STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

Lock Data on Form

1. REPORT NUMBER
CA16-2872

2. GOVERNMENT ASSOCIATION NUMBER

3. RECIPIENT'S CATALOG NUMBER

4. TITLE AND SUBTITLE
From LOS to VMT, VHT, and Beyond through Data Fusion: Application to Integrated Corridor Management

5. REPORT DATE
April 30, 2016

6. PERFORMING ORGANIZATION CODE

7. AUTHOR
Qijian Gan, Ph.D., Gabriel Gomes, Ph.D., Alexandre Bayen, Ph.D.

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
University of California Center for Economic Competitiveness in Transportation
University of California, Berkeley
2150 Shattuck Avenue, Suite #300
Berkeley, CA 94 704-5940

10. WORK UNIT NUMBER

11. CONTRACT OR GRANT NUMBER
65A0529 TO 022

12. SPONSORING AGENCY AND ADDRESS
California Department of Transportation
Division of Research, Innovation, and Systems Information
1120 N Street, MS-38
Sacramento, CA 95814

13. TYPE OF REPORT AND PERIOD COVERED
Final Report

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT
This work supports an implementation of some of the objectives of SB-743 for operations and planning. First, the project will investigate the current gaps in the traffic data sources commonly used in the computation of standard metrics such as Level of Service (LOS). It will investigate the data requirements necessary for the computation of metrics suggested by SB-743, in particular Vehicle Miles Traveled (VMT) and Vehicle Hours Traveled (VHT). Second, it will propose two algorithms to integrate new sources of data in the computations of these metrics. The first algorithm aims to improve the computations of VHT in arterial networks by integrating available GPS data from existing infrastructure. The second algorithm will then investigate the use of cell tower phone data, i.e., call data records (CDR), to solve the route inference problem, through compressed sensing and convex optimization based least square optimization. The project will suggest how to use the enhanced information to produce more reliable VMT and VHT estimates. The contributions will be illustrated by leveraging the Connected Corridors I-210 Pilot in the San Gabriel Valley, Los Angeles, California. The research result may present significant improvement over current approaches, and can support the aspects of SB-743 that require this data.

17. KEY WORDS
traffic, data, Level of Service,Vehicle Miles Traveled,data records , cell tower,SB-743

18. DISTRIBUTION STATEMENT
No Restrictions

19. SECURITY CLASSIFICATION *(of this report)*
Unclassified

20. NUMBER OF PAGES
43

21. COST OF REPORT CHARGED

Disclaimer

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the STATE OF CALIFORNIA or the FEDERAL HIGHWAY ADMINISTRATION. This report does not constitute a standard, specification, or regulation

University of California
Center for Economic
Competitiveness in
Transportation

UC CONNECT

From LOS to VMT, VHT and Beyond Through Data Fusion: Application to Integrate Corridor Management

Final Report UCCONNECT - 2016 - TO 022 - 65A0529

Alexandre Bayen; Chancellor's Professor; University of California, Berkeley
Qijian Gan; Ph.D; University of California, Irvine
Gabriel Gomes; Ph.D; University of California, Berkeley

Sponsored by

Caltrans

2016

Berkeley UC Irvine UCLA UCR UCSB CAL POLY POMONA

From LOS to VMT, VHT, and Beyond through Data Fusion:

Application to Integrated Corridor Management

April 30, 2016

A report to the

University of California Center on Economic Competitiveness in Transportation

Qijian Gan, Ph.D. (UC Irvine)

Gabriel Gomes, Ph.D. (UC Berkeley)

Alexandre Bayen, Ph.D. (UC Berkeley)

Partners for Advanced Transportation Technology (PATH)

University of California at Berkeley

Berkeley, CA 94720

Executive Summary

Traffic performance metrics such as delay and Level Of Service (LOS), which are well documented in the *Highway Capacity Manual (HCM)*, have been widely used by most of the transportation consulting companies, public agencies, and etc. For arterial delay analysis, prevailing commercial tools like *Synchro* have adopted the method proposed by the HCM, which is rooted in the Webster's delay calculation proposed more than 50 years ago. The LOS is obtained using a lookup table that assigns a certain grade (from A to F) to the estimated delay according to its value. Without knowing detailed vehicles trajectory profiles, this kind of delay calculation method relies on macroscopic queueing theory and assumes certain types of arrival patterns. As mentioned in the *State Bill 743 (SB743)* and the memo entitled *Preliminary Evaluation of Alternative Methods of Transportation Analysis* issued by Governor Brown's Office of Planning and Research on December 30, 2013, current calculation of LOS is difficult and expensive. Particularly, as will be illustrated in Section 2, the state-of-the-practice calculation of delay and LOS for local intersections is very complicated and has the following limitations.

i. Invalid assumptions. First, the delay calculation relies on macroscopic queueing theory and assumes certain types of vehicle arrival patterns. However, vehicle's arrival patterns are very complicated, which are related to the demand patterns, signal phase settings, signal coordination, and etc. Second, the delay calculation for a certain lane group is independent from others'. However, lane group movements are interacted with each other. Phenomena such as queue spillbacks and left-/right-turn blockages can be observed frequently under heavy traffic conditions or with bad signal settings. Even though improvements and updates have been made to the latest version of HCM (i.e., HCM 2010), such problems mentioned above still have not been fully addressed.

ii. Multi-step procedure. The method provided in the HCM is very complicated in calculating vehicle's average delay. There are specific procedures to account for the following aspects:

a) Geometric design. While calculating the saturation flow-rate for a given lane group, 11 parameters are introduced to take into account the following factors: lane width, heavy vehicles, grade, parking activities, bus stops, area types, lane utilization, left-/right-turn movements, existence of pedestrians and bicycles. The adjustment of these factors is very engineering-oriented and difficult to validate.

b) Signal control. In the HCM, for a given intersection, delay calculation varies among different types of control, e.g., four-way stops, fixed-time control, and actuated control.

Furthermore, LOS is not a practical metric since it only provides a certain grade of the system performance and doesn't have any physical meanings. Delay is also not a good metric for regional planning analysis since the total travel distance and total travel time may increase while minimizing the travel delay. Therefore, alternative performance metrics such as Vehicle Miles Traveled (VMT) and Vehicle Hours Traveled (VHT) are suggested.

In practice, signal design varies from place to place. For a given intersection, if the signal timing plan is traffic responsive, loop detectors are usually installed to detect incoming flow-rates. Currently, loop detector data is the main data source for arterial traffic study. However, it is reported that as of July 2015, 68% of American adults have a smartphone[1]. With advanced sensors equipped in the smartphones, now it is getting easier to collect additional data. Detailed vehicle trajectories can be collected from smartphones equipped with GPS. Vehicles can be re-identified and travel times can be obtained if Bluetooth connection is enabled. Compared to the loop detector data, this additional data source provides more detailed information of traffic. Therefore, it is possible to reconstruct real-world traffic conditions with this additional data source. In the literature, studies have shown that it is possible to reconstruct traffic on freeways with a low penetration rate of probe vehicles, e.g., 2-3%. However, it is still unknown how much data we need for local arterials since traffic patterns are more complicated than those on freeways.

[1] http://www.pewinternet.org/data-trend/mobile/device-ownership/. Last visited: 4-20-2016

The motivation of this research is to explore the benefits of utilizing current available data sources, including loop detector data and probe data, in the following aspects:

i. to improve the current calculation of delay and LOS performance metrics;

ii. to provide estimates of alternative performance metrics such as VHT.

To calculate delay and LOS, traditional methods such as the one proposed by the HCM are model-driven. However, in this research, we propose a data fusion approach that tries to use vehicle counts from loop detectors and travel times from probe vehicles for the estimation. At the meantime, more physically meaningful metrics such as VHT can be calculated. To perform our research, we have to first address the following issues:

i. Ideally, if all vehicle travel times can be retrieved, the calculation of VHT, delay, and LOS would be rather simple. However, in reality, the penetration rate is not 100%, and not all probe vehicles can be re-identified. Therefore, the accuracy of VHT, delay, and LOS estimates highly relies on the percentage of observed probe vehicles. In this study, we use statistical models instead of physical ones while estimating these metrics, and we also investigate how much probe data is needed to guarantee the estimation accuracy.

ii. It is very difficult to collect enough loop detector data from the field. First, not all intersections are equipped with loop detectors. Second, even for those intersections equipped with detectors, not all of them are connected to the central database. That means there exists a portion of detectors working locally and the detected vehicle volumes will be discarded as time elapses. In this study, we use microsimulations to generate a set of synthetic data for our study network. Compared with collecting data from the field, there are advantages of using microsimulations.

    a) The cost of running simulations is much cheaper than collecting data from the field.

    b) To assess the performance of the proposed method, it is possible to generate datasets that contain various traffic conditions since we can adjust the demands, signal timings, signal coordination, and etc. in the study network. However, such datasets are hard to obtain from the field.

For our proposed method, we assume the midblock loop detector data and the travel time data are given. Then VHT, delay and LOS can be estimated as follows:

i. VHT estimation. VHT for a given lane group / approach / intersection, is estimated as the product of the mean of sampled travel times and the observed midblock traffic count.

ii. Delay and LOS calculation. Delay for a given lane group / approach / intersection, is estimated as the difference between the mean of sampled travel times and the free-flow travel time. Then LOS is obtained using the delay-LOS lookup table provide by the HCM.

As a comparison, we also estimate the delay and LOS using the HCM method. Our proposed method is simple, but has the following advantages over the HCM method:

i) It is non-parametric which doesn't rely on any particular models. This is a very important feature since it is very challenging to develop a single model that can work under various traffic conditions at signalized networks.

ii) It doesn't require inputs of geometric design and signal settings. Different from the HCM method, our proposed method is generic that can work under various types of traffic control, e.g., four-way stops, fixed-time control, and actuated control. For the estimation of VHT, the proposed method is valid as long as the midblock detector provides accurate estimates of traffic counts. For the estimation of delay, the proposed method turns out to be the unbiased estimator with minimum variance under a wide range of travel time distributions.

iii) The estimation accuracy is in a direct relation with the availability of vehicle travel times, which will definitely increase in the near future.

As it is difficult to collect traffic counts and travel times from the field, a set of synthetic trajectory data is collected using microsimulations in Aimsun. The study network is an arterial network of Huntington Dr. and Colorado Blvd, which run parallel to the I-210 freeway. Traffic demands as well as the signal settings are tuned in order to generate various traffic congestion profiles. After collecting vehicle's trajectory data, we apply the following method to collect the data required for this study.

i. Demand data. In the delay calculation proposed by the HCM, the input of vehicle volume is actually the vehicle demand. In practice, vehicle demand is very difficult to obtain, especially during heavy traffic conditions. Therefore, in our study, we use the entering flow as a proxy for the demand of a given approach.

ii. Traffic counts. To obtain the traffic counts, we virtually place midblock detectors on the approaches at each intersection.

iii. Vehicle travel times. We assume the penetration rate of probe vehicles is relatively constant over time inside our study network. To obtain a set of sampled travel times, we use the following procedure:

    a) Probe vehicles are first sampled from the whole population in the synthetic dataset.

    b) Vehicle trajectories for these selected probe vehicles are extracted, and then, the set of sampled travel times is obtained.

    This step is conducted for 500 times in order to obtain statistically meaningful estimation results.

For a given intersection, there are three different levels of analyses: lane group level, approach level, and intersection level. Correspondingly, we can obtain three different levels of estimates. For the estimation of approach VHT, delay, and LOS, there are two different ways. The first way is to calculate the estimates of lane group VHT, delay, and LOS, and then aggregate them into the approach level. The second way is to obtain the estimates of approach VHT, delay, and LOS directly using the proposed method. As a comparison, estimates of lane group delay and LOS are also calculated using the HCM method. Then estimates at the approach and intersection levels are obtained by aggregating the ones at the lane group level.

According to our analysis, we have the following findings:

i. The proposed method provides better estimates at the approach level than at the lane group level for all the estimates of VHT, delay and LOS. The reasons may be:

   a) The approach based volume is higher than the lane group based one. With a given penetration rate, the chance to observe a probe vehicle is higher at the approach level than at the lane group level.

   b) For a given approach, the distributions of travel times are similar among lane groups since they interact with each other.

ii. A penetration rate of 7% is enough to obtain reliable and accurate estimates of VHT, delay and LOS. This penetration rate probably can be reached in the near future.

iii. For the estimation of delay and LOS, the proposed method generally outperforms the HCM method. Using the synthetic dataset, we show that the HCM method provides inaccurate estimates of LOS 40% of the time. However, if the penetration rate of probe vehicles is about 7%, the proposed method can improve the accuracy up to 83% at the approach level.

According to our findings, the required penetration rate (7%) for arterial networks is higher than that for freeway networks. In the future, it is necessary to validate the actual penetration rate on local arterials. We may expect such a requirement may not be satisfied given certain periods of time or for minor streets. Therefore, such an analysis can help us better understand what we should do if we want to apply our proposed method to the field.

In this study, we haven't involved the estimation of VMT. There are several reasons listed below:

i) Ideally, if detectors are installed at all approaches of a given intersection and are healthy, the estimation of VMT is rather simple. The appropriate proxy would be the product of the midblock detector counts and the corresponding link length.

ii) However, not all approaches at a given intersection are equipped with loop detectors, and not all loop detectors are healthy all the time. Therefore, there is no enough field data for the estimation of VMT.

iii) There are other data sources, such as Call Data Records (CDRs), available. It is possible to apply optimization schemes to reassign the cellular flows to local arterials to match the observed link counts. However, we need to have a better understanding on this new data type before calculating estimates of VMT for arterial networks.

Table of Contents

Tables and Figures

Tables

Figures

Acknowledgements

This work was fully supported by UCCONNECT and Caltrans. The authors would like to thank Andrew Campbell and Prof. Alexei Pozdnukhov for the data support. The authors would also like to thank Francois Belletti and Matthew Wright for their useful comments.

About the Authors

Qijian Gan received the Engineering Degree in automatic control from the University of Science and Technology of China in 2009, the M.S. and Ph.D. in Civil Engineering from the University of California, Irvine, in 2010 and 2014 respectively. He is now an Assistant Project Scientist in the Institute of Transportation Studies at UC Irvine. His main expertise includes network traffic flow theory, network modeling and simulation, traffic signal control, and data analysis.

Gabriel Gomes is an Assistant Research Engineer with the California PATH program. Dr. Gomes holds a BS in mechanical engineering from the Simon Bolivar University in Venezuela (1996) and a doctorate degree in systems and control theory from the Mechanical Engineering department at U.C. Berkeley (2004). His work at PATH focuses on problems of modeling, control, and simulation of transportation systems.

Alexandre Bayen is the Liao-Cho Professor of Engineering at UC Berkeley. He is a Professor of Electrical Engineering and Computer Science, and Civil and Environmental Engineering. He is currently the Director of the Institute of Transportation Studies (ITS). He is also a Faculty Scientist in Mechanical Engineering, at the Lawrence Berkeley National Laboratory (LBNL). He received the Engineering Degree in applied mathematics from the Ecole Polytechnique, France, in 1998, the M.S. and Ph.D. in aeronautics and astronautics from Stanford University in 1998 and 1999 respectively. He was a Visiting Researcher at NASA Ames Research Center from 2000 to 2003. Between January 2004 and December 2004, he worked as the Research Director of the Autonomous Navigation Laboratory at the Laboratoire de Recherches Balistiques et Aerodynamiques, (Ministere de la Defense, Vernon, France), where he holds the rank of Major. He has been on the faculty at UC Berkeley since 2005. Bayen has authored two books and over 200 articles in peer reviewed journals and conferences. He is the recipient of the Ballhaus Award from Stanford University, 2004, of the CAREER award from the National Science Foundation, 2009 and he is a NASA

Top 10 Innovators on Water Sustainability, 2010. His projects Mobile Century and Mobile Millennium received the 2008 Best of ITS Award for Best Innovative Practice, at the ITS World Congress and a TRANNY Award from the California Transportation Foundation, 2009. Mobile Millennium has been featured more than 200 times in the media, including TV channels and radio stations (CBS, NBC, ABC, CNET, NPR, KGO, the BBC), and in the popular press (Wall Street Journal, Washington Post, LA Times). Bayen is the recipient of the Presidential Early Career Award for Scientists and Engineers (PECASE) award from the White House, 2010. He is also the recipient of the Okawa Research Grant Award, the Ruberti Prize from the IEEE, and the Huber Prize from the ASCE.

Section 1: Introduction

The evaluation of performance metrics for urban networks is an important task for both urban planners and traffic operators. In the planning context, metrics are used to estimate the impacts of proposed long-term projects at the neighborhood or city level. The criteria used in California for evaluating these projects have shifted recently with the passage of State Bill 743 in 2013. This bill seeks to upgrade the process for evaluating transportation projects under the California Environmental Quality Act (CEQA) such that it aligns with the broader goal of reducing environmental impacts, specifically by reducing greenhouse gas emissions. The guidelines that have emerged from this process [1] argue against the use of the Level Of Service (LOS) as a guiding metric, and for alternatives such as Vehicle Miles Traveled (VMT) and Vehicle Hours Traveled (VHT). Among the stated reasons for this is the fact that LOS forecasts tend to favor development in lightly populated areas over urban areas and city centers. Standard LOS evaluations fail to promote mode shifts away from cars and into cleaner alternatives such as bikes, buses, and walking. It is also argued that delay-based metrics in general are not adequate for evaluating projects that involve changes in trip lengths. It is possible, for example, to decrease average delay while increasing travel times, if trips are made longer. Hence, the trend in urban planning is currently towards aggregate metrics such as VHT and VMT, and away from delay-based metrics such as LOS.

The case is somewhat different in the short-term operations context. Here the objective is to use performance metrics to evaluate an intersection or group of intersections in order to improve their signal control settings. In this context, trip lengths are largely fixed - some portion of drivers may be coaxed into taking a different route, but only by increasing the delay on their nominal route. Thus, it can be argued that, at the operations level, the goals of greenhouse gas reduction and delay reduction are aligned.

The aim of this research is to investigate the extent to which the estimation quality of both VHT and LOS can be improved by using travel time samples from onboard devices such as smartphones.

Several technologies now exist that enable the collection of travel times in urban settings. These include toll tag readers, Bluetooth devices, and smartphones. Additionally, travel times can be obtained through vehicle re-identification techniques, such as that of [2]. The availability of probe devices is growing rapidly. It was reported in [3] that as of July 2015, 68% of American adults own a smartphone. The potential for using probe data to infer traffic states has been established. Herrera et al. [4] found that a 2-3% penetration rate of cell phones is sufficient to provide accurate measurements of speeds on a freeway. Patire et al. [5] further confirmed that relatively low penetration rates for GPS-based probes are sufficient to significantly improve the estimation of traffic states on freeways. However few studies have been conducted on the potential uses of probe data on local streets. The existence of signalized intersections on these streets, along with the complexities of urban routing, makes it likely that a higher rate of probes will be needed.

Important contributions to the problem of travel time estimation include those of Hofleitner [6] and Feng [7]. Both of these works are concerned with fitting the parameters of a candidate model -- a mixture Gaussian in the case of [7] and a model derived from traffic theory in [6]. While these serve useful purposes, for example for the route selection problem, the goals of the present effort require only the estimation of the mean of travel times, and thus a more simple non-parametric approach is sought.

This report introduces a method for combining travel time samples with loop detector counts to estimate VHT, delay, and LOS. The method is contrasted with the formulas of the Highway Capacity Manual (HCM) [8], which take lane group volumes, queues, capacities, and signal parameters as inputs. These formulas are provided in Section 2. Section 3 presents the proposed method. The framework for studying the method and comparing it to the HCM formulas is described in Section 4. This framework consists of a micro-simulation model built with the Aimsun software [9]. The model covers 11 intersections along streets running parallel to the eastbound I-210 freeway in Arcadia, California. The analysis of the simulation results is provided in Section 5. In Section 6, we draw our final conclusions with some recommendations.

Section 2: The HCM Method

The method currently used by both urban planners and traffic engineers to calculate delay and LOS is that of the Highway Capacity Manual [8]. This method, illustrated in Figure **1**, requires information about the geometric design, traffic demands (forecasts in the planning case), and signal settings. The details of the calculations are provided below. These are applied for comparison to our proposed method in Section 3.

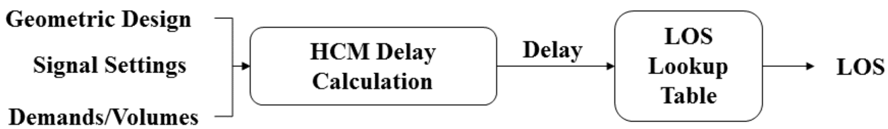


Figure 1. HCM procedure for calculating delay and LOS.

Section 2.1: HCM delay calculations

Conceptually, as illustrated in Figure **2**, the delay experienced by a vehicle at a fixed-time traffic signal is measured between point A at which it first begins to decelerate as it approaches the queue, and point B where it has accelerated to its original speed after exiting the queue [10]. However, in the HCM, the control delay is measured slightly different which only takes into account the delay between point A and the stop line (point B') (See Exhibit 31-5 in [8]).

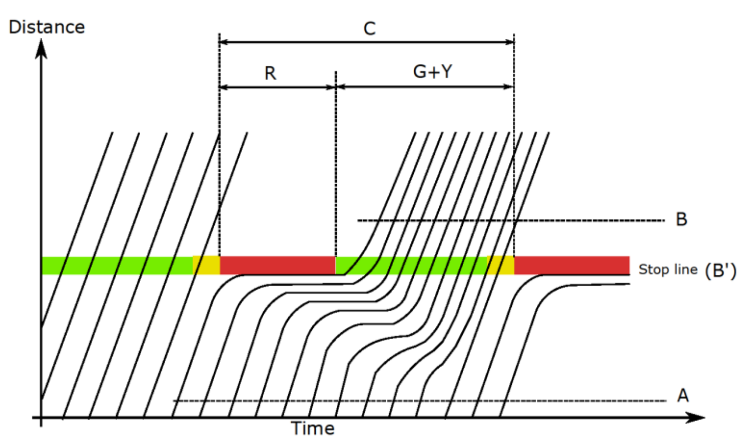


Figure 2. Distance-time diagram for a signalized intersection.

The HCM identifies three components of delay: uniform delay (d_1), incremental delay (d_2), and initial queue delay (d_3). The average delay for the lane group, d_g, is computed as a sum of these three parts:

$$d_g = d_1 + d_2 + d_3. \tag{1}$$

The uniform delay represents the delay computed under an idealized assumption of uniform arrivals. It can be computed using the Incremental Queue Accumulation (IQA) method in [8], which is provided below.

$$d_1 = \frac{0.5 \sum_{i=1}(Q_{i-1} + Q_i)t_{t,i}}{qC} \tag{2}$$

with

$$t_{t,i} = \min\{t_{d,i}, \frac{Q_{i-1}}{w_q}\} \tag{3}$$

where

$t_{t,i}$: duration of trapezoid or triangle in interval i (s),

$t_{d,i}$: duration of time interval i during which the arrival flow-rate and saturation flow-rate are constant (s),

Q_i: queue size at the end of interval i (veh),

v: demand flow-rate (veh/h),

q: arrival flow-rate $= v/3600$ (veh/s),

w_q: queue change rate (veh/s),

C: the cycle length (s).

The second component (d_2) accounts for delay caused by random arrivals or oversaturation during the analysis period. This component is calculated with,

$$d_2 = 900T((X-1) + \sqrt{(X-1)^2 + \frac{8kIX}{cT}}) \tag{4}$$

Here T is the duration of the analysis period during which traffic conditions are assumed to remain steady. The selection of T typically ranges from 15 minutes to 1 hour. X is the volume-to-capacity ratio ($\frac{v}{c}$), where c is the capacity. k is a factor that depends on controller settings and varies from 0.04 to 0.5. For pre-timed phases, coordinated phases, and phases set to " recall-to-maximum", a value of $k = 0.5$ is recommended. I is an adjustment factor that accounts for the effect of an upstream signal on vehicle arrivals. The selection of I is typically within $[0.09, 1]$. A value of $I = 1$ is used for isolated intersections, while smaller values are recommended for interacting intersections.

A third component d_3 is included whenever a queue exists at the beginning of the analysis period. This term is calculated with,

$$d_3 = \frac{3600}{vT}(t_A \frac{Q_b + Q_e - Q_{eo}}{2} + \frac{Q_e^2 - Q_{eo}^2}{2c} - \frac{Q_b^2}{2c}) \tag{5}$$

with

$$Q_e = Q_b + t_A(v - c) \tag{6}$$

$$Q_{eo} = \begin{cases} 0, & v < c \\ T(v-c), & Otherwise \end{cases} \tag{7}$$

$$t_A = \begin{cases} \min\left\{\frac{Q_b}{c-v}, T\right\}, & v < c \\ T, & Otherwise \end{cases} \tag{8}$$

where

t_A: duration of unmet demand in the analysis period (h),

Q_b: initial queue at the beginning of the analysis period (veh),

Q_e: queue at the end of the analysis period (veh),

Q_{eo}: queue at the end of the analysis period when $v \geq c$ and $Q_b = 0$ (veh).

The average delays for approach a (d_a) and for intersection i (d_i) are calculated by aggregating the lane group delays,

$$d_a = \frac{\sum_{g \in \chi_a} d_g v_g}{v_a},$$
$$d_i = \frac{\sum_{a \in \chi_i} d_a v_a}{\sum_{a \in \chi_i} v_a}. \tag{9}$$

Here χ_a is the set of lane groups belonging to approach a, χ_i is the set of approaches belonging to intersection i, and $v_a = \sum_{g \in \chi_a} v_g$ is the volume on approach a.

Once the average delays for lane groups, approaches, and intersection have been calculated, the corresponding LOS's are found with the lookup table provided in Table **1**.

Table 1: LOS criteria for signalized intersections in the HCM.

LOS	Average delay (s/veh)
A	≤ 10
B	$(10, 20]$
C	$(20, 35]$
D	$(35, 55]$
E	$(55, 80]$
F	≥ 80

Section 2.2: Commentary

Although the HCM delay estimation methodology is widely used by traffic engineers and has been tested and validated in numerous field studies, there are several limitations that are worth noting. First, the method requires a large amount of information regarding the geometric design, signal timing parameters, demand volumes, estimated capacities, and even estimated queues. This makes its application time consuming and error prone. Second, the delay calculation for actuated traffic signals is very complicated since the cycle lengths and green times depend on arrival patterns and may vary during the analysis period. Third, the formulas assume that traffic conditions downstream of the intersection are clear. For this reason the results become less reliable as the levels of congestion increase, which is precisely when delay estimates are most important. Finally, the methodology assumes that lane groups exist independently of one another, such that their delays can be computed separately with simple equations. Hence, interaction between different lane group movements for the same approach is not considered.

Section 3: The proposed method

In this section we develop a method for computing link delays and VHTs that makes use of travel time measurements obtained from probe vehicles or by some re-identification techniques. As illustrated in Figure **3**, it is assumed that travel times are collected for some portion of the vehicle population from the entrance of an approach section (point A) to the exit of the intersection (points B). The goal is to combine this data with existing loop detector measurements so as to obtain estimates of delay and VHT.

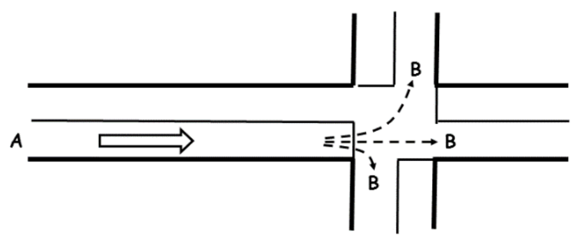


Figure 3. Measurement of link travel times.

Figure **4** illustrates the proposed approach. The inputs are volumes obtained from mid-block loop detectors, the set of travel time samples, and the geometric characteristics of the intersection, specifically the segment lengths and free-flow speeds (or, equivalently, the free flow travel times). The mid-block volumes and the probe travel times are provided to the VHT estimation algorithm. This component does not require system parameters. Delay is calculated directly from the sampled travel times and the free flow travel time. Delay is then used in Table **1** to obtain LOS. Further details are provided below.

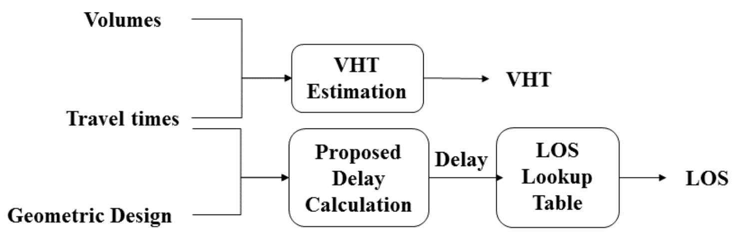


Figure 4. A data fusion method to calculate VHT, delay, and LOS.

Section 3.1: Travel time measurements

Travel time measurements are taken from the beginning of one road segment (point A) to the beginning of the next road segment (points B), as shown in Figure **3**. Thus the measurements capture the free-flow travel time, plus the delays induced by the traffic signal and by congestion spilling back from downstream segments. As a result, neither the details of the signal controller nor the presence of downstream congestion need to be provided to the estimator, since both of these are recorded in the travel time measurements.

We assume that travel times are recorded over a time period T (15 minutes, for example). Use n to denote the total number of arrivals to points B during this period. The set of travel times is $tt = \{tt_i\}_{i=1,\ldots,n}$. We assume that only a small portion p (the penetration rate) of these are observed. The set of observed travel times $\overline{tt}$ is a subset of the actual ones ($\overline{tt} \subseteq tt$). The number of collected samples, $\bar{n}$, is a binomial random variable with n trials and probability p, i.e., $\bar{n} \sim B(n, p)$.

Section 3.2: VHT estimation

VHT is typically defined as the time spent by vehicles in some portion of a transportation network during some given period of time, e.g., the horizontal length of trajectories within the box in Figure **5**. This concept can be applied to any portion of the network: a lane group, a road segment, an intersection, etc. Here we will adopt a slightly different definition of VHT as the total time spent by vehicles who "exit" the study area during the given time interval, e.g., the horizontal lengths of solid trajectories, which cross point B during the observation period in Figure **5**. These definitions are similar whenever vehicle trips do not start or end within the study area, and the observation time period is sufficiently large. The VHT thus defined can be expressed as,

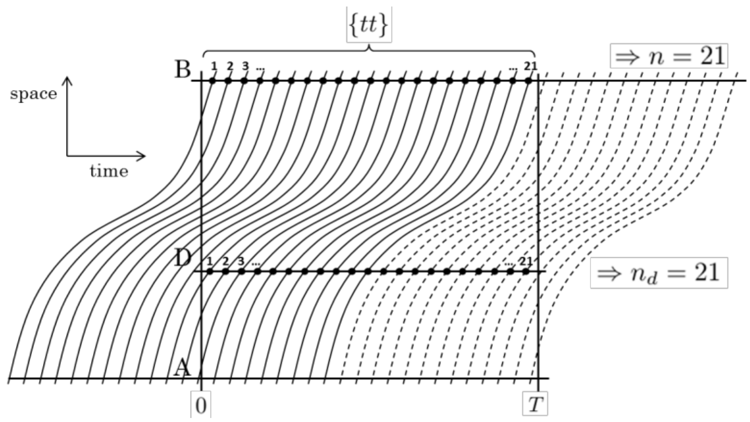


Figure 5. Two different definitions of VHT.

$$VHT = \sum_{i=1}^{n} tt_i = n \times mean(tt) \tag{10}$$

where tt is the set of trips that exit the study area during the sample period, and $mean(tt)$ is their average.

The proposed strategy for estimating VHT is to approximate the two terms in Eq. (10) using the available measurements. Specifically we use the vehicle count of the mid-block loop detectors as a proxy for n, and the mean of the reported travel times as an estimate of the average travel time. Thus we obtain the following estimator for VHT,

$$\widehat{VHT} = n_d \times mean(\overline{tt}) \tag{11}$$

where n_d is the total number of vehicles observed by the mid-block detectors (shown in Figure **5**) during time interval T.

Despite its simplicity, this estimator has some useful properties. First, the fact that it is non-parametric implies that it does not rely on the assumptions of any particular model. This is in contrast to the HCM methodology which has assumptions on vehicle's arrival patterns, traffic conditions, etc. This property allows the estimator to adapt to changing conditions with little or no additional tuning. Given the required measurements of travel time, Eq. (11) is valid for any signal control algorithm (fixed-time or actuated). It is also valid over a range of demands and queueing states, as long as the mid-block loop detector provides a good estimate of the number of trips completed during the time interval T. Second, the precision of the estimator is in a direct relation with the availability of probe data, which is likely to increase in the coming years. In Section 5 we investigate the dependency between estimation error and penetration rate. Finally, if we assume n_d to provide a good estimate of the total number of completed trips, then the estimator will be consistent, meaning that it will converge to the true VHT as the penetration rate reaches 1. One can observe in Figure **5** that n_d becomes a better representation of the number of trips as the length of the observation time increases.

Section 3.3: Delay estimation

Delay for a vehicle is defined as the difference between its actual travel time and its free flow travel time,

$$d_v = tt_v - tt_f \tag{12}$$

Here the sub-index v represents the vehicle. The free flow travel time tt_f assumes that the vehicle is not delayed by traffic or the signal. It is typically calculated as the ratio of the length of the segment to the free flow speed. Here we assume that it is given and equal for all vehicles in the segment. In this case the lane group and approach delays can be expressed as,

$$d_g = mean(tt_g) - tt_f \tag{13}$$

$$d_a = mean(tt_a) - tt_f \tag{14}$$

where tt_g and tt_a are the collection of travel times for vehicles in the lane group and the approach segment respectively. These definitions imply the following relationship between the lane group delays and the approach level delay,

$$n_a d_a = \sum_{g \in \chi_a} n_g d_g \tag{15}$$

where n_g is the number of trips completed for lane group g, and $n_a = \sum_{g \in \chi_a} n_g$ is the number of completed trips for the approach.

Delays are estimated similarly to VHT, by taking the sample mean of travel times recorded for either the lane group or the approach,

$$\hat{d}_g = mean(\hat{tt}_g) - tt_f \tag{16}$$

$$\hat{d}_a = mean(\hat{tt}_a) - tt_f \tag{17}$$

These are both consistent and unbiased estimators of delay, as they are sample means taken from the underlying delay distribution. Also, the sample mean has the property of being the minimum variance estimator amongst all linear estimators of the mean, regardless of the underlying distribution (assuming i.i.d. samples). This fact provides important support for the use of non-parametric estimators for this purpose.

The estimated LOS is obtained by evaluating Table **1** using estimated delay values.

Section 4: Study framework

The goals of the experiments are i) to test the various assumptions that have been made in constructing the VHT and delay estimators, ii) to assess the performance of the estimators under various traffic scenarios, iii) to compare that performance with the HCM methodology, and iv) to study the dependency of the estimation error on the penetration rate of probes.

To do this it is necessary to collect a “ground truth" set of travel times. For this we used a micro-simulation model, as described next.

Section 4.1: Study site

The experiments were done using an Aimsun model of Huntington Dr. and Colorado Blvd. in Arcadia, California. These two streets run parallel to the I-210 freeway and have been extensively studied as possible detour routes for corridor management [11]. The model includes eleven signalized intersections, of which six were chosen for detailed study: @San Clara St., @Santa Anita Ave., @1st Ave., @2nd Ave., @Gateway Dr., and @5th Ave. (see Figure **6**). Table **2** provides additional parameters used in the calculations: link lengths and free-flow speeds per intersection approach.

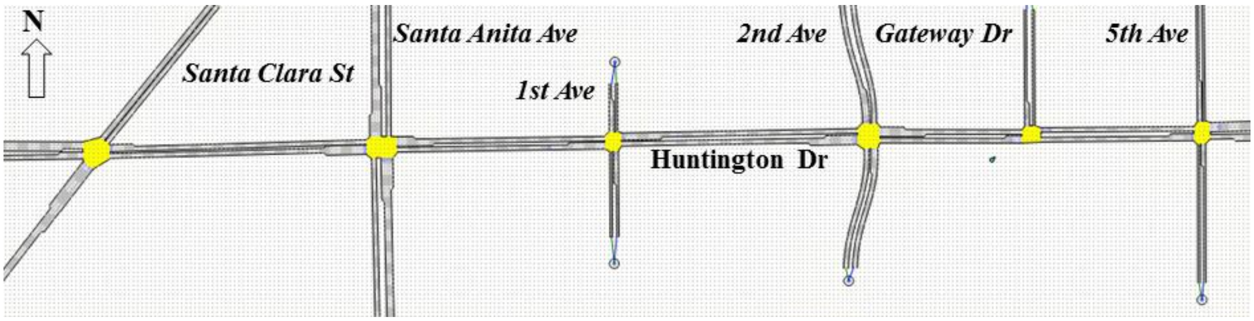


Figure 6. Six selected signalized intersections along the Huntington Dr. in the City of Acadia, CA.

The lane groups were defined as the set of traffic movements sharing the same lanes. Thus, an approach in which all movements share all lanes is considered to have only one lane group. An approach in which left-turns, through movements, and right-turns are segregated into different lanes is considered to have three lane groups. The lane group information is also provided in Table **2**.

The simulation model includes fixed-time signal controllers with protected left turns for all intersections. The specific settings (green times, offsets, etc.) were modified from the original values in order to produce a more varied array of queueing scenarios. A common cycle length of 120 seconds was applied to all intersections. Yellow and all-red times were set to 3 seconds and 2 seconds respectively. Detailed signal settings are provided in Table **2**.

Demands applied to the sources of the network were updated every five minutes. Aimsun generates a uniform stream of vehicles at the sources as long as the first road segment has sufficient space to accommodate them.

Table 2: Intersection geometric design and signal settings for the six signalized intersections.

Intersection	Approach	L_l (ft)	v_f (mph)	Lane Group ID	No. of Lanes	Phase ID	g (sec)
5th Ave.	Eastbound	624	30	1: L	1	1	25
				2: T & R	2	2	35
	Southbound	589	30	1: L	1	3	10
				2: T & R	1	4	30
	Westbound	477	30	1: L	1	1	25
				2: T	2	2	35
				3: R	1	2	35
	Northbound	565	30	1: L	1	3	10
				2: T & R	1	4	30
Gateway Dr.	Eastbound	578	30	1: L	1	1	15
				2: T	2	2	60
	Southbound	473	30	1: L	1	3	30
				2: R	1	3	30
	Westbound	623	30	1: T & R	2	2	60
2nd Ave.	Eastbound	964	30	1: L	1	1	20
				2: T	2	2	30
				3: R	1	2	30
	Southbound	717	30	1: L	1	3	25
				2: T & R	2	4	25
	Westbound	578	30	1: L	1	1	20
				2: T	2	2	30
				3: R	1	2	30
	Northbound	495	30	1: L	1	3	25
				2: T & R	2	4	25
1st Ave.	Eastbound	848	30	1: L	1	1	15
				2: T & R	2	2	45
	Southbound	195	30	1: L	1	3	15
				2: T & R	1	4	25
	Westbound	964	30	1: L	1	1	15
				2: T & R	2	2	45
	Northbound	346	30	1: L	1	3	15
				2: T & R	1	4	25
Santa Anita Ave.	Eastbound	1048	30	1: L	1	1	15
				2: T	2	2	35
				3: R	1	2	35
	Southbound	734	30	1: L	2	3	20
				2: T	2	4	30
				3: R	1	4	30
	Westbound	849	30	1: L	1	1	15
				2: T	2	2	35
				3: R	1	2	35
	Northbound	1506	30	1: L	2	3	20
				2: T	2	4	30
				3: R	1	4	30
Santa Clara St.	Eastbound	380	30	1: L	1	1	43
				2: T	2	1	80
	Southbound	717	30	1: R	2	3	30
	Westbound	1046	30	1: T	2	2	32
				2: R	1	2	32
	Northbound	1953	30	1: L	1	3	30
				2: T	2	3	30
				3: R	2	3	30

Note: L: left-turn movement; T: through movement; and R: right-turn movement. "1: L" means the left-turn movement is defined as lane group 1.

Section 4.2: Experimental design

The steps of the experiments are as follows,

1) **Model calibration**. The model was calibrated by a previous effort of the Connected Corridors group at U.C. Berkeley [11].
2) **Vehicle trajectories**. The model was run over the four-hour afternoon peak period, between 4:00 PM and 8:00 PM. The simulation was observed to produce a variety of queueing conditions at different intersections. Data was gathered from six of these including demands, link counts, signal settings, and geometric designs. Detailed trajectories were gathered for all vehicles in the system. From these it is possible to calculate the travel time between any two points for a given vehicle.
3) **Loop detector data**. The midblock loop detectors were placed in the middle of the links. Synthetic measurements were gathered for each lane group in every approach segment. These were aggregated into $T = 15$ minute observation periods.
4) **HCM estimates**. The signal settings, geometric designs, and demand volumes were used in Eqs. (1) through (8) and Table **1** to calculate HCM estimates of delay and LOS for all lane groups and for each 15-minute period. These were aggregated using Eq. (9) to the approach and intersection levels. Here the demand volume is computed as the arrival rate at the entrance of each approach segment. Such a measurement is a good proxy for the actual demand when no queue spills back to the entrance point and the analysis time period is long enough. For real-world applications, the midblock detector counts can be used as the demand volumes.
5) **Travel times**. For each period k and approach link i, the number of vehicles that exited the intersection during k after transiting through i was determined. These were organized into travel time sets indexed by k and i: $\{tt\}_{k,i}$.

6) **Ground truth calculations**. True values of VHT and delay were computed from the synthetic data for all lane groups, approaches, and intersections that were considered. The delays were used to find ground truth LOS values from Table **1**.

7) **Sampled travel times**. Vehicles that report travel times are sampled from the whole population in the network at a penetration rate p. The sampling follows a Binomial distribution, i.e., $\bar{n} \sim B(n, p)$, where n is number of unique vehicles in the network, and $\bar{n}$ is the number of sampled vehicles. Then the set of sampled travel times $\{\overline{tt}\}_{k,i}$ is formed using these sampled vehicles.

8) **Estimates of VHT, delay, and LOS**. Lane group VHT and delay estimates were computed with Eqs. (11) and (16) for all lane groups and time intervals. There are two possible methods for estimating VHT and delay for an approach. The first is to aggregate the lane-group VHTs to the approach level and apply Eq. (9) to get the approach delay. The other is to directly take the mean of the collected travel times for the approach and apply Eqs. (11) and (17) to get the approach VHT and delay. The former is only possible if lane-by-lane counts are available, which is not always the case. Here we consider both of these possibilities. Estimated LOS values were found using Table **1**.

9) Errors calculations. Steps 7) and 8) were repeated 500 times for each penetration rate, and for 18 different penetrations rates in [0, 1]. The selected penetration ratios are listed in Figure **7**. Deviations of the various estimates from the ground truth values were computed using the MAPE (mean absolute percentage error) metric.

Section 5: Results

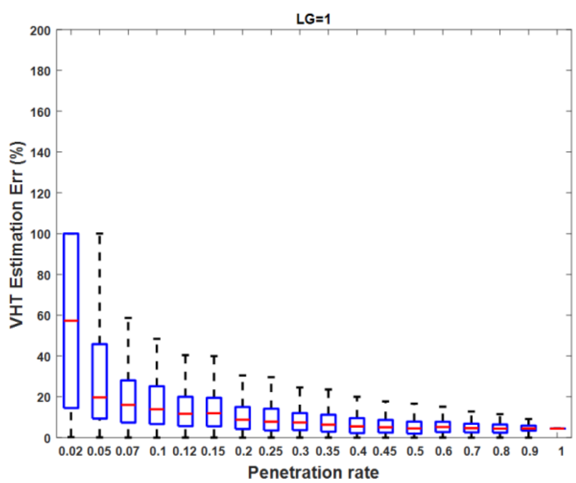
In this section we compare the errors obtained with the proposed method to those of the HCM.

Section 5.1: VHT estimation errors

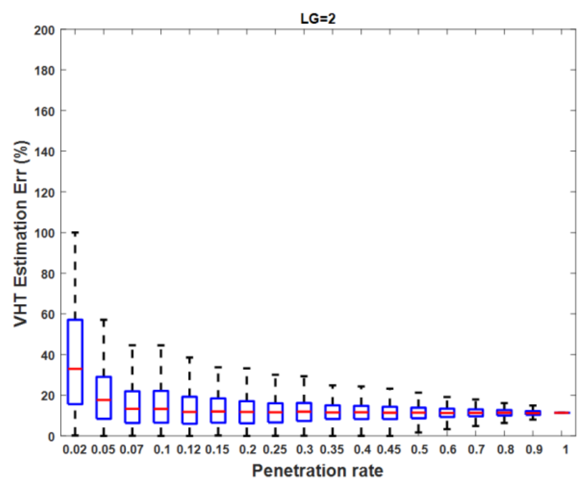
Section 5.1.1: Lane group based

Figure **7** provides box plots of lane group based VHT estimation errors for the eastbound direction of the intersection at Santa Anita Ave. from 4:15PM to 4:30PM. Analogous plots for other approaches, intersections, and time intervals exhibit similar patterns.

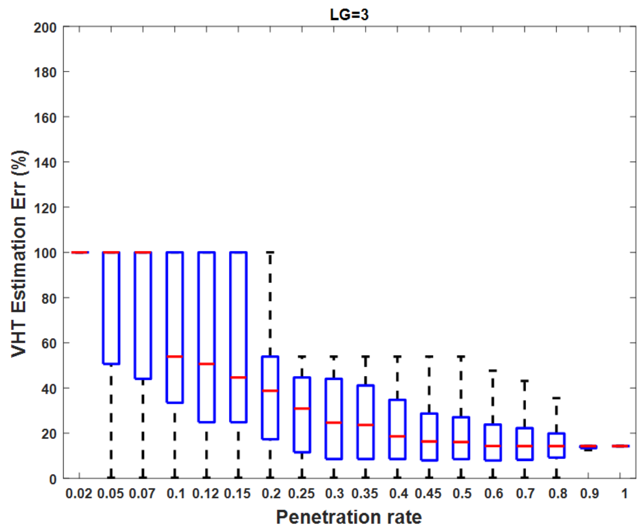
The following observations can be made. First, the estimation error as well as its variance reduces as the penetration rate increases. This is true for all lane groups and population sizes. Second, for a fixed penetration rate, the estimation error reduces as the population size grows. Thus we can identify two important factors for VHT inference: penetration rate and population size. Either a high penetration rate or a large population size is needed for a good estimation of VHT and delay from sampled travel times.



(a) 45 vehicles for LG 1 (left-turn movement)



(b) 88 vehicles for LG 2 (through movement)

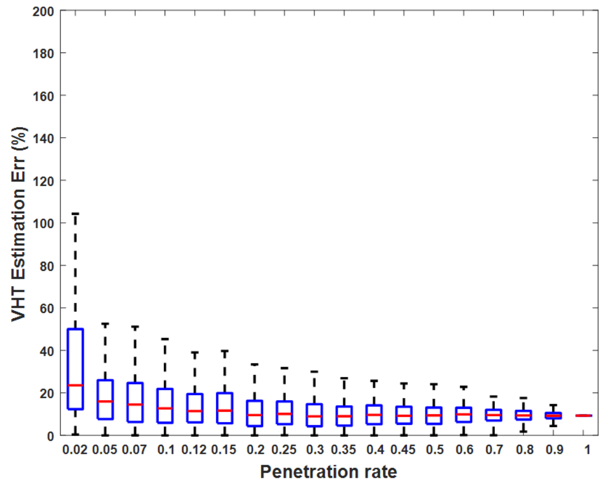


(c) 7 vehicles for LG 3 (right-turn movement)

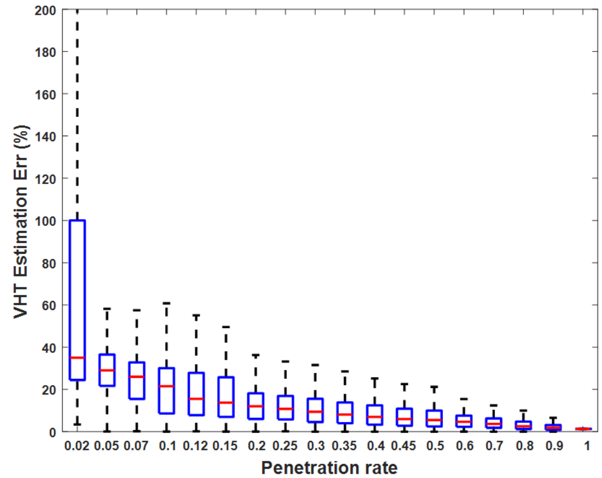
Figure 7. Lane group based VHT estimation errors in the eastbound direction of the intersection at Santa Anita Ave. from 4:15 PM to 4:30 PM.

Section 5.1.2: Approach based

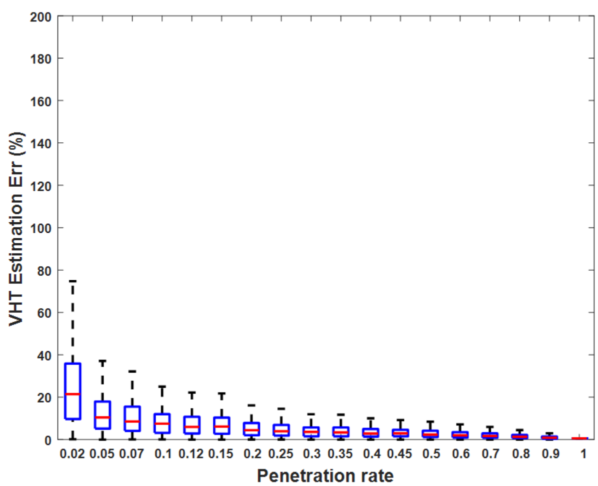
Figure **8** provides the box plots of approach based VHT estimation errors for the same intersection and time period. Similar patterns as those in the lane group case are apparent. In addition, compared with the lane group based case, VHT estimates at the approach level turn out to be more reliable with smaller variance. This can be explained by the fact that a larger number of samples is taken for the approach than for the individual lane-groups, while the travel time patterns are similar across lane groups in the same approach segment.



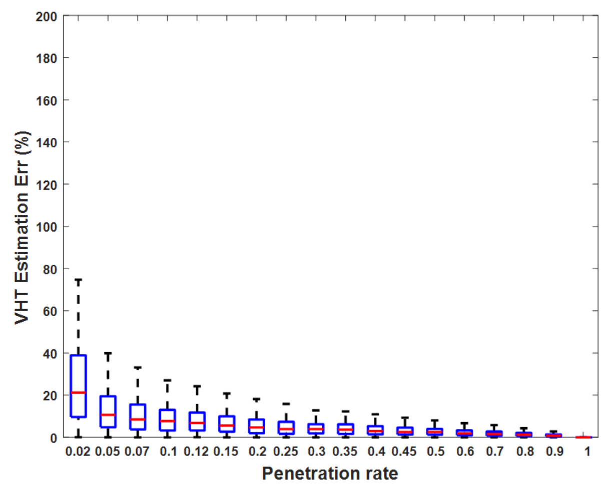
(a) 140 vehicles for the EB direction



(b) 77 vehicles for the WB direction



(c) 170 vehicles for the SB direction



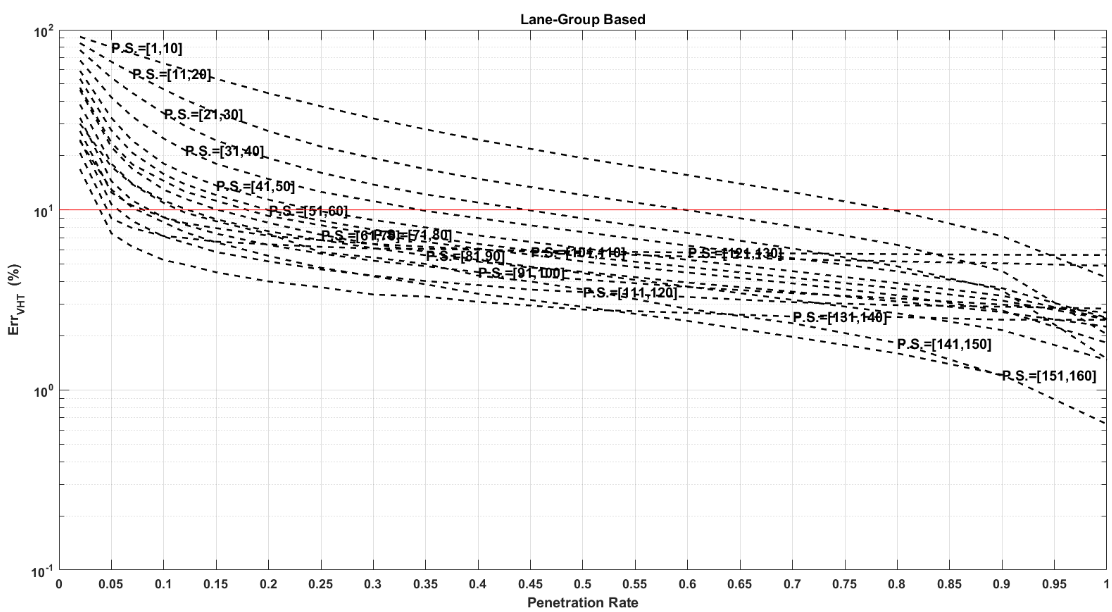
(d) 124 vehicles for the NB direction

Figure 8. Approach based VHT estimation errors at the intersection of Santa Anita Ave. from 4:15 PM to 4:30 PM.

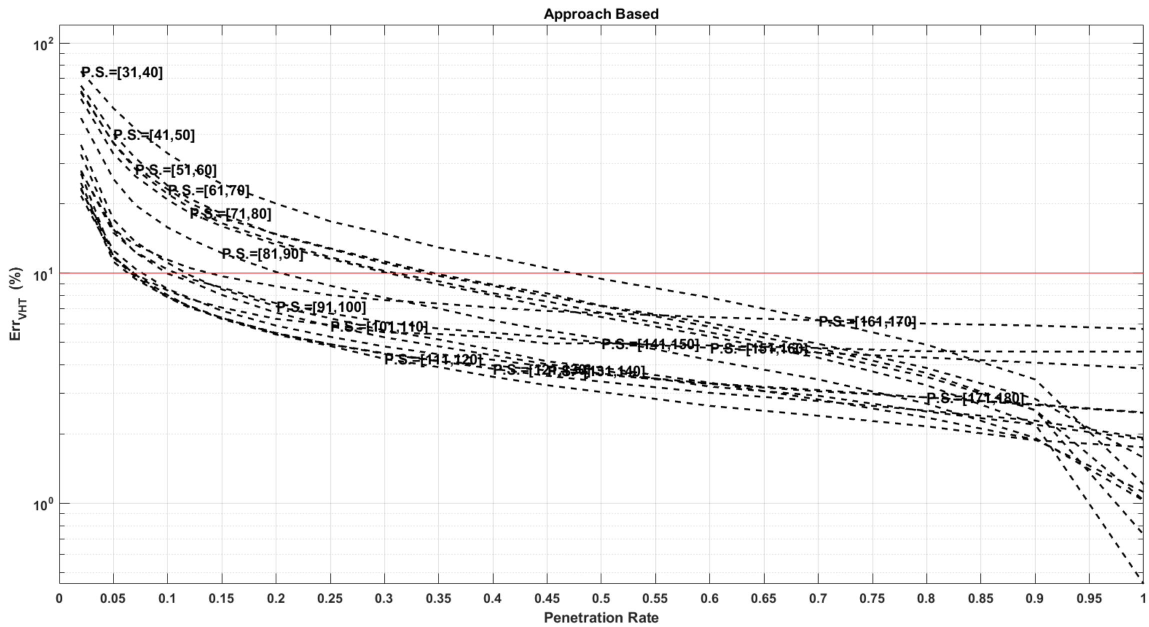
Section 5.1.3: Impact of penetration rates

Based on the above analysis, we further analyze the impact of penetration rates on the estimation errors. For both lane group and approach based methods, we first group the population size (P.S.) into different bins. Then in each bin, we obtain the average VHT estimation error for each given penetration rate. Figure

9 illustrates the relation between the average VHT estimation error and the penetration rate under different bins of population sizes. Regardless of lane groups, approaches, and population sizes, it is clear to find a monotonic decreasing trend for the estimation error as the penetration rate increases. The improvement is significant especially when the penetration rate is less than 10%. When the penetration rate is about 5-7%, it generally requires more than 90 vehicles (per 15 minutes) for one lane group to guarantee the average estimation error less than 10%. For the approach based method, more than 110 vehicles (per 15 minutes) are needed to guarantee the same estimation error. Since it is relatively easier for one approach to satisfy such a requirement, it is better to estimate the performance metrics at the approach level. Note that when the penetration rate reaches one, the VHT estimation error is very small but not zero. That is because the vehicle counts from midblock detectors ($\bar{n}$) generally have small deviations from the actual demands (n).



(a) Lane group based



(b) Approach based

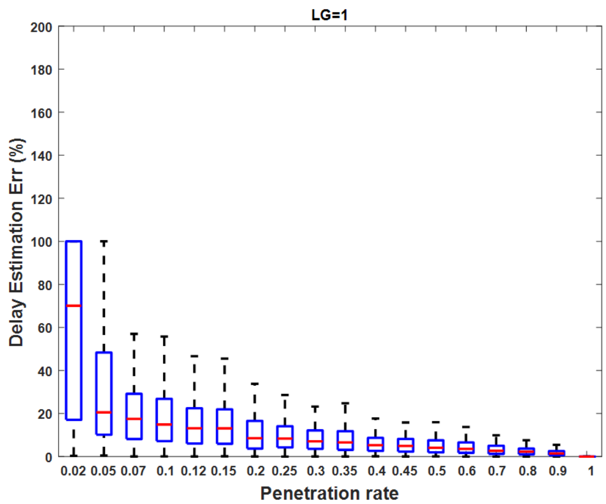
Figure 9. The impact of penetration rates on the estimation errors.

Section 5.2: Delay and LOS calculation errors

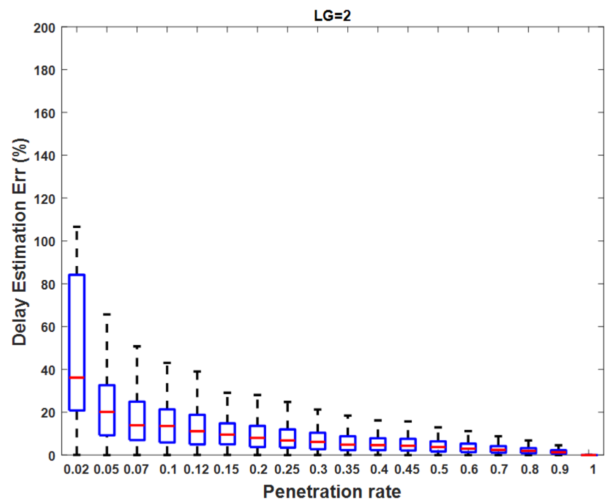
The ground-truth delay is calculated using Eqs. (16) and (17) by replacing the sets of observed travel times $\overline{tt}_g$ and $\overline{tt}_a$ with the ground truth ones tt_g and tt_a. The HCM delay is calculated using Eqs. (1) to (9). The delay for the proposed method is calculated using Eqs. (16) and (17). The same LOS lookup table (Table **1**) is used to map the delays to the corresponding LOS's.

Section 5.2.1: Delay

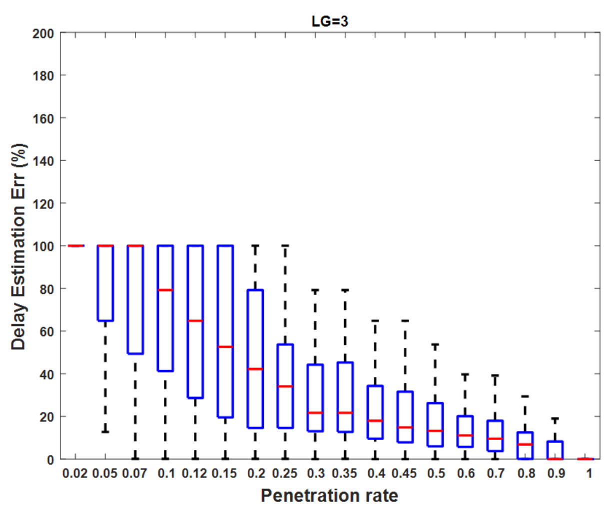
Figure **10** provides the lane group based delay estimation errors using the proposed method for the eastbound approach of the intersection at Santa Anita Ave. From the figure, it is not surprising to find that the trends are similar to those in Figure **7** since the delays are calculated from the estimates of average travel times, which are also used in the VHT estimation. As a comparison, we provide the corresponding errors using the HCM delay calculation. We find that the proposed method performs better than the HCM if the population size is large, even with low penetration rates. However, it requires a higher penetration rate if the population size is small. For example, the right-turn movement has only 7 vehicles, and it requires a penetration rate of more than 30% in order to outperform the HCM method. Figure **11** provides the approach based delay estimation errors for the same intersection and time period. We can find that in order to perform better than the HCM, a penetration rate of 7-10% is required.



(a) Left-Turn (HCM Delay Err: 49.5%)

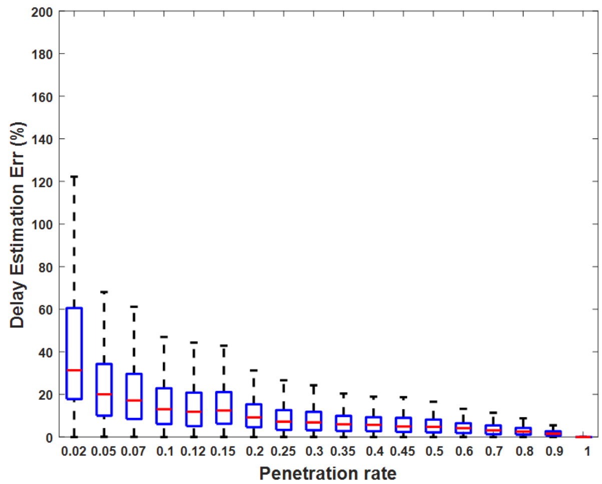


(b) Through (HCM Delay Err: 14.9%)

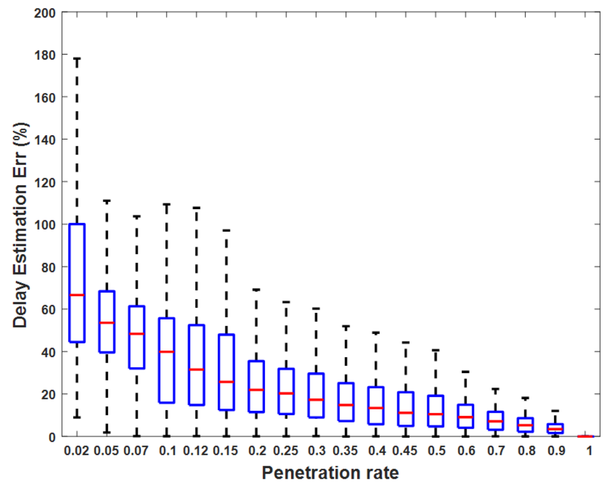


(c) Right-turn (HCM Delay Err: 28.5%)

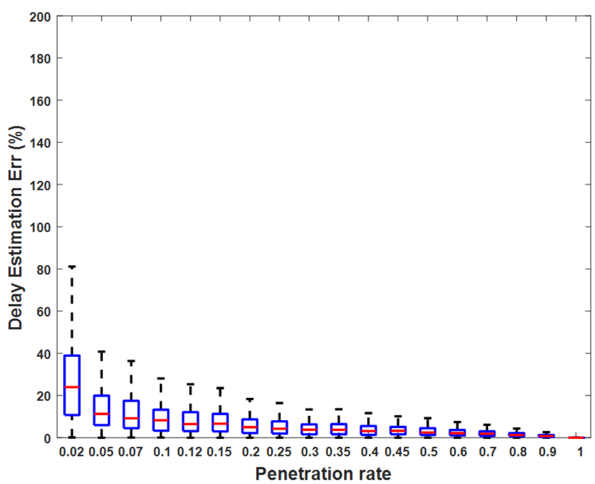
Figure 10. Lane group based delay estimation errors in the eastbound direction of the intersection at Santa Anita Ave. from 4:15 PM to 4:30 PM.



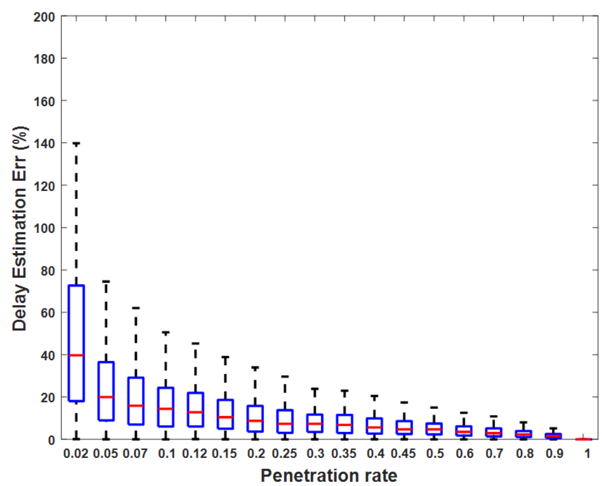
(a) EB direction: (HCM Delay Err: 36.1%)



(b) WB direction: (HCM Delay Err: 90.9%)



(c) SB direction: (HCM Delay Err: 15.4%)



(d) NB direction: (HCM Delay Err: 7.5%)

Figure 11. Approach based delay estimation errors at the intersection of Santa Anita Ave. from 4:15 PM to 4:30 PM.

Section 5.2.2: LOS

Table **3** provides the estimated approach LOS's for all intersections from 4:00PM to 4:30PM. The estimation accuracy for all approaches, intersections, and time intervals (Totaling 368 samples) is also provided under different penetration rates. From the table, we can find that about 40% of the LOS estimates using the HCM method are incorrect. When the penetration rate is low, e.g. 2%, the estimation accuracy using the proposed method is very low, even lower than that of the HCM method. However, as the penetration rate increases, the estimation gets more accurate. For example, if the penetration rate reaches 7%, the accuracy is 73.6% for the lane group based case, and 82.6% for the approach based one. Furthermore, we find that the proposed method always performs better at the approach level than the lane group level for any given penetration rate. Note that, for some approaches, the proposed method provides incorrect estimates even when the penetration rate reaches 10%, which is a result of a low population size.

Table 3: Approach LOS estimation results for all intersections.

Intersection	Approach	Time	LOS									
			Ground Truth	HCM	VHT Inference							
					pr=2%		pr=5%		pr=7%		pr=10%	
					LG Based	App. Based	LG Based	App. Based	LG Based	App. Based	LG Based	App. Based
5th Ave.	EB	4:00-4:15PM	B	D	A	A	A	A	A	A	A	A
		4:15-4:30PM	B	D	A	A	A	A	A	A	A	A
	WB	4:00-4:15PM	F	D	F	F	F	F	F	F	F	F
		4:15-4:30PM	F	F	E	F	F	F	F	F	F	F
	SB	4:00-4:15PM	D	D	B	C	C	D	C	D	D	D
		4:15-4:30PM	D	D	B	C	C	D	D	D	D	D
	NB	4:00-4:15PM	D	D	B	C	C	D	C	D	D	D
		4:15-4:30PM	D	E	B	C	C	D	C	D	D	D
Gateway Dr.	EB	4:00-4:15PM	D	D	B	B	C	C	C	C	D	D
		4:15-4:30PM	E	D	B	C	C	C	D	D	D	D
	WB	4:00-4:15PM	C	B	B	B	B	B	B	B	C	C
		4:15-4:30PM	C	B	B	B	C	C	C	C	C	C
	SB	4:00-4:15PM	C	D	B	A	B	B	C	C	C	C
		4:15-4:30PM	C	D	B	B	C	C	C	C	C	C
2nd Ave.	EB	4:00-4:15PM	E	E	E	E	E	E	E	E	E	E
		4:15-4:30PM	E	F	D	E	E	E	E	E	E	E
	WB	4:00-4:15PM	C	D	A	A	A	B	A	B	B	B
		4:15-4:30PM	B	D	A	A	A	A	A	A	A	A
	SB	4:00-4:15PM	D	D	C	C	D	D	D	D	D	D
		4:15-4:30PM	D	D	C	C	C	D	D	D	D	D
	NB	4:00-4:15PM	D	E	D	D	D	D	D	D	D	D
		4:15-4:30PM	E	E	D	D	E	E	E	E	E	E
1st Ave.	EB	4:00-4:15PM	C	C	C	C	C	C	C	C	C	C
		4:15-4:30PM	C	C	C	C	C	C	C	C	C	C
	WB	4:00-4:15PM	D	C	D	D	D	D	D	D	D	D
		4:15-4:30PM	D	C	D	D	D	D	D	D	D	D
	SB	4:00-4:15PM	D	D	B	C	C	D	D	D	D	D
		4:15-4:30PM	D	D	B	C	C	D	D	D	D	D
	NB	4:00-4:15PM	D	F	C	C	D	D	D	D	D	D
		4:15-4:30PM	D	F	C	C	D	D	D	D	D	D
Santa Anita Ave.	EB	4:00-4:15PM	D	D	C	D	D	D	D	D	D	D
		4:15-4:30PM	E	D	D	E	E	E	E	E	E	E
	WB	4:00-4:15PM	C	D	B	B	B	C	B	C	B	C
		4:15-4:30PM	C	D	A	B	B	C	B	C	B	C
	SB	4:00-4:15PM	F	F	F	F	F	F	F	F	F	F
		4:15-4:30PM	F	F	F	F	F	F	F	F	F	F
	NB	4:00-4:15PM	D	D	C	D	D	D	D	D	D	D
		4:15-4:30PM	D	D	C	D	D	D	D	D	D	D
Santa Clara St.	EB	4:00-4:15PM	F	E	F	F	F	F	F	F	F	F
		4:15-4:30PM	F	F	F	F	F	F	F	F	F	F
	WB	4:00-4:15PM	F	F	F	F	F	F	F	F	F	F
		4:15-4:30PM	F	F	F	F	F	F	F	F	F	F
	SB	4:00-4:15PM	D	D	C	C	D	D	D	D	D	D
		4:15-4:30PM	D	D	D	D	D	D	D	D	D	D
	NB	4:00-4:15PM	E	E	E	E	E	E	E	E	E	E
		4:15-4:30PM	F	F	F	F	F	F	F	F	F	F
Accuracy (for all approaches, intersections, and time intervals; 368 samples)				59.0%	35.3%	46.7%	60.1%	76.4%	73.6%	82.6%	86.1%	87.0%

Section 6: Conclusions and recommendations

Section 6.1: Conclusions

This study has proposed a data-fusion based method for computing performance metrics for intersections. Under the assumption that travel time data can be obtained from a sub-population of vehicles, we proposed to combine the sample mean of travel times with the vehicle count obtained from mid-block loop detectors to obtain an estimate of VHT. This estimator has several advantages as compared to the current state-of-practice. First, it is data-driven rather than model-driven and therefore, it does not rely on any modeling assumptions. For this reason it can be applied in a variety of scenarios, including congestion and spillback. It is also very simple to compute as compared to the delay formulas of the HCM. The method also does not require signal timing parameters to be known. Although the travel time distributions were assumed to be stationary, it remains to be tested whether the method works well under actuated or adaptive signal control. The simple structure of the estimator make it, in our opinion, a good candidate for deployment, provided the travel times can be obtained.

A microsimulation-based traffic model was used to evaluate the performance of the estimator and compare it to the methodology of the HCM currently used by most traffic analysts in the United States. The simulation model provided a set of ground truth data, on which both methods were applied.

The complexity of the travel time distribution produced by the simulator strengthened the case for a data-driven approach. Two possibilities for data collection were considered: lane-group level and approach level, and the results showed that better results were obtained in the latter case. The study also identified the penetration rate and the population size as the two main factors influencing the estimation error. During peak hours, when the population size is large, only about 7% of probe vehicles are needed in order to obtain VHT, delay, and LOS estimated that improve upon the HCM. However a larger percentage is needed during off-peak hours.

The study also showed that the HCM formulas often failed to produce the correct value of LOS (40% of the time in Table **3**. The approach based VHT inference formula performed best at LOS estimation, producing the correct value 82.6% of the time at a penetration rate of 7%. These numbers hold promise for the use of probe data for estimating performance metrics for signalized intersections.

Section 6.2: Recommendations

According to our research findings, we have the following recommendations:

i. Our proposed method is very generic which doesn't rely on any particular models and can work under any distribution of travel times. Therefore, the research findings obtained from microsimulations can be applied to the field directly. For next steps, it would be better to validate what the current penetration rate of probe vehicles is on local arterials. We can expect that the requirement of 7% may not be met given certain time periods or for some minor streets. A better understanding on this would help us decide whether to switch from the HCM calculation method to the proposed one or not.

ii. From our analysis, we need two different types of data sources: traffic counts from loop detectors, and travel times from probe vehicles. It is very important to make sure: a) the local intersections are fully covered with loop detectors, and b) all these loop detectors are connected to a central database so that we are able to obtain real-time or historical traffic counts. That means we need to update the current infrastructure where it is needed. It is also beneficial of doing so since another performance metric, VMT, can be easily obtained if traffic count data is available for all approaches at a given intersection.

References

[1] E. Brown and K. Alex, “Preliminary evaluation of alternative methods of transportation analysis,” Governors Office of Planning and Research, December 30, 2013.

[2] K. Kwong, R. Kavaler, R. Rajagopal, and P. Varaiya, “A practical scheme for arterial travel time estimation based on vehicle re-identification using wireless sensors.” in Proceedings of the 89th Meeting of the Transportation Research Board.

[3] Mobile technology fact sheet. [Online]. Available: http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/

[4] J. C. Herrera, D. B. Work, R. Herring, X. J. Ban, Q. Jacobson, and A. M. Bayen, “Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment,” Transportation Research Part C: Emerging Technologies, vol. 18, no. 4, pp. 568–583, 2010.

[5] A. D. Patire, M. Wright, B. Prodhomme, and A. M. Bayen, “How much gps data do we need?” Transportation Research Part C: Emerging Technologies, 2015.

[6] A. Hofleitner, R. Herring, and A. Bayen, “Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning,” Transportation Research Part B: Methodological, vol. 46, no. 9, pp. 1097–1122, 2012.

[7] Y. Feng, J. Hourdos, and G. A. Davis, “Probe vehicle based real-time traffic monitoring on urban roadways,” Transportation Research Part C: Emerging Technologies, vol. 40, pp. 160–178, 2014.

[8] “Highway Capacity Manual 2010,” Washington, DC: Transportation Research Board, 2010.

[9] Aimsun website. [Online]. Available: https://www.aimsun.com/

[10] F. V. Webster, “Traffic signal settings,” Tech. Rep., 1958.

[11] F. Dion, “Connected corridors: I-210 pilot integrated corridor management system concept of operations,” California PATH, Berkeley, CA