Queue Length Estimation for Signalized Intersections under Partially Connected Vehicle Environment

Lu Wei,1 Jin-hong Li,1 Li-wen Xu,2 Lei Gao,1 and Jian Yang1

1Beijing Key Lab of Urban Road Traffic Intelligent Control Technology, North China University of Technology, Beijing 100144, China
2College of Science, North China University of Technology, Beijing 100144, China

Correspondence should be addressed to Lu Wei; wltop001@sina.com

Received 18 January 2022; Revised 21 March 2022; Accepted 21 April 2022; Published 2 May 2022

1. Introduction

Queue length is one of the most important performance measures for traffic signal control at urban road intersections. Many researchers have conducted queue length estimation based on various types of vehicle detector data [1–3]. The existing conventional methods including cumulative curve theory [4], shockwave theory [5–7], and queuing theory [8] mainly rely on fixed-location detectors (e.g., loops, magnetometer sensors, and virtual loops in cameras). The excellent data quality of these detectors could help produce well estimated results of lane-based queue length. For example, Ma et al. [9, 10] proposed a method to estimate lane-based queue length using the travel time data collected by video imaging detectors based on Ban et al. methods [11]. Two models were combined to derive the maximum queue length model. The results show that the proposed method performs a better precision compared to the existing methods in similar concepts. However, the disadvantage of these fixed-location detectors is expensive installation and maintenance costs, which leads to a low spatial coverage of urban intersections [12]. Moreover, the fixed-location detectors can only capture vehicle status at a specific point of the road section and cannot continuously track the vehicle trajectory change and running speed to support a more cost-effective, higher accuracy, and robust queue length estimation method. The installed location of these detectors is also an influence factor of the estimation results. For example, in oversaturated situation, when the queued vehicles exceed the detect zone, the total queue length usually cannot be acquired.

The rapid development of intelligent connected vehicles (CV) provides massive high-resolution vehicle trajectory data for urban traffic management and control. Different
from fixed-location detector data collecting, which requires additional equipment installation and maintained, the vehicle trajectory data is generated from the onboard GPS devices, while the moving CVs cost much less. Besides, with a larger spatial coverage, the trajectory data can capture dynamic spatial-temporal characteristics of traffic flow in the entire road network. Therefore, queue length estimation for signalized intersections based on CV data has gradually become a hot research topic and gains more and more recognition and attentions [13, 14]. Most of the existing queue length estimation methods can be classified into two categories: deterministic and stochastic methods [15]. The deterministic methods estimate queue length by reconstructing traffic flow models such as shockwaves models [16] or input-output model [17]. Based on high-precision trajectory data, Cheng et al. [18] developed a threshold-based critical point extraction algorithm to capture the changes in vehicle dynamics and then reconstruct shockwaves to estimate queue length cycle-by-cycle. The method was evaluated on both simulation and NGSIM data and the results indicated that promising outcome. Instead of using traffic volume or occupancy as the input as most queue length estimation methods based on fixed-location detectors, Ban et al. [11] estimated real-time queue length at signalized intersections using sample travel times from mobile traffic sensors. By recognizing the delay pattern changes such as nonsmoothness and discontinuities, the real-time queue length was constructed. Yin et al. [19] also proposed a queue length estimation method using low penetration vehicle trajectory data as the only input. Based on the combination of Kalman Filtering and shockwave theory, a state-space model with two state variables and the system noise was established to characterize the stochastic property of queue forming. By calculating queue-forming and queue-discharging waves, the maximum queue length can be estimated. Wang et al. [20] developed a Kalman filtering method based on the idea of input-output models to estimate the real-time queue length for an isolated intersection with CV data. The simulation results showed that the proposed method is suitable for different traffic demand levels without being affected by volume/capacity ratio. Ramezani and Geroliminis [21] integrated the traffic flow shockwave analysis and trajectory data mining techniques to extract joining and leaving critical points and then modeled queue formation process by a piecewise linear function. The significance of this study is its applicability in oversaturated conditions. These deterministic methods could produce promising estimates when the penetration rate of vehicle trajectories is sufficient high. However, due to the low penetration rates at present or even in the near future (for example, the penetration rate of vehicles with detectable trajectories from map navigation companies such as Didi, Baidu, is often less than 10% on average), these methods face the challenge of inaccurate and unstable estimation in practical applications.

Stochastic methods consider the traffic arrival as a stochastic process and estimate the distribution of queue length based on probabilistic or statistical models under sampled trajectory data environment [22]. The expectation value of the estimated distribution is then derived as total queue length. For example, Comert and Cetin [23] developed an analytical formulation based on conditional probability distributions for estimating the expected queue length and its variance from the location information of probe vehicles at signalized intersections. Various numerical results showed that only the information of the last stopped probe vehicle in the queue is sufficient for queue length estimation. The relationship between penetration rate and estimation accuracy is also analyzed in this study. To further improve the accuracy and generalization of the model estimation, Comert and Cetin incorporated the stop-line detector data [24] and the time when the last probe vehicle joined the queue [25] into the queue length estimation model. Subsequently, they further proposed a series of estimators for primary parameters of the queue length estimation model, such as the penetration rate and the arrival rate under the Poisson distribution assumption [26]. However, these models assume that vehicle arrivals follow uniform Poisson distribution, which is an indispensable a priori knowledge. In order to compensate the weakness of sparse vehicle trajectory data, Tan et al. [12] fuse license plate recognition (LRP) data and probe vehicle data for lane level queue length estimation after fully analyzing the advantages of the two data sources. Historical trajectory data and LPR data are used to calibrate two-dimensional probability density distribution for discharge headway and stop-line crossing time for queued and nonqueued vehicles based on kernel density estimation. Then, the lane-based queue length is obtained by using the Bayesian theory. To extend the queue length estimation model, Zhao et al. [27] studied the estimation of two parameters: the probe vehicle penetration and the queue length distribution using historical trajectory data. A maximum likelihood estimator is adopted, and EM algorithm is applied to solve the problem iteratively. Validation results on simulated vehicle trajectory data show that the proposed method could estimate the parameters accurately.

In summary, for stochastic methods, a priori knowledge such as penetration rate, queue length distribution, and arrival pattern is quite important. Noticeably, though some researchers have conducted the estimation of penetration rate and queue length distribution, the arrival pattern has not been well studied. Motivated by this issue and the queue length estimation model proposed by Comert and Cetin [23], this paper proposed a queue length estimation method with unknown arrival pattern under partially connected vehicle environment. The main contributions of this study include the following:

1. By utilizing sampled CV trajectory information in the queue at signalized intersection, a method of estimating the arrival rate of all vehicles was proposed based on empirical Bayes.

2. To estimate the hyperparameters of the prior Gamma distribution, we constructed the marginal distribution of CV arrivals and then applied maximum likelihood estimator to solve the hyperparameters estimation. Finally, the estimated queue length is derived as the expectation of conditional queue length distribution.
In order to validate the proposed method, the secondary development of VISSIM simulation software based on Python is conducted to establish a partially CV data environment with adjustable penetration rates.

The rest of this paper is organized as follows. Section 2 describes the problem and provides the notations used in this paper. Section 3 introduces the queue length estimation model and presents the empirical Bayesian method for arrival rate estimation and maximum likelihood estimation of hyperparameters. In Section 4, a partially connected vehicles simulation scenario is established, and experiments analysis is also conducted in VISSIM simulation software. Conclusions and future works are summarized in Section 5.

2. Problem Statement and Notation

At urban signalized intersections, vehicles traveling on the road start to form a queue behind the stop line when the traffic signal turns red. Figure 1 shows the snapshot of the queued vehicles on an approach to the intersection. The blue rectangles indicate the connected vehicles whose position, speed, and other information can be collected in real-time. The white rectangles indicate normal vehicles, and their states cannot be collected without other additional traffic sensors. The primary objective of this research is estimating the total queue length based on the observable connected vehicle information (e.g., the number and stopped positions of the connected vehicles in the queue). Some assumptions are made to simplify the discussion. First, the intersection traffic is assumed to be undersaturated and vehicle arrivals follow Poisson distribution. Second, the signal timing of the intersection is known. Third, at least one CV exists in the queue to make sure that the proposed method can be applied.

For a better illustration the proposed method, a summary of the notation is provided in Table 1.

Take the scenario described in Figure 1 as example, the total queue length $N = 7$, the number of CVs in the queue $M = 2$, and the position of the last position of CV stopped in the queue $L = 5$.

3. Methodology

3.1. Queue Estimation Model. The process of queue length estimation under partially connected vehicle data environment can be concluded in the following steps:

1. Deriving the number of stopped CVs in the queue based on whether the speed is less than the stopping speed threshold.
2. Determining the position of the last stopped CV in the queue, that is, the farthest stopped CV from the stop line.
3. Calculating the distance from the last stopped CV to the intersection stop line and then deriving an initial estimated queue length based on the standard spacing headway. The initial queue length also can be treated as the lower bound of the total queue length of all vehicles.

(4) Estimating the number of vehicles that may exist after the last stopped CV based on other prior knowledge, such as queue length and (or) vehicle arrival distribution, the penetration rate of CV, etc.

Based on the observable information including the positions and number of connected vehicles, Wang et al. [14] introduce two cycle-by-cycle queue length estimators from the perspective of the probability theory: maximum likelihood estimator and expected queue length conditional on the observed partial queue, and the latter follows the model proposed by Comert and Cetin [23]. In this research, we continue to use it as the basic model for queue length estimation.

Assume that the probability of a vehicle being a connected vehicle follows Bernoulli distribution, and the probability of success is denoted as $p$, as the penetration rate of CV. Given the total queue length of all vehicles, denote $N = n$, and the number of CVs in the queue $M$ follows binomial distribution and can be written as

\[ P(M = m| N = n) = \binom{n}{m} p^m (1-p)^{n-m}, \]  

where $p$ is the penetration rate of connected vehicles, and $0 \leq p \leq 1$. $m$ represents the number of CV in the queue, and the value range is $m = 0, 1, 2, \ldots, n$.

In addition, given $N = n$ and $M = m$, the probability distribution for the position of last stopped CV can be derived from the number of possible combinations [23].

Table 1: Summary of notation.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>Total queue length</td>
</tr>
<tr>
<td>$M$</td>
<td>Number of CVs in the queue</td>
</tr>
<tr>
<td>$L$</td>
<td>Last position of stopped CVs in the queue</td>
</tr>
<tr>
<td>$p$</td>
<td>Penetration rate of CV</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Arrival rate of all vehicles</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Arrival rate of CVs</td>
</tr>
<tr>
<td>$t_R$</td>
<td>Duration of red signal</td>
</tr>
</tbody>
</table>
where \( l \) denotes the position of last stopped CV and \( l = m, m + 1, m + 2, \ldots, n \).

According to Bayes’ rule, the total queue length distribution condition on the position of the last stopped CV and the number of CVs in the queue can be written as follows:

\[
P(N = n \mid L = l, M = m) = \frac{\binom{l - 1}{m - 1}}{\binom{n}{m}}
\]

Substituting equations (1) and (2) into equation (3), and applying the Total Probability Law, equation (3) can be calculated as

\[
P(N = n \mid L = l, M = m) = \frac{(1 - p)^{\eta}P(N = n)}{\sum_{k=l}^{n} (1 - p)^{k} P(N = k)}
\]

Equation (4) implies that the total queue length distribution only depends on the position of the last stopped CV and the penetration rate.

In this paper, the penetration rate is assumed to be known. Actually, in many engineering applications, the penetration rate usually can be obtained from the analysis of the connected vehicle trajectory data and other fixed detection data. For example, Figure 2 shows the volume comparison between license plate recognition (LPR) data collected by camera and trajectory data provided by the navigation company in an area of Lianyungang city in January 2019. The penetration rate can be estimated as the ratio of sampled trajectory volume and camera volume. In the case of Figure 2, the result of penetration rate calculation is about 0.061, which is consistent with the penetration rate provided by the map navigation company. Therefore, this method is verified with high practicality and accuracy.

The difference between this paper and previous studies is that the arrival rate of total vehicles is unknown and needs to be estimated based on the observable CV information. Vehicle arrivals are usually assumed to be Poisson distribution with a rate of \( \lambda \) in undersaturated scenarios [23]. The probability density function of Poisson arrivals can be expressed as

\[
P(N = n) = \frac{(\lambda t_R)^n e^{-\lambda t_R}}{n!}, \lambda > 0, n = 0, 1, 2, \ldots
\]

where \( \lambda \) is the average arrival rate, and \( t_R \) is the duration of the red signal. For the conditional probability distribution function of equation (6) can be expressed as

\[
P(M = m \mid \eta) = \frac{\eta^m e^{-\eta}}{m!}, m = 0, 1, 2, \ldots
\]

where \( \eta \) is the unknown arrival rate of CV and needs to be estimated.

Conventional approaches for arrival rate estimation include simply computing the mean of observable samples or the Maximum Likelihood estimation. However, these approaches perform effectively only when the arrival rate sample is sufficient, and the time interval is large [28]. However, for cycle-by-cycle queue length estimation, the sample size is usually small. Also, due to cycle constraints, the time interval is not large. For this case, the Bayesian estimation approach could be used by considering that the process would benefit from prior information about the probability distribution of \( \eta \). However, the assumption that the prior distribution of \( \eta \) is fully known is rather strong and may not be applicable in realistic scenarios. To overcome this limitation, this paper adopts the Empirical Bayes approach in which only the class of the prior is known. Compared to the classical approach, the Empirical Bayes approach improves...
generality and flexibility to some extent. According to the knowledge of probability and statistics, it is known that Gamma distribution is a conjugate prior over the rate of the Poisson distributions [16]. Based on this, we assume that \( \eta \sim \text{Gamma}(\nu, \sigma) \) and its probability density function can be written as

\[
g(\eta; \nu, \sigma) = \frac{e^{-\eta \sigma}}{\sigma \Gamma(\nu)} \eta^{\nu-1}, \quad \eta > 0,\tag{8}
\]

where \( \nu \) and \( \sigma \) are the hyperparameters to be estimated.

The hyperparameters can be estimated via a Maximum Likelihood procedure since the arrival of the connected vehicles during the red light is observable. Therefore, the Empirical Bayes estimation of arrival rate can be carried out in the following two steps: constructing the marginal distribution and hyperparameter estimation.

3.2.1. Constructing the Marginal Distribution. First, based on the properties of the Gamma function, for any positive number \( c \) and \( d \), we have

\[
\int_{0}^{\infty} \eta^{c-1} e^{-\eta d} d\eta = d^{c} \Gamma(c).	ag{9}
\]

By using equations (6) and (7), the marginal distribution can be derived as

\[
f_{\nu, \sigma}(m) = \int_{0}^{\infty} P(M = m|\eta) g(\eta) d\eta = \int_{0}^{\infty} \frac{\eta^{\nu+m-1} e^{-\eta} (\sigma + \eta)^{-\nu}}{\sigma \Gamma(\nu)m!} d\eta.	ag{10}
\]

Combining the conclusion of equation (9), equation (10) can be rewritten as

\[
f_{\nu, \sigma}(m) = \frac{\sigma^{m} \Gamma(\nu + m)}{(\sigma + 1)^{\nu} \Gamma(\nu)m!}.	ag{11}
\]

Equation (11) shows that the hyperparameters of Gamma distribution can also be found in the marginal distribution of the connected vehicle arrivals. Therefore, the parameter estimation problem can be transformed into a Maximum Likelihood estimation procedure according to the observable connected vehicles arrivals.

3.2.2. Hyperparameter Estimation via MLE. We suppose that the arrivals of connected vehicles during red light are independent of each other. It is denoted the observed arrivals in \( K \) signal cycles as \( \{m_1, m_2, \ldots, m_K\} \), and due to independence, the likelihood function is the product of the marginal distributions of all sampled arrivals

\[
L(m_1, \ldots, m_K|\nu, \sigma) = \prod_{i=1}^{K} f_{\nu, \sigma}(m_i).	ag{12}
\]

The above expression for the total probability is actually quite a pain to solve. Therefore, we work with the simpler log-likelihood instead of the original likelihood.

\[
L(m_1, \ldots, m_K|\nu, \sigma) = \sum_{i=1}^{K} \ln f_{\nu, \sigma}(m_i).	ag{13}
\]

Thus, the estimator \( \hat{\nu} \) of \( \nu \) and \( \hat{\sigma} \) of \( \sigma \) can be solved by the following optimization problem:

\[
\hat{\nu}, \hat{\sigma} = \arg \max_{\nu, \sigma} \sum_{i=1}^{K} \ln f_{\nu, \sigma}(m_i).	ag{14}
\]

Take the partial derivatives with respect to \( \nu \) and \( \sigma \), and the following equations can be derived:

\[
\begin{align*}
\frac{\partial L}{\partial \nu} & = \sum_{i=0}^{m-1} \frac{1}{\nu+i} - \ln(\sigma + 1), \\
\frac{\partial L}{\partial \sigma} & = \sum_{i=0}^{m-1} \frac{m}{\sigma + \nu + m}.
\end{align*}
\]

This optimization problem cannot be solved in closed-form from the above equations and requires a numerical optimization method. Therefore, this paper adopts the gradient ascent method to solve it. The gradient update equation is as follows:

\[
\begin{align*}
\nu_j & = \nu_j + \alpha \frac{\partial L}{\partial \nu} \\
\sigma_j & = \sigma_j + \alpha \frac{\partial L}{\partial \sigma},
\end{align*}
\]

where \( \alpha \) is the learning rate. After the optimal estimators of \( \nu \) and \( \sigma \) are obtained, we can further derive the posterior distribution of \( \eta \). Then, the mean of the posterior distribution \( \bar{\eta} \) is used to solve the following equation:

\[
\bar{\eta} = \lambda pt_R.	ag{17}
\]

Based on equation (17), the estimated total arrival rate \( \hat{\lambda} \) can be derived as \( \hat{\lambda} = \bar{\eta}/pt_R \).

With the derived \( \hat{\lambda} \), the arrival distribution of total vehicles can be obtained. With the known penetration rate of connected vehicles \( P(N = n) \), substituting it into equation (4) and then the conditional probability distribution of total queue length can be calculated, and the mean of the distribution can be derived as the total queue length.

4. Simulation Experiments

Queue data generated from an isolated simulated intersection in VISSIM is used to validate the proposed methodology. VISSIM is a microscopic road traffic simulator based on the individual behavior of vehicles provided by PTV in German and has been widely used for various scenarios by traffic engineers in practice as well as by researchers for developments related to urban road traffic analysis, management, and control. Despite VISSIM has offered a friendly user graphical interface (GUI) for simply and quickly establishing dedicated traffic simulation networks, for the case in this paper, the GUI is not satisfied. For this end, we use Python programming language to collect the objects, manipulate the attributes of internal objects, and specify simulation parameters dynamically. VISSIM COM
interface defines a hierarchical model in which the attributes and functions of the internal objects provided by GUI can be manipulated by external programming. Yao and Jiang [29] developed an integrated connected vehicle simulation platform of VISSIM and Python to deal with the problems of difficult implementation of complex control algorithms and time-consuming. Inspired by Yao’s works, a simulation evaluation platform based on Python for partially connected vehicle data environment is established in the following steps.

4.1. Simulation Settings. A pseudoisolated intersection with fixed signal timing in VISSIM is used to test the queue length estimator in the proposed method. The intersection has four approaches: and the western approach is selected to conduct on queue length estimation. In order to collect the total queue length of all vehicles as a ground truth baseline, a queue counter is set on the western approach. The random seed is set to a single fixed value (42) to ensure a trustworthy comparison and reproducibility.

Modeling connected and autonomous vehicles using VISSIM internally or externally has been stated by PTV Group [30]. Since the focus of this study is to estimate queue length from data generated by connected vehicles with different penetration rates, simultaneously considering the simplicity and convenience, this paper implements the internal way, which is to modify the VISSIM default vehicle types, traffic compositions, and driving behavior parameters [31].

To analyze the impact of different CV penetration rates, this paper establishes simulation scenarios with different penetration rates. First, a new vehicle type should be defined with the name set to CV, and then, adjust its attributes such as speed and acceleration distribution as needed. If a vehicle is set as a CV, its real-time locations and speeds can be acquired. By adding the CV vehicle type to the traffic composition and setting the percentage of each vehicle type, the scenarios with different CV penetration rates can be established. The specific process is shown in Figure 3.

4.2. Data Collection. As mentioned above, the VISSIM COM interface provides a strict object hierarchy for dynamic accessing. In this paper, we use the Python language to collect the real-time moving information of CV, including simulation timestamp, vehicle ID, position, and speed. Meanwhile, in order to evaluate the estimation accuracy of the proposed method, the real total queue length is also recorded by the queue counter cycle by cycle. The data interaction process is shown in Figure 4.

A sample of the CV moving data and the real total queue length recorded from VISSIM is shown in Table 2, where the allqueued field represents the actual total queue length, which is used as the benchmark of the proposed queue estimation method. Cvqueued represents the number of CVs stopped in the queue, cvpos is the positions of the CVs, and if there is more than one CV, it is separated by “-” to indicate their different stopping positions, and last_cv_pos denotes the position of last stopped CV, which is located farthest from the stop line. Among of them, cvqueued, cvpos, and last_cv_pos are the corresponding observable information for real-world scenarios.

4.3. Results and Discussion. The estimation results can be obtained by applying the method proposed in this paper on the CV data recorded via VISSIM COM interface by Python.
For sensitivity analysis, penetration rates of 5%, 10%, 20%, 50%, and 80% are chosen. The comparison results of the estimated queue length, the position of the last stopped CV, and the actual total queue length are presented in Figure 5. For each plot, the blue solid line with star markers represents the ground truth queue length collected by queue counter set in VISSIM, the orange dotted line with dot markers represents the estimated queue length, and the green dashed line with dot markers represents the position of the last stopped CV.

**Figure 5:** The estimated results with different CV penetration rates. (a) Penetration rate of 5%. (b) Penetration rate of 10%. (c) Penetration rate of 20%. (d) Penetration rate of 50%. (e) Penetration rate of 80%.
length by using the method of this study, and the green dotted line with triangle markers represents the position of the last stopped CV, which can be seen as a naïve estimation of the queue length or the lower bound of the real queue length.

As shown in Figure 5, compared with other penetration rates, the estimated queue length deviates far from the ground truth queue length in the case of low penetration rates (e.g., 5% penetration rate). The reason is that when the penetration rate is low, the observed information of CV is rather limited, which results in a rough estimation of the queue length. However, when the penetration rate increases, the queue length estimated by the proposed method is approaching the real queue length. It also can be seen that the position of the last stopped CV, which represents a naïve estimation of the ground truth queue length, also performs more accurately with the increasing penetration rates.

In order to further demonstrate the results, two indicators, that is, the normalized mean absolute error (NMAE) and the normalized root mean square error (NRMSE), are used as the metrics to measure the accuracy of the proposed method. The formulas of NMAE and NRMSE are given by equations (18) and (19):

\[
\text{NMAE} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i| / \frac{1}{N} \sum_{i=1}^{N} y_i
\]

\[
\text{NRMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 / \frac{1}{N} \sum_{i=1}^{N} y_i^2}
\]
NMAE \( (n, \tilde{n}) \) = \frac{\text{MAE} (n, \tilde{n})}{\text{mean} (|n|)} = \frac{\sum_{i=1}^{C} |n_i - \tilde{n}_i|}{\sum_{i=1}^{C} |n_i|}, \quad (18)

NRMSE \( (n, \tilde{n}) \) = \frac{\text{RMSE} (n, \tilde{n})}{\text{mean} (|n|)} = \frac{\sqrt{1/C \sum_{i=1}^{C} (n_i - \tilde{n}_i)^2}}{1/C \sum_{i=1}^{C} |n_i|}, \quad (19)

where \( C \) denotes the number of cycles, \( n_i \) denotes the ground truth of the queue length of cycle \( i \), and \( \tilde{n}_i \) denotes the corresponding estimator.

As Figure 6 shows, when the penetration rate gets higher, the performance measures in terms of estimation accuracy are better, and both NMAE and NRMSE indicators are decreasing.

When the penetration rate is determined, this paper also attempts to analyze the different stop positions of the last CV in the queue with the deviation and the estimation results. As shown in Figure 7, the deviation becomes smaller when the position of the last CV stopped in the queue appearing farther from the stop line.

Figure 7 implies that when applying the proposed method in engineering, different confidence levels can be assigned according to the positions of the last stopped CV in the queue. Thus, the reliability of the estimation method could be improved practically.

To further investigate the impact of cycle time and red time on queue length estimation, four different cycle and red time scenarios are designed as follows:

1. Cycle time: 60 s, red time: 20 s
2. Cycle time: 80 s, red time: 30 s
3. Cycle time: 80 s, red time: 40 s
4. Cycle time: 120 s, red time: 40 s

The red time indicated the red signal duration of the phase for the investigated western approach. For each scenario, the penetrations of 5%, 10%, 20%, 50%, and 80% are selected, respectively. The mean value of total real queue length and the estimated results under different penetrations are calculated as Figure 8 shows. From Figure 8, it could be easily observed that the real queue length increases with the cycle time and red time becoming larger. Note that the highest real queue length appears in cycle 80 and red time 40, because the red ratio is largest among the four scenarios, which causes more stopped vehicles and delays. In addition, as the penetration rate increases, the mean value of estimated queue length gets closer to the real queue length.

NMAE and NRMSE are also calculated and analyzed to examine the performance of the proposed method for different cycle and times under the penetration rates of 5%, 10%, 20%, 50%, and 80%. As shown in Figures 9 and 10, it is obvious that the NMAE and NRMSE benefit from the increasing penetration rates for each cycle and red time scenario. Besides, for the same penetration, the scenario of cycle 80 and red time 40 performs better. Note that the mean value of real queue length for this scenario is also the highest shown in Figure 8. Therefore, the main reason may be that when the number of the queue vehicles increases, the number of CVs also increases, which leads to more CVs being detected during a signal cycle. The last stopped CV in the queue also tends to be closer to the actual end of the entire queue.

To elaborate the effectiveness of the proposed method, we compare the performance measures of the proposed method to the existing model proposed by Comert and Cetin [23], and the arrival rate estimation is implemented from Comert et al. [26]. Figures 11(a) and 11(b) show the estimation results in terms of NMAE and NRMSE provided by the two methods under different penetration rates, respectively.

From the comparison results shown in Figure 11, the proposed method could achieve more accurate estimates than the baseline method. The means of NMAE and NRMSE of the proposed method are 0.243 and 0.307, respectively, while those of the baseline method are 0.258 and 0.333. The proposed method performs better especially when the
penetration rate is not sufficient high. The potential reason for that can be attributed to the empirical Bayesian method used in this paper, which can obtain the prior distribution of arrival rate by using historical observed data compared with the baseline method, which is based on the current observation data only. However, when the penetration rates are higher than 70%, the NMAEs and NRMSRs calculated by the two methods become approximately equivalent.

5. Conclusion

Cycle-by-cycle queue length estimation methods under partially connected vehicle data environments require the knowledge of the penetration rate and the arrival rate in the existing studies. However, the estimation of these prior parameters has not been adequately and extensively studied. This paper proposes an empirical Bayes estimation method to estimate the arrival rate of total vehicles (assuming Poisson arrivals). However, we do not have any observable samples of the total queued vehicles. Fortunately, with the help of connected vehicle technology, the information such as positions and numbers of CV in the queue can be acquired in real-time. Therefore, based on the conclusions of previous research, the estimation problem can be transformed into CVs arrival rate estimation. For the hyperparameters in the prior distribution of empirical Bayes, the Maximum Likelihood estimation is applied to solve this. In order to validate the proposed methods, a partially connected vehicle environment is established by using the VISSIM COM interface and Python. The validation results have shown that the proposed method could estimate queue length accurately. The impacts of the penetration rates and the positions of the last stopped CV in the queue are also studied. We also explored the effect of cycle time and red time on the achieved outcomes and provided sensitivity analysis.

By accurately estimating queue length using the proposed method, the signal timing design and evaluation at signalized intersections could be improved. For instance, based on the estimated queue length of each direction, the optimal duration of green phase can be derived to reduce the total delay of an isolated intersection. For traffic signal coordination along arterial, queue length could help determine the offset and green wave band width. Besides, accurate queue length can also prevent queue spillovers effectively, which would lead to more serious congestion at intersections under oversaturated traffic conditions. However, some issues should be considered such as the existing GPS data that can not reach lane level when applying the proposed method into real-world engineering cases.

There are a few directions that have not been explored in this paper and may be conducted in the future work. First, the dynamic characteristics and imbalance in spatial-temporal can be explored in the queue length estimation model. Second, the nonhomogeneous Poisson process can model the vehicle arrival more accurately. Third, this paper only focuses on queue length estimation in undersaturated traffic conditions, the overflow queue estimation that considering correlations will be investigated in future work.

Data Availability

The data generated by the VISSIM simulation and recorded by Python via the VISSIM COM interface are used in this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was financially supported by the project "National Key R&D Program of China" under grant no. 2018YFB1601003. The authors are grateful to Lianyungang Jari Electronics Company, Ltd. and Baidu Map team for providing data used in this study.

References

[11] X. Ban, P. Hao, Z. Sun, and Z. Sun, “Real time queue length estimation for signalized intersections using travel times from


