WILEY | Hindawi

Research Article

Freeway Traffic Speed Estimation of Mixed Traffic Using Data from Connected and Autonomous Vehicles with a Low Penetration Rate

Shanglu He [1,2] **Xiaoyu Guo,**[3] **Fan Ding,**[4,5] **Yong Qi,**[1] **and Tao Chen**[2]

[1]*School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China*
[2]*Key Laboratory for Automotive Transportation Safety Enhancement Technology of the Ministry of Communication, Chang'an University, Xi'an 710084, China*
[3]*Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77843-3135, USA*
[4]*Joint Research Institute on Internet of Mobility, Southeast University and University of Wisconsin-Madison, Madison, WI, USA*
[5]*School of Transportation, Southeast University, Nanjing 211189, China*

Correspondence should be addressed to Shanglu He; slhemickey@126.com

Connected and autonomous vehicles (CAVs) are on the way to the field application. In the beginning stage, there will be a mixed traffic flow, containing the regular human-driven vehicles and CAVs with a low penetration rate. Recently, the discussion about the impact of a small proportion of CAVs in the mixed traffic is controversial. This paper investigated the possibility of applying the limited data from these lowly penetrated CAVs to estimate the average freeway link speeds based on the Kalman filtering (KF) method. First, this paper established a VISSIM-based microsimulation model to mimic the mixed traffic with different CAV penetration rates. The characteristics of this mixed traffic were then discussed based on the simulation data, including the sample size distribution, data-missing rate, speed difference, and fundamental diagram. Accordingly, the traditional KF-based method was introduced and modified to adapt data from CAVs. Finally, the evaluations of the estimation accuracy and the sensitive analysis of the proposed method were conducted. The results revealed the possibility and applicability of link speed estimation using data from a small proportion of CAVs.

1. Introduction

Autonomous vehicle (AV) technology is a hot and practical research spot. When AVs embedded with the feature to communicate with others including vehicles, roadside infrastructures, or traffic control centers, they are defined as connected and autonomous vehicles (CAVs). It is expected that CAVs can provide faster responses and keep shorter headways, which lead to an increased overall roadway capacity [1]. Other expected benefits of CAVs include improved mobility to people with disabilities, enhanced productive use of travel time, better fuel efficiency, fewer emissions, and flexibility in parking [2, 3]. However, it is estimated that the market penetration rate of CAVs might be able to reach between 24% and 87% by 2045 [4, 5]. Therefore,

there will be a long period of mixed traffic condition comprising CAVs and regular human-driven vehicles (RVs).

A majority of research works have been dedicated to the impact analysis of AVs/CAVs in the mixed traffic. Some focused on the impact on the traffic efficiency, i.e., capacity and throughput. For instance, Davis explored the contribution of adaptive cruise control (ACC) vehicles to the reduction in the jam formation [6]. Shladover et al. proved that the Cooperative Adaptive Cruise Control (CACC) technology has the potential to increase lane throughput from the average 2000 veh/h to approximately 4000 veh/h with high market penetrations [7]. Friedrich found that the traffic volume could be increased to about 3900 veh/h/lane when AVs are in application compared with current designed capacity values of a lane of 2200 veh/h [8]. Both

Zhou et al. and Xiao et al. found that a cooperative control of AVs would improve the traffic efficiency of the merging area [9, 10]. Some focused on modelling the different traffic behaviours of CAVs, such as fundamental diagram and longitudinal and latitudinal movements. For example, Baskar et al. demonstrated that RVs and ACC-equipped intelligent vehicles had the different fundamental diagrams [11]. Liu et al. changed the lane-changing rules in the cellular automata to simulate the autonomous vehicles [12]. Lu and Aakre proposed a smart driver model to simulate the car-following behaviour of CAVs [13]. Moreover, some discussed the influence on other aspects, i.e., safety and environment. For example, Morando et al. investigated the safety performance of AVs with varying penetration rates in two different cases, i.e., a roundabout and a signalized intersection [14]. Lu et al. improved the ACC model of CAVs and validated that these CAVs performed better than RVs in fuel economy [15, 16].

Obviously, these works admitted that CAVs had different behaviours compared with RVs. Most of them expected the CAVs to have a faster reaction time, and thus, CAVs could keep a smaller distance with the front vehicle and be safer. These works have validated that the application of CAVs is definitely beneficial when CAVs take a high proportion of traffic, but the impact of CAVs with a low penetration rate is controversial. If the penetration rate of CAVs is high in the mixed traffic, the information from CAVs is definitely sufficient to identify the traffic state. What if CAVs only take a low proportion of mixed traffic, will their information be enough to acquire or estimate the traffic state? Since the penetration rate of CAVs grows slowly, it is meaningful to explore whether these CAVs in low penetration rates are a new data source to assist the surveillance of the traffic condition.

Data provided by CAVs resemble the data collected through the traditional human-driven probe vehicles, such as global positioning system- (GPS-) based data and cell-phone-based data. Traffic state estimation based on these probe vehicles is one of the most effective methods because probe vehicles have a wide coverage over space and time [17–21]. There are two common categories of traffic state estimation methods, i.e., the model-based methods and the data-driven methods. The model-based methods are made up of two parts. The first part is the traffic flow model, such as the Lighthill-Whitham-Richards (LWR) model [22], Payne model [23], and their successors. The second part is a data assimilation method to realize the estimation, such as Kalman filtering (KF) and its extensions [19, 24]. The data-driven methods mine the relationship between estimates and observations from the historical big data. The commonly used data mining techniques include the statistical analysis algorithms for the time-series data and the artificial intelligence models [20]. However, it should be noted that the traditional probe-based methods are under the human-driven mode, and the probe and non-probe vehicles are supposed to have similar driving behaviour. As mentioned before, the driving behaviours of CAVs are expected to be different from those of RVs, so the applicability of the traditional probe-based estimation methods is uncertain. The data-driven methods require vast amount of historical data.

However, the CAVs have not been put into the market officially, so it is hard to achieve sufficient historical CAV data. Regarding these factors, this study would like to focus on the model-based estimation method using CAV data. There has been some research using the model-based method. For instance, Wang et al. compared the first-order and second-order models to estimate the mixed traffic state with different AV penetration rates [25], but they did not discuss the low penetration specifically. Considering the controversy under the low penetration condition, this study aims to furtherly discuss how to use the model-based estimation method with information from a small proportion of CAVs in mixed traffic.

More specifically, this study would firstly contribute to set up a simulation platform. Hereafter, this study would explore the sampling characteristics of CAV probes under a low penetration rate, such as their sample size, data-missing rate, and their speed difference with the average link speeds. Furthermore, whether their limited information is supportive to the traffic state estimation would be discussed. Afterwards, although the KF technique is widely used, this study makes the following adjustments to adapt the lowly penetrated CAVs: a recursive model to fulfil the missing parts, calculation methods for state, and measurement noise. Its performance and accuracy are going to be evaluated.

Accordingly, the rest of this paper is organized as follows: Section 2 introduces a simulation platform of mixed traffic to generate the data for the following investigations. The different characteristics of traffic with CAVs are discussed in Section 3. Section 4 presents the exploration of KF-based estimation. Finally, Section 5 summarizes the main conclusions and provides some plans for the improvement and future study.

2. Simulation Platform for Mixed Traffic

2.1. Simulation Settings

2.1.1. Assumptions. The highly or fully automated CAVs referring to Level 4 or Level 5 in the SAE autonomy level definitions [26] are still in development or test. The simulation method provides a possibility to studying the mixed traffic condition with CAVs. There is a bunch of expected types of highly or fully automated CAVs. Different types of CAVs would lead to different influences on traffic. Therefore, this study made some preceding assumptions to clarify the studied object and situation.

First, CAVs are supposed to behave more assertively than RVs, and thus, they can maintain a shorter distance with the front vehicle.

Second, CAVs have a stronger ability to sense the traffic environment compared with RVs. This ability could be enhanced either by the communication with everything (roadside unit, other vehicles, traffic management center, and so on) or by the advance sensing facilities. As a result, this sensing range is supposed to be within ± 500m in this study.

Third, since this study is based on the simulation, the latency and packet loss of the communication between CAVs and everything (roadside unit, other vehicles, traffic management center, and so on) would not be considered this time.

2.1.2. Simulation Parameters. This study uses VISSIM (version 9) to simulate the mixed traffic containing CAVs and RVs. PTV Group has stated that CAV behaviour could be modelled using VISSIM internally or externally [27]. This study implements the internal way, which is to modify the VISSIM default driving behaviour parameters. Comparatively speaking, the internal way is simpler and more convenient to use, whereas the external approach is used when researchers want to define their own driving behaviour models in VISSIM. Since the focus of this study is to estimate speeds from data generated by CAVs with a low penetration rate, the internal way is more suitable and achievable.

PTV Group has given some recommendations to set the internal model by changing the car-following and lane-changing behaviour parameters for the CAVs [28]. In application, there have been some works that are based on the internal model in VISSIM to explore the impact of CAVs. Table 1 summarizes their adjusted parameters as well as the corresponding default value in VISSIM 9. It should be noted that both this study and the works in Table 1 use the Wiedemann 99 model as the car-following model for the freeway traffic.

Since no empirical data are available, these applications have indicated the possibility of modelling CAVs in VISSIM internally, to some extent. Although it seems that each study has made different adjustment to the default values, they have something in common. For instance, they would let the CAV keep a shorter distance with the front vehicle, have faster and smoother reactions, observe more around vehicles, and realize the cooperative lane changing. Some differences might be caused by the different versions of VISSIM. For example, the maximum speed difference is different between VISSIM version above 9 and below 9. Within the threshold present in these existing studies, this study made the following modifications to the internal models in VISSIM 9, as shown in Table 2. RVs use the default values, while some parameters are adjusted for CAVs. Besides, the desired speed is reset as well, which is 80 km/h for RVs and 90 km/h for CAVs.

2.2. Simulation Scenarios. A simplified freeway is simulated, which contains a 6-km three-lane mainline in one travel direction, a one-lane on-ramp, and a one-lane off-ramp, as shown in Figure 1. The simulation duration is 15300 s with a 900 s warm-up period. Data collected from 900 s to 15300 s are used for analysis.

To analyse the impact from CAV penetration rates, this study proposes six scenarios with different compositions of RVs and CAVs, as shown in Table 3. To indicate a traffic condition with a low proportion of CAVs, the largest ratio of CAVs in mixed traffic is set as 10%. In each scenario, the mixed traffic is loaded on mainline and on-ramp, which is varying over time, as shown in Table 4. The input traffic is set to approach the designed freeway lane capacity from the simulation time 8100 s and last to 10700 s. Besides, in all scenarios, 15% of mainline traffic is assigned to leave the freeway at off-ramp.

3. Discussions on Mixed Traffic

The 6 km mainline is divided by 500 m into 12 links. Those links are then labelled from Link 1 to Link 12 same as the travel direction, as shown in Figure 1. Data are integrated by the time interval of one minute. The average ground-truth link speeds could be calculated by the ratio of the link length to the average travel time of all vehicles. The average speed of CAVs on a link during a time interval is calculated using the position and the timestamps of CAVs. Here are some statistical findings about the simulated mixed traffic.

3.1. Sample Size and Data-Missing Rate. The boxplot in Figure 2 shows the distribution of the sample size per minute under different CAV penetrations. The median sample size per minute under the penetration rate of 1%, 3%, 5%, 7%, and 10% is 1, 2, 4, 5, and 7, respectively. When the penetration rate is 1%, sample size per minute would mostly appear as a number within [1, 2]12. Similarly, it can be seen that the most frequent sample size for the penetration rate of 3%, 5%, 7%, and 10% is [1, 3], [2, 5], [3, 7], and [5, 10], respectively. Besides, it seems that the variation of the sample size adds with the increase of penetration rate.

Except for the sample size, another very concerned issue in the discussion of traffic probe with low penetration rate is the missing data rate. This study defines the data-missing rate on a link as the rate between the number of time intervals that have collected CAV data and the total number of time intervals. Figure 3 presents the data-missing rate on different links and under different CAV penetrations.

Figure 3 shows that if the CAV penetration rate is small, the sample size is really small and there will be a serious data loss. Especially, when the proportion of CAVs is 1%, the data-missing rate almost reaches fifty percent. It requires that the estimation method is capable of filling the missing parts.

3.2. Speed Difference. Afterwards, this study looks into the difference between the average speeds of CAVs and the mixed traffic speeds on a link. This difference is calculated by the following equation:

$$d_{ij} = v_{cij} - v_{ij}, \qquad (1)$$

where d_{ij} is the difference between the average speed of CAVs and the average link speed at the *i*th time interval and *j*th link, v_{cij} is the average speed of CAVs at the *i*th time interval and *j*th link, and v_{ij} is the average link speed at the *i*th time interval and *j*th link.

Table 5 summarizes the speed differences and their variances. The desired speed of CAV is higher than that of RV, so it could be referred that the average speeds of CAVs would be most likely higher than the average link speeds. It is proved by the average and median speed differences in Table 5. The maximum and minimum differences indicate that the average speeds of CAV might also overestimate and underestimate the link speeds. It would be vital to establish a

TABLE 1: Existing studies on CAV simulation by the internal model.

Parameters	Default value	Atkins [29]	Aria et al. [30]	Tibljaš et al. [31]	Stanek et al. [32]	Sukennik [28]	Asadi et al. [33]
Version of VISSIM	VISSIM 9	VISSIM 8	DK	DK	VISSIM 9	VISSIM 11	VISSIM 10
CC0, standstill distance (m)	1.50	—	—	1.00	1.25	1.00	1.00–1.50
CC1, headway time (gap between vehicles) (s)	0.90	—	0.30	0.50	0.25	0.60	0.50–1.50
CC2, car-following distance/following variation (m)	4	0	—	1.00	3	0	0
CC3, threshold for entering following (s)	−8	—	—	—	−12	−6	−8
CC4, negative following threshold (m/s)	−0.35	0	−0.56	−0.10	—	−0.10	0
CC5, positive following threshold (m/s)	0.35	0	0.56	—	—	0.10	0
CC6, speed dependency of oscillation (1/(m/s))	11.44	0	—	0	0	0	0
CC7, oscillation during acceleration (m/s^2)	0.25	—	—	0.40	—	0.10	0.15–0.45
CC8, standstill acceleration (m/s^2)	3.50	—	—	4	—	4	3.30–3.90
CC9, acceleration at 80 km/h (m/s^2)	1.50	—	—	2	—	2	1.30–1.90
Look-ahead distance	0 to 250 m	—	150–200 m	0–105 m	0–500 m	0–250 m	0–800 m
Look-back distance	0 to 150 m	—	150–200 m	0–280 m	0–250 m	0–150 m	0–800 m
Observed vehicles	2	10	6–8	—	10	99	10
Smooth close-up behaviour	No	—	—	Yes	—	—	Yes
Accepted deceleration, trailing vehicle (m/s^2)	−1	—	—	—	—	−1.50	—
Minimum headway, front/rear (m)	0.50	—	—	—	0.40	0.50	0.20–0.70
Safety distance reduction factor	0.60	—	—	—	0.45	0.75	0.30–0.80
Maximum deceleration for cooperative braking (m/s^2)	−3	—	—	—	−4	−6	—
Cooperative lane change	No	—	—	—	Yes	Yes	Yes
Maximum speed difference (km/h)	3	—	3	—	—	10.80	10.80

right relationship model between speeds of CAVs and link speeds. Accordingly, the variance of speed differences is calculated as shown in Table 5, which could be applied to calibrate the relationship model. The variance shows the expectation of the squared deviation of the speed difference from its mean difference, and Table 5 indicates the deviation decreases with the increase in the CAV penetration rate.

3.3. Fundamental Diagram of Mixed Flow.

This section aims to discuss the impact of CAVs on the fundamental diagram. Taking Link 7 as an example, Figure 4 shows the speed-flow diagrams under different penetration rates of CAVs. It seems that an increase in the CAV penetration rate has a slight impact on the shape of the speed-flow fundamental diagram. The largest traffic volume (approaching the link capacity) is 7860, 7620, 7980, 7560, 7920, and 8040, respectively, when CAVs account for 0%, 1%, 3%, 5%, 7%, and 10%. It indicates that an increase in the CAV penetration rate would not definitely contribute to the increase in traffic flux, when the penetration rate is under 10%. Moreover, when the penetration rate increases, the number of the scatter dots on the left side reduces. To some extent, it indicates that the increase of CAVs in the mixed flow could relief the traffic congestion.

Besides, the critical speed to identify the free-flow state seems to remain the same at 80 km/h, as shown in Figure 4. Since the lowly penetrated CAVs do not have a significant impact on the critical speed and volume, this study would assume that the traditional estimation method (i.e., Kalman filtering-based estimation method) might be effective when the proportion of CAV in the mixed flow is low.

4. Kalman Filtering-Based Estimation Method for Mixed Traffic

4.1. Basic Kalman Filtering Algorithm.

The traditional Kalman filtering-based estimation method is applied. For application in this study, the discrete form of the KF in the linear speed model is given by

$$x_t = A_t x_{t-1} + \omega_t, \qquad (2)$$

$$y_t = H_t x_t + v_t, \qquad (3)$$

where x_t is the average link speed at the tth time interval. For simplicity, it is originally assumed that it has a linear relationship with the speed value at the previous time interval. y_t is the collected speed which is the average speed of CAVs at the tth time interval. Similarly, the CAV speed is supposed

TABLE 2: Car-following and lane-change parameters of RVs and CAVs in this study.

| Parameters	Default value for RVs	Adjusted value for CAVs
CC0, standstill distance (m)	1.5	1.25
CC1, headway time (gap between vehicles) (s)	0.9	0.5
CC2, car-following distance/following variation (m)	4	3
CC3, threshold for entering following (s)	−8	−12
CC4, negative following threshold (m/s)	−0.35	−0.1
CC5, positive following threshold (m/s)	0.35	0.1
CC6, speed dependency of oscillation (1/(m/s))	11.44	0
CC7, oscillation during acceleration (m/s^2)	0.25	0.25
CC8, standstill acceleration (m/s^2)	3.5	3.5
CC9, acceleration at 50 miles per hour (m/s^2)	1.5	1.5
Look-ahead distance	0 to 250 m	0–500 m
Look-back distance	0 to 150 m	0–500 m
Observed vehicles	2	10
Smooth close-up behaviour	Checked	Checked
General behaviour	Free lane selection	Free lane selection
Maximum deceleration, own vehicle (m/s^2)	−4	−4
Maximum deceleration, trailing vehicle (m/s^2)	−3	−3
−1 m/s^2 per distance, own vehicle and trailing vehicle (m)	100	100
Accepted deceleration, own vehicle (m/s^2)	−1	−1
Accepted deceleration, trailing vehicle (m/s^2)	−1	−1
Waiting time before diffusion (s)	60.00	60.00
Minimum headway, front/rear (m)	0.5	0.37
Safety distance reduction factor	0.6	0.45
Maximum deceleration for cooperative braking (m/s^2)	−3	−4
Cooperative lane change	Not checked	Checked
Maximum speed difference (km/h)	3	3
Maximum collision time (s)	10	10
Collision time gain (s)	2	2
Minimum longitudinal speed (km/h)	1	1
Time before direction changes (s)	0	0
Overtake same lane vehicle, minimum lateral distance standing (m) at 0 km/h	1	0.75
Overtake same lane vehicle, minimum lateral distance driving (m) at 50 km/h	1	0.75
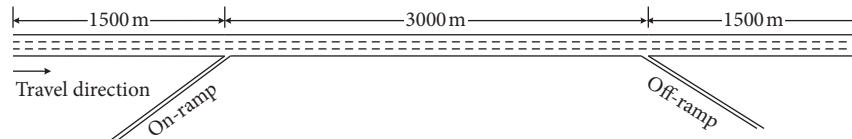
FIGURE 1: Simulated freeway network diagram.

TABLE 3: Simulation scenarios of traffic composition.

Scenarios	Proportion of RVs (%)	Proportion of CAVs (%)
1	100	0
2	99	1
3	97	3
4	95	5
5	93	7
6	90	10

to have a linear relationship with the average link speed. A_t and H_t are the linear coefficients. ω_t and v_t represent the state and measurement noises, respectively. Usually, $E(\omega_t) = 0$, $E(v_t) = 0$, $E(\omega_s, \omega_t') = Q_t$, $E(v_s, v_t') = R_t$, and $E(\omega_s, v_t') = 0$. The state equation (2) shows the behaviour of an n-dimensional state vector x_t, and the measurement

equation (3) describes how the state vector is related to an m-dimensional measurement vector y_t. Obviously, in this study, m and n are mostly not equal. Especially, when CAVs account for 1%, m is far less than n. In the presence of incomplete data, the following recursive formula is used to solve the previous discrete model in this study:

TABLE 4: Traffic input in a scenario.

Simulation time interval (s)	Mainline traffic input (three-lane) (veh)	On-ramp traffic input (one-lane) (veh)
0~2700	3000	300
2700~4500	3600	300
4500~6300	4500	300
6300~8100	5100	300
8100~9900	6000	300
9900~11700	6300	300
11700~12600	5100	300
12600~13500	4500	300
13500~14400	3600	300
14400~15300	3000	300

○ Number of CAVs under 1% penetration

□ Number of CAVs under 3% penetration

◇ Number of CAVs under 5% penetration

▽ Number of CAVs under 7% penetration

☆ Number of CAVs under 10% penetration
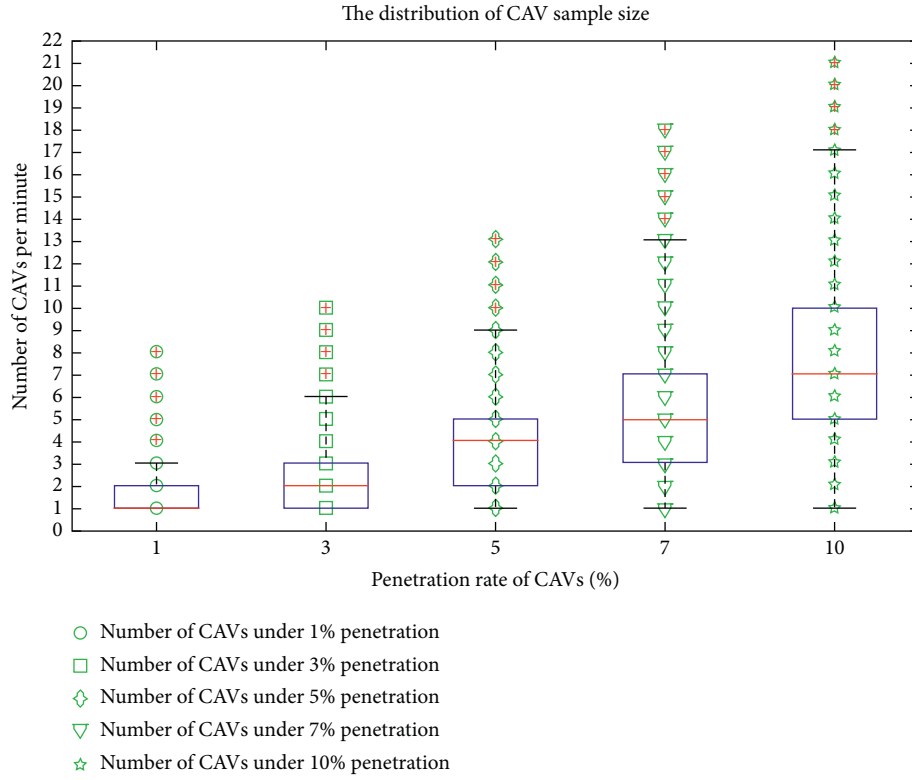
FIGURE 2: Distribution of sample size under different CAV penetration rates.

$$
\widehat{x}_t^t = \begin{cases} \widehat{x}_t^{t-1} + \dfrac{P_t^{t-1} \cdot H_t'}{H_t \cdot P_t^{t-1} \cdot H_t' + R_t}\left(y_t - H_t \cdot \widehat{x}_t^{t-1}\right), & \text{if } y_t \text{ is collected,} \\[2ex] \widehat{x}_t^{t-1}, & \text{if } y_t \text{ is missing,} \end{cases}
$$
$$
P_t^t = \begin{cases} P_t^{t-1}\left(1 - \dfrac{P_t^{t-1} \cdot H_t' \cdot H_t}{H_t \cdot P_t^{t-1} \cdot H_t' + R_t}\right), & \text{if } y_t \text{ is collected,} \\[2ex] P_t^{t-1}, & \text{if } y_t \text{ is missing,} \end{cases}
$$

(4)

where P_t^t is the error covariance matrix of state x_t and $P_t^{t-1} = A_t \cdot P_{t-1}^{t-1} \cdot A_t' + Q_t$. Usually, A_t, H_t, Q_t, and R_t are calibrated using the historical data. According to the small average speed difference in Table 5, this study sets both A_t and H_t as 1. As for the state noise Q_t, this study calibrates it separately based on the traffic condition. With a mixed traffic of 1% CAV penetration rate as an

example shown in Figure 5, it is obvious that the variation of state error enlarges when speed falls below the critical speed of 80 km/h. From the observations among all fundamental diagrams across all penetration rates, the application of CAVs does not have a significant impact on the critical speeds when CAVs have a low penetration rate. Therefore, the same critical speed is used to identify
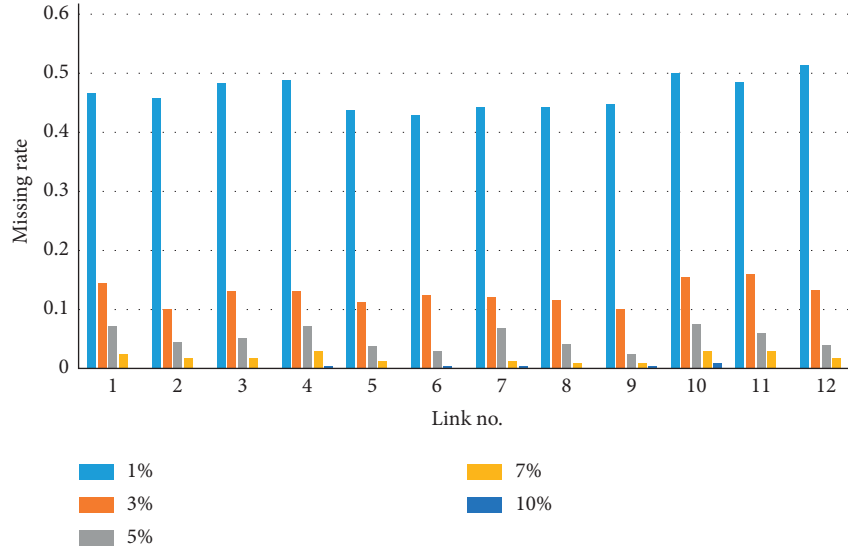
FIGURE 3: Data-missing rate on different links.

TABLE 5: Speed difference.

Speed difference	CAV penetration rate (%)	Link no.											
		1	2	3	4	5	6	7	8	9	10	11	12
Average difference (km/h)	1	3	1	1	2	2	1	1	1	2	3	2	2
	3	3	2	2	2	2	2	1	1	1	3	3	3
	5	3	2	2	2	2	2	2	2	2	4	3	3
	7	3	2	2	2	2	2	2	2	2	4	3	3
	10	3	2	3	2	2	2	2	2	2	3	3	3
Median difference (km/h)	1	3	1	1	1	2	1	1	1	2	2	1	1
	3	3	1	1	1	1	1	1	1	1	3	2	2
	5	3	2	2	2	1	1	1	1	1	4	3	3
	7	3	2	2	2	2	1	1	1	1	3	3	3
	10	3	2	2	2	2	2	2	1	2	3	3	3
Maximum difference (km/h)	1	21	32	12	17	29	19	19	20	22	18	14	14
	3	16	16	15	22	27	19	16	27	23	23	27	18
	5	14	11	16	16	13	22	20	22	22	21	17	14
	7	14	12	15	16	17	21	19	19	16	18	16	15
	10	15	17	14	13	26	15	16	18	22	15	23	14
Minimum difference (km/h)	1	−10	−24	−13	−14	−8	−9	−20	−13	−12	−8	−7	−6
	3	−9	−6	−8	−12	−9	−9	−10	−15	−39	−9	−4	−6
	5	−3	−5	−8	−8	−4	−9	−12	−15	−12	−12	−9	−5
	7	−6	−5	−4	−8	−4	−5	−8	−21	−20	−4	−6	−3
	10	−3	−5	−5	−6	−6	−14	−10	−5	−18	−7	−4	−4
Variance (km/h)^2	1	23	32	20	30	27	20	26	27	33	22	19	17
	3	14	13	14	27	18	16	19	24	41	18	22	19
	5	9	8	11	14	11	13	14	18	20	16	14	13
	7	8	7	7	12	9	10	12	16	19	12	10	10
	10	6	6	6	8	8	8	8	8	17	8	9	10

the traffic state. The state noise Q_t is calibrated under free-flow condition and non-free-flow condition separately. The measurement noise R_t could be calculated using the variance of difference between CAV speeds and ground-truth speeds, as shown in Table 5. Finally, the initial state values are set as $x_0 = 85$ and $P_0 = Q_t$ under the free-flow condition.

4.2. Estimation Results. Using the proposed KF-based estimation method, the speed estimates are obtained. Taking the scenarios of 1% and 10% as an example, Figure 6 exhibits the ground-truth link average speeds, the average speeds of CAVs, and the estimated speeds on Link 7. First, the results indicate that the estimation method interpolates the missing parts of CAV data. Second, the estimation method smooths and
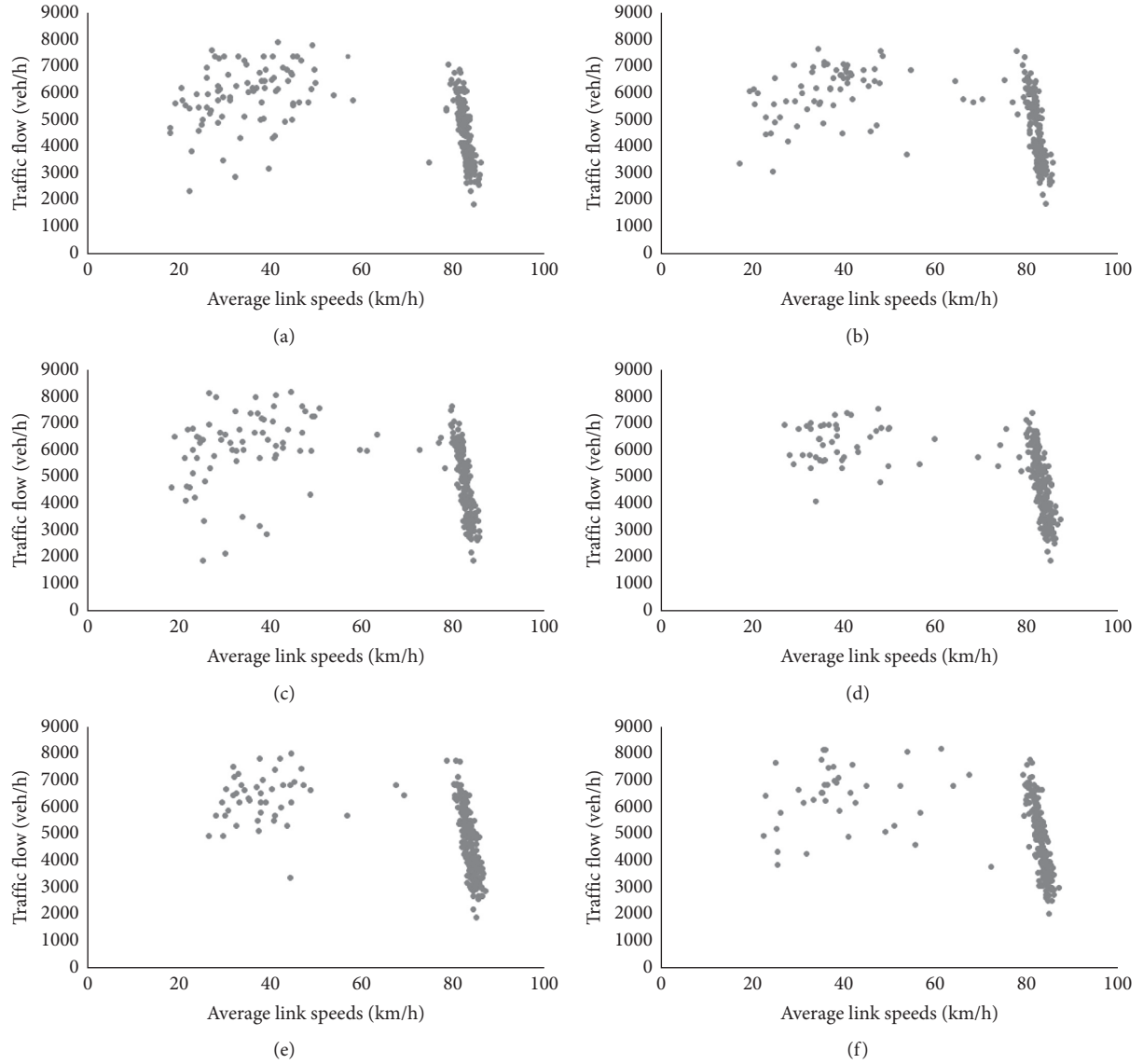
Figure 4: Speed-flow diagrams on Link 7. (a) No CAVs. (b) CAVs account for 1%. (c) CAVs account for 3%. (d) CAVs account for 5%. (e) CAVs account for 7%. (f) CAVs account for 10%.
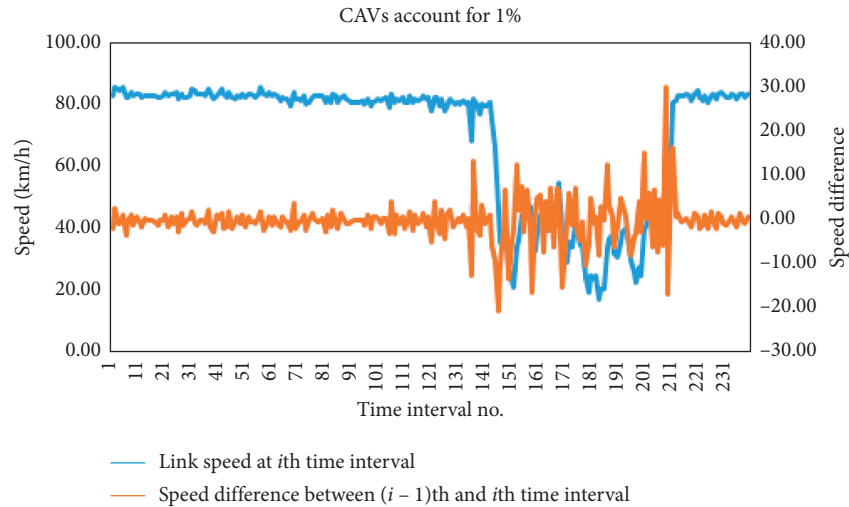

Figure 5: An example for speed difference between $(i-1)$th and ith time interval.
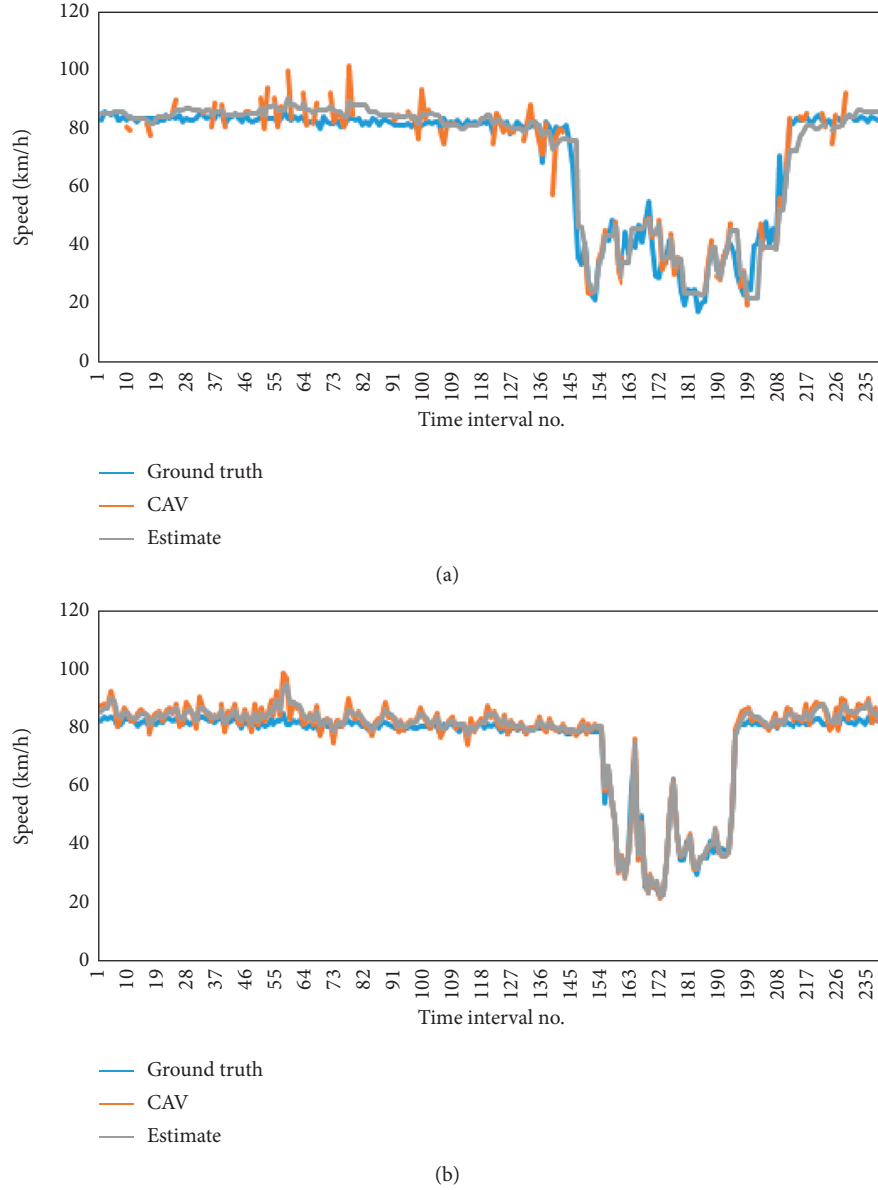
(a)

(b)

FIGURE 6: Ground-truth speeds, CAV speeds, and estimated speeds on Link 7. (a) CAVs account for 1%. (b) CAVs account for 10%.

modifies the CAV speeds, and thus, the estimated speeds are closer to the ground truth.

Furtherly, the ground-truth speeds, estimated speeds, and CAV speeds are illustrated in the time-space form, as shown in Figure 7. According to the speed values, traffic state is divided into three conditions which are represented by three different colors, i.e., green, yellow, and red. Figure 7 shows that the estimates (i.e., Figures 7(a) and 7(d)) almost copy the ground-truths (i.e., Figures 7(b) and 7(e)). If not for the data missing, CAV speeds could almost tell the traffic condition, as shown in Figures 7(c) and 7(f). Especially, 10% CAVs (i.e., Figure 7(f)) seem to be able to visualize the traffic state in a rough three-color map compared with the ground-truth speed map (i.e., Figure 7(e)).

4.3. Accuracy. This section will further evaluate the estimation accuracy. It is measured by RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error). They can be obtained by

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (x_i - \widehat{x}_i)^2},$$

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^{m} |x_i - \widehat{x}_i|. \tag{5}$$

Figure 8 presents the RMSE and MAE of estimated speeds at each penetration rate. Comparatively, the scenario
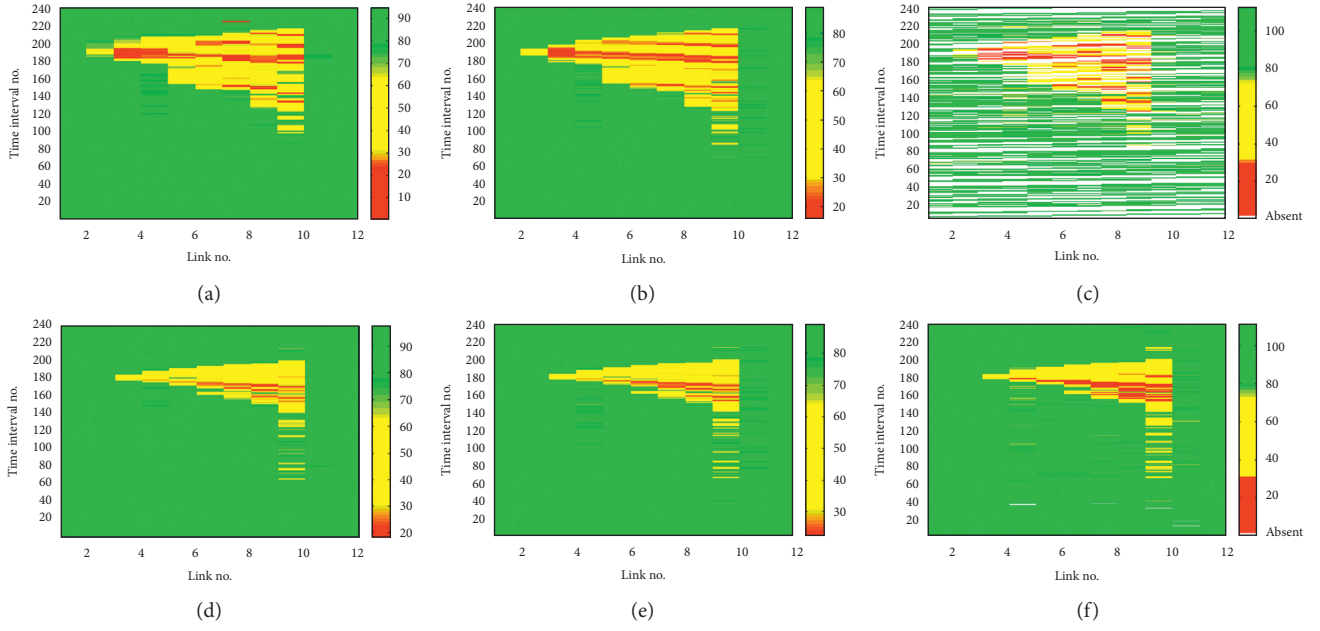
FIGURE 7: Estimated speeds, the ground-truth speeds, and the CAV speeds in the color map. (a) Estimated speeds with 1% penetration rate of CAVs (km/h). (b) Ground-truth speed when CAVs account for 1% (km/h). (c) CAV speeds with 1% CAVs in mixed traffic (km/h). (d) Estimated speeds with 10% penetration rate of CAVs (km/h). (e) Ground-truth speeds when CAVs account for 10% (km/h). (f) CAV speeds with 10% CAVs in mixed traffic (km/h).
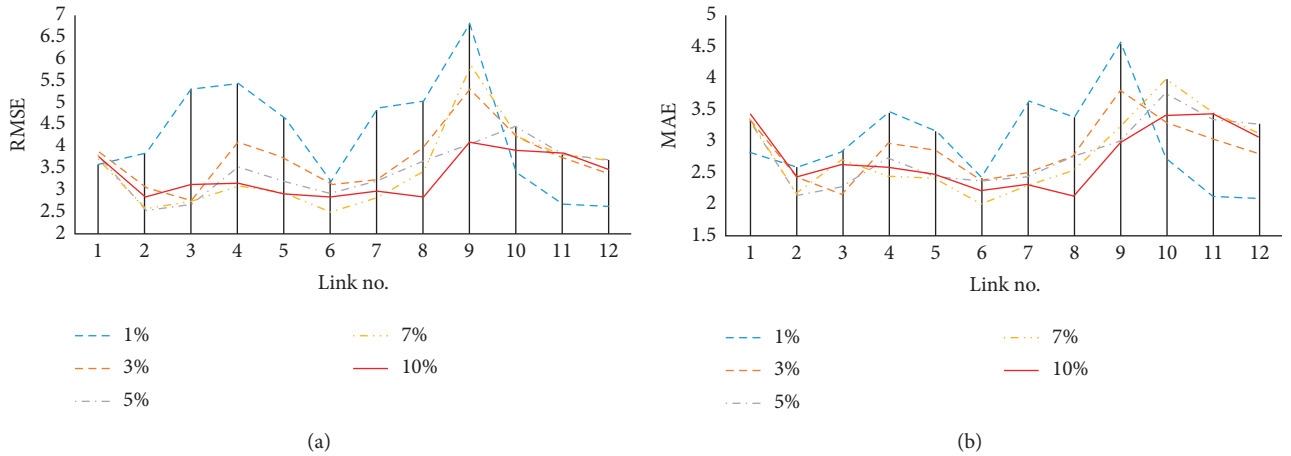

FIGURE 8: (a) RMSE and (b) MAE of estimated speeds.

with 10% CAVs has smaller RMSE and MAE than other penetration scenarios. However, RMSE and MAE of the 3%, 5%, and 7% scenarios are quite approaching those of the 10% scenario. Although RMSE and MAE of the 1% scenario are a bit larger than other scenarios, their values remain within a small value, i.e., RMSE is less than 7 and MAE is less than 5. In general, the estimation method with limited CAV data has a reasonable performance, even when the proportion of CAVs in mixed traffic is only 1%.

Moreover, this study would like to compare the accuracy of estimates and CAV speeds. Since there are missing parts in CAV speeds, RMSE and MAE are calculated using the

data that eliminates the data-missing time intervals. Taking the data from Link 7 as an example, the accuracy comparison results are shown in Figure 9. It is obvious that estimates reduce the speed error compared with the CAV speeds.

4.4. Sensitivity Analysis. In the application of this KF-based estimation method, some parameters might play an important role in the estimation accuracy. They are the state and measurement noises. As mentioned in the KF-based estimation method, the measurement noise is calibrated by the historical data of the CAV speeds and the ground-truth speeds on each link. In practice, the ground-truth speeds are
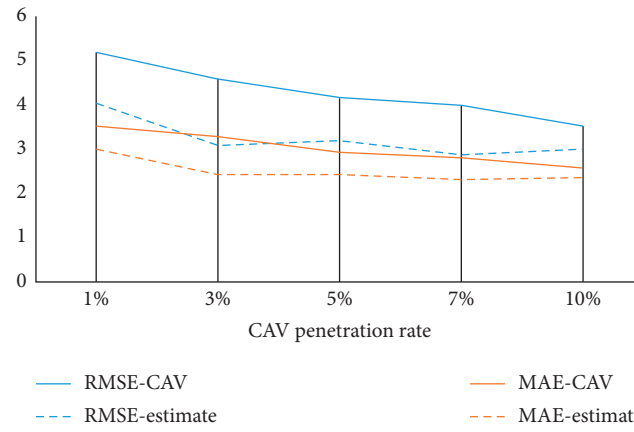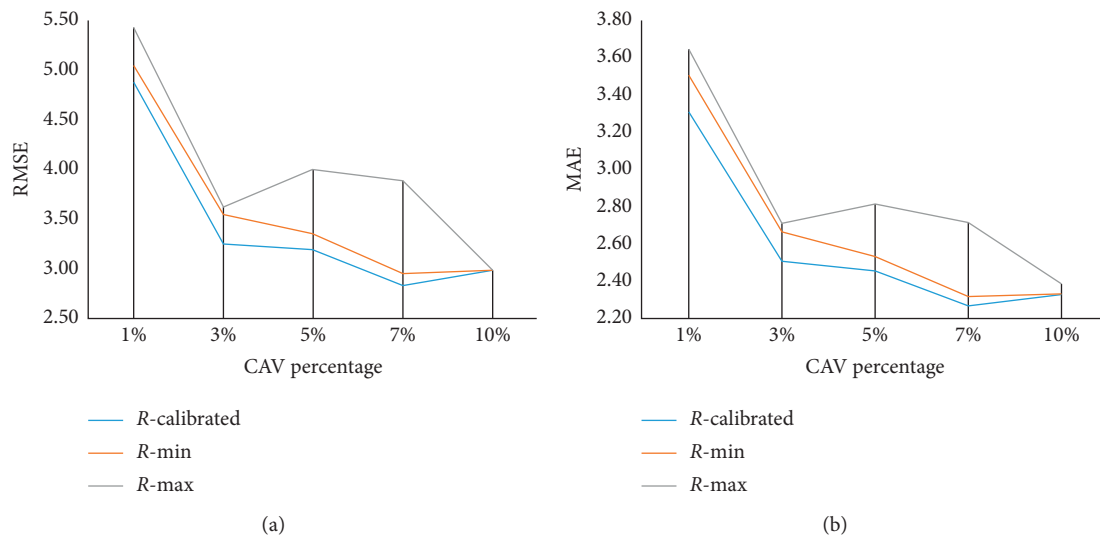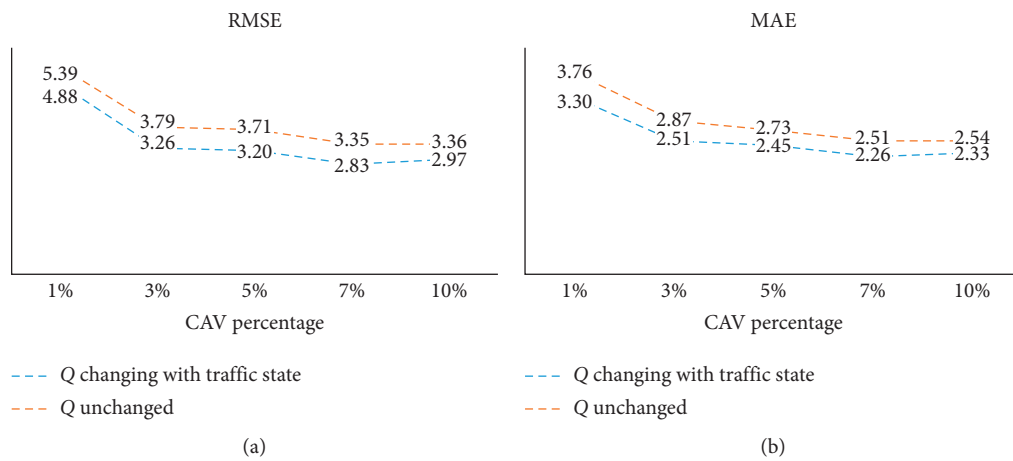
FIGURE 9: Accuracy comparison between estimated speeds and CAV speeds.

FIGURE 10: The impact of measurement error (R) on accuracy. (a) Impact on RMSE. (b) Impact on MAE.

FIGURE 11: The impact of state error (Q) on accuracy. (a) Impact on RMSE. (b) Impact on MAE.

usually not available on all links. Therefore, this study selects out the minimum and maximum R_t from Table 5 and uses these values in the estimation. Their estimation accuracy is compared with those using the calibrated R_t, as shown in Figure 10. Obviously, the proposed method is the optimal, but if R_t could not be calibrated on each link, a small value of R_t calibrated on other links is suggested.

Another parameter is the state noise Q_t. In the proposed method, both the calibration and application of the state noise would be separated based on the traffic condition. If this separation is eliminated, this study finds that it will lead to larger estimation errors as the comparison results in Figure 11. It indicates the outperformance of the proposed method.

5. Conclusions and Future Works

It seems to be inevitable that CAVs will come into the market and travel on the regular roads in the near future. It also could be imaged that there will be mixed traffic consisting of CAVs and RVs, and the proportion of CAVs will be low at the beginning stage. This study discussed the application of the limited CAV data to estimate traffic state at this beginning stage. At first, this study set up a microsimulation platform of the mixed traffic flow using the VISSIM. Five simulation scenarios with the CAV penetration increasing from 1% to 10% were set to generate the testing data. Then, a step-by-step discussion on the characteristics of mixed traffic was conducted based on the simulation data. The sample size distribution under different CAV penetrations was found, and the data-missing rate was calculated which was especially large when CAVs only account for 1% of mixed traffic. The analysis on the speed difference between CAV speeds and the ground-truth link speeds was an assistant in the following calibration of the proposed estimation model . The speed-flow diagrams of mixed traffic indicated the possibility of applying the traditional estimation method. Accordingly, the simple KF-based estimation method was used and adjusted to adapt the incomplete CAV data. The estimation results, accuracy evaluations, and sensitivity analysis validated the applicability and precision of the proposed estimation method using limited CAV data.

Since the Level 4 and Level 5 CAVs are not ready to enter the market, the simulation method is an alternative way to make these investigations. With the developing technology of CAVs, the driving behaviour model of CAVs needs to be updated accordingly in the future. Besides, the simulated roadway is oversimplified. The complex merging and weaving area from on/off-ramp to the mainline could be discussed in detail. Moreover, the fusion of the measurements from other existing sensors, such as loop detectors and GPS probe vehicles, and the testing of other existing estimation methods are also useful for the field application.

Data Availability

The data used to support the findings of this study were generated from VISSIM simulation software based on the simulation settings described within the article.

Conflicts of Interest

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

Acknowledgments

References

[1] R. Hoogendoorn, B. V. Arem, and S. Hoogendoorn, "Automated driving, traffic flow efficiency, and human factors: literature review," *Journal of the Transportation Research Board*, vol. 2422, no. 1, pp. 113–120, 2014.

[2] S. Childress, B. Nichols, B. Charlton, and S. Coe, "Using an activity-based model to explore the potential impacts of automated vehicles," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2493, no. 1, pp. 99–106, 2015.

[3] D. Milakis, B. V. Arem, and B. V. Wee, "Policy and society related implications of automated driving: a review of literature and directions for future research," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 21, no. 4, pp. 324–348, 2015.

[4] P. Bansal and K. M. Kockelman, "Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies," *Transportation Research Part A: Policy and Practice*, vol. 95, pp. 49–63, 2017.

[5] T. Litman, "Autonomous vehicle implementation predictions: implications for transport planning," in *Proceedings of the Transportation Research Board 94th Annual Meeting*, Washington, DC, USA, January 2015.

[6] L. C. Davis, "Effect of adaptive cruise control systems on traffic flow," *Physical Review E*, vol. 69, no. 6, Article ID 066110, 2004.

[7] S. Shladover, D. Su, and X.-Y. Lu, "Impacts of cooperative adaptive cruise control on freeway traffic flow," *Journal of the Transportation Research Board*, vol. 2324, pp. 63–70, 2012.

[8] B. Friedrich, "The effect of autonomous vehicles on traffic," in *Autonomous Driving*, Springer, Berlin, Germany, 2016.

[9] M. Zhou, X. Qu, and S. Jin, "On the impact of cooperative autonomous vehicles in improving freeway merging: a modified intelligent driver model-based approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1422–1428, 2017.

[10] L. Xiao, M. Wang, W. Schakel, and B. van Arem, "Unravelling effects of cooperative adaptive cruise control deactivation on traffic flow characteristics at merging bottlenecks," *Transportation Research Part C: Emerging Technologies*, vol. 96, pp. 380–397, 2018.

[11] L. D. Baskar, B. D. Schutter, and J. Hellendoorn, "Optimal routing for intelligent vehicle highway systems using a macroscopic traffic flow model," in *Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems*, pp. 576–581, St. Louis, MO, USA, October 2009.

[12] Y. Liu, J. Guo, J. Taplin, and Y. Wang, "Characteristic analysis of mixed traffic flow of regular and autonomous vehicles using cellular automata," *Journal of Advanced Transportation*, vol. 2017, Article ID 8142074, 10 pages, 2017.

[13] C. Lu and A. Aakre, "A new adaptive cruise control strategy and its stabilization effect on traffic flow," *European Transport Research Review*, vol. 10, no. 49, pp. 1–11, 2018.

[14] M. M. Morando, Q. Tian, L. T. Truong, and H. L. Vu, "Studying the safety impact of autonomous vehicles using simulation-based surrogate safety measures," *Journal of Advanced Transportation*, vol. 2018, Article ID 6135183, 11 pages, 2018.

[15] C. Lu, J. Dong, L. Hu, and C. Liu, "An ecological adaptive cruise control for mixed traffic and its stabilization effect," *IEEE Access*, vol. 7, pp. 81246–81256, 2019.

[16] C. Lu, J. Dong, and L. Hu, "Energy-efficient adaptive cruise control for electric connected and autonomous vehicles," *IEEE Intelligent Transportation Systems Magazine*, vol. 11, no. 3, pp. 42–55, 2019.

[17] 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.

[18] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobson, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile century field experiment," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568–583, 2010.

[19] S. He, Y. Cheng, F. Ding, J. Zhang, and B. Ran, "Extended Kalman filter-based freeway traffic state estimation using cellphone activity data," in *Proceedings of the Transportation Research Board 95th Annual Meeting*, Washington DC, USA, January 2016.

[20] S. He, J. Zhang, Y. Cheng, X. Wan, and B. Ran, "Freeway multisensor data fusion approach integrating data from cellphone probes and fixed sensors," *Journal of Sensors*, vol. 2016, Article ID 7269382, 13 pages, 2016.

[21] S. He, Y. Cheng, G. Zhong, and B. Ran, "A data-driven study on the sample size characteristics of cellular handoff probe system," *Advances in Mechanical Engineering*, vol. 9, no. 4, pp. 1–9, 2017.

[22] M. J. Lighthill and G. B. Whitham, "On kinematic waves II. A theory of traffic flow on long crowded roads," *Proceedings of The Royal Society A Mathematical, Physical and Engineering Sciences*, vol. 229, no. 1178, pp. 317–345, 1955.

[23] H. J. Payne, *Models of Freeway Traffic and Control (Mathematical Models of Public Systems)*, Simulation Councils, Inc., Vista, CA, USA, 1971.

[24] T. Seo, T. Kusakabe, and Y. Asakura, "Traffic state estimation with the advanced probe vehicles using data assimilation," in *Proceedings of the IEEE 18th International Conference on Intelligent Transportation Systems*, IEEE Computer Society, Washington, DC, USA, 2015.

[25] 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.

[26] SAE, "Automated driving: levels of driving automation are defined in new SAE international standard J3016," 2017, https://web.archive.org/web/20170903105244/https://www.sae.org/misc/pdfs/automated_driving.pdf.

[27] B. Cisco, "PTV VISSIM & connected autonomous vehicles," in *Proceedings of the 18th Annual North America PTV Vision Traffic User Group Meeting*, Portland, OR, USA, 2017.

[28] P. Sukennik, *Micro-Simulation Guide for Automated Vehicles*, 2018, https://www.h2020-coexist.eu/wp-content/uploads/2018/11/D2.5-Micro-simulation-guide-for-automated-vehicles.pdf.

[29] Atkins, *Research on the Impacts of Connected and Autonomous Vehicles (CAVs) on Traffic Flow*, Atkins, Epsom, UK, 2016.

[30] 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.

[31] 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, p. 1060, 2018.

[32] D. Stanek, E. Huang, R. T. Milam, and Y. Wang, "Measuring autonomous vehicle impacts on congested networks using simulation," in *Proceedings of the Transportation Research Board 97th Annual Meeting*, Washington, DC, USA, 2018.

[33] F. E. Asadi, A. K. Anwar, and J. C. Miles, "Investigating the potential transportation impacts of connected and autonomous vehicles," in *Proceedings of the 2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE)*, Graz, Austria, November 2019.