



Speed profile estimation using license plate recognition data



Baichuan Mo^a, Ruimin Li^{a,*}, Xianyuan Zhan^b

^a Department of Civil Engineering, Tsinghua University, Beijing, China

^b Lyles School of Civil Engineering, Purdue University, USA

ARTICLE INFO

Article history:

Received 14 February 2017

Received in revised form 11 July 2017

Accepted 14 July 2017

Keywords:

Speed profile estimation

License plate recognition

Emission estimation

Car-following model

License plate number matching procedure

Minimum cost bipartite matching

ABSTRACT

Vehicle speed profile is a fundamental data support for calculating vehicular emission using the micro-emission model. However, achieving accuracy and breadth for the speed profile estimation is difficult. This study proposes a new vehicle speed profile estimation model using license plate recognition (LPR) data. This model allows speed profile estimation of every individual vehicle between two consecutive intersections. A systematic LPR data-mending method is developed to infer the information of unmatched vehicles. Using the complete arrival and departure information as boundary conditions, a customized car-following model combined with dummy-overtaking hypothesis and boundary constraints is then applied to estimate the speed profile of vehicles. The proposed model is validated using ground truth speed information from a field experiment conducted in Langfang City in China. Results show that the model can fully capture the pattern of ground truth speed profile. A complementary model validation using the Next Generation Simulation dataset and a model application for calculating emissions are also conducted. The numerical results indicate the effectiveness of the proposed model in estimating vehicle speed profile and emissions.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Motor vehicle exhaust emissions are important contributors to air pollution. Vehicular emission is heavily affected by vehicle driving conditions, particularly the speed profile. The speed profile means the curve of speed change versus time. Over the past few decades, well-established macroscopic/mesosopic emission estimation methods have been widely used on freeways or highways with uninterrupted flow using the aggregated traffic characteristics (e.g., average speed, traffic volume, and distance traveled) (Evans and Herman, 1978; Barth et al., 1999; Barth and Boriboonsomsin, 2008; Zegeye et al., 2013; Shang et al., 2014). However, for roadways with interrupted flow (e.g., arterial roads with traffic signals), emission estimation is difficult and cannot use average traffic speed alone (Yang et al., 2011; Sun et al., 2015); the reason is that vehicles with the same average speed present significantly different speed profiles in arterial roads from those on freeway segments (Nesamani et al., 2007). As a result, microscopic approaches that can reasonably estimate speed profile in the individual vehicle level are necessary to estimate the emissions of arterial roads accurately.

Abbreviations: LPR, license plate recognition; MRE, mean relative error; NGSIM, Next Generation Simulation; CMEM, comprehensive modal emissions model; FVD, full velocity difference; GA, genetic algorithm; RMSE, root-mean-square error; MAE, mean absolute error.

* Corresponding author.

E-mail addresses: mbc14@mails.tsinghua.edu.cn (B. Mo), lrmin@tsinghua.edu.cn (R. Li), zhanxianyuan@purdue.edu (X. Zhan).

<http://dx.doi.org/10.1016/j.trc.2017.07.006>

0968-090X/© 2017 Elsevier Ltd. All rights reserved.

Recently, several microscopic models with reconstructed vehicle speed profiles are proposed. Yang et al. (2011) estimated vehicle trajectories by explicitly modeling driving modes, such as steady-state cruise, acceleration, deceleration, and idle modes. Then, the estimated vehicle trajectories are used to calculate emissions. This method obtains improved emission estimation results compared with macroscopic/mesososcopic methods but is limited by its incapability to capture detailed behavior of individual vehicles, such as the acceleration and deceleration processes (Sun et al., 2015). Following Yang et al.'s (2011) study, Sun et al. (2015) extended the vehicle trajectory estimation method proposed by Sun and Ban (2013) to a modal decomposition speed profile estimation method for emission estimation using mobile sensing data. Uniform deceleration, state-dependent acceleration, and random-noise-added cruise process are considered to provide a realistic speed profile. Sun et al.'s (2015) method overcomes the limitations of Yang et al.'s (2011) method and obtains reasonable emission estimation results. However, the estimated speed profiles in Sun et al.'s (2015) study are still different from the ground truth speed profiles because of the rigidly modal decomposition of the traveling process. In addition, the estimated number of vehicles may not match the ground truth data because of the limitation of mobile sensing data, thereby leading to emission estimation errors. Therefore, a microscopic method that can model the fine vehicle speed profiles of the entire traffic flow must be developed. License plate recognition (LPR) data provide an alternative solution to this end.

LPR is an emerging technology that is widely used in urban and highway transportation systems. In recent years, large-scale LPR systems have been rapidly deployed in many countries. For example, Beijing has an LPR system with 1958 LPR cameras covering the entire metropolitan area (Beijing Traffic Management Bureau, 2017); this number is expected to increase in the next few years (Beijing Municipal Committee, 2016). The data obtained from the LPR system can provide information about the departing timestamps of vehicles at the stop line as well as their license plate number. LPR systems have been employed in many different transport applications involving enforcement, vehicle monitoring, and access control (Nakanishi and Western, 2005; Rossetti and Baker, 2001). With the unique identification of vehicle license plate, LPR data are used for detailed traffic information estimation, such as link travel time estimation (Bertini et al., 2005; Yasin et al., 2009) and queue length estimation (Zhan et al., 2015). Compared with other decentralized data (e.g., mobile sensing data and GPS data), LPR data do not suffer from the issue of low penetration rate because LPR cameras can record the information of nearly all vehicles departing from an intersection. Therefore, LPR data can be used to model the entire traffic flow and overcome the aforementioned limitations of Sun et al.'s (2015) method. However, LPR data are seldom available to the research community due to privacy issues, and research efforts to develop data analytics for complex transportation applications, including vehicle speed profile estimation, are limited. Despite the increasing recognition accuracy of the LPR system, LPR data still suffer from a low matching rate of license plate number between upstream and downstream intersections (Section 2). Some LPR data-mending methods have been proposed but are flawed. Oliveira-Neto et al. (2012) proposed license plate-matching procedures based on the probability model; these procedures improve the matching rate of LPR data but fail to recover the complete information of LPR data. Zhan et al. (2015) used an interpolation model based on the Gaussian process to infer unobserved and erroneously recognized data; however, this model filters out a large amount of prior information and thus results in low accuracy.

In this study, we propose a hybrid framework for speed profile estimation that fully utilizes the information provided in LPR data. This framework allows the detailed estimation of the speed profile of every individual vehicle between two consecutive intersections. An LPR data-mending model combining the advantages of Oliveira-Neto et al.'s (2012) method and Zhan et al.'s (2015) method is first developed to infer the information of all unmatched vehicles. This data-mending model overcomes the drawbacks of Oliveira-Neto et al.'s (2012) method and Zhan et al.'s (2015) method and fully utilizes valid data information. Next, we use a highly customized car-following model to estimate the speed profile of vehicles. This model can reconstruct vehicle trajectories under the boundary constraints provided in LPR data. A field experiment is conducted in Langfang, Hebei Province, China to collect ground truth data for validation. The ground truth vehicle speed profile data are collected using two probe vehicles equipped with GPS devices. The results show that the model can fully capture the pattern of the ground truth speed profile. The mean relative error (MRE) of speed profile estimation is 12.98%. A complementary validation is conducted using the Next Generation Simulation (NGSIM) Peachtree Street dataset (US Department of Transportation—FHWA, 2008) and shows that the proposed model is robust. Then, the estimated speed profile is applied to estimate emissions with comprehensive modal emissions model (CMEM; Barth et al., 2001). The estimation relative errors of fuel consumption, CO emissions, and NO_x emissions are approximately 16% in the individual vehicle level.

The contributions of this work are summarized as follows:

1. A systematic LPR data-mending method is proposed to infer the complete information of unmatched vehicles, thereby overcoming the drawbacks of previous research and maximizing the use of valid data information.
2. A customized car-following model fully corresponding to LPR data characteristics is proposed. The model can be extended to other micro-traffic information estimation methods using similar data (e.g., RFID data).
3. The speed profile of vehicles on road segments is estimated using LPR data.

The rest of the paper is organized as follows. Section 2 presents the proposed data-mending method. Section 3 presents the speed profile estimation using the proposed customized car-following model. Section 4 presents the field experiment design, and Section 5 shows the numerical results. Section 6 elaborates the conclusions of the study.

2. LPR data mending

Prior to formally describing the LPR data-mending method, we first define the following concepts:

- **Arrival time:** The time when vehicles enter the investigated road segment from upstream intersection. Notably, the record of arrival time is unaffected by the upstream queuing process (Hao et al., 2015) because it is the time when vehicles enter the road after queuing.
- **Departure time:** The time when vehicles depart from the investigated road segment in downstream intersection.
- **Cumulative index:** The time sequence of certain vehicle passing through the intersection.
- **Cumulative arrival/departure curve:** The corresponding relationship between the cumulative indices of vehicles and their arrival or departure time.
- **Consistent vehicle:** Vehicles with the same cumulative index in upstream and downstream LPR record.
- **Inconsistent vehicle:** Vehicles with different cumulative indices in upstream and downstream LPR record.
- **Matching:** Consider an undirected graph $G = (V, E)$. A *matching* M of G is a subset of the edges E , such that no vertex in V is incident to more than one edge in M .
- **Bipartite graph:** A graph $G = (V, E)$ is bipartite when V is formed by two disjoint subsets S_1 and S_2 , and all edges $[i, j] \in E$ have an endpoint in S_1 and the other in S_2 .
- **Perfect matching:** A matching in which every vertex of the graph is incident to exactly one edge of the matching.
- **Minimum cost bipartite matching:** A matching with minimum cost between S_1 and S_2 among all those with maximum cardinality.

In typical LPR systems used for red-light violation enforcement, a virtual detection zone is set right behind the stop line of each lane (Fig. 1); this zone has the size usually smaller or equal to the area occupied by a vehicle. The LPR camera will take a picture when a vehicle passes the stop line. Thus, the accurate timestamps and their recognized license plate number can be obtained. However, errors can occur in the data because of the weather, plate configuration, and inaccuracies associated with recognition algorithms. The most common resulting errors are (1) erroneous records of timestamps, (2) erroneous recognition of license plates, and (3) failed recognition of license plate numbers. The first error exists only in the first vehicle of each queue. When the vehicle stops in front of the stop line but with body over the stop line (Fig. 1(b)), the camera will also take a picture of it, thereby making an incorrect record of actual leaving time. An example of the second error is when the system recognizes “Z” as “2”. The third error results in vehicles with plate numbers recorded as “unrecognized” (the timestamp is however still recorded accurately). In addition, vehicles entering from the right turning lane are not monitored as “right turning on red” is often permitted in many countries (e.g., China), thereby resulting in information loss on these vehicles. However, given that the timestamps of nearly all vehicles (including the vehicles entering from the right turning lane) passing through the downstream intersection are recorded precisely, the recorded downstream license plates can be defined as the reference and used to match with the upstream recorded license plates for flow conservation. Directly matching license plates leads to vehicle records with unmatched license plates (referred to as *unmatched vehicle*) because of the limitation of the LPR data, as mentioned previously. Fig. 2 presents the proportion of unrecognized and unmatched vehicles from 8:00 to 17:00 in accordance with the field experiment data in Section 4. Each point is calculated for every 15 min interval. The average unmatched vehicles account for 50% of the total data records. For unknown reasons, the unrecognized rate in upstream suddenly increases abnormally and results in an unusually high proportion of unmatched vehicles between 10:00 and 13:00 (such poor performance of LPR system is unusual). The missing information of unmatched vehicles prohibits the direct implementation of LPR data for speed profile estimation, such that data mending becomes necessary.

We propose a systematic LPR data-mending method for inferring unmatched vehicles information. We first use signal timing information to correct the erroneous timestamps in LPR data (i.e., Error 1), which can provide a realistic link travel time information for the subsequent steps. Then, simplified Oliveira-Neto’s et al. (2012) procedure is used to match the same vehicles recorded in downstream and upstream. The matched vehicles are used as the prior information in cumulative arri-

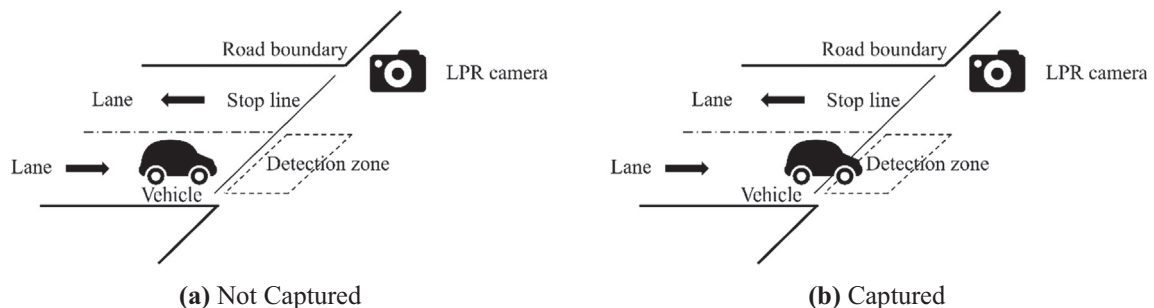


Fig. 1. Illustration of the LPR system.

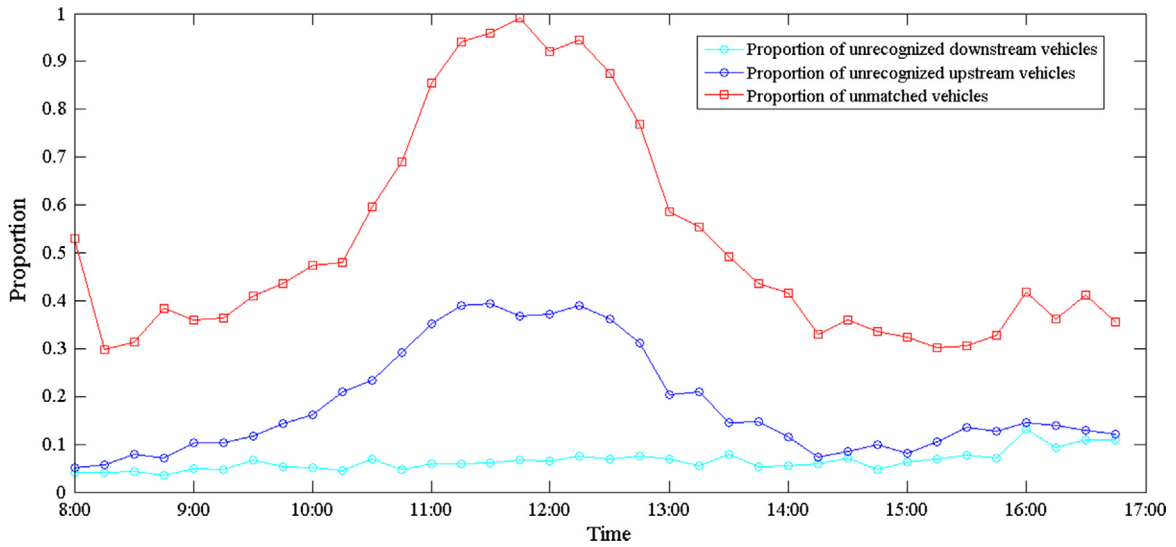


Fig. 2. Proportion of unrecognized and unmatched vehicles.

val/departure curve reconstruction. A Gaussian interpolation model (Zhan et al., 2015) and a minimum cost perfect bipartite matching method are used to infer the unknown arrival time and arrival indices of unmatched vehicles. In this manner, the complete information of all unmatched vehicles can be recovered; the drawbacks of Error 2, Error 3, and no monitoring of right turning vehicles of the LPR system can be overcome.

2.1. Erroneous timestamp correction

To provide realistic link travel time information for the subsequent steps, the signal timing information is used to correct erroneous records of timestamps. A typical upstream signal timing plan with four phases is illustrated in Fig. 3. All the arrows represent the allowable passing directions within a phase plotted below. The green arrows represent the directions of vehicles that can enter the investigated road segment. These vehicles can be considered to obtain a combination of three arrival processes: (1) arriving vehicles from left turning movement during the first phase (from 0 to T_1); (2) arriving vehicles from through movement during the second phase (from T_1 to T_2); and (3) arriving vehicles from right turning movement during the entire cycle (from 0 to T_{cyc}). The upstream LPR system does not record the third arrival process; thus, our discussion of erroneous timestamp fixing procedure is restricted to only the first and second arrival processes.

We denote V_1 as the first vehicle of a queue and V_2 as the vehicle in the same queue right behind V_1 . T_{V_1} represents V_1 's arrival (if V_1 in upstream) or departure time (if V_1 in downstream) derived from the timestamp TS_{V_1} . The procedure in Fig. 4 can be summarized by the following steps. (1) For vehicles arriving from left turning movement, if $T_{V_1} \notin [0, T_1]$, which means the recorded timestamp TS_{V_1} is not the actual time for leaving the stop line, then correct the timestamp TS_{V_1} to TS'_{V_1} by performing piecewise cubic Hermite interpolation (Powell et al., 1981) (Algorithm 1). Considering that this error exists only in

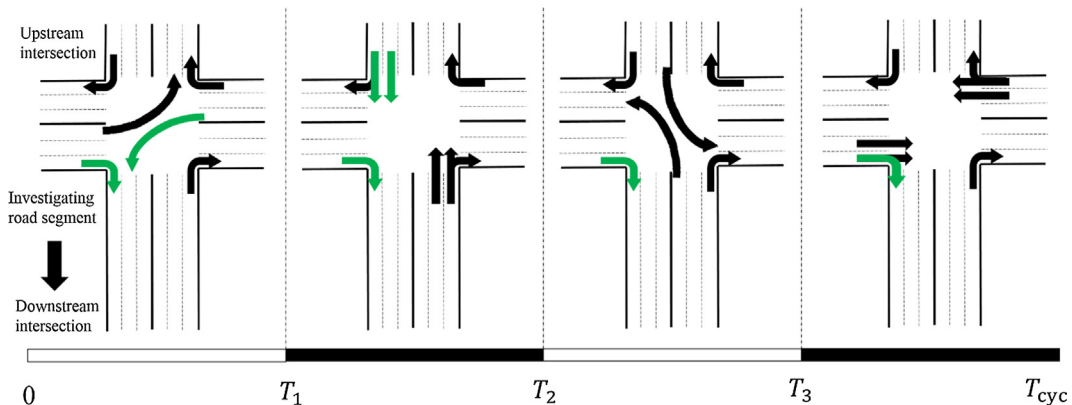


Fig. 3. Illustration of a typical upstream signal timing plan.

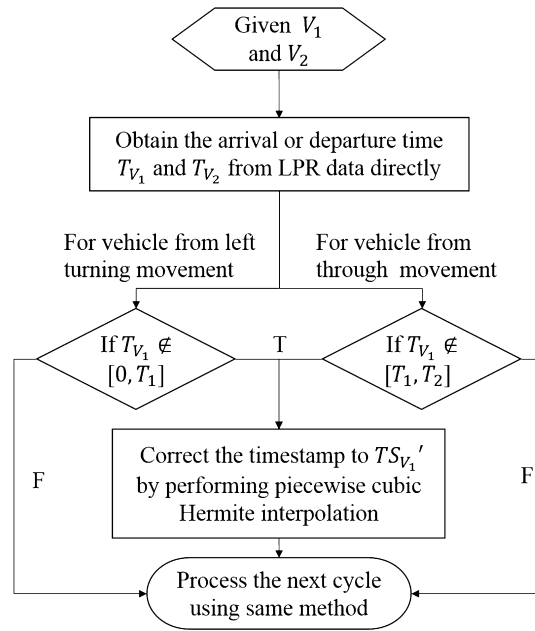


Fig. 4. Procedure of repairing erroneous timestamps.

the first vehicle of each queue, other vehicle timestamp in the same queue is considered to be accurate and can be used as reference. Given the traffic dynamics of the departure process, the cubic Hermite interpolation can infer the most realistic timestamp of the only unknown point $(T_{V_1}, 1)$ in this queue compared with other interpolation algorithms because of numerical testing. (2) For vehicles arriving from through movement, if $T_{V_1} \notin [T_1, T_2]$, then correct the timestamp to TS'_{V_1} the same as step 1. (3) Process the next cycle using the same method. The schematic in Fig. 5 provides a better illustration for vehicles arriving from left turning movement. The blue vehicle V_1 satisfies $T_{V_1} \notin [0, T_1]$, which should correct the timestamp to TS'_{V_1} . The cycle that V_1 belongs to may change after the correction procedure. Thus, the correction process is significant because the cycle that every vehicle belongs to is important for cumulative arrival/departure curve reconstruction in Section 2.3.

Algorithm 1 (Interpolation-based correction of erroneous timestamps.).

-
- Step1: Suppose (T_{V_i}, i) $(i \neq 1)$ are vehicles in the same queue as $(T_{V_1}, 1)$, where T_{V_i} represents the i th vehicle arrival or departure time in the queue. Indicate $(T_{V_0}, 0) = (0, 0)$.
 - Step2: Use (T_{V_i}, i) $(i \neq 1)$ to perform piecewise cubic Hermite interpolation (add a dummy vehicle $(T_{V_2}, 2) = (T_1, 2)$ if only V_1 is present in this queue), thereby obtaining the interpolated first vehicle $(T'_{V_1}, 1)$ in the queue.
 - Step3: Accept T'_{V_1} as V_1 's real arrival or departure time and transform it to TS'_{V_1} .
-

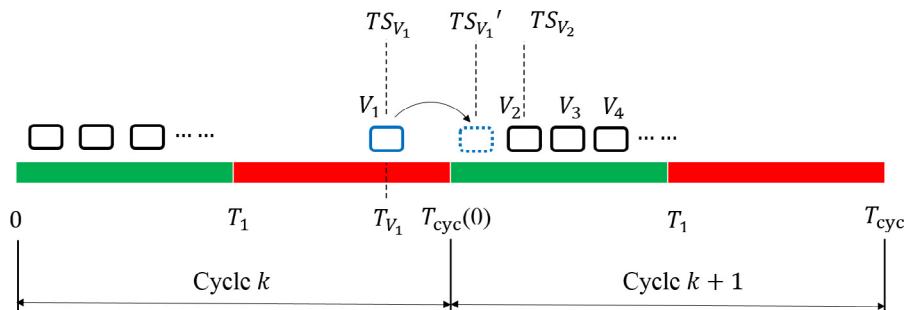


Fig. 5. Schematic of repairing erroneous timestamps for vehicles arriving from left turning movement.

2.2. Vehicle-Matching procedure

Fig. 2 shows that directly matching upstream and downstream vehicles using license plate number produces high fraction of unmatched vehicles. Erroneous recognition (i.e., Error 2) contributes to the main source of this problem (Oliveira-Neto et al., 2012). In fact, misrecognition probability differs among characters. For example, the system is more likely to misread “5” to “S” than to “1”. As a result, Oliveira-Neto et al. (2012) proposed a probabilistic vehicle-matching procedure that can reduce the proportion of unmatched vehicles.

The key component of the said procedure is to find which two license plate sequences present the highest probability to match even if they are different. We simplify Oliveira-Neto’s method into a string matching problem of the same length, with the assumption that the recorded downstream license plate is a true reference string for matching and that the misreading only occurs in the upstream. According to the numerical test, considering the misreading error of the dual LPR setup (i.e., using Oliveira-Neto’s method) will cost a computation time two times that of the simplified one but will only result in slight improvement in the matching results.

We let $u = u_1u_2 \dots u_i \dots u_l$ and $d = d_1d_2 \dots d_i \dots d_l$ be the two sequences of license plate numbers read at upstream and downstream intersections, respectively, with string lengths equal to l . We denote $p(u_i|d_i)$ as the conditional probability of a given downstream string character d_i being recognized as u_i . Detecting the length of each license plate number in our real-world LPR data shows that the recognition error seldom results in the change of string length. Thus, we assume that two strings with different lengths are unlikely to be matched. The probability of downstream license plate sequence d matched with upstream license plate sequence u can thus be given by the following equation:

$$p(d \rightarrow u) = \prod_{i=1}^l p(u_i|d_i). \tag{1}$$

The most likely match MLM of d is given by the sequence in upstream with the highest probability and is equivalent to minimizing the negative natural logarithm of Eq. (2) (Oliveira-Neto et al., 2012).

$$MLM(d \rightarrow u) = \min_{TT \in [TT_{\min}, TT_{\max}]} \left\{ \sum_{i=1}^l \log \left(\frac{1}{p(u_i|d_i)} \right) \right\}, \tag{2}$$

where TT is the travel time through the investigated road segment. $TT_{\max}(TT_{\min})$ is the maximum (minimum) value of TT , which is derived from the sample travel times of directly matched vehicles (i.e., matching based on exactly the same license plate). Notably, TT_{\max} and TT_{\min} are constants in this study, which defines an approximate range of matching candidates to simplify the searching process. MLM has two thresholds, namely, $MLM_{\text{threshold}}^{\max}$ and $MLM_{\text{threshold}}^{\min}$. Thus, if $MLM(d \rightarrow u) > MLM_{\text{threshold}}^{\max}$, then d is unlikely to match u . Conversely, if $MLM(d \rightarrow u) < MLM_{\text{threshold}}^{\min}$, then d is very likely to match u . Apart from the magnitude of $MLM(d \rightarrow u)$, a secondary evidence, TT , is also used in the matching procedure. TT of cycle k is assumed to be normally distributed with parameters (μ_k, σ_k) . μ_k and σ_k can be estimated using the travel times of directly matched vehicles in cycle k and the adjacent two cycles ($k - 1$ and $k + 1$). The probabilistic matching process, as shown in Fig. 6, can be illustrated by the following steps. (1) Given TT_{\max} and TT_{\min} , for downstream license plate d , calculate μ_k and σ_k from the sample of travel times in the direct matching procedure. (2) Compute $MLM(d \rightarrow u)$. (3) If $MLM(d \rightarrow u) < MLM_{\text{threshold}}^{\min}$, then accept u as the genuine match of d . (4) If $MLM(d \rightarrow u) > MLM_{\text{threshold}}^{\max}$, then denote the matching result of d as “unmatched.” (5) If $MLM_{\text{threshold}}^{\min} \leq MLM(d \rightarrow u) \leq MLM_{\text{threshold}}^{\max}$, then accept u as the genuine match of d only if $TT \in [\mu_k - \delta_k, \mu_k + \delta_k]$ (where δ_k is the amplitude of TT for cycle k). Otherwise, denote the matching result of d as “unmatched.” δ_k is calculated using the following equation (Oliveira-Neto et al., 2012):

$$\delta_k = \sqrt{9 \times \frac{MLM_{\text{threshold}}^{\max} - MLM(d \rightarrow u)}{MLM_{\text{threshold}}^{\max} - MLM_{\text{threshold}}^{\min}}} \times \sigma_k. \tag{3}$$

Eq. (3) shows that, when $MLM(d \rightarrow u)$ is large, the constraint of TT becomes strict to reduce the false match.

In actual application, the conditional probability $p(u_i|d_i)$ trained by Oliveira-Neto et al. (2012) is directly used in the matching procedure because of the lack of training samples on recognition errors. As an example, Table 1 presents the calculating process of our matching procedure. $TT_{\min} = 30$, $TT_{\max} = 250$ and $MLM_{\text{threshold}}^{\max} = 13$, $MLM_{\text{threshold}}^{\min} = 6.5$ are used in the matching procedure. Notably, the values of $MLM_{\text{threshold}}^{\max}$ and $MLM_{\text{threshold}}^{\min}$ are determined from many numerical tests, which show that this pair of values has good results.

Using the probabilistic vehicle-matching procedure, unmatched vehicles decrease but some of them still remain. The remaining unmatched vehicles are caused by several reasons. For example, the unmonitored right turning vehicles are only recorded in the downstream intersection and thus cannot be matched. In addition, if a vehicle is recorded as “unrecognized” in either upstream or downstream intersection (i.e., Error 2), then it cannot be matched. Nevertheless, this process increases the amount of prior information for the reconstruction of the cumulative arrival curve presented in Section 2.3.

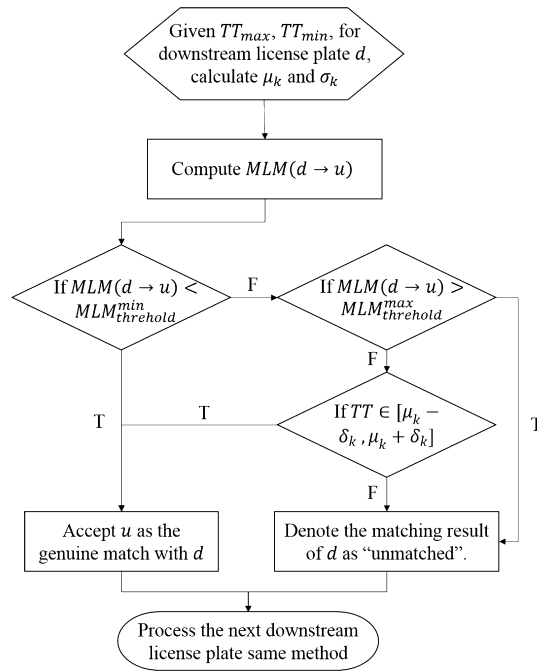


Fig. 6. Process of the proposed probabilistic vehicle-matching procedure.

2.3. Cumulative arrival/departure curve reconstruction

The cumulative departure curve can be plotted by LPR data directly owing to the complete record of vehicle departure time in downstream. Our key task is to reconstruct the cumulative arrival curve and infer the information for unmatched vehicles (i.e., the arrival time and cumulative arrival index). Zhan et al. (2015) proposed a Gaussian process interpolation model to reconstruct the arrival curve; however, this model suffers from a large predicted variance because of filtering out all inconsistently matched vehicles. We retain the information of inconsistent vehicles and perform a minimum cost bipartite matching to reconstruct a more realistic arrival curve than previous study, thereby accurately optimizing the reconstruction method.

The Gaussian process interpolation model can find the most likely function (arrival curve) by inputting a finite sample set. The cumulative arrival curve is modeled as follows:

$$y_u(x|\theta) = \mu(x|\theta) + \eta, \tag{4}$$

where y_u is the upstream arrival index, x is the vehicle arrival timestamp, $\mu(x|\theta)$ is the number of mean cumulative arriving vehicles (i.e., the arrival indices) modeled by Zhan et al. (2015), η is the disturbance term that is assumed to be normally distributed, and θ is used to specify the mean cumulative arrival process. Then, the points of matched vehicles in the cumulative arrival curve can be plotted directly (with $\eta = 0$). The points of unmatched vehicles (including the unmonitored right turning vehicles) can be estimated using the methods presented in the subsequent paragraphs.

The upstream index of consistently matched vehicles (which is equal to the downstream index) and their arrival timestamp are already known and can directly be used for the Gaussian process interpolation model. A reference cumulative arri-

Table 1
Example of most likely license plate matching procedure.

d	u	TT (s)	$\sum_{i=1}^l \log(\frac{1}{p(u d_i)})$	$MLM(d \rightarrow u)$	Judgment	Matching result
R9A930	RYS560	40	∞	8.836	$MLM(d \rightarrow u) \in [MLM_{threshold}^{min}, MLM_{threshold}^{max}]$ and $TT \in [\mu_k - \delta_k, \mu_k + \delta_k]$	R91730
	R91730	58	8.836			
	ROY085	72	∞			
			
R9980D	R99803	68	5.528	5.528	$MLM(d \rightarrow u) < MLM_{threshold}^{min}$	R99803
			
R9A93G	R91775	99	17.011	17.011	$MLM(d \rightarrow u) > MLM_{threshold}^{max}$	Unmatched
			

val can thus be obtained after execution. On the basis of the reference curve, the upstream index of inconsistently matched vehicles can be inferred by matching the arrival timestamp. An iteration method of the arrival timestamp matching process is used to ensure the reasonability. Then, the information of all matched vehicles (consistent and inconsistent) is input into the Gaussian process interpolation model, and the final cumulative arrival curve is reconstructed. The reconstruction process is illustrated in [Algorithm 2](#). Detailed background information can be found in [Zhan et al. \(2015\)](#), [Bishop \(2006\)](#), and [Roberts et al. \(2013\)](#).

Algorithm 2 (Cumulative arrival/departure curve reconstruction).

Step 1: Construct departure curve and initialize

Draw the departure curve using LPR data directly. As for cycle k , create the vehicle downstream index $\mathbf{y}_d = (1, 2, \dots, N_k)$ on the basis of the departure timestamp sequence, where N_k is the number of vehicles entering the investigated segment in cycle k . Input the arrival timestamps of matched vehicles into ordered set \mathbf{x}^m . Denote \mathbf{y}_d^m as the corresponding ordered downstream index vector (e.g., $\mathbf{y}_d^m = (1, 2, 4, 6)$, $\mathbf{x}^m = (28152, 28165, 28157, 28168)$), where $\mathbf{y}_d^m(1)$ and $\mathbf{y}_d^m(4)$ ($\mathbf{y}_d^m(2)$ and $\mathbf{y}_d^m(3)$) are the downstream indices of consistent (inconsistent) vehicles).

Step 2: Initialize \mathbf{y}_u^m

Denote \mathbf{y}_u^m as the ordered upstream index vector based on the arrival time sequence of \mathbf{x}^m , with $\mathbf{y}_u^m(i)$ and $\mathbf{y}_d^m(i)$ representing the same matched vehicle i . (e.g., $\mathbf{y}_u^m = (1, 4, 2, 6)$, where indices “4” and “2” in \mathbf{y}_u^m are assigned only to distinguish the inconsistent vehicles and need to be estimated in step 4).

Step 3: Initialize \mathbf{x}_0 and \mathbf{y}_0

As for $\mathbf{y}_u^m(i)$, if $\mathbf{y}_u^m(i) \neq \mathbf{y}_d^m(i)$, assign $\mathbf{y}_u^m(i)$ as null. Input all non-null elements of \mathbf{y}_u^m into \mathbf{y}_0 as well as their corresponding arrival timestamps into \mathbf{x}_0 . (e.g., $\mathbf{y}_0 = (1, 6)$, $\mathbf{x}_0 = (28152, 28168)$)

Step 4: Perform the Gaussian process interpolation

Use the pair $(\mathbf{x}_0, \mathbf{y}_0)$ as known samples to perform the Gaussian process interpolation model. Obtain the reference curve $\hat{\mathbf{y}}_u(\mathbf{x}|\theta)$.

Step 5: Estimate the upstream indices of inconsistently matched vehicles

As for $\mathbf{x}^m(j)$, if $\mathbf{x}^m(j) \notin \mathbf{x}_0$, then find the upstream index $\mathbf{y}_u^m(j)$ through $\hat{\mathbf{y}}_u$. (e.g., $\mathbf{x}^m(2) = 28165$ and $\mathbf{x}^m(3) = 28157 \notin \mathbf{x}_0$. Thus, $\mathbf{y}_u^m(2) = \hat{\mathbf{y}}_u(\mathbf{x}^m(2)|\theta) = 5$ and $\mathbf{y}_u^m(3) = \hat{\mathbf{y}}_u(\mathbf{x}^m(3)|\theta) = 2$)

If no duplicate elements are present in \mathbf{y}_u^m , then proceed to step 5. (e.g., $\mathbf{y}_u^m = (1, 5, 2, 6)$)

Otherwise, replace the duplicate elements to null. Return to step 3.

Step 6: Reconstruct the final curve

Re-perform the Gaussian process interpolation model using $(\mathbf{x}^m, \mathbf{y}_u^m)$ to reconstruct the final cumulative arrival curve.

After the arrival/departure curve reconstruction, the arrival time of matched vehicles can be easily obtained by inputting the corresponding arrival indices. However, the arrival indices of unmatched vehicles are still unknown as a result of the unknown consistency. A minimum cost bipartite matching method ([Ahuja et al., 1993](#)) is proposed to estimate the arrival time and cumulative arrival indices of unmatched vehicles.

We assign the downstream indices of all unmatched vehicles into \mathbf{y}_d^{um} . Then, we obtain the upstream indices of all unmatched vehicles on the basis of the reconstructed curve and input them into \mathbf{y}_u^{um} . The key task is to match every index in \mathbf{y}_d^{um} with that in \mathbf{y}_u^{um} . Considering the same dimension of two vectors, the problem can be considered a minimum cost perfect bipartite matching. We denote M as a possible matching in this case. A sample graph G is plotted in [Fig. 7](#). The cost of edges $(\mathbf{y}_u^{um}(i), \mathbf{y}_d^{um}(j))$ is defined as $c_{ij} = e^{|\mathbf{y}_d^{um}(i) - \mathbf{y}_u^{um}(j)|}$ because vehicles are more likely to follow their predecessor and thus keep the same index in the upstream and downstream intersections ([Hao et al., 2013](#)). The Hungarian method ([Kuhn, 1955](#)) is applied to find a perfect matching of minimum cost; the cost of a matching M is given by $C(M) = \sum_{i,j \in M} c_{ij}$. Thus, the upstream indices of unmatched vehicles and their arrival time can be obtained through the matching results.

Using the proposed cumulative arrival/departure curve reconstruction process, the complete information of unmatched vehicles (caused by Error 2, Error 3 and no monitoring of right turning vehicles) can be recovered and used for vehicle speed profile estimation in [Section 3](#).

3. Vehicle speed profile estimation

The arrival and departure times of every vehicle can be obtained from the proposed LPR data-mending method, and they can be regarded as the boundary constraints of vehicle trajectories. A customized car-following model combined with boundary conditions and dummy-overtaking hypothesis is proposed for speed profile estimation. This model is derived from full velocity difference (FVD) for a car-following theory ([Jiang et al., 2001](#)). The FVD model is based on an extension of the optimal velocity model ([Bando et al., 1995](#)) and exhibits superiority in capturing the traveling characteristics of vehicles in detail. However, the FVD model only considers the car-following motion in a single lane without any overtaking behavior. In this study, overtaking behavior, although only a small proportion, cannot be neglected because of its significance to speed

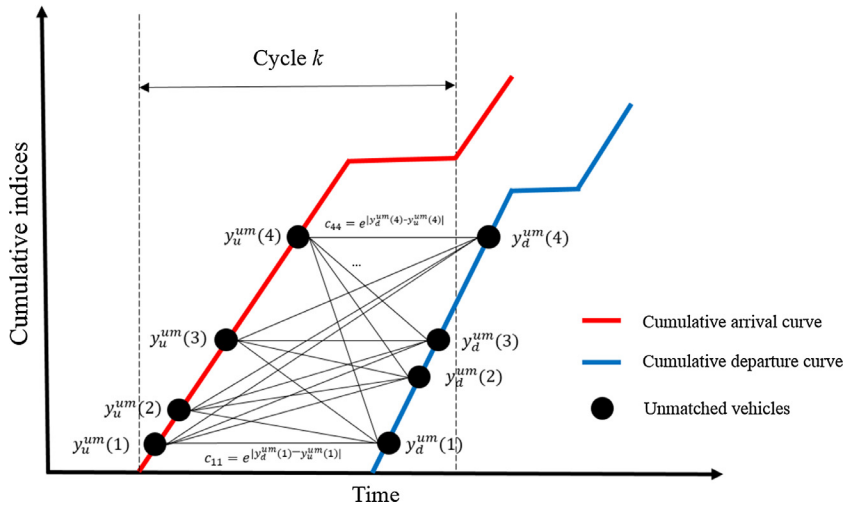


Fig. 7. Sample graph for illustrating the minimum cost perfect bipartite matching problem.

estimation. Thus, we propose a dummy-overtaking hypothesis to simulate the overtaking behavior. The boundary constraints of two intersections are also added to this model for simulating the travel behavior of vehicles.

For vehicle n , if $y_u(n) > y_d(n)$, where $y_u(n)$ is the arrival index and $y_d(n)$ is the departure index, then vehicle n can be regarded as an *overtaking vehicle*, which usually travels at a high speed with overtaking and car-following behavior. Other vehicles are regarded as *normal vehicles*, which only present car-following behavior. Overtaking behavior must occur with lane change behavior in real situation. This study assumes that the dummy-overtaking and car-following processes occur on the same divided lane of the investigated road and that the lane change behavior is ignored. This hypothesis is illustrated in Fig. 8.

For overtaking vehicle i , we assume the overtaking behavior starts as soon as it enters the individual lane. During the overtaking process, vehicle i does not affect the following behavior of other vehicles by marking a “dummy” sign and will follow vehicle k in the same lane with an overtaking optimal velocity function until it comes to the middle of vehicles j and k to overtake (where $y_u(k) = y_u(i) + 1$, $y_u(j) = y_u(i) - 1$). Meanwhile, the “dummy” sign is removed, and vehicle i becomes a normal vehicle that finishes the overtaking process and starts the car-following process. The dummy-overtaking process simulates an overtaking vehicle traveling on the near lane at high speed and changing to the investigated lane to finish overtaking. Thus, the overtaking behavior is hypothetical but is equivalent to the real situation in terms of speed profile estimation.

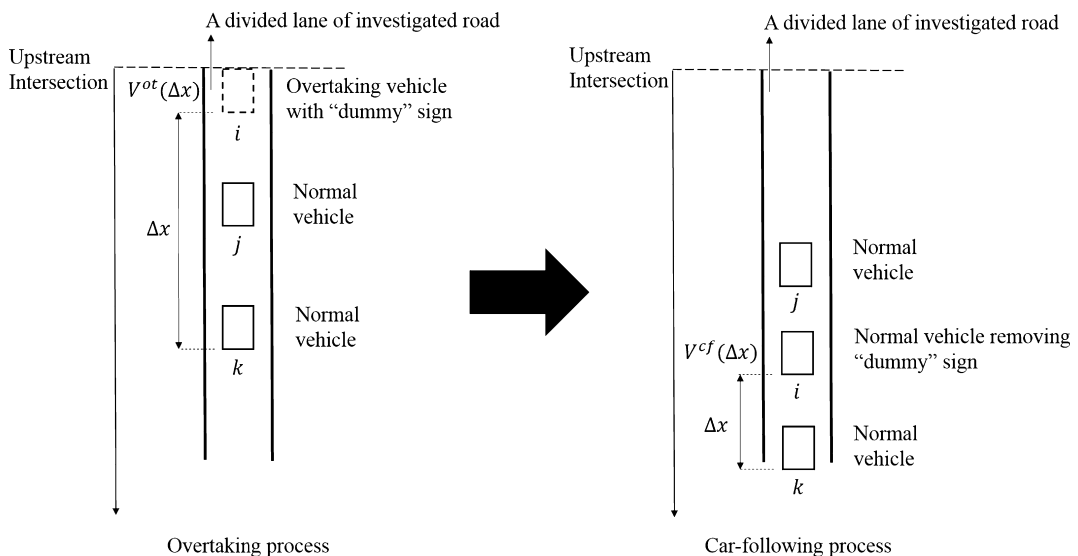


Fig. 8. Dummy-overtaking hypothesis.

The dummy-overtaking hypothesis is a simplification and is somewhat subjective. However, this assumption is adopted for following reasons. (1) Given the limitation of LPR data, no additional information is utilized to constrain the overtaking behavior perfectly. Thus, the simplification is necessary. (2) The simplest treatment is used to handle overtaking to avoid making many assumptions under limited data. (3) Boundary constraints are important bases for speed profile estimation, and the dummy-overtaking model can more easily satisfy the boundary constraints than the complex lane change behavior. (4) According to the real-world LPR data, the proportion of overtaking vehicles only accounts for approximately 10%. Thus, the dummy-overtaking hypothesis will insignificantly affect the speed profile estimation. (5) This study focuses on vehicle speed profile estimation with LPR data. The dummy-overtaking hypothesis can adapt to limited data information and lead to good results. Thus, the dummy-overtaking hypothesis is adopted even if it is unrealistic. (6) Several lane change models (e.g., probabilistic discretionary lane change model; Toledo et al., 2009) are also tested. However, embedding the lane change behavior causes severe violation of the boundary conditions and leads to inaccurate speed profile estimation.

Jiang's et al. (2001) car-following model has the following form:

$$a_n(t) = \kappa[V(\Delta x) - v_n(t)] + \lambda \Delta v, \tag{5}$$

where $a_n(t)$ and $v_n(t)$ are the acceleration and speed of vehicle n at time t . κ is a sensitivity constant, and V is the optimal velocity that drivers prefer. Δx is the headway, i.e., $\Delta x = x_{n-1}(t) - x_n(t)$, where x_n is the position of the vehicle n . $\Delta v = v_{n-1}(t) - v_n(t)$, and λ is the sensitivity defined in Eq. (6), where $b_2 = 0$ and $s_c = 120$ m are used in this study and b_1 is calibrated in Section 5.2.

$$\lambda = \begin{cases} b_1, & \Delta x < s_c \\ b_2, & \Delta x \geq s_c \end{cases}. \tag{6}$$

The optimal velocity function (Helbing and Tilch, 1998) is expressed as follows:

$$V(\Delta x) = V_1 + V_2 \tanh[C_1(\Delta x - l_c) - C_2], \tag{7}$$

where l_c is the length of the vehicles and can be taken as 5 m in simulations. $V_1, V_2, C_1,$ and C_2 are the parameters that need to be calibrated. The following two explicit Euler scheme equations are used to update the position of each vehicle and obtain the real-time vehicle speed:

$$v_n(t + \delta t) = v_n(t) + \delta t \times a_n(t), \tag{8}$$

$$x_n(t + \delta t) = x_n(t) + \delta t \times v_n(t), \tag{9}$$

where δt denotes the updating time step size. In actual implementation, δt is taken as 0.5 s. From the formulation above, the current vehicle speed at the next time step can be efficiently obtained by simply evaluating the headway and speed difference between the current vehicle and its predecessor.

As mentioned above, overtaking vehicles follow a fixed vehicle in a higher speed than normal vehicles owing to the dummy-overtaking hypothesis. Thus, two optimal velocity functions are adopted to describe overtaking and car-following behavior, respectively, i.e., $V^{ot}(\Delta x)$ and $V^{cf}(\Delta x)$. They have the same formation as Eq. (7); however, the values of $V_1, V_2, C_1,$ and C_2 are different. For distinction, we define $V_1^{ot}, V_2^{ot}, C_1^{ot},$ and C_2^{ot} as the parameters for V^{ot} , and $V_1^{cf}, V_2^{cf}, C_1^{cf}, C_2^{cf}$ as the parameters for V^{cf} .

The boundary conditions for each vehicle n include the exact arrival and departure times T_n^{arr}, T_n^{dep} as well as the cumulative index at the upstream and downstream $y_u(n), y_d(n)$. In particular, to ensure the customized car-following model is correctly specified, we introduce an initial arrival speed function v_n^{ini} based on the upstream cumulative index of vehicles as follows:

$$v_n^{ini} = \begin{cases} \alpha \times (y_u(n) - 1), & \text{if } \alpha \times (y_u(n) - 1) < v_{max}^{ini} \\ v_{max}^{ini}, & \text{otherwise} \end{cases}, \tag{10}$$

where α is a constant indicating the rate of speed increasing in arrival flow, and v_{max}^{ini} is the maximum arrival speed. A positive arrival acceleration a^{ini} is also introduced for all the vehicles. Notably, α and v_{max}^{ini} should be calibrated, and $a^{ini} = 1$ m/s² is used empirically. The numerical results show that a^{ini} has a minor effect on the results because it only influences the initial behavior of a vehicle when it enters the lane and will be offset by its interaction with other vehicles. The extra or less time spent on acceleration or deceleration at the initial stage is negligible compared with the total link travel time if the link is sufficiently long. The boundary conditions for each vehicle can be summarized as follows:

- Initial condition ($t = T_n^{arr}$)

$$a_n(t) = a^{ini}, v_n(t) = v_n^{ini}, x_n(t) = 0; \tag{11}$$

- Final condition ($t = T_n^{dep}$)

$$x_n(t) = L_R, \tag{12}$$

where L_R is the length of the investigated segment.

A pseudo vehicle remaining stationary at location $L_R + l_c$ is created to finish the simulation and incorporate the boundary constraints. When an actual vehicle predecessor leaves the investigated segment, it will follow the pseudo vehicle until the departure time to satisfy the boundary conditions. Thus, the pseudo vehicle can directly dominate the travel behavior of the first vehicle to ensure that the proposed car-following model corresponds well to the boundary conditions (Zhan et al., 2015). This pseudo vehicle-constrained method may be slightly arbitrary. However, this study adopts this method for the following reasons. (1) Given that only the boundary information in LPR data is employed, deriving a better method to incorporate the boundary conditions is difficult. (2) This pseudo vehicle-constrained method is proven effective for LPR data by existing literature. Zhan et al. (2015) used a similar method to incorporate the boundary conditions in the car-following model and obtained good results in queue length estimation.

The overall outline of the speed profile estimation procedure using the proposed vehicle travel simulation model is presented in Algorithm 3.

Algorithm 3 (*Speed profile estimation procedure*).

Step 1: Create a pseudo vehicle

Create a pseudo vehicle that remains stationary at location $L_R + l_c$. After creating a stationary pseudo vehicle, the first vehicle entering the investigated segment at T_{min}^{arr} can follow.

Step 2: Vehicle travel simulation.

Implement the proposed customized car-following model combined with boundary conditions and dummy-overtaking hypothesis to perform the simulation process.

Step 3: Vehicle-based speed estimation.

Obtain vehicle speed from step 3 using Eq. (8) and plot the speed profile of vehicles.

4. Field experiment

The proposed speed profile estimation method was tested using real-world LPR data and validated using ground truth vehicle speed measured by GPS devices fixed on two probe vehicles. The field experiment was designed in a unidirectional segment of Heping Road in Langfang City, Hebei Province, China (Fig. 9). The total length of the road L_R was 720 m. The upstream (at XiangYun Street) and downstream intersection (at BeiFeng Street) were equipped with four LPR cameras for each approach, and the LPR data were obtained from the local traffic management agency. These data contained the following information for each passing vehicle: (1) an identified license plate number, (2) the timestamp when a vehicle passes the stop line, (3) approach of intersection, and (4) the lane from which the vehicle exits. The upstream and downstream intersections shared the same lane configurations, from which the leftmost to rightmost lane were dedicated left turning lane (*Lane 1*), through lane (*Lane 2*), and through and right turning lane (*Lane 3*). The two intersections adopted fixed timing plans with the same cycle length of 120 s (7:00–9:00) and 100 s (9:00–16:30). However, different phase intervals were used for the upstream and downstream intersections, and these intervals were also recorded in the field experiment.

Our field experiment was conducted from 9:30 to 16:30 on March 10, 2016. Two probe vehicles equipped with GPS devices repeatedly passed across the investigated segment many times, and the ground truth GPS trajectories were recorded once per second. Considering the unknown malfunction of the LPR system between 10:00 and 13:00 (Fig. 2), the LPR data between 14:30 and 16:30 were used to perform the validation by comparing the ground truth speed with the estimated speed. Unrecognized and erroneously recognized LPR records existed; thus, only 17 ground truth GPS trajectories that traverse the testing road segment were used for comparison. The poor performance of the LPR system indicates that inferring the missing information from the unmatched arrival vehicles is critical in this study.

5. Numerical results

The real-world LPR data during the field experiment are used to test the model. We first reconstruct the cumulative arrival/departure curve and obtain the cumulative indices and arrival time of all vehicles. Then, we use parts of ground truth vehicle speed against the estimation ones to calibrate our customized car-following model by genetic algorithm (GA) (Holland, 1975). The speed profile of the estimation results are plotted in Fig. 12 and are compared with the ground truth vehicle speed and the state-dependent naive speed. A complementary model validation is also conducted using the NGSIM Peachtree Street dataset (US Department of Transportation–FHWA, 2008) to prove the robustness of the proposed model. Finally, the proposed LPR-based model is applied to estimate emissions with CMEM (Barth et al., 2001).

5.1. Cumulative arrival/departure curve reconstruction result

Compared with previous study (Zhan et al., 2015), the amount of sample input in the Gaussian interpolation model improves nearly 30% of all cycles because the information of inconsistently matched vehicles is retained. Thus, the recon-



Fig. 9. Field experiment design.

struction accuracy can be promoted with considerable priori information. Fig. 10 shows the reconstruction and matching results of Lane 3. The blue curve and dots represent the departure curve and departing vehicles. The red curve and dots represent the reconstructed arrival curve and arriving vehicles. The green line represents the estimated posterior mean of the cumulative arrival curve. The cyan-blue line represents the matching edge of two vehicles. The points with crossing cyan-blue line represent the overtaking vehicles.

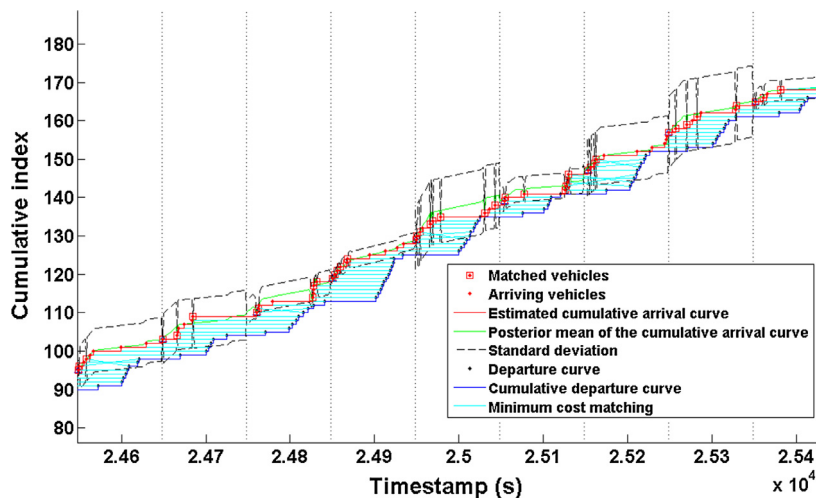


Fig. 10. Reconstructed cumulative arrival/departure curve and matching results of Lane 3.

5.2. Calibration of the customized car-following model

With the addition of the dummy-overtaking hypothesis and change in the traffic conditions, the proposed customized car-following model should be calibrated prior to simulation. We use the ground truth speed data of vehicles in March 10 from 14:30 to 15:30 for calibration. Loss function is defined as follows:

$$Q(\phi) = \sum_{i=1}^{N_c} RMSE_i, \quad (13)$$

$$RMSE_i = \sqrt{\frac{\sum_{t=0}^{TT_i} (v_i(t, \phi) - v_i^a(t))^2}{TT_i}}, \quad (14)$$

where $RMSE_i$ is the root-mean-square error (RMSE) of the estimated speed of vehicle i , N_c is the number of vehicles used for calibration, and $v_i(t, \phi)$ is the estimated speed of vehicle i in time t given ϕ ($\phi = (b_1, \alpha, v_{\max}^{\text{ini}}, V_1^{\text{ot}}, V_2^{\text{ot}}, C_1^{\text{ot}}, C_2^{\text{ot}}, V_1^{\text{cf}}, V_2^{\text{cf}}, C_1^{\text{cf}}, C_2^{\text{cf}}, \kappa)$). $v_i^a(t)$ is the actual speed of vehicle i in time t , and TT_i is the total travel time of vehicle i . ϕ can thus be calibrated by optimizing the minimum $Q(\phi^*)$. GA (Holland, 1975) is used for optimization. The final calibration results are shown in Table 2, and they indicate that the optimal velocity of overtaking vehicles is higher than that of normal vehicles.

5.3. Vehicle speed profile estimation results

A state-dependent naive approach is used as benchmark to demonstrate the effectiveness of the proposed approach. This naive approach is similar to the modal decomposition trajectory estimation method used by Yang et al. (2011). The traveling process is explicitly divided into four states (Yang et al., 2011), namely, acceleration, cruise, deceleration, and idle modes. A typical state-dependent speed profile curve is shown in Fig. 11, where TT is the total travel time. An acceleration state (from 0 to $p_1 \times TT$) is observed as the vehicle enters from the upstream intersection at initial speed $v_{\text{ini}}^{\text{naive}}$. This state is followed by a cruise state (from $p_1 \times TT$ to $p_2 \times TT$), in which the vehicle travels at free-flow speed $v_{\text{naive}}^{\text{ini}} + p_1 \times TT \times a_1^{\text{naive}}$ (where a_1^{naive} is the acceleration of the first acceleration state). The vehicle maintains this speed until it reaches the next traffic signal or queue of traffic. A deceleration mode (from $p_2 \times TT$ to $p_3 \times TT$) occurs, followed by an idle mode (from $p_3 \times TT$ to $p_4 \times TT$) until the traffic signal turns green or the queue disperses. Then, the vehicle accelerates to depart from the downstream intersection with the acceleration of a_2^{naive} (from $p_4 \times TT$ to TT). Thus, the vehicle speed in time t (i.e., $V_{\text{naive}}(t)$) can be expressed as the piecewise function in Eq. (15):

$$V_{\text{naive}}(t) = \begin{cases} v_{\text{ini}}^{\text{naive}} + t \times a_1^{\text{naive}}, & 0 \leq t \leq p_1 \times TT \\ V_{\text{naive}}(p_1 \times TT), & p_1 \times TT < t \leq p_2 \times TT \\ V_{\text{naive}}(p_2 \times TT) \times \left(1 - \frac{(t - p_2 \times TT)}{(p_3 - p_2) \times TT}\right), & p_2 \times TT < t \leq p_3 \times TT \\ 0, & p_3 \times TT < t \leq p_4 \times TT \\ (t - p_4 \times TT) \times a_2^{\text{naive}}, & p_4 \times TT < t \leq TT \end{cases} \quad (15)$$

The parameters of $v_{\text{ini}}^{\text{naive}}$, p_1 , p_2 , p_3 , p_4 , a_1^{naive} , and a_2^{naive} are calibrated using the ground truth speed profile of all sample vehicles. The calibration method is similar to that presented in Section 5.2.

The estimation results using real-world LPR data from 14:30 to 16:30 in March 10, 2016, Langfang, China, are plotted in Fig. 12 (speed profile resolution = 0.5 s). The naive speed profile and ground truth speed profile of all sample vehicles are compared. The RMSE and mean absolute error (MAE) of each sample are computed to evaluate the accuracy of the model. The $RMSE_i$ and MAE_i of vehicle i are defined in Eqs. (14) and (16), respectively. Notably, $RMSE1$ and $MAE1$ are calculated for the LPR data-based speed profile estimation model, whereas $RMSE2$ and $MAE2$ are calculated for the naive approach. The MRE of the entire model is defined in Eq. (17), where V_{mp} is the average speed at road midpoint (17.82 m/s) calculated from the sample vehicle trajectories and N_s is the number of sample vehicles ($N_s = 17$).

$$MAE_i = \frac{\sum_{t=0}^{TT_i} |v_i(t, \phi) - v_i^a(t)|}{TT_i} \quad (16)$$

$$MRE = \frac{\sum_{i=1}^{N_s} \frac{MAE_i}{V_{mp}}}{N_s} \quad (17)$$

Table 2
Parameter calibration results.

Parameters (unit)	Value	Parameters (unit)	Value
b_1 (m^{-1})	0.203	α (m/s)	2.816
v_{max}^{ini} (m/s)	11.548	V_1^{df} (m/s)	8.514
V_1^{ot} (m/s)	12.528	V_2^{df} (m/s)	7.912
V_2^{ot} (m/s)	8.412	C_1^{df} (1)	0.122
C_1^{ot} (1)	0.131	C_2^{df} (1)	1.577
C_2^{ot} (1)	1.443	κ (1)	0.142

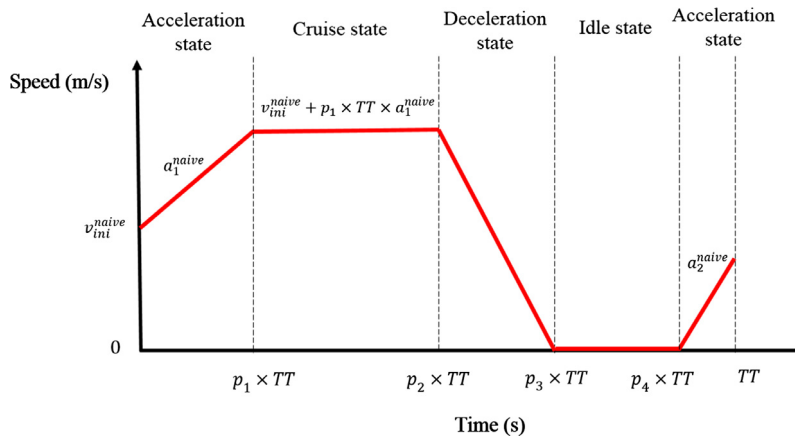


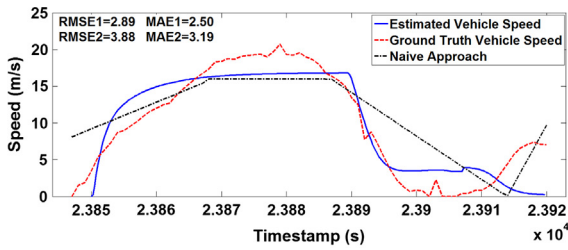
Fig. 11. Typical state-dependent naïve speed profile curve.

The figure reveals that the estimated vehicle speed curves (blue line) capture the pattern of ground truth vehicle speed curve (red dashed line) well. The RMSE1 and MAE1 of the 17 samples are controlled below 3.945 and 3.285 m/s, respectively. In addition, the average values of RMSE1 and MAE1 are 2.896 and 2.313 m/s, respectively, and the MRE is 12.98%, which is generally better than that of the naïve approach (the average values of RMSE2 and MAE2 are 4.183 and 3.351 m/s, respectively, MRE = 18.7%).

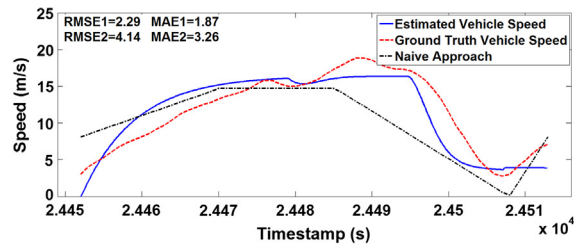
At the different scenarios of the speed profiles, the proposed model achieves high accuracy in vehicles with obvious acceleration, constant speed, and deceleration stages (e.g., Samples 1, 2, and 14) but performs poorly in vehicles with gradual speed change (e.g., Samples 6, 12, and 13). Besides, the proposed model has difficulty in capturing the peak of the ground truth vehicle speed curve (e.g., Sample 8). And it may cause the rapid deceleration in the deceleration phase (e.g., Samples 3, 11, and 13). The inconsistent performance of our model may be due to the inherent attributes of the car-following model. Jiang’s car-following model assumes that vehicles tend to drive at a desired and constant speed; however, this assumption does not fully capture the complex and nonhomogeneous interactions among vehicles as well as some complicated driving behavior, such as vehicle peak speed and sudden speed change. In addition, the minimum cost perfect bipartite matching is based on the hypothesis that vehicles are likely to keep the same index in upstream and downstream. Thus, the estimated arrival indices may not be the real one such that the overtaking behavior of some vehicles may be neglected, and their speed may be underestimated. Notably, Sample 4 is an overtaking vehicle and its ground truth speed curve peak is successfully captured because of the overtaking hypothesis. This finding helps improve the proposed model by considering nonhomogeneous and complicated driving behavior in the future.

5.4. Complementary model validation with NGSIM data

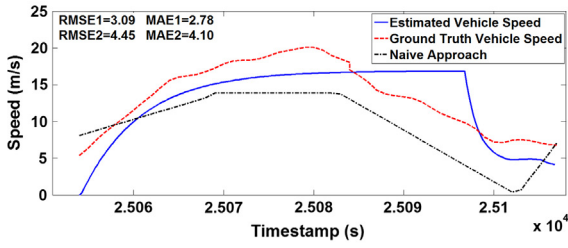
Given the small size of sample vehicles in the LPR data, a complementary model validation is conducted using the NGSIM Peachtree Street dataset (US Department of Transportation—FHWA, 2008) to prove the robustness of the proposed speed profile estimation model. The southbound road of Peachtree Street Section 2 is selected to simulate the road pattern with LPR system for our study because it is a closed road (satisfying flow conservation) with traffic lights. The total length of Section 2 is 150 m. The schematic of the investigated area is plotted in Fig. 13. Then, the arrival and departure times of 209 vehicles from 15 min NGSIM trajectory data (from 4:00 pm to 4:15 pm) are extracted. Unmatched vehicles at a proportion of 35% (same as that of the real-world LPR data) are randomly selected to simulate the form of the LPR data. Then, the cumulative arrival/departure curve is reconstructed with the proposed Gassing interpolation model and minimum cost bipartite matching method. The reconstructed curve of the left-most lane is shown in Fig. 14.



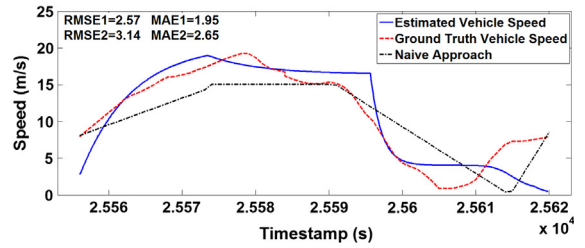
(1) Sample 1 (Lane 2, Travel Time = 72 s)



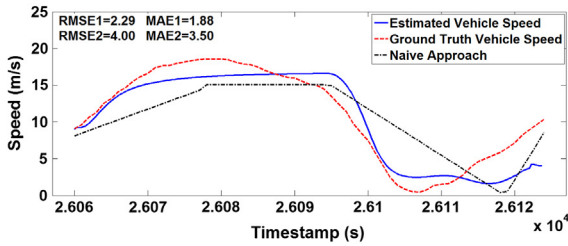
(2) Sample 2 (Lane 3, Travel Time = 62 s)



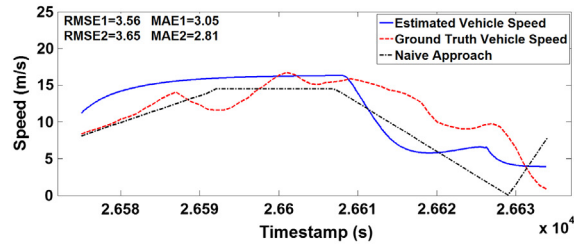
(3) Sample 3 (Lane 3, Travel Time = 54 s)



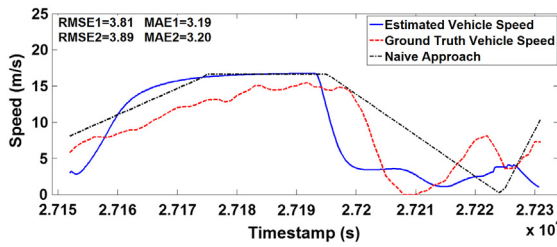
(4) Sample 4 (Lane 3, Travel Time = 54 s)



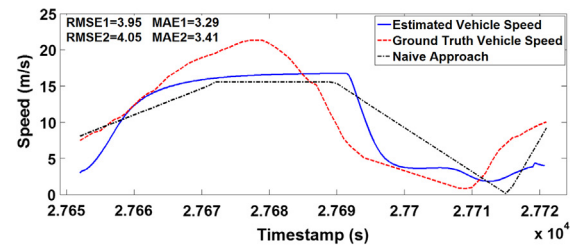
(5) Sample 5 (Lane 2, Travel Time = 65 s)



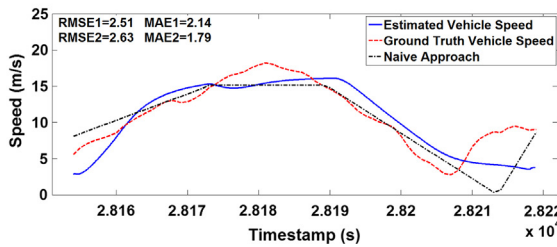
(6) Sample 6 (Lane 3, Travel Time = 65 s)



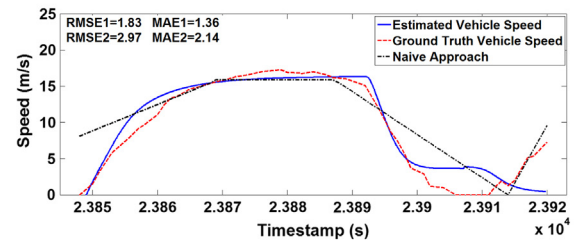
(7) Sample 7 (Lane 2, Travel Time = 60 s)



(8) Sample 8 (Lane 3, Travel Time = 80 s)



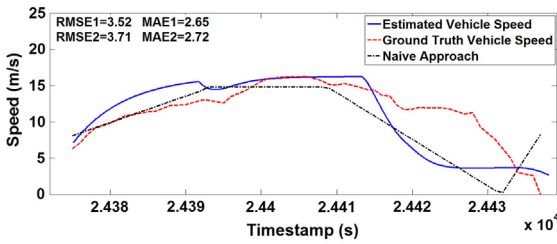
(9) Sample 9 (Lane 3, Travel Time = 66 s)



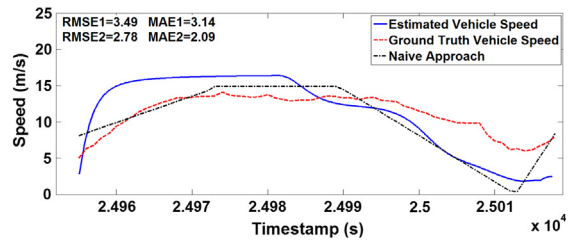
(10) Sample 10 (Lane 2, Travel Time = 51 s)

Fig. 12. Speed estimation results compared with ground truth data and naive approach.

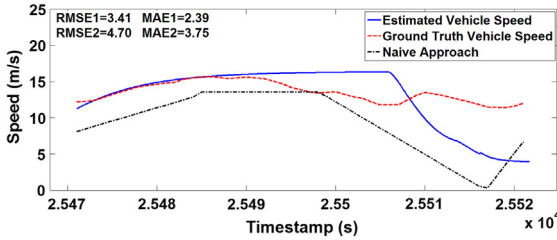
Given that the traffic conditions in USA are different from that in China, the parameters are recalibrated using the same method presented in Section 5.2. Then, the arrival and departure times are applied in the proposed speed profile estimation model. The NGSIM data contain the ground truth trajectory of all 209 vehicles. Therefore, a wide range of model validation



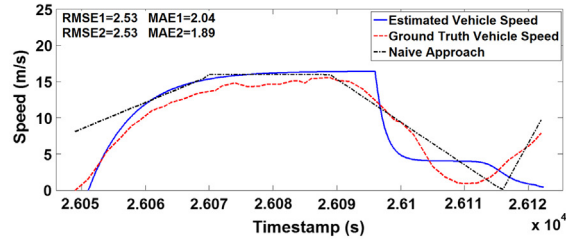
(11) Sample 11 (Lane 1, Travel Time = 48 s)



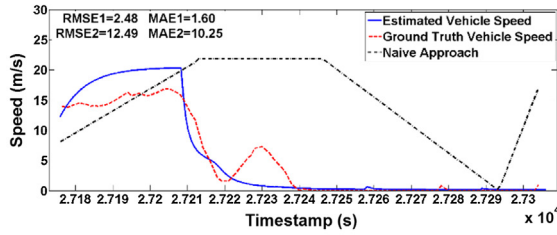
(12) Sample 12 (Lane 2, Travel Time = 59 s)



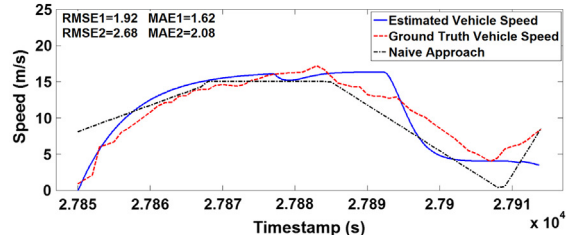
(13) Sample 13 (Lane 3, Travel Time = 41 s)



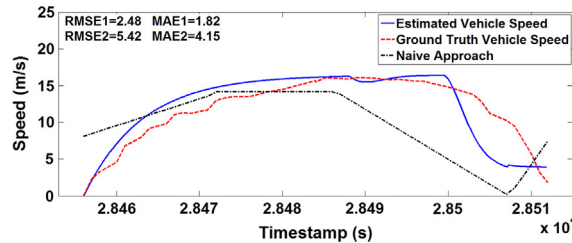
(14) Sample 14 (Lane 2, Travel Time = 50 s)



(15) Sample 15 (Lane 2, Travel Time = 94 s)



(16) Sample 16 (Lane 2, Travel Time = 55 s)



(17) Sample 17 (Lane 3, Travel Time = 44 s)

Fig. 12 (continued)

can be conducted. The distribution of the estimation RMSE and MAE is shown in Fig. 15. The average values of RMSE and MAE are 2.44 and 1.73 m/s, respectively. Considering that the average speed at the midpoint of this road section is 10.50 m/s, the MRE is 16.48%, which is similar to that of the real-world LPR data (12.98%). Therefore, the proposed speed estimation model is robust.

5.5. Model application in emission calculation

As mentioned previously, the estimated speed profile can be directly used in the micro-emission model. Therefore, the CMEM (Barth et al., 2001) is used to validate the application of the proposed model in emission estimation with the real-world LPR data. Acceleration information is also important for emission calculation; thus, the estimated acceleration profile is extracted from Eq. (5). The ground truth acceleration of probe vehicles is calculated through the slope of the speed curve. The RMSE and MAE are used to quantify the estimated error of acceleration. For simplification, all vehicles are assumed to belong to type 6 (three-way catalyst, fuel-injected car, accumulated miles <50,000, low power/weight) defined by Barth et al. (2001). Then, the emission of all vehicles can be determined using the second-by-second speed and acceleration data. The

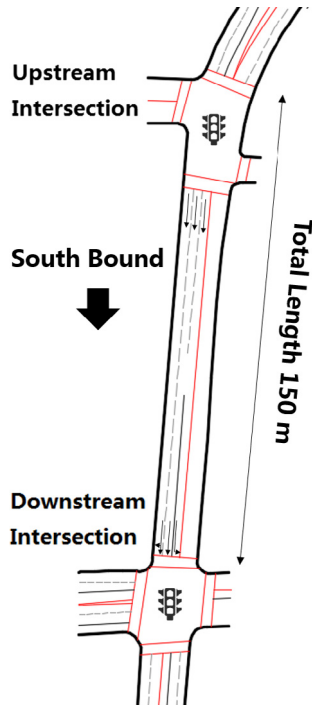


Fig. 13. Schematic of Peachtree Street Section 2.

estimation results are compared against the state-dependent naïve approach (benchmark) and ground truth data (notably, the ground truth emission is calculated by inputting ground truth speed and acceleration data in CMEM).

The acceleration and emission estimation results of Samples 1, 6, 8, and 11 are shown in Fig. 16. These sample vehicles correspond to the different scenarios of speed profiles mentioned in Section 5.3 and are thus used in the discussion. The proposed model can barely capture the fluctuation of acceleration because the car-following model assumes a smooth acceleration change. From the estimated emissions, relatively good results of the emission calculation of Sample 1 (vehicle with obvious acceleration, constant speed, and deceleration stages) can be obtained, which indicates that speed profiles and emissions of this category of vehicles can be estimated. Moreover, the emission results of Sample 8 (vehicle with a peak in the speed curve) are reasonably estimated. This finding indicates that, even if the peak of the ground truth vehicle speed curve cannot be captured, it has a minor effect on emission estimation. However, the emissions of HC of Sample 6 (vehicle with

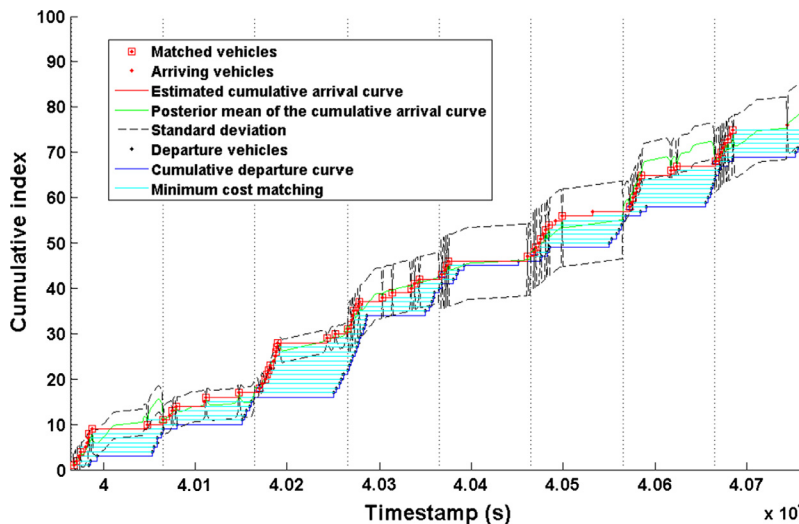


Fig. 14. Reconstructed cumulative arrival/departure curve of the left-most lane in Peachtree Street.

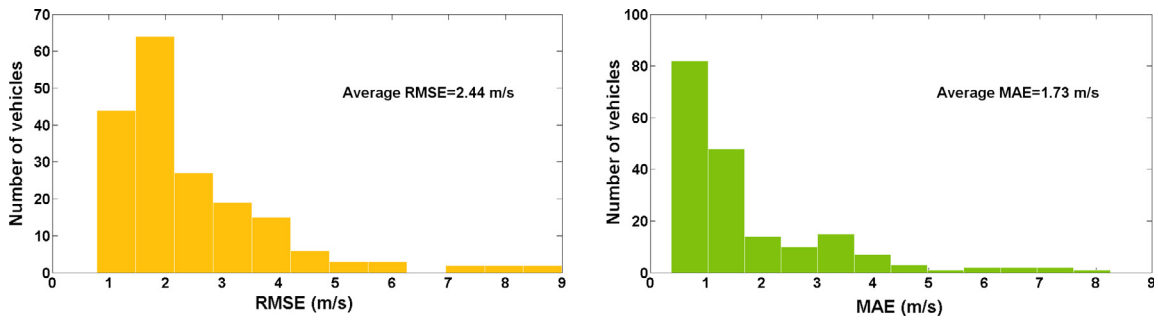


Fig. 15. Distribution of the estimation RMSE and MAE of the NGSIM data.

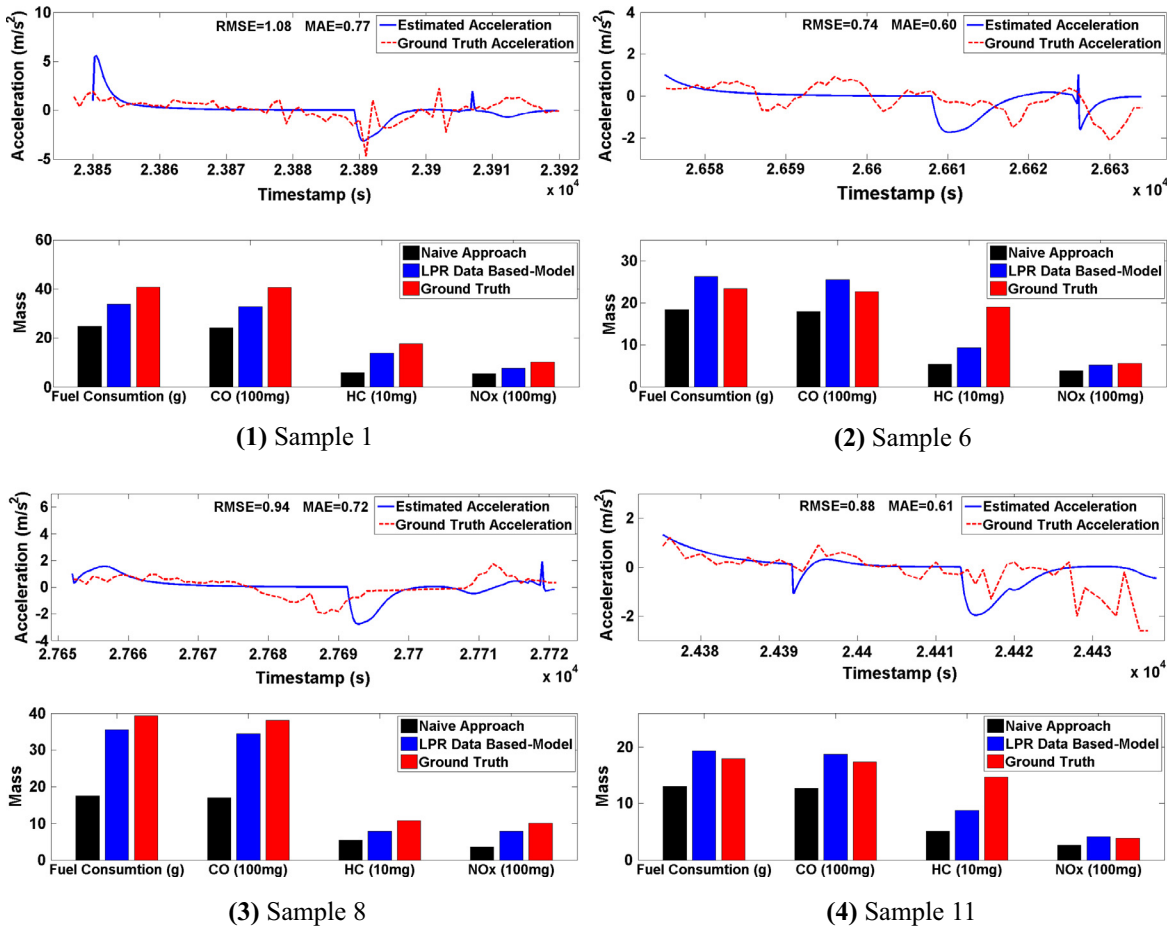


Fig. 16. Acceleration and emission estimation results of Sample 1, 6, 8, and 11.

gradual speed change) and Sample 11 (vehicle with rapid deceleration) are underestimated. These findings may be attributed to underestimation of the deceleration duration. According to An et al. (1999), HC emissions associated with deceleration episodes are significant. Rapid deceleration events result in speed change in a short time (i.e., underestimated the deceleration duration) and thus cause underestimation of HC emissions.

The overall estimation results for the sample vehicles are listed in Table 3. The average values of acceleration estimation RMSE and MAE are 0.79 and 0.54 m/s^2 , respectively, and the MRE is 12.98%. The average relative estimation errors of fuel consumption, CO, HC, and NO_x are 15.52%, 15.26%, 35.22%, and 17.34%, respectively, which are better than those of the naïve method (i.e., 42.06%, 42.16%, 55.03%, and 49.43%, respectively). In general, the LPR data-based model provides reasonable emission estimation in the individual vehicle level. The estimation of fuel consumption, CO, and NO_x is more accurate than

Table 3
Estimation results of acceleration and emission.

Samples	Acceleration (m/s ²)		Fuel consumption (g)			CO (100 mg)			HC (10 mg)			NO _x (g)		
	RMSE	MAE	Estimated	Ground truth	Relative error (%)	Estimated	Ground truth	Relative error (%)	Estimated	Ground truth	Relative error (%)	Estimated	Ground truth	Relative error (%)
1	1.08	0.77	33.79	40.69	-16.95	32.77	40.55	-19.19	13.80	17.66	-21.89	7.55	10.14	-25.52
2	0.62	0.49	30.81	33.72	-8.62	29.89	32.71	-8.62	12.66	9.41	34.51	7.01	8.05	-12.99
3	1.03	0.55	33.94	29.20	16.25	32.92	28.32	16.25	6.86	13.33	-48.55	7.71	7.58	1.80
4	0.90	0.59	22.49	33.36	-32.60	21.81	32.36	-32.60	10.67	10.35	3.08	6.21	7.72	-19.52
5	0.50	0.38	30.96	31.88	-2.87	30.03	30.92	-2.87	10.54	7.53	39.98	6.43	7.60	-15.36
6	0.74	0.60	26.28	23.38	12.43	25.49	22.68	12.43	9.28	18.98	-51.09	5.17	5.59	-7.53
7	1.16	0.82	38.93	35.40	9.96	37.76	34.34	9.96	13.93	21.42	-34.94	8.15	7.61	7.04
8	0.94	0.72	35.47	39.25	-9.64	34.40	38.08	-9.64	7.81	10.68	-26.90	7.85	10.05	-21.91
9	0.58	0.45	26.86	31.51	-14.76	26.05	30.57	-14.76	8.46	10.28	-17.69	6.11	7.57	-19.35
10	0.66	0.43	23.11	23.38	-1.16	22.42	22.68	-1.16	5.78	13.85	-58.29	5.18	5.59	-7.37
11	0.88	0.61	19.35	17.93	7.93	18.77	17.39	7.93	8.77	14.72	-40.44	4.14	3.84	7.88
12	0.68	0.43	22.76	24.68	-7.78	24.21	23.94	1.14	7.08	15.91	-55.49	5.15	5.11	0.82
13	0.65	0.38	17.60	14.42	22.04	17.07	13.99	22.04	6.57	11.39	-42.34	3.61	2.99	20.73
14	0.77	0.49	21.94	25.24	-13.07	21.28	24.48	-13.07	8.73	7.80	12.02	4.86	5.52	-12.02
15	0.66	0.33	23.08	14.78	56.11	22.39	14.34	56.11	10.45	18.03	-42.01	5.79	3.12	85.58
16	0.72	0.53	28.27	30.10	-6.10	27.42	29.20	-6.10	11.47	9.53	20.43	6.40	6.75	-5.14
17	0.82	0.56	25.09	19.98	25.57	24.34	19.38	25.57	7.03	13.84	-49.20	5.68	4.57	24.18
Average absolute value of error	0.79	0.54			15.52			15.26			35.22			17.34

the estimation of HC. This observation can be explained by the proposed model and existing literature. According to [Ross and An \(1993\)](#), fuel consumption is directly associated with typical driving characteristics (e.g., average speed and attempted maximum speed). Therefore, fuel consumption can be reasonably estimated in all conditions because the estimated speed curve can capture the pattern of ground truth vehicle speed curve. In addition, the NO_x emission rate is insensitive to the level of acceleration ([Rakha and Ding, 2003](#)). Thus, the NO_x emission is also reasonably estimated. For CO, given that the rates of CO emissions are closely related to fuel use ([An and Ross, 1996](#)), reasonable estimation results of CO can also be obtained. However, for HC emissions, given that the said value is highly sensitive to the level of acceleration at high speeds ([Ahn et al., 2002](#); [Bokare and Maurya, 2013](#)) and the deceleration events significantly contribute to this value ([An et al., 1999](#)), it is poorly estimated because the proposed model fails to capture the fluctuation of acceleration at high speeds and underestimates the deceleration duration.

6. Conclusion

This study proposes a vehicle speed profile estimation method based on simulation using LPR data in lane level. A LPR data-mending method is proposed to infer the information of unmatched vehicles prior to simulation. Then, the complete arrival and departure information is used as the boundary conditions in a customized car-following model, which adds dummy-overtaking hypothesis and boundary constraints in estimating the speed profile of vehicles. A field experiment is conducted in March 10, 2016 in Langfang, Hebei Province, China. The LPR data during the experiment are used to test the validation. The estimation results show that our model can fully capture the variation patterns of ground truth data. The average values of RMSE and MAE are 2.896 and 2.313 m/s, respectively, and the MRE is 12.98%. A complementary model validation is conducted using the NGSIM Peachtree Street dataset, and shows that the proposed model is robust. The estimated speed profile is applied to estimate emissions with CMEM. The estimation errors of fuel consumption, CO emissions, and NO_x emissions are approximately 16% in the individual vehicle level. The numerical results demonstrate the effectiveness of the proposed model in estimating speed profile and emission. As a result, considerable information support can be provided for traffic management and control using the proposed model.

This study has several limitations, which can be further improved in future works. (1) We apply Oliveira-Neto's condition probability $p(u_i|d_i)$ into the matching procedure directly (Section 2.2) because of the lack of training samples on recognition errors. As a result, only 3% of the proportion of matched vehicles is enhanced. In actual application, $p(u_i|d_i)$ needs to be calibrated using new ground truth samples of recognition errors. (2) The dummy-overtaking hypothesis and pseudo vehicle boundary conditions have several drawbacks, as mentioned in Section 3. (3) The real-world traffic situation is more complex than our model; thus, our model has difficulty in estimating certain speed variations of complicated travel behavior.

The application prospects of this LPR data-based model should be discussed. As mentioned previously, the proposed model can adapt to any kind of data with the same characteristics as the LPR data (e.g., RFID data). With the development of the vehicle-to-infrastructure cooperation system in China, the RFID facilities are expected to be implemented nationwide ([Ministry of Transport of China, 2016](#)) and thus provide considerable prospect for the application of the proposed model. On the other hand, other direct techniques for speed profile measurement (e.g., wide-scale GPS) may become available in a few years. However, given the difficulty in gathering and processing the decentralized GPS data, obtaining the complete trajectory information for each vehicle is still difficult ([Ahmed et al., 2017](#)). The combination of wide-scale GPS data and LPR data will also enable elaborate simulation of the total traffic flow.

From a technological perspective, (1) future works can apply a realistic simulation model to capture complicated travel behavior with additional data (e.g., GPS data), particularly considering the complex and nonhomogeneous interactions among vehicles (e.g., lane change behavior). (2) We propose an arrival time estimation model for LPR data correction; thus, additional traffic information can be obtained from the LPR data. (3) We select only a single investigated road as a case study, but the current method can be extended to a network-wide traffic condition.

Acknowledgment

This research was supported by the National Natural Science Foundation of China (71361130015) and Beijing Natural Science Foundation (8162024). The authors thank the Bureau of Traffic Management of Langfang for providing the data, and Mr. Zhiyong Liu and Mr. Xiaolong Liu from Tsinghua University, China, for helping with the field experiment.

References

- Ahmed, U., Sahin, O., Cetin, M., 2017. Minimizing GPS Dependency for a Vehicle's Trajectory Identification by Utilizing Data from Smartphone Inertial Sensors and OBD Device. In: *Transportation Research Board 96th Annual Meeting*.
- Ahn, K., Rakha, H., Trani, A., Van Aerde, M., 2002. Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels. *J. Transport. Eng.* 128 (2), 182–190.
- Ahuja, R.K., Magnanti, T.L., Orlin, J.B., 1993. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall.
- An, F., Barth, M., Scora, G., Ross, M., 1999. Modal-based intermediate soak-time emissions modeling. *Transport. Res. Rec.: J. Transport. Res. Board* 1664 (1), 58–67.
- An, F., Ross, M., 1996. A simple physical model for high power enrichment emissions. *J. Air Waste Manag. Assoc.* 46 (3), 216–223.
- Bando, M., Hasebe, K., Nakayama, A., Shibata, A., Sugiyama, Y., 1995. Dynamical model of traffic congestion and numerical simulation. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids*, 51 (2), 1035.

- Barth, M., An, F., Younglove, T., Scora, G., Levine, C., Ross, M., Wenzel, T., 2001. Comprehensive modal emissions model (CMEM), version 2.02. User's guide. <http://www.cert.ucr.edu/cmем/docs/CMEM_User_Guide_v3.01d.pdf> (accessed 05.05.17).
- Barth, M., Boriboonsomsin, K., 2008. Real-World CO₂ Impacts of Traffic Congestion, Transportation Research Board Conference, pp. 036807–036807.
- Barth, M., Scora, G., Younglove, T., 1999. Estimating emissions and fuel consumption for different levels of freeway congestion. *Transport. Res. Rec.: J. Transport. Res. Board* 1664, 47–57.
- Beijing Municipal Committee, 2016. The Thirteenth Five-Year Plan for the Transportation Development of Beijing <http://www.bjijtw.gov.cn/xxgk/ghjh/gh/201607/t20160704_131635.html>, (accessed 05.05.17).
- Beijing Traffic Management Bureau, 2017. Location Table of Fixed Traffic Monitoring Equipment <<http://www.bjijgl.gov.cn/zhuanti/09sxt/040.html>>, (accessed 05.05.17).
- Bertini, R.L., Lasky, M., Monsere, C.M., 2005. Validating predicted rural corridor travel times from an automated license plate recognition system: Oregon's frontier project. *Proceedings. In: Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE. IEEE*, pp. 296–301.
- Bishop, C., 2006. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer.
- Bokare, P.S., Maurya, A.K., 2013. Study of effect of speed, acceleration and deceleration of small petrol car on its tail pipe emission. *Int. J. Traffic Transport Eng.* 3 (4), 465–478.
- Evans, L., Herman, R., 1978. Automobile fuel economy on fixed urban driving schedules. *Transport. Sci.* 12 (2), 137–152.
- Hao, P., Ban, X., Yu, J.W., 2015. Kinematic equation-based vehicle queue location estimation method for signalized intersections using mobile sensor data. *J. Intell. Transport. Syst. Technol. Plan. Operat.* 19 (3), 256–272.
- Hao, P., Sun, Z., Ban, X., Guo, D., Ji, Q., 2013. Vehicle index estimation for signalized intersections using sample travel times. *Transport. Res. Part C Emerg. Technol.* 36 (11), 513–529.
- Helbing, D., Tilch, B., 1998. Generalized force model of traffic dynamics. *Phys. Rev. E* 58 (1), 133.
- Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence*. U Michigan Press.
- Jiang, R., Wu, Q., Zhu, Z., 2001. Full velocity difference model for a car-following theory. *Phys. Rev. E* 64 (1), 017101.
- Kuhn, H.W., 1955. The Hungarian method for the assignment problem. *Naval Res. Logist. Quart.* 2 (1–2), 83–97.
- Ministry of Transport of China, 2016. The Thirteenth Five-Year Plan for Transportation Science and technology development plan of China <http://www.moc.gov.cn/zhuanti/shisanwujitysfzgh/guihua-wenjian/201702/t20170213_2163741.html>, (accessed 05.05.17).
- Nakanishi, Y., Western, J., 2005. Ensuring the security of transportation facilities: evaluation of advanced vehicle identification technologies. *Transport. Res. Rec.: J. Transport. Res. Board* 1938, 9–16.
- Nesamani, K., Chu, L., McNally, M.G., Jayakrishnan, R., 2007. Estimation of vehicular emissions by capturing traffic variations. *Atmos. Environ.* 41 (14), 2996–3008.
- Oliveira-Neto, F.M., Han, L.D., Jeong, M.K., 2012. Online license plate matching procedures using license-plate recognition machines and new weighted edit distance. *Transport. Res. Part C: Emerg. Technol.* 21 (1), 306–320.
- Powell, M.J.D., Powell, M.J.D., Powell, M.J.D., Mathématicien, G.-B., Powell, M.J.D., Mathematician, G.B., 1981. *Approximation Theory and Methods*. Cambridge University Press, Cambridge.
- Rakha, H., Ding, Y., 2003. Impact of stops on vehicle fuel consumption and emissions. *J. Transport. Eng.* 129 (1), 23–32.
- Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., Aigrain, S., 2013. Gaussian processes for time-series modelling. *Phil. Trans. Roy. Soc. A* 371 (1984), 20110550.
- Ross, M., An, F., 1993. *The Use of Fuel by Spark Ignition Engines*. SAE Technical Paper.
- Rossetti, M.D., Baker, J., 2001. Applications and evaluation of automated license plate reading systems. In: *Proceedings of the Intelligent Transportation Society of America 11th Annual Meeting and Exposition*.
- Shang, J., Zheng, Y., Tong, W., Chang, E., Yu, Y., 2014. Inferring gas consumption and pollution emission of vehicles throughout a city. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 1027–1036.
- Sun, Z., Ban, X.J., 2013. Vehicle trajectory reconstruction for signalized intersections using mobile traffic sensors. *Transport. Res. Part C: Emerg. Technol.* 36, 268–283.
- Sun, Z., Hao, P., Ban, X.J., Yang, D., 2015. Trajectory-based vehicle energy/emissions estimation for signalized arterials using mobile sensing data. *Transport. Res. Part D: Transport Environ.* 34, 27–40.
- Toledo, T., Koutsopoulos, H.N., Ben-Akiva, M., 2009. Estimation of an integrated driving behavior model. *Transport. Res. Part C Emerg. Technol.* 17 (4), 365–380.
- US Department of Transportation—FHWA, 2008. NGSIM: Next Generation Simulation <<https://ops.fhwa.dot.gov/trafficanalysis/tools/ngsim.htm>>, (accessed 07.05.17).
- Yang, Q., Boriboonsomsin, K., Barth, M., 2011. Arterial roadway energy/emissions estimation using modal-based trajectory reconstruction. In: *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on. IEEE*, pp. 809–814.
- Yasin, A.M., Karim, M.R., Abdullah, A.S., 2009. Travel time measurement in real-time using automatic number plate recognition for Malaysian environment. In: *Proceedings of the Eastern Asia Society for Transportation Studies. Eastern Asia Society for Transportation Studies*, pp. 324–324.
- Zegeye, S.K., Schutter, B.D., Hellendoorn, J., Breunese, E.A., Hegyi, A., 2013. Integrated macroscopic traffic flow, emission, and fuel consumption model for control purposes. *Transport. Res. Part C: Emerg. Technol.* 31 (2), 158–171.
- Zhan, X., Li, R., Ukkusuri, S.V., 2015. Lane-based real-time queue length estimation using license plate recognition data. *Transport. Res. Part C: Emerg. Technol.* 57, 85–102.