Enhanced Traffic Signal Operation using Connected Vehicle Data

by:

Ehsan Bagheri

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Tony Qiu Associate Professor
Supervisor(s)	Dr. Bruce Hellinga Professor
Internal Member	Dr. Liping Fu Professor
Internal Member	Dr. Carl Haas Professor
Internal-external Member	Dr. Clarence Woudsma Associate Professor

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Chapter 3 of this thesis contains materials from a published journal paper co-authored by myself, my supervisor, Dr. Bruce Hellinga, and a post-doctoral fellow, Dr. Babak Mehran; and a conference paper co-authored by myself, and my supervisor, Dr. Bruce Hellinga. I developed and documented the methodology and Dr. Hellinga and Dr. Mehran assisted me in writing of the papers.

Abstract

As traffic on urban road network increases, congestion and delays are becoming more severe. At grade intersections form capacity bottlenecks in urban road networks because at these locations, capacity must be shared by competing traffic movements. Traffic signals are the most common method by which the right of way is dynamically allocated to conflicting movements.

A range of traffic signal control strategies exist including fixed time control, actuated control, and adaptive traffic signal control (ATSC). ATSC relies on traffic sensors to estimate inputs such as traffic demands, queue lengths, etc. and then dynamically adjusts signal timings with the objective to minimize delays and stops at the intersection.

Despite, the advantages of these ATSC systems, one of the barriers limiting greater use of these systems is the large number of traffic sensors required to provide the essential information for their signal timing optimization methodologies.

A recently introduced technology called connected vehicles will make vehicles capable of providing detailed information such as their position, speed, acceleration rate, etc. in real-time using a wireless technology. The deployment of connected vehicle technology would provide the opportunity to introduce new traffic control strategies or to enhance the existing one. Some work has been done to-date to develop new ATSC systems on the basis of the data provided by connected vehicles which are mainly designed on the assumption that all vehicles on the network are equipped with the connected vehicle technology. The goals of such systems are to: 1) provide better performance at signalized intersections using enhanced algorithms based on richer data provided by the connected vehicles; and 2) reduce (or eliminate) the need for fixed point detectors/sensors in order to reduce deployment and maintenance costs. However, no work has been done to investigate how connected vehicle data can improve the performance of ATSC systems that are currently deployed and that operate using data from traditional detectors. Moreover, achieving a 100% market penetration of connected vehicles may take more than 30 years (even if the technology is mandated on new vehicles). Therefore, it is necessary to provide a solution that is capable of improving the performance of signalized intersections during this transition period using connected vehicle data even at low market penetration rates.

This research examines the use of connected vehicle data as the only data source at different market penetration rates aiming to provide the required inputs for conventional adaptive signal control systems. The thesis proposes various methodologies to: 1) estimate queues at signalized intersections; 2) dynamically estimate the saturation flow rate required for optimizing the timings of traffic signals at intersections; and 3) estimate the free flow speed on arterials for the purpose of optimizing offsets between traffic signals.

This thesis has resulted in the following findings:

1. Connected vehicle data can be used to estimate the queue length at signalized intersections especially for the purpose of estimating the saturation flow rate. The vehicles' length information provided by connected vehicles can be used to enhance the queue estimation when the traffic composition changes on a network.
2. The proposed methodology for estimating the saturation flow rate is able to estimate temporally varying saturation flow rates in response to changing network conditions, including lane blockages and queue spillback that limit discharge rates, and do so with an acceptable range of errors even at low level of market penetration of connected vehicles. The evaluation of the method for a range of traffic Level of Service (LOS) shows that the maximum observed mean absolute relative error (6.2%) occurs at LOS F and when only 10% of vehicles in the traffic stream are connected vehicles.
3. The proposed method for estimating the Free Flow Speed (FFS) on arterial roads can provide estimations close to the known ground truth and can respond to changes in the FFS. The results also show that the maximum absolute error of approximately 4.7 km/h in the estimated FFS was observed at 10% market penetration rate of connected vehicles.
4. The results of an evaluation of an adaptive signal control system based on connected vehicle data in a microsimulation environment show that the adaptive signal control system is able to adjust timings of signals at intersections in response to changes in the saturation flow rate and free flow speed estimated from connected vehicle data using the proposed methodologies. The comparison of the adaptive signal control system against a fixed time control at 20% and 100% CV market penetration rates shows improvements in average vehicular delay and average number of stops at both market penetration rates and though

improvements are larger for 100% CV LMP, approximately 70% of these improvements are achieved at 20% CV LMP.

Acknowledgement

First, I would like to thank my supervisor, Professor Bruce Hellinga for his inspiration, guidance, mentorship, and support throughout my PhD studies. It has been my privilege to work with him and to learn from his professionalism, knowledge, and great personality.

I would like to thank my committee members, Dr. Tony Qiu, Dr. Clarence Woudsma, Dr. Carl Haas, and Dr. Liping Fu for their time and valuable comments toward improving this research.

I am grateful to all my colleagues and friends at University of Waterloo for their support and great friendship that kept me motivated throughout the course of this research.

Special thanks go to Sara Ghaffarian for her support and unwavering belief in me. Thank you for always being by my side. I could not have completed this journey without you.

Finally, I would like to extend my deepest gratitude to my family. My brothers Dr. Hossein Bagheri, Amin, and Moeen, and my sister-in-law Dr. Atieh Mehdizadeh for supporting me throughout this endeavour. Last but not least, my parents Tahereh and Reza whom I cannot thank enough. Without you I would not be the person that I am today. I love you both.

Dedication

To my beloved parents

Tahereh Nassirian, and Reza Bagheri

for their unconditional love and support…

Table of Contents

List of Figures

List of Tables

1

Introduction

The transportation system is a key component of the economy that provides the movement of people and goods between the locations of different activities. The road network as a part of the transportation system carries a high portion of the traffic and the transportation demand in an urban system. Signalized intersections are used within the road network to manage the right of way of different movements that conflict over shared segments of roads.

The role of intersections is important from three aspects. First, intersections are the points where traffic experiences most of its delay. Therefore, an appropriate and well-designed traffic signal control can reduce the imposed delays. Second, different modes of traffic conflict at intersections resulting in significant safety risk both for motorized and non-motorized movements. As reported by the Ministry of Transportation Ontario (2013), 43.2 percent of all motor vehicle accidents in 2013 in Ontario occurred at intersections or were related to the operation of intersections. It is also reported that among the total fatal accidents, 29 percent happened at or were related to the operation of intersections. Finally, from the environmental point of view, delays and traffic congestion at intersections result in increased fuel consumption and tailpipe emissions.

The focus of this thesis is on examining how the operation of signalized intersections can be improved through the use of new sources of data; specifically data that will be available from connected vehicles (vehicles which have the ability to wirelessly communicate a defined set of data between suitably equipped roadside infrastructure and other connected vehicles). The next section provides background material before the research goals and objectives are defined in more detail.

1.1 Background

This section provides a brief introduction to the operation of traffic signals as well as to key traffic engineering definitions that are used later in this document.

1.1.1 Traffic Signals

The existence of traffic signals can be traced back to the first traffic signal that was installed in London, U.K. back in 1868. This device manages the traffic by altering the right of way at an intersection between conflicting traffic movements. The Manual on Uniform Traffic Control Devices (MUTCD, 2009) defines a "phase" as "the right-of-way, yellow change, and red clearance intervals in a cycle that are assigned to an independent traffic movement or combination of movements." The "interval" in the above definition is the duration of time that traffic signals remain unchanged (e.g. red, amber, green intervals).

A traffic signal controller is a device that can be programmed to switch between phases. The controller starts with a phase for a specific movement and continues switching between the phases until it returns to the same (first) phase. A complete sequence of phases is called a "cycle" and its duration is known as the cycle length. Figure 1-1 shows typical phase numbers for a four leg intersection. In practice these phases are grouped into a diagram called "Ring-Barrier Diagram" as shown in Figure 1-2. Each ring shows a sequence of phases (respective right of way followed by an amber and all-red intervals) that follow after each other. Barriers are used in this diagram to separate conflicting groups of phases. Phases from ring 1 and 2 can be combined as long as they are within the same group.

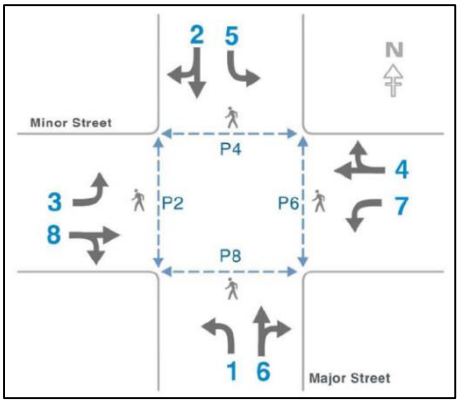


Figure 1-1: Typical phases for a four-leg intersection (Traffic Signal Timing Manual, 2008)

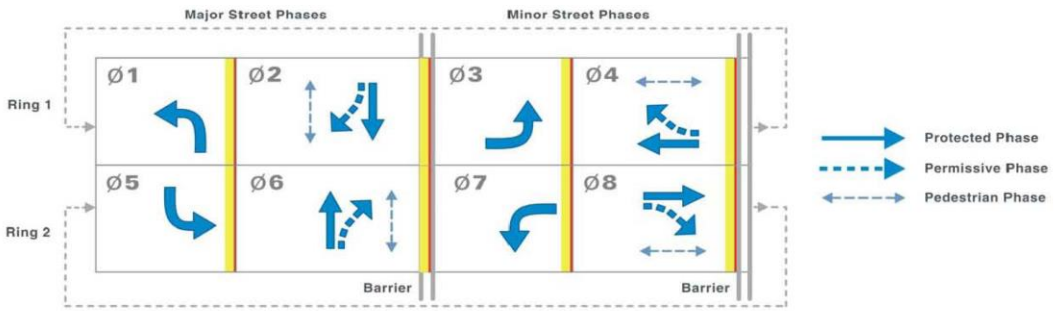


Figure 1-2: Ring-barrier diagram (Traffic Signal Timing Manual, 2008)

During the red interval a queue of vehicles forms at an intersection. When the signal switches to green, it will take a while for the first few vehicles in the queue to observe and react to the change in the signal plus additional time required for the vehicles to accelerate and to start moving. This delay in release of the queue is called start-up lost time (Figure 1-3). After this period, traffic will discharge at the maximum possible flow rate called the saturation flow rate *S*. When the queue has been dissipated, the flow that passes the stop-line decreases and becomes equal to the flow approaching the intersection called the arrival flow rate *V*.

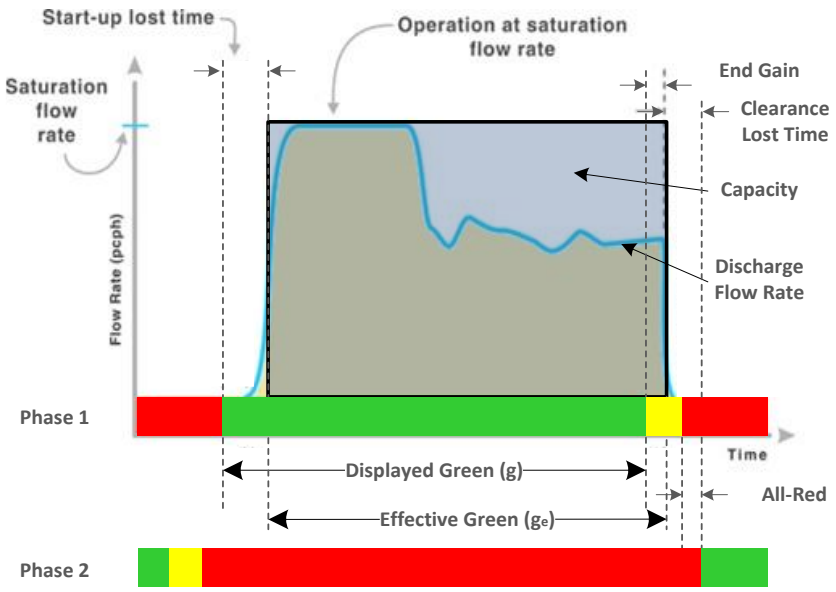


Figure 1-3: Typical traffic flow profile at a signalized intersection (Traffic Signal Timing Manual, 2008).

After the end of the green interval, a few vehicles typically still continue traveling through the intersection during some portion of the amber interval. This time is called “end gain”. The period of time between the end of one green interval and the beginning of the next green interval of the next phase is called the intergreen. This period is typically composed of an amber and an all-red

intervals. The portion of the intergreen during which vehicles do not discharge is called the "clearance lost time". Effective green g_e is the period over which vehicles actually discharge and is equal to displayed green, g, minus start-up loss time plus end gain. The effective green time is typically estimated as displayed green time plus 1 second (Teply et al., 2008). The capacity of a phase is defined as the total traffic volume that can be served during the effective green at the saturation flow rate as shown in the blue box in Figure 1-3.

When signalized intersections are closely spaced, traffic from the upstream intersection to the downstream intersection tends to travel as a platoon. Coordinating traffic signals is an approach to program several traffic signals timings at adjacent intersections in such a way that platoons of vehicles can travel along the arterial with the minimum delay and stops. In order to provide coordination, a specific interval (usually green) for a reference phase at consecutive intersections should start at a specific time relative to each other. This difference in start time is called the offset. An example of coordination is shown in Figure 1-4. The blue lines in the figure represent the portion of the space-time diagram in which vehicles are able to traverse this section of arterial without experiencing delays at the signalized intersections assuming there are no queues at these intersections.

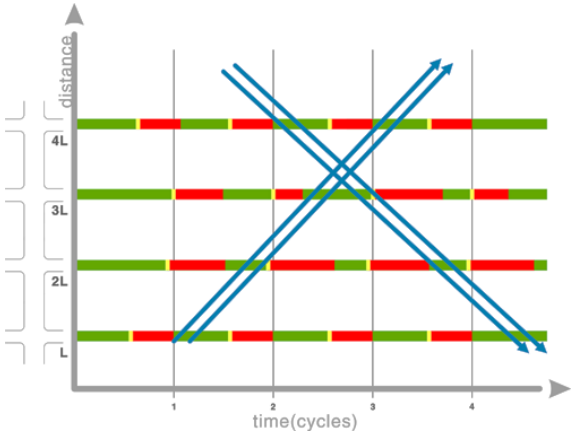


Figure 1-4: Signal coordination (Traffic Signal Timing Manual, 2008).

1.1.2 Delay at Signalized Intersections

Delay is an important measure of performance of signalized intersections. Two types of delays are commonly used in analyzing the performance of signal control systems namely: control delay, and stopped delay.

Figure 1-5 illustrates the definitions of these delays. The figure shows the trajectory of a single vehicle as it passes through a signalized intersection. The vehicle approaches the intersection at free flow speed, which is defined as the speed at which vehicles travel when not impeded by traffic control devices . At time t_1 the vehicle begins to decelerate because the traffic signal is red and at time t_3 comes to a stop. At time t_5 the signal turns green and at t_6 the vehicle begins to accelerate. At t_7 the vehicle passes the stop line and at t_8 reaches the free flow speed again. The dashed line represents the trajectory of the vehicle in the absence of the signal.

Control Delay is defined as the total delay experienced by the vehicles because of the traffic control system (Equation (1-1)). Stopped delay is defined as the time that a vehicle spends at a stationary mode caused by the presence of a traffic control device (Equation (1-2)). Approach delay is the portion of the control delay that occurs upstream of the stop line, and is equal to the time that the vehicle passes the stop line (t_7) minus the time at which the vehicle would have passed the stop line if the traffic signal had not been in place t_2.

$$\text{Control Delay} = t_8 - t_4 \tag{1-1}$$

$$\text{Stopped Delay} = t_6 - t_3 \tag{1-2}$$

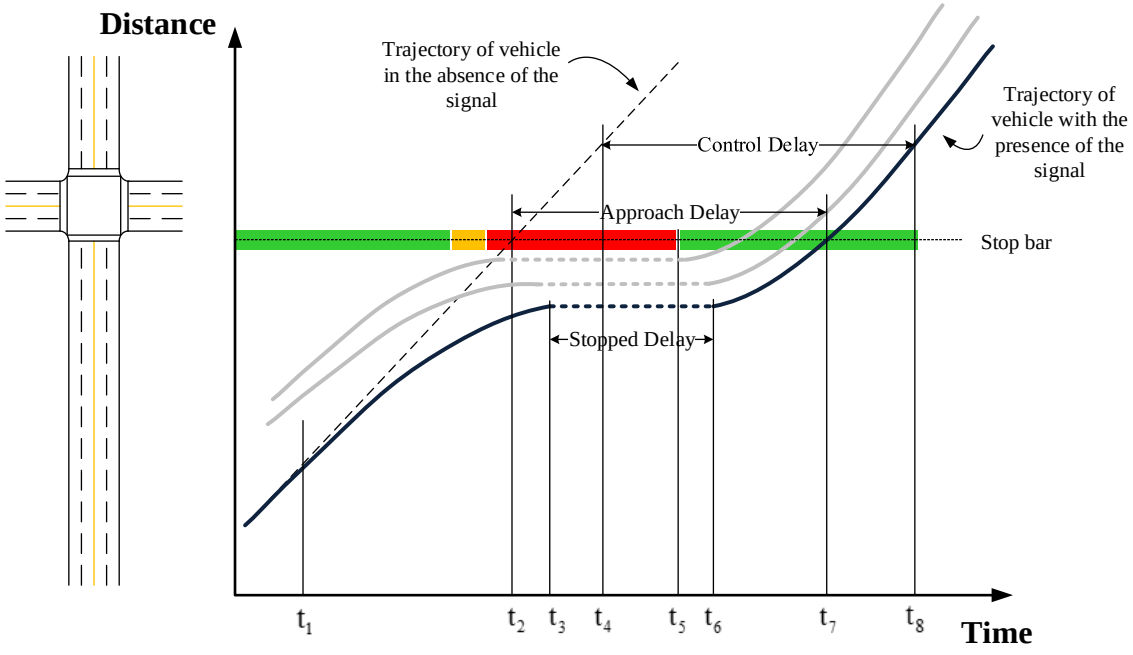


Figure 1-5: Different types of delay.

An extensive body of research has been developed during the last 60 years to develop models to estimate delay at signalized intersections. A deterministic queuing model assumes a uniform arrival and departure rate of vehicles at intersections (D/D/1). If we assume that demand is less than capacity (i.e. intersection is under-saturated), then delay is calculated using Equation (1-3).

$$d = \frac{C(1-\lambda)^2}{2(1-\lambda x)} \tag{1-3}$$

Where,

d = Average delay per vehicle, (s/veh)
C = Cycle length, (s)
λ = Effective green to cycle length ratio (g_e/C)
x = Degree of Saturation which is defined as the ratio of arrival flow to the capacity of the approach (V/c)
V = Arrival flow rate, (veh/hr)
c = Capacity of the approach ($S \times g_e/C$), (veh/hr)
S = Saturation flow rate, (veh/hr)

Details on how to derive the above equation can be found in Roess, Prassas, & McShane (2004, pp. 470-494).

However, the assumption of deterministic arrivals is frequently violated in practice. Webster (1958) proposed a delay function assuming Poisson arrivals and incorporating an empirical correction term (Equation (1-4)).

$$d = \frac{C(1-\lambda)^2}{2(1-\lambda x)} + \frac{x^2}{2V(1-x)} - 0.65\left(\frac{C}{V^2}\right)^{\frac{1}{3}} x^{2+5\lambda} \tag{1-4}$$

The first term in the Webster's delay function estimates delay assuming uniform traffic arrival as it was presented in Equation (1-3). The second term accounts for the random nature of the traffic arrival assuming Poisson distribution, and the third term is an empirical correction expression. Webster's model is valid only when traffic at an intersection is at the steady state ($x < 1$). As x approaches 1, d increases asymptotically and therefore, Webster's explanation is not suitable for applications in which x is greater than approximately 0.85 and when traffic is close to the saturation condition. Therefore, Kimber & Hollis (1979) presented a coordinate transformation technique to address this problem with the Webster's model. In their method as shown in Figure 1-6, the

Webster model is transformed in such a way to be asymptotic to the queue length line produced by the deterministic theory.

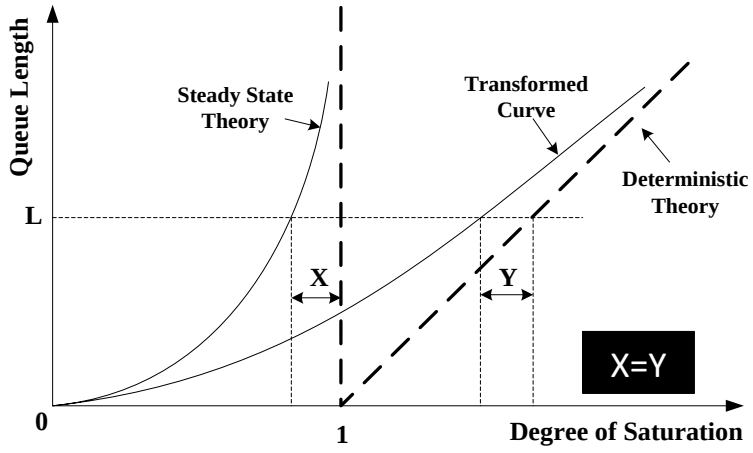


Figure 1-6: Co-ordinate transformation (Kimber & Hollis, 1979)

Equation (1-5) presents the delay function used in (HCM, 2010). The first term d_1 represents the uniform delay; d_2 is taking into account the random delay; d_3 accounts for the delay due to the existing queue at the start of the analysis, and the Progression Factor, *PF*, is used to capture the effect of coordination on delay when vehicles arrive in platoons.

PF has a range of values between zero to around 2.5 with zero when there is an excellent coordination on an arterial and higher values when the progression becomes poor. Therefore, during under-saturated conditions providing coordination on an arterial can significantly reduce the delay as the d_1 term of the HCM delay function becomes small and close to zero. However, when an intersection is operating at the oversaturated condition, queues form at the intersection which cannot be cleared during the provided green interval. Therefore, queues start to grow resulting in the d_3 term becoming large and dominant in the delay function. This fact indicates the importance of managing queues at oversaturated conditions in order to reduce delays.

$$d = d_1(PF) + d_2 + d_3 \tag{1-5}$$

Where,

$$d_1 = \frac{0.5C(1-\lambda)^2}{1-[\min(1,x)\lambda]}$$

$$d_2 = 900T\left[(x-1) + \sqrt{(x-1)^2 + \frac{8kIx}{cT}}\right]$$

$$d_3 = \frac{1800 Q_b (1+u) t'}{\mathrm{cT}}$$

PF = Progression Factor,

T = Analysis period (h),

k = Calibration term incorporating the effect of controller type,

I = Adjustment that accounts for the impact of upstream signal on arrival rate,

Q_b = initial queue at the start of the analysis period T (veh),

$$t' = \begin{cases} 0 & \text{if } Q_b = 0 \\ \min\left\{T, \dfrac{Q_b}{c[1-\min(1,x)]}\right\} & \text{if } Q_b \neq 0 \end{cases}$$

$$u = \begin{cases} 0 & \text{if } t' < T \\ 1 - \dfrac{cT}{Q_b[1-\min(1,x)]} & \text{if } t' = T \end{cases}$$

The Canadian Capacity Guide (Teply, et al., 2008) uses a similar delay function presented by Equation (1-6). This guide, in contrast to the HCM, does not consider the delay due to the presence of an initial queue at the beginning of the analysis period.

$$d = PF \cdot d_1 + d_2 \tag{1-6}$$

Where,

$$d_1 = \frac{C\left(1 - \frac{g_e}{C}\right)^2}{2\left[1 - \frac{g_e}{C} \cdot \min(x, 1.0)\right]}$$

$$d_2 = 15T\left[(x-1) + \sqrt{(x-1)^2 + \frac{240x}{cT}}\right]$$

T = Evaluation Period, (sec)

Delay at signalized intersections, as presented by the Highway Capacity Manual and Canadian Capacity Guide in equations 1-5 and 1-6, can be influenced by several factors. Two of these factors include:

1) Capacity of the intersection which is influenced in turn by the saturation flow rate.
2) The progression factor which has impact on the d_1 term as the dominant factor in the delay function during under-saturated conditions. Better progression or signal coordination is

achieved by adjusting signal offsets calculated knowing the free flow speed travel time between consecutive intersections.

Consequently, proper signal timing design and signal offsets on the basis of true saturation flow rates and the free flow speeds between intersections, respectively, can improve the performance of traffic on arterials. The next section presents different types of signal control strategies with the goal to reduce delay at signalized intersections.

1.2 Categories of Traffic Signal Control Systems

Different traffic signal control algorithms have been developed all with a common goal to keep the traffic moving through the intersection with the minimum delay and the highest safety. These systems can be classified into four categories as they are presented in the following sections (Traffic Signal Timing Manual, 2008).

1.2.1 Fixed-time or Pre-timed Control

Signal timing plans in pre-timed controls are based on the historical traffic flows at intersections. These types of controls are suitable for locations with consistent traffic patterns or closely spaced intersections. In this type of control, several signal timing plans that are programmed and stored in traffic signal controllers are scheduled to be selected by the controller for different times of a day and days of a week considering average historical traffic demands. This "pre-timed" and "time-based" type of signal control cannot accommodate real-time or unpredicted changes in the traffic demand. In this method, the optimum cycle length is calculated using the well-known Webster's (1958) model presented in Equation (1-7). Then the calculated optimum cycle length is split between different phases based on the ratio of arrival flow rate to saturation flow rate (*V/S*) of each approach as shown in Equation (1-8).

$$C_o = \frac{1.5L + 5}{1 - \sum_{p=1}^{P} y_p} \tag{1-7}$$

$$g_{ep} = (C_o - L).\frac{y_p}{\sum_{p=1}^{P} y_p} \tag{1-8}$$

Where,

C_o = Optimum cycle length, (s);
L = Total lost time within a cycle, (s);
y_p = Maximum (critical) flow ratio (*V/S*) of all lane groups that are served in phase *p*;
P = Total number of phases;
g_{ep} = Effective green time for the phase of interest *p*, (s).

Pre-timed signal control systems are popular since they have the lowest implementation and maintenance costs. However, in addition to their weakness in responding to fluctuations in the traffic demand as mentioned before, their signal timings must be updated every few years to account for any changes in the traffic pattern and demand.

1.2.2 Actuated Control

The majority of actuated traffic signals operate using data provided by Inductive Loop Detectors (ILDs). An ILD is a wire loop that carries an electrical current and is installed beneath the surface of the road. When a vehicle passes the ILD, the electrical current in the loop changes (an actuation) and a record of the vehicle is created. The data provided by ILDs can include vehicle counts, presence, speed, and some can do vehicle classification.

In order to operate these types of controls, traffic sensors are installed only on the minor approach (semi-actuated) or on all the approaches (fully actuated) to detect gaps in the traffic flow or presence of vehicles. The most common use of semi-actuated control is on arterials with coordinated signals for the major movement intersecting with minor streets with low traffic demands. Using the semi-actuated mode, traffic signals stay green for the major arterial movement unless demands are detected on the minor streets. This type of operation, in comparison to the fixed-time method, helps in reducing delays and stops on the major movement of an arterial by preventing unnecessary changes of signal to minor streets when there is no demand.

The actuated control system also has the capability to adjust the green interval between some minimum and maximum values. If the time between successive detections of vehicles is less than a defined value (e.g. 2 seconds), the green interval is extended by a specified amount (e.g. 4 seconds). In this way, the green interval for the actuated phase can be extended as long as vehicles are present and the maximum green time has not been reached. Maximum and minimum phase times are calculated with a method similar to the fixed-timed control and stored in the controller

providing the flexibility to increase and decrease the phases by the controller within that range. Furthermore, the controller can skip phases with no detected demand.

The advantages of using this control are: First, it reduces the delay at intersections in comparison to the fixed-time control; Second, green intervals and the signal cycle length can be adjusted based on the variations in the traffic demand; and third, a phase with no call (a phase with no detected vehicle) can be skipped.

However, there are several disadvantages associated with the actuated signal control. First, this technology requires installing traffic sensors, and therefore initial installation and maintenance costs are higher than for the pre-timed signal control. Second, the data provided by the traffic detectors contain information related to the traffic condition only at the intersection and the controller does not consider the traffic conditions at nearby adjacent intersections. Finally, the response time to the change in the traffic demand is still an issue in these systems. As the traffic needs to be detected by the sensors located near the intersections, movements still may experience unnecessary stops and delays at the intersection.

1.2.3 Traffic Responsive Control

Traffic responsive controls are developed to address some of the disadvantages related to the time-based controls. In these systems, traffic signal plans that are preprogrammed and stored in the signal controllers are selected based on the traffic flow information obtained from detectors rather than time of day. Traffic responsive systems require additional detectors called “system detectors” to be installed along an arterial to collect traffic data (usually volume and occupancy) required for the system. These systems usually change the timing plan of several coordinated intersections on an arterial at the same time.

1.3 Adaptive Traffic Signal Control (ATSC)

Adaptive traffic signal controls determine signal plans based on the traffic information they obtain from detectors located near to the stop line or some distance upstream of the intersection on the approach link. When data from upstream traffic sensors are used, a prediction model is usually used to predict the arrival of the measured traffic demands at the downstream intersection. Potential changes to the signal timings are evaluated using the predicted future traffic demands

and quantified using a Performance Index (PI). The PI can consider a combination of vehicular delays, stops, and travel times at intersections.

ATSCs respond to the change in the traffic demand before traffic arrives at the intersection and therefore, these systems are considered proactive. Signal plan retiming is not required in these systems since ATSCs can adapt themselves to the changes in the traffic demand by continuously updating their signal timings. The following provides description of three widely used adaptive signal control systems.

The Sydney Coordinated Adaptive Traffic System (SCATS)

SCATS was developed in Australia but has been deployed widely around the world. It works at two optimization levels called Strategic and Tactical. Strategic control optimizes the network at a higher level, while the tactical operates locally within the constraints defined by the strategic control. SCATS tries to provide coordination for a group of intersections called a subsystem. Each subsystem has a critical intersection with the highest *V/c* ratio for which the cycle length and splits are optimized by a regional computer at the strategic level. Several subsystems can also be coordinated to expand the progression through a network. Each intersection can operate in actuated mode which provides the ability to skip or shorten or increase a phase within the cycle length equal to the critical intersection.

As Figure 1-7 shows, detectors in SCATS are installed close to the stop line (Lowrie, 1982). It optimizes the cycle length and splits of the critical intersection by maintaining the degree of saturation *x* close to 90% for the lane with the highest *x* within a lane group. However, SCATS is considered as a reactive system since it adjusts the signal timings based on the observed instead of predicted traffic demand. The system also lacks of a traffic model since it does not predict traffic conditions for the future. During oversaturated conditions SCATS increases the cycle length up to a maximum value and if the oversaturated condition continues, the green interval for the phase associated with the major approach (usually the one that requires coordination) will be increased, which is called “Stretch Effect”. However, the lengths of queues on approaches are unknown to the SCATS system.

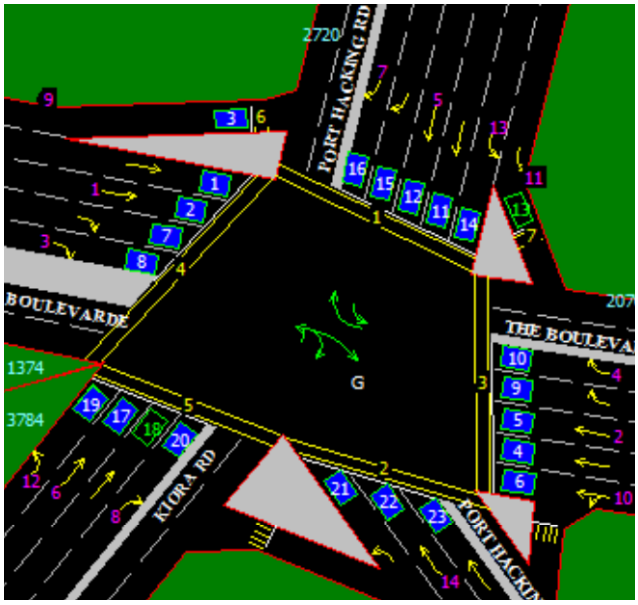


Figure 1-7: Layout of detectors in SCATS system (SCATS, 2013).

The Split Cycle Offset Optimization Technique (SCOOT)

SCOOT has been used in many cities around the world (e.g. Toronto). It uses detectors installed upstream of an intersection to create a shape of the traffic demand called traffic Cyclic Flow Profile (CFP). It uses a platoon dispersion model to predict the shape of the CFP when it arrives at the intersection (Dunn Engineering Associates, 2005, pp. 3-37). Using CFPs it also predicts the length of queues for that intersection as shown in Figure 1-8. Similar to SCATS, SCOOT groups several traffic signals and each group has a critical intersection. The degree of saturation at the critical intersection is maintained at 90%.

SCOOT uses a prediction model to proactively adjust the signals. Signal settings are optimized by applying small changes to current signal settings and to estimate the impact of these changes on a PI. Green splits are optimized a few seconds before the end of a phase by determining whether it is better to delay or advance the termination of the current phase. Offsets are optimized once every cycle by altering the start of the cycle. The cycle length is optimized once every few minutes (2.5 minutes when demand is changing rapidly or 5 minutes during the normal operation).

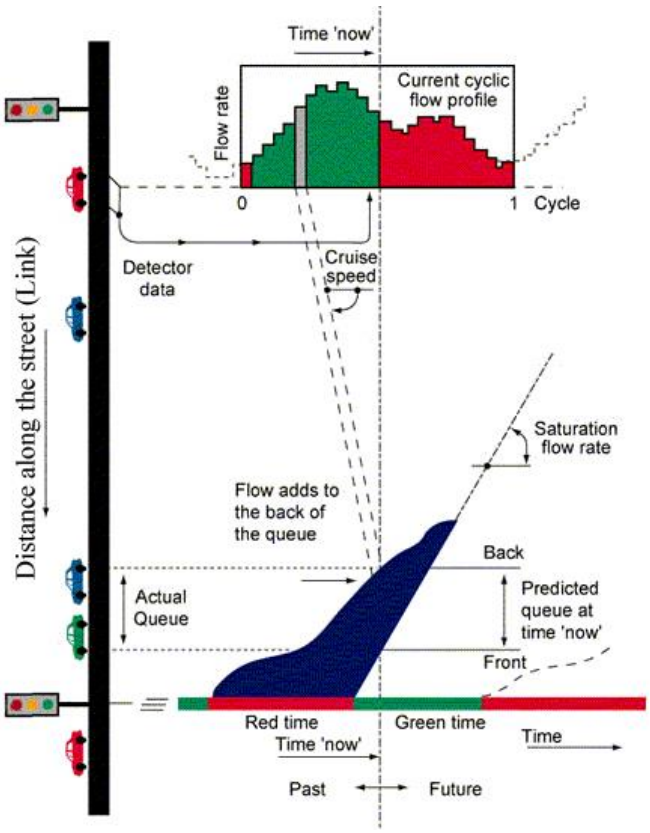


Figure 1-8: Queue prediction in SCOOT system (How SCOOT Works, 2008)

However, SCOOT has a few drawbacks:

- The operation of the system is dependent on fixed point traffic sensors. Therefore, with even few detectors failures, the performance of the system degrades to a pre-timed control.
- The system does not perform well during over-saturated condition. Queues can reach the upstream detectors resulting in failure to create CFPs (Jhaveri, 2003).

Adaptive Control Software Lite (ACS Lite)

ACS Lite (Shelby, et al., 2008) uses stop-line detectors to measure the degree of saturation for each phase and adjusts their timing to maintain *x* for all the phases under 95%. It also uses upstream detectors to build CFPs and to adjust coordination. The optimization is performed every 5 to 10 minutes. However, ACS Lite does not optimize the cycle length and the cycle length is selected based on the time of day plans.

The fact that it cannot optimize the cycle length makes it unsuitable for locations where the traffic pattern can change due to special events and where there are fluctuations in the traffic demand.

In summary, there are several limitations associated with ATSC systems:

1) The operation of these systems is highly dependent on a relatively large number of dedicated traffic sensors on each approach, and thus these systems are relatively costly to deploy and maintain. Different sensor technologies including cameras, microwave, infrared radars, and micro loops have been developed to collect traffic information to be used in signal control algorithms. However, concerns about the accuracy and the type of information provided by these sensors, as well as the installation and maintenance costs act as a deterrent to wider deployment of these systems (Dunn Engineering Associates, 2005).
2) Current adaptive signal control systems use static traffic flow parameters in their optimization process. Some assume a static value for the saturation flow rate and the free flow speed of vehicles on the links (SCOOT FAQ, 2008). The free flow speed is used to predict the location of vehicles in the future when they arrive at the stop line and the saturation flow rate is required to estimate the number of vehicles that depart from the intersection. Knowing these two parameters, the queue lengths and utilization of capacity of an intersection can be estimated. However, as shown by Edwards (1999) and Niittymäki & Pursula (1997) inclement weather can have an impact on these key parameters that are used in these methods to predict the future traffic conditions. This misrepresentation of the model in comparison to the actual traffic condition on the roads may have additional negative impact on traffic for instance during icy road conditions as result of unsuitable signal plans (Gillam & Wilhill, 1992).
3) Obtaining real-time turning movement ratios is critical for operation of ATCS. Traffic models use these ratios to predict the condition of the traffic at the stop-line. Turn ratios at intersections varies by time of day and day of the week. However, the majority of ATCS based on traditional traffic sensors use predefined (static) turning movement ratios in their model which may result in misrepresentation of the condition in the field.

1.4 Introduction to the Connected Vehicles Technology

An efficient and effective signal control algorithm requires sensors that can provide the required data with a good quality and spatial coverage. Recently, a new technology called “Connected Vehicles” (CV) has been introduced, which has the potential to provide accurate and real-time data over a wide spatial area. CV is a system in which vehicles are able to communicate with each other

or with road side infrastructure using a wireless technology. This technology can bring benefits to different aspects of the transportation system as the real-time data provided are more accurate, and with new pieces of information about individual vehicles such as speed, acceleration rate, position, etc. As a result of this improvement in the data quality, exchange, and availability, new features and applications in terms of traffic management can be provided to both road users and transportation agencies. Traditional traffic signal control systems also can be modified to incorporate the data provided by the connected vehicles technology with the potential to improve the performance of the signalized intersections.

The standards for the wireless technology required for implementing connected vehicles have been developed (SAE, 2016). A standard developed by the Society of Automotive Engineers (SAE) called J2735 DSRC Message Set Dictionary defines the type of data provided by connected vehicles (SAE, 2016). Currently there are 231 data elements included in 16 different message sets. A mandatory message set called *Basic Safety Message (BSM) Part I*, contains basic information (data elements) about vehicles such as: Latitude; Longitude; Elevation; Positional Accuracy; Transmission status; Speed; Heading; Steering Wheel Angle; 4 way acceleration; Brake System Status; and Vehicle Size (SAE, 2016). This message set is transmitted with maximum rate of 10 times every second. Appendix A provides more information on the connected vehicle technology including available standards in North America and Europe, message sets and several envisioned applications of connected vehicles.

The National Highway Traffic Safety Administration (NHTSA) has provided guidance and rules for implementation of the connected vehicle technology. The latest Notice for Proposed Rulemaking (NPRM) announced by the agency in December 2016 states that:

> "The agency is proposing that the effective date for manufacturers to begin implementing these new [connected vehicles] requirements would be two model years after the final rule is adopted, with a three year phase-in period to accommodate vehicle manufacturers' product cycles. Assuming a final rule is issued in 2019, this would mean that the phase-in period would begin in 2021, and all vehicles subject to that final rule would be required to comply in 2023." (NHTSA, 2016)

NHTSA does not mandate the connected vehicle technology nor provide an implementation schedule for existing manufactured vehicles; though states that any aftermarket device for such vehicles should meet the connected vehicles requirements.

The National Electrical Manufacturers Association (NEMA) jointly with the Institute of Transportation Engineers (ITE) and the American Association of State Highway and Transportation Officials (AASHTO) have developed NTCIP (National Transportation Communications for ITS Protocol) standards. The NTCIP 1202 defines the standards for communicating to traffic signal controllers. Currently, a new version of this standard (version 3) is under development since October 2014, which will provide information on how signal controllers will interface with the connected vehicle technology and how connected vehicle data can be captured by the signal controller and communicated to a central management system using standard communication protocols. Therefore, it is expected that traffic signal controller manufacturers will comply with the new release of this standard once issued (initial release date was the end of 2016), providing an easier path to use connected vehicle data in traffic signal operations (Johnson et al., 2016).

1.5 Problem Definition and Research Goals and Objectives

Conventional signal control systems typically rely on fixed sensor data (i.e. loop detectors, radar, etc.). The deployment of Connected Vehicles provides an opportunity to use this new data source as an input to the signal control system. However, complete adaptation of the connected vehicles technology on new and existing vehicles will take several years. An earlier forecast (John A. Volpe , 2008) suggested that approximately 50%, 90%, and 100% CV market penetration rate would be achieved around 9, 20, and more than 30 years after the start of the phase-in period respectively. In order to take advantage of this new technology in improving the operation of signalized intersections, developed methods for signal timing optimization based on connected vehicle data should also be able to operate during early implementation stages of this technology when only a portion of vehicles in the traffic stream are equipped with this new technology.

There are many deployments of adaptive signal control systems around the world that operate using data from traditional detectors. Three of these systems (SCATS, SCOOT, and ACS lite) are described in the previous section. An approach to take advantage of connected vehicles to enhance

the operation of signalized intersections is to develop methodologies to provide required inputs for these existing signal control systems that typically are not available from traditional sensors.

As mentioned earlier in this chapter, two important inputs to existing adaptive signal control systems are the stop line saturation flow rate and the free flow speed. It is proposed that the data from connected vehicles can be used to provide these two inputs to these systems in real-time to enhance their operation while reducing or eliminating their dependency on traditional sensors. Thus, this thesis has the following main objectives:

1. Develop a method to estimate the saturation flow rate using data obtained from connected vehicles. The method must be applicable when the portion of connected vehicles in the traffic stream is relatively small.
2. Evaluate the performance of the proposed saturation flow rate estimation method for different market penetration rates of connected vehicles, and for different scenarios including different traffic volumes, different approach geometries, traffic incidents, and road surface conditions.
3. Develop an algorithm to extract arterial segment free flow speed using data from connected vehicles.
4. Evaluate the performance of the developed free flow speed estimation method for different levels of market penetration of connected vehicles and road surface conditions.
5. Evaluate the impact that the proposed algorithms have in terms of the operation of the signalized intersection.
6. Determine the conditions for which the proposed algorithms provide superior performance.

1.6 Dissertation Outline

The rest of this proposal is organized into 5 chapters:

Chapter 2 presents the methodology proposed for the estimating the number of queued vehicles at intersection using connected vehicle data.

Chapter 3 describes the proposed methodology for estimating the signalized intersection lane group saturation flow rate using data from connected vehicles.

Chapter 4 presents the methodology proposed for estimating free flow speed on arterials using connected vehicle data.

Chapter 5 demonstrates and evaluates using the proposed methodologies in the previous chapters in an adaptive signal control system for optimizing arterial signal timings in real-time.

Chapter 6 provides a summary of conclusions, and research contributions as well as directions for future studies.

2

Queue Estimation for the Purpose of Estimating Saturation Flow Rate using Connected Vehicles

This chapter presents methodologies to estimate queue length at signalized intersections using data from connected vehicles for the purpose of estimating the saturation flow rate.

Adaptive Traffic Signal Control (ATSC) systems such as SCOOT (Siemens, 2012) are designed to provide real-time signal timing adjustments on arterial streets in reaction to changes in the traffic conditions. These systems use models to optimize the signal timing and these models require several inputs or assumptions about the traffic stream such as arrival rate, saturation flow rate, turn ratios, etc. These inputs are provided by traffic sensors, usually Inductive Loop Detectors (ILD) or other technologies such as radar or video image processing. A recent initiative called Connected Vehicles (CV) technology facilitates vehicles to broadcast their real-time status data such as their position, speed, acceleration, etc. along their travel path, which can potentially be used instead of traditional sensors to provide the required inputs to ATSCs.

One of the inputs required by ATSCs is the saturation flow rate, which is used to estimate the capacity of different movements at signalized intersections. As discussed in more detail in the next chapter, one of the requirements in the process of estimating the saturation flow rate is the number of stopped (queued) vehicles behind (i.e. upstream) the intersection stop line. This chapter presents queue estimation methodologies using connected vehicles that are suitable for estimating the saturation flow rate. Evaluations and performance comparison of these proposed techniques against different traffic conditions and road geometries are also provided.

2.1 Background

Queue length is one of the measures of effectiveness used for evaluating the performance of signalized intersections. Different methodologies have been developed by researchers to estimate

the queue length. These methodologies can be categorized into three main groups: input-output models, shockwave theory, and probe based approaches.

The input-output (I/O) model was first introduced by Webster (1958). This model estimates the number of vehicles accumulated behind the stop line using the difference between the uniform arrival and departure rates at an intersection. Several improvements were proposed to this model including those by Miller (1968) and Akçelik (1988) to take into account the stochastic nature of the arrival flow on arterial streets. The model was later improved by Akçelik (1999) and adopted for the Highway Capacity Manual (HCM) (2000, 2010). In the methodology provided by the HCM, the back-of-queue is estimated using Equation (2-1) by considering the queue caused by the accumulation of vehicles during the red interval, the queue caused by the random arrival of vehicles, and the remaining queue from previous analysis period. The HCM model is mostly used in determining the average queue length and queue spill-back during an analysis period.

$$Q = Q_1 + Q_2 + Q_3 \tag{2-1}$$

Where,

Q = Back-of-queue size, (*veh/ln*);

Q_1 = Queue caused at each cycle due to the signal sequencing through phases, (*veh/ln*);

Q_2 = Random queue term; the queue caused due to the fluctuation in the arrival flow, (*veh/ln*);

Q_3 = Queue due to oversaturation condition at the traffic signal, (*veh/ln*).

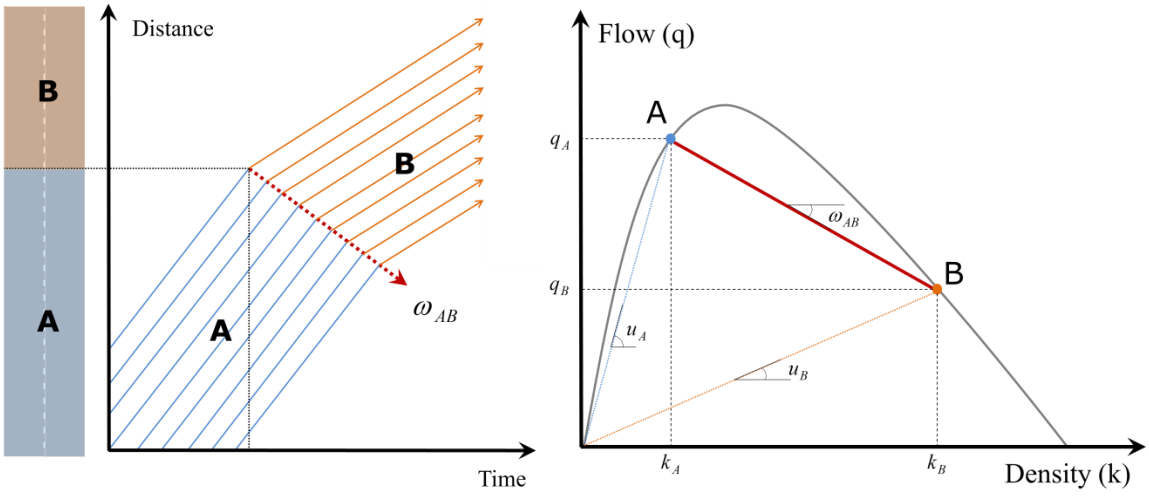
An example of the real-time queue length estimation using the input-output model is SCOOT (Siemens, 2012) which is based on the approach used in the TRANSYT software (Robertson, 1969). The model uses traffic detectors located upstream of intersections to obtain the vehicle arrival profiles. The difference between the arrival and an assumed departure flows are used to estimate the number of queued vehicles behind the stop line.

The input-output models are able to only estimate the number of queued vehicles and therefore, the spatial context of the queue is unknown. These models also have difficulties to estimate a queue when the queue extends upstream of the location of the advance detector (Islam, 2013).

Another group of queue estimation methods is based on shockwave theory (also known as the LWR model) (Lighthill & Whitham, 1955) (Richards, 1956). The early version of the model was

developed for uninterrupted flows, which later was improved by Stephanopoulos and Michalopoulos (1977, 1979, and 1981) to accommodate the behaviour of traffic flow on signalized roads.

Shockwave theory uses the traffic flow fundamental theory to describe the boundary of two traffic states on a section of a road. The model is based on the notion that the flow rate on any point along a road is correlated to its traffic density. Therefore, the model is able to describe a dynamic boundary between two traffic states having different traffic flows and densities. The boundary between two traffic states, which can propagate through time and space, is called a shockwave. Figure 2-1(a) shows trajectories of vehicles traveling through two traffic states (A, and B) on a space-time diagram. The diagram shows that over time, the boundary between traffic states A and B moves upstream away from the congested area with a speed = ω_{AB}. Figure 2-1(b) represents the same traffic states within a flow-density diagram. The shockwave speed (ω_{AB}) in these diagrams can be calculated using the slope of the line that connects the two states as presented in Equation (2-2).



(a) (b)

Figure 2-1: Shockwave representations on (a) space-time and (b) fundamental diagrams.

$$\omega_{AB} = \frac{q_B - q_A}{k_B - k_A} \qquad (2\text{-}2)$$

Where,

ω_{AB} = The speed of shockwave between states A and B, (*km/h*);

q_A = Flow rate of traffic state A, (*veh/h*);

q_B = Flow rate of traffic state B, (*veh/h*);

k_A = Density of traffic state A, (*veh/km*);

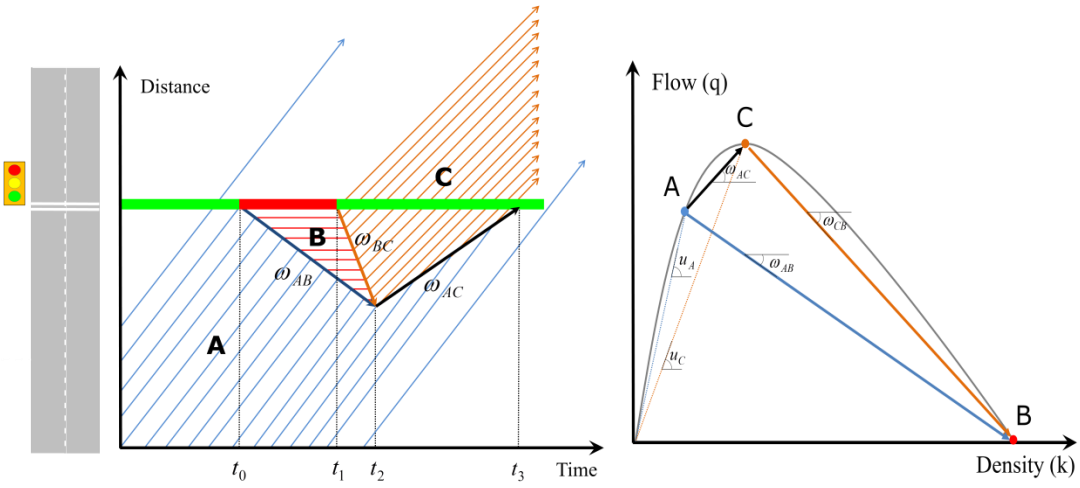
k_B = Density of traffic state B, (*veh/km*).

Shockwave analysis can be used at signalized intersections to estimate queue length (Stephanopoulos et al., 1979). In contrast to queuing theory that is unable to describe the spatial aspect of a queue, shockwave theory can define a queue in both time and space dimensions.

Figure 2-2(a) shows a space-time diagram with trajectories of vehicles at a signalized intersection. In this figure, traffic state A represents the arrival flow at the intersection. Starting the red interval at time (t_0), vehicles start to form a queue and the queue grows at a speed equal to (ω_{AB}). This is a backward forming shockwave. Traffic state B represents those vehicles in a stationary mode in a queue. At the start of the green interval (t_1) vehicles start discharging at the saturation flow rate (traffic state C) and the head of the queue starts to move upstream of the stop line with the speed of (ω_{BC}). This is a backward recovery shockwave. The queue dissipates when the queue formation and dissipation shockwaves (ω_{AB} and ω_{BC}) meet at the time (t_2). It is important to note that, in this figure, the maximum queue length occurs at t_1, and the maximum extent of the queue happens at t_2. Traffic starts to pass the stop line at the arrival rate (state A) when the forward moving shockwave between traffic states A and C (ω_{AC}) reaches the stop line.

Figure 2-2(b) illustrates the above traffic states on the fundamental diagram, in which the speeds of the three discussed shockwaves can be calculated using the slope of the lines that connects each pair of traffic states and as previously presented by Equation (2-2).

Yi et al. (2001) proposed a shockwave theory based queue estimation model which uses a general speed-density relationship to formulate a mathematical model for estimating queue lengths. The model has the ability to adapt to different traffic conditions and geometries by updating the parameters used in the general speed-density relationship using real-time data (speed and headway) obtained from fixed traffic sensors.



(a) (b)

Figure 2-2: Shockwave representations for a signalized intersection.

The real-time queue length estimation methods based on shockwave theory require upstream detectors to measure the arrival flows at intersections. Therefore, these methods perform poorly when queues extend upstream of the upstream detectors, because under these conditions the arrival flow cannot be measured. To address this issue, Mück (2002) developed an enhanced queue length estimation method based on shockwave theory. The method estimates queue lengths using the queue formation shockwave speed estimated by knowing the location of the upstream (also called the advance) detector and the difference in start time of the red interval and the time that the queue reaches the advance detector. However, the methodology assumes uniform headways for vehicle arrivals and therefore assumes the speed of the backward forming shockwave is constant. To address this issue, Liu et al. (2009) developed a queue estimation methodology based on the queue discharge process and break points in occupancy and vehicle headway data collected from loop detectors installed at nearby signalized intersections. The method uses high resolution traffic signal and loop detector data (Sturdevant et al. 2012) to estimate the end of queue both at under and oversaturated conditions.

Probe vehicles or mobile sensors data have recently received increased attention and have been used to estimate queue length at signalized intersections. Comert & Cetin (2009) proposed a methodology that relies only on the location of the last probe vehicle in a queue and a conditional probability distribution of a queue to estimate the expected queue length. Ban et al. (2011) and Hao et al. (2015) used travel times obtained from probe vehicles traveling through an intersection

to estimate queue lengths by detecting sudden changes in the travel times of probe vehicles and applying shockwave theory. Cetin (2012) used coordinates of the first and last probe vehicles joining the back of queue in each cycle to reconstruct queue representations in a space-time diagram and to estimate the speed of shockwaves required to approximate the length of a queue during over-saturated conditions. Other methodologies proposed by Badillo et al. (2012), Li et al. (2013), and Cia et al. (2014) tried to incorporate both fixed sensors and probe vehicle data to estimate queue length on the basis of shockwave theory. Goodall et al. (2014) presented a method that uses spatial gaps between probe vehicles in a stopped queue to infer the presence of non-probe vehicles by assuming an average distance between vehicles. Goodall also made use of a microscopic model and compared the reported behaviour of probe vehicles to the simulation (when vehicles were not in a stopped queue) to reduce errors in inferring non-probe vehicles. Details on the four-step process used to detect the position of non-probe vehicles on arterials are as follow:

1. In the first step, the methodology examines the spatial gaps between all vehicles (including both probe and non-probe vehicles inserted in the model in the previous interval) in a stopped queue within a distance from the stop line to infer the presence of new non-probe vehicles. Spatial gaps between probe (equipped) vehicles in a stopped queue are used as an indication of the existence of a new non-probe (unequipped) vehicle (Figure 2-3).
2. In the second step, the methodology inserts new vehicles in the detected gaps identified in step 1 by assuming an average vehicle length in the network.

.

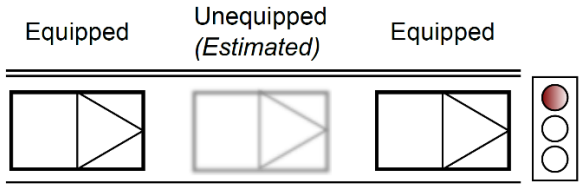


Figure 2-3: Estimation of an unequipped vehicle using gaps between equipped vehicles .

3. The third step involves simulating the movement of all vehicles (both probe and non-prove vehicles) in a microsimulation. This step is used to simulate the movement of non-probe vehicles after they leave the queue.
4. The methodology compares the reported locations of probe vehicles in the field against the positions of non-probe vehicles in the simulation. If the location of a probe vehicle overlaps with the simulated location of a non-probe vehicle, then it is assumed that the estimated location of the non-probe vehicle is incorrect and the vehicle is removed from the simulation.

However, the goal of Goodall's methodology is not to estimate the queue but to detect non-probe vehicles on arterials. Furthermore, the method is only applied to the region within 50 metres from the stop line and does not examine if the stopped queue is contiguous.

For the purpose of estimating the saturation flow rate using connected vehicle data, two key parameters are required:

1. The location of the last (most upstream) probe vehicle in the queue and the number of queued vehicles in front of it and behind the stop line (in the same lane).
2. The time that the last probe vehicle in the queue crosses the stop line.

Therefore, estimating the actual queue length is not necessary as the passing time of the last vehicle in the queue will be unknown if it is not a probe vehicle. Within this context, the queue estimation methods presented above have one or more of the following disadvantages which make them unsuitable for the purpose of estimating the saturation flow rate:

1. Their goal is to estimate the actual queue length;
2. They try to estimate queue length in the form of distance from the stop line instead of the number of queued vehicles. Although the length can be converted to the number of vehicles assuming an average spacing between vehicles, using a constant (assumed) value for jam density (i.e. density of vehicles in a stopped queue) and only the location of the last probe vehicle in the queue can result in large estimation errors;
3. They are designed for isolated intersections and do not consider the possibility that vehicles which are stopped on the approach link do not form a continuous queue to the stop line. For example, there may be two queues on the approach link. The downstream queue is formed as a result of the traffic signal but the upstream queue may result from a temporary

lane blockage (e.g. a stopped transit vehicle; a vehicle stopped to complete a turn into a mid-block side street or driveway, etc.).

4. They limit the length of the approach link over which they consider a queue can exist;
5. They rely on data from both probe vehicles and fixed sensors.

Therefore, the goal in this chapter is to present a methodology to estimate the number of queued vehicles behind a stop line using only probe (Connected Vehicles) data and that can be used for estimating the saturation flow rate.

2.2 Methodology

The methodology presented in this section consists of the following two major tasks:

1. Detect the furthest upstream CV in a stopped queue on a lane behind the stop line.
2. Estimate the total number of vehicles (CV and non-CVs) in the above detected queue using CV data as the only input to the system.

The following assumptions are made for the methodology presented in this section:

1. CV data (i.e. BSM Part I) is broadcasted at the rate of 10 times per second.
2. Data provided by BSM Part I are accurate and include vehicle size, speed, longitudinal position on the approach (i.e. distance from the stop line to the front bumper of the vehicle), and lateral position on the approach (i.e. lane currently in).
3. There is no significant delay in the transmission and receipt of the BSM Part I.
4. Signal timings are available (i.e. outputs of signal controllers).
5. CV are considered to be stopped (for the purposes of being part of a stopped queue), when their speed falls below a pre-determined threshold.
6. The DSRC wireless system provides spatial coverage sufficient to be able to receive information from all vehicles in the stopped queue at the intersection.

As mentioned above, the goal is to detect CVs and the total number of vehicles in the queue. At 100 percent penetration rate of connected vehicles, the number of vehicles on a lane in a queue easily can be obtained by detecting and counting all stopped CVs behind the stop line on the lane. However, When the LMP is less than 100% we have information from only a sample of the traffic stream (i.e. just the CVs). This leads to two challenges:

1. How to detect a queue and CVs that are part of the queue;
2. How to estimate the number of non-CVs in a queue.

Sections 2.2.1 and 2.2.2 present methodologies to address these two challenges.

2.2.1 Queue detection

The queue detection algorithm aims to identify connected vehicles that are queued behind the stop line. Connected vehicles are assumed to be part of a queue when their speeds are less than a threshold (S_Q). The goal is to detect CVs that are part of a queue in a manner that non-CVs can also be estimated using the information provided by CVs. Details about the non-CV estimation are presented in section 2.2.2.

Figure 2-4(a) presents a situation when the queue estimation starts right after the start of a phase. Vehicles are queued on a single lane behind a stop line when the green interval starts (t_0). At this time, the queue detection algorithm performs a detection procedure to detect all stopped CVs behind the stop line using a step-wise queue search technique. As presented in the next section, since gaps between CVs are used to estimate the number of non-CVs, using a step-wise queue detection approach prevents accumulation of errors in the estimation of non-CVs that may have occurred if the sum of all gaps between CVs in the whole queue was used to estimate the number of non-CVs. The step-wise queue search technique is also used to prevent detecting a discontinuous queue as a single continuous queue.

Starting from the stop line, the algorithm searches for the first stopped CV (speed $\leq S_Q$) within a defined searching distance (Ω). If a CV is found to be within this range and its speed is $\leq S_Q$, then it is added to the queue (in this case CV_1) and the non-CV estimation is performed by knowing the distance between CV_1 and the stop line. The algorithm then resets the searching point from the stop line to the location of the recently detected CV in the queue (CV_1), the search for additional stopped connected vehicles continues. As shown in Figure 2-4(a), both CV_1 and CV_2 are stopped with speeds $\leq S_Q$, and therefore, CV_2 is added to the queue and the searching point is reset to the location of CV_2. At the start of the green interval no additional stopped CVs are detected so the search process terminates for the present time.

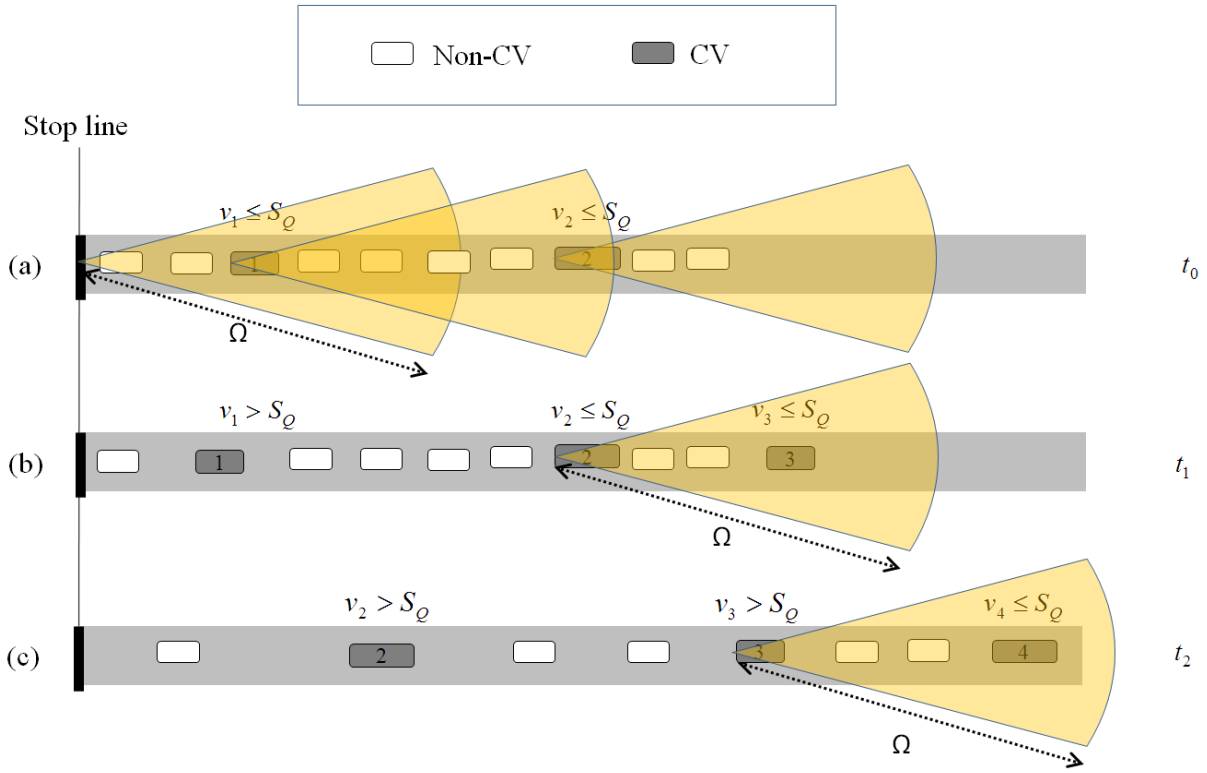


Figure 2-4: Step-wise queue detection.

After the initial queue detection at the start of the green, the algorithm continues to update the queue (and the searching point) during the green interval and to add new CVs to the queue when the following three conditions apply:

1) The speed of the last known CV in the queue $\le S_Q$ (i.e. the CV at the current searching point)
2) The speed of any new CV $\le S_Q$;
3) The new CV is within the searching distance (Ω).

Figure 2-4(b) illustrates a situation in which at some time during the green interval (t_1), CV_3 joins the end of the queue. In this case, since CV_2 is still in the stationary mode with its speed $\le S_Q$, CV_3 is added to the queue and the searching point resets to the location of CV_3. However, as shown in Figure 2-4(c), at time t_2, CV_4 enters the searching distance of CV_3 with its speed less than S_Q, but in this case, the speed of CV_3 (the end of the queue) > S_Q, and therefore CV_4 is not added to the queue. Both CVs must be stationary because this provides a high likelihood that all the non-CVs between the two CVs are also stationary and at jam density. If the downstream CV has already

begun to discharge, then the vehicles between the downstream and upstream CV will not be at jam density, increasing the error for estimating the number of non-CVs between the upstream and downstream CVs.

The goal of using a search distance threshold in the step-wise queue detection method is to prevent errors in the non-CV estimation caused by detecting discontinuous queues while at the same time to be able to identify the furthest CV in a queue from the stop line. At low level of market penetration (LMP) of CVs the distance between stopped CVs in a queue can be large, and therefore a large search distance is required to detect all queued CVs. However, a large search distance may result in errors in the estimation of non-CVs due to the presence of multiple queues for instance, when there is a shockwave in the queue or stopped CV in mid-block. On the other hand, at high LMP, the average distance between consecutive CVs in the stopped queue will be much shorter and therefore a short search distance is sufficient to detect all CVs. Thus, the searching distance threshold should vary with the LMP to reduce errors in identifying vehicles that are part of a contiguous queue.

The above discussion is better illustrated in Figure 2-5. Figure 2-5 (a) and (b) present an intersection approach with a midblock street where some vehicles executing left turn movements. Figure 2-5 (a) shows an example when the market penetration of connected vehicles is low and a larger search distance is used in the queue detection method because the average distance between consecutive CVs is large. As shown in this figure, using a large search distance may result in the detection of CV_3, which should not be included as it is not part of the queue created by the traffic signal operations, rather it is stopped to complete a left-turn. Figure 2-5 (b) presents the same geometry but when the LMP is high, then the average distance between consecutive CVs in the queue is shorter. Using a shorter search distance in this case prevents the incorrect detection of CV_6, which is waiting in the midblock to execute a left turn.

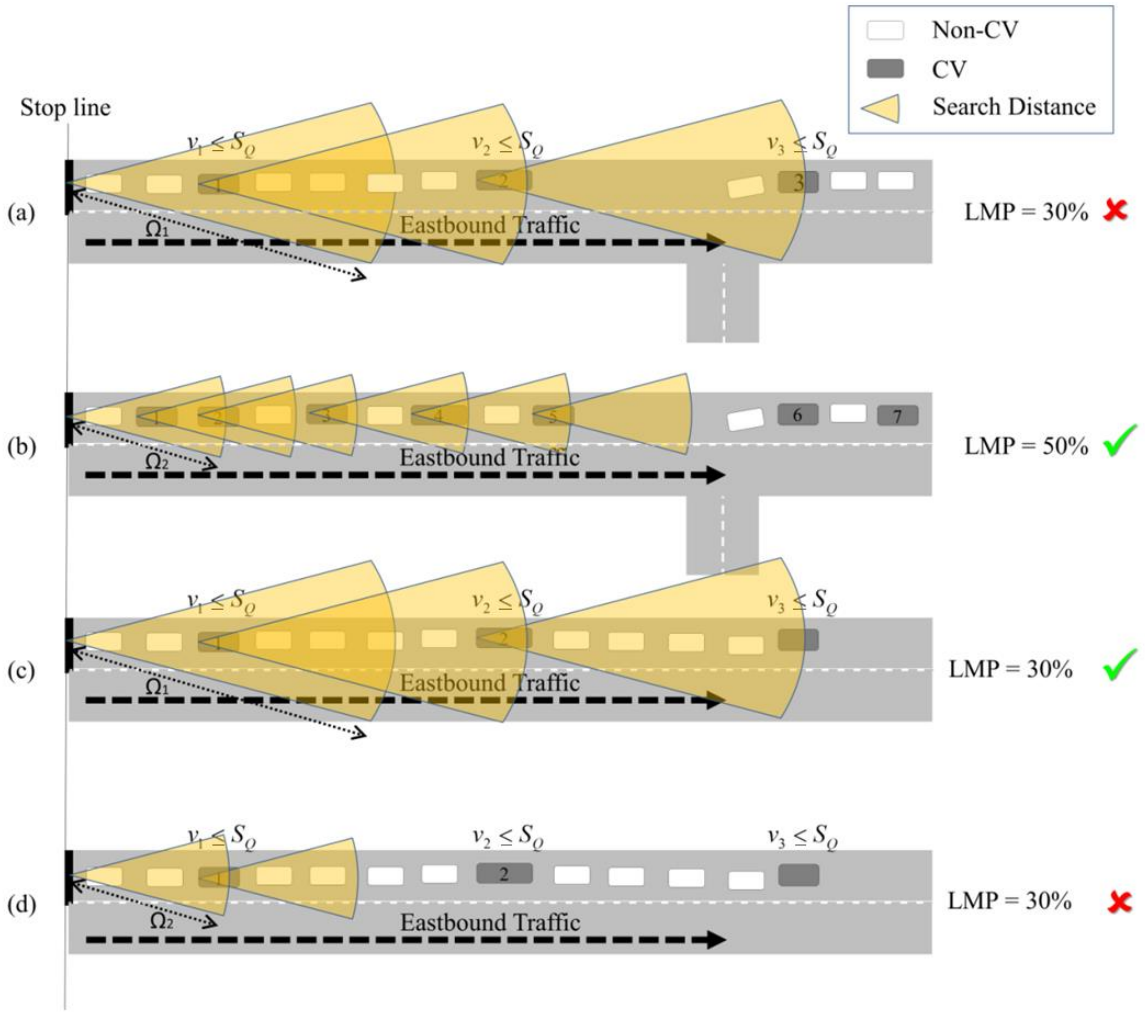


Figure 2-5: The impact of using different search distance thresholds at different CV LMPs on the queue detection accuracy.

Figure 2-5 (c) and (d) present an intersection approach with no mid-block street. Figure 2-5 (c) uses a large search distance to detect CVs in a queue at low market penetration rate. The large search distance allows the method to detect all CVs that are part of the queue. However, as shown in Figure 2-5 (d), using a shorter search distance can result in underestimating the queue as CVs are spaced far apart.

We consider an algorithm that is based on CV data to be robust if it is capable of operating at low LMP while being able to improve its performance and estimation accuracy as the LMP increases. Therefore, it is desirable to enable the algorithm to dynamically adjust its search distance threshold based on a dynamically estimated penetration rate of connected vehicles in the field.

To achieve the above goal, a methodology was developed to estimate the CV penetration rate in the network. As presented in Equation (2-3), the algorithm estimates the real-time LMP using the known number of detected CVs in queues and the estimated number of non-CVs in the same queues.

$$CV\ LMP_t = \frac{N_t^{CV}}{N_t^T} \times 100 \tag{2-3}$$

Where,

$CV\ LMP_t$ = The estimated level of market penetration (%) of connected vehicles at the time interval *t*;
N_t^{CV} = Total number of connected vehicles detected in all stopped queues in the area under study at the time interval *t*, (*veh*);
N_t^T = Total number of vehicles (CVs and estimated non-CVs) in all stopped queues in the area under study at the time interval *t*, (*veh*).

The above estimated CV LMP at each time interval is smoothed using a conventional smoothing technique such as the exponential smoothing method presented in Equation (2-4). Note that any smoothing technique suitable for real-time application could be applied.

$$CV\ LMP_t^{Smoothed} = (\alpha) CV\ LMP_t + (1-\alpha) CV\ LMP_{t-1}^{Smoothed} \tag{2-4}$$

Where,

$CV\ LMP_t^{Smoothed}$ = Smoothed estimated CV LMP at the time interval *t*;
$CV\ LMP_{t-1}^{Smoothed}$ = Smoothed estimated CV LMP at the previous time interval (*t*-1);
α = Smoothing weight ($0 < \alpha < 1$).

A value of $\alpha = 0.5$ was chosen for this methodology.

Section 2.2.2 provides details on using CV data to estimate non-CVs and hence total number of vehicles in a queue.

Using the estimated CV LMP, the search distance is dynamically adjusted accordingly as presented in Equation (2-5). In this equation, the minimum search distance (Ω_{min}) is defined by the operator and should be selected such that all CVs in a queue are detected when CV LMP is 100%. The maximum search distance (Ω_{max}) is also defined by the operator considering the geometry of the approach such as the location of mid-blocks, bus stops, and the distance to the upstream

intersection. The impact of different values of $\Omega_{\min}$ and $\Omega_{\max}$ on the search distance is shown in Figure 2-6.

$$\Omega_t = \min\left(\frac{\Omega_{\min}}{CV\ LMP_t^{Smoothed}}, \Omega_{\max}\right) \quad (2\text{-}5)$$

Where,

Ω_t = The search distance at time interval *t*;

$\Omega_{\min}$ = The minimum search distance defined by the user;

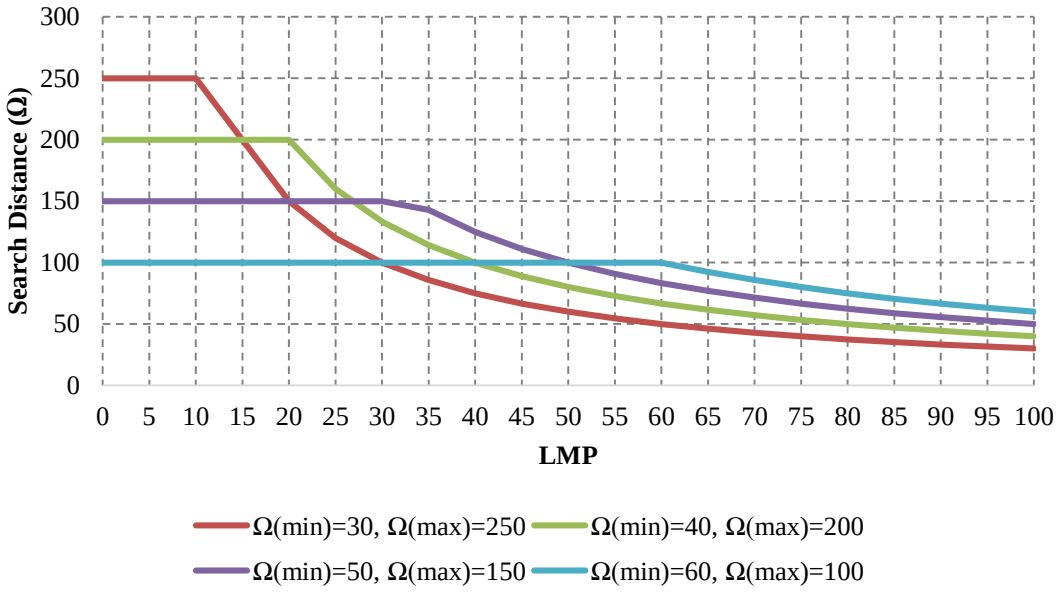


Figure 2-6: The impact of Ω_{min} and Ω_{max} on the search distance.

2.2.2 Non-CV estimation

The objective of the non-CV estimation method is to estimate the number of non-CVs in a queue behind a stop line and the most upstream detected CV in a queue when CV LMP is less than 100 percent.

In this methodology, CVs in a queue on a lane are labelled sequentially as $j = 1, 2, \ldots, N$ beginning from the stop line and moving upstream as they are detected by the queue detection algorithm presented earlier. Then the number of vehicles (CVs and non-CVs) in the queue between the stop line and each CV is estimated using Equation (2-6) by assuming an average distance headway between vehicles (γ) in the queue.

$$N_j^T = \left(\frac{P_j - P_{j-1} - \Delta_{j-1}}{\gamma} \right) + 1 + N_{j-1}^T \quad (2\text{-}6)$$

Where,

N_j^T = Number of CVs and estimated non-CVs in the stopped queue in the current lane measured from the stop line up to and including the j^{th} CV, (*veh*);

N_{j-1}^T = Number of CVs and estimated non-CVs in the stopped queue in the current lane measured from the stop line up to and including the j-1^{th} CV, (*veh*);

P_j = The distance from the stop line to the front of the j^{th} CV in the stopped queue, (*m*);

P_{j-1} = The distance from the stop line to the front of the j-1^{th} CV in the stopped queue, (*m*);

Δ_{j-1} = Length of the j-1^{th} CV, (*m*);

γ = Assumed average distance headway for vehicles in a stationary queue, (*m*).

The non-CV estimation is performed at the start of the green interval when the initial queue is detected by the queue detection algorithm and continues during the green interval. Recall that the speed of each identified CV is used to determine if the CV is part of a stationary queue (i.e. if the speed is less than a threshold S_Q). Two methods are proposed to estimate the number of non-CVs in a queue.

The first methodology uses a small value of S_Q (S_Q = 2 km/h) to detect stationary CVs. As presented in Equation (2-7), the parameter γ consists of an average vehicle size and an average separation distance between stopped vehicles and is used to estimate the number of non-CVs between two consecutive CVs in the queue. In the first methodology, one value of γ is used for estimating the number of non-CVs both at the start of the green and subsequent non-CVs detections during the green interval.

$$\gamma = \Delta' + \varepsilon \quad (2\text{-}7)$$

Where,

Δ' = The assumed average vehicles size in a queue, (*m*);

ε = The assumed average separation distance between vehicles in a queue, (*m*).

The second methodology incorporates a larger value of S_Q (S_Q =5 km/h) which increases the chance of detecting a queue when vehicles do not decelerate fast enough to become completely stationary. However, using a larger value of S_Q may result in larger errors in the non-CV estimation as vehicles are moving in the queue. Therefore, the second methodology considers two different values of γ for non-CV estimation performed at the start of the green interval when the queue is stationary and during the green interval when new CVs join the tail of the queue.

Figure 2-7 shows a space-time diagram and depicts the trajectories of vehicles close to an intersection. The signal turns green at time t_0 and the queue detection algorithm identifies CV 1, 2, and 3 as part of the initial stationary queue. This is also shown in Figure 2-8 (a). At this moment vehicles including CVs are in a stationary queue with their speed close to 0 km/h and therefore a separation distance of ε_1 corresponding to the jam density can be assumed to estimate the number of non-CVs.

As time continues, the queue detection method is repeatedly applied to determine if additional CVs have joined the tail of the queue. As described in Section 2.2.1, a new CV is detected and added to a queue when: 1) it is within the searching distance; 2) has speed ≤ S_Q; and 3) the last known furthest upstream CV in the queue is also traveling with speed ≤ S_Q.. Assuming that condition 1 applies to all of the vehicles in Figure 2-7, then at time t_1, the speed of CV_4 reaches 5km/h and CV_3 has also started accelerating with its speed = 3.1 km/h. In this situation the first methodology with S_Q =2 km/h is unable to detect and add CV_4 to the queue as queue detection conditions 2 and 3 are not satisfied. On the other hand, these queue conditions are satisfied if the second methodology with S_Q = 5 km/h is used.

However, as shown in Figure 2-8 (b), using (ε_1) may result in overestimation of the number of non-CV between CV_3 and CV_4 as vehicles are not completely stationary. Therefore, to compensate for the larger spacing that vehicles have at a higher speed, a larger value of average separation distance (ε_2) is used to calculate γ in the model.

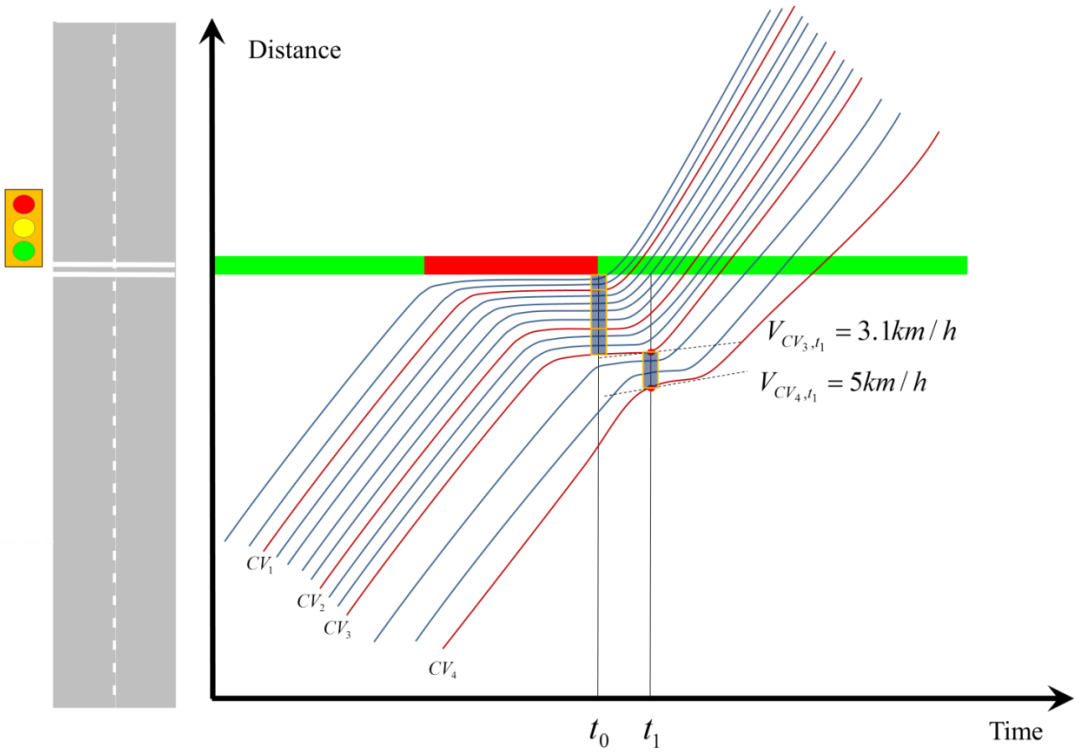


Figure 2-7: Vehicle trajectories close to stop line

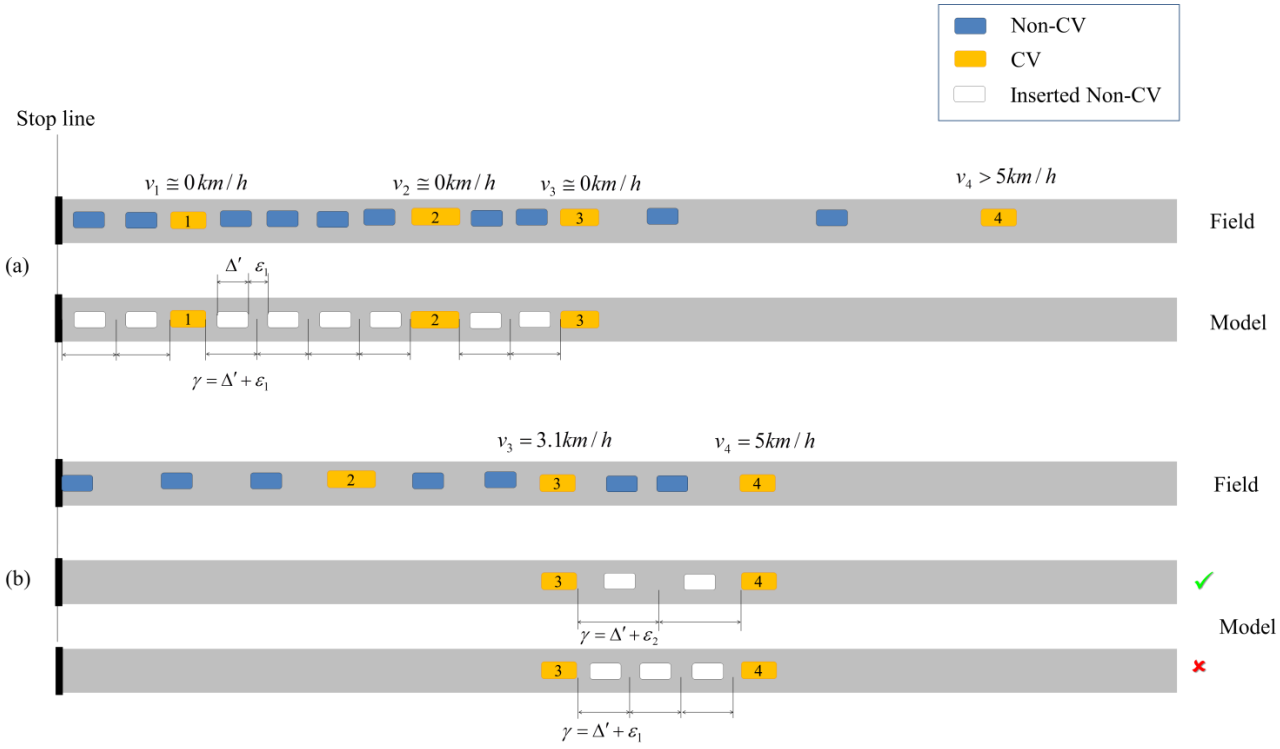


Figure 2-8: Using two values of vehicle spacing for Non-CV estimation.

The value of γ plays an important role in the level of accuracy of the non-CV estimation methodology. Changes in this value can result in over or under estimation of the number of vehicles in a queue. The value of γ may change due to various reasons, for instance, time of day restrictions for heavy vehicles for accessing specific roads, or diversion of traffic from a highway with high commercial vehicles volume to a local street due to an incident or construction on a section of the highway. In such situations changes in the vehicle stream composition will result in increased differences between the actual average vehicle length and the assumed average vehicle length (Δ') and increased errors in the estimated number of vehicles in the stationary queue.

One of the data elements included in the Basic Safety Message part I is the vehicle size, which includes the vehicle's length and width. Therefore, there is the potential to overcome the above issue by using vehicle size information obtained from CVs as a random sample of the traffic.

As presented in Equation (2-7), the parameter γ includes two terms: Δ', which represents the average vehicle size, and ε which is the average separation distance between vehicles. Using the vehicle size data element included in the BSM part I the value of Δ' can be updated in real-time. In the presented methodology, lengths of connected vehicles are recorded when they cross the stop line on each lane. An average of all the recorded lengths is calculated every 5 minutes for each lane, and as presented in Equation (2-8) an exponentially weighted average ($\alpha = 0.5$) is used to smooth the average vehicle lengths obtained at each interval. This value is then used to estimate the number of non-CVs in a queue on each lane.

$$\Delta'_{lt} = (\alpha) \sum_{t-1<t'<t} \Delta_{jlt'} + (1-\alpha)\Delta'_{l(t-1)} \qquad (2\text{-}8)$$

Where,

Δ'_{lt} = The average of vehicle sizes on lane *l* calculated for the time interval *t*, (*m*);

$\Delta'_{l(t-1)}$ = The average of vehicle sizes on lane *l* calculated for the time interval *t-1*, (*m*);

$\Delta_{jlt'}$ = The length of the j^{th} CV crossing the stop line on lane *l* at time *t´*, (*m*);

α = Smoothing weight ($0 < \alpha < 1$).

2.3 Evaluation

The proposed methodology is evaluated using a microscopic traffic simulation environment (VISSM version 5.4). VISSIM is a discrete time simulation model that models the behaviour of individual traffic signals and vehicles with a time step of $1/10^{th}$ of a second.

For these evaluations, the detailed trajectory data (i.e. speed and position at each $1/10^{th}$ of a second) from each vehicle in the simulation were used and post processed to extract the measures of interest. Vehicles from the traffic stream were randomly designated as connected vehicles depending on the desired LMP. CVs and non-CVs were simulated using the same behavioural models. The only difference is that we assumed data (equivalent to the data contained in the BSM part 1) were available from the vehicles designated as CVs and could be used within the proposed model. No BSM data was assumed available from the non-CVs. Ground truth measures were computed using data from all simulated vehicles.

To assist in the presentation of the results in this section, a naming convention is used to refer to combinations of evaluated methodologies as shown in Figure 2-9. In this figure, the dynamic and static search distance refer to different search distances used to detect CVs in a queue (refer to section 2.2.1 and equation (2-5)); constant distance headway refers to case when a constant jam density is used for a queue and the dynamic distance headway is when the jam density is updated in real-time using vehicle length information obtained from CVs (refer to section 2.2.2 equation (2-8)). Finally, methods are differentiated based on the speed thresholds used to detect CVs as part of a queue (refer to section 2.2.2).

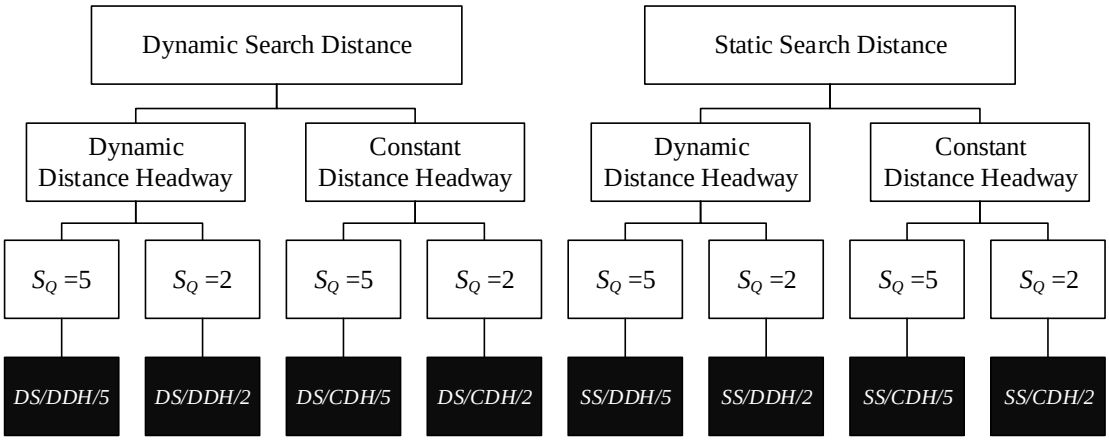


Figure 2-9: Naming conventions used for proposed methodologies.

The following sections present evaluations of the presented methodologies. In all analyses, the ground truth queue length is calculated in VISSIM when speeds of vehicles become ≤ 5 km/h.

2.3.1 Dynamic vs. Constant Distance Headway

An evaluation was performed to examine the impact of using dynamic versus constant values of distance headway in estimating the number of non-CVs in a queue. Compared methodologies in this section are highlighted in Figure 2-10. As presented in Figure 2-11, the evaluation was performed for an isolated intersection controlled by a fully actuated signal controller with two phases each with a minimum and maximum green times of 4 and 35 seconds, and an intergreen (amber + all red) of 5 seconds.. All approaches of the intersection included two lanes (one through and one shared through/right). No left turns were modelled.

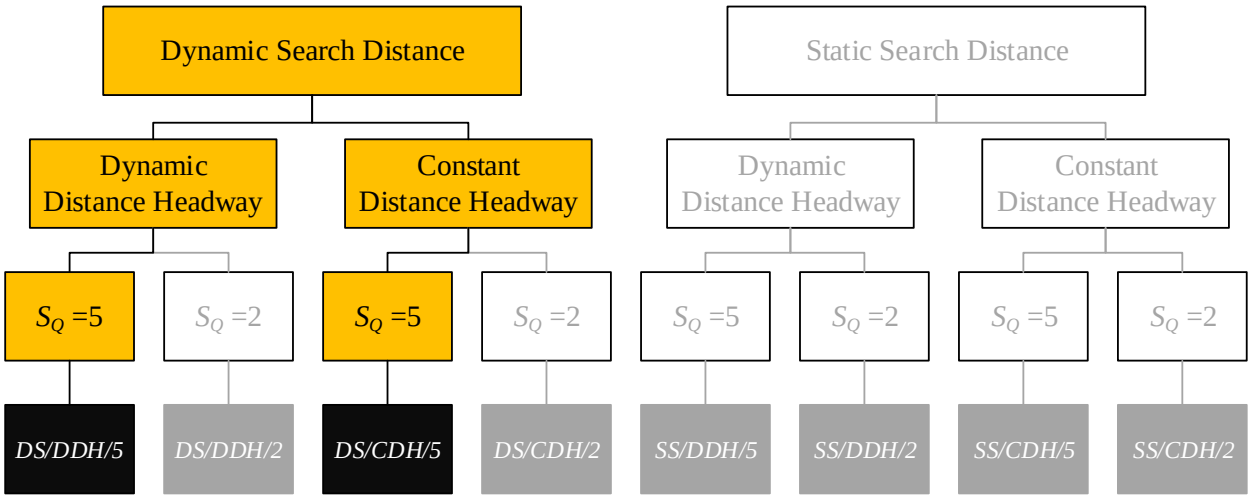


Figure 2-10: Compared methodologies in section 2.3.1.

Ten levels of CV penetration rate (10% to 100%) were considered, and for each level, 10 simulation runs (with different random seeds) were conducted each for a duration of 4800 seconds. The first 600 seconds of each run was considered a warm-up period and data were collected from the remaining 4200 seconds. This resulted in a total of 100 simulation runs for a total of 117 hours of simulated traffic conditions.

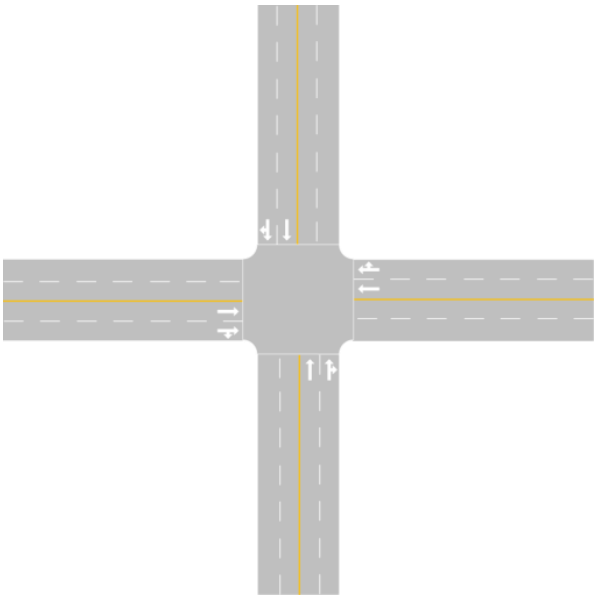


Figure 2-11: Study area used in the evaluation of dynamic versus constant distance headway.

As presented in Table 2-1, the traffic composition on each approach includes passenger cars (4.11 m to 4.76 m long) and truck trailers (TT) (17.22 m long). During each simulation run, the percentage of truck trailers in the network was increased at each 10 minutes interval. To calculate γ, in Equation (2-7), the average separation distance (ε) of 3 metres was used during the non-CV estimation for the initial queue, and 6 metres used when the estimation was performed for the subsequent updates in the queue during the green interval. When DS/CDH/5 was used, the average vehicle length (Δ') was assumed to be 5 metres and when DS/DDH/5 was used this value was updated using CV data.

Table 2-1: Traffic composition and volume used on approaches.

Interval (s)	**0-600**	**600-1200**	**1200-1800**	**1800-2400**	**2400-3000**	**3000-3600**	**3600-4200**	**4200-4800**
Volume (vph)	2000	2000	1700	1500	1500	1500	1500	1500
Comp. (%)	TT=0 Cars=100	TT =0 Cars=100	TT =10 Cars=90	TT =20 Cars=80	TT =30 Cars=70	TT =40 Cars=60	TT =50 Cars=50	TT =50 Cars=50

The goal of this evaluation was to examine the performance and accuracy of the two proposed methods for estimating the number of non-CVs in the stationary queue between the stop line and the most upstream CV. This is not the same as estimating the total number of vehicles in the stationary queue. Consequently, ground truth was also determined as the number of vehicle in the

queue from the stop line to the most upstream CV in the queue. This is illustrated in Figure 2-12. Ground truth is determined as the number of vehicles between the stop line and the most upstream CV that is determined to be part of the stationary queue. At the start of green (time t_0) all the vehicles between the stop line and CV_2 are part of the stationary queue. Later in the green interval, additional vehicles, including CV_3 join the tail of the queue. However, only those vehicles which occur between the stop line and the most upstream CV (in this case CV_3) are counted as parted of the stationary queue and considered to be the ground truth against which the proposed methods are compared.

Figure 2-13 presents the results of the estimated number of vehicles using DS/DDH/5 and DS/CDH/5 methodologies. The graphs present the Root Mean Squared Error (RMSE) of the estimations using each of the methodologies against the ground truth calculated at each cycle for each LMP using an average of all 10 simulation runs (Equation (2-9)). RMSE is calculated using the average of all four approaches of the intersection as all have the same characteristics.

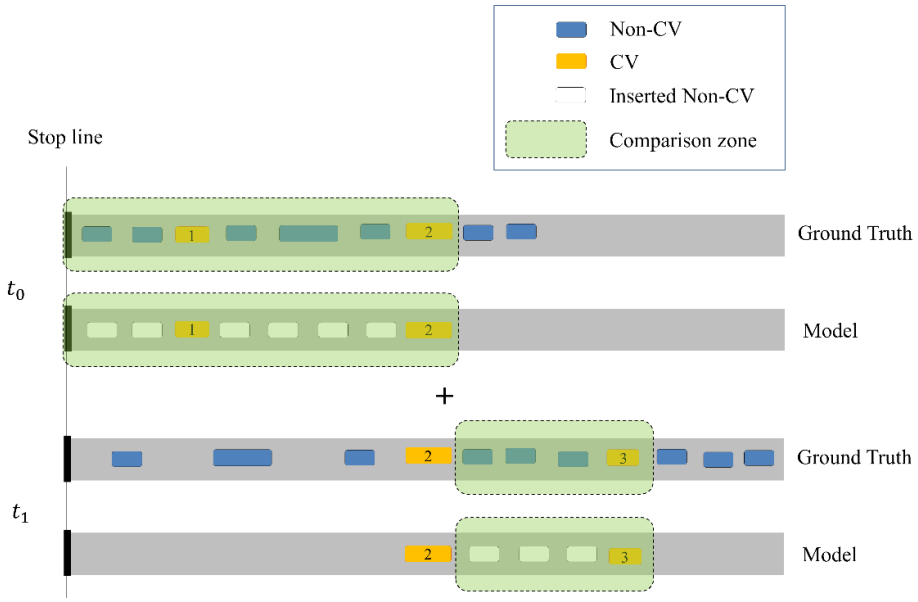


Figure 2-12: Comparison between the ground truth and the estimated number of vehicles in a queue using proposed methodologies.

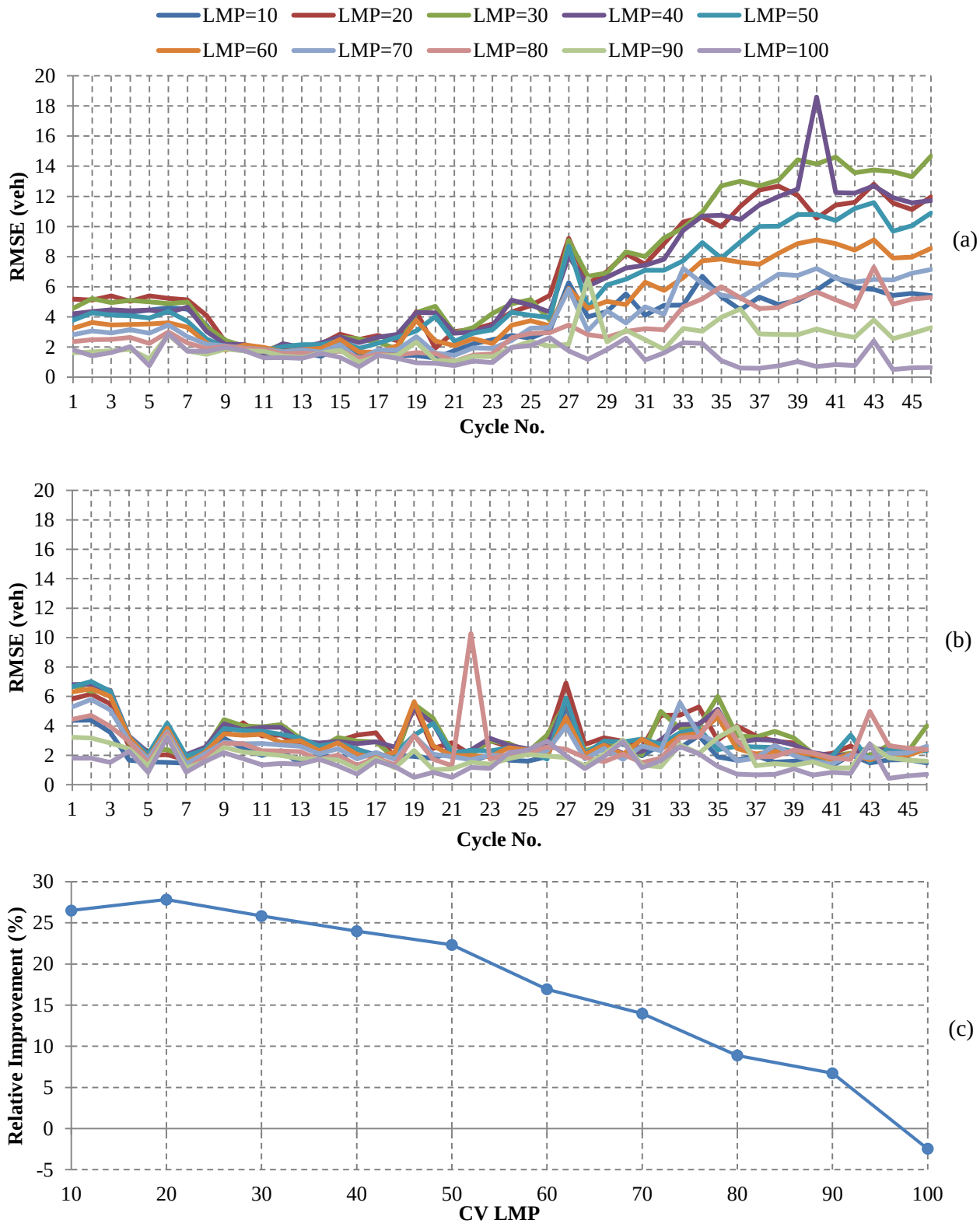


Figure 2-13: (a) Queue estimation RMSE using DS/CDH/5, (b) Queue estimation RMSE using DS/DDH/5 method, (c) Relative improvement of DS/DDH/5 over DS/CDH/5 at each LMP.

$$RMSE_C^{LMP} = \frac{1}{m}\sum_{a=1}^{m}\sqrt{\frac{\sum_{R=1}^{10}\left(N'_{C,LMP,R} - \hat{N}_{C,LMP,R}\right)^2}{10}} \quad (2\text{-}9)$$

Where,

$RMSE_C^{LMP}$ = Root Mean Squared Error calculated at cycle *C* when CV penetration rate is LMP averaged over 4 intersection approaches, (*veh*);

$N'_{C,LMP,R}$ = Ground truth number of queued vehicles between the stop line and CV_J at cycle *C* obtained during simulation run *R*, (*veh*);

$\hat{N}_{C,LMP,R}$ = Estimated number of queued vehicles between the stop line and CV_J using one of the proposed methodologies at cycle *C*, when CV penetration rate is *LMP* and obtained during simulation run *R*, (*veh*);

m = Number of approaches (m=4).

Figure 2-13 (a) presents the estimation results using the DS/CDH/5 methodology. The results show the DS/CDH/5 methodology fails to correctly estimate the number of vehicles when there is an unexpected change in the traffic composition and when the penetration rate of connected vehicles is low. Changes in the traffic composition resulted in deviations of the estimations using the CV data from the ground truth beyond the 20th cycle when the average vehicle lengths in the simulation started to increase and deviated from the assumed one (5m). The figure shows that RMSE can increase to about 15 vehicles when half of the traffic composition in the network includes truck trailers. However, as shown in Figure 2-13 (b) the DS/DDH/5 methodology can successfully adapt to the changes in the traffic composition during simulations and estimate the number of vehicles in a queue using CV data with low RMSE at all penetration rates of connected vehicles. The figure shows that majority of the estimations have RMSE < 6 vehicles.

The relative improvement in the queue estimation error using DS/DDH/5 over DS/CDH/5 was also investigated. Equation (2-10) was used to calculate the percentage of improvement in the RMSE at each LMP as presented in Figure 2-13. In this graph, positive values indicate improvements in the estimation errors using the dynamic headway distance methodology. The results indicate that the vehicle size data element included in BSM I can maintain the accuracy of the non-CV estimation at different traffic compositions. Consequently, DDH is recommended over CDH for locations where changes in the traffic composition may occur as the result of, for instance,

diversion of traffic from nearby routes, or where time varying traffic stream composition is expected.

$$RI_{LMP} = \frac{1}{n} \times \sum_{C=1}^{n} \frac{RMSE_{CDH,LMP,C} - RMSE_{DDH,LMP,C}}{RMSE_{CDH,LMP,C}} \times 100 \quad (2\text{-}10)$$

Where,

RI_{LMP} = Relative improvement in RMSE of estimation using DS/DDH/5 over DS/CDH/5

$RMSE_{CDH,LMP,C}$ = Queue estimation RMSE using DS/CDH/5 method when CV penetration rate is *LMP*, (*veh*);

$RMSE_{DDH,LMP,C}$ = Queue estimation RMSE using DS/DDH/5 method when CV penetration rate is *LMP*, (*veh*);

n = Total number of cycles.

2.3.2 Estimation of CV LMP

The simulation scenarios in the previous section were used to evaluate the accuracy of estimating the CV LMP. Using Equations (2-3) and (2-4), CV LMP was estimated every 5 minutes by aggregating data from all 4 approaches of the intersection obtained from the DS/DDH/5 and DS/CDH/5 methods.

The objective in estimating the CV LMP is to be able to dynamically track changes in the LMP over time so that the system can automatically adjust as LMP increases when CV technology becomes more widely adopted. As we are not attempting to capture short-term fluctuations in CV LMP in this analysis, the last smoothed estimated value of the LMP using each method (at cycle 45) was extracted and compared to the actual LMP.

Figure 2-14 presents the LMP estimation results using DS/DDH/5 and DS/CDH/5 methodologies vs the ground truth LMP (45 degree line). The results show that both methodologies overestimate the LMP by a relatively small amount when the ground truth CV LMP is low (LMP < 20%). This was expected as at low LMPs, queues that do not consist of any connected vehicles cannot be captured by any of the methodologies resulting in undercounting non-CVs in the network and therefore overestimation of the LMP.

DS/DDH/5 shows an improvement in the accuracy of the estimated LMP as the ground truth CV LMP increases. However, DS/CDH/5 tends to underestimates the CV LMP (LMP > 20%) and has a greater deviation from the ground truth LMP. These results are consistent with the non-CV estimation results from the previous section. As presented in the previous section, DS/CDH/5 overestimates the number of non-CVs in queues when the number of trucks increased in the network causing in underestimation of the CV LMP in DS/CDH/5.

In general, the results show that LMP estimation methodology implemented using DS/DDH/5 can be used to estimate the percentage of vehicles equipped with the connected vehicle technology in a signalized network.

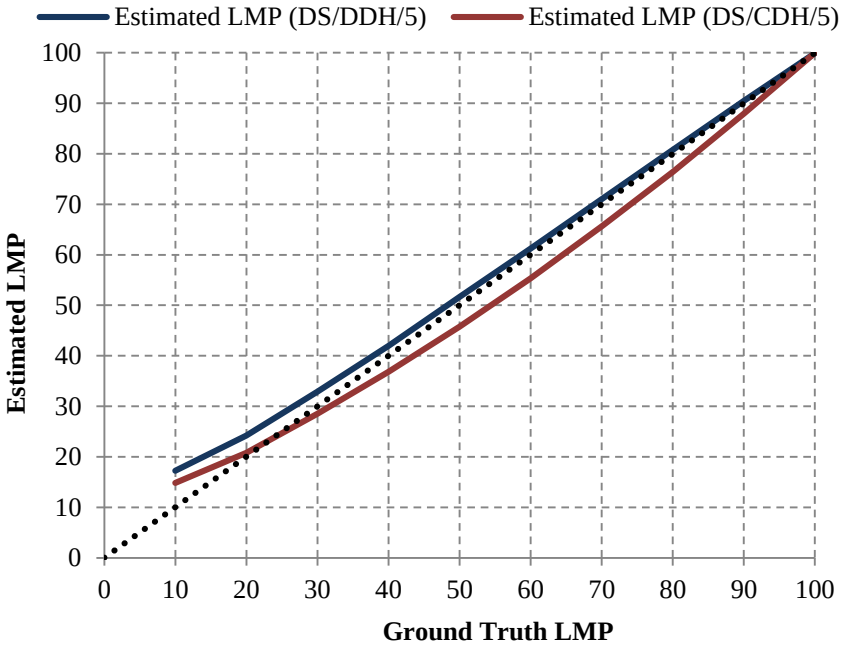


Figure 2-14: Estimating CV penetration rate.

2.3.3 Comparison of the Proposed End of Queue Detection Methodologies.

This section presents an evaluation of the two proposed tail of queue detection methodologies using CV data shown in Figure 2-15. The goal of the analysis is to determine which of the two proposed methods performs better and provides higher queue length estimation accuracy as well as how the accuracy is affected by changes in the traffic composition.

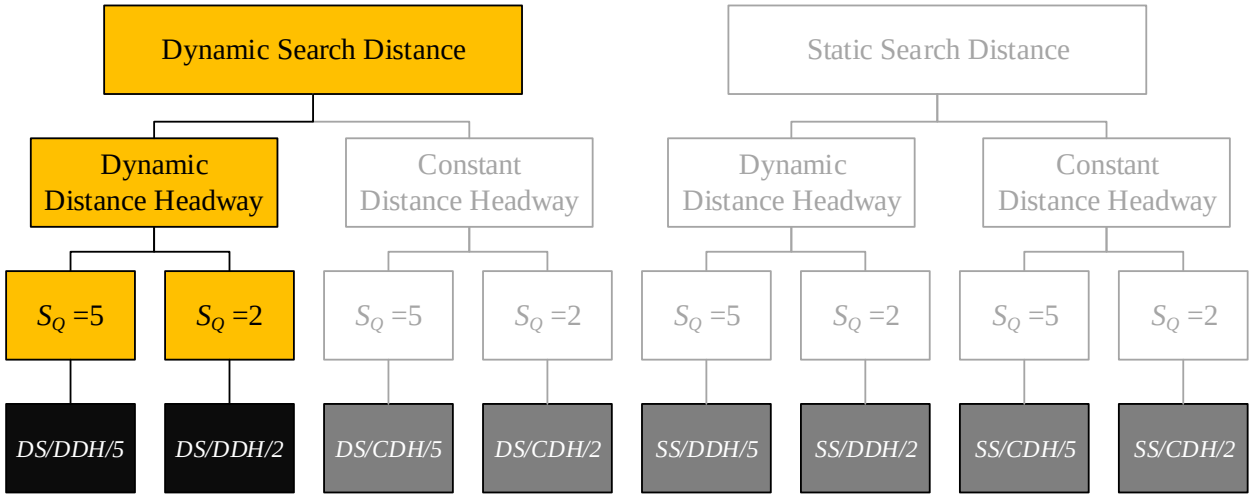


Figure 2-15: Compared methodologies in section 2.3.3.

The study area, simulation, and the number of runs used for this analysis were identical to the one presented in the previous sections (Figure 2-11), but queues were estimated using two variations of the Dynamic Distance Headway (DDH) method, by incorporating each of two proposed speed threshold methods. When DS/DDH/2 was used, the average separation distance (ε) was assumed to be 3 metres and for DS/DDH/5 the average separation distance was assumed to be 3 metres during the non-CV estimation for the initial queue (ε_1 = 3m) and 6 metres for subsequent queue updates during the green interval (ε_2 = 6m).

The goal of the proposed tail of queue detection methodologies is to find the tail of a queue as close as possible to its actual one. However, these methods do not tend to estimate the actual location of the tail of the queue as this is not useful in the saturation flow rate estimation methodology based only on CV data as the stop line crossing time of only CVs are known. However, to examine which of the methods estimate the number of vehicles in a queue closer to the actual queue, the evaluation of these methods in this section is done against the ground truth tail of queue.

The root mean squared error of the estimated queue (number of vehicles) versus the ground truth at each cycle was calculated using Equation (2-11).

$$RMSE_C^{LMP} = \frac{1}{m}\sum_{a=1}^{m}\sqrt{\frac{\sum_{R=1}^{10}\left(N_{C,R} - \hat{N}_{C,LMP,R}\right)^2}{10}} \qquad (2\text{-}11)$$

Where,

$RMSE_C^{LMP}$ = Root Mean Squared Error of the estimated queue (number of vehicles) calculated at cycle *C* when CV penetration rate is LMP, (*veh*);

$N_{C,R}$ = Ground truth queue length (number of vehicles) at cycle *C* obtained during simulation run *R*, (*veh*);

$\hat{N}_{C,LMP,R}$ = Estimated queue length using one of the proposed methodologies at cycle *C*, when CV penetration rate is *LMP* and obtained during simulation run *R*, (*veh*);

m = Number of approaches (m=4).

Figure 2-16 (a) and (b) present the results using DS/DDH/5 and DS/DDH/2 methodologies respectively. The actual queue length at each cycle using average of all 4 approaches is shown in Figure 2-16(c). In general, the results show that using DS/DDH/5 with two values of separation distance performs better than using DS/DDH/2 with a single constant value of separation distance. The RMSE of the queue estimations prior to cycle 20th in Figure 2-16 (a) are in the range of 0 to 5 vehicles for LMP $\geq$ 30%, while this range is between 5 and 10 in Figure 2-16 (b) using DS/DDH/2. The results also show that at LMP=100% and beyond cycle 20 when the percentage of truck trailers is 30% and more, DS/DDH/2 exhibits larger queue estimation errors. This is because heavy vehicles have different driving behaviours (i.e. decelerate and accelerate more slowly) which makes them difficult to be detected by the DS/DDH/2 methodology. However, DS/DDH/5 method can capture these vehicles and performs well throughout the simulations with errors close to zero.

Increase in the error beyond cycle 20 was observed for both methodologies when CV LMP < 100%. In addition to the influence of slow moving vehicles, the increase in the error at low LMPs is also due to longer queues beyond cycle 20 when shockwaves start to form in the queue. However, in contrast to DS/DDH/2, the increase in the error using DS/DDH/5 is proportional to the decrease in CV LMP and the errors generally tend to be lower when compared to DS/DDH/2.

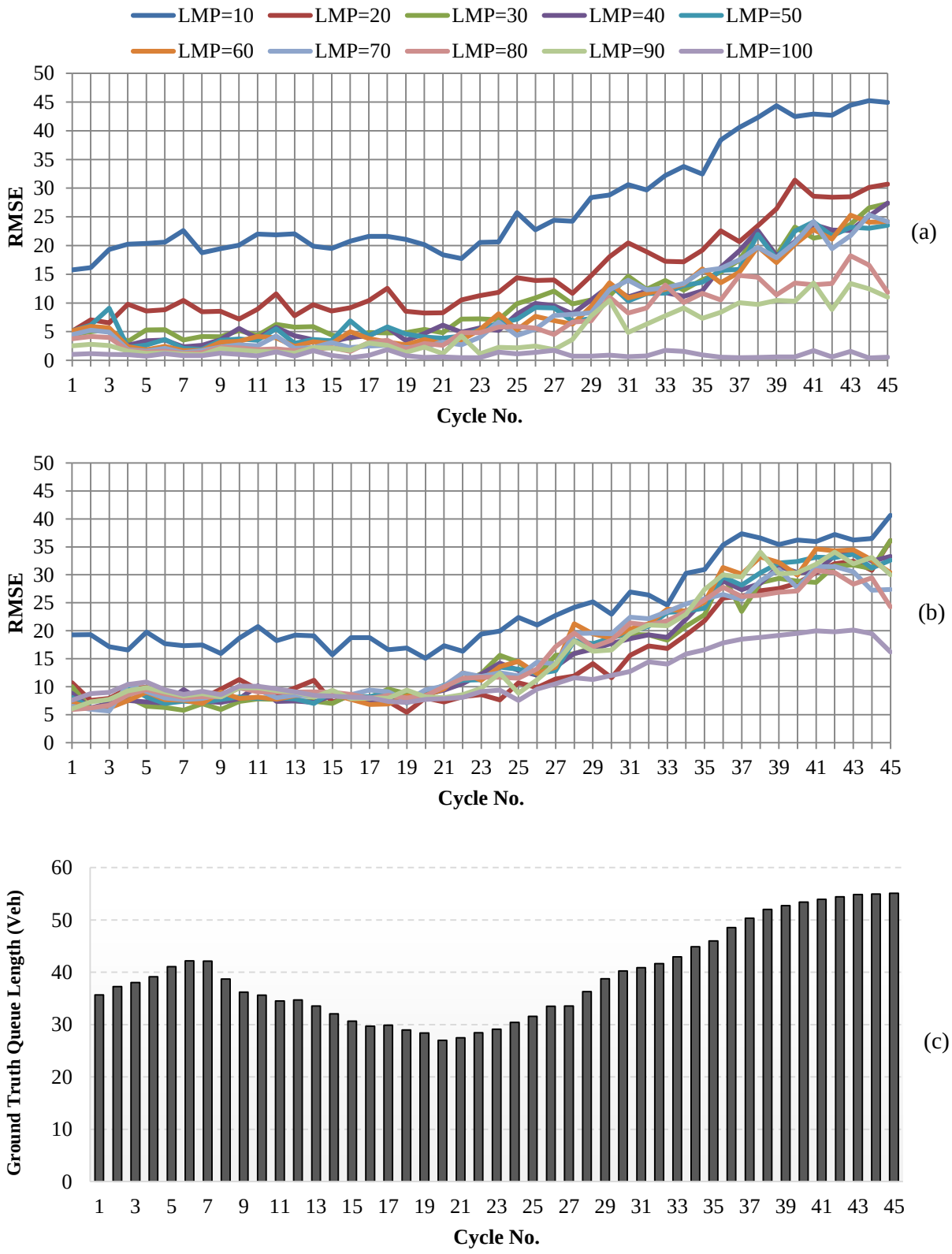


Figure 2-16: Queue estimation error using (a) DS/DDH/5, vs. (b) DS/DDH/2, (c) average ground truth queue length at each cycle

Figure 2-17 presents the relative improvement in the errors achieved by using DS/DDH/5 instead of DS/DDH/2 calculated using Equation (2-12). As the graph shows, DS/DDH/5 performs better than DS/DDH/2 when CV LMP > 20% but slightly poorer when CV LMP ≤ 20% which indicates that using a higher speed threshold for queue detection can increase the error at very low LMPs. However, an increase in value of ε_2 may reduce this error. In general, the results suggest that DS/DDH/5 performs better than DS/DDH/2.

$$RI_{LMP} = \frac{1}{n} \times \sum_{C=1}^{n} \frac{RMSE_{DS/DDH/2,LMP,C} - RMSE_{DS/DDH/5,LMP,C}}{RMSE_{DS/DDH/2,LMP,C}} \times 100 \quad (2\text{-}12)$$

Where,

RI_{LMP} = Relative improvement in RMSE of estimations using DS/DDH/5 over DS/CDH/5 when CV penetration rate is *LMP, (%);*

$RMSE_{DS/DDH/5,LMP,C}$ = RMSE of queue estimation using DS/DDH/5 method when CV penetration rate is *LMP*, (*veh)*;

$RMSE_{DS/DDH/2,LMP,C}$ = RMSE of queue estimation using DS/DDH/2 method when CV penetration rate is *LMP*, (*veh)*;

n = Total number of cycles.

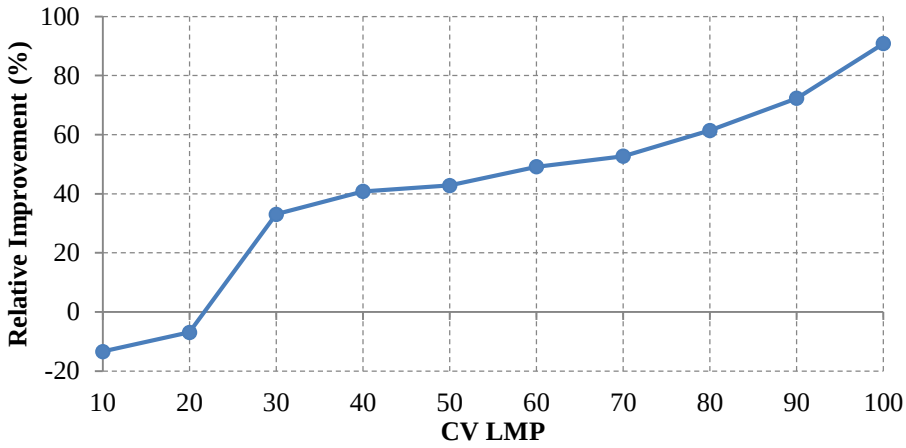


Figure 2-17: Relative improvement in the estimation accuracy using DS/DDH/5 vs. DS/DDH/2.

2.3.4 Dynamic vs Static Queue Search Distance

The two proposed methodologies for detecting the end of queue are examined in this section. To compare the two approaches, DS/DDH/5 and SS/DDH/5 combination of methodologies were chosen for the analyses as presented in Figure 2-18.

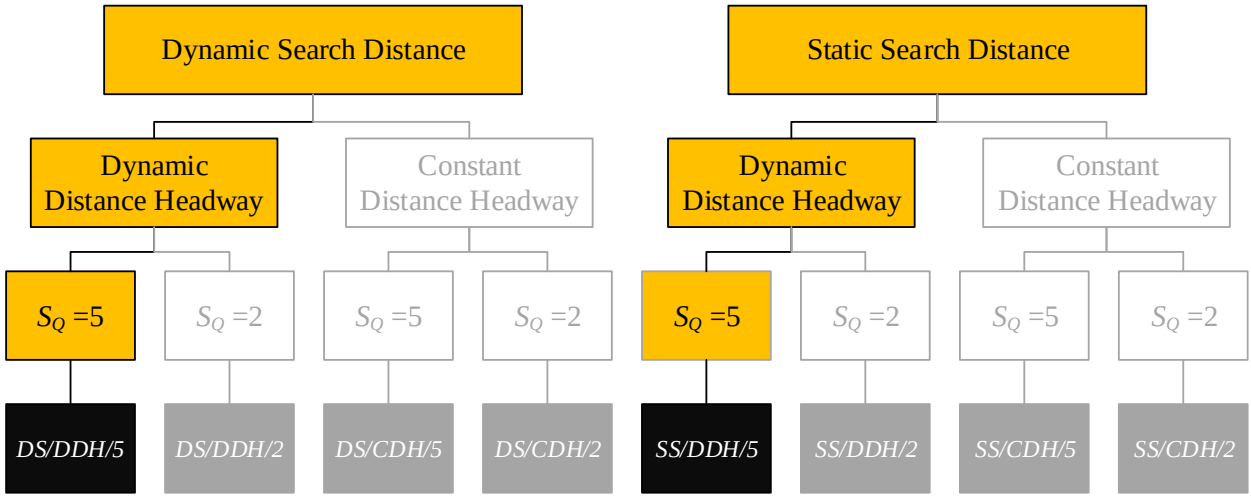


Figure 2-18: Compared methodologies in section 2.3.4.

As shown in Figure 2-19, the evaluation was performed in VISSIM microsimulation software for a hypothetical network. The hypothetical network was constructed to specifically create situations in which midblock queues could form as a result of a bus stop and vehicle stopping to make a left turn. The study area includes an off-ramp A, and a mid-block street B that connects to approach 4. Vehicles approaching the intersection can make left turns to street B located 200 metres upstream of the intersection using gaps in the opposing traffic approaching from the off-ramp A. The traffic flow on off-ramp A was set to 2000 vph, and 700 vph were simulated for approaches 2, 4, and 6. The lengths of approaches 2, 4, and 6 were 950 metres.

The traffic composition was assumed to be constant throughout the simulation with 10% of trucks (VISSIM default HGV vehicle type). The proportion of the left turners from approach 4 to street B was set to 10%. A bus stop was simulated on approach 2 located 300 metres upstream of the stop line with a bus frequency of every two minutes and dwell time distribution of N(40,5) seconds. The intersection was controlled using an actuated traffic signal controller with two phases (Figure 2-19). For each phase, the minimum and maximum green times were chosen to be 4 and 35 seconds, and the intergreen (amber + all red) was 5 seconds. Figure 2-20 presents sample

trajectories of vehicles on each approach at CV LMP=20% and shows the conditions of the simulated queues in the network (distance is measured from the signalized intersection stop line). In this figure CVs are shown in blue, non-CVs in orange and buses in grey colours.

Figure 2-20 (a) shows the impact of a transit vehicle (bus) stopping to board and discharge passengers at the mid-block bus stop on approach #2. When the transit vehicle stops, upstream vehicles on the single lane approach must also stop and consequently, a queue forms. Vehicles downstream of the transit vehicle are able to continue to proceed along the approach but may form a queue at the downstream intersection when the signal turns red. In this case, there are two distinct queues on the approach link.

Figure 2-20 (b) illustrates the trajectories of vehicle on approach #4. In this case, some vehicles on the approach attempt to make a left-turn onto link B. When these vehicles are unable to find a gap in the opposing traffic stream, then they are required to stop and wait, and cause upstream vehicles to also queue.

Figure 2-20 (c) shows the traffic condition on approach #6. On this approach there are no mid-block traffic interruptions and therefore the only queues that form do so at the stop line as a result of the traffic signal operations.

The SS/DDH/5 method was run with three different static search distances (40m, 150m, and 250m) to investigate the impact of the search distance on the accuracy of the method. The search distance for DS/DDH/5 was dynamically determined based on the estimated LMP and Equation (2-5). The minimum and maximum search distances were set to 40m and 250m respectively resulting in the relationship shown in Figure 2-21. Each of the four methods was run using 10 different seeds. Each seed was run for 3000 seconds after 1200 seconds warm-up period used by the dynamic model to have an initial estimation of CV LMP in the network (total 4200 seconds).

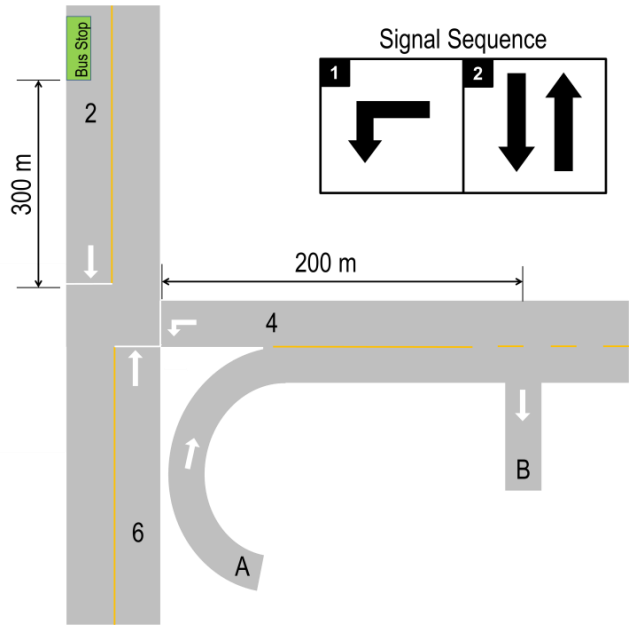


Figure 2-19: Hypothetical study area.

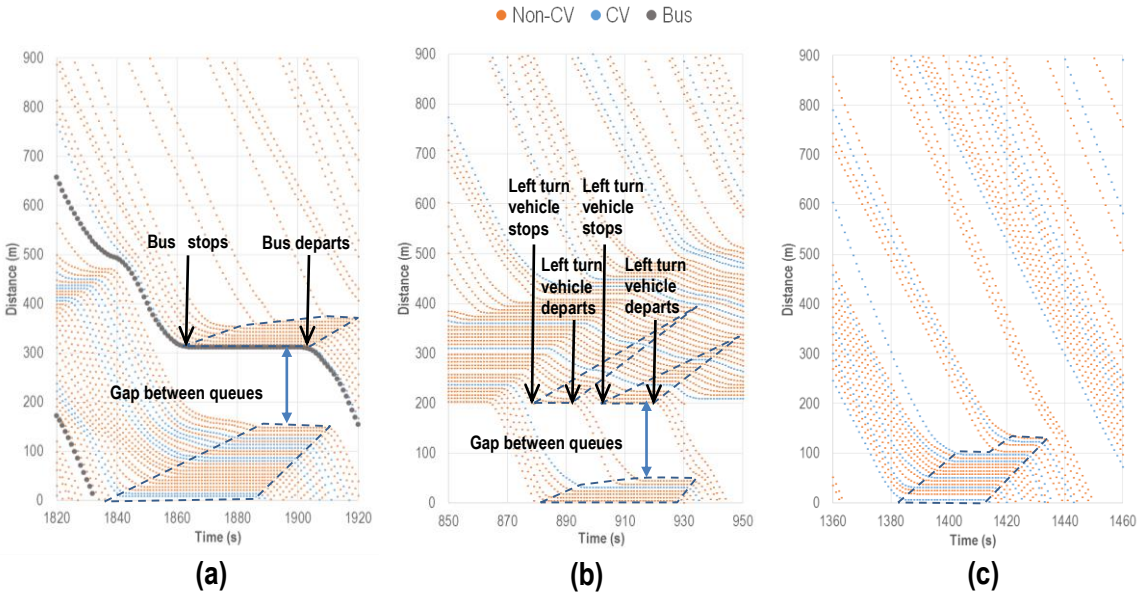


Figure 2-20: Sample vehicle trajectories on: (a) approach 2, (b) approach 4, and (c) approach 6.

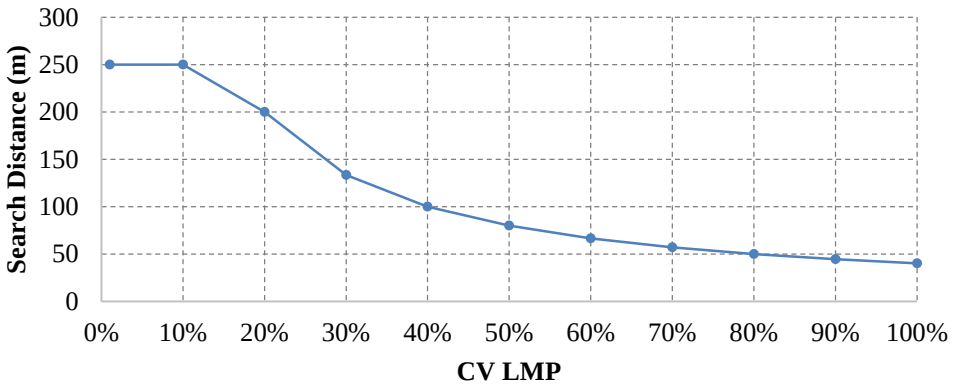


Figure 2-21: Search distance relationship for evaluation of dynamic versus static search distance method.

The results for each approach are presented in Figure 2-22. The figures present the RMSE of location of the end of queue calculated using Equation (2-13) for each of the examined methodologies as an average of number of runs performed for each LMP.

$$RMSE_{Avg}^{LMP} = \frac{1}{n}\sum_{C=1}^{n} RMSE_{C}^{LMP} \qquad (2\text{-}13)$$

Where,

$RMSE_{Avg}^{LMP}$ = Average RMSE of the estimated queues (number of vehicles) when CV penetration rate is LMP over all cycle *C* and simulation seeds, (*veh)*;

$RMSE_{C}^{LMP}$ = RMSE of the estimated queue (number of vehicles) calculated at cycle *C* when CV penetration rate is LMP (Equation (2-11) with m=1), (*veh)*;

n = Total number of cycles.

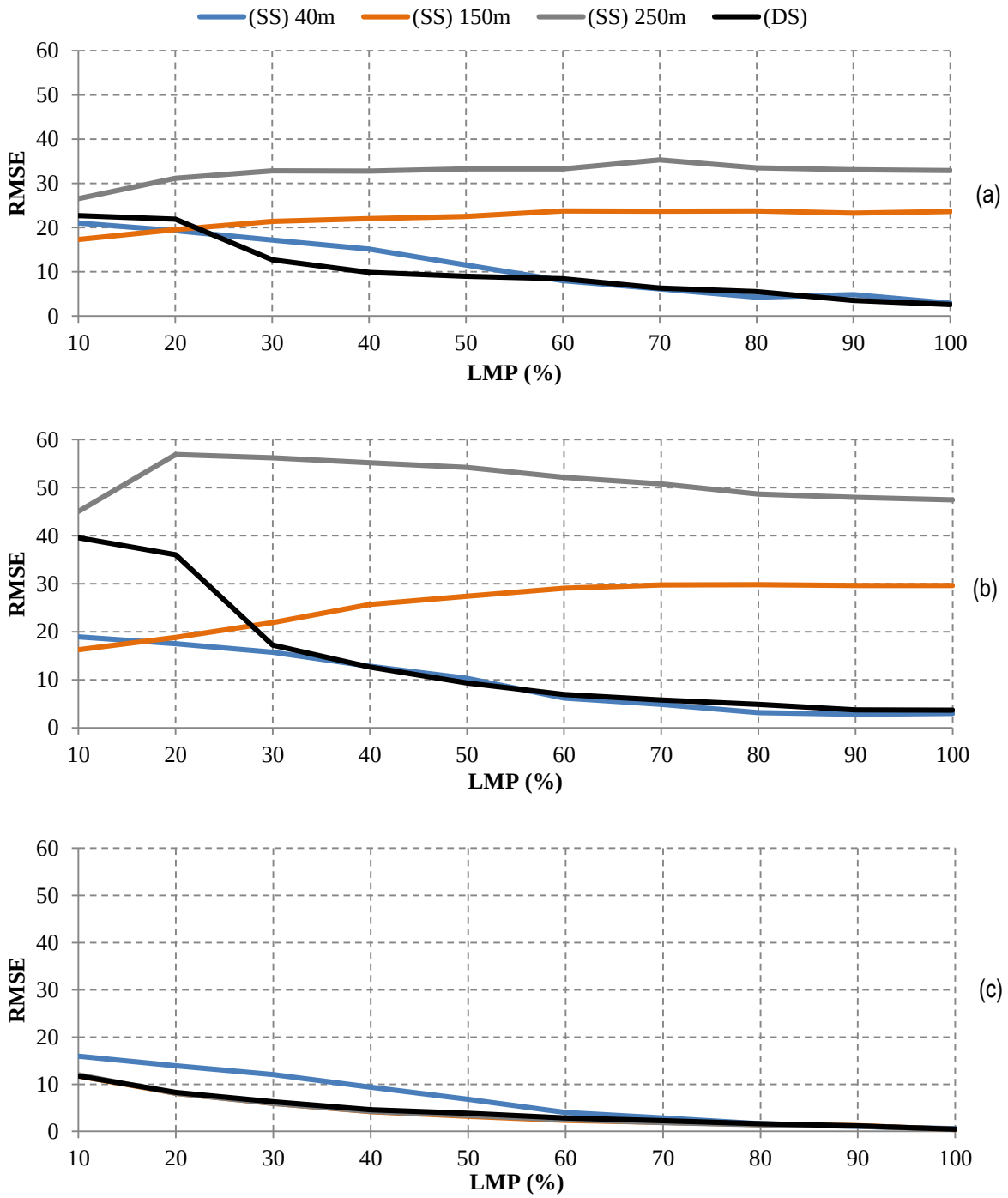


Figure 2-22: End of queue estimation errors using dynamic vs static search methodologies. (a) Approach 2, (b) Approach (4), (c) Approach (6)

Figure 2-22 (a) presents the results for approach 2. The results show that, when the search distance is short, the SS/DDH/5 methodology performs better in comparison to using longer search distances. This was expected as using a shorter search distance reduces the chance of detecting

queued vehicles behind a bus at the bus stop as part of the queue at the intersection. At 40 m search distance the accuracy of the SS/DDH/5 method is comparable to the DS/DDH/5 methodology which provided the best queue estimation accuracy.

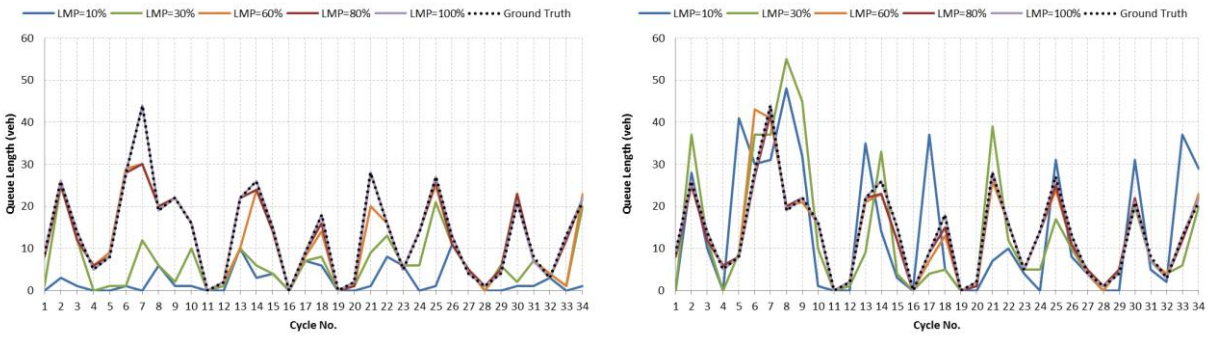
Figure 2-23 shows a sample run of estimated queues using static with search distance = 40 metres and dynamic methods over all cycles and CV LMPs. These graphs show that as the LMP reduces in the network, the static method with search distance = 40 metres underestimates the queue length. This underestimation resulted in increase in the errors especially at low LMPs. On the other hand, the dynamic method can adapt to the change in the LMP by increasing the search distance, and as presented in Figure 2-22 (a), it has a better accuracy than the static method when the LMP is between 30 and 60 percent. However, as the search distance is further increased at lower LMPs (less than 30%) queued vehicles behind the bus stop are detected as part of the stop line queue and it starts to overestimate the queue. Therefore, a sudden increase in the error can be observed when LMP<30%.

Figure 2-22 (b) presents the results for approach 4. Similar to the results obtained on approach 2, the accuracy of the SS/DDH/5 methodology was highest when the search distance was 40m. This improvement in the accuracy occurs because of a shorter search distance that prevents the method from erroneously including vehicles stopped in a mid-block queue caused by a vehicle attempting to turn left to link B. The results also show that the SS/DDH/5 with 40m search distance performs better than DS/DDH/5 methodology when CV LMP is lower than approximately 30%.

Figure 2-24 shows a sample of estimated queues using both dynamic and static (search distance = 40 m) methods when CV LMP = 20%. Similar to the results presented for approach 2, the static method underestimated the queue length at low LMPs while the dynamic method overestimates the queue as the LMP decreases below 30%. When LMP is low, the DS/DDH/5 method increases the search distance up to the maximum defined by the user (250 m) which increases the queue estimation errors as mid-block turning vehicles are more likely to be detected by the method and considered to be part of the stop line queue. As the CV LMP increases, the method reduces the search distance, resulting in smaller estimation errors. However, as shown in Figure 2-22 (b), at low LMPs, queue overestimations using the dynamic method on this approach result in larger errors (due to long queues formed behind the mid-block left turning vehicles) in comparison to queue under-estimation errors caused by using the static method.

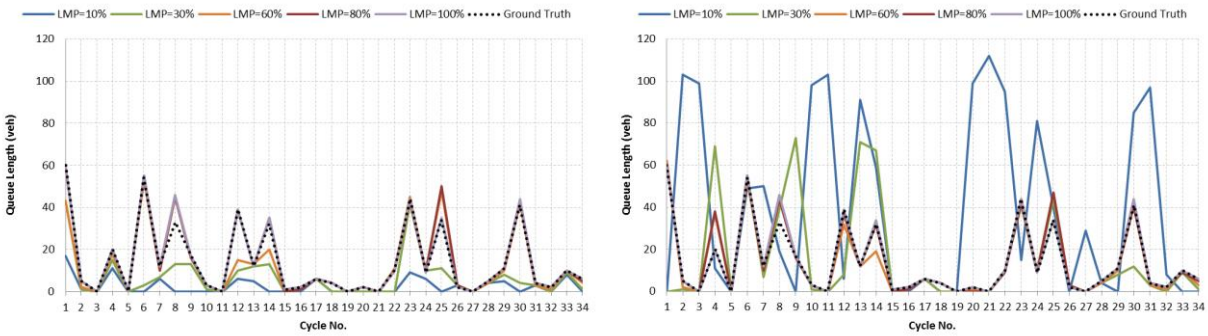
Figure 2-22 (c) presents the results for approach 6. The results for approach 6 show a different trend in comparison to those presented for approaches 2 and 4. Mid-block queues do not form on approach 6. The only queues that form are those caused by the traffic signal operation and consequently we do not expect any advantage to using a short search distance. As the results show for the static method, using a larger search distance produces better queue estimation results when CV LMP is low and the best accuracy is obtained when 250m search distance is used. This is because using a large search distance more CVs in the queue can be detected during low CV LMPs as the distances between CVs in the queue are larger. Therefore, as expected using 40m search distance cannot capture all CVs in the queue and results in higher estimation errors. On this approach the dynamic method performed well and could adapt its search distance based on the estimated LMP.

Figure 2-25 compares sample queue estimation results performed by both static (search distance =40m) and dynamic methods on approach 6. The results show that at low LMPs the static method underestimates the queue length and the dynamic method performs much better.



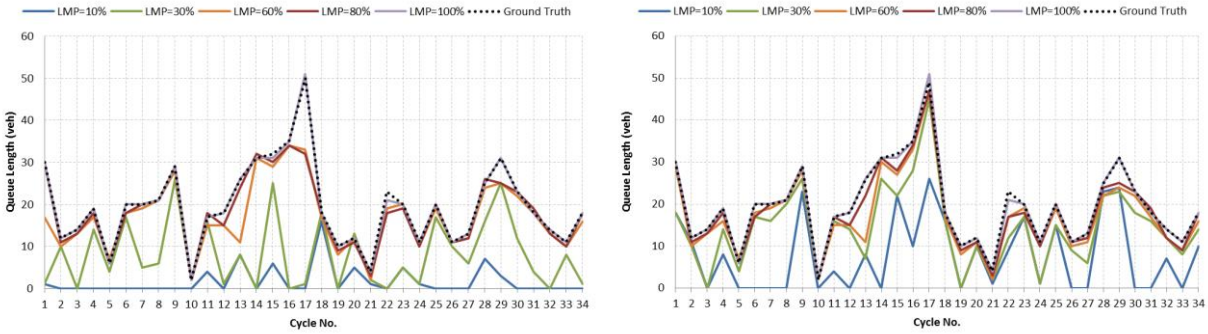
(a) **(b)**

Figure 2-23: Sample estimated queues versus the ground truth at different LMPs on approach 2 using (a) Static method with search distance =40 m, and (b) dynamic method



(a) **(b)**

Figure 2-24: Sample estimated queues versus the ground truth at different LMPs on approach 4 using (a) Static method with search distance =40 m, and (b) dynamic method



(a) **(b)**

Figure 2-25: Sample estimated queues versus the ground truth at different LMPs on approach 6 using (a) Static method with search distance =40 m, and (b) dynamic method

We have quantified the accuracy of the proposed method for estimating the location of the most upstream CV in the queue. However, we also wish to compare this performance to other exiting methods. Unfortunately, at the time of completing this research, the author was not able to find any other method that estimates the most upstream CV in the queue. The most similar approach is a method proposed by Tiaprasert et al. (2015) who proposed a methodology to estimate the actual queue length using CV data. The methodology requires knowing estimates of possible maximum and minimum queue length on an approach. The minimum queue length is defined by the location of the stopped connected vehicle that is farthest upstream of the stop line, and the maximum possible queue length is determined by the location of the moving connected vehicle that is closest to the stop line but upstream of the farthest upstream stopped connected vehicle. By knowing the possible minimum and maximum queue lengths, the methodology constructs a joint probability function to find the probability of having a queue with N vehicles given k detected connected vehicles in the queue. The methodology also uses a discrete wavelet transform method to smooth estimates over time and thus avoid large and unrealistic changes in the estimated queue length. The model has been tested in a microsimulation model for different traffic conditions (under-saturated vs over saturated conditions) and fixed versus actuated traffic signals for an isolated intersection with no mid-block traffic. We were not able to replicate their method and apply it to our network and/or simulation runs. Instead, we used the performance results they

report from their study for under-saturated conditions and actuated traffic signal and we compare these to the results of the proposed methodology obtained on approach 6 where the traffic state is under-saturated and there is no mid-block traffic.

Tiaprasert et al. (2015) quantified performance accuracy as the average RMSE divided by the maximum of the 2-second ground truth queue length occurring over a simulation run. RMSE was calculated as the root mean squared errors between the ground truth queue and the estimated queues every 2 seconds, and summed across the simulation time. This was done for each simulation run and then the average RMSE was computed as the average of the root mean squared error computed for each of the simulation runs.

Note this is different from the evaluation presented in this section as we calculated the RMSE between the maximum ground truth and the maximum estimated queues within each signal cycle. The RMSE calculated by Tiaprasert et al. (2015) is done every 2 seconds, and therefore includes many intervals for which the estimated and actual queue is zero (i.e. for those time interval when the queue has been served). Computing the average RMSE across all time intervals, including those for which no queue exists, results in a lower average RMSE than if the average RMSE was computed only over intervals for which a queue existed.

Therefore, in order to be able compare the results presented in this chapter to those reported by Tiaprasert et al. (2015), an analysis was performed to investigate the relationship between the average RMSE presented in this chapter ($RMSE_1$) and the one calculated by Tiaprasert et al. (2015) ($RMSE_2$).

Assuming a single lane approach with a fixed time traffic signal and deterministic vehicle arrival, D/D/1 queuing theory was used to model the queue for the approach. The signal cycle was set to 85 seconds as the average of all cycles observed in the VISSIM simulation analysis presented earlier; and the green time for the approach was assumed 35 seconds which is the maximum green time for approach 6 in the simulation. The departure rate was calculated based on the average measured saturation flow rate of 2090 vphpl in VISSIM simulation for approach 6.

The modeled queue was assumed as the ground truth queue and a set of "estimated" queue lengths were computed by multiplying the ground truth queue lengths by a factor, F. The value of F was varied from a minimum of 0.2 to a maximum of 2.0 to reflect queue estimation methodologies having different accuracies. When $F = 0.2$, the estimated queue is always much smaller than the

true queue and when F = 2.0, the estimated queue is twice as large as the true queue. The arrival flow was also varied to model a range of degree of saturation (x) between 0.1 to 1 for the approach. The ground truth queue length was calculated every 2 seconds and the maximum extent of the queue at each cycle was obtained. The objective is to evaluate the impact of the different ways of computing overall queue estimation accuracy (i.e. $RMSE_1$ vs $RMSE_2$) as a function of F and x. Therefore, the ratio $RMSE_2/RMSE_1$ was computed for each combination of F and x using Equation (2-14).

$$R_{F,x} = \frac{RMSE_2^{F,x}}{RMSE_1^{F,x}} \tag{2-14}$$

Where,

$R_{F,x}$ = Ratio of RMSEs calculated using the two methods at degree of saturation x and for a value of F;

$RMSE_1^{F,x}$ = Calculated RMSE using the method presented in this chapter (Equation (2-13)) at degree of saturation x and for a value of F;

$RMSE_2^{F,x}$ = Calculated RMSE using the method presented by Tiaprasert et al. (2015) at degree of saturation x and for a value of F;

Figure 2-26 presents $R_{F,x}$ for different combinations of F and x. The results show that there is a constant relationship between the RMSE calculations made by the two methods for different values of F. However, $R_{F,x}$ decreases as the degree of saturation on the approach increases. Figure 2-27 presents $RMSE_2/RMSE_1$ as a function of x, which depicts a linear relationship between these two parameters with $R^2 \cong 1$. Therefore, to convert $RMSE_1$ to $RMSE_2$ for the evaluation condition used in the VISSIM simulation, we must find the ratio of $RMSE_2/RMSE_1$ that corresponds to the degree of saturation on approach 6.

Knowing the average cycle length = 85 seconds, green interval = 35 seconds, arrival flow = 700 vehicle per hour, and average saturation flow rate = 2090 vehicle per hour per lane, we can compute x = 0.81. Therefore, using Figure 2-27, we can conclude that the RMSE calculated using the method presented in this chapter can be factored down by 1.85 to be comparable to the RMSE results presented by Tiaprasert et al. (2015).

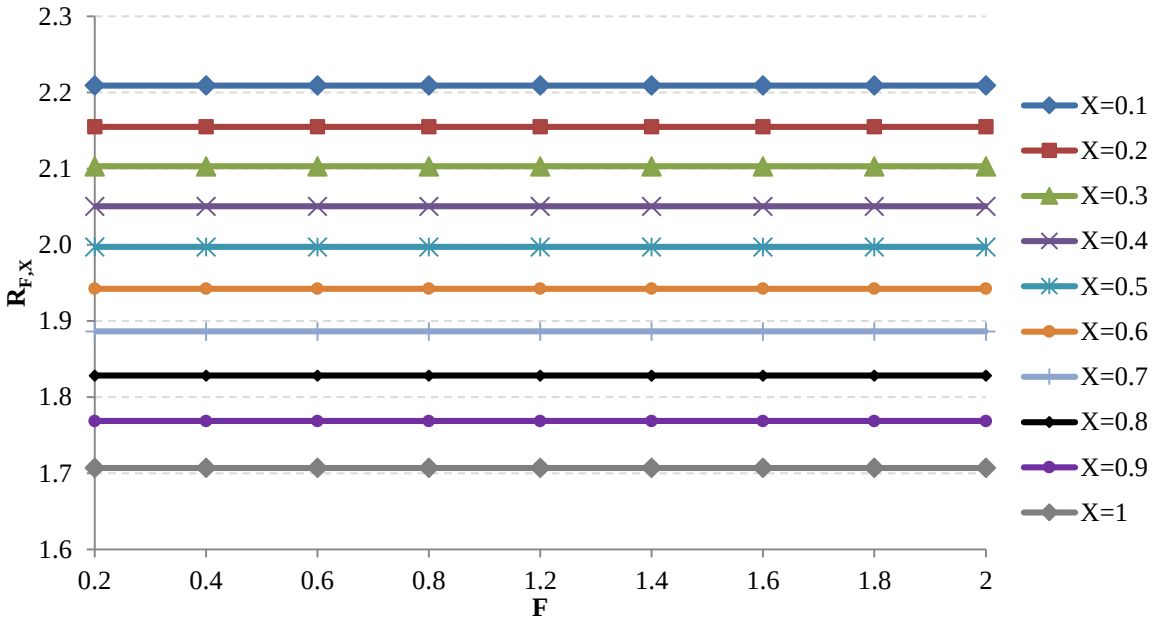


Figure 2-26: $R_{F,X}$ for a range of values of F and X.

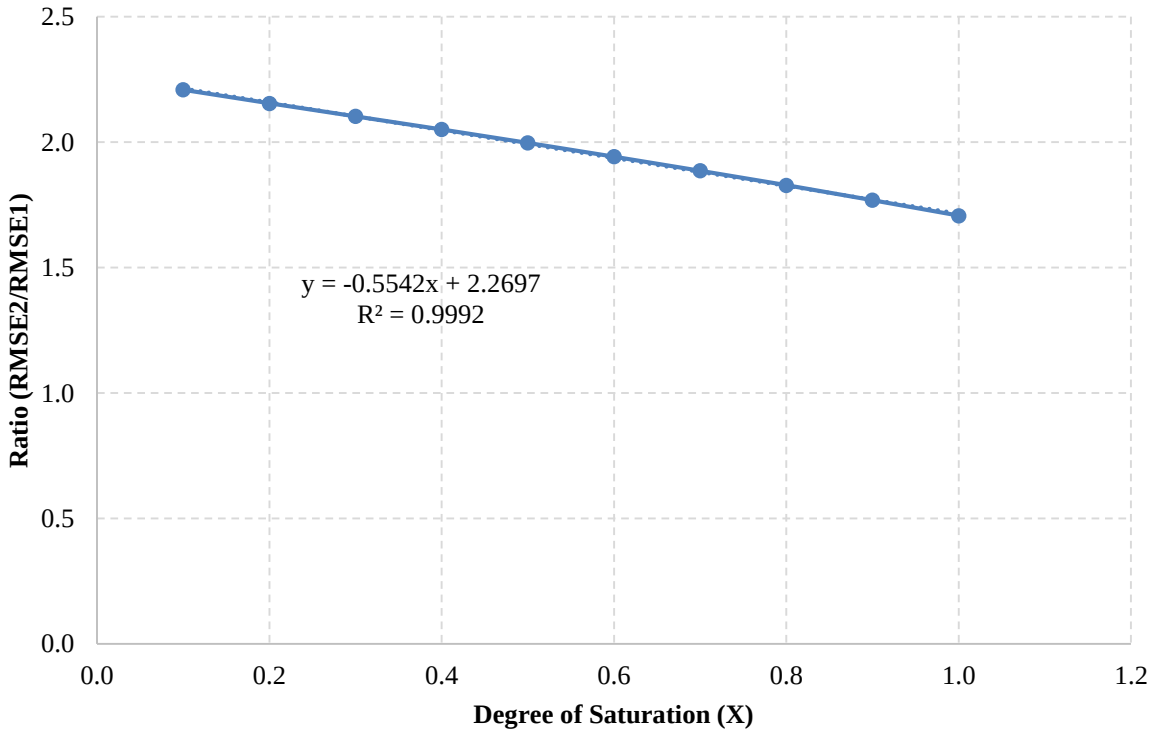


Figure 2-27: $RMSE_2/RMSE_1$ as the function of degree of saturation (X).

Tiaprasert et al. (2015) reported that at LMP=10% their queue estimation methodology resulted in average RMSE = 4.8 vehicles and the maximum ground truth queue length = 17.13 vehicles

indicating that the relative error is approximately 28%. For the model proposed in this thesis, the average RMSE at LMP = 10% was 11.74 vehicles. Using the conversion factor presented above, this results in RMSE = 6.35 if calculated using the method presented by Tiaprasert et al. (2015). Thus, knowing the converted RMSE and the average maximum ground truth queue length of 29.45 vehicles (on approach 6), the relative estimation error for the proposed method at LMP = 10% is approximately 22%. Thus the method proposed in this chapter appears to perform as well as or better than the method proposed by Tiaprasert et al (2015).

2.4 Summary and Conclusions

This chapter presented and evaluated methodologies for estimating queue length at signalized intersections for the purpose of estimating the saturation flow rate. The goal was to detect as much of the stationary queue as possible. Given that data are assumed to be available only from CVs, the goal is to identify all CVs that join the stationary queue caused by the traffic signal operations and then to estimate the number of non-CVs between the known CVs.

The methodology presented in this chapter is distinct from similar methodologies based on CV data such as the one proposed by Goodall et al. (2013) in terms of the following enhancements:

1- The methodology checks for discontinuous queues by using a search distance which can be dynamically adjusted based on the estimated LMP in the field or selected by a user.
2- The methodology can adjust the distance headway value used for inferring non-CVs between detected CVs in a queue by incorporating vehicle length information obtained from CVs in the field. Therefore, it can adapt itself to changes in the traffic composition.
3- The methodology uses two speed thresholds for detecting CVs as part of a queue and in order to do that it uses two values of vehicle spacing for the initial queue detection at the end of the red interval when vehicles are stationary and further queue detection during the green intervals when additional CVs join a queue.

Different combinations of the proposed methodologies were evaluated using the VISSIM microsimulation software. From these evaluations, we can make the following conclusions:

1. The queue estimation method requires an estimate of the average vehicle length. An evaluation of using dynamically estimated average vehicle length versus assuming a static average vehicle length demonstrated that the vehicle length data element in BSM I can be

used to estimate the average length of the traffic stream where the traffic composition varies over time. The results show that the dynamic estimation method (DS/**DDH**/5) provides a better accuracy in comparison to using the method that uses a constant value for vehicle spacing (DS/**CDH**/5). The improvement in the accuracy was observed during low CV LMPs with the maximum improvement of 25% at LMP=10%. Therefore, the dynamic method (DDH) is preferred over the static method (CDH) especially for locations where the traffic stream composition is expected to change over time of day or as a result of diversion. The DDH method has the additional advantage that there is no need to calibrate an average vehicle length as these data are obtained from connected vehicles. In contrast, the static method (CDH) requires that the system operator specify an appropriate average vehicle length for each location.

2. Two methodologies for queue detection (DS/DDH/2 versus DS/DDH/5) were evaluated and the results showed that DS/DDH/**5** methodology provided superior performance when CV LMP > 20%. Using DS/DDH/5, the chance of detecting CVs in a queue is increased resulted in an increase in the queue estimation accuracy between 30 to 90 percent for a range of LMPs between 30 and 100 percent.
3. An evaluation of using a dynamic search distance versus a constant (static) search distance was also conducted. The results show that the dynamic method (**DS**/DDH/5) is better for locations where mid-block queues are unlikely to occur and the static method (**SS**/DDH/5) with a short distance is better suited for locations where mid-block queues are more likely to occur.
4. Finally, the evaluation showed that the proposed methodology for estimating the market penetration rate of connected vehicles in the field can perform well especially when the LMP is above 20% and when the dynamic vehicle spacing (DDH) model is used.

The next chapter describes how the methodologies presented in this chapter are used for estimating the saturation flow rate and the impacts of using some of these methods on the accuracy of the estimations.

3

Estimation of Saturation Flow Rate using Data from Connected Vehicles[1]

This chapter describes the proposed methodology for estimating the saturation flow rate using CV data. The chapter is divided into three main sections. An overview of methods to estimate the saturation flow rate is presented in the first section. The second section presents the proposed methodology for estimating the saturation flow rate using CV data; and finally, evaluation of the algorithm for a range of different traffic conditions and scenarios are presented in section 3.3.

3.1 Background

At signalized intersections, traffic is interrupted periodically as the right of way is sequentially assigned to conflicting movements. Figure 3-1 shows sample trajectories of vehicles on a time-space diagram for a movement at an intersection controlled by a traffic signal. During the red interval a queue of vehicles forms behind the stop line. When the signal turns green vehicles start to discharge from the stop line. The time difference between consecutive vehicles crossing the stop line (or a reference point) is defined as the time headway, and is calculated using Equation (3-1). The passing time of vehicles can be recorded relative to a reference point on vehicles such as front/rear bumper or front/rear axle.

$$h_{i,i-1} = t_i - t_{i-1} \tag{3-1}$$

[1] The content of this chapter has been published in the following journal and conference papers:

Bagheri, E., Mehran, B., & Hellinga, B. (2015). Real-Time Estimation of Saturation Flow Rates for Dynamic Traffic Signal Control Using Connected-Vehicle Data. Transportation Research Record: Journal of the Transportation Research Board, (2487), 69-77.

Bagheri, E., & Hellinga, B. (2016). Real-Time Saturation Flow Rate Estimation Using Connected Vehicle Data. In Transportation Research Board 95th Annual Meeting (No. 16-1480).

Where,

$h_{i,i-1}$ = Headway of vehicle *i*, (*s*);

t_i = The time at which vehicle *i* passes the stop line, (*s*);

t_{i-1} = The time at which vehicle *i-1* passes the stop line, (*s*);

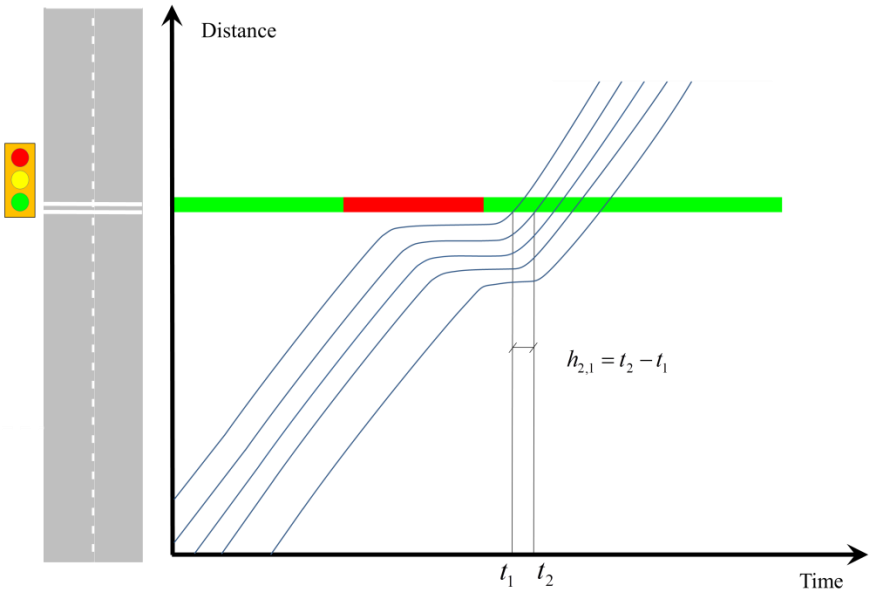


Figure 3-1: Discharge of queue at signalized intersections

During the green interval, vehicles in a queue tend to cross the stop line with the maximum possible rate, which is referred to as the saturation flow rate (*S*). Webster and Cobbe (1966) and the Highway Capacity Manual (HCM) (2010) define the saturation flow rate as the flow on a lane that would be observed discharging from the stop line assuming a queue always existed and the green signal was displayed all the time.

Equation (3-2) presents the cycle-by-cycle field measurement of saturation flow rate (S_C) for a lane based on the saturation headway ($h_{S,C}$), which is calculated by dividing the number of seconds in an hour by the average measured headways of queued vehicles crossing the stop line at cycle *c*.

$$S_C = \frac{3600}{h_{S,C}} \qquad (3\text{-}2)$$

However, as mentioned in Chapter 1, when the signal turns green, it takes some time for the drivers of the first few vehicles in the queue to observe and react to the green light. As shown in Figure

3-2 these vehicles experience larger headways in comparison to the following vehicles. Hence, the saturation flow rate is usually "achieved after about 10 to 14 seconds of green, which corresponds to the front axle of the 4th to 6th passenger car crossing the stop line after the beginning of green" (HCM, 2010). This is referred to as start-up loss. Therefore, the saturation headway is calculated using Equation (3-3) where $t_{C,n}$ and $t_{C,4}$ represent the time when the front axles of the nth and 4th vehicles (the last and 4th vehicles that were part of the queue) cross the stop line in cycle *C*. The total start-up lost time for a cycle (l_c) is also calculated using Equation (3-4) where; l_i is the lost time for the i^{th} vehicle in the queue. In this equation *n* is usually selected to be equal to 4 to correspond to the 4th vehicle in the queue. In the absence of direct measurement, l_c is typically assumed to be 2 seconds (HCM, 2010)

$$h_{S,C} = \frac{t_{C,n} - t_{C,4}}{n_C - 4} \tag{3-3}$$

$$l_C = \sum_{i=1}^{n} l_i \tag{3-4}$$

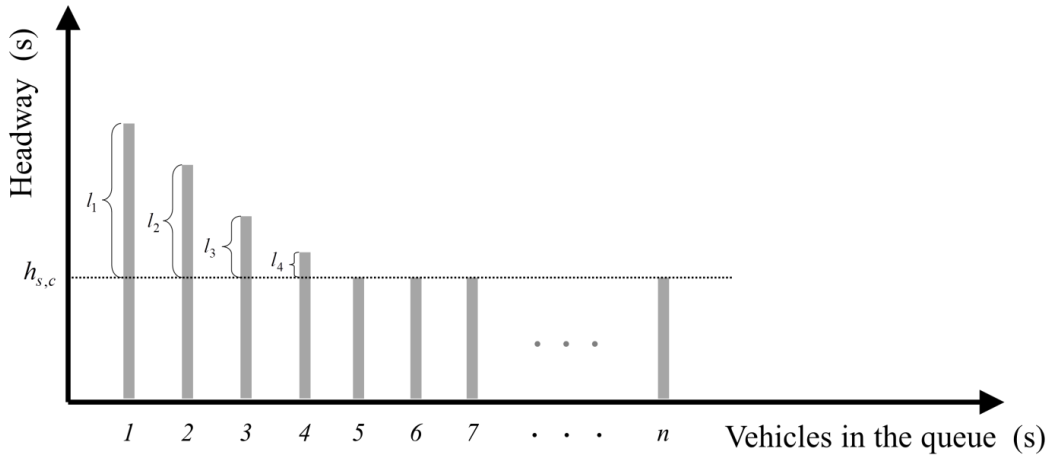


Figure 3-2: Illustration of discharge headways at a signalized intersection

The HCM (2010) provides an analytical procedure for estimating the saturation flow rate. As presented in Equation (3-5), the method estimates the saturation flow rate as a function of a base saturation rate and a range of multiplicative adjustment factors that account for the influence of site specific geometry, traffic stream composition, parking, transit vehicles, etc. The base saturation flow rate in Equation (3-5) is the saturation flow rate for a lane with a width of 3.66

metres (12 ft) with no bus stops, no turning vehicles, no vertical grade, no parking, no pedestrians, and no heavy vehicles.

$$S_C = S_0 \, f_w \, f_{HV} \, f_g \, f_p \, f_{bb} \, f_a \, f_{LU} \, f_{LT} \, f_{RT} \, f_{Lpb} \, f_{Rpb} \quad (3\text{-}5)$$

S_C = Adjusted saturation flow rate, (veh/h/ln);
S_0 = Base saturation flow rate, (pc/h/ln);
f_w = Adjustment factor for lane width;
f_{HV} = Adjustment factor for number of heavy vehicles in the traffic stream;
f_g = Adjustment factor for approach grade;
f_p = Adjustment factor for parking activity;
f_{bb} = Adjustment factor for the influence of bus stop in vicinity of an intersection;
f_a = Adjustment factor for area type (i.e. downtown area)
f_{LU} = Adjustment factor for lane utilization;
f_{LT} = Adjustment factor for presence of left turning vehicles in the lane group;
f_{RT} = Adjustment factor for presence of right turning vehicles in the lane group;
f_{Lpb} = Adjustment factor for the influence of pedestrian on left turning vehicles in the lane group
f_{Rpb} = Adjustment factor for the influence of pedestrian and bicycle on right turning vehicles in the lane group

The above method is well-known and is used in many offline traffic signal timing optimization methods. However, it was not designed for and is not well suited for real-time applications. In particular, it does not have the ability to capture changes in the saturation flow rate, for instance, as a result of a lane closure caused by traffic incidents (accidents, construction, etc.), or changes in the road surface condition caused by inclement weather, etc. Furthermore, the adjustment factors are typically computed on the basis of information that may not be readily available in real-time applications (e.g. number of pedestrians, parking utilization, etc.).

The HCM also presents a methodology for directly measuring the saturation flow rate in the field. The HCM recognizes that direct observation will produce more accurate results than the analytical estimation procedure. In this method, observers in the field identify the last (most upstream) vehicle in a queue on a lane at the beginning of the green interval. Then, starting at the beginning of the green interval, vehicles are counted and their passing times from a reference point (i.e. stop

line) are recorded until the last vehicle in the queue passes. Then, the saturation flow rate for the lane is calculated using Equations (3-2) and (3-3).

The direct measurement methodology is expected to provide more accurate estimations of the saturation flow rate in comparison to the analytical method, however, implementation of the method in the field is challenging. An effort for field implementation of the method is presented by Henderson and Wood (2005) who proposed a method for estimating the saturation flow rate using data from Induction Loop Detectors (ILDs). As shown in Figure 3-3(a) the method counts the number of vehicles discharging from the stop line until the time headway between successive vehicles exceeds a specified threshold which is assumed to be the time that the queue has been served and the detectors are measuring the arrival flow rate[2]. This method has been evaluated for simple intersections and works well for normal conditions. However, for some conditions it is not valid to assume that a headway larger than a threshold means the end of a queue. For example consider Figure 3-3 (b), which shows a queue that forms at the traffic signal. When the signal turns green, the queue begins to discharge, but then the discharge flow drops to zero when a bus stops to board/alight passengers and blocks the traffic. The queue has not yet been discharged, but the recorded headway will exceed the critical gap and the system will ignore the remaining vehicles when determining the saturation flow rate. The average (or effective) saturation flow must include the effect of these temporary blockages.

It is anticipated that using the data provided by connected vehicles, will provide the ability to better identify those vehicles discharging at the saturation rate. Therefore, the next section presents a new methodology to estimate the saturation flow rate using data provided by connected vehicles. The proposed methodology is expected to take into account all the factors that affect the saturation flow rate.

[2] This is essentially the same method used in actuated and semi-actuated control systems in which time headways between successive vehicles are recorded and when the time headway exceeds a threshold value, the initial queue is assumed to have been served.

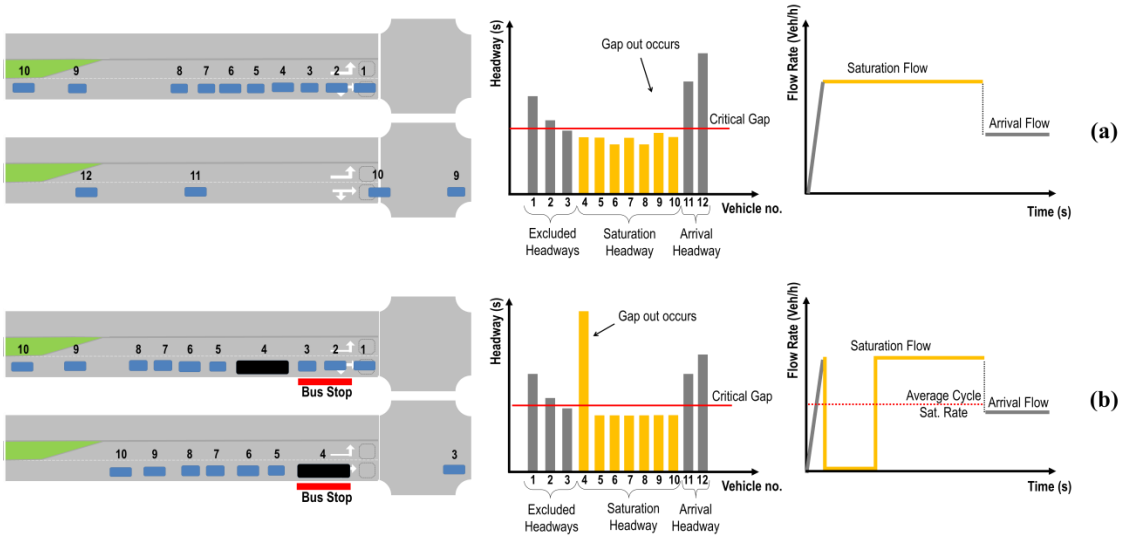


Figure 3-3: Saturation flow rate estimation based on a gap threshold

3.2 Methodology

The proposed methodology is intended to estimate the saturation flow rate for a lane group at a signalized intersection using CV data. Several assumptions are used when developing the proposed method. The main assumption is that connected vehicles in the network broadcast the BSM Part I with the frequency of 10 Hz or 10 times per second. The BSM contains information about each connected vehicle at a minimum including a temporary ID, the vehicle size, speed, acceleration, and position (longitude and latitude). Real-time positions of vehicles are mapped to an available survey of the site containing information such as locations of intersection stop lines, lane configurations, and intersection geometry. Therefore, the distance to the stop line and the lane that each connected vehicle is traveling in is available to the system.

It is also assumed that there is no significant delay in the transmission and receipt of the BSM data packets by the receiver. Real-time signal timings information is also provided to the system as conventional outputs of the traffic signal controllers.

The proposed method is based on the HCM method for direct measurement of the saturation flow rate in the field. Therefore, we must perform the following tasks in order to be able to estimate the saturation flow rate:

1. Identify the vehicles in the current lane that were part of the stopped queue.

2. Identify the vehicles which were part of the stopped queue and have crossed the stop line and made their manoeuver during the current green interval.
3. Record (or estimate) the times when the fourth vehicle and the last vehicle that were part of the stopped queue have crossed the stop line. The crossing time of the fourth vehicle is used as the start point of saturation flow rate estimation to eliminate the impact of start-up lost time, and the time that the last vehicle, which was part of the stopped queue, crosses the stop line is used as the upper bound to end the estimation of saturation flow rate for a given phase.

Chapter 2 presented methodologies to identify queued vehicles in a queue on a lane using CV data. Since we are not able to obtain the time that a non-CV crosses the stop line, we only estimate a queue up to the location of the most upstream CV in the queue. During the green and the following amber intervals, the estimates of the number of vehicles in queue, and the number of vehicles that cross the stop line on each lane are updated every 0.1 second. Once the signal phase has ended, the saturation flow rate for the lane group in that cycle is estimated. An exponentially smoothed running average saturation flow rate is computed for each lane group at the end of each cycle.

Satisfying the above three requirements and estimating the saturation flow rate is relatively straightforward when: the LMP of CV is 100%; the approach consists of only a single lane; the approach discharges in a single phase; and the approach is under-saturated. However, when these conditions do not hold, then the estimation becomes more challenging as discussed below.

When an intersection approach has multiple lanes and/or lane groups (which is typically the case), vehicles have the potential to change lanes. This is usually not an issue when CVs change lanes as we are aware of these changes. However, when non-CVs change lanes, particularly after having been estimated to be part of the queue, this may introduce errors to the accuracy of the estimated saturation flow rate.

We address different possible traffic conditions in the field by identifying six scenarios. These scenarios consider the combination of three factors, namely:

1. whether the lane group is under or over saturated;
2. the number of CVs that have been observed to discharge during the green interval in the existing signal phase; and

3. whether or not we are able to identify the 4th vehicle to discharge at the beginning of green and thereby capture the effect of start-up loss.

The scenarios are identified in Table 3-1. Details on each of scenarios are provided in the following sections. For each of these six scenarios, we propose a methodology for estimating the cycle saturation flow rate. We note that in Scenario 3, the approach is under-saturated but no CVs were observed to pass the stop-line and therefore no estimate of the saturation flow rate can be made for that cycle. In that case, the estimated saturation flow rate for the current cycle will be equal to the saturation flow rate estimated for the previous cycle.

Table 3-1: Scenarios considered by the proposed methodology.

Scenario	Saturation		Number of CVs which discharged in current phase		4th vehicle in queue is CV?	
	Under	Over	0	≥ 1	No	Yes
1	✓			✓	✓	
2	✓			✓		✓
3	✓		✓		✓	
4		✓	✓		✓	
5		✓		✓	✓	
6		✓		✓		✓

The proposed methodology uses the following criteria to determine the status of each of the three factors that define the six scenarios:

Degree of Saturation:

For each phase, the algorithm records the time when each CV, which was part of the queue, passes the stop line. If all the CVs in the queue pass the stop line before the end of the amber interval, then the phase is considered to be under-saturated and the algorithm uses one of the methods designed for under-saturated conditions (i.e. Scenarios 1, 2 or 3). Otherwise, the phase is considered oversaturated and an oversaturated method is used (i.e. Scenarios 4, 5, or 6).

Number of CVs that discharged during the current phase:

At the end of each phase, the algorithm determines the number of CVs that were part of the queue and passed the stop line (for each lane in a lane group). We distinguish between zero vehicles and one or more vehicles. For scenarios in which one or more CVs discharge, the methodology

identifies the last discharged CV which was part of the stopped queue and denotes this vehicle as *l*.

Fourth vehicle in queue:

Consistent with the HCM, the proposed methodology attempts to eliminate the impact of start-up lost time by ignoring from the calculations the first four discharged vehicles. When the LMP is substantially less than 100%, it is unlikely that the first four vehicles in the queue are CVs. Consequently, we must use a method to estimate if a given CV is in the 4th position. We assume that at most ten seconds are required for the first four vehicles to pass the stop line (HCM, 2010). Consequently, a CV is assumed to be in the 4th position (and is designated as $V_{4^{th}}^{CV}$) if the following two conditions are met:

1. The CV passes the stop line within 10 seconds from the start of green; and
2. The number of vehicles estimated to be in queue from the stop line to (and including) the CV is equal to 4 ($N_j^T = 4$).

The method can detect CVs that have changed their lanes during the green interval by removing them from the detected queue of the origin lane, and adding them to the number of CVs detected on the destination lane when they pass the stop line.

3.2.1 Scenario 1

In Scenario 1 (illustrated in Figure 3-4), one or more CVs were observed to be part of the queue, all of these vehicles discharged during the phase (i.e. lane group is under-saturated), and the 4th vehicle in the queue could not be identified (i.e. was not a CV). A start-up lost time (*L*) equal to two seconds is assumed and used to take into account the differences between the headways of the first four vehicles and the saturation headway. The saturation flow rate for the cycle is estimated using Equation (3-6). Note that the numerator is an estimate of the number of vehicles that have passed the stop line over the period from the start of the green interval until the time the rear bumper of the last CV from the queue crosses the stop line. If lane changes do not occur, then we expect this number to be simply N_l^T. However, vehicles may make lane changes and therefore enter or exit the lane of interest (as indicated by the red vehicle in Figure 3-4). The impact of these lane changes are captured by the inclusion of the N_l^{CV} and φ_l terms in Equation (3-6).

$$S_C = 3600 \times \frac{\left(N_l^T - N_l^{CV}\right) + \varphi_l}{t_l - t_G - L} \tag{3-6}$$

Where,

S_C = Estimated saturation flow rate for the cycle, (*vph*);

N_l^T = Total number of vehicles (CVs and estimated non-CVs) in the stopped queue in the current lane measured from the stop line up to and including the last CV, (*veh*);

N_l^{CV} = Total number of CVs in the stopped queue in the current lane measured from the stop line up to and including CV_l, (*veh*);

φ_l = Total number of CVs that have passed the stop line from the start of the green interval until time t_l, (*veh*);

t_l = Time at which vehicle CV_l passed the stop line, (*s*);

t_G = Start time of green interval, (*s*);

L = Start-up lost time = 2 s

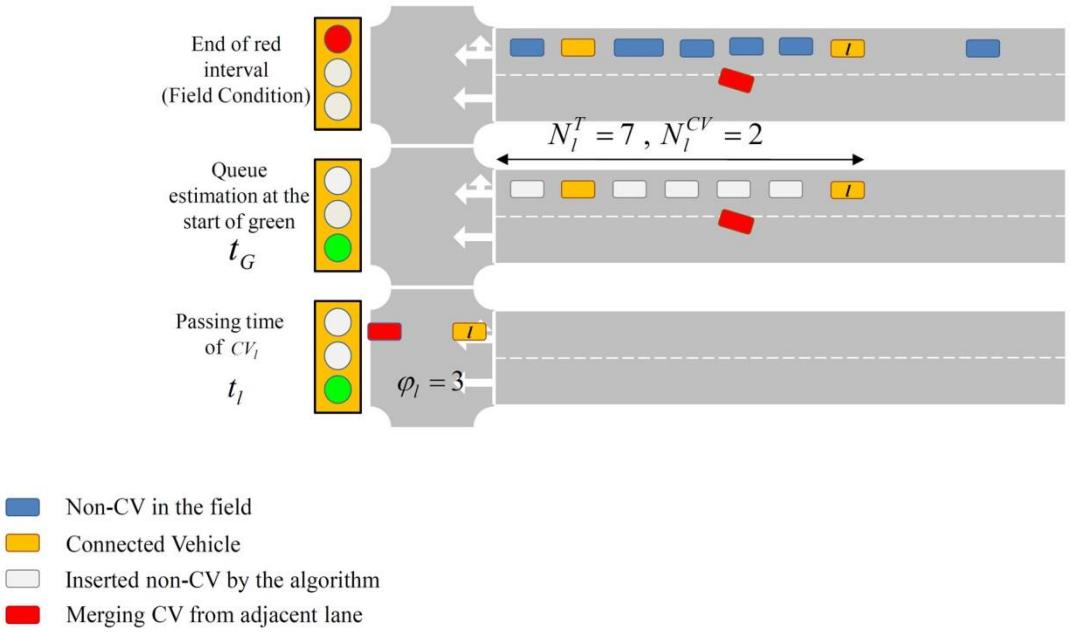


Figure 3-4: Illustration of scenario 1

3.2.2 Scenario 2

As the market penetration of CVs increases, it is more likely that the fourth vehicle in the queue is a connected vehicle ($V_{4^{th}}^{CV}$). In Scenario 2 (illustrated in Figure 3-5), one or more CV were observed to be part of the queue, all of these vehicles discharged during the phase (i.e. lane group is under-saturated), and the 4th vehicle in the queue could be identified (i.e. was a CV).

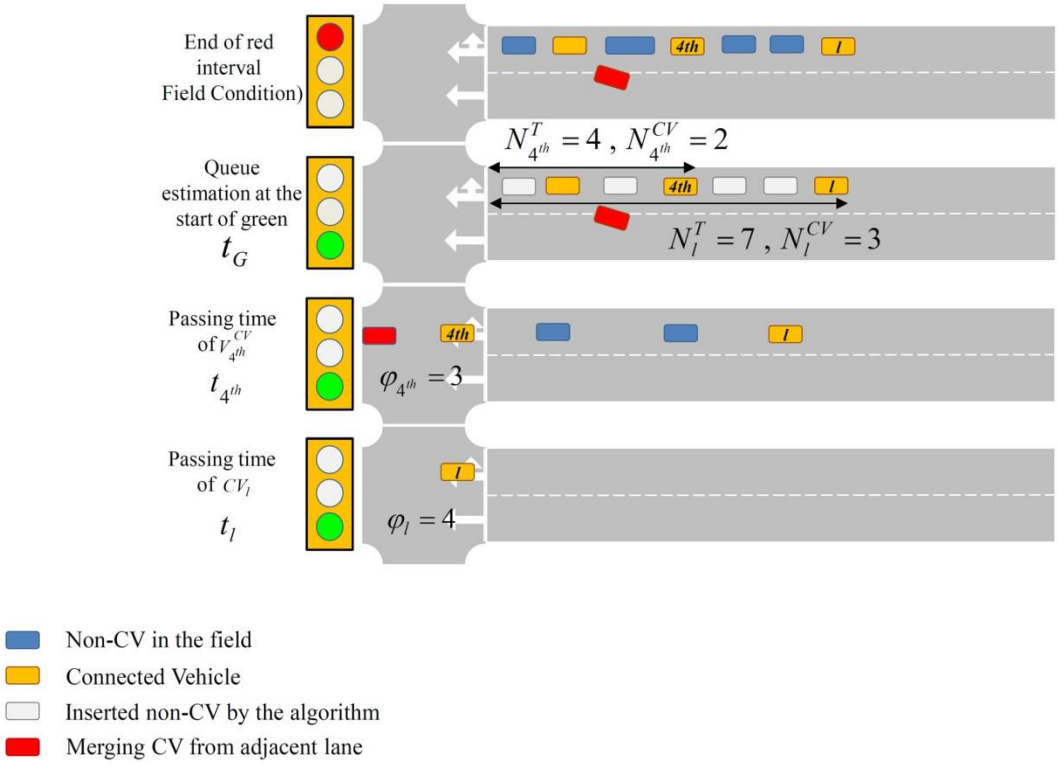


Figure 3-5: Illustration of scenario 2

Since $V_{4^{th}}^{CV}$ can be identified, Equation (3-7) is used to estimate the saturation flow rate and there is no need to assume a start-up loss time as in Scenario 1. The effect of lanes changes are captured using the same logic as in Scenario 1. However, in this case the algorithm needs to observe at least one more CV (CV_l) beyond the $V_{4^{th}}^{CV}$ to perform its estimations. In another words, if $V_{4^{th}}^{CV}$ is the last CV in the queue ($V_{4^{th}}^{CV}=CV_l$) then no estimation can be made for the current cycle. Note that only those variables which have not already been defined for previous scenarios are defined after each equation.

$$S_C = 3600 \times \frac{\left(N_l^T - N_{4^{th}}^T\right) - \left(N_l^{CV} - N_{4^{th}}^{CV}\right) + \left(\varphi_l - \varphi_{4^{th}}\right)}{t_l - t_{4^{th}}} \quad (3\text{-}7)$$

Where,

$N_{4^{th}}^T$ = Total number of vehicles (CVs and estimated non-CVs) in the stopped queue in the current lane measured from the stop line up to and including $V_{4^{th}}^{CV}$ (This value is always equal to 4), (*veh*);

$N_{4^{th}}^{CV}$ = Total number of CVs in the stopped queue in the current lane measured from the stop line up to and including $V_{4^{th}}^{CV}$,(*veh*);

$\varphi_{4^{th}}$ = Total number of CVs that have passed the stop line from the start of the green interval until time $t_{4^{th}}$, (*veh*);

$t_{4^{th}}$ = Time at which vehicle $V_{4^{th}}^{CV}$ passed the stop line, (*s*).

3.2.3 Scenario 3

Scenario 3 is characterized as under-saturated, but no CVs are observed to cross the stop-line. In this case, we have no information and are not able to estimate a saturation flow rate for the cycle.

3.2.4 Scenario 4

Scenario 4 is characterized as over-saturated, with one or more CVs in the stopped queue, but no CVs that were part of the queue passed the stop line during the phase. Figure 3-6 illustrates this case in which only a single CV was part of the queue. We label the first CV in the unserved portion of the queue as CV_r . We are able to estimate the number of vehicles between the stop line and CV_r at the beginning of the green interval. However, since no CVs in the queue actually cross the stop-line during the phase, we are not able to use Equation (3-6) as we do for Scenario 1. Instead, we must estimate the number of non-CVs that discharged during the phase on the basis of the position of CV_r at the beginning and end of the phase. When the queue begins to discharge at the start of green, a shockwave propagates upstream representing the boundary between the vehicles that have begun to accelerate and those that are still stationary. By definition, when the lane group is over-saturated, the queue is not dissipated before the end of the green. Consequently, it may take some time for vehicles in the portion of the queue that is not served, to move downstream to the stop-line and then form a new stationary queue. It is necessary to wait until the vehicles once again

form a stationary queue in order to determine the position of CV_r. Therefore, we identify CV_r and obtain its new position at the start of green for this phase in the next cycle. At this time the unserved queue is estimated using Equation (3-8) by knowing the new position of CV_r.

$$N_r'^T = \frac{P_r'}{\gamma} + 1 \qquad (3\text{-}8)$$

Where,

$N_r'^T$ = Number of vehicles measured from the stop-line to and including CV_r estimated in the next cycle, (*veh*);

P_r' = The position of CV_r relative to the stop-line, (*m*).

γ = Assumed average distance headway for vehicles in a stationary queue, (*m*).

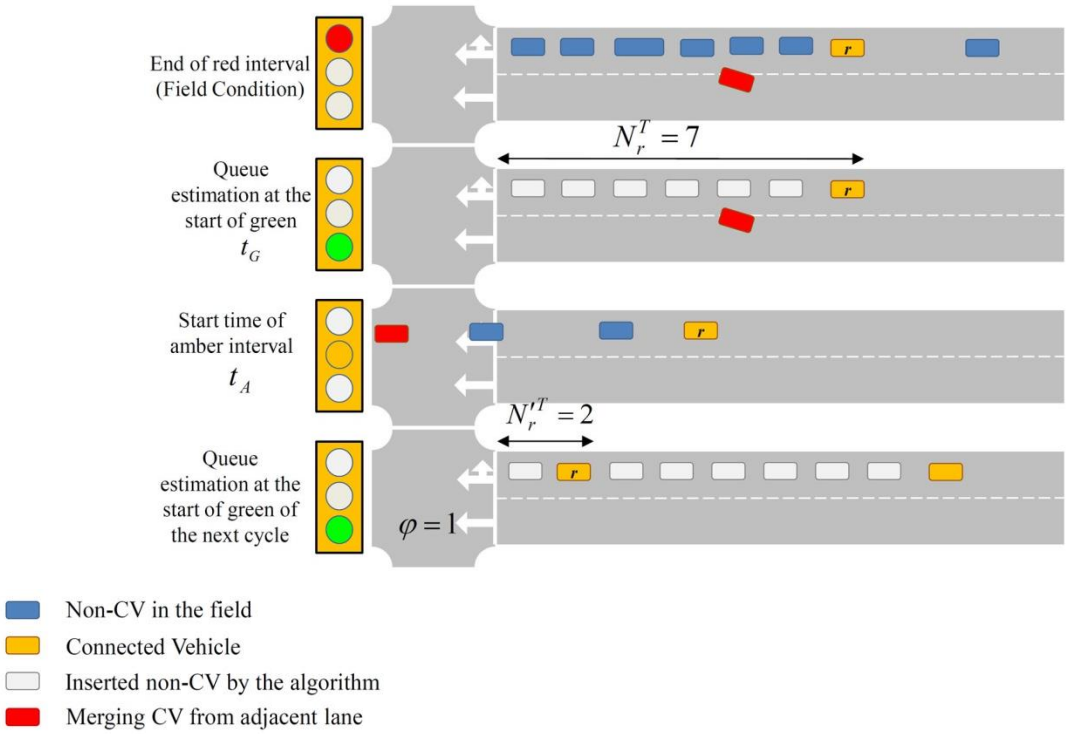


Figure 3-6: Illustration of scenario 4

Finally, Equation (3-9) is used to estimate the saturation flow rate for the cycle. To account for the possible few discharging vehicles during the amber interval, an end gain equal to 1 second is assumed in this algorithm. The number of discharged non-CVs ($N_r''^{non-CV}$) is estimated by considering N_r^T and $N_r'^T$ as is explained later Section 3.2.7 and Equation (3-13).

$$S_C = 3600 \times \frac{N_r''^{non-CV} + \varphi}{t_A - t_G - L + E} \quad (3\text{-}9)$$

Where,

φ = Total number of CVs that have passed the stop line since the start of green, (*veh*);

t_A = Start time of amber interval, (*s*);

E = End gain = 1 s.

3.2.5 Scenario 5

Scenario 5 is characterized as over-saturated, with one or more CVs in the stopped queue passing the stop line during the green interval. Figure 3-7 illustrates this case in which there are two CVs in the stationary queue. The last CV to be able to discharge during the green interval is designated as CV_l and, as in Scenario 4, the first CV in the unserved portion of the queue is designated as CV_r.

Similar to Scenario 4, $N_r''^{non-CV}$ is computed using Equation (3-13); however $N_r''^{non-CV}$ is estimated considering the remained queue upstream of CV_l. Finally, the saturation flow rate for the cycle is estimated using Equation (3-10).

$$S_C = 3600 \times \frac{N_l^T - N_l^{CV} + \varphi + N_r''^{non-CV}}{t_A - t_G - L + E} \quad (3\text{-}10)$$

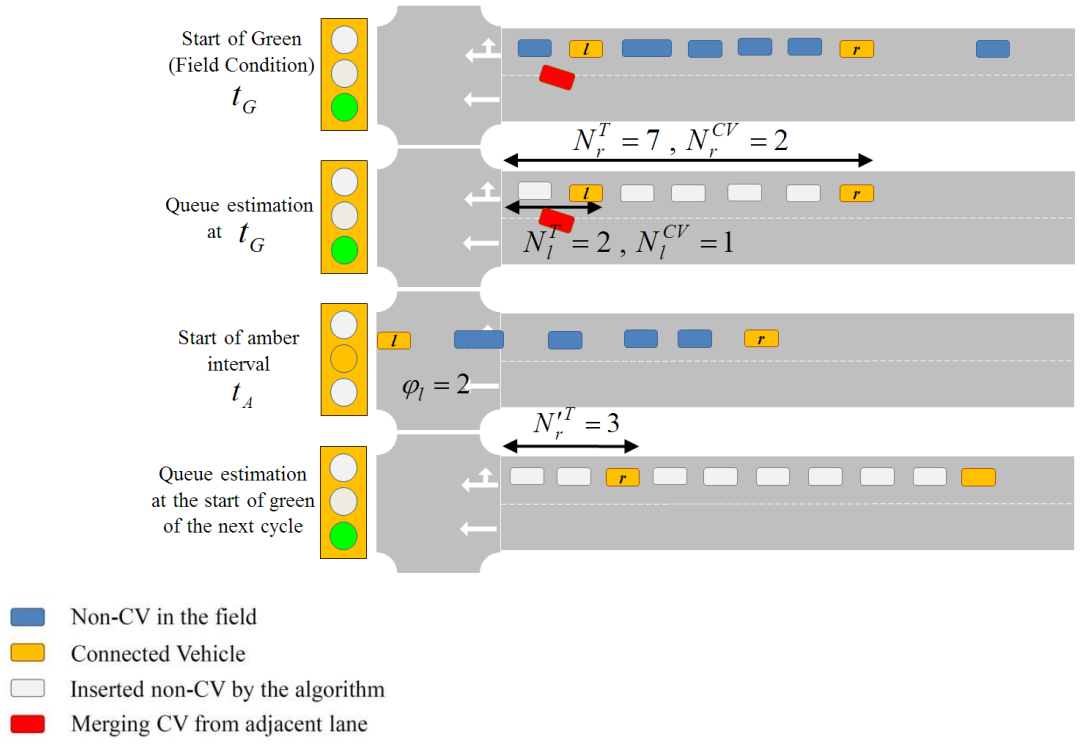


Figure 3-7: Illustration of scenario 5

3.2.6 Scenario 6

Scenario 6 is characterized as over-saturated, with one or more CVs passing the stop line during the green interval and the fourth vehicle in the queue is able to be identified. Figure 3-8 illustrates this case. The saturation flow rate for this Scenario is estimated using Equation (3-11).

$$S_C = 3600 \times \frac{\left(N_l^T - N_{4^{th}}^T\right) - \left(N_l^{CV} - N_{4^{th}}^{CV}\right) + \left(\varphi - \varphi_{4^{th}}\right) + N_r''^{non-CV}}{t_A - t_{4^{th}} + E} \quad (3\text{-}11)$$

It should be noted that, based on the above methodology, the estimation of saturation flow rate for an oversaturated phase is lagged by one cycle. This means that the saturation flow rate for an oversaturated phase on cycle (*C*) will be available and estimated by the methodology at the start of the same phase on cycle (*C+1*).

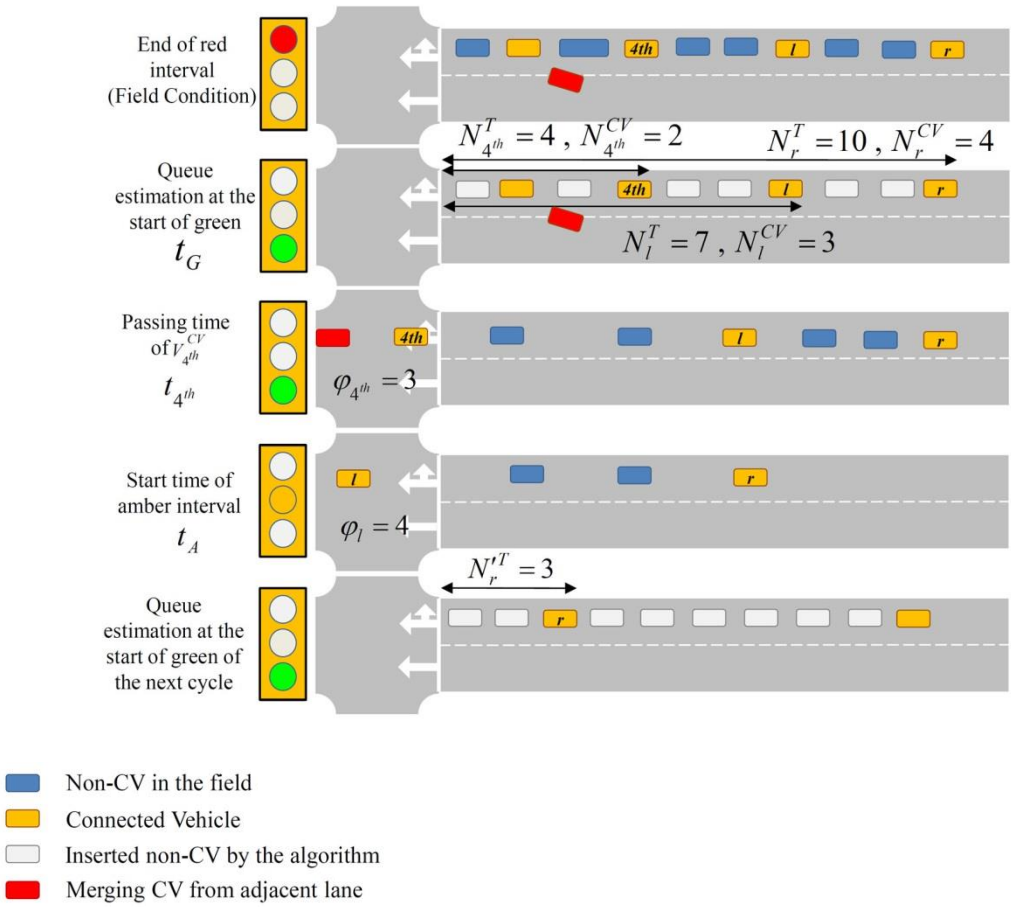


Figure 3-8: Illustration of scenario 6

3.2.7 Exclusive Turning Bays

One of the challenges in estimating the saturation flow rate at low CV LMP is the existence of exclusive turn bays adjacent to the target lane. When the stationary queue at the end of the red interval extends upstream of the turning bay, some of the vehicles in the queue will access the turning bay to make their turning movement and should not be considered when estimating the saturation flow rate for the curb lane. This is illustrated in Figure 3-9 which depicts a signalized intersection approach consisting of a single shared through and right-turn lane, and an exclusive left-turn lane. The left-turn lane has length of equal to α. In this figure, CVs are shaded grey and are numbered. Non-CVs which will remain in the curb lane to make either a through or right-turn movement are shaded black. Non-CVs which will make a left-turn are shaded white and labelled with a letter. Figure 3-9(a) through (d) illustrate the location of the vehicles on the approach at different instances in time as follows:

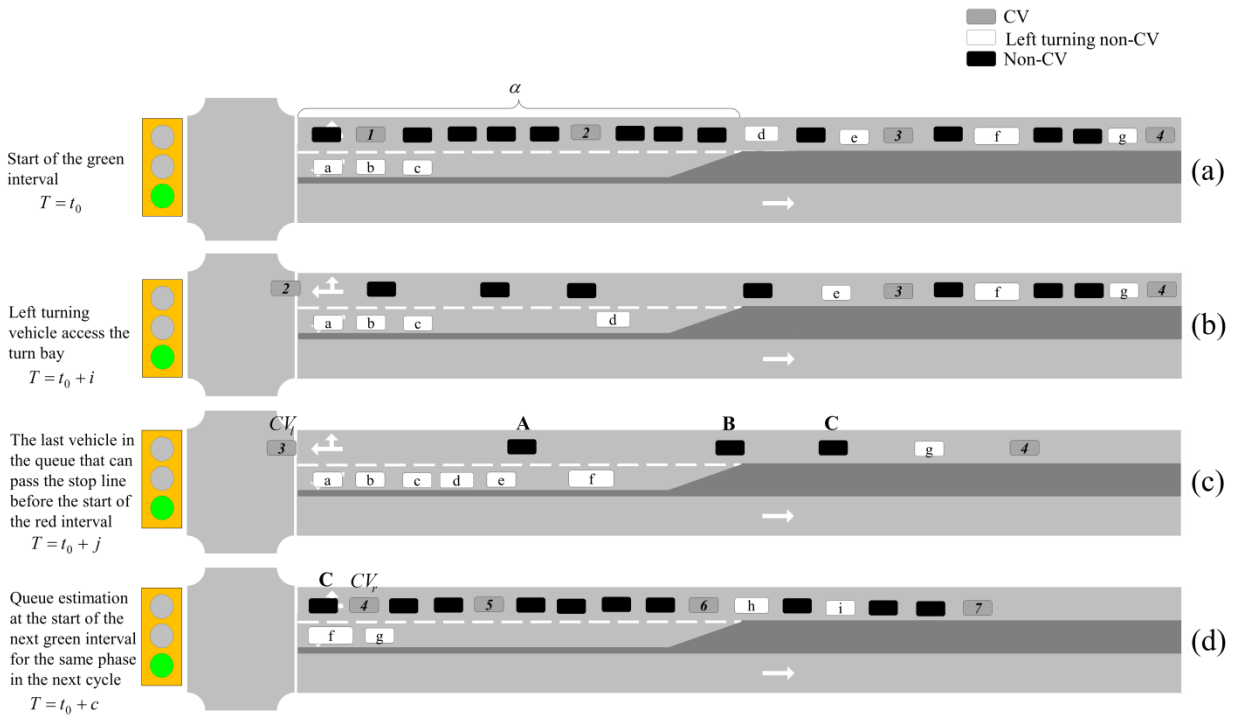


Figure 3-9: Non-CV estimation on lanes with adjacent turn bays; Traffic condition at (a) the start of the green interval, (b) some time during the green interval, (c) close to the end of the green interval, (d) start of the same phase in the next cycle

Figure 3-9(a) depicts the position of vehicles at the start of the green interval (having duration g) at time t_0. The vehicles on the approach are in a stationary queue and there is a mix of CVs and non-CVs. Furthermore, some of the vehicles in queue in the curb lane upstream of the left-turn bay intend to access the left-turn lane (i.e. vehicles labelled *d* through *g*).

Figure 3-9(b) depicts the position of vehicles at some time (t_0+i; $0 \leq i < g$) within the green interval, when a number of vehicles that were part of the queue have already crossed the stop line.

Figure 3-9(c) depicts the position of vehicles at some time (t_0+j; $i \leq j < g$) close to the end of the green interval. The most recent CV to pass the stop line before the start of the red interval (in this case vehicle 3) is labelled as CV_l. Note that until this moment, the vehicles between CV_3 and CV_4 have not been served but may cross the stop line sometime between (t_0+j) and the end of the amber interval. In this scenario vehicle *A* will cross the stop line before the end of the phase. Vehicles *B* and *C* arrive on red and subsequently vehicle *B* completes a right turn during the red interval. Figure 3-9(d) depicts the position of vehicles at the start of the next green interval for the same phase.

In Chapter 2, we proposed methodologies to estimate the number of non-connected vehicles between two consecutive CVs. Thus, from Figure 3-9(a), we would estimate 4 vehicles between the first and second CVs; 6 vehicles between the second and third CVs; and 5 vehicles between the third and fourth CVs. However, as indicated in Figure 3-9(b), some of the vehicles upstream of the left turn bay will access the left turn lane and will not cross the stop line in the curb lane, and therefore should not contribute to the calculation of the saturation flow rate for the curb lane.

In this section we propose an extension to the methodology presented in Chapter 2 to improve the saturation flow rate estimation accuracy for lanes with adjacent exclusive turning lanes.

We begin by making the assumption that non-CVs positioned less than α from the stop line will travel in the curb lane until they pass the stop line. However, some of the vehicles located upstream of the turn bay may use the turn bay. We can estimate the number of non-CV that use the left turn bay by using the turn movement ratios obtained from CV data as a random sample of the traffic stream. Consequently, the number of non-CVs between two consecutive CVs on a lane with an adjacent turn bay is estimated and corrected using Equation (3-12).

Considering the situation shown in Figure 3-9(c) and (d), if an unserved queue remains at the end of the phase, then the number of non-CVs discharged between CV_l and the first CV remaining in the unserved queue (CV_r) (shown as CV_4 in Figure 3-9(d)) is calculated using Equation (3-13). The second term of this equation estimates the number of unserved vehicles by approximating the number of non-CVs in front of CV_r at its new position (i.e. P_r'/γ), and multiplying that by a factor that takes into account those remaining non-CVs that have completed a right turn on red in a lane with shared right turn movement. The third term in this equation removes detected CVs completing a right turn on red from the saturation flow rate estimation.

$$N_j^{T(Corrected)} = \left[\left(N_j^T - N_{j-1}^T - 1\right) \times \left(1 - \beta \times \omega\right)\right] + 1 + N_{j-1}^{T(Corrected)} \quad (3\text{-}12)$$

Where,

$N_j^{T(Corrected)}$ = Corrected total number of vehicles (CVs and estimated non-CVs) in the stopped queue measured from the stop line up to and including the CV_j that will pass the stop line using the curb lane, (*veh*);

N_j^T = Total number of vehicles (CVs and estimated non-CVs) in the stopped queue on the curb lane measured from the stop line up to and including CV_j, (*veh*);

N_{j-1}^T = Total number of vehicles (CVs and estimated non-CVs) in the stopped queue on the curb lane measured from the stop line up to and including CV_{j-1}, (*veh*);

$$\omega = \begin{cases} R'_{Right}: & \text{If the adjucent bay is for right turn movement} \\ R'_{Left}: & \text{If the adjucent bay is for left turn movement} \end{cases}$$

R'_{Right} = Exponentially smoothed right turn ratio, ($0 \leq R'_{Right} \leq 1.0$);

R'_{Left} = Exponentially smoothed left turn ratio, ($0 \leq R'_{Left} \leq 1.0$);

$$\beta = \begin{cases} \left(1 - \dfrac{\alpha - P_{j-1}}{P_j - P_{j-1}}\right) & P_{j-1} < \alpha, P_j > \alpha \\ 1 & P_{j-1} \geq \alpha, P_j > \alpha \\ 0 & P_{j-1} < \alpha, P_j \leq \alpha \end{cases}$$

P_j = The distance from the stop line to the front of the CV_j in the stopped queue, (*m*);

P_{j-1} = The distance from the stop line to the front of the CV_{j-1} in the stopped queue, (*m*);

$$N_r''^{non-CV} = \left[N_r^{T(Corrected)} - N_l^{T(Corrected)} - 1\right] - \left[\left(\frac{P_r'}{\gamma}\right) \times \left(\frac{1}{1 - \chi \times R'_{Right}}\right)\right] - \chi \times RTOR^{CV} \qquad (3\text{-}13)$$

Where,

$N_r''^{non-CV}$ = Number of non-CVs estimated to discharge from the curb lane during the green interval between CV_l and CV_r, (*veh*);

$N_r^{T(Corrected)}$ = Corrected total number of vehicles measured from the stop-line to and including CV_r (the first CV in the unserved portion of the queue) in the initial queue, (*veh*);

$N_l^{T(Corrected)}$ = Corrected total number of vehicles measured from the stop-line to and including CV_l (the last CV that have passed the stop line during the green interval) in the initial queue, (*veh*);

P_r' = The position of CV_r relative to the stop line at the start of the next cycle, (*m*);

γ = Assumed average distance headway for vehicles in a stationary queue, (*m*);

$$\chi = \begin{cases} 1 & \text{If the lane consists of a shared right movement} \\ 0 & \text{Otherwise} \end{cases}$$

$RTOR^{CV}$ = Number of detected CVs completing right turn maneuvers during a red interval, (*veh*);

3.3 Evaluation

The evaluation of the proposed methodology is performed within the VISSIM microsimulation software (PTV, 2012). As presented in Figure 3-10, signal timings, loop detector data, and vehicle data (including vehicle ID, speed, current lane/link, coordinates along the link), are exported to a Microsoft SQL database in real-time at the rate of 10 times per second. These data are used to compute the true (ground truth) saturation flow rates and also to generate BSM like data required as the inputs to the proposed methodology. Custom software was developed in Visual Basic.Net that controls the VISSIM software and also implements the methodology for calculating the saturation flow rates on the basis of only the CV data.

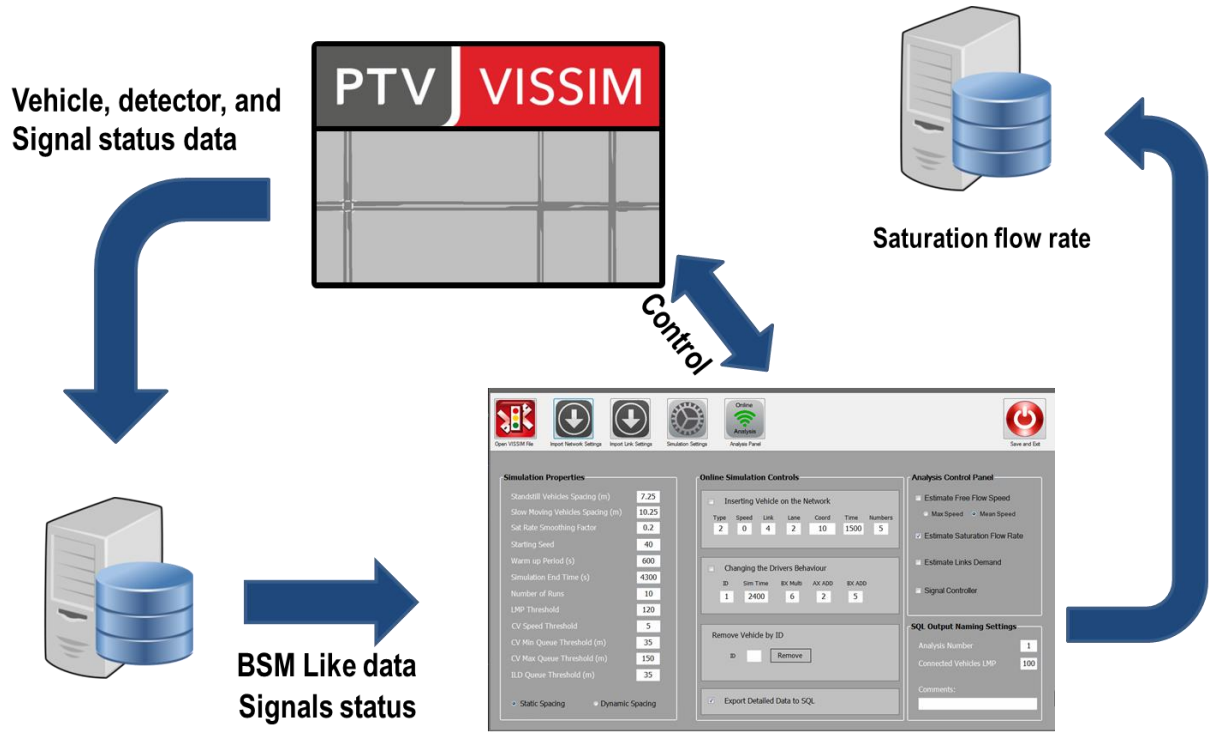


Figure 3-10: Simulation setup

The "true" (ground truth) saturation flow rates for individual lanes in VISSIM were calculated by extracting data from the simulation and applying the direct measurement of saturation flow rate presented in the HCM (HCM, 2010). The ground truth method determines the end of queue on a lane by detecting vehicles with speeds of less than 5 km/h. During the green interval, vehicles which are part of the queue are counted crossing the stop line until the last vehicle in the queue. It should be noted that the tail of the queue is continuously updated during the green and amber intervals. The saturation flow rate is obtained by dividing the number of discharged vehicles by the observation period. To remove the impact of start-up lost time, estimations start when the 4th vehicle in the queue crosses the stop line. As per the HCM the time when the 4th vehicle crosses the stop line is usually achieved within 10 seconds from the start of green. Therefore, in the ground truth saturation flow rate estimation method, if the 4th vehicle does not cross the stop line within 10 seconds, then the software assumes that there is some other reason the vehicle is delayed in discharging (e.g. insufficient discharge space) and a start-up lost time of 2 seconds is assumed and estimation of the saturation flow rate is made from the start of the green interval by including all vehicles in the queue.

The proposed methodology was evaluated for different traffic situations and geometries. The following sections provide details and evaluation results for each of evaluation scenarios.

3.3.1 Isolated Intersection

A four-legged isolated intersection, as shown in Figure 3-11, was modelled in VISSIM. The intersection was controlled by a fixed-timed signal with a cycle length of 90 seconds. The traffic stream consisted of 90% passenger cars and 10% trucks (VISSIM default HGV) which was constant during the whole simulation. The SS/CDH/2 queue estimation method was used for this analysis. The static search distance was set to 250 metres, and the distance headway (γ) was assumed to be 8 meters. The following four cases were considered for evaluating the algorithm:

1. The left turn lane on the westbound approach: a queue spillback is simulated on the downstream link after 15 minutes into the simulation, which lasts for 15 minutes.
2. The southbound approach permits through, right turn, and left turn movements. The lane group containing the shared through/right and through lanes is examined. This lane group is also affected by the queue spillback on the downstream link.

3. The eastbound approach permits through, right, and left turn movements and simulates an incident 50m along the downstream link. The lane group containing the shared through/right and through lanes is examined.
4. The northbound approach permits only left turn and through movements. The examined lane group includes the two through lanes and there is no impediment on the downstream link.

All four cases were tested for two levels of traffic demands (high and low) as shown on Figure 3-11 and for two LMP (20% and 100%). For each case and combination of traffic demand and LMP, ten simulation runs were performed using different seeds (40 runs in total for each case). Each simulation was run for 40 cycles (1 hour) after a 10-minute warm up period (total duration of 4200 seconds). In each cycle, the saturation flow rate for the studied lane group was estimated using the proposed method and the ground truth saturation flow rate was computed. The estimates were also averaged using the exponential smoothing method in which a weighting of 20% was applied to the current estimate and 80% on the smoothed value from the previous cycle.

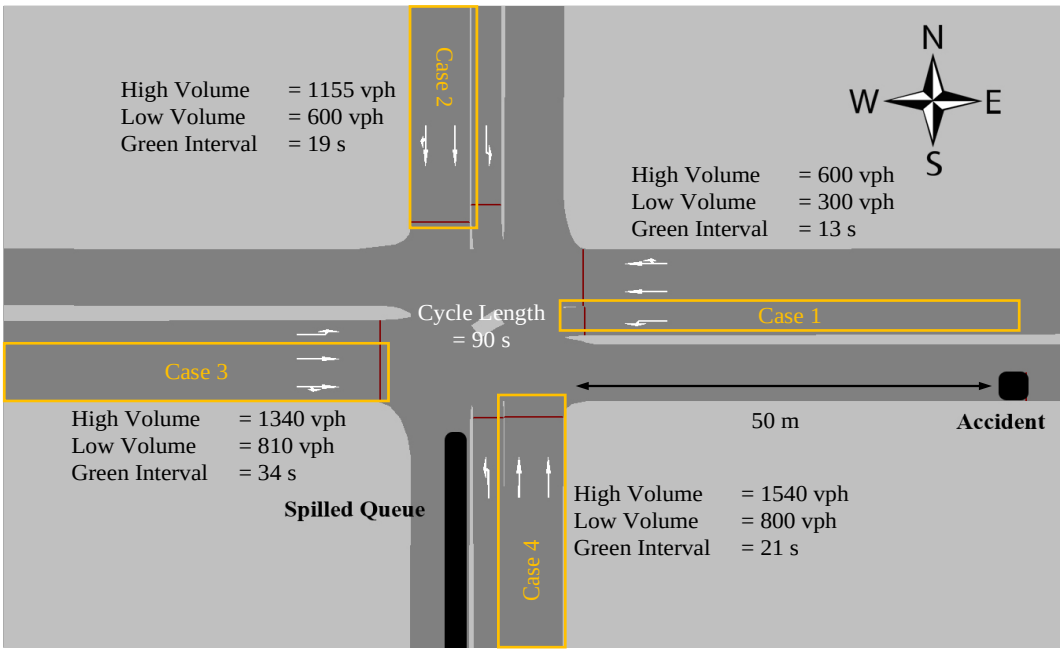


Figure 3-11: The simulated intersection, lane closures, and examined lane groups

Results

Figures 3-12(a) to (h) illustrate the ground truth and estimated saturation flow rate for each cycle of one simulation run for each of the four cases under the heavy traffic demand and for 100% and 20% LMPs.

It should be noted that ground true results are not the same for the cases with LMP = 100% and LMP = 20% even though the same simulation seeds are used. This occurs because for each model run, we simulate four classes of vehicles; CV passenger cars; non-CV passenger cars; CV heavy trucks; and non-CV heavy trucks. For LMP = 100%, the vehicle class proportions are 90%, 0%, 10%, 0% respectively. For the cases with LMP = 20%, the vehicle class proportions are 18%, 72%, 2%, 8%, respectively. Thus, even when the simulation seed remains constant, the change in vehicle class proportions, leads to small changes in the traffic stream composition and therefore in the ground truth results.

In Case 1, an incident is modeled to create a queue that spills back onto the outbound leg. The spillback begins 15 minutes into the simulation and lasts for 15 minutes. The presence of the queue prevents vehicles from discharging from the studied lane group (WB LT) and effectively reduces the saturation flow rate. We can observe this temporal variation in the ground truth saturation flow rates in Figures 3-12(a) and 3-12(b). Note that, as mentioned before, each simulation run is subject to different stochastic processes and therefore the ground truth curves for the LMP = 20% and LMP = 100% are not the same, even for the same case. We can also observe that the saturation flow rate estimated using the proposed methodology closely replicates the ground truth saturation flow rate. As expected, the accuracy of the estimates is higher for LMP = 100% than for LMP = 20%, but even with LMP = 20%, the estimates are very close to the ground truth.

The remaining three cases exhibit less severe temporal variations in the ground truth saturation flow rates and for all cases, the proposed methodology provides estimates of the saturation flow rate that closely replicate the ground truth.

In order to provide quantitative measure of performance, the absolute relative error at each cycle is calculated using Equation (3-14).

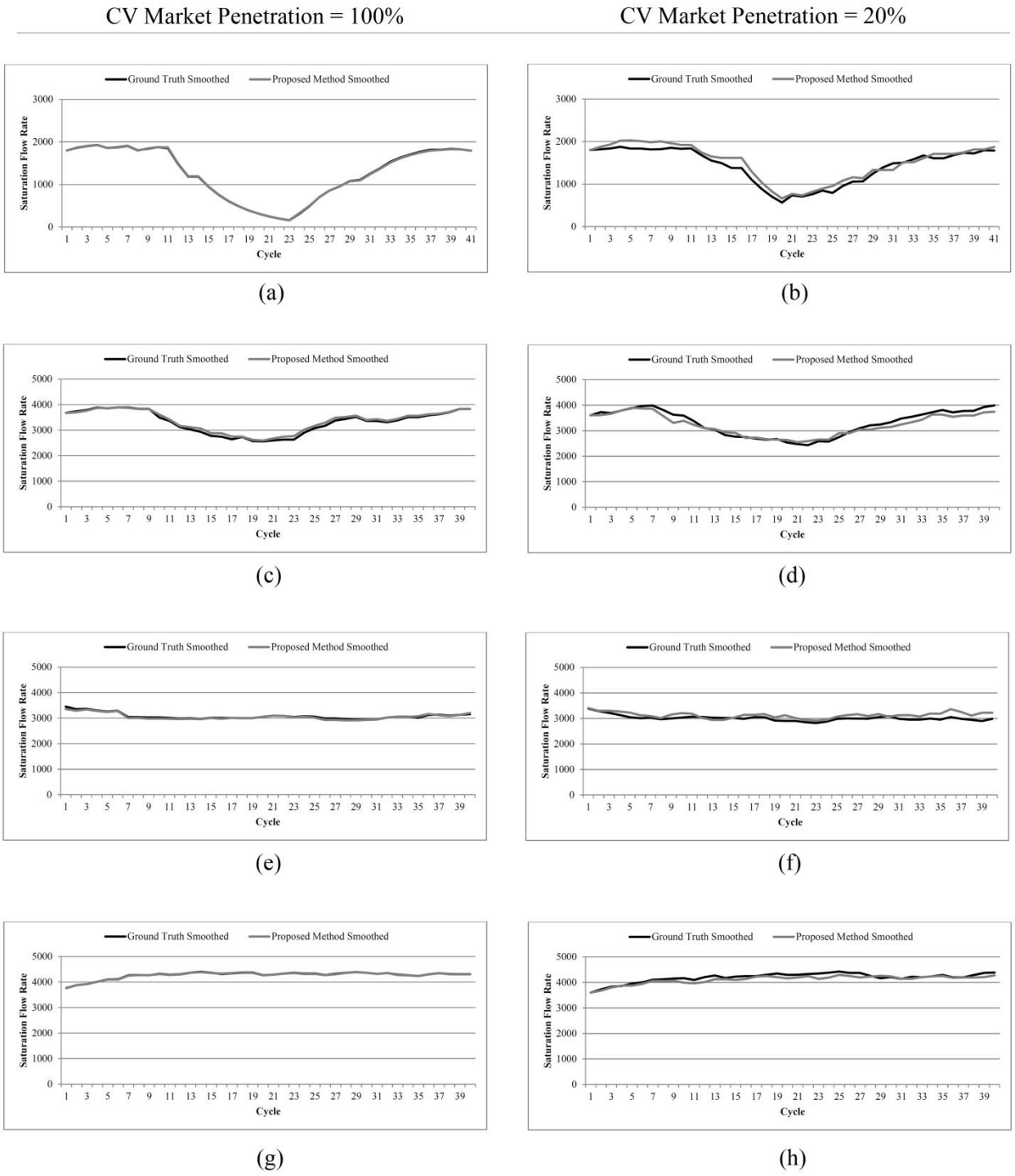


Figure 3-12: Ground truth and estimated saturation flow rates for a single simulation run for each scenario and LMP combination (high traffic demands). (a) Scenario 1 LMP=100; (b) Scenario 1 LMP=20; (c) Scenario 2 LMP=100; (d) Scenario 2 LMP=20; (e) Scenario 3 LMP=100; (f) Scenario 3 LMP=20; (g) Scenario 4 LMP=100; (h) Scenario 4 LMP=20

$$ARE_C = \left| \frac{S_C^{CV} - S_C^{ILD}}{S_C^{ILD}} \right| \quad (3\text{-}14)$$

Where,

ARE_C = Absolute relative error for cycle *C;*

S_C^{CV} = Smoothed saturation flow rate obtained using the proposed method at cycle *C*, (*vph*);

S_C^{ILD} = Smoothed ground truth saturation flow rate at cycle *C*, (*vph*).

Table 3-2 provides aggregate measures of the performance of the proposed methodology for each combination of case, traffic demand, and LMP. For each combination, the statistics reported are based on 10 simulation runs. The table provides the minimum ARE, maximum ARE, mean ARE, and 85th percentile ARE.

Table 3-2: Aggregated absolute relative error (%) for each scenario using 10 runs.

Traffic Demand	Scenario no.	LMP = 100%				LMP = 20%			
		Min ARE	Max ARE	85th Percentile ARE	Mean ARE	Min ARE	Max ARE	85th Percentile ARE	Mean ARE
High	1	0.03	5.35	1.98	1.04	0.34	23.74	15.08	8.97
	2	0.04	7.50	4.48	2.11	0.10	9.31	5.66	3.37
	3	0.12	4.93	3.05	1.81	0.54	13.32	8.83	5.70
	4	0.02	1.24	0.63	0.37	0.11	5.61	3.86	2.23
Low	1	0.06	10.43	3.42	1.79	0.64	34.71	18.27	9.37
	2	0.07	7.95	4.34	2.29	0.26	18.77	12.24	6.27
	3	0.08	5.96	3.29	1.74	0.44	14.83	9.93	6.33
	4	0.03	3.59	1.97	1.05	0.18	7.87	4.66	2.74

All values are in percent (%)

In general, the proposed method shows high performance in estimating the saturation flow rate. For LMP = 100%, the largest ARE was 10.4%, but the largest 85th percentile error for any case was only 4.5% (Case 2 – High traffic demand) and the largest mean ARE was only 2.3%. This high level of performance is encouraging, but is not unexpected. For LMP = 20%, errors are larger, but overall the proposed methodology still performs quite well. The mean ARE ranges from a low of 2.2% to a maximum of 9.4%. The largest mean error occurs for Case 1 under low traffic

demands. This is expected as Case 1 experiences the most significant temporal variation in the saturation flow rate and the low traffic demand combined with a LMP of 20% provides the fewest CVs in the traffic stream from which to obtain data. However, calculated errors for this case include saturation flow rates with values of less than a 1000 vph, where small differences in the estimations and the ground truth result in a high relative error. Yet, for this case, the results show that on average estimations deviate from the ground truth by only approximately 100 vehicles per hour.

3.3.2 Corridor

Section 3.3.1 provided a limited evaluation of the proposed methodology for an isolated intersection, two levels of traffic volumes, two levels of CV market penetration rates, and using a fixed-timed signal control. Therefore, it is not clear how the methodology performs when it is implemented for more than one intersection, various traffic conditions or other CV market penetration rates, and if other types of signal controls are used.

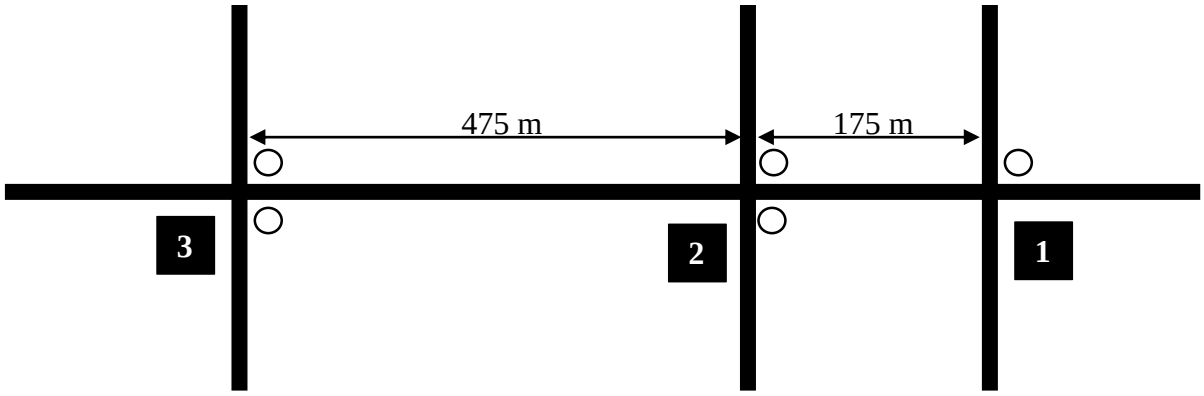
Thus, the objective of this section is to overcome the limitations of the previous evaluation, by evaluating the proposed methodology on a signalized corridor which contains a wider variety of intersection approach geometries; considering a range of operating conditions from Levels of Service (LOS) A through F; considering a range of CV LMP from 10% - 100%; and using actuated signal controllers.

As shown in Figure 3-13(a), the methodology was evaluated on a hypothetical corridor with three intersections. The geometries, the lane configurations, and signal scheme of these intersections are presented in Figure 3-13(b) and (c). Pedestrian crossings are modeled at intersection 3 with constant volume of 100 person per hour, but intersections 1 and 2 consist of only vehicular traffic. Near-side/far-side bus stops are modeled on the corridor (shown with white circles) with bus service every 5 minutes for both eastbound and westbound directions and with dwell time distribution of N(20,2) seconds. The traffic stream consists of 90% passenger cars and 10% heavy vehicles. The DS/DDH/2 queue estimation method was used for this analysis.

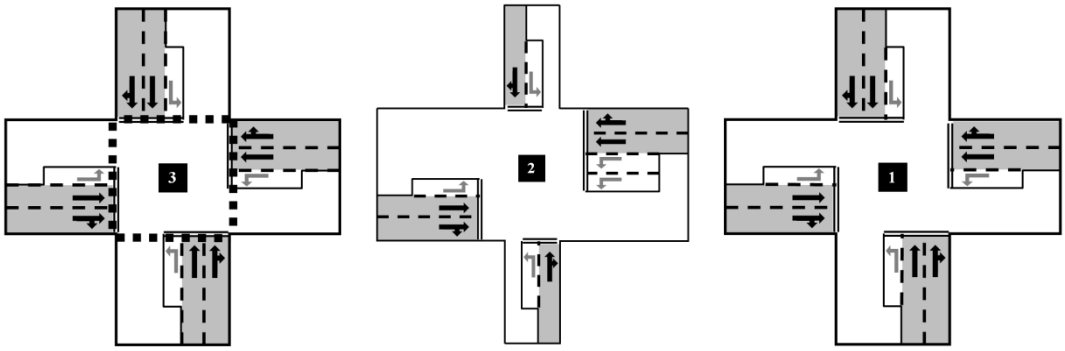
Synchro (Trafficware, 2006) was used to optimize semi-actuated signal timings and coordination on the corridor for LOS C at all intersections with the cycle length of 100 seconds, and the traffic demand on the input links were varied to achieve other levels of service . Turning movement

proportions for all approaches were set to 10% right and left, and 80% through traffic. The CV LMP was varied from 10% to 100%. Ten runs with different random number seeds were performed for each combination of LOS and CV LMP. Each simulation was run for 4200 seconds including a 600-seconds warm-up period following by an analysis period of 3600 seconds. Vehicle demand rate remained temporally constant throughout each simulation run. An exponential smoothing method was used to average the estimates at each cycle by applying weighting of 20% to the current estimate and 80% to the last smoothed estimate from the previous cycles. Thus an estimated and ground truth saturation flow rate value was obtained for each through lane and each shared through/right turn lane on each approach at the end of each signal cycle.

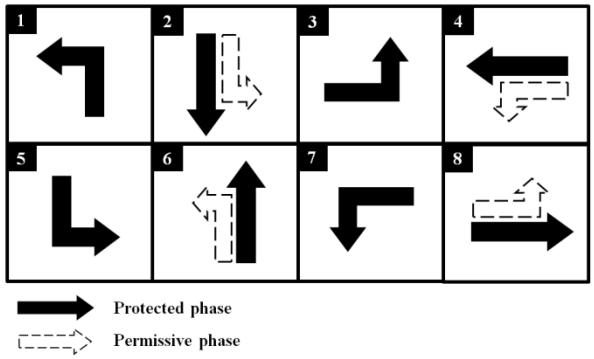
Figure 3-14 illustrates the raw and smoothed ground truth and estimated saturation flow rates as a function of the simulation time for the westbound Through + shared Through/Right turn lane group at intersection 1 for a single simulation run (LMP = 30%; LOS = F; Repetition 1). The temporal variation in the saturation flow rate occurs as a result of limited discharge storage space and queue spillback from intersection 2. Note, the HCM method for estimating the ground truth saturation flow rate requires that there be more than four vehicles in the queue at the end of the red interval. Consequently, if four or fewer vehicles were in queue, no ground truth saturation flow rate value and no estimation error could be determined for that cycle. The LOS for each saturation flow rate value was determined on the basis of average vehicle delay for the lane group computed over the entire simulation run.



(a)



(b)



(c)

Figure 3-13: (a) Study Network; (b) Lane group configurations; (c) Signal phasing scheme

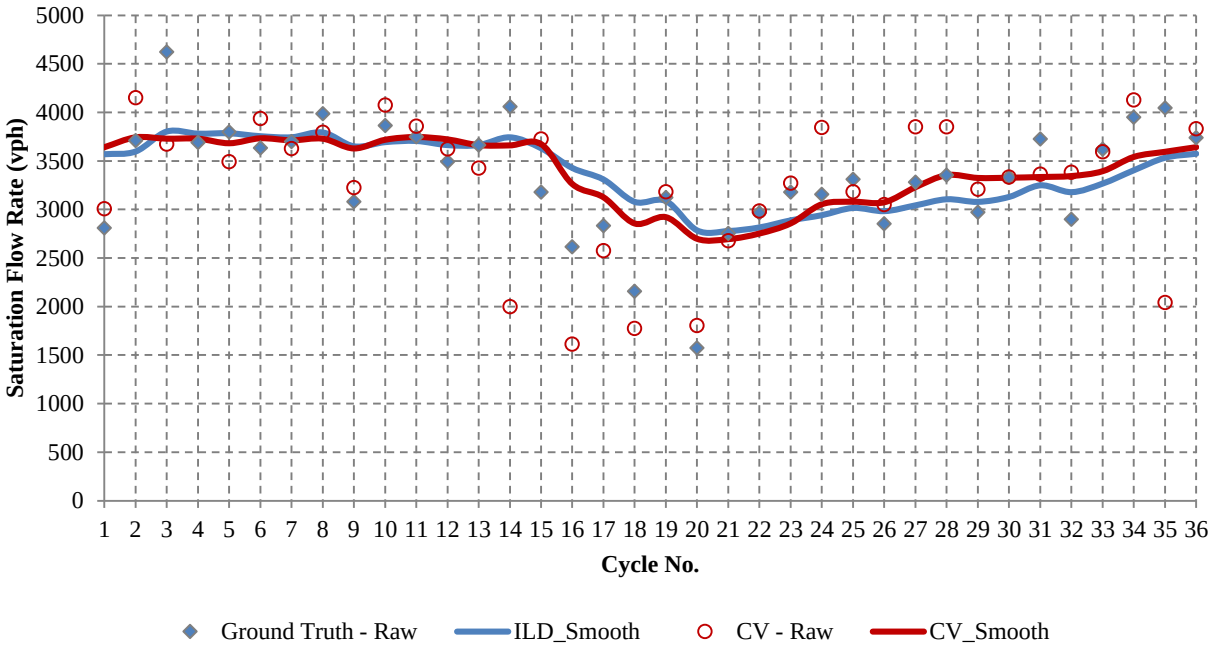


Figure 3-14: Ground truth vs. proposed method saturation flow estimations for a study approach at LMP 30% and LOS F.

Results

The evaluation was performed to answer the following questions:

1. What is the accuracy of the saturation flow rate estimations for through and shared through/right turn movements using the proposed methodology for different levels of services, market penetration of connected vehicles, and using actuated signals?
2. Does the proposed method provide a level of estimation accuracy that is expected to be sufficient for practical purposes?

The results are presented and discussed to address each of the above questions in turn.

Accuracy of the Saturation Flow Rate Estimation Model:

Table 3-3 presents the Mean Absolute Relative Errors (MARE) of the smoothed estimated and ground truth saturation flow rate values at each cycle aggregated across all approaches on the corridor and categorized by LOS and LMP, calculated using Equation (3-15). Absolute Relative Error (ARE) is calculated using Equation (3-16). The sample size associated with each combination of LOS and LMP is also presented in this Table.

$$MARE_{LOS}^{LMP} = \frac{1}{N \times C_{LOS}} \times \sum_{C \in LOS} \sum_{n}^{N} ARE_{C,n}^{LMP} \quad (3\text{-}15)$$

$$ARE_{C,n}^{LMP} = \left| \frac{S_{C,n}^{CV} - S_{C,n}^{ILD}}{S_{C,n}^{ILD}} \right| \quad (3\text{-}16)$$

Where,

$MARE_{LOS}^{LMP}$ = Mean Absolute Relative Error calculated using N lane groups, and all estimations performed in cycles with CV penetration rate of *LMP* and level of service of *LOS*.

$ARE_{C,n}^{LMP}$ = Absolute relative error calculated in cycle C for lane group n while CV penetration rate was *LMP;*

$S_{C,n}^{CV}$ = Smoothed value of the saturation flow rate obtained from the proposed methodology at cycle C for lane group n, (*vph*);

$S_{C,n}^{ILD}$ = Smoothed ground truth saturation flow rate calculated at cycle C for lane group n, (*vph*);

C_{LOS} = Total number of cycles which are part of the LOS.

N = Total number of lane groups

The results show that as the LOS improves, the sample size becomes dramatically smaller with an average of only 4.5 observations for LOS A, 290 for LOS B, and 2,000 or more for LOS C through F. The very small number of observations for LOS A occurs because under these conditions, the queue rarely exceeds four vehicles and then the ground truth saturation flow rate cannot be estimated. Given the very small sample sizes LOS A was excluded from the rest of the analyses.

Table 3-3: Mean absolute relative errors and sample sizes for each combination of LOS and LMP

LMP	MARE (%)						Sample Size					
	LOS						LOS					
	A	B	C	D	E	F	A	B	C	D	E	F
10	3.3	4.2	3.9	4.1	3.6	6.1	10	261	1805	1231	832	5968
20	2.6	5.2	4.0	4.3	3.2	5.7	7	275	1937	1043	843	5930
30	3.6	4.5	3.5	3.8	3.2	5.4	1	233	1853	1289	956	5686
40	3.9	4.7	4.0	3.8	3.0	4.8	2	285	1983	1102	1019	5676
50	7.7	3.8	3.2	3.9	2.8	4.6	3	280	1764	1217	887	5836
60	3.2	4.3	3.1	3.4	2.7	4.1	3	279	1924	1105	843	5903
70	4.1	4.6	2.8	3.0	2.8	3.6	4	270	1854	1176	930	5824
80	2.2	3.3	2.7	2.7	2.3	2.8	3	272	2353	3211	1778	5870
90	2.4	2.2	2.3	2.2	1.9	2.1	3	261	2240	3001	2124	5774
100	1.6	2.2	1.6	1.6	1.2	1.4	7	203	2491	2808	2062	5828

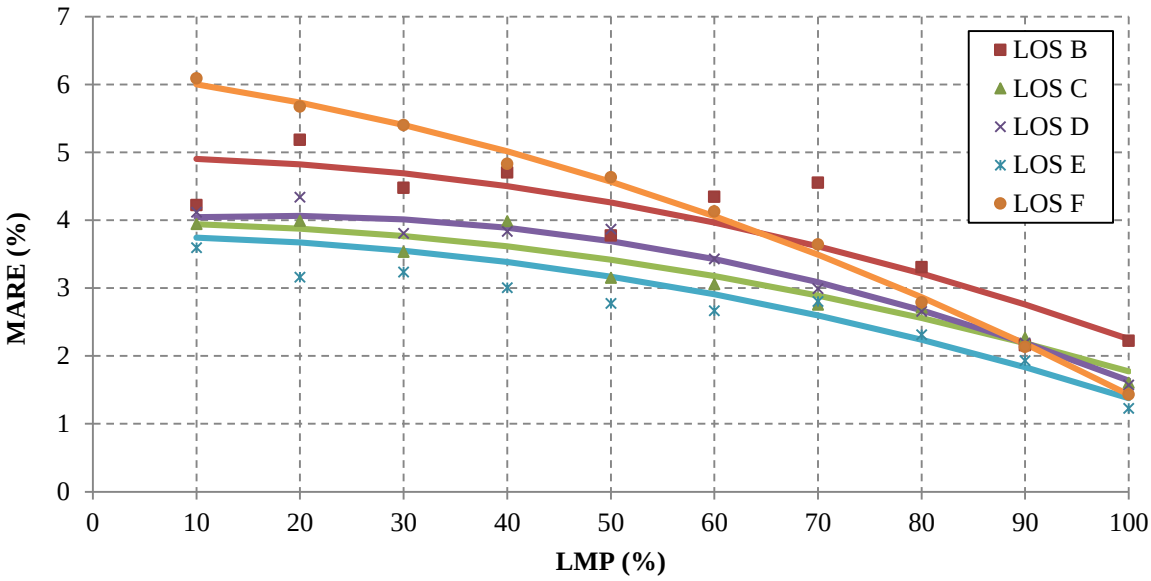
Regression analyses were performed on the data shown in Table 3-3 to develop models to represent the performance of the proposed methodology at different LOS and LMP. The dependent variables used in these regressions were the MARE of estimations weighted by the number of available observations at each combination of LOS and LMP. A range of model structures were tested and evaluation considered the explanatory power of the model and the statistical validity of the coefficients (at the 95% confidence interval). The best regression function was a second order polynomial (Equation (3-17)) and corresponding plots are shown in Figure 3-15.

$$MARE_{LOS(\phi)} = C_0 + C_1 \times LMP + C_2 \times LMP^2 \tag{3-17}$$

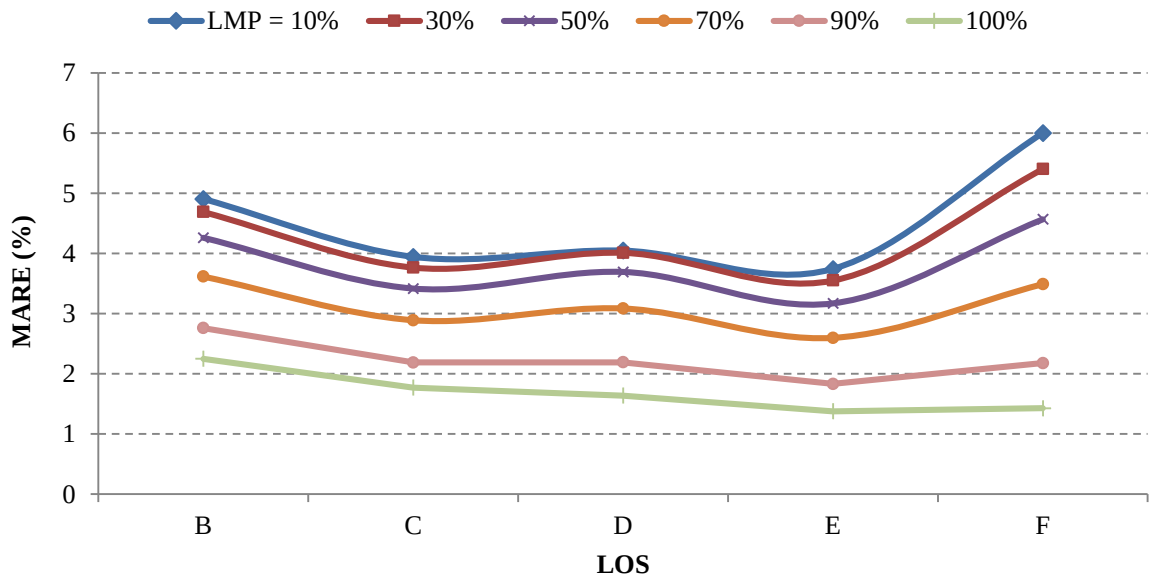
Where,

$$0 < LMP < 1$$

$$\phi = \begin{cases} B: & C_0 = 4.932 & C_1 = 0 & C_2 = -2.682 \\ C: & C_0 = 3.965 & C_1 = 0 & C_2 = -2.193 \\ D: & C_0 = 3.956 & C_1 = 1.266 & C_2 = -3.584 \\ E: & C_0 = 3.767 & C_1 = 0 & C_2 = -2.389 \\ F: & C_0 = 6.211 & C_1 = -1.791 & C_2 = -2.990 \end{cases}$$



(a)



(b)

Figure 3-15: Regression plots of MARE as a function of LOS and LMP.

Several observations can be made on the basis of Figure 3-15:

1. The maximum MARE is approximately 6% and occurs for LOS F at LMP = 10%.

2. The minimum MARE is approximately 1.2% and occurs for LMP = 100% and LOS E and F. It is important to note that even at LMP = 100% estimations by the proposed method are slightly different from the ground truth obtained using the HCM method. There are two reasons for this. First, the different methodologies use two different ways to compute the saturation flow rate. Second, the proposed CV based method cannot know if all vehicles passing the approach are connected vehicles and therefore, the module used to estimate the number of non-CVs between known CVs is always applied. Nevertheless, estimation errors at high LMP are very small.
3. The MARE decreases non-linearly as LMP increases. The relationship between MARE and LMP are quite similar for LOS C, D, and E. In contrast, the relationships for LOS B and F appear to be somewhat different. These differences can be observed more clearly in Figure 3-15 (b). At high LMP, the MARE is highest for LOS B and decreases as traffic becomes more congested (i.e. LOS F). This occurs because at LOS B queues are short and therefore both the ground truth saturation flow rate and the CV based saturation flow rate are estimated from a small number of vehicles and/or for relatively few cycles within the simulation. At lower LMP, we observe that MARE increases for LOS F. This occurs because under LOS F queues are longer and shockwaves in the queue may still exist at the beginning of the green interval when the non-CV estimation occurs, resulting in increases in the estimation errors.

Is the Accuracy Sufficient for Practical Applications?

The results show that in the worst case the MARE of estimations was about 6%. However, the question is whether this level of accuracy is sufficient for practical purposes. Answering this question is rather challenging, primarily because there is no existing benchmark or standard which defines the required level of accuracy. An ultimate test would be to carry out a field evaluation using actual CV data, estimating saturation flow rates using the proposed method, and then using these estimates within an ATSC. There are several practical challenges with such an approach. The first is the lack of CV data (with the exception of a limited number of testbeds). The second is the fact that different ATSC are likely to have different levels of sensitivity to errors in the estimated saturation flow rates and therefore the evaluation results would be specific to the ATSC used. The third is that a field evaluation would need to be able to assess the level of performance

of the ATSC which would have been achieved using conventional saturation flow rate estimation methods.

Given these challenges, we have considered an initial assessment of the suitability of the proposed method to estimate saturation flow rates with sufficient accuracy by making two different comparisons.

The first compares the results to the performance of the ILD gap-out method proposed by Henderson and Wood (2005). Henderson and Wood evaluated their model using microsimulation for an isolated intersection with no turning bays, no flared approaches, and by using only through traffic with an arrival flow rate of 600 veh/h. They computed a mean estimated saturation flow rate of 2029 vphpl and compared this to the mean true saturation flow rate of 2012 vphpl for an absolute relative error of 0.85%.

The CV methodology proposed in this chapter was evaluated for the same isolated intersection geometry and traffic characteristics with the arrival rate of 600 vph, corresponding to LOS C. Ten simulation runs with different seeds were performed for each LMP of 10% to 100% and with the duration of 3600 seconds after a 600 seconds warm-up period (i.e. 100 simulations in total). Figure 3-16 shows the absolute relative errors (computed the same way as done by Henderson and Woods) of the proposed CV method for each LMP and the error reported by Henderson and Woods (2005) for the gap-out method. The results show that the proposed CV methodology performs better than the gap-out method when LMP $\geq$ 40%, but errors are larger for LMP < 40%.

We make the following additional observations:

1. The maximum error in Figure 3-16 occurs for LMP = 10% and compares to the MAREs observed for the arterial corridor for LOS C, D, and E at LMP = 10% (Figure 3-15(a)).
2. The relationship of error as a function of LMP in Figure 3-16 differs from those shown in Figure 3-15(a) because Figure 3-16 presents the ARE of the average estimated saturation flow rates using all cycles at each LMP, in contrast to Figure 3-15(a) in which the MARE of all estimated saturation flow rates at individual cycles are presented.

Though the results indicate that the proposed CV methodology performed better than the gap-out method for LMP $\geq$ 40%, this does not directly answer the question of what level of accuracy is actually required for practical applications.

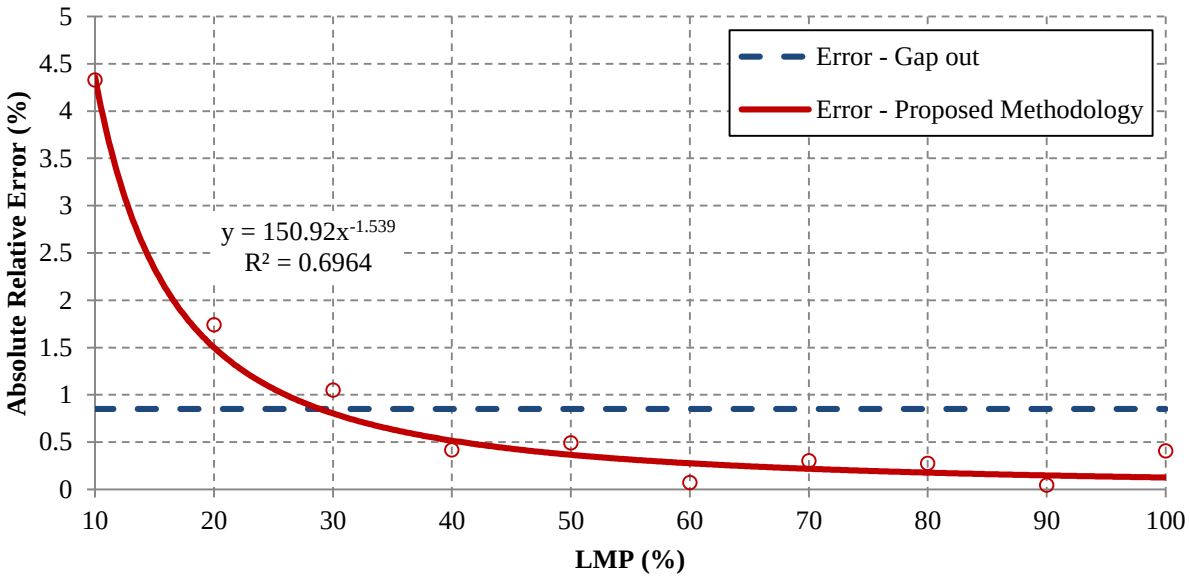


Figure 3-16: Absolute relative errors of the estimated mean saturation flow rates made by the proposed methodology at different LMP in comparison to error reported (Henderson & Wood, 2005) using the gap-out method.

The second comparison examines the distribution of the estimation errors rather than just the mean. Akcelik (1998) reports that the field sites used in developing the base saturation flow rate for each class of road in the Australian Road Capacity Guide (Miller A. J., 1968) had standard deviations of close to 200 vphpl. He also reports that according to Miller (1968) about 10% of these field sites had saturation flow rates with differences of more than 300 vphpl from the base value and states that this amount of deviation is "significant in practice". Using this as a guide, we can examine the proportion of the estimated saturation flow rates for which the absolute estimation error exceeds 300 vphpl.

Figure 3-17 shows the cumulative distribution of absolute error of estimations calculated using Equation (3-15), for all lane groups, LOS, and LMP. From the distribution we can observe that 95% of estimations have errors less than 175 vphpl and 99.9% of errors fall below 300 vphpl.

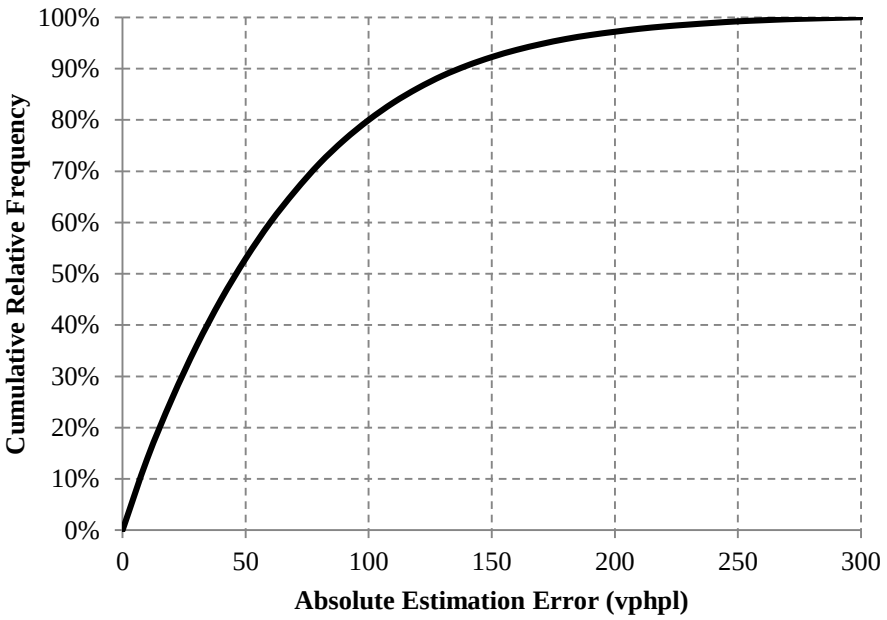


Figure 3-17: Cumulative distribution of absolute saturation flow rate errors

3.3.3 Traffic Composition

As previously mentioned earlier in Section 3, one of the factors that have influence on the saturation flow rate is the percentage of heavy vehicles in the traffic stream. Heavy vehicles usually have lower acceleration and deceleration rates. They also have less maneuver capabilities in comparison to passenger vehicles causing reduction in stop bar discharge rate and saturation flow rate. This section provides an evaluation of the proposed saturation flow rate estimation methodology against changes in the traffic composition.

The study area used in the previous section is also used for the evaluation purpose in this section. The traffic composition was changed during each simulation run as presented in Table 3-4. Ten different levels of CV market penetration rates (10% to 100%) were evaluated and 10 runs with different random seeds were performed for each CV LMP each for a duration of 4200 seconds after 600 seconds of warm-up period (100 simulation runs in total).

Two queue estimation methodologies (DS/DDH/5 and DS/CDH/5) were used and compared in this section in terms of the errors resulted from using each of the queue estimation methodologies in the estimated saturation flow rate.

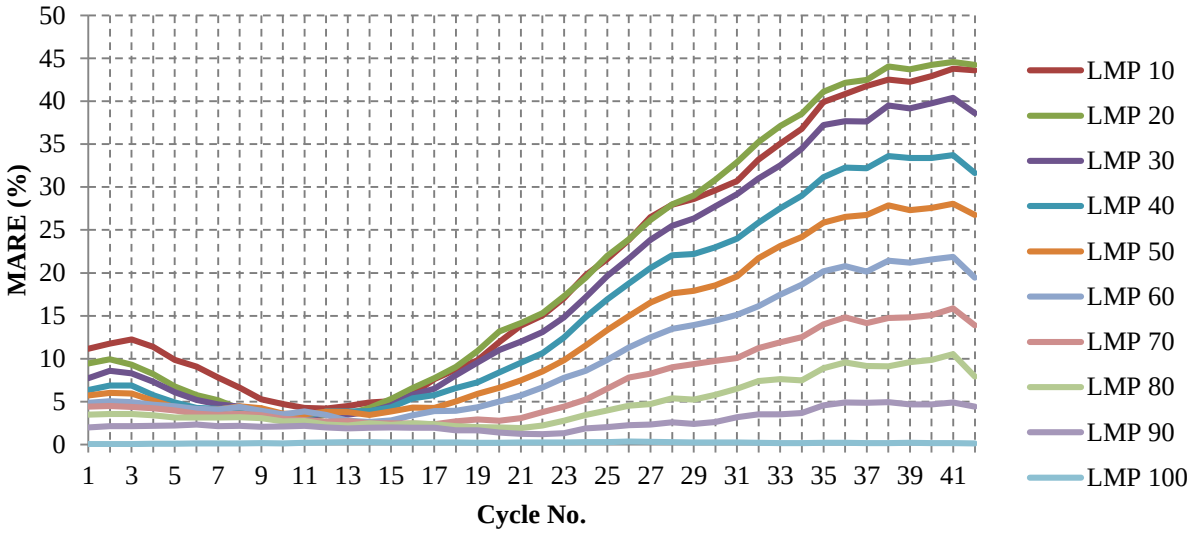
The traffic composition includes passenger cars (4.11 m to 4.76 m long) and truck trailers (TT) (17.22 m long). The percentage of truck trailers in the network was increased during each simulation run at each 10 minutes interval. To calculate γ when DS/DDH/5 was used (as it was presented in the previous chapter in Equation (2-7)), the average separation distance (ε) was assumed to be 3 metres for the non-CV estimation for the initial queue, and 6 metres for the subsequent updates in the queue during the green interval. The average vehicle length (Δ') was updated using CV data. When DS/CDH/5 was used, average distance headway (γ) was assumed to be 7.75 metres for the initial queue estimation and 10.75 metres for subsequent queue updates.

Table 3-4: Traffic composition and volume used on approaches.

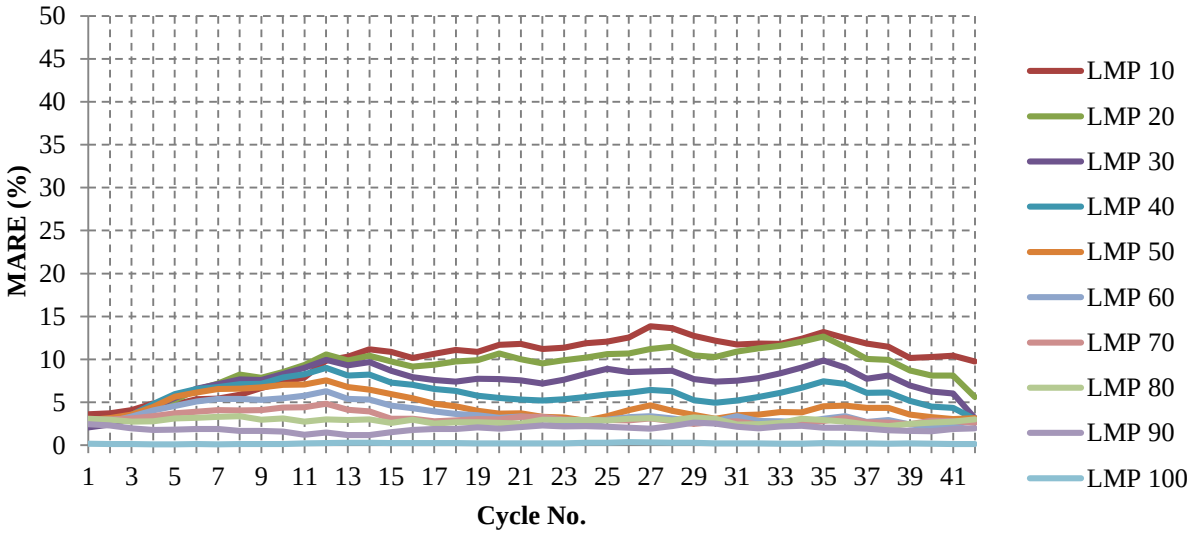
Interval (s)	0-600	600-1200	1200-1800	1800-2400	2400-3000	3000-3600	3600-4200	4200-4800
Volume (vph)	2000	2000	1700	1500	1500	1500	1500	1500
Comp. (%)	TT=0 Cars=100	TT =0 Cars=100	TT =10 Cars=90	TT =20 Cars=80	TT =30 Cars=70	TT =40 Cars=60	TT =50 Cars=50	TT =50 Cars=50

Figure 3-18 presents the saturation flow rate estimation results using constant and dynamic vehicle spacing methods at different CV market penetration rates. As shown in Figure 3-18 (a), by using a constant value for γ, an increase in the number of heavy vehicles in the network results in larger errors in the estimated saturation flow rate. The errors also increase as the CV LMP decreases in the network. The maximum estimation error (MARE close to 45) is observed when 50% of traffic consists of heavy vehicles and market penetration of connected vehicles is close to 10%. This trend in the errors was expected as changes in the traffic composition results in the actual vehicle spacing to deviate from the assumed spacing.

However, as shown in Figure 3-18 (b), errors in the estimated saturation flow using dynamic vehicle spacing are significantly lower. The dynamic spacing method uses observed CV data and the vehicle length information included in BSM part I to adapt itself to changes in the traffic composition, and therefore, the method updates the average vehicle length as the number of heavy vehicles increases in the network. This adaptation results in lower errors (MARE < 15) than obtained when using a constant value of the average distance headway, even when 50% of the traffic stream consists of heavy vehicles.



(a)



(b)

Figure 3-18: Saturation flow rate errors (MARE) at different CV LMPs using (a) constant vehicle spacing (DS/CDH/5); and (b) dynamic vehicle spacing (DS/DDH/5).

3.4 Summary and Conclusions

This section presented a technique to estimate the saturation flow rate for lane groups at signalized intersections using data from connected vehicles rather than fixed dedicated traffic sensors such as loop detectors. The methodology was evaluated for an isolated intersection as well as for a corridor considering different geometries, traffic conditions, levels of services, and a range of CV LMP. A summary of findings is as follows:

1. The proposed method is able to estimate temporally varying saturation flow rate in response to changing network conditions, including lane blockages and queue spillback that limit discharge rates, and do so with an acceptable range of errors even at low level of market penetration of connected vehicles.
2. The evaluation of the method for a range of traffic level of service show that the maximum observed mean absolute relative error (6.2%) occurs at LOS F and when only 10% of vehicles in the traffic stream are connected vehicles. The regression analysis also showed that there is a non-linear relationship between the LMP and LOS and the accuracy of the estimated saturation flow rate.
3. The adequacy of the proposed CV saturation flow rate methodology for practical applications was addressed, albeit not definitively, by comparing the performance of the proposed CV methodology to an ILD gap-out method and by examining the distribution of the absolute estimation errors. The comparison to the ILD gap-out method by Henderson and Wood (2005) showed that the proposed methodology has lower estimation error when LMP $\geq$ 40% but errors are larger than the gap-out method when LMP $\leq$ 30%. There does not appear to be an accepted standard of the estimation accuracy required for practical applications. However, according to Akcelik (1998), absolute errors in the estimated saturation flow rate exceeding 300 vphpl have practical significance. An examination of the cumulative distribution of the absolute saturation flow rate estimation errors resulting from the proposed CV methodology indicate that 99.9% of errors are less than 300 vphpl, and 95% are less than 175 vphpl. These results suggest that the proposed saturation flow rate estimation method provides a level of accuracy that is sufficient for practical applications and performs better than the ILD gap-out method when LMP $\geq$ 40%.

4. The methodology was evaluated by assuming a constant value of average distance headway as well as dynamic one derived from CV data (queue estimation methodologies (DS/CDH/5) and DS/DDH/5)). The results for a corridor with a varying traffic composition show that, in contrast to the dynamic spacing, assuming a constant value for average distance headway can causes large errors if the traffic composition deviates from the assumed one due to traffic incidents and traffic diversion. Using dynamically estimated average distance headway can help the saturation flow rate methodology to adapt to the changes in the traffic composition and to provide more reliable estimations.

The proposed method has the potential to be used in adaptive signal control systems for dynamically estimating the saturation flow rate to respond to unexpected changes in the traffic caused by road incidents such as temporary lane blockages, incidents, adverse weather conditions, etc. and does not require the deployment of dedicated traffic sensors such as loop detectors. In Chapter 5 we present an application of the methodology in retiming traffic signals in real-time to adapt to the changes in the saturation flow rate due to inclement weather conditions.

4

Estimation of Free Flow Speed on Arterial Roads using Connected Vehicles Data

This chapter presents a methodology for estimating the free flow speed on arterials using data from connected vehicles. The next section provides background information on free flow speed and a review of literature followed by sections presenting the methodology, simulation and results, and the conclusions.

4.1 Background

Free Flow Speed (FFS) is an important parameter widely used in many traffic applications such as capacity analysis of highways or signal coordination on arterial streets. The Highway Capacity Manual (HCM) (2010) defines the free flow speed on highways as "the theoretical speed when the density and flow rate on the study segment are both zero." The HCM also specifies that in practice, free flow speed on highways is expected to be observed when the flow rate is less than 1000 passenger cars per hour per lane (pc/h/ln). HCM (2010) also defines FFS on arterial roads as "the average running speed of through automobiles traveling along a segment under low-volume conditions and not delayed by traffic control devices or other vehicles."

An important application of free flow speed is in traffic signal coordination. Coordination is mainly used to reduce traffic delay and number of stops on arterial streets. In traffic signal coordination, offset is defined as "the time relationship, expressed in either seconds or as a percent of the cycle length, between [the start of] coordinated phases at subsequent traffic signals." (Traffic Signal Timing Manual, 2008). The offset between two traffic signals is calculated using Equation (4-1) (Roess et al., 2004). As presented in this equation, two factors have substantial influence on the accuracy of the offset estimation including saturation headway (h_s), and the platoon speed between two intersections (v').

$$offset = \frac{L'}{v'} - (qh_s + l) \quad (4\text{-}1)$$

Where,

$offset$ = Offset between two subsequent intersections, (*s*)

L' = Distance between subsequent intersections measured from stop line, (*m*);

v' = Average speed of vehicles, (*m/s*);

q = Number of queued vehicles per lane, (*veh*);

h_s = Saturation headway, (*s/veh*);

l = Start-up lost time

The average speed of vehicles v' is defined as the average running speed in the HCM, and is "computed as the length of the segment divided by the average running time. The running time is the time taken to traverse the street segment, less any stop-time delay" (HCM, 2000). Details on a procedure to estimate the running time for a segment of an arterial street is presented in Chapter 17 of the Highway Capacity Manual (2010). In this procedure, the running time is calculated as a function of the free flow speed adjusted by factors including vehicle proximity, delay due to turning vehicles, and delay due to other sources (i.e. parallel parking, mid-block pedestrian crossing, bicycles, etc.).

In addition to the above method, platoon dispersion models are also used for optimizing traffic signal coordination on arterial roads. One of the well-known platoon dispersion models was developed by Robertson (1969), and is used in the TRANSYT software. As shown in Figure 4-1 the model estimates the shape of a platoon of vehicles represented in a format called Cycle Flow Profile (CFP) discharged from an intersection and predicts its shape when it is expected to arrive at the downstream intersection. This prediction is based on the assumption that vehicles in a platoon travel at different speeds resulting in the platoon to disperse as vehicles travel along an arterial segment. The Robertson platoon dispersion model is presented in Equation (4-2) (Gordon & Tighe, 2005)

$$q'_{(i+t)} = F \times q_i + (1 - F) \times q'_{(i+t-1)} \quad (4\text{-}2)$$

$$F = \frac{1}{1+\alpha t} \tag{4-3}$$

Where,

$q'_{(i+t)}$ = Flow rate at the i[th] interval of the predicted platoon arriving at downstream intersection at time *t*, (vphpl);

q_i = Flow rate recorded at i[th] interval of the initial platoon, (vphpl);

t = $\beta \times$ travel time at cruise speed, (measured in units of time interval) ;

α = Platoon dispersion factor, (unitless);

β = Speed factor, (unitless)

F = Smoothing Factor, (time interval $^{-1}$)

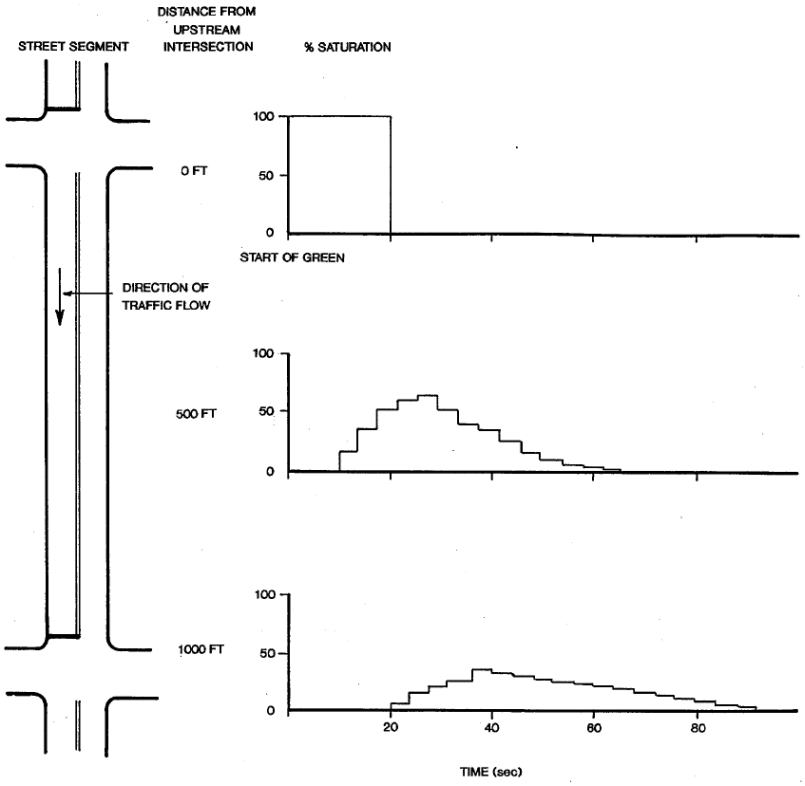


Figure 4-1: platoon dispersion between two intersections (Dunn Engineering Associates, 2005)

Aside from required calibration for parameters α and β, the model relies on an assumption about the cruise speed to correctly predict the cyclic flow profile at the downstream intersection. Platoon dispersion models have been adopted in Adaptive Traffic Signal Control (ATSC) systems such as

SCOOT for optimizing signal timings and offsets using a performance index that is based on predicted queue lengths and number of stops (Hunt et al., 1981). In SCOOT, the cruise speed is defined as "the average speed of vehicles travel at between the [upstream] detector and the stop line when not constrained by queues or red traffic signals." This definition is similar to the definition of the free flow speed used in the HCM. In SCOOT, the cruise speed is typically a constant value, obtained through field measurements or an assumed value. When field conditions are impacted by adverse weather, road surface conditions, special events, etc. that impact the cruise speed, then the use of a constant cruise speed results in incorrect flow predictions, and as a result, sub-optimal signal timing and offsets. Therefore, it is desirable to update cruise speeds in real-time to reflect changing field conditions. Free flow speed can be influenced by several factors including road surface condition (rain or snow), visibility reduction (snow, rain, and fog), wind, narrow lanes (construction), etc.

The Highway Capacity Manual (2000) indicates that on freeway facilities free flow speed (FFS) can be reduced by 2 km/h during light rain, 5 to 7 km/h when there is a heavy rain, and between 37 and 42 km/h in heavy snow conditions. Ibrahim & Hall (1994) studied free flow speed on sections of the Queen Elizabeth Way (QEW) in Mississauga, Canada for different weather categories. The study reports that during light rain precipitation, FFS drops only by a maximum of 2 km/h; light snow causes FFS to reduce by a maximum of 3 km/h; heavy rain reduces FSS between 5 to 10 km/h; and heavy snow causes FFS to reduce between 38 and 50 km/h. Rakha et al. (2008) also investigated the impact of adverse weather on FSS using traffic data collected from Baltimore, Minneapolis-Saint Paul and Seattle in the United States. Their study concluded that light rains causes FSS to drop between 2% and 3.6%, 6% and 9%, 5% and 16%, and 5% and 19% during light rain, heavy rain, light snow, and heavy snow respectively. The study also concluded that reduction in visibility caused by snow contributes to the reduction in the free flow speed. Kyte et al. (2001) examined the impact of wind speed and visibility distance on free flow speed. The study showed that the average wind speed of greater than 24 km/h and visibility of less than 300 metre cause reductions in FFS.

The impact of adverse weather on arterial speed was examined by Maki (1999). The study used traffic and weather data collected from Trunk Highway 36 in Minnesota including five signalized intersections. Data were collected during PM peak for both dry and snowy conditions with at least

3 inches of snow. The study reported that arterial speed reduced by 40% (dropping from about 71 km/h to 42 km/h) during inclement weather in comparison to the normal dry condition. Perrin et al. (2001) used traffic and weather data collected from two intersections in Salt Lake City, Utah and examined the impact of inclement weather on saturation flow rate, start-up lost time and free flow speed. The study concluded that FFS reduces by 10%, 13%, 25%, and 30% when road surface condition is wet, wet and snowing, wet and slushy, and slushy in wheel paths respectively.

Gillam & Wilhill (1992) developed traffic parameters for SCOOT system during inclement weather conditions. As part of the study, link journey time data were collected from several cities in the United Kingdoms during both dry and wet road surface condition. The study showed that link journey time increased by about 11% during wet condition in comparison to dry condition.

Mathematical models have been developed to estimate the free flow speed on arterial roads. The HCM (2010) provides a methodology which takes into account the impact of several factors on the free flow speed including the speed limit, median type, density of mid-block right or left turn access points, and presence of a curb. Similar models have also been proposed by Ye et al. (2001), Moses & Mtoi (2013) and Fazio et al. (2014). However, all these analytical models aim to estimate FFS for existing conditions on arterial roads or for design purposes and cannot dynamically estimate FFS if these conditions change.

There have been some efforts to estimate real-time travel time on arterial roads using fixed point traffic sensors such as inductive loop detectors (e.g. Zhang, 1999; Chi et al., 2001; Skabardonis & Nikolaos, 2005; Wu et al., 2012). However, the goal of these models is to estimate the time taken by vehicles to traverse a corridor and this time includes the delay experienced due to traffic control devices (as well as other sources of delay) and not the travel time at free flow speed required for signal coordination. In fact, these models require knowing the free flow speed to estimate the free flow travel time and to add delay at signalized intersections to this time to estimate the expected travel time. The FFS can be assumed or estimated using measurements from dedicated traffic sensors such as a single loop detector. FFS can be estimated from loop detector measurements using Equation (4-4) (Athol, 1965). This approach requires:

1. an assumption about average effective vehicle length (L) in the traffic stream; and
2. detectors must be located so that vehicles can reach FFS while detectors are not occupied by queued vehicles.

$$v_i = \frac{N_i}{T \cdot O_i \cdot g} \quad (4\text{-}4)$$

Where,

v_i = Spot speed at interval *i*, (*m/s*);

N_i = Traffic volume during interval *i*, (*veh*);

T = Duration of interval *i*, (*s*);

O_i = Average detector occupancy during interval *i*, (*%*);

g = Constant factor (g-factor) (g=1/L);

L = Effective vehicle length, (*m*).

Several studies have used probe vehicle data to estimate average speed or travel time on arterial and highway segments. For instance, Herring et al. (2010) used GPS data obtained from a fleet of 500 taxis in the city of San Francisco and a statistical model to estimate and predict travel times on arterial roads. Researchers have used probe data to estimate link travel time or average speed by measuring the difference in detection times of probe vehicles at entry and exit of a link segment and by knowing the length of the segment. (e.g. Cheu, et al., 2002; Argote, et al., 2011). Astarita, et al. (2006) used combined probe vehicles and loop detector data to estimate traffic flow parameters (flow and density) and derived average speed by dividing estimated flow by density. De Fabritiis, et al. (2008) developed two methods based on artificial neural networks and pattern matching to estimate and predict freeway and arterial speeds using GPS probe data. Remias, et al. (2013) used probe data to evaluate the performance of arterial roads and identify the potential to improve traffic flow by retiming traffic signals. In their method, free flow speed was used as part of estimating control delay and assumed to be the speed equivalent to the 5^{th} percentile of all travel times obtained from probe vehicles during the entire observation period.

The following observations can be made from the review of state of the art and practice:

1) Coordination of traffic signals on arterial roads and the offset between two consecutive intersections are influenced by vehicles' travel time at free flow speed. Adaptive traffic signal control systems use a predetermined value of free flow speed in their traffic model when optimizing offsets on a signalized corridor.

2) The free flow speed is impacted by several factors including on street parking, lane width, road surface condition, visibility, etc. and these factors are temporally variable.
3) The performance of adaptive traffic signal control degrades when free flow speed on a corridor deviates from the assumed one, resulting in signal timings (particularly offsets) which are not optimal for the existing traffic conditions.
4) There are several mathematical models to estimate or predict free flow speed on a particular road segment. However, these models are designed for offline analyses and are not suitable for real-time applications.
5) Methods have been developed to estimate real-time vehicle travel times on arterial roads using traditional fixed point detectors and/or probe vehicle data. These travel times are not suitable for coordinating traffic signals because the travel time includes delays that vehicles experience in queues.

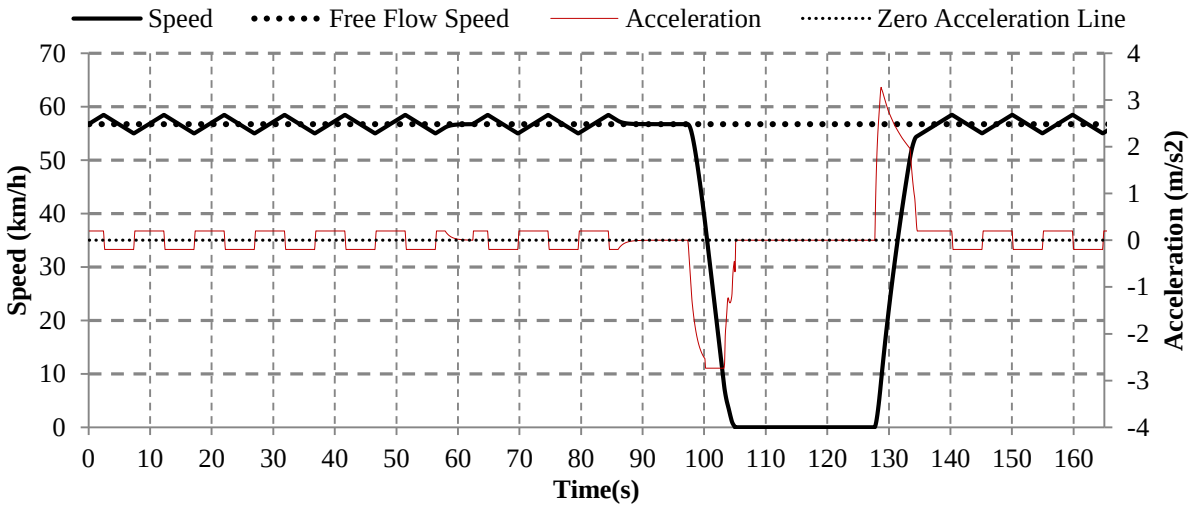
Therefore, this chapter aims to presents a methodology to estimate real-time free flow speed on arterials using CV data. The goal is use the output of this methodology as the inputs to adaptive traffic signal control systems for better optimization of signal timings.

4.2 Methodology

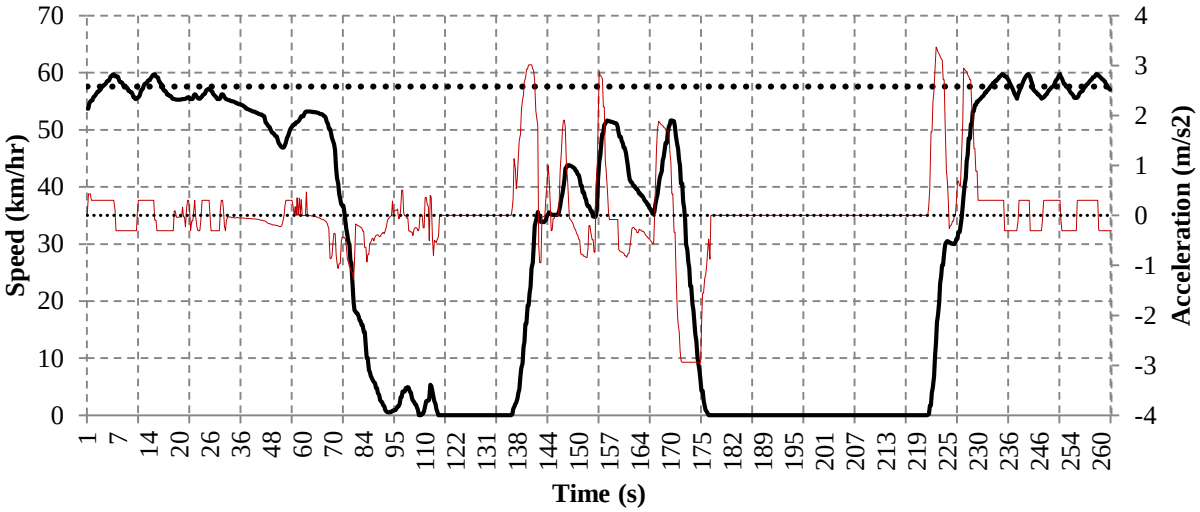
The goal of the presented methodology is to estimate the average speed that drivers of vehicles are willing to achieve on an arterial when their speed choice decision is not influenced by the traffic signal control system. As explained in the previous section, this speed is referred to as the free-flow speed. The goal is to estimate the free-flow speed in real-time using the data provided by connected vehicles. The following assumptions are made regarding the data provided by connected vehicles:

1) The data include BSM part I which is transmitted by individual vehicles at the rate of 10 times per second.
2) BSM part I includes the vehicle ID, real-time speed, acceleration, and position.
3) There is no significant latency between transmitting and receiving the connected vehicles data
4) The connected vehicle data are available and received along the corridor of interest.

The major challenge in developing the proposed methodology is to detect when a connected vehicle is traveling at free flow speed. Consider Figure 4-2(a) showing a sample speed and acceleration profiles of a vehicle simulated in VISSM 5.4 (PTV, 2012) on a hypothetical corridor shown in Figure 4-3. The vehicle of interest is traveling from East to West. Figure 4-2(a) also displays the actual free flow speed (known as the desired speed in VISSIM) of the vehicle on the corridor. In Figure 4-2(a), there are no other vehicles in front of the vehicle of interest and it can travel at free flow speed where possible. Figure 4-2(b) shows speed and acceleration profiles of another vehicle traveling through the same network, but this time with other vehicles. Figure 4-2(b) shows that the vehicle enters the network with small fluctuations in its speed around the free-flow speed. The vehicle travels at around free flow speed for about 25 seconds and its acceleration during this time is slightly positive or negative; however, it does not maintain positive or negative accelerations for a long period as it tries to maintain a constant speed. At second 25, the vehicle starts reducing its speed and maintaining negative acceleration for some time with a short period of positive acceleration followed by another long period of negative acceleration until it become stationary at about second 95. The vehicle then moves in a queue behind the stop line until it joins the stopped queue. At second 135, the vehicle accelerates and its speed increases, however, as the distance to the downstream intersection is short the vehicle encounters the next queue and stops at second 175. During this time the vehicle has large positive and negative acceleration values. At second 220, the vehicle starts to increase its speed with large positive acceleration until it reaches free flow speed at about second 230 when its acceleration starts oscillating around the zero line.



(a)



(b)

Figure 4-2: Sample vehicle speed and acceleration profiles on a corridor: (a) a single vehicle traveling through the corridor with no other traffic, (b) a vehicle traveling through the corridor with traffic.

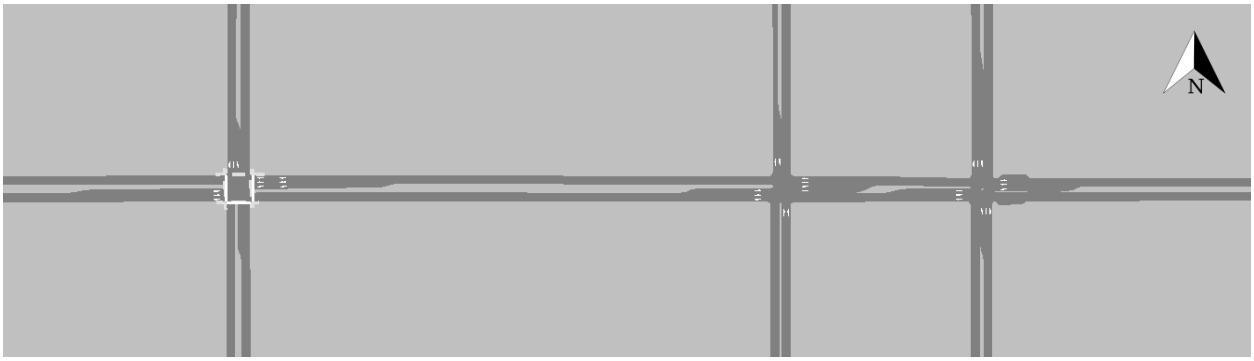


Figure 4-3: A Hypothetical corridor used to generate and study sample vehicles' speed and acceleration profiles.

The challenge is to develop a method which relies only on the CV data to estimate the free flow speed. We make several observations about vehicle behaviour, as illustrated in Figure 4-2, and then describe the logic of the proposed method:

a) A vehicle travelling at a constant speed (including free flow speed) will exhibit small fluctuations in speed and consequently small changes in the acceleration rate (i.e. small magnitude with either positive or negative sign, and for short duration) due to imperfect control of the engine power by a driver.

b) Vehicles which are slowing down to stop for a traffic signal or because of some downstream impedance (e.g. a slower moving vehicle, etc.), may do so by applying a

relatively large deceleration rate for a short period of time or by applying a smaller deceleration rate but for a longer period of time.

c) Similarly, a vehicle wishing to increase its speed may do so using a relatively large acceleration rate for a short period of time or a more moderate acceleration rate for a longer period of time. The combination of rate and duration is influenced by a range of factors including driver characteristics, vehicle characteristics, distance to the downstream obstacle, etc.

Based on the above observations, the following describes the logic used in developing the proposed algorithm. The flowchart of the algorithm is presented in Figure 4-4.

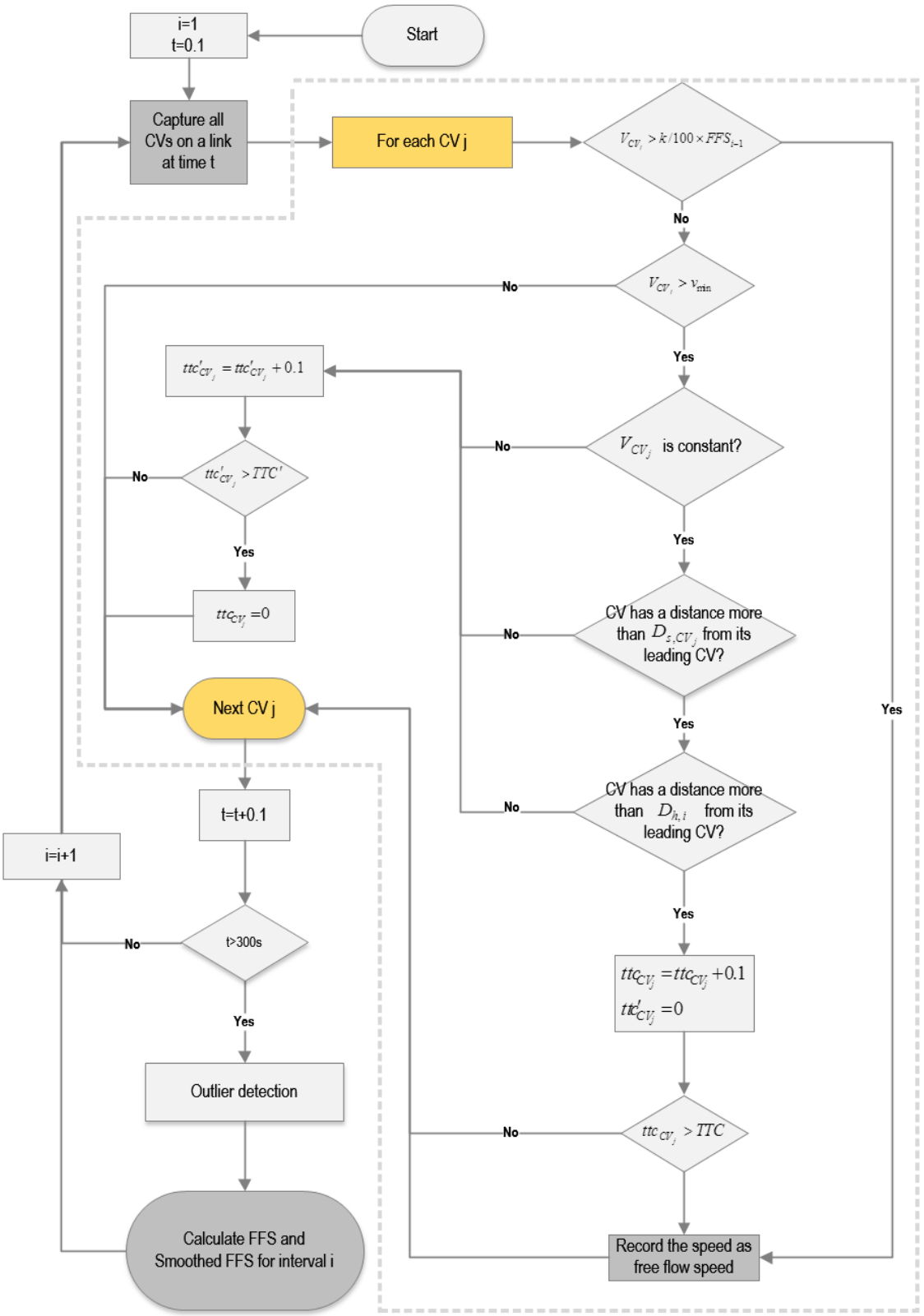


Figure 4-4: Proposed algorithm flowchart for estimating free flow speed using CV data.

4.2.1 Detecting a Vehicle Travelling at a Constant Speed

As shown in Figure 4-2, a vehicle travelling a constant speed (including free flow speed) does not exhibit large magnitude of acceleration/deceleration rates, or moderate rates for long durations. Therefore, the proposed algorithm monitors the rate of change in the speed of individual connected vehicles. It calculates an accumulated change in the speed for each connected vehicle, which is the real-time sum of the changes in the speed of each connected vehicle as new data become available. If a change in the sign of the acceleration is observed the value of the accumulated change in the speed is reset to zero. A connected vehicle is considered to be traveling at a constant rate when its calculated accumulated speed change is within a defined threshold.

A connected vehicle may also accelerate to a certain speed and keep traveling at that speed for only a short period of time before reducing its speed due to a short distance left to the next traffic signal or reacting to a leading non-connected vehicle for instance. In this case, although the vehicle may travel at a constant speed for a few seconds, it may not reach the free flow speed. To avoid such a situation, the algorithm assumes a connected vehicle is traveling at a constant speed only when the above condition is detected for at least a minimum time threshold (TTC).

In summary, at this step, the algorithm tries to detect vehicles

- with accumulated speed changes within a defined threshold (maintaining their speed at a specific rate);
- which have maintained their speed rate within the threshold for a minimum time; and
- which are travelling at a speed greater than a defined minimum threshold speed (to avoid detecting vehicles which are travelling at a constant speed but in a slow moving queue)

Using this approach, the algorithm also tries to detect connected vehicles that are reacting to non-connected vehicles which are not visible to the algorithm by monitoring their speed change and acceleration change rates.

4.2.2 Determining if Constant Speed is Free Flow Speed

Traveling at constant speed does not necessary mean that a vehicle is traveling at free flow speed. A vehicle may travel at a constant speed less than the free speed when following another vehicle. Therefore, the algorithm tries to further improve its accuracy by checking if the speed of the

vehicle of interest is influenced by a downstream vehicle. This is done by defining two separate distance headway thresholds as described in the following sections.

Threshold 1:

The first threshold is defined making use of the concept of a car following model.

Consider two connected vehicles travelling in the same direction in a single lane. Define vehicle A as the downstream vehicle and B as the upstream vehicle. There are no vehicles between A and B. Thus A is the *leader* and B is the *follower*. In a car following model, the behaviour (i.e. speed) of B is dictated by the speed of A and the separation distance between A and B (i.e. distance measured from the front of vehicle B to the rear of vehicle A).

Many different car following models have been described in the literature. Car following models form the core of traffic simulation models and are also the fundamental element in intelligent/adaptive cruise control (technology that is now widely available on new vehicles). For many of these models, there is some set of conditions when the separation distance between A and B is sufficiently large that the speed of vehicle B (the follower) is not impacted by the speed of vehicle A.

The Wiedemann 74 car following model (Wiedemann, 1974) is a psychophysical model which uses certain thresholds to understand the behaviour of the following driver under different situations. Figure 4-5 shows the different behaviour regimes captured by the Wiedemann 74 car following model. The x-axis is the difference in the speeds of the lead and following vehicle (ΔV = speed follower – speed leader). The y-axis is the distance headway.

The free driving region represents the conditions for which the following driver's choice of speed is not impacted by the leading vehicle. If we make the assumption that the lead and following vehicles are travelling at approximately the same speeds, then if the distance headway exceeds the value of SDX, we can conclude that the following vehicle is travelling at the free speed.

SDX ranges between 1.5 to 2.5 times ABX (Rakha & Gao, 2010) and ABX can be computed from Equation (4-5).

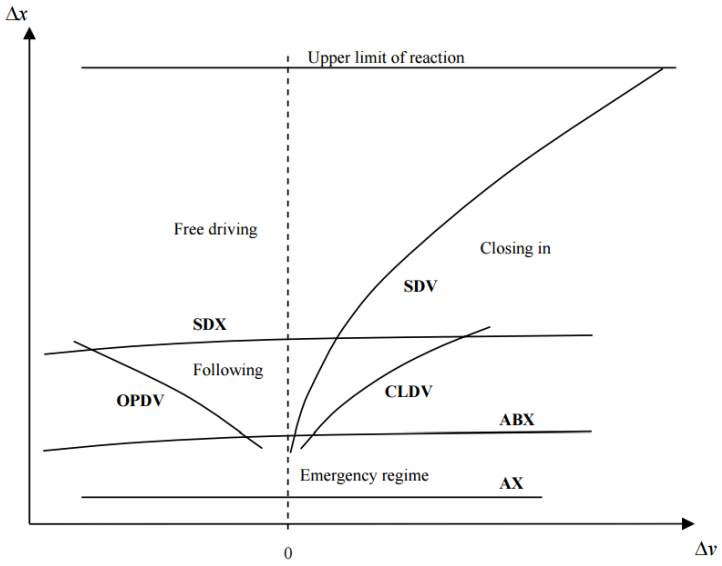


Figure 4-5: Graphical representation of the Wiedemann car following model (Olstam & Tapani, 2004).

$$ABX = AX + \left(BX_{add} + BX_{multi} \times \beta\right) \times \sqrt{v} \quad (4\text{-}5)$$

Where,

ABX = Desired minimum following distance, (m);

AX = Average standstill distance including the length of the leader vehicle + the desired separation distance, (m);

BX_{add} = Additive part of desired safety distance, (m);

BX_{multi} = Multiplicative part of desired safety distance, (m);

β = A value between 0 and 1 that is normally distributed with $\mu = 0.5$ and $\sigma = 0.15$;

v = Speed of the vehicle, (m/s).

The algorithm considers a connected vehicle is traveling at free flow speed if the distance headway to the next downstream vehicle is greater than the distance threshold calculated using Equation (4-6).

$$D_{h,i} = 2 \times ABX \quad (4\text{-}6)$$

Where,

$D_{h,i}$ = Minimum required distance headway between consecutive CVs in time interval *i* if traveling at the free flow speed estimated in the previous time interval *i-1*, (*m*).

Threshold 2:

Threshold 1 is defined assuming the lead and following vehicles are travelling at approximately the same speed. Threshold 2 is introduced assuming the following vehicle is travelling at a higher speed than the lead vehicle. We expect that drivers select their speeds with the constraint that they are able to decelerate to a stop if the lead vehicle decelerates to a stop. Similarly, if we know the speed of the leading vehicle, we can determine the minimum distance headway required between the leading and following vehicles to ensure the safe stopping distance is met. Distance threshold 2 is calculated using Equation (4-7).

$$D_{s,CV_j} = FFS_{i-1} \times t_{pr} + \frac{FFS_{i-1}^2 - V_{CV_{j-1}}^2}{2 \times \mu \times g} + \Delta_{j-1} \quad (4\text{-}7)$$

Where,

D_{s,CV_j} = Total stopping distance for CV_j if traveling at free flow speed, (*m*);

$FFS_{(i-1)}$ = Estimated free flow speed at interval *i-1*, (*m/s*);

$V_{CV_{j-1}}$ = Speed of the leading connected vehicle (CV_{j-1}) , (*m/s*);

μ = Coefficient of friction ($\mu = 0.35$)

g = Acceleration due to gravity ($g = 9.81\ m/s^2$)

t_{pr} = Perception and reaction time ($t_{pr} = 1.5s$)

Δ_{j-1} = Length of CV_{j-1}, (*m*);

Finally, we determine that the connected vehicle of interest is travelling at the free flow speed if the connected vehicle has been determined to be travelling at a constant speed and the distance headway to the next downstream connected vehicle in the same lane is greater than the maximum of distance threshold 1 ($D_{h,i}$) and distance threshold 2 (D_{s,CV_j}).

It must be mentioned that the development of thresholds 1 and 2 assumed that both the leading and following vehicles are connected vehicles. This assumption is only true when the LMP of

connected vehicles is 100%. When the LMP is much less than 100%, then it is likely that there may be non-connected vehicles in between the two connected vehicles. When this occurs, the proposed method for identifying free speed will become less accurate.

While the speed of a connected vehicle is being recorded as the free-flow speed, the vehicle is in active monitoring mode by the algorithm. During the active monitoring mode, if one of the above conditions is not satisfied, the algorithm stops recording the speed of the vehicle but continues to monitor the vehicle for a period (TTC'). If during this monitoring time period the free-flow speed conditions are satisfied again, then the algorithm starts recording the speed of the vehicle immediately without observing the minimum time (TTC) as described above. If the monitoring period expires and the free flow speed conditions have not been satisfied, then the connected vehicle is released from the active monitoring state.

Regardless of the above conditions, the speed of a connected vehicle is recorded as the free flow speed if it reaches at least specified percentage (k) of the last known value of free flow speed and ignored if it is less than a user defined threshold ($v_{\min}$).

4.2.3 Detecting Outliers

The proposed heuristic method for acquiring free flow speed data may result in the mislabelling of connected vehicle speeds (i.e. labelling speeds as free flow speeds when they are not). A sample output of the heuristic method is presented in Figure 4-6 illustrating the distribution of calculated average speed of individual connected vehicles that are labelled to be traveling at free flow speed during an estimation interval. The figure shows that the majority of speeds range from 48 to 54 km/h. However, a small number of the speeds are much lower and these data have been incorrectly labelled as free flow speeds. As such it is necessary to apply an outlier detection technique. In this methodology, since we can assume that the distribution of the free flow speeds follows a normal distribution (Fazio, et al., 2014), the Chauvenet's criterion method (1908) can be used to remove outliers from the recorded data.

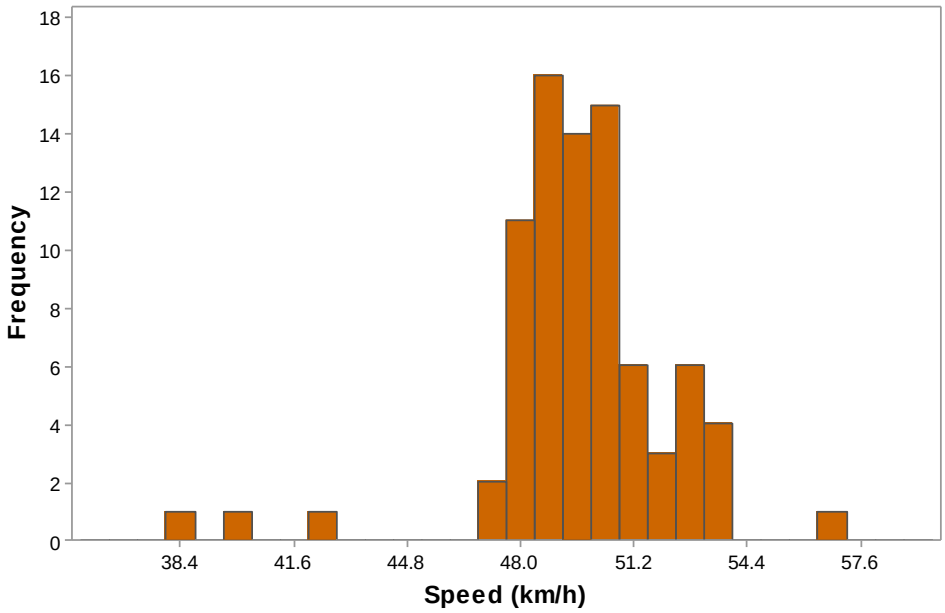


Figure 4-6: Sample distribution of speeds labelled as free flow speed during an estimation interval

The outlier detection process is shown in Figure 4-7. This process is performed at the end of each data collection interval (e.g. 5 minutes). At the end of each interval, the algorithm calculates mean (μ) and standard deviation (σ) of all the recorded free flow speeds from all CVs on a link. A probability band is calculated using Equation (4-8), which defines a boundary around the mean. Any data point outside this boundary is considered as an outlier.

$$p = \frac{1}{2n} \tag{4-8}$$

Where,

p = Probability band with mean at its centre;

n = Number of data points in the dataset (sample size).

To identify if a data point is within the probability band, the z-score of the data point is calculated using Equation (4-9) and compared to the two sided z-score associated with the probability band. If the z-score of the data point is larger than the z-score of the probability band then it is considered as an outlier and removed from the dataset. This process is performed on all data points in the whole dataset and repeated until no more outliers are detected by the Chauvenet's criterion method or the maximum of 3 iterations.

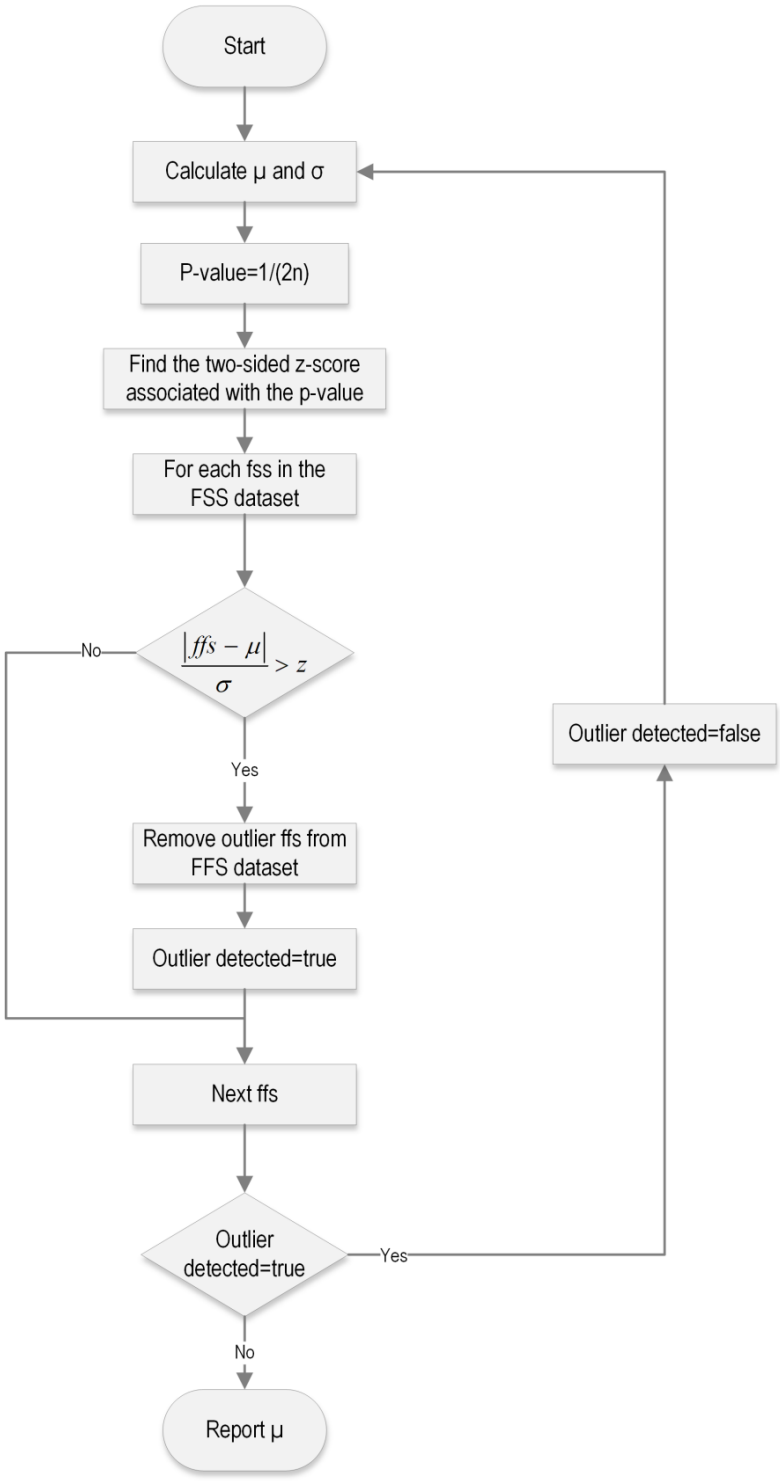


Figure 4-7: Outlier detection process using Chauvenet's criterion.

$$z = \frac{|x_i - \mu|}{\sigma} \quad (4\text{-}9)$$

Where,

z = z-score;

x_i = The value of the data point being examined.

Finally, at the end of each interval, the estimated mean FFS is smoothed using an exponential smoothing method presented in Equation (4-10).

$$FFS_i^{Smoothed} = (\alpha)FFS_i + (1-\alpha)FFS_{i-1} \quad (4\text{-}10)$$

Where,

$FFS_i^{Smoothed}$ = Smoothed free flow speed at interval *i*, (*km/h*);

FFS_i = Free flow speed at interval *i*, (*km/h*);

FFS_{i-1} = Free flow speed at interval *i-1*, (*km/h*);

α = Smoothing weight ($0 < \alpha < 1$).

4.3 Evaluation and Results

4.3.1 Simulation Setup

The methodology was evaluated using VISSIM microsimulation software (PTV, 2012). A hypothetical corridor as shown in Figure 4-8 was modeled in VISSIM using lane configuration, signal timings, and constant volumes shown in the figure. Turning ratios on all approaches were set to 80% through, 10% right and 10% left turns. All signals are semi-actuated with a 100 seconds cycle length. Ten levels of CV LMP ranging from 10% to 100% were simulated with each level simulated using 5 different random number seeds. Each simulation was 4900 seconds long with the first 900 seconds used as the warmup period, thus providing 13 five minute intervals for which data were extracted. For the simplicity of the evaluation and ease of calculating the ground truth free flow speed from the simulation, the traffic consisted of only one class of vehicles (passenger cars only).

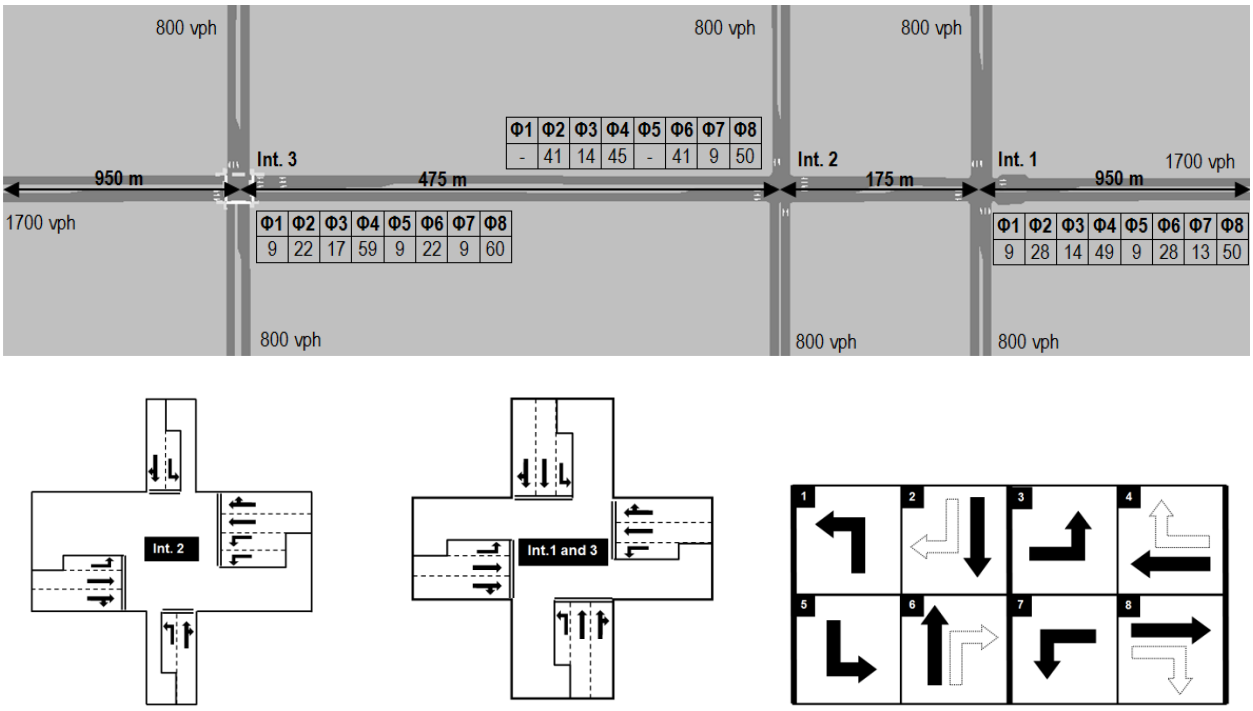


Figure 4-8: Study corridor

Two weather conditions were modeled during each simulation run. A dry road surface condition was modelled between simulation time 900 and 2900 seconds using VISSIM default settings, and a heavy snow condition was modeled after 2900 seconds into the simulation. The simulation parameters used for both conditions are presented in Table 4-1. During the snowy conditions, we expect reductions in the free flow speed in comparison to the dry conditions. It was presented earlier in this chapter that Perrin, et al. (2001) and Maki (1999) reported 30% and 40% reductions in the original free flow speed in the snowy conditions respectively. Therefore, for the purpose of this evaluation, we assumed that the desired speeds of vehicles decreased by 40% from 50 km/h to 30 km/h. Garber and Hoel (1988) stated that in snowy conditions, the deceleration rate of vehicles reduces from 27% to 20% of gravitational acceleration (74% of the dry surface condition) and Asamer, et al. (2013) reported that accelerations of vehicles during snowy condition can be reduced to 71% of the accelerations during dry road conditions. Therefore, the distributions of the acceleration and deceleration rates of vehicles in VISSIM were adjusted accordingly. Asamer, et al. (2013) also presented required parameters for modelling snowy conditions in VISSIM using Wiedemann 74 car-following model. Based on their study, assuming the desired speed of 30km/h and an acceleration rate of 71% of the original values the appropriate parameters were selected for snowy conditions as presented in Table 4-1.

Table 4-1: Desired speed distributions and car-following parameters used in VISSIM to model dry and snowy road surface conditions.

Parameter	Dry	Heavy Snow
Desired Speed	50 km/h (48 to 58 km/h)	30 km/h (30 to 35 km/h)
Car-following Parameters	Multiplicative part of safety distance (BX_{multi}) = 3 Additive part of safety distance (BX_{add}) =2	BX_{multi} = 6 BX_{add} =5

For the evaluation of the algorithm, the user configurable parameters presented in Table 4-2 were used.

Table 4-2: User configurable parameters used in the evaluation of the algorithm.

Variable	Value	Description
AX_{add}	2 m (VISSIM default)	Additive part of safety distance
BX_{add}	2 m (VISSIM default)	Additive part of desired safety distance
BX_{multi}	3 m (VISSIM default)	Multiplicative part of desired safety distance
β	1	
TTC	3 s	Minimum monitoring time of a vehicle at free flow speed before its speed can be recorded.
TTC'	3 s	Maximum time that a vehicle is dropped from active monitoring state if it does not satisfy the FFS conditions again.
$v_{\min}$	20 km/h	The minimum acceptable value as the free flow speed.
α	0.5	Smoothing weight

These values were chosen on the basis of engineering judgement, an understanding of the impact that these parameters have on the proposed method for identifying free flow speeds, and the characteristics of the test network. In general, selecting parameters values that increase threshold 1 (i.e. AX, BX_{add}, and BX_{multi}) or 2 will tend to reduce the likelihood of mislabelling low speeds as the free flow speed. However, this will also reduce the number of speed observations that are

labelled as free flow speed and consequently increases the time for the system to response when the free flow speeds change.

The appropriate value for TTC depends on the length of the links as choosing a large value for this parameter may result in no vehicle being detected on a link if the travel time is short on that link.

Finally choosing large values for α will decrease the response time in the reported estimated free flow speed by the algorithm, while increase the fluctuations in the estimations. On the other hand, small values of α will results in a slow response from the algorithm if the free flow speed changes on a corridor.

4.3.2 Results and Discussion

Figure 4-9 identifies the links on the corridor for which results are presented in this section.

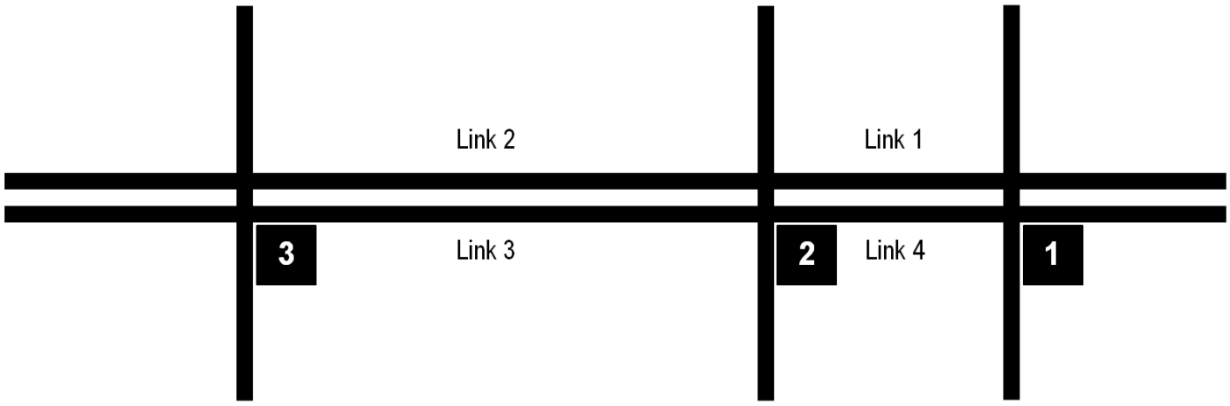


Figure 4-9: Links used in the evaluation of the methodology.

Figure 4-10 shows a sample of estimated free flow speeds for vehicles traveling on link 1 using the proposed methodology and the data obtained from connected vehicles when the market penetration is 60%. The figure shows the individual connected vehicles average recorded free flow speeds and the cleaned free flow speeds after outliers are removed from the dataset. The figure also shows the average free flow speed calculated at each interval using the cleaned free flow speed data from all connected vehicles and the smoothed value.

Figure 4-11 presents the smoothed estimated FFS results for the approach links numbered in Figure 4-9 as an average of all smoothed FFS at each corresponding interval using 5 simulation runs. Figure 4-12 presents the absolute relative error of the smoothed free flow speed estimations in

comparison to the ground truth calculated for each link at each interval and using 5 simulation runs. The absolute relative error for each interval during each simulation run is calculated using Equation (4-11). The estimated free flow speeds were examined at the end of the 6^{th} estimation time interval, which is the last interval for which the dry weather condition applies; and the end of the 13^{th} time interval when the last estimation is performed for the snowy condition at the end of the simulation run. The ground truth FFS was assumed to be 53 km/h and 32.5 km/h for dry and snowy conditions respectively which are the mean value of ranges shown in Table 4-1. Based on the results the following observations can be made:

1. The estimated FFS values correspond quite closely to the expected FFS values for the dry and snowy conditions (the majority of estimations have absolute errors of less than 4 km/h);
2. The accuracy of the estimated FFS values does not appear to be very sensitive to the LMP of connected vehicles for the dry road surface condition; and
3. The accuracy of the FFS estimates for the snowy condition increases as the LMP increases.

During the dry road surface condition, the traffic in the network is less congested since traffic signals and offsets are designed for these conditions. Therefore, vehicles with constant speeds are usually traveling at free flow speed and with larger distance headways. Under these conditions, even a few connected vehicles can be used to estimate the link free flow speed. However, during the snowy condition, capacity is reduced and congestion and slowdowns occur on the corridor resulting in speed choice of more vehicles being influenced by their leading vehicles. Consequently, as the LMP of connected vehicles increases, the algorithm can better identify these following vehicles and as result provides more accurate estimations.

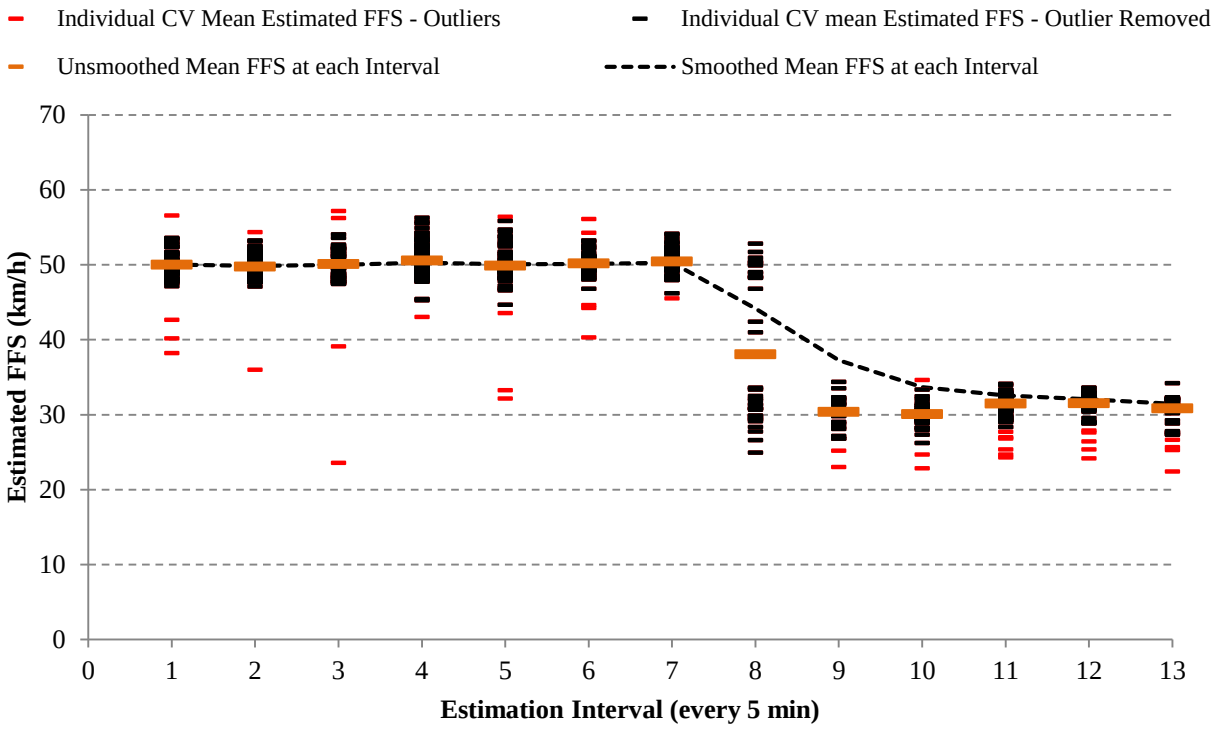


Figure 4-10: Sample estimated free flow speed data on link 1 (CV LMP = 60%)

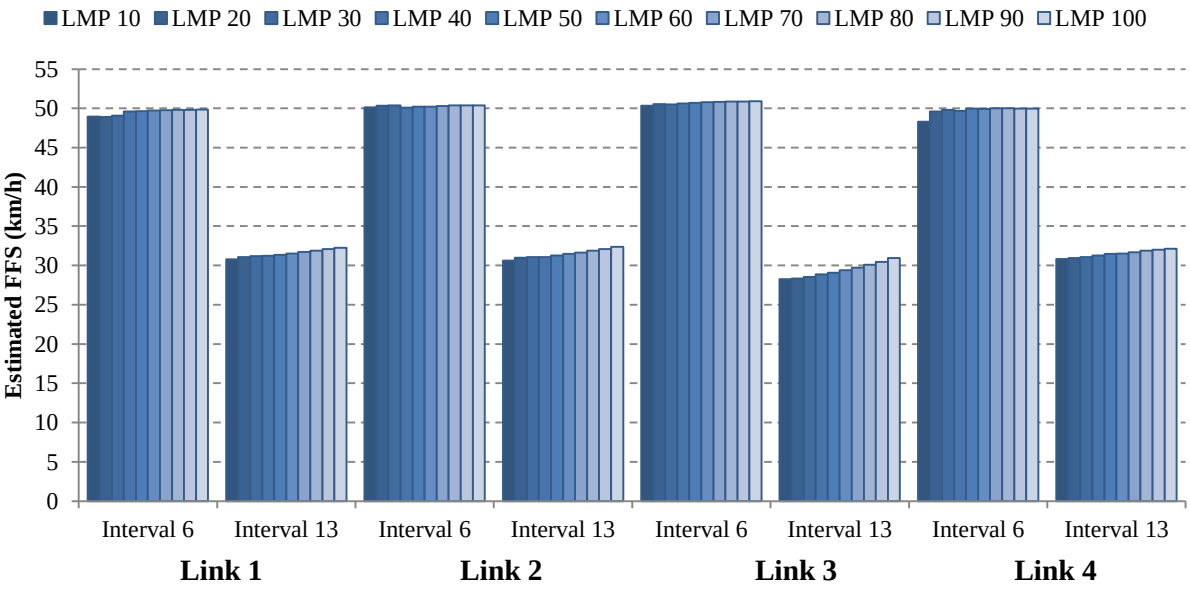


Figure 4-11: Estimated free flow speed at intervals 6 and 13 for different links and range of CV LMP.

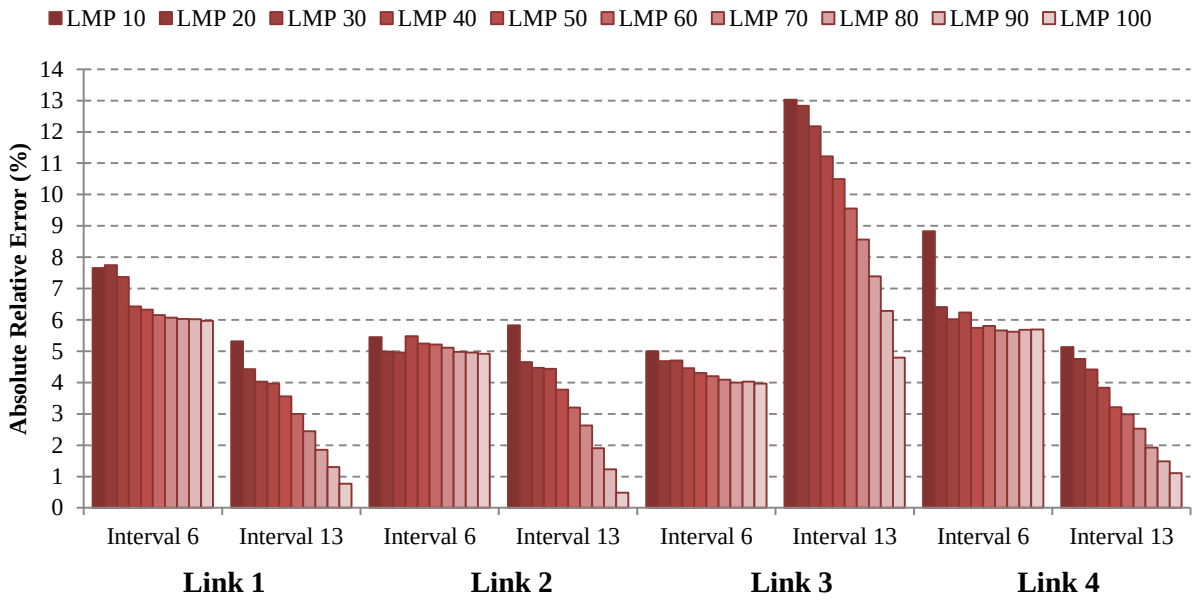


Figure 4-12: Absolute relative error between the estimated and ground truth FFS at intervals 6 and 13 for different links and range of CV LMP.

$$ARE_i^{LMP} = \left| \frac{FFS_i^{estimated} - FFS_i^{ground\_truth}}{FFS_i^{ground\_truth}} \right| \quad (4\text{-}11)$$

Where,

ARE_i^{LMP} = Absolute Relative Error of the estimated free flow speed at interval *i;*

$FFS_i^{estimated}$ = Smoothed free flow speed estimated at interval *i, (km/h);*

$FFS_i^{ground\_truth}$ = Ground truth free flow speed at interval *i, (km/h).*

There is a challenge with respect to assessing whether or not the proposed method provides an acceptable level of accuracy. First, there are no established performance benchmarks. The second challenge is that there are no other works in the literature which attempt to estimate the FFS on arterials on the basis of connected vehicle data and therefore a direct comparison of the performance of the proposed method with another existing method is not possible. However, Cheu et al. (2002) examined the problem of estimating average arterial link speeds from probe vehicles. They were interested in determining the minimum probe vehicle sample size necessary to estimate the average arterial link speed with a maximum absolute error of 5 km/h at least 95% of the times. Though their goal was different than ours, their work does provide one means of determining an

acceptable level of accuracy, and therefore we define absolute errors less than 5 km/h as acceptable. Examining the results presented in Figure 4-11, the estimations made at intervals 6 and 13 for all range of CV LMPs have errors less than this accuracy threshold and therefore, we can assume the results of the algorithm are within the acceptable range of required accuracy.

The results presented in Figure 4-12 show that during snowy condition, estimations made for free flow speed on link 3 have larger absolute relative errors in comparison to other links. Examining the simulations during the snowy condition, it was observed that shockwaves and queues form on link 3 during snowy condition due to unsuitable signal offsets which are optimized for the dry condition; and therefore resulting in vehicles having a much smaller likelihood of achieving the free flow speed on this link.

Figure 4-13 shows the smoothed estimated mean free flow speeds aggregated across all 4 study links over the simulation at different CV LMPs. This figure shows that the proposed methodology estimates almost constant FFS during the dry condition and adapts to the change in the free flow speed caused by the change in the weather condition. Furthermore the model appears to provide relatively consistent estimates regardless of the LMP.

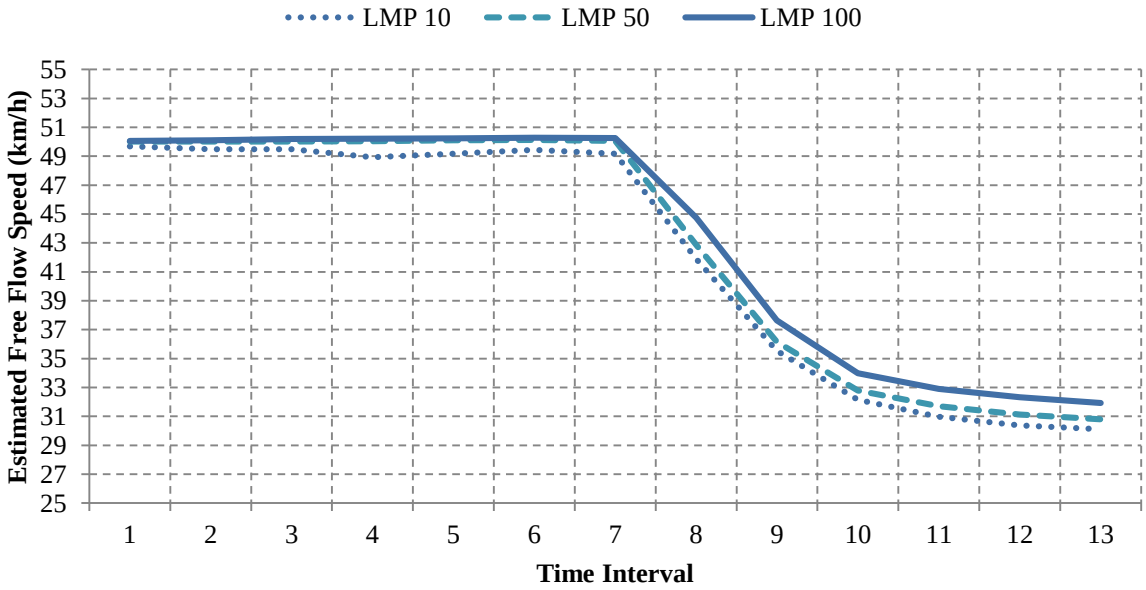


Figure 4-13: Estimated smoothed FFS produced by the proposed methodology as a function of LMP

4.4 Conclusions

This chapter presented a methodology to estimate arterial free flow speed using data from connected vehicles. The goal of the proposed methodology is to dynamically estimate the free flow speed and to provide inputs to real-time signal timing control systems such as adaptive or responsive traffic signal control methods. Such systems currently assume a predetermined free flow speed (or cruise speed) for different links on a network resulting sub-optimal signal offsets when changes in the free flow speed occur as a result of inclement weather conditions (or other factors).

The proposed methodology was evaluated on a hypothetical corridor. The results show that the methodology can provide FFS estimations close to the known ground truth and can respond to changes in the FFS when simulated snowy conditions occur during the simulation. The results also show that the maximum absolute error of about 4.7 km/h in the estimated FFS was observed and that the highest absolute relative errors occurred on a highly congested link where vehicles are less likely to achieve their free flow speed.

5

Evaluation of Signal Timing Optimization using Connected Vehicle Data

The previous chapters presented methodologies for real-time estimation of saturation flow rate and free flow speed on signalized arterials using data from connected vehicles. The role of these two parameters and their influence on determining the traffic signal timing (split, cycle, and offset) were highlighted. The proposed methodologies are envisioned to be used in the existing adaptive traffic signal control systems to adjust traffic signal timings in real-time according to changes in the traffic condition using connected vehicle data as the only source. However, it is not clear how much improved performance, if any, a signalized corridor can receive from using such a system. Therefore, the goal of this chapter is to evaluate a real-time signal timing optimization system using dynamically estimated saturation flow rate and free flow speed obtained from the methodologies proposed in the previous chapters.

The next section presents the evaluation methodology used in this chapter, followed by sections presenting simulation setup, evaluation, results, and conclusions.

5.1 Evaluation Methodology

This section describes the proposed adaptive signal control system platform used for the purpose of the evaluations required in this chapter. Figure 5-1 presents the overall system architecture. VISSIM (PTV, 2012) is used as a microsimulation platform to simulate a signalized corridor and connected vehicles. Connected vehicle data including position, speed, acceleration rate, and the status of traffic signal controllers are obtained from VISSIM at the rate of 10 times per second. The methodologies proposed and described in Chapters 2, 3, and 4 are then used to estimate the stop line saturation flow rates and free flow speeds using the real-time data obtained from the VISSIM model.

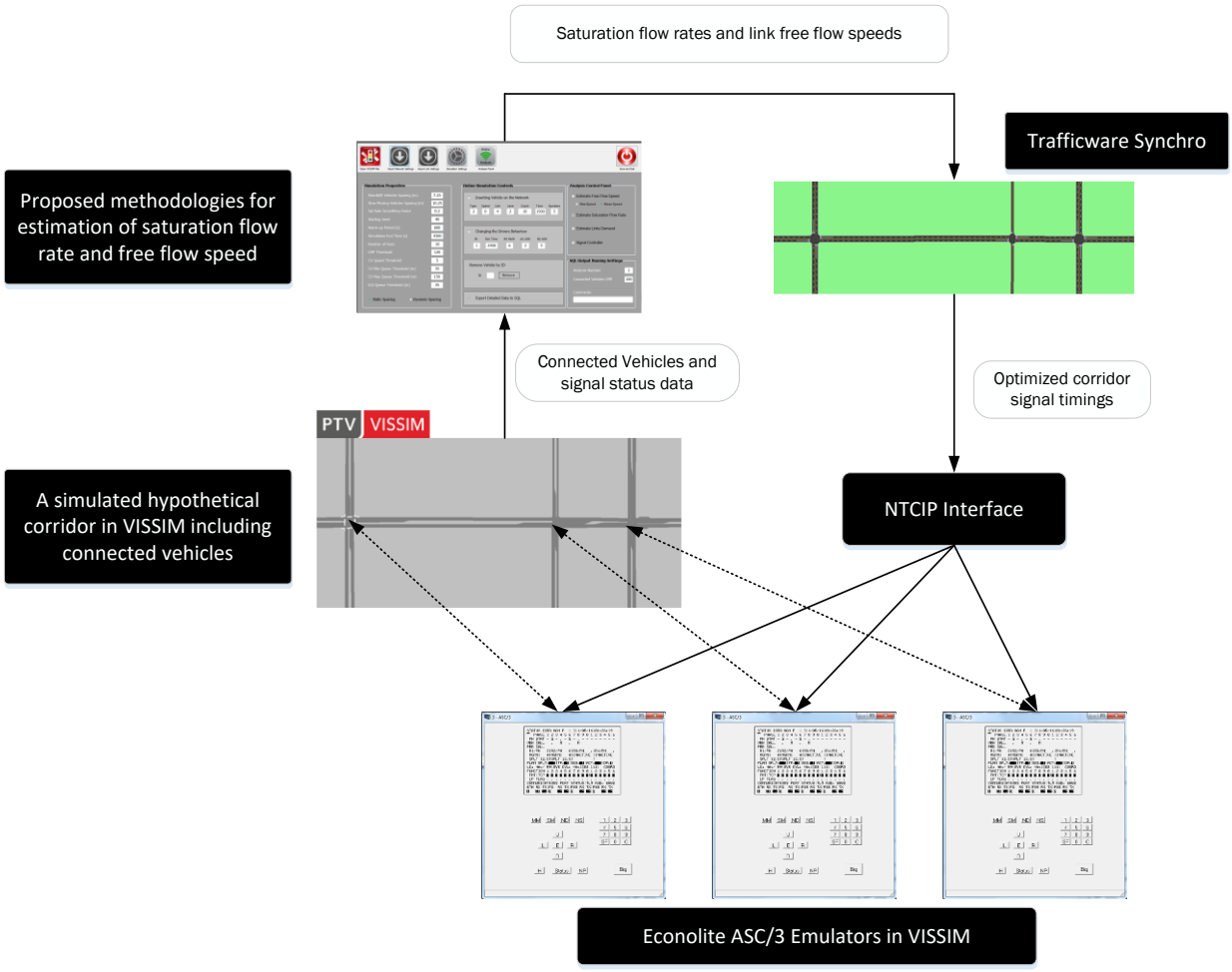


Figure 5-1: Overview of developed adaptive signal control system.

At fixed time intervals (*T*), the most recent estimated saturation flow rates and free flow speeds obtained from the previous step are used as inputs to a model of the network in Synchro (Trafficware, 2006). Synchro is a commercially available software used to develop optimal fixed time (e.g. time of day) traffic signal timing plans (splits, cycle, offsets) for arterial corridors.

Once signals are optimized in Synchro, the new timings are transferred to the signal controllers in VISSIM. Traffic signals in VISSIM are controlled using emulated Econolite traffic signal controllers (Econolite, 2016). This feature in VISSIM provides access to fully functional traffic signal controllers with communication capabilities based on NTCIP 1202 (The National Transportation Communications for ITS Protocol) standard (NTCIP 1202, 2005). NTCIP provides a standard communication protocol for communicating with traffic signal controllers, and is used

in this evaluation to download new optimized signal timings to the controllers during the simulation.

In practice, most real-time traffic adaptive signal control systems restrict the magnitude of changes that can be made to the timings at a given intersection in each optimization interval. For example, they may restrict changes in the cycle length to not more than 10% of the existing cycle length. These constraints help to reduce the impact of the transition period and by reducing the rate at which the signal timings can change, it is also possible to reduce the optimization period interval duration and optimize signal timings more frequently. The development of a new adaptive optimization algorithm was beyond the scope of this thesis, and the adaptive system used for this evaluation does not have a mechanism to control the magnitude of changes to the existing timings in the controllers.

When new signal timings are transferred to the signal controllers in VISSIM, the controller uses a transition algorithm to transition from the existing signal timings to the new signal timings. The transition algorithm ensures that all timing constraints (e.g. minimum green times, intergreen intervals, etc) are satisfied when transitioning to the new timing plan. Given that implementing new signal timings can cause disruption to signal co-ordination, time interval *T* was chosen to be 10 minutes to reduce the impact of the transition period while still providing a relatively quick rate of response to changing traffic conditions.

5.2 Simulation Setup

The evaluation was performed on a hypothetical corridor modeled in VISSIM microsimulation software as shown in Figure 5-2. The corridor consists of three coordinated signalized intersections that were operated in fixed-time modes (not in actuated mode) at the local level; however their signal timings were updated every 10 minutes using the procedure explained earlier in this chapter. Temporally constant traffic demands, as shown on the figure, were simulated for the whole duration of each simulation run. For the purposes of demonstrating impacts, no right or left turn traffic were allowed at the intersections (only through traffic was modeled). Although, no pedestrian traffic was modeled, a total walk and clearance time of 16 seconds was used and honoured at all intersections. The amber and all-red intervals for all approaches at all intersections were set to 3 and 2 seconds respectively.

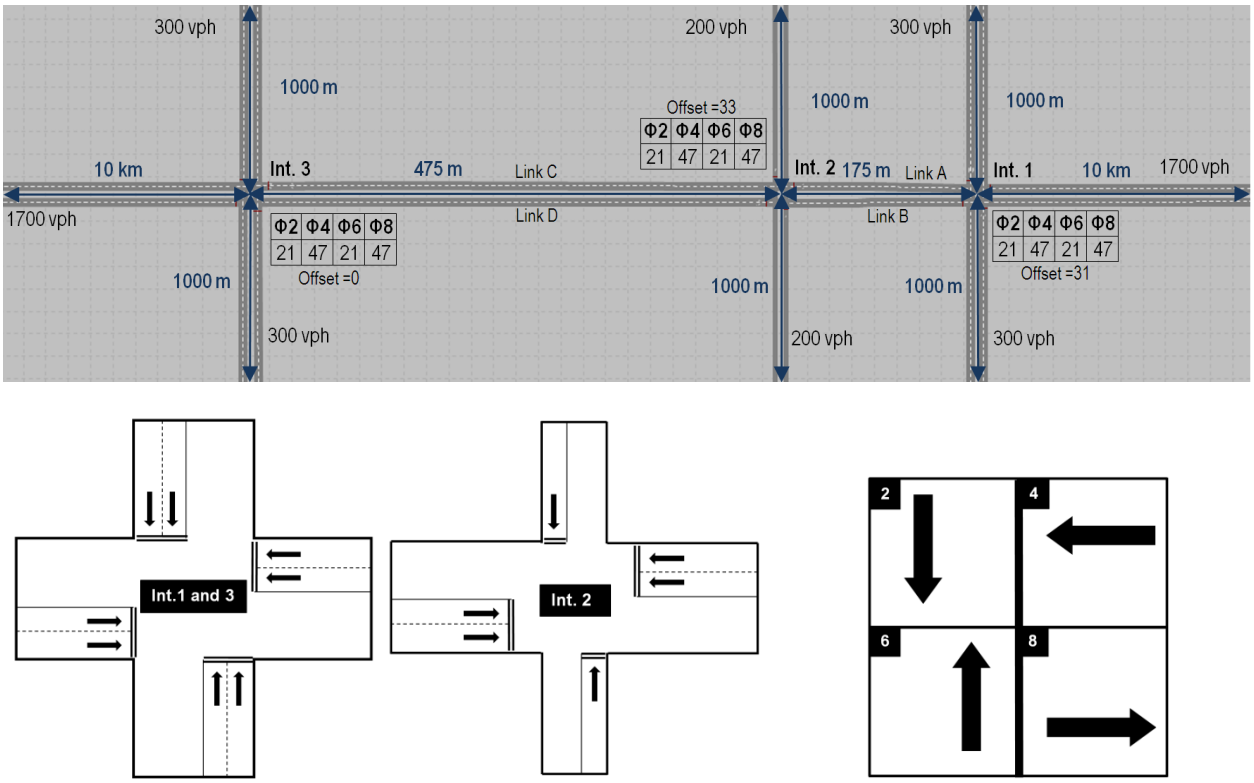


Figure 5-2: Study corridor

Note that the WB approach link to Intersection 3 and the EB approach link to Intersection 1 had a length of 10km. This long length was used to ensure that any queues that might form at the signalized intersections would not spill back and exceed the length of the approach link. When queues extend beyond the length of the approach link, vehicles are unable to enter the network and the delays that they experience are not captured in the VISSIM model outputs.

Connected vehicles were modeled in VISSIM as a separate class of vehicles. Different levels of market penetration of connected vehicles (CV LMP) were possible to model by changing the proportion of the connected vehicles class in the simulation model.

The evaluation was performed by comparing network vehicular average delay and total stops for the following scenarios:

1- Using fixed-time signal timings (non-adaptive) at all intersections during the simulation runs;
2- Using the adaptive signal control system at 100% market penetration rate of connected vehicles;

3- Using the adaptive signal control system when market penetration rate of connected vehicles is 20%.

Five simulations with different random seeds were performed for each of the above scenarios. The duration of each simulation run was 8000 seconds with 1000 seconds used as the warm-up period. Dry road surface condition was simulated for the first 3000 seconds of a simulation using VISSIM's default settings; and a snow storm was modeled for the rest of the simulation by modifying drivers' behaviour and vehicles characteristics as presented in Chapter 4, section 4.3.1.

Saturation flow rates and free flow speed obtained from the simulation for the dry road surface condition were used in Synchro to optimize the traffic signals on the network. These optimized timings, as presented in Figure 5-2, were used for the evaluation of Scenario 1 and were used as the initial signal timings for Scenarios 2 and 3 (in these scenarios the adaptive signal control system was used so that signal timings were updated every 10 minutes during each simulation run when new estimated saturation flow rates and free flow speeds became available).

5.3 Results

This section presents the evaluation results for the three scenarios defined above. The performance of the corridor under these scenarios was measured using average delay per vehicle and average number of stops per vehicle as the two most commonly used measures of performance in arterial performance evaluation.

5.3.1 Scenario 1: Non-Adaptive Control

We begin the discussion of the results by examining the traffic conditions that result from the use of the non-adaptive traffic signal control strategy. In this strategy, the signal timing plans have been optimized for the existing traffic demands and dry-clear weather conditions.

Figure 5-3 shows the average delay per vehicle as a function of simulation time using 5 simulation runs. During the first 2000 seconds of the simulation data collection period (i.e. simulation time 1000 to 3000), the actual conditions within the simulation are those for which the signal timings have been optimized and as expected the timings perform well. The signalized intersections operate in an under-saturated mode (e.g. demand is less than the capacity) and as a result, vehicles experience relatively little delay (average delay is approximately 15 seconds/vehicle) with average LOS = B.

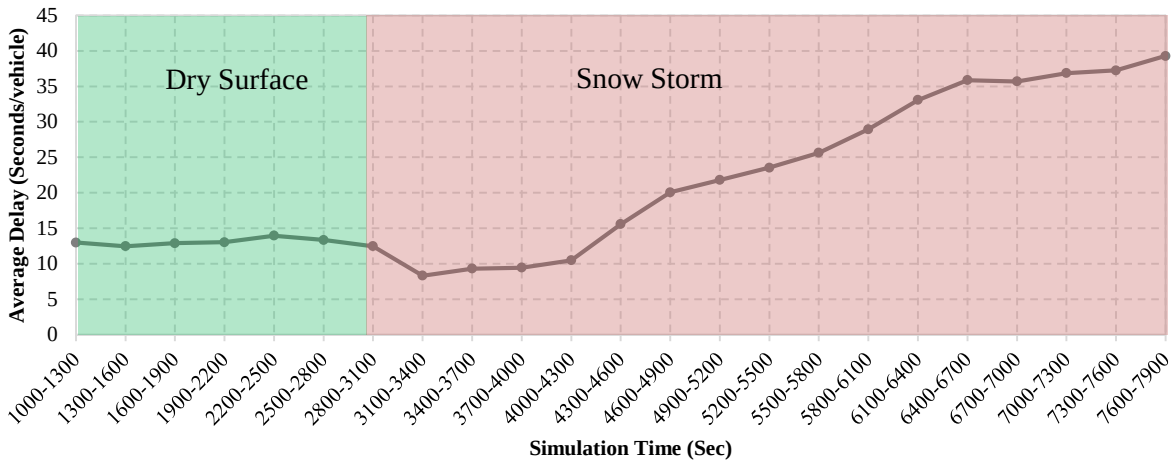


Figure 5-3: Average vehicle delay as a function of simulation time for non-adaptive control

At simulation time 3000, the VISSIM simulation parameters are altered to simulate the effects of a snow storm. These effects are to: (i) alter the speed flow relationship (i.e. reduce the free speed) at which vehicles travel along the roadway; and (ii) increase vehicle headways which reduces the saturation flow rate.

The speed reduction during the snow storm is modelled in VISSIM by assigning a new distribution of desired speed (30 to 35 km/h) to all the existing and new vehicles in the network. Although the impact of the reduced speeds during a snow storm can be modelled using this approach, lowering the desired speeds can result in VISSIM reporting a decrease in delay. In VISSIM, delay for individual vehicles is calculated by subtracting the travel time of a vehicle at ideal (desired) speed from the actual travel time of the vehicle. For example, consider a link that is 0.5 km in length. If the desired speed is 50 km/h then the associated desired travel time is 36 seconds. If the actual speed is 35 km/h, then the actual travel time is 51 seconds. Then the delay = 51 – 36 = 15 seconds. However, if the desired speed is reduced from 50 to 35 km/h (for example because we are modelling a snow storm event), then the actual travel time remains at 51 seconds and the desired travel time also becomes 51 seconds and the delay is reduced to zero. Normally, we don't expect such as large reduction in the delay because the effect of modelling the snow conditions also reduces capacity and therefore increases the travel time (and delay) that vehicles experience. Nevertheless, as can be seen in Figure 5-3, following the start of the snow storm, a reduction in delay is observed due to the lower desired speed assigned to the vehicle. However, since the goal of this evaluation is to compare the relative performance of the three scenarios, these reductions

in the delay do not impact the evaluation as all three scenarios are impacted by this reduction with the same magnitude. Furthermore, we observe that after a relatively short period of time, the delay increases because the decreased capacity that results from the snow storm, cause some approaches to become oversaturated.

5.3.2 Scenario 2: Adaptive Control

We now consider the performance of the proposed adaptive control strategy (Scenarios 2 and 3). When presenting the performance measure for the adaptive control strategy, we also provide the results from the non-adaptive strategy (i.e. those presented in the previous section) as the benchmark against which the adaptive control performance can be compared.

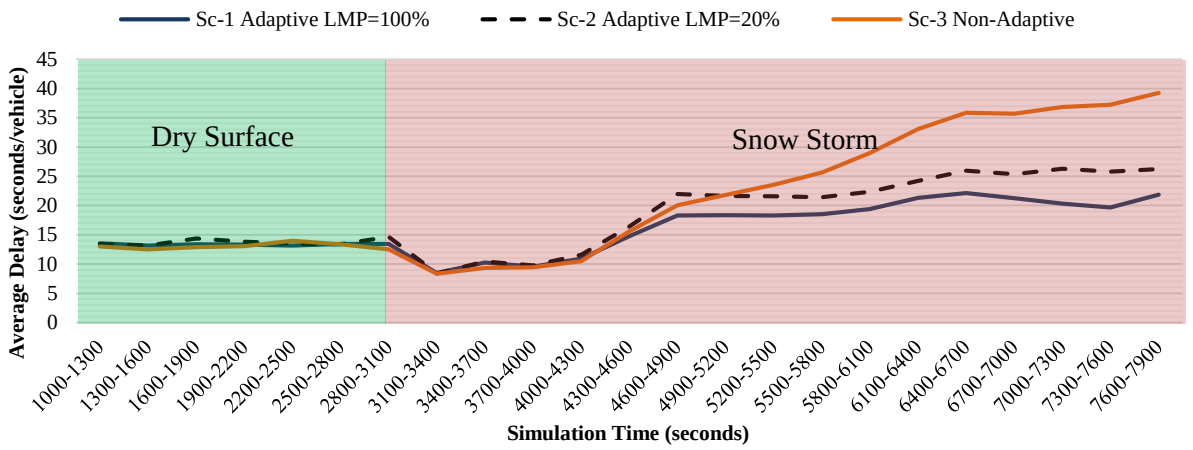
Figure 5-4 (a) and (b) present the average delay per vehicle and average number of stops per vehicle in the network for all three scenarios. We saw in the previous section, that under non-adaptive control, the network became over-saturated during the snow storm and average vehicle delays continued to increase over time. These results show that the adaptive control performs better in comparison to the non-adaptive scenario by adjusting the signal timings to the new condition (snow storm) and preventing oversaturated conditions. The results also show that the average number of stops/vehicle decreases as a result of using the adaptive system.

The comparison of the improvements obtained from using the adaptive system under 100% and 20% CV LMPs shows that the system performs better when 100% vehicles are connected vehicles; however, substantial performance improvements are achieved in the network even at 20% CV LMP. The results show that at 100% CV LMP, the average delay calculated at the end of simulation runs is reduced by approximately 45% (17.2 seconds) and the average number of stops is reduced by 20% (0.048 stops/vehicle). The corresponding improvements for 20% CV LMP are a reduction in delay of 31% (11.7 seconds/vehicle) and reductions in stops of 15.7% (0.038 stops/vehicle).

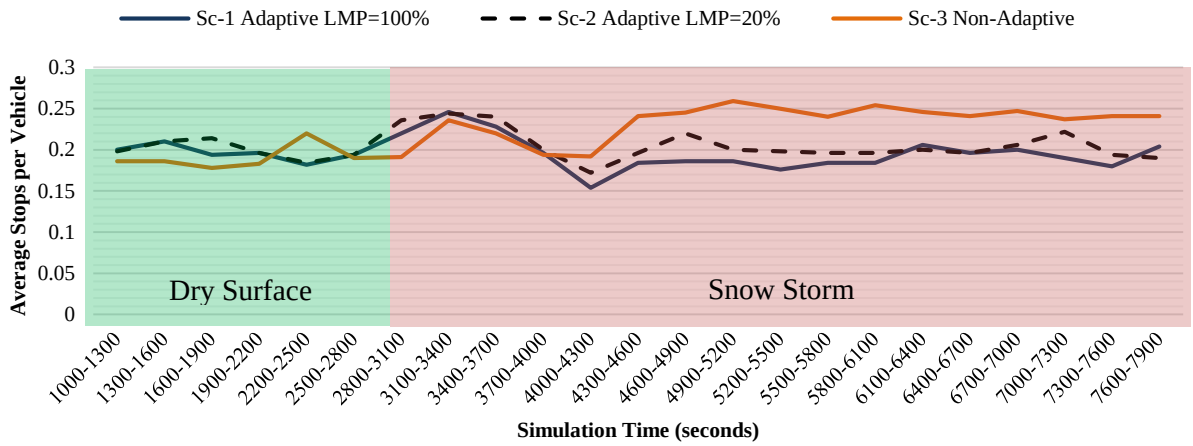
Figure 5-5 and Figure 5-6 present the variation in the signal timing for coordinated movements (phase 4 and 8) at all intersections under the adaptive signal control scenarios in response to changes in the saturation flow rate and the free flow speed at 100% and 20% CV LMPs respectively. The figures show that when the saturation flow rate and free flow speed are reduced due to the snow storm, the adaptive signal control system increases the time allocated to the major

movements. This increase in the signal timing is mainly in response to the loss of capacity caused by reduction in the saturation flow rate.

The offsets at intersections 1 and 2 are also increased and optimized as result of the reduction in vehicles speeds. This correction in offset is in response to the increased travel time between intersections. However, some fluctuations in the offset at intersection 1 can be observed when the system is operating using 20% LMP of connected vehicles.



(a)



(b)

Figure 5-4: (a) Average vehicle delay, and (b) total number of stops under different scenarios

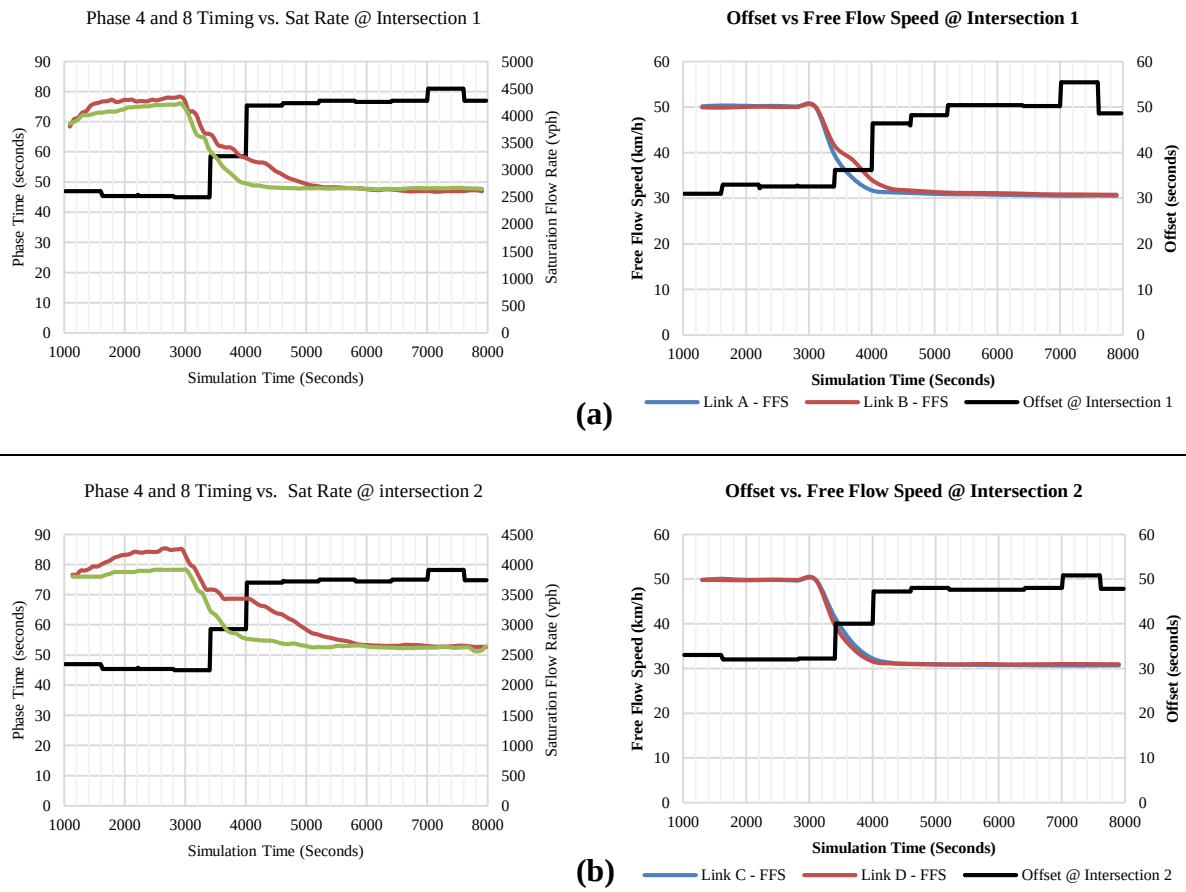
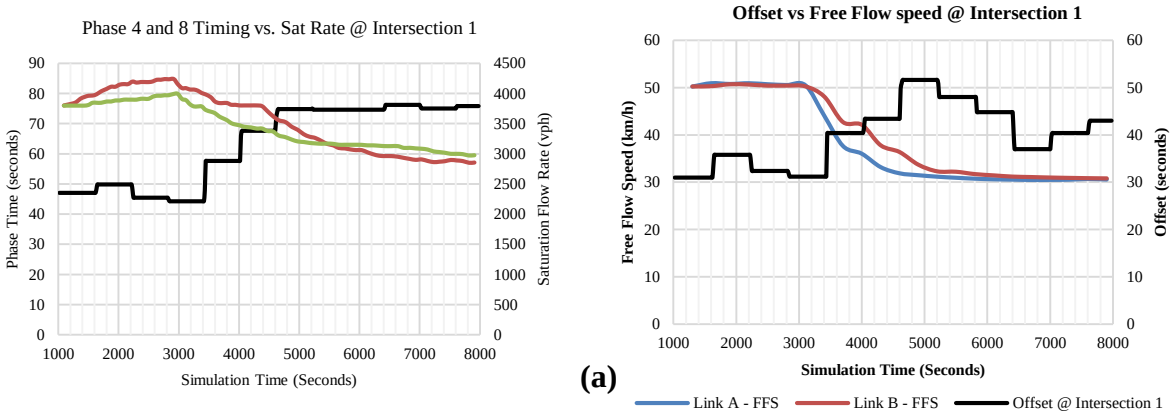
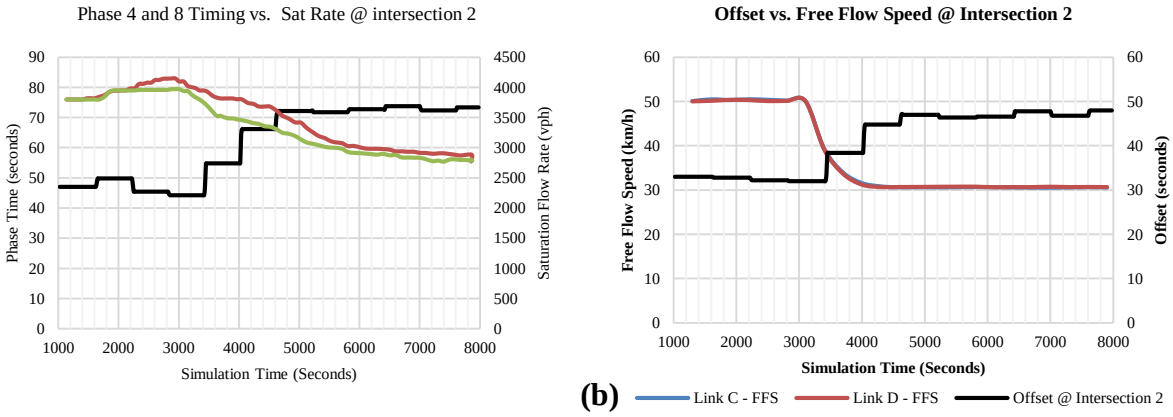


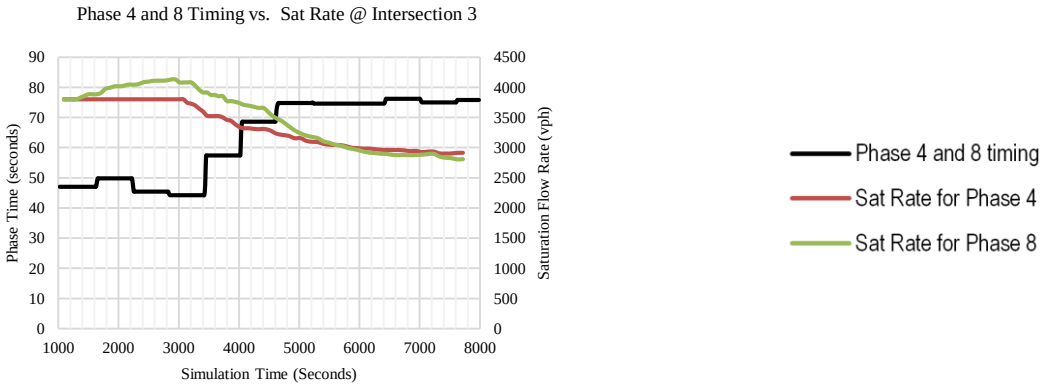
Figure 5-5: Adaptive signal system performance at 100% CV LMP.



(a)



(b)



(c)

Figure 5-6: Adaptive signal system performance at 20% CV LMP.

5.4 Conclusions

This chapter presented a demonstration of using the methodologies proposed in the previous chapters for estimating the signalized intersections saturation flow rate and the arterial free flow speed using data from connected vehicles. The goal was to show and evaluate the performance of an adaptive signal control system that uses connected vehicle data as the only data source and does not rely on data from conventional traffic sensors.

The adaptive system was evaluated on a hypothetical corridor and performance was compared to using non-adaptive (fixed time) signal timings. From the evaluation results the following conclusions can be made:

1- The adaptive signal control is able to adjust intersection signal timings in response to changes in the saturation flow rate and free flow speed estimated from connected vehicle data.
2- The adaptive signal control system based on the connected vehicle data can automatically respond to changes in the weather/road surface conditions and as demonstrated, provides improved performance along the signalized corridor as compared to a fixed time signal control when these conditions occur.
3- Both average vehicle delay and average number of stops are reduced using the adaptive signal control system in comparison to the fixed time control.
4- The adaptive signal control system improves the network performance at both 20% and 100% CV levels of market penetration and though improvements are larger for 100% CV LMP, approximately 70% of these improvements are achieved at 20% CV LMP.
5- The evaluation presented in this chapter shows that there is the potential to use the proposed methodologies for estimation of the saturation flow rate and the free flow speed on signalized arterials in the existing and conventional adaptive signal control systems. The proposed methodologies could help in improving the performance of existing adaptive signal control systems.

6

Conclusions

6.1 Conclusions

The connected vehicle technology is expected to revolutionize the road network transportation by providing real-time connectivity and exchange of information between vehicles, road infrastructures, and mobile devices. This technology will enable vehicles to continuously broadcast their speed, heading, position, acceleration, and other information at the rate of 10 times per seconds. It is expected that the connected vehicle technology will improve transportation safety, mobility, and reduce the environmental impacts of road transportation by providing an extensive set of data that will enable transportation professionals to better manage and control the traffic and road infrastructure.

One of the main challenges facing the deployment of different applications based on the connected vehicles data is the market penetration rate of this technology. Many applications of connected vehicles require 100% of vehicles to be connected vehicles, which likely will not be achieved for at least a decade after this technology is mandated on new vehicles.

There is the potential for some applications, particularly those focussed on transportation mobility rather than safety, to use the connected vehicle data during this transition period, when only a portion of the traffic stream consists of connected vehicles. Enhancing the operation of traffic signals is one of these applications. Existing adaptive signal control systems (e.g. SCOOT, SCATS) rely on traditional traffic sensors installed at intersections or sometimes upstream of signals to obtain the required inputs to their signal timing optimization models. While providing limited information on the traffic, these detectors account for a substantial portion of the installation and the maintenance costs of these systems. For instance, the cost to install SCATS per intersection is around $61,161; however, average cost per intersection to install inductive loops detection system is around $42,751 (Lodes & Benekohal, 2013).

Therefore, this thesis describes the development and evaluation of methods to use the connected vehicle data to provide some of the required inputs for the existing adaptive signal control systems with the goal to potentially enhance their operation by providing better information on traffic conditions (i.e. real-time saturation flow rate and free flow speeds) to these systems.

6.2 Contributions

The following describes the four most significant contributions from this research:

1) *Developed and evaluated a method to estimate the number of stopped vehicles behind the stop line at signalized intersection using only connected vehicle data.*

   The goal of the proposed methodologies is to estimate the number of queued vehicles behind the stop line at a signalized intersection as one of the inputs required for estimating the stop line saturation flow rate. As part of developing these methodologies, we proposed to check for continuity of a queue using a stepwise queue search and a search distance. We also proposed an enhancement to the above method by dynamically changing the radius that the algorithm searches for stopped connected vehicles based on a dynamically estimated market penetration rate of connected vehicles. Dynamically changing the search radius results in better performance and makes the method more robust as it can automatically adapt to changes in the level of market penetration of connected vehicles (including both short term changes such as those that occur with a day or between the days of the week, but also longer term changes such as those associated with gradual increases in the CV LMP). We also proposed using the vehicle length information provided by connected vehicles to estimate the average vehicle length on different lanes at an intersection and showed that this information improves the accuracy of queue estimation where traffic composition can change.

2) *Developed and evaluated a method to estimate market penetration rate of connected vehicles on signalized arterials.*

   This research demonstrated that the queue information data provided by the proposed queue estimation methodology can be used to estimate the proportion of traffic equipped with the connected vehicle technology. The results showed that estimated CV LMP is close to the ground truth when at least 20% of traffic is equipped with the connected vehicle technology.

3) *Developed and evaluated a method to estimate the real-time saturation flow rate at signalized intersections using only connected vehicle data.*

   The results of the evaluation of the developed methodology in this research showed that the data from connected vehicles can be used to dynamically estimate the saturation flow rate at signalized intersections in response to changes in this parameter causes by a range of factors such as changes in the road surface condition, traffic composition, lane blockage, etc. The significance of the developed methodology is that, while it does not rely on any traditional fixed point detectors, it has the ability to incorporate the estimated queue behind the stop line when estimating the saturation flow rate and also operate at low penetration rate of connected vehicles. The dynamically estimated saturation flow rate can be used to adjust the cycle length and splits at signalized intersections in response to changes in the capacity of the intersections.

4) *Developed and evaluated a method to estimate free flow speed on arterials using only CV data*

   Free flow speed on arterials is used to adjust signal offsets between consecutive intersections. Existing adaptive signal control systems, such as SCOOT for example, assume a constant free flow speed in their signal optimization models. However, assuming a constant free flow speed can result in sub-optimal signal timings when this parameter changes especially during inclement weather conditions. Therefore, in this research we proposed a methodology to estimate the free flow speed on arterials using connected vehicle data. The methodology relies only on information provided by the connected vehicles to infer when these vehicles travel at free flow speed. The results from the proposed methodology for estimating the free flow speed show the data provided by connected vehicles can be used to estimate changes in the free flow speed on arterials.

Finally, we also demonstrated that the proposed methodologies in this research based on only connected vehicle data can provide substantial improvements to the performance of a signalized arterial by providing the real-time information to the signal optimization model and ultimately adjusting the cycle length, splits, and offset of the traffic signals in response to changes in the traffic condition.

6.3 Directions for Future Research

Based on the methodologies and results presented in this research, the following topics are suggested for future research:

1. Proposed methodologies in this research were evaluated within a simulation environment and on hypothetical study areas. Therefore, it is recommended to evaluate these methodologies using real connected vehicle data once these data become available and in real world traffic and road geometry conditions.
2. This research proposed methodologies for estimating saturation flow rate, and free flow speed required for analytical signal timing optimization techniques. However, estimating traffic demand as one of the required inputs was not studied in this research. Therefore, it is recommended to investigate developing methodologies for estimating traffic demand on arterial roads using only connected vehicle data as the only data source.
3. Finally, it is recommended to investigate incorporating the output of the developed algorithm into existing adaptive signal controller systems and evaluate the impact that these inputs have on the performance of these systems.

References

Akçelik, R. (1988). The Highway Capacity Manual Delay Formula for Signalized Intersections. *ITE Journal, 58*(3), 23-27.

Akcelik, R. (1998). *Traffic Signals: Capacity and Timing Analysis.* Australian Road Research Board. Research Report ARR No. 123.

Akçelik, R. (1999). *A queue model for HCM 2000.* Vermont South, Australia: ARRB Transportation Research Ltd.

Argote, J., Christofa, E., Xuan, Y., & Skabardonis, A. (2011). Estimation of Measures of Effectiveness Based on Connected Vehicle Data. *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, (pp. 1767-1772). Washington, DC.

Asamer, J., van Zulen, H., & Heilmann, B. (2013). Calibrating car-following parameters for snowy road conditions in the microscopic traffic simulator VISSIM. *The Institution of Engineering and Technology - Intelligent Transport Systems, 7*(1), 114-121.

Astarita, V., Bertini, R., d'Elia, S., & Guido, G. (2006). Motorway traffic parameter estimation from mobile phone counts. *European Journal of Operational Research, 175*(3), 1435-1446.

ASTM E2213-03. (2010). *Standard Specification for Telecommunications and Information Exchange Between Roadside and Vehicle Systems — 5 GHz Band Dedicated Short Range Communications (DSRC) Medium Access Control (MAC) and Physical Layer (PHY) Specifications.* West Conshohocken, PA: ASTM International.

Athol, P. (1965). Interdependence of certain operational characteristics within a moving traffic stream. *Highway Research Record 72, Transportation Research Board*, 58-87.

Badillo, B., Rakha, H., Rioux, T., & Abrams, M. (2012). Queue length estimation using conventional vehicle detector and probe vehicle data. *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference*, 1674 - 1681. doi:10.1109/ITSC.2012.6338891

Ban, X. (., Hao, P., & Sun, Z. (2011). Real time queue length estimation for signalized intersections using travel times from mobile sensors. *Transportation Research Part C: Emerging Technologies, 19*(6), 1133-1156. doi:10.1016/j.trc.2011.01.002

Cai, Q., Wang , Z., Zheng , L., Wu , B., & Wang, Y. (2014). Shock Wave Approach for Estimating Queue Length at Signalized Intersections by Fusing Data from Point and Mobile Sensors. *Transportation Research Record: Journal of the Transportation Research Board, 2422*, 79-87. doi:10.3141/2422-09

Cetin, M. (2012). Estimating Queue Dynamics at Signalized Intersections from Probe Vehicle Data. *Transportation Research Record: Journal of the Transportation Research Board, 2315*, 164-172. doi:10.3141/2315-17

Chauvenet, W. (1908). *A manual of spherical and practical astronomy: embracing the general problems of spherical astronomy, the special applications to nautical astronomy, and the theory and use of fixed and portable astronomical instruments, with an appendix on the method of* (Vol. 2). JB Lippincott.

Cheu, L., Chi Xie, R., & Lee, D.-H. (2002). Probe vehicle population and sample size for arterial speed estimation. *Computer-Aided Civil and Infrastructure Engineering, 17*(1), 53-60.

Chi, X., Cheu, R., & Lee, D.-H. (2001). Calibration-free arterial link speed estimation model using loop data. *ournal of Transportation Engineering, 127*(6), 507-514.

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. doi:10.1016/j.ejor.2008.06.024

de Fabritiis, C., Ragona, R., & Valenti, G. (2008). Traffic Estimation And Prediction Based On Real Time Floating Car Data. *2008 11th International IEEE Conference on Intelligent Transportation Systems* (pp. 197-203). Bejing: IEEE.

Dunn Engineering Associates. (2005). *Traffic Control Systems Handbook.* Washington, D.C.: Federal Highway Administration.

Econolite. (2016). *ASC/3-2070 SOFTWARE.* Retrieved from Econolite: http://www.econolite.com/

Edwards, J. B. (1999). Speed adjustment of motorway commuter traffic to inclement weather. *Transportation Research Part F: Traffic Psychology and Behaviour, 2*(1), 1-14. doi:10.1016/S1369-8478(99)00003-0

ETSI. (2009). *ETSI TR 102 638 V1.1.1 - Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Definitions.* European Telecommunications Standards Institute.

ETSI. (2012). *ETSI EN 302 665 V1.2.0 - Intelligent Transport Systems (ITS); Communications Architecture.* European Telecommunications Standards Institute.

ETSI. (2012). *ETSI TS 101 556-1 V1.1.1 - Intelligent Transport Systems (ITS); Infrastructure to Vehicle Communication; Electric Vehicle Charging Spot Notification Specification.* European Telecommunications Standards Institute.

ETSI. (2012). *ETSI TS 102 724 V1.1.1 - Intelligent Transport Systems (ITS); Harmonized Channel Specifications for Intelligent Transport Systems operating in the 5 GHz frequency band.* European Telecommunications Standards Institute.

ETSI. (2013). *ETSI EN 302 663 V1.2.1 - Intelligent Transport Systems (ITS); Access layer specification for Intelligent Transport Systems operating in the 5 GHz frequency band.* European Telecommunications Standards Institute.

ETSI. (2014). *ETSI EN 302 636-5-1 V1.2.1 - Intelligent Transport Systems (ITS); Vehicular Communications; GeoNetworking; Part 5: Transport Protocols; Sub-part 1: Basic Transport Protocol.* European Telecommunications Standards Institute.

ETSI. (2014). *ETSI EN 302 637-2 V1.3.2 - Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service.* European Telecommunications Standards Institute.

ETSI. (2014). *ETSI EN 302 637-3 V1.2.2 - Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 3: Specifications of Decentralized Environmental Notification Basic Service.* European Telecommunications Standards Institute.

ETSI. (2014). *ETSI TS 101 556-3 V1.1.1 - Intelligent Transport Systems (ITS); Infrastructure to Vehicle Communications; Part 3: Communications system for the planning and reservation of EV energy supply using wireless networks.* European Telecommunications Standards Institute.

ETSI. (2014). *ETSI TS 102 894-2 V1.2.1 - Intelligent Transport Systems (ITS); Users and applications requirements; Part 2: Applications and facilities layer common data dictionary.* European Telecommunications Standard Institute.

ETSI. (2016). *ETSI EN 302 571 V2.0.0 - Intelligent Transport Systems (ITS); Radiocommunications equipment operating in the 5 855 MHz to 5 925 MHz frequency band; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU.* European Telecommunications Standards Institute.

Fazio, J., Wiesner, B., & Deardoff, M. (2014). Estimation of free-flow speed. *KSCE Journal of Civil Engineering, 18.2*, 646-650.

FCC 47 Part 95. (2009). *Personal Redio Services.* Federal Communications Commission.

Garber, N., & Hoel, L. (1988). *Traffic and Highway Engineering.* St. Paul: West Publishing Company.

Gillam, W. J., & Wilhill, R. A. (1992). UTC and inclement weather conditions. *Road Traffic Monitoring, 1992 (IEE Conf. Pub. 355)*, (pp. 85-88).

Goodall, N. J., Park, B. (., & Smith, B. (2014). Microscopic Estimation of Arterial Vehicle Positions in a Low-Penetration-Rate Connected Vehicle Environment. *Journal of Transportation Engineering, 140*(10), 04014047-1-9. doi:10.1061/(ASCE)TE.1943-5436.0000716

Goodall, N., Smith, B. L., & Park, B. (2013). Traffic Signal Control with Connected Vehicles. *In TRB 92nd Annual Meeting*, 15p.

Gordon, R., & Tighe, W. (2005). *Traffic Control System Handbook.* Federal Highway Administration.

Hao , P., Ban , X., & Yu , J. (2015). Kinematic Equation-Based Vehicle Queue Location Estimation Method for Signalized Intersections Using Mobile Sensor Data. *Journal of Intelligent Transportation Systems, 19*(3), 256-272. doi:10.1080/15472450.2013.857197

HCM. (2000). *Highway Capacity Manual.* Washington, D.C.: Transportation Research Board.

HCM. (2010). *Highway Capacity Manual.* Washington, D.C.: Transportation Research Board.

Henderson, I., & Wood, K. (2005). *Computerised real-time saturation flow measurement for signalised traffic junctions.* London: Transport Research Laboratory.

Herring, R., Hofleitner, A., Abbeel, P., & Bayen, A. (2010). Estimating arterial traffic conditions using sparse probe data. *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on* (pp. 929-936). Funchal, Portugal: IEEE.

How SCOOT Works. (2008). Retrieved 11 4, 2012, from SCOOT: http://www.scoot-utc.com/DetailedHowSCOOTWorks.php?menu=Technical

Hunt, P., Robertson, D., Bretherton , R., & Winton, R. (1981). *SCOOT - a traffic responsive method of coordinating signals.* Crowthrone: Transport and Road Research Labratory.

Ibrahim, A., & Hall, F. (1994). Effect of adverse weather conditions on speed-flow-occupancy relationships. *Transportation research record: Journal of the transportation research board, 1457*, 184-191.

IEEE 1609.0. (2013). *IEEE Guide for Wireless Access in Vehicular Environments (WAVE) - Architecture.* IEEE Standards Association.

IEEE 1609.3. (2016). *IEEE Standard for Wireless Access in Vehicular Environments (WAVE) - Networking Services.* The Institute of Electrical and Electronics Engineers.

IEEE 802.2 . (1998). *Standard for Information technology — Telecommunications and information exchange between systems—Local and metropolitan area networks — Specific requirements — Part 2: Logical Link Control.* IEEE Standards Association.

IEEE Std 802.11. (2012). *IEEE Standard for Information technology--Telecommunications and information exchange between systems Local and metropolitan area networks--Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications.* IEEE Standards Association.

Islam, M. K. (2013). *Real-Time Queue Length Estimation Using Probe Vehicles with Cumulative Input-Output Technique.* (Master of Science Thesis). Department of Civil Engineering. University of Calgary.

ISO 21210. (2012). *Intelligent transport systems -- Communications access for land mobiles (CALM) -- IPv6 Networking.* International Organization for Standardization.

ISO 21217. (2014). *Intelligent transport systems -- Communications access for land mobiles (CALM) -- Architecture.* International Organization for Standardization.

ISO TS 19321. (Under development). *ntelligent transport systems -- Cooperative ITS -- Dictionary of in-vehicle information (IVI) data structures.* International Standard Organization.

Jhaveri, C. S. (2003). SCOOT Adaptive Signal Control: An Evaluation of its Effectiveness over a Range of Congestion Intensities. *TRB 82nd Annual Meeting Compendium of Papers.* Washington, D.C.: Transportation Research Board of the National Academies.

John A. Volpe . (2008). *Vehicle-Infrastructure Integration (VII) Initiative Benefit-Cost Analysis Version 2.3 (Draft).* Washington, DC: U.S. Department of Transportation - National Transportation Systems Center.

Johnson, J., Thai, J., Tarico, D., & Chan, P. (2016, June 12). *Updating NTCIP 1202 (Actuated Signal Controllers) to Support a Connected Vehicle Environment.* Retrieved from ITS America 2016: https://itsamerica2016.org/wp-content/uploads/2016/08/Connected-Autonomous-Technology-Making-it-Work.Patrick-Chan.pdf

Kimber, R., & Hollis, E. M. (1979). *Traffic queues and delays at road junctions.* Laboratory Report 909. Crowthorne, UK.: Transport and Road Research Labratory.

Kyte, M., Khatib, Z., Shannon, P., & Kicthener, F. (2001). Effect of weather on free-flow speed. *Transportation research record: Journal of the transportation research board, 1776*, 60-68.

Li, J.-Q., Zhou , K., Shladover , S., & Skabardonis, A. (2013). Estimating Queue Length Under Connected Vehicle Technology. *Transportation Research Record: Journal of the Transportation Research Board, 2356*, 17-22. doi:10.3141/2356-03

Lighthill, M. J., & Whitham, G. B. (1955). On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. *the Royal Society of London A: Mathematical, Physical and Engineering Sciences. 229*, pp. 317-345. London: The Royal Society.

Liu, H., Wu, X., Ma, W., & Hu, H. (2009). Real-time queue length estimation for congested signalized intersections. *Transportation Research Part C: Emerging Technologies, 17*(4), 412-427. doi:http://dx.doi.org/10.1016/j.trc.2009.02.003

Lodes, M., & Benekohal, R. (2013). *Safety Benefits of Implementing Adaptive Signal Control Technology: Survey Results.* Springfield: Illinios Center for Transportation.

Lowrie, P. R. (1982). The Sydney coordinated adaptive traffic system - principles, methodology, algorithms. *International Conference on Road Traffic Signalling* (pp. 67-70). London, United Kingdom: IEE.

Maki, P. (1999). Adverse Weather Traffic Signal Timing. *Transportation Frontiers for the Next Millennium: 69th Annual Meeting of the Institute of Transportation Engineers.*

Michalopoulos, P. G., Stephanopoulos, G., & Stephanopoulos, G. (1977). Oversaturated signal systems with queue length constraints—I: Single intersection. *Transportation Research, 11*(6), 413-421.

Michalopoulos, P., Stephanopoulos, G., & Stephanopoulos, G. (1981). An application of shock wave theory to traffic signal control. *Transportation Research Part B: Methodological, 15*(1), 35-51. doi:http://dx.doi.org/10.1016/0191-2615(81)90045-X

Microsoft. (n.d.). *he OSI Model's Seven Layers Defined and Functions Explained*. Retrieved from support.microsoft.com: https://support.microsoft.com/en-us/kb/103884

Microsoft TechNet. (n.d.). *TCP/IP Protocol Architecture*. Retrieved from technet.microsoft.com: https://technet.microsoft.com/en-us/library/cc958821.aspx?f=255&MSPPError=-2147217396

Miller, A. J. (1968). *Australian road capacity guide: Provisional introduction and signalized intersections.* Australian Road Research Board.

Moses, R., & Mtoi, E. (2013). *valuation of Free Flow Speeds on Interrupted Flow Facilities.* Tallahassee, United States: Technical Report. Department of Civil Engineering, FAMU-FSU College of Engineering.

MTO. (2013). *Ontario Road Safety Annual Report (ORSAR) 2013*. Retrieved May 30, 2012, from Ontraio Ministry of Transportation: http://www.mto.gov.on.ca/english/publications/pdfs/ontario-road-safety-annual-report-2013.pdf

Mück, J. (2002). Using detectors near the stop-line to estimate traffic flows. *Traffic engineering & control., 43*(11), 429-434.

MUTCD. (2009). *Manual on Uniform Traffic Control Devices for Streets and Highways.* Washington, D.C.: U.S. Department of Transportation Federal Highway Administrartion.

National Marine Electronics Association. (2008). *NMEA 0183 standard.* National Marine Electronic Association Publications/Standards.

NHTSA. (2016). *NOTICE OF PROPOSED RULEMAKING ON V2V COMMUNICATIONS.* Washington, DC: National Highway Traffic Safety Administration.

Niittymäki, J., & Pursula, M. (1997). Saturation Flows at Signal-Group-Controlled Traffic Signals. *Transportation Research Record: Journal of the Transportation Research Board, 1572*(1), 24-32. doi:10.3141/1572-04

NTCIP 1202. (2005). Retrieved from National Transportation Communication for ITS Protocol: www.ntcip.org

Olstam, J., & Tapani, A. (2004). *Comparison of Car-following models.* Swedish National Road and Transport Research Institute.

OST-R. (2017, January 28). *Development Activities - International Standards Harmonization*. Retrieved from United States Department of Transportation - Office of the Assistand Secretart for Research and Technology - Intelligent Transportation Systems Joint Program Office: https://www.standards.its.dot.gov/DevelopmentActivities/IntlHarmonization

Perrin, H., Martin, P., & Hansen, B. (2001). Modifying Signal Timing During Inclement Weather. *ransportation Research Record: Journal of the Transportation Research Board, 1748*, 66-71.

PTV. (2012). VISSIM 5.40 - User Manual. Planung Transport Verkehr.

Rakha, H., & Gao, Y. (2010). *Calibration of steady-state car-following models using macroscopic loop detector data.* Blacksburg, VA: Virginia Tech Transportation Institute.

Rakha, H., Farzaneh, M., Arafeh, M., & Sterzin, E. (2008). Inclement weather impacts on freeway traffic stream behavior. *Transportation Research Record: Journal of the Transportation Research Board, 2071*, 8-18.

Remias, S., Hainen, A., Day, C., Brennan, T., Li, H., Rivera-Hernandez, E., . . . Bullock, D. (2013). Performance Characterization of Arterial Traffic Flow with Probe Vehicle Data. *Transportation Research Record: Journal of the Transportation Research Board, 2380*, 10-21.

Richards, P. I. (1956). Shock Waves on the Highway. *Operations Research, 4*(1), 42-51. doi:10.1287/opre.4.1.42

RITA. (2012). *International Deployment of Cooperative intelligent Transportation Systems - Bilateral Efforts of the European Commisson and United State Department of Transportation.* Washington, DC: Federal Highway Administration.

RITA. (2014, March 18). *Harmonization of International Standards and Architecture around the Vehicle Platform - See more at: http://www.its.dot.gov/research/harmonization.htm#sthash.CUf8Ji46.dpuf*. Retrieved April 14, 2014, from U.S. Department of Transportation - Research and Innovative Technology Administration: http://www.its.dot.gov/research/harmonization.htm

RITA. (2015, April 6). *DSRC: The Future of Safer Driving*. Retrieved from U.S. Department of Transportation Research and Innovative Technology Administration: http://www.its.dot.gov/factsheets/dsrc_factsheet.htm

Robertson, D. I. (1969). *Transyt: A traffic network study tool.* Berkshire, England: Transport Research Laboratory.

Roess, R. P., Prassas, E. S., & McShane, W. R. (2004). *Traffic engineering* (3 ed.). Upper Saddle River, N.J: Pearson/Prentice Hall.

SAE. (2016). *J2735 Dedicated Short Range Communications (DSRC) Message Set Dictionary.* Society of Automotive Engineers.

SCATS. (2013). *SCATS 6 Functional Description.* Retrieved from SCATS: http://www.scats.com.au/files/an_introduction_to_scats_6.pdf

SCOOT FAQ. (2008). Retrieved May 11, 2013, from SCOOT: http://www.scoot-utc.com/FAQ.php?menu=Reference#Q13

Shelby, S., D.M., B., D., G., R.S., G., Z.A., S., & N., S. (2008). Overview and Performance Evaluation of ACS Lite: Low-Cost Adaptive Signal Control System, Presented at Transportation Research Board 87th Annual Meeting. Washington, D.C.

Siemens. (2012). *SCOOT USER GUIDE.* SIEMENS PLC.

Skabardonis, A., & Nikolaos, G. (2005). Real-time estimation of travel times on signalized arterials. *Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory*. (pp. 387-406). Maryland: Elsevier.

Stephanopoulos, G., Michalopoulos, P. G., & Stephanopoulos, G. (1979). Modelling and analysis of traffic queue dynamics at signalized intersections. *Transportation Research Part A: General, 13*(5), 295-307.

Sturdevant, J., Overman, T., Raamot, E., Deer, R., & Miller, D. (2012). *Indiana Traffic Signal Hi Resolution Data Logger Enumerations.* Indiana: Indiana Department of Transportation and Purdue University.

Teply, S., Allingham, D. I., Richardson, D. B., & Stephenson, B. W. (2008). *Canadian Capacity Guide for Signalized Intersections.* Canada: Institute of Transportation Engineers (ITE).

Tiaprasert, K., Zhand, Y., Wang, X., & Zeng, X. (2015, August). Queue Length Estimation Using Connected Vehicle Technology for Adaptive Signal Control. *IEEE Transactions on Intelligent Transportation Systems, 16*(4), pp. 2129-2140.

Traffic Signal Timing Manual. (2008). Washington, D.C.: U.S. Department of Transportation Federal Highway Administrarion.

Trafficware. (2006). Synchro Studio 7 User Guide.

Webster, F. (1958). *Traffic Signal Settings.* Scientific and Industrial Research. London: Her Majestry's Stationary Office.

Webster, F., & Cobbe, B. (1966). *Traffic Signals - Road Research Technical Paper No. 56.* London: Her Majestry's Stationary Office.

Wiedemann, R. (1974). *Simulation des Straßenverkehrsflusses.* Karlsruhe: Univ., Inst. für Verkehrswesen.

Wu, Y.-J., Zhang, G., & Wang, Y. (2012). Link-Journey Speed Estimation for Urban Arterial Performance Measurement Using Advance Loop Detector Data under Congested Conditions. *Journal of Transportation Engineering, 138*(11), 1321-1332.

Ye, Q., Tarko, A., & Sinha, K. (2001). Model of free-flow speed for Indiana arterial roads. *Transportation Research Record: Journal of the Transportation Research Board, 1776*, 189-193.

Yi, P., Xin, C., & Zhao, Q. (2001). Implementation and field testing of characteristics-based intersection queue estimation model. *Networks and Spatial Economics, 1*(1-2), 205-222. doi:10.1023/A:1011589313578

Zhang, H. (1999). Link-journey-speed model for arterial traffic. *ransportation Research Record: Journal of the Transportation Research Board*, 109-115.

Appendix A
An Overview of the Connected Vehicle Technology

Over recent years, a new Intelligent Transportation System (ITS) technology called Connected Vehicles has received a lot of attention. This technology aims to address some of the problems associated with traffic safety and management. The connected vehicles technology facilitates two types of communications namely Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I). Using V2V, vehicles are able to communicate with each other and constantly exchange their travel data such as speed, position, length, etc. In V2I communication, data exchange is performed between vehicles and roadside facilities. V2V and V2I can bring new opportunities to improve traffic safety and mobility, for instance, by warning drivers about hazardous situations, or by constantly adapting the operation of traffic control devices to the real-time traffic conditions.

Data exchange in connected vehicles technology is provided using a wireless technology called Dedicated Short-Range Communication (DSRC). DSRC is a short to medium range wireless communication technology that allows a minimum communication range of 300 metres (NHTSA, 2016). It operates at 75 MHz spectrum in the 5.9 GHz frequency band. This spectrum is licensed and allocated to mobility and safety applications of Intelligent Transportation Systems (ITS) in order to prevent interruptions or noises from non-traffic related sources. Since vehicular safety is one of the main intended applications of the connected vehicles, this platform requires a fast and reliable communication in order to detect and to provide necessary actions in case of hazardous situations. In addition to satisfying this need, the DSRC technology has the ability of working in different weather conditions with a low latency as well as operating at high speed conditions (RITA, DSRC: The Future of Safer Driving, 2015).

Standards have been developed for DSRC to provide interoperability between devices operating based on this wireless communication technology. Since the early 2000s, studies have been undertaken in Europe, Unites States, and Japan to develop related standards. These efforts have

led to the development of three major standardizations in these three regions, which are different in some aspects. Thus, this appendix provides information on DSRC standards developed in the U.S. known as Wireless Access in Vehicular Environment (WAVE), and in Europe called Cooperative ITS (C-ITS) as two major comparable standards. Communication architecture and standards used at each level are presented and differences are highlighted. Messages used in both standards are presented and some applications related to connected vehicles are provided.

A.1. DSRC Standards in the U.S. and Europe

System Architecture

The European C-ITS is based on a system architecture presented in ETSI[3] EN 302 665 (2012) and ISO[4] 21217 (2014) standards. Similarly, the system architecture of the U.S. WAVE is specified in IEEE[5] 1609.0 (2013). Figure A-1 (a) and (c) compares both architectures against the general Open Systems Interconnection (OSI) (Microsoft, n.d.) and Transmission Control Protocol (TCP) (Microsoft TechNet, n.d.) reference models presented in Figure A-1 (b).

The European C-ITS access layer includes the layers 6 and 7 from the OSI model. The access layer contains Physical, MAC, and Logical Link Control (LLC) sublayers. Specifications related to the C-ITS access layer are presented in the European Standard ETSI EN 302 663 (2013). Similarly, the U.S. WAVE protocol stack includes physical, WAVE MAC, and LLC layers that correspond to the physical and data link layers of the OSI model. Functionalities of physical and MAC layers for both C-ITS and WAVE are presented in IEEE 802.11 (2012), and the LLC layer is specified in IEEE 802.2 (1998).

Both DSRC designs use 5.9 GHz frequency band. The details of the DSRC 5.9 GHz allocation scheme have been specified by the ETSI EN 302 571 (2016) for Europe, whereas the FCC 47

[3] In Europe, the DSRC standardization is performed by the European Telecommunications Standards Institute (ETSI).

[4] International Standard Organization (ISO)

[5] Institute of Electrical and Electronics Engineers (IEEE)

CFR[6] Part 95 (2009) and ASTM[7] E2213-03 (2010) have defined the channel allocation scheme for the United States. The 5.9 GHz frequency band includes several 10MHz channels as presented in Figure A-2.

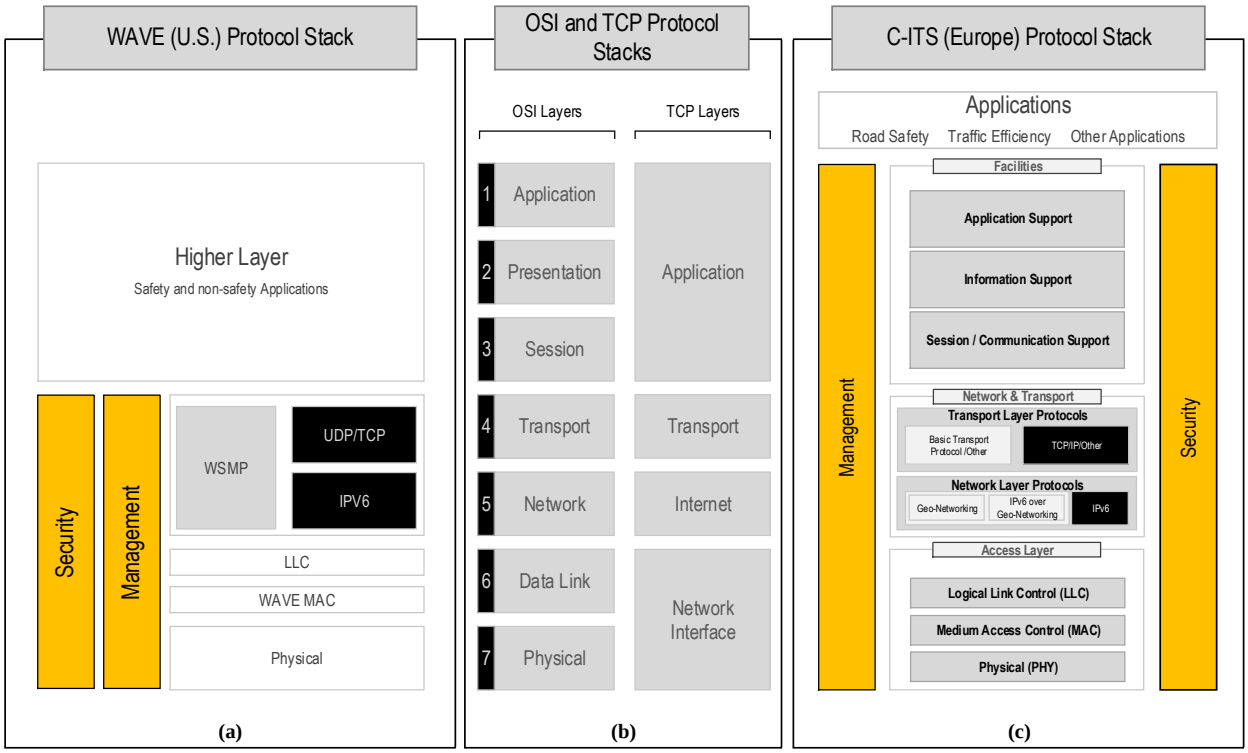


Figure A-1: The U.S. and European DSRC protocol stacks vs OSI and TCP layers

In the European Union, the frequencies between 5855 MHz and 5875 MHz are designated to ITS non-safety applications also known as ITS G5B. The frequency range of 5875 MHz to 5905 MHz is assigned to the ITS road safety applications (ITS G5A), whereas future ITS applications (ITS G5D) will use 5905 MHz to 5925 MHz. The Control Chanel (CCH) is used as the default channel for cooperative road safety. If the CCH is congested with safety messages, Service Channels (SCH) 178 and 176 will be used to announce ITS safety and road efficiency services (ETSI, 2012). In the U.S., SCH 172 is dedicated to V2V road safety collision avoidance. SCH 184 is dedicated for public safety applications and has a longer communication range and a higher power. Similar to C-ITS, WAVE uses the CCH as the default channel for its operation. However, no specific

[6] Federal Communications Commission (FCC) - Code of Federal Regulations (CFR)

[7] American Society for Testing and Materials

channel range is assigned for safety applications; therefore, any of the channels including the control channels can be used for both safety and non-safety applications (IEEE 1609.0, 2013).

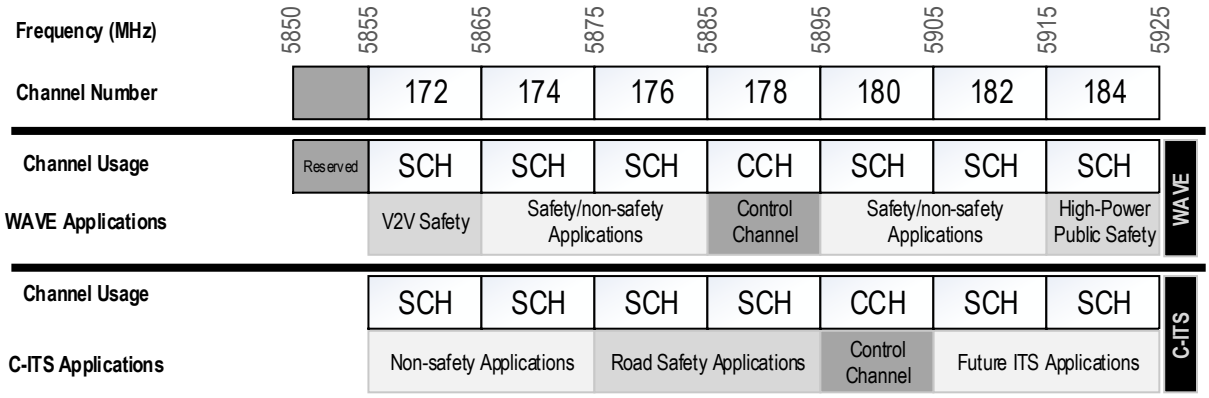


Figure A-2: Frequency allocation in C-ITS (Europe) and WAVE (U.S.)

In both protocol stacks, the Network and Transport layers (Layers 4 and 5 of the OSI model) work together. These layers are responsible for providing addressing and delivery services between upper and lower layers. Services such as these can be performed either with the standard IPv6 protocol specified in ISO 21210 (2012), or one of the following protocols designed for the ITS application.

As identified in IEEE 1609.3 (2016), the WAVE Short Message Protocol (WSMP) is designed specifically for the operation of WAVE. This protocol provides faster communication between layers using a lower payload and latency.

Similarly, C-ITS uses an ITS specific communication protocol called Basic Transport Protocol (BTP), which is presented in ETSI EN 302 636-5-1 (2014). BTP in combination with Geo-networking in network layer acts similarly to WSMP and provides a fast and low latency ad-hoc communication between vehicles. The role of Geo-networking is to disseminate data packets based on geographical locations of vehicles using several methods including GeoUnicast, GeoBroadcast, and topologically-scoped broadcast. Details about the geo-networking are defined in ETSI EN 302 636 standards parts 1 to 4.

The top layer of the TCP protocol stack is called application, which corresponds to the top three layers of the OSI protocol stack (application, presentation, and session). These three layers are treated differently in European and the American DSRC designs. In the U.S. WAVE standard, applications directly interact with the Transport and Network layers, and they contain all the

necessary functions to work with these layers. A standard called SAE J2735 (2016) is developed by the Society of Automotive Engineers, which provides details about the type of information that applications can exchange using the WAVE protocol stack. Details about this standard are provided later in this appendix.

In contrast to the WAVE design, in the C-ITS European design, common functionalities, communication tools, and information required for applications are included in a layer called Facilities, which is located between Application and Transport/Network layers. The C-ITS facilities layer includes three main entities as described below and presented in Figure A-3. Detailed description of these entities are provided in ETSI TR 102 638 (2009) .

1) ***The Application Support*** includes common message management service used for exchanging data between ITS stations. Two important common massages include Decentralized Environmental Notification Message (DENM), and Cooperative Awareness Message (CAM), which are explained later in this paper. The application support provides information such as position (latitude, longitude, altitude) to different requesting applications.

2) ***The information support*** corresponds to the presentation layer of the OSI model and contains information such as maintenance of applications and their versions. It also decodes and encodes messages for applications and includes a Local Dynamic Map (LDM) database, which is a detailed map containing road geometry and other information obtained from sensors.

3) ***The communication support*** covers the OSI session layer and manages communication with the lower layers.

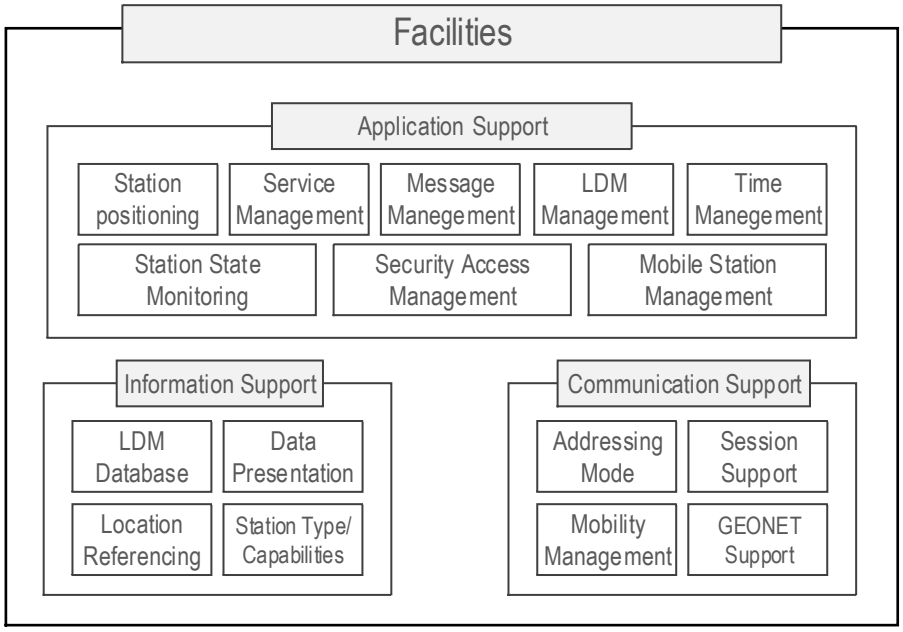


Figure A-3: Entities included in the C-ITS facility layer (ETSI, 2009)

As shown in Figure A-1, both C-ITS and WAVE protocol stacks use management and securities entities to manage the communication, security and privacy of the exchanged data over DSRC devices. Details about these entities are beyond the scope of this appendix. In the following, more details are provided about the message types that can be sent using the DSRC technology in both C-ITS and WAVE standards.

WAVE message sets

SAE J2735 DSRC Message Set Dictionary defines the type of data that can be transmitted by WAVE. In this manual there are 231 data elements identified that can be sent using 16 different message sets. Message sets contain several data elements and data frames that are transmitted as a package. Brief descriptions of these 16 messages are provided below (SAE, 2016)

- ***Basic Safety Message (BSM)***

  The basic safety message (Figure A-4) (also known as "heartbeat message") includes basic but essential information about individual vehicles. This message set has two parts. Part I of the BSM includes crucial information that is considered mandatory to be sent with each broadcast of the message at the frequency of 10 times per second (10MHz). Part II may include additional information required by certain applications. BSM Part II is optional and can be sent at different frequencies or as required (e.g. event based).

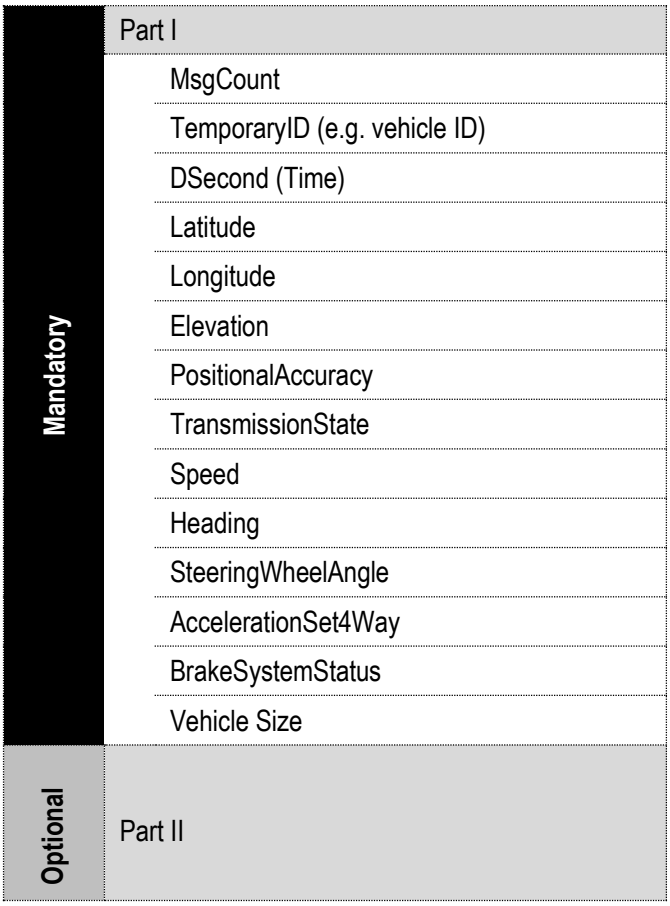


Figure A-4: Basic Safety Message – Data Elements (SAE, 2016)

- **Map Data**

  The map data includes different maps such as intersection geometry, lane information, road segment information, etc. that are necessary for some applications (e.g. intersection collision avoidance).

- **Signal Phase and Timing Message (SPAT)**

  SPAT contains information about active signal phases of one or several intersections including the remaining green time, pre-emptions, and priorities.

- **Common Safety Request**

  The Common Safety request refers to a unicast message sent from one vehicle to another asking for certain information required for its safety application. The message is received by the second vehicle and the requested information (if supported by that vehicle) will be included in Part II of its safety message.

- **Emergency Vehicle Alert**

This message is broadcasted by an emergency vehicle during an emergency operation. It is a warning message to other drivers about the operation of the emergency vehicle and to produce appropriate recommendations to drivers. This message includes information such as the response type, status of siren and light bar, vehicle mass and type, etc.

- **Intersection Collision Avoidance**

  This message is used for the intersection collision avoidance applications. It includes the recent path, and acceleration of a vehicle, as well as the ID of the approaching intersection.

- **NMEA Correction**

  Correction parameters used in GPS devices to correct errors and increase the accuracy of the estimated position are broadcasted using this message. The differential corrections for GPS/GNSS (Global Navigation Satellite System) systems are defined in National Marine Electronic Association (NMEA) 0183 (2008) standard.

- **Personal Safety Message**

  Includes information on pedestrians, cyclists, and road construction workers. This information includes position, heading, acceleration, path history, etc.

- **Probe Data Management**

  This message set helps in identifying the Road Side Unit (RSU) coverage. It specifies the time and the distance that a vehicle can start joining and exchanging data with a RSU.

- **Probe Vehicle Data**

  The message includes travel information of a vehicle along a segment of a road. Collected data are transmitted to other DSRC devices (usually RSU) at a time defined in the message. The message may include several elements defined in the SAE J2735 manual.

- **Road Side Alert**

  This message set is used to alert drivers about different hazardous situations such as an icy road section, an approaching train, etc.

- **RTCM Corrections**

  It is used to provide the differential corrections for vehicles' on-board GPS systems defined in various standards developed by Radio Technical Commission for Maritime (RTCM) special committee number 104. The message helps in improving the estimated positional accuracy of GPS systems.

- **Signal Request Message**

This message is sent by a vehicle requesting a pre-emption (e.g. an emergency vehicle) or priority (e.g. Transit vehicles)

- **Signal Status Message**

  This is a response to a vehicle sending the signal request message. It includes the status of signals, pending pre-emptions, and the priority and the rank of the requesting vehicle among other requests.

- **Traveller Information**

  The message includes advisories for drivers (e.g. traffic information, incidents, road signs, work zones, speed limits, etc.). Advisories can be in forms of audio, text, graphics, or others.

- **Test Messages**

  Designed for developing customized messages. In total, there are 16 test messages available that can be used.

C-ITS Message Sets

ETSI TS 102 894-2 (2014) standard specifies 132 Data Elements (DE) for the DSRC technology. These data elements can be sent as different message sets presented in Table A-1. Due to their importance in safety applications, in the following, the Cooperative Awareness Message (CAM), and the Decentralized Environment Notification Message (DENM) are presented in more details.

Table A-1: C-ITS message sets (ETSI, 2014)

Message	Standard
Cooperative Awareness Message (CAM)	(ETSI, 2014)
Decentralized Environmental Notification Message (DENM)	(ETSI, 2014)
Point of Interest (POI)	(ETSI, 2012)
Signal Phase And Timing (SPAT)	(SAE, 2016)
MAP	(SAE, 2016)
In Vehicle Information (IVI)	(ISO TS 19321, Under development)
Electric vehicle recharging spot reservation (ev-rsr)	(ETSI, 2014)

Cooperative Awareness Message (CAM)

ETSI EN 302 637-2 (2014) standard is dedicated to CAM, which is a mandatory message broadcasted by vehicles. The CAM message is broadcasted 10 times per second at its maximum

rate similar to BSM in SAE J2735, and every second at its minimum frequency depending on the dynamic of the broadcasting unit. The standard also identifies the criteria that affect the rate of the broadcast. The general structure of CAM is presented in Figure A-5. The message includes a header, a basic container and a High Frequency (HF) container, which are mandatory to be sent by each message. The message also contains optional components namely a Low Frequency (LF) container and a special vehicle container.

Mandatory			Optional	
Header	Basic Container	High Frequency (HF) Container	Low Frequency (LF) Container	Special Vehicle Container

Figure A-5: Structure of CAM (ETSI, 2014)

The basic container provides basic information about the broadcasting unit such as its type and latest position with its accuracy. The HF container provides additional information about the vehicle as presented in Figure A-6.

Mandatory	Heading
	Speed
	Driving Direction
	Longitudinal Acceleration
	Curvature
	Curvature Calculation Mode
	Yaw Rate
	Vehicle Length
	Vehicle Width
Optional	Performance Class
	Acceleration Control
	Lane Number
	Steering Wheel Angle
	Lateral Acceleration
	Vertical Acceleration
	cenDsrcTollingZone

Figure A-6: Cooperative Awareness Message (CAM) high frequency container ETSI EN 302 637-2 (2014)

The LF container includes less dynamic information about the broadcasting unit that do not require frequent updates including vehicle role, exterior light status, and the path history. The special container includes additional information about the unit such as if the unit is carrying dangerous goods, or if it is an emergency vehicle in operation, etc. These containers are included in the CAM message with the frequency of equal or larger than twice per second (ETSI, 2014).

Decentralized Environment Notification Message (DENM)

DENM is a safety event based message used to inform drivers about road incidents. Details and standards related to this message are provided in ETSI EN 302 637-3 (2014), in which defines 58 data elements that can be included in this message. DENM is usually broadcasted within a specific area, and unlike CAM that is broadcasted with a higher rate along a vehicle's path, the message may not be broadcasted by the originating vehicle when it leaves the incident area. In addition to this difference, DENM can be forwarded by other vehicles using the GeoReferencing technique to deliver the message to a destination beyond the transmission radius of the originating point, or to keep dissemination of a message within an area when the originating vehicle does not present in that zone anymore.

Figure A-7 shows DENM data structure. The mandatory parts include a header and a management container. The management container includes information about the safety situation such as its occurrence time and position, and some parameters associated to the DENM including its validity duration, and transmission interval. The situation container includes more information about the incident such as the quality of the obtained information, type of the event, any other events resulted from or related to the current event (e.g. an accident caused by a slippery road surface condition), etc. The location container includes information about the event, for instance, its speed, trace, and heading. Finally, A la carte container may carry additional information that is not covered by above containers.

Mandatory		Optional		
Header	Management Container	Situation Container	Location Container	A la carte Container

Figure A-7: Structure of the DENM

Applications

Several applications related to the connected vehicle technology are proposed both by the North American and the European organizations involved in the development of this technology. These applications are under development and investigation and standards try to identify the use of defined message sets or data elements for these proposed applications as well as identify the minimum requirements in terms of the data accuracy, broadcast frequency, latency, etc. for such applications. Table A-2 provides a list of proposed connected vehicles applications in the North

America, and Table A-3 presents a list of applications envisioned for the connected vehicle technology in the Europe.

Table A-2: Connected Vehicles applications in the North America

V2I Safety

- Red Light Violation Warning
- Curve Speed Warning
- Stop Sign Gap Assist
- Spot Weather Impact Warning
- Reduced Speed/Work Zone Warning
- Pedestrian in Signalized Crosswalk Warning (Transit)

V2V Safety

- Emergency Electronic Brake Lights (EEBL)
- Forward Collision Warning (FCW)
- Intersection Movement Assist (IMA)
- Left Turn Assist (LTA)
- Blind Spot/Lane Change Warning (BSW/LCW)
- Do Not Pass Warning (DNPW)
- Vehicle Turning Right in Front of Bus Warning (Transit)

Probe Data

- Probe-based Pavement Maintenance
- Probe-enabled Traffic Monitoring
- Vehicle Classification-based Traffic Studies
- CV-enabled Turning Movement & Intersection Analysis
- CV-enabled Origin-Destination Studies
- Work Zone Traveler Information

Environment

- Eco-Approach and Departure at Signalized Intersections
- Eco-Traffic Signal Timing
- Eco-Traffic Signal Priority
- Connected Eco-Driving
- Wireless Inductive/Resonance Charging
- Eco-Lanes Management
- Eco-Speed Harmonization
- Eco-Cooperative Adaptive Cruise Control
- Eco-Traveler Information
- Eco-Ramp Metering
- Low Emissions Zone Management
- AFV Charging / Fueling Information
- Eco-Smart Parking
- Dynamic Eco-Routing (light vehicle, transit, freight)
- Eco-ICM Decision Support System

Road Weather

- Motorist Advisories and Warnings (MAW)
- Enhanced MDSS
- Vehicle Data Translator (VDT)
- Weather Response Traffic Information (WxTINFO)

Mobility

- Advanced Traveler Information System
- Intelligent Traffic Signal System (I-SIG)
- Signal Priority (transit, freight)
- Mobile Accessible Pedestrian Signal System (PED-SIG)
- Emergency Vehicle Preemption (PREEMPT)
- Dynamic Speed Harmonization (SPD-HARM)
- Queue Warning (Q-WARN)
- Cooperative Adaptive Cruise Control (CACC)
- Incident Scene Pre-Arrival Staging Guidance for Emergency Responders (RESP-STG)
- Incident Scene Work Zone Alerts for Drivers and Workers (INC-ZONE)
- Emergency Communications and Evacuation (EVAC)
- Connection Protection (T-CONNECT)
- Dynamic Transit Operations (T-DISP)
- Dynamic Ridesharing (D-RIDE)
- Freight-Specific Dynamic Travel Planning and Performance
- Drayage Optimization

Smart Roadside

- Wireless Inspection
- Smart Truck Parking

Table A-3: Connected Vehicles applications in Europe (ETSI TR 102 638, 2009)

Co-operative road safety	Traffic Efficiency	Others
• Emergency electronic brake lights • Safety function out of normal condition warning • Emergency vehicle warning • Slow vehicle warning • Motorcycle warning • Vulnerable road user Warning • Wrong way driving warning • Stationary vehicle warning • Traffic condition warning • Signal violation warning • Roadwork warning • Decentralized floating car data • Overtaking vehicle warning • Lane change assistance • Pre-crash sensing warning • Co-operative glare reduction • Across traffic turn collision risk warning • Merging Traffic Turn Collision Risk Warning • Co-operative merging assistance • Hazardous location notification • Intersection Collision Warning • Co-operative forward collision warning • Collision Risk Warning from RSU	• Regulatory/contextual speed limits • Traffic light optimal speed advisory • Traffic information and recommended itinerary • Enhanced route guidance and navigation • Intersection management • Co-operative flexible lane change • Limited access warning, detour notification • In-vehicle signage • Electronic toll collect • Co-operative adaptive cruise control • Co-operative vehicle-highway automation system (Platoon)	• Point of interest notification • Automatic access control/parking access • Local electronic commerce • Car rental/sharing assignment/reporting • Media downloading • Map download and update • Ecological/economical drive • Instant messaging • Personal data synchronization • SOS service • Stolen vehicle alert • Remote diagnosis and just in time repair notification • Vehicle relation management • Vehicle data collect for product life cycle management • Insurance and financial Services • Fleet management • Vehicle software/data provisioning and update • Loading zone management • Vehicle and RSU data calibration

A.2. Harmonization of International Standards

The U.S. Department of Transportation (USDOT) Research and Innovative Technology Administration (RITA) (now known as the Office of the Assistant Secretary for Research and Technology (OST-R)), and the European Commission Directorate General for Communication Networks, Content and Technology (CONNECT) have signed an agreement in 2009 to collaborate on a research program focused on the cooperative vehicles. The goal is to harmonize the currently available and future standards related to the connected vehicles technology in order to ease the development and adaptation of this technology in both regions and to reduce the deployment costs. The following six working groups focus on different aspects of standards in the EU and the US (RITA, 2012).

1. ***Safety applications working group***

   This working group collaborates on harmonizing the over-the air data and communication interfaces to pave the path for the development of applications and hardware that can operate in both regions.

2. ***Sustainability application working group***

   The focus of this research group is to develop connected vehicle applications with the goal of improving the quality of the environment in a transportation network. Furthermore, the work group works on developing a common standard and an environmental message set to facilitate the development of such applications.

3. ***Standards harmonization working group***

   This working group aims at developing and harmonizing connected vehicle standards in EU and the US, and preventing the development of redundant standards in both regions.

4. ***Assessment tools working group***

   This work group is dedicated to developing common evaluation methodologies and test fields for performance assessment of connected vehicle systems.

5. ***Driver distraction and human-machine interaction (HMI) working group***

   This is research collaboration on driver distractions related to HMI and related road safety issues.

6. ***Glossary working group***

   This working group focuses on developing common terms and keywords in both regions to ease discussions related to connected vehicles.

The standard harmonization working group further divides into several Harmonization Task Groups (HTG), working on different aspects of the standards harmonization between the United States and Europe. These HTGs include (OST-R, 2017):

- **HTG 1** and **HTG 3**: ITS Security and Communication Protocols
- **HTG 2**: Harmonization of Basic Safety Message (BSM) and Cooperative Awareness Message (CAM)
- **HTG 4** and **HTG 5**: Infrastructure (V2I) Messages
- **HTG 6**: Cooperative ITS Security Policy

- **HTG 7**: Standards Selection, Gap Analysis, and Identifiers for Connected Vehicle architectures

To date, as part of the achievements of the above harmonization task groups, several aspects of standards are harmonized and some are ongoing. As part of this harmonization, as an example, contents of the CAM and the BSM have been harmonized (although they are not identical) so that systems and applications can work using both two messages with a simple translation (RITA, 2014).

A.3. Summary and Conclusions

This appendix presented an overview of the DSRC technology in the United States and Europe. The protocol stacks of both regions were examined and the related standards were introduced. The comparison of the DSRC protocol stacks demonstrated that these two platforms are very similar at the lower layers. However, there are some differences in the upper layers in terms of how communication is handled and messages are formatted. The major difference between the two protocol stacks is in the use of a facilities layer in the Europe in contrast to the U.S. which applications directly communicate with the networking and transport layers. Collaboration is underway between European commission and US DOT to harmonize standards in these two regions and to adapt similar approaches in designing their DSRC technology. It is expected that this harmonization will ease the development of the hardware and applications that can operate in both regions, and will reduce the associated production costs, which will pave the path for faster deployment of the cooperative vehicles technology.