Paper No. 12-1493

Estimation of Arterial Measures of Effectiveness with Connected Vehicle Data

Juan Argote (corresponding author)
Department of Civil and Environmental Engineering
Institute of Transportation Studies
109 McLaughlin Hall, University of California, Berkeley, CA 94720
Phone: +1(510) 857-4547, Email: juan.argote@berkeley.edu

Eleni Christofa
Department of Civil and Environmental Engineering
Institute of Transportation Studies
109 McLaughlin Hall, University of California, Berkeley, CA 94720
Phone: +1(510) 604-0080, Email: christofa@berkeley.edu

Yiguang Xuan
Department of Civil and Environmental Engineering
Institute of Transportation Studies
109 McLaughlin Hall, University of California, Berkeley, CA 94720
Phone: +1(614) 209-1430, Email: xuanyg@berkeley.edu

and
Alexander Skabardonis, Ph.D.
Department of Civil and Environmental Engineering
Institute of Transportation Studies
109 McLaughlin Hall, University of California, Berkeley, CA 94720
Phone: +1(510) 642-9166, Fax: +1(510) 642-1246, Email: skabardonis@ce.berkeley.edu

91st Annual Meeting
Transportation Research Board
Washington, D.C.
January 22–26, 2012

Word Count: 4,314
Figures/Tables: 12
Total: 7,314

Revised November 8th, 2011

ABSTRACT

The Connected Vehicle technology is a promising mobile data source for improving real-time traffic conditions monitoring and control. This paper presents estimation methods for a variety of measures of effectiveness both at the arterial and intersection level (e.g., average speed, acceleration noise, queue length). These performance measures are essential for determining traffic conditions and improving signal control strategies in real-time. The estimation methods are tested with two datasets comprising of various traffic conditions: undersaturated and oversaturated and the minimum penetration rate requirements for accurate MOE estimates are determined. These reveal that for oversaturated conditions the required penetration rates are smaller than those for undersaturated. Finally, an application of the queue length estimation as a trigger for alternative real-time control strategies is presented. The developed method determines the last observed vehicle distance threshold that activates the alternative signal plan as a function of the Connected Vehicle penetration rate.

INTRODUCTION

The Connected Vehicle (CV) technology is a promising mobile data source that can provide real-time information necessary for evaluating traffic conditions on a network. Such real-time information can be used to design and evaluate more efficient control strategies for signalized networks. However, the ability to use such data depends on the penetration rate of equipped vehicles. It is therefore critical to develop a comprehensive methodology that uses CV data under different penetration rates to estimate Measures of Effectiveness (MOEs). In addition, there is a need to estimate minimum penetration rate requirements that allow accurate estimates under different traffic conditions. This is particularly useful to determine which MOEs can be used for a specific real-time application under different market penetration of the traffic conditions and the CV technology.

Several studies have been conducted in estimating MOEs with the help of CV or probe data. The majority of these studies have focused on the estimation of queue length (*1, 2, 3, 4*), 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 (*1*) who studied the distribution of average queue lengths assuming known penetration rate, while Venkatanarayana et al. (*2*) estimated queue lengths by using the position of the last IntelliDrive^{SM} equipped vehicle in queue (i.e., maximum likelihood estimation). Hao and Ban (*3*) and Cheng et al. (*4*) used kinematic wave theory to estimate queue lengths. However, only Comert and Cetin (*1*) considered the impact of the penetration rate of the equipped vehicles on the estimation accuracy.

However, none of the studies provides a comprehensive description of the estimation process for a wide range of MOEs and traffic conditions. In addition, the penetration rate thresholds that ensure accurate MOEs estimates for real-time traffic control applications have not been identified in the literature. In this paper we explicitly address these issues. We developed and tested algorithms for estimating MOEs from CV trajectory data, and identified penetration rate requirements for accurate MOE estimates in both undersaturated and oversaturated traffic conditions.

The rest of the paper is organized as follows. First, the two datasets used in the analysis are briefly described. Next, the methods for estimating arterial MOEs are presented and the minimum penetration rate requirements for accurate estimates are identified. Estimation of the main intersection level MOE, i.e. queue length, follows and minimum penetration rates for accurate estimates are also identified. Next, we present a method for determining distance from the intersection thresholds for the last observed vehicle, which can be used to identify queue spillovers, is investigated. The last section summarizes the study findings and outlines future research.

TEST SITES

The impact of the CV technology penetration rate on the accuracy of the MOEs has been estimated with the use of actual vehicle trajectories obtained from the Next Generation Simulation (NGSIM) program (*5*) and simulated trajectories from El Camino Real, a major arterial in the San Francisco Bay Area. The NGSIM dataset was used to test the proposed estimation procedures for undersaturated traffic conditions, the conditions where the queues are completely dissipated at within each signal cycle. Since no real data was available for oversaturated traffic conditions, when not all the demand can be served within the cycle and residual queues form, the El Camino simulated data was used. A more detailed description of the two datasets follows.

NGSIM Data: Peachtree Street, Atlanta, GA

NGSIM data collected at Peachtree Street, Atlanta, GA for four signalized intersections have been chosen for this study. The NGSIM dataset 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 and present only undersaturated conditions. Since the MOEs estimated in this paper describe the traffic conditions for a given direction, only the trajectories of vehicles traveling northbound of the main arterial were considered. This resulted in a total of 228 vehicle trajectories processed.

Simulated Data: El Camino Real, San Francisco Bay Area, CA

An existing microsimulation model of El Camino Real in VISSIM was used to obtain data for oversaturated traffic conditions. The selected arterial section includes five intersections starting at Cambridge and ending at Page Mill for the southbound direction. The simulation included 30 minutes of warm-up period so that the arterial could be loaded gradually and the next 15 minutes were used in the analysis. The trajectories were exported with the same fields and resolution as the NGSIM dataset. A total of 1,332 vehicle trajectories were processed.

ARTERIAL MEASURES OF EFFECTIVENESS

This section presents the estimation of arterial-based MOEs for different market penetration rate values, p, for the CV equipped vehicles. For each of the p values considered, 10,000 samples were obtained where each vehicle trajectory was sampled following a Bernoulli trial with probability p. Box plots were used to reveal the arterial MOE estimate variation and identify the minimum p required to accurately estimate a given MOE for both undersaturated and oversaturated conditions. On each box, the central mark represents the median of the estimates, while the edges of the box show the 25th and 75th percentiles (see for example Figure 1). These plots also include a whisker range that extends 1.5 times the distance between the 25th and 75th percentiles at both extremes of the box. A penetration rate was considered acceptable if the whisker range, which corresponds to ± 2.7 the standard deviation of the estimate, laid within a 10% of the average value for the ground truth. If the estimation errors are assumed to be normally distributed, this criteria ensures that the MOE estimate would lay within a 10% of the ground truth with a probability of 0.9965.

Average Speed

The average speed for all lanes in the observed direction for a single sample was obtained by using Edie's generalized average speed definition (*6*), including all the vehicles in the sample:

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

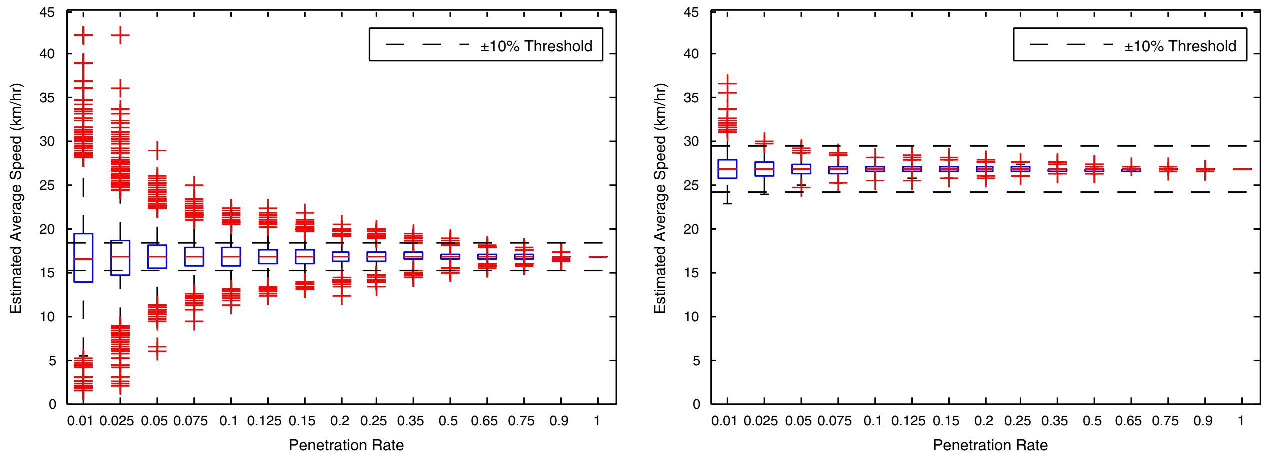
Where:

S is the total number of vehicle trajectories sampled,

l_i is the total distance traveled by vehicle i,

t_i is the total time spent by that vehicle to traverse l_i.

Figure 1 shows the boxplot of the average speeds for different penetration rates for both the NGSIM and the simulated data. Based on the threshold of $\pm 2.7\sigma$ defined above, Figure 1 reveals that penetration rates higher than 35% are necessary to provide accurate speed estimates for an arterial for undersaturated traffic conditions and higher than 5% for ovesaturated traffic condtions.



(a) Undersaturated Traffic Conditions

(b) Oversaturated Traffic Conditions

FIGURE 1 : Box Plots of Average Speed Estimates vs Penetration Rate.

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 was obtained by the following expression:

$$\bar{D} = \frac{1}{S}\sum_{i=1}^{S}\frac{1}{l_i}\left(t_i - \frac{l_i}{v_f}\right). \tag{2}$$

Where:

$\bar{D}$ is the average delay for a single sample of trajectories,

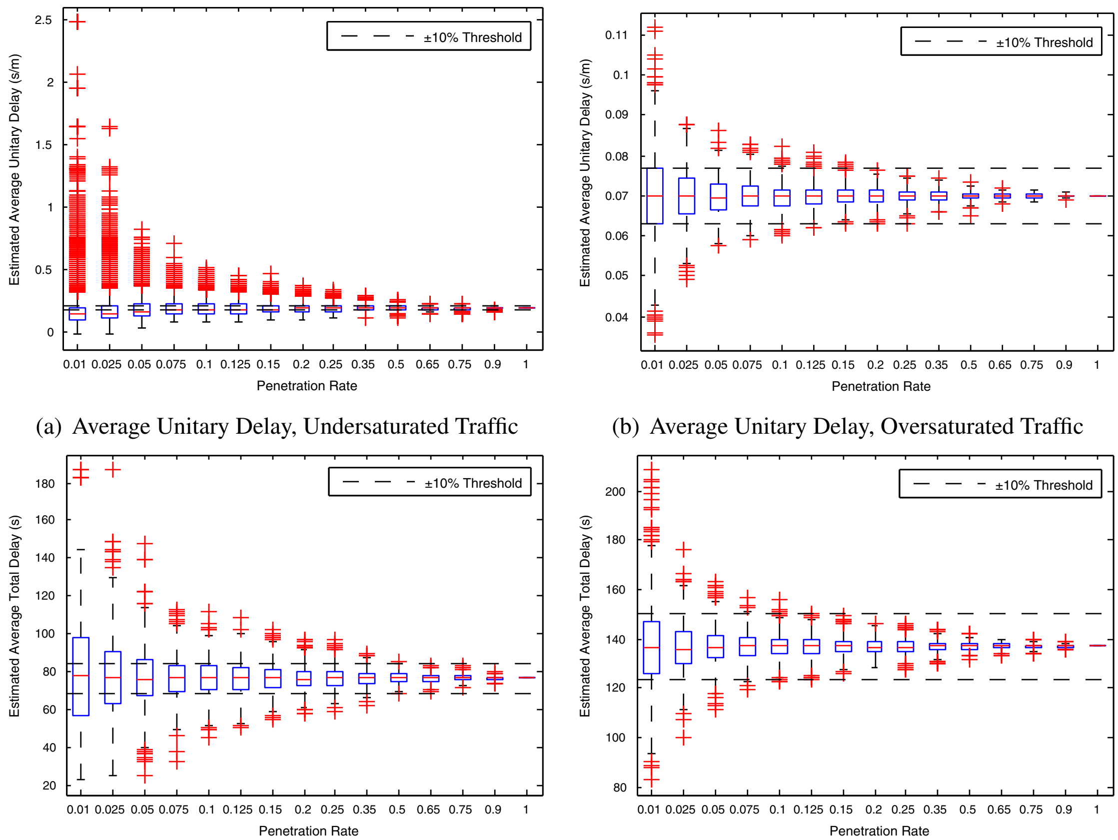
v_f is the free flow speed on the arterial corridor.

The free flow speed used for Peachtree street and El Camino Real 40 km/hr (25 mph) and 64 km/hr (40 mph) respectively. This MOE can also be interpreted as the difference between the sampled average pace and the free flow pace. Again, the box plot in Figure 2(a) 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 75% would be necessary for accurate estimates in undersaturated conditions. On the other hand, Figure 2(b) shows that only a 15% penetration rate would be required in oversaturated conditions.

Alternatively, we can simply compute the average total delay, without considering a distance normalization. In that case the algorithm formulation would be:

$$\bar{D} = \frac{1}{S}\sum_{i=1}^{S}\left(t_i - \frac{l_i}{v_f}\right). \tag{3}$$

The results presented in Figure 2(c) and Figure 2(d) show that in this case penetration rates higher than 50% would be necessary for accurate estimates in undersaturated conditions and 10% in oversaturated conditions.



(a) Average Unitary Delay, Undersaturated Traffic

(b) Average Unitary Delay, Oversaturated Traffic

(c) Average Total Delay, Undersaturated Traffic

(d) Average Total Delay, Oversaturated Traffic

FIGURE 2 : Box Plots of Delay Estimates vs Penetration Rate.

Average Number of Stops

The average number of stops was obtained as the mean number of stops of all vehicles within the sample that were traveling through the arterial on the direction under consideration. A stopped vehicle was defined as a vehicle with speed less than 5 km/h (3.1 mph). There is a high variability in the estimates for the average number of stops for low p. In this case, p values higher than 50% and 20% would be necessary to obtain results that would lie close to the ground truth value for the undersaturated and oversaturated case respectively.

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 was estimated as the mean of the vehicle accelerations' standard deviation (7). The results show that penetration rates on the order of 10% and above should be sufficient to obtain reliable standard deviations for the acceleration for undersaturated conditions but only 1% is necessary for oversaturated conditions.

INTERSECTION MEASURE OF EFFECTIVENESS: QUEUE LENGTH

Queue length is the most used MOE for development and evaluation of signal control strategies at an intersection. This section presents the methodology for identifying stopped vehicles at an approach and estimating the queue length from the available information. Three methods are tested for estimating queue length and the most accurate one is identified.

Discretization of Time and Space

First, the time-space of interest is discretized in different parallelogram regions accounting for the different links in the arterial l=$\{1, 2, \ldots\}$, the different signal cycle times k=$\{1, 2, \ldots\}$, and the backward wave speed of the stopped and accelerating vehicles (Figure 3). This discretization allows us to identify several cells noted by the pairs (k,l) with their corresponding queues. Such unconventional discretization of time and space allows to capture the cases where queue propagates so long that cars have to still decelerate and stop even after the onset of the green time for the next cycle. This is particularly useful for oversaturated traffic conditions. The backward wave speed was estimated from the El Camino Real dataset as 19.44 km/h (12.08 mph).

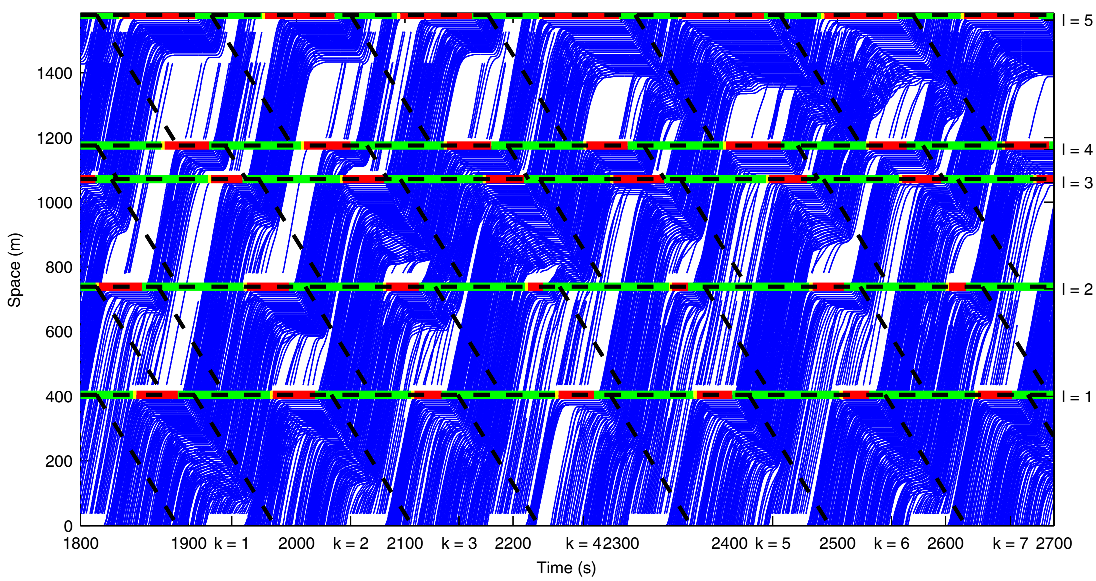


FIGURE 3 : Discretization of Space and Time.

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 so that the positions of the 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:

$$t_i^d = min\{t|v_i(t) > v_c \;\&\; v_i(t+\Delta T) \leq v_c \;\&\; (t, y_i(t)) \in \text{cell}(k,l)\} \tag{4}$$

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 (3 mph)), 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 when it starts decelerating at a lower speed than the specified threshold. For this reason, it is necessary to consider the minimum of the times (or the earliest deceleration point) that satisfies these conditions, so that non-desirable deceleration points are not taken into account.

All the times that satisfy (4), the set $\{t_i^d\}$, can now be rearranged defining a new set of times $U = \{t_{(i)}^d\}$ so that $y_{(1)}(t_{(1)}^d) \geq y_{(2)}(t_{(2)}^d) \geq \ldots \geq y_{(s)}(t_{(s)}^d)$. Note that all the times included in this unfiltered set

U do not necessarily correspond to deceleration points within the queue. As shown in Figure 4, 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 need to impose additional constraints to eliminate these undesired points.

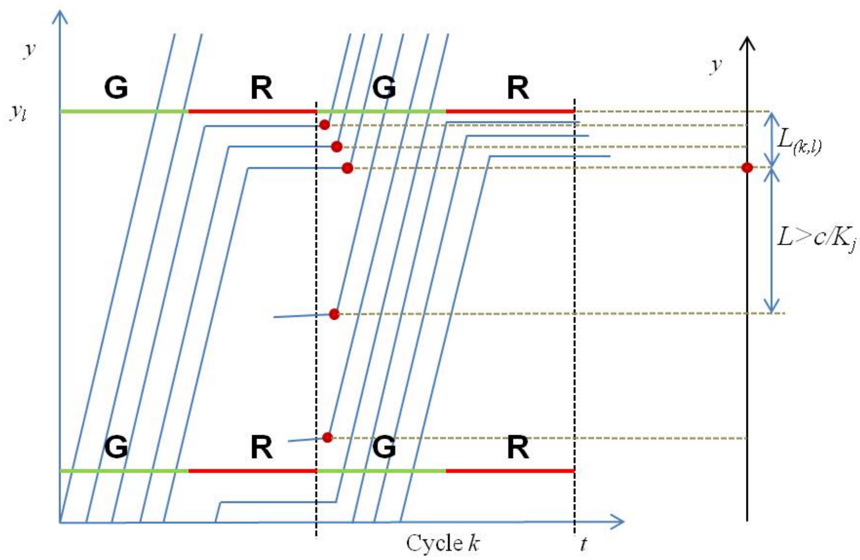


FIGURE 4 : Identification of Deceleration and Acceleration Points within the Queue.

Figure 4 indicates that the undesired deceleration points are far from the deceleration points within the queue. Thus, the distance between such points could be used as a filtering criterion. If vehicles within the same queue are sampled independently with probability p, given by the CV market penetration rate, 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 study, which corresponds to $\ln(0.1)/ln(1-p)$). 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:

$$R = \left\{ (i) \middle| d_{(i)} \triangleq |y_{(i)}(t^d_{(i)}) - y_{(i+1)}(t^d_{(i+1)})| > \frac{c}{K_j} = \max\left\{ \frac{\ln(\varepsilon)/\ln(1-p)}{N_L K_j}, 1/K_j \right\} \right\} \tag{5}$$

After applying (5) to trajectories $(i) \in \{(1), \ldots, (s-1)\}$, the set of vehicle indices to be considered in the queue length estimation is given by:

$$A = \begin{cases} \{(1)\} & \text{if } s = 1 \\ \{(1), \ldots, (s)\} & \text{if } R = \{\emptyset\} \\ \{(1), \ldots, \min\{R\}\} & \text{otherwise} \end{cases} \tag{6}$$

Finally, F defines the set of all deceleration points that belong to vehicles within the queue of interest.

$$F = \{t^d_{(i)} | (i) \in A\} \tag{7}$$

Note that F is a subset of U, and if two deceleration points are too far apart, we would suspect that the most upstream one is not actually in the queue. A similar method can be applied to identify the acceleration points, which are denoted similarly as $t^a_{(i)}$ and are always paired with deceleration points. Figure 5 shows the deceleration (red asterisks) and acceleration (green asterisks) points at intersection 2 of Peachtree Street after filtering.

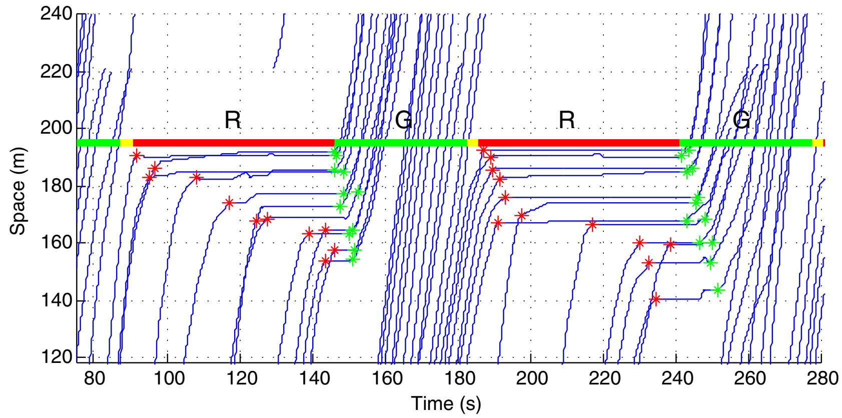


FIGURE 5 : Filtered Deceleration and Acceleration Points.

Queue Length Estimation Methods

Queue length can be easily estimated if the penetration rate of CVs is 100%. But with only a fraction of 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.

Maximum Likelihood (ML) Queue Length Estimation

Assuming that the position of the vehicles in queue can be characterized by a discrete uniform distribution, we can use Maximum Likelihood (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, i.e., the maximum order statistic for the relative position of the vehicles within the queue with respect to the intersection (Figure 6). From equation (6), if we define the number of elements in A (or its cardinality) to be n_A we have:

$$\hat{L}_{ML} = |y^l - y_{(n_A)}(t^d_{(n_A)})| \tag{8}$$

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 straightforward. If this algorithm is applied assuming 100% penetration rate, the estimate obtained gives the ground truth queue length.

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 tends to underestimate the actual queue length. For low penetration rates it is very unlikely that the sampled vehicles will include the last vehicle in the queue and therefore the queue length will be underestimated.

Method of Moments (MM) 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:

$$\hat{L}_{MM} = 2\left|y^l - \frac{1}{n_A}\sum_{i=1}^{n_A} y_{(i)}(t^d_{(i)})\right| \tag{9}$$

This method is theoretically unbiased. 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 estimates than the previous one.

Queue Length Estimation based on Kinematic Wave Theory (KWT)

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). Figure 6 shows how the shockwaves (the threshold between the free flowing and queued states) predicted by KWT can be estimated. The intersection point of the two shockwaves determines the maximum possible queue length, $\hat{L}_{KW}$:

$$\hat{L}_{KW} = \left|y^l - \frac{\frac{y_{(1)}(t^d_{(1)})}{w^d} - \frac{y_{(1)}(t^a_{(1)})}{w^a} + t^a_{(1)} - t^d_{(1)}}{\frac{1}{w^d} - \frac{1}{w^a}}\right| \tag{10}$$

where:

$$\frac{1}{w^d} = \frac{t^d_{(n_A)} - t^d_{(1)}}{y_{(n_A)}(t^d_{(n_A)}) - y_{(1)}(t^d_{(1)})} \tag{11}$$

$$\frac{1}{w^a} = \frac{t^a_{(n_A)} - t^a_{(1)}}{y_{(n_A)}(t^a_{(n_A)}) - y_{(1)}(t^a_{(1)})} \tag{12}$$

For undersaturated traffic conditions the beginning of the first shockwave does not necessarily match the beginning of the red phase in the signal cycle. So it is necessary to sample at least two vehicle trajectories in at least one of the travel lanes 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:

$$P(s_L \geq 2|N_L = m) = 1 - P(s_L < 2|N_L = m) = 1 - \sum_{j=0}^{1}\binom{m}{j} p^j (1-p)^{m-j} \tag{13}$$

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:

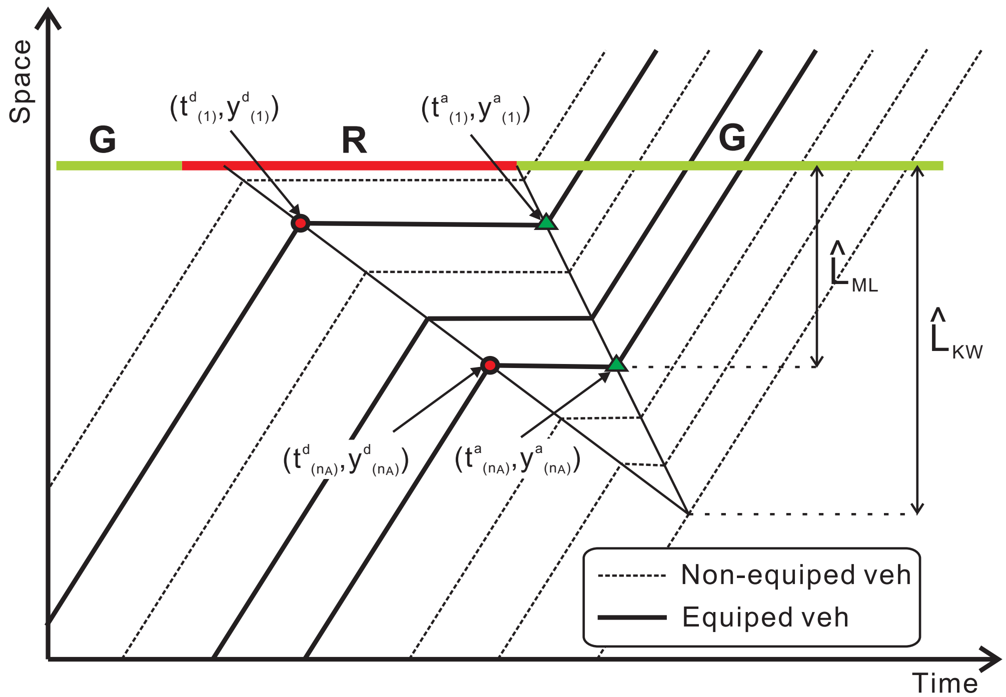


FIGURE 6 : ML and KWT Queue Length Estimation.

$$P(s_L \geq 2) = 1 - \sum_{m=N_{min}}^{N_{max}} P(s_L < 2 | N_L = m) P(N_L = m) =$$
$$1 - \frac{1}{N_{max} - N_{min} + 1} \sum_{m=N_{min}}^{N_{max}} \left(\sum_{j=0}^{1} \binom{m}{j} p^j (1-p)^{m-j} \right) \tag{14}$$

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 Figure 7 provides the probability $P(s_L \geq 2)$ for different penetration rates. It can be seen 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.

Figure 8 compares the average absolute relative errors obtained by the three queue length estimation methods. Average absolute relative errors are used here instead of box plots because all three estimators are biased and the results from different intersections are combined. Unfortunately, the undersaturated conditions in the NGSIM dataset 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. For oversaturated conditions, the KWT estimator still has the highest absolute error, which however is much smaller than the one for undersaturated conditions.

The maximum likelihood estimation method 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% for undersaturated traffic conditions. However, for oversaturated conditions, a penetration rate of only 20% is sufficient to provide an absolute relative error lower than 10%.

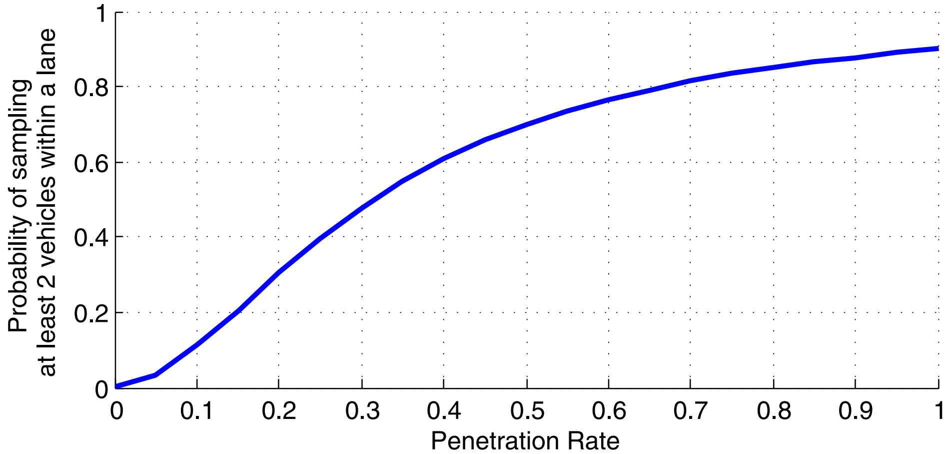
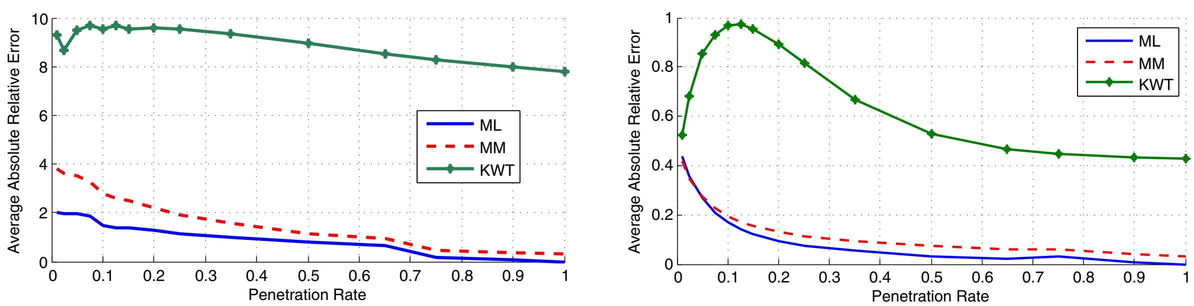


FIGURE 7 : Probability of Sampling at Least Two Vehicles for a Given Penetration Rate, with N_{min} = 1 and N_{max} = 10).



(a) Undersaturated Traffic Conditions

(b) Oversaturated Traffic Conditions

FIGURE 8 : Average Absolute Relative Error vs Penetration Rate for the Three Queue Length Estimation Methods.

SPILLBACK DETECTION

In this section, we present an algorithm for identifying back of the queue thresholds to be used as trigger points for real-time signal control strategies. When queues spill back into an intersection, vehicles cannot enter the intersection during their green time. The green time is then wasted, also known as "de facto red time" (*8*) . To avoid this situation, it is common practice to switch to an alternative signal timing plan (e.g., allocate more green time to the crossing traffic). The alternative plan is triggered when the distance from the end of queue to the downstream intersection is longer than a threshold value L_{th}, e.g., 80% of the link length (*2*). L_{th}, is site specific and is generally defined as a function of the link length.

However, if not all vehicles are equipped with CV technology, the last CV-equipped vehicle in the queue may not be in reality the last vehicle in the queue. We assume that L_{th} is given, and develop a method to estimate the distance between the last CV-equipped vehicle and the last vehicle in queue, which we term gap length, X, with respect to different penetration rates p. Figure 9 provides a graphic illustration of the threshold value and the gap length.

We define K_j to be the jam density (in veh/km) on the link, typically corresponding to the spacing of stopped vehicles (e.g. 25 ft/veh). Note that the jam density is defined over all lanes within the link. We seek the minimal gap length $X^* = \min_x \{X | \Pr(X \leq x | L_{th}) \geq 1 - \alpha\}$, given a false negative rate α. Note

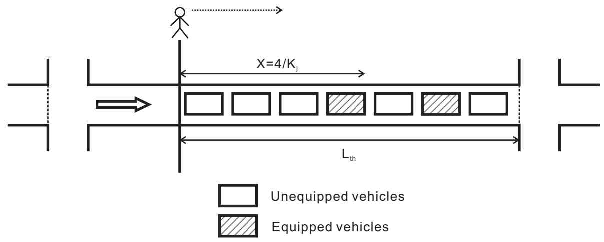


FIGURE 9 : Graphic Illustration of the Threshold Value and Gap Length.

that the term negative refers to the situation where a queue does not spill over the threshold, while positive corresponds to the opposite situation. Therefore, if the end of queue based on CV-equipped vehicles is within X^* of the threshold, we predict that queue spills over. Given L_{th}, the distribution of X follows the probability mass function below:

$$\Pr(X = x|N) = \begin{cases} p(1-p)^{K_j x} & \text{when } 0 \le x < L_{th} \\ (1-p)^{K_j L_{th}} & \text{when } x = L_{th} \end{cases} \tag{15}$$

Note that this distribution is very similar to a geometric distribution. The second situation ($x = L_{th}$) occurs when no vehicle is sampled. Note that when the penetration rate is so low that $(1-p)^{(K_j L_{th})} > \alpha$, the minimum gap length equals the threshold value ($X^* = L_{th}$). This means that for these penetration rates, if any CV-equipped vehicle is observed to join the queue, the alternative signal plan should be triggered. If this is not the case, the expression of X^* can be derived as follows:

$$\Pr(X \le x|L_{th}) = 1-(1-p)^{(K_j x)} \ge 1-\alpha \tag{16}$$

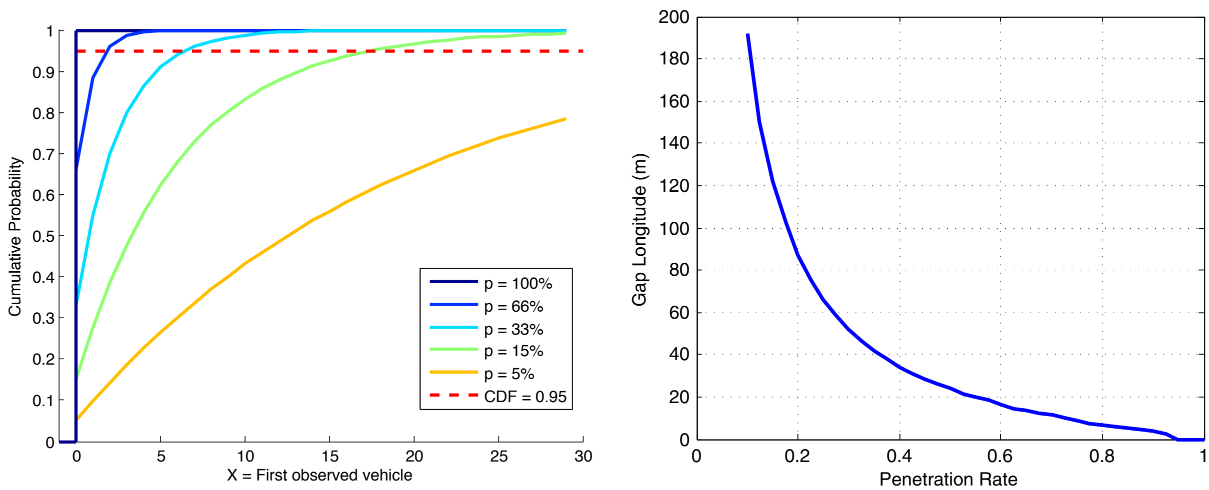
$$K_j x \ge \frac{\ln(\alpha)}{\ln(1-p)} \tag{17}$$

$$X^* = \min\{x\} = \frac{1}{K_j}\left\lceil \frac{\ln(\alpha)}{(\ln(1-p)}\right\rceil \tag{18}$$

where $\lceil x \rceil$ produces the smallest integer that is greater than or equal to x. Figure 10(a) shows the cumulative distribution function of X given N for different penetration rates, as derived in equation (16). Figure 10(b) shows the minimal gap length X^* versus penetration rate, as derived in equation (18). In the figure, we assume a jam density of K_j = 1/7 vehicle/meter, and a threshold value of $L_{th} = 30/K_j$ (which corresponds to a link length of about 265 meters).

This method was applied to the El Camino Real oversaturated dataset. Two different intersections and five different cycles were considered, where half of the cases were positive. As before, we define positive to be the situation that the end of queue spills beyond the threshold, and negative the opposite situation. We then randomly sample vehicles with probability p to be the CV-equipped vehicles, and determine the deceleration point of the last CV-equipped vehicle in queue. If the location of the deceleration point is within X^* from the threshold line, we say the queue spills beyond the threshold, and vice versa. Figure 11 shows the false positive rate, false negative rate, and correct rate versus different penetration rates. Because X^* is calculated with a false negative rate of 0.05, the false negative rate remains at that level, and eventually drops for very high penetration rates. We also see that there is a sharp drop in false positive rate and a sharp surge in success rate as penetration rate increases, but both of them level off after penetration rate reaches 0.15.

Note that the correct rate depends on the proportion of positive cases. If an intersection is frequently congested, the proportion of positive cases is high, and the correct rate is heavily influenced by the false



(a) (b)

FIGURE 10 : Estimation of the Minimal Gap Length. (a) The CDF of X Given N for Different Penetration Rates. (b) Minimal Gap Length X^* vs Penetration Rate.

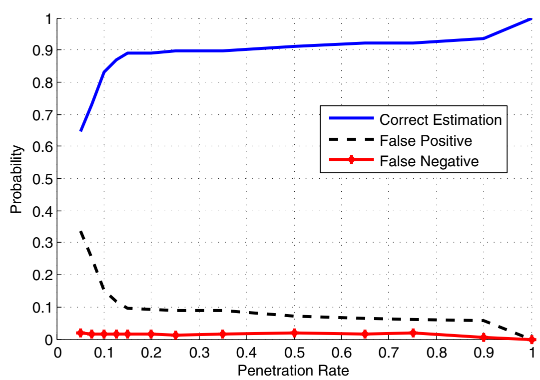


FIGURE 11 : False Positive Rate, False Negative Rate, and Correct Rate vs Penetration Rate.

negative rate. If an intersection is not very busy, the proportion of positive cases is low, and the correct rate is heavily influenced by the false positive rate. For demonstration purposes, we fix our choice of the false positive rate α to be 0.05. However, the analysis can be easily extended to maximize the correct rate (true positive plus true negative). In this situation, the value of α should depend on the proportion of positive and negative cases.

DISCUSSION

We developed methods for estimating common arterial MOEs, and queue lengths at signalized intersections for different penetration rates of the CV technology. Moreover, we identified minimum penetration rates that allow accurate estimates for each performance measure.

Testing of the proposed estimation methods in two different datasets concludes that in general lower penetration of CV is needed in order to obtain accurate estimate when in oversaturated traffic conditions in comparison with undersaturated ones (Table 1). This result is logical since in oversaturated conditions we have a greater amount of vehicles, which increases the number of CV sampled. This implies that for congested networks, in which real-time control strategies are more critical, accurate estimation of MOEs can be done with lower penetration rates.

TABLE 1 : Minimum Penetration Rate for MOE estimates within 10% of the ground truth

MOE	Undersaturated Required p	Oversaturated Required p
Average Speed (km/hr)	35%	5%
Average Total Delay (s/m)	50%	10%
Average Number of Stops	50%	20%
Acceleration Noise (m/s^2)	10%	1%

Three queue length estimation methods were tested, the Maximum Likelihood, the Method of Moments, and one based on Kinematic Wave Theory. The results on both test sites indicate that 80% of the vehicles need to be equipped to obtain accurate estimates for undersaturated conditions but only 10% for oversaturated conditions. This is promising in the sense that accurate estimates of queue length can be obtained for low penetration rates and as a result, improved control strategies can be designed for oversaturated conditions even at the first stages of the CV implementation.

We also developed a threshold strategy to avoid queue spillovers in a given intersection based on the location of the last CV. This concept has been tested in the oversaturated dataset obtaining over 90% correct estimates for penetration rates as small as 25%. This shows the great potential that this technology could have on real-time signal control strategies. Future steps include fusing CV with existing loop detector data to obtain accurate estimates for low penetration of CV vehicles. The estimates can then be used to develop advanced signal control strategies for signalized networks both at the microscopic and the macroscopic level.

ACKNOWLEDGMENTS

This work was partly sponsored by the Federal Highway Administration (FHWA) of the United States Department of Transportation (US DOT) as part of the Exploratory Research Program (EAR) - Advanced Signal Control Algorithms project. The authors would like to thank Dr. David Yang FHWA contract monitor and Professor Carlos Daganzo of UC Berkeley for their comments and suggestions.

References

[1] Comert, G. 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, 2009, pp. 196–202.

[2] Venkatanarayana, R., 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.

[3] Hao, P. 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.

[4] Cheng, Y., 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.

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

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

[7] Jones, T. and R. Potts, The Measurement of Acceleration Noise-A Traffic Parameter. *Operations Research*, Vol. 10, No. 6, 1962, pp. 745–763.

[8] Abu-Lebdeh, G. and R. Benekohal, Development of traffic control and queue management procedures for oversaturated arterials. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1603, 1997, pp. 119–127.