

## Research Article

# Vehicle Speed and Position Estimation considering Microscopic Heterogeneous Car-Following Characteristics in Connected Vehicle Environments

HanQing Xia <sup>1,2</sup> XiaoMing Liu <sup>1,2</sup> ZeChao Ma <sup>1,2</sup> Fang Zhu <sup>1</sup> Lin Zhang <sup>1</sup>  
YingJie Zhao <sup>1</sup> and YuanRong Wang <sup>3</sup>

<sup>1</sup>School of Electrical and Control Engineering, North China University of Technology, Beijing 100144, China

<sup>2</sup>Beijing Key Laboratory of Urban Traffic Control Technology, North China University of Technology, Beijing 100144, China

<sup>3</sup>College of Applied Science and Technology of Beijing Union University, Beijing Union University, Beijing 100012, China

Correspondence should be addressed to HanQing Xia; [hqx\\_its@163.com](mailto:hqx_its@163.com)

Received 23 May 2023; Revised 7 November 2023; Accepted 27 November 2023; Published 8 December 2023

Academic Editor: Indrajit Ghosh

Copyright © 2023 HanQing Xia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a method for estimating the speed and position of unsampled vehicles using sampled data from connected automated vehicles (CAVs). The determination of vehicle speed and position on the road is a challenging and crucial task, as they can effectively reflect traffic flow characteristics and contribute to traffic state estimation and intersection signal timing optimization. Connected automated vehicles have the capability to upload their own trajectory data while also capturing trajectory data of surrounding vehicles through onboard sensors. Therefore, this paper proposes a novel approach to estimate the speed and position of unsampled vehicles. Firstly, using real vehicle trajectory data, the correlation between the velocity of following vehicles and the velocity of leading vehicles under different densities is analyzed, leading to the development of a velocity estimation model incorporating a speed correction factor. Secondly, the correlation between time headway, the rate of change of following vehicle acceleration, and traffic density is examined. To address the issue of heterogeneous behavior in vehicle following described by the Intelligent Driver Model (IDM), a real-time optimization model for estimating vehicle position by optimizing IDM parameters is proposed. The velocity estimation model and the position estimation model are summarized as two nonlinear optimization problems. Finally, the proposed method is validated using actual vehicle trajectory data. Experimental results demonstrate that when the number of connected automated vehicles (CAVs) is 2, the proposed method reduces the average absolute error by 30.73% and the standard deviation of the average absolute error by 42.8% compared to a linear model-based speed estimation method under different density conditions. Compared to a method that estimates vehicle position by calibrating desired gaps, the proposed method reduces the average absolute error by 38.2% and the standard deviation of the average absolute error by 41.7% under different density conditions. Furthermore, the proposed method exhibits good practicality under different CAV penetration rates.

## 1. Introduction

Vehicle trajectory information accurately reflects the spatiotemporal evolution characteristics of traffic flow and has been widely used by many scholars to estimate queue length [1, 2], travel time [3], traffic delay [4], and traffic volume [5]. It also helps optimize signal timing at intersections and evaluate control strategies [6], providing a solid foundation for intelligent management and control of traffic systems.

Currently, there are two main ways to obtain vehicle trajectory data: fixed position sensors and mobile sensors. Fixed sensors include video cameras, loop detectors, and microwave radar detectors. They can only detect limited data of fixed sections at fixed times, including traffic flow and vehicle types. However, the detection range of these sensors is small and the missed detection rate is high, making it impossible to record detailed vehicle spatiotemporal information [7]. Moreover, due to their high installation and

maintenance costs and the high coverage of urban road networks, these fixed detectors are difficult to cover the entire urban road network [8].

Therefore, the development of connected and automated vehicle (CAV) technologies have created the possibility of solving the above problems. CAVs can provide safer, more environmentally friendly, more energy-efficient, and more convenient travel methods and comprehensive solutions. They are believed to have the potential to improve traffic efficiency, reduce energy consumption, and reduce the occurrence of traffic accidents. CAVs are vehicles that achieve assisted driving or even autonomous driving without human intervention, using advanced sensing technology [9]. Equipped with advanced sensors and various wireless technologies, it can not only share its own trajectory data with other CAVs and upload it to the Roadside Unit (RSU) but also acquire trajectory data from human-driven vehicles (HVs) without sensors within its surrounding detection range. Therefore, CAVs can be considered a new type of mobile sensor and can be used to collect spatiotemporal information of other vehicles, thus providing comprehensive vehicle trajectory data [10]. However, CAVs will not appear on daily roads in a highly penetrated form in the next few decades. Even if CAVs can provide complete trajectory data of vehicles within their sampling range, the sparse spatiotemporal data are caused by the low penetration rate.

Wireless transmission of data for traffic management and control has been a hot research topic in recent years. Most of the work has focused on using wireless transmission of data for traffic state estimation and single-vehicle position estimation, rather than position estimation of unsampled vehicles in a queue. Bar-Gera et al. [10] estimated vehicle speed and travel time based on triangulation from signal towers. The literature [11–13] proposed linear state estimation algorithms based on Kalman filtering to estimate the position of a single vehicle using GPS location data. The literature [14, 15] used particle filtering algorithms to estimate vehicle positions. Liu et al. [16] improved the accuracy of single-vehicle position estimation by using high-resolution data from 5G base stations. Although the above methods can estimate the position of a single vehicle well, weak signals in remote mountainous areas, tunnels, and overlapping overpasses can easily cause GPS and 5G signal loss, making it impossible to estimate the position of a single vehicle using wireless transmission of data. Qin et al. [17] used passive RFID tags deployed on the road to perform real-time position estimation of a single vehicle, but this method relies too heavily on special road infrastructure and communication equipment. Therefore, determining the state of unsampled vehicles based on partially sparse sampled data is the key to solving the above problem.

In recent years, with the development of CAV technologies, it has become a trend to estimate the state of unsampled vehicles based on partially sampled data. Numerous studies, such as literature [18], estimate the travel time of this type of vehicle based on partial CAVs passing through the intersection, assessing traffic delays. Feng et al. [19] reconstructed vehicle trajectories in the road network using Particle Filter theory and five correction factors,

including path consistency, travel time consistency, flow model, etc., based on data from Automatic Vehicle Identification (AVI) and traditional detectors. Chen et al. [20] proposed a method for reconstructing highway vehicle trajectories based on CAV sampling data using the Intelligent Driver Model (IDM) and introducing expected headway calibration factors. Some scholars also use a data-driven approach to complete unsampled data by analyzing the correlation among sampled data. Lint et al. [21] estimated vehicle speed using a State-Related Filter. Ji [22] used Long Short-Term Memory (LSTM) to train on NGSIM data and estimated the speed and position of vehicles closely following the front vehicle on highways. Nanthawichit et al. [23] estimated the travel time of vehicles through the fusion of data from mobile sensors and fixed-point detectors. However, when the sampled data are sparse, data-driven methods (such as LSTM) cannot obtain sufficient data for training, which can lead to anomalous driving behavior in estimating the state of unsampled vehicles, such as negative estimated speed or overlapping vehicle positions.

In order to address the issue of sparse sampling data affecting the estimation accuracy of data-driven methods, using a car-following model to estimate vehicle positions can fully consider the interdependence between the leading and following vehicles [24]. Goodall et al. [25] utilized GPS data from highway vehicles and developed a car-following position estimation algorithm based on the Wiedemann model with preset parameters. By comparing the actual acceleration of vehicles with the expected acceleration calculated based on the Wiedemann model, it was determined that there were undetected manually driven vehicles between two CAVs when the difference in acceleration exceeded a certain threshold [26]. Yao et al. [27] employed the IDM to optimize the insertion position of human-driven vehicles, with the primary objective of minimizing the mean squared error between the actual acceleration of the leading vehicle and the estimated acceleration. This method, based on car-following models, not only addresses the limitations of data-driven approaches but also significantly enhances the accuracy of estimating the position of artificially driven vehicles.

In general, current research on vehicle position and velocity estimation primarily falls into two categories: data-driven methods and model-based methods. Data-driven approaches face a major challenge stemming from the sparsity of sampled data, resulting in insufficient training data. On the other hand, while model-based methods for vehicle position estimation using car-following models address issues related to vehicles deviating significantly from these models, they also encounter accuracy challenges when dealing with sparsely sampled data. One significant reason for this is that existing research often utilizes predefined parameter models (e.g., Wiedemann and IDM) for vehicle estimation, disregarding the heterogeneity within traffic flow. For instance, car-following behaviors exhibit variability under different traffic densities, even within the same vehicle platoon, as local density fluctuations lead to variations in car-following behavior.

Considering the limitations of prior research, this paper presents a model-based method for estimating the speed and position of human-driven vehicles. Unlike conventional estimation methods that rely on fixed-parameter models, this study formulates the selection of parameters for vehicle speed and position estimation models as two optimization problems. The objective is to utilize real-time traffic data from connected and autonomous vehicles (CAVs) to adjust model parameters to better align with the current traffic conditions, thus reducing estimation errors. The results demonstrate that this proposed method exhibits strong estimation performance, even in environments with low CAV penetration rates. The major contributions are as follows:

- (1) In response to the issue of low estimation accuracy when using fixed-parameter models for vehicle speed and position estimation, this study, based on real-world traffic data, analyzes the model parameters that influence the precision of estimating the speed and position of human-driven vehicles.
- (2) Considering the influence of unstable car-following behavior, based on the analysis of model parameters that impact estimation performance, a method is proposed to optimize the speed estimation model and corresponding parameters of the IDM using CAV detection data, thereby enhancing the precision of speed and position estimation for human-driven vehicles.
- (3) Under various CAV penetration rates and different traffic densities, a comprehensive evaluation of the proposed method was conducted using real-world traffic scenario data. The results demonstrate that the proposed method consistently outperforms other baseline models, leading to significant improvements in both speed and position estimation accuracy as well as stability.

The rest of this paper is organized as follows. Section 2 introduces the research problem. Section 3 describes the speed and position estimation modeling method. Section 4 validates the proposed method using actual data. Section 5 concludes the paper and proposes future research directions.

## 2. Representative Scenario

First, we provide a representative scenario to explain our work, as shown in Figure 1. For modeling simplicity, we only focus on the car-following process without lane changing. The mixed traffic flow studied in this paper consists of two types of vehicles: human-driven vehicles (HVs) and connected and automated vehicles (CAVs). We assume that each CAV is equipped with a Mobile Object Detection and Tracking System (MODAT) to detect the surrounding traffic conditions [24]. CAVs are equipped with various sensors, including GPS, stereo vision cameras, and LiDAR. LiDAR is used to acquire point cloud data in the surrounding space, and the data from stereo vision cameras and GPS are fused to obtain real-time positions of surrounding vehicles. The acquired data include vehicle ID, timestamp, and location.

Referencing to [10, 20], the detection range for CAVs is set at 100 m. CAVs use their communication capabilities to upload this information to Roadside Units (RSUs). RSUs are equipped with data processing systems and computing units. The data processing system can generate high-resolution trajectory maps (speed, position, and acceleration) for CAVs and vehicles within the detection range. The computing unit estimates the speed and position information of undetected human-driven vehicles based on velocity and position estimation models and sends this information to all CAVs to assist them in evaluating the surrounding traffic conditions.

## 3. Methodology

Figure 2 illustrates a typical following scenario. In the following model, it is assumed that the driving behavior of the following vehicle is influenced by the changes in the motion of the leading vehicle, which is reflected in the speed, position, and acceleration of the following vehicle. Therefore, the following model can be used to define the behavior of the following vehicle under the influence of the leading vehicle. When the following vehicle significantly deviates from the expected behavior defined by the following model, it indicates that the following vehicle is influenced by other vehicles [26], and therefore, the position of these vehicles needs to be estimated.

*3.1. Acceleration Estimation.* In 2002, Helbing et al. [28] proposed the Intelligent Driver Model (IDM) based on empirical observations, which differs from most existing car-following models that separate free-flow and congested states. The IDM provides a unified framework to describe various states of vehicles, ranging from free flow to complete congestion, with a concise set of interpretable parameters. Furthermore, the IDM has been widely utilized to capture the effects of driving behavior between CAVs and HVs [7, 20, 24]. Therefore, for the sake of modeling convenience, this study assumes that the interactions between HVs and CAVs, as well as between HVs, conform to the IDM. The IDM takes into account both the position and velocity differences between the leading and following vehicles, and a typical IDM is shown in the following equation:

$$a_n(t) = a_{\max} \left[ 1 - \left( \frac{v_n(t)}{v_{\max}} \right)^\delta - \left( \frac{S^*(t)}{\Delta s(t)} \right)^2 \right]. \quad (1)$$

where  $n$  represents vehicle  $n$ , which is the following vehicle in a group of adjacent cars;  $a_{\max}$  is the maximum acceleration;  $v_n$  represents the velocity of the following vehicle;  $v_{\max}$  represents the desired velocity of the vehicle under free-flow conditions;  $t$  represents the current time;  $\delta$  is the acceleration exponent; and  $\Delta s$  represents the actual distance between two vehicles, defined as  $\Delta s = p_{n-1} - p_n - L$ , in which  $p_{n-1}$  represents the position of the leading vehicle  $n - 1$ ,  $p_n$  represents the position of the following vehicle  $n$ , and  $L$  represents the length of the vehicle.  $S^*$  represents the desired distance between two vehicles, which is expressed as a function of the velocities of both vehicles in the following equation:

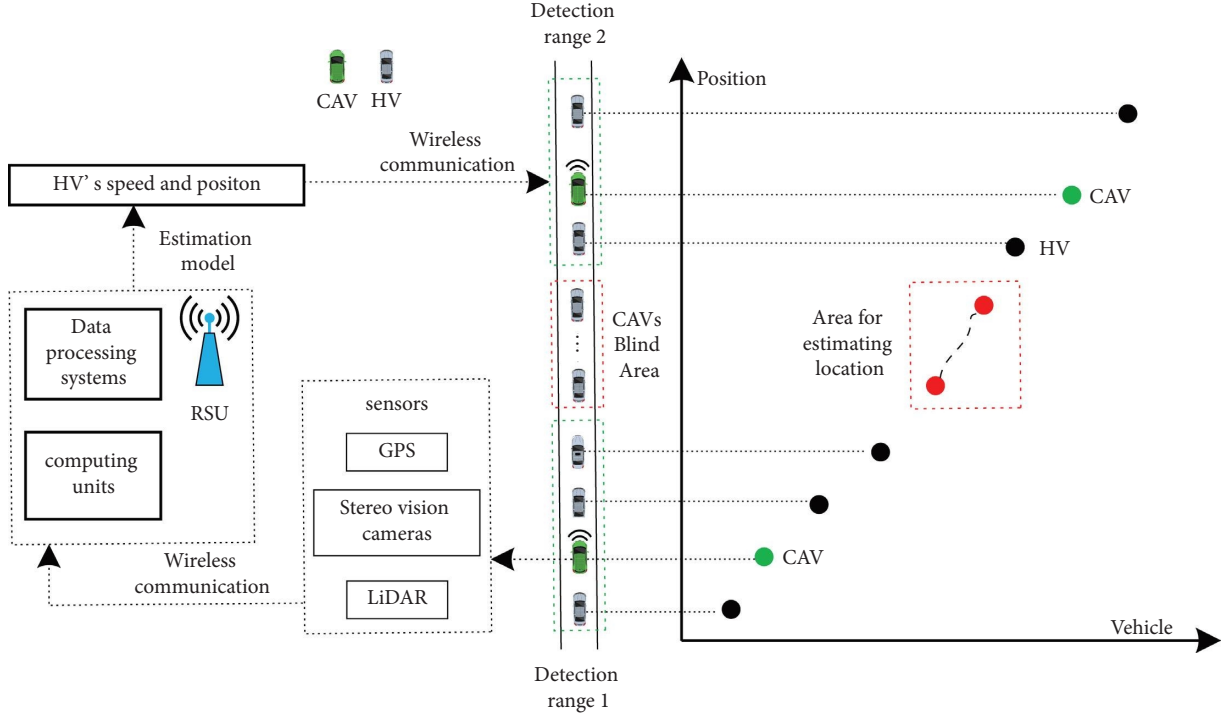


FIGURE 1: Mixed traffic flow with CAVs and HVs.

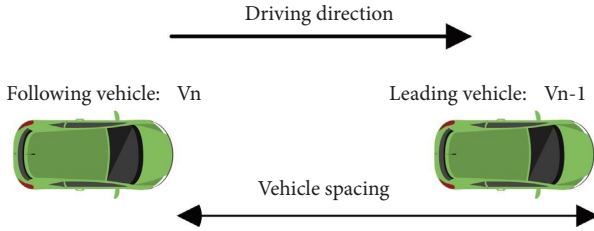


FIGURE 2: Typical car-following scenario.

$$S^*(t) = s_a + \max \left\{ v_n(t)T + \frac{v_n(t)[v_n(t) - v_{n-1}(t)]}{2\sqrt{a_{\max} \cdot b}}, 0 \right\}, \quad (2)$$

where  $S_a$  represents the minimum safe distance between two cars;  $T$  represents the safe time headway between the following car and the leading car; and  $b$  represents the comfortable deceleration.

### 3.2. Checking for the Existence of Unsampld HVs.

Figure 3 shows a typical following scenario in a mixed traffic flow. The first HV within the detection range of CAV2 is designated as the following vehicle  $V_n$ , and the last vehicle within the detection range of CAV1 is designated as the leading vehicle  $V_{n-1}$ . Assuming that there are two vehicles on the road, CAV1 and CAV2, with  $V_n$  following  $V_{n-1}$  within the detection range and responding based on  $V_{n-1}$ 's behavior, at time  $t$ , if the difference between the theoretical acceleration of  $V_n$  calculated by IDM and the actual acceleration  $a_n(t)_{\text{actual}}$  exceeds a preset threshold value  $\sigma$ , it is assumed that there exist unsampled HVs in the detection blind zone and their traffic parameters need to be estimated. For the selection of threshold value  $\sigma$ , we refer to the study

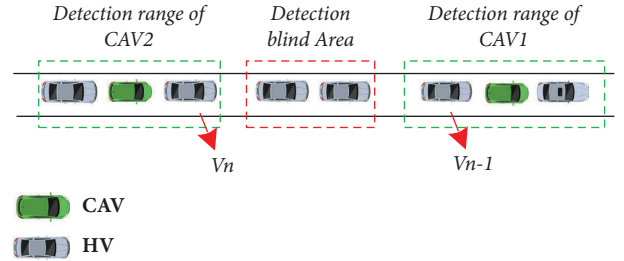


FIGURE 3: Mixed traffic flow following scenario 1.

by Goodall et al. [26], who calibrated the NGSIM data in the United States and found that when  $\sigma = 1.96$ , it can describe various actual situations. The theoretical acceleration of  $V_n$  is calculated by the following equation:

$$a_n(t)_{\text{estimate}} = a_{\max} \left[ 1 - \left( \frac{v_n(t)}{v_{\max}} \right)^\delta - \left( \frac{S^*(t)}{\Delta S(t)} \right)^2 \right]. \quad (3)$$

Therefore, the difference between theoretical and actual acceleration can be calculated by the following equation:

$$\sigma = |a_n(t)_{\text{estimate}} - a_n(t)_{\text{actual}}|. \quad (4)$$

After determining the existence of unsampled HVs between CAV1 and CAV2, it is necessary to estimate the positions of these unsampled vehicles. At any time, the HVs whose positions need to be estimated must satisfy the minimum safe distance between  $V_{n-1}$  and  $V_n$ . We denote the  $i$ -th vehicle inserted in front of  $V_n$  as  $V_{n,i}$ , and therefore, the position of  $V_{n,i}$ , denoted as  $p_{n,i}$ , must be kept within the safe distance range between  $V_{n-1}$  and  $V_n$ , as shown in the following equation:

$$p_n(t) + s_a + l < p_{n,i}(t) < p_{n-1}(t) + s_a + l, \quad (5)$$

where  $s_a$  represents the minimum safe distance between  $V_{n-1}$  and  $V_n$ .

**3.3. Estimation of the Lead Vehicle Speed.** As shown in Figure 4, we assume that  $V_n$  is following the first human-driven vehicle HV1 in the detection blind zone. The acceleration of the following vehicle  $V_n$  is affected by its current speed  $v_n(t)$  and the speed of a leading vehicle HV1.

Therefore, the following conditions hold:

$$v_{n-1}(t)' = \max\{v_n(t) + \lambda a_n(t), 0\}, \quad (6)$$

where  $v_{n-1}(t)'$  represents the estimated speed of the leading vehicle  $V_n$ . The coefficient  $\lambda$  was calibrated by Goodall et al. [27] in 2013 to be 0.162 based on NGSIM data. Using a preset parameter model to estimate the speed of the leading vehicle in free flow and stable following situations may work well. However, when the following vehicle is sufficiently close to the leading vehicle, the randomness of the following vehicle's acceleration increases, and a fixed  $\lambda$  cannot accurately capture the effect of the leading vehicle's speed changes on the following vehicle's acceleration. To verify this idea, we randomly extracted trajectory data of 1000 vehicle convoys for each density between 30 and 60 veh/km at intervals of 10 veh/km, considering the majority of density scenarios in actual road conditions using the NGSIM data collected on the US-101 highway. We analyzed the effect of different densities on the value of  $\lambda$ , and the calibration formula used is shown in the following equation:

$$\lambda = \begin{cases} \frac{v_{n-1}(t) - v_n(t)}{a_n(t)}, & a_n(t) \neq 0, \\ 0, & a_n(t) = 0. \end{cases} \quad (7)$$

In Figure 5, the change trend of the  $\lambda$  value under densities of 30–60 veh/km is shown, and with the increase of the number of selected vehicle platoons, the  $\lambda$  values under different densities finally converge to different values. Table 1 provides the statistical values of  $\lambda$  under different densities. It can be found that when the traffic density is between 30 and 50 veh/km, the absolute value of the average of coefficient  $\lambda$  increases with the growing density. This can be easily explained because as the traffic density increases, the headway distance between vehicles becomes smaller, and the impact of the preceding vehicle on the following vehicle becomes more severe. When the traffic flow enters the range of 50–60 veh/km,  $\lambda$  becomes smaller. At this point, the traffic flow gradually transitions from free flow to congestion, and most vehicles in the platoon will follow the preceding vehicle's speed closely. The influence of the preceding vehicle's velocity change on the following vehicle's velocity and acceleration gradually decreases. Therefore, it can be verified that the degree to which the acceleration of the following vehicle is affected by its own current speed and the preceding vehicle's velocity depends on the traffic density. Using the preset  $\lambda$  value method to estimate the velocity of the leading vehicle cannot well reflect the real traffic situation.

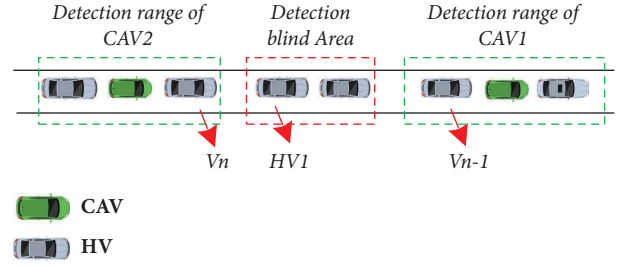


FIGURE 4: Mixed traffic flow following scenario 2.

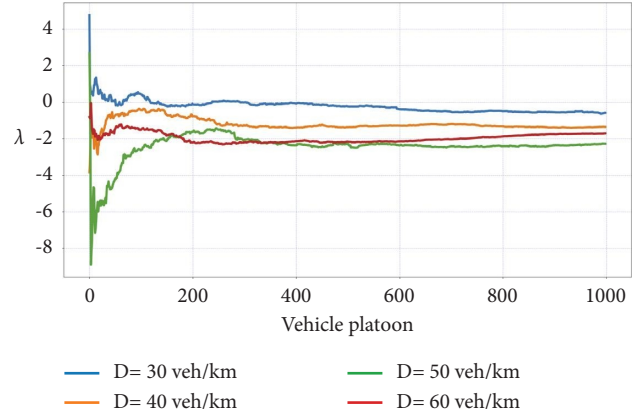


FIGURE 5: The values of the coefficient  $\lambda$  under different densities.

TABLE 1: Statistical values of parameter  $\lambda$  under different traffic densities.

Traffic density (veh/km)	Parameters	Mean
30	$\lambda$	-0.23316
40	$\lambda$	-1.20608
50	$\lambda$	-2.43141
60	$\lambda$	-1.96361

To address this issue, this study incorporates an optimization factor  $\Theta$  into the model as shown in equation (8), where  $\Theta$  is calculated through optimization based on the detection data within the CAV detection range. In addition, it is considered that the estimated speed of the lead vehicle should not exceed the free-flow speed, as shown in equation (9). Based on these considerations, a method is proposed to use real-time CAV sample data to optimize the value of  $\Theta$  (the optimization method is introduced later and is also applied to improve the IDM) in order to improve the accuracy of HV speed estimation.

$$v_{n-1}(\Theta(t))' = \max\{\Theta(t)[v_n(t) + \lambda a_n(t)], 0\}, \quad (8)$$

$$0 \leq v_{n-1}(\Theta(t))' \leq v_{\max}. \quad (9)$$

Once the speed of HV1 immediately leading the vehicle  $V_n$  has been determined, it is reasonable to assume that the speeds of other HVs ahead of HV1 and between HV1 and the vehicle  $V_{n-1}$  can be estimated as the average of the speeds of HV1 and  $V_{n-1}$ , considering that in stable traffic flow, the speed difference between the leading and following vehicles is small. The speed is calculated by the following equation:

$$v_{\text{insert}_i}(t) = \frac{\max\{\Theta(t)[v_n(t) + \lambda a_n(t)], 0\} + v_{n-1}(t)}{2}. \quad (10)$$

3.4. *Estimation of HV Position.* The essence of using a car-following model to estimate the position of the lead vehicle is to calculate the headway between the following vehicle and

the leading vehicle based on the following vehicle's speed, position, and acceleration. Solving and transforming equation (1) yields equation (11), and substituting equation (2) into equation (11) results in equation (12).

$$\Delta s(t) = \frac{S^*(t)}{\sqrt{1 - (v_n(t)/v_{\max})^\delta - (a_n(t)/a_{\max})}} \quad (11)$$

$$\Delta s(t) = \frac{s_a + \max\left\{v_n(t)T + \left(v_n(t)[v_n(t) - v_{n-1}(t)]/2\sqrt{a_{\max} \cdot b}\right), 0\right\}}{\sqrt{1 - (v_n(t)/v_{\max})^\delta - (a_n(t)/a_{\max})}}. \quad (12)$$

When the traffic flow reaches a stable state, it can be assumed that the following vehicle will follow the leading vehicle at a speed close enough to the front vehicle and the leading vehicle's acceleration is 0. Therefore,  $v_n(t) - v_{n-1}(t)$  approaches 0 infinitely, and equation (2) can be simplified to equation (13). Equation (10) can be transformed to equation (14), and the estimated position of the leading vehicle can be calculated by equation (15).

$$S^*(t) = S_a + v_n(t)T, \quad (13)$$

$$\Delta s(t) = \frac{s_a + v_n(t)T}{\sqrt{1 - (v_n(t)/v_{\max})^\delta}}, \quad (14)$$

$$p_{n-1}(t) = p_n(t) + \Delta s(t). \quad (15)$$

Based on equations (1)–(15), we can estimate the position and velocity of unsampled HVs in the detection blind zone at any time. When an HV is inserted into the undetected range, it can be treated as a known vehicle within the CAV detection range, and its information can be used to continue estimating the position of the leading vehicle. The entire estimation process is repeated until it exceeds the predetermined estimation interval. To verify the effectiveness of using the IDM with preset parameters to estimate the position of unsampled HVs, we evaluate the mean absolute error between the estimated position of the HV (denoted by  $p_{\text{insert}_i}(t)_{\text{estimate}}$ ) and its actual position (denoted by  $p_{\text{insert}_i}(t)_{\text{actual}}$ ) by the following equation:

$$P, \text{error}_{\text{MAE}} = \frac{1}{k} \sum_{i=0}^k |p_{n,i}(t)_{\text{estimate}} - p_{n,i}(t)_{\text{actual}}|, \quad (16)$$

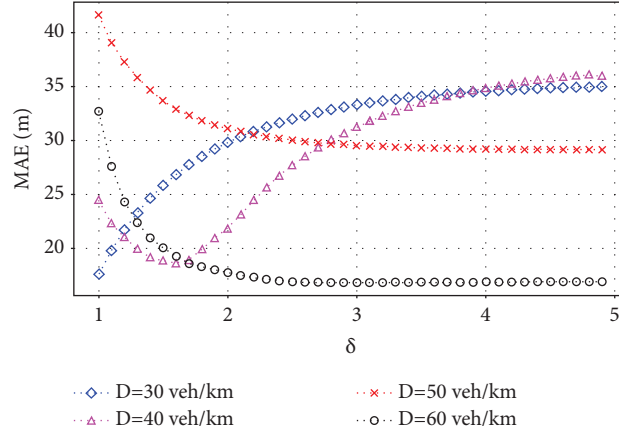
where  $k$  represents the number of HVs to be inserted between  $V_n$  and  $V_{n-1}$ .

However, when the traffic density is moderate, using equations (1)–(16) may not be a problem, as the mutual influence between vehicles remains stable and the following car can maintain a stable following rule with the leading car. When the traffic density is low, the mutual influence between

vehicles is weak, and the IDM with preset parameters may not accurately estimate the headway distance between two cars. On the other hand, as the traffic density increases, the following car may approach the leading car with a smaller safe headway distance. At this time, the randomness of the influence of the leading car on the following car increases. For example, if the leading car suddenly decelerates, its influence on the following car mainly manifests in the significant oscillation of its acceleration.

In the IDM,  $\delta$  is the acceleration exponent, which characterizes how the acceleration decreases. A larger  $\delta$  means a higher rate of change of the acceleration of the following vehicle, and in previous studies [28], it was usually set to 4. To verify the correlation between the acceleration exponent and traffic density, we extracted the trajectory data of four different traffic density scenarios with intervals of 10 veh/km between 30 and 60 veh/km from the NGSIM dataset and specified the first and last vehicles in the convoy as CAV1 and CAV2, respectively. Based on previous studies, we first set the safe time headway  $T$  in the IDM as a constant (e.g., 1.98 s) and selected different values of  $\delta$  to estimate the position of HV1 using equations (1)–(15). The changes in MAE under different  $\delta$  values are shown in Figure 6. It can be seen that the MAE of estimating vehicle positions increases with increasing  $\delta$  at a density of 30 veh/km, and at a density of 40 veh/km, the MAE reaches a minimum value at a value of 1.6. This indicates that when the traffic density is moderate, the mutual influence between vehicles is weak, and the rate of change of the acceleration of the following vehicle is less affected by the preceding vehicle and remains relatively stable. At traffic densities between 50 veh/km and 60 veh/km, the MAE of estimating vehicle positions decreases with increasing  $\delta$ , indicating that as traffic density increases, the degree of mutual influence between vehicles increases, and the rate of change of the acceleration of the following vehicle increases accordingly, especially when traffic density is between congested flow and free flow.

In the IDM, parameter  $T$  represents the safe time headway between the following vehicle and the leading vehicle. The IDM assumes that the vehicle maintains a fixed

FIGURE 6: Variation of MAE with respect to  $\delta$  at different densities.

safe headway when following the leading vehicle, which is obviously unrealistic. Similar to the analysis method of the  $\delta$  value in the IDM, based on the NGSIM data, we extracted 1000 different platoon trajectory data in the density range of 30–60 veh/km with an interval of 10 veh/km, aiming to analyze the evolution of the average headway of platoon under different densities. The headway calculation formula used is shown in equation (17), where  $T_{\text{HDW}}(t)$  represents the time headway between the following vehicle  $V_n$  and the leading vehicle  $V_{n-1}$ . Figure 7 shows the change of the average headway of platoon under different densities, and Table 2 provides statistical values for the average headway of platoons under different density conditions. It is evident that under density conditions ranging from 30 veh/km to 60 veh/km, the average platoon time headway decreases as traffic density increases. This conforms to the fundamental characteristics of traffic flow. Figure 7 illustrates the variations in time headway under different density conditions. It can be observed that, when density is held constant, the average time headway within different platoons exhibits fluctuations due to the heterogeneity in following behaviors. For instance, within the same vehicle platoon, variations in driving behaviors among different drivers result in these fluctuations. Cautious drivers tend to maintain a larger gap behind the leading vehicle during the following process, while more aggressive drivers tend to closely tail the leading vehicle. Thus, even within the same platoon, differences in driving behavior can lead to local changes in traffic density.

$$T_{\text{HDW}}(t) = \begin{cases} \frac{p_{n-1}(t) - p_n(t)}{v_n(t)}, & v_n(t) \neq 0, \\ 0, & v_n(t) = 0. \end{cases} \quad (17)$$

Therefore, we can reasonably assume that the acceleration index  $\delta$  and the safe time headway  $T$  are related to traffic density. Previous studies have assumed them as fixed values, which differs from the randomness of vehicle

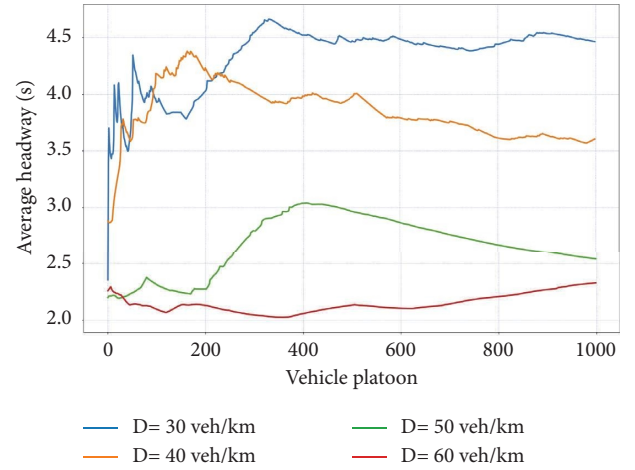


FIGURE 7: Changes in average time headway under different traffic densities.

TABLE 2: Statistics of time headway at different traffic densities.

Traffic density (veh/km)	Parameters	Max	Mean
30	$T_{\text{HDW}}$	4.663	4.343
40	$T_{\text{HDW}}$	4.379	3.846
50	$T_{\text{HDW}}$	3.040	2.669
60	$T_{\text{HDW}}$	2.328	2.147

behavior found in the above experiments. Therefore, before estimating the position of the preceding vehicle using the IDM, we introduce correction factors  $\beta$  and  $\kappa$  in the formula for  $\Delta s(t)$  to, respectively, correct the acceleration exponent  $\delta$  and headway  $T$  based on the sampling data obtained within the CAV detection range. The aim is to obtain IDM parameters that are more in line with the actual situation of the current vehicle fleet. The correction process, similar to the previous speed correction factor  $\Theta$ , will be discussed in detail in the following. Therefore, equation (12) is transformed into equations (14) and (18) is transformed into equation (19):

$$\Delta s(\kappa(t), \beta(t)) = \frac{s_a + \max\left\{v_n(t)\kappa(t)T + \left(v_n(t)[v_n(t) - v_{n-1}(t)]/2\sqrt{a_{\max} \cdot b}\right), 0\right\}}{\sqrt{1 - (v_n(t)/v_{\max})^{\beta(t)\delta} - (a_n(t)/a_{\max})}}, \quad (18)$$

$$p_{n-1}(t)' = p_n(t) + \Delta s(\kappa(t), \beta(t)). \quad (19)$$

The flowchart of the proposed vehicle position estimation method is shown in Figure 8. This method includes four parts: determining the difference between expected and actual behavior, optimizing model parameters, estimating traffic parameters, and estimating vehicle position. The specific details will be discussed in the following sections.

**3.5. The Calculation Procedure for the Correction Factors ( $\Theta$ ,  $\beta$ ,  $\kappa$ ).** In this section, we will provide a detailed explanation of how to calculate the three correction factors using the platoon data extracted from the NGSIM dataset. Assuming that the platoon scenario extracted is as shown in Figure 9, when any vehicle in the platoon is assumed to be a CAV, as depicted in Figure 10, there are a total of  $j$  vehicles within its detection range, including the CAV itself, all sequentially numbered as  $\text{veh}_1, \text{veh}_2, \dots, \text{veh}_{j-1}, \text{veh}_j$ ; with the aid of various CAV sensors, information about the  $j$  vehicles (including speed, position, and acceleration) can be acquired. Let  $Vc_i$  be defined as the  $i$ -th group of leading and following vehicles within these  $j$  vehicles (e.g.,  $Vc_1 = [\text{veh}_1, \text{veh}_2]$ ). As a result, the platoon comprises a total of  $j - 1$  vehicle combinations ( $Vc_1, Vc_2, \dots, Vc_{j-1}$ ). Taking  $Vc_1$  as an example, based on the information of  $\text{veh}_2$  (velocity, position, and acceleration), it is possible to estimate the information of  $\text{veh}_1$  and then compare it with the information of  $\text{veh}_1$ . The process can be repeated for vehicle combinations  $Vc_2$  to vehicle combinations  $Vc_{j-1}$ .

**3.5.1. Speed Correction Factor  $\Theta$ .** When optimizing the value of  $\Theta$ , it is necessary to impose constraints on the range of  $\Theta$ . In order to ensure that the estimated speed of  $v_i(\Theta(t))'$  does not exceed the free-flow speed and is not negative,  $v_i(\Theta(t))'$  is constrained as follows:

$$0 \leq v_i(\Theta(t))' \leq v_{\max}. \quad (20)$$

Substituting (20) into (21),  $\Theta(t)$  is constrained as follows:

$$0 \leq \Theta(t) \leq \frac{v_{\max}}{v_i(t) + \lambda a_i(t)}. \quad (21)$$

Taking vehicle combination  $Vc_1$  as an example, assuming that the real speed, position, and acceleration of  $\text{veh}_1$  and  $\text{veh}_2$  obtained are  $p_1(t), v_1(t), a_1(t)$  and  $p_2(t), v_2(t), a_2(t)$ , by substituting the speed  $v_2(t)$  and acceleration  $a_2(t)$  of  $\text{veh}_2$  into equation (8), the estimated speed  $v_1(\Theta(t))'$  of vehicle  $\text{veh}_1$  can be calculated. While varying  $\Theta(t)$  within the feasible range leads to different results for  $v_1(\Theta(t))'$ , when the estimated speed  $v_1(\Theta(t))'$  of vehicle  $\text{veh}_1$  is closest to its real speed  $v_1(t)$ ,

the corresponding value of  $\Theta(t)$  is the result obtained using vehicle combination  $Vc_1$ . Expanding this calculation process to all vehicle combinations in the entire platoon, in all vehicle combinations ( $Vc_1, Vc_2, \dots, Vc_{j-1}$ ), the leading vehicles are  $\text{veh}_1, \text{veh}_1, \dots, \text{veh}_{j-1}$ , respectively, each with different values of  $\Theta(t)$  within the feasible range, and their respective estimated speeds  $v_1(\Theta(t))', v_2(\Theta(t))', \dots, v_{j-1}(\Theta(t))'$  also change. When the overall error in the estimated speeds of all vehicles is minimized, the corresponding  $\Theta(t)$  is the final result after the calculations are completed.

Thus, an objective function can be constructed to solve for the optimal speed correction factor  $\Theta(t)$ , taking into consideration all vehicle combinations ( $Vc_1, Vc_2, \dots, Vc_{j-1}$ ) within the detection range that can be used for calculations. The  $i$ -th vehicle in the platoon used for estimation is denoted as  $\text{veh}_i$ , and the optimization objective is to minimize the root mean square error between  $\text{veh}_i$ 's real speed and estimated speed, as shown in the following equation:

$$\begin{aligned} \min \text{ velocity error } (v', \theta) &= \sqrt{\frac{1}{j-1} \sum_{i=1}^{j-1} [v_i(\Theta(t))' - v_i(t)]^2}, \\ \text{s.t. } v_i(\Theta(t))' &= \max\{\Theta(t)[v_{i+1}(t) + \lambda a_{i+1}(t)], 0\}, \\ 0 &\leq \Theta(t) \leq \frac{v_{\max}}{v_i(t) + \lambda a_i(t)}. \end{aligned} \quad (22)$$

**3.5.2. Time Headway Correction Factor  $\kappa(t)$  and Acceleration Exponent Correction Factor  $\beta(t)$ .** When adjusting the value of the correction factor  $\kappa(t)$ , it is essential to take into account the time headway situation in the actual traffic scenario. Therefore, the values of the integral term  $\kappa(t)T$  should satisfy the following equation:

$$T_{\min} \leq \kappa(t)T \leq T_{\max}, \quad (23)$$

where  $T_{\min}$  and  $T_{\max}$  represent the minimum and maximum time headways, respectively. Based on equation (5), after calibrating the extracted NGSIM data, we obtain  $T_{\min} = 0.8$  s and  $T_{\max} = 5$  s. Equation (22) can be rearranged to obtain the range of values for  $\kappa(t)$ , as shown in the following equation:

$$\frac{T_{\min}}{T} \leq \kappa(t) \leq \frac{T_{\max}}{T}. \quad (24)$$

Referring to Helbing et al. [28],  $1 \leq \delta \leq 5$ . The integral term  $\beta(t)\delta$  should satisfy  $1 \leq \beta(t)\delta \leq 5$ ; during the optimization process, we set  $\delta$  to a fixed value of 5, and  $\beta(t)$  is constrained as follows:



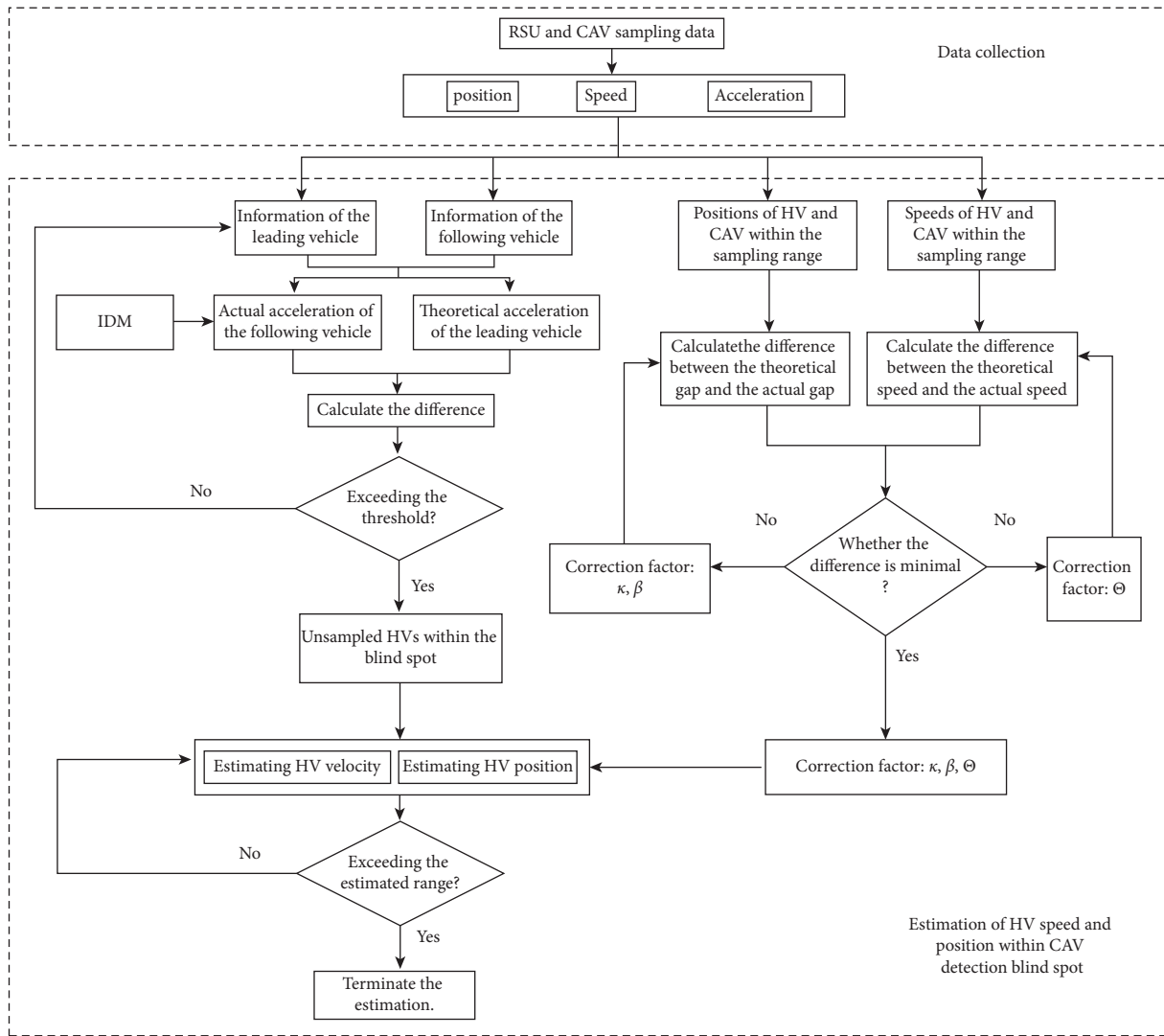


FIGURE 8: Flowchart of vehicle position estimation method using CAV sampling data.

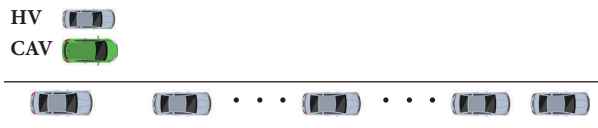


FIGURE 9: Vehicle queue scenarios extracted from NGSIM Data.

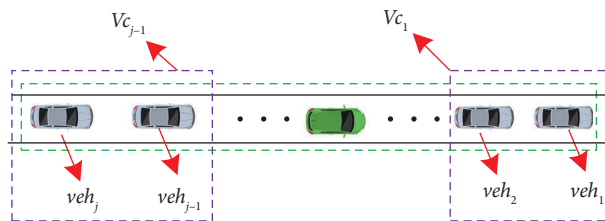


FIGURE 10: Scenario with HV assumed as CAV.

$$0 \leq \beta(t) \leq 1. \quad (25)$$

Taking vehicle combination  $V_{c_1}$  as an example, by substituting the speed  $v_2(t)$ , position  $p_2(t)$ , and acceleration  $a_2(t)$  of  $\text{veh}_2$  into equation (19), the estimated position  $p_1((\kappa(t), \beta(t)))'$  of vehicle  $\text{veh}_1$  can be calculated. While varying  $\kappa(t)$  and  $\beta(t)$  within the feasible range leads to different results for  $v_1(\Theta(t))'$ , when the estimated position  $p_1((\kappa(t), \beta(t)))'$  of vehicle  $\text{veh}_1$  is closest to its real position  $p_1(t)$ , the corresponding value of  $\kappa(t)$  and  $\beta(t)$  is the result obtained using vehicle combination  $V_{c_1}$ . Expanding this calculation process to all vehicle combinations in the entire platoon, in all vehicle combinations ( $V_{c_1}, V_{c_2}, \dots, V_{c_{j-1}}$ ), the leading vehicles are  $\text{veh}_1, \text{veh}_2, \dots, \text{veh}_{j-1}$ , respectively, each with different values of  $\kappa(t)$  and  $\beta(t)$  within the feasible

range, and their respective estimated positions  $p_1((\kappa(t), \beta(t)))', p_2((\kappa(t), \beta(t)))', \dots, p_{j-1}((\kappa(t), \beta(t)))'$  also change. When the overall error in the estimated positions of all vehicles is minimized, the corresponding value of  $\kappa(t)$  and  $\beta(t)$  is the final result after the calculations are completed.

Thus, an objective function can be constructed to solve for the optimal speed correction factor  $\Theta(t)$ , taking into consideration all vehicle combinations ( $V_{c_1}, V_{c_2}, \dots, V_{c_{j-1}}$ ) within the detection range that can be used for calculations. The  $i$ -th vehicle in the platoon used for estimation is denoted as  $\text{veh}_i$ , and the optimization objective is to minimize the root mean square error between  $\text{veh}_i$ 's real speed and estimated speed, as shown in the following equation:

$$\begin{aligned} \min \text{ position error} \left( P', \kappa, \beta \right) &= \sqrt{\frac{1}{j-1} \sum_{i=1}^{j-1} \left[ p_i((\kappa(t), \beta(t)))' - p_i(t) \right]^2}, \\ \text{s.t. } p_i(\kappa(t), \beta(t))' &= p_{i+1}(t) + \Delta s(\kappa(t), \beta(t)), \\ 0 &\leq \beta(t) \leq 1, \\ \frac{T_{\min}}{T} &\leq \kappa(t) \leq \frac{T_{\max}}{T}. \end{aligned} \quad (26)$$

In summary, the determination of the three correction factors has been summarized into two nonlinear optimization problems. Based on equations (1)–(26), we are able to estimate the information of the HVs in the detection blind zone between CAV1 and CAV2, including position and speed. Once a vehicle is estimated, it is treated as a known vehicle and used to continue estimating the positions of the HVs in lead. The position estimation stops when the position estimation interval is exceeded, and the complete convoy position information is obtained. To better describe the entire position estimation process, we provide Algorithm 1 as an explanation.

## 4. NGSIM Data-Based Experiments

**4.1. Experimental Setup.** Firstly, we set up various experimental scenarios involving different densities and penetration rates, as well as nonideal conditions more in line with real traffic scenarios, to evaluate the performance of the proposed method.

Many scholars choose NGSIM (the Next-Generation Simulation) data as their data source when researching real vehicle trajectories. These data encompass the southbound lanes of US-101, the eastbound lanes of California's Emeryville I-80, Peachtree Street in Atlanta, Georgia, and others. To validate the practicality of the proposed methodology in this study, we selected vehicle trajectory data from the US-101 segment as experimental data to assess the performance of the proposed method and for comparative experiments. The characteristics of the experimental segment are illustrated in Figure 11.

The experiments were conducted under varying traffic densities ranging from 30 veh/km to 65 veh/km, with density intervals of 5 veh/km, as this range covers most real-world traffic scenarios. In total, 400 traffic scenarios were extracted from the experimental data, with 50 scenarios extracted for each density interval. Each scenario represents a continuous queue of vehicles on a lane, with the number of vehicles in the queue corresponding to the density condition. The data include vehicles' IDs, timestamps, positions, speeds, accelerations, and lane IDs.

**4.2. Optimization Problem Calculation.** To solve the optimization problem of calculating the correction factors mentioned in Section 3.5, it is essential to solve two sets of nonlinear optimization functions in equations (20) and (21). In this paper, we used the Particle Swarm Optimization (PSO) algorithm to solve such optimization problems, with the following parameters: population size of 100, inertia weight of 0.4, individual weight of 0.7, social weight of 0.9, maximum iteration of 500, minimum particle movement step of  $1e-8$ , and minimum change value of objective function of  $1e-8$ . Since the initial positions of the population are different, which can lead to different results for each run, we repeated the optimization process 10 times and selected the correction factors corresponding to the minimum error as the optimization result. We no longer assign specific IDM parameters to CAVs but instead assume a human-like behavior for CAVs. Therefore, both CAVs and HVs utilize the IDM parameters of HVs. The IDM

```

(1) %Step 1: Obtain the speed, position, and acceleration information of the vehicles within the detection range of CAV1 and CAV2.
(2) Vehicle information:  $vehicles = \{veh_1, \dots, veh_2, \dots, veh_j\}$ ;
(3) Speed information:  $v = \{v_1(t), \dots, v_n(t), v_{n+1}(t), \dots, v_j(t)\}$ ;
(4) Position information:  $p = \{p_1(t), p_2(t), \dots, p_j(t)\}$ ;
(5) Acceleration information:  $a = \{a_1(t), a_2(t), \dots, a_j(t)\}$ ;
(6) %Step 2: Calculate the speed correction factor  $\Theta(t)$ , acceleration index correction factor  $\kappa(t)$ , and headway distance correction factor  $\beta(t)$  at the current time.
(7) While  $veh_i \in \{veh_1, veh_2, \dots, veh_{j-1}\}$ :
(8)   Optimization Correction Factors  $\Theta(t)$ :
(9)      $\min \sqrt{(1/j - 1) \sum_{i=2}^j (v_i(\Theta(t))' - v_i(t))}$ 
(10)   Optimization Correction Factors  $\kappa(t)$  and  $\beta(t)$ :
(11)      $\min \sqrt{(1/j - 1) \sum_{i=2}^j (p_i(\kappa(t), \beta(t))' - p_i(t))^2}$ 
(12)   Output:  $\Theta(t), \kappa(t), \beta(t)$ 
(13) %Step 3: Estimating the position of the HV
(14) Determine if there are any undetected HVs that need to be estimated:
(15) Calculate the theoretical acceleration of  $veh_j$ :
(16)  $a_{anticipate}(t) = a_{max} [1 - (v_n(t)/v_{max})^\delta - (S^*(t)/\Delta s(t))^2]$ 
(17) If  $a_{estimated}(t) - a_{actual}(t) > \sigma$ :
(18)   %There are HVs that have not been detected and need to be estimated.
(19)   While  $p_{estimate,i}(t) \in [p_n(t) + s_a + l, p_{n-1}(t) - s_a - l]$ :
(20)     %Calculate the velocity of the HVs within the undetected range
(21)      $v_i(t) = \max\{\Theta(t)[v_n(t) + \lambda a_n(t)], 0\} + v_{n-1}(t)/2$ 
(22)     %Calculating the vehicle spacing
(23)      $\Delta s(t) = s_a + \max\{v_n(t)\kappa(t)T + (v_n(t)[v_n(t) - v_{n-1}(t)]/2\sqrt{a_{max} \cdot b}), 0\} / \sqrt{1 - (v_n(t)/v_{max})^{\beta(t)\delta} - (a_n(t)/a_{max})}$ 
(24)     %Calculate the position of the undetected HV.
(25)      $p_{i-1,estimate}(t) = p_i(t) + \Delta s(t) + l$ 
(26)   Else:
(27)     Continue
(28) Output:  $\{p_{estimate}(t)\}$ 

```

ALGORITHM 1: Speed and position estimation for unsampled human vehicles.

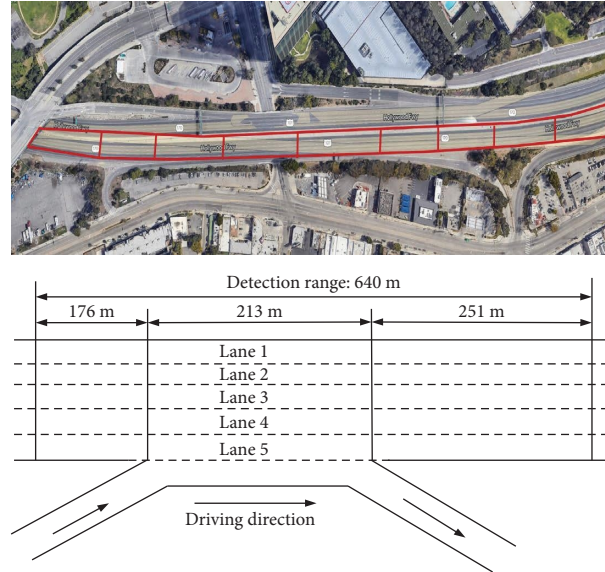


FIGURE 11: Layout of the studied road section.

parameters used in the calculations are based on the research by Chen et al. [20] as shown in Table 3. The other experimental parameters are shown in Table 4, and the optimization results of the correction factors for each density are shown in Table 5.

**4.3. Performance under Different Densities.** In this section, we consider a fixed number of CAVs in the platoon, which is set to 2 since the proposed estimation method requires at least two CAVs. To evaluate the performance of the proposed method under nonideal conditions, we assume that

TABLE 3: Selected IDM parameters.

Model parameters	Definition	Unit	Value
$s_a$	Static safety distance	m	2.48
$v_{\max}$	Free-flow speed	m/s	32.8
$a_{\max}$	Maximum acceleration	m/s <sup>2</sup>	2.78
$b_c$	Comfortable deceleration	m/s <sup>2</sup>	2.35
$T$	Safe time headway	s	1.98

TABLE 4: Other experimental parameters.

Parameters	Definition	Unit	Experimental values
$l$	Vehicle length	m	4.5
$\sigma$	Estimation threshold	m/s <sup>2</sup>	1.96
$\kappa$	Correction factor	—	
$\beta$	Correction factor	—	
$\Theta$	Correction factor		
$\lambda$	Correction factor		0.162
Dec	CAV detection range	m	100
$P$	CAV penetration rate	%	
$D$	Traffic density	veh/km	

TABLE 5: Statistical results of optimized correction factors at various densities.

Traffic density (veh/km)	$\Theta$			$\beta$		$\kappa$		
	Max	Mean	MAE (m/s)	Max	Mean	Max	Mean	MAE (m)
30	1.49	1.08	$1.2 \times 10^{-8}$	0.85	0.66	1.18	0.69	1.18
35	1.29	1.03	$1.44 \times 10^{-8}$	0.82	0.61	1.0	0.71	1.37
40	1.32	1.01	$1.27 \times 10^{-8}$	0.83	0.62	1.08	0.66	1.39
45	1.43	0.94	$1.21 \times 10^{-8}$	0.86	0.57	0.82	0.63	1.06
50	2.01	0.98	$7.94 \times 10^{-4}$	0.83	0.62	0.99	0.68	0.3
55	1.24	0.85	$7.4 \times 10^{-9}$	0.78	0.56	1.33	0.90	0.46
60	3.78	1.09	$3.9 \times 10^{-3}$	0.82	0.60	1.36	0.87	1.41
65	3.47	1.05	$1.1 \times 10^{-2}$	0.73	0.58	1.29	0.83	1.12

the lead and rear vehicles in the platoon are CAVs, which enables us to estimate more vehicles. Figure 12 shows a typical scenario of estimating the positions of HVs in the unsampled blind zone of the platoon with a fixed number of CAVs. The positions of the vehicles sampled by CAVs are denoted by green dots, and the unsampled vehicles are denoted by red dots. Our goal is to estimate the velocities and positions of the unsampled vehicles based on the observed information. To evaluate the velocity estimation performance of unsampled vehicles, we calculate the mean absolute error between the true velocity and the estimated velocity at the current time for each unsampled vehicle, as shown in equation (26). To evaluate the position estimation performance of unsampled vehicles, we calculate the mean absolute error between the true position and the estimated position for each unsampled vehicle, as shown in the following equation:

$$V, \text{error}(t)_{MAE} = \frac{1}{k} \sum_{i=0}^k |v_i(t)_{\text{estimate}} - v_i(t)_{\text{actual}}|, \quad (27)$$

$$P, \text{error}(t)_{MAE} = \frac{1}{k} \sum_{i=0}^k |p_i(t)_{\text{estimate}} - p_i(t)_{\text{actual}}|.$$

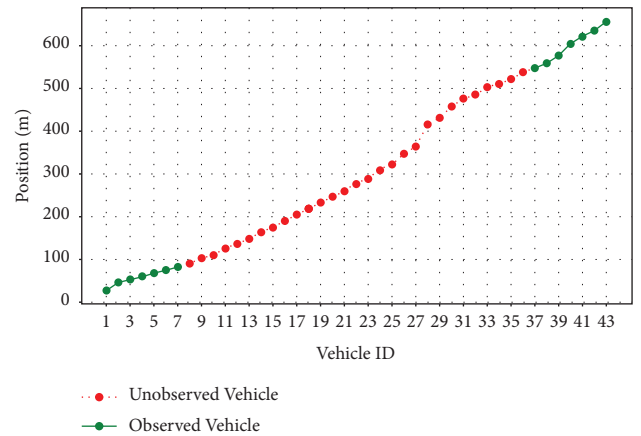


FIGURE 12: Example of experimental scene.

Figures 13(a)–13(h) present the average absolute errors between the estimated speeds and the true speeds of the vehicles under 50 different traffic scenarios within the density range of 30–65 veh/km for the two methods. Table 6 provides the statistical values of the errors. As shown in the figure, the proposed method has the best speed estimation accuracy. To further verify the stability of the proposed speed estimation method, the standard deviations of the mean

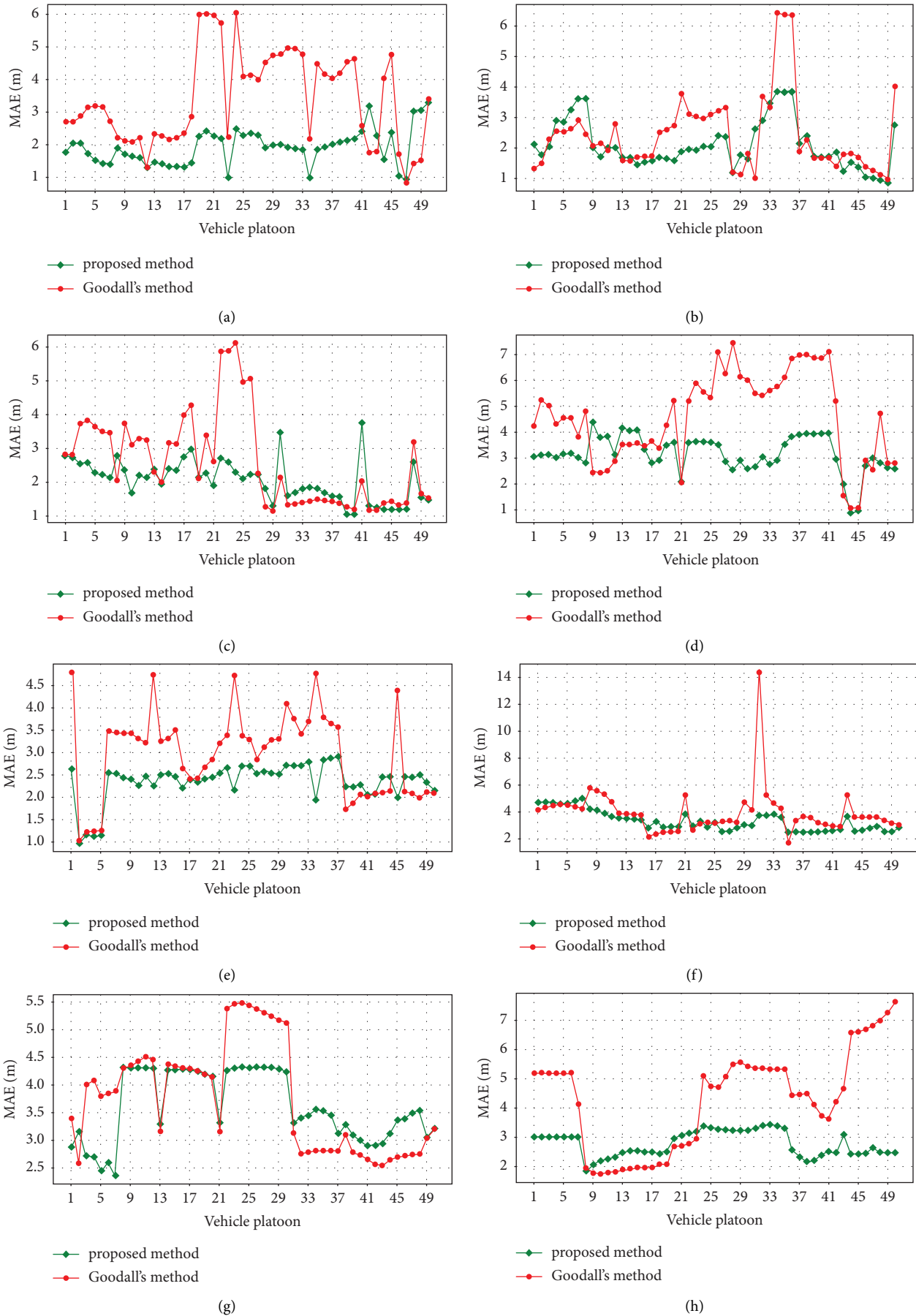


FIGURE 13: MAEs of speed estimation under different densities. (a)  $D = 30$  veh/km. (b)  $D = 35$  veh/km. (c)  $D = 40$  veh/km. (d)  $D = 45$  veh/km. (e)  $D = 50$  veh/km. (f)  $D = 55$  veh/km. (g)  $D = 60$  veh/km. (h)  $D = 65$  veh/km.

TABLE 6: The statistical values of the MAE of speed estimation under different densities.

Traffic density (veh/km)	MAE (m)				Performance improvement (%)
	Proposed method		Goodall's method		
	Mean	SD	Mean	SD	
30	2.28	0.87	3.40	1.27	+32.9
35	2.02	0.79	2.43	1.17	+16.9
40	2.06	0.62	2.75	1.42	+25.1
45	2.81	1.06	4.58	2.06	+38.6
50	2.36	0.87	2.95	1.24	+20
55	2.61	0.39	3.92	0.68	+33.4
60	3.31	0.62	3.76	1.01	+11.97
65	2.96	0.44	4.35	1.17	+47

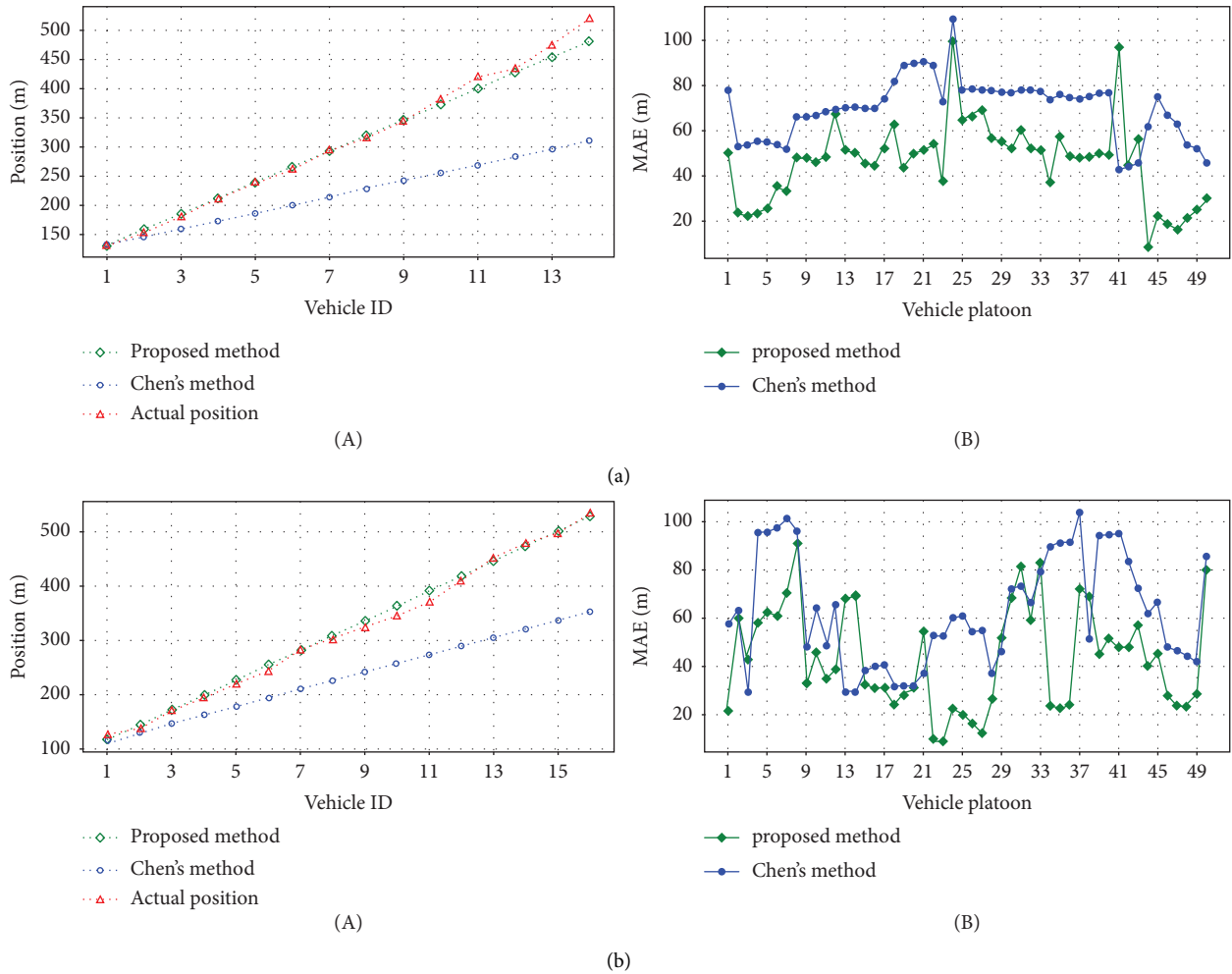
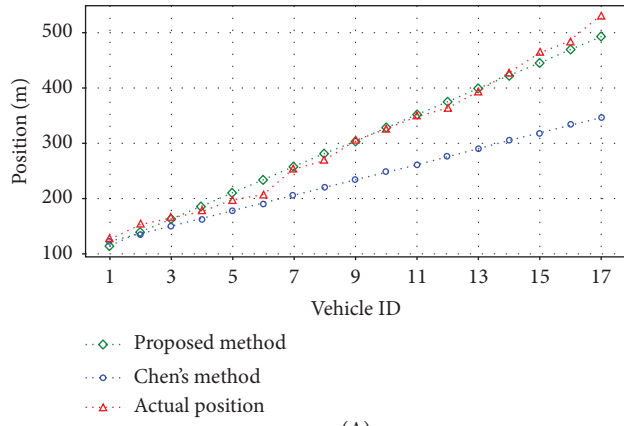
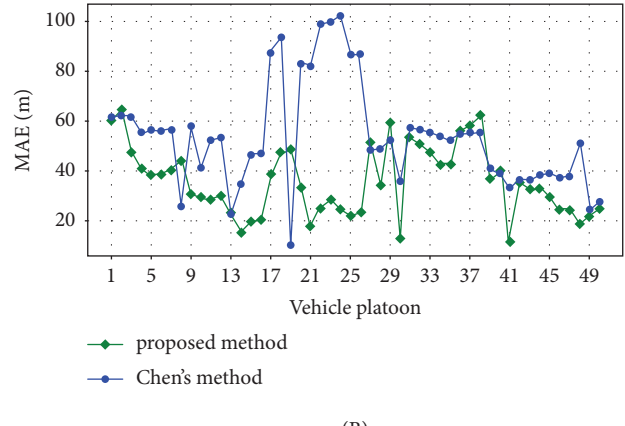


FIGURE 14: Continued.

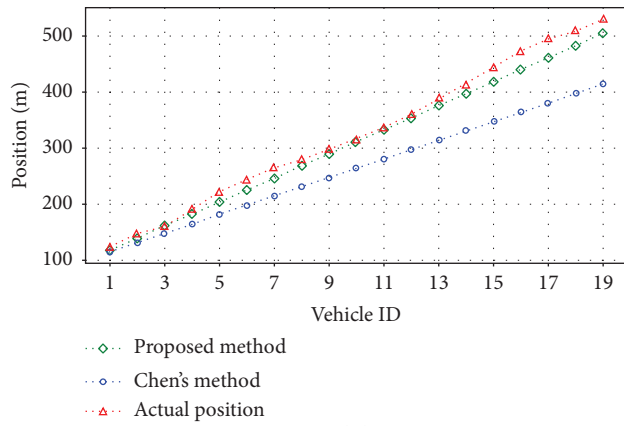


(A)

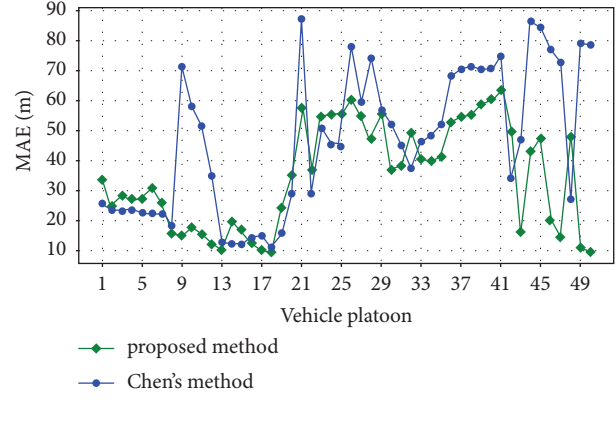


(B)

(c)

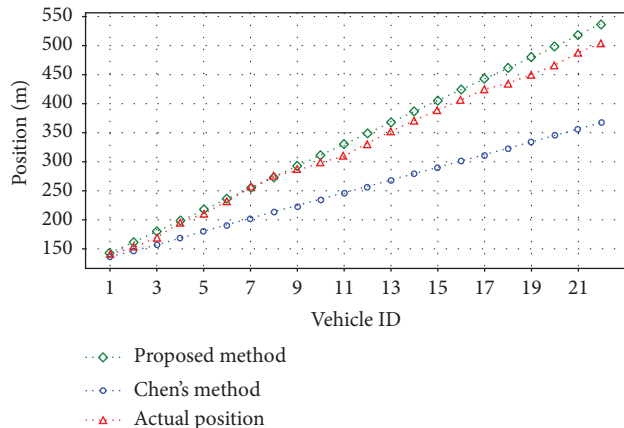


(A)

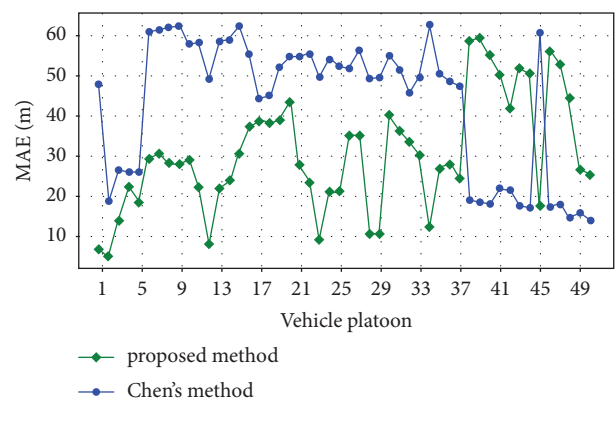


(B)

(d)



(A)



(B)

(e)

FIGURE 14: Continued.

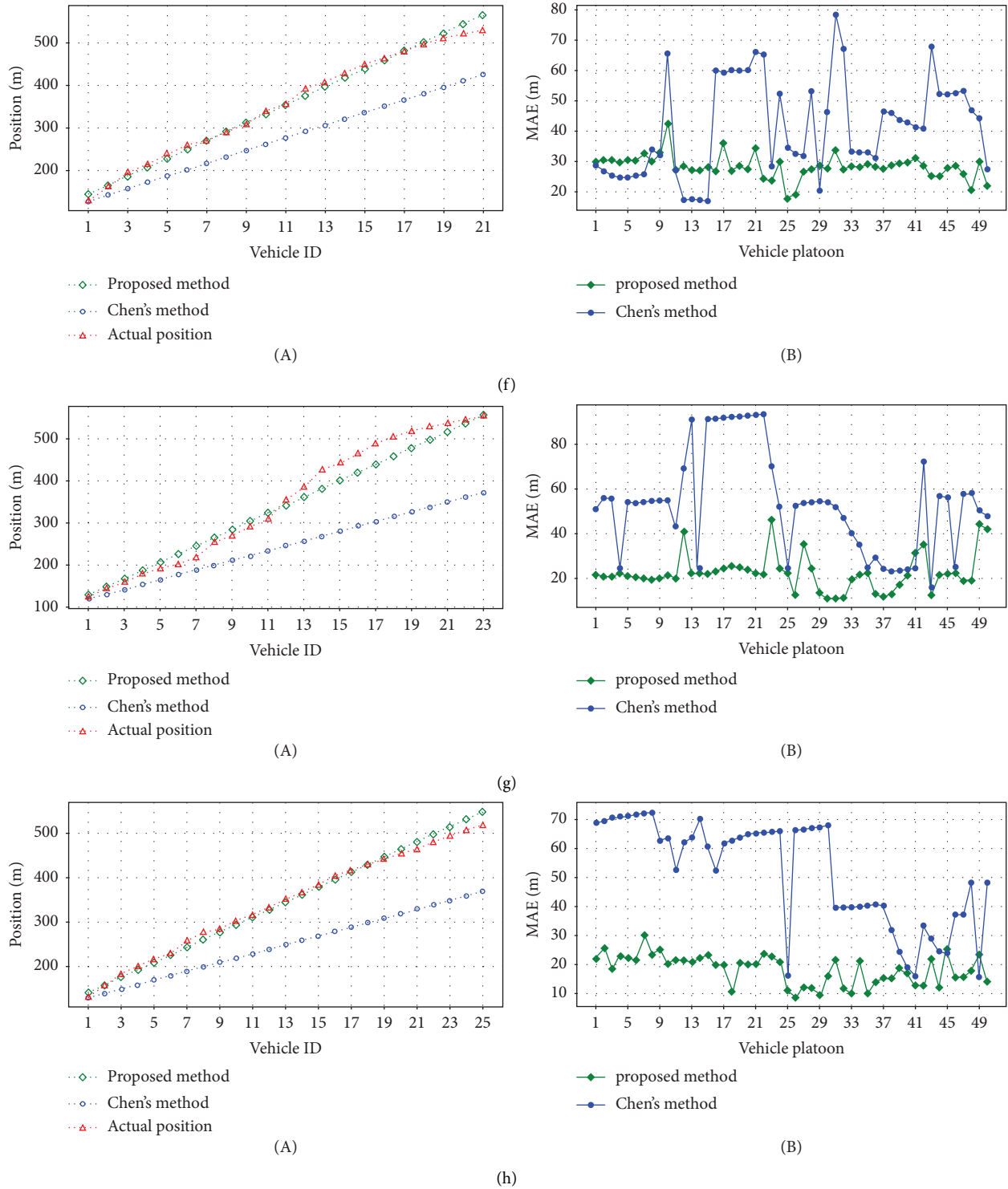


FIGURE 14: MAEs of position estimation under different densities. (a)  $D = 30$  veh/km. (b)  $D = 35$  veh/km. (c)  $D = 40$  veh/km. (d)  $D = 45$  veh/km. (e)  $D = 50$  veh/km. (f)  $D = 55$  veh/km. (g)  $D = 60$  veh/km. (h)  $D = 65$  veh/km.

absolute errors are also provided in Table 6. A lower value indicates stronger stability of the method. According to the statistical results in Table 6, the stability of the proposed method is better than that of the Goodall method at all densities because the proposed method considers the heterogeneity of the following behavior caused by different

traffic densities, which affects the vehicle speed and can better reflect the speed of vehicles in different traffic scenarios, based on the Goodall algorithm.

Figure 14 shows the estimated vehicle positions and the true vehicle positions under 50 different traffic scenarios within the density range of 30–65 veh/km for both the



TABLE 7: The statistical values of the MAE of position estimation under different densities.

Traffic density (veh/km)	Proposed method		Chen’s method		Performance improvement (%)
	Mean	SD	Mean	SD	
30	44.09	19.99	69.97	22.62	36.9
35	42.51	21.30	62.94	23.06	32.5
40	37.33	15.83	53.62	21.27	30.4
45	34.87	14.84	46.80	24.13	25.5
50	31.19	14.63	42.46	17.38	26.5
55	27.31	4.38	41.5	16.11	34.2
60	24.88	4.29	53.45	22.70	53.5
65	17.43	3.81	51.87	18.18	66.4

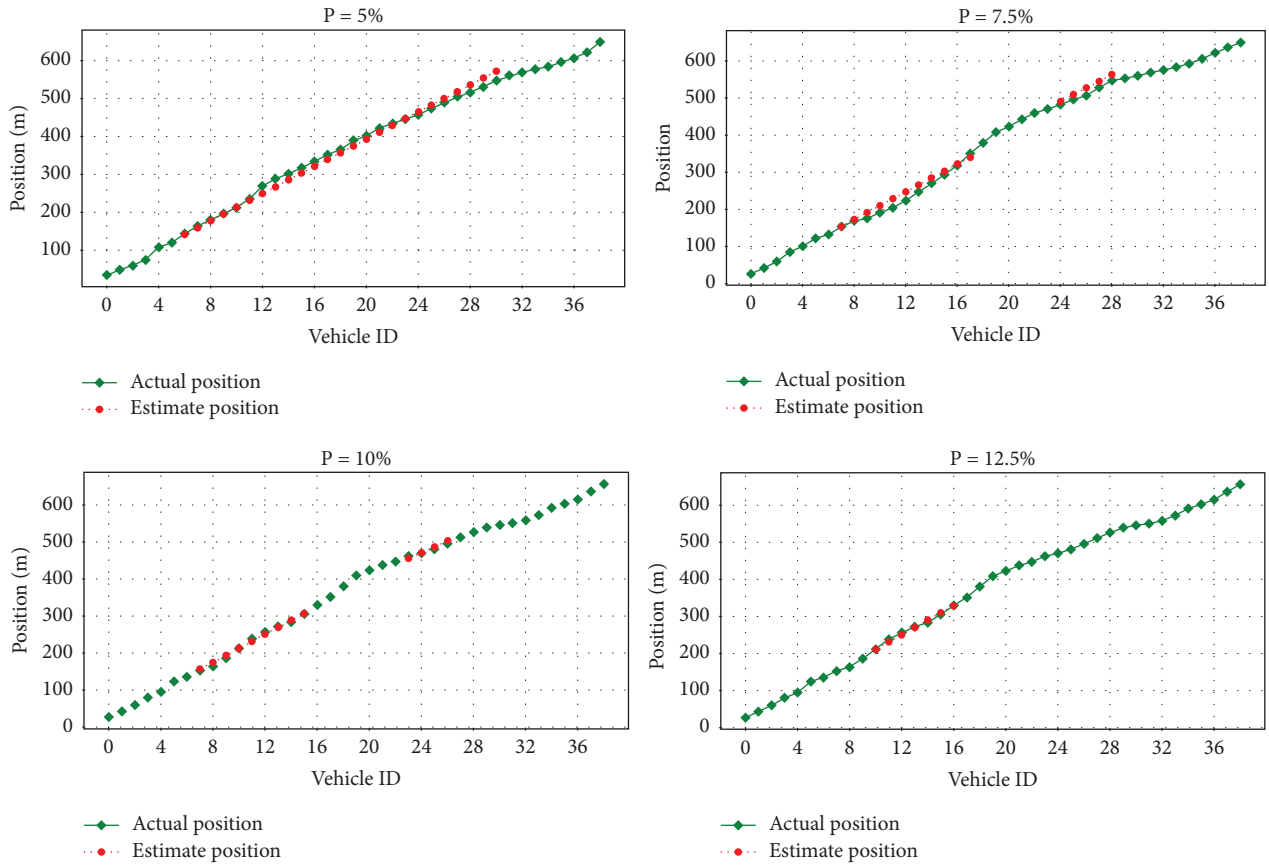


FIGURE 15: Estimated positions under different penetration rates.

proposed method and the method proposed by Chen, which uses calibrated expected spacing such as (a-1). It also shows the average absolute errors between the estimated positions and the true positions for both methods such as (a-2). The red dashed line represents the true vehicle position, the green dashed line represents the vehicle position estimated by the proposed algorithm, and the blue dashed line represents the vehicle position estimated by Chen’s method. From the subplots, it can be observed that the number of vehicles to be estimated increases with the increase in traffic density. Moreover, under various density conditions, the proposed method achieves the best estimation accuracy.

Table 7 presents the average mean absolute error (MAE) for position estimation of the two methods under different densities. The standard deviation of the average MAE is also

provided to assess the stability of the proposed position estimation method. The results demonstrate that the proposed method outperforms Chen’s method in terms of accuracy and stability across various density conditions. The error in vehicle position estimation decreases as the traffic density increases, which is expected. In low-density traffic, the larger vehicle spacing and weaker interaction between vehicles pose challenges for the proposed method to capture detailed driving behavior accurately. However, as the traffic density increases, the reduced vehicle spacing and enhanced following characteristics make it easier for the proposed method to interpret the driving behavior of each vehicle, leading to improved position estimation accuracy. On the other hand, Chen’s method exhibits an increasing MAE from 55 veh/km to 60 veh/km. This can be attributed to Chen’s method

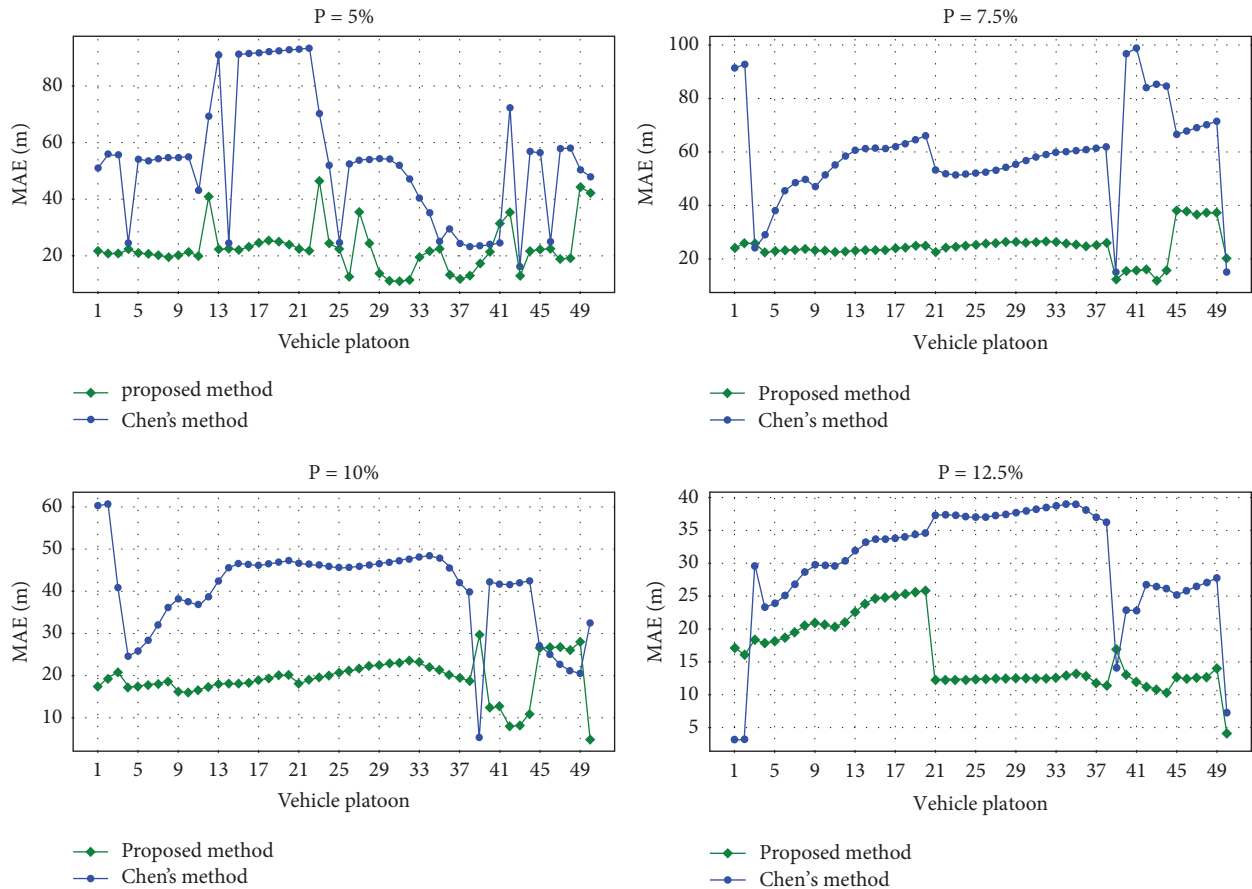


FIGURE 16: MAEs of position estimation under different penetration rates.

primarily considering the impact of traffic density through calibrating the desired headway while overlooking important factors such as variations in vehicle speeds, headways, and acceleration rates of change, which affect the accuracy of position estimation. In summary, in real-world traffic environments, calibrating the desired headway method is more suitable for medium to low traffic densities. However, the proposed method, incorporating three correction factors of speed, time headway, and acceleration rate of change, provides more accurate position estimation for unsampled vehicles in a variety of traffic scenarios.

**4.4. Performance under Different CAV Penetration Rates.** In this section, the traffic density is set as a fixed constant, and we conducted experiments using data from a traffic density of 60 vehicles/km, which falls between free flow and congested flow. At this density, there are 40 vehicles on the selected US-101 road segment, and considering that the proposed method requires a minimum of 2 CAVs, the minimum required CAV penetration rate should be 5%. Therefore, we set the CAV penetration rates to 5%, 7.5%, 10%, and 12.5% and randomly selected the corresponding number of vehicles in the platoon to be CAVs. The experiments were conducted on 50 randomly selected traffic scenarios using different random seeds.

Figure 15 shows the estimated vehicle positions (represented by red dashed lines) and the actual vehicle positions (represented by green solid lines) for the proposed method under different penetration rates. It is evident that as the penetration rate of CAV increases, the number of human-driven vehicles (HVs) to be estimated decreases. This indicates that the proposed method effectively captures the characteristics of real traffic scenarios.

Figure 16 shows the variation of mean absolute error (MAE) in position estimation between the proposed method (represented by green solid lines) and Chen's method (represented by blue solid lines) across 50 different convoy scenarios under varying penetration rates. Similarly, Figure 17 shows the MAE variation in speed estimation between the proposed method (represented by green solid lines) and Goodall's method (represented by red solid lines) under different penetration rates. To provide a comprehensive performance evaluation, the average MAE values for position and velocity estimation were computed and are presented in Table 8. Notably, the results demonstrate a consistent decrease in MAE values for both position and velocity estimation as the penetration rate increases. This can be attributed to the increased availability of HV trajectory information within the CAV sampling range, which contributes to enhanced optimization of the correction factors

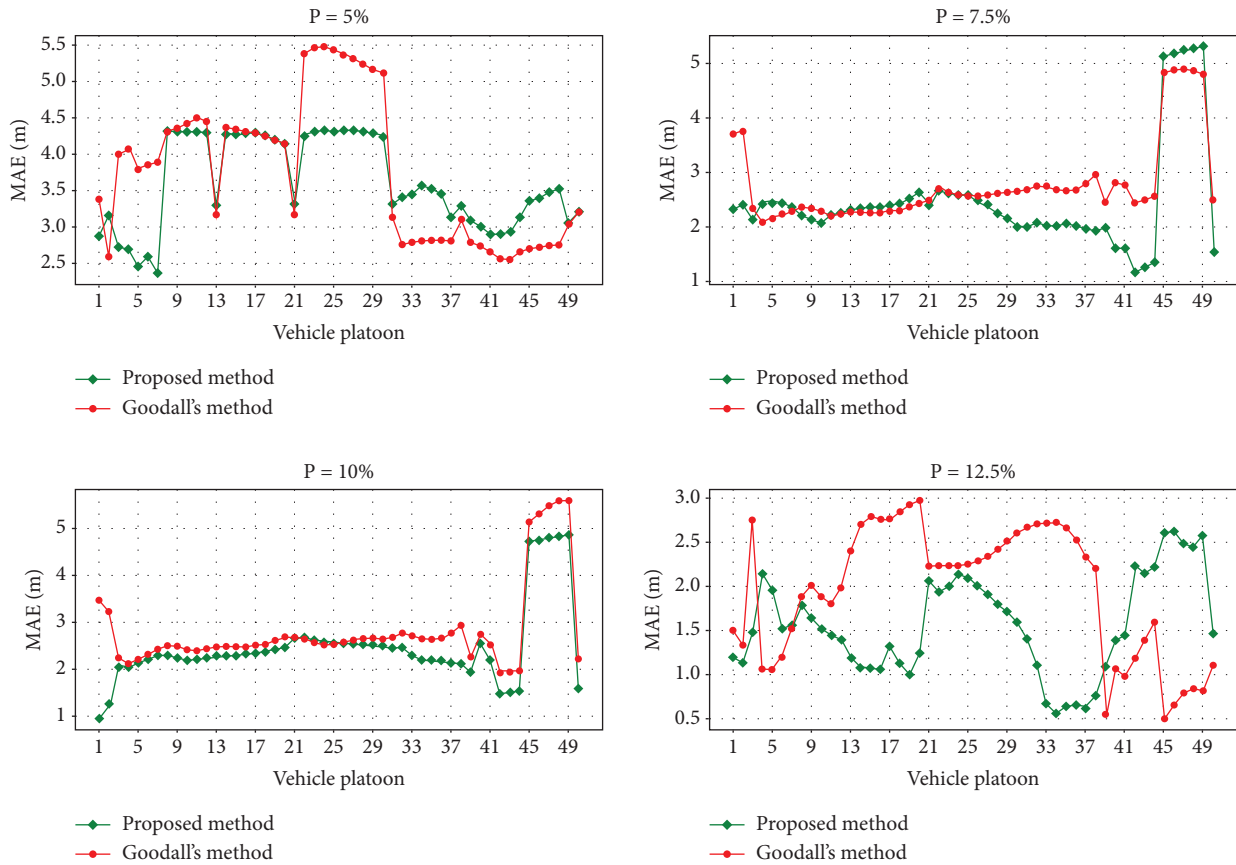


FIGURE 17: MAEs of speed estimation under different penetration rates.

TABLE 8: Statistical values of MAE under different penetration rates.

Penetration rate (%)	Speed estimation MAE (m/s)			Position estimation MAE (m)		
	Proposed method	Goodall's method	Performance improvement (%)	Proposed method	Chen's method	Performance improvement (%)
5	3.31	3.86	14.3	26.02	53.4	51.3
7.5	2.36	2.79	15.4	22.58	49.4	54.3
10	2.11	2.62	19.5	18.27	40.43	54.8
12.5	1.56	2.34	33.3	13.25	30.19	56.1

and subsequently improves the accuracy of velocity and position estimation. Furthermore, when compared to Goodall's method, the proposed velocity estimation method consistently outperforms it across all penetration rate scenarios. Moreover, the proposed position estimation method significantly reduces the error in HV position estimation compared to Chen's method.

### 5. Conclusion and Future Work

This paper presents a method for velocity and position estimation of vehicles based on sampled data from CAVs and its validation using real-world datasets. Firstly, vehicle trajectory data under different densities are extracted from the NGSIM dataset. Based on the model proposed by Goodall for estimating the preceding vehicle speed, the influence of the

preceding vehicle's speed variation on the following vehicle's acceleration is analyzed under different densities. The results show that this influence increases with increasing traffic density under noncongested conditions but decreases when the density is between congested and free-flowing states. Therefore, a velocity correction factor is introduced based on the Goodall model, and the correction factors are optimized using CAV sampled data, transforming the determination of the correction factors into a nonlinear optimization problem. Next, the variations of average time headway and acceleration rate of change are analyzed under different densities. The experimental results demonstrate that when density is held constant, the average time headway within different platoons exhibits fluctuations due to the heterogeneity in following behaviors, while the acceleration rate of change exhibits higher randomness under low densities. Therefore, time

headway correction factors and acceleration exponent correction factors are introduced in IDM, and their values are determined using CAV detection data. The determination of these correction factors is transformed into a nonlinear optimization problem involving two parameters. Then, based on the proposed speed estimation model and position estimation model, the speed and position of unsampled HVs are estimated in a CAV platoon. Finally, the performance of the proposed method is experimentally validated using the NGSIM dataset. The results show that even under the extreme condition where there are only two CAVs in the platoon, serving as the lead and trailing vehicles, the proposed method can more accurately reflect the true speeds of vehicles under different densities compared to the linear model's method. Furthermore, compared to the method that estimates the positions of unsampled vehicles using calibrated desired headways, the proposed method significantly reduces the estimation errors of undetected vehicle positions. Additionally, the CAV penetration rate has a minimal impact on the estimation results, and in general, the error in position estimation decreases with an increase in the penetration rate.

Although the influence of unstable following behavior has been considered in this study, the range of consideration is limited to the number of vehicles within the detection range of CAV because the optimization process of IDM parameters is based on the vehicle data within the detection range of CAV. In future research, the proposed method can be further extended by considering the time-variant factors in the CAV detection blind zone to reproduce more realistic HV information. However, the proposed estimation method is based on the moment level, which greatly improves the effect of the lane change behaviors of multilane.

## Data Availability

The dataset is available at <https://data.transportation.gov/Automobiles/Next-Generation-Simulation-NGSIM-Vehicle-Trajectory/8ect-6jqj>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] Y. Zhao, S. Y. Shen, and H. X. Liu, "A hidden Markov model for the estimation of correlated queues in probe vehicle environments," *Transportation Research Part C: Emerging Technologies*, vol. 128, Article ID 103128, 2021.
- [2] C. Tan, J. Yao, K. Tang, and J. Sun, "Cycle-based queue length estimation for signalized intersections using sparse vehicle trajectory data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 91–106, 2021.
- [3] L. Li, R. Jiang, Z. He, X. Chen, and X. Zhou, "Trajectory data-based traffic flow studies: a revisit," *Transportation Research Part C: Emerging Technologies*, vol. 114, pp. 225–240, 2020.
- [4] M. Ramezani and N. Geroliminis, "Queue profile estimation in congested urban networks with probe data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 30, no. 6, pp. 414–432, 2015.
- [5] P. Wang, J. Lai, Z. Huang, Q. Tan, and T. Lin, "Estimating traffic flow in large road networks based on multi-source traffic data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 9, pp. 5672–5683, 2021.
- [6] T. Ghoul and T. Sayed, "Real-time signal-vehicle coupled control: an application of connected vehicle data to improve intersection safety," *Accident Analysis and Prevention*, vol. 162, no. 2021, Article ID 106389, 2021.
- [7] X. Liu, J. Guan, R. Jiang, S. Ge, and B. Chen, "Finite-horizon URTSS-based position estimation for urban vehicle localization," *IEEE Sensors Journal*, vol. 23, no. 4, pp. 4011–4021, 2023.
- [8] Y. Zhou, Y. Lin, S. Ahn, P. Wang, and X. Wang, "Platoon trajectory completion in a mixed traffic environment under sparse observation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16217–16226, 2022.
- [9] Y. Wang, L. Wei, and P. Chen, "Trajectory reconstruction for freeway traffic mixed with human-driven vehicles and connected and automated vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 135–155, 2020.
- [10] H. Bar-Gera and Hillel, "Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel," *Transportation Research Part C: Emerging Technologies*, vol. 15, no. 6, pp. 380–391, 2007.
- [11] A. Georgiev and P. K. Allen, "Localization methods for a mobile robot in urban environments," *IEEE Transactions on Robotics*, vol. 20, no. 5, pp. 851–864, 2004.
- [12] M. E. El Najjar, P. Bonnifait, and P. Bonnifait, "A road-matching method for precise vehicle localization using belief theory and kalman filtering," *Autonomous Robots*, vol. 19, no. 2, pp. 173–191, 2005.
- [13] J. C. Herrera and A. M. Bayen, "Incorporation of Lagrangian measurements in freeway traffic state estimation," *Transportation Research Part B: Methodological*, vol. 44, no. 4, pp. 460–481, 2010.
- [14] P. Vadakkepat and L. Jing, "Improved particle filter in sensor fusion for tracking randomly moving object," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 5, pp. 1823–1832, 2006.
- [15] S. Perea, A. Morell, J. Toledo, L. Acosta, and Daniel, "GNSS integration in the localization system of an autonomous vehicle based on particle weighting," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3314–3323, 2019.
- [16] A. Liu, L. Lian, V. Lau, G. Liu, and M. Zhao, "Cloud-assisted cooperative localization for vehicle platoons: a turbo approach," *IEEE Transactions on Signal Processing*, vol. 68, pp. 605–620, 2020.
- [17] H. Qin, Y. Peng, and W. Zhang, "Vehicles on RFID: error-cognitive vehicle localization in GPS-less environments," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 9943–9957, 2017.
- [18] X. Jeff Ban, P. Hao, and Z. Sun, "Real time queue length estimation for signalized intersections using travel times from mobile sensors," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1133–1156, 2011.
- [19] Y. Feng, J. Sun, and P. Chen, "Vehicle trajectory reconstruction using automatic vehicle identification and traffic count data," *Journal of Advanced Transportation*, vol. 49, no. 2, pp. 174–194, 2015.
- [20] P. Chen, T. Wang, and N. Zheng, "Reconstructing vehicle trajectories on freeways based on motion detection data of connected and automated vehicles," *Journal of Intelligent Transportation Systems*, vol. 26, no. 6, pp. 639–654, 2022.

- [21] J. W. C. H. Van Lint, "Empirical evaluation of new robust travel time estimation algorithms," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2160, no. 1, pp. 50–59, 2010.
- [22] X. Ji, "Research on vehicle position estimation method in intelligent connected traffic environment," Master's thesis, Southeast University, Dhaka, Bangladesh, 2019.
- [23] C. Nanthawichit, T. Nakatsuji, and H. Suzuki, "Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1855, pp. 49–59, 2003.
- [24] X. Chen, J. Yin, K. Tang, Y. Tian, and J. Sun, "Vehicle trajectory reconstruction at signalized intersections under connected and automated vehicle environment," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 17986–18000, 2022.
- [25] N. J. Goodall, B. Smith, and B. Park, "Microscopic estimation of freeway vehicle positions using mobile sensors," in *Proceedings of the 91st Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2012.
- [26] N. J. Goodall, B. Smith, and B. Park, "Traffic signal control with connected vehicles," *Transportation Research Record*, vol. 2381, no. 1, pp. 65–72, 2013.
- [27] Z. Yao, M. Liu, Y. Jiang, Y. Tang, and B. Ran, "Trajectory reconstruction for mixed traffic flow with regular, connected, and connected automated vehicles on freeway," *IET Intelligent Transport Systems*, 2022.
- [28] D. Helbing, A. Hennecke, V. Shvetsov, and M. Treiber, "Micro-and macro-simulation of freeway traffic," *Mathematical and Computer Modelling*, vol. 35, no. 5-6, pp. 517–547, 2002.