

Research Article

Automatic Estimation Method for Intersection Saturation Flow Rate Based on Video Detector Data

Linhong Wang¹, Yunhao Wang², and Yiming Bie¹

¹School of Transportation, Jilin University, Changchun 130022, China

²School of Economics, Northeast Normal University, Changchun 130117, China

Correspondence should be addressed to Yiming Bie; yimingbie@126.com

Received 30 April 2018; Revised 20 June 2018; Accepted 5 July 2018; Published 17 July 2018

Academic Editor: David Z. W. Wang

Copyright © 2018 Linhong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Saturation flow rate (SFR) is a fundamental parameter to the level of service evaluation, lane capacity calculation, and signal timing plan optimization at signalized intersections. It is affected by a variety of factors including weather conditions, lane width, and the type of the driver. How to accurately estimate the SFR remains one of the most important tasks in traffic engineering. Existing studies generally rely on the field measurement method which requires a large number of people collecting data at the intersection. As a result, the method incurs a high economic cost and cannot adapt to the dynamic change of SFR. In recent years, video detectors have been widely installed at intersections which are capable of recording the time each vehicle passes the stop line, the number plate of each vehicle, and the vehicle type. This paper therefore aims to propose an automatic estimation method for the SFR based on video detector data in order to overcome the limitation of the field measurement method. A prerequisite for estimating the SFR is to recognize the saturation headway. We consider the actual vehicle headway as time series and build an auxiliary regression equation whose parameters are estimated through the ordinary least squares method. We employ the Dickey-Fuller test to verify whether the headways in the time series are saturation headways. An iterative method using quantiles is proposed to filter out abnormal data. The SFR is finally calculated using the average value of saturation headways. To demonstrate the proposed method, we conduct a case study using data from an intersection with three entrance lanes in Qujing city, Yunnan Province, China. The overall estimation process is displayed and the impacts of quantile selection and data duration on the estimation accuracy are analyzed.

1. Introduction

In traffic engineering domain, the saturation flow rate (SFR) of an approaching lane at the intersection is a fundamental parameter for signal timing optimization, capacity calculation, level of service evaluation, and intersection design. As defined by the Highway Capacity Manual (HCM), the SFR is the flow in vehicles per hour that can be accommodated by the lane group assuming that the green phase was displayed 100 percent of the time [1]. The SFR cannot be measured directly and hence needs to be estimated by traffic engineers. While different intersections may have different SFRs, the same intersection may also exhibit a changing SFR because it is affected by a variety of factors including weather conditions, lane width and slope, the type of the driver, and the ratio of heavy vehicles [2, 3].

Given the importance of the SFR, a number of studies have been devoted to its estimation methods. Two methods are provided by HCM. The first method is called the adjustment method. Computations begin with the selection of a base SFR, usually 1900 passenger cars per hour per lane (pcu/h/ln). This value is adjusted for a variety of factors, such as number of lanes in lane group, lane width, heavy vehicles, approach grade, parking, bus blockage, type of area, lane utilization, left turns, right turns, pedestrian, and pedestrian-bicycle. The adjustment formula for each factor is given. However, adjusting the SFR considering such a variety of factors would result in a huge amount of work [4]. HCM therefore provides the second method which is called the field measurement method. Field-measured values of SFR will produce more accurate results than the estimation procedure described in the adjustment method and can be used directly

without further adjustment. SFR is the maximum discharge rate during the green time. It is usually achieved after about 10 to 14 s of green, which corresponds to the front axle of the fourth to sixth passenger car crossing the stop line after the beginning of green. The period of saturation flow begins when the front axle of the fourth vehicle in the queue crosses the stop line or reference point and ends when the front axle of the last queued vehicle crosses the stop line. The last queued vehicle may be a vehicle that joined the queue during the green time. The times when the fourth to the last queued vehicle cross the stop line are recorded and used to calculate the average headway per vehicle. The SFR is 3,600 divided by this value. In order to obtain a statistically significant value, a minimum of 15 signal cycles with more than eight vehicles in the initial queue is typically required. An average of the SFR values in individual cycles represents the prevailing local SFR for the surveyed lane [5, 6].

Shanteau [7] proposed a method to estimate SFR based on cumulative curve method. In the cumulative curve figure, the horizontal axis represented time and the vertical axis represented the number of the vehicle crossing the stop line. A diagram was presented to show the average number of vehicles to enter the intersection versus time. Then a straight line was selected to best fit the data. The straight line portion of the plot could be determined by eye (the easiest and usually sufficient way) or by some formalized method such as linear regression. Then the SFR could be determined according to the slope of the straight line. Hossain [8] developed a microsimulation model to depict the traffic flows in developing cities, where the traffic comprised both motorized and nonmotorized vehicles. The saturation flows at signalized intersections were investigated in an aggregate form of vehicles per hour. It was found that the saturation flows were dependent on the approach width, turning proportion, and composition of the traffic mix. A regression model was proposed to estimate the saturation flow from the influencing variables. Shao and Liu [9] found that using the average value of observed queue discharge headways to estimate the saturation headway might lead to underestimate SFR. Based on the field data, normal and log-normal distributions were used to fit the headways. The median value of discharge headways was suggested to estimate the saturation headway. Zhao et al. [10] studied the headway distribution function at signalized intersections using field collected data. The relationship between headway and vehicle position were modeled and the 78th percentile of headway at each position was selected to calibrate the SFR.

The SFR is equal to 3,600 divided by the saturation headway. When the saturation headway is obtained, the SFR is automatically determined. Therefore, a few methods for estimating the saturation headway have been proposed. Tong and Hung [11] proposed a neural network approach to simulate the queued vehicle discharge headway. A computer-based three-layered neural network model was developed for the estimation of discharge headway. Hung et al. [12] found that discharge headway of individual vehicles at different queue position follows the Type I Extreme Value Distribution. They proposed a method to estimate site-specific parameters for this distribution. Murat and Gedizlioglu [13] investigated the

vehicle time headways at signalized intersection using the data in Turkey. Seven distribution models were taken into account to fit the headway data and log-normal, Weibull, gamma, or Pearson distribution models performed better than other models. Jin et al. [14] found that distributions of the departure headways at each position (except the first position) in a queue were revealed to approximately follow a certain log-normal distribution. Then they proposed a new car-following model from the driver's perspective to explain why the departure headways follow such a log-normal distribution. Tan et al. [15] compared the departure headway collected from intersections in Beijing and generated by VISSIM and TransModeler. Results showed that slight difference in car-following models and parameters would result in significant difference of the distribution. Yin et al. [16] compared the departure headways in Beijing and Atlanta. Urban traffic flows in Beijing, China, were often hindered by pedestrians at intersections and queuing vehicles often cannot be totally discharged during the green time. It was found that the departure headways of the first and second vehicles and after the 5th vehicle in Beijing were significantly different from those in Atlanta. Tan et al. [17] developed two distribution models to fit startup lost time and effective flow rate, respectively. The dependence of departure headways in three types of simulation software (VISSIM, PARAMICS, and TransModeler) was also examined. Hao and Ma [18] tried to validate the assumption that the mean headway of saturation flow was saturated headway. They investigated the distribution of departure headways based on the field data in Shanghai, China. It was found that shifted log-normal was more suitable to capture the distribution of steady-flow headways, and the mode of steady-flow headway distribution could be considered as theoretical saturation headway and the mean could be considered as practical saturation headway.

While quite some research effort has been made to estimate the intersection SFR, all the aforementioned methods are essentially the field measurement method provided by HCM. These methods require traffic engineers to conduct filed surveys at intersections. The number of vehicles queuing behind the stop line is recorded manually and the times when the fourth to the last queued vehicle cross the stop line are extracted to calculate the saturation headways. Although the field measurement method can estimate the saturation headway and hence the SFR, it has the following two limitations:

(i) High labor and economic costs: as suggested by HCM, two surveyors are needed to estimate the SFR for one entrance lane [1]. While the field measurement method may be feasible when the number of entrance lanes is small, it can hardly be applied in a city where there are hundreds of intersections and hence thousands of entrance lanes. Therefore, to be able to estimate the SFR, a large number of surveyors are required to collect data for each lane resulting in a high economic cost.

(ii) Inability to adapt to the dynamic change in SFR: traffic data collected by the field measurement method and the resulting estimations of the SFR only represent a single or a few typical traffic scenarios. Since the SFR is affected by a number of factors, there can be hundreds of traffic scenarios and, hence, the method can hardly be applied to estimate

the SFR under all possible scenarios. Moreover, the real-time SFR is an input parameter for adaptive signal control at intersections. When one of the influencing factors changes (e.g., weather or the ratio of heavy vehicles), the SFR changes accordingly. While it may take at least 30 minutes to send traffic engineers to the site and obtain the adjusted SFR, this obviously cannot fulfill the real-time requirement of adaptive signal control.

In recent years, video detectors have been widely installed at signalized intersections in many cities. They can provide fruitful traffic information to advanced traffic management system, such as traffic volumes of different types of vehicles, volume of pedestrian, vehicle trajectory, and traffic OD (origin-destination) matrix. For example, with the help of video detectors, adaptive traffic control system can be developed to adjust signal timing plans in real time and estimate the queue length behind the stop line. In addition, video detectors can also provide time headways by recording in real time the number plate and time of each vehicle crossing the stop line and the type of the vehicle. The advantages of such an automatic estimation method for the SFR using video detector data include the: (i) vehicle discharge headways during the green phase can be automatically recorded without the need for manual surveys, which reduces the economic cost; (ii) vehicle operating data under different traffic scenarios can be recorded in real time, which fulfills the real-time requirement of traffic control; (iii) the method can be applied to the entire urban road network to automatically estimate the SFR for each entrance lane under each traffic scenario.

However, vehicle discharge headways obtained from video detector data cannot be used directly to estimate the SFR. This is because, as shown in HCM, the SFR can only be estimated using the saturation headways from the queuing vehicles. While the field measurement method allows traffic engineers to identify the queuing vehicles on-site, video detectors only record the time each vehicle crosses the stop line rather than automatically identifying the queuing vehicles. Therefore, further analysis on video detector data is needed.

Different video detectors can record different traffic flow parameters. The video detectors used in this study cover the area surrounding the stop line and record the following parameters: the number plate, speed, direction, and time of each vehicle crossing the stop line, the type of the vehicle, and the number of the entrance lane. Such a technique has already been widely applied at intersections in a number of cities in China. Therefore, the objective of this study is to propose an automatic estimation method for the intersection SFR based on video detector data that can overcome the limitations of the field measurement method (i.e., high labor cost and inability to fulfill the real-time requirement of traffic control). The rest of the paper is organized as follows. Section "Automatic Estimation Method for Saturation Flow Rate" proposes the automatic estimation methodology, the core of which is the automatic recognition of the saturation flow. Section "Case Study" conducts a case study on a field intersection and analyzes the effects of different parameters on the SFR. The last section concludes the paper.

2. Automatic Estimation Method for Saturation Flow Rate

2.1. Recognition of Saturation Flow. As described in the HCM, the period of saturation flow begins when the front axle of the fourth vehicle in the queue crosses the stop line or reference point and ends when the front axle of the last queued vehicle crosses the stop line. The first step of estimating the SFR is to recognize the saturation flow during the green phase, i.e., to recognize the fourth to the last queued vehicles. Since video detectors record the time each vehicle crosses the stop line, it is rather easy to obtain the time the fourth vehicle crosses the stop line. However, it is impossible to tell which vehicle is the last in the queue because video detectors do not record the real-time queue length. Therefore, a new method is needed to recognize the saturation flow.

During the green phase, the queuing vehicles cross the stop line successively. If the traffic flow is saturated, vehicle headways should fluctuate around the saturation headway. Therefore, from a statistics perspective, vehicle headways should fluctuate around a certain constant and the fluctuations are irregular. In this sense, the error series of the vehicle headway series should be independently and identically distributed (IID) with zero mean. Let X_m denote the time series of all sampled saturation headways during the peak period, with observations denoted by $X_m(1), X_m(2), \dots, X_m(n), \dots, X_m(N_m)$, where $1 \leq n \leq N_m$ and N_m is the number of samples. The n -th time headway $X_m(n)$ is therefore expressed as follows:

$$X_m(n) = \mu + e(n) \quad (1a)$$

where μ is the true saturation headway and $e(n)$ is the random error term of the n -th headway. We assume that e follows a normal distribution with zero mean and variance σ_e^2 and that any two of the random errors are IID. Since μ is unknown and to be estimated, the fluctuations of the observed headways around μ are unknown. Thus, σ_e^2 is also unknown.

Since e follows a normal distribution and μ is a constant, the expectation and variance of X_m can be obtained based on the following two equations:

$$E[X_m(n)] = \mu \quad (1b)$$

$$\text{var}[X_m(n)] = \sigma_e^2 \quad (1c)$$

That is, the saturation headway series X_m follows a normal distribution with mean equal to μ and variance equal to σ_e^2 .

According to (1a), the $(n-1)$ -th time headway in X_m can be expressed as follows:

$$X_m(n-1) = \mu + e(n-1) \quad (1d)$$

Both (1a) and (1d) contain the real saturation flow rate μ ; thus we can subtract (1d) from (1a) to eliminate μ and then obtain the following relation:

$$X_m(n) - X_m(n-1) = e(n) - e(n-1) \quad (2a)$$

TABLE 1: Critical values of the *DF* distribution.

Number of samples	Quantile α							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
25	-2.66	-2.26	-1.95	-1.60	0.92	1.33	1.70	2.16
50	-2.62	-2.25	-1.95	-1.61	0.91	1.31	1.66	2.08
100	-2.60	-2.24	-1.95	-1.62	0.90	1.29	1.64	2.03
250	-2.58	-2.23	-1.95	-1.62	0.89	1.29	1.63	2.01
500	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00
∞	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00

By setting $\Delta e(n) = e(n) - e(n-1)$ and $\Delta X_m(n) = X_m(n) - X_m(n-1)$, (2a) is further simplified:

$$\Delta X_m(n) = \Delta e(n) \quad (2b)$$

Let Δe and ΔX_m denote the time series comprised $\Delta e(n)$ and ΔX_m , respectively. Since any two of the random errors are IID, the expectation and variance of $\Delta e(n)$ are expressed as follows:

$$\begin{aligned} E[\Delta e(n)] &= E[e(n) - e(n-1)] \\ &= E[e(n)] - E[e(n-1)] = 0 \end{aligned} \quad (3a)$$

$$\begin{aligned} \text{var}[\Delta e(n)] &= \text{var}[e(n) - e(n-1)] \\ &= \text{var}[e(n)] + \text{var}[e(n-1)] = 2\sigma_e^2 \end{aligned} \quad (3b)$$

Equations (3a) and (3b) show that Δe follows a normal distribution with zero mean and $2\sigma_e^2$ variance, which provides a first method to test whether the vehicle headways in the time series are saturation headways. Specifically, we first obtain ΔX_m and Δe through the observed X_m and calculate the sample mean and variance. We then perform a hypothesis test of the population mean (i.e., *t*-test with unknown population variance) by checking whether there is a significant difference between the sample mean and zero. If there is no significant difference, the expectation of Δe is considered zero. While the sample variance of Δe can be calculated, σ_e^2 cannot be obtained through (1a). As a result, we cannot check whether the sample variance of Δe is equal to $2\sigma_e^2$ and the method cannot work properly.

To better test whether the observed time headways are saturation headways, we set $k(n) = \Delta e(n)$ and $\sigma_k^2 = 2\sigma_e^2$ and propose the following auxiliary regression equation:

$$X_m(n) = \rho X_m(n-1) + k(n) \quad (4a)$$

where ρ is a parameter to be estimated.

After (4a) has been developed, the next step is to perform a hypothesis test with the null hypothesis of $\rho = 1$. If the null hypothesis cannot be rejected, the observed headways can be considered as saturation headways. In such condition, $X_m(n) = X_m(n-1) + k(n)$. Because $k(n) = \Delta e(n)$ and $E[\Delta e(n)] = E[k(n)] = 0$ (please refer to (3a)), then we can get $E[X_m(n)] = E[X_m(n-1)]$. Thus, it can be concluded that the headways in X_m are saturation headways.

The parameters ρ and $k(n)$ in (4a) are estimated using the ordinary least squares (OLS) method. The resulting estimations are denoted by $\hat{\rho}$ and $\hat{k}(n)$ as follows:

$$\hat{\rho} = \frac{\sum_{n=2}^{N_m} X_m(n-1) X_m(n)}{\sum_{n=2}^{N_m} [X_m(n-1)]^2} \quad (4b)$$

$$\hat{k}(n) = X_m(n) - \hat{\rho} X_m(n-1) \quad (4c)$$

According to classical regression theory, $(\hat{\rho} - \rho)/SE(\hat{\rho})$ should follow a *t* distribution with a $(N_m - 2)$ degree of freedom. If $\rho = 1$ is satisfied, $(\hat{\rho} - 1)/SE(\hat{\rho})$ should follow a *t* distribution with a $(N_m - 2)$ degree of freedom where $SE(\hat{\rho})$ is the standard error of $\hat{\rho}$. A smaller standard error indicates more closeness between the sample statistics and the population parameters. The samples are therefore a better representative of the population which results in more reliable estimations of the population parameters.

$$SE(\hat{\rho}) = \frac{\hat{\sigma}_k^2}{\sum_{n=2}^{N_m} [X_m(n-1) - \bar{X}_m]^2} \quad (4d)$$

$$\bar{X}_m = \frac{1}{N_m - 1} \sum_{n=2}^{N_m} X_m(n-1) \quad (4e)$$

$$\hat{\sigma}_k^2 = \frac{1}{N_m - 2} \sum_{n=2}^{N_m} [\hat{k}(n)]^2 \quad (4f)$$

Since (4a) involves both the *n*-th and (*n*-1)-th vehicle headways, the characteristics of the constructed statistic for the hypothesis test of ρ may change. It is proved by Dickey and Fuller [19] that when a hypothesis test is conducted on ρ in (4a), the constructed “*t*” statistic based on the null hypothesis of $\rho = 1$ no longer follows the *t* distribution. Instead, it follows the *DF* distribution. To distinguish between the two statistics, we construct the following *DF* statistic:

$$DF = \frac{(\hat{\rho} - 1)}{SE(\hat{\rho})} \quad (4g)$$

The constructed *DF* statistic in (4g) follows the *DF* distribution whose critical values are needed during the hypothesis test. Table 1 shows the critical values of the *DF* distribution corresponding to different quantiles.

Table 1 shows that when we perform a two-sided test with a significance level of 5%, the acceptance region lies

between -2.25 and 1.66 given that the number of samples is larger than 50. That is, when the DF statistic lies within $[-2.25, 1.66]$, the null hypothesis of $\rho = 1$ cannot be rejected. The observed headways can be considered as saturation headways. In statistics, the time series X_m satisfying $X_m(n) = X_m(n-1) + k(n)$ is termed a random walk.

2.2. Exclusion of Abnormal Data. While section “Recognition of Saturation Flow” proposes a method to test whether the observed headways are saturation headways, this section aims to propose a method to filter out abnormal data.

Estimating the SFR usually needs traffic data collected during a peak period. Let t_{start} and t_{end} denote the beginning and ending times of the peak period, respectively. Based on the outputs of video detectors (i.e., the time each vehicle crosses the stop line), we can obtain the time series of the observed headways within $[t_{start}, t_{end}]$, denoted by X_0 . The number of samples is N_0 and the n -th sample is $X_0(n)$ where $1 \leq n \leq N_0$. When we construct X_0 , various factors that may affect the headway are not taken into account. As a result, the headways in X_0 may not necessarily be at saturation.

There are three main types of abnormal headways in X_0 : (a) due to the red light, the headway may be large between the first vehicle crossing the stop line in current cycle and the last vehicle crossing the stop line in the previous cycle; (b) at the beginning of the green phase, the headways between the first few vehicles crossing the stop line may be large because they are still accelerating; (c) during the green phase, there may be a large headway as a result of the inattention of drivers or the random arrivals of vehicles.

When the calculated headway is larger than or equal to the red time of the signal phase, the (a) type of headways occurs and is excluded from X_0 . The time series after the exclusion is denoted by X_1 and the number of samples is N_1 . Since X_1 still includes the (b) and (c) types of headways, the next step is to find a threshold θ whereby the headways smaller than θ are considered as saturation headways (i.e., through the DF test shown in section “Recognition of Saturation Flow”).

Since X_1 is calculated using traffic data collected during a peak period, the number of abnormal samples of types (b) and (c) only accounts for a small proportion of all the samples, and the majority of observed headways are fluctuating around the saturation headway. The mode of the sampled data in X_1 remains around the saturation headway rather than becoming larger due to abnormal data. If the mode exists for X_1 containing abnormal data, the mean value should be larger than the mode; i.e., X_1 is a right-skewed distribution. As a result, using the mean value of X_1 as the threshold θ is inappropriate. Using the mode as the threshold is, however, also inappropriate because, under some circumstances, the mode does not exist for X_1 . Moreover, the mode of some headway samples may be smaller than the saturation headway. Therefore, this paper uses sample quantiles to set the threshold θ .

When the number of samples is large, we can use the 60%-80% quantile to set θ which excludes more data from X_1 . When the number of samples is small, we can use the 80%-95% quantile as the threshold which excludes fewer data from

X_1 . The detailed steps of filtering out abnormal data are as follows.

Step 1. Initialize and set $i = 1$.

Step 2. Calculate the headway corresponding to the quantile β in X_i and set as the threshold θ .

Step 3. Exclude data larger than θ from X_i and generate the new time series X_{i+1} (the number of samples becomes N_{i+1}). Conduct a DF test on X_{i+1} . If the test passes, terminate the iterative algorithm; otherwise go to Step 4.

Step 4. Set $i = i + 1$ and go to Step 2.

The time series generated at the last iteration is denoted by X_m consisting of N_m samples. The algorithmic steps show that each iteration reduces the number of samples by a factor of β ($\beta < 1$) and that the headway corresponding to the quantile β changes accordingly. The number of samples in X_m , i.e., N_m , is equal to $\beta^{m-1} \times N_1$.

2.3. Calculation of the Saturation Flow Rate. From section “Exclusion of Abnormal Data”, we obtain the time series X_m where the headways are denoted by $X_m(1), X_m(2), \dots, X_m(N_m)$, respectively. Since X_m passes the DF test, it contains saturation headways and should follow a normal distribution as proved by (1a), (1b), (1c), and (1d). In this sense, the mean value, median, and mode of X_m are nearly equal.

We use the sample mean value to estimate the true saturation headway μ . The estimation $\hat{\mu}$ is expressed as follows:

$$\hat{\mu} = \frac{1}{N_m} \sum_{n=1}^{N_m} X_m(n) \quad (5a)$$

By taking the expectation on both sides of (5a) and plugging it into (1b), we obtain $E[\hat{\mu}] = \mu$. Therefore, $\hat{\mu}$ is an unbiased estimation of μ and the estimation accuracy can be expressed by $\text{var}[\hat{\mu}]$ where $\text{var}[\hat{\mu}] = \sigma_{\hat{\mu}}^2/N_m$. $\sigma_{\hat{\mu}}^2$ is the unknown variance of the true saturation headway μ . We can use the variance of X_m , i.e., $S_{\hat{\mu}}^2$, as the estimation of $\sigma_{\hat{\mu}}^2$:

$$S_{\hat{\mu}}^2 = \frac{1}{N_m - 1} \sum_{n=1}^{N_m} [X_m(n) - \hat{\mu}]^2 \quad (5b)$$

The interval estimation of μ with a significance level of α is as follows:

$$\begin{aligned} \hat{\mu} - t_{\alpha/2}(N_m - 1) \frac{S_{\hat{\mu}}^2}{\sqrt{N_m}} &\leq \mu \\ &\leq \hat{\mu} + t_{\alpha/2}(N_m - 1) \frac{S_{\hat{\mu}}^2}{\sqrt{N_m}} \end{aligned} \quad (6)$$

where $t_{\alpha/2}(N_m - 1)$ is the critical value of the t distribution with a $(N_m - 1)$ degree of freedom and a significance level of α .

The limit error of the estimated μ , i.e., δ , is expressed as follows:

$$\delta = t_{\alpha/2}(N_m - 1) \frac{S_{\hat{\mu}}^2}{\sqrt{N_m}} \quad (7a)$$

TABLE 2: Statistics of the raw time headways.

Statistical parameters	Lane 1	Lane 2	Lane 3
Number of samples	810	1247	986
Mean	11.029	8.634s	9.526s
Median	3.063s	2.521s	2.874
Mode	2.867s	2.400s	2.704s
Standard deviation	15.292	12.619s	18.835s
Minimum	1.597s	1.259s	1.332s
Maximum	163.006s	145.572s	145.384s

Assuming that the minimum number of samples to ensure that δ is smaller than ε is N_ω , we obtain the following relation:

$$t_{\alpha/2}(N_\omega - 1) \frac{S_\mu^2}{\sqrt{N_\omega}} \leq \varepsilon \quad (7b)$$

We can use (7b) to solve the interval for N_ω . The minimum value of N_ω is achieved when the error precision ε is reached. When the number of samples N_m is much larger than N_ω , the estimated SFR using the mean value of the samples can achieve a very high precision. If N_m is smaller than N_ω , we can use headway data in peak hours of multiple days (e.g., a week or a month) to increase N_m .

3. Case Study

We apply the proposed method to a signalized intersection in Qujing, China, and present the overall estimation process of the SFR. Qujing is the second largest city in Yunan province, with the main urban area of about 130 square kilometers and a population of 1.0 million. The intersection is made up of Qilin South Road and Wenchang Street and the layout is shown in Figure 1. There are three approaching lanes and three exiting lanes on each leg. To illustrate the estimation of SFR, the three approaching lanes on the north leg are selected as examples: lane 1 dedicated to left turn vehicles, lane 2 for through vehicles, and lane 3 shared by through and right-turn vehicles. The lane width is 3.5 meters. Video detectors are installed for each lane recording data such as the number plate, type, and time of each vehicle crossing the stop line. The cycle length of the signal timing during the morning peak is 180 seconds. The green time, red time, and amber time for lane 1 are 19 seconds, 158 seconds, and 3 seconds, respectively. The green time, red time, and amber time for lanes 2 and 3 are the same because they are controlled by the same signal phase, which are 36 seconds, 141 seconds, and 3 seconds, respectively.

3.1. Data Description. We use data collected during the morning peak (7:00 – 10:00) on October 30, 2017, to estimate the SFRs for the three lanes. The weather was fine without rains. The numbers of raw samples of headways for the three lanes are 810, 1,247, and 986, respectively. Table 2 shows the resulting statistics. It is obvious that the raw samples contain quite a few abnormal data. For example, the maximum headway of lane 2 reaches 145.572 seconds, which actually

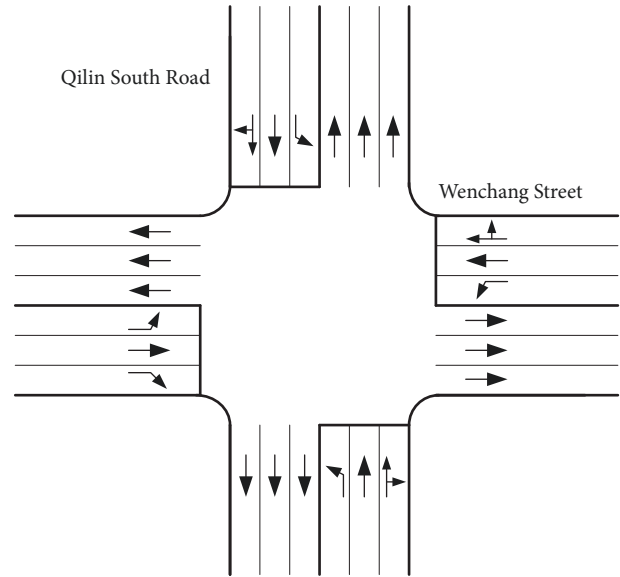


FIGURE 1: The layout of the survey intersection.

represents the headway between the first vehicle crossing the stop line in the current cycle and the last vehicle crossing the stop line in the previous cycle.

3.2. Saturation Flow Estimation Using the 80% Quantile. We first use the red time for each lane as the threshold and exclude the (a) type of abnormal data shown in the section “Exclusion of Abnormal Data”. That is, we exclude headways larger than 158 seconds and 141 seconds for lane 1 and lanes 2 and 3, respectively. The numbers of samples for the three lanes after the exclusion become 743, 1,183, and 922, respectively. Table 3 shows the resulting statistics. By comparing Table 2 with Table 3, we observe that the mean values of the samples reduce after the (a) type of abnormal data is excluded. However, the differences between the mean values, medians, and modes are still large. The maximum values of the samples for the three lanes reach 10.256 s, 8.884 s, and 9.857 s, respectively.

To exclude the (b) and (c) types of abnormal data, we apply the algorithmic steps proposed in Section “Exclusion of Abnormal Data” with $\beta = 80\%$. We take lane 2 as an example and show the process in Table 4 where H_0 represents the null hypothesis. The samples pass the DF test after six

TABLE 3: Statistics of the samples after excluding the (a) type of abnormal data.

Statistical parameters	Lane 1	Lane 2	Lane 3
Number of samples	743	1184	922
Mean	4.643s	3.648s	4.257s
Median	2.937s	2.441s	2.661s
Mode	2.867s	2.400s	2.704s
Standard deviation	7.421s	5.037s	6.633s
Maximum	1.597s	1.259s	1.332s
Minimum	10.256s	8.884s	9.857s

TABLE 4: Applying the iterative method using samples from lane 2.

Number of iteration	Number of samples	DF statistic	Reject or not $H_0 : \rho = 1$	80% quantile (θ)
1	1184	-20.77	Reject	3.73s
2	947	-5.23	Reject	2.88s
3	759	-3.63	Reject	2.49s
4	607	-2.82	Reject	2.28s
5	486	-2.27	Reject	2.16s
6	392	-1.83	Do not reject	/

TABLE 5: Saturation headway estimations based on the 80% quantile for the three lanes.

Data	Lane 1	Lane 2	Lane 3
T.N.I.	3	6	5
N.S.	380	392	302
Final DF statistic	-1.91	-1.83	-1.87
Mean value of SH(s)	2.421	1.790	2.232
Median of SH (s)	2.404	1.801	2.229
Mode of SH (s)	2.419	1.796	2.224
S.D. of SH (s)	0.544	0.251	0.362
Limit error of SH (s)	0.055	0.025	0.041
I.E. of SH (s)	[2.366, 2.476]	[1.765, 1.815]	[2.191, 2.273]
Estimated SFR (pcu/h)	1487	2011	1613
I.E. of SFR (pcu/h)	[1454, 1522]	[1983, 2039]	[1583, 1643]

* T.N.I. represents total number of iterations; N.S. denotes number of the samples; S.D. represents standard deviation; I.E. denotes interval estimation; SH is the abbreviation of saturation headways.

iterations and the number of samples reduces to 392. The mean value, median, and mode of these samples are 1.790 s, 1.801 s, and 1.796 s, respectively, which are nearly equal. This implies that the 392 samples obtained after six iterations follow a normal distribution and that the value of each sample fluctuates around a certain constant (i.e., the true saturation headway).

We therefore use the mean value of the samples to estimate the saturation headway for lane 2; i.e., $\hat{\mu}_2 = 1.790$, where the subscript 2 represents the number of the lane. Hence, the SFR of lane 2 is estimated to be $\hat{\lambda}_2 = 2011$ pcu/h. The standard deviation of the samples is 0.251 s. Given $\alpha = 0.05$ and $t_{\alpha/2}(391) = 1.97$, the limit error of the saturation headway of lane 2 is $\delta_2 = 1.97 \cdot 0.251/\sqrt{392} = 0.025$ s. We are therefore 95% confident that the saturation headway of lane 2 lies within $[\hat{\mu}_2 - \delta_2, \hat{\mu}_2 + \delta_2]$, i.e., [1.765 s,

1.815 s]. The corresponding interval estimation of the SFR is [1983 pcu/h, 2039 pcu/h]. The method is also applied to lanes 1 and 3 and the results are shown in Table 5. The results of lane 2 are included for the comparison.

3.3. Effect of Quantile Selection on the Estimated Saturation Flow Rate. When filtering out abnormal data in section “Saturation Flow Estimation using the 80% Quantile”, we assume $\beta = 80\%$ indicating that the number of samples reduces by 20% during each iteration. In this section, we aim to analyze the effect of different quantiles on the iterative process and the estimated saturation headway.

We choose β to be 60%, 70%, and 90%, respectively, and estimate the corresponding SFRs based on the collected headway samples of lane 2. The results are compared with what we have obtained in section “Saturation Flow Estimation

TABLE 6: SFR estimations based on the 60%, 70%, 80%, and 90% quantiles for lane 2.

Data	$\beta = 60\%$	$\beta = 70\%$	$\beta = 80\%$	$\beta = 90\%$
T.N.I.	3	4	6	12
N.S.	258	407	392	383
Final <i>DF</i> statistic	-1.34	-1.89	-1.83	-1.79
Mean value of SH(s)	1.661s	1.804s	1.790s	1.782s
Median of SH (s)	1.720s	1.810s	1.801s	1.800s
Mode of SH (s)	1.920s	1.805s	1.805s	1.805s
S.D. of SH (s)	0.210s	0.256s	0.251s	0.247s
Limit error of SH (s)	0.026s	0.025s	0.025s	0.025s
I.E. of SH (s)	[1.635, 1.687]	[1.779, 1.829]	[1.765, 1.815]	[1.757, 1.807]
Estimated SFR (pcu/h)	2167	1995	2011	2020
I.E. of SFR (pcu/h)	[2133, 2201]	[1968, 2023]	[1983, 2039]	[1992, 2049]

TABLE 7: SFR estimations of lane 2 under different data durations.

Data duration (min)	30	60	90	120	150	180
N.R.S.	166	399	644	855	1054	1247
T.N.I.	3	4	4	5	6	6
N.S.	99	197	315	335	338	392
Final <i>DF</i> statistic	-1.47	-1.52	-1.92	-1.82	-1.75	-1.83
Mean value of SH(s)	2.276s	1.977s	1.924s	1.838s	1.788s	1.790s
Median of SH (s)	2.285s	2.040s	1.921s	1.880s	1.800s	1.801s
Mode of SH (s)	2.400s	2.400s	1.796s	1.796s	1.796s	1.796s
S.D. of SH (s)	0.459s	0.354s	0.324s	0.280	0.246	0.251
Limit error of SH (s)	0.091	0.049	0.036	0.030	0.027	0.025
I.E. of SH (s)	[2.185, 2.367]	[1.928, 2.026]	[1.888, 1.960]	[1.808, 1.868]	[1.761, 1.815]	[1.765, 1.815]

*N.R.S. denotes number of raw samples.

using the 80% Quantile”, which are further summarized in Table 6.

Table 6 shows that when β is equal to 70%, 80%, or 90%, the mean value, median, and mode of the saturation headways are nearly equal. However, there is a big difference when β is equal to 60% which implies that the saturation headways do not follow a normal distribution. This is because when a small quantile is selected, a large number of samples are excluded by the iterative method which, however, may not all be abnormal data. As a result, the remaining headways tend to be smaller. When excluding the (a) type of abnormal data for lane 2 in Table 3, we are left with 1,184 samples. In Table 6, the numbers of iterations corresponding to β equal to 60%, 70%, 80%, and 90% are 3, 4, 6, and 12, respectively, and the numbers of the remaining samples are 258, 407, 392, and 383, respectively. It is obvious that when β is equal to 60%, a large number of samples are excluded and, hence, a small β is not recommended.

A larger β , however, may result in a larger number of iterations which reduces the real-time capability of the estimation method without increasing its accuracy. Therefore, we suggest β be set at 80%.

3.4. Effect of Data Duration on the Estimated Saturation Flow Rate. While video detectors are capable of continuously recording traffic data for estimating the SFR, excessive data

do not necessarily result in higher accuracy. However, the accuracy may reduce if the number of samples is insufficient. This section therefore aims to analyze the effect of data duration on the estimated SFR.

The data used in the two previous subsections were collected during the 7:00-10:00 morning peak on October 30, 2017. We take lane 2 as an example and estimate the SFRs corresponding to 30 min, 60 min, 90 min, 120 min, 150 min, and 180 min of data duration. The estimated SFRs are summarized in Table 7 with β equal to 80%.

Table 7 shows that when the data duration is small, the mean value, median and mode of the saturation headways exhibit a large difference which violates the assumption of a normal distribution. When the data duration increases to 150 min and 180 min, all the three values are nearly equal. Therefore, we suggest using the data duration no less than 150 min to ensure the estimation accuracy and robustness.

4. Conclusion

By using video detector data, this study proposes an automatic estimation method for the intersection SFR consisting of the following three sections: automatic recognition of the saturation flow, exclusion of abnormal data, and calculation of the SFR based on the saturation headway. The following conclusions are made.

(i) Recognition of saturation flow is the key to the estimation of the SFR. In this study, saturation flow is recognized by constructing the DF statistic and conducting the DF test. We find that the time series of the saturation headways is a random walk in statistics.

(ii) While headways can be obtained through video detector data, there exist quite a few abnormal data that need to be filtered out to ensure the estimation accuracy. The iterative method based on quantiles proposed in this study is able to exclude these abnormal data and maintain a high level of flexibility.

(iii) Both quantile selection and data duration have an impact on the estimation accuracy of the SFR. We suggest using the quantile of 80% and the data duration of no less than 150 min.

Data Availability

The data supporting the conclusions of the study can be found in the section of "Case Study".

Conflicts of Interest

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

Acknowledgments

This study is supported by the National Natural Science Foundation of China (nos. 71771062, 71771050, and 71501038), the Postdoctoral Startup Foundation of Heilongjiang Province (LBH-Q15050), and the Opening Foundation of the Intelligent Traffic Engineering Technology Research Center of Yunan Province Universities.

References

- [1] Transportation Research Board, "Highway Capacity Manual," *National Research Council*, 2000.
- [2] S. Wang and X. Qu, "Rural bus route design problem: Model development and case studies," *KSCE Journal of Civil Engineering*, vol. 19, no. 6, pp. 1892–1896, 2015.
- [3] Y. Bie, X. Gong, and Z. Liu, "Time of day intervals partition for bus schedule using GPS data," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 443–456, 2015.
- [4] Z. Liu, S. Wang, W. Chen, and Y. Zheng, "Willingness to board: a novel concept for modeling queuing up passengers," *Transportation Research Part B: Methodological*, vol. 90, pp. 70–82, 2016.
- [5] Z. Liu, S. Wang, B. Zhou, and Q. Cheng, "Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 58–72, 2017.
- [6] Z. Liu, Y. Yan, X. Qu, and Y. Zhang, "Bus stop-skipping scheme with random travel time," *Transportation Research Part C: Emerging Technologies*, vol. 35, pp. 46–56, 2013.
- [7] J. Bared, G. L. Giering, and D. L. Warren, "Safety evaluation of acceleration and deceleration lane lengths," *ITE Journal*, vol. 69, no. 5, pp. 50–54, 1999.
- [8] M. Hossain, "Estimation of saturation flow at signalised intersections of developing cities: a micro-simulation modelling approach," *Transportation Research Part A: Policy and Practice*, vol. 35, no. 2, pp. 123–141, 2001.
- [9] C. Shao and X. Liu, "Estimation of Saturation Flow Rates at Signalized Intersections," *Discrete Dynamics in Nature and Society*, vol. 2012, pp. 1–9, 2012.
- [10] Y. Zhao, W. Zhang, J. Lu, W. Zhang, and Y. Ma, "Estimation of Saturation Flow Rate and Start-Up Lost Time for Signal Timing Based on Headway Distribution," *Discrete Dynamics in Nature and Society*, vol. 2015, pp. 1–7, 2015.
- [11] H. Y. Tong and W. T. Hung, "Neural network modeling of vehicle discharge headway at signalized intersection: Model descriptions and results," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 1, pp. 17–40, 2002.
- [12] W. T. Hung, F. Tian, and H. Y. Tong, "Discharge headway at signalized intersections in Hong Kong," *Journal of Advanced Transportation*, vol. 37, no. 1, pp. 105–117, 2003.
- [13] Y. S. Murat and E. Gedizlioglu, "Investigation of vehicle time headways in Turkey," *Proceedings of the Institution of Civil Engineers: Transport*, vol. 160, no. 2, pp. 73–78, 2007.
- [14] X. Jin, Y. Zhang, F. Wang et al., "Departure headways at signalized intersections: a log-normal distribution model approach," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 3, pp. 318–327, 2009.
- [15] J. Tan, L. Li, Z. Li, and Y. Zhang, "A comparison of departure headway at signalized intersection produced by VISSIM and TransModeler," in *Proceedings of the Paper presented at the IEEE International Conference on Vehicular Electronics and Safety*, Beijing, China, July 2011.
- [16] S. Yin, Y. Su, C. Wang, D. Yao, L. Li, and Y. Zhang, "Comparison of vehicle departure headways in Beijing and Atlanta," *Tsinghua Science and Technology*, vol. 16, no. 3, pp. 332–336, 2011.
- [17] J. Tan, L. Li, Z. Li, and Y. Zhang, "Distribution models for start-up lost time and effective departure flow rate," *Transportation Research Part A: Policy and Practice*, vol. 51, pp. 1–11, 2013.
- [18] H. Hao and W. Ma, "Revisiting distribution model of departure headways at signalised intersections," *Transportmetrica B: Transport Dynamics*, vol. 5, no. 1, pp. 1–16, 2017.
- [19] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, vol. 74, no. 366, part 1, pp. 427–431, 1979.



Hindawi

Submit your manuscripts at
www.hindawi.com

