Loop detector data error diagnosing and interpolating with probe vehicle data

Li Jie

[a] *Department of Traffic and Road Engineering, Civil Engineering College, Hunan University, Lushan South Road, 410082 Changsha, Hunan Province, PR China*
Tel.:+86 13017312752; fax:+86 731 88822667.
email: ljlj369@msn.com

Henk J. van Zuylen

Department of Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands,
Tel.: +31 13626158161
e-mail: h.j.vanzuylen@tudelft.nl
Corresponding author

Wei Guorong

CCCC First Highway Consultants Co.,LTD, Polymaths South Building 703, Xiameilin, 518000 Shenzhen, Guangdong Province, PR China
Tel.:+86 13528741048 fax:+86 755 83938471
Email: gd_huli@126.com

Submitted for presentation and publication to the 93th Annual Meeting of the Transportation Research Board January

Word count: 4354 words + (10 Figures+2 tables) * 250 = 7354 equivalent words

Paper 14-5349

ABSTRACT

Traffic volume and queue length are two important variables to identify traffic states in urban areas. Loop detectors are often used for monitoring traffic. A prominent weakness of loop detector data is the limited reliability due to equipment malfunctions and communication faults. This paper presents a method to identify the validity of loop detector measurements by analyzing the ratio of counts on adjacent lanes.

GPS data from probe vehicles are an alternative for monitoring traffic states. They include instantaneous speed, acceleration and position; individual vehicle trajectories can be grossly estimated in real time. Consequently, the queue length can be measured approximately. This can be used to estimate traffic volumes and the percentage of probe vehicles. By combining loop detector and GPS data, more information regarding traffic status can be extracted from both data sources.

This paper discusses two methods to check the volume counts and to fill in missing or invalid data: one method uses the ratio of loop detector counts on adjacent links, the other one uses data from probe vehicles to estimate the traffic volumes from the dynamics of the queue length at signalized intersections. Real data from the city Changsha in China are used to validate these two methods. The developed methods provide a way for the online monitoring of detectors' performance and level of service at signalized intersections.

Keywords: Signalized intersection; Loop detector; Data correction; GPS data; Queues

INTRODUCTION

The inductive single loop detector is one of the most popular instruments for collecting traffic data over an extended period in large areas. Traffic volume, vehicle speed and occupancy can be obtained from these detectors. Especially data related to volumes are widely used in traffic management practice, including real-time traffic information and traffic control.

The accuracy and reliability of the data obtained from loop detectors are critical for the quality of the applications. However, a prominent drawback of loop detector data is the reliability failures due to equipment malfunctions and communication faults. Especially in urban areas road works often affect the loop detectors. Loop data diagnostics have been extensively studied for decades. In the literature, data errors were mostly identified by using threshold checks on speed, volume, or occupancy, either individually or combined with adjacent locations and with historical data, the so-called Temporal-Spatial Method (*1, 2*). Still limited attention has been given to the application of data from Probe Vehicles with Global Positioning System (GPS) as an additional traffic data source. Currently, most research about GPS data application in transportation is related to travel time and traffic state estimation (*3, 5, 6*).

Besides traffic volume, speed and occupancy, the queue length is another important characteristic of the traffic state in urban areas. Until now, little effort has been given to the estimation of queue lengths (*7, 1, 10*). Probe vehicles are a good instrument to obtain information about queue lengths, but less usable for the direct estimation of traffic volumes. Peng Hao et al. (*9*) showed that traffic volumes and the percentage of probe vehicles can be estimated from the position of probe vehicles in a queue. Even though this method is not very accurate, it is possible to obtain additional information for traffic volumes and to verify uncertain loop detector data. A simple estimation method is developed in this paper.

This study focuses first on the development of an approach for validating the traffic volume data collected by single loop detectors at signalized intersections. Because GPS data include instantaneous speed and location, individual trajectories can be grossly estimated in real time. Consequently, queue lengths can be estimated as a time-dependent variable. This can be used for a macroscopic quality check for the traffic counts. By combining loop detector and GPS data, one can improve the relative quality of both data sources.

Section 2 first analyzes the quality of loop detector data in the center of a Chinese city Changsha. Due to several causes the validity of these data appears to be limited. A method is introduced to check the validity of traffic volumes by comparing the counts on adjacent lanes. Section 3 discusses a methodology to identify failing detectors. Section 4 describes a method to estimate the dynamics of queues from trajectories of probe vehicles and to use this for the estimation of volumes. Real data from the city Changsha in China are used in section 5 to validate the methodology. In Changsha detailed traffic counts, loop detector occupancy and GPS data were available for research purposes.

LOOP DETECTOR ANALYSIS IN CHANGSHA

The loop detector data used in this study were collected in 2010 in Changsha city, China. At that time 99 intersections were controlled by a SCATS traffic management system. The SCATS system monitors the traffic in a network and controls the signals – based on volumes and occupancy – and imposes coordination control strategy between neighboring intersections based on certain criteria. Traffic volumes and occupancies, essential inputs to the SCATS system, are measured by loop detectors located on a short distance before the stop lines.

In order to verify the quality of loop detectors, the traffic counts from the SCATS detectors were compared with visual observations of the traffic flow in this study. The SCATS detectors are short loops (4.0 m long, 2 m wide) installed in each lane on 1.50 m from the stop line. Video observations were the source for visual counts. Thirteen intersections in the center of Changsha with 208 lanes in total were chosen for this survey; 168 of the lanes have loop detectors. The survey was done on May 14, 2010. On that day only 142 detectors (69%) were operational.

The error rates were statistically analyzed, as shown in Table 1 and Figure 1. It shows that 25% of investigated detectors, excluding fully failed ones, show errors of more than 20%.

The character of the detector data errors is a little different on different lanes and the comparison between lanes with only through going traffic and lanes that carry both going through and turning traffic (mixed traffic) are shown in the following FIGURE 1. We also found differences in detector errors between larger intersections (more than 3 lanes on all approaches) and smaller ones, as shown in TABLE 1.

TABLE 1 Error distributions for the loop detectors on different lanes and different intersections.

	< 5% error	5 to 10% errors	10 to 20% errors	> 20% errors
All intersections	28 %	26 %	21 %	25%
Exclusive straight on traffic Lanes	34 %	28 %	17 %	21 %
Mixed straight on and lefts turning lanes	20 %	22 %	27 %	31 %
Lanes on large intersections	28 %	23 %	21 %	28 %
Lanes on minor intersections	31 %	25 %	19%	15 %

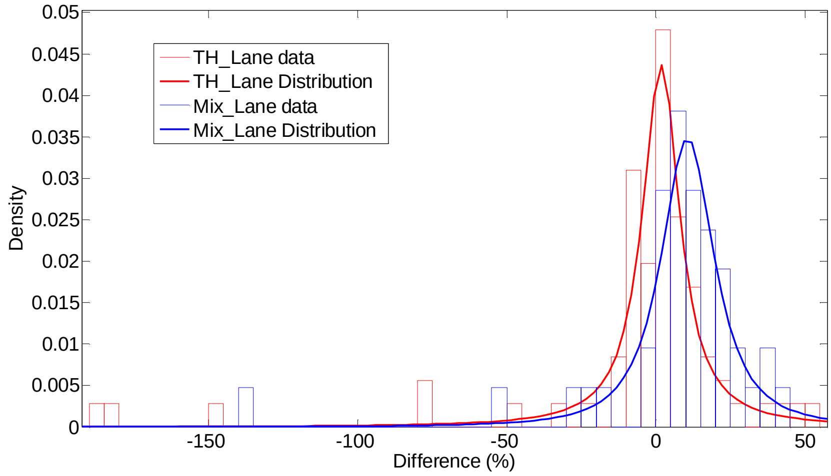
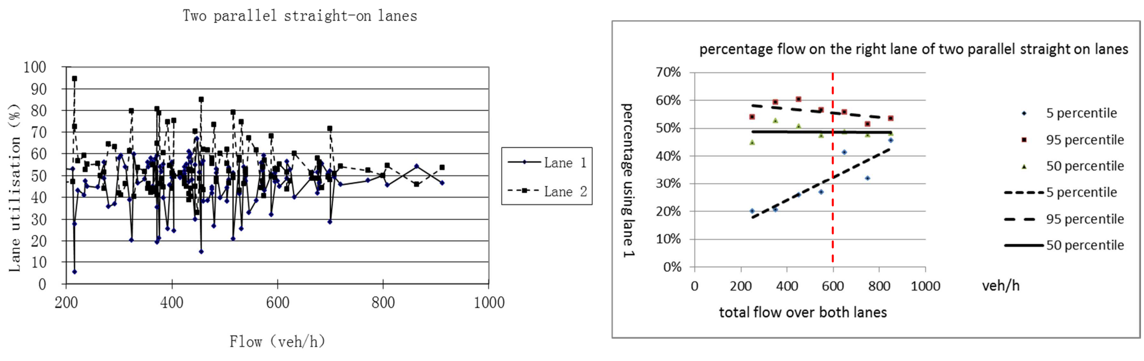


FIGURE 1 Distribution of the errors (observed errors and matching analytical distribution) for exclusive straight-on lanes and mixed traffic lane).

THE IDENTIFICATION OF INVALID LOOP DETECTOR DATA

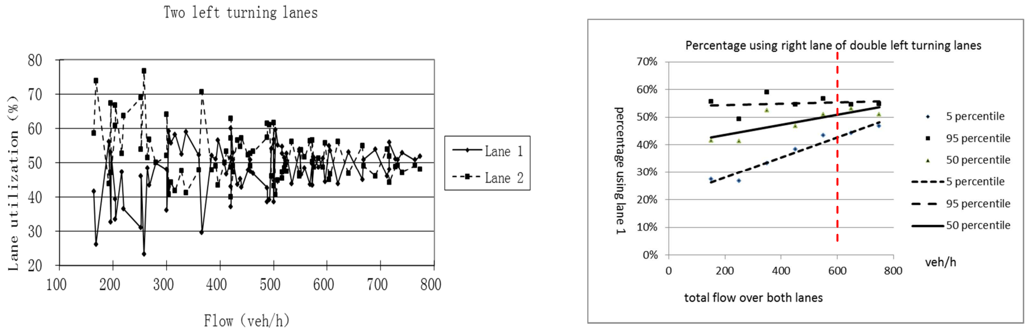
Errors in the automatically collected loop detector data are not always easy to be identified. The extreme failures from malfunctioning detectors are clear when they give zero or very unlikely counts. It is more difficult to identify wrong counts when the difference between the real volumes and the detector counts are relatively small. In that case such detector failures can be detected if different lanes for the same directions have an illogical ratio of volumes.



(a) **(b)**

FIGURE 2 Observed distribution of traffic over two parallel straight-on lanes for different total volumes (lane 1 is the right lane, lane 2 the left one). Left figure (a) gives the observed distribution, the right figure (b) the 5, 95 and 50 percentiles for the left lane (lane 1)

Figure 2 (a) gives the empirical distribution of traffic volumes between parallel straight-on lanes at a specific intersection. As can be seen, the distribution over the different parallel lanes becomes uniform for higher volumes. At lower volumes the distribution may become unbalanced but no systematic difference exists between the two lanes. For left turning lanes a similar pattern can be seen in Figure 3 (b).



(a) (b)

FIGURE 3a Observed distribution of traffic over two parallel left turning lanes; FIGURE 3b shows the 5, 95 and 50 percentiles for the left lane (lane 1)

As shown in Figure 3b, the left lane is generally used more often than the right one at lower flow rates, but at higher flow rates the distribution becomes uniform over the lanes. From this kind of analysis it is possible to derive rules for the range of the distribution between lanes and recognize detector failure. Then the ratio between the flow on a left lane and an inner lane can be presented with a function of the total flow with lower and higher limits (5 and 95 percentiles), as shown in Figure 2b and Figure 3b. Furthermore, the range of the volume distribution can be used to identify the loop detector errors. For instance, for two parallel lanes with total through going traffic 600 veh./h, the percentage p_1 on the right lane should satisfy 32%< p_1<56%, while for two left turning lanes the range for the percentage p_1 on the right lane should satisfy 42 %< p_1<56% .

The boundary values of the volume distribution depend on the total volumes. Taking the left straight-on lane (Lane 1) as an example, the relationship between volume distribution and total volume can be expressed as:

$$7.6 + 0.04\, q < p_1 < 60.0\text{-}0.0077q,$$

where q is the total flow in veh./h over two lanes.

For left turning traffic, the boundaries for data validity are given by

$$0.2067+0.0004\, q < p_1 < 0.54 + 0.00002\, q.$$

With such rules, which have to be calibrated for the different types of intersections, an error detection algorithm can be developed to identify whether the loop detectors are likely to be correct. For instance, for two parallel straight-on lanes, 240 veh./h and 360 veh./h are counted on the right lane (lane 1) and on the inner lane respectively. Then the percentage of total flow rate on right lane is 40%. That is within the boundaries as shown in Figure 2b. However, if the same flow rates were counted on two left turning lanes, Figure 3b shows that it is likely that one or both detectors are not working well.

This method has been applied on the detector data from the SCATS loop detectors on 24 pairs of parallel lanes (both trough going or both left turning), where the ground truth traffic counts were obtained by visual observations. When the difference between the detector counts and the visual observations was larger than 20%, the detector counts were assumed to be incorrect. The number of correct approvals was 12, while in 2 cases the approval was incorrect. The number of correct rejections was 2, and no incorrect rejection was made. Thus, only in 8% of the cases the criterion did not give the correct indication.

If all of the lanes with the same flow direction fail to work, or the lane doesn’t have any parallel lane, the total flow rate cannot be directly obtained from the detectors. The next section explains a method to estimate the total flow rate from probe vehicles. This method does not only provide an estimated flow rate, but also can be used for more purposes, such as the monitoring of queue lengths and the estimation of the percentage of probe vehicles.

In Changsha, the city where we did the detector reliability survey, more than 6000 taxis drive around with GPS. The speed, driving direction and position of these taxis are recorded every 30 seconds. This gives a possibility to estimate traffic flow rates from the queuing of probe vehicles at signalized intersections.

USING GPS DATA TO CHECK AND ENHANCE LOOP DETECTOR COUNTS

Some attempts have been made to derive traffic flow rates from probe vehicle data. Comert and Cetin (*7*) developed a model to estimate the queue length at the end of the red-phase from probe vehicle data, assuming Poisson arrivals. They ignored the effect of the increasing queue during the red phase. In principle this method can be used to calculate the traffic flow rates from the queue length. Kuwahara et al. (*11*) developed a method to estimate vehicle trajectories from a few probe vehicle data. They assume that the traffic flow is according to a fundamental diagram, while no overtaking occurs (first-in-first-out). Peng Hao et al. (*9*) derived a method to estimate the position of a probe vehicle among a set of waiting or driving vehicles. Under certain conditions their method can estimate the flow rate between two probe vehicles. Rahmazani and Geroliminis (*10*) developed a classification method for stopping and accelerating vehicles on the assumption that drivers are a uniform group and drive according to the LWR traffic model. As will be shown in this section, the assumption of no overtaking and a uniformly behaving driver population is not realistic, but such a simplified driving model still provides a good framework for the development of the method to deduce some traffic characteristics.

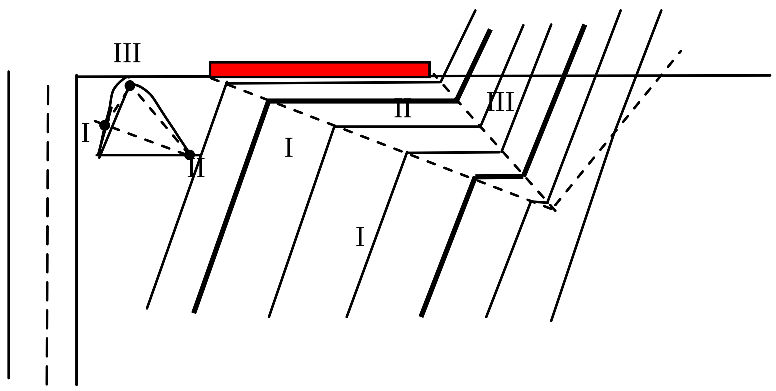


FIGURE 4 Theoretical trajectories of vehicles approaching a signalized intersection. The thick trajectories are from probe vehicles.

Figure 5 shows that several vehicles including two probe vehicles with theoretical trajectories are approaching a signalized intersection (*12*). The points where the vehicles join the back of the queue (i.e. the transition to zero speed) give the transition between the free flow regime (I) and the queue (region II). The back of the queue – counted as the number of vehicles- propagates with a speed equal to the arrival rate.

Since the back of the queue moves backwards against the driving direction, the arrival rate at the back of the queue is slightly higher than the arrival rate at a stationary point. This leads to the following equation for the dynamics of the number of vehicles in the queue N

$$N(t+\Delta t) = N(t) + \Delta t.q(t) + \Delta t.q(t).l\ .d_I(t) \qquad (1)$$

Where

$q(t)$: the arrival flow rate at time t,

l : the average stand still distance between two consecutively queued vehicles, including vehicle length and gap,

$d_I(t)$: the density of the arriving flow at time t.

For l we can fill in $l = 1/d_{II}$, with d_{II} the vehicle density of a queue in one single lane (i.e. the inverse of the length of the road occupied by a single queued vehicle).

The number of lanes close to the intersection is often higher than the number of lanes further upstream from the stop line. The expression for $L(N)$, the length of the queue of n vehicles can be expressed as:

$$L(N) = \frac{N}{n_1 d'_{II}} \qquad for\ N < {n_1 d'_{II}}/{D_1}$$
$$L(N) = D_1 + \frac{(N - {n_1 d'_{II}}/{D_1})}{n_2 d'_{II}} \qquad for\ N > {n_1 d'_{II}}/{D_1} \tag{2}$$

Where

D_1 : the length of the road before the stop line with additional lanes;

n_1: the number of the lanes on the approach;

n_2: the number of the lanes on the upstream section.

Equation (2) simply expresses the fact that the physical length of the queue grows relatively slowly in the wide road section with additional lanes. When the area with additional lanes is filled, the arriving vehicles queue up in the narrower road section with fewer lanes and the queue growth rate becomes larger.

Equation (2) gives the transformation from queue length to number of queued vehicles. Combined with equation (1), it is possible to calculate the arrivals between two moments t_1 and t_2 on which the queue length has been measured. If the length of the queue at both moments is less than D_1, i.e. if the queue builds up in the road section with many lanes, the number of arrivals between t_1 and t_2 is

$$N(t_1, t_2) = \int_{t_1}^{t_2} q(t')\{1 + d_I(t') / n_1 d'_{II}\} dt' \qquad L_Q(t_1) < D_1 \text{ and } L_Q(t_2) < D_1 \tag{3}$$

$$L(t_2) - L(t_1) = N(t_1,t_2) / n_1 d'_{II} \tag{4}$$

A similar expression is applicable when the queue length at t_1 and t_2 is both larger than D_1. In the case that the transition occurs in the time interval that the queue grows from the wide road section with many lanes to the narrower section with fewer lanes, the equations (3) and (4) become slightly more complicated.

The simplest case is a uniform arrival rate so that the density and flow rates are constants in the time and that d_I can be approximately estimated by:

$d_I = q(t)/v\ (t) \approx q/v$

with v is the average speed derived from a fundamental diagram, as shown in FIGURE 4.

FIGURE 5 shows the stopped probe vehicles along a road in a time – distance diagram on a macroscopic scale. The data were extracted from taxis GPS data set in Changsha. The traffic control cycles are clearly visible in the periodicity of the queue, the propagating waves of the back, and the front of the queues. The pattern is best visible on a macroscopic scale, like in FIGURE 5. If the figure is analyzed in more detail, the boundary around the queued vehicles appears to be fuzzier than as it appears on the macroscopic figure.

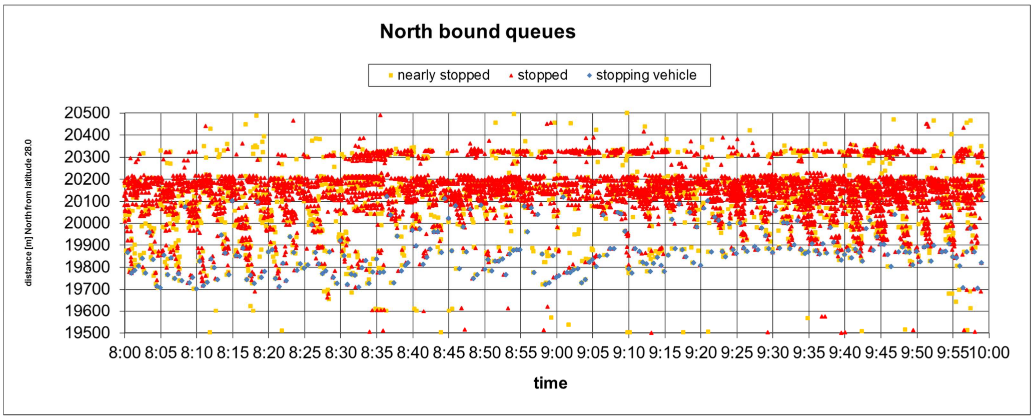


FIGURE 5 Time – distance diagram of slowly driving, stopped probe vehicles and vehicles that make the transition between driving and being stopped

One reason is that the probe vehicles don't behave as the idealized model as shown in FIGURE 4. In FIGURE 6 we see the trajectories of probe vehicles of a half an hour on the same road section in more detail. Looking more closely to the trajectories, we can observe that the drivers don't all behave in the same way: some drivers stop abruptly, others decelerate slowly at the approach of the back of the queue. These vehicles join the queue at a later moment than the abruptly stopping vehicles. Several trajectories show that some drivers just decelerate without stopping. For these vehicles the queue length at arrival should be estimated through redrawing the trajectories: replacing them by pieces of full speed and zero speed.

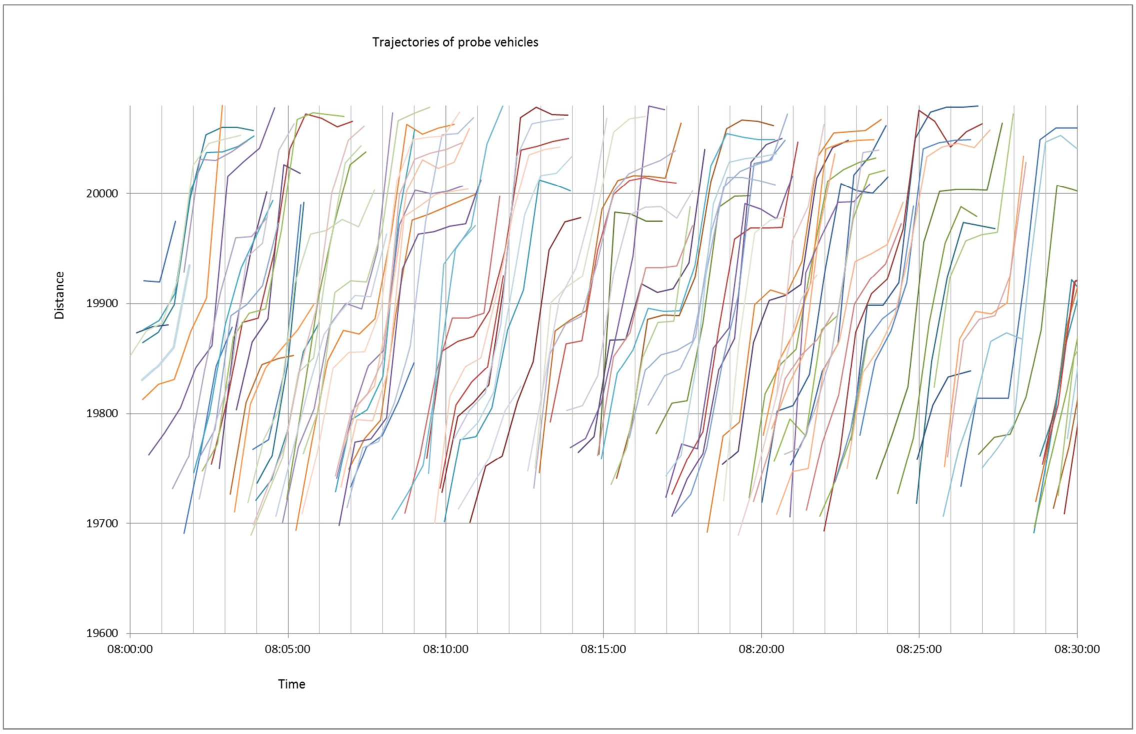


FIGURE 6 Trajectories of probe vehicles with a polling time interval of 30 seconds. The accuracy of the position is about 6 m on average.

FIGURE 7 is a subsection of FIGURE 6. It shows that the fuzzy boundary between driving vehicles becomes much sharper when the trajectories are replaced by straight lines corresponding to two states: full speed moving and fully standing still. This means that the 'back of the queue' is only a theoretical concept, not clearly visible in reality. Trajectories can be replaced with 'ideal' ones with only going and stopping phases. Of course, such an adaptation of the trajectories can be done with a smart computer program, but it is also possible to estimate the most likely back of the queue visually.

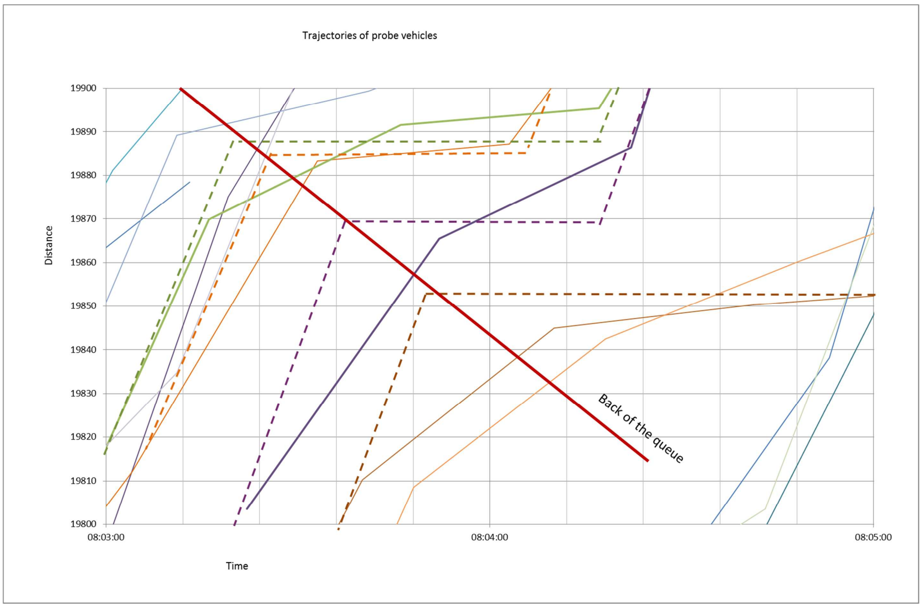


FIGURE 7 Observed trajectories (solid lines) and reconstructed trajectorie (dashed lines)

5. Empirical test

For the empirical test of the method, we selected another road (Laodong road) in Changsha. A video camera was used to record the traffic around the intersections for one hour from a high building. The number of lanes of the relevant road section on one direction was two. The traffic volumes were counted manually from the video observation, as shown in TABLE 2. The GPS data were selected by map matching and a sample of the raw taxi GPS trajectories is presented in Figure 9.

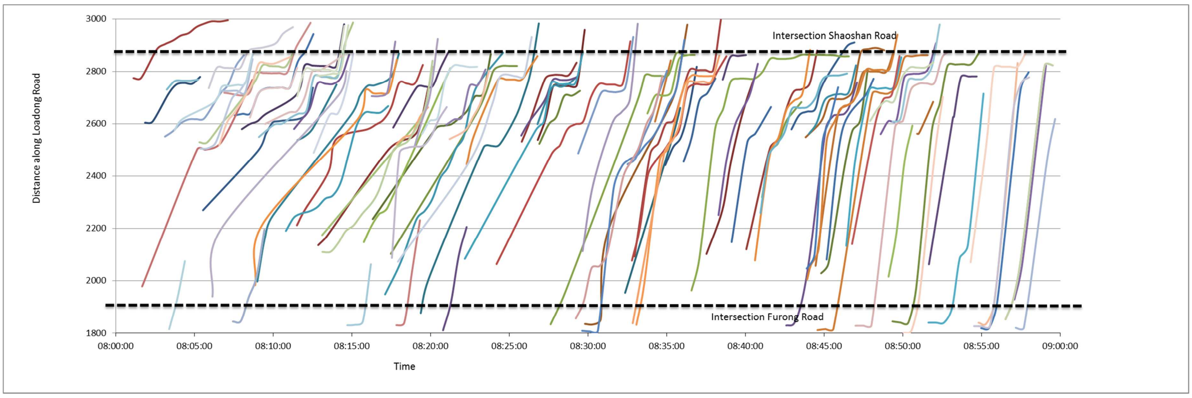


FIGURE 8 Trajectories of taxis along the Laodong Road in Changsha eastwards between Furong Road (in the West) and Shaoshan Road (in the East).

Given the fact that the back of the queue cannot directly be determined by checking the stopped vehicles only, we used the visual method to estimate the dynamics of the back of the queue. From FIGURE 8, several irregularities in driving can be observed: vehicles in the same traffic conditions travel at different speeds, overtake each other and travel on different lanes, even on the parallel road (dedicated for bicycles but also allowed for motor vehicles that turn right are the next intersection). Furthermore, the driving speed of vehicles is significantly lower in the time period 8:00 to 8:30 (visible in the slope of the trajectories) when spill back of the queue occurred. Trajectories are redrawn by extending the free driving parts and by interpolating the stationary parts. In FIGURE 9 the extrapolated points of stopped taxis are connected with dotted lines. The slope of this line should be an indication of the flow rate, as estimated in TABLE 2. Even though the percentage of taxi probe vehicles was rather low, we could still identify at least two taxis in each red phase and we could use the time and location of the stopping taxis to estimate the traffic flow between these two moments.

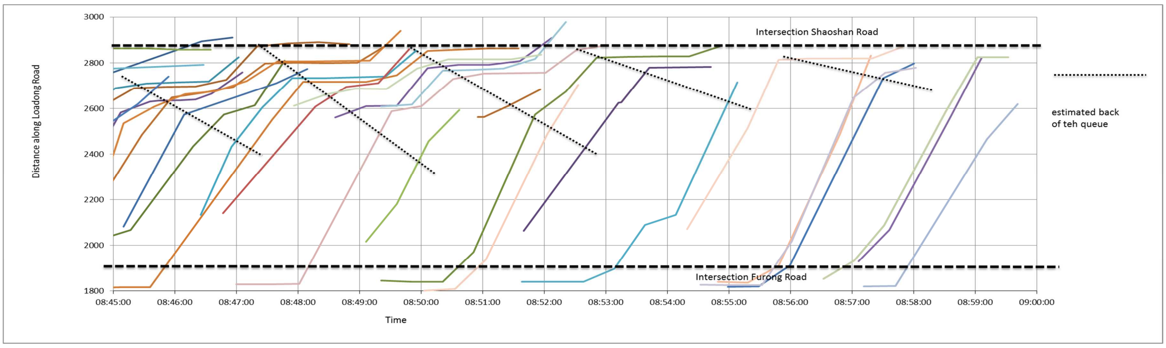


FIGURE 9 Trajectories of taxis and estimated length of the queue for a period of 45 minutes.

For the purpose of estimating the traffic volume from GPS data we assume that equation (4) can be approximated by:

$$L(t_2) - L(t_1) \cong (t_2 - t_1)\{q / n_1 d'_{II} + (q / n_1 d'_{II})^2 / v\} \qquad (5)$$

From equation (5) the flow rate q can directly be solved.
The comparison of traffic volume between video manually counted and GPS estimated is shown in Table 2.

TABLE 2 Counted volumes and volumes derived from the queues

Time of the day	8:00-8:15	8:15-8:30	8:30-8:45	8:45-9:00
Counted flow in veh/s	0.485	0.522	0.5033	0.4622
Observed growth in ($N(t)$ / s (± standard deviation)	1.542 (±0.3)m/s	1.375 (±0.3)	1.5417 (±0.2)	1.0333(±0.14)
Flow derived from queues	0.555 (±0.102)	0.495(±0.099)	0.555 (±0.075)	0.372 (±0.052)
Error	16%	-5%	10%	-19%

In the conversion between number of vehicles in the queue and queue length, the jam density d_{II} of 360 veh/km was assumed (2 lanes). It is obvious that the traffic volume derived from GPS data does not give very accurate results with errors between 5 and 19%. There are two reasons for this:

- The volumes are not so regular and constant over the time due to stochastic effects in the arrival process and the influence of platooning of arriving traffic (*4,13*); this is visible in the fact that the flows derived from the queue have a rather large standard deviation,
- The position of the back of the queue cannot be determined accurately because of the driving style, i.e. the early deceleration from a long distance to the back of the queue.

Furthermore, some turning lanes might have a different queue length than straight-on lanes. GPS of the taxis cannot give sufficiently accurate positions to distinguish different lanes. On section 4 we assume that all lanes fill up simultaneously, which is a kind of simplification of the real traffic and can result in some errors in estimation. Notwithstanding the limitations of the method, the correlation between manually counted flows and flows estimated from the GPS queue length is still not too bad: R=0.59.

CONCLUSIONS

As an indication for the traffic state, traffic volumes are very important in most traffic control systems. Traffic volumes are both essential parameters for the evaluation of the traffic and for simulations. Through a traffic survey made in Changsha, we found that volumes recorded by loop detectors are uncertain due to detector failures or irregular driving paths of vehicles. The regularities in the distribution of volumes over parallel lanes make it possible to detect detector failures if the ratio of counts between two lanes is not logical. Errors can be found and repaired.

An innovative method was developed to derive traffic volumes from probe vehicles. Even though the actual traffic did not behave like the simple fundamental diagram, we could derive a simple method to estimate traffic flow rates from the dynamics of the queue. Further refinements of this method, especially the reconstruction of trajectories to eliminate the irregularity of the drivers, can lead to a traffic flow estimation that will enrich the loop detector data and can be used as an independent flow measure. The method of the trajectories analysis and queue length estimation can provide relevant additional information for traffic monitoring and control.

ACKNOWLEDGEMENTS

This research has been made possible by grants for Changsha Science and Technology Commission under number K1001010-11 and K1106004-11. Also the financial support of vZC is acknowledged. The authors thank the anonymous reviewers for their useful comments.

REFERENCES

1. Ramezani, A., Moshiri, B. (2011) The traffic condition likelihood extraction using incomplete observation in distributed traffic loop detectors. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC , art. no. 6082919 , pp. 1291-1298
2. Sun, S., Zhang, C.(2005), Using a random subspace predictor to integrate spatial and temporal information for traffic flow forecasting. Lecture Notes in Computer Science 3611 (PART II) , pp. 652-655.
3. Zhan, X., Hasan, S., Ukkusuri, S.V., Kamga, C. (2013) Urban link travel time estimation using large-scale taxi data with partial information. Transportation Research Part C: Emerging Technologies 33 , pp. 37-49
4. Zheng, F., Van Zuylen, H., (2011). Estimating the delay distribution for urban trips based on sample measurements. In: Transportation Research Board 90th Annual Meeting. No. 11-0831.
5. Zheng, F., Van Zuylen, H. (2013) Urban link travel time estimation based on sparse probe vehicle data. Transportation Research Part C: Emerging Technologies 31 , pp. 145-157.
6. Jenelius, E., Koutsopoulos, H.N. (2013) Travel time estimation for urban road networks using low frequency probe vehicle data. Transportation Research Part B: Methodological 53 , pp. 64-81
7. Comert, G., Cetin, M., 2009. Queue length estimation from probe vehicle location and the impacts of sample size. European Journal of Operational Research 197, 196–202.
8. Ban, X, Hao, P. and Sun, Z. (2011) Real time queue length estimation for signalized intersections using sampled travel times. Transportation Research C, 19, 1133-156.
9. Peng Hao, Zhanbo Sun, Xuegang Ban, Dong Guo, Qiang Ji (2013) Vehicle index estimation for signalized intersections using sample travel times. In S.P. Hoogendoorn, V.L. Knoop and H. van Lint (eds) Proceedings 20th International Symposium on Transportation and Traffic Theory. Procedia Social and Behavioural Science 80 (2013)
10. Ramezani, Mohsen and Nikolas Geroliminis (2013) Exploiting Probe Data to Estimate the Queue Profile in Urban Networks. Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, October 6-9, 2013
11. Kuwahara, Masao, Koichi Abe, Takeshi Imai, 2013. Estimating vehicle trajectories on a signalized urban arterial and a motorway by data fusion of probe and detector data. Optimum 2013, International Symposium on Recent Advance in Transport Modelling.
12. Newell, G. F., 1982. Applications of queuing theory, 2nd Edition.
13. Van Zuylen, H.J., S.P. Hoogendoorn (2007) A probabilistic calculation of queue lengths at signalized intersections. Proceedings ITS World Congress Beijing.