

# Queue Length Estimation at Traffic Signals: Connected Vehicles with Range Measurement Sensors

Gurcan Comert<sup>a,\*</sup>, Mecit Cetin<sup>b</sup>

<sup>a</sup>*Department of Computer Sc., Phys., and Eng., Benedict College, 1600 Harden St., Columbia, SC 29204 USA*  
<sup>b</sup>*Department of Civil and Env. Engineering, Old Dominion University, 135 Kaufman Hall, Norfolk, VA 23529 USA*

## Abstract

Today vehicles are becoming a rich source of data as they are equipped with localization or tracking and with wireless communications technologies. With the increasing interest in automated- or self-driving technologies, vehicles are also being equipped with range measuring sensors (e.g., LIDAR, stereo cameras, and ultrasonic) to detect other vehicles and objects in the surrounding environment. It is possible to envision that such vehicles could share their data with the transportation infrastructure elements (e.g., a traffic signal controller) to enable different mobility and safety applications. Data from these connected vehicles could then be used to estimate the system state in real-time. This paper develops queue length estimators from connected vehicles equipped with range measurement sensors. Simple plug-and-play models are developed for queue length estimations without needing ground truth queue lengths by extending the previous formulations. The proposed method is simple to implement and can be adopted to cyclic queues at traffic signals with known phase lengths. The derived models are evaluated with data from microscopic traffic simulations. From numerical experiments, the QLE model with range sensors improves the errors as much as 25% in variance-to-mean ratio and 5% in coefficient of variation at low  $\leq 20\%$  market penetration rates.

**Keywords:** Applied probability, queue length estimation, connected and autonomous vehicles, range measurement sensor, sensor fusion.

## 1. Introduction

Tracking individual vehicles traversing the transportation network in real time has gained significant momentum lately due to recent developments in Connected Vehicles (CVs) technologies. Real-time data from CV enable observing complex system dynamics and support new applications to improve system efficiency and safety. However, not all vehicles may be equipped with CV technologies nor all drivers are willing to share their locations. Consequently, predicting the overall system behavior based on data from a fraction of vehicles has been an area of active research for more than three decades now. Queue Length (QL) is a fundamental performance criteria for any system with multiple users demanding service from a server as for real-time control of signalized intersections or port terminal gates. It is critical for applications such as emergency management and reliability of the network. Models based on performance measures can be input for traffic control, facility design, and construction planning purposes. Specifically, the real-time (adaptive) control demands high accuracy from such models. Today, vehicles are becoming a rich data and computational resource as they are equipped with localization or tracking, wireless communication technologies, and microprocessors. Furthermore, with the increasing interest on automated- or self- driving technologies, vehicles are also being equipped with range measuring sensors (e.g., LIDAR, stereo cameras, and ultrasonic) to detect other vehicles and objects in the surrounding environment. It is possible to envision that such vehicles could share their data with the transportation infrastructure elements (e.g., a signal controller) to enable various traffic mobility and safety applications. Data from these CVs could then be used to estimate the system state in real-time. Consequently, this

\*Corresponding author

Email addresses: [gurcan.comert@benedict.edu](mailto:gurcan.comert@benedict.edu) (Gurcan Comert), [mcetin@odu.edu](mailto:mcetin@odu.edu) (Mecit Cetin)

paper is focused on estimation of QL from such CVs that may represent partial vehicles in the traffic or other systems (Nam et al. (2009)). Connected and autonomous vehicles (CAVs) or autonomous vehicles can also be treated similarly.

### 1.1. *Background*

Queue length estimation (QLE) from connected vehicle at traffic signals is a well-investigated problem. In basic form, there is no feedback from environment, thus, true queue length values are not known. This makes critical difference with the prediction formulations with series of true queue lengths are available where researchers used filtering, fit models, and deep learning methods (Tiaprasert et al. (2015); Yin et al. (2018); Gao et al. (2020); Comert et al. (2019)).

In the literature, researchers obtained CV based queues via reconstructing trajectories (Argote et al. (2011); Cetin (2012); Cheng et al. (2012); Xu et al. (2017); Rompis et al. (2018); Zhao et al. (2019b)). Stochastic formulations are desirable as increase in number of related paper show. Researchers developed important models using similar approach. Within the scope and space, for more information on general queue length estimation, readers are referred to recent survey studies on QLE methods from CAVs (Zhao et al. (2019b); Guo et al. (2019)) and queue estimation bibliography by (Asanjarani et al. (2017)). Hence here, we focus on related previous works on stochastic QLEs from CVs are discussed.

Stochastic models of QLEs from CVs mainly concerns with stationary (behavior after sufficiently long time interval) or non-stationary (short-time intervals, by cycle, and red phase). Such classification is impacted by treatment of arrival rate, distribution, and market penetration rate (MPR) (probe percentage). For instance, cycle-to-cycle estimation of QLs including green arrivals was modeled in (Tan et al. (2019)) and real-time applications with high frequency connected vehicle data was presented by (Mei et al. (2019)).

In queueing, detection the stopped vehicles accurately is crucial as it will be impacting the back of queue. This problem was addressed in (Liu et al. (2019)). Reported error levels in similar QLE studies may exceed 10% of true queue lengths (Argote et al. (2011)) which would require 80% CV penetration rate. In order to improve the accuracy, data fusion from different traffic surveillance technologies are also considered such as loops (Li et al. (2013); Comert (2013a)), license plate recognition (Tan et al. (2020)), and inclusion of radio frequency identification Wu and Yang (2013) which acts as filtered QLE would give higher errors. Queue lengths for constant arrival rate at the end of red phase act like autoregressive (2) time series with large coefficients with opposite signs or moving average (1) with positive high coefficient. Therefore, filtering for a true QLs would over-smooth the series. With back of queue estimation and convex optimization, similar error levels at  $p = 20\%$  was observed (Yang and Menendez (2018)). Range sensor applications in traffic in some capabilities are discussed in (Mandal et al. (2020)). Authors, placed ultrasonic sensors capable Arduinos on lane markings and reported distance that vehicles could be detected for queue lengths. Although we seek higher penetration levels, efficient estimators performing at low MPRs would be desirable as we will need to address efficient data management. Question of how much data we need is important not to overload systems or security reasons. Partially supporting these, in an interesting report, researchers found in a simulation environment, over 200 vehicles at 100% penetration level signal control from CVs becomes difficult (University of Arizona (2016)) due to computational load.

In this study, simple plug-and-play expressions are developed for QLE from CVs in the case of connected vehicles are also equipped with a range sensor by extending the authors' previous formulations in (Comert and Cetin (2009, 2011); Comert (2013b, 2016)). Derived formulate in this paper incorporate the measured range from each CV as an additional source of real-time data supplementing the typical CV information (e.g., location and time). In a standing queue, the range will inform whether there is a vehicle behind a CV. Steady-state behavior of estimated QLs and their errors are formulated. These are then compared against and validated with results from point queue as well as microscopic simulations.

### 1.2. *Contributions of this study*

Our modular expectation-based approach enables application of filtering algorithms to deal with measurement or low MPR related errors. Note also that the estimation problem is interesting where

only assume partially observed system. Thus, for any similar problem such as autonomous vehicles, our approach can be adopted. One of limitations for connected vehicles with low MPRs, as we discussed in (Comert (2016)), is distinguishing whether the last connected vehicle is the last vehicle in the queue. This information would be available from a range sensor (ultrasonic, LIDAR) Cetin et al. (2017) showed the accuracy levels in LIDAR-based estimations. Moreover, presented formulae do not depend on any primary parameter such as arrival rate or MPR. So, method can be used for time dependent arrival rates or  $p$ . Estimation of these primary parameters are becoming interest also (Comert (2016); Zhao et al. (2019a); Van Phu and Farhi (2018)).

Impact of additional sensors in terms of estimation errors for different arrival rates  $\lambda$  and MPR (probe percentage)  $p$  are presented. This research gives insights about the benefit of additional data from range measuring sensors that can be common with the upcoming autonomous vehicles. The results can be incorporated for more accurate queue lengths and delay calculations in new control heuristics (Comert et al. (2009)).

## 2. Problem Statement and Notation

Fig. 1 shows a view of an approach lane right before turning green, vehicles with darker colors represent CVs and shaded ones depict the detected vehicles via range sensors. Location of the connected vehicle information in queue  $L$  are inferred as number of vehicles. Number of CVs ( $M$ ) is 3, location of the last CV ( $L$ ) is 9, and queue time tag of the last probe  $T$  is 35 seconds (s). Moreover, last CV is able to detect the vehicle behind it, so,  $L' = (9 + 1)$  and queuing time of the following vehicle detected by sensor on the last CV is  $T' = 38$  s.

Basic problem is defined as deriving estimators for the total queue length  $N$  given the information from CVs and other detected vehicles by CVs' range measuring sensors such as LIDARs (i.e., CV locations and time tags). Range measuring sensors provide the critical information of whether the last CV actually is the last vehicle in the queue (i.e., there is at least one more vehicle joined the queue after it). The queue length ( $N$ ) is 11 which is estimated in real-time at the end of red phase  $R$  using the location and the time of CVs ( $L' = 10$ ,  $T' = 38$ ,  $M = 3$ , and  $R$ ). Estimators are developed for known  $R$ , but they can be used for actuated or adaptive signal phase lengths at the time instance of interest.

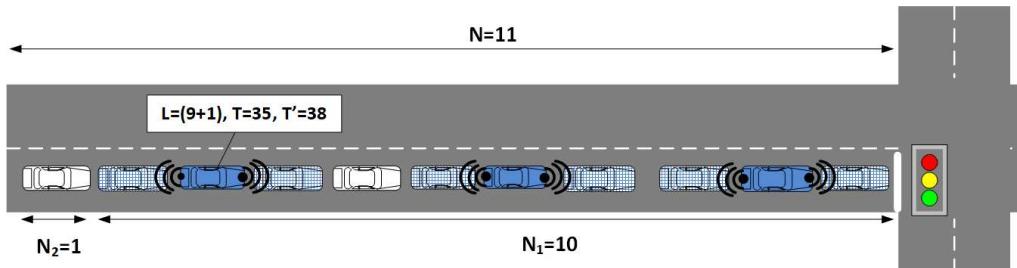


Fig. 1: Approach lane right before the green phase

### 2.1. Limitations of the study

In this study, we aim to develop analytical input-output models. For simplifications we assume following.

- (i) Approach lane has no capacity limit and there is no loss time due to acceleration. These assumptions are for adopted overflow queue models (Haight (1959); Ohno (1978); Medhi (1991); Heidemann (1994)). Usefulness of the models in the proposed estimators are shown via comparing with errors from microscopic traffic simulations (i.e., VISSIM) which provide more realistic arrivals and queuing at the signal.
- (ii) Poisson arrivals to the approach lane are assumed which can describe isolated intersections, low demand scenarios of more complex intersections, and enable simple closed-form derivations.

- (iii) Lane level accuracy for CV locations is assumed. Position and time tags are accurately received. Connected vehicle architectures especially at intersections will satisfy this by map information and timing passed by roadside units. When location is provided as latitude and longitude, using average vehicle length or passenger car equivalencies, spot in number of vehicles can be estimated. Basic safety messages also include vehicle type, using vehicle lengths by type locations can also be estimated.
- (iv) Uniform vehicle type is assumed. Mixed traffic could be incorporated to the models by arrival rates in Poisson arrivals (Walraevens et al. (2017)). Impact of different vehicle types would impact intersection capacity and location information.
- (v) Connected vehicles are able to sense accurately if there is a vehicle behind with the help of range sensors.

Estimators presented are unbiased under these assumptions. They would hold for estimating real-time queues at isolated signals, signals with low to moderate volume-to-capacity ratios, Poisson, or wherever approximately random arrivals can be assumed (Newell (1982); Fiems and De Turck (2019)). Such QLEs are going to be unbiased. In fact, any approach queue length can be estimated given formulae. Amount of deviations from random arrivals will be reflected into the estimation errors as bias.

### 3. Analysis for the Case without Overflow Queue

First, it is assumed that vehicles that have arrived in one given cycle clear the intersection in that cycle (the overflow queue ( $Q$ ) is zero). At each cycle, the total number of vehicles in the queue  $N$  is estimated given  $L$ ,  $T$ , and primary parameters arrival rate  $\lambda$ ,  $p$ , and  $R$ . Throughout the examples, signal related parameters are chosen to be effective red duration of 45 s, green time of 43.2 s, and discharge rate 1.8 seconds per vehicle ( $spv$ ). Other selected parameter values are  $\lambda = \{0.111, 0.133, 0.163, 0.190, 0.218, 0.239, 0.267\}$   $vps$  which correspond to volume-to-capacity ratios of  $\rho = \{0.41, 0.49, 0.60, 0.70, 0.80, 0.88, 0.98\}$  and  $0.001 \leq p \leq 1.00$  (Comert (2016)).

As shown in Eq. 1, estimator of the queue length  $N$  can be written as summation of expected values of the number of vehicles known until last CV  $N_1$  and inferred  $N_2$  after this CV for given conditions: (i) whether the last CV is also the last vehicle  $I(L = l_v) = 1$  and (ii) the last CV is not the last vehicle  $I(L \neq l_v) = 1$ . Here,  $I()$  is the indicator function.

$$E(N|L = l, T = t) = \begin{cases} [E(N_1|L = l, T = t) + E(N_2|L = l, T = t)]I(L = l_v) + \\ [E(N_1|L = l, T' = t') + E(N_2|L = l, T' = t')]I(L \neq l_v) \end{cases} \quad (1)$$

In Eq. (1), for condition (i)  $I(L = l_v)$ ,  $N_1$  is known and  $N_1 = l$  and  $E(N_2|L = l, T = t)$  is number of arrivals after the last CV during  $(R - T) = \delta$  which is zero leading to estimator with no error  $N = l$ . For condition (ii) ( $L \neq l_v$ ), after the last CV, only non-CVs would arrive, thus  $N_1 = l + 1$  and  $\delta = (R - T')$ . Arrival rate during  $\delta$  is  $(1-p)\lambda$ . Denoting  $(1-p)\lambda$  by  $\theta$ , we can simply write  $N_2|L = l, T = t' \sim \text{Poisson}(\theta\delta)$ . Simplified estimator is given in Eq. (2).

For the estimation error when first condition, queue length in Eq. (2) is known ( $N = l$ ), thus, the error is zero. For the second condition, first term is known ( $N_1 = l + 1$ ) and there is no error. The second term is  $\sim \text{Poisson}(\theta\delta)$  and the variance is  $\theta\delta$ . Thus, the conditional variance of Eq. (1) for  $L \geq 0, T \geq 0$  can simply be written as Eq. 3. Practically, given the time information and if there is a vehicle behind or not, we can calculate the estimation errors under the assumptions listed in 2.

$$E(N|L = l, T = t) = \begin{cases} l, & L = l_v \\ l + 1 + \theta\delta, & \delta = (R - t'), L \neq l_v \end{cases} \quad (2)$$

$$V(N|L = l, T = t) = \theta(R - t')I(L \neq l_v), L \geq 0, T \geq 0 \quad (3)$$

Formula for the errors can be expressed as deviations from  $D = N - E(N|L = l, T = t)$ . Thus, the variance of  $D$  for  $L \geq 0$  can be written as  $V(D) = E[V(N|L = l, T = t)]$  and given in Eq. (4). Suppose

that  $E(T)$  and  $E(T')$  are known (without and with range sensor, respectively), errors can be evaluated and simplified to Eq. 4.

$$V(D) = E[\theta\delta] = \theta[R - E(T')]P(L \neq l_v) = \frac{p(1 - e^{-p\lambda R(1+p)})}{(1+p)}, \quad t < R \quad (4)$$

where,  $P(L \neq l_v) = E(1 - e^{-\lambda(R-t)})$ ,  $t < R$  and  $P(L = l_v) = 1 - P(L \neq l_v)$  can be interpreted as  $L = l_v$  probability of last CV is the last vehicles meaning no arrival during  $\delta = (R-t)$  (i.e.,  $e^{-\lambda(R-t)}$ ). Using distribution of  $T$  (Comert (2013b))  $f(t) = \frac{\lambda p e^{-p\lambda(R-t)}}{(1-e^{-\lambda p R})}$ ,  $t < R$  is input for scenario probabilities  $P(L \neq l_v)$  as well as for the expected value of  $T'$  respectively. They can then be incorporated to compute the variance given in Eq. (4).

### 3.1. Derivation of the expectations

Revising the the probability distributions derived in Comert (2013b), queue length estimators and the errors are derived. Corresponding  $E(L)$  and  $E(L')$  revised for range sensor incorporation which are given in Eq. (5). The impact of sensor information is simply in the denominator adding one non-CV  $(1-p)$ .

$$\begin{aligned} E(L) &= \lambda R - [(1 - e^{-\lambda p R})/p], \quad m > 0 \\ E(L') &= \lambda R - [(1 - e^{-\lambda p R})/(p + (1 - p))], \quad m > 0 \end{aligned} \quad (5)$$

Similarly, in Eq. (6),  $E(T)$  and  $E(T')$  are provided. Impact is simple again by adding the term  $(1-p)\lambda$ . The revised variance of  $D$  ( $V(D)$ ) can be calculated using Eq. (6) in Eq. (4) and shown in Eq. (7). Note that the QLE error formula without sensors is  $V(D) = E[\theta\delta] = \theta[R - E(T)] = (1 - p)(1 - e^{-p\lambda R})/p$ .

$$\begin{aligned} E(T) &= R - [(1 - e^{-\lambda p R})/\lambda p], \quad m > 0 \\ E(T') &= R - [(1 - e^{-\lambda p R})/\lambda p + (1 - p)/\lambda], \quad m > 0 \end{aligned} \quad (6)$$

One of the main contribution of Comert (2013b) was to be able to compute  $E(L)$  and  $E(T)$  from one another, thus,  $E(T')$  can be utilized to calculate  $E(L')$  and shows the impact of range sensor in Fig. (2) which are derived using Eqs. (6) and (5). This error (Eq. 7) can be calculated using estimators for  $p$  and  $\lambda$  and signal timing information.

$$\begin{aligned} V(D) &= E[\theta\delta] = \theta[R - E(T')]P(L \neq l_v) \\ &= (1 - p)\lambda[R - (R - (1 - e^{-p\lambda R})/\lambda p + (1 - p)/\lambda)]P(L \neq l_v) \\ &= [1 - p][(1 - e^{-p\lambda R})/p - (1 - p)]P(L \neq l_v) \end{aligned} \quad (7)$$

where,  $P(L \neq l_v) = [1 - p(1 - e^{-(1+p)\lambda R})/(1 + p)]$ .

To verify the formulation above, the results are compared with simple point queue simulations with Poisson arrivals coded in R. From Fig. 2-a, it can be seen that  $E(T')$  is following the simulated values very closely. As  $p$  is approaching 100%, the  $E(T')$  is getting closer to the red duration (e.g., 45 seconds). In addition, Fig. 2-b shows the values of  $V(D)$ s computed with the analytical models with and without range measuring sensors. These are then compared against simulations with 100,000 replications. As it is evident analytical results follow simulation results very closely.

Figs. 3 a-b are given to illustrate how much improvement is achieved when CVs are also equipped with range measuring sensor, i.e., these graphs show the additional reduction in QLE error when  $N$  is estimated from CVs that also provide information about their following vehicles (as compared to CVs that provide only their own positions). In Fig. 3 a, this improvement is shown via graphing percent differences of  $V(D)/E(N)$ s (i.e., variance-to-mean ratio or dispersion index) with respect to probe percentages  $p$ .

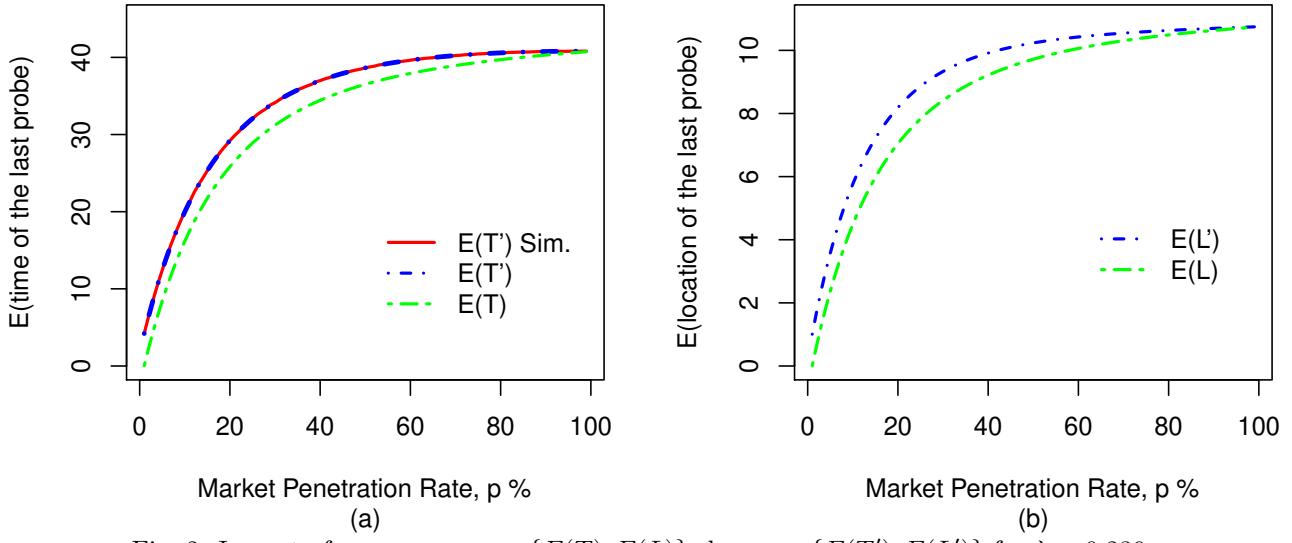


Fig. 2: Impact of range sensor on  $\{E(T), E(L)\}$  shown as  $\{E(T'), E(L')\}$  for  $\lambda = 0.239$  vps

Figure depicts maximum gain (e.g., close to 20% at  $\lambda = 0.133$  vps) at about  $p = 10\%$  and increasing as  $\lambda$  decreases. Figure suggests that about 20% probe vehicles are need to experience the maximum impact ( $\approx 10\%$ ). Lower  $p$  values are still showing 8 to 10% improvements relative to mean QLs. And, the impact is diminishing as  $p$  increases where  $P(L = lv)$  also get higher.

Figs. 4 a-b demonstrate the impact of range sensors in terms of % difference of  $\sqrt{V(D)}/E(N)$ s (i.e., coefficient of variations) and difference of  $\sqrt{V(D)}$ s changing as functions of market penetration rate  $p$  at different  $\lambda$  values. The maximum improvement (e.g., close to 7% for  $\lambda = 0.133$  vps) at about  $p = 30\%$  and increasing as  $\lambda$  decreases. Similarly, estimation error differences are higher for lower  $\lambda$  values. They get very close to each other after about 30%  $p$  level. Difference of  $\sqrt{V(D)}$  values are given in Fig. 4 b that are QLE error differences in terms of number of vehicles. Again, maximum impact 0.35 vehicle per cycle is observed at  $p \approx 30\%$ . This behavior would make such estimator more suitable for intersections with low demand or rural intersections. Desirably, this would also satisfy the assumptions for the derivations.

#### 4. Analysis for the Case with Overflow Queue

When the overflow queue from signal cycles is incorporated in the queue length estimation, we can express queue length  $N$  as the sum of the overflow queue  $Q$  and the new red duration arrivals  $A$ . Then, for any cycle, based on the last CV information (e.g., CV in  $Q$  or  $A$ ) and whether it is the last vehicle ( $L = lv$ ) (i.e., detected by range sensor), we can model the estimator with the help of five possible scenarios:

- (i) The last CV is within new arrivals and it is the last vehicle,  $I(l \in A \wedge L = lv)$ .
- (ii) The last CV is within new arrivals and it is not the last vehicle,  $I(l \in A \wedge L \neq lv)$ .
- (iii) The last CV is in the overflow queue and it is the last vehicle,  $I(l \in Q \wedge L = lv)$ .
- (iv) The last probe may be within the overflow queue and it is not the last vehicle,  $I(l \in Q \wedge L \neq lv)$ .
- (v) There is no CV in the queue,  $I(l = 0)$ .

Considering these scenarios, we can write  $N|L, T = I(L = lv)[Q|L, T + A|L, T] + I(L \neq lv)[Q|L, T + A|L, T]$ . Then, the estimator for the total queue length given CV information can be written as Eq. (8). When utilized, Fig. 5 was given for the average estimation errors  $V(D)$ s from with and without range

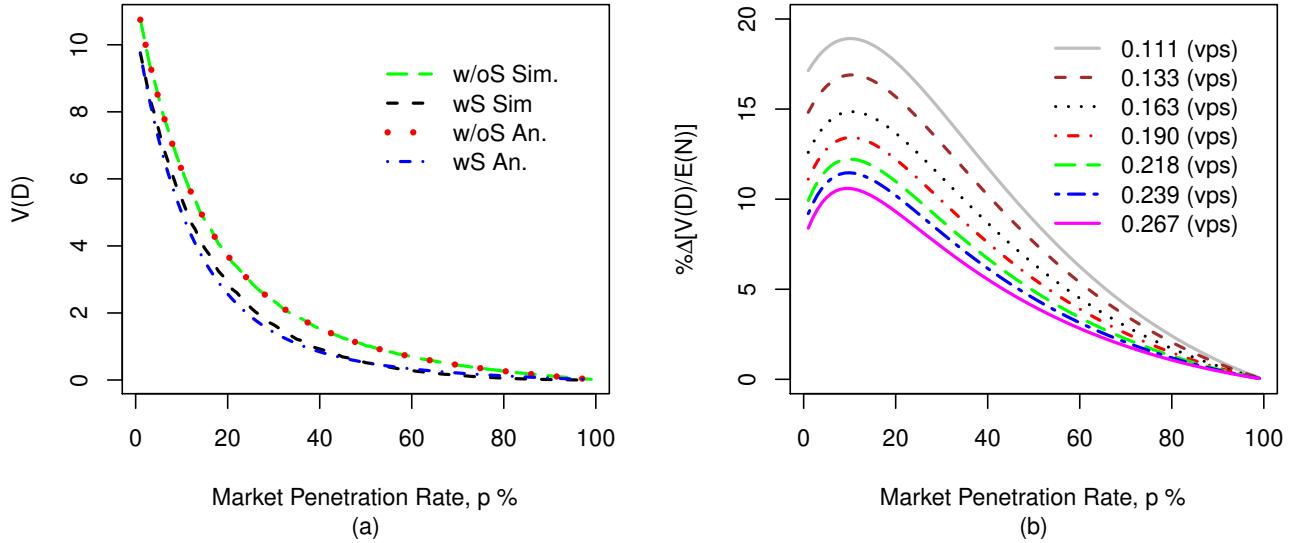


Fig. 3: Improvements in QLE errors when each CV also provides its following vehicle's position in the queue (various  $\lambda$  levels)

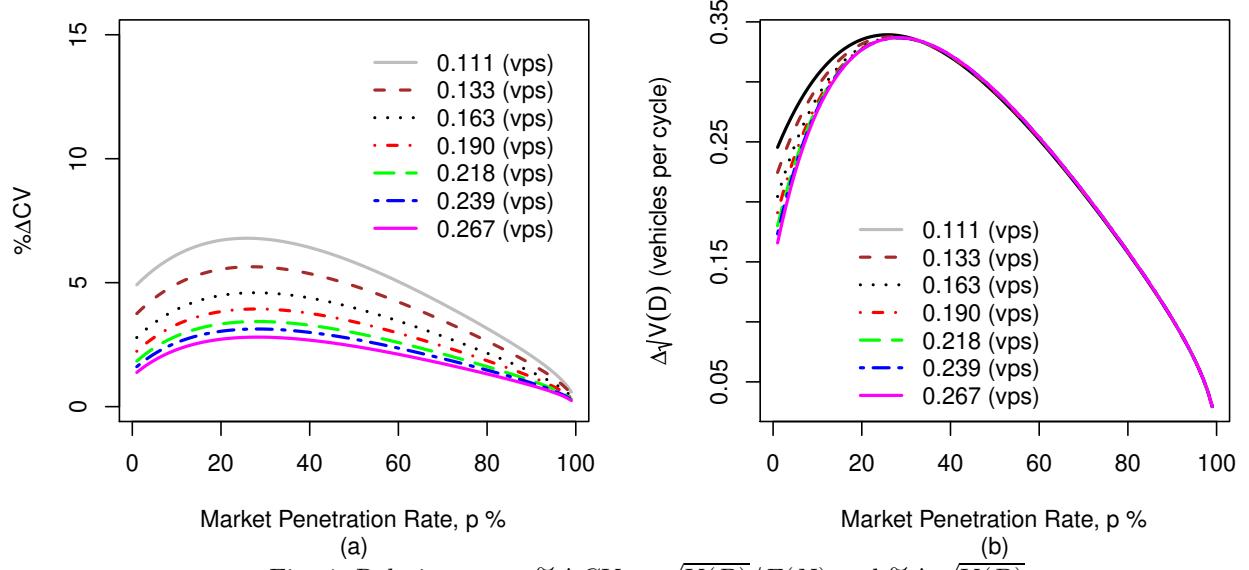


Fig. 4: Relative errors  $\% \Delta CV = \sqrt{V(D)}/E(N)$  and  $\% \Delta \sqrt{V(D)}$

sensors. From the figure, similar to  $Q = 0$  case, improvements with sensors are higher in between  $p=0.05 - 30\%$ . The estimator with overflow queue also contains a revised queue joining time  $\tau$  that defined as the remaining last CV's joining time from previous green duration. These observations are desired and expected. Benefit at lower CV penetration levels would be more cost saving. As  $p$  gets higher, it is expected that more CVs will be the last vehicle in the queue, thus, saving would be less from the

range sensor fusion.

$$E(N|L = l, L = lv, T = t) = \begin{cases} I(l \in Q \wedge L = lv)[l + \theta R] + \\ I(l \in Q \wedge L \neq lv)[l + 1 + \theta(C - \tau') + \theta R] + \\ I(l \in A \wedge L = lv)[l] + \\ I(l \in A \wedge L \neq lv)[l + 1 + \theta(R - \tau')] + \\ I(l = 0)[(1 - p)(E(Q) + \theta R)]. \end{cases} \quad (8)$$

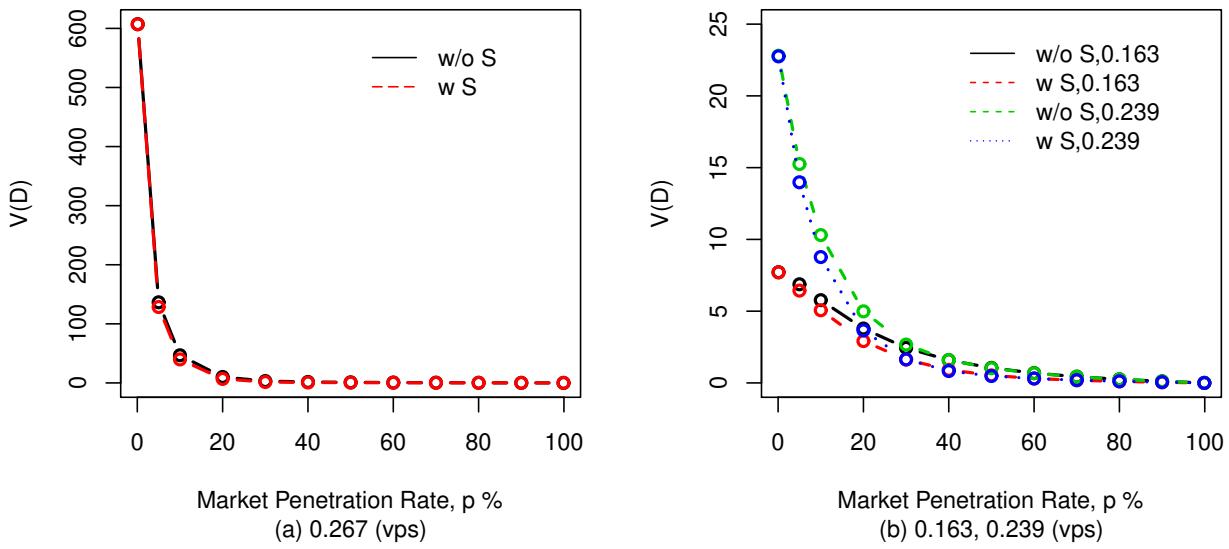


Fig. 5: Estimation errors from VISSIM with and without  $Q$  results for different  $\lambda$  and  $p$  values

Average overflow queue in Eq. (8),  $E(Q) = \frac{\rho^2}{2(1-\rho)}$ , is adopted from (Akçelik (1980)) where  $\rho = \lambda C/X$  with  $X = 24$  vehicles per cycle in our examples,  $\lambda$  arrival rate per second, and  $C$  is cycle time 90 seconds. Alternatively, from (Viti (2006); Medhi (1991)) different average overflow queue formula can be adopted  $E(Q) = 1.5(\rho - \rho_o)/(1 - \rho)$ ,  $E(Q) = \frac{(2\rho - 1)\rho^4}{2(1 - \rho)}$ , and  $E(Q) = [\frac{\rho}{2(1 - \rho)}]e^{-[\frac{(1 - \rho)\sqrt{R/2 - R(1 - \rho)^2}}{4}]}$  where  $\rho_o = 0.67 + X/600$ . The queue length on the approach lane is assumed to be zero for all examples in the study. Using the scenario probabilities in Eq. (9), we can obtain Eq. (8).

$$E(N) = \begin{cases} P(l \in Q \wedge L \neq lv)[E(L) + \theta(C - E(\tau')) + \theta R] + \\ P(l \in Q \wedge L = lv)[E(L) + \theta R] + \\ P(l \in A \wedge L \neq lv)[E(L) + 1 + \theta(R - \tau)] + \\ P(l \in A \wedge L = lv)[E(L)] + \\ P(l = 0)[(1 - p)(E(Q) + \theta R)] \end{cases} \quad (9)$$

Average or steady-state estimation error,  $V(D)$ s, can be given as in Eq. (10) with  $V(Q)$  in Eq. (12) from (Medhi (1991)).

$$V(D) = \begin{cases} P(l \in Q \wedge L \neq lv)[\theta(C - E(\tau')) + \theta R] + \\ P(l \in Q \wedge L = lv)[\theta R] + \\ P(l \in A \wedge L \neq lv)[(1 - p)(1 - e^{-(1+p)\lambda R})/(1 + p)] + \\ P(l = 0)[(1 - p)(V(Q) + \theta R)] \end{cases} \quad (10)$$

Probability distributions of  $Q$  and  $\tau$  are needed to calculate Eq. (10). These distributions can be obtained numerically (Olszewski (1990); Viti and Van Zuylen (2010)) or by solving the probability generating function (Gray et al. (1992); Tarko and Routhail (1994); Mung et al. (1996); Van Den Broek et al. (2006)). But, no exact simple form for  $V(D)$  is available. Thus, in Section 5 we used approximations in the estimators and the errors and evaluated them with microsimulation results. Figs. 6 a-b are derived using the approximation formulae and microsimulations using Eq. (14). Details of the approximate estimator is explained next by utilizing steady-state  $V(Q)$ s by the formulae adopted from (Medhi (1991); Fu and Hellinga (2000)). The approximate  $V(D)$ s are also given against VISSIM results in Figs. 5-10. Relative errors are expressed in coefficient of variations (i.e.,  $100\sqrt{V(D)}/E(N)$  or  $\% \Delta CV$ ). Figures also given as  $\% \Delta \sqrt{V(D)}$  from the two evaluations.

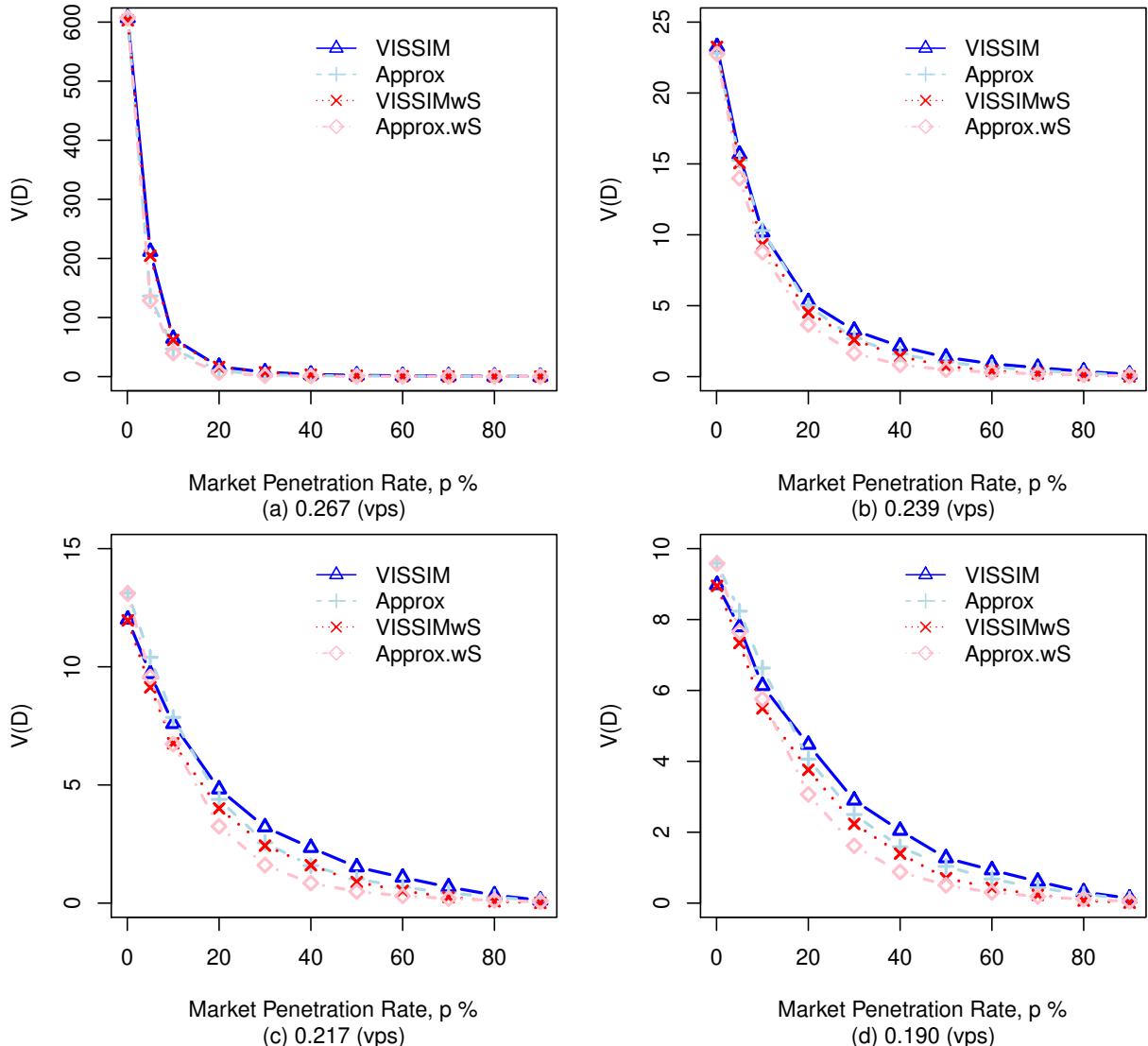


Fig. 6: Comparing estimation errors obtained from VISSIM and approximations

## 5. Approximate Inference for $E(N)$ and $V(D)$

Approximation formulas for the estimators can be derived simply quantifying the impact of overflow queue in terms of  $E(Q)$  and  $V(Q)$  in Eq. (11) and Eq. (12), respectively. Note that these formulas assume  $M/G/1$  queue and they are based on the Pollaczek-Khintchine formula (Haight (1959); Heidemann

(1994); Ohno (1978)) and adopted from (Medhi (1991)). Numerical values are determined and discussed in the simulation set up (Comert (2016)), mean discharge rate from the intersection is determined as 1.8 seconds per vehicle with the capacity for effective green of 43.20 seconds ((43.20/1.80)=24 vehicles per cycle). Demand levels of {600, 700, 800, 900, 985} vph that corresponds to volume-to-capacity ratios of  $\rho=\{0.60, 0.70, 0.80, 0.88, 0.98\}$  are used in the numerical examples. VISSIM is run 1000 cycles (25 hours) for 3 random seeds for 11  $p = \{0.1\%, 5\%, 10\%, \dots, 90\%\}$  CV percentages at each demand level. Average of 3 random seeds are reported for estimation errors. Default queue definition of VISSIM in terms of minimum speed  $\leq 10$  kilometers per hour is used. For detailed discussions, readers are referred to (Akçelik (1980); Medhi (1991); Fu and Hellinga (2000); Viti (2006)).

Note that we are presenting the steady-state QLE errors incorporating overflow queue moments, however, cycle-to-cycle formulas were also given  $E(Q_i)$  and results are given as figures in Section 6.  $E(Q_i)$  and  $V(Q_i)$  are adopted from (Akçelik (1980)) and (Viti (2006)), respectively.

$$E(Q) = \frac{\rho^2}{2(1 - \rho)} \quad (11)$$

$$V(Q) = [4(1 - \rho)\rho^3 + 3\rho^4]/(12(1 - \rho)^2) \quad (12)$$

Simply, queue length estimation approximations for expected value and errors are obtained by incorporating Eqs. (11) and (12) into appropriate cases in Eqs. (8) and (10). Resulting approximations with range sensors are then expressed as Eqs. (13) and (14). Scenario probabilities can be simply approximated by quadratic equations.

$$E(N) = \begin{cases} P(l \in Q)[E(Q) + \theta R] + \\ P(l \in A)[E(Q) + \lambda R] + \\ I(l = 0)[(1 - p)(E(Q) + \theta R)] \end{cases} \quad (13)$$

$$V(D) = \begin{cases} P(l \in Q)[(1 - p)(1 - e^{-pE(Q)})/p + \theta R] + \\ P(l \in A)[(1 - p)(1 - e^{-p\lambda R})/p] + \\ P(l = 0)[(1 - p)(V(Q) + \theta R)] \end{cases} \quad (14)$$

where, approximate probabilities are given in Eq. 15.

$$P(L = l) = \begin{cases} e^{-p\lambda R} - e^{-p(\lambda R + (9.87p^2 - 4.62p + 0.991)E(Q))}, L \in Q \\ 1 - e^{-p\lambda R}, L \in A \\ e^{-p(\lambda R + (9.87p^2 - 4.62p + 0.991)E(Q))}, L = 0 \end{cases} \quad (15)$$

$$V(D) = \begin{cases} P(l \in Q)[(1 - p)p(1 - e^{-(1+p)E(Q)})/(1 + p) - p\theta R] + \\ P(l \in A)[(1 - p)(1 - e^{-p\lambda R})/p - (1 - p)][1 - p(1 - e^{-(1+p)\lambda R})/(1 + p)] + \\ P(l = 0)[(1 - p)(V(Q) + \theta R)] \end{cases} \quad (16)$$

Performance of the approximations are presented in Figs. 6 a-d. It can be seen that approximate errors are following VISSIM values closely for both with and without range sensors at all  $\rho$  and  $p$  levels. Moreover, Fig. 7 presents linear regression lines without intercepts for  $V(D)$ s from VISSIM simulations and approximate models without and with range sensor. Regression lines are given for estimated and simulation results. The lines show high correlations between analytical approximations and VISSIM errors without and with range sensor at  $R^2=0.986$  and  $R^2=0.985$ , respectively.

Fig. 8-a shows Eq. (14) values along with VISSIM QLs for various  $\rho$  levels. Developed approximation for QL is a simple formula that needs  $\rho$ ,  $\lambda$ , and signal timing. Fig. 8 b shows the evolution of errors with respect to cycles for  $p \leq 0.05$  and  $\rho \geq 0.80$ . These errors are calculated from microsimulations. We can see that approximate errors are able to match simulation values even at high  $\rho \geq 0.80$  and low probe proportion  $p \leq 0.05$  values. When we consider the error differences with respect the mean QLs, they are very close.

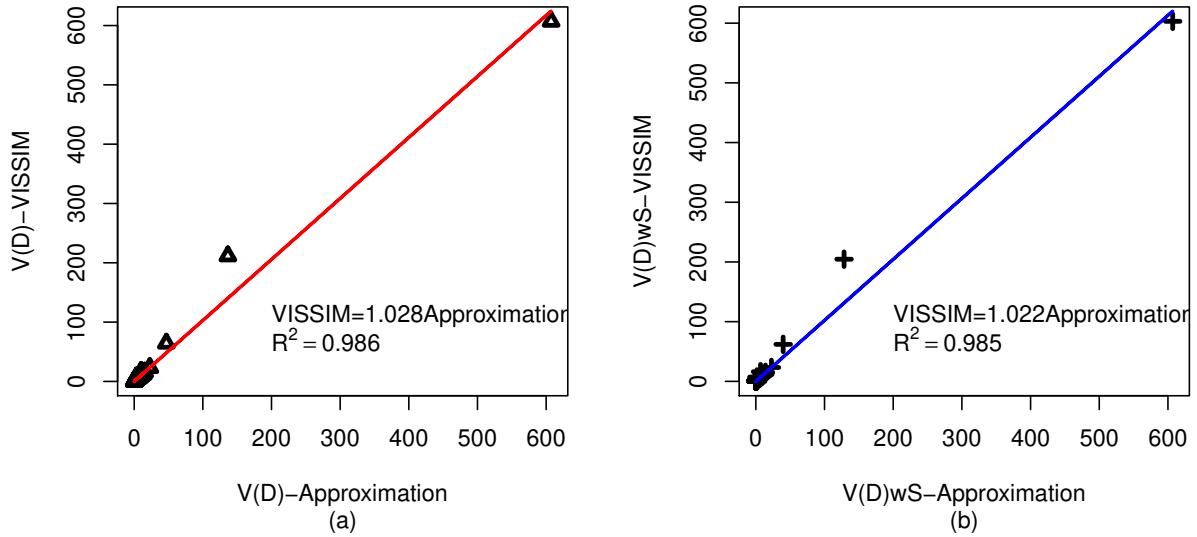


Fig. 7: Comparing estimation errors obtained from VISSIM and approximations

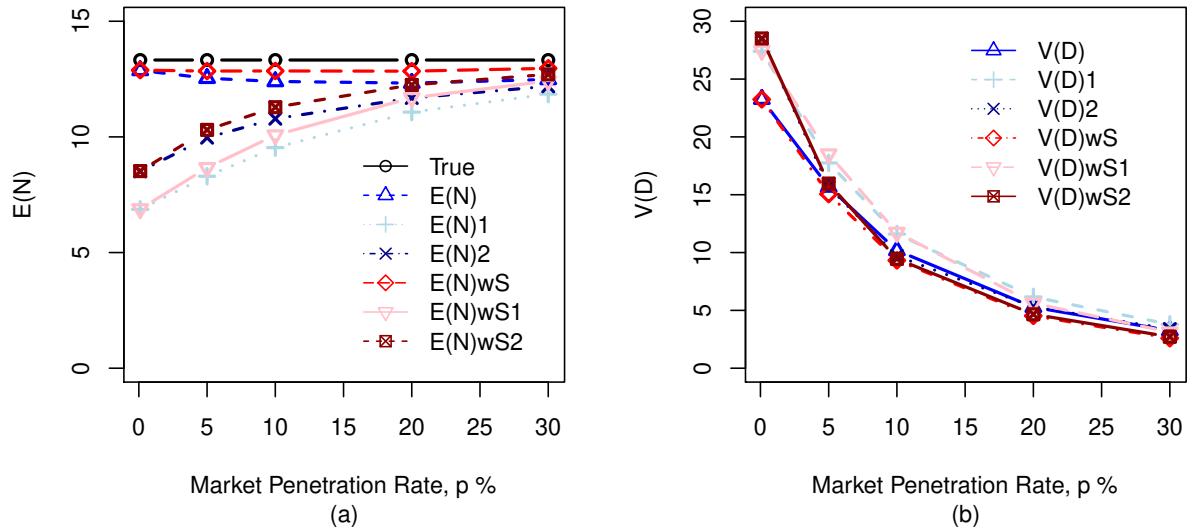


Fig. 8: Estimation errors with and without results for different  $\lambda$   $vps$  and  $ps$

## 6. Estimation with Range Sensor, Unknown $\lambda$ , and $p$

Known arrival  $\lambda$  and market penetration rate  $p$  parameters in Eqs. (2) and (8) can be relaxed by replacing them with estimators for these low level parameters. Detailed derivations of such estimators were given in (Comert (2016)). Best performing combinations were adopted from this earlier study. Without  $Q$ , the QL with unknown arrival rate and probe proportion can simply be estimated using Eq. (17). For  $E(N|l, t, m, Q = 0)$ , referred as *Estimator 1* with combination of  $\hat{p}_1 = m/l$  and  $\hat{\lambda}_1 = l/R$  simply Eq. (18) is derived. *Estimator 2* uses  $\hat{p}_2 = \frac{mt}{(mt+(l-m)R)}$  and  $\hat{\lambda}_2 = \frac{l-m}{t} + \frac{m}{R}$  which simplifies to Eq. (19).

$$E(N|l, t, m, Q = 0) = l + (1 - \hat{p})\hat{\lambda}(R - t) \quad (17)$$

$$E(N|l, t, m, Q = 0) = l + (l - m)(1 - t/R) \quad (18)$$

$$E(N|l, t, m, Q = 0) = l + (1 - \frac{mt}{mt + (l - m)R})((l - m)/t + m/R)(R - t) \quad (19)$$

Similarly, estimation of queues real-time and steady-state including  $Q$  are also presented. In this case, QLs can be considered as sum of random variables given in Eq. (20). Indicator function  $I(l \in Q)$  represents the case if the last CV is in the overflow queue,  $I(l \in A)$  indicates the last CV in new red arrivals, and  $I(l = 0)$  indicates that there is no CV in the queue. At any given cycle, these terms represent disjoint events, hence, only one term will be positive where others will be zeros.

$$E(N_i|L = l, T = t, Q_i \geq 0) = \begin{cases} I(l \in Q)[l + \hat{\theta}(C - t') + \hat{\theta}R] + \\ I(l \in A)[l + \hat{\theta}\delta] + \\ I(l = 0)[(1 - \hat{p})(E(Q_i) + \hat{\theta}R)] \end{cases} \quad (20)$$

$$E(N_i|L = l, L = lv, T = t, Q_i \geq 0) = \begin{cases} I(l \in Q \wedge L = lv)[l + \hat{\theta}R] + \\ I(l \in Q \wedge L \neq lv)[l + 1 + \hat{\theta}(C - \tau') + \hat{\theta}R] + \\ I(l \in A \wedge L = lv)[l] + \\ I(l \in A \wedge L \neq lv)[l + 1 + \hat{\theta}(R - t')] + \\ I(l = 0)[(1 - \hat{p})(E(Q_i) + \hat{\theta}R)]. \end{cases} \quad (21)$$

Overflow queue  $E(Q_i) = \frac{Xi(\hat{\rho}-1)}{4} \sqrt{(\hat{\rho}-1)^2 + \frac{12(\hat{\rho}-\rho_o)}{Xi}}$  is adopted from (Akçelik (1980)) where  $\rho_o=0.67+X/600$ ,  $X = 24$  vehicles per cycle,  $\hat{\rho} = \hat{\lambda}C/X$ , and  $i = 1, 2, 3\dots$  denotes the cycle index. For  $E(Q_i) = E(Q)(1 - e^{-\beta i})$  from (Viti (2006)) can also be used which gives very close results (Comert (2013b)) where  $E(Q) = \frac{3(\hat{\rho}-\rho_o)}{2(1-\hat{\rho})}$ .

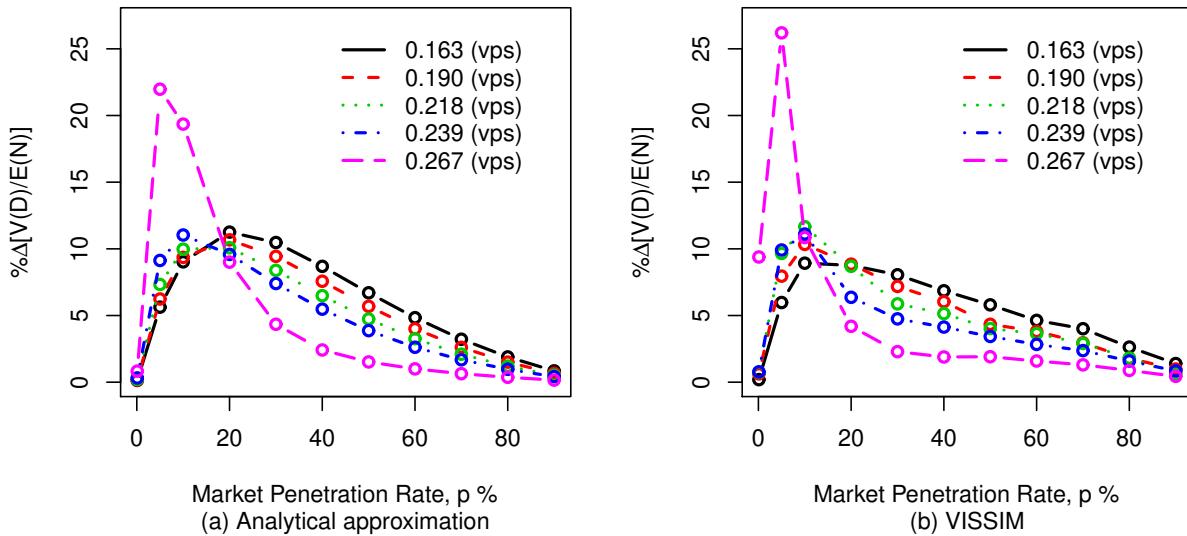


Fig. 9: Estimation errors  $\%ΔV(D)/E(N)$  with and without range sensor results for different  $\lambda$  vps and  $ps$

Average true and estimated QLs derived from VISSIM runs. Note that QLs depend on VISSIM queue definition where default *in queue*= $I(speed \leq 10$  kilometers per hour). From Fig. 10, at 600 vehicles per hour, 7.34 vehicles per 45 seconds red duration is expected. Similarly others 8.57, 9.79, 10.78, and 12.01 vehicles per red. However, overflow queue becomes more important when  $\rho > 0.80$  on average 0.68, 2.13,

and 20.68 vehicles in length for  $\rho = 0.80$ ,  $\rho = 0.88$ , and  $\rho = 0.98$ , respectively. Impact of overflow queue also observed accuracy of the estimators start to decrease after  $\rho > 0.88$  from  $\pm 1.5$  to  $\pm 3.45$  vehicles. Overall, *Estimator 2* can provide accuracy between 12% to 19% of the true average QLs.

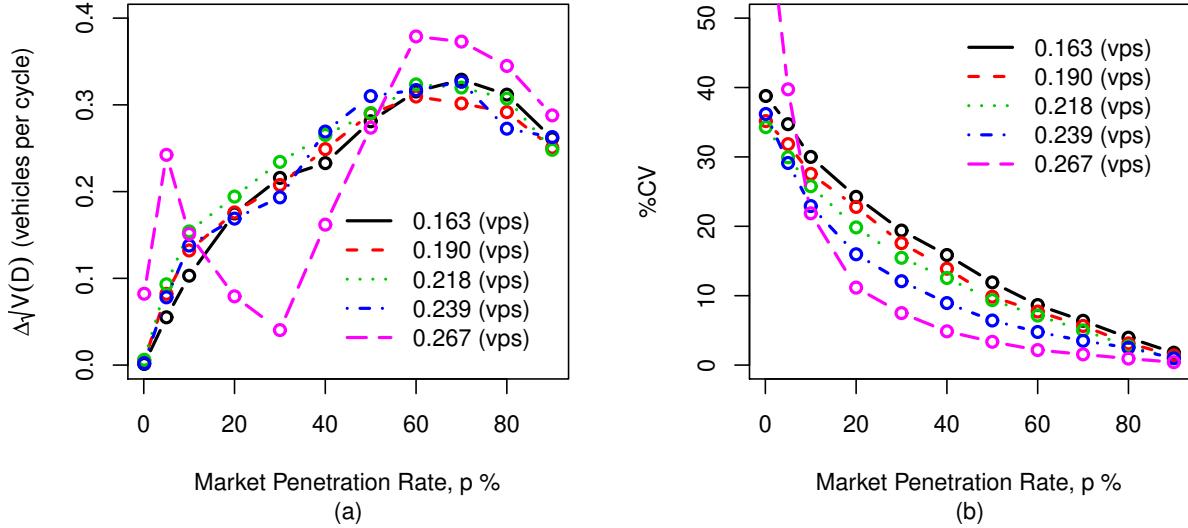


Fig. 10: Relative estimation errors  $\% \Delta \sqrt{V(D)}$  and  $\% \sqrt{V(D)}/E(N)$  from VISSIM simulations

With overflow queue combination of parameter estimators were input where results are shown in Fig. 8. As queue length estimators closely follow true queue lengths by  $p=30\%$ , queue length estimators with  $\lambda$  and  $p$  estimators also gets very close to the behavior of estimator with known parameters.

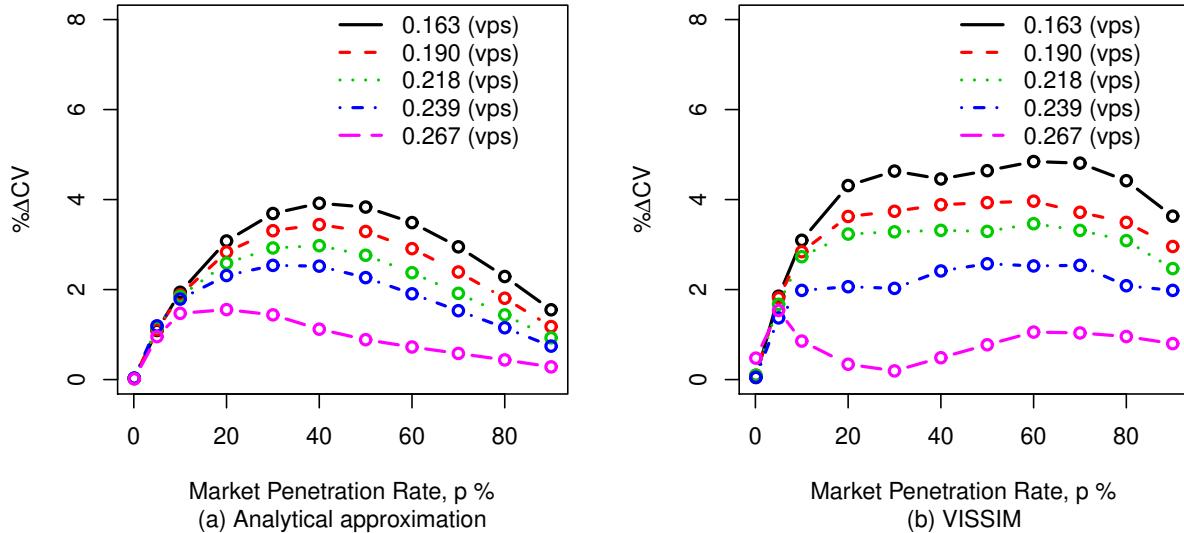


Fig. 11: Estimation errors  $\% \Delta \sqrt{V(D)}/E(N)$  from VISSIM and approximations

$$E(N_i|l, t, m, Q_i = 0) = l + (l - m/l)(l/R)(R - t) = l + (l - m)(1 - t/R) \quad (22)$$

$$E(N_i|l, t, m, Q_i = 0) = l + \left(1 - \frac{mt}{mt + (l - m)R}\right)((l - m)/t + m/R)(R - t) = m + \frac{R(l - m)}{t} \quad (23)$$

Figs. 9-11 summarize the QLE errors with the overflow queue. The errors are expressed in  $\% \Delta V(D)$  and  $\% \Delta CV$  where,  $CV = \sqrt{V(D)}/E(N)$ . Since estimators are able to capture overflow queue, errors are declining to zero as connected vehicle MPR increases. Results from the literature cited above also confirmed that  $p \geq 30\%$  queue length can be predicted within  $\pm 10\%$  across all all  $\rho$ levels. Similarly at  $p = 50\%$ , queue length can be predicted within  $\pm 2\%$  in  $\% \Delta CV$  and  $\% \Delta EN$ .

Fig. 12 confirms these results with good agreement with the true QLs. The figures depict about 20 cycles for illustration of performance of estimators. Based on the numerical results, for  $p \geq 30\%$  QLEs with unknown estimators shows good agreement with true average QLs. Thus, any of the *Estimator 1* or *Estimator 2* can be used in accurate queue length estimation in signal control. Estimator with range sensor can also be seen that more closely following the true queue values.

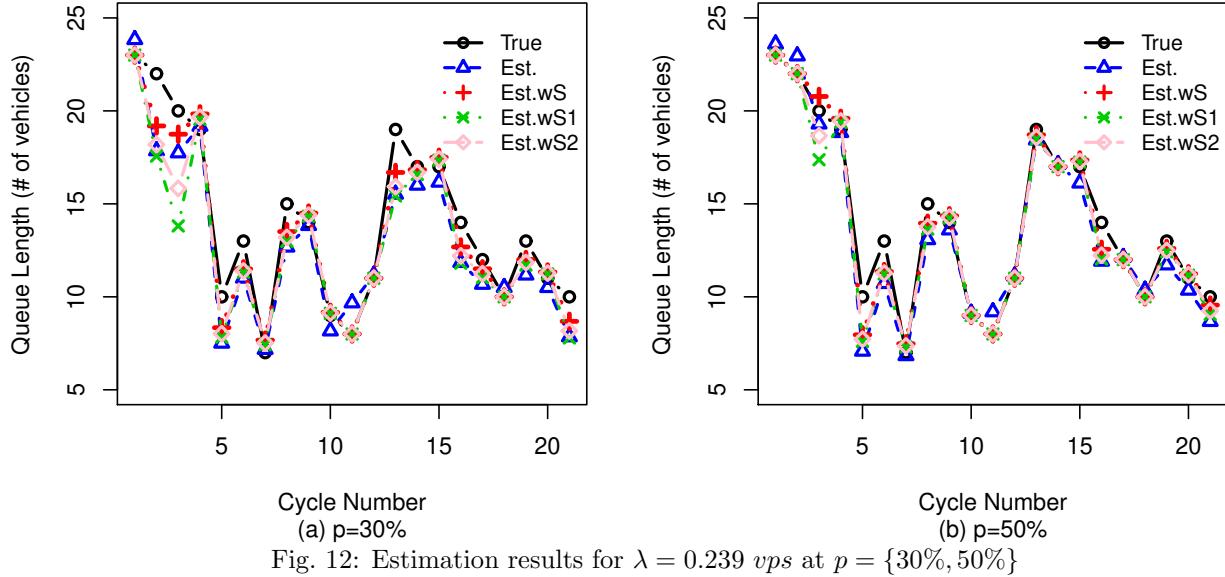


Fig. 12: Estimation results for  $\lambda = 0.239$  vps at  $p = \{30\%, 50\%\}$

## 7. Conclusions

In this study, models for queue length estimation from connected vehicles with range sensors at signalized intersections are derived. Fully analytical models without overflow queue and approximations with the overflow queue are presented. Contribution of the proposed models are simple input-output models without requiring true queue length observations, as a step towards generalizing estimation problem with partially observed entities in a cycling queueing system, and leverage of simple variations to previously derived models to express additional complexity.

When range sensor included, estimators are more accurate as high as 25% in estimation errors relative to average queue lengths for high volume-to-capacity ratio  $\rho=0.98$  at market penetration ratio of  $p = 5\%$ . About 10% improvement is observed when volume-to-capacity ratio is  $\leq 0.88$ . Moreover, approximately 0.25 vehicle per cycle was improved. For example, at constant arrival rate in 10 hours, total estimation error of 100 vehicles (i.e., energy consumption, emissions, and delay) would be improved.

## Acknowledgments

This study is supported by the Center for Connected Multimodal Mobility ( $C^2M^2$ ) (USDOT Tier 1 University Transportation Center) Grant headquartered at Clemson University, Clemson, South Carolina, USA. The authors would also like to acknowledge U.S. Department of Homeland Security (DHS)

Summer Research Team Program Follow-On, FY19 US Department of Education MSEIP Grant Award P120A190061, and National Science Foundation (NSF, No. 1719501, 1436222, and 1400991) grants. Any opinions, findings, conclusions or recommendations expressed in this study are those of the author(s) and do not necessarily reflect the views of ( $C^2M^2$ ), USDOT, DHS, U.S. Department of Education, or NSF and the U.S. Government assumes no liability for the contents or use thereof.

## References

Akçelik, R., 1980. Time-dependent expressions for delay, stop rate and queue length at traffic signals. Australian Road Research Board Melbourne, Australia.

Argote, J., Christofa, E., Xuan, Y., Skabardonis, A., 2011. Estimation of measures of effectiveness based on connected vehicle data. In: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1767–1772.

Asanjarani, A., Nazarathy, Y., Pollett, P. K., 2017. Parameter and state estimation in queues and related stochastic models: A bibliography. arXiv preprint arXiv:1701.08338.

Cetin, M., 2012. Estimating queue dynamics at signalized intersections from probe vehicle data: Methodology based on kinematic wave model. *Transportation research record* 2315 (1), 164–172.

Cetin, M., Sazara, C., Nezafat, R. V., et al., 2017. Exploring the use of lidar data from autonomous cars for estimating traffic flow parameters and vehicle trajectories.

Cheng, Y., Qin, X., Jin, J., Ran, B., 2012. An exploratory shockwave approach to estimating queue length using probe trajectories. *Journal of intelligent transportation systems* 16 (1), 12–23.

Comert, G., 2013a. Effect of stop line detection in queue length estimation at traffic signals from probe vehicles data. *European Journal of Operational Research* 226 (1), 67–76.

Comert, G., 2013b. Simple analytical models for estimating the queue lengths from probe vehicles at traffic signals. *Transportation Research Part B: Methodological* 55, 59–74.

Comert, G., 2016. Queue length estimation from probe vehicles at isolated intersections: Estimators for primary parameters. *European Journal of Operational Research* 252 (2), 502–521.

Comert, G., Cetin, M., 2009. Queue length estimation from probe vehicle location and the impacts of sample size. *European Journal of Operational Research* 197 (1), 196–202.

Comert, G., Cetin, M., 2011. Analytical evaluation of the error in queue length estimation at traffic signals from probe vehicle data. *IEEE Transactions on Intelligent Transportation Systems* 12 (2), 563–573.

Comert, G., Cetin, M., Nichols, A. P., 2009. Incorporating queue length measurements into actuated signal control: evaluation of efficiency benefits at an intersection. Tech. rep.

Comert, G., Khan, Z., Rahman, M., Chowdhury, M., 2019. Grey models for short-term queue length predictions for adaptive traffic signal control. arXiv preprint arXiv:1912.12676.

Fiems, D., De Turck, K., 2019. Taylor-series approximations for queues with arrival correlation. *Applied Mathematical Modelling* 69, 113–126.

Fu, L., Hellinga, B., 2000. Delay variability at signalized intersections. *Transportation Research Record* 1710, 215–221.

Gao, K., Huang, S., Han, F., Li, S., Wu, W., Du, R., 2020. An integrated algorithm for intersection queue length estimation based on iot in a mixed traffic scenario. *Applied Sciences* 10 (6), 2078.

Gray, W. J., Wang, P., Scott, M., 1992. An m/g/1-type queuing model with service times depending on queue length. *Applied mathematical modelling* 16 (12), 652–658.

Guo, Q., Li, L., Ban, X. J., 2019. Urban traffic signal control with connected and automated vehicles: A survey. *Transportation research part C: emerging technologies*.

Haight, F. A., 1959. Overflow at a traffic flow. *Biometrika* 46 (3-4), 420–424.

Heidemann, D., 1994. Queue length and delay distributions at signalized intersections. *Transportation Research Part B* 28 (5), 624–635.

Li, J.-Q., Zhou, K., Shladover, S. E., Skabardonis, A., 2013. Estimating queue length under connected vehicle technology: Using probe vehicle, loop detector, and fused data. *Transportation research record* 2356 (1), 17–22.

Liu, H., Liang, W., Rai, L., Teng, K., Wang, S., 2019. A real-time queue length estimation method based on probe vehicles in cv environment. *IEEE Access* 7, 20825–20839.

Mandal, A., Sadhukhan, P., Gaji, F., Sharma, P., 2020. Measuring real-time road traffic queue length: A reliable approach using ultrasonic sensor. In: *Proceedings of the 2nd International Conference on Communication, Devices and Computing*. Springer, pp. 391–398.

Medhi, J., 1991. Stochastic models in queuing theory. Academic Press, Inc. Harcourt Brace Jovanovich Publishers.

Mei, Y., Gu, W., Chung, E. C., Li, F., Tang, K., 2019. A bayesian approach for estimating vehicle queue lengths at signalized intersections using probe vehicle data. *Transportation Research Part C: Emerging Technologies* 109, 233–249.

Mung, K. S. G., Poon, A. C. K., Lam, W. H. K., 1996. Distributions of queue lengths at fixed time traffic signals. *Transportation Research Part B:Methodological* 30 (6), 421–439.

Nam, S. Y., Kim, S., Sung, D. K., 2009. Estimation of available bandwidth for an m/g/1 queueing system. *Applied mathematical modelling* 33 (8), 3299–3308.

Newell, G. F., 1982. Applications of queueing theory, 2nd Edition.

Ohno, K., 1978. Computational algorithm for a fixed cycle traffic signal and new approximate expressions for average delay. *Transportation Science* 12 (1), 29–47.

Olszewski, P., 1990. Modeling of queue probability distribution at traffic signals. In: M.Koshi (Ed.), *Transportation and Traffic Theory*. Elsevier, Amsterdam.

Rompis, S. Y., Cetin, M., Habtemichael, F., 2018. Probe vehicle lane identification for queue length estimation at intersections. *Journal of Intelligent Transportation Systems* 22 (1), 10–25.

Tan, C., Liu, L., Wu, H., Cao, Y., Tang, K., 2020. Fuzing license plate recognition data and vehicle trajectory data for lane-based queue length estimation at signalized intersections. *Journal of Intelligent Transportation Systems*, 1–18.

Tan, C., Yao, J., Tang, K., Sun, J., 2019. Cycle-based queue length estimation for signalized intersections using sparse vehicle trajectory data. *IEEE Transactions on Intelligent Transportation Systems*.

Tarko, A., Roushail, N., 1994. Distribution-free model for estimating random queues in signalized networks. *Transportation Research Record* 1457, 192–197.

Tiaprasert, K., Zhang, Y., Wang, X. B., Zeng, X., 2015. Queue length estimation using connected vehicle technology for adaptive signal control. *IEEE Transactions on Intelligent Transportation Systems* 16, 2129–2140.

University of Arizona, University of California PATH Program, S. N. I. E., 2016. Multi-modal intelligent traffic signal system—phase ii: System development, deployment and field test.

Van Den Broek, M., Van Leeuwaarden, J., Adan, I., Boxma, O., 2006. Bounds and approximations for the fixed-cycle traffic-light queue. *Transportation science* 40 (4), 484–496.

Van Phu, C. N., Farhi, N., 2018. Estimation of urban traffic state with probe vehicles. *arXiv preprint arXiv:1811.05394*.

Viti, F., 2006. The dynamics and the uncertainty of delays at signals. Ph.D. thesis, Delft University of Technology, Delft, the Netherlands.

Viti, F., Van Zuylen, H., 2010. Probabilistic models for queues at fixed control signals. *Transportation Research Part B* 44 (1), 120–135.

Walraevens, J., Bruneel, H., Fiems, D., Wittevrongel, S., 2017. Delay analysis of multiclass queues with correlated train arrivals and a hybrid priority/fifo scheduling discipline. *Applied Mathematical Modelling* 45, 823–839.

Wu, A., Yang, X., 2013. Real-time queue length estimation of signalized intersections based on rfid data. *Procedia-Social and Behavioral Sciences* 96, 1477–1484.

Xu, H., Ding, J., Zhang, Y., Hu, J., 2017. Queue length estimation at isolated intersections based on intelligent vehicle infrastructure cooperation systems. In: 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 655–660.

Yang, K., Menendez, M., 2018. Queue estimation in a connected vehicle environment: A convex approach. *IEEE Transactions on Intelligent Transportation Systems* 20 (7), 2480–2496.

Yin, J., Sun, J., Tang, K., 2018. A kalman filter-based queue length estimation method with low-penetration mobile sensor data at signalized intersections. *Transportation Research Record* 2672 (45), 253–264.

Zhao, Y., Zheng, J., Wong, W., Wang, X., Meng, Y., Liu, H. X., 2019a. Estimation of queue lengths, probe vehicle penetration rates, and traffic volumes at signalized intersections using probe vehicle trajectories. *Transportation Research Record* 2673 (11), 660–670.

Zhao, Y., Zheng, J., Wong, W., Wang, X., Meng, Y., Liu, H. X., 2019b. Various methods for queue length and traffic volume estimation using probe vehicle trajectories. *Transportation Research Part C: Emerging Technologies* 107, 70–91.