WILEY | Hindawi

Research Article

Vehicle Trajectory Reconstruction for Signalized Intersections with Low-Frequency Floating Car Data

Hua Wang ⬤,[1] **Changlong Gu** ⬤,[1] **and Washington Yotto Ochieng**[2]

[1]*School of Transportation Science and Engineering, Harbin Institute of Technology, No. 73 Huanghe Road, Nangang District, Harbin 150090, China*
[2]*Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK*

Correspondence should be addressed to Hua Wang; wanghua@hit.edu.cn

Floating car data are beneficial in estimating traffic conditions in wide areas and are playing an increasing role in traffic surveillance. However, widespread application is limited by low-sample frequency which makes it hard to get a complete picture of a vehicle's motion. An accurate and reliable reconstruction of a vehicle's trajectory could effectively result in a higher sampling frequency enabling a more accurate estimation of road traffic parameters. Existing methods require additional information such as nearby vehicles, signal timing strategies, and queue patterns which are not always available. To address this problem, this paper presents a method used with low-sample frequency data to reconstruct vehicle trajectories through intersections, without the need for extra information. Furthermore, the additional parameters for the speed-time curve distributions for deceleration rate and acceleration rate are generated. A piecewise deceleration and acceleration model is developed to calculate the acceleration rate for different travel modes in the trajectory. The distribution parameters of the acceleration data for each travel mode are then estimated using a new Expectation Maximization (EM) algorithm. The acceleration statistics are then used to reconstruct the corresponding parts of the trajectory. Compared to the reference trajectories (truth), the test results show that the method developed in this paper achieves improvement in accuracy ranging from 16 to 67% over the commonly used linear interpolation method. In addition, the proposed method is not very sensitive to the sampling interval of the floating car data, unlike the linear interpolation method where the error grows rapidly with increasing sampling interval.

1. Introduction

The potential for the contribution of floating car data to traffic state monitoring has attracted significant research in the past decade [1–3]. However, constrained by the penetration rate of equipped vehicles and cost of data storage and transmission, the sampling interval of floating car data can be low, restricting their application. Liu *et al.* evaluated the sensitivity of delay measurement to sampling frequency of floating car data, showing that delays estimated from data sampled at a 10-second interval are consistent with those from a 5-second interval for 74% of the cases. However, when the sampling interval is 60 seconds, the consistency drops to 37% [4]. Patire *et al.* investigated the question of how much GPS data are needed for a traffic information system capable of providing accurate speed (and thus travel time) information. They point out that this question must address issues of data quality in terms of sample rate and penetration rate [5]. Bucknell *et al.* investigated the relationship between penetration rate and sampling rate, concluding that, in general, increasing sample frequency is more beneficial when the current penetration rate is low [6]. High-resolution trajectory data are ideal for traffic state monitoring. However, the majority of trajectory data (especially those for commercial use) today are collected at a relatively low sampling frequency. Hence, it is necessary to develop methods to reconstruct vehicle trajectories from low-frequency floating car data. Researchers have conducted various studies on vehicle trajectory reconstruction, classified into two categories. One category aims at eliminating the trajectory noise caused by GPS positioning error, while the other attempts are to reconstruct the vehicle's trajectory from low-frequency floating car data.

For trajectory error correction, Huang *et al.* propose an approach for predicting vehicle trajectory by using information from nearby vehicles to construct the local driving environment [7]. Fard *et al.* present a simple two-step method based on wavelet analysis for filtering errors and reconstructing trajectories. Wavelet transform is employed to identify and modify the outliers [8]. Marcello *et al.* develop an approach with a multistep procedure based on the information of traffic kinematics and vehicle dynamics [9]. Since high-frequency data cannot be collected at a large scale, it is impossible to provide high-accuracy traffic monitoring of the road network.

To reconstruct the trajectory between spares probes, Sun *et al.* apply the Variation Formulation (VF) method to obtain the complete picture of traffic flow using sample vehicle trajectories. However, to properly apply the VF method, the shockwave boundaries have to be known, which is estimated from the fixed sensors data [10]. Wan *et al.* propose Expectation Maximization (EM) and Maximum Likelihood (ML) methods to reconstruct trajectories between two consecutive transit bus GPS updates. The path is divided into short segments and the segment travel time is allocated iteratively with the EM method. The maximum likelihood trajectory is generated based on travel time statistics [11]. In the method, it is assumed that the traffic signal timing is known. Hao *et al.* propose a model investigating all possible driving mode sequences between data points. Detailed trajectories are reconstructed based on the optimal driving mode sequences [12]. Hao *et al.* develop a modal activity-based stochastic model for reconstructing the trajectories from sparse data [13]. However, in their approach, the distribution parameters of four types of modal activity are estimated with the second-by-second trajectory in the NGSIM Lankershim dataset. The NGSIM data consist of detailed vehicle trajectory and are collected using digital video camera. The precise location of each vehicle on a 0.5 to 1.0 kilometer section of roadway is recorded every one-tenth of a second. The applicability of this method to different locations and traffic conditions is still to be proven. Wang *et al.* proposed an approach for trajectory reconstruction with sparse probe data for estimating control delay. The method describes the deceleration and acceleration as a piecewise constant deceleration process and acceleration process, respectively. This method has the limitation of poor stability [14].

The problem for trajectory reconstruction is that only a few discrete sampling points are available. Furthermore, the correlation between sampling points is weak as a result of low-frequency sampling. Therefore, in the current methods, additional information is required for reconstruction. For example, Huang *et al.* exploit information from proximate vehicles. Wan *et al.* use signal timing strategies and queue patterns to assist in the trajectory reconstruction. In the approach proposed by Hao *et al.*, the parameters for deceleration and acceleration are estimated from second-by-second vehicle trajectory data. However, in reality, the extra data are not always available. Hence, this paper proposes a trajectory reconstruction method for signalized intersections with low-frequency floating car data.

In this paper, it is assumed that the vehicle driving mode through an intersection evolves in a given pattern, that is, cruise (1) – deceleration (2) – idle (3) – acceleration (4) – cruise (5). More precisely, a deceleration (acceleration) rate-based piecewise process is used to model the driving mode. Based on this assumption (i.e., the 5 driving modes), a model is developed to calculate deceleration and acceleration rate from selected historical data. An Expectation Maximization (EM) algorithm is specified to estimate the corresponding distribution parameters. For given low-frequency sample points, when the distribution parameters of deceleration and acceleration are available, the time period and distance for each of the five modes are calculated by solving a Constrained Quadratic Programming (CQP) problem.

In this paper, unlike the current methods, only low-frequency floating car data are used, with no additional or extra information. Furthermore, in addition to position-time curve, speed-time curve, and distribution parameters of acceleration (deceleration) rate are obtained. By reconstructing the trajectory of the low-frequency trajectory, more accurate and valuable information could be extracted from trajectories, enabling floating car data to play a more important role in traffic state monitoring.

The paper is structured as follows. The next section presents the methodology followed by the results and their analysis and conclusions.

2. Methodology

Different from the current methods, the method developed in this paper reconstructs the trajectory in two parts, deceleration process and acceleration process (i.e., not the trajectory between two consecutive updates). Traffic signal timing and historical queue length data are not needed. The main factors that may influence trajectory reconstruction include traffic conditions, number of sample points, and a vehicle's stop position relative to the intersection. To better reconstruct the low-frequency trajectory, the numbers of sample points and stop positions are investigated in the first instance. It should be noted that in our statistics even if there are many sample points with zero speed, the number of points is recorded as 1. For an arterial street with signalized control, we generate the number of sample points and stop positions of floating cars for different periods of the day over a timespan of three weeks. The results are shown in Figure 1.

As shown in Figures 1(a) and 1(b), whether in the peak hour (7:00-9:00 am) or off-peak hour (10:00-12:00 am), the distribution is skewed to the left. This means that, for the floating car through the intersection, there are more cases generating 3, 4, and 5 sample points than 6, 7, and 8 points, with the most likely being 4. Figures 1(c) and 1(d) show the vehicle stop position histogram, exhibiting a bimodal distribution. There is no obvious difference between the peak hour and off-peak hour. However, in both figures, the vehicles rarely stopped between 100 and 300m. The data used for the analysis here were collected by taxis. Hence, the gap in the stop position histograms is attributed to the traffic volume. When the traffic volume was low, the taxi drivers preferred to stop close to the intersection by changing lane. However,

(a) 7:00-9:00 AM

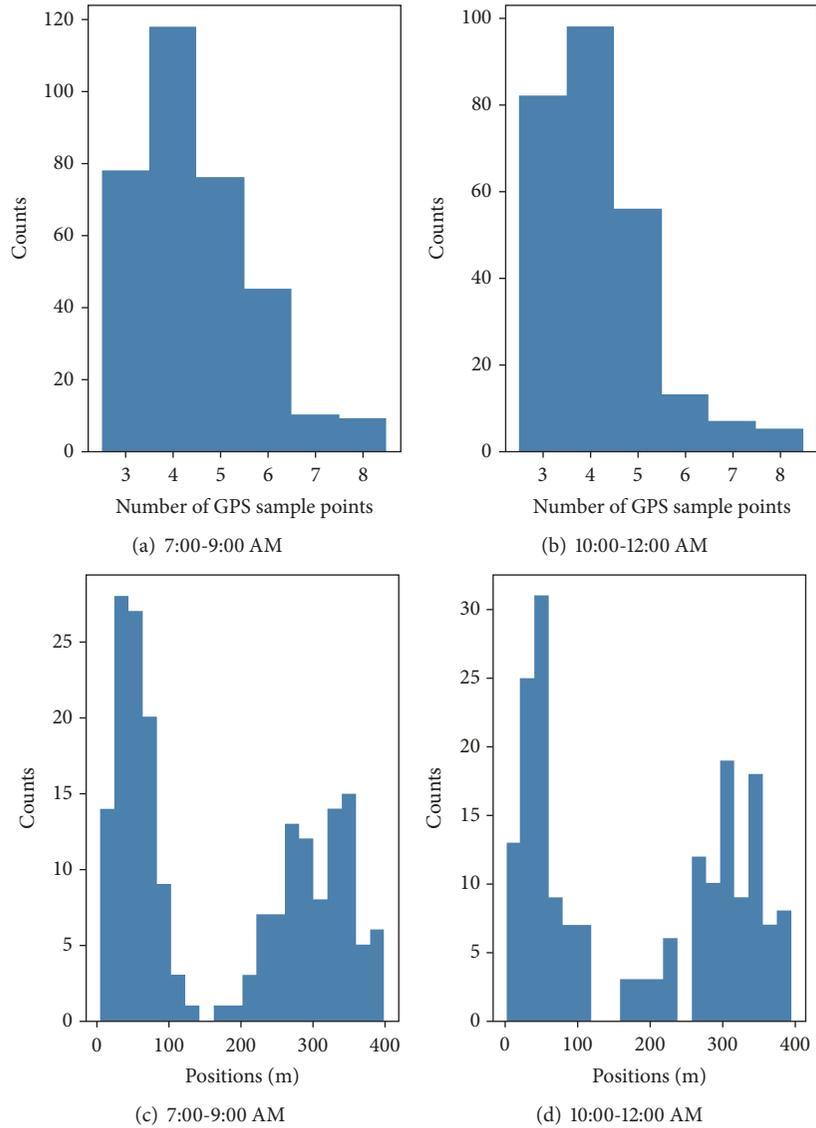(b) 10:00-12:00 AM

(c) 7:00-9:00 AM

(d) 10:00-12:00 AM

FIGURE 1: Histogram of GPS sample points and stop position.

when the traffic volume was high, all the lanes were mostly occupied with other vehicles, so that the taxis had to queue. Hence, in most cases, the taxis stopped relatively far from the intersection. This led to the vehicle stopping either close to or far from the intersection. From the histogram, it could be inferred that the maximum queue length of the intersection was about 400m.

A vehicle's motion is disrupted at the point where the speed is zero. Hence, the trajectory reconstruction is divided into two processes: deceleration and acceleration.

The schematic of this research is shown in Figure 2, for reconstructing the maximum likelihood trajectory from all possible trajectory sets {1, 2, 3,..} with sparse sample points A, B, and C. Figure 3 presents the reconstruction steps, with the main elements of estimation of the distribution parameters for different modes and maximum likelihood trajectory reconstruction.

2.1. Acceleration Rate Estimation for Different Modes. The model for calculating the acceleration and deceleration rates with the low-frequency floating car data originates from previous work by the authors on the estimation of vehicle control delay [13]. The model is developed on the basis of a piecewise constant deceleration model and a Simple Platoon Advancement (SPA) model. As illustrated in Figure 4, deceleration and acceleration are both divided into three parts. For the deceleration process, the first stage represents a vehicle travelling at the free-flow speed. The second and third stages represent vehicle deceleration rates a_1 and a_2 in m/s^2. Similarly, for the acceleration process, the first and second stages are represented by acceleration rates a_3 and a_4. The third stage represents a vehicle travelling at the free-flow speed. Hence, the vehicle driving pattern through the intersection is as follows: cruise (1) – deceleration1 (2) – deceleration2 (3) – idle (4) – acceleration1 (5) – acceleration2 (6) – cruise (7).
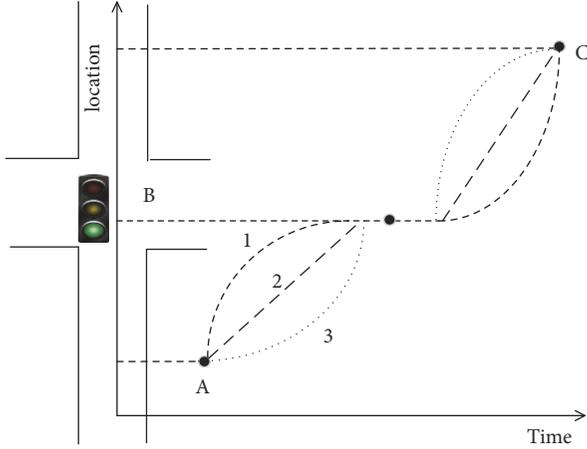
FIGURE 2: Schematic of trajectory reconstruction with low-frequency probe data.

Based on the model and driving pattern for a low-frequency trajectory, deceleration rates a_2 for mode2 (deceleration1) and a_3 for mode3 (deceleration2) and acceleration rates a_5 for mode5 (acceleration1) and a_6 for mode6 (acceleration2) can be calculated by solving a constrained nonlinear programming problem as follows.

For a typical low-frequency trajectory of the probe vehicle shown in Figure 5, the vehicle travels at the free-flow speed at Point 1 and then decelerates between the point in the dotted box and Point 2. At Point 3, the vehicle has stopped and is waiting for the traffic light to turn green. From Point 4 to Point 5, the vehicle accelerates to reach the free-flow speed. Points 1-5 cover the entire motion process of the vehicle through the intersection. The deceleration onset and acceleration end points correspond to the points in the dotted boxes.

The deceleration rates a_2 and a_3 are calculated with the developed model.

Let t_{dy} be the time interval between t_1 and t_1' and let t_{ay} be the time interval between t_4' and t_5. The time and distance between Point 1 and Point 2 are expressed by the following equations:

$$t_{dy} + \frac{v_1 - v_m}{a_2} + \frac{v_m - v_2}{a_3} = T \qquad (1)$$

$$v_1 * t_{dy} + \frac{v_1^2 - v_m^2}{2a_2} + \frac{v_m^2 - v_2^2}{2a_3} = L_{12} \qquad (2)$$

$$0 \le t_{dy} \le 30, \ 0 \le a_2 \le 9, \ 0 \le a_3 \le 9 \qquad (3)$$

where T is the sampling interval of the probe vehicle data and is also the time travelled between Point 1 and Point 2. v_1 and v_2 are the instantaneous velocities at Point 1 and Point 2, respectively. In this paper, the sample interval is 30 seconds. L_{12} is the distance between Point 1 and Point 2; $v_m = (v_1 + v_2)/2$. In (1) and (2), there are three unknown parameters. The only solution cannot be obtained from this nonlinear equation group. Therefore, to obtain a_2 and a_3, expressions

1 and 2 are converted into a nonlinear programming problem with the objective function as follows:

$$z = \min\left[\left(t_{dy} + \frac{v_1 - v_m}{a_2} + \frac{v_m - v_2}{a_3} - 30\right)^2 \right.$$
$$\left. + \left(v_1 * t_{dy} + \frac{v_1^2 - v_m^2}{2a_2} + \frac{v_m^2 - v_2^2}{2a_3} - L_{12}\right)^2\right] \qquad (4)$$

subject to

$$0 \le t_{dy} \le 30,$$
$$0 \le a_2 \le 9, \qquad (5)$$
$$0 \le a_3 \le 9$$

By solving the nonlinear programming problem, the values of a_2 and a_3 could be calculated. For the acceleration model, the acceleration rates of a_5 and a_6 are calculated as follows:

$$z = \min\left[\left(t_{ay} + \frac{v_n - v_4}{a_5} + \frac{v_5 - v_n}{a_6} - 30\right)^2 \right.$$
$$\left. + \left(v_5 * t_{ay} + \frac{v_n^2 - v_4^2}{2a_5} + \frac{v_5^2 - v_n^2}{2a_6} - L_{45}\right)^2\right] \qquad (6)$$

subject to

$$0 \le t_{ay} \le 30,$$
$$0 \le a_5 \le 9, \qquad (7)$$
$$0 \le a_6 \le 9$$

where v_4 and v_5 are the instantaneous speeds at Point 4 and Point 5, respectively, $v_n = (v_4 + v_5)/2$, and L_{45} is the distance between Point 4 and Point 5.

With a_2, a_3, a_5, and a_6, the trajectory can be reconstructed. However, the model only works for the scenarios where there is a sample point before deceleration and in deceleration. For the acceleration process, the requirement is that there is a sample point in acceleration and after acceleration. Nevertheless, probe vehicle updates normally occur at random positions and times, so there may be no sample point in or before deceleration in a trajectory. To address these issues, we propose an approach to reconstruct the maximum likelihood trajectory to alleviate the effect caused by the randomness of sampling. The essential parameters of our method are distribution parameters of deceleration rate and acceleration rate for different modes.

2.2. Distribution Parameters Estimation of Acceleration Rate for Different Modes.
In this section, the valid trajectories are selected from the historical trajectory data. A valid trajectory has a sample point before deceleration and in deceleration and in acceleration and after acceleration. By modeling and solving the nonlinear programming problem (see (4) and (6)), the acceleration rate for modes 2, 3, 5, and 6 for each
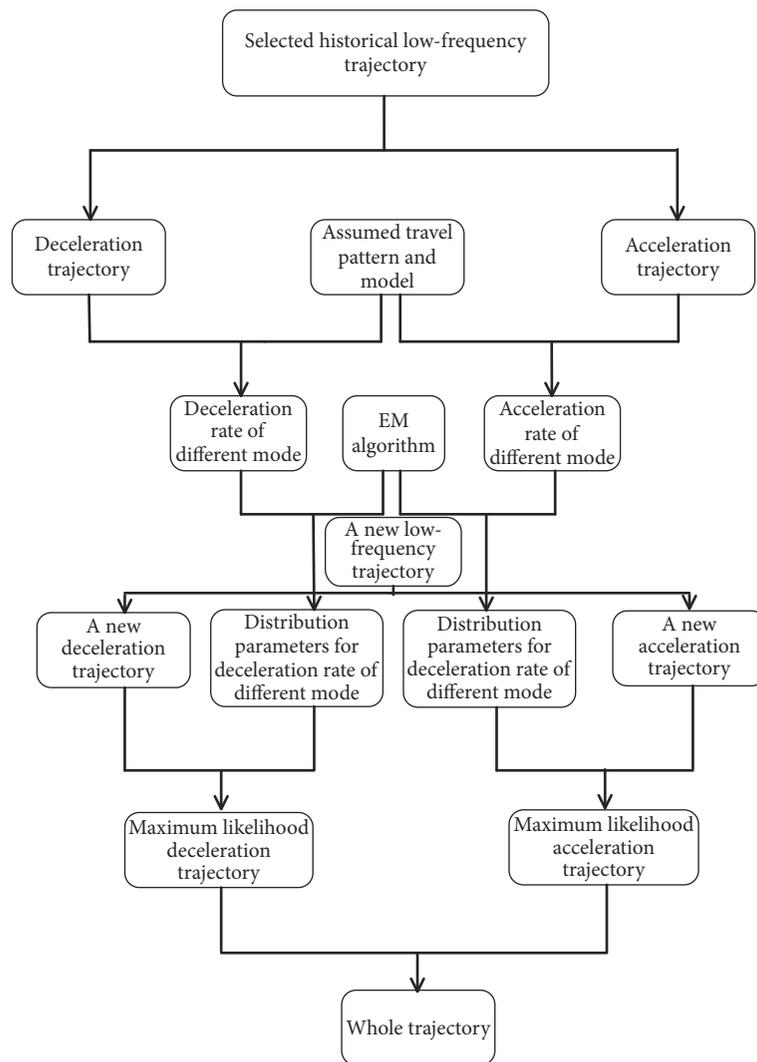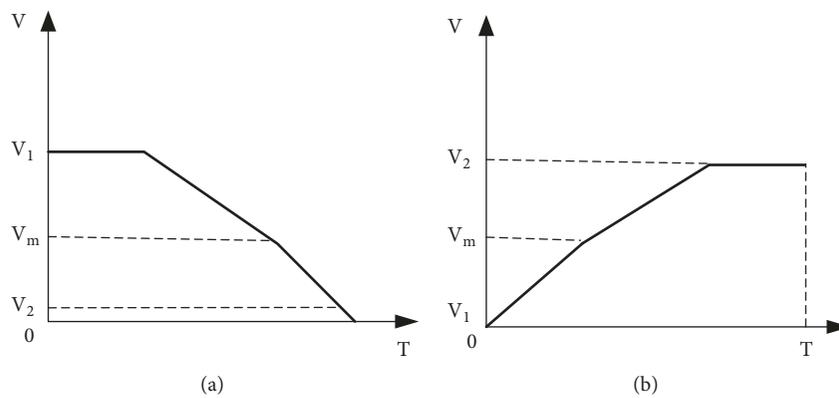
FIGURE 3: Trajectory reconstruction process.

FIGURE 4: Relationship between speed and time in deceleration model (a) and acceleration model (b).
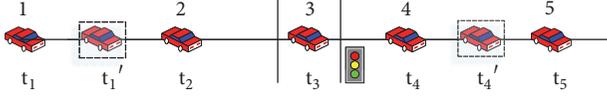
FIGURE 5: Low-frequency vehicle trajectory at intersection.

valid trajectory is obtained. To avoid errors by using these directly to calculate the distribution parameters of acceleration rate for different modes, we adopt an Expectation Maximization method to iteratively calculate the acceleration rate for each mode. It is assumed that the deceleration rate and acceleration rate are independent from each other. For the idle and cruise modes, the mode time mainly depends on the signal plan, traffic condition, and intersection placement. In this paper, it is assumed that the idle and cruise mode travel times t are uniformly distributed.

$$t_i \sim U(b_i, c_i) \tag{8}$$

where $i = 1, 4, 7$ correspond to cruise (1), idle (4), and cruise (7) modes.

For the deceleration and acceleration modes, it is assumed that the acceleration rate follows a Gaussian distribution. The travel time t that a vehicle spends under the deceleration and acceleration modes is the product of speed variation Δv divided by corresponding time period:

$$a_j = \frac{\Delta v_j}{t_j} \sim N(\mu_j, \sigma_j^2) \tag{9}$$

where j = 2,3,5,6 correspond to deceleration1 (mode2), deceleration2 (mode3), acceleration1 (mode5), and acceleration2 (mode6). Each valid trajectory provides an observation of deceleration rate and acceleration rate for different modes. Assume that there are M floating cars passing through the intersection. Suppose that k⩽M is the trajectory sequence number. The accelerate rate for each mode is calculated.

E Step. For mode j (j=2, 3, 5, 6), we calculate the mean μ_j and standard deviation σ_j of the corresponding acceleration. Under the assumption that the acceleration rate follows a Gaussian distribution, the probability density function for a_j^k is determined as

$$p\left(a_j^k \mid \mu_j, \sigma_j\right) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-(a_j^k - \mu_j)^2/2\sigma_j^2} \tag{10}$$

where a_j^k is the acceleration rate for mode i of k*th* trajectory, j=2,3,5,6. For ease of calculation, the log-likelihood function of a_j^k is used.

$$\log\left[p\left(a_j^k \mid \mu_j, \sigma_j\right)\right] = -\frac{\left(a_j^k - \mu_j\right)^2}{2\sigma_j^2} - \log\left(\sqrt{2\pi}\sigma_j\right) \tag{11}$$

M Step. In this step, for each valid trajectory, deceleration rates and acceleration rates for different modes a_j^k are recalculated such that the likelihood function is maximized. For the

deceleration process of each trajectory, its likelihood function can be represented as

$$\arg\max\left(\sum_{i=1, j=2,3} \log\left[p\left(a_j^k \mid \mu_j, \sigma_j\right) + p\left(t_i^k \mid U_i\right)\right]\right)$$

$$= \arg\min\left(\sum_{j=2,3} \frac{\left(a_j^k - \mu_j\right)^2}{2\sigma_j^2}\right) \tag{12}$$

$$= \arg\min\left(\sum_{j=2,3}\left(\frac{a_j^{k2}}{2\sigma_j^2} - \frac{a_j^k\mu_j}{\sigma_j^2}\right)\right)$$

subject to the equality

$$t_1^k + \sum_{j=2,3} \frac{\Delta v_j^k}{a_j^k} = T \tag{13}$$

$$t_1^k * v_1^k + \frac{\left(v_1^k\right)^2 - \left(v_m^k\right)^2}{2 * a_2^k} + \frac{\left(v_m^k\right)^2 - \left(v_2^k\right)^2}{2 * a_3^k} = L_{12}^k \tag{14}$$

where t_1^k is travel time of mode1 for the k*th* trajectory, v_1^k and v_2^k are the speeds at the first sample point and second sample point for k*th* trajectory, respectively, $v_m^k = (v_1^k + v_2^k)/2$, $\Delta v_2^k = v_1^k - v_m^k$, $\Delta v_3^k = v_m^k - v_2^k$, and L_{12}^k is the distance between the first sample point and the second sample point. In the M step, μ_j and σ_j are considered to be constants, and $p(t_i^k \mid U_i)$ is fixed. Therefore, the terms containing them can be ignored when maximizing the likelihood function. This is a Constrained Quadratic Programming (CQP) problem. The solution of this CQP problem is the travel time of mode1 and the distribution parameters of acceleration rate for mode 2 and mode 3.

In a similar way, the likelihood function for the acceleration process is formulated as

$$\arg\max\left(\sum_{i=7, j=5,6} \log\left[p\left(a_j^k \mid \mu_j, \sigma_j\right) + p\left(t_i^k \mid U_i\right)\right]\right)$$

$$= \arg\min\left(\sum_{j=5,6}\left(\frac{a_j^{k2}}{2\sigma_j^2} - \frac{a_j^k\mu_j}{\sigma_j^2}\right)\right) \tag{15}$$

subject to the equality

$$t_7^k + \sum_{j=5,6} \frac{\Delta v_j^k}{a_j^k} = T \tag{16}$$

$$t_7^k * v_5^k + \frac{\left(v_5^k\right)^2 - \left(v_n^k\right)^2}{2 * a_6^k} + \frac{\left(v_n^k\right)^2 - \left(v_4^k\right)^2}{2 * a_5^k} = L_{45}^k \tag{17}$$

where t_7^k is the travel time of mode7 for the k*th* trajectory, v_4^k and v_5^k are the speeds at the fourth and the fifth sample points for the k*th* trajectory, $v_n^k = (v_4^k + v_5^k)/2$, $\Delta v_5^k = v_n^k - v_4^k$,

$\Delta v_6^k = v_5^k - v_n^k$, and L_{45}^k is the distance between the fourth and the fifth sample points. It should be noted that the M step is run separately for each trajectory. When the M step ends, the method goes back to the E step to update the mean and variance based on all trajectories. When the difference between consecutive iterations is below a threshold, the iteration process stops.

2.3. Maximum Likelihood Trajectory Estimation.

With the distribution parameters of acceleration for each mode being known, the maximum likelihood trajectory of a given low-frequency trajectory is estimated. As a result of low-frequency sampling, the sample points may be located at any position of the intersection. Besides, the number of sample points in a trajectory is uncertain. To better capture the difference, we define scenarios for the deceleration stage of a low-frequency trajectory each with different numbers and reconstruct the maximum likelihood trajectory for each.

Scenario 1. 1 sample point before deceleration, none in the deceleration stage.

Scenario 2. 1 sample point before deceleration, 1 in deceleration.

Scenario 3. Multiple sample points before deceleration, none in the deceleration stage.

The scenario definitions for the acceleration stage are similar to those for deceleration.

Scenario 1. 1 sample point after acceleration, none in the acceleration stage.

Scenario 2. 1 sample point after acceleration, 1 in acceleration.

Scenario 3. Multiple sample points after acceleration.

Here the sample points refer to the points where the speed is above zero. Different trajectories contain different numbers of sample points and different combinations of scenarios. The trajectory reconstruction is divided into two parts: deceleration process and acceleration process. The reconstruction method is the same for both processes; hence, only the deceleration process reconstruction with different sample points is illustrated here.

(1) Scenario 1 for Deceleration Trajectory. It is assumed that P1 is the sample point where the speed is above zero; P2 is the sample point where the speed is zero. The distribution parameters of deceleration rate for mode 2 and mode 3 are (μ_2, σ_2) and (μ_3, σ_3). In order to reconstruct the maximum likelihood trajectory of deceleration process, the following is formulated and solved:

$$
\begin{aligned}
\arg\max(&\log\left(p\left(a_2 \mid \mu_2, \sigma_2\right)\right) \\
&+ \log\left(p\left(a_3 \mid \mu_3, \sigma_3\right) + p\left(t_1 \mid U_1\right)\right) \\
&= \arg\min\left(\sum_{j=2,3}\left(\frac{a_j^{k2}}{2\sigma_j^2} + \frac{a_j^k \mu_j}{\sigma_j^2}\right)\right)
\end{aligned}
\tag{18}
$$

subject to

$$
t_1 + \frac{v_1}{2 * a_2} + \frac{v_1}{2 * a_3} < T
\tag{19}
$$

$$
t_1 * v_1 + \frac{\left(v_1\right)^2 - \left(v_m\right)^2}{2 * a_2} + \frac{\left(v_m\right)^2}{2 * a_3} = L_{12}
\tag{20}
$$

In scenario 1, $v_2 = 0$, $v_1 - v_m = v_m - v_2 = v_1/2$. The solution to this problem generates the travel time for mode1 and deceleration rates of mode2 and mode3. $t_m = t_1 + v_1/(2a_2)$; $t_{2'} = t_1 + v_1/(2a_2) + v_1/(2a_3)$. This information is used to construct the speed-time curve and the position-time curve, that is, trajectory of the vehicle. v is the speed of the vehicle and y is the position of the vehicle at time t. According to the model, the expressions for v and y are as follows:

$$
v = \begin{cases}
v_1 & 0 < t < t_1 \\[2mm]
v_1 + a_2 * \left(t - t_1\right) & t_1 < t < t_1 + \dfrac{v_1}{2a_2} \\[3mm]
\dfrac{v_1}{2} + a_3 * \left(t - t_1 - \dfrac{v_1}{2a_2}\right) & t_1 + \dfrac{v_1}{2a_2} < t < t_1 + \dfrac{v_1}{2a_2} + \dfrac{v_1}{2a_3}
\end{cases}
\tag{21}
$$

$$
y = \begin{cases}
v_1 * t & 0 < t < t_1 \\[2mm]
v_1 * t_1 + v_1 * \left(t - t_1\right) + \dfrac{a_2}{2} * \left(t - t_1\right)^2 & t_1 < t < t_1 + \dfrac{v_1}{2a_2} \\[3mm]
v_1 * t_1 + \dfrac{3v_1^2}{8a_2} + \dfrac{v_1}{2} * \left(t - t_1 - \dfrac{v_1}{2a_2}\right) + \dfrac{a_3}{2} * \left(t - t_1 - \dfrac{v_1}{2a_2}\right)^2 & t_1 + \dfrac{v_1}{2a_2} < t < t_1 + \dfrac{v_1}{2a_2} + \dfrac{v_1}{2a_3}
\end{cases}
\tag{22}
$$

(2) Scenario 2 for Deceleration Trajectory. As shown in Figure 6(b), there is a sample point P1 before deceleration and another, P2, in the deceleration stage of a trajectory. In this scenario, the trajectory between

P1 and P2 is reconstructed with the following expressions:

$$\arg \max(\log{(p(a_2 \mid \mu_2, \sigma_2)}$$

$$+ \log{(p(a_3 \mid \mu_3, \sigma_3) + p(t_1 \mid U_1))}$$

$$= \arg \min \left(\sum_{j=2,3} \left(\frac{a_j^{k2}}{2\sigma_j^2} + \frac{a_j^k \mu_j}{\sigma_j^2} \right) \right) \tag{23}$$

subject to

$$t_1 + \frac{v_1 - v_m}{a_2} + \frac{v_m - v_2}{a_3} = T \tag{24}$$

$$t_1 * v_1 + \frac{(v_1)^2 - (v_m)^2}{2 * a_2} + \frac{(v_m)^2 - (v_2)^2}{2 * a_3} = L_{12} \tag{25}$$

However, a part of the deceleration is unclear. To reconstruct the entire deceleration process, it is assumed that a vehicle decelerates at rate a_3 from Point P2 to Point $P2'$ as shown in Figure 6(b). The time and position when the vehicle is stationary can be calculated and this estimated point is regarded as a virtual sample point $P2'$. If the position of P3 is different from that of point $P2'$, this indicates that the estimated trajectory has an obvious error. However, the estimated trajectory could be corrected and the estimation error reduced. Keep the obtained acceleration rates a_2 and a_3 unchanged and assume that a vehicle travels at speed v_1':

$$t_1 * v_1' + \frac{(v_1')^2 - (v_m)^2}{2 * a_2} + \frac{(v_m)^2}{2 * a_3} = L_{13} \tag{26}$$

$$t_1 + \frac{v_1' - v_m}{a_2} + \frac{v_m}{a_3} = t_{12'} \tag{27}$$

The solution of this problem generates the travel time of mode1 and acceleration rate of mode2 and mode3. The speed and position at time t of the vehicle are obtained as in scenario 1. The trajectory reconstruction for scenario 2 is then completed.

(3) Scenario 3 for Deceleration Trajectory. For scenario 3 as shown in Figure 6(c), there is no sample point in deceleration and there are two sample points before deceleration. The trajectory reconstruction between P2 and P3 is the same as scenario 1. However, for the trajectory between P1 and P2, the assumption that a vehicle travels at a constant speed may not hold. Because the probe data are based on a fixed time frequency, more sample points represent a longer time at the intersection. Under this circumstance, a vehicle may experience fluctuation in speed before deceleration, especially in congested traffic. Hence, we propose a piecewise constant acceleration rate method to construct the trajectory between P1 and P2. This method adapts to various traffic conditions and the constant speed motion is a special case. Firstly, another virtual point P is assumed with the corresponding speed v_{12} which is the average speed between P1 and P2. It is assumed that a vehicle's speed at point P1 is

v_1; the vehicle's speed reaches v_{12} at time t_{12} at a constant acceleration rate. Then the vehicle's speed becomes v_2 at the same acceleration rate at timestamp T. The speed and position are then determined by the following expressions:

$$\frac{|v_1 - v_{12}|}{|v_2 - v_{12}|} = \frac{t_{12}}{T - t_{12}} \tag{28}$$

$$\frac{(v_{12} + v_1) * t_{12}}{2} + \frac{(v_{12} + v_2) * (T - t_{12})}{2} = L_{12} \tag{29}$$

T is the sample interval. After solving the problem, the speed and position of the vehicle between P1 and P2 are obtained. Combined with the trajectory between P2 and P3, the trajectory of the deceleration process is constructed.

The trajectory reconstruction of the acceleration process is similar to that for deceleration. For a trajectory, when deceleration and acceleration reconstructions are completed, the stop position and duration can be calculated. The low-frequency trajectory reconstruction is then completed.

3. Field Experiments

In this paper, the site for experiment site was a portion of the south bound carriageway on Songshan Road, spanning the intersection of Huanghe and Songsha Roads in the Nangang District of the city of Harbin in China (Figure 7). To estimate the distribution parameters of deceleration rate and acceleration rate, accumulated historical data collected by taxi from January to June 2017 are used. There are about 66700 taxis in Harbin in 2017, proving a good coverage of the road network. The typical sample frequency of the taxi is 1/30 Hz. As the type of the floating car in this paper is a taxi, to guarantee the statistical significance of the results of the distribution parameter estimation, the trajectories that included pick-up and drop-off points were excluded. The low-frequency trajectory passing through the targeted intersection for scenario 2 is selected to estimate the distribution parameters of deceleration rate and acceleration rate.

To test the accuracy of the proposed method, a field experiment was conducted in the experiment site. Four probe vehicles were equipped with GPS receivers to collect high-frequency GPS data at 1Hz. The vehicles were driven in the north-south direction (red arrow direction in Figure 7) traversing straight through the intersection repeatedly during the periods of 06:00-11:00 and 13:30-19:00 on a normal weekday, capturing the morning and evening peaks and off-peaks, respectively. To obtain the possible trajectories of the vehicles through the intersection, the vehicles joined the flow at different locations. The process lasted for six hours generating 189 valid high-frequency GPS tracks.

The statistical results for acceleration are shown in Table 1.

As can be seen in Table 1, in the deceleration process, the vehicle's acceleration in the first stage is lower than that in the second stage. However, in the acceleration process, the vehicle's acceleration in the first stage is higher than that in the second stage. The standard deviation for mode4 is the smallest. The standard deviation is a function amongst others of differences in driving behavior and habits of the driver.
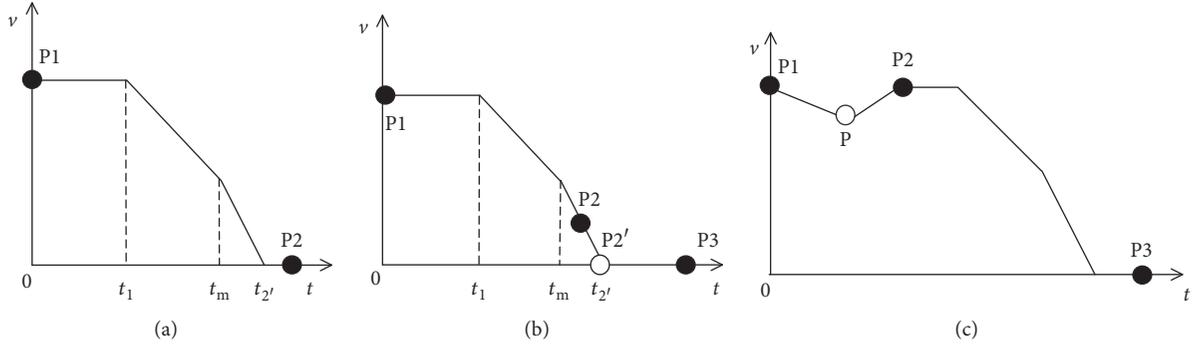
Figure 6: Different sample scenarios for deceleration process.

Table 1: Parameters for each mode.

Modes	Mean (m/s^2)	Standard Deviation (m/s^2)
Deceleration (Mode2)	0.6916	0.203
Deceleration (Mode3)	0.894	0.202
Acceleration (Mode5)	0.961	0.239
Acceleration (Mode6)	0.688	0.141

Figure 7: The experiment site.

4. Result Analysis and Discussion

4.1. Trajectory Reconstruction for Different Sample Points with the Proposed Method. With the distribution parameters of acceleration rate for each mode, the trajectory from the low-frequency floating car data can be reconstructed with the proposed method. To test the proposed method's performance, the estimated trajectories are compared with the observed trajectories achieved in a field experiment. Figure 8 shows this comparison for three typical low-frequency trajectories, each with a different number of sample points and combinations of sample scenarios. Figure 8(a) represents the trajectory with 4 sample points, scenario 1 for deceleration trajectory and scenario 3 for acceleration trajectory. Figure 8(b) illustrates trajectories with 5 sample points, scenario 2 for deceleration trajectory and scenario 3

for acceleration trajectory. Figure 8(c) illustrates trajectories with 7 sample points, scenario 3 for deceleration trajectory and scenario 3 for acceleration trajectory.

From Figure 8, it can be seen that in general the error between the estimated trajectory and the observed trajectory is small for the most part of the trajectory, which indicates the effectiveness of our method. For the trajectory in Figure 8(b), the performance of the method is the best. For the trajectory in Figure 8(a), the trajectory error between the third and fourth sample points is relatively large, while for the trajectory in Figure 8(c), the large estimation error occurs in the first and third sample points. To explain this phenomenon, Figure 9 shows the estimated and observed speed-time curve of the three trajectories. It can be seen that, for the trajectories in Figures 8(b) and 8(c), the vehicle experiences another deceleration after accelerating. For the trajectory in Figure 8(c), the vehicle even experiences another stop (inconsistent with our assumption) and experiences speed fluctuation before decelerating resulting in the relatively large trajectory estimation error. In general, the proposed method provides a good reconstruction of the low-frequency trajectory in all the cases investigated.

To further demonstrate the significance of the proposed method, the estimated trajectories were compared with trajectories that were created with the linear interpolation method. In the linear method, it is assumed that the distance the vehicle travels between any consecutive updates is fixed. As illustrated in Figure 10, the trajectories estimated by the proposed method have a better performance than the trajectories created by the linear interpolation, especially for the estimation of the stop period. In addition, the proposed method captures the change in the trajectory caused by the change in speed, while the linear-interpolation method treats all the trajectories in a similar way. From Figure 10, it can be seen that the proposed method's performance

(a) Estimated and observed trajectory for trajectory with 4 sample points

(b) Estimated and observed trajectory for trajectory with 5 sample points

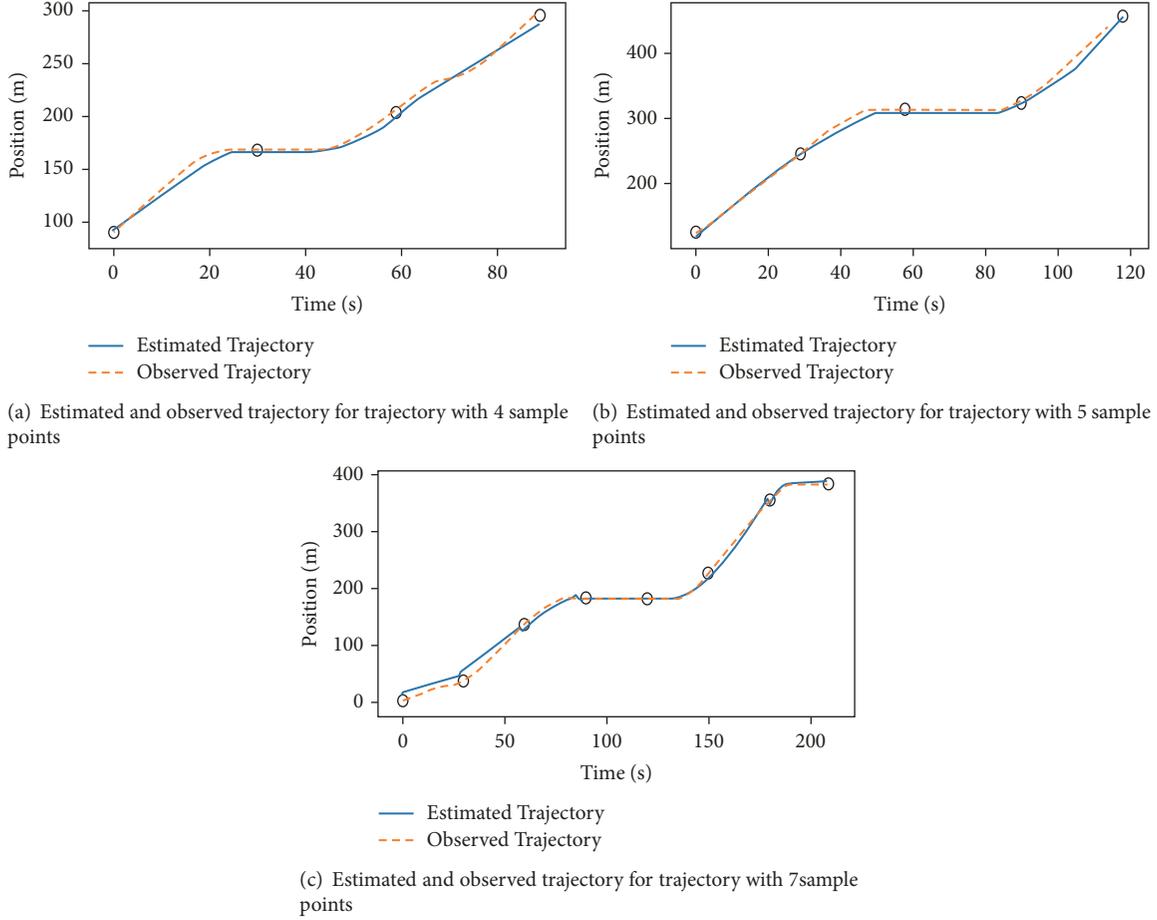(c) Estimated and observed trajectory for trajectory with 7sample points

FIGURE 8: Estimated and observed trajectories for different types of trajectory.

is obviously better than the linear-interpolation method, especially for the deceleration and acceleration processes of each trajectory, which indicates the effectiveness and practicality of our method. Note that, even in the cases where the selected trajectory is relatively complex, with fluctuations in acceleration and deceleration, the estimated trajectory is close to the observed trajectory for the most part.

To quantitatively analyze the estimation results, the Mean Absolute Error (MAE) between the estimated and ground truth trajectories is calculated. The MAE is defined as follows:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|e_i| = \frac{1}{n}\sum_{i=1}^{n}|\widehat{x_i} - x_i| \qquad (30)$$

where $\widehat{x_i}$ is the estimated position and x_i is the actual position at time step i. Hence, the estimation error of a group of trajectories, MMAE, is defined as

$$\text{MMAE} = \frac{1}{m}\sum_{i=1}^{m}|\text{MAE}_i| \qquad (31)$$

where MAE_i is the estimation error of the ith trajectory and m is the number of trajectories. The MAE between the estimated and ground truth trajectories is calculated. Besides, the stop period is also an important evaluation indicator. TAE is defined as the absolute error of the estimated time period and the actual time period for each trajectory. Hence, MTAE is defined as the mean of the TAE of a group of trajectories. To test the ability of the proposed method to reconstruct trajectories generated in different traffic conditions and stop positions, the trajectories were classified and counted according to the speed of deceleration onset point and stop position. The MMAE and MTAE values for each category of trajectory are listed in Table 2.

Compared with the linear interpolation method, the proposed method generally provides better estimation in terms of both MMAE and MTAE. For the proposed method, when the speed at the deceleration onset point ranges from 0 to 15 km/h, which indicates that the traffic is congested, the MMAE and MTAE are obviously larger than those for other speed ranges. However, for the linear-interpolation method, there is no obvious difference between each speed range. For the proposed method, the MMAE ranges from 3.47 to 10.53 with MTAE from 3.8 to 9.9. For the linear-interpolation method, the MMAE range is 9.35-13.9 and the MTAE range is 9.7-13.1. When the speed range at the deceleration onset point is 45-60 km/h, the proposed method achieves smaller MMAE and MTAE than the other speed intervals. This may be because when a vehicle travels at a relatively high speed, the travel pattern is in line with

(a) Estimated and observed speed for trajectory with 4 sample points

(b) Estimated and observed speed for trajectory with 5 sample points

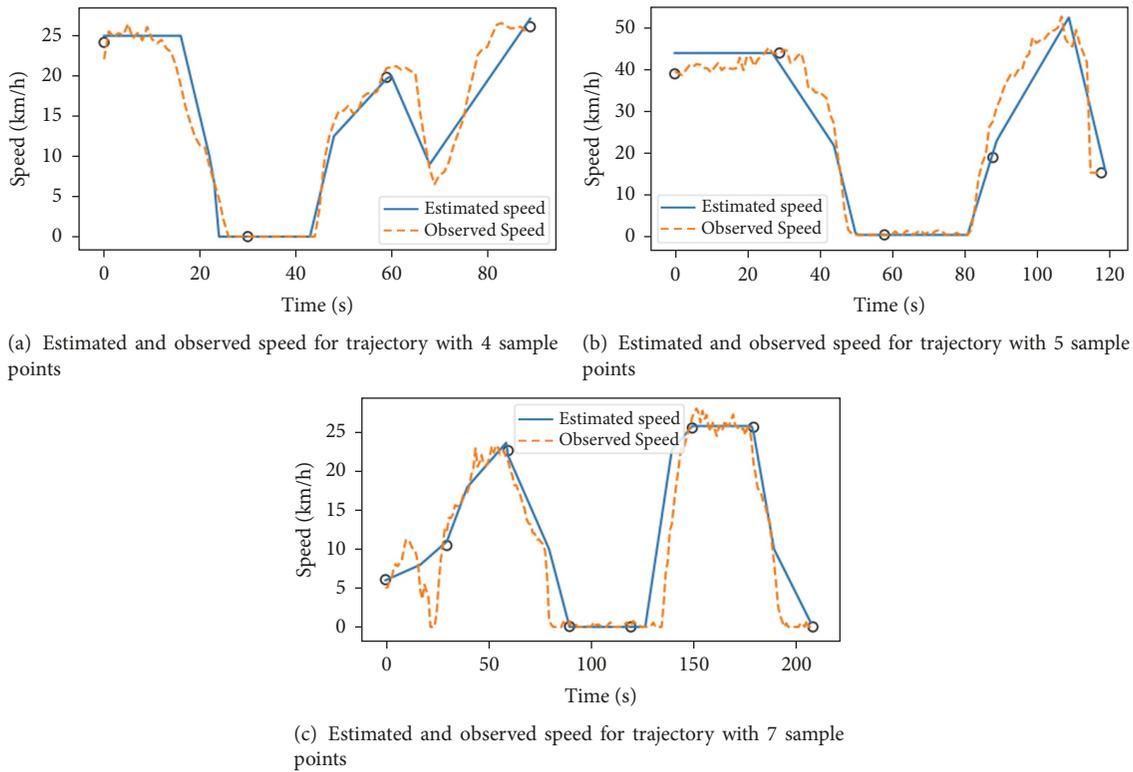(c) Estimated and observed speed for trajectory with 7 sample points

FIGURE 9: Estimated and observed speed for different types of trajectory.

TABLE 2: Trajectory estimation error for proposed method and benchmark method.

Deceleration onset Speed (Km/h)	Stopped position (m)	Number of trajectories	Proposed method		Linear Interpolation Method	
			MMAE (m)	MTAE (s)	MMAE (m)	MTAE (s)
[0-15)	[0, 100)	5	10.53	6.6	11.3	9.7
[0-15)	[100, 200)	7	9.34	8.9	9.48	13.1
[0-15)	[200, 300)	7	9.86	9.5	12.51	12.7
[0-15)	[300, 400)	16	10.26	9.9	13.9	11.6
[15, 30)	[0, 100)	14	8.23	6.3	11.1	10.3
[15, 30)	[100, 200)	10	6.11	7.4	9.95	10.5
[15, 30)	[200, 300)	9	7.57	6.0	10.66	11.2
[15, 30)	[300, 400)	7	8.35	6.2	11.38	9.9
[30, 45)	[0, 100)	20	7.39	6.5	10.17	10.4
[30, 45)	[100, 200)	16	6.46	5.8	9.35	11.7
[30, 45)	[200, 300)	13	6.32	6.3	11.36	10.9
[30, 45)	[300, 400)	10	7.49	6.1	9.77	9.8
[45, 60]	[0, 100)	23	6.49	4.6	9.59	10.5
[45, 60]	[100, 200)	15	5.93	5.7	9.84	9.9
[45, 60]	[200, 300)	17	3.47	3.8	10.97	10.7
[45, 60]	[300, 400)	-	-	-	-	-

(a) Estimated and observed trajectory for trajectory with 4 sample points

(b) Estimated and observed trajectory for trajectory with 5 sample points

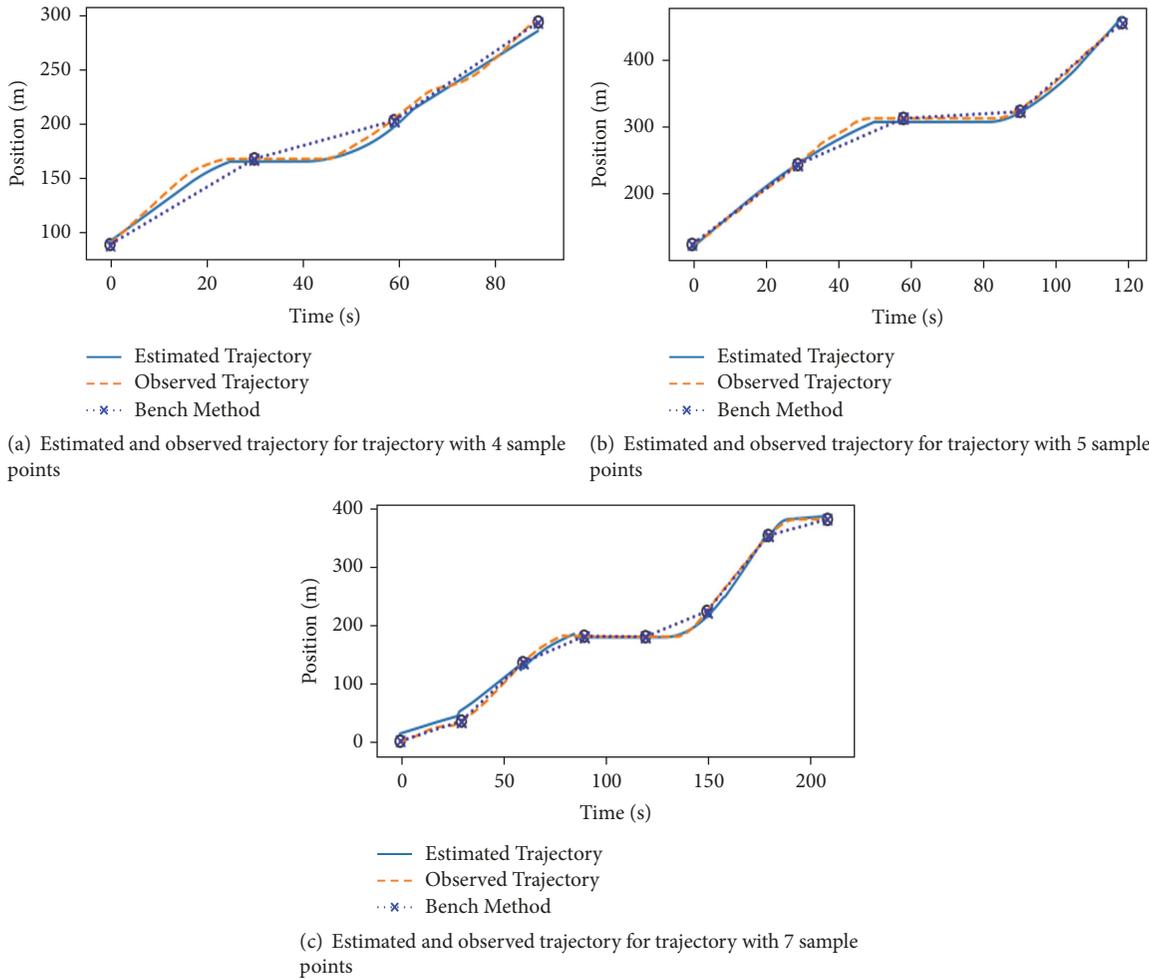(c) Estimated and observed trajectory for trajectory with 7 sample points

FIGURE 10: Comparison of estimated trajectory with the observed and linear method based trajectories.

the modal structure in our proposed method. Conversely, when the vehicle travels at a low speed, more sample points are generated and the vehicle may experience stop and go and even pass through the intersection in two or more signal cycles. This circumstance is not consistent with our model's assumption. This could explain the differences in performance of the proposed method under different travel speed ranges. It should be noted that the worst performance appears when the vehicle travels at low speed and stops at the furthest distance from the intersection. This is the case for both the proposed and linear interpolation methods.

In this paper, the number of sample points of the trajectory is an important factor for trajectory reconstruction. To give more insight into the proposed method's performance with different number of sample points, the MMAE for different number of sample points and travel speed ranges can be visualized in Figure 11.

Figure 11 shows that, for a specific stop position range, when the number of sample points is below 6, the MMAE is not affected by the number of sample points. However, when the number is above 6, as the number increases, the MMAE increases. The further the stop position from the intersection, the larger the error. When the number of sample points is
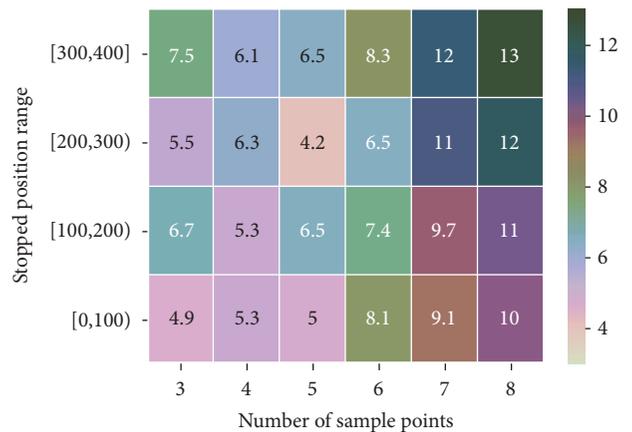

FIGURE 11: MMAE for different number of sample points and stop position ranges.

below 7, the stop position range has no effect on the MMAE. However, when the number is above 6, the MMAE increases with the distance from the stop position to the intersection.

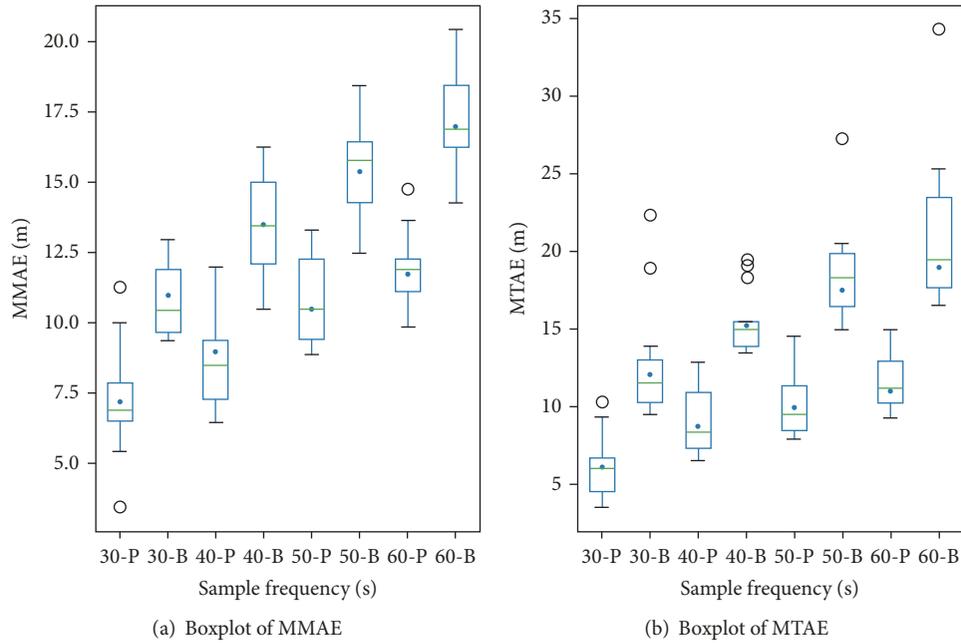(a) Boxplot of MMAE

(b) Boxplot of MTAE

FIGURE 12: Sensitivity analysis with different sample interval.

In this paper, the low-frequency floating car data is sampled at 30s. The effect of higher sampling intervals is investigated through a sensitivity analysis employing both the proposed and linear interpolation methods.

In Figure 12, "P" represents the proposed method and "B" the linear-interpolation method, with the number before them being the sample interval in seconds. The point inside each box represents the mean value. In Figure 12(a), it is shown that the linear-interpolation method is more sensitive to the sample interval. By contrast, the proposed method has a better performance in stability. For the sample intervals 30, 40, 50, and 60 the corresponding MMAE values from the proposed method are 7.2, 9.0, 10.5, and 11.7m. The corresponding MMAE values for the linear-interpolation method are 11.0, 13.5, 15.4, and 17.0m. It can be seen that the difference in performance between the proposed and linear-interpolation methods increases with the length of the sample interval. Figure 12(b) presents the MTAE for different sample intervals for the proposed method and the linear-interpolation method. As the length of the sample interval increases, the MTAE for the linear-interpolation method shows greater growth and includes frequent abnormal values. Hence, the estimation results from the proposed method are more reliable than the linear-interpolation method.

5. Conclusions

In this paper, a novel method is proposed to reconstruct a vehicle's trajectory from low-frequency floating car data. Under the assumption that the vehicle travels in a certain pattern, a model is developed to estimate the distribution parameters of deceleration rate and acceleration rate with selected historical data. With the distribution parameters as prior information, trajectories for different scenarios are constructed. Compared to a reference trajectory (truth), the proposed method achieves encouraging accuracy, with MMAE ranging from 3.47m to10.53m, higher than the commonly used linear-interpolation method which that MMAE ranging from 9.35m to 12.51m. In our method, extra information is not required and, in addition, the speed curve and distribution parameters of the deceleration rate and acceleration rate are obtained. The proposed method, therefore, paves the way for a widespread application of floating car data in transport.

The presented work has established a framework for vehicle trajectory reconstruction that could be modified or solve other traffic problems in the future. To improve and extend the application of our model, future work will include the following cases:

(1) Vehicles experiencing stop-and-go behavior when the queue at an intersection is long. In this case, the vehicle movement may deviate from our assumed driving pattern, increasing trajectory estimation error. We will extend our model to account for the limitations of our assumptions to improve its robustness.

(2) Queue length at an intersection is long such that vehicles pass through the intersection in two or more signal cycles. In this case, we do not recommend the application of our method. We mine multicycle vehicle trajectories to capture the vehicle movement patterns and use them to improve our model.

(3) Queue profile at the intersection. Combined with the traffic shockwave theory, the queue length and control delay of the intersection could be effectively estimated. We will update our model to incorporate traffic shockwave models.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] S. Pan, B. Jiang, N. Zou, and L. Jia, "Average travel speed estimation using multi-type floating car data," in *Proceedings of the International Conference on Information and Automation (ICIA '11)*, pp. 192–197, IEEE, June 2011.

[2] M. Rahmani, E. Jenelius, and H. N. Koutsopoulos, "Non-parametric estimation of route travel time distributions from low-frequency floating car data," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 343–362, 2015.

[3] S. Y. R. Rompis, M. Cetin, and F. Habtemichael, "Probe vehicle lane identification for queue length estimation at intersections," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 22, no. 1, pp. 10–25, 2018.

[4] K. Liu, T. Yamamoto, and T. Morikawa, "Estimating delay time at signalized intersections by probe vehicles," in *Proceedings of the 5th International Conference on Traffic and Transportation Studies (ICTTS '06)*, pp. 644–655, August 2006.

[5] A. D. Patire, M. Wright, B. Prodhomme, and A. M. Bayen, "How much GPS data do we need?" *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 325–342, 2015.

[6] C. Bucknell and J. C. Herrera, "A trade-off analysis between penetration rate and sampling frequency of mobile sensors in traffic state estimation," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 132–150, 2014.

[7] J. Huang and H.-S. Tan, "Vehicle future trajectory prediction with a DGPS/INS-based positioning system," in *Proceedings of the American Control Conference*, pp. 5831–5836, IEEE, June 2006.

[8] M. Rafati Fard, A. Shariat Mohaymany, and M. Shahri, "A new methodology for vehicle trajectory reconstruction based on wavelet analysis," *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 150–167, 2017.

[9] M. Montanino and V. Punzo, "Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns," *Transportation Research Part B: Methodological*, vol. 80, pp. 82–106, 2015.

[10] Z. Sun and X. J. Ban, "Vehicle trajectory reconstruction for signalized intersections using mobile traffic sensors," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 2680-2283, 2013.

[11] N. Wan, A. Vahidi, and A. Luckow, "Reconstructing maximum likelihood trajectory of probe vehicles between sparse updates," *Transportation Research Part C: Emerging Technologies*, vol. 65, pp. 16–30, 2016.

[12] P. Hao, K. Boriboonsomsin, G. Wu, and M. Barth, "Probabilistic model for estimating vehicle trajectories using sparse mobile sensor data," in *Proceedings of the 17th IEEE International Conference on Intelligent Transportation Systems (ITSC '14)*, pp. 1363–1368, October 2014.

[13] P. Hao, K. Boriboonsomsin, G. Wu, and M. J. Barth, "Modal activity-based stochastic model for estimating vehicle trajectories from sparse mobile sensor data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 701–711, 2017.

[14] Z. Zhang, Y. Wang, G. Zhang, and H. Wang, "Estimating control delays at signalised intersections using low-resolution transit bus-based global positioning system data," *IET Intelligent Transport Systems*, vol. 10, no. 2, pp. 73–78, 2016.