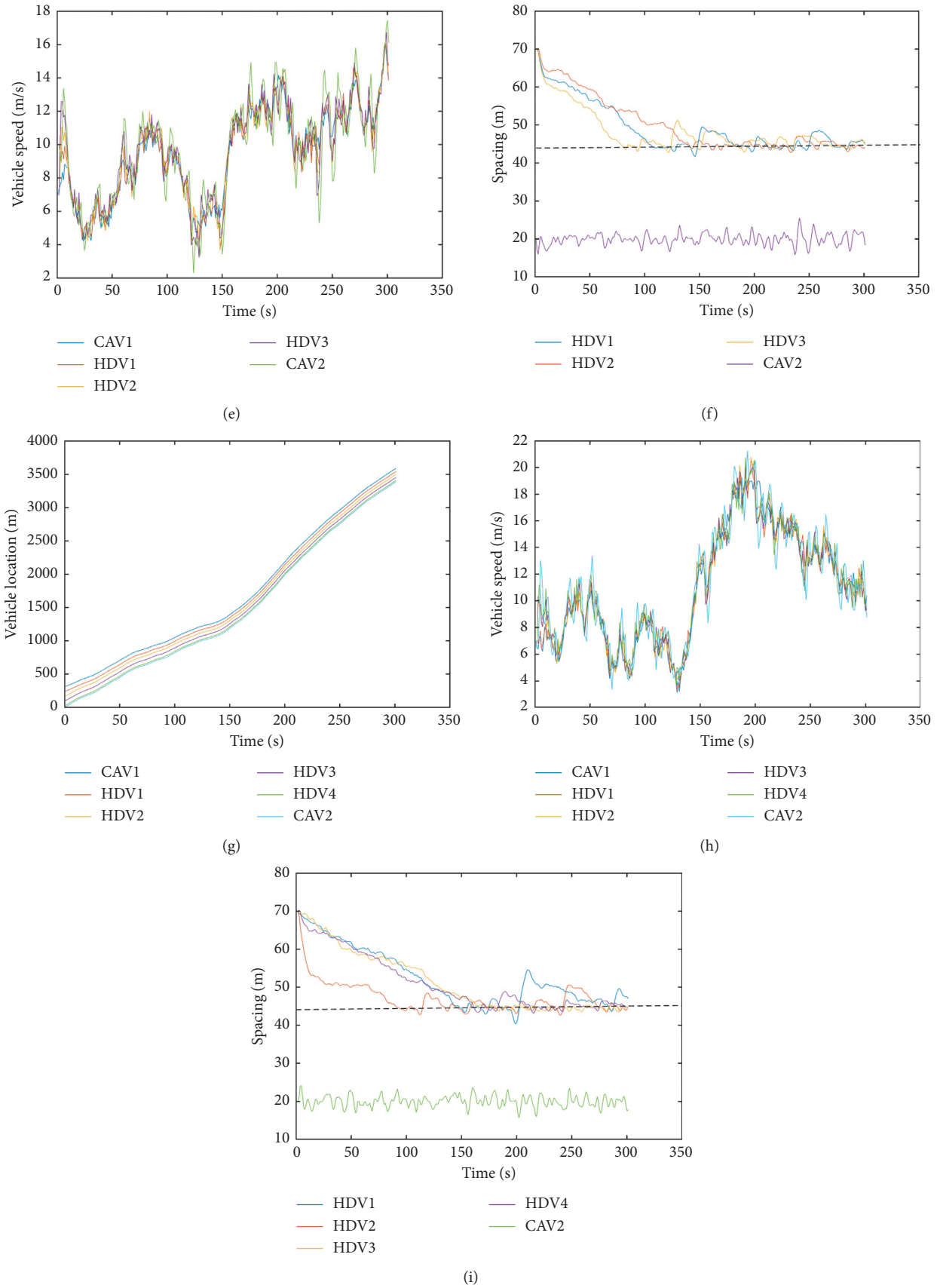



Figure 11: The driving situations of the fleet in change 1, change 2, and change 3. (a) The vehicle location of CAVs and HDVs (change 1). (b) The vehicle speed of CAVs and HDVs (change 1). (c) The spacing of formation (change 1). (d) The vehicle location of CAVs and HDVs (change 2). (e) The vehicle speed of CAVs and HDVs (change 2). (f) The spacing of formation (change 2). (g) The vehicle location of CAVs and HDVs (change 3). (h) The vehicle speed of CAVs and HDVs (change 3). (i) The spacing of formation (change 3).

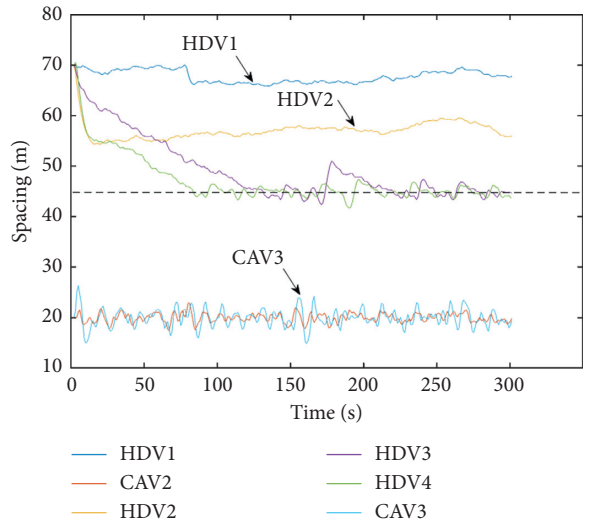



Figure 12: The spacing of formation (change 4).

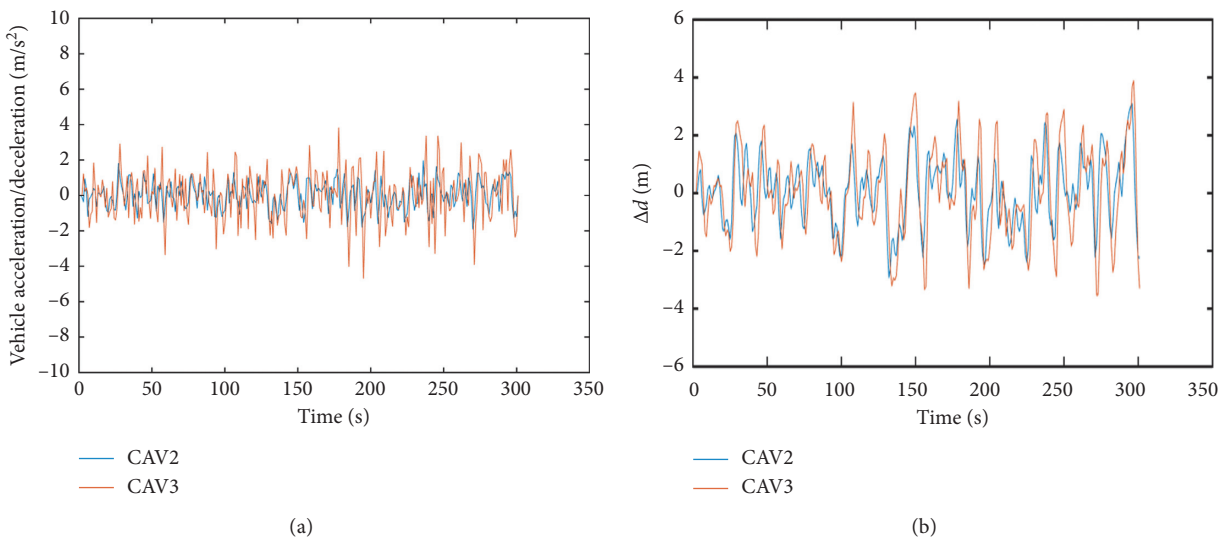



Figure 13: The fleet control under low stability requirement. (a) The acceleration of CAVs. (b) The control error of CAVs.

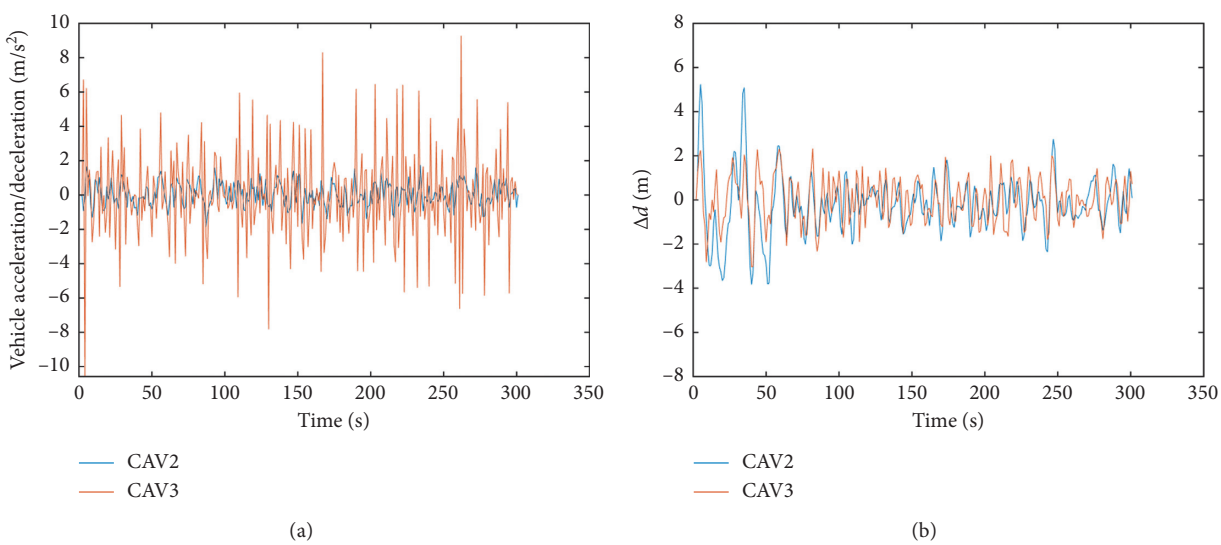



Figure 14: The fleet control under high stability requirement. (a) The acceleration of CAVs. (b) The control error of CAVs.

corresponding parameters are set as in Table 2. The experiments are checked under the low control quality and high control quality for the fleet in dense fog environment, respectively. The experiments results are shown in Figures 13 and 14.

As shown in Figure 13(a), CAV(3) and CAV(2) are not significantly difference in acceleration. However, the control error of CAV(3) is greater than the CAV(2), as shown in Figure 13(b). In Figure 14(a), CAV(3) is greater than the CAV(2) in acceleration, but the control error of CAV(3) is smaller than the CAV(2), as shown in Figure 14(b). The reason is that when the fleet is controlled under low stability requirements, the system has low control requirements for the CAV(3). CAV(3) performs the control strategy without pursuing the purpose of reducing errors in the fleet. In this situation, due to the randomness of HDV, the control error of CAV(3) will be greater than that of the CAV(2). However, when the fleet is controlled under high stability requirement, the CAV(3) must ensure the control error is smaller than the CAV(2). As to achieve this requirement and offset the randomness of HDVs, the CAV(3) needs to perform larger acceleration or deceleration. It can be concluded that the string stability of fleet is not easy to pursue. However, it is feasible to seek the balance of control efficiency and stability in fleet control based on DMPC by using a reasonable acceleration range for CAVs to make the fleet safety in passing through the dense fog environment.

## 5. Conclusions

This paper provides a control model for the fleet mixed with HDVs and CAVs in dense fog environment on highway based on distributed model prediction control (DMPC). Firstly, the state space model of CAVs is proposed to describe the state of CAVs within the fleet in dense fog environment. The paper used the discretization equation to describe the dynamical changing of state of CAVs in the fleet control. Secondly, a car-following model of HDVs in dense fog environment on highway is presented. It is used to describe the characteristic of the driver in dense fog environment as to follow the AGV within the scope of visibility and keep safety distance from the AGV. The simulation results show that the phenomenon shown in the simulation is consistent with the assumptions of HDV driver characteristics and fall into the relationship of speed and visibility as studied by Siebert and Wallis [45]. Thirdly, the distributed serial model predictive control (DMPC) model is used to control the fleet in dense fog environment. The controlling procedure is divided into two parts: (1) the fleet state prediction and (2) the fleet control based on DMPC. Predicting the states of the system based on DMPC focuses on the HDVs perception and prediction by CAVs and transfer between CAVs. Rolling optimization based on DMPC is used to optimize the local object of fleet control in dense fog environment with constraints. Rolling implementation of the control function is used to ensure the local equilibrium state of CAVs in the fleet control. Then, the proposed fleet control model is analyzed to meet the characteristics of the system asymptotic stability, and it can provide a nonstrict string stability during passing through dense fog environment on the highway. Finally, numerical experiments under different visibility in dense fog environment on highway are used to verify the effectiveness of the proposed model. The experiments results show that the proposed DMPC control algorithm can make CAVs consistent with HDVs under different visibilities. The spacing between CAVs and HDVs can be controlled within a predetermined safety distance. The effectiveness of the DMPC-based is highly related to the penetration rates of CAVs in the fleet and the alignment of the fleet. Furthermore, the systems of the fleet controlled based on the proposed model can guarantee the local asymptotic stability and global nonstrict string stability, simultaneously.

This paper provides the approach to control the fleet mixed with HDVs and CAVs in dense fog environment on highway. However, the method and model of the present study are not free from limitations. The first limitation is that that the present study only considers the impact of visibility. However, in dense fog environment, restrictions such as road friction and driver specificity also greatly affect the fleet driving on the highway. The second limitation is that the arrangement of the fleet vehicles in this study cannot guarantee the safety of the system that the HDV follows CAV. A third limitation is that this study did not consider the effect of the delay between CAVs in DMPC control.

Furthermore, there are some topics that remain to be studied. Further research work includes the following. (1) Other restrictions, such as communication delay, road friction coefficient, and drivers' specificity in dense fog environment, should be considered in the fleet. (2) Strict string stability of the fleet can be achieved through improving CAV prediction methods or communication technology. (3) At present, only the simulator-based tests are performed; the hardware experiments, in-vehicle tests, and scenarios' tests for the proposed modeling methods should be carried out in the future.

## Data Availability

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

## Conflicts of Interest

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

## Acknowledgments

The authors acknowledge the National Natural Science Foundation of China (Grant nos. 51408257 and 51308248), the Youth Scientific Research Fund of Jilin (Grant no. 20180520075JH), and the Science and Technology Project of Jilin Provincial Education Department (Grant nos. JJKH20170810KJ and JJKH20180150KJ) that partly support this work.

## References


[1] A. Alim, A. Joshi, F. Chen, and C. T. Lawson, "Techniques for efficient detection of rapid weather changes and analysis of their impacts on a highway network," *Geoinformatica*, vol. 24, p. 31, 2020.

[2] C. Li, *"Research on expressway traffic flow characteristics, traffic guidance and control under adverse weather,"* Beijing University of Technology, Beijing, China, Ph.D. Dissertation, 2015.

[3] Z. Lu, T. J. Kwon, and L. Fu, "Effects of winter weather on traffic operations and optimization of signalized intersections," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 6, no. 2, pp. 196–208, 2019.

[4] C. Xu, W. Wang, and P. Liu, "Identifying crash-prone traffic conditions under different weather on freeways," *Journal of Safety Research*, vol. 46, pp. 135–144, 2013.

[5] C. Chen, X. Zhao, H. Liu, G. Ren, and X. Liu, "Influence of adverse weather on drivers' perceived risk during car following based on driving simulations," *Journal of Modern Transportation*, vol. 27, no. 4, pp. 282–292, 2019.

[6] J. O. Brooks, M. C. Crisler, N. Klein et al., "Speed choice and driving performance in simulated foggy conditions," *Accident Analysis & Prevention*, vol. 43, no. 3, pp. 698–705, 2011.

[7] Z. Xuguang and G. Jianping, "Research on the fixation transition behavior of drivers on expressway in foggy environment," *Safety Science*, vol. 119, pp. 70–75, 2019.

[8] M. M. Ahmed, M. Abdel-Aty, J. Lee, and R. Yu, "Real-time assessment of fog-related crashes using airport weather data: a feasibility analysis," *Accident Analysis & Prevention*, vol. 72, pp. 309–317, 2014.

[9] Y. Wang, L. Liang, and L. Evans, "Fatal crashes involving large numbers of vehicles and weather," *Journal of Safety Research*, vol. 63, pp. 1–7, 2017.

[10] Q. Q. Shangguan, T. Fu, and S. Liu, "Investigating rear-end collision avoidance behavior under varied foggy weather conditions: a study using advanced driving simulator and survival analysis," *Accident Analysis and Prevention*, vol. 139, p. 14, Article ID 105499, 2020.

[11] J.-h. Tan, "Impact of risk illusions on traffic flow in fog weather," *Physica A: Statistical Mechanics and Its Applications*, vol. 525, pp. 216–222, 2019.

[12] F. Rosey, I. Aillerie, S. Espie, and F. Vienne, "Driver behaviour in fog is not only a question of degraded visibility-a simulator study," *Safety Science*, vol. 95, pp. 50–61, 2017.

[13] Y. Wu, M. Abdel-Aty, J. Park, and R. M. Selby, "Effects of real-time warning systems on driving under fog conditions using an empirically supported speed choice modeling framework," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 97–110, 2018.

[14] B. Zhai, J. Lu, Y. Wang, and B. Wu, "Real-time prediction of crash risk on freeways under fog conditions," *International Journal of Transportation Science and Technology*, 2020, In press.

[15] H. M. Hassan and M. A. Abdel-Aty, "Predicting reduced visibility related crashes on freeways using real-time traffic flow data," *Journal of Safety Research*, vol. 45, pp. 29–36, 2013.

[16] T. Winkle, C. Erbsmehl, and K. Bengler, "Area-wide real-world test scenarios of poor visibility for safe development of automated vehicles," *European Transport Research Review*, vol. 10, no. 2, p. 15, 2018.

[17] Y. Wu, M. Abdel-Aty, J. Park, and J. Zhu, "Effects of crash warning systems on rear-end crash avoidance behavior under fog conditions," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 481–492, 2018.

[18] Z. Li, Y. Li, P. Liu, W. Wang, and C. Xu, "Development of a variable speed limit strategy to reduce secondary collision risks during inclement weathers," *Accident Analysis & Prevention*, vol. 72, pp. 134–145, 2014.

[19] 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.

[20] S. Bahrami and M. J. Roorda, "Optimal traffic management policies for mixed human and automated traffic flows," *Transportation Research Part A: Policy and Practice*, vol. 135, pp. 130–143, 2020.

[21] G. Munster and A. Bohlig, "Auto outlook 2040: the rise of fully autonomous vehicles," 2020, https://loupventures.com/auto-outlook-2040-the-rise-of-fully-autonomous-vehicles.

[22] Ministry of Sicence and Technology of the People's Republic of China, *"Research And Development Guideline- Comprehensive Transportation And Intelligent Transportation,"* Ministry of Sicence and Technology of the People's Republic of China, Beijing, China, 2020, http://www.most.gov.cn/.

[23] C. Jianyang, "Dense fog and freeway traffic accidents," *Journal of Highway and Transportation Research and Development*, vol. 15, no. 2, pp. 38–40, 1998.

[24] W. Wang, Q. Cheng, C. Li, D. André, and X. Jiang, "A cross-cultural analysis of driving behavior under critical situations: a driving simulator study," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 62, pp. 483–493, 2019.

[25] R. Mu and T. Yamamoto, "An analysis on mixed traffic flow of conventional passenger cars and microcars using a cellular automata model," *Procedia-Social and Behavioral Sciences*, vol. 43, pp. 457–465, 2012.

[26] G. H. Bham and R. F. Benekohal, "A high fidelity traffic simulation model based on cellular automata and car-following concepts," *Transportation Research Part C: Emerging Technologies*, vol. 12, no. 1, pp. 1–32, 2004.

[27] L. Yang, J. Zheng, Y. Cheng, and B. Ran, "An asymmetric cellular automata model for heterogeneous traffic flow on freeways with a climbing lane," *Physica A: Statistical Mechanics and Its Applications*, vol. 535, Article ID 122277, 2019.

[28] S. Maerivoet and B. De Moor, "Cellular automata models of road traffic," *Physics Reports*, vol. 419, no. 1, pp. 1–64, 2005.

[29] S. Gong and L. Du, "Cooperative platoon control for a mixed traffic flow including human drive vehicles and connected and autonomous vehicles," *Transportation Research Part B: Methodological*, vol. 116, pp. 25–61, 2018.

[30] Y. Zhou, M. Wang, and S. Ahn, "Distributed model predictive control approach for cooperative car-following with guaranteed local and string stability," *Transportation Research Part B: Methodological*, vol. 128, pp. 69–86, 2019.

[31] 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.

[32] Y. Wu, M. Abdel-Aty, Q. Cai, J. Lee, and J. Park, "Developing an algorithm to assess the rear-end collision risk under fog conditions using real-time data," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 11–25, 2018.

[33] X. Yu, "Study of zero-order holder discretization in single input sliding mode control systems," in *Proceedings of the IEEE International Symposium on Circuits & Systems*, Florence, Italy, 2008.

[34] S. Barnett, "Linear system theory and design," *Automatica*, vol. 22, no. 3, pp. 385-386, 1986.

[35] F. Rosey, I. Aillerie, S. Espié, and F. Vienne, "Driver behaviour in fog is not only a question of degraded visibility - a simulator study," *Safety Science*, vol. 95, pp. 50–61, 2017.

[36] K. Gao, H. Tu, H. Shi, and Z. Li, "Effects of low visibility in haze weather on driving behaviors in different car following states," *Journal of Jilin University (Engineering Edition)*, vol. 47, no. 6, pp. 1717–1727, 2017.

[37] A. Hegyi, B. De Schutter, and H. Hellendoorn, "Model predictive control for optimal coordination of ramp metering and variable speed limits," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 185–209, 2005.

[38] W. B. Dunbar and R. M. Murray, "Distributed receding horizon control for multi-vehicle formation stabilization," *Automatica*, vol. 42, no. 4, pp. 549–558, 2006.

[39] J. Ploeg, E. Semsar-Kazerooni, G. Lijster, N. van de Wouw, and H. Nijmeijer, "Graceful degradation of cooperative adaptive cruise control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 488–497, 2015.

[40] D. Chen, J. Laval, Z. Zheng, and S. Ahn, "A behavioral car-following model that captures traffic oscillations," *Transportation Research Part B: Methodological*, vol. 46, no. 6, pp. 744–761, 2012.

[41] S. V. Rakovic, E. C. Kerrigan, K. I. Kouramas, and D. Q. Mayne, "Invariant approximations of the minimal robust positively Invariant set," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 406–410, 2005.

[42] Z. Sun and S. S. Ge, *Stability Theory of Switched Dynamical Systems (Communications and Control Engineering)*, Springer-Verlag, London, UK, 2011.

[43] A. Bacciotti and L. Rosier, *Lyapunov Functions and Stability in Control Theory (Communications and Control Engineering)*, Springer-Verlag, Berlin, Germany, 2005.

[44] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert, "Constrained model predictive control: stability and optimality," *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.

[45] F. W. Siebert and F. L. Wallis, "How speed and visibility influence preferred headway distances in highly automated driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 64, pp. 485–494, 2019.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8816615, 12 pages
https://doi.org/10.1155/2020/8816615

*Research Article*

# Data-Driven Modeling of Systemic Air Traffic Delay Propagation: An Epidemic Model Approach

**Shanmei Li**[1], **Dongfan Xie**[2], **Xie Zhang**[1], **Zhaoyue Zhang**[3] and **Wei Bai**[4]

[1]*College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China*
[2]*School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*
[3]*School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China*
[4]*Air Traffic Control Department, North China Air Traffic Management Bureau, Beijing 100621, China*

Correspondence should be addressed to Shanmei Li; amy820203@163.com

Received 10 April 2020; Revised 12 June 2020; Accepted 13 July 2020; Published 3 August 2020

Academic Editor: Chunjiao Dong



To better understand the mechanism of air traffic delay propagation at the system level, an efficient modeling approach based on the epidemic model for delay propagation in airport networks is developed. The normal release rate (NRR) and average flight delay (AFD) are considered to measure airport delay. Through fluctuation analysis of the average flight delay based on complex network theory, we find that the long-term dynamic of airport delay is dominated by the propagation factor (PF), which reveals that the long-term dynamic of airport delay should be studied from the perspective of propagation. An integrated airport-based Susceptible-Infected-Recovered-Susceptible (ASIRS) epidemic model for air traffic delay propagation is developed from the network-level perspective, to create a simulator for reproducing the delay propagation in airport networks. The evolution of airport delay propagation is obtained by analyzing the phase trajectory of the model. The simulator is run using the empirical data of China. The simulation results show that the model can reproduce the evolution of the delay propagation in the long term and its accuracy for predicting the number of delayed airports in the short term is much higher than the probabilistic prediction method. The model can thus help managers as a tool to effectively predict the temporal and spatial evolution of air traffic delay.

## 1. Introduction

Flight delays are one of the most important performance indicators of air transportation system. It has become increasingly more serious, which directly causes huge damage to the quality of civil aviation services, such as declines in operational safety, increases in operating costs, and more serious environmental pollution. Notably, commercial aviation players understand delays as the difference between scheduled and real times of departing or arriving fights [1]. According to the Federal Aviation Administration (FAA), a flight can be considered as delayed if the operation takes place 15 minutes after schedule [2].

The delay of an individual flight seems to be random at a glance. A flight delay may be transferred and amplified by consequent operations. Some delays that originate from upstream flights spread to downstream flights, which is particularly evident when an aircraft flies multiple flight legs. This phenomenon is defined as delay propagation (DP) [3–6]. Actually, DP causes delays to obey certain statistical laws [7] when long-term delay records for a large number of flights are taken into account, which form airport delay propagation. Consequently, a congested airport may propagate delays to connecting airports through the delayed flights, which eventually has an impact on the performance of a significant part of the entire network [8]. The delay can be magnified when it is examined in multiairport networks [9].

There are many factors affecting airport delays. The factors can be divided into two categories: propagation factors (PF) (airport delays caused by connected delayed airports) and nonpropagation factors (NPF) (airport delays caused by original factors, such as extreme weather and equipment trouble) [10]. As the number of flights increases,

increasingly more airport delays are caused by the PF [11]. If there are many delayed flights in one airport, the connected airports may become delayed, which can affect further operations in a cascade-like effect. Due to the complexity of air transportation, the mechanisms of delay propagation at the airport level are not fully understood. Therefore, research on the mechanism of delay propagation is timely yet challenging.

Delay spreading has received lots of attention from the Air Traffic Management (ATM) community during the last decade. Some studies [12–14] have established flight delay propagation models based on Bayesian Networks and analyzed the internal factors influencing air traffic delay propagation. Pablo Fleurquin [15] introduced an agent-based model that reproduces the delay propagation patterns observed in U.S. performance data and identified passenger and crew connectivity as the most relevant internal factor contributing to delay spreading. Qiu [16] constructed a joint distribution of continuous flight delays by using the 2D copula function. Wong and Tsai [17] established a cox proportional hazard model for flight delay propagation based on survival analysis theory. Kalfe and Zou [3] promoted a delay propagation pattern based on an econometric method and analyzed factors by using the Heckman two-step method.

To the best of our knowledge, gaps still remain in understandings of the delay propagation in airport networks. The process of delay propagation needs to be analyzed from a broader and network-based perspective because flight scheduling for airlines and airport operations is increasingly synchronized from the perspective of network operation. The linkage in the airport networks is the direct operation by the airlines linking the airports [18]. It is an essential feature of the network structure of the air traffic system. Therefore, the propagation dynamics cannot be understood without referring to the underlying complex network structure. Some scholars have used complex network theory to characterize transportation [19–22]. The initial studies [23, 24] have identified a high heterogeneity in the traffic sustained by each edge. Most of the airport network exhibits a heavy-tailed degree distribution, which is often well approximated for a significant range of values of degree $k$ by a power-law behavior ($P(k) \sim k^{-\gamma}$), from which the name "scale-free network" originated [25–29]. Further studies have used the complex network characteristics to explain the propagation of air traffic congestion and flight delay [29–34]. From a macroperspective, the airport delay is usually caused by air traffic congestion. Thus, the application of the complex network theory to the air traffic problems is feasible. And most of them focused on the delay propagation between sequence flights from the perspective of single airport operation.

As aforementioned, due to the large number of airports and complex interactions, the features of delay propagation cannot be understood based on the information of an individual airport. Complex network theory and its associated metrics and tools present an opposite approach to study the air transport system beyond what is offered by classical techniques. To further understand the effects of delay propagation, most of the existing achievements focused on the delay propagation between sequence flights from the perspective of single airport operation. In this paper, we propose a network-based approach to modeling airport delay propagation. There are several classical epidemic models in complex network theory, such as Susceptible-Infected (SI) model, Susceptible-Infected-Susceptible model (SIS), Susceptible-Infected-Recovered (SIR) model, and Susceptible-Infected-Recovered-Susceptible (SIRS) model. Because the propagation mechanisms of SIRS is the most similar to that of airport delay propagation(the details can be seen in Section 4.1), the SIRS model is utilized to understand the process of air traffic delay propagation in the context of an airport network and explain the spreading characteristics between different airports in this paper. The SIRS model has been normally used to simulate the process of how diseases [35], safety risk [36], or computer virus [37] spread.

To the best of the authors' knowledge, this study is the first to apply the SIRS model to air traffic delay propagation. Firstly, the metrics for measuring airport delays are introduced. Then, the fluctuation of airport delay is studied from different time scales based on complex network theory in order to find out the propagation factors, PF or NPF. Then, an integrated airport-based SIRS (ASIRS) model is developed. At last, the effective and accuracy of this model is demonstrated using empirical data of China.

The outline of the remainder of this paper is as follows. The data sources and a measurement of airport delay are provided in Section 2. Section 3 is devoted to fluctuations of the average flight delay and the determination of airport delay status. The ASIRS model is established in Section 4. The data-driven description of the ASIRS model is presented in Section 5. Finally, Section 6 contains the conclusions.

## 2. Data Description and Airport Delay Measurement

The dataset analyzed in this paper is provided by the Civil Aviation Administration of China (CAAC) and consists of all flight information of China from June to December in 2015. There are totally 93630 records of flights. The number of carriers in the data is 295. The information from the historical flight data consists of flight ID, date of flight, real and scheduled departure (arrival) times (Beijing Time), origin, and destination. A short sample of the original database is shown in Table 1.

There are 205 domestic airports in the database. A sample of airports are shown in Table 2.

The identification of the airport delay state is the first problem to be solved. As we know, the delay state of an airport is the concentrated performance of individual flight delay. Thus, the airport delay is measured by the delays of arriving and departing flights.

In our study, the normal release rate (NRR) and average flight delay (AFD) are considered to measure airport delay:

(1) The NRR of an airport is the ratio of the number of normal released flights to the total number of departure flights. According to the Federal Aviation

Table 1: A sample of the original database.

| Flight ID | Scheduled departure airport | Scheduled departure time | Scheduled arrival airport | Scheduled arrival time | Real departure airport | Real departure time | Real arrival airport | Real arrival time |
|---|---|---|---|---|---|---|---|---|
| CES2265 | ZLXY | 201510071525 | ZGHA | 201510071700 | ZLXY | 201510071606 | ZGHA | 201510071737 |
| CES2266 | ZGHA | 201510071740 | ZLXY | 201510071920 | ZGHA | 201510071848 | ZLXY | 201510072030 |
| CES2266 | ZLXY | 201510072045 | ZLIC | 201510072150 | ZLXY | 201510072202 | ZLIC | 201510072256 |
| HXA2689 | ZUCK | 201510071415 | ZLLL | 201510071555 | ZUCK | 201510071421 | ZLLL | 201510071549 |
| GCR7567 | ZGOW | 201510071530 | ZJHK | 201510071650 | ZGOW | 201510071533 | ZJHK | 201510071648 |
| JYH1032 | ZBMZ | 201510071625 | ZYCC | 201510071755 | ZBMZ | 201510071621 | ZYCC | 201510071750 |
| CHH7702 | ZGSZ | 201510072140 | ZBAA | 201510080040 | ZGSZ | 201510072225 | ZBAA | 201510080106 |
| UEA2755 | ZSLG | 201510071105 | ZYTX | 201510071250 | ZSLG | 201510071109 | ZYTX | 201510071241 |
| CSN6633 | ZHCC | 201510071035 | ZLXN | 201510071225 | ZHCC | 201510071051 | ZLXN | 201510071242 |

Table 2: A sample of airports.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Airport code | ZBAA | ZBMZ | ZGGG | ZGHA | ZGOW | ZGSZ | ZHCC | ZJHK |
| City | Beijing | Inner Mongolia | Guangzhou | Changsha | Shantou | Shenzhen | Zhengzhou | Haikou |
| Airport code | ZPPP | ZSHC | ZSLG | ZSPD | ZSSS | ZUCK | ZUUU | ZLXY |
| City | Kunming | Hangzhou | Lianyungang | Shanghai | Shanghai | Chongqing | Chengdu | Xi'an |

Administration (FAA), a flight is considered as abnormal if the departure operation takes place more than 15 minutes after schedule:

$$\mathrm{NRR} = \frac{N - n}{N}, \tag{1}$$

where $N$ represents the number of departure flights and $n$ represents the number of abnormal flights.

(2) The AFD of an airport is the ratio of the total delay time to the total number of all the departure and arrival flights of the airport:

$$\mathrm{AFD} = \frac{\sum_{i=1}^{N+M} d_i}{N + M}, \tag{2}$$

where $M$ represents the number of arrival flights and $d_i$ represents the delay of flight $i$.

The NRR and AFD are counted every 1 hour using the above database. Partial statistical results are shown in Tables 3 and 4.

## 3. Fluctuations of Airport Delay and Determination of Airport Delay Status

*3.1. Fluctuations of Airport Delay.* To explore the propagation laws of airport delay, the delay fluctuations should be studied from the perspective of airport networks. Fluctuations can be considered by investigating the coupling between the average flux and the fluctuations, which is actually the mean and standard deviation analysis, as developed in [38–42]. It is found that the standard deviation $\sigma_i$ and the average flux $\langle f_i \rangle$ on individual nodes obey a unique scaling law as

$$\begin{aligned} \sigma_i &\propto \langle f_i \rangle^{\alpha}, \\ \langle f_i \rangle &= \sum_{t=1}^{T} f_i^{\Delta t}(t), \\ \sigma_i &= \sqrt{\sum_{t=1}^{T} \left( f_i^{\Delta t}(t) - \langle f_i \rangle \right)^2}, \end{aligned} \tag{3}$$

where $f_i^{\Delta t}(t)$ denotes the flux of node $i$ in time interval $[(t - \Delta t), t]$.

As the strength of the external driving force increases, the value of $\alpha$ gradually increases.

A method to separate the internal dynamics from the external fluctuations of complex systems is also promoted in [38–42].

The dynamical variable $f_i(t)$ can be separated into two components:

$$f_i(t) = f_i^{\mathrm{int}}(t) + f_i^{\mathrm{ext}}(t), \tag{4}$$

where $f_i^{\mathrm{ext}}(t)$ is generated by external factors and $f_i^{\mathrm{int}}(t)$ is generated by internal factors. They can be described as follows:

$$\begin{aligned} f_i^{\mathrm{ext}}(t) &= A_i \sum_{i=1}^{N} f_i(t), \\ f_i^{\mathrm{int}}(t) &= f_i(t) - A_i \sum_{i=1}^{N} f_i(t), \\ A_i &= \frac{\sum_{t=1}^{T} f_i(t)}{\sum_{i=1}^{N} \sum_{t=1}^{T} f_i(t)}. \end{aligned} \tag{5}$$

Table 3: Samples of average flight delay (minute).

| | 0:00–1:00 | 1:00–2:00 | 2:00–3:00 | 3:00–4:00 | 4:00–5:00 | 5:00–6:00 | 6:00–7:00 | 7:00–8:00 |
|---|---|---|---|---|---|---|---|---|
| ZGSZ | 152.5 | 104 | 113 | 61.78 | 95 | 43.66 | 13.84 | 24.2187 |
| ZJHK | 0 | 0 | 0 | 0 | 0 | 0 | 28.37 | 14.5714 |
| ZPPP | 86.5 | 22 | 22 | 0 | 0 | 0 | 9.5 | 22.1395 |
| ZBAA | 40.36 | 31.27 | 30.75 | 27 | 52 | 89.91 | 31.5 | 30.7857 |
| ZLXY | 39 | 162 | 68 | 0 | 0 | 0 | 0 | 14.76 |

Table 4: Samples of airport normal release rate (%).

| | 0:00–1:00 | 1:00–2:00 | 2:00–3:00 | 3:00–4:00 | 4:00–5:00 | 5:00–6:00 | 6:00–7:00 | 7:00–8:00 |
|---|---|---|---|---|---|---|---|---|
| ZGSZ | 81.81 | 87.5 | 75 | 40 | 80 | 25 | 24 | 5.88 |
| ZJHK | 100 | 0 | 0 | 0 | 0 | 0 | 27.27 | 6.66 |
| ZPPP | 33.33 | 50 | 0 | 0 | 0 | 0 | 33.33 | 10.41 |
| ZBAA | 6.66 | 21.42 | 7.69 | 20 | 50 | 16.66 | 28.94 | 5.40 |
| ZLXY | 40 | 0 | 50 | 0 | 100 | 0 | 0 | 24.24 |

Furthermore, whether or not the fluctuations are mainly internally or externally imposed can be determined:

$$\eta_i = \frac{\sigma_i^{\text{ext}}}{\sigma_i^{\text{int}}}, \tag{6}$$

$$\sigma_i^{\text{ext}} = \sqrt{\langle f_i^{\text{ext}}(t)^2\rangle - \langle f_i^{\text{ext}}(t)\rangle^2}, \tag{7}$$

$$\sigma_i^{\text{int}} = \sqrt{\langle f_i^{\text{int}}(t)^2\rangle - \langle f_i^{\text{int}}(t)\rangle^2}. \tag{8}$$

If $\eta_i \gg 1$, the system dynamics are dominated by the network-wide factors, while for $\eta_i \ll 1$, local dynamics overshadow the network-imposed changes.

We aggregate the data and carry out the scaling law analysis at different time scales. In our study, $\langle f_i^{\Delta t}\rangle$ represents the AFD in the time scale $\Delta t$ of airport $i$ and $\sigma_i^{\Delta t}$ represents the standard deviation.

Figure 1 shows the relationship between $\langle f_i^{\Delta t}\rangle$ and $\sigma_i^{\Delta t}$, with time scale $\Delta t = 1$ h, 3 h, and 6 h. The scaling law between $\sigma$ and $\langle f\rangle$ can be clearly observed.

It can be seen that the value of the scaling exponent $\alpha$ increases as $\Delta t$ increases, suggesting that the system may have an inhomogeneous influence, as pointed out by Eisler and Kertesz [43]. The reason for this result is that the fluctuations of AFD are due mainly to network-wide factors such as the PF when $\Delta t$ is much bigger; on the contrary, the fluctuations of AFD are due mainly to local factors such as the NPF when $\Delta t$ is much smaller. When $\Delta t$ is much bigger, airport delays may be caused by connected delayed airports. Some delays that originate from upstream flights spread to downstream flights, which is particularly evident when an aircraft flies multiple flight legs. When $\Delta t$ is much smaller, airport delays may be caused by original factors, such as extreme weather and equipment trouble. Additionally, the ratio $\eta_i$ for the 1 hour interval is calculated using the above method. The result reveals that the average $\eta_i$ is 5.665353, which shows that the dynamic of airport delay of every 1 hour is dominated by PF. And the larger the time scale is, the bigger the value of $\eta_i$ is. When the time scale is 3 hours, the value of $\eta_i$ is 6.431234. When the time scale is 6 hours, the value of $\eta_i$ is 8.534778. Thus, we have to study how does the delay originating from an airport propagate to other airports in a large time scale.

*3.2. Determination of Airport Delay Status.* NRR and AFD are used to determine if the airport is in the delay state. The specific criteria is as follows:

If $\{\text{AFD}_i(t) > \xi_1 \& \text{NRR}_i(t) < \xi_2\}$, airport $i$ is in the delay state in the time interval $t$.

To explore the characteristics of airport delay propagation, the value of the time interval $t$ is 1 hour, as the dynamic of airport delay of every 1 hour is dominated by PF. When the delay propagates in the airport network, the airport delay is usually severe. The airport delay deduced by NPF is much small in the time scale of one hour. Thus, in order to eliminate the influence of NPF, the values of threshold $\xi_1$ should be relatively large and the values of threshold $\xi_2$ should be relatively small. Here, $\xi_1 = 15$ min and $\xi_2 = 0.7$.

## 4. Epidemic Model of Airport Delay Propagation

*4.1. Airport-Based Susceptible-Infected-Recovered-Susceptible Model.* From the discrimination of delay propagation, we find that the epidemic model in a complex network is a valuable research tool for the exploration of fundamental laws and trends of delay propagation in airport networks. There are three kinds of individuals in the SIRS model: susceptible ones ($S$), infected ones ($I$), and recovered ones ($R$). The susceptible ones are currently in a healthy state, and when they contact the source of the infection, they will become infected ones with an infection rate $\alpha$. The infected ones are unhealthy ones, and they can infect susceptible ones. The infected ones will be cured with a cure rate $\beta$ and become recovered ones, and the recovered ones become healthy ones with an immunity ability. The immunity will disappear under some certain situations, and the recovered ones will become susceptible ones with an immunity-loss rate $\gamma$. The infectious mechanism is described in Figure 2.

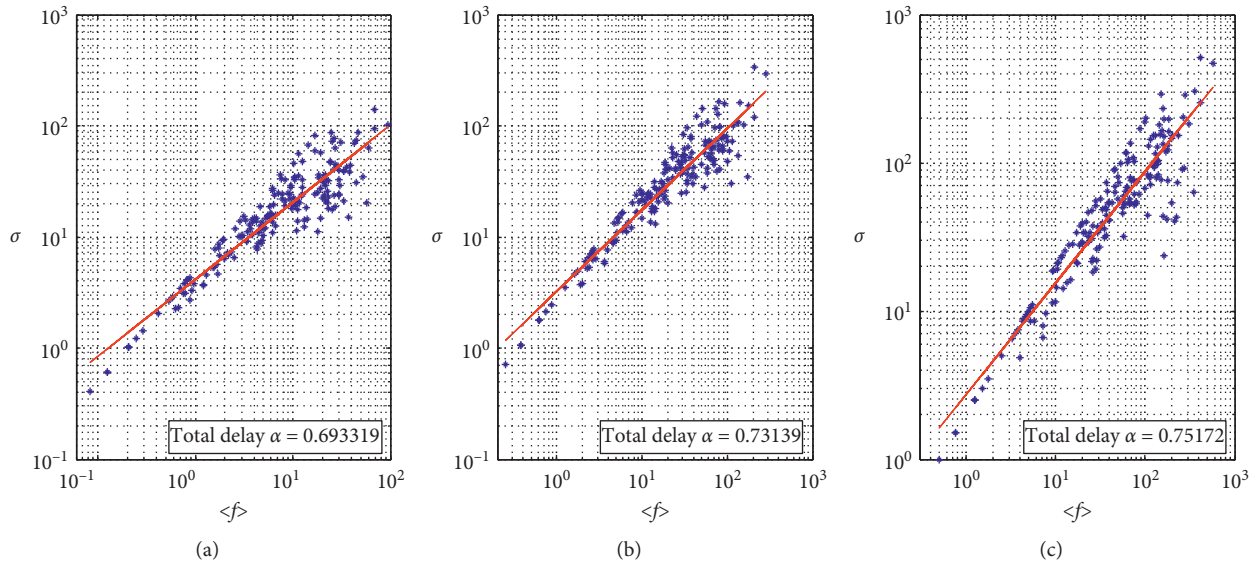



FIGURE 1: Scaling relations between fluctuations $\sigma$ and AFD $\langle f \rangle$ at different time scales. (a)$\Delta t = 1$ h. (b)$\Delta t = 3$ h. (c)$\Delta t = 6$ h. The estimated values of scaling exponent $\alpha$ are also indicated.

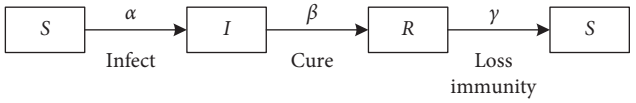



FIGURE 2: Infectious mechanism of SIRS.

In an airport network, the original airport delay may be due to capacity reduction, airport equipment trouble, and extreme weather. In the process of delay propagation for resource-shared flights, delays are propagated from an upstream flight at the departure airport to the arrival airport. As shown in Figure 3, the airports with "delay root" represent the susceptible ones, the airports with "delay propagation" represent the infected ones, and the airports with "delay termination" represent the recovered ones. The propagation of airport delay has traditionally been described as graphs with vertices representing airports and edges representing connectivity. When the delay is serious in one airport, the delay of its connected airports may be increased due to the delay spreading. Furthermore, the delay of spread airports may be absorbed in the subsequent operations, and they would not be influenced again by the same initial airport delay. However, they may be affected by another original airport delay later. Because of the complexity of airport networks [32], the evolution of delay within them possesses the characteristics of propagation in complex networks.

From the above analysis, we find that the infectious mechanism of SIRS is similar to the propagation characteristics of airport delay discussed before. Suppose there are three kinds of airports in the network at time $t$: non-delayed airports ($S$) which are easily infected, delayed airports ($I$), recovered airports ($R$) which used to be delayed but are back to normal. The recovered airports only have immunity to the current delay spread and may become susceptible ones later. As the stochastic process is applied to all flights, the airports are affected by probability.

The dynamics of ASIRS model can be written as

$$\begin{cases} \dfrac{dS(t)}{dt} = -\alpha S(t) I(t) + \gamma R(t), \\ \dfrac{dI(t)}{dt} = \alpha S(t) I(t) - \beta I(t), \\ \dfrac{dR(t)}{dt} = \beta I(t) - \gamma R(t), \\ S(t) + I(t) + R(t) = 1, \end{cases} \tag{9}$$

where $S(t), I(t)$, and $R(t)$ represent the fraction of susceptible airports, infected airports, and recovered airports, respectively, at time $t$; $\alpha$ is the infection rate; $\beta$ is the cure rate; and $\gamma$ is the immunity-loss rate.

Assume that the proportion of infected airports, susceptible airports, and recovered airports at the initial moment $t_0$ is $I_0 (I_0 > 0)$, $S_0 (S_0 > 0)$, and $R_0 (R_0 = 0)$, respectively:

$$\begin{cases} \dfrac{dI(t)}{dt} = \alpha S(t) I(t) - \beta I(t), \quad I(0) = I_0, \\ \dfrac{dS(t)}{dt} = -\alpha S(t) I(t), \qquad\qquad S(0) = S_0. \end{cases} \tag{10}$$

The phase trajectory of the ASIRS model is analyzed. The S-I plane is called the phase plane, and the domain of the phase trajectory is $(S, I) \in D$:

$$D = \{(S, I) \mid S(t) \geq 0, I(t) \geq 0, I(t) + R(t) \leq 1\}. \tag{11}$$

Let $\delta = \beta/\alpha$. The following equation can be obtained:

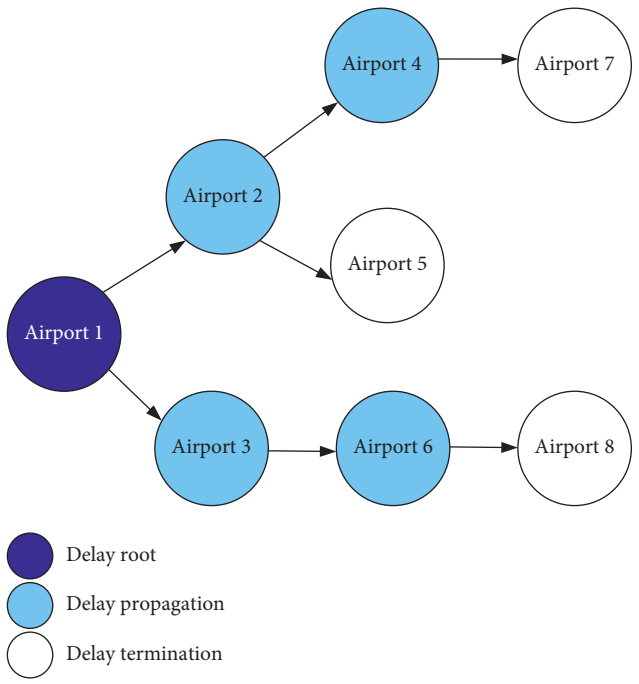



Figure 3: Delay propagation tree.

$$I(t) = I_0 + S_0 + \delta \quad \ln\frac{S(t)}{S_0} - S(t). \tag{12}$$

The phase trajectory diagram is shown in Figure 4.

When $t \longrightarrow \infty$, the limit values of $I(t), S(t)$, and $R(t)$ are $I_\infty, S_\infty$, and $R_\infty$, respectively:

(1) No matter how the initial values of $S_0$ and $I_0$ change, the airport delay situation will eventually disappear, $I_\infty = 0$.

(2) In equation (10), let $I(t) = 0$. The value of $S_\infty$ can be calculated, which is the root value of equation (11) in the range of $(0, \delta)$. $S_\infty$ is the abscissa of the intersection point between the phase trajectory and transverse axis in the range of $(0, \delta)$:

$$I_0 + S_0 + \delta \quad \ln\frac{S(t)}{S_0} - S_\infty = 0. \tag{13}$$

(3) If $S_0 < \delta$, $I(t)$ increases first; if $S(t) = \delta$, $I(t)$ reaches its maximum and then decreases to zero. At the same time, $S(t)$ is monotonically reduced to $S_\infty$:

$$I_m = I_0 + S_0 - \delta\left(\ln\frac{1}{\delta}S_0 + 1\right), \tag{14}$$

where $I_m$ is the maximum of $I$.

According to the above analysis, the following conclusions can be drawn:

(1) If $\alpha S(t) < \beta$, $I(t)$ increases and the airport delay will spread to more airports

(2) If $\alpha S(t) > \beta$, $I(t)$ decreases, the delay situation of the airport network will be alleviated, and the airport delay will not spread to others

(3) If $\alpha S(t) = \beta$, $I(t)$ reaches the maximum and the delay situation of the airport network is the most serious

*4.2. Parameter Analysis.* As mentioned above, the ASIRS model contains three parameters: $\alpha$,$\beta$, and $\gamma$. To investigate the change of $I(t), S(t)$, and $R(t)$ over time, suppose $\alpha = 1$, $\beta = 0.1$, and $\gamma = 0.1$. The change of $I(t), S(t)$, and $R(t)$ is shown in Figure 5.

*4.2.1. Analysis of Parameter $\alpha$.* First, the influence of $\alpha$ on airport delay propagation is discussed. Figure 6 shows the changes of $I(t), S(t)$, and $R(t)$ under different values of $\alpha$ with the assumptions that $\beta = 0.1$ and $\gamma = 0.1$.

As seen, the higher $\alpha$ is, the earlier $I(t)$ and $R(t)$ reach their peak value and the higher the peak value is. The higher $\alpha$ is, the earlier $S(t)$ reaches its minimum and the lower the minimum is. Thus, it can be concluded that, as the speed at which airport delay propagates increases, more airports will be infected more quickly, and more airports will be recovered.

*4.2.2. Analysis of Parameter $\beta$.* Figure 7 shows the changes of $I(t), S(t)$, and $R(t)$ under different values of $\beta$ with the assumption that $\alpha = 1$ and $\gamma = 0.1$.

When the value of $\beta$ becomes smaller, the peak of $I(t)$ appears later and the trough of $S(t)$ appears later also. The trends of $R(t)$ with $\beta = 0.5/1$ and $\beta = 0.1/0.2$ are not the same. Thus, $R(t)$ is much more complex than $I(t)$ and $S(t)$.

*4.2.3. Analysis of Parameter $\gamma$.* Figure 8 shows the changes of $I(t), S(t)$, and $R(t)$ under different values of $\gamma$ with the assumption that $\alpha = 1$ and $\beta = 0.1$.

The higher $\gamma$ is, the earlier $I(t)$ reaches its peak value and the higher the peak value is. The lower $\gamma$ is, the later $S(t)$ reaches its minimum, the lower the minimum is, and the higher $R(t)$ is. Parameter $\gamma$ has little effect on $I(t)$ and $S(t)$ in the early stage and mainly affects the later stage.

## 5. Case Study

*5.1. Statistical Calculation of $I(t), S(t)$, and $R(t)$.* According to Section 3.2, we calculate $I(t), S(t)$, and $R(t)$ based on the following criteria:

(1) If $\{\mathrm{NRR}_i(t) > 15\,\mathrm{min}\ \&\ \mathrm{ADF}_i(t) < 0.7\}$, airport $i$ is infectious at time $t$

(2) If $\{\mathrm{NRR}_i(t) \le 15\,\mathrm{min} \mid \mathrm{ADF}_i(t) \ge 0.7\}\&\{\mathrm{NRR}_i(t-1) > 15\,\mathrm{min}\ \&\ \mathrm{ADF}_i(t) < 0.7\}$, the airport is recovered at time $t$ and infectious at time $t-1$

(3) Apart from the above situations, the airport is susceptible and can be easily infected by infectious airport

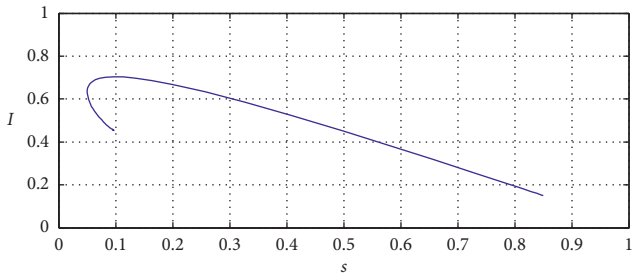



Figure 4: Phase trajectory diagram.

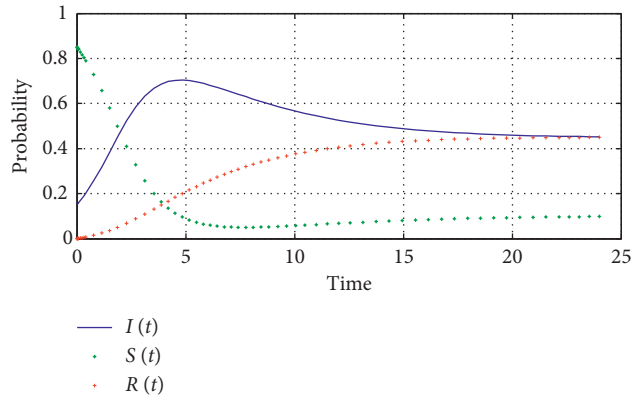



Figure 5: Change of $I(t), S(t)$, and $R(t)$.

*5.2. Determination of ASIRS Parameters.* The hourly states of all the airports in the network can be identified based on the above criteria using Tables 1 and 2. Taking the airport ZBNY in October 1, 2015, for example, Figure 9(a) shows the time varying state of airport ZBNY. It can be seen that the airport ZBNY is infected in {6 : 00–10 : 00, 14 : 00–20 : 00}, susceptible in {0 : 00–6:00, 11 : 00–14 : 00, 21 : 00–24 : 00}, and recovered in {10 : 00–11 : 00, 20:0–21 : 00}. We also investigate the air traffic flow of ZBNY, which is shown in Figure 9(b). Comparing Figures 9(a) and 9(b), we find that the larger the traffic flow is, the more the airport tends to be infected. It should be noted that, although the airport has a small number of flights, the flight delay is serious, which may be deduced by the delay propagation in the airport network.

In addition, we find that there are 7 infectious airports in 4 : 00–5:00: ZBAA, ZGSZ, ZUCK, ZGGG, ZSHC, ZSPD, and ZHCC. Six of these are the top 10 airports by throughput in China (the throughput of airports is provided by the CAAC), as shown in Table 5. There are many flights in the six airports. Large departure delays may influence the operations of the arrival airports, and the flights will arrive at or depart this airport with big delays. Thus, the phenomenon of delay propagation appears.

The simulation model is ASIRS established in Section 4.1, as shown in equation (7). The simulation method is that we use the real information of flights in China to calculate the values of $I(t), S(t)$, and $R(t)$ and then calibrate the parameters of $\alpha$, $\beta$, and $\gamma$. Thus, the ASIRS model for simulating the real flight data can be obtained.

Next, we calculate the values of $I(t), S(t)$, and $R(t)$. Partial statistical results are shown in Table 6.

The parameters in the ASIRS model are calibrated based on the statistical values of $I(t), S(t)$, and $R(t)$ by using the numerical simulation method. For every day's traffic situation, we can always build an excellent ASIRS model. Taking October 1, 2015 (there are 16804 scheduled flights connecting 199 different commercial airports), for example, we find that the model fits the actual operation situation to the highest level when $\alpha = 0.47, \beta = 0.175$, and $\gamma = 0.45$.

*5.3. Accuracy of the Long-Term Simulation of Delay Propagation.* The values of $I(t), S(t)$, and $R(t)$ for October 1, 2015, in Figure 10 are the actual values, which are counted based on the database. The values of $I(t), S(t)$, and $R(t)$ in Figure 11 are the simulations of the ASIRS model for that day.

Comparing Figure 10 with Figure 11, it can be seen that the long-term trends of the actual and predicted values of $I(t), S(t)$, and $R(t)$ are similar, although there are some differences between local values. Thus, from a qualitative point of view, we can conclude that the ASIRS model is reasonable and can describe the real situation of airport delay propagation to a certain extent.

To further examine the application of the ASIRS model, we also study the characteristics of airport delay propagation derived from the degree distribution.

First, the ASISR models for the peak period of 12 : 00–19 : 00, October 1, 2015, and the nonpeak period of 00 : 00–07 : 00, October 1, 2015, are constructed. The parameters' values of the two models are determined based on trial calculation using the actual data:

(1) 12 : 00–19 : 00: $\alpha = 0.4, \beta = 0.25, \gamma = 0.8$

(2) 00 : 00–07 : 00: $\alpha = 0.38, \beta = 0.28, \gamma = 0.6$

The $\alpha$ and $\gamma$ values of 12 : 00–19 : 00 are bigger than those of 00 : 00–07 : 00, and the $\beta$ values of 12 : 00–19 : 00 are smaller than those of 00 : 00–07 : 00. According to the parameter analysis mentioned above, the conclusion can be obtained that airport delays propagate faster and wider in 12 : 00–19 : 00 than in 00 : 00–07 : 00.

The relative cure rate $\rho$ is the ratio of the infection rate $\alpha$ to the cure rate $\beta$:

$$\rho = \frac{\alpha}{\beta}, \tag{15}$$

where $\rho$ reflects the recovery speed of delay propagation. The larger $\rho$ is, the more the delay propagation tends to alleviate. The value of $\rho$ for 12 : 00–19 : 00 is 0.625; the value of $\rho$ for 00 : 00–07 : 00 is 0.737. The relative recovery rate of delayed propagation in 00 : 00–07 : 00 is higher than that in 12 : 00–19 : 00. This is consistent with the actual situation.

Next, the degree $k$ and the degree distribution $p(k)$ of the Chinese airport network in the above two periods are counted. The degree of airport A represents the number of airports with flights to or from A in the statistical time period. Thus, the same airport may have different degrees at

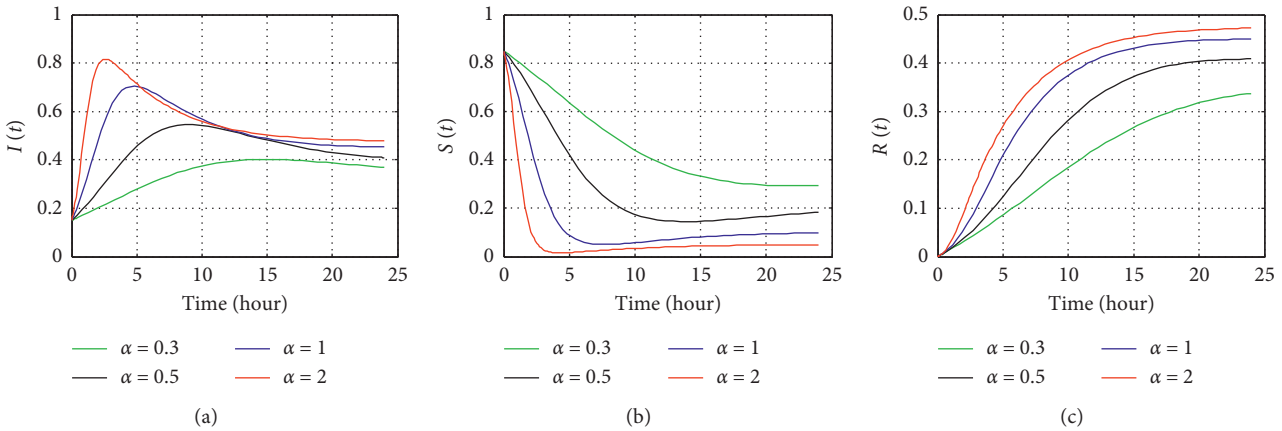



Figure 6: $I(t), S(t)$, and $R(t)$ under different $\alpha$.

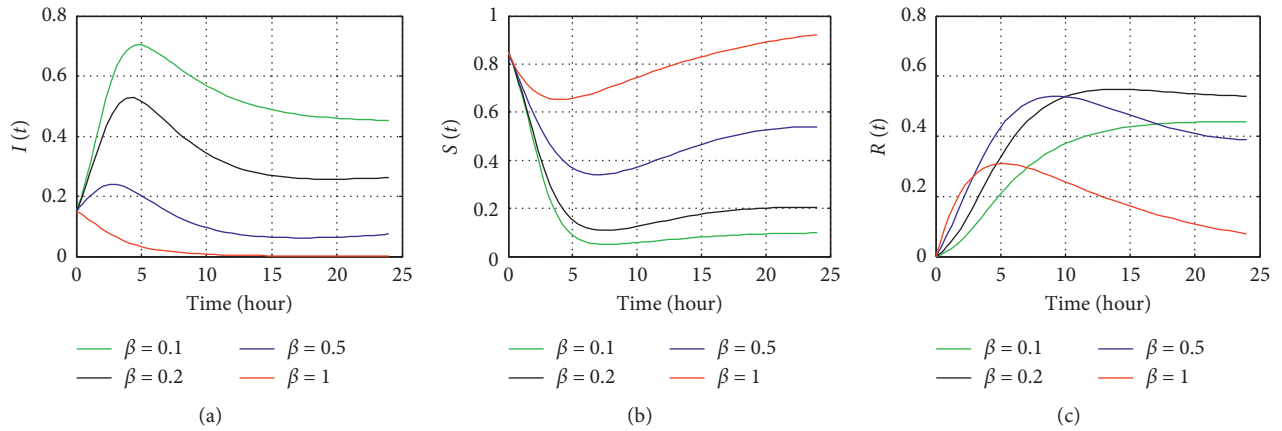



Figure 7: $I(t), S(t)$, and $R(t)$ under different $\beta$.

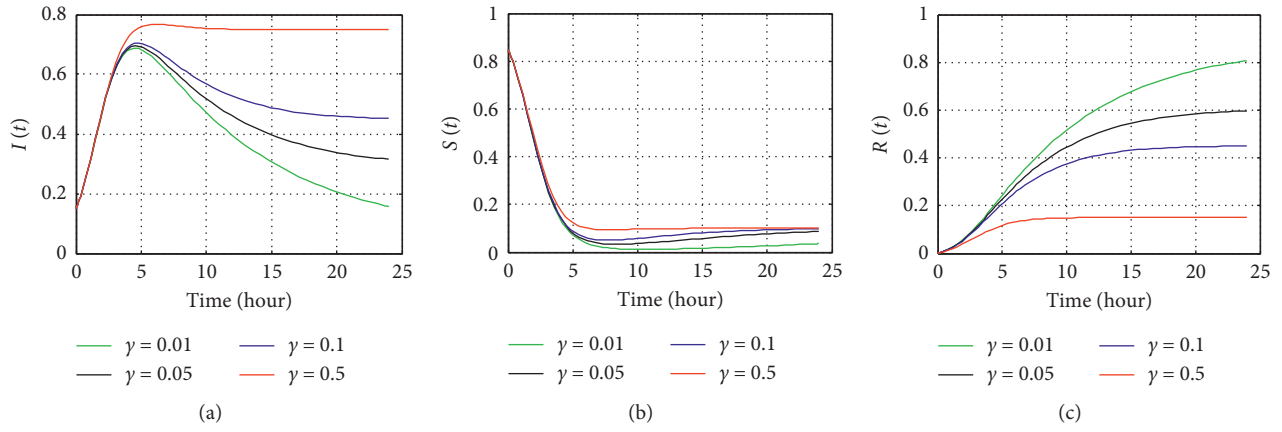



Figure 8: $I(t), S(t)$, and $R(t)$ under different $\gamma$.

different time periods. Partial results of the statistics are shown in Table 7.

As seen in Table 7, the maximum airport degrees of the nonpeak period and the peak period are 92 and 275, respectively. Figure 12 shows the relationship between $k$ and $p(k)$, with time periods 00 : 00–07 : 00 and 12 : 00–19 : 00, respectively. A scaling law between $k$ and $p(k)$ can be observed clearly.

Figures 12(a) and 12(b) illustrate two segments that follow the power laws:

$$(1)\ 00:00\text{–}07:00\ p(k) \sim \begin{cases} k^{-1.1966}, & k \le 53, \\ k^{-0.4054}, & k > 53. \end{cases}$$

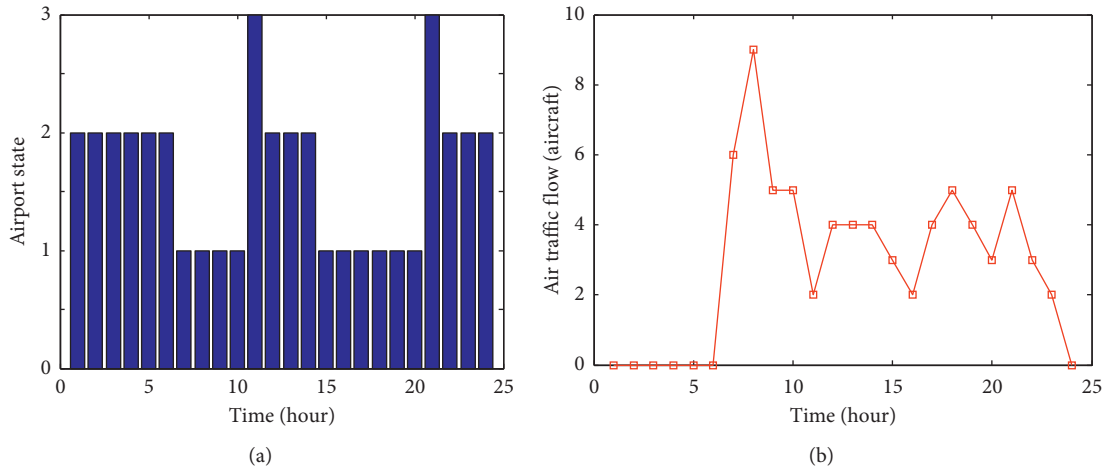



Figure 9: (a) Time varying airport states of airport ABNY. (b) Time varying air traffic flow volume of airport ABNY. In (a), "1" represents infected state, "2" represents susceptible state, and "3" represents recovered state.

Table 5: The top 10 airports.

| Airport | Takeoffs and landings | Ranking |
|---|---|---|
| ZBAA | 590199 | 1 |
| ZSPD | 449171 | 2 |
| ZGGG | 409679 | 3 |
| ZGSZ | 305461 | 4 |
| ZPPP | 300406 | 5 |
| ZUUU | 293643 | 6 |
| ZLXY | 267102 | 7 |
| ZSSS | 256603 | 8 |
| ZUCK | 255414 | 9 |
| ZSHC | 232079 | 10 |

Table 6: Samples of $I(t), S(t)$, and $R(t)$ statistics of October 1, 2015.

| | 0:00–1:00 | 1:00–2:00 | 2:00–3:00 | ... | 4:00–5:00 | 5:00–6:00 |
|---|---|---|---|---|---|---|
| $I(t)$ | 0.105528 | 0.080402 | 0.085427 | ... | 0.381910 | 0.201005 |
| $S(t)$ | 0.894472 | 0.879397 | 0.884422 | ... | 0.497487 | 0.582915 |
| $R(t)$ | 0 | 0.040201 | 0.030151 | ... | 0.120603 | 0.216080 |

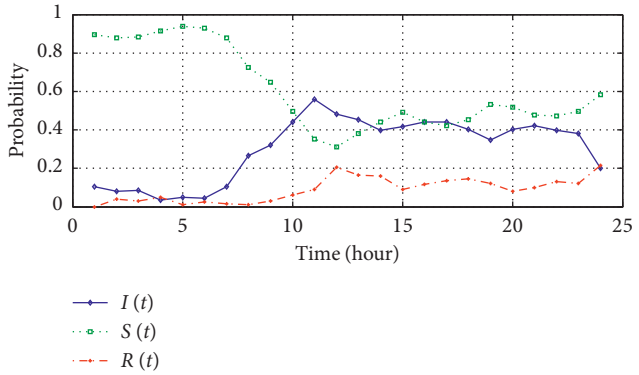



Figure 10: Actual value of $I(t), S(t)$, and $R(t)$ of Chinese airport network.

$$(2)\ 12:00\text{–}19:00\ P(k) \sim \begin{cases} k^{-2.0460}, & k \le 83, \\ k^{-0.4497}, & k > 83. \end{cases}$$

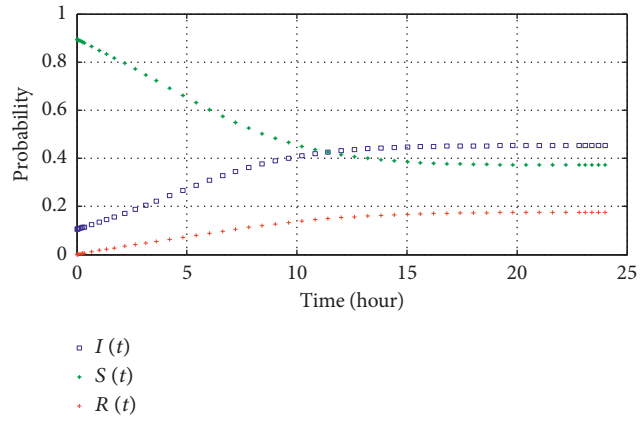



Figure 11: Simulated value of $I(t), S(t)$, and $R(t)$ by the ASIRS model.

Table 7: Airport degree distribution.

| 00:00–07:00 | | | 12:00–19:00 | | |
|---|---|---|---|---|---|
| $k$ | $N$ | $p(k)$ | $k$ | $N$ | $p(k)$ |
| 1 | 33 | 0.340206 | 1 | 19 | 0.104972 |
| 2 | 16 | 0.164948 | 2 | 23 | 0.127072 |
| 3 | 8 | 0.082474 | 3 | 21 | 0.116022 |
| 4 | 7 | 0.072165 | 4 | 12 | 0.066298 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 53 | 1 | 0.010309 | 185 | 1 | 0.005525 |
| 55 | 1 | 0.010309 | 201 | 1 | 0.005525 |
| 60 | 1 | 0.010309 | 220 | 1 | 0.005525 |
| 92 | 1 | 0.010309 | 275 | 1 | 0.005525 |

The degree distributions are strikingly different from those of random graphs, small-world networks, and scale-free networks. First, the degree distributions of the Chinese airport network display two segments and follow the Double Pareto Law. The degree of the critical airports is approximately 53 for 00:00–07:00 and 83 for 12:00–19:00. The smaller the exponent is, the stronger the heterogeneous characteristics of the network are [44]. The exponents of 12:

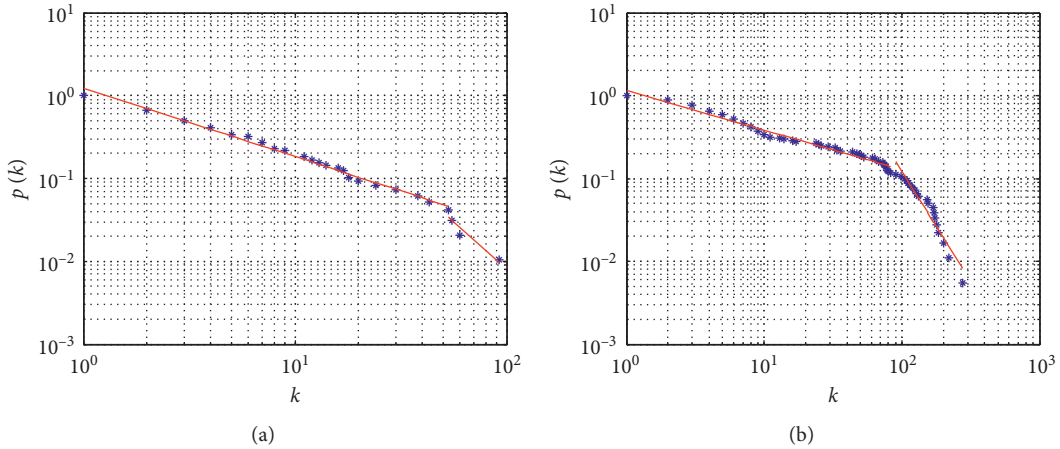



FIGURE 12: Scaling relations between $k$ and $p(k)$ at different time periods. (a) 00:00–07:00 and (b) 12:00–19:00.

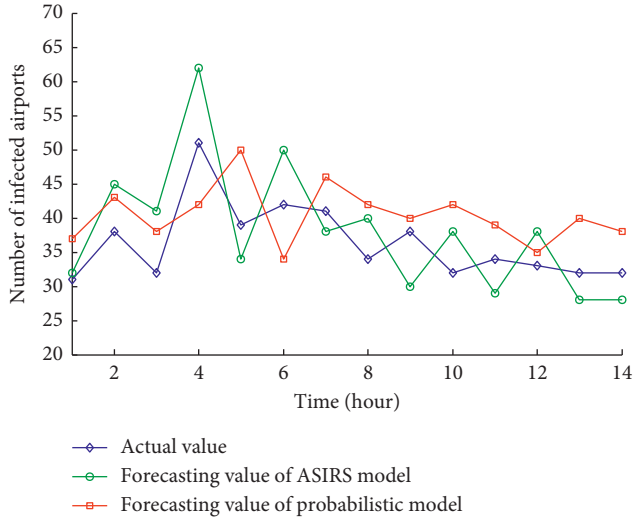



FIGURE 13: Comparison of actual and forecasting values.

00–19:00 are smaller than those of 00:00–07:00, which indicates that the heterogeneous characteristics of the Chinese airport network in 12:00–19:00 are stronger than those in 00:00–07:00. Thus, airport delays during 12:00–19:00 spread over the airport network more easily, quickly, and widely.

Obviously, the results from the ASIRS model and the degree distribution are consistent, which further indicates that the ASIRS model is well suited for characterizing airport delay propagation.

*5.4. Accuracy of the Short-Term Prediction of Delay Propagation.* As mentioned above, the ASIRS model can describe the characteristic of delay propagation in the long term, but the forecast accuracy is not very good (seen in Figures 10 and 11). How about the forecast accuracy of the ASIRS model in the short term?

To gauge the forecast accuracy, we introduce the probabilistic prediction method, which is common to delay forecast [45]. Intercepting the delay period (the period is usually no more than 4 hours) on a typical day, we forecast the number of delayed airports by constructing ASIRS models for every 4 hours. The prediction results are shown in Figure 13.

We use the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) to compare the accuracy of the two methods.

$$\text{MAPE} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{A_t - F_t}{A_t}\right| \times 100\%,$$
$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(A_t - F_t)^2}, \tag{16}$$

where $A_t$ is the actual value in the time interval $t$, $F_t$ is the forecast value in the time interval $t$, and $T$ is the number of time intervals.

Table 8 shows a comparison of the two methods. It can be seen that the MAPE and RMSE of the SIRS model are smaller

Table 8: MAPE and RMSE of the two models compared with historical data.

| Models | MAPE (%) | RMSE |
|---|---|---|
| ASIRS model | 15.9 | 6.3 |
| Probabilistic model | 18.1 | 7.1 |

than those of the probabilistic model. Thus, the ASIRS model is more accurate than the model based on probability in forecasting the number of delayed airports in the short term.

## 6. Conclusion

Understanding the process and evolution of airport delay propagation is very important for both air traffic management and aviation planning. In this study, we investigated the mechanism of delay propagation among airports from a new perspective:

(1) The delay fluctuations of airport networks are studied. To quantify the delay dynamics, we collected the airport delay at 199 Chinese airports and identified the existence of a certain scaling law, which indicates that the dynamic of airport delay is dominated by a propagation factor.

(2) The ASIRS model for airport delay propagation is presented to reveal the macroscopic appearance of delay propagation. The modeling approach is data driven in the sense that it is based on real China performance data.

(3) The long-term characteristics of delay propagation is described through building the ASIRS model. The accuracy of the short-term prediction of delay propagation is also examined.

It is remarkable that the airport delay is the result of the coupling of different factors, and there is no information on delay factors in the datasets. We cannot determine a delay is caused by which factors. Thus, we study the delay propagation from the overall delay data and simulate the overall delay without considering the specific factors.

Our ongoing work involves further calibration and validation of the ASIRS model. It is interesting to compare the epidemic model of airport delay propagation in different countries and investigate the practices of the countries. We will come up with insights for mitigating airport delay from such international comparisons.

## Data Availability

The data used to support the findings of this study have not been made available because the data also form part of an ongoing study.

## Conflicts of Interest

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

## Acknowledgments

This study was supported by the National Natural Science Foundation of China (nos. 71801215, 71671014, and U1833103), Natural Science Foundation of Tianjin (nos. 18JCQNJC04300), and Fundamental Research Funds for the Central Universities (nos. 3122016C009, 3122019129, and 3122018D026).

## References

[1] F. Wieland, "Limits to growth: results from the detailed policy assessment tool (air traffic congestion)," in *Proceedings of the 16th Digital Avionics Systems Conference (DASC16)*, Irvine, CA, USA, 1997.

[2] Faa.gov (Internet), Ishington DC: FAA; (up-dated 2018 January 3), http://aspmhelp.faa.gov/index.php/APM:_Definitions_of_Variables.

[3] N. Kafle and B. Zou, "Modeling flight delay propagation: a new analytical-econometric approach," *Transportation Research Part B: Methodological*, vol. 93, pp. 520–542, 2016.

[4] S. Ahmadbeygi, A. Cohn, Y. Guan, and P. Belobaba, "Analysis of the potential for delay propagation in passenger airline networks," *Journal of Air Transport Management*, vol. 14, no. 5, pp. 221–236, 2008.

[5] C. Dong, C. Shao, J. Li, and Z. Xiong, "An improved deep learning model for traffic crash prediction," *Journal of Advanced Transportation*, vol. 2018, Article ID 3869106, 13 pages, 2018.

[6] P. T. R. Wang, L. A. Schaefer, and L. A. Wojcik, "Flight connections and their impacts on delay propagation," in *Proceedings of the 22nd Digital Avionics Systems Conference (DASC22)*, Indianapolis, Indiana, 2003.

[7] P. Fleurquin, J. J. Ramasco, and V. M. Eguíluz, "Characterization of delay propagation in the US air-transportation network," *Transportation Journal*, vol. 53, no. 3, pp. 330–344, 2014.

[8] B. Yu, Z. Guo, S. Asian, H. Wang, and G. Chen, "Flight delay prediction for commercial air transport: a deep learning approach," *Transportation Research Part E: Logistics and Transportation Review*, vol. 125, pp. 203–221, 2019.

[9] A. Kondo, "Delay propagation and multiplier," in *Proceedings of the 51st Annual Transportation Research Forum*, pp. 11–13, Arlington, Virginia, March 2010.

[10] Y. J. Wang, Y. K. Cao, C. P. Zhu et al., "Characterizing departure delays of flights in passenger aviation network of United States," 2017, https://arxiv.org/abs/1701.05556.

[11] W.-B. Du, M.-Y. Zhang, Y. Zhang, X.-B. Cao, and J. Zhang, "Delay causality network in air transport systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 118, pp. 466–476, 2018.

[12] J. S. Li and J. L. Ding, "Analysis of flight delay propagation using Bayesian networks," *Acta Aeronaut Astronaut Sin*, vol. 29, no. 6, pp. 1598–1604, 2008, in Chinese.

[13] Y. J. Liu, W. D. Cao, and S. Ma, "Estimation of arrival flight delay and delay propagation in a busy hub-airport," in *Proceedings of the 4th IEEE Computer Society (ICNC4)*, Jinan, China, October 2008.

[14] W. Wu, C.-L. Wu, T. Feng, H. Zhang, and S. Qiu, "Comparative analysis on propagation effects of flight delays: a case study of China airlines," *Journal of Advanced Transportation*, vol. 2018, Article ID 5236798, 10 pages, 2018.

[15] P. Fleurquin, Ramasco, J. José, and V. M. Eguiluz, "Systemic delay propagation in the US airport network," *Scientific Reports*, vol. 3, no. 1, p. 1159, 2013.
[16] S. P. Qiu, "Modeling and analysis of the correlation betweenflight delays and flight delay propagation based on copula," Dissertation, Nanjing University of Aeronautics and Astronautics, Nanjing,China, 2015.
[17] J.-T. Wong and S.-C. Tsai, "A survival model for flight delay propagation," *Journal of Air Transport Management*, vol. 23, no. 7, pp. 5–11, 2012.
[18] X. Zheng, C.-M. Liu, and P. Wei, "Air transportation direct share analysis and forecast," *Journal of Advanced Transportation*, vol. 2020, Article ID 8924095, 12 pages, 2020.
[19] C. J. Dong, A. J. Khattak, C. F. Shao, and K. Xie, "Exploring the factors contribute to the injury severities of vulnerable roadway user involved crashes," *International Journal of Injury Control and Safety Promotion*, vol. 26, no. 3, pp. 302–314, 2019.
[20] W.-B. Du, Y. Gao, C. Liu, Z. Zheng, and Z. Wang, "Adequate is better: particle swarm optimization with limited-information," *Applied Mathematics and Computation*, vol. 268, pp. 832–838, 2015.
[21] Y. Cao, C. Zhu, Y. Wang, and Q. Li, "A method of reducing flight delay by exploring internal mechanism of flight delays," *Journal of Advanced Transportation*, vol. 2019, Article ID 7069380, 8 pages, 2019.
[22] C. Dong, C. Shao, H. Huang, X. Chen, and N. N. Sze, "Advances in traffic safety methodologies and technologies," *Journal of Advanced Transportation*, vol. 2018, Article ID 4129582, 2 pages, 2018.
[23] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences*, vol. 101, no. 11, pp. 3747–3752, 2004.
[24] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, "The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences*, vol. 102, no. 22, pp. 7794–7799, 2005.
[25] L. P. Chi, R. Wang, H. Su et al., "Structural properties of US flight network," *Chinese Physics Letters*, vol. 20, no. 8, p. 1393, 2003.
[26] W. Li and X. Cai, "Statistical analysis of airport network of China," *Physical Review E*, vol. 69, no. 4, Article ID 046106, 2004.
[27] G. Bagler, "Analysis of the airport network of India as a complex weighted network," *Physica A: Statistical Mechanics and Its Applications*, vol. 387, no. 12, pp. 2972–2980, 2008.
[28] S. W. Cheng, Y. P. Zhang, S. Q. Hao, R. W. Liu, X. Luo, and Q. Luo, "Study of flight departure delay and causal factor using spatial analysis," *Journal of Advanced Transportation*, vol. 2019, Article ID 3525912, 11 pages, 2019.
[29] M. Zanin, S. Belkoura, and Y. Zhu, "Network analysis of Chinese air transport delay propagation," *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 491–499, 2017.
[30] W. B. Du, "Research on complex network and network dynamicsfor air transportation systems," Dissertation, Anhui: University of Science and Technology of China, Huainan, Chi, 2010.
[31] S. M. Li, X. H. Xu, and L. H. Meng, "Fluctuations in airport arrival and departure traffic: a network analysis," *Chinese Physics B*, vol. 21, no. 8, Article ID 088901, 2012.
[32] H. Wang, Z. Song, R. Wen, and Y. Zhao, "Study on evolution characteristics of air traffic situation complexity based on complex network theory," *Aerospace Science and Technology*, vol. 58, pp. 518–528, 2016.
[33] A. Gautreau, A. Barrat, and M. Barthelemy, "Microdynamics in stationary complex networks," *Proceedings of the National Academy of Sciences*, vol. 106, no. 22, pp. 8847–8852, 2009.
[34] X. P. Wu, H. Y. Yang, and S. C. Han, "Analysis of properties and delay propagation of air traffic based on complex network," *Acta Aeronaut Astronaut Sinica*, vol. 38, no. S1, pp. 1–6, 2017.
[35] L. Zheng and L. Tang, "A node-based SIRS epidemic model with infective media on complex networks," *Complexity*, vol. 2019, Article ID 2849196, 14 pages, 2019.
[36] T. Q. Chen, B. C. Ma, and J. N. Wang, "SIRS contagion model of food safety risk," *Journal of Food Safety*, vol. 38, no. 1, Article ID e12410, 2018.
[37] Y. Muroya, Y. Enatsu, and H. Li, "Global stability of a delayed SIRS computer virus propagation model," *International Journal of Computer Mathematics*, vol. 91, no. 3, pp. 347–367, 2014.
[38] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
[39] A.-L. Barabasi, M. A. De Menezes, S. Balensiefer, and J. Brockman, "Hot spots and universality in network dynamics," *The European Physical Journal B—Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 169–175, 2004.
[40] M. Argollo de Menezes and A. L. Barabasi, "Separating internal and external dynamics of complex systems," *Physical Review Letters*, vol. 93, Article ID 068701, 2004.
[41] Z. Eisler, J. Kertész, S.-H. Yook, and A.-L. Barabási, "Multiscaling and non-universality in fluctuations of driven complex systems," *Europhysics Letters (EPL)*, vol. 69, no. 4, pp. 664–670, 2005.
[42] R. N. Mantegna and H. E. Stanley, "Scaling behaviour in the dynamics of an economic index," *Nature*, vol. 376, no. 6535, pp. 46–49, 1995.
[43] Z. Eisler and J. Kertesz, "Random walks on complex networks with inhomogeneous impact," *Physical Review E*, vol. 71, Article ID 057104, 2005.
[44] J.-F. Zheng, Z.-Y. Gao, and X.-M. Zhao, "Properties of transportation dynamics on scale-free networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 373, pp. 837–844, 2007.
[45] E. R. Mueller and G. B. Chatterji, "Analysis of aircraft arrival and departure delay characteristics," in *Proceedings of the AIAA Aircraft Technology, Integration and Operations (ATIO) Conference*, p. 5866, Los Angeles, CA, USA, October 2002.