2011 14th International IEEE Conference on
Intelligent Transportation Systems
Washington, DC, USA. October 5-7, 2011

Estimation of Measures of Effectiveness Based on Connected Vehicle Data

Juan Argote, Eleni Christofa, Yiguang Xuan, and Alexander Skabardonis

Abstract— **Vehicle-infrastructure cooperation via the Connected Vehicle initiative is a promising mobile data source for improving real-time traffic management applications such as adaptive signal control. This paper focuses on developing estimation methods with the use of Connected Vehicle data for several measures of effectiveness (e.g., queue length, average speed, number of stops), essential for determining traffic conditions on urban signalized arterials for real-time applications. This research systematically determines minimum penetration rates that allow accurate estimates for a wide range of measures of effectiveness in undersaturated traffic conditions. The estimation of these measures and minimum penetration requirements has been tested using Next Generation Simulation (NGSIM) data.**

I. INTRODUCTION

Efficient design of control strategies on signalized arterials are crucial for mitigating congestion in urban areas. Traffic signals are the main mechanism for managing arterial network capacity, yet the control of traffic signals has not significantly changed over the past several decades, despite rapid and profound changes in electronics, sensor and communication technologies, and software. The main impediment to improvements in traffic signal systems has been the limited ability of available fixed-point sensors (e.g., loop detectors, cameras) to measure the true state of the traffic network and its response to signal changes. Vehicle-infrastructure cooperation via the Connected Vehicle initiative could provide this comprehensive real-time information on the movements of vehicles throughout the entire road network and allow a transformational change in how we control traffic and, finally, significantly reduce congestion along arterials.

This paper focuses on developing algorithms for estimating Measures Of Effectiveness (MOEs) from sampled Connected Vehicle (trajectory) data and identifying the penetration rate requirements for accurate estimates in undersaturated traffic conditions. These MOE estimates can be used in order to improve the performance of traffic management applications such as adaptive signal control. First, three queue length estimation processes, Maximum Likelihood (ML), Method of Moments (MM) and one based on Kinematic Wave Theory (KWT) [1], [2] are proposed and their accuracy for different penetration rates is investigated. Next, the process for estimating several other MOEs is presented and the minimum penetration rate requirements for achieving accurate estimates are identified.

This work was supported by the FHWA Exploratory Advanced Research Program, US DOT

J. Argote, E. Christofa, and Y. Xuan are with the Department of Civil and Environmental Engineering, Institute of Transportation Studies, University of California, Berkeley, CA, 94720, USA {juan.argote; christofa; xuanyg}@berkeley.edu

A. Skabardonis is a Professor in the Department of Civil and Environmental Engineering, Institute of Transportation Studies, University of California, Berkeley, CA, 94720, USA skabardonis@ce.berkeley.edu

II. BACKGROUND

Several studies have been conducted in estimating MOEs with the help of Connected Vehicle or probe data. The majority of these studies have focused on the estimation of queue length [3], [4], [5], [6], which is a commonly used measure for designing and evaluating control strategies and is difficult to estimate with static sensors. Probe vehicle data were used by Comert and Cetin [3] who studied the distribution of average queue lengths assuming known penetration rate, while Venkatanarayana et al. [4] estimated queue lengths by using the position of the last IntelliDrive^{SM} equipped vehicle in queue (i.e., ML estimation). Hao and Ban [5] and Cheng et al. [6] used KWT to estimate queue lengths. However, only Comert and Cetin [3] considered the impact of the penetration rate of the equipped vehicles on the estimation accuracy.

Other studies have focused on estimating travel time and identifying the number of probe vehicles or penetration rate necessary for reliable travel time estimation on urban arterials [7], [8]. Both studies concluded that even low penetration rates can provide satisfactory travel time estimates for freeways and congested arterials. Li et al. [9] and Cetin et al. [10] also developed travel time estimation algorithms with the use of Vehicle-Infrastructure Integration (VII) and probe vehicle data respectively and investigated the impact of penetration rate on the estimate accuracy. VII data were also used by Vasudevan [8] for the estimation of mid-block flow rates and speeds.

None of the studies has managed to provide a comprehensive description of the estimation process for a wide range of MOEs and identify penetration rate thresholds that ensure accurate estimates. This paper presents the first study that systematically determines minimum penetration rates that allow accurate estimates for a variety of MOEs essential for real-time traffic management applications.

III. DATA

The impact of penetration rate on the accuracy of the MOEs has been estimated with the use of real-world vehicle trajectories obtained by the Next Generation Simulation (NGSIM) program [11]. More specifically, NGSIM data collected at Peachtree Street, Atlanta, GA for four signalized intersections have been chosen for this study. The NGSIM

978-1-4577-2197-7/11/$26.00 ©2011 IEEE

data set consists of vehicle trajectories; the format includes vehicle ID, time, position, lane, speed, and acceleration. All data are available at a resolution of $\Delta T = 0.1$ seconds. In addition, vehicle type, vehicle length, and their origins and destinations are available. The traffic signals' settings (cycle length, green, yellow, and offsets) are also known. The data used in this study were collected from 12:45 to 13:00. A total of 1,115 vehicles were processed. However, the number of vehicles reduced to 228 after only considering the ones that travel on the northbound direction of the main arterial.

IV. Queue Length Estimation

Estimation of queue length with the use of vehicle trajectory data can be done by different methods. However, all these methods require identifying the position of the vehicles in queue at an approach. As a result, in order to estimate the queue length for one intersection approach during a single cycle we need to determine the spatio-temporal domain of the estimation and consequently the sections of the sampled trajectories considered in the estimation. In addition, we need to identify the appropriate deceleration points of the sampled trajectories, i.e., the points where the vehicles start decelerating when they are joining the queue.

A. Discretization of Space and Time

First, the time-space of interest is discretized in different rectangular regions accounting for the different links in the arterial l={1, 2, 3, 4} and the different signal cycle times k={1, 2, ..., 9} (Fig. 1). This allows us to identify several cells noted by the pairs (k,l) with their corresponding queues.

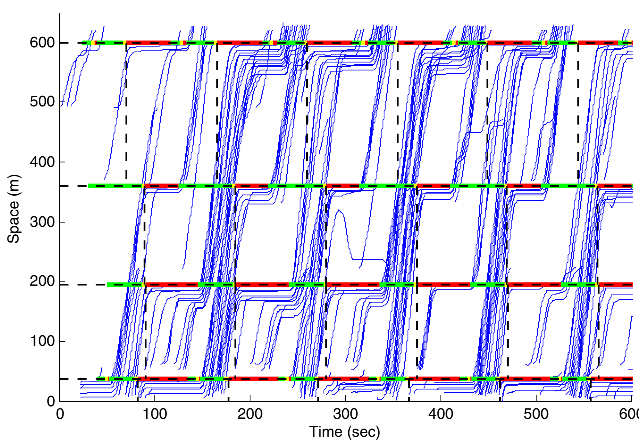

Fig. 1. Discretization of space and time

B. Identification of Deceleration and Acceleration Points within a Queue

Let us now focus on a specific cell (k,l) and drop the notation (k,l) to facilitate expression. Next, all the deceleration points corresponding to that cell have to be identified. By doing so, the positions of stopped vehicles can be determined. The time corresponding to the deceleration point of vehicle i (i=1, ...,s) within a cell t_i^d is determined by selecting the time that satisfies the following condition:

$$t_i^d = min\{t|v_i(t) > v_c \ \& \ v_i(t+\Delta T) \le v_c \\ \& \ (t, y_i(t)) \in \text{cell}(k,l)\} \quad (1)$$

where $v_i(t)$ is the speed of vehicle i at time t, v_c is a constant speed threshold (e.g., 5 km/hr), i.e., the vehicle is practically stopped, ΔT is the sampling interval and $y_i(t)$ is the position of vehicle i at time t. These conditions identify the points in time and space within the cell where vehicle i joins the queue. A vehicle is considered to be in queue from the moment it starts decelerating with a speed that is lower than the specified threshold. For this reason, it is necessary to consider the minimum of the times (or the earliest deceleration point) that satisfies those conditions, so that non-desirable deceleration points are not taken into account.

All the times that satisfy (1), the set $\{t_i^d\}$, can now be rearranged defining a new set of times $U = \{t_{(i)}^d\}$ so that $y_{(1)}(t_{(1)}^d) \ge y_{(2)}(t_{(2)}^d) \ge ... \ge y_{(s)}(t_{(s)}^d)$. Note that all the times included in this unfiltered set U do not necessarily correspond to deceleration points within the queue. As shown in Fig. 2, some of the points included in U may correspond to vehicles which are not in the queue, e.g., stopping within the link in order to park. Therefore, we will need to impose more constraints to eliminate these undesired points.

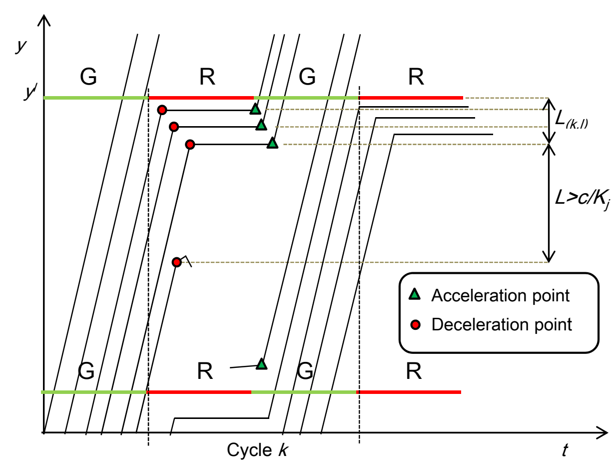

Fig. 2. Identification of deceleration and acceleration points within the queue

We can see in Fig. 2 that the undesired deceleration points are far from the deceleration points within the queue. Thus, the distance between deceleration points should be used for filtering. If vehicles within the same queue are sampled independently with probability p, the number of vehicles between adjacent sampled vehicles is a random variable satisfying a geometric distribution with parameter p. The 100(1-ε)th percentile of a geometric distribution with parameter p is $\ln(\varepsilon)/ln(1-p)$ (we used the 90th percentile in this paper, which is $\ln(0.1)/ln(1-p)$). Therefore, the filtering process is defined by the next condition, with N_L

being the number of lanes on the approach and K_j being the jam density per lane:

$$R = \left\{ (i) \middle| d_{(i)} \triangleq |y_{(i)}(t^d_{(i)}) - y_{(i+1)}(t^d_{(i+1)})| > \frac{c}{K_j} = \max\left\{ \frac{\ln(\varepsilon)/\ln(1-p)}{N_L K_j}, 1/K_j \right\} \right\} \quad (2)$$

After applying (2) to trajectories $(i) \in \{(1), \ldots, (s-1)\}$, the set of vehicle indices to be considered in the queue length estimation is given by:

$$A = \begin{cases} \{(1)\} & \text{if } s = 1 \\ \{(1), ..., (s)\} & \text{if } R = \{\emptyset\} \\ \{(1), ..., \min\{R\}\} & \text{otherwise} \end{cases} \quad (3)$$

Finally, F defines the set of all deceleration points that are within the queue of interest.

$$F = \{t^a_{(i)} | (i) \in A\} \quad (4)$$

Note that F is a subset of U, and if two deceleration points are too far apart, we would suspect that the upstream one is not actually in the queue.

A similar method can be applied to identify the acceleration points. The acceleration points will be denoted similarly as $t^a_{(i)}$ and they are always paired with deceleration points. Fig. 3 shows the deceleration (dark asterisks) and acceleration (light asterisks) points at intersection 2 after filtering.

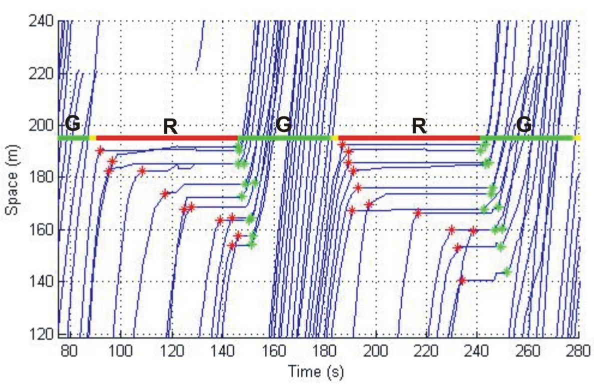

Fig. 3. Detailed view of deceleration and acceleration points (Int. 2)

C. Queue Length Estimation Methods

Queue length can be easily estimated if the penetration rate is 100%. But with only some of the vehicles sampled, queue length estimation is a challenge. We analyze three possible methods to tackle this problem. For each estimation method 2,000 samples are obtained where each vehicle trajectory is sampled following a Bernoulli trial with probability p.

1) Maximum Likelihood Queue Length Estimation: Under the assumption that the position of the vehicles in queue can be characterized by a discrete uniform distribution, we can use ML estimation. In this case, the estimator of the queue length, $\hat{L}_{ML}$ will be the relative position of the vehicle located further apart from the intersection: the maximum order statistic for the relative position of the vehicles within the queue with respect to the intersection (Fig. 4). From equation (3), if we define the number of elements in A (or its cardinality) to be n_A we have:

$$\hat{L}_{ML} = |y^l - y_{(n_A)}(t^d_{(n_A)})| \quad (5)$$

where y^l is the location of the downstream intersection of link l. As we saw in the previous section the implementation of this algorithm is fairly simple. If this algorithm is applied assuming 100% penetration rate, the estimate obtained gives the ground truth queue length.

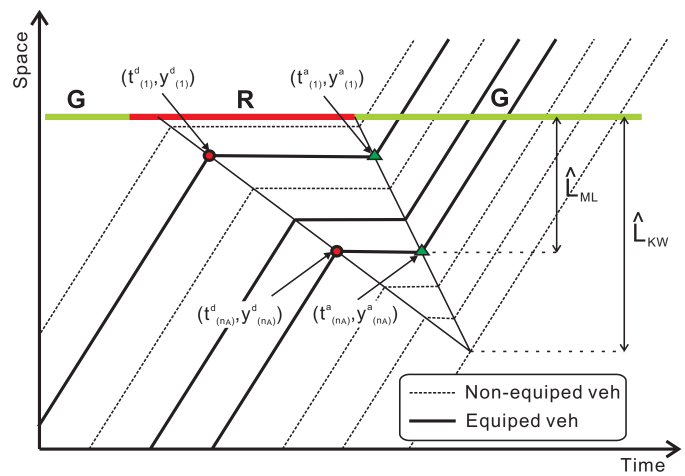

Fig. 4. ML and KWT queue length estimation

However, for penetration rates different than 100% this estimator is biased, i.e., the average value given by the estimator will not be equal to the real value of the queue length. Instead, this estimator will tend to underestimate the actual queue length. For low penetration rates it is very likely that our sample of vehicle positions within the queue will not include its last vehicle and therefore, we will tend to underestimate queue lengths.

2) Method of Moments Queue Length Estimation: The method of moments is also used under the assumption that the vehicles in queue are characterized by a discrete uniform distribution. In this case, the queue length estimator, $\hat{L}_{MM}$ is equal to two times the mean of the sampled relative positions:

$$\hat{L}_{MM} = 2\left| y^l - \frac{1}{n_A}\sum_{i=1}^{n_A} y_{(i)}(t^d_{(i)}) \right| \quad (6)$$

This method is unbiased theoretically. However, in reality the cars in a queue will not be uniformly spaced, and even with a 100% penetration rate we would not observe the actual queue length, which means that this method yields worse results than the previous one.

3) Queue Length Estimation based on Kinematic Wave Theory: The application of queue length estimation based on KWT requires at least two measurements of deceleration-acceleration points from all the vehicle trajectories in a given cell (k,l). Fig. 4 shows how the shockwaves predicted by KWT can be estimated. The intersection point of the two shockwaves determines the maximum possible queue length, $\hat{L}_{KW}$:

$$\hat{L}_{KW} = \left| y^l - \frac{\frac{y_{(1)}(t^d_{(1)})}{w^d} - \frac{y_{(1)}(t^a_{(1)})}{w^a} + t^a_{(1)} - t^d_{(1)}}{\frac{1}{w^d} - \frac{1}{w^a}} \right| \quad (7)$$

where:

$$\frac{1}{w^d} = \frac{t^d_{(n_A)} - t^d_{(1)}}{y_{(n_A)}(t^d_{(n_A)}) - y_{(1)}(t^d_{(1)})} \quad (8)$$

$$\frac{1}{w^a} = \frac{t^a_{(n_A)} - t^a_{(1)}}{y_{(n_A)}(t^a_{(n_A)}) - y_{(1)}(t^a_{(1)})} \quad (9)$$

In undersaturated conditions the beginning of the first shock-wave (the threshold between the free flowing and queued states) does not necessarily match the beginning of the red phase in the signal cycle. This makes it necessary to sample at least two vehicle trajectories within a given cycle, in order to be able to estimate the maximum potential queue length. Thus, before we apply the proposed estimation algorithm we need to know which the minimum acceptable penetration rate p is. A possible metric to determine p would consist of calculating the value that guarantees that at least 90% of the time two vehicle trajectories on the same lane will be sampled. Let us assume that the number of vehicles traveling on one lane can be modeled as a discrete Uniform(N_{min}, N_{max}) distribution. Then, for a given number of cars traveling on one lane, approaching an intersection, the probability of selecting at least two of them, based on a binomial distribution, is:

$$P(s_L \geq 2 | N_L = m) = 1 - P(s_L < 2 | N_L = m) = 1 - \sum_{j=0}^{1} \binom{m}{j} p^j (1-p)^{m-j} \quad (10)$$

where s_L is the number of vehicles sampled in lane L and N_L is the total number of vehicles at that lane for a given cycle. This formula in combination with the assumption that N_L can be modeled as a discrete Uniform distribution, allows us to compute the probability of sampling at least two vehicles traveling on lane L in a given cycle as:

$$P(s_L \geq 2) = 1 - \sum_{m=N_{min}}^{N_{max}} P(s_L < 2 | N_L = m) P(N_L = m) = 1 - \frac{1}{N_{max} - N_{min} + 1} \sum_{m=N_{min}}^{N_{max}} \left(\sum_{j=0}^{1} \binom{m}{j} p^j (1-p)^{m-j} \right) \quad (11)$$

To illustrate this, let us assume that the number of vehicles arriving at an intersection in a specific lane ranges from 1 to 10 vehicles, then Fig. 5 provides the probability $P(s_L \geq 2)$ for different penetration rates. The figure indicates that penetration rates higher than 50% ensure sampling of at least two vehicles that travel on the same lane within a cell more than 75% of the time.

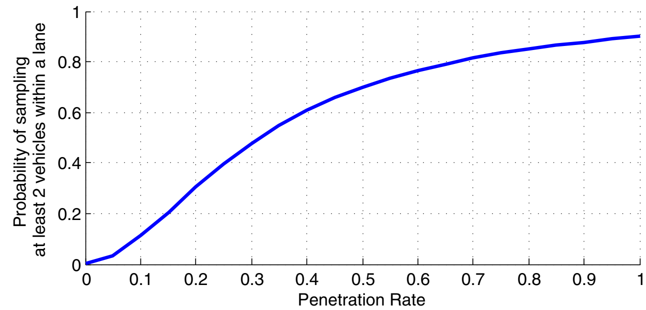

Fig. 5. Probability of sampling at least two vehicles for a given penetration rate, with N_{min} = 1 and N_{max} = 10)

Unfortunately, the undersaturated conditions in the NGSIM data set are not appropriate for the use of the estimator based on the KWT. The low number of vehicles and the non-uniformity of their arrivals at the intersections result in the overestimation of the queue length. Fig. 6 compares the average absolute relative errors obtained by the three queue length estimation methods.

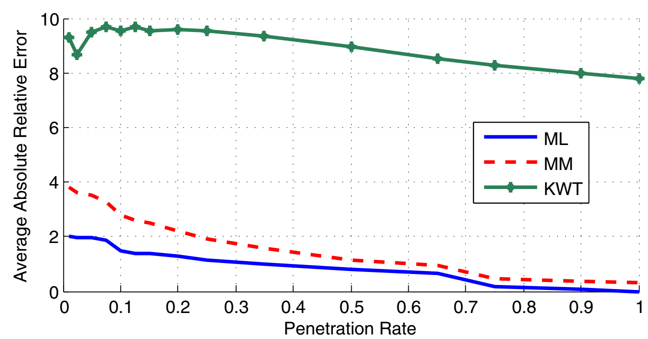

Fig. 6. Average absolute relative error vs penetration rate for the three estimation methods

The maximum likelihood estimation clearly provides the best results, although it would require a penetration rate of at least 80% to guarantee that the average absolute relative error of the estimation does not exceed 10%.

V. Estimation of Other Measures of Effectiveness

This section presents the estimation of various MOEs considering different p values. For each of these values, 10,000 samples are obtained where each vehicle trajectory is sampled following a Bernoulli trial with probability p. The results presented allow to determine the minimum p required to accurately estimate a given MOE.

A. Average Speed

The average speed for a given sample is obtained by using Edie's generalized average speed definition [12], including all the vehicles in the sample:

$$\bar{v} = \frac{\sum_{i=1}^{S} l_i}{\sum_{i=1}^{S} t_i} \tag{12}$$

where S is the total number of vehicle trajectories sampled, l_i is the total distance traveled by vehicle i and t_i is the total time spent by that vehicle to traverse l_i. After computing the average speed for each one of the 10,000 samples it is possible to obtain a mean value and the standard deviation of those average speeds for each of the penetration rates considered. Fig. 7 shows the mean value of the average speed estimate for different penetration rates.

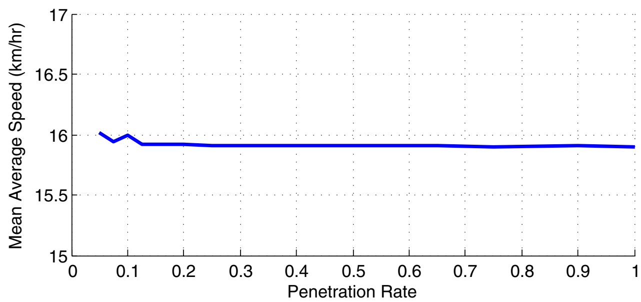

Fig. 7. Average speed estimates vs penetration rate

Although the results presented in Fig. 7 show a very small variability for the estimate of the average speed, it is necessary to recall that these results consider the mean value across all 10,000 samples for a given p. In reality, with low p the average speed estimate could vary greatly from one sample to another. This variability can be shown with the use of a box plot (Fig. 8). A penetration rate is considered acceptable if $\pm 2.7\sigma$ of the estimated average value (whisker range) lays within a 10% of the average value for the ground truth. Under this consideration, penetration rates higher than 50% provide accurate enough speed estimates.

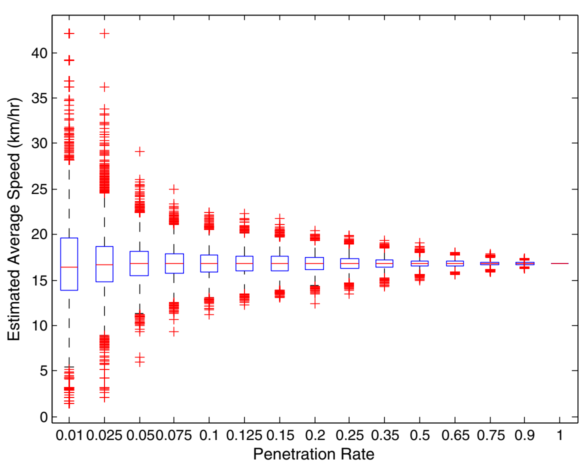

Fig. 8. Box plot of average speed estimates vs penetration rate

B. Average Delay per Unit Distance

Delay is defined as the difference between the actual travel time to traverse an arterial section and the travel time under free flow conditions. The average delay per unit distance for a given sample is obtained by the following expression:

$$\bar{D} = \frac{1}{S} \sum_{i=1}^{S} \frac{1}{l_i} \left(t_i - \frac{l_i}{v_f} \right) \tag{13}$$

where $\bar{D}$ is the average delay for a single sample of vehicles, and v_f is the free flow speed on the arterial corridor (a free flow speed of 40 km/h was used in this study). Again, the box plot (Fig. 9) indicates that samples obtained under low penetration rates present a significant number of outliers, which could lead to a misinterpretation of the traffic conditions. For this case penetration rates higher than 80% would be necessary for accurate estimates.

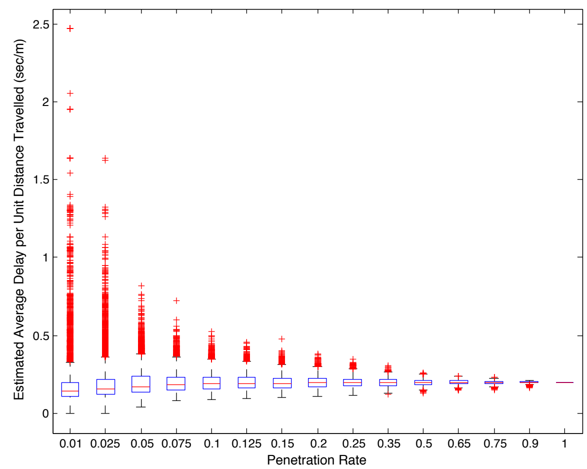

Fig. 9. Box plot of average delay estimates vs penetration rate

C. Average Number of Stops

The average number of stops is obtained as the mean number of stops of all vehicles within the sample that are traveling through the arterial on the direction under consideration. A stopped vehicle has been defined as a vehicle with speed less than 5 km/h. As in the case of average speed and delay estimates, the mean of the average number of stops does not vary a lot even for low penetration rates. However, a more detailed look at the box plot (Fig. 10) shows a high variability in the average number of stops for low penetration rates. In this case, penetration rates higher than 50% would be necessary to obtain results that would lie close enough to the ground truth value.

D. Average Acceleration Noise

Acceleration noise is defined as the standard deviation of a vehicle's acceleration and is used as a measure of the smoothness of traffic flow along signalized arterials. The average acceleration noise is estimated as the mean of each of the vehicle accelerations' standard deviation. Fig. 11 shows again that variability is much higher for low p; values of p on the order of 10% and above should be sufficient to obtain reliable standard deviations for the acceleration.

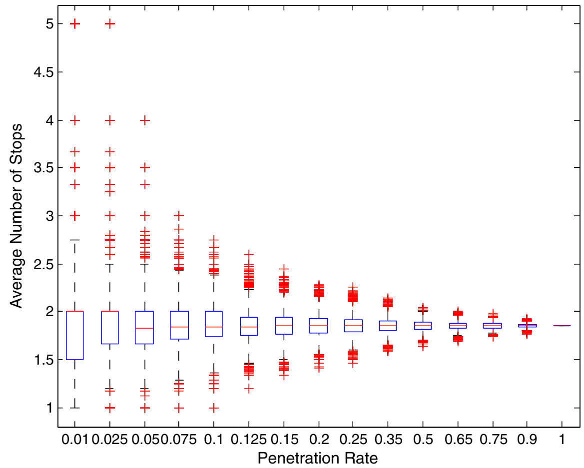

Fig. 10. Box plot of average number of stops estimates vs penetration rate

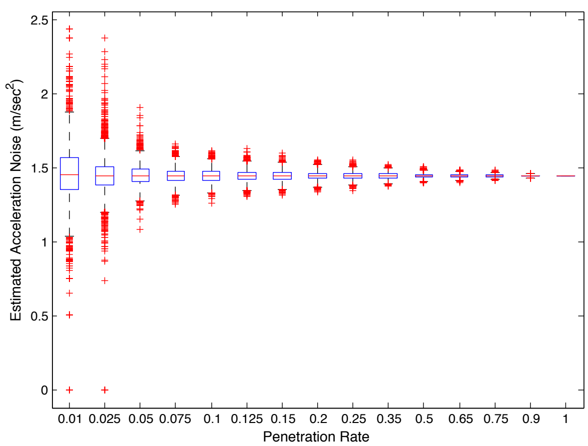

Fig. 11. Box plot of acceleration noise estimates vs penetration rate

VI. DISCUSSION

The study conducted reveals that the required penetration rates for accurate estimates vary across the different MOEs. The analysis of three different methods to estimate queue length revealed that the most effective approach for undersaturated conditions is the simplest one: the maximum likelihood estimation. However, even that method requires a minimal 80% penetration rate to guarantee an average absolute relative error smaller than 10%.

For the rest of the MOEs considered, the minimum penetration rates were estimated by analyzing the whisker ranges in the box plots. As indicated in Table I, average speed and delay can be estimated accurately with relatively low penetration rates (50%). In addition, average acceleration noise requires only a 10% penetration rate to provide estimates within a 10% of the ground truth. Therefore, this study reveals the MOEs that can be effectively used in the initial and middle stages of the Connected Vehicle technology implementation, when penetration rates will be lower than 100%.

TABLE I

MINIMUM PENETRATION RATE FOR MEASURES OF PERFORMANCE ESTIMATES WITHIN 10% OF THE GROUND TRUTH

Performance Measure	Required Penetration Rate (%)
Average Speed (km/hr)	50%
Average Delay (sec/m)	80%
Average Number of Stops	50%
Acceleration Noise (m/s2)	10%

Next steps include testing these MOEs estimation methods for oversaturated conditions and fusing connected vehicle data with loop detector data to obtain more accurate estimates. Moreover, these estimates (e.g. queue length) can be used as a trigger to adjust the signal settings at upstream intersections in order to avoid network gridlock. This study is part of a greater effort to develop a framework to use Connected Vehicle data for advanced signal control strategies.

REFERENCES

[1] M. Lighthill and G. Whitham, "On kinematic waves. II. A theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pp. 317–345, 1955.

[2] P. Richards, "Shock waves on the highway," *Operations research*, vol. 4, no. 1, pp. 42–51, 1956.

[3] G. Comert and M. Cetin, "Queue length estimation from probe vehicle location and the impacts of sample size," *European Journal of Operational Research*, vol. 197, no. 1, pp. 196–202, 2009.

[4] R. Venkatanarayana, H. Park, B. Smith, C. Skerrit, Jr., and N. Ruhter, "Application of IntelliDrive^{SM} to address oversaturated conditions on arterials," in *Transportation Research Board 90th Annual Meeting*, 2011.

[5] P. Hao and X. Ban, "Vehicle Queue Location Estimation for Signalized Intersections Using Sample Travel Times from Mobile Sensors," in *Transportation Research Board 90th Annual Meeting*, 2011.

[6] Y. Cheng, X. Qin, J. Jin, B. Ran, and J. Anderson, "Cycle by cycle queue length estimation for signalized intersections using sampled trajectory data," in *Transportation Research Board 90th Annual Meeting*, 2011.

[7] K. Srinivasan and P. Jovanis, "Determination of number of probe vehicles required for reliable travel time measurement in urban network," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1537, pp. 15–22, 1996.

[8] M. Vasudevan, "VII Data Characteristics for Arterial Applications," in *VII/CICAS Workshop, TRB Signal Systems Committee Mid-Year Meeting*, 2008.

[9] M. Li, Z. Zou, F. Bu, and W. Zhang, "Application of vehicle-infrastructure integration (VII) data on real-time arterial performance measurements," in *Transportation Research Board 87th Annual Meeting*, 2008.

[10] M. Cetin, G. List, and Y. Zhou, "Factors Affecting the Minimum Number of Probes Required for Reliable Travel Time Estimation," *Transportation Research Record*, vol. 1917, pp. 37–44, 2005.

[11] V. Alexiadis, J. Colyar, J. Halkias, R. Hranac, and G. McHale, "The Next Generation Simulation Program," *ITE Journal*, vol. 74, pp. 22–27, 2004.

[12] L. Edie, "Traffic Stream Measurements and Definitions," *Proceedings of the 2nd International Symposium on the Theory of Traffic Flow*, pp. 139–154, 1963.