

# Queue Profile Estimation in Congested Urban Networks with Probe Data

Mohsen Ramezani & Nikolas Geroliminis\*

Urban Transport Systems Lab. (LUTS), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

**Abstract:** *Queues at signalized intersections are the main cause of traffic delays and travel time variability in urban networks. In this article, we propose a method to estimate queue profiles that are traffic shockwave polygons in the time-space plane describing the spatiotemporal formation and dissipation of queues. The method integrates the collective effect of dispersed probe vehicle data with traffic flow shockwave analysis and data mining techniques. The proposed queue profile estimation method requires position and velocity data of probe vehicles; however, any explicit information of signal settings and arrival distribution is indispensable. Moreover, the method captures interdependencies in queue evolutions of successive intersections. The significance of the proposed method is that it is applicable in oversaturated conditions and includes queue spillover identification. Numerical results of simulation experiments and tests on NGSIM field data, with various penetration rates and sampling intervals, reveal the promising and robust performance of the proposed method compared with a uniform arrival queue estimation procedure. The method provides a thorough understanding of urban traffic flow dynamics and has direct applications for delay analysis, queue length estimation, signal settings estimation, and vehicle trajectory reconstruction.*

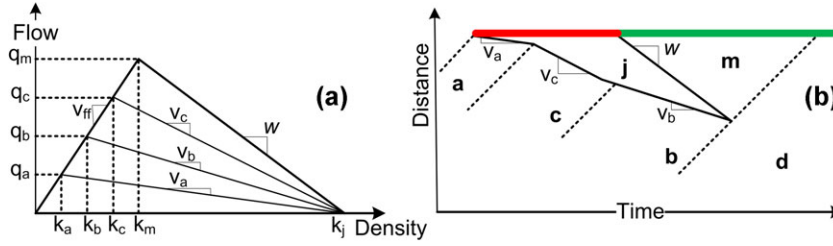
## 1 INTRODUCTION

Vehicle queues at intersections are the crucial cause of variability in drivers' experienced travel time in urban networks. Hence, an accurate and practical queue estimation method is of great importance for intelligent transportation systems (ITS) (i) to provide a better understanding of urban flow dynamics, (ii) to be utilized for traffic state estimation, and (iii) to be integrated in

a traffic signal control framework. In this article, we aim at estimating the evolution of queues by utilizing probe data. We introduce the concept of queue profile that describes the formation and dissipation of queues in the time-space ( $x-t$ ) plane. Queue profile is a polygon in the  $x-t$  plane that its edges designate traffic shockwave boundaries (see Figure 1b), based on kinematic LWR theory (Lighthill and Whitham, 1955; Richards, 1956). By estimating the queue profile for each cycle, one can derive performance measures including queue length, position of back and front of the queue, and delays per vehicle (average and distribution) that can be further used to estimate travel time distributions as a travel time reliability index (Ramezani and Geroliminis, 2012). In addition, given the queue profile vehicle trajectories can be reconstructed (see for example, Sun and Ban, 2013), which are fundamental for fuel consumption and emission estimation (Skabardonis et al., 2013) and travel time decomposition (Hellinga et al., 2008).

Literature of queue estimation can be categorized to two modeling classes: (i) models based on the cumulative traffic input–output (Webster, 1958; Akçelik, 1998; Erera et al., 1998; Viti and Van Zuylen, 2010), with the limitation that these models are insufficient to provide the spatial distribution of queue dynamics (Michalopoulos et al., 1981) and suffer from measurement errors, and (ii) models based on LWR shockwave theory (Skabardonis and Geroliminis, 2008; Ban et al., 2011; Wu and Liu, 2011), which provide the temporal–spatial dynamics of the queuing process with input data from loop detectors or probe vehicles. Further, many pivotal studies of queue and delay estimation in transportation research (Newell, 1960; Darroch et al., 1964; Newell, 1965; McNeill, 1968) are based (fully or partially) on rather limiting assumptions to result in closed-form theoretical solutions. These assumptions can be summarized as: (i) known signal settings, (ii) known

\*To whom correspondence should be addressed. E-mail: [nikolas.geroliminis@epfl.ch](mailto:nikolas.geroliminis@epfl.ch).



**Fig. 1.** The fundamental diagram and a queue profile at a signalized intersection. (See color figure in online version.)

initial queue size at the start of the cycle, (iii) known arrival pattern (e.g., uniform, Poisson), and (iv) undersaturated conditions that expected value of arrivals do not (constantly) exceed the signal capacity (Dion et al., 2004).

The main contribution of this article is to introduce a queue profile estimation method that relaxes the aforementioned assumptions. Because signal settings, that is, cycle length and splits, might not be readily available specifically from actuated traffic signals and network-wide systems (Hao et al., 2012). Note that signal settings are crucial for capturing queue spatiotemporal characteristics, as the spatial extent of queue should be estimated at the correct time instance. For example, if the exact queue length is estimated with a time lag the estimation error would be significant. Whereas in the analyses of maximum queue length estimation (or studies with assumption of known signal settings), temporal term of estimation error is zero. In addition, a known arrival traffic flow distribution might be considered valid in case of an isolated intersection; whereas such an assumption is not valid in arterials, as it neglects the impact of upstream intersections that alter the arrival pattern further downstream (Ramezani and Geroliminis, 2012). The other significance of the proposed queue estimation method is to cope with oversaturated conditions and identify the occurrence of spillovers in arterials with probabilistic inference.

The basic ingredient for the new wave of smart cities that has emerged during the last decade is massive data sets concerning human mobility, fostered by the widespread distribution of sensors, such as Global Positioning System (GPS) devices in many modes of transport, smart phones, and traffic fixed sensors (e.g., loop detectors). These network infrastructures allow for sensing and collecting substantial spatiotemporal data, such as the trajectories of many vehicles from navigation devices, which represent proxies for human mobility patterns. This “big mobility data” provides a unique social observatory that can help us understand how congestion develops and evolves, and discover hidden patterns and identify models that can contribute to efficient

traffic management systems to improve cities’ mobility and accessibility.

Prevailing queue estimation methods employ several monitoring technologies, e.g., loop detectors (Skabardonis and Geroliminis, 2008; Geroliminis and Skabardonis, 2011; Wu et al., 2011) and event-based signal and vehicle data (Wu et al., 2010). It has been reported in many publications that input–output diagrams have strong difficulties in estimating queue lengths accurately under congested conditions. The main reason is that detector errors (even unbiased) accumulate over time (see for example, Vigos et al., 2008). Furthermore, in case of detectors in the beginning and end of a link, stop-and-go traffic and non-smooth flow due to upstream departures creates strong disturbances in the estimation. For mid-block detectors when queues are overpassing the detector, the accuracy of estimation degrades. The emergence and a steadily increase of public deployment of user-based data collection systems, e.g., GPS-equipped vehicles (Herrera et al., 2010; Bhaskar et al., 2011; Jenelius and Koutsopoulos, 2013) and vehicle re-identification using cellphones or wireless magnetic sensors (Kwong et al., 2009), provides a great potential for probe vehicles in ITS applications.

We introduce a method that is based on probe data as the single source of information. Utilizing probe vehicles for traffic observation and estimation raises challenges because of (i) errors in raw measurements and post-processing algorithms, e.g., map-matching, and (ii) the fact that the probe data are inherent samples of a true traffic state. Thus, probe data cannot be readily applied to well-established traffic estimation methods. In this study, the position and instantaneous speed of probe vehicles are the input to the queue profile estimation method and the robustness of the method to measurement errors is investigated.

Kwong et al. (2009) and Hofleitner et al. (2012) represent two of the few efforts to estimate performance measures without the need for signal settings information. The former requires wide deployment of sensors and the latter proposes a probabilistic framework for arterial traffic state estimation using sparse

probe data that infer the average signal settings and arrival rates from historical data. Utilizing probe data, Ramezani and Geroliminis (2012) develop a method based on the link travel times of probe vehicles to estimate the arterial travel time distribution by considering spatiotemporal nonlinear correlations. Comert and Cetin (2011) provide an analytical model for queue length estimation of an isolated intersection with the assumption of Poisson arrival distribution and examined the effect of probe vehicles penetration rate. In addition, Izadpanah et al. (2009) and Cheng et al. (2012) study a queue estimation method based on the identification of shockwaves from probe vehicle trajectories. Although queue estimation is straightforward given trajectory of probe vehicles with large penetration rates, for realistic cases with low penetration rates and high sampling intervals, an approach that combines data mining, optimization techniques, and physical properties of traffic flow is needed (e.g., see Hao et al. (2012) and Hofleitner et al., (2012) also see Vlahogianni and Karlaftis (2013) for incident duration modeling, Ghosh-Dastidar and Adeli (2006) for delay and queue length estimation at freeway work zones, and Jiang and Adeli (2005) for traffic flow forecasting). A brief description of the developed method without considerations for spillovers, capturing the interdependencies of queuing dynamics of adjacent links and less robust optimization framework to noisy measurements is presented in Ramezani and Geroliminis (2013).

The remainder of the article is organized as follows. Section 2 presents the preliminaries of the traffic flow modeling utilized for the queue profile estimation. Next, we introduce the proposed method and elaborate on its details in Section 3. Section 4 presents the results for two case studies based on field data and simulation, comparisons with a queue estimation procedure based on the uniform arrival assumption, and the effect of probe data penetration rates, sampling intervals, and measurement errors on the performance of the proposed method. The spillover probabilistic inference method is introduced in Section 5. Finally, Section 6 draws the conclusions and discusses future research directions.

## 2 BACKGROUND

This article aims at estimating the queue shockwave profile that is a polygon in the  $x$ - $t$  plane such that its every edge designates a traffic shockwave. The shockwaves model the formation and dissipation of the queue during signal cycles. Based on LWR theory, shockwaves are boundaries determining flow-density discontinuities in the traffic stream, that is, points on the border be-

tween two different traffic states. Thus, crossing traffic shockwaves, the slope of vehicle trajectory in the  $x$ - $t$  plane (i.e., speed of the vehicle) changes abruptly. In this study, for each urban network link we assume a triangular fundamental diagram (FD) characterized by the maximum flow (capacity),  $q_m$ , the free flow speed,  $v_{ff}$ , and the jam density,  $k_j$  (see Figure 1a). Although LWR theory can integrate more complicated shapes of FDs, we choose the widely used triangular FD as it can represent well the development and dissipation of shockwaves in arterials and allows for elegant data mining techniques consistent with the physics of traffic. In this case, the platoon dispersion effect cannot be estimated. We refer to Michalopoulos et al. (1981) or Geroliminis and Skabardonis (2005) for non-triangular FDs for arterial estimation.

Figure 1b depicts a queue profile at a signalized intersection. The queue discharging shockwave represents the front of queue and a 3-segment piecewise linear shockwave represents the back of queue. It is apparent that the queue discharging shockwave separates the traffic state at capacity (**m**) from the jammed state (**j**). Likewise, the 3-segment shockwave of the back of queue separates the jammed state from 3 different traffic states ( $q_i, k_i$ ), in the flow-density plane with  $i \in \{a, b, c\}$ . Note that traffic state (**d**) may represent a different state than state (**b**). The slope of the shockwave between traffic states  $s$  and  $s'$  and the slope of the queue-discharging shockwave ( $w$ ) are, respectively:

$$v_{s's'} = \frac{q_s - q_{s'}}{k_s - k_{s'}}, \quad w = \frac{q_m}{\frac{v_{ff}}{q_m} - k_j} \quad (1)$$

It is straightforward to estimate the queue shockwave profile (the extent and slope of shockwaves) given the arrival traffic flow and the signal settings (which both impose measurement difficulties). However, the proposed queue profile estimation method does not require either arrival traffic distribution or signal settings. Note that one application of queue profile is that by reverse modeling, one can estimate the attributes of shockwaves, and consequently the characteristics of arrival traffic state can be determined. In addition, the estimated queue profiles approximate the signal setting, that is, the start and end time of signal phases. A recent study (Hao et al., 2012) estimates signal settings utilizing vehicle travel time observations from upstream and downstream of an intersection.

In this article, we try to estimate the queue profile polygon from probe vehicles. Probe vehicles provide samples of their individual traffic state and we aim at leveraging the collective information of temporally and spatially dispersed probe data. As we stated earlier, traffic shockwaves designate the points in the  $x$ - $t$  diagram associated with the sudden change of vehicles speeds.

In urban networks and particularly at signalized intersections, this associates with how probe vehicles join and leave the queue, which has to be estimated. The proposed method utilizes kinematic analysis to approximate the time and the position that a probe vehicle joins and leaves the queue (joining and leaving points) from its reported data by assuming constant acceleration ( $a_{\text{acc}}$ ) and deceleration ( $a_{\text{dec}}$ ). This assumption is not expected to influence the accuracy of the model (Skabardonis and Geroliminis, 2005).

### 3 METHODOLOGY

In this section, first a brief overview of the proposed queue profile estimation method is given, and then the principal steps of the method are elaborated in detail. Let us assume that probe vehicle  $i$  reports its position (the distance from upstream of the link),  $x_i^k$ , and velocity,  $v_i^k$ , along with the time stamp,  $t_i^k$ , at time step  $k$ . In the sequel, the link index is omitted for simplicity, and we assume periodic reporting of the above information with sampling interval  $T$  (s). However, the method can be easily applied with non-periodic probe data collected from various types of probe vehicles, e.g., taxis, transit fleet, and cars. Note that the queue profile estimation is a link-wise procedure and in case of possible spillovers (or spillbacks), the queue profile should be estimated from downstream to upstream links to be able to capture the spillover development and propagation. Note that the occurrence of spillbacks is not considered in this section, whereas in Section 5, we relax this assumption and develop a spillover identification procedure based on Bayesian inference.

An example of the queue profile estimation with  $T = 30$  (s) and penetration rate ( $\phi$ ) of 40% is depicted in Figure 2, where ground truth and estimated queue profiles are, respectively, denoted by black and red polygons. Even if this penetration rate is not realistic, it is chosen for the explanation of the method. Later, smaller values of  $\phi$  are analyzed. Note that probe data points are the input to the method and trajectories in Figure 2 are shown for illustration purposes of the queue dynamics. The details shown in Figure 2 will be fully elaborated in the following subsections. In addition, effects of measurements or map-matching errors on the performance of the proposed estimation method are scrutinized in Section 4.

A conceptualization of the queue profile estimation method is as follows. Because signal settings are assumed to be unavailable, we need to identify stopped data points and assign them to various groups so that each group represents a cycle. Therefore in Section 3.1, we introduce a classifier based on the velocity attribute

to classify all probe data to two groups, stopped and moving. In Figure 2, red squares denote the stopped data points whereas black diamonds denote the moving data points. Afterwards in Section 3.2, we cluster the stopped data points into different signal cycles based on the projection profile algorithm. Subsequently in Section 3.3, the moving data are assigned to different cycles. The magenta lines provide linear boundaries that group the moving data in cycles. Figure 2 depicts the cycle number of each stopped and moving data point. As queued vehicles discharge at the capacity flow, based on LWR theory, the queue discharging shockwave is a line with slope  $w$  in the  $x$ - $t$  plane. In addition, the leaving points of vehicles from the queue (filled blue points in Figure 2) are on the queue discharging shockwave. Hence, the estimation of front of queue for each cycle can be formulated as a constrained least squares problem, which is presented in Section 3.4. The proposed method does not require the knowledge of arrival patterns to estimate queue profiles. Thus to model the back of queue, we consider a piecewise linear function that fits the joining points of vehicles to the queue (filled green, cyan, and yellow points in Figure 2). In Section 3.5, a curve-fitting nonlinear optimization method is introduced to identify the number and attributes of the piecewise linear function (Groot et al., 2013).

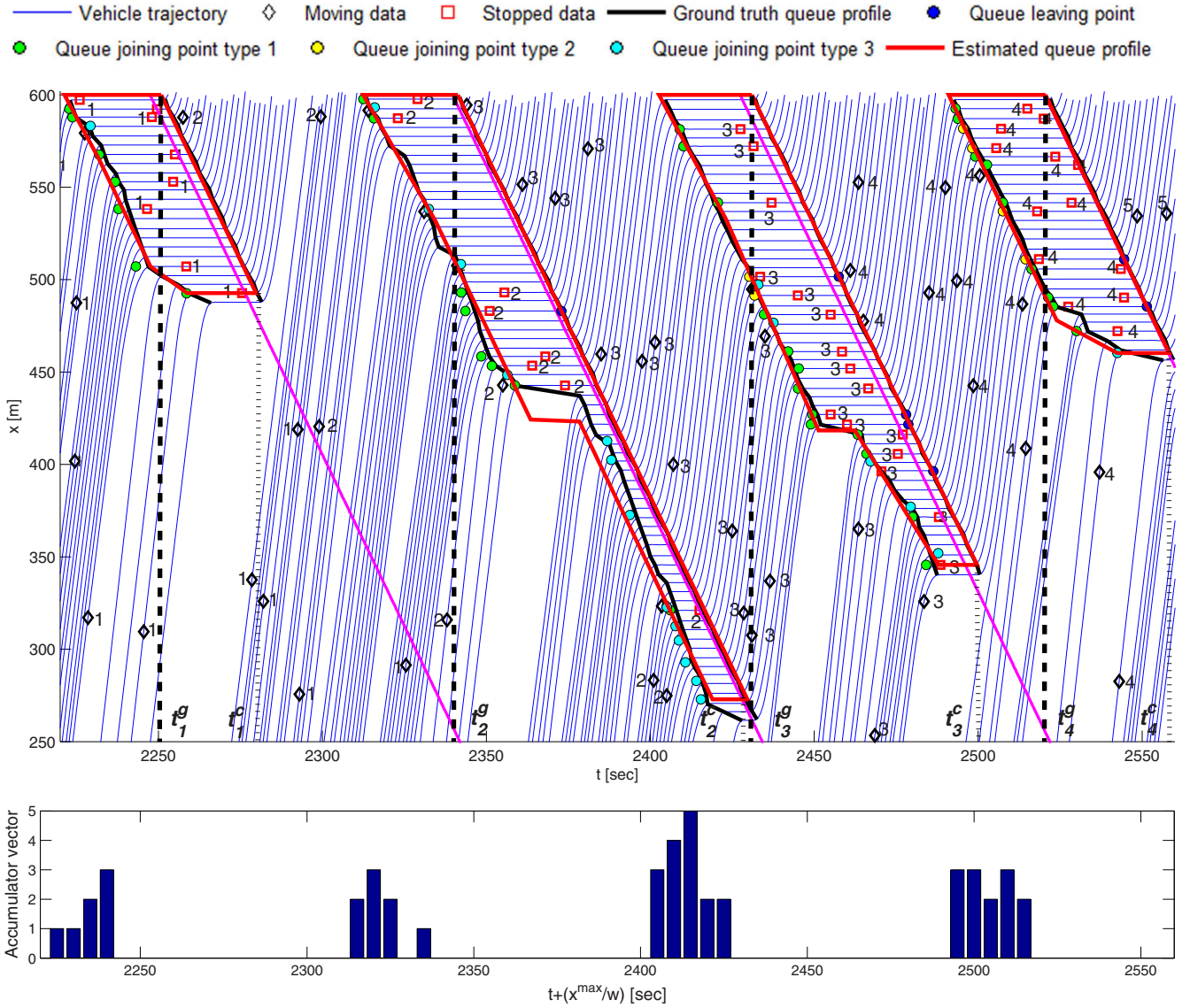
#### 3.1 Classification of probe data to moving and stopped classes

The first step of the procedure is to classify the probe data into two classes of moving and stopped vehicles  $\{m, s\}$ . This can be done with a simple threshold-based classifier:

$$c_i^k = \begin{cases} m & \text{if } v_i^k > v_{\text{th}} \\ s & \text{if } v_i^k \leq v_{\text{th}} \end{cases} \quad (2)$$

where  $v_{\text{th}}$  (m/s) is a predefined threshold parameter to designate the vehicles with velocity close to zero as stopped.

The correct classification rate of the proposed classifier is 100% in case of exact measurements. Though errors in velocity measurements degrade the classifier performance, measurement errors do not significantly affect the method (The robustness of method to measurement errors is examined later in Section 4). However, a more precise velocity measurement, e.g., utilizing differential GPS or fusion of GPS speed data with vehicle speedometer seems essential, given also that signal settings are considered unknown. The classification of probe data points to stopped (red square) and moving (black diamond) is illustrated in Figure 2.



**Fig. 2.** An illustrative example of the queue profile estimation method. The black diamonds and red squares respectively show input probe data classified to moving and stopped points (Section 3.1). Their associated numbers are the corresponding cycle index (Sections 3.2 and 3.3), where magenta lines group the moving points to cycles. The filled blue circles are the queue leaving points of probe vehicles that designate the front of the queue profile (Section 3.4). The front of queue shockwave determines the estimated start of green phase,  $t^g$  (Section 3.4). The filled green, yellow, and cyan circles are different types of queue joining points of probe vehicles that designate the back of queue (Section 3.5).  $t^c$  denotes the time that the queue clears. The bottom plot shows the corresponding accumulator projection vector (Section 3.2) depicting four groups of stopped data associated with four cycles. All the details are elaborated in corresponding subsections. (See color figure in online version.)

### 3.2 Clustering the stopped vehicles to cycles

Our motivation in this article is to relax the assumption of known signal settings. With no information of signal settings, the number and duration of cycles are unknown, especially for actuated signals. As queues are associated with cycles, in the second step of the procedure a clustering technique is needed to cluster the

stopped data into various groups that represent cycles (of the same or different duration). By considering the temporal and spatial distribution of stopped data in the  $x$ - $t$  plane, the stopped data of two consecutive cycles can be linearly separated with a straight line with slope  $w$  that lies properly between the two cycles. Therefore, we propose a clustering technique based on the

projection profile method (Jelaca et al., 2012) to cluster the stopped data into cycles. The purpose of the projection profile method is to extract a scalar feature from the (2D) dispersed data of stopped vehicles in the  $x$ - $t$  plane. Accordingly, we introduce a projection function,  $\mathcal{P}$ , which projects every stopped point along a line with slope  $w$  into an accumulator projection vector over the line  $x = 0$  (the upstream of link). At this location, time is divided into equal intervals (bins) such that each bin is associated with the number of projected points lying inside the bin, see Figure 2.

$$\text{if } c_i^k = s \Rightarrow \mathcal{P}(t_i^k, x_i^k) = t_i^k - \left(\frac{x_i^k}{w}\right) \quad (3)$$

It is expected that the projection profile method with projection function  $\mathcal{P}$  results in an accumulator vector that contains (i) clusters of adjacent bins with positive values, representing the red interval of a cycle, and (ii) clusters of adjacent bins with zero values, representing the green duration of a cycle (see values of the accumulator projection vector in Figure 2). Consequently, the stopped data associated with every cluster of stopped bins (adjacent bins with positive value) correspond to a common cycle. It is remarkable that no prior information on number of clusters, that is, number of cycles, is needed for the clustering procedure. This property is important for actuated traffic signals or for spillback identification.

The accuracy of projection profile algorithm depends on the bin size ( $s$ ). To determine the size of each bin, one should consider that large values cannot discriminate between cycles; while small values may decrease the chance that stopped data points of the same cycle are projected into adjacent bins. (Note: In case of small-sized bins, the number of bins with zero value during the red interval increases. This might create group(s) of zero-valued bins during the red interval. So, the clustering procedure should disregard small number of zero-valued bins inside a cluster of stopped bins; see for example, cycle 2 in Figure 2.) We test the sensitivity of the projection profile algorithm to various bin sizes and values between 4 and 10 ( $s$ ) show satisfactory outcomes. In this study, we set the bin size equal to 5 ( $s$ ). Result of this step is apparent in Figure 2, where the cycle number of each stopped data point is depicted. Note that this method is still valid in case of active spillovers that block departures during green times of upstream intersections. An enhancement to the clustering procedure concerning occurrence of spillovers will be discussed in Section 5. In case of very low penetration rates ( $<5\%$ ), this method can be improved by also considering the moving points during green durations, which can be considered as a future work.

### 3.3 Associating the moving vehicles to cycles

The previous step identifies the number of cycles (not signal settings) and assigns stopped data to their corresponding cycle. Likewise, the proposed queue profile estimation method needs the moving data ( $c_i^k = m$ ) to be associated to cycles to formulate the estimation problem of the front and back of queue. Let  $S_j$  denote the set of stopped data assigned to cycle  $j$ . One can fit a line with slope  $w$  to every data point that belongs to  $S_j$ . The line with the maximum intercept (traffic progresses in direction of  $x$ -axis) can be regarded as a lower envelope of the “true” discharging shockwave, see magenta lines in Figure 2. This line can be considered as a linear boundary to distinguish between moving data of every cycle. In other words, the moving data of cycle  $j$  are on the left side of this line for cycle  $j$ , and right side of the corresponding line for cycle  $j - 1$ . We denote the set of moving data associated with cycle  $j$  as  $M_j$ . Figure 2 depicts the number of the cycle that each moving point is assigned to. Note that the error between the magenta lines and the “true” discharging shockwaves only influences assigning cycles to moving points and does not intervene on cycle identification, signal setting estimation, and queue profile estimation.

### 3.4 Estimation of the front of queue

The queuing process can be divided into two distinct fundamental processes, the formation and dissipation of queues, which in urban networks is associated with the traffic signal alternation. This step of the queue profile estimation is to determine the queue dissipation for every cycle ( $j$ ). Based on LWR theory, the queued vehicles discharging at the capacity flow result in a shockwave designated by a line with slope  $w$  in the  $x$ - $t$  diagram. Hence, there is only one parameter to estimate the front of queue shockwave (discharging line), that is, the line intercept. Initially, we need to estimate the leaving point of probe vehicles, that is, at which time and position probe vehicles leave the queue. The estimation of leaving points of probe vehicles is important because these points are ideally supposed to be on the discharging shockwave line. From the formulas of kinematics on one dimension and based on two consecutive data points of the same probe vehicle, where the first one is stopped and the other one is moving (i.e.,  $c_i^k = s$ ,  $c_i^{k+1} = m$ ,  $(t_i^k, x_i^k) \in S_j$ ), the time that probe vehicle  $i$  leaves the queue can be estimated as

$$\tau_i^1 = \begin{cases} t_i^{k+1} - \frac{v_i^{k+1}}{2a_{\text{acc}}} - \frac{x_i^{k+1} - x_i^k}{v_i^{k+1}} & \text{if } v_i^{k+1} > \eta \times v_{\text{ff}} \quad (4a) \\ t_i^{k+1} - \frac{2(x_i^{k+1} - x_i^k)}{v_i^{k+1}} & \text{if } v_i^{k+1} \leq \eta \times v_{\text{ff}} \quad (4b) \end{cases}$$

Equation (4a) corresponds to a vehicle that reaches its desired speed, which is in a vicinity of the free flow speed, specified by  $\eta \in (0, 1)$ .  $\eta$  is the parameter to designate whether a probe vehicle is in acceleration (or deceleration) mode or it has reached the free flow speed. Likewise, (4b) corresponds to a vehicle that its instantaneous speed at time step  $k + 1$  is not its desired speed, because the vehicle is still accelerating. Note that, there is another possibility that the vehicle has a speed below its desired speed ( $v_i^{k+1} \leq \eta \times v_{\text{ff}}$ ) and be in the deceleration stage. In this case, (4b) gives a value less than  $t_i^k$  which is incorrect because,  $\tau_i^1 \in (t_i^k, t_i^{k+1})$ . Therefore, we need to modify (4b) to correctly estimate the leaving time of probe vehicle  $i$  that has a speed below the desired speed and is in the deceleration stage. So, if  $v_i^{k+1} \leq \eta \times v_{\text{ff}}$  and  $\tau_i^1$  computed by (4b) is smaller than  $t_i^k$ , the time that probe vehicle  $i$  leaves the queue is

$$\tau_i^1 = t_i^{k+1} - \frac{v_{\text{ff}}}{2a_{\text{acc}}} - \frac{x_i^{k+1} - x_i^k}{v_{\text{ff}}} + \frac{(v_i^{k+1} - v_{\text{ff}})^2}{2 \cdot v_{\text{ff}} \cdot a_{\text{dec}}} \quad (4c)$$

Equation (4c) implies that the vehicle accelerates with acceleration  $a_{\text{acc}}$ , reaches its desired speed which is equal to  $v_{\text{ff}}$ , and then decelerates with deceleration  $a_{\text{dec}}$  to its speed at time step  $k + 1$ ,  $v_i^{k+1}$ .

Ultimately, the leaving point is estimated as  $(\hat{t}_i, \hat{x}_i) = (\tau_i^1, x_i^k)$ . In Figure 2, filled blue points represent leaving points. Although time spent in acceleration and deceleration mode is a minor component of the delay under congested conditions and can be ignored, it is a necessary step in our approach to classify the different types of vehicles (joining or leaving the queue).

The discharging line has slope  $w$  such that all the moving data points of the next cycle,  $M_{j+1}$ , and stopped data points of the corresponding cycle,  $S_j$ , are on its right and left side, respectively. Hence, the estimation of front of queue is formulated as a constrained least squares problem:

$$\min_{B_j} \sum_{l=1}^{N_L} (\hat{x}_l - w\hat{t}_l - B_j)^2 \quad (5)$$

$$\text{subject to } \begin{cases} x - wt - B_j \leq \varepsilon_1 & \forall (t, x) \in S_j \\ x - wt - B_j \geq -\varepsilon_1 & \forall (t, x) \in M_{j+1} \end{cases} \quad (6)$$

where  $N_L$  denotes the total number of estimated leaving points and  $B_j$  is the intercept of the discharging line shockwave of cycle  $j$ . Because problem (5) is convex, the solution is global although not necessarily unique. Note that given  $B_j$ , the start time of the current cycle green phase, denoted by  $t_j^g$ , can be estimated as:

$$t_j^g = \frac{x^{\text{max}} - B_j}{w} \quad (7)$$

where  $x^{\text{max}}$  is the position of intersection stop line from the entrance (upstream) of the link.

The objective of Equation (5) is to minimize the sum of squared errors between the leaving points and the estimated discharging shockwave line. Regarding constraints (6), introduction of  $\varepsilon_1$  provides a tuning parameter to regulate the extent of hardness of the constraints, e.g.,  $\varepsilon_1 = 0$  represents hard-constraints that obliges the solution of Equation (5) to fully separate data points of  $S_j$  and  $M_{j+1}$ , whereas a positive value of  $\varepsilon_1$  relaxes the constraints, which is necessary in case of noisy input data, because there might be no line with slope  $w$  that fully separates noisy data of  $S_j$  and  $M_{j+1}$ . It is worth mentioning that, there is no restriction that two consecutive stopped and moving data points ( $c_i^k = s$  and  $c_i^{k+1} = m$ ) are on the same link, that is, the moving point can belong to the downstream link. Nevertheless, the chance of finding two such data points in the most downstream link is less compared to the other links. Thus, in case there is no estimated leaving point (see the first cycle in Figure 2), the discharging shockwave line is estimated based on the best generalization concept which provides the maximum margin from both stopped ( $S_j$ ) and moving vehicles ( $M_{j+1}$ ). In other words, the discharging line should have an equal distance from the nearest data points of both  $S_j$  and  $M_{j+1}$ . This concept is widely utilized in support vector machine (SVM) in computer science literature (Cortes and Vapnik, 1995) and reads:

$$B_j = \frac{\max_{S_j} B + \min_{M_{j+1}} B}{2} \quad (8)$$

where  $\max_{S_j} B$  ( $\min_{M_{j+1}} B$ ) denotes the maximum (minimum) intercept of lines with slope  $w$  that are fitted to the data points of set  $S_j$  ( $M_{j+1}$ ).

### 3.5 Estimation of the back of queue

The goal of this step is to estimate the queue formation for each cycle ( $j$ ). The back of queue in the  $x$ - $t$  diagram can be modeled by a piecewise linear function of several segments so that each one represents a shockwave. Without any explicit information of arrival patterns, the back of queue estimation procedure should determine the number and extent of segments of the piecewise linear function. Initially, similar to the previous step, we need to estimate the joining point of probe vehicles, that is, at which time and position probe vehicles join the queue, because the joining points ideally belong to the piecewise linear shockwave. We consider the following types of two consecutive data points for probe

vehicle  $i$  that contribute to determining the joining points:

(Type 1) the first point is moving and the other one is stopped, that is,  $(t_i^{k-1}, x_i^{k-1}) \in M_j$  and  $(t_i^k, x_i^k) \in S_j$ , (filled green points in Figure 2)

(Type 2) both points are stopped, that is,  $(t_i^{k-1}, x_i^{k-1}) \in S_{j-1}$  and  $(t_i^k, x_i^k) \in S_j$ , (filled yellow points in Figure 2)

(Type 3) both points are moving, that is,  $(t_i^{k-1}, x_i^{k-1}) \in M_j$  and  $(t_i^k, x_i^k) \in \{M_{j+1} \cup ND\}$ , (filled cyan points in Figure 2)

(Type 4) first point is stopped and the second one is moving, that is,  $(t_i^{k-1}, x_i^{k-1}) \in S_{j-1}$  and  $(t_i^k, x_i^k) \in \{M_{j+1} \cup ND\}$ .

$ND$  denotes the set of moving points in the downstream link that are not delayed, that is, the points that do not belong to the queue discharged traffic state (state  $\mathbf{m}$  in Figure 1b). Note that because the queue profile estimation method starts from the downstream link,  $ND$  is already available. By considering moving points of  $ND$ , the methodology implicitly correlates the queuing state of successive links. For Types 3 and 4, first we should assess if probe vehicle  $i$  encounters any significant delay between the two consecutive observations. This is straightforward by comparing the average speed between two successive observations and the free flow speed. Afterwards, if the probe vehicle is delayed (as it joins and leaves a queue), the data at time step  $k$  belonging to  $M_{j+1}$  or  $ND$  can be converted to a queue leaving point according to Equation (4), because the discharging line of the cycle  $j$  is already estimated in the fourth step. Therefore, data points of Types 3 and 4 can be readily converted to Types 1 and 2, respectively.

For Type 1, the time that probe vehicle  $i$  joins the queue can be calculated as

$$\tau_i^j = \begin{cases} t_i^{k-1} - \frac{v_i^{k-1}}{2a_{\text{dec}}} + \frac{x_i^k - x_i^{k-1}}{v_i^{k-1}} & \text{if } v_i^{k-1} > \eta \times v_{\text{ff}} \text{ (9a)} \\ t_i^{k-1} + \frac{2(x_i^k - x_i^{k-1})}{v_i^{k-1}} & \text{if } v_i^{k-1} \leq \eta \times v_{\text{ff}} \text{ (9b)} \end{cases}$$

Equations (9a, 9b) have the same logic as (4a, 4b), so that Equation (9a) corresponds to a vehicle that reaches its desired speed in the vicinity of free flow speed, and Equation (9b) corresponds to a vehicle that its instantaneous speed at time step  $k-1$  is not its desired speed, because the vehicle is decelerating. Similarly, the third possibility is that the vehicle has a speed below its desired speed ( $v_i^{k-1} \leq \eta \times v_{\text{ff}}$ ) and is in the acceleration stage. In this case, Equation (9b) gives a value greater

than  $t_i^k$  which is incorrect because,  $\tau_i^j \in (t_i^{k-1}, t_i^k)$ . Therefore, we need to modify (9b) to correctly estimate the joining time of probe vehicle  $i$ . So, if  $v_i^{k-1} \leq \eta \times v_{\text{ff}}$  and  $\tau_i^j$  computed by (9b) is greater than  $t_i^k$ , the time that probe vehicle  $i$  joins the queue is

$$\tau_i^j = t_i^{k-1} - \frac{v_{\text{ff}}}{2a_{\text{dec}}} + \frac{x_i^k - x_i^{k-1}}{v_{\text{ff}}} + \frac{(v_{\text{ff}} - v_i^{k-1})^2}{2 \cdot v_{\text{ff}} \cdot a_{\text{acc}}} \text{ (9c)}$$

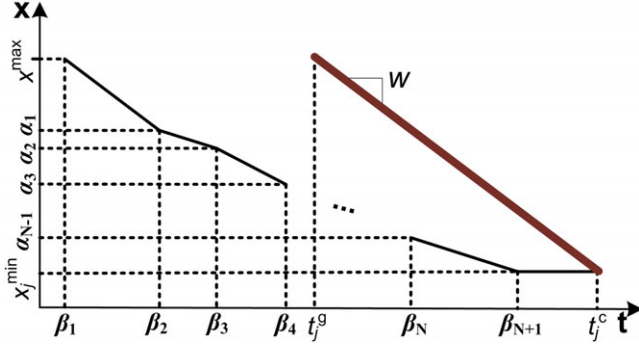
Equation (9c) implies that the vehicle with speed  $v_i^{k-1}$  accelerates with acceleration  $a_{\text{acc}}$ , reaches its desired speed which is equal to  $v_{\text{ff}}$ , and then decelerates with deceleration  $a_{\text{dec}}$  to the stopping state. Ultimately, the joining point is estimated as  $(\check{t}_i, \check{x}_i) = (\tau_i^j, x_i^k)$ .

Note that for data Type 2, the leaving point of probe vehicle  $i$  at cycle  $j-1$  can be estimated by utilizing the stopped data point  $(t_i^{k-1}, x_i^{k-1})$  and the discharging shockwave of the previous cycle,  $j-1$ . (The discharging shockwave of the previous cycle is already known, because queue profile estimation advances cycle by cycle in time.) Thus with the estimated leaving point that is certainly at the acceleration stage and by applying (9c) the corresponding joining point can be estimated. Remarkably, because the data point at time step  $k-1$  of Type 2 is discharging from the queue of the previous cycle, it arrives at capacity state to the queue of the current cycle. Hence, the slope of the corresponding segment of back of queue shockwave should be  $w$ .

Given the estimated joining points,  $J_j$ , the back of queue piecewise linear shockwave,  $f_b(x)$ , should be estimated according to the following objectives: (i) to be as close as possible to the joining points, and (ii) to separate all the moving data points associated with cycle  $j$ ,  $M_j$ , and stopped ones of cycle  $j$ ,  $S_j$ , on its left and right side, respectively (see the definition of function  $\mathcal{F}$ ). Further, based on LWR theory, the slope of each segment of the piecewise linear function should be between zero and  $w$  (see Equation (13)). Note that, if a leaving point of Type 2 or 4 is associated with a segment of  $f_b(x)$ , the segment slope should be equal to  $w$ . To incorporate all the aforementioned physical properties, the estimation of the back of queue by an  $N$ -segment piecewise linear function is formulated as the following nonlinear program:

$$\min_{\substack{\alpha_1, \dots, \alpha_{N-1} \\ \beta_1, \dots, \beta_{N+1}}} \frac{\sum_{l=1}^{N_j} (\mathcal{D}(f_b(\cdot), (\check{t}_l, \check{x}_l)))^2}{N_j} + \lambda N + C \left[ \sum_{S_j} \mathcal{F}(t, x) + \sum_{M_j} \mathcal{F}(t, x) \right] \text{ (10)}$$





**Fig. 3.** A schematic of  $N$ -segment piecewise linear back of queue at cycle  $j$ , annotated with parameters of optimization Equation (10). (See color figure in online version.)

$$f_b(x) = \begin{cases} \beta_1 + \frac{x - x^{\max}}{\alpha_1 - x^{\max}}(\beta_2 - \beta_1) & \alpha_1 < x \leq x^{\max} \\ \vdots \\ \beta_n + \frac{x - \alpha_{n-1}}{\alpha_n - \alpha_{n-1}}(\beta_{n+1} - \beta_n) & \alpha_n < x \leq \alpha_{n-1} \\ \vdots \\ \beta_N + \frac{x - \alpha_{N-1}}{x_j^{\min} - \alpha_{N-1}}(\beta_{N+1} - \beta_N) & x_j^{\min} \leq x \leq \alpha_{N-1} \end{cases} \quad (11)$$

$$x_j^{\min} \triangleq \min_x (S_j, J_j) < \alpha_{N-1} < \dots < \alpha_1 < x^{\max} \quad (12a)$$

$$t_{j-1}^g < \beta_1 < \dots < \beta_{N+1} \leq t_j^c = \frac{x_j^{\min} - B_j}{w} \quad (12b)$$

$$\mathcal{F}(t, x) = \frac{\beta_{n+1} - \beta_n}{\alpha_n - \alpha_{n-1}} \leq \frac{1}{w}; n = 1, \dots, N \quad (13)$$

$$\begin{cases} \mathcal{D}(f_b(\cdot), (t, x)) & \text{if } (t, x) \in S_j \wedge f_b(x) - t > 0 \\ \mathcal{D}(f_b(\cdot), (t, x)) & \text{if } (t, x) \in M_j \wedge f_b(x) - t < 0 \end{cases} \quad (14)$$

where  $N_j$  denotes the total number of estimated joining points, and function  $\mathcal{D}$  determines the distance of a point from a piecewise line. Hence, the first term of the objective function (10) is to minimize the sum of squared distance between the joining points and the desired piecewise linear function. Given that  $f_b(x)$  comprises  $N$  segments, (10) optimizes  $2N$  parameters defining slope and extent of segments, that is,  $\alpha_1, \dots, \alpha_{N-1}$  that are associated with the spatial extent of segments and  $\beta_1, \dots, \beta_{N+1}$  that are associated

with the temporal extent of segments, see Figure 3. It is apparent that as  $N$  increases,  $f_b(x)$  becomes more complex and can attain better optimization results as opposed to higher chance of overfitting. To prevent overfitting, we need to balance the number of parameters, as an indicator of complexity, and the value of the objective function. To this end, we add a complexity penalty term to the objective function,  $\lambda N$ , where  $\lambda$  is the regularization parameter. In addition, function  $\mathcal{F}$  calculates the distance of misclassified moving data of  $M_j$  and stopped data of  $S_j$  and penalizes them with a relatively large constant,  $C$ . Finally to find the optimum solution of Equation (10), we restrict  $N \leq \sqrt{N_j}$  and execute the optimization algorithm for several random initial points to prevent reaching local optima.

Furthermore, constraints (12) ensure that  $f_b(x)$  has to be connected and to be in proper intervals, that is,  $x_j^{\min} \leq x \leq x^{\max}$  and  $t_{j-1}^g < t \leq t_j^c$ . In other words, (12a) ensures that the spatial extent of the back of queue shockwave is between the position of intersection stop line,  $x^{\max}$ , and the position of the furthest queued vehicle from the intersection for cycle  $j$ ,  $x_j^{\min}$ . To estimate  $x_j^{\min}$ , we consider the furthest position of available probe information, as the most conservative estimator. Thus,  $x_j^{\min}$  is the minimum between the furthest position from the stop line among stopped data  $S_j$  (see cycles 1 and 3 in Figure 2), and the furthest position among the estimated joining points  $J_j$  (see cycles 2 and 4 in Figure 2). Note that,  $x_j^{\min}$  is the position where the estimated discharging shockwave and the piecewise shockwave of back of queue are connected to each other with a horizontal line (slope equals to zero) to connect the queue profile polygon (see Figures 2 and 3). Also (12b) implies that the estimation of the start time of red phase of current cycle,  $\beta_1$ , should be after the start time of green phase of the previous cycle denoted by  $t_{j-1}^g$ . (If there is any information regarding a minimum duration of red phases, this can be easily incorporated in the optimization framework.) In addition, the back of queue piecewise shockwave should be always (with respect to time) before the front of queue discharging shockwave in the  $x$ - $t$  plane, that is, the utmost time extent of the piecewise shockwave,  $\beta_{N+1}$ , must be less than the estimated time of queue full clearance, which is denoted by,  $t_j^c$ .

In case there is no estimated joining point, the piecewise back of queue shockwave can be estimated similarly based on the best generalization concept which has the maximum margin from both stopped ( $S_j$ ) and moving vehicles ( $M_j$ ). In other words, the discharging piecewise shockwave should have the maximum distance from the nearest data points of both  $S_j$  and  $M_{j+1}$ .

This reads:

$$\max_{\substack{\alpha_1, \dots, \alpha_{N-1} \\ \beta_1, \dots, \beta_{N+1}}} \left( \min_{S_j} \mathcal{D}(f_b(\cdot), (t, x)) + \min_{M_j} \mathcal{D}(f_b(\cdot), (t, x)) - \lambda N - C \left[ \sum_{S_j} \mathcal{F}(t, x) + \sum_{M_j} \mathcal{F}(t, x) \right] \right) \quad (15)$$

where  $\min_{S_j} \mathcal{D}(f_b(\cdot), (t, x))$  or  $(\min_{M_j} \mathcal{D}(f_b(\cdot), (t, x)))$  denotes the distance of the nearest data point of  $S_j$  or  $(M_j)$  to  $f_b$ . To solve the nonlinear problem (15), the conditions (11)–(14) should be also satisfied.

## 4 RESULTS

In this section, we test and discuss the proposed queue profile estimation method using (i) NGSIM field data and (ii) a more congested micro-simulated arterial with four signalized intersections. We also investigate the effect of probe data sampling interval, penetration rates, and measurement noise on the performance of method. Moreover, for further comparison the proposed method is compared with a uniform arrival queue estimation procedure, which requires signal settings data and aggregated data of an upstream (entrance) loop detector. For uncongested conditions, the uniform arrival assumption leads to a triangular queue profile for each cycle (see Figure 4a), such that the red interval is the one edge, the second edge is the discharging shockwave starting at the end of red phase with slope  $w$ , and the third edge starts from the beginning of red phase with a slope based on Equation (1), where  $s'$  is the jammed traffic state and state  $s$  is:

$$q_s = \frac{\text{count}_j}{C_j}, \quad k_s = \frac{q_s}{v_{\text{ff}}} \quad (16)$$

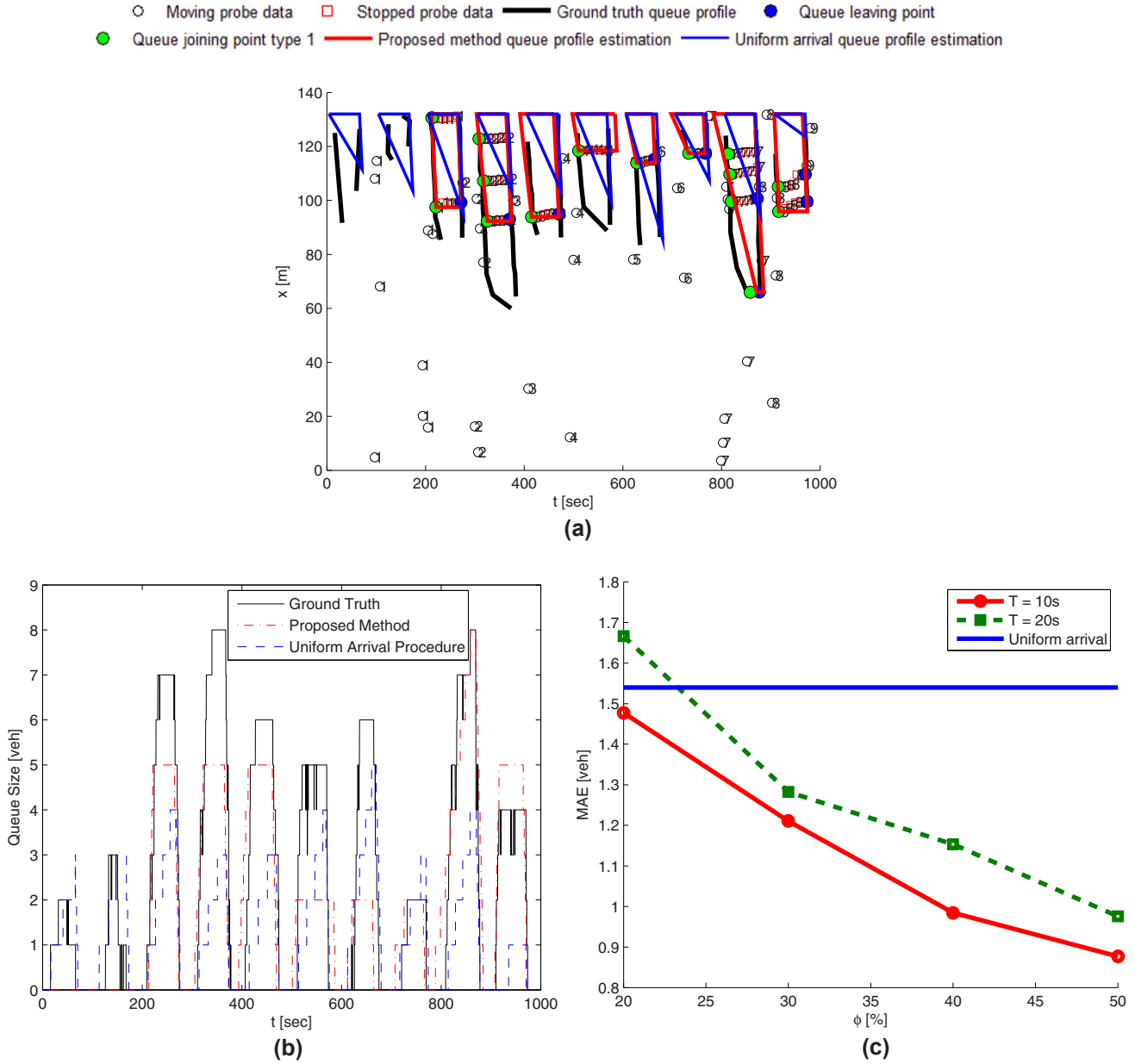
$C_j$  denotes the duration of cycle  $j$  and  $\text{count}_j$  is number of vehicles that enter the link and with the free flow speed would reach to the intersection stop line during cycle  $j$ . In case of oversaturated conditions and residual queues at the end of the green time, the queue profile has a trapezoidal form (see for example, Skabardonis and Geroliminis, 2008). An approach where the uniform arrival is applied without considering residual queues (i.e., a triangular queue profile) would result in much worse performance.

The method is tested on NGSIM data set. NGSIM program (2006) was managed by the Federal Highway Administration to provide a data set of vehicle trajectories for traffic behavioral analyses. The data set contains comprehensive individual vehicle information such as instantaneous position and velocity along with the time,

link, and direction stamps at the resolution of 0.1 second. We use NGSIM data set of the through lane of the southbound link between 11th Street and 10th Street on Peachtree Street, Atlanta, from 16:00 to 16:15. This link is approximately 130 (m) long and the speed limit is 56 (km/h). The FD parameters are set to  $v_{\text{ff}} = 15.65$  (m/s),  $q_m = 2,100$  (veh/h), and  $w = -3.7$  (m/s), and the parameters of the proposed method are chosen as:  $v_{\text{th}} = 1$  (m/s),  $\eta = 0.8$ ,  $a_{\text{acc}} = 2$  (m/s<sup>2</sup>),  $a_{\text{dec}} = -3.5$  (m/s<sup>2</sup>),  $\varepsilon_1 = 0$ ,  $C = 10^4$ , and  $\lambda = 1$ .

Figure 4a illustrates the estimated queue profiles for 10 cycles based on the proposed method (red polygons,  $\varphi = 30\%$  and  $T = 10\text{s}$ ) and the uniform arrival assumption procedure (blue polygons) along with the ground truth queue profiles (black polygons). Given the estimated queue profiles, signal settings, queue length, average delay, etc. can be readily estimated. For quantitative comparison, the queue size is derived based on the estimated queue profiles and depicted in Figure 4b. Furthermore, Figure 4c shows the mean absolute error (MAE) between the ground truth queue size and the estimated queue size averaged over 10 runs for various penetration rates ( $\varphi = 20, 30, 40, 50\%$ ) and sampling intervals ( $T = 10, 20\text{s}$ ). Note that during the data collection interval, there are only 80 vehicles crossing the link. Hence with a low penetration rate, the number of probe vehicles and consequently probe data would be insufficient to have an accurate estimation. It is apparent in Figure 4a where the first two cycles are missed with the proposed method. Likewise, sampling intervals greater than 20s yields inadequate number of probe data because the link free flow travel time is below 10s. However, it is apparent in Figure 4c that the proposed queue profile estimation method without any information of arrival distribution and signal settings is more accurate than the uniform arrival estimation procedure. Evidently with higher penetration rates or lower sampling intervals of probe vehicles, the outcomes of the proposed queue profile estimation method would be more precise.

A fundamental property of probe data is that as congestion increases, which is the case when the traffic estimation and control are crucial, the number of probe data increases (consider that congestion increases the total vehicle hours traveled by all vehicles, which is directly related to the number of available measurements). Hence in case of congestion, the proposed method performance is significantly better compared to light conditions. The NGSIM data do not include oversaturated conditions (note maximum queue size is 8 vehicles in Figure 4b). The results are not surprising as at undersaturated conditions, simple and elegant methods can provide decent quality of results. Travel time estimation methods have reached similar conclusions about

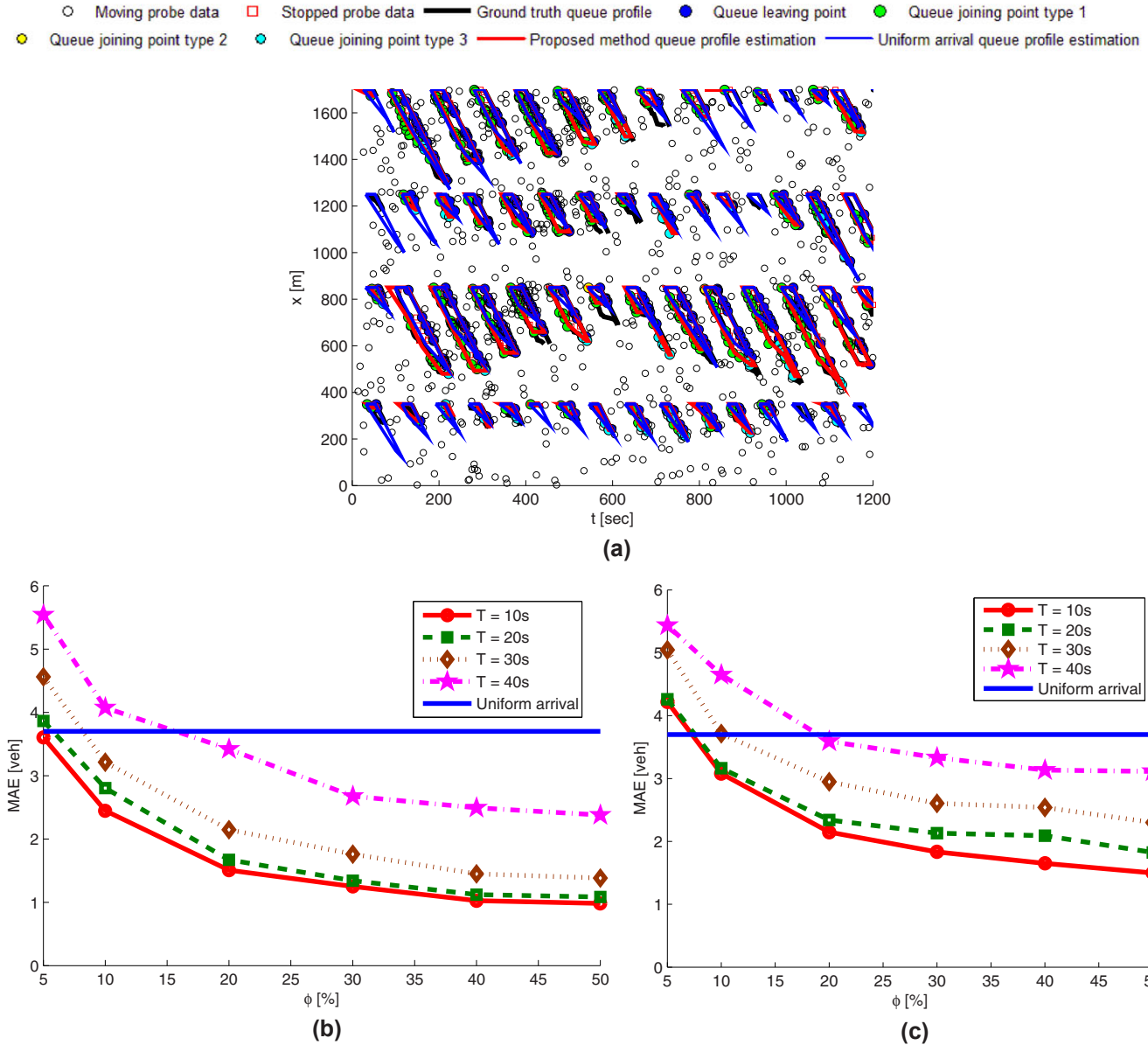


**Fig. 4.** The NGSIM ground truth (black) and estimated ( $\phi = 30\%$ ,  $T = 10s$ ) queue profiles based on the proposed method (red) and based on the uniform arrival assumption (blue). The probe input data consist of time, position, and velocity of probe vehicles (a); the estimated and ground truth queue length (b); the MAE between the ground truth and estimated queue length with different penetration rates and sampling intervals (c). (See color figure in online version.)

the accuracy of a simple model of uniform arrivals. Nevertheless, these models fail when congestion and long queues are present (see for example, Skabardonis and Geroliminis, 2008).

To test the developed methodological framework on more challenging scenarios, we analyze a micro-simulated case study with relatively high traffic volumes and long queues that do not exist in the NGSIM

data. This also enables us to investigate the effect of low penetration rates, high sampling intervals, and measurement noise. The simulated case study is an arterial with four signalized intersections where length of links are respectively 350, 500, 400, and 450 (m), with time-varying demand for 20 minutes. The signals are traffic-actuated each with different cycle length. Note that no spillovers occur in this case study, while in the

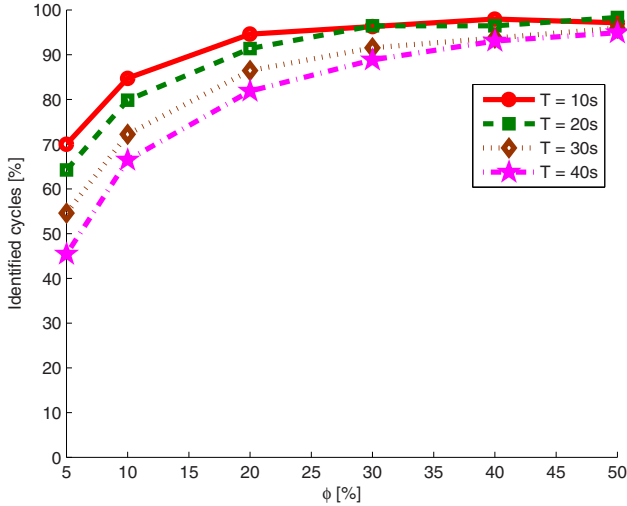


**Fig. 5.** The ground truth (black) and estimated ( $\phi = 20\%$ ,  $T = 20s$ ) queue profiles for all 4 intersections based on the proposed method (red), and based on the uniform arrival assumption (blue) (a). The MAE between the ground truth and estimated queue length with different penetration rates and sampling intervals in case of no noise (b). The MAE between the ground truth and estimated queue length with different penetration rates and sampling intervals in case of noisy measurements (c). (See color figure in online version.)

following section an extension of the proposed methodology handling spillovers is presented.

Figure 5a illustrates the estimated ( $\phi = 20\%$  and  $T = 20s$ ) and ground truth queue profiles for all intersections and demonstrates that the performance of proposed queue profile estimation method is promising and can capture the fundamental characteristics of queue profiles without any information of arrival distribution and signal settings. Further, to scrutinize the proposed

method robustness to measurement noise and map-matching errors, we add an error term to position and velocity measurements. The position and velocity error terms are assumed as a normal random variable with zero mean and standard deviations equal to 2.5 (m) and 1 (m/s), respectively. (These values are relatively pessimistic considering technological advancements and developments of filtering, estimation and fusion algorithms, see Xu et al., 2010 and Witte and



**Fig. 6.** The percentage of identified cycles in step 2 of the method with different penetration rates and sampling intervals. (See color figure in online version.)

Wilson, 2004.) Figures 5b and c depict the MAE between the ground truth queue size and the estimated queue size averaged over 10 runs for penetration rates,  $\phi = 5, 10, 20, 30, 40, 50\%$ , and sampling intervals,  $T = 10, 20, 30, 40$ s for two cases without and with measurement errors. It is apparent in Figure 5b that even in case of sparse probe data the proposed method results are more accurate than the uniform arrival estimation (with the integration of residual queues modeling of Skabardonis and Geroliminis, 2008). It is also clear that higher penetration rates and lower sampling intervals are not necessary as the error is close to one vehicle for  $T = 10$ s and  $\phi = 30\%$ . Moreover, Figure 5c reveals that the proposed method is robust to measurement errors. We also assess robustness of the proposed method to measurement errors closer to the state-of-the-art with standard deviations equal to 1 (m), and 0.2 (m/s) and the outcomes are statistically similar to no error case. Note that the proposed method parameters are similar in both case studies, which demonstrates the method is insensitive to a sensible range of parameters.

To scrutinize the method performance with very sparse probe data (e.g., 5% penetration rate), Figure 6 shows the percentage of identified cycles, as the number of identified cycles in step 2 of the method divided by total number of cycles (a similar performance measure is also discussed in Ban et al., 2011). For  $T = 20$ s and  $\phi = 20\%$ , the percentage of correctly identified cycles is above 90%, while evidently cycles with lower queue size and associated with less congested traffic states have higher probability to be missed in case of very

sparse probe data (see 5 missed cycles in Figure 5a). In this example, the maximum queue size is 29 [veh] and the mean queue size is 7.9 [veh] (that is the MAE of a zero queue estimation, equivalent to no estimated queue profile or missed cycles). In case prior signal setting information is available (e.g., lower and upper bounds of red and green phases), the performance of the method can be improved. In Section 6, a method based on historical data is discussed to tackle missed cycles.

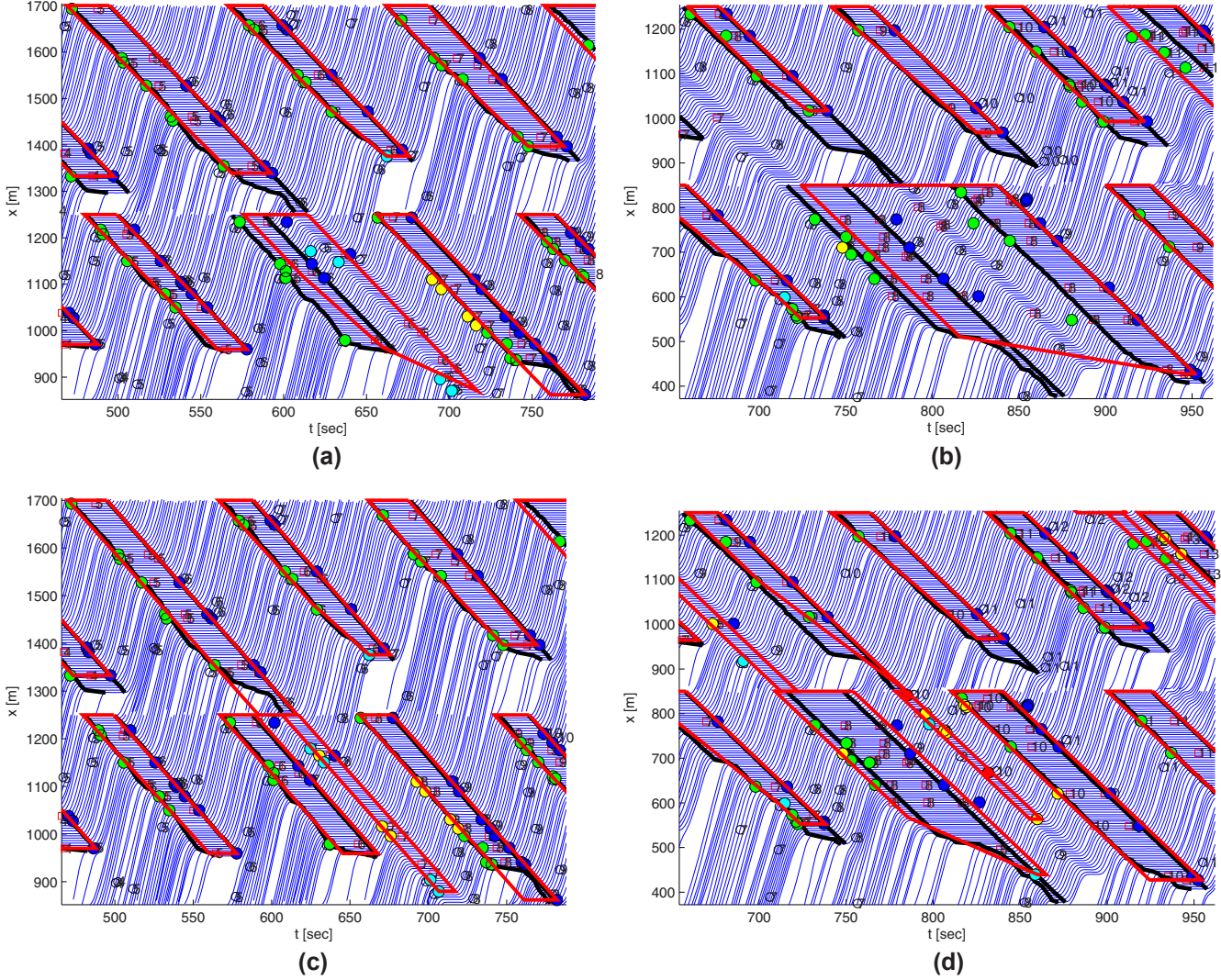
## 5 SPILLOVER BAYESIAN INFERENCE

In this section, we explore the effect of spillovers on the proposed queue profile estimation method, subsequently introduce an elegant probabilistic spillover identification procedure based on Bayesian inference and probe data observations, and finally incorporate the spillover identification within the queue profile estimation method.

### 5.1 Spillover inference

Spillovers occur during oversaturated conditions when exceeding queue at the downstream link impedes the arrivals from the upstream link, while the signal phase is green. This creates a *de facto* red phase that decreases the outflow of the link and consequently the network outflow. Recent studies have addressed the problem of spillover identification using conventional loop detectors (e.g., Geroliminis and Skabardonis, 2011; Wu et al., 2010) and control strategies to avoid spillovers (e.g., Christofa et al., 2013). However, the literature of methods identifying queue spillovers in urban networks with probe data is limited.

Given that in this article signal settings are unknown, occurrence of spillovers (and hence the *de facto* red) causes difficulties for step 2 of the proposed method (Section 3.2). In step 2, the stopped data points are clustered to cycles based on the projection profile method, while due to randomness of probe data, the clustering procedure disregards small number of zero-valued bins inside a cluster of positive-valued bins. Note that zero-valued bins of the accumulator vector represent intervals (the time between two stripes in x-t plane with slope  $w$ ) without any stopped data points. So, if the *de facto* red occurs, step 2 might consider this red as a part of the nearest signal red phase (see Figure 7a), or even as a whole with both adjacent red phases (see Figure 7b) depending on the *de facto* red spatiotemporal extent. However in both cases, step 2 results in a longer red phase with a long interval (inside the long red phase) without any stopped data point observation, and also a



**Fig. 7.** The ground truth (black) and estimated ( $\varphi = 10\%$ ,  $T = 20$ s) queue profiles based on the proposed method (red). The adverse effect of spillover and de facto red results in longer than usual red phases,  $R_{r_j}$ , with long  $Z_{r_j}$ , and also a systematic error in leaving and joining point estimation (a). The adverse effect of the de facto red when it combines with two adjacent red phases (b). The two modifications of the spillover inference module on  $x_j^{\min}$  and  $r_j$  (c). The adjustments of the spillover inference module when the de facto red combines with two adjacent red phases (d). Note that blue lines are vehicle trajectories and are depicted for illustration purposes. (See color figure in online version.)

systematic error in the estimation of leaving and joining points. The key idea of the proposed spillover identification method is to integrate these two indications from the upstream link along with the length of the queue into a probabilistic inference system to identify the occurrence of spillovers.

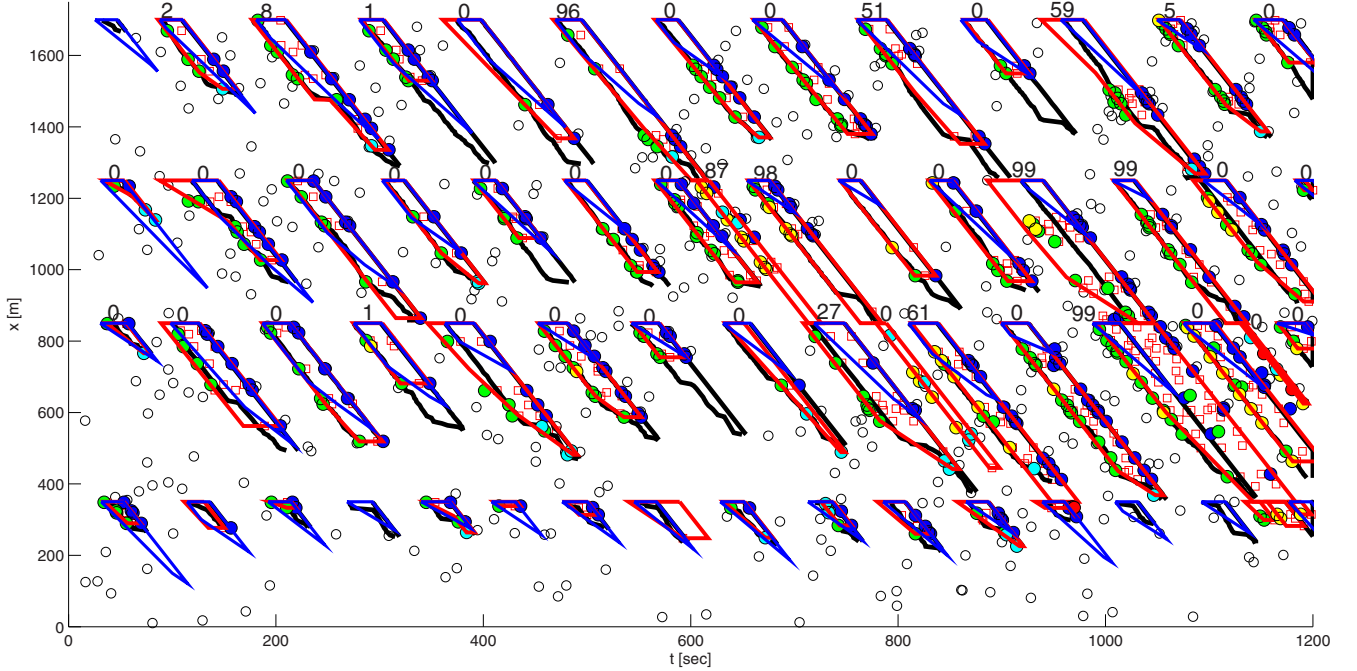
Bayesian inference method provides a probabilistic framework to update the prior probability, that is,  $pr(H)$ , of a set of hypotheses,  $\{H_1, H_2, \dots, H_h\}$ , while further evidence regarding the hypotheses is acquired. Bayesian inference method estimates the poste-

rior probability of the hypotheses given the evidences, that is,  $pr(H|E)$ , as

$$pr(H_i|E) = \frac{pr(E|H_i)}{\sum_h pr(E|H_h) pr(H_h)} pr(H_i) \quad (17)$$

where for a set of independent and identically distributed evidences,  $\{e_1, e_2, \dots, e_k\}$ ,

$$pr(E|H_i) = \prod_k pr(e_k|H_i) \quad (18)$$



**Fig. 8.** The ground truth (black) and estimated ( $\varphi = 10\%$ ,  $T = 20s$ ) queue profiles for all 4 intersections based on the proposed method (red), and based on the uniform arrival assumption (blue). The values (in %) are the spillover probability associated for each queue profile. (See color figure in online version.)

To infer the probability of a spillover occurrence at cycle  $j$  of the downstream link, we need to pre-process steps 1 and 2 for the upstream link to identify corresponding red phases. Note that by definition, spillovers occur during green phases, whereas this duration will be accounted as a (*de facto*) red phase in the proposed queue profile estimation method. So, if queue profile  $j$  spills over, it must reach to the upstream intersection during a *de facto* red phase,  $r_j$ . As mentioned above, duration of  $r_j$ ,  $R_{r_j}$ , is longer than other (normal) red phases. Likewise, the maximum interval (the time between two stripes in  $x$ - $t$  plane with slope  $w$ ) within  $r_j$  without any stopped point,  $Z_{r_j}$ , is also longer than the corresponding value of other red phases. Therefore, the magnitude that  $R_{r_j}$  and  $Z_{r_j}$  are outliers and greater than the rest of samples in their corresponding set (respectively  $R$  and  $Z$  values for red phases of the upstream link) can be regarded as evidence for spillover occurrence. An intuitive equation to designate the extent of being outlier is as follows:

$$Q_a = \frac{a - \bar{A}}{\sigma(A)}, \quad a \in A \quad (19)$$

where  $a$  is the sample and  $A$  denotes the set with mean and range (difference between the largest and smallest values), respectively as  $\bar{A}$  and  $\sigma(A)$ .

Lastly, we consider the evidences in Bayesian inference as,  $e_1$ : the position of the furthest queued vehicle

from the intersection,  $x_j^{\min}$ ;  $e_2$ :  $Q_{R_{r_j}}$ ;  $e_3$ :  $Q_{Z_{r_j}}$ . The two complementary possible hypotheses are  $H_1$ : occurrence of spillover and  $H_2$ : no occurrence of spillover. Furthermore, the likelihood functions,  $pr(e_k|H_i)$ ;  $i = 1, 2$ ;  $k = 1, 2, 3$ , are

$$pr(e_1|H_1) = \frac{1 - e^{-\left(\frac{x^{\max} - x^{\min}}{x^{\max}}\right)}}{1 - e^{-1}} \quad (20a)$$

$$pr(e_2|H_1) = \max(\delta, Q_{R_{r_j}}) \quad (20b)$$

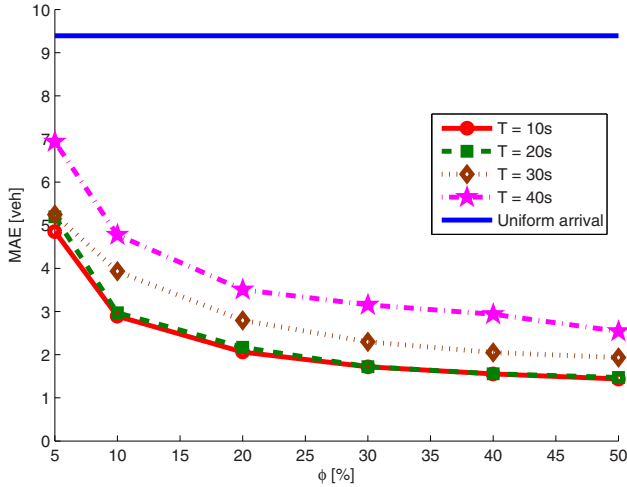
$$pr(e_3|H_1) = \max(\delta, Z_{R_{r_j}}) \quad (20c)$$

$$pr(e_k|H_2) = 1 - pr(e_k|H_1); \quad k = 1, 2, 3 \quad (20d)$$

where  $\delta$  is a small predefined positive value near zero to avoid negative and zero values. More sophisticated inference methods or optimizing the likelihood functions are a future research direction.

## 5.2 Integration of spillover identification into queue profile estimation

The Bayesian spillover inference is an intermediate step between steps 2 and 3 of the queue profile estimation method that assigns a spillover probability to each queue profile. With this inference method, queue



**Fig. 9.** The MAE between the ground truth and estimated queue length with different penetration rates and sampling intervals in case of spillovers. (See color figure in online version.)

profiles are grouped into two distinct sets, ones with spillover probability close to zero and ones close to 1, see Figure 8. Thus, a prescribed threshold distinguishes the two sets. Two modifications should be applied to queue profile  $j$  that is labeled as spillover, that is, its associated spillover probability is higher than the predefined threshold. First,  $x_j^{\min}$  should be set equal to the position of the link upstream (instead of Equation (12a)). Second,  $r_j$  should be divided into the *de facto* red and the normal signal red phase during step 2 for the upstream link. The division should be placed at the interval corresponding to  $Z_{r_j}$ . Figure 7c illustrates the two modifications that have been applied to the example in Figure 7a. In case the *de facto* red was considered with the two adjacent red phases (e.g., Figure 7b),  $r_j$  should be split into two intervals accordingly. A simple procedure is to check if the second largest interval without a stopped point observation is close enough to  $Z_{r_j}$ . Figure 7d illustrates the same example in Figure 7b with modification on  $x_j^{\min}$  and dividing  $r_j$  into two signal red phases and a *de facto* red. The spillover inference procedure acquires the correlation of queue spatiotemporal extent between two successive links and can identify the spillbacks that propagate to more than one upstream link.

Figure 8 illustrates the queue profiles of a case study with the same topology but higher demand with oversaturated conditions and spillovers. It is apparent that the proposed queue profile estimation method with sparse probe data captures the evolution and fundamental attributes of queues and spillovers in the signalized arterial. Furthermore, Figure 9 shows the MAE between the ground truth queue size and the estimated queue size

averaged over 5 runs for various penetration rates and sampling intervals. It is clear that the outcome of the proposed method is more precise than the estimation of uniform arrival procedure. Even if the effect of oversaturated condition has been integrated in the uniform arrival estimation (see for example the blue trapezium in Figure 8), the developed method is superior even with 5% penetration rates. If oversaturated conditions are not embedded in the uniform arrival estimation and only triangular queue profiles are applied, the errors would be much larger.

## 6 DISCUSSION

This article has presented a method to estimate queue shockwave profiles at signalized intersections in urban networks, based on LWR traffic theory and dispersed data of position and velocity from probe vehicles. The significance of the proposed method is the estimation of shockwaves without any explicit information of signal settings and arrival traffic flow patterns. The method incorporates the physics of traffic, longitudinal kinematics, and interdependencies between adjacent links into an optimization framework. The outcome of the proposed method for two case studies with field and simulated data seems promising, especially for congested conditions. The method is robust to noisy measurements and more accurate once compared to a uniform arrival queue profile estimation procedure. The effect of penetration rates and sampling interval of probe data on the performance of the method is also investigated. In addition, the method is applicable to oversaturated conditions and with integration of Bayesian inference; the method infers the probability of spillover occurrence.

The unique feature of queue shockwave profile estimation is that it provides the comprehensive dynamics of the urban traffic flow, which can be further applied for performance measurement and control applications. For instance, the estimated queue shockwave profile can be readily utilized for delay analysis (mean and distribution) and queue length estimation. Moreover taking into account that the arrival traffic flow distribution can be estimated from the estimated queue profile, a noteworthy application of the proposed method is vehicle trajectory reconstruction, which is not limited to probe vehicles and particularly can be applied for all vehicles. In other words, the proposed method estimates trajectories of all vehicles given sample data of only sensor-equipped probe vehicles. Estimation of detailed trajectory of vehicles can be further applied for (rough) emission or fuel consumption estimation models. Nevertheless, further research is needed to scrutinize and improve the accuracy of these models.



Estimation of vehicles trajectories also enables the method to tackle the travel time decomposition problem (Hellinga et al., 2008), which is about how to decompose the travel time between two successive probe data with high sampling interval (e.g., 3 min) to individual link travel time. Link travel time (mean and distribution) is a crucial performance index that can be applied for route travel time distribution estimation and travel time reliability models (see for example, Ramezani and Geroliminis, 2012). Although the proposed method requires a sampling interval that produces enough probe data to estimate the queue profiles, notably it can provide the link travel time for all the vehicles and not only for probe vehicles. This is another future direction of research.

The queue profile estimation method also provides signal settings, that is, phase durations, at each intersection. However, the method cannot identify a cycle without any probe data. So the proposed method needs to be further explored to tackle missing cycles, which might occur in case of low penetration rates and high sampling intervals. To this end, prior assumptions on minimum or maximum phase duration can enhance the method regarding missing cycles, while the remainder of the methodology remains the same. Also, given the estimated queue shockwaves, one can calculate a historical value for phase durations based on the average length of red and green intervals (except the outliers). Then long (outlier) green intervals can be detected, which probably include missed cycles, and be decomposed to several green and red phases, based on the historical phase durations. This provides a proxy for number of missed cycles and, assuming uniform arrival rate, an estimator of back of queue length. However, this is not accurate for queue length estimation as biased information of signal settings (e.g., several seconds lag) leads to significant queue measurement error.

A future research priority is to extend the queue profile estimation model to consider multilane roads, where in case of lanes with different channelization characteristic (e.g., left turn), the queue development and dissipation varies among different lanes. The challenge is that although the input data of the proposed method include location and velocity of probe vehicles, the state-of-practice technology is immature to provide accurate enough lane specific measurements. Further research can attempt to infer the lane features of probe vehicles based on their path characteristics (Rahmani and Koutsopoulos, 2013) or possible data fusion with loop detector data that are able to provide lane information. In addition, for queue shockwave estimation in multimodal networks, vehicle classifications based on loop detector data (Wei et al., 2013) or probe data are an intermediate step.

Even though the method does not require arrival traffic flow distribution, outflow measurements from upstream input links can provide significant information for queue estimation in case of limited probe data. Data fusion of sparse probe data with loop detector data (e.g., see Treiber et al., 2011; van Lint and Hoogendoorn, 2010) can also facilitate the developed method in case of small penetration rates (e.g., less than 5%) and improve the accuracy of the model.

Integration of the proposed methodology in real time traffic control is a tall order, but a feasible solution given the new era of data for vehicle positioning and crowdsourcing. Existing traffic control strategies that require queue size information (e.g., see Varaiya, 2013; Aboudolas et al., 2009) can utilize this work.

## ACKNOWLEDGMENTS

This research was supported by Swiss National Science Foundation, Grant #200021-130165. The authors would like to thank the anonymous reviewers for their constructive comments.

## REFERENCES

- Aboudolas, K., Papageorgiou, M. & Kosmatopoulos, E. B. (2009), Store-and-forward based methods for the signal control problem in large-scale congested urban road networks, *Transportation Research Part C*, **17**(2), 163–74.
- Akçelik, R. (1998), A queue model for HCM 2000, Technical Note. ARRB Transport Research Ltd, Vermont South, Australia.
- Ban, X., Hao, P. & Sun, Z. (2011), Real time queue length estimation for signalized intersections using travel times from mobile sensors, *Transportation Research Part C*, **19**(6), 1133–56.
- Bhaskar, A., Chung, E. & Dumont, A. (2011), Fusing loop detector and probe vehicle data to estimate travel time statistics on signalized urban networks, *Computer-Aided Civil and Infrastructure Engineering*, **26**, 433–50.
- Cheng, Y., Qin, X., Jin, J. & Ran, B. (2012), An exploratory shockwave approach to estimating queue length using probe trajectories, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, **16**(1), 12–23.
- Christofa, E., Argote, J. E. & Skabardonis, A. (2013), Arterial queue spillback detection and signal control based on connected vehicle technology, *Transportation Research Record* **2356**, 61–70.
- Comert, G. & Cetin, M. (2011), Analytical evaluation of the error in queue length estimation at traffic signals from probe vehicle data, *IEEE Transactions on Intelligent Transportation Systems*, **12**(2), 563–73.
- Cortes, C. & Vapnik, V. (1995), Support-vector networks, *Machine Learning*, **20**(3), 273–97.
- Darroch, J. N., Newell, G. F. & Morris, R. W. J. (1964), Queues for a vehicle-actuated traffic light, *Operations Research*, **12**, 882–94.

- Dion, F., Rakha, H. & Kang, Y. (2004), Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections, *Transportation Research Part B*, **38**(2), 99–122.
- Erera, A. L., Lawson, T. W. & Daganzo, C. F. (1998), Simple, generalized method for analysis of traffic queue upstream of a bottleneck, *Transportation Research Record*, **1646**(1), 132–40.
- Geroliminis, N. & Skabardonis, A. (2011), Identification and analysis of queue spillovers in city street networks, *IEEE Transactions on Intelligent Transportation Systems*, **12**(4), 1107–15.
- Ghosh-Dastidar, S. & Adeli, H. (2006), Neural network-wavelet microsimulation model for delay and queue length estimation at freeway work zones, *Journal of Transportation Engineering*, **132**(4), 331–41.
- Groot, N., De Schutter, B. & Hellendoorn, H. (2013), Integrated model predictive traffic and emission control using a piecewise-affine approach, *IEEE Transactions on Intelligent Transportation Systems*, **14**(2), 587–98.
- Hao, P., Ban, X., Bennett, K. P., Ji, Q. & Sun, Z. (2012), Signal timing estimation using sample intersection travel times, *IEEE Transactions on Intelligent Transportation Systems*, **13**(2), 792–804.
- Hellinga, B., Izadpanah, P., Takada, H. & Fu, L. (2008), Decomposing travel time measured by probe-based traffic monitoring systems to individual road segments, *Transportation Research Part C*, **16**(6), 768–82.
- Herrera, J. C., Work, D. B., Herring, R., Ban, X. & Bayen, A. (2010), Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile century field experiment, *Transportation Research Part C*, **18**(4), 568–83.
- Hofleitner, A., Herring, R., Abbeel, P. & Bayen, A. (2012), Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network, *IEEE Transactions on Intelligent Transportation Systems*, **13**(4), 1679–93.
- Izadpanah, P., Hellinga, B. & Fu, L. (2009), Automatic traffic shockwave identification using vehicles' trajectories, in *Proceedings of the 88th Annual Meeting of the Transportation Research Board (CD-ROM)*, January, Washington, D.C.
- Jelaca, V., Castaneda, J. O. N., Pizurica, A. & Philips, W. (2012), Image projection clues for improved real-time vehicle tracking in tunnels, in *Proceedings of the SPIE – The International Society for Optical Engineering*, 8301, Intelligent Robots and Computer Vision XXIX: Algorithms and Techniques, art. no. 83010C, January, California, USA, DOI: 10.1117/12.908873.
- Jenelius, E. & Koutsopoulos, H. N. (2013), Travel time estimation for urban road networks using low frequency probe vehicle data, *Transportation Research Part B*, **53**, 64–81.
- Jiang, X. & Adeli, H. (2005), Dynamic wavelet neural network model for traffic flow forecasting, *Journal of Transportation Engineering ASCE*, **131**(10), 771–79.
- Kwong, K., Kavalier, R., Rajagopal, R. & Varaiya, P. (2009), Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors, *Transportation Research Part C*, **17**(6), 586–606.
- Lighthill, M. J. & Whitham, G. B. (1955), On kinematic waves. I: Flood movement in long rivers. II: A theory of traffic flow on long crowded roads, *Proceedings of Royal Society A* **229**, 281–345.
- McNeill, D. R. (1968), A solution to the fixed cycle traffic light problem with compound Poisson arrivals, *Journal of Applied Probability*, **5**(3), 624–35.
- Michalopoulos, P. G., Stephanopoulos, G. & Stephanopoulos, G. (1981), An application of shock wave theory to traffic signal control, *Transportation Research Part B*, **15**(1), 31–51.
- Newell, G. F. (1960), Queues for a fixed-cycle traffic light, *The Annals of Mathematical Statistics*, **31**, 589–97.
- Newell, G. F. (1965), Approximation methods for queues with application to fixed-cycle traffic light, *SIAM Review* **7**, 223–40.
- NGSIM, (2006). Next Generation Simulation. <http://ngsim.fhwa.dot.gov/>.
- Rahmani, M. & Koutsopoulos, H. (2013), Path inference from sparse floating car data for urban networks, *Transportation Research Part C*, **30**, 41–54.
- Ramezani, M. & Geroliminis, N. (2012), On the estimation of arterial route travel time distribution with Markov chains, *Transportation Research Part B*, **46**(10), 1576–90.
- Ramezani, M. & Geroliminis, N. (2013), Exploiting probe data to estimate the queue profile in urban networks, in *Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 1817–22.
- Richards, P. J. (1956), Shock waves on the highway, *Operations Research* **4**, 42–51.
- Skabardonis, A. & Geroliminis, N. (2005), Real-time estimation of travel times on signalized arterials, *16th International Symposium on Transportation and Traffic Theory*, University of Maryland.
- Skabardonis, A. & Geroliminis, N. (2008), Real-time monitoring and control on signalized arterials, *Journal of Intelligent Transportation Systems*, **12**(2), 64–74.
- Skabardonis, A., Geroliminis, N. & Christofa, E. (2013), Prediction of vehicle activity for emissions estimation under oversaturated conditions along signalized arterials, *Journal of Intelligent Transportation Systems*, **17**(3), 191–99.
- Sun, Z., & Ban, X. (2013), Vehicle trajectory reconstruction for signalized intersections using mobile traffic sensors, *Transportation Research Part C*, **36**, 268–83.
- Treiber, M., Kesting, A. & Wilson, R. E. (2011), Reconstructing the traffic state by fusion of heterogeneous data, *Computer-Aided Civil and Infrastructure Engineering*, **26**, 408–19.
- van Lint, J. W. C. & Hoogendoorn, S. P. (2010), A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways, *Computer-Aided Civil and Infrastructure Engineering*, **25**, 596–612.
- Varaiya, P. (2013), Max pressure control of a network of signalized intersections, *Transportation Research Part C*, **36**, 177–95.
- Vigos, G., Papageorgiou, M. & Wang, Y. (2008), Real-time estimation of vehicle-count within signalized links, *Transportation Research Part C*, **16**, 18–35.
- Viti, F. & Van Zuylen, H. J. (2010), Probabilistic models for queues at fixed control signals, *Transportation Research Part B*, **44**(1), 120–35.
- Vlahogianni, E. I. & Karlaftis, M. G. (2013), Fuzzy-entropy neural network freeway incident duration modeling with single and competing uncertainties, *Computer-Aided Civil and Infrastructure Engineering*, **28**(6), 420–33.
- Webster, F. V. (1958), *Traffic Signal Settings*, Road Res. Lab. Tech. Paper 39, London, U.K.

- Wei, H., Liu, H., Ai, Q., Li, Z., Xiong, H. & Coifman, B. (2013), Empirical innovation of computational dual-loop models for identifying vehicle classifications against varied traffic conditions, *Computer-Aided Civil and Infrastructure Engineering*, **28**(8), 621–34.
- Witte, T. H. & Wilson, A. M. (2004), Accuracy of non-differential GPS for the determination of speed over ground, *Journal of Biomechanics*, **37**, 1891–98.
- Wu, X. & Liu, H. X. (2011), A shockwave profile model for traffic flow on congested urban arterials, *Transportation Research Part B*, **45**(10), 1768–86.
- Wu, X., Liu, H. X. & Geroliminis, N. (2011), An empirical analysis on the arterial fundamental diagram, *Transportation Research Part B*, **45**(1), 255–66.
- Wu, X., Liu, H. X. & Gettman, D. (2010), Identification of oversaturated intersections using high-resolution traffic signal data, *Transportation Research Part C*, **18**(4), 626–38.
- Xu, H., Liu, H., Tan, C. & Bao, Y. (2010), Development and application of an enhanced Kalman filter and global positioning system error-correction approach for improved map-matching, *Journal of Intelligent Transportation Systems*, **14**(1), 27–36.