Real-Time Traffic Performance Measurement of Signalized Intersections Using Connected Vehicle Data: A Simulation-Based Study

by

Yubo Wang

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Civil Engineering

Waterloo, Ontario, Canada, 2023

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

Traffic congestion has long become a major concern in many cities in Canada and around the world. It has been estimated that the annual total economic loss due to traffic congestion in major Canadian urban centers has reached nearly $4 billion. Real-time monitoring of traffic conditions and measurement of the performance of the underlying traffic management systems is a critical requirement for mitigating and minimizing the impact of traffic congestion in an urban road network. The latest advance in the Connected Vehicle (CV) technology has afforded a new opportunity for developing solutions that make use of high-resolution trajectory data for real-time urban traffic monitoring and performance measurement, such as Automated Traffic Signal Performance Measures (ATSPM). However, many critical issues still need to be addressed before the potential of CV can be fully realized. For example, in the context of ATSPM, what traffic performance measures could be derived from the CV data? Can non-recurrent congestion be detected in real-time and at what latency? What would be the optimal spatial and temporal data aggregation resolutions of CV data? What would be the effect of the CV market penetration rate on the reliability of specific performance measures? This research attempts to address some of these questions through a simulation study of a real-world signalized urban arterial corridor from Broward County, Florida, US, consisting of 17 signalized intersections with a wide range of layouts and congestion levels. An extensive set of simulation experiments have been conducted under a range of scenarios varying by facility types (single intersection vs. corridor), congestion level (from undersaturated to oversaturated), CV market penetration rates (1%-25%), and signal timing plans. Under each scenario, samples of vehicles at specific market penetration rates are randomly drawn from the simulated traffic population to represent the CVs and their trajectory data are used to calculate various signal performance measures, including average overall delay, percentile queue length, percentage of stopped vehicles and an average number of stops, at the spatial aggregation levels of movements, approaches, intersections, and corridor. A sensitivity analysis is subsequently conducted to assess the accuracy and reliability of the performance measures derived from CV data as related to some specific external conditions and factors. The results from the simulation experiments have underscored the significant potential of CV data, even under the current relatively low market penetration rate, for estimating various important traffic performance measures and detecting non-recurrent events or bottlenecks - a basic requirement for implementing ATSPM.

Acknowledgements

First and foremost, I would like to extend my deepest gratitude to my supervisor, Professor Liping Fu, for his unwavering guidance and mentorship throughout my thesis writing and my entire master's program. His invaluable support and benevolence have been instrumental in enabling me to complete this research. I consider myself incredibly fortunate to have had the opportunity to become his student. Professor Fu has been the most respectful mentor in my life, inspiring and motivating me with his profound professionalism, extensive knowledge, and remarkable personality.

Furthermore, I would like to extend my heartfelt gratitude to my committee members, Professor Mahesh Pandey and Assistant Professor Pejoohan Tavassoti-Kheiry, for dedicating their time to reviewing my thesis.

I am also grateful to my colleagues from the iTSS lab: Pirooz Gahleh, Reza Aminghafouri, and Ryoto Miyako for their constructive feedback during our group meetings. Special thanks to my peers, Alex (Jihao) Deng, Ce Zhang, Jeffrey Dai, Qiangqiang Shangguan, Yubo Jiao, and Yuzhe You. Their camaraderie and insights have been indispensable. May each fellow in my lab has a shiny and bright future in both academia and career.

Lastly, my heartfelt appreciation goes to my parents, whose unwavering love and support have been pivotal in moulding me into the person I stand as today. Their encouragement has been my guiding light throughout this academic adventure. I am also deeply grateful to Jiazhen Lin and his family for the warmth and care they provided during my stay in Canada. Additionally, I would like to extend my thanks to Jiaozi and Gaolegao for the moments of joy and happiness they have added to my life.

Table of Contents

Author's Declaration ii

Abstract iii

Acknowledgements iv

List of Figures viii

List of Tables ix

Chapter 1 Introduction 1

1.1 Research Background 1

1.2 Real-time traffic monitoring and management systems and ATSPM 2

1.3 Research Problem 5

1.4 Research Objectives 6

1.5 Thesis Structure 7

Chapter 2 Literature Review 9

2.1 Studies on The ATSPM Derived from CV Data 9

2.2 Studies on The Effect of CVMPR 11

2.3 Studies on The Congestion Detections Based on CV Data 13

2.4 Summary 14

Chapter 3 Methodology 16

3.1 Overview of the Proposed Methodology 16

3.2 VISSIM Simulation Study 18

3.2.1 The Broward Blvd 18

3.2.2 Basic Information on Traffic Conditions 20

3.3 Simulation Experiments 21

3.3.1 Simulation Experiment Design 22

3.3.2 Simulation Model Calibration..24

3.3.3 Vehicle Sampling Procedure and Result Analysis Techniques 25

3.4 Performance Measure Calculation using CV Data from Simulation 27

3.4.1 Identification of Deceleration Starting Point 27

3.4.2 Average Overall Delay..30

3.4.3 95th Percentile Queue Length...31

3.4.4 Percentage of Stopped Vehicles (PSV)...34

3.5 Sensitivity Analysis...35

Chapter 4 Analysis of Performance Measurement and Congestion Detection 38

4.1 Performance Measure Accuracy Analysis 38

4.1.1 Results at the Movement Level...38

4.1.2 Result at the Approach Level..41

4.1.3 Results at the Intersection Level 44

4.1.4 Results at the Corridor Level 47

4.2 Congestion Detection 49

4.2.1 Detection of Recurrent Congestions 50

4.2.2 Detection of Non-recurrent Congestions under Capacity Constraints 52

4.2.3 Detection of Non-recurrent Event-driven Congestions...55

Chapter 5 Conclusions 58

5.1 Summary of The Key Findings...58

5.2 Significance of the Study for Urban Traffic Monitoring and Management...............................60

5.3 Future Research Directions and Recommendations 61

5.4 Limitations 62

References 63

Appendix A Code..70

Appendix B Signal Timing of Each Intersection 74

Appendix C: Lane Configuration and Movements of the Individual Intersections 77

List of Figures

Figure 1-1 The Top Ten Most Congested Cities in North America (INRIX, 2022) 1
Figure 1-2 Intersection Performance in Toronto during Peak Hours (City of Toronto, 2022) 2
Figure 1-3 Five Basic ATSPM System Components (USDOT, 2020) 3
Figure 3-1 The General Framework of the Proposed Simulation-Based Study 17
Figure 3-2 Overview of the Study Corridor 19
Figure 3-3 A Sample of Real-World Vehicle Trajectories 19
Figure 3-4 Eastbound AADT of Several Major Intersections along The Broward Blvd. 21
Figure 3-5 Corridor Layout in the VISSIM Simulation Model 22
Figure 3-6 CV reports in the Simulation 28
Figure 3-7 Time Sequence within An Individual Vehicle Trajectory 29
Figure 3-8 The Time-Distance Diagram with Delay Terms at The Signalized Intersection 30
Figure 3-9 Identification of Acceleration and Deceleration Points in the Queue 32
Figure 3-10 Impact of W-74 Standstill Distance 35
Figure 3-11 Impact of W-74 Safety Distance (ft.) 36
Figure 3-12 Impact of Acceleration Threshold Settings 36
Figure 4-1 Movement-level Accuracy Analysis of Performance Measure Estimations 41
Figure 4-2 Approach-level Accuracy Analysis of Performance Measure Estimations 42
Figure 4-3 Intersection-level Analysis of Performance Measures Derived from Simulated CV Data 47
Figure 4-4 Corridor-level Analysis of Performance Measures Derived from Simulated CV Data 49
Figure 4-5 Recurrent Congestion Detections 51
Figure 4-6 Graphic Illustration of the Capacity Constraint Scenario 53
Figure 4-7 Non-recurrent Congestion Detection Using Simulated CV Data (Capacity-constraint Scenario) 54
Figure 4-8 Non-recurrent Congestion Detection Using Simulated CV Data (Event-driven Scenario) 57

List of Tables

Table 2-1 Summary of Proposed Performance Measures in Literature 10
Table 3-1 Summary of the 17 Intersections 20
Table 3-2 Eastbound AADT of Several Major Intersections along The Broward Blvd. 21
Table 3-3 The Traffic Volume of Different Scenarios at Movement Level 22
Table 3-4 The Traffic Volume of Different Scenarios at Approach Level 23
Table 3-5 The Traffic Volume of Different Scenarios at the Intersection Level 23
Table 3-6 The Number of Trajectories Collected from the Corridor Level 23
Table 3-7 The Simulation Settings of Other Parameters 24
Table 3-8 Number of Runs and Performance Reliability 25
Table 3-9 Key Output Variables from VISSIM CV Trajectory Data 26
Table 4-1 Traffic Volume Variations during The Recurrent Congestion Detections 50
Table 4-2 Capacity Variations during The Capacity-Constraint Scenario 53
Table 4-3 Traffic Volume Variations during The Event-driven Congestion Scenario 55

Chapter 1
Introduction

1.1 Research Background

Traffic congestion has drawn major concern in many cities in Canada and around the world. The estimated total loss generated by unexpected congestion reaches nearly $4 billion every year in major Canadian urban centers. According to the Canadian Automobile Association (CAA), congestion in major Canadian cities collectively costs drivers over 11.5 million hours and drains an extra 22 million litres of fuel per year (CAA, 2021). In the Greater Toronto and Hamilton area, the social and economic costs of congestion were approximately $3.3 billion per year (Metrolinx, 2021). It is worth noting that the City of Toronto is the third among the top ten North American cities experiencing the most acute traffic congestion, as depicted in Figure 1-1 (INRIX, 2022). Traffic congestion has been a persistent source of disturbance in the operational aspects of traffic management within the city.

Figure 1-2 shows the traffic conditions of diverse intersections operating at the highest capacity during AM peak hours and PM peak hours in the metropolitan city of Toronto (City of Toronto, 2020). The unfavourable traffic conditions exert extremely negative influences on traffic efficiency and increase the travel time of each driver.

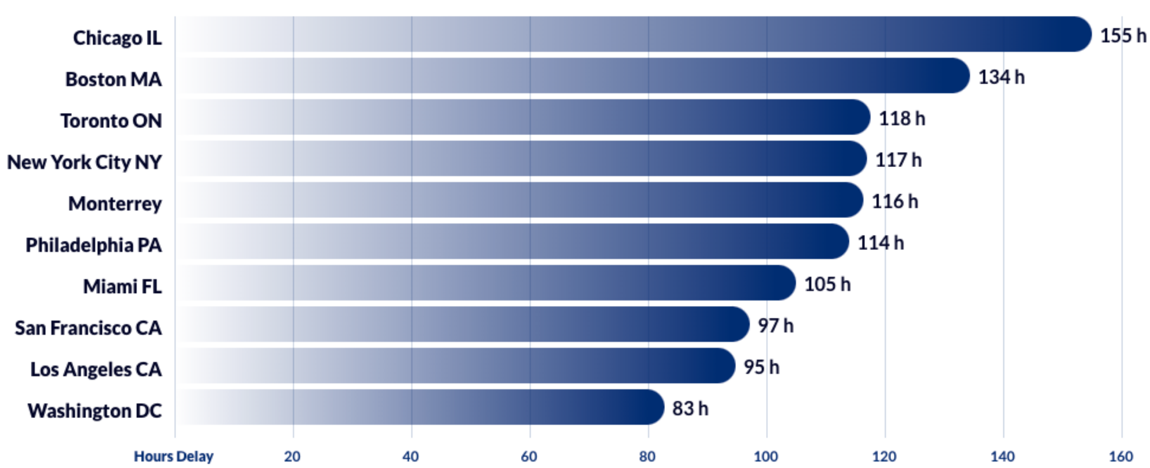


Figure 1-1 The Top Ten Most Congested Cities in North America (INRIX, 2022)

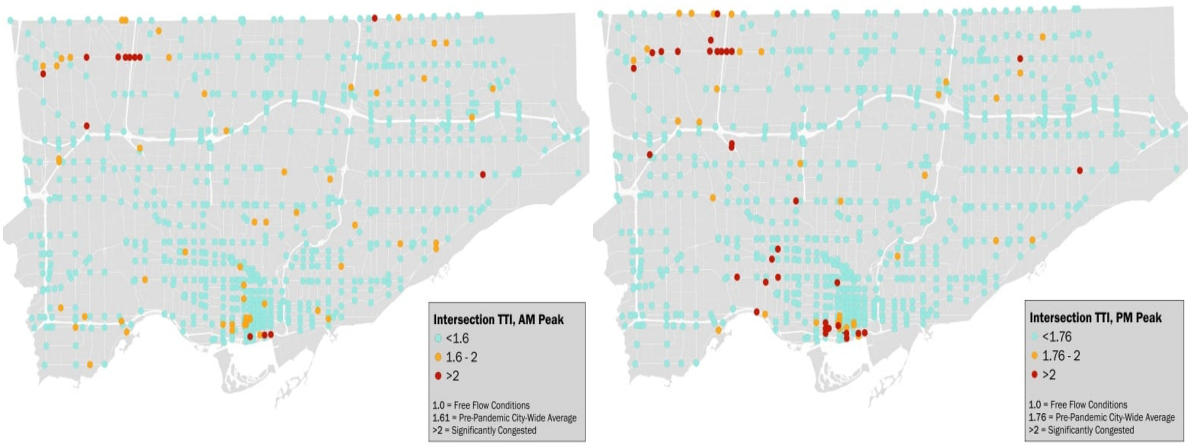


Figure 1-2 Intersection Performance in Toronto during Peak Hours (City of Toronto, 2022)

In response to the challenge of traffic congestion and the need to improve traffic performance, innovative technologies are being developed as important tools for real-time traffic monitoring and management.

1.2 Real-time traffic monitoring and management systems and ATSPM

Effective traffic management is crucial for mitigating and minimizing the influence of traffic congestion in urban areas. Central to any traffic management system is the ability to monitor traffic conditions and measure the performance of the underlying traffic management solutions such as signal control in real-time, which enables prompt detection and immediate response to variations in traffic conditions, especially under non-recurrent events. Through the continuous gathering and analysis of traffic data, real-time traffic monitoring and management systems can provide expeditious and precise information on traffic performance measurements and identifications of potential areas where traffic flow optimization is required.

Traffic management systems deploy diverse strategies and solutions targeting congestion problems, such as adaptive signal control and incident management (Brandon Nevers et al., 2020). Among various types of urban traffic management solutions, the Automated Traffic Signal Performance Measure (ATSPM) is one of the most widely adopted technologies due to its effectiveness in identifying signal operational deficiencies, improving roadway capacity and managing signal timing, which is defined as a technology that uses performance measures, data collection and data analysis tools to support objectives and performance-based approaches to traffic signal operations, maintenance, management and design to improve the safety, mobility and efficiency of signalized intersections for all users (Federal Highway Administration (FHWA), 2021). The ATSPMs contain five essential components as shown in Figure 1-3.

ATSPMs encompass methods for analyzing real-time traffic data obtained through sensors and a data logging-capable traffic signal controller, which provide a range of functionalities, including the generation of real-time and historical data on measures such as vehicle delay, volume, speed, and travel time. This technology also allows for monitoring of various traffic performance measures, such as the platoon ratio, split failure, number of stops per mile and Green Occupancy Ratio (GOR). Among these measures, the platoon ratio represents the individual phase progression performance derived from the percentage of arrivals on green (Smaglik et al., 2007). Split failure is employed to identify instances where the intended allocation of green time to different phases of a traffic signal cycle deviates from expectations, resulting in an unexpected distribution of signal timings (Freije et al., 2014). The number of stops is a count of how many times vehicles come to a complete stop at a given location, such as a traffic signal intersection, which provides insights into the level of congestion, delays, and the overall efficiency of traffic movement at the intersection (Argote-Cabañero et al., 2015). GOR is simply the ratio of the detector occupancy during the green phase to the total green available in the split (Dakic et al., 2017). These enhanced insights enable transportation agencies to consistently assess the performance of their traffic signal controllers and take proactive measures to optimize signal timing, reduce congestion, and enhance overall road safety and efficiency.

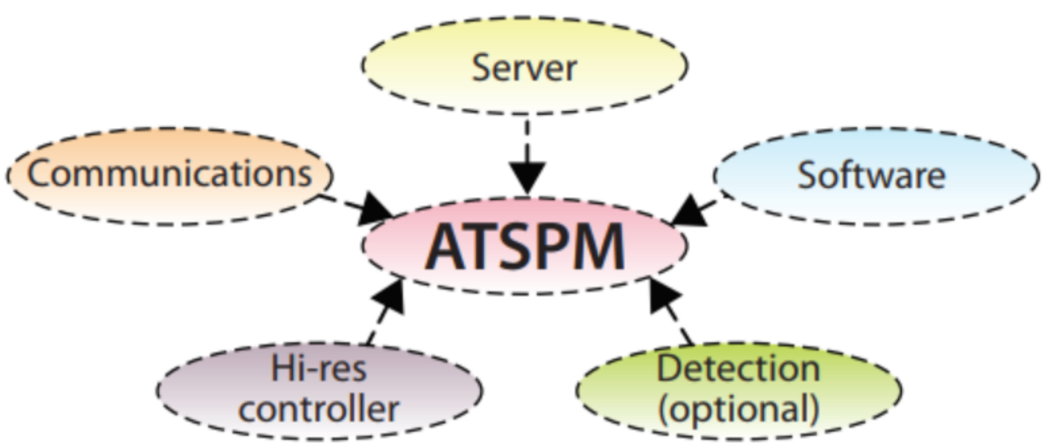


Figure 1-3 Five Basic ATSPM System Components (USDOT, 2020)

Conventional application approaches of ATSPMs are established on the collection and process of various data sources including radar data (Santiago-Chaparro et al., 2012), video surveillance data (Tang Hua et al., 2013), loop detector data (Shin-Ting Jeng et al., 2014), GPS data (P. Stuckmann et al., 2002; Christopher M. et al., 2008; Vaughan et al., 2013; Z. Arbi et al., 2017). However, the installation of loop detectors and radar devices can be costly, presenting a significant hurdle for widespread use. Meanwhile, GPS data, though widely available, may not provide sufficient resolution for certain applications, such as determining a vehicle's exact lane position.

Continuing along this trajectory, the emergence of new technology in real-time traffic performance measurement has been a significant milestone in the evolution of ATSPMs, which allows for instantaneous traffic condition evaluations in comparison with the historic traffic data (Zaigham Mahmood, 2020). Connected Vehicle (CV) technology, which involves vehicles that can communicate with each other and with road infrastructure, has unlocked new possibilities in this regard. CV data can be leveraged to derive real-time performance measures, providing a more timely and dynamic understanding of traffic conditions. More importantly, CV data can offer critical, high-precision information such as the exact positioning of a vehicle within a lane, its speed, direction, and heading. Furthermore, unlike loop detectors or radar systems, CV data collection has no requirement for pre-installed infrastructure. Instead, it leverages real-time vehicle-to-vehicle communications, making it potentially more scalable and cost-effective.

This real-time functionality is particularly beneficial in situations that require swift action, such as the detection of non-recurrent congestion and traffic incidents. The ability to detect and respond to these events as they occur, rather than relying on historical trends, which can offer more granular insights into traffic operations, capturing the fine-scale variations in queue lengths (Peng Hao et al., 2014), space-mean speed (Ahsani et al., 2019), and vehicle density (Aljamal et al., 2020) that might be overlooked when using historical data alone. With the advances in CV technology, more measures are proposed. The arrival on the green is a depiction of vehicle arrival patterns at the movement of interest during green phase phases to identify the vehicle headway distributions (Howell Li et al., 2016). CV data can also enable the comprehensive validation scheme for arterial travel time estimations (Xuechi Zhang et al., 2019). Split failure is proposed to indicate the allocated green time for a specific movement at an intersection is inadequate to accommodate the actual traffic demand (Saldivar-Carranza et al., 2021). However, the reliability of such measures is yet to be explored as the

CV market penetration rates (CVMPRs) are unknown in the real-world practice of ATSPM applications in most circumstances.

Meanwhile, concerns exist as the penetration rate of CVs has significant influences on the accuracy of such performance results, which has long been a subject of practice focus (Day et al., 2016, 2017). Recent research has revealed that the median CVMPR was approximately 4.5% according to a study on the state highway performance monitoring systems in Indiana, Ohio and Pennsylvania (Margaret Hunter et al., 2021). Therefore, the low CVMPR will remain to be an important factor influencing the reliability of ATSPM results.

To enhance the accuracy and reliability of non-recurrent congestion identification, besides penetration rates, careful consideration of optimal spatial and temporal data resolutions becomes crucial. While some researchers have explored data resolutions in performance measure estimations (Juan Argote-Cabañero et al., 2015), their investigations were limited to a specific analysis time interval and spatial aggregation, lacking a comprehensive understanding of the overall impact of data resolution. Moreover, the influence on the results of applying ATSPM derived from CV data to congestion detections has not yet been explored with optimized data resolutions.

This research is motivated by the opportunity to integrate the new form of data from the emerging technology, namely CV, into ATSPM, focusing on some of the critical issues such as the accuracy and reliability of the derived performance measures, the effect of CVMPR, and optimal spatial and temporal aggregation of data.

1.3 Research Problem

Presently, a considerable amount of research is being conducted to investigate the viability of utilizing CV data for traffic performance measurements and state monitoring. This topic is elaborated upon in the subsequent chapter. However, there are still existing gaps in the following areas that necessitate additional exploration.

1. Accuracy and reliability of performance measurements

The measurements derived from CV data have the potential to estimate traffic performance, such as the estimations on vehicle space-mean speed, travel time and split failure (Howell Li et al., 2016, Xuechi Zhang et al., 2019, Saldivar-Carranza et al., 2021). However, the accuracy and reliability of these measurements remain to be illustrated as the real-world CVMPR is unknown (Saldivar-

Carranza et al., 2021), and the levels of congestion are limited (Howell Li et al., 2016, Xuechi Zhang et al., 2019). Further research and investigations are required to establish the validity of these estimates and their applicability in a broader range of traffic conditions and CVMPRs.

2. Effect of CVMPR on the performance measurements

The influence exerted by different CVMPRs is not fully understood in the reliability analysis of performance measurement results (Argote-Cabañero et al., 2015; Day et al., 2016, 2017), especially the low penetration rates. Specifically, the minimum requirements of CVMPR for different performance measurements are yet to be explored to attain accurate and reliable performance results.

3. Optimal spatial and temporal data aggregation resolutions

Determining the optimal resolutions for aggregating CV data is pivotal for accurate performance measurements. Essential factors, including suitable temporal and spatial consolidations for various scenarios, play a vital role in accuracy and reliability analysis. Presently, the scope of aggregation levels explored in current research remains limited (Argote-Cabañero et al., 2015), highlighting the potential for further enhancement of estimation accuracy.

4. Feasibility of detecting non-recurrent congestion

The application of CV data in non-recurrent congestion detection needs to be examined for the latency in evaluating the real-time traffic state. At present, the primary emphasis of congestion detection predominantly centers on identifying recurrent congestion patterns, such as traffic density estimations (Gayah et al., 2013), instances of intersection overflow (Hiribarren et al., 2014) and vehicle arrival depictions (Waddell et al., 2020). However, there exists a broader spectrum of congestion scenarios and underlying causes, ranging from event-driven congestion to unexpected incidents. The feasibility of detecting such congestions through performance measures has not yet been explored with the application of CV data.

1.4 Research Objectives

In light of the research gaps in the field of urban traffic performance management and monitoring based on Connected Vehicle technology, the main objective of this thesis is to address some of the critical issues for supporting Automated Traffic Signal Performance Measures (ATSPMs). Therefore, the specific objectives of this research are as follows:

1. To explore the potential of CV data in obtaining various traffic performance measurements, including the average control delay, percentile queue length, percentage of stopped vehicles and average number of stops, which contains the accuracy and reliability assessment of such measures under different scenarios.
2. To investigate the feasibility of real-time non-recurrent congestion detections using CV data and explore the latency in the detection process and implications for traffic state monitoring.
3. To identify the optimal spatial and temporal data aggregation for CV data application in the context of ATSPMs, considering external factors such as congestion level and CV market penetration rates (CVMPR).
4. To evaluate the effect of CVMPR based on the specific performance measurements derived from CV data in terms of its accuracy and determine the minimum requirements for CVMPR to achieve an accurate and reliable estimation of the proposed measurements.

By addressing the above-mentioned objectives, the research aims to contribute to the current knowledge on the application of CV data in urban traffic performance measurements and state monitoring, which can present insights for practitioners and policymakers in the design and implementation of ATSPM tools. Moreover, the findings of this thesis can lay the groundwork for future research in the traffic engineering field, focusing on unlocking the potential of CV technology as well as addressing the challenges in the engineering practice.

1.5 Thesis Structure

this thesis is structured and organized into five chapters. The basic summary of each chapter is listed below.

Chapter 1: Introduction

This chapter summarizes the synopsis of the research background, as well as an introduction to the Automated Traffic Signal Performance Measures (ATSPMs) and real-time traffic monitoring and management systems. The research problems and gaps, research objectives and thesis structure are presented in this chapter.

Chapter 2: Literature Review

In this chapter, the applications of CV data in traffic performance measurement and state monitoring are discussed through an overview of ATSPMs. Meanwhile, the existing research on CV data for traffic performance measurements is discussed to identify the research gaps and limitations existing in the current literature.

Chapter 3: Methodology

The methodology chapter is outlined for explanations of the proposed framework including procedures of data collection and data preprocessing. The calculation of proposed performance measurements derived from CV data and the error evaluation techniques are detailed as well.

Chapter 4: Results and Analysis

This chapter provides the accuracy and reliability analysis results based on the proposed performance measurements under the influence of CVMPR, ranging from 1% to 25%. Moreover, various spatial and temporal aggregations are applied to the investigation in the context of ATSPMs with time intervals of 5min, 10min, 20min, 30min and 60min and study subjects of movements, approaches, intersections, and the corridor. The sensitivity analysis is also conducted to evaluate the impact of different levels of congestion.

Chapter 5: Discussion, Conclusion, and Future Work

The final chapter discusses the major findings and contributions of this research, including the possible implications for practitioners and policymakers. Furthermore, the study limitations and research recommendations are presented, which focus on the prospective applications of CV data and the challenges associated with its widespread adoption in the field of traffic engineering.

Chapter 2
Literature Review

In this chapter, a comprehensive literature review and theoretical background are presented for the research conducted in this study. The subsequent sections delve into the application of Connected Vehicle (CV) data in the context of Automated Traffic Signal Performance Measures (ATSPMs), highlighting the challenges associated with applying CV data for traffic management. Firstly, the attempts to examine the influence of CV market penetration rates (CVMPRs) are discussed. Following this, an overview of the existing methodologies and techniques for assessing the accuracy and reliability of ATSPMs derived from CV data is provided, including the effects of CVMPRs and different levels of data aggregation. Additionally, the literature related to congestion detection approaches based on CV data is listed. This chapter ends with a summary of research gaps, setting the stage for the methodology and findings presented in the subsequent chapters.

2.1 Studies on The ATSPM Derived from CV Data

In recent years, a significant number of studies have investigated the potential of applying CV data in the context of ATSPMs based on the CV data collected from the real world. However, these studies often rely on specific assumptions that need further refinement to be effectively generalized for real-world applications. Moreover, the reliability of such performance measure estimation results remains unknown as the CVMPR is undetectable on most occasions. Table 2-1 summarizes the performance measures proposed by researchers, which are derived from CV data. The definitions of each measure and the limitations of their studies are discussed in the table.

Beyond the studies mentioned in the table, additional endeavours have emerged in the exploration of harnessing CV trajectories for performance measurement purposes. Zhang et al. (2019) developed a comprehensive validation scheme for measuring arterial travel time using GPS trajectories and Bluetooth data as two independent sources. While their approach is centred on travel time, it unveils potential for further exploration and expansion into broader performance measures. Khadka et al. (2022) presented traffic performance metrics with queue length and propagation, vehicle's actual delay time on arterials and time-space diagram (TSD) on arterials based on the Internet-Connected Vehicle (ICV) data. Despite encompassing data from approximately 10%-15% of all vehicles in motion, concerns about the data's representativeness persist due to the unknown nature of the real-world CVMPR. Lloret-Batlle et al. (2023) utilized the trajectory data to estimate the arrival speed and traffic volumes for both undersaturated and oversaturated signalized intersections. However, their simulation primarily focused on scenarios with medium to high volume-to-capacity ratios, and the undersaturated scenario was disregarded. Wang et al. (2023) proposed algorithms for calculating different mobility measures indices including vehicle delay, number of stops, space-mean speed, and coordination measurements to evaluate urban traffic performance. However, their research was conducted in the real world as well without the exploration of the effect of CVMPRs.

Table 2-1 Summary of Proposed Performance Measures in Literature

Author	Year	Measures	Definition	Limitations
Argote-Cabañero et al.	2015	average speed	the average speed of all CVs for all lanes in the observed direction	The influence of CVMPRs, traffic scenarios and data resolutions has not been fully explored.
		average number of stops	the average number of times that CVs traveling in the direction of interest have to halt forward motions	
		average acceleration noise	the standard deviation of a CV's acceleration measurements along its trajectory	
		average delay per unit distance	a measure of the difference between the sampled average pace and the free-flow pace	
Day et al.	2016, 2017	control delay	the difference between the time a CV arrives at the intersection and the time it can proceed through the intersection	The study omitted the representation of overflow queues, and the effectiveness of their proposed methodology was examined under highly constrained data availability conditions, which requires at least 2 weeks to collect data under the CVMPR of 0.08%.
		travel time	the duration it takes for a vehicle to travel between two specific points or along a certain route	
		green occupancy ratio	the proportion of time that a traffic signal phase displays a green signal within a specific period, typically the signal cycle length	
		maximum queue length	the longest line of vehicles that has formed and come to a complete stop on a roadway measured by the CV trajectory	
Li et al.	2016	percent on green	a measure of how many vehicles pass through an intersection during the green phase of a traffic signal	The challenges arise as the limited real-world CVMPR falling between 0.6% and 3%.
		purdue coordination diagram	a visual tool used to analyze the coordination and effectiveness of traffic signals along a corridor	

		arrival flow profile	a graphic illustration of the rate at which vehicles arrive at a certain point over time	
Saldivar-Carranza et al.	2021	split failure	a reference of the inability of a signal's green phase to serve all queued vehicles	The research has not yet revealed the real-world CVMPR and its associated impacts, leaving the reliability of the performance measures in a state of uncertainty.
		downstream blockage	a condition where vehicles cannot proceed through an intersection despite a green signal due to congestion downstream	
		quality of progression	a measure of the effectiveness of signal coordination in allowing vehicles to move through multiple intersections without stopping	

Given that gaining a complete and accurate understanding of the present CVMPR in real-world settings proves to be a challenging task, a simulation-based study can provide insight for applying performance measures under different CVMPRs as well as examining the effectiveness of proposed measures to estimate the traffic performance within a near real-time context, which has ample potential to be explored. As such, the accuracy and reliability of these measures often remain elusive, leaving gaps in the current knowledge base. Therefore, it becomes critical to conduct studies that delve into these areas, investigating the fidelity of these measures.

2.2 Studies on The Effect of CVMPR

Recently, a multitude of research initiatives have attempted to examine the effect of CVMPRs on traffic performance measurements. Recently, researchers have explored a wide array of methodologies and models to better understand how different CVMPRs influence performance measurements and condition monitoring. However, limitations still exist in the literature as discussed below.

Much research has narrowed the focus to one single measure in their studies when examining the influence of CVMPRs. Comert and Cetin (2009, 2011, 2013) have developed various analytical models for cycle-by-cycle queue length estimation to examine the estimation errors under different CVMPRs with a dataset retrieved from CV trajectory through a simulation study. The proposed methodology is validated solely with consideration for the overflow queue. Peng Hao et al. (2014) have constructed Bayesian Network (BN) models for queue length estimation and explored the result accuracy under specific CVMPR. They have concluded that the Mean Absolute Error (MAE) is 2.8

when the CVMPR is 5% based on the proposed methodology. However, their approach required travel time information to identify the traffic scenario before the queue estimation procedure. Bucknell et al. (2014) conducted an analysis of the influence of CVMPR on the accuracy of traffic flow estimation based on the travel time information. They have managed to figure out the optimal penetration rate and the minimum requirement of sampling frequency under the current CVMPRs, while neglecting a broader range of potential traffic scenarios. Liberis et al. (2016) conducted the estimation of traffic states based on the CV data to describe the dynamics of the percentage of CVs and employed a Kalman filter to determine the CVMPR. However, the methodology required the densities and flows of the CVs as known information at the studied location. Zheng et al. (2017) developed an approach to estimate traffic volume using GPS trajectory data from CV or navigation devices under low market penetration rates, while the investigation on the effect of different CVMPRs was limited. Gao et al. (2019) proposed a queue length sensing model through CV data, which consists of two sub-models based on shockwave sensing and backpropagation (BP) neural network sensing under different CVMPRs. However, the queue length sensing methodology required data from roadside units and the corresponding traffic volumes of the studied road network in the simulation experiment were not illustrated. Tang et al. (2020) proposed a non-parametric model based on Bayes' theorem and a resampling process to predict short-term urban link travel time under different CV market penetration rates. However, the efficiency of the proposed methodology was not proved in the undersaturated scenario and under the CVMPR of 25%. Majstorović et al. (2022) explored how the CV data can be used for the creation of Speed Transition Matrices (STMs) at an isolated intersection and analyzed the impacts of CVMPRs on the accuracy of the created STM. However, their research was specific to an isolated signalized intersection and limited traffic scenario, omitting performance measures at alternate spatial levels. Chen et al. (2022) developed a Connected Autonomous Vehicle (CAV) data-based trajectory reconstruction method for freeway traffic state estimations and validated the proposed method under different traffic densities and penetration rates of CAVs. However, their approach required assumptions on the location of non-CAVs based on the detection of CAVs, leading to estimation errors, particularly in undersaturated scenarios.

Some research has succeeded in constructing performance metrics. Argote-Cabañero et al. (2015) presented estimation methods for performance measures, followed by an initial evaluation of different penetration rates' impact on the estimation results in undersaturated and oversaturated conditions based on the 15-minute time interval and movement-level analysis. However, the result was restricted to the extreme traffic scenarios and the data resolution was limited.

In summary, the question of the effect of CVMPR on the accuracy and reliability of various performance measures has yet to be fully answered. Specifically, the minimum requirement of CVMPRs across diverse traffic scenarios and data resolutions remains further explorations for various traffic performance measures.

2.3 Studies on The Congestion Detections Based on CV Data

Recently, the detection of congestions based on the CV trajectory data has also been a focused topic in the traffic management area. Various approaches have been proposed to detect congestion based on estimations of queue length, traffic density, arrival profiles and travel time. However, these approaches can be challenging to implement, which require information from graphic profiles instead of providing direct estimations from trajectory data. The detection of different congestion types and the exploration of utilizing more measures derived from CV data are not yet fully understood.

Liu et al. (2009) have proposed the real-time queue length estimation method for congested signalized intersections based on high-resolution data, while error exists when the arrival traffic is at a saturated flow rate. Argote-Cabañero et al. (2011) utilized the time-space diagram to determine the traffic conditions on urban signalized arterials for real-time applications. However, the variability in traffic conditions is not considered in their research. Gayah et al. (2013) presented a method to estimate the average vehicle densities in real-time, which was shown to be accurate only when the network is congested. Hiribarren et al. (2014) proposed a method to capture the traffic dynamics including traffic density and flow and predict the queue length at intersections and travel times along a road section at both congested and uncongested conditions based on trajectory data. However, the result is attained from limited scenarios and the effect of the CV market penetration rate is not examined. Hao et al. (2015) estimated the queue profile and maximum queue length of one cycle based on the short vehicle trajectories. However, limitations exist as the method is not validated under different traffic conditions. Wan et al. (2016) proposed an approach to reconstruct trajectories using sparse probe vehicle data with estimations on the travel time to small segments for congestion identification purposes, while the approaches rely on historic data and the statistics of estimated travel time require validation before practice.

Generally, there is a need to explore a broader range of traffic scenarios and performance measurements to assess the feasibility of congestion detection more thoroughly. Remias et al. (2018) utilized the automated vehicle location data consisting of GPS-based vehicle trajectories to measure signal performance and coordination at urban corridors, as well as demonstrate the feasibility of identifying signal coordination issues. Ahsani et al. (2019) proposed a statistical analysis of speed data bias and investigated the accuracy and reliability of INRIX for congestion detection purposes. Cao et al. (2019) developed an improved queue length estimation method for vehicles in the left-turn bay based on a general queue length estimation method with CVs to detect congestion. Brennan et al. (2019) proposed the performance index of travel time inflation to characterize the corridor travel time and delay based on probe vehicle data. Zhao et al. (2019) proposed a novel method for scaling up the number of probe vehicles to estimate the queue length and traffic volumes at signalized intersections. The detection is conducted through the observation of the distribution of the vehicle stopping positions at the intersections to indicate the possible signal failure. Luo et al. (2019) proposed an analytical method for traffic flow estimation in urban arterial corridors to capture the vehicle delay and queuing dynamics based on CV trajectories collected through V2C communication. Waddell et al. (2020) quantified signal performance on 19 intersections in Michigan and Ohio to scale a performance assessment using crowdsourced trajectory data, which focused on capturing the vehicle arrival and departure to detect the congestions through signal effectiveness evaluations. However, these studies primarily focus on congestion stemming from signal timings. Yet in real-world scenarios, non-recurrent congestion—often triggered by events such as capacity loss and large-scale happenings—is also a common occurrence. This aspect merits equal consideration in our approach to effective traffic management.

2.4 Summary

From the prior section, most research is established on the reconstruction and analysis of real-world trajectories, while the CVMPR in the real world is unknown. Moreover, many existing studies have primarily focused on a single measure, rather than establishing a more comprehensive set of metrics derived from CV data. While some studies have sought to incorporate a wide range of CVMPR in the reliability analysis of ATSPMs with simulation-based experiments, it becomes evident that analysis under 25% CVMPR deserves more concentrated focus. Further, the specific performance measures that can be obtained from CVs to depict traffic performance at signalized corridors need further exploration.

In traffic management practice, temporal and spatial aggregations are important factors for consideration in performance analysis. While most existing studies have acknowledged this to some extent, the detailed impacts of different levels of temporal and spatial aggregation on performance measurement and reliability are yet to be explored. For instance, further discussion is required on how different aggregation time intervals may influence the reliability of traffic performance measurements. In addition, the varying scenarios that might affect the impact of the CVMPR on result accuracy have not been comprehensively investigated.

At present, congestion detection predominantly relies on queue length and traffic flow data. However, the estimation methods are often bound by specific assumptions about vehicle arrivals. The influence of CVMPRs on the reliability of congestion detection outcomes remains unexplored. In addition, while much of the current research focuses on signal failure detection, non-recurrent congestion stemming from capacity loss or demand fluctuations also represents a widespread issue that merits due attention in real-world contexts. Apart from the gaps already discussed, the potential of using CV data for detecting the onset and end of traffic congestions, especially, non-recurrent ones including the congestions generated by capacity-constraint conditions and event-driven conditions, across various scenarios and different CVMPRs is an important aspect of traffic condition monitoring in urban areas.

Chapter 3
Methodology

This section describes the research methodologies employed in the study, with particular emphasis on the approach of deriving Automated Traffic Signal Performance Measures (ATSPMs) from simulated Connected Vehicle (CV) trajectories and the steps involved in the design of simulation-based experiments. Firstly, the studied corridor is introduced, including a description of the characteristics of simulated CV data, as well as the intersection layout and turning movement counts. The details of the simulation experiment are then followed upon, discussing the calibration of the model and the sampling procedure. Subsequently, this chapter provides detailed definitions of the traffic performance measures derived from simulated CV data and their estimation processes. Lastly, the sensitivity analysis is demonstrated to investigate the impacts of external conditions and factors.

3.1 Overview of the Proposed Methodology

Accuracy evaluation of the performance measurement system using real-world CV data is considerably challenging, if not impossible, due to the unknown ground truth about real-world traffic operations. To address this issue, this research proposes a simulation-based study in which controlled experiments can be conducted with simulated CV data under a wide range of road traffic conditions, signal control options, and specific levels of CV market penetration rate (CVMPR) as shown in Figure 3-1.

In the simulation study, random samples of vehicles are extracted from the simulated traffic (i.e., vehicle population) to represent the CV. The trajectories of the selected vehicles can be exported and used subsequently for calculating various performance measures such as control delay, 95th percentile queue length and percentage of stopped vehicles. The key procedure of this study includes:

1. Preparation of a case-study signalized arterial corridor: In this step, an urban arterial corridor will be selected for the simulation study, which should have specific characteristics relevant to the research objectives, such as a set of signalized intersections, representative intersection configurations, and varying levels of traffic congestion.

2. Development of a simulation model: a simulation model will be built and calibrated for the study site identified in the previous step. The model incorporates relevant traffic parameters, signal timings, and vehicle dynamics to ensure a realistic representation of real-world conditions.

3. Formulation of ATSPMs: With the simulation models in place, the appropriate traffic performance metrics that could be derived from the simulated CV data will be constructed for evaluating the traffic signal operations. These performance measures are designed to provide meaningful insights into the effectiveness of signal timing plans.

4. Conduct simulation experiments: Simulation experiments will be performed using the developed models under various alternative scenarios. The experiments should be constructed in such a way that would allow exploration of the impact of the various factors on ATSPMs and the overall performance of the traffic management system.

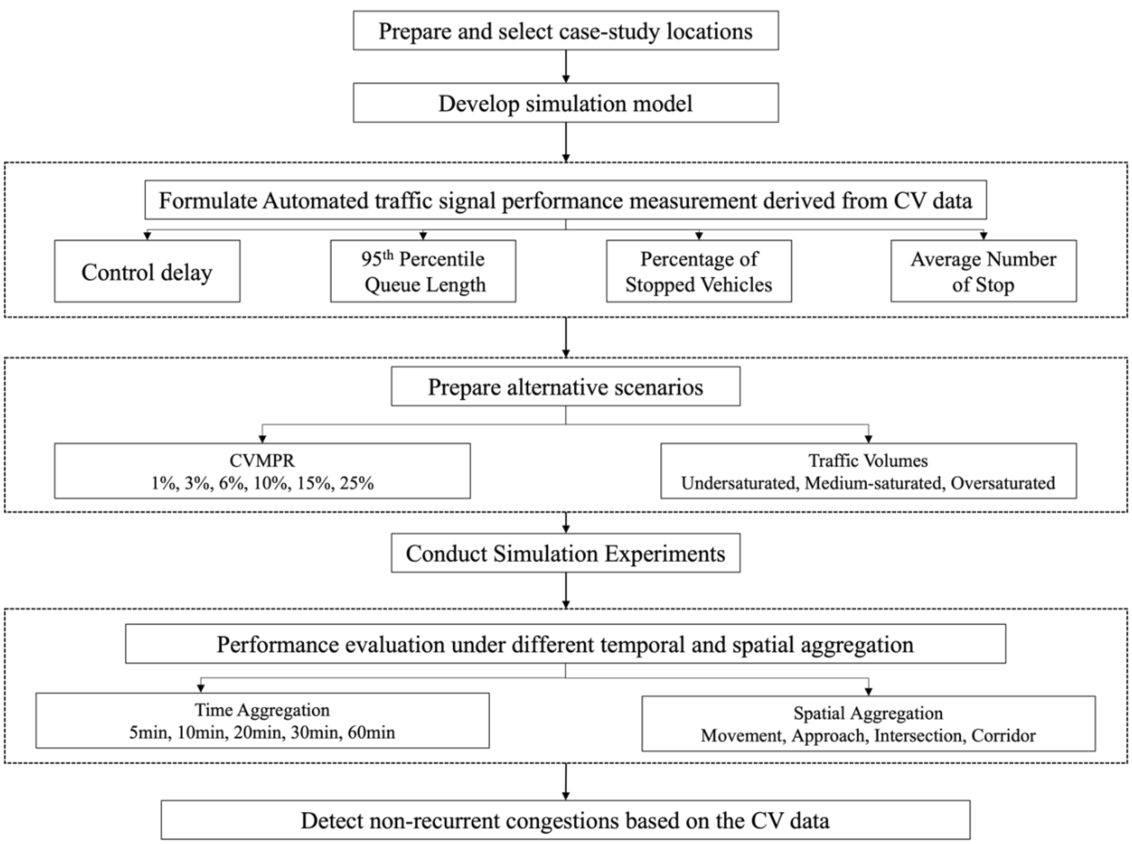


Figure 3-1 The General Framework of the Proposed Simulation-Based Study

5. Congestion detection: Non-recurrent congestions caused by special events will be simulated by introducing sudden influxes of demand at certain intersections along the corridor in the simulation model. The study will also investigate capacity loss due to events such as construction work and accidents.

This setup allows for the exploration of the feasibility of non-recurrent congestion detection based on CV data, under realistic traffic scenarios.

3.2 VISSIM Simulation Study

This section provides the details of the VISSIM simulation study, including the study area and the simulation scenarios. The study corridor is first introduced with an introduction of the study area, followed by basic information on traffic conditions. One critical intersection is identified for further analysis. Finally, the simulation model is demonstrated in this section with the model calibration process and vehicle sampling procedures.

3.2.1 The Broward Blvd

The urban arterial corridor, West Broward Blvd., located in Broward County, Florida, U.S., was selected for the simulation modelling and analysis of this research due to the availability of real-world CV data for this area and other traffic-related data such as TMC and signal timings. A VISSIM simulation model was constructed based on the model imported from Synchro software provided by the County, which is introduced in the following subsection. The geometric layout of the studied corridor is shown in Figure 3-2. A sample of the real-world CV trajectories is illustrated in Figure 3-

3.

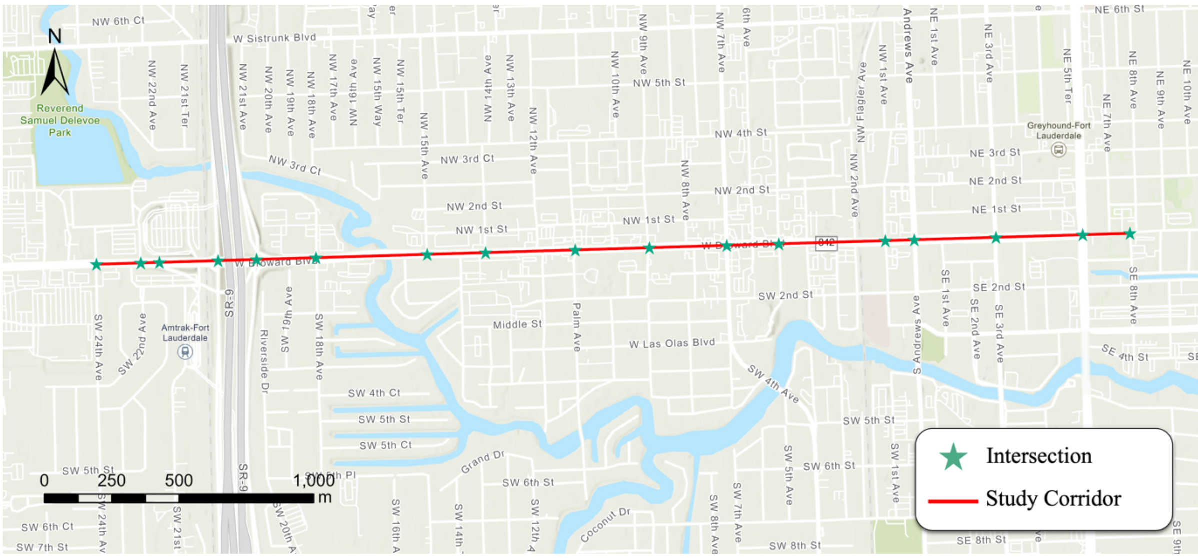


Figure 3-2 Overview of the Study Corridor

Figure 3-3 A Sample of Real-World Vehicle Trajectories

The Broward Boulevard corridor is a major thoroughfare located in the heart of Broward County, Florida. The study area holds numerous commercial buildings, residence locations and community service facilities, which generates a large amount of commuting demand every day.

To provide more detailed information about the intersection, the basic information of each intersection along the corridor is presented in the following section, including the intersection layout, volumes of different scenarios, and congestion levels.

3.2.2 Basic Information on Traffic Conditions

This West Broward Boulevard corridor consists of 17 intersections, as summarized in Table 3-1, including the minor street names and intersection types for each intersection. The traffic signal control is set as the fixed-time control with time-of-day plans. A common cycle length of 160 seconds is used for all intersections, which benefits the coordination along the arterial corridor. The signal timing plans and geometric layouts of each intersection are detailed in Appendix B and Appendix C respectively.

Table 3-1 Summary of the 17 Intersections

No.	Minor Street	Intersection Type
1	W 24th Ave.	Major to Minor
2	SW 22nd Ave.	Major to Major
3	Highway Ramp 1	Major to Ramp
4	Highway Ramp 2	Major to Ramp
5	Highway Ramp 3	Major to Ramp
6	W 18th Ave.	Major to Minor
7	W 15th Ave.	Major to Minor
8	SW 14th Ave.	Major to Minor
9	NW 11th Ave. / Palm Ave.	Major to Minor
10	W 9th Ave.	Major to Minor
11	W 7th Ave.	Major to Major
12	W 5th Ave.	Major to Minor
13	W 1st Ave.	Major to Minor
14	NS Andrews Ave.	Major to Major
15	NE 3rd Ave.	Major to Major
16	Federal Hwy 1	Major to Major
17	SE 8th Ave.	Major to Minor

Figure 3-4 depicts the Average Annual Daily Traffic (AADT) data of the Broward Blvd. provided by the county (Florida Department of Transport, 2023). The statistics include the AADT data of several major intersections along the corridor, spanning the years from 2018 to 2022. The detailed AADT information is demonstrated in Table 3-2. Based on the AADT data, it is apparent that traffic volumes experienced a slight decline after 2020 and have since stabilized starting in 2021.

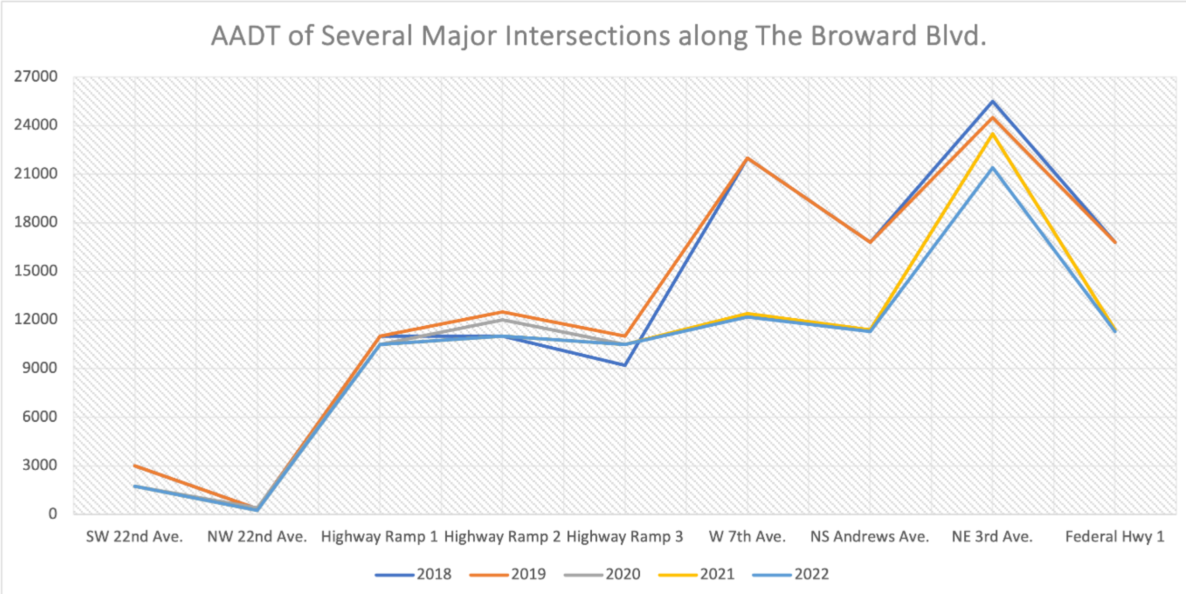


Figure 3-4 Eastbound AADT of Several Major Intersections along The Broward Blvd.

Table 3-2 Eastbound AADT of Several Major Intersections along The Broward Blvd.

Street Name.	2018	2019	2020	2021	2022
SW 22nd Ave.	3000	3000	1750	1750	1750
NW 22nd Ave.	350	350	400	250	250
Highway Ramp 1	11000	11000	10500	10500	10500
Highway Ramp 2	11000	12500	12000	11000	11000
Highway Ramp 3	9200	11000	10500	10500	10500
W 7th Ave.	22000	22000	12400	12400	12400
NS Andrews Ave.	16800	16800	11400	11400	11300
NE 3rd Ave.	25500	24500	23500	23500	21400
Federal Hwy 1	16800	16800	11400	11400	11300

3.3 Simulation Experiments

In this section, the design of simulation experiments is discussed for evaluating the performance of traffic signal operations using CV data. These experiments account for different facility types, congestion levels, and CVMPRs, which are critical factors influencing the accuracy and reliability of the derived traffic signal performance measures. The purpose of these experiments is to assess the effectiveness of performance measures derived from simulated CV data and identify optimal traffic management strategies under diverse conditions.

3.3.1 Simulation Experiment Design

The simulation model has been derived based on the depicted sketch, as presented in Figure 3-5. This model is then showcased in the corridor view.

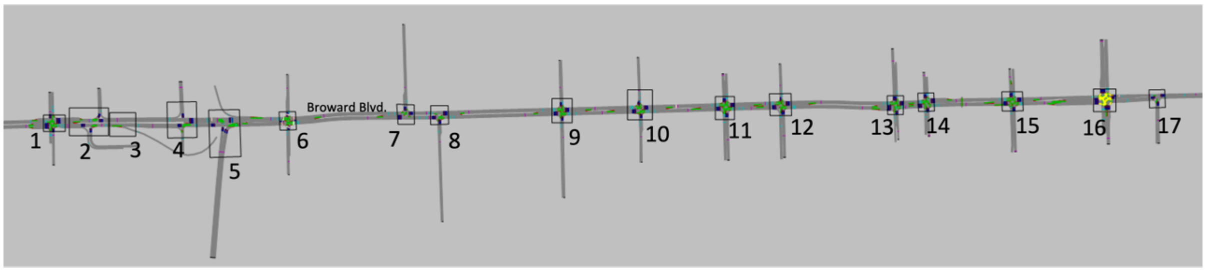


Figure 3-5 Corridor Layout in the VISSIM Simulation Model

To ensure a robust simulation experiment and reduce errors associated with varying simulation durations, the duration of each simulation run is kept consistent. In order to accurately simulate real-world traffic conditions throughout a typical weekday, the duration of each simulation scenario is set to 9000 seconds including 1800 seconds of warm-up period at the beginning of the simulation. Excluding the warm-up period, each simulation run is divided into three distinct 2-hour periods, each representing a different traffic scenario. These include the weekday off-peak, weekday midday and weekday peak, representing undersaturated, medium-saturated, and oversaturated scenarios, respectively. Therefore, the corridor overall input volumes are set as three individual levels, including 3 000 veh/h (weekday off-peak scenario), 6 000 veh/h (weekday midday scenario), and 10 000 veh/h (weekday off-peak scenario).

Specifically, study locations are chosen according to total hourly traffic volumes (f) and degrees of saturation (x) to conduct the performance estimations at the movement and the approach levels under each traffic scenario, which are detailed in Table 3-3 and Table 3-4 separately.

Table 3-3 The Traffic Volume of Different Scenarios at Movement Level

Scenario	Type	Location	Total Hourly Traffic Volume (veh/h) (f)	Degree of Saturation (x)
Off-peak	Undersaturated	SB right-turn of Intersection No.16	243	0.30
Midday	Medium-saturated	NB through-movement of Intersection No.11	350	0.62
Peak	Oversaturated	NB left-turn of Intersection No.16	250	1.05

Table 3-4 The Traffic Volume of Different Scenarios at Approach Level

Scenario	Type	Location	Total Hourly Traffic Volume (veh/h) (*f*)	Degree of Saturation (*x*)
Off-peak	Undersaturated	WB of Intersection No.9	1 188	0.32
Midday	Medium-saturated	EB of Intersection No.1	2 423	0.63
Peak	Oversaturated	EB of Intersection No.6	3 356	1.08

Among all intersections, the critical intersection for the eastbound direction is identified as the one of W. Broward Blvd. and the Federal Hwy. 1 (Intersection No.16). The distribution of traffic volume at this intersection displays significant variation across its approaches and movements as shown in Figure 3-5. Therefore, the intersection-level analysis is conducted at the selected study objective and its traffic volume information is listed in Table 3-5.

Given that the volume inputs for the entire corridor are different across each traffic scenario, the trajectories of each vehicle that traverses the eastbound and westbound directions through the corridor are collected as the subject of corridor-level analysis. And the total amount of such trajectories under each scenario is illustrated in Table 3-6.

Table 3-5 The Traffic Volume of Different Scenarios at the Intersection Level

Scenario	Type	Location	Total Hourly Traffic Volume (veh/h) (*f*)	Degree of Saturation (*x*)
Off-peak	Undersaturated	Intersection No.16	2 159	0.34
Midday	Medium-saturated		4 185	0.65
Peak	Oversaturated		6 761	1.05

Table 3-6 The Number of Trajectories Collected from the Corridor Level

Scenario	Type	Total Numbers of Trajectories (Eastbound) (vehs)	Total Numbers of Trajectories (Westbound) (vehs)	Degree of Saturation (*x*)
Off-peak	Undersaturated	813	764	0.33
Midday	Medium-saturated	1 455	1 510	0.64
Peak	Oversaturated	2 102	1 984	1.01

Therefore, the study location selected in this research is demonstrated at different levels of spatial aggregations.

3.3.2 Simulation Model Calibration

1. Turning volume

In order to accurately model the traffic flow in the simulation study, the turning volumes are calibrated using real-world vehicle trajectories. The turning movement counts for each intersection are assigned based on the data provided in the Synchro model.

Considering the inherent uncertainties associated with the real-world CVMPRs (Connected Vehicle Market Penetration Rates), the calibration of turning volumes is primarily based on the collected trajectories from different movements. This approach ensures that the simulation accurately reflects the observed traffic patterns and behaviour in the study area.

2. Other Parameters

Table 3-7 summarizes the settings of other essential parameters in the simulation model of the corridor, including the lane width, link behaviour, vehicle class, link speed, the velocity of reduced speed area and default random seed, which are validated through the simulation experiments.

Table 3-7 The Simulation Settings of Other Parameters

Parameter	Settings
Lane width	3.5m
Link behavior	Urban (Motorized)
Vehicle class	Car
Link speed	60 km/h
Reduced speed area	20 km/h
Default random seed	112

3. Number of runs

To measure the performance reliability across different times of runs, an assessment was conducted based on the system delay and corridor delay. The robustness of simulation runs is assessed by comparing the results to 100 separate simulation runs, encompassing run times of 20, 40, 60, 80, and 100 (Table 3-8). And the calculation is conducted based on the comparison of system delay and corridor delay between the values obtained from each simulation run setting and that from 100 run times. Equation 3-1 describes the calculation in detail.

Table 3-8 Number of Runs and Performance Reliability

Run times (n)	System delay (s)		Corridor delay (s)			
			Eastbound		Westbound	
	Value	*R*	Value	*R*	Value	*R*
20	3 287 170	-1.05%	244	-1.57%	195	1.68%
40	3 346 139	0.76%	250	0.98%	193	0.74%
60	3 314 833	-0.32%	249	0.34%	192	0.22%
80	3 339 884	0.51%	247	-0.73%	191	-0.53%
100	3 325 347	-	248	-	192	-

$$\boldsymbol{R} = \frac{Run_n - Run_{100}}{Run_{100}} \times 100\% \qquad \text{(Equation 3-1)}$$

Where, Run_n represents the system delay or corridor delay attained from total n run times; n=20, 40, 60, 80, 100 in this research. Run_{100} represents the system delay or the corridor delay attained from 100 run times.

Based on the table above, it can be observed that as the number of simulation runs increases to 20 or more, the delay values obtained for both the system and the corridor exhibit only minor variations compared to the results obtained from 100 simulation runs. Therefore, the number of runs is set at 60 times in this research to ensure robust and reliable results across different iterations.

3.3.3 Vehicle Sampling Procedure and Result Analysis Techniques

With the purpose of calculating the individual performance measure, all vehicle trajectories are exported from each simulation run to perform the CV sampling procedure, which includes two algorithms as discussed below. The identification of specific movement, approach and intersection is essential to investigate the optimal analysis interval and minimum CVMPR requirement for different spatial resolutions.

1. The CV Sampling Procedure

To study the effect of various CVMPRs on the reliability of different performance measures, the following sampling procedure is conducted to extract the desired amount of CV trajectories from the vehicle population:

Step 1 Denote the set of all vehicles in the simulation run as V and the total number of vehicles as $N_v = |V|$. Define the CVMPR, denoted as p. In this research, p = 1%, 3%, 6%, 10%, 15%, and 25%.

Step 2 Select a subset C of vehicles from V such that the number of vehicles in C, denoted N_c, is equal to $p \times N_v$. The selection is performed randomly to ensure a fair representation of all vehicles, where $N_c = p \times N_v$ and $C \subseteq V$. Therefore, the vehicles in C are treated as the connected vehicles.

Step 3 Export the trajectory data of each CV from the VISSIM simulation platform. The original CV data contains information including timestamp, vehicle positioning, link identification and lane identification as shown in Table 3-9.

Table 3-9 Key Output Variables from VISSIM CV Trajectory Data

Variables	Usage
Vehicle ID	Unique identifier for each vehicle.
Timestamp	The time when the data point was recorded.
Position	Indication of the distance that vehicle travels along its current link.
Lane ID/Link ID	Identifier for the lane or link (road segment) that the vehicle is on.
Vehicle Type	Type of the vehicle, such as car, truck, bus, etc.
Lane Index	The direction of the vehicle's movement at an intersection (left turn, right turn, straight ahead).

2. Identification of the Specific Movement, Approach, and Intersection

For each simulated CV, the trajectory T_C is expressed through the geometric identifications G and detailed travel information P. Let the intersection be defined by a set of lanes L, approaches A, and movements M. The intersection geometry is represented as $G = \{L, A, M\}$, where L is determined by the vehicle Lane ID, and A is determined by the link ID obtained from the original simulated CV report. For each CV, the sequence of data packets is represented as $P = \{p_1, p_2, ..., p_m ...\}$, where each data packet p_m contains the information on at the data logging instance m, that is, $p_m = (t_m, x_m, s_m, a_m)$, where x_m represents the vehicle position on the current link at t_m; s_m represents the vehicle speed at t_m; and a_m represents the vehicle acceleration at t_m.

To perform performance estimation at different spatial and temporal resolutions, the trajectory data T_C is organized in a structured format suitable for analysis across various temporal intervals (denoted as K). In this research, K is set to 5 minutes, 10 minutes, 15 minutes, 20 minutes, and 60 minutes. Additionally, different spatial aggregation levels, such as movement, approach, or intersection, are considered. Repeat the entire procedure N times, where N represents the number of simulations applied in this research. For each repetition, calculate the mean values of both the simulated population and the simulated CV samples and then compare them with the analysis technique as mentioned below.

3. Measure of Estimation Errors

To measure the difference between performance results attained at various CVMPRs and that calculated from the population, the Root Mean Square Error (RMSE) is introduced in this section as **Equation 3-2**.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - \widehat{M}_i)^2} \qquad \text{(Equation 3-2)}$$

where, $\widehat{M}_i$ represents the value attained from the estimation method. M_i represents the value of the entire population, $RMSE \in [0, +\infty)$. And the analysis is conducted based on the mean value attained from 60 simulation runs, $N = 60$ in this research.

3.4 Performance Measure Calculation using CV Data from Simulation

This section introduces the methods that are applied for determining the individual performance measures based on vehicle trajectory data. It starts with the methods for calculating the speed of a vehicle and identifying its state in terms of stopping, which are used in the later stage for determining other performance measures such as queue length and percentage of stopped vehicles.

3.4.1 Identification of Deceleration Starting Point

To identify the acceleration and deceleration points in the simulated CV trajectory data, a sequence of example CV reports collected from the simulation experiments is presented in Figure 3-6. In this figure, each CV trajectory is organized as the sequences of timestamps and locations attained from the individual reports in the simulation, which are denoted as “SIMSEC” and “POSITION” respectively. The speed and acceleration are derived from the timestamp and positioning in the simulated CV report. Thus, the identification hinges on the substantial shifts in vehicle acceleration.

To calculate the acceleration of a connected vehicle, its trajectory data is first organized based on the vehicle ID, positioning on the link of interest and the corresponding timestamps. The travel information of each simulated CV is stored in the data packet *P* as mentioned in the previous sections, which can be obtained from the change in distance and time for each consecutive pair of data points, as given in Equation 3-3 and Equation 3-4.

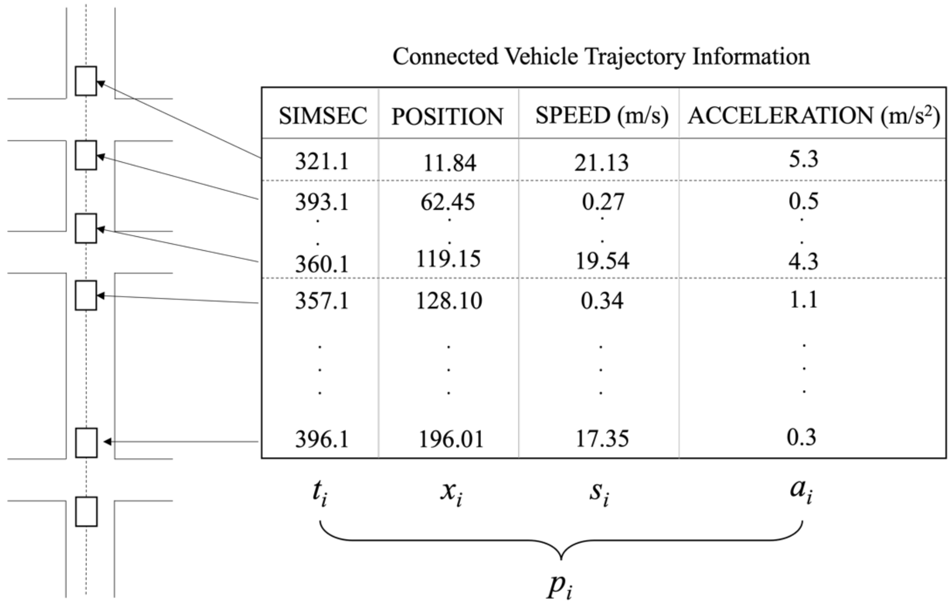


Figure 3-6 CV reports in the Simulation

$$s_i = \frac{x_{i+1} - x_i}{t_{i+1} - t_i} \qquad \text{(Equation 3-3)}$$

$$a_i = \frac{s_{i+1} - s_i}{t_{i+1} - t_i} \qquad \text{(Equation 3-4)}$$

Where, s_{i+1} and s_i represent the vehicle speed associated with the time instance i+1 and i respectively; a_{i+1} and a_i represent the vehicle acceleration at time instance i+1 and i; x_{i+1} and x_i represent the vehicle position at the time instance i+1 and i. Note that in real-world CV trajectory reports, x_i denotes the coordinate. In contrast, simulated trajectories offer relative distances on the link as vehicle positioning data. For clarity in subsequent sections, l_i will be used in place of x_i to represent this vehicle positioning information at time instance i.

The travel status of a vehicle (deceleration or acceleration) can be detected by assessing changes in vehicle acceleration over time. As the literature indicates (Quiroga et al., 1999), the identification of deceleration starting point is associated with the vehicle acceleration and speed in the original report as shown in Figure 3-7.

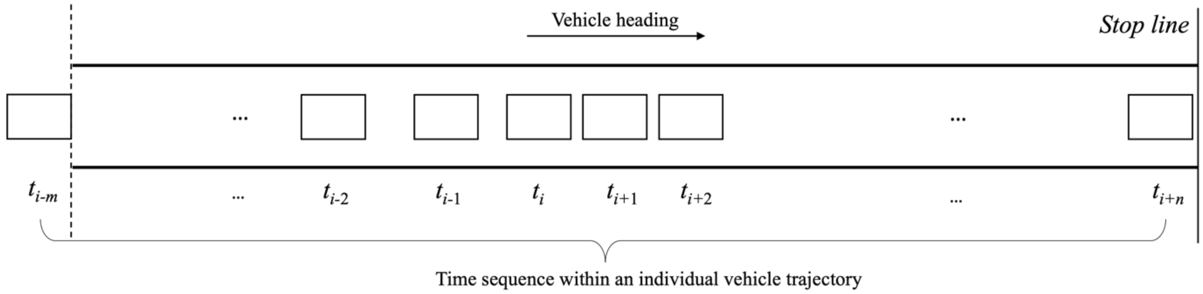


Figure 3-7 Time Sequence within An Individual Vehicle Trajectory

According to the time sequence in the figure above, m and n represent the number of timestamps before and after the considered point, respectively. the average acceleration in the forward direction is calculated using (Equation 3-5) for a sequence of timestamp pairs (t_{i-m}, t_{i-m+1}), …, (t_i, t_{i+1}), (t_i, t_{i+2}), (t_i, t_{i+3}), …, (t_i, t_{i+n}). And the average speed of the studied vehicle before the deceleration timestamp is calculated according to Equation 3-6 based on the sequence of timestamp pairs generated before (t_i, t_{i+1}). If the absolute value of the average acceleration exceeds the threshold and the absolute value of the average speed is also below the threshold, it indicates significant deceleration at timestamp t_i. In this case, t_i is added to the set $\{T^D\}$, indicating that a deceleration point is identified.

$$a_i^D = \begin{cases} a_i, \ if \frac{1}{n}\left|\sum_i^{i+n} a_i\right| > threshold \\ 0, \ if \frac{1}{n}\left|\sum_i^{i+n} a_i\right| \leq threshold \end{cases} \quad \text{(Equation 3-5)}$$

$$s_i^D = \begin{cases} s_i, \ if \frac{1}{m}\left|\sum_{i-m}^{i} s_i\right| > threshold \\ 0, \ if \frac{1}{m}\left|\sum_{i-m}^{i} s_i\right| \leq threshold \end{cases} \quad \text{(Equation 3-6)}$$

In idealized trajectories, vehicle deceleration points can be interpreted as stopping points directly. However, in real-world scenarios, it is more accurate to consider the beginning of deceleration as the stopping point. Therefore, when the vehicle's speed and deceleration meet the identification criteria, a deceleration point is established, indicating that the vehicle has come to a stop.

In the algorithm above, a threshold must be set in advance. Based on the findings in the literature, the threshold in Equation 3-5 is set as 2.0 m/s^2 (Quiroga et al., 1999). The threshold in speed is set as 5 km/h in association with vehicle speed (Argote-Cabañero et al., 2015). Given that the simulated CV data can capture the changes in the acceleration from consecutive points, the effective number of points is examined in the subsection of sensitivity analysis to make further explanation on the determination of the acceleration threshold.

3.4.2 Average Overall Delay

Delay is generally defined as the difference between the travel time that would have occurred in the absence of the intersection control, and the travel time that results because of the presence of the intersection control (National Academies of Sciences, 2022). This thesis focuses on the average movement delay derived from the simulated CV data, which can then be used to estimate the average approach and intersection delay. Following the literature (Quiroga et al., 1999), the calculation of average movement delay is conducted based on the position and timestamp information in the CV trajectory to identify the crucial timestamps as demonstrated in Figure 3-8. To simplify the explanation, one vehicle trajectory is analyzed to attain the overall delay.

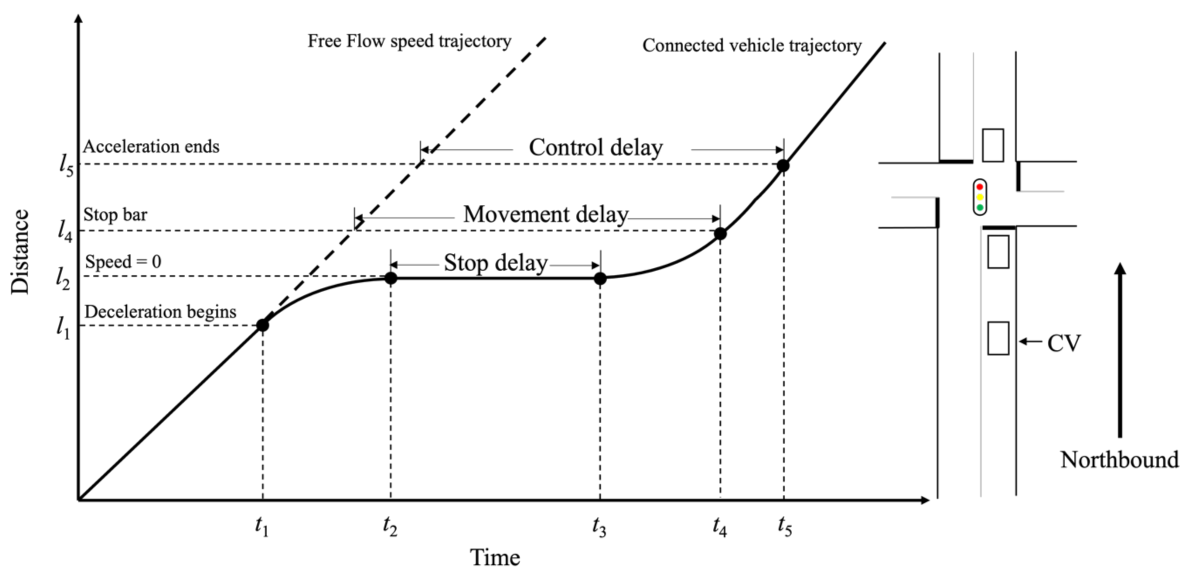


Figure 3-8 The Time-Distance Diagram with Delay Terms at The Signalized Intersection

In Figure 3-8, all delay terms are shown with the difference between the actual trajectory and the free flow speed trajectory. Generally, the movement delay can be calculated for one single vehicle trajectory as Equation 3-7.

$$D_{movement} = t_4 - t_1 - \frac{l_4 - l_1}{s_f} \quad \text{(Equation 3-7)}$$

Where, t_4 represents the timestamp that the vehicle passes the stop bar; t_1 represents the timestamp that the vehicle starts to decelerate; l_4 represents the position that the vehicle passes the stop bar; l_1 represents the position that the vehicle starts to decelerate; S_f represents the free flow travel speed on the movement of interest. Note that the l_4 and l_1 are obtained based on the position in the CV report.

The average overall delay used in this research is obtained from simulated CV samples to estimate the delay experienced by the vehicle population at the specific movement of interest within the defined time intervals, which is attained from Equation 3-8.

$$\overline{D}_{movement} = \frac{1}{N_C} \sum_{i=1}^{N_C} D_{movement} \qquad \text{(Equation 3-8)}$$

Where N_C represents the total number of CV trajectories in the sample. As the free flow speed usually varies among individual vehicles, an examination should be conducted to determine the vehicle free-flow speed. Therefore, the speed of a CV is calculated based on the average speed of the studied vehicle before the t_i as shown in the following calculations (Equation 3-9).

$$S_f = \frac{1}{m} \times \sum_{i-m}^{i} s_i \qquad \text{(Equation 3-9)}$$

Through the above-mentioned procedures, the calculations of average overall delay are demonstrated based on the individual trajectories.

3.4.3 95th Percentile Queue Length

The 95th percentile queue length is defined as the queue length (usually measured in vehicles or feet per meter) below which 95% of all observed queue length falls (National Academies of Sciences, 2022). In this research, calculating the 95th percentile queue length from a sample of CV data is conducted based on the identification of vehicles in the queue.

In Figure 3-9, the deceleration points of each vehicle at the spatial level of interest can be identified through the methodology described in 3.4.1. Thus, the distance between the last CV in the queue of interest and the stop bar can be calculated through the identification of the deceleration point of the last vehicle in the queue. After applying the filter procedure to all the simulated CVs at the spatial level of interest, the 95th percentile queue length is obtained based on the observations of maximum queue length estimations.

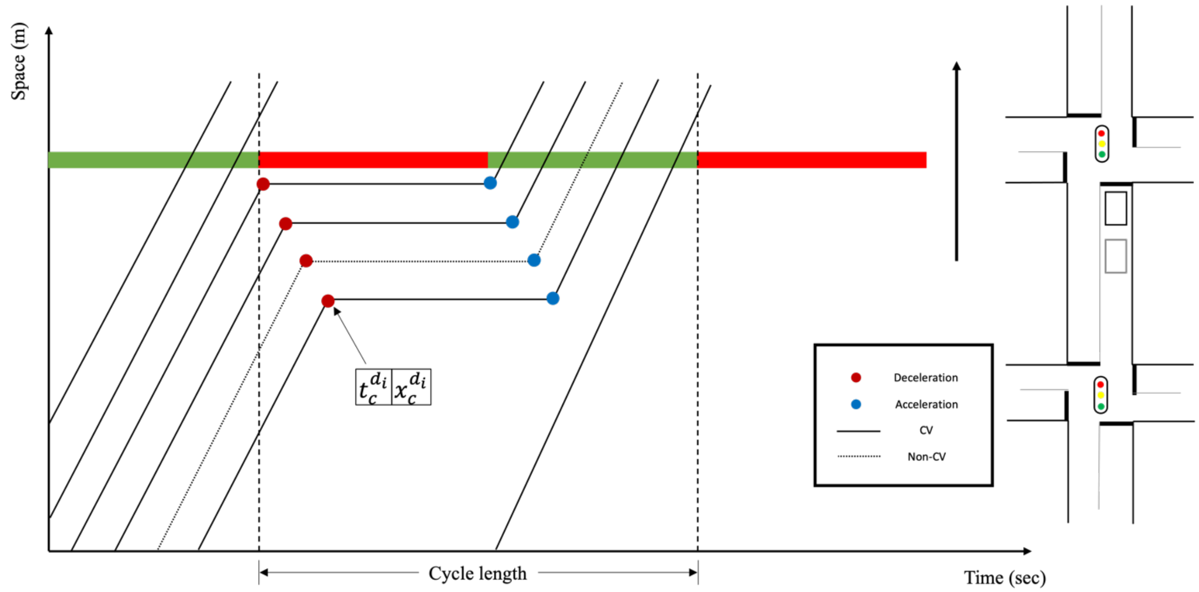


Figure 3-9 Identification of Acceleration and Deceleration Points in the Queue

Under the CVMPR of p%, the vehicle distribution follows the successive Bernoulli trial (Papoulis A., 1984), which can be represented by the vehicle geometric positions in the queue. Following this distribution, p can be considered as the probability of selecting a vehicle at a given position in the queue and the geometric distribution determines the number of vehicles between two consecutively sample vehicles, which can be expressed as Equation 3-10. Therefore, Equation 3-10 and Equation 3-11 are utilized to determine whether a queue is forming on the lane of interest based on the identification parameter F, which acts as a threshold based on the likelihood of identifying a vehicle and the jam density of the movement.

$$P(X = k) = (1 - p)^{k-1} p. \quad \text{(Equation 3-10)}$$

$$F = \max\{\frac{\mathrm{p}}{\rho_0}, \frac{1}{\rho_0}\} \quad \text{(Equation 3-11)}$$

Where, X is a random variable representing the number of CVs in the queue. ρ_0 is the jam density of the studied movement.

Thus, the deceleration points of each CV trajectory are identified in the time sequence as shown in Equation 3-12. And the CVs involved in the queue are recognized as well.

$$S = \{c \mid d_i = |l_{i+1}t_{i+1}^d - l_i t_i^d| > F\} \quad \text{(Equation 3-12)}$$

Where, S represents the time sequence in the filtering process based on the deceleration point consists of elements c where the change in position (or deceleration) between two consecutive times exceeds the threshold F; d_i represents the change in position or difference between two consecutive positions of a vehicle, which might be indicative of a deceleration or a stop; l_{i+1} and l_i represent the positions of a vehicle at two consecutive times, with l_{i+1} being the position at a later time and l_i at an earlier time; t_{i+1}^d and t_i^d represents the specific time instances related to the deceleration or change in position of the vehicle; $|l_{i+1}t_{i+1}^d - l_i t_i^d|$ calculates the difference between two consecutive positions of a vehicle, factoring in the associated times. The result is a measure of how much the vehicle has moved or decelerated between these two times.

In Equation 3-12, the position of the last detected control vehicle (CV) in the lane of interest is utilized to determine whether the queue being formed extends beyond the current lane. This determination depends on the identification parameter defined by both the CV occurrence probability (attained from Equation 3-10) and the jam density described in Equation 3-11.

After applying the filtering process to the trajectories of interest in the sample, all the vehicles in the queue are identified as:

$$Z = \begin{cases} 0 & , \ S = \{\emptyset\} \\ \{1,2,\dots,c\}, & otehrwise \end{cases} \quad \text{(Equation 3-13)}$$

Where Z represents the set of vehicles is observed in the queue at the spatial level of interest. Note that $\{T^D\}$ is the subset of $\{Z\}$ and is obtained from the deceleration identification procedure and c represents the last CV in the queue, which is identified from the sequence of deceleration point in $\{Z\}$.

Then, the 95th percentile queue length is calculated based on the sorted maximum queue length. Note that for idealized trajectories as shown in the figure above, vehicle deceleration points are identified as stopping point directly. However, in the practice, the start of deceleration should be considered as the stopping point.

The maximum queue length estimation is applied in this paper to attain the 95th percentile queue length as in (16).

$$\hat{L}_Q = |x^s - x_c^{d_i}| \quad \text{(Equation 3-14)}$$

Where, x^s is the location of the stop bar, which is measured in meters per lane. $x_c^{d_i}$ is the last CV in the queue of interest.

Therefore, sorting the $\{L_{Q_k}\}$ in ascending order within the time analysis interval of interest, the 95th percentile queue length is calculated as (17).

$$L_{Q_{95}} = L_Q[0.95 \times n] \qquad \text{(Equation 3-15)}$$

Then, the 95th percentile queue length is calculated based on the sorted queue length obtained from the movement and approach of interest. Furthermore, the vehicle headway's value also influences the performance measure estimation, a factor examined in the analysis section to determine the appropriate standstill distance.

3.4.4 Percentage of Stopped Vehicles (PSV)

The Percentage of Stopped Vehicles (PSV) is intended to assess the efficiency of signal timing at intersections. This metric is of particular interest, as multiple stops experienced by a vehicle can signify potential cycle failure in the existing signal timing plan. Such a situation arises when the green signal duration is insufficient, causing vehicles to halt multiple times. From the set of simulated CV trajectories, the average number of stops is attained from the identification of vehicle decelerations. These stopped CVs are denoted as CV_i^{stop} and stored in the set of CV^{stop} as $\{CV^{stop}{}_1, CV^{stop}{}_2, \ldots, CV^{stop}{}_n\}$.

Therefore, the percentage of stopped vehicles is calculated as Equation 3-16 for the spatial and temporal aggregation levels of interest.

$$PSV = \begin{cases} \frac{\sum_{i=1}^{n} CV_i^{stop}}{\sum_{i=1}^{N_C} CV_i} \times 100\% & (N_C \neq \emptyset) \\ 0 & (N_C = \emptyset) \end{cases} \qquad \text{(Equation 3-16)}$$

Where, CV represents for the total number of CVs observed in the cycle of interest. In the context of this calculation, each CV_i^{stop} is counted as the individual connected vehicles that have experienced stopping within the temporal and spatial levels of interest by summarizing CV_i^{stop} to CV_n^{stop}. And the total number of connected vehicles under the same temporal and spatial level is calculated by summarizing CV_i to CV_N in the set of N_C as obtained from the previous subsections. Note that the CV^{stop} is a subset of N_C.

3.5 Sensitivity Analysis

The influence of different parameters on the performance results is evaluated through the examination of four aspects. These aspects encompass both system-wide and corridor-wide analysis, including system delay, system average speed, arterial westbound delay, arterial eastbound delay and the 95th percentile queue length of Intersection No. 7 Northbound. As the system average speed influences identification of vehicle stops, it is included as one of the aspects of analysis. The influence of the selected parameters is evaluated with the percentage of differences compared to the measure value attained from the default value. Initially, the sensitivity analysis focuses on the Wiedemann-74 Standstill distance as shown in Figure 3-10.

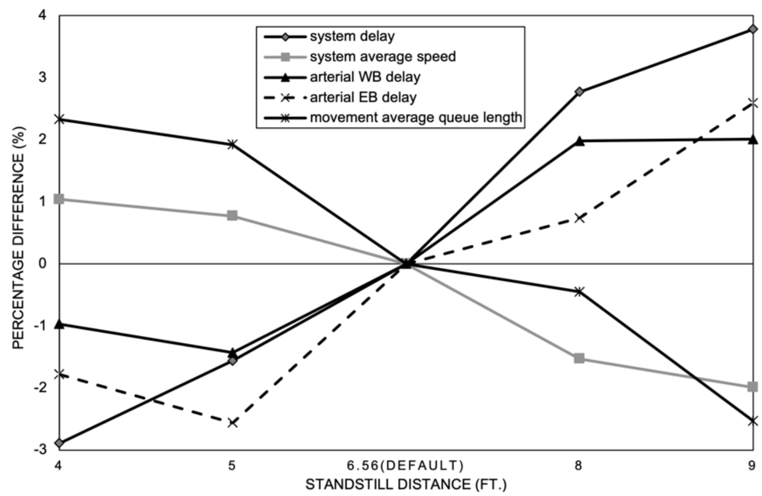
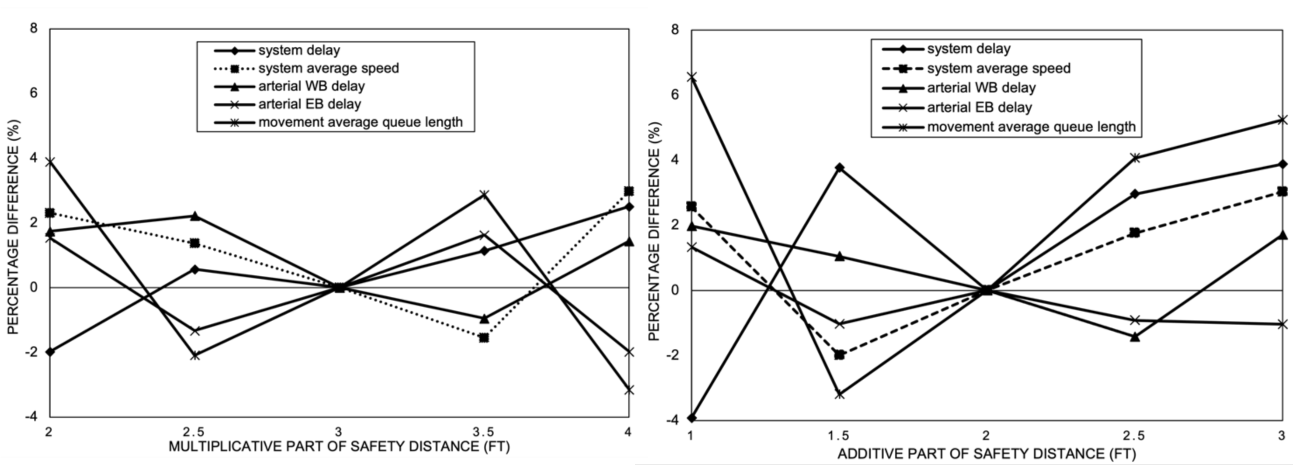


Figure 3-10 Impact of W-74 Standstill Distance

The standstill distance represents the desired average distance between vehicles as they form a queue in front of the stop bar. Generally, a higher value corresponds to a larger standstill distance and potentially lower capacity. In Figure 3-10, as the standstill distance varies from 4ft. to 6.56ft. (default value) and from 6.56ft. to 9 ft., the system delay, corridor-level delay and movement-level delay have increased significantly. Based on ground truth observations, the median distance between stationary cars (inter-car following distance) in America is 8.51 ft (Houchin et al., 2015). For the purposes of this research, this standstill distance is rounded to 8.5 ft. The sensitivity analysis suggests that as the standstill distance approaches the observed 8.5ft median value, the delays across the system tend to escalate, underlining the importance of considering real-world inter-car distances in traffic modeling.

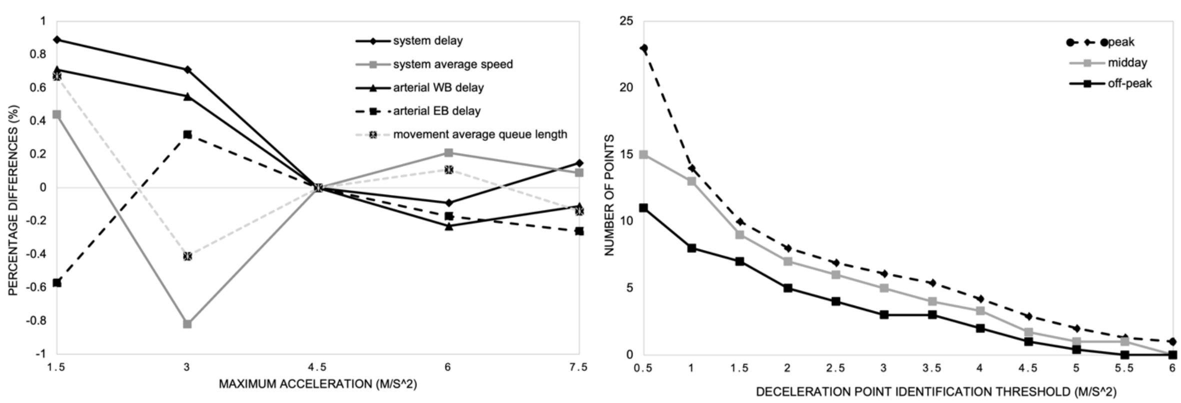
To evaluate the impact of safety distance, the additive part and multiplicative part of safety distance is assessed separately for the sensitivity analysis in Figure 3-11 (a) and Figure 3-11 (b). Figure 3-12 shows the impact of maximum deceleration (measured in m/s^2).

The default value of the additive part of the safety distance is initially set to 2 feet. However, when this value is increased to 3 feet, the average speed shows only minor variations. In contrast, the delay measure exhibits noticeable variations at different levels in response to the increased safety distance. Similar trends exist in the analysis of multiplicative part of safety distance when the default value of that is set to 3 feet.



(a) Multiplicative Part (b) Additive Part

Figure 3-11 Impact of W-74 Safety Distance (ft.)



(a)Maximum Deceleration (b)Identification Threshold

Figure 3-12 Impact of Acceleration Threshold Settings

When the value exceeds 4.5 m/s^2, the percentage difference becomes negligible. To reflect real-world traffic conditions, the additive safety distance is set to 2.5 ft for weekday off-peak scenarios and 2.0 ft for weekday peak and midday scenarios. These values are selected based on the analysis of percentage differences in delay, queue length, and vehicle speed compared to the default value. The chosen values of the additive safety distance ensure that the differences in performance measures are within 5% compared to the default setting. Similarly, the multiplicative part of the safety distance is set to 3.5 ft for weekday off-peak scenarios and 3.0 ft for weekday peak and midday scenarios.

According to literature (Quiroga et al., 1999), the acceleration threshold should be set between 2.0 m/s^2 and 3.5 m/s^2. Based on the findings in Figure 12(b), it can be observed that the rate of false positives for the deceleration point identification decreases as the threshold increases, reaching a minimum around the 2.0 m/s^2 mark. Beyond this point, there's a slight increase or plateau in the rate for all three traffic conditions (peak, midday, and off-peak).

Setting the threshold at 2.0 m/s^2 appears to be the optimal choice as it minimizes the rate of false positives across different traffic conditions. This ensures that the deceleration point identification is most accurate, with the least number of false detections, during peak, midday, and off-peak times.

Chapter 4
Analysis of Performance Measurement and Congestion Detection

This chapter presents results derived from simulation experiments conducted to investigate the reliability and accuracy of the performance estimations obtained from simulated Connected Vehicle (CV) trajectory data. In the following sections, the accuracy of different measures is analysed under a variety of traffic scenarios, Connected Vehicle Market Penetration Rates (CVMPRs), and temporal aggregations. The discussions on the reliability of different measures are extended based on the results from different levels of spatial aggregations, including movements, approaches, intersections, and corridors. The feasibility of detecting different types of congestion is lastly explored.

4.1 Performance Measure Accuracy Analysis

As discussed previously, the accuracy and reliability of a performance measure derived from CV data are generally a function of CVMPRs, spatial and temporal aggregation levels, and traffic conditions. On the other hand, the required accuracy and reliability on a performance measure depends on the application of each derived performance measure, ranging from real-time condition monitoring to performance evaluation for the purpose of intersection planning, design and control. For the former, high temporal granularity would be required, that is, conditions are expected to be measured at a time interval of as low as 5-10 minutes. However, the spatial resolution could be lower, e.g., at the approach and intersection levels. In contrast, for the later applications, the common time interval is 15 minutes as defined by HCM; however, the spatial resolution required would be higher, i.e., at the movement level. In the following section, the discussions are structured by four aspects: a) spatial aggregation levels (movement, approach, intersection and corridor) ; b) types of performance measures (overall delay, 95^{th} percentile, …); c) temporal aggregation levels or analysis time interval (5-10 min vs. 15 min); d) CVMPR (current state (~3%) vs. future states).

4.1.1 Results at the Movement Level

Figure 4-1 presents the accuracy analysis of the performance measure estimations derived from the simulated CV data at the movement level. Generally, a CVMPR of 3%-6% is essential for the estimations obtained from the movement level according to the requirements of different measures when the temporal aggregation is set to 10 minutes and below. For the average overall delay, a minimum temporal aggregation of 10 minutes is essential during weekday peak scenario when the

CVMPR reaches 3% and above. If the CVMPR is 6% and below, the analysis interval should be set to 5 minutes and longer. For the 95th percentile queue length, a 10-minute analysis interval is recommended for performance estimations at the movement level. The detailed analysis is discussed as follows.

- Average overall delay:
    - Under the weekday peak scenario (x = 1.05) with a 5-minute analysis interval, when the CVMPR increases from 3% to 6% and 10%, the Root Mean Square Error (RMSE) of the estimated average overall delay decreases from 6.83 s/veh to 6.37 s/veh and 5.63 s/veh, respectively, which correspond to the true overall delay of 68.82 s/veh. The corresponding relative errors are 9.81%, 9.23% and 8.16%, respectively.
    - The weekday midday scenario (x = 0.65) represents the moderately congested conditions with the true average overall delay of 44.33 s/veh. It was found that the temporal analysis interval should be 5 minutes at least to keep the relative estimation errors to 10% or lower. The corresponding RMSEs of the estimated average overall delay are 4.52 s/veh (6% CVMPR), 4.02 s/veh (10% CVMPR) and 3.57 s/veh (15% CVMPR).
    - Under the weekday off-peak scenario (x = 0.34) in which the true average overall delay is 22.31 s/veh, a 10-minute analysis interval is required to achieve a relative estimation error of 10% and below. Under the CVMPR and analysis interval of interest, the RMSE decrease from 2.17 s/veh (6% CVMPR) to 2.02 s/veh (10% CVMPR) and 1.53 s/veh (15% CVMPR). The corresponding relative errors are 9.64% (6% CVMPR), 9.08% (10% CVMPR) and 6.84% (15% CVMPR).
- 95th percentile queue length:
    - Under the weekday peak scenario (x = 1.05), the 95th percentile queue length attained from the population is 29.11 veh/lane with a 5-minute analysis interval. With the same interval, the errors vary from 4.65 veh/lane (6% CVMPR) to 3.59 veh/lane (10% CVMPR) and 3.13 veh/lane (15% CVMPR) with the relative errors below 10%.
    - Under the weekday midday scenario (x = 0.65), the actual value of queue length measure is 18.39 veh/lane when applying the temporal aggregation of 10 minutes. With a 10-minute analysis interval, the errors vary from 1.78 s/veh (6% CVMPR) to 1.35 s/veh

(10% CVMPR) and 1.28 s/veh (15% CVMPR). And the relative errors are calculated as 9.71% (6% CVMPR), 7.52% (10% CVMPR) and 6.91% (15% CVMPR).

- Under the weekday off-peak scenario (x = 0.34), as the CVMPR increases and the temporal aggregation extends, the estimation error has minor changes. However, with the analysis interval of 10 minutes, it is still possible to obtain reliable estimations of the relative errors are 10% and below when the CVMPR is 10%. The corresponding RMSE is 0.85 veh/lane in comparison with the actual value of 9.63 veh/lane and the relative error is 8.66% when the CVMPR climbs to 10% under the 10-minute interval.

- Percentage of stopped vehicles:
  - Under the weekday peak scenario (x = 1.05), there are 65.73% vehicles that are observed to have stopped before the stop bar when applying the 5-minute intervals when considering the entire population of interest. From the error evaluation results, when the CVMPR rises to 6%, this measure can provide estimation results with relative error below 10%. Under the 5-minute interval, the RMSEs are 6.82% (3% CVMPR), 6.09% (6% CVMPR) and 5.58% (10% CVMPR) respectively, and the corresponding relative errors are 10.32% (3% CVMPR), 9.29% (6% CVMPR) and 8.15% (10% CVMPR).
  - Under the weekday midday scenario (x = 0.65), after 3% CVMPR, the estimation error has evident drop with 10-minutes interval, with the relative errors below 10%. Under the CVMPR of 6% and the selected analysis interval, the RMSEs are 5.47% (3% CVMPR), 5.19% (6% CVMPR) and 4.33% (10% CVMPR). And the relative errors are 9.92% (3% CVMPR), 9.42% (6% CVMPR) and 7.86% (10% CVMPR).
  - Under the weekday off-peak scenario (x = 0.34), generally a 10-minute analysis interval is intended to provide accurate results with relative errors below 10%. When the CVMPR escalates to 3% and above, the estimation error reduces significantly as the RMSEs are 3.04% (6% CVMPR), 2.93% (10% CVMPR) and 2.76% (15% CVMPR), which are obtained in comparison with the ground truth of 30.55% under the analysis interval of 10 minutes. Moreover, the relative errors are 9.91% (6% CVMPR), 9.58% (10% CVMPR) and 9.12% (15% CVMPR) respectively.

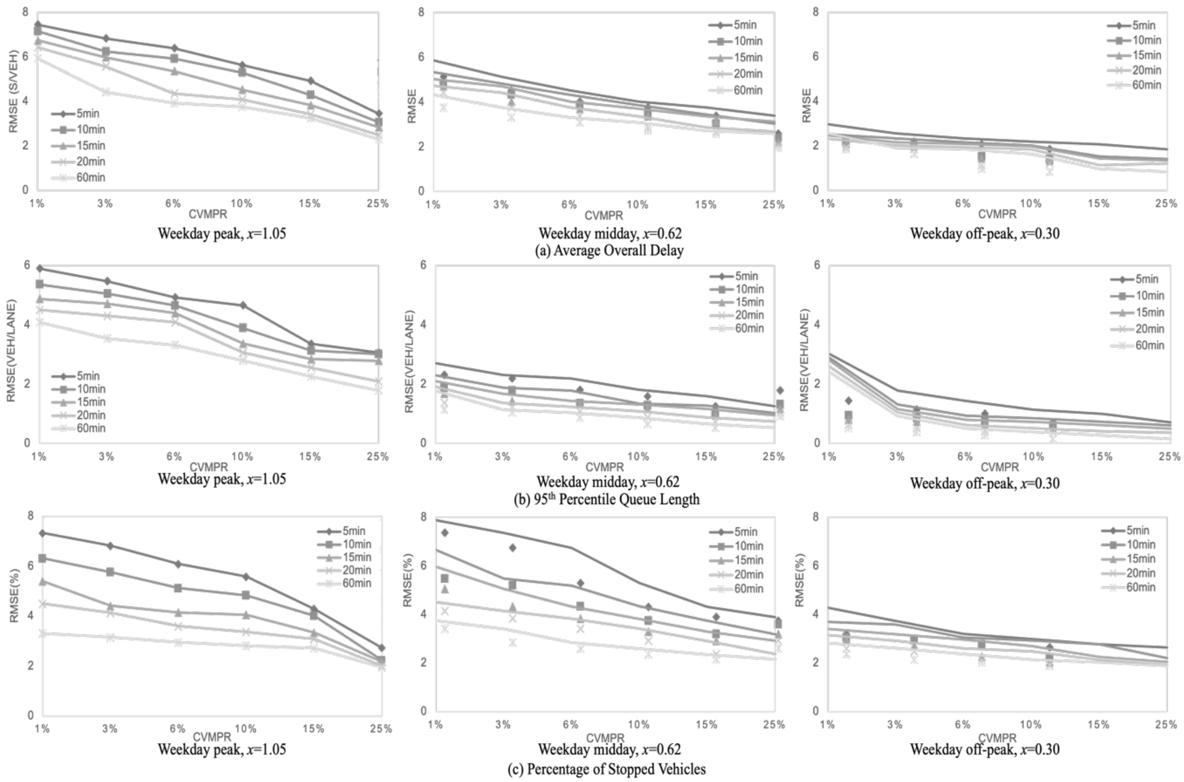


Figure 4-1 Movement-level Accuracy Analysis of Performance Measure Estimations

From the analysis above, the average overall delay and percentage of stopped vehicles can obtain performance estimations with consistently low RMSE results based on a 5-minute analysis interval when the CVMPR becomes greater as 3% (average overall delay) or 6% (percentage of stopped vehicles) and above at the weekday peak scenario. For the weekday midday scenario and off-peak scenario, the CVMPR should be at least 6% (average overall delay) and 10% (percentage of stopped vehicles) to achieve the reliable performance estimations (relative error below 10%) under the analysis interval of 10 minutes at the movement level. However, the requirements for minimum CVMPR and analysis interval should be 6% and 10 minutes to obtain reliable estimations under the weekday peak scenario when applying the 95th percentile queue length measure.

4.1.2 Result at the Approach Level

In **Figure 4-2**, the requirements of minimum CVMPR and optimal temporal analysis interval are investigated at the approach level. Although the amount of simulated CV data collected from the approach level is higher than that from the movement level, in general, the estimation accuracy at the approach level shows slight improvement compared to that at the movement level, particularly during the weekday off-peak scenario given that the variation in performance is larger at the approach level. However, the minimum CVMPR can be 10% and the temporal aggregation can be 10 minutes when the percentage of stopped vehicle measure is applied under the weekday off-peak scenario. More discussions are listed below.

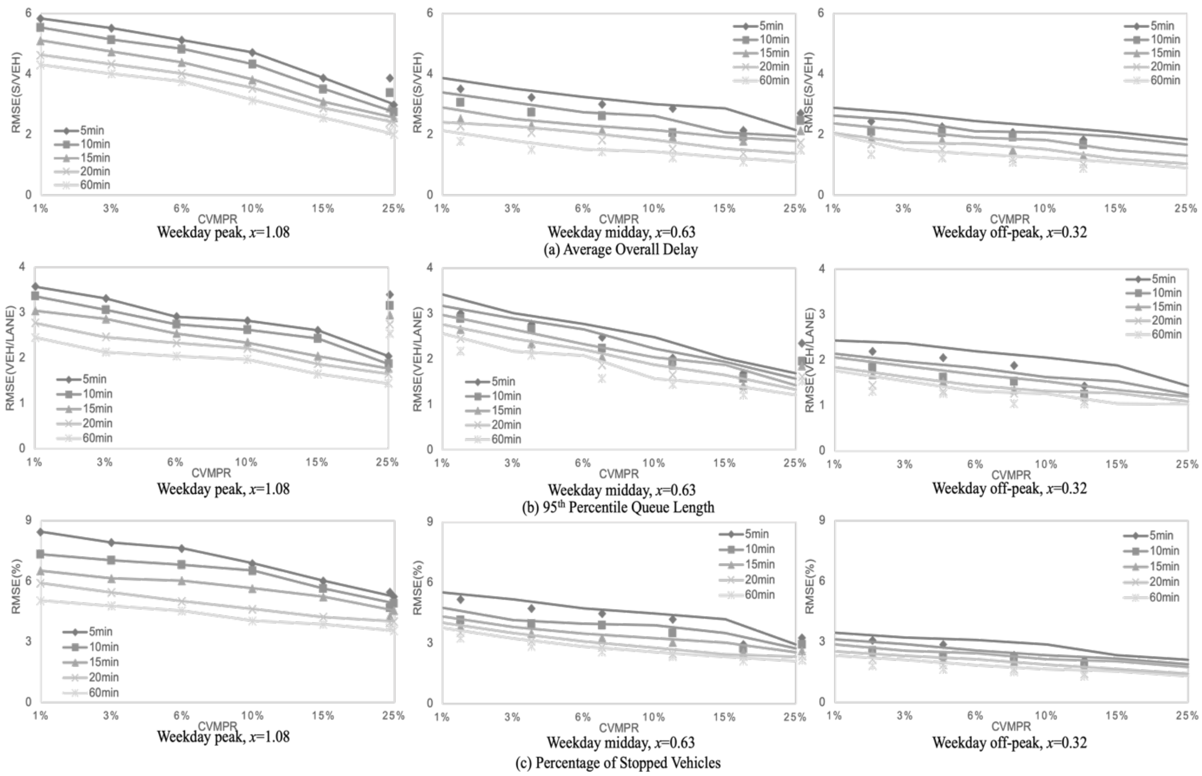


Figure 4-2 Approach-level Accuracy Analysis of Performance Measure Estimations

- Average overall delay:
  - Under the weekday peak scenario (x = 1.08), when the temporal aggregation is set as 5 minutes, the relative errors are observed as below 10% in comparison with the actual value of 54.17 s/veh, including 9.64 s/veh (6% CVMPR), 8.31 s/veh (10% VMPR) and 7.09 s/veh (15% CVMPR). And the RMSEs are 5.32 s/veh (6% CVMPR), 4.72 s/veh (10% VMPR) and 3.87 s/veh (15% CVMPR).
  - Under the weekday midday scenario (x=0.63), the relative error reduces to 10% and below in comparison with the actual value of 32.29 s/veh when the CVMPR surpasses

3% and above under the analysis interval of 10 minutes. The relative errors are 9.32% (3% CVMPR), 8.31% (6% CVMPR) and 6.03% (15% CVMPR). Moreover, the RMSEs are 3.07 s/veh(3% CVMPR), 2.74 s/veh (6% CVMPR) and 2.07 s/veh (15% CVMPR) respectively.

  - Under the weekday off-peak scenario (x=0.32), the estimations are reliable when the CVMPR surmounts 6% and above with the analysis interval is set as 10 minutes given that the relative errors are 9.83% (6% CVMPR), 9.07% (10% CVMPR) and 8.81% (15% CVMPR). Meanwhile, the RMSEs are calculated as 2.11 s/veh (6% CVMPR), 2.07 s/veh (10% CVMPR) and 1.93 s/veh (15% CVMPR) in comparison with the actual value of 21.27 s/veh.

- 95th percentile queue length:
  - Under the weekday peak scenario (x = 1.08), the relative estimation errors are reduced to 10% and below when the CVMPR is 6% and above with the analysis interval of 5 minutes. Given the actual value of 31.37 veh/lane, the RMSEs are 2.91 veh/lane (6% CVMPR), 2.82 veh/lane (10% CVMPR) and 2.61 veh/lane (15% CVMPR). And the corresponding relative errors are 9.27% (6% CVMPR), 8.99% (10% CVMPR) and 8.33% (15% CVMPR).
  - Under the weekday midday scenario (x=0.63), estimations obtained from the 5-minute interval are regarded as reliable when the CVMPR increase % CVMPR) and 2.02 veh/lane (15% CVMPR) and the relative errors are 9.90% (10% CVMPR) and 8.08% (15% CVMPR) respectively.
  - Under the weekday off-peak scenario (x=0.32), the relative estimation errors attained from 10-minute analysis interval are below 10% when the CVMPR overtakes 10% in comparison with the actual value of 11.87 veh/lane. The corresponding RMSEs are 1.16 veh/lane (10% CVMPR) and the relative error is 9.78%. If the CVMPR passes 15%, the RMSE reduces to 1.03 veh/lane with the relative error of 8.69%.

- Percentage of stopped vehicles:
  - Under the weekday peak scenario (x = 1.08), the error reduces significantly when the interval increases to 5 minutes and above. The RMSE is 7.63% (6% CVMPR), 6.89%

(10% CVMPR) and 6.02% (25% CVMPR), which is calculated based on the measure results attained from the population as 77.89%. The relative errors are 9.80% (6% CVMPR), 8.85% (10% CVMPR) and 7.63% (25% CVMPR) respectively.

- Under the weekday midday scenario (x=0.63), the RMSEs are 3.96% (3% CVMPR), 3.89% (10% CVMPR) and 3.50% (15% CVMPR) in comparisons with the actual value of 41.95% when applying 10-minute interval. Moreover, the relative errors are 9.47% (3% CVMPR), 9.19% (10% CVMPR) and 8.55% (15% CVMPR) respectively.
- Under the weekday off-peak scenario (x=0.32), the values of RMSE have minor differences when the CVMPRs and temporal aggregations change. However, when the CVMPR rises above 10% and analysis interval extends to 10 minutes, the averaged RMSE is 2.33% in comparison to the actual value of 24.47%, where the relative error is calculated as 9.35%. If the CVMPR is 15%, the RMSE and relative error are 2.17% and 8.91% respectively.

In Figure 4-2 (a) and (c), the measures including average overall delay and percentage of stopped vehicles can provide accurate estimation when the CVMPR goes beyond 6% and the analysis time interval increases to 10 minutes under the weekday midday scenario. However, the 95th percentile queue length is less reliable as the penetration rates should reach at least 10% to attain accurate estimations when the temporal aggregation is extended to 20 minutes and above under the same scenario.

Under the weekday peak scenario, the 3% CVMPR and a 10-minute analysis interval can be sufficient when applying the measure of average overall delay. When the percentage of stopped vehicles is taken into consideration, the 3% CVMPR and a 10-minute analysis interval are essential.

4.1.3 Results at the Intersection Level

In this subsection, the accuracy analysis of performance results attained from the critical intersection is conducted at multiple temporal aggregations and under various traffic scenarios (**Figure 4-3**). Generally, the 3%-6% CVMPR and a 10-minute analysis interval is necessary to guarantee accurate performance measurement estimations when applying the three measures under the weekday peak scenarios at the intersection level. Detailed discussions are listed as below.

- Average overall delay:

- Under the weekday peak scenario (x = 1.04), the RMSEs of average overall delay have evident decrease from 5.03 s/veh to 4.79 s/veh when the CVMPR increases from 3% to 6% under the 10-minute analysis interval in comparison with the actual value of 50.95 s/veh. Meanwhile, the corresponding relative errors are 9.87% (3% CVMPR) and 9.67% (6% CVMPR). With the 10% CVMPR and the 10-minute temporal aggregation level, the RMSE is 3.66 s/veh as the relative error is 9.22%.
- Under the weekday midday scenario (x=0.61), even with a CVMPR as low as 3%, accurate estimations can be achieved, provided that the analysis interval is set to 15 minutes. However, if the temporal aggregation is limited to 10 minutes, a CVMPR of 6% becomes necessary to ensure reliability. The RMSE values are 3.29 s/veh (3% CVMPR, 15 minutes) and 3.16 s/veh (6% CVMPR, 10 minutes), with corresponding relative errors of 9.83% (3% CVMPR, 15 minutes) and 9.77% (6% CVMPR, 10 minutes).
- Under the weekday off-peak scenario (x=0.30), the 3% CVMPR can still provide reliable performance estimations with the analysis interval of 15 minutes. Under such conditions, the RMSE error is 1.52 s/veh, which is associated with the relative error of 9.69%. To perform real-time traffic monitoring, a 6% CVMPR is required with the 10-minute analysis interval as the relative error is 9.88% and the RMSE is 1.72 s/veh.

- 95th percentile queue length:
  - Under the weekday peak scenario (x = 1.04), the temporal aggregations should be set to 10 minutes when the CVMPR reaches 10% to obtain reliable estimations. As the actual value attained from the population of interest is 33.95 veh/lane, the RMSE is calculated as 3.24 veh/lane. And the corresponding relative error is 9.68%. If the CVMPR is 6%, the relative errors are 10.04% and 9.93% for the 10-minute and 15-minute analysis interval separately. The RMSE is 3.67 veh/lane (6% CVMPR, 10-minute) and 2.98 veh/lane (6% CVMPR, 15-minute).
  - Under the weekday midday scenario (x=0.61), when the CVMPR reaches 10%, a 10-minute analysis interval is necessary for performance estimation. The RMSE is 1.69 veh/lane based on the actual value of 17.44 veh/lane given the relative error of 9.71%. Meanwhile, reliable results can be achieved with a 6% CVMPR using only a 15-minute interval. The RMSE is 1.67 veh/lane. And the relative error is 17.66 veh/lane.

- Under the weekday off-peak scenario (*x*=0.30), the minimum requirement for temporal aggregation is 15 minutes when the CVMPR reaches 10% as the relative error reduces to 9.51%, which indicates that the queue length measure cannot provide real-time monitoring at the intersection level. The RMSE is 1.22 veh/lane under 10% CVMPR and the 15-minute interval, which is calculated based on the actual value of 13.13 veh/lane.

- Percentage of stopped vehicles:
  - Under the weekday peak scenario (*x* = 1.04), the performance estimations attained from the 10-minute analysis interval are reliable when the CVMPR is 3%. The relative error is calculated as 9.44% based on the actual value of 45.66%. Meanwhile, the RMSE is 4.19%. When the CVMPR reaches 6%, a shorter analysis interval can still guarantee accurate results with relative error below 10% (9.24%). The RMSE is 4.33%.
  - Under the weekday midday scenario (*x*=0.61), the analysis interval can be set as 10 minutes to attain reliable results with relative errors of 10% and below when the CVMPR ascend beyond 6%. The actual value is 36.81% and the RMSE is calculated as 3.52%.
  - Under the weekday off-peak scenario (*x*=0.30), when the CVMPR achieves 10%, it is possible to obtain accurate performance estimations with the analysis interval of 10 minutes. The RMSE is calculated as 2.03% based on the actual value of 21.95%, which is associated with the relative error of 9.34%.

Generally, the percentage of stopped vehicles perform better than the average overall delay and 95th percentile queue length given that its requirement for CVMPR is lower under the weekday off-peak scenarios. Under the weekday midday and the weekday peak scenario, the percentage of stopped vehicles can present estimation results close to the actual value with relative errors below 10% when the CVMPR reaches to 3% and above. The corresponding time analysis interval can be set as 5 minutes or 10 minutes to conduct real-time traffic performance measurements.

Moreover, the 95th percentile queue length can provide reliable results with relative errors below 10% under different traffic scenarios according to the specific requirements as discussed above. However, under the weekday off-peak scenario (undersaturated scenario), it is noticeable that the queue length measure is not suitable for real-time monitoring at the intersection level as its minimum requirement for analysis interval is 15 minutes when the CVMPR reaches 10%.

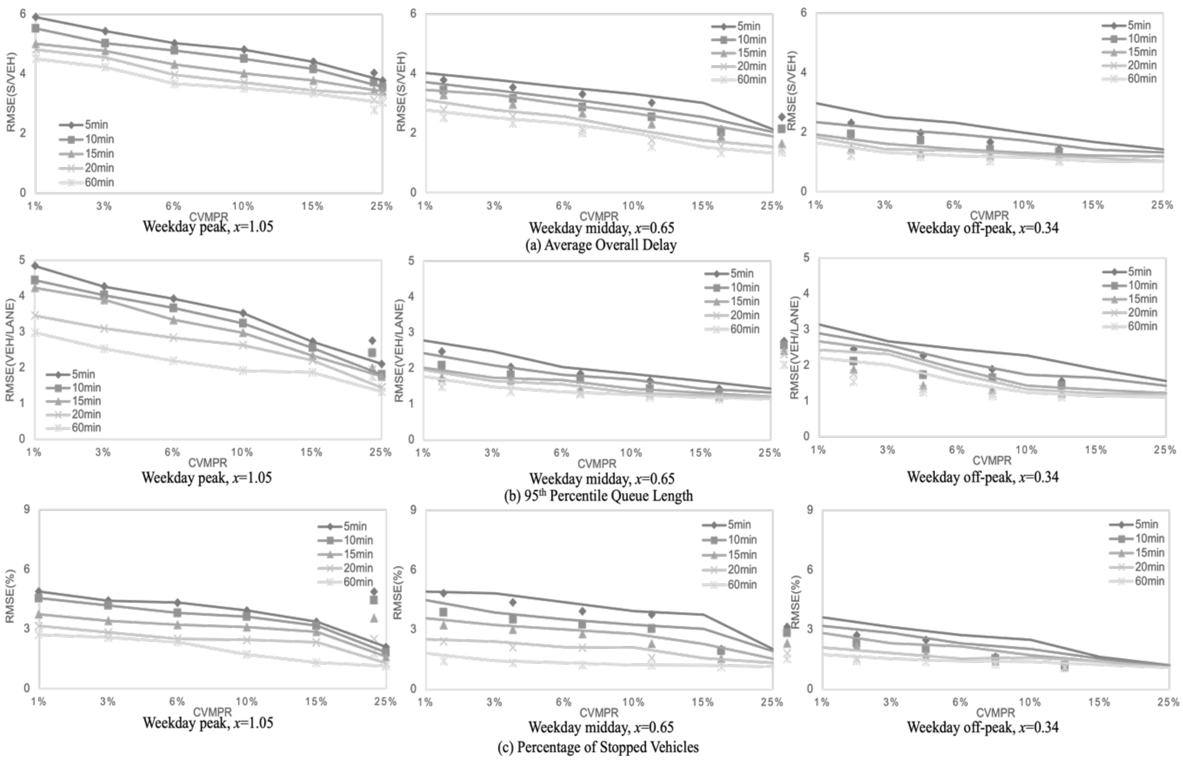


Figure 4-3 Intersection-level Analysis of Performance Measures Derived from Simulated CV Data

4.1.4 Results at the Corridor Level

To assess the impact of CVMPR and investigate the most suitable temporal aggregations, an analysis of estimation accuracy is conducted at the corridor level, as depicted in Figure 4-4. This analysis encompasses the simulated eastbound CVs that traverse the entire corridor in the simulation.

Furthermore, as the queue length measure lacks relevance at the corridor level, the focus is narrowed to estimating only the average overall delay and the percentage of stopped vehicles along the corridor.

- Average overall delay:
    - Under the weekday peak scenario ($x = 1.12$), the relative estimation errors decrease to 10% and below as the CVMPR achieves 6% and above with the 5-minute analysis interval. The RMSEs under different market penetration rates are 10.66 s/veh (6% CVMPR), 9.89 s/veh (10% CVMPR) and 9.25 s/veh (15% CVMPR) respectively in comparison with the actual value of 108.91 s/veh.

- Under the weekday midday scenario (x = 0.67), when the CVMPR reaches 6% and above, the result shows the reliability of the relative error below 10% with a 10-minute analysis interval. The RMSE is 7.17 s/veh and the average relative error is 9.52%, which is calculated based on the actual value of 74.33 s/veh attained from the population of interest.
- Under the weekday off-peak scenario (x = 0.35), the average overall delay can still provide reliable results when the CVMPR is 6% under a 15-minute interval. With the averaged RMSE of 3.47 s/veh, the corresponding relative error is calculated as 9.84% based on the results attained from the simulation experiments, in comparison with the actual value of 35.35 s/veh.

- Percentage of stopped vehicles:
  - Under the weekday peak scenario (x = 1.12), with the 5-minute interval, the percentage of stopped vehicles can provide accurate estimations under even 3% CVMPR. Based on the actual value of 85.41%, calculations reveal that the RMSE is 8.19% and the relative error is 9.64%.
  - Under the weekday midday scenario (x = 0.67), when the CVMPR is 3%, a 10-minute analysis interval can be applied to guarantee reliable performance estimations. Based on the actual value of 51.12%, the RMSE is 5.08, which is associated with a relative error of 9.97%. If the CVMPR rises to 6%, the RMSE is 4.99% with the relative error of 9.83%.
  - Under the weekday off-peak scenario (x = 0.35), the relative estimation error falls below 10% when using a 10-minute analysis interval, provided that the CVMPR reaches 6%. In this scenario, performance estimations exhibit an absolute error of 2.73% (RMSE) and a relative error of 9.72% when compared to the actual value of 28.09%.

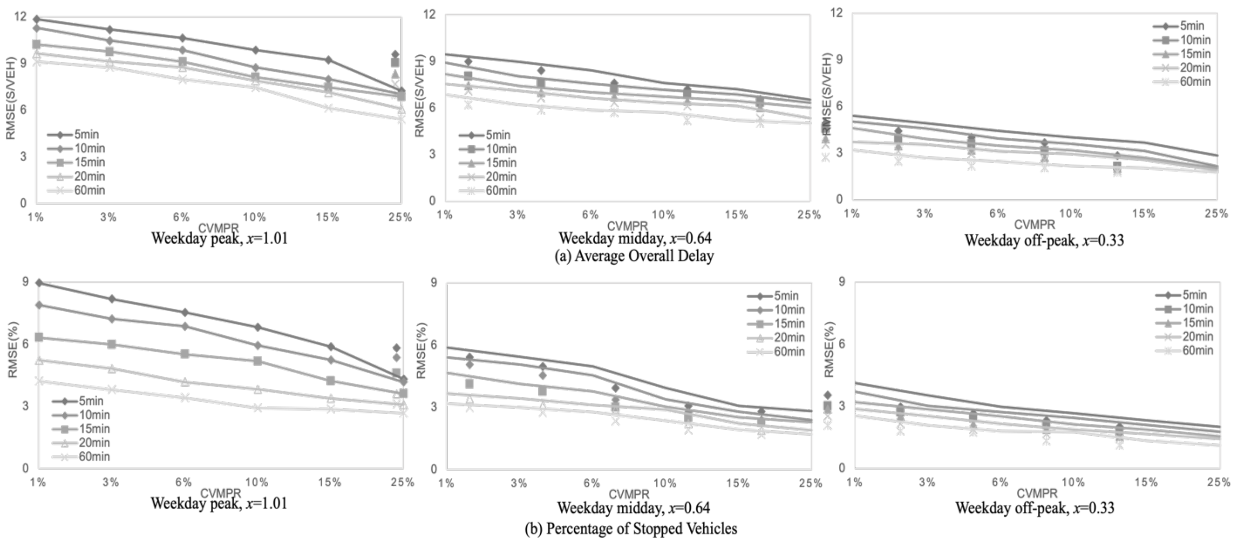


Figure 4-4 Corridor-level Analysis of Performance Measures Derived from Simulated CV Data

4.2 Congestion Detection

In this section, the feasibility of applying CV data in the congestion detections is explored based on the simulation experiments. Both recurrent congestion and non-recurrent congestion detections are conducted.

The analysis first focuses on identifying recurring instances of traffic congestion. Specifically, the study examines congested traffic conditions arising from the daily commute within the urban area. This examination is conducted by modelling both weekday off-peak and weekday peak scenarios. Then, the analysis is conducted based on the non-recurrent congestions with capacity constraints and event-driven peak conditions, which is designed to explore the feasibility of applying traffic performance measures derived from the CV data. The settings of each scenario are detailed in the following subsections.

The investigation of the feasibility of congestion detection is performed based on the performance measures attained from the simulated CV data collected under each scenario. The analysis is performed to detect both the recurrent and the non-recurrent congestions with a 5-minute analysis interval to observe the changes in performance estimations under various CVMPRs.

4.2.1 Detection of Recurrent Congestions

During the simulation, the total traffic volume is gradually raised from 3,000 vehicles per hour to 10,000 vehicles per hour. This change occurs precisely at the 30-minute mark (which corresponds to 1 800 seconds) and ends at the 90-minute mark (or 5 400 seconds) within the overall 2-hour duration (equivalent to a 7200-second period). The traffic volume variations are detailed in Table 4-1.

Table 4-1 Traffic Volume Variations during The Recurrent Congestion Detections

Spatial level	0min-30min	30min-90min	90min-120min
West Broward Blvd.	3 000 veh/h	10 000 veh/h	3 000 veh/h
Intersection No.10 Eastbound Right-turn	3 veh/h	9 veh/h	3 veh/h

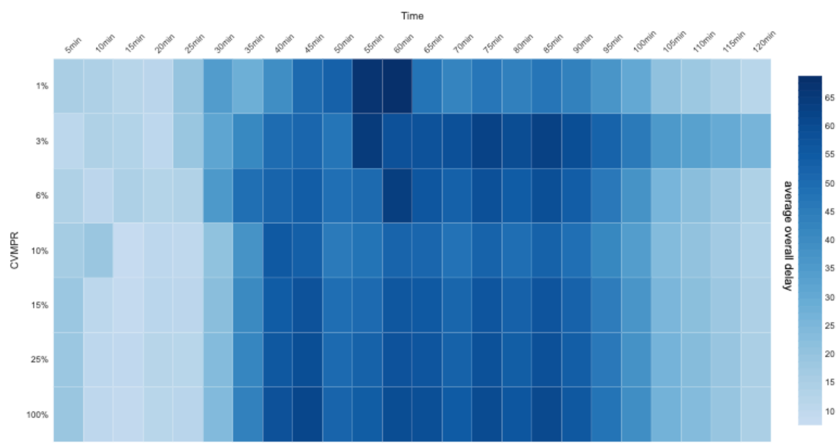
As depicted in Figure 4-5, the performance measure – average overall delay, 95th percentile queue length and percentage of stopped vehicles - extracted from the simulated CV data, considering various CVMPRs, effectively captures shifts in traffic performance. The performance estimations are obtained from the eastbound right-turning movement of Intersection No. 10, which includes a clear identification of the occurrence and dissipation of congestion as traffic volume fluctuates. The average hourly volume of the selected movement is 3 veh/h during 0min-30min, 9 veh/h during 30min-90min and 3 veh/h during 90min-120min.

Typically, the influence of changes in traffic volume becomes evident after the simulation has been running for 35 minutes, which has a 5-minute lag compared to the initiation of such changes. The alterations in traffic performance become discernible through these measures within a span of 10 minutes when the traffic volume turns from 10 000veh/h to 3 000 veh/h.

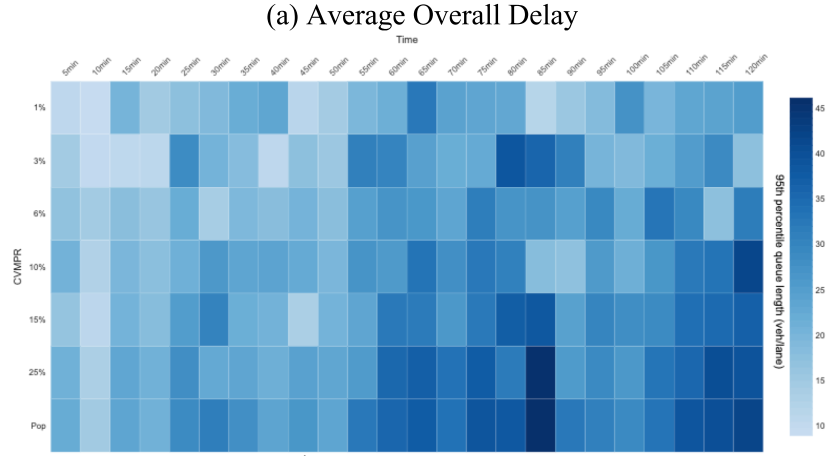
From Figure 4-5(a), the peak value of delay appears at 45 minutes (61.9 s/veh) and 85 minutes (61.12 s/veh) with respect to the beginning of the simulation. When the CVMPR reaches 10% and above, the trend of performance changes is clearly reflected in the change of average overall delay. Despite a CVMPR of 6%, the delay measure successfully identifies the commencement and termination of recurrent congestion but falls short in accurately estimating the peak value.

Referring to Figure 4-5(b), it is observed that the onset of congestion can be successfully detected when the CVMPR reaches 15% and above. However, the estimation still falls short at the 45-minute mark. Notably, neither the 10% CVMPR nor the 6% CVMPR can effectively meet the requirement for detecting congestions with the 5-minute interval.

In Figure 4-5(c), the performance estimations regarding the percentage of stopped vehicles effectively capture changes in traffic conditions, even with a CVMPR as low as 6%. Furthermore, it should be noted that a minimum CVMPR of 10% is required for the detection of the entire extent of recurrent congestion.

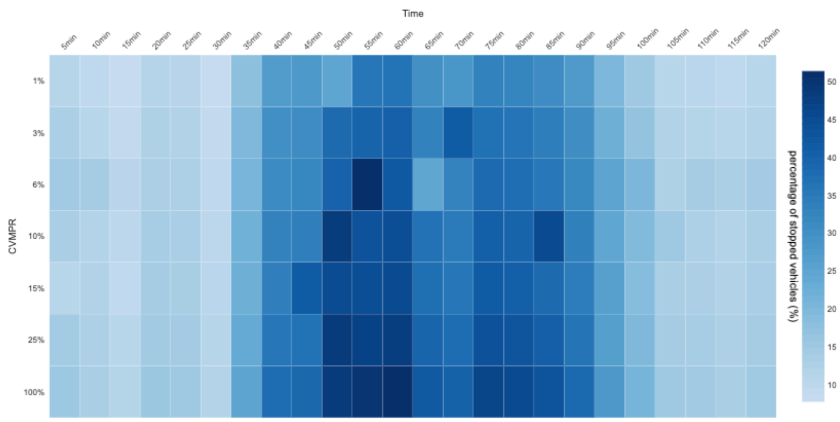


(a) Average Overall Delay



(b) 95th Percentile Queue Length

Figure 4-5 Recurrent Congestion Detections



(c) Percentage of Stopped Vehicles

Figure 4-5 Recurrent Congestion Detections

4.2.2 Detection of Non-recurrent Congestions under Capacity Constraints

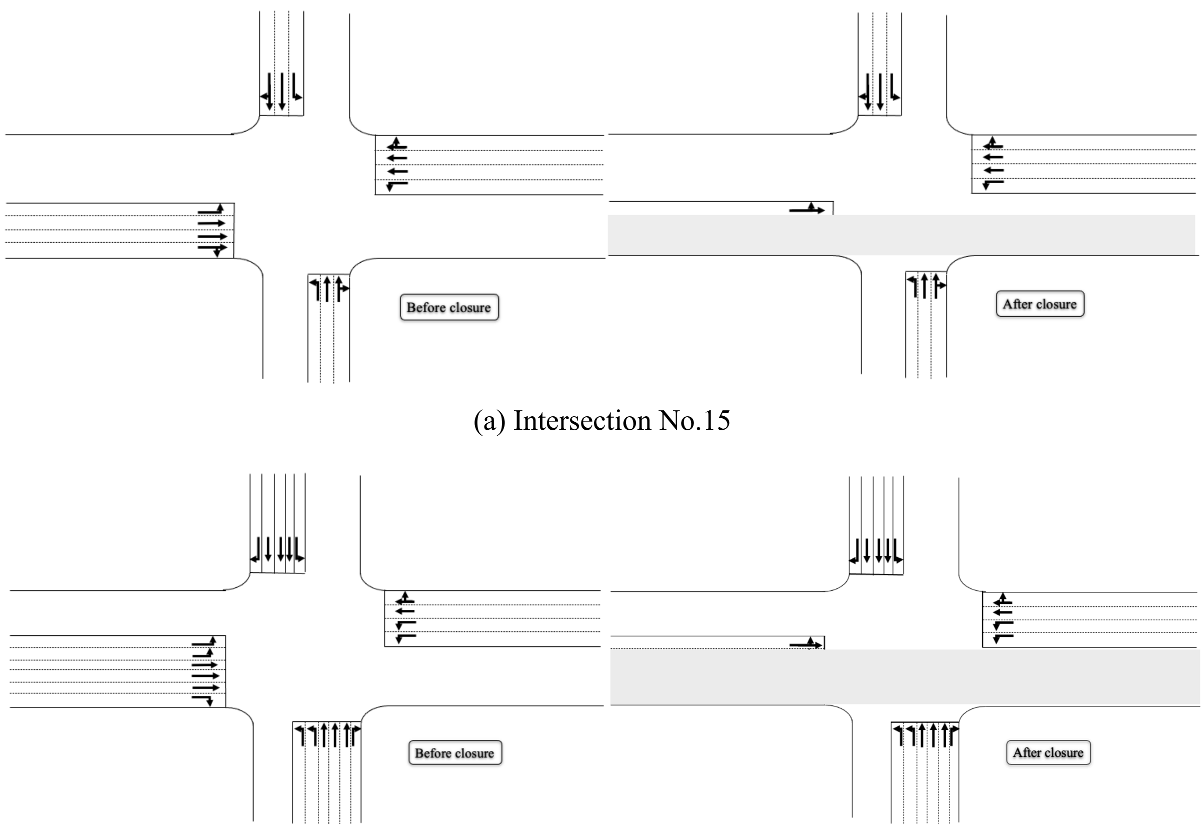
Table 4-2 summarizes the characteristics of the scenario to simulate congestions caused by capacity loss due to a hypothetical construction project at the eastbound direct along the West Broward Blvd. corridor.

This capacity constraint scenario entails the closure of three lanes within the eastbound corridor stretch, spanning a total length of 350 meters. The affected area includes Intersection No.15 and Intersection No.16.

To ensure the continuous flow of through movement traffic along the eastbound corridor, a modification is made: the eastbound left-turn at both Intersection No.15 and Intersection No.16 transitions into a merged lane, accommodating both left-turn and through movements. The geometric alterations are visually depicted in Figure 4-6. For the newly merged lane, its capacity remains unchanged. However, the capacity distribution between the left-turn and through movements is determined based on the pre-alteration ratio of capacities for these two movements, which is demonstrated in Table 4-2 as well.

Table 4-2 Capacity Variations during The Capacity-Constraint Scenario

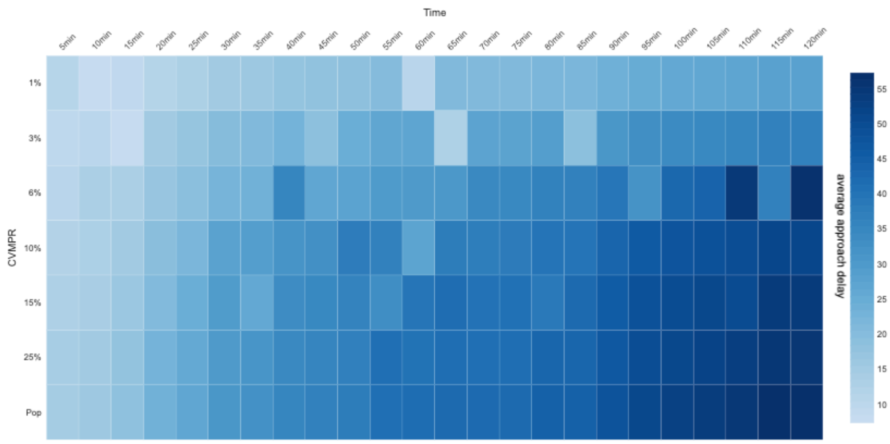
Spatial level	Capacity (veh/h)			
	0s-1800s		1800s-7200s	
Intersection No.15	Eastbound Left-turn	855 veh/h	Eastbound Left-turn	283 veh/h
	Eastbound Through	1723 veh/h	Eastbound Through	570 veh/h
	Eastbound Right-turn	314 veh/h	Eastbound Right-turn	0 veh/h
	Total capacity loss		2039 veh/h	
Intersection No.16	Eastbound Left-turn	513 veh/h	Eastbound Left-turn	324 veh/h
	Eastbound Through	890 veh/h	Eastbound Through	561 veh/h
	Eastbound Right-turn	374 veh/h	Eastbound Right-turn	0 veh/h
	Total capacity loss		892 veh/h	



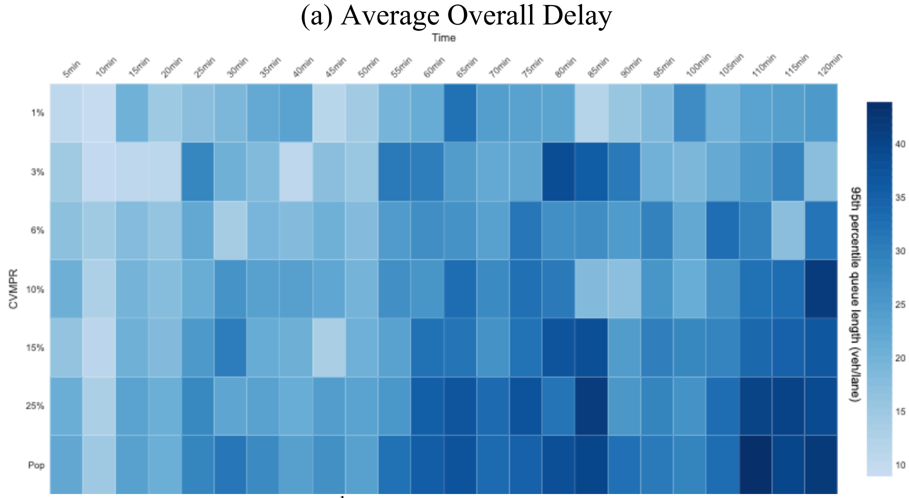
(a) Intersection No.15

(b) Intersection No.16

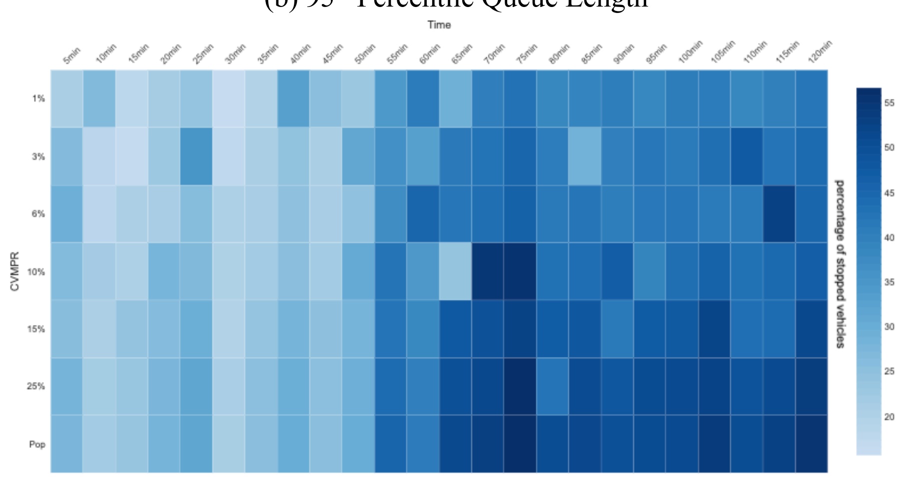
Figure 4-6 Graphic Illustration of the Capacity Constraint Scenario



(a) Average Overall Delay



(b) 95th Percentile Queue Length



(c) Percentage of Stopped Vehicles

Figure 4-7 Non-recurrent Congestion Detection Using Simulated CV Data (Capacity-constraint Scenario)

The congestion detection is conducted at the merged lane of Intersection No.15 with the performance measures derived from the CV trajectory data obtained from the simulation experiment. In Figure 4-7(a), the congestion caused by the capacity loss is forming gradually as the vehicles experiencing delay accumulate before the stop bar. On the movement of interest, even at a CVMPR of as low as 1%, 3% and 6% congestion can be detected with the variations in the average overall delay attained from CVs. To perform exact detection of the severity of the congestion, the CVMPR is required to be 10% at least when the aggregated interval is 5 minutes to observe the onset of the congestion.

In Figure 4-7(b), the utilization of the 95th percentile queue length measure proves effective in detecting the initiation of non-recurrent congestions once the CVMPR reaches 15% and above. However, it is important to note that the detection result of this measure might experience significant errors. Yet, its advantage lies in the fact that it avoids overestimating performance results. Despite these challenges, the queue length measure still manages to aptly depict variations in traffic demand under the CVMPR of 15% and above with the application of 5-minute analysis intervals.

In Figure 4-7(c), the percentage of stopped vehicles measure effectively identifies the emergence of congestions, particularly in the capacity-constraint scenario with a 5-minute analysis interval. Notably, this measure adeptly captures variations in traffic demand when the CVMPR is at 10% and above, providing performance estimations that closely align with the actual values.

4.2.3 Detection of Non-recurrent Event-driven Congestions

Under the event-driven congestion condition, it is assumed that an event results in the increase of traffic demand at the Intersection No.16 with an influx of 3206 veh/h at a specific time. The congestion detection is conducted at the eastbound left-turn based on the proposed measure according to a 5-minute analysis interval with the performance measures derived from the simulated CV data. A detailed illustration on the traffic volume variations is listed in Table 4-3.

Table 4-3 Traffic Volume Variations during The Event-driven Congestion Scenario

Spatial level		Traffic Volumes (veh/h)		
		0s-2400s	2400s-6000s	6000s-7200s
Intersection No.16	Northbound	1416	6416	1416
	Southbound	1789	1789	1789
	Eastbound	1281	6281	1281

Westbound	1032	1032	1032
Total volume increase during event		1000 veh/h	

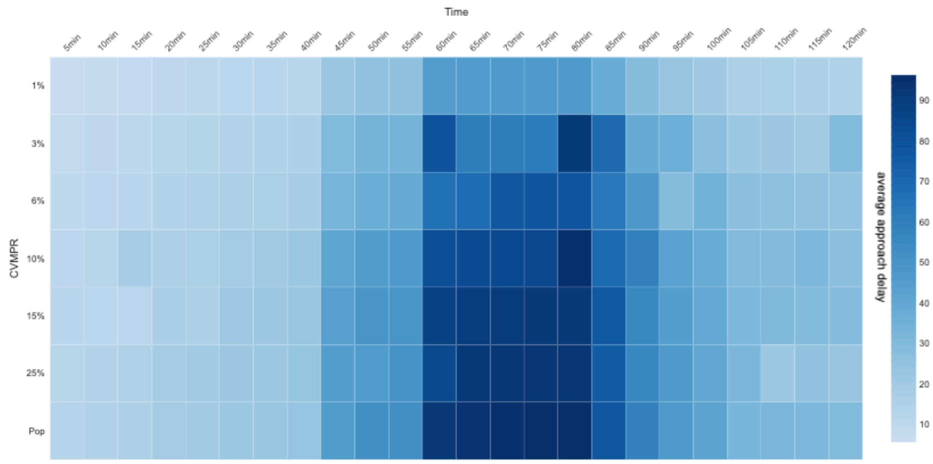
In the scenario with event occurrence, three individual measures are applied to detect the congestion with a 5-minute analysis interval as shown in Figure 4-8. From Figure 4-8(a) to Figure 4-8(c), a noticeable change is observed in the traffic demand, beginning around the 45-minute mark (2700 seconds) and ending around the 110-minute mark (6600 seconds) with respect to the beginning of simulation. In contrast to the simulation experiment's configuration, the impact of heightened traffic demand takes approximately 5 minutes to manifest when the performance measures are employed to pinpoint these fluctuations. Moreover, the dissipation of the congestions takes approximately 10 minutes as observed in the simulation experiment.

Furthermore, during the window of 40 to 45 minutes, the average total delay, the 95th percentile of queue length, and the proportion of stationary vehicles exhibit lower values compared to those measured from 45 minutes onwards. Similarly, during the dissipation phase, spanning 100 to 110 minutes, the performance measures manifest lower values than those recorded during the period preceding 100 minutes. Beyond the 110-minute juncture, congestion progressively diminishes over time.

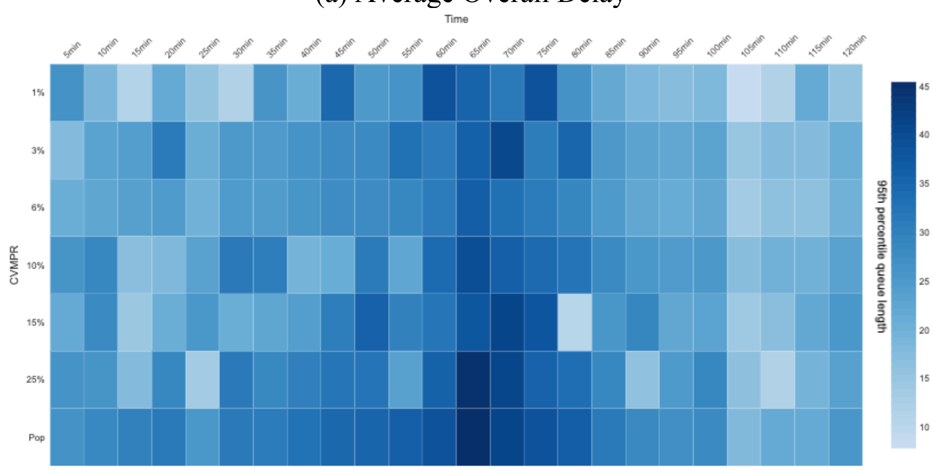
In Figure 4-8(a), the average overall delay effectively reflects the fluctuations in traffic demand when the CVMPR is 10% and above. To accurately represent the extent of delay, a minimum CVMPR of 15% is necessary for reliable application.

In Figure 4-8(b), the outcomes obtained from the queue length measure exhibit considerable errors, failing to capture the demand fluctuations accurately via queue changes. As the CVMPR increases to 25%, a pronounced peak value of the queue is estimated using the CV data, potentially indicating the onset of severe congestion. While the 95th percentile queue length measure might not offer dependable detection capabilities, estimations derived from the simulated CV data of queue length measure can suggest potential traffic jams across the road network when the event happens along the corridor, even if not precisely quantified.

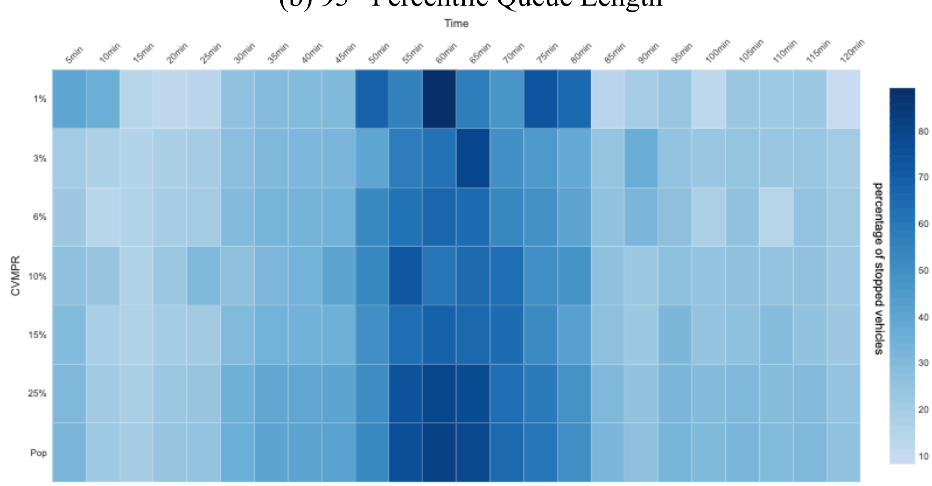
In Figure 4-8(c), the performance estimations derived from the percentage of stopped vehicles become reliable when the CVMPR exceeds 10% and higher. Specifically, to accurately depict congestion peaks, a CVMPR of at least 15% is recommended.



(a) Average Overall Delay



(b) 95th Percentile Queue Length



(c) Percentage of Stopped Vehicles

Figure 4-8 Non-recurrent Congestion Detection Using Simulated CV Data (Event-driven Scenario)

Chapter 5 Conclusions

In this chapter, the key findings attained from the analysis in the previous chapters are presented accordingly with discussions on different performance measures and the optimal spatial and temporal data resolutions. Moreover, the impacts of CV market penetration rate (CVMPR) on the accuracy and reliability of the performance measures are summarized. And the feasibility of detecting congestions is summarized as well. Following this, the significance of this research is highlighted with recommendations for the future research.

5.1 Summary of The Key Findings

This research explored the potential of applying the Connected Vehicle (CV) data for traffic performance measurement and investigated the feasibility of detecting non-recurrent congestions through a simulation-based study. An extensive simulation study was conducted on a model of the Broward Blvd. arterial located in Florida, U.S., which includes three congestion scenarios of weekday off-peak, midday and peak periods. The analysis concluded with the following main findings:

- On the reliability of different traffic performance measures:
    - For average overall delay, CV data of a wide range of market penetration rates was found to be sufficient in capturing traffic conditions across all levels of congestion at both movement and arterial levels. As the CVMPR increases, the accuracy of the estimate on the average overall delay improves, providing reliable measurements at the 10-minute or 15-minute aggregation intervals.
    - For the 95th percentile queue length, it generally requires higher CVMPR and longer temporal aggregations for performance measurements. It was found that the 95th percentile queue length could not be estimated reliably using CV data during the low traffic scenarios (e.g., weekday off-peak) at the arterial-level analysis, thus not recommended for performance measurement propose.
    - The estimates on the percentage of stopped vehicles using CV data were shown to be more reliable than that of queue length. Furthermore, this measure shows acceptable accuracy even when the collected data amount is extremely low.
- On the effects of spatial and temporal data resolution:

  - Based on the movement-level analysis, insights into traffic performance during undersaturated scenarios can be obtained with as low as 3%-6% CVMPR when the temporal aggregation is set to 10 minutes and above with the application of average overall delay and percentage of stopped vehicles. For other performance measures, a higher CVMPR of at least 15% is required, and the analysis time interval should be 20 minutes or longer under the low saturated scenario being examined (i.e., x = 0.34). For the intermediately saturated condition (i.e., x = 0.65), a CVMPR of 10% and an analysis time interval of 15 minutes are sufficient to estimate the average overall delay and percentage of stopped vehicles.
  - For the scenario of oversaturated conditions (i.e., x=1.05), a CVMPR of 6% is necessary to obtain reliable performance estimations at the approach level with an analysis interval of 10 minutes. These findings highlight the importance of a reasonable CVMPR and appropriate analysis time interval for accurate and meaningful performance assessment.
  - At the corridor level, an aggregation interval of 10 minutes is required to obtain reliable estimations for weekday peak scenario when the CVMPR reaches 6% or above with the performance measures of average overall delay and percentage of stopped vehicles. The CVMPR should be 10% or higher in order to achieve reliable estimations at the analysis interval of 15 minutes.
  - Typically, longer analysis intervals yield more reliable results than shorter ones. The results attained from the arterial level are more reliable than those from the movement level. Results obtained from the arterial level encompass a broader perspective, considering the collective behaviour of multiple movements and intersections. This aggregation reduces the impact of random variations and provides a more stable representation of overall traffic conditions, making the results more reliable compared to individual movement-level assessments.
- On the feasibility of congestion detection:
  - The results from the simulation analysis suggest that it is feasible to detect the onset and dissipation of non-recurrent traffic congestions based on CV data at different spatial aggregation levels including movements, approaches, intersections, and the entire corridor. The detection requires appropriate temporal analysis resolutions and CVMPRs.
- On the effect of CVMPR on performance measurement:

- Generally, the effect of CVMPR on performance measurement is evident from the analysis. Naturally, the minimum required CVMPR is different for different traffic conditions and required spatial and temporal resolutions. When the CVMPR increases to 15% and above, the error in performance estimation can be reduced significantly for all scenarios with appropriate temporal resolutions, which can be considered acceptable for practical applications.

5.2 Significance of the Study for Urban Traffic Monitoring and Management

The use of connected vehicle (CV) data for automated signal performance measurements has significant implications for urban traffic monitoring and management. This study focuses on the potential of applying CV data for automated traffic signal performance measures (ATSPM).

Firstly, the ability to capture changes in traffic patterns and performance using CV data enables transportation professionals to gain valuable insights into various traffic scenarios and their effects on signal performance. Traffic engineers can identify and respond to minor fluctuations in traffic flow or obtain a more generalized overview of traffic performance. Furthermore, the use of CV data in automated signal performance measurements can support the development of data-driven transportation policies and guidelines. By leveraging CV data, transportation agencies can better understand the impacts of various traffic scenarios on signal performance, enabling them to design more effective strategies and allocate resources more efficiently.

Additionally, the real-time nature of CV data facilitates the identification of potential safety hazards, such as intersections with high accident rates or traffic flow patterns contributing to congestion. With the CV data, transportation agencies can continuously monitor traffic patterns and proactively address areas in need of improvement. For instance, by analyzing CV data, sudden changes in traffic flow that could lead to accidents can be detected. This information can then be used to adjust signal timings or implement other safety measures to minimize accident risk.

Moreover, future advancements in connected vehicle technology, coupled with the integration of data from other sources, such as traffic sensors or social media, could further improve the accuracy and reliability of performance measures and congestion detection algorithms. This would enable transportation agencies to make more informed decisions, improving overall traffic flow and safety.

In conclusion, the utilization of CV data in automated signal performance measurements offers significant potential for improving traffic flow, reducing congestion, and enhancing safety on urban roads. By further exploring and implementing the applications of CV data, transportation agencies can make more informed decisions, leading to more efficient, safe, and sustainable transportation systems.

5.3 Future Research Directions and Recommendations

This study has provided valuable insights into the use of connected vehicle data for traffic performance measurement and congestion detection. However, there are still many potential venues for future research to further improve the understanding and application of connected vehicle data in transportation planning and management. Here, several research directions and recommendations are suggested for future work:

1. Exploration of alternative performance measures: While this study focused on a select set of performance measures, future research could investigate additional or alternative performance measures that may provide more comprehensive insights into traffic conditions and the impact of connected vehicles.

2. Integration of data from other sources: Combining connected vehicle data with data from other sources, such as traffic sensors, video cameras, or social media, could enhance the accuracy and reliability of performance measures and congestion detection algorithms.

3. Development of advanced data processing and analysis techniques: Further research could focus on the development of new methodologies for processing and analyzing connected vehicle data, potentially utilizing machine learning, artificial intelligence, or big data techniques to improve the accuracy and efficiency of traffic performance measurements.

4. Evaluation of the impact of connected vehicle technologies on traffic operations: As connected vehicle technologies continue to evolve, future studies should assess the impact of these technologies on traffic flow, safety, and overall transportation system performance.

By pursuing these research directions, the understanding of the potential benefits and challenges associated with the use of connected vehicle data in transportation planning and management can be further enhanced, ultimately contributing to the development of more efficient, safe, and sustainable transportation systems.

5.4 Limitations

This research has several limitations that could be explored in future research. Firstly, this study has focused primarily on evaluating the accuracy of various ATSPMs using average estimation errors (e.g., RMSE) as a performance indicator, future research should also investigate the reliability of the performance measures such as how the RMSE obtained from multiple experiments vary by variation in traffic demand, CVMPR and measurement errors. Secondly, the scenarios applied in the non-recurrent congestion detections are limited as only the capacity-constraint conditions and event-driven conditions are considered. More research is needed to evaluate the feasibility of detecting a wider variety of non-recurrent events under varying road and traffic conditions. Thirdly, the current study has assumed a maximum CVMPRs of 25%, which could be exceeded in near future with the current fast pace of CV development and adoption. Fourthly, the proposed methodology for performance measurements only considers signalized intersections, which should be extended to cover all types of intersections and road infrastructure.

References

Ahsani, Vesal & Amin-Naseri, Mostafa & Knickerbocker, Skylar & Sharma, Anuj. (2018). Quantitative analysis of probe data characteristics: Coverage, speed bias and congestion detection precision. Journal of Intelligent Transportation Systems. 23. 1-17. 10.1080/15472450.2018.1502667.

Aljamal, Mohammad A., Hossam M. Abdelghaffar, and Hesham A. Rakha. 2020. "Estimation of Traffic Stream Density Using Connected Vehicle Data: Linear and Nonlinear Filtering Approaches" Sensors 20, no. 15: 4066. https://doi.org/10.3390/s20154066

Bin Masud, Saumik Sakib, et al. A Sensitivity Analysis of Freeway Connected and Automated Vehicle Platooning Parameters to Use in Microsimulation. No. TRBAM-23-04161. 2023.

Brennan, T. M., Remias, S. M., & Manili, L. (2015). Performance Measures to Characterize Corridor Travel Time Delay Based on Probe Vehicle Data. Transportation Research Record, 2526(1), 39–50. https://doi.org/10.3141/2526-05

Canadian Automobile Association. Breaking the Bottlenecks: Congestion Solutions for Canada, Executive Summary, 2021. https://www.caa.ca/app/uploads/2021/01/Congestion-solutions-Summary-ENG-V2.pdf

Cao, X., Jiao, J., Zhang, Y., & Wang, X. (2019). Left-Turn Spillback Probability Estimation in a Connected Vehicle Environment. Transportation Research Record, 2673(4), 753–761. https://doi.org/10.1177/0361198119837178

Chauhan, B.P., Joshi, G.J., Parida, P. (2020). Speed Trajectory of Vehicles in VISSIM to Recognize Zone of Influence for Urban-Signalized Intersection. In: Arkatkar, S., Velmurugan, S., Verma, A. (eds) Recent Advances in Traffic Engineering. Lecture Notes in Civil Engineering, vol 69. Springer, Singapore. https://doi.org/10.1007/978-981-15-3742-4_31

Christopher Bucknell, Juan C. Herrera, A trade-off analysis between penetration rate and sampling frequency of mobile sensors in traffic state estimation, Transportation Research Part C: Emerging Technologies, Volume 46, 2014, Pages 132-150, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2014.05.007.

City of Toronto, 10 Most Congested Locations in the City, 2020. https://www.toronto.ca/legdocs/mmis/2013/pw/bgrd/backgroundfile-58159.pdf

Day, C. M., & Bullock, D. M. (2016). Detector-Free Signal Offset Optimization with Limited Connected Vehicle Market Penetration: Proof-of-Concept Study. Transportation Research Record, 2558(1), 54–65. https://doi.org/10.3141/2558-06

Day, C. M., D. M. Bullock, H. Li, S. M. Remias, A. M. Hainen, R. S. Freije, A. L. Stevens, J. R. Sturdevant, and T. M. Brennan. Performance Measures for Traffic Signal Systems: An Outcome-Oriented Approach. Purdue University, West Lafayette, Indiana, 2014. https://doi.org/10.5703/1288284315333

Day, C. M., Li, H., Richardson, L. M., Howard, J., Platte, T., Sturdevant, J. R., & Bullock, D. M. (2017). Detector-Free Optimization of Traffic Signal Offsets with Connected Vehicle Data. Transportation Research Record, 2620(1), 54–68. https://doi.org/10.3141/2620-06

Ehsan Bagheri (2017). Enhanced Traffic Signal Operation using Connected Vehicle Data. UWSpace. http://hdl.handle.net/10012/12261

Fakharian Qom, S., Hadi, M., Xiao, Y., & Al-Deek, H. (2017). Queue Length Estimation for Freeway Facilities: Based on Combination of Point Traffic Detector and Automatic Vehicle Identification Data. Transportation Research Record, 2616(1), 19–26. https://doi.org/10.3141/2616-03

Freije, R. S., Hainen, A. M., Stevens, A. L., Li, H., Smith, W. B., Summers, H., Day, C. M., Sturdevant, J. R., & Bullock, D. M. (2014). Graphical Performance Measures for Practitioners to Triage Split Failure Trouble Calls. Transportation Research Record, 2439(1), 27–40. https://doi.org/10.3141/2439-03

G. Comert and M. Cetin, "Analytical Evaluation of the Error in Queue Length Estimation at Traffic Signals From Probe Vehicle Data," in IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 2, pp. 563-573, June 2011, doi: 10.1109/TITS.2011.2113375.

Gabriel Hiribarren, Juan Carlos Herrera, Real time traffic states estimation on arterials based on trajectory data, Transportation Research Part B: Methodological, Volume 69, 2014, Pages 19-30, ISSN 0191-2615, https://doi.org/10.1016/j.trb.2014.07.003.

Gao, F., You, S., Wang, J., Zhang, S., Xie, D.: Whether and how congested is a road? Indices, updating strategy and a vision-based detection framework. IET Intell. Transp. Syst. 00, 1– 14 (2023). https://doi.org/10.1049/itr2.12342

Gurcan Comert, Mecit Cetin, Queue length estimation from probe vehicle location and the impacts of sample size, European Journal of Operational Research, Volume 197, Issue 1, 2009, Pages 196-202, ISSN 0377-2217, https://doi.org/10.1016/j.ejor.2008.06.024.

Gurcan Comert, Simple analytical models for estimating the queue lengths from probe vehicles at traffic signals, Transportation Research Part B: Methodological, Volume 55,

H. Li, C. M. Day and D. M. Bullock, "Virtual Detection at Intersections using Connected Vehicle Trajectory Data," 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 2016, pp. 2571-2576, doi: 10.1109/ITSC.2016.7795969.

Houchin, Andrew; Dong, Jing; Hawkins, Neal; and Knickerbocker, Skylar, "Measurement and Analysis of Heterogenous VehicleFollowing Behavior on Urban Freeways: Time Headways and Standstill Distances" (2015).Civil, Construction and EnvironmentalEngineering Conference Presentations and Proceedings. 99. https://lib.dr.iastate.edu/ccee_conf/99

Hunter, M., Mathew, J.K., Li, H. and Bullock, D.M. (2021) Estimation of Connected Vehicle Penetration on US Roads in Indiana, Ohio, and Pennsylvania. Journal of Transportation Technologies, 11, 597-610. doi: 10.4236/jtts.2021.114037.

Igor Dakic, Aleksandar Stevanovic, Milan Zlatkovic, Cameron Kergaye, Refinement of Performance Measures based on High-Resolution Signal and Detection Data, Transportation Research Procedia, Volume 22, 2017, Pages 372-381, ISSN 2352-1465, https://doi.org/10.1016/j.trpro.2017.03.055.

INRIX, Ten Highest Traffic Delay Times by City, INRIX Global Traffic Scorecard, 2022. https://inrix.com/scorecard/

Jianfeng Zheng, Henry X. Liu, Estimating traffic volumes for signalized intersections using connected vehicle data, Transportation Research Part C: Emerging Technologies, Volume 79, 2017, Pages 347-362, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2017.03.007.

Juan Argote-Cabañero, Eleni Christofa, Alexander Skabardonis, Connected vehicle penetration rate for estimation of arterial measures of effectiveness, Transportation Research Part C: Emerging

Technologies, Volume 60, 2015, Pages 298-312, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2015.08.013.

Khadka, Swastik, Li, Pengfei “Taylor”, and Wang, Qichao. Developing Novel Performance Measures for Traffic Congestion Management and Operational Planning Based on Connected Vehicle Data. United States: N. p., 2022. Web. doi:10.1061/(asce)up.1943-5444.0000835.

L. Shao, C. Wang, Z. Li and C. Jiang, "Traffic condition estimation using vehicular crowdsensing data," 2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC), Nanjing, China, 2015, pp. 1-8, doi: 10.1109/PCCC.2015.7410307.

Liu, H. X., 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. https://doi.org/10.1016/j.trc.2009.02.003

Liwen Vaughan, Rongbin Yang, Web traffic and organization performance measures: Relationships and data sources examined, Journal of Informetrics, Volume 7, Issue 3, 2013, Pages 699-711, ISSN 1751-1577, https://doi.org/10.1016/j.joi.2013.04.005.

Lloret-Batlle, R., Wang, Z.-H., & Zheng, J. (2023). Traffic Volume Estimation for both Undersaturated and Oversaturated Signalized Intersections With Stopbar Location Estimation Using Trajectory Data. Transportation Research Record, 2677(3), 343–354. https://doi.org/10.1177/03611981221115073

Mahmood, Z. (2020). Connected Vehicles in the IoV: Concepts, Technologies and Architectures. In: Mahmood, Z. (eds) Connected Vehicles in the Internet of Things. Springer, Cham. https://doi.org/10.1007/978-3-030-36167-9_1

Metrolinx and HDR Corp., Costs of Road Congestion on the Greater Toronto and Hamilton Area: Impact and Cost Benefit Analysis of the Metrolinx Draft Regional Transportation Plan, 2008. https://www.metrolinx.com/en/projects-and-programs/regional-transportation-plan

Moghaddam, S. S., & Hellinga, B. (2013). Quantifying Measurement Error in Arterial Travel Times Measured by Bluetooth Detectors. Transportation Research Record, 2395(1), 111–122. https://doi.org/10.3141/2395-13

National Academies of Sciences, Engineering, and Medicine. 2020. Performance-Based Management of Traffic Signals. Washington, DC: The National Academies Press. https://doi.org/10.17226/25875.

Nianfeng Wan, Ardalan Vahidi, Andre Luckow, Reconstructing maximum likelihood trajectory of probe vehicles between sparse updates, Transportation Research Part C: Emerging Technologies, Volume 65, 2016, Pages 16-30, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2016.01.010.

Nicholson, A. & Daniel, Basil & Koorey, Glen. (2023). INVESTIGATING SPEED PATTERNS AND ESTIMATING SPEED ON TRAFFIC-CALMED STREETS.

P. Stuckmann, N. Ehlers and B. Wouters, "GPRS traffic performance measurements," Proceedings IEEE 56th Vehicular Technology Conference, Vancouver, BC, Canada, 2002, pp. 1289-1293 vol.2, https://doi: 10.1109/VETECF.2002.1040813.

Papoulis, A. "Bernoulli Trials." §3-2 in Probability, Random Variables, and Stochastic Processes, 2nd ed. New York: McGraw-Hill, pp. 57-63, 1984.

Quiroga, C. A., & Bullock, D. (1999). Measuring control delay at signalized intersections. Journal of Transportation Engineering, 125(4), 271-280. https://doi.org/10.1061/(ASCE)0733-947X(1999)125:4(271)

Remias, S. M., Day, C. M., Waddell, J. M., Kirsch, J. N., & Trepanier, T. (2018). Evaluating the Performance of Coordinated Signal Timing: Comparison of Common Data Types with Automated Vehicle Location Data. Transportation Research Record, 2672(18), 128–142. https://doi.org/10.1177/0361198118794546

S. -T. Jeng and L. Chu, "A high-definition traffic performance monitoring system with the Inductive Loop Detector signature technology," 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, 2014, pp. 1820-1825, doi: 10.1109/ITSC.2014.6957957.

Saldivar-Carranza, E., Li, H., Mathew, J., Hunter, M., Sturdevant, J., & Bullock, D. M. (2021). Deriving Operational Traffic Signal Performance Measures from Vehicle Trajectory Data. Transportation Research Record, 2675(9), 1250–1264. https://doi.org/10.1177/03611981211006725

Santiago-Chaparro, Kelvin & Chitturi, Madhav & Bill, Andrea & Noyce, David. (2012). Real-time performance measures from radar-based vehicle detection systems. 19th Intelligent Transport Systems World Congress, ITS 2012. AM-00051.

Smaglik, Edward & Bullock, Darcy & Sharma, Anuj. (2007). Pilot Study on Real-Time Calculation of Arrival Type for Assessment of Arterial Performance. Journal of Transportation Engineering-asce - J TRANSP ENG-ASCE. 133. 10.1061/(ASCE)0733-947X(2007)133:7(415).

Smaglik, Edward J.; Sharma, Anuj; Bullock, Darcy M.; Sturdevant, James R.; and Duncan, Gary, "EventBased Data Collection for Generating Actuated Controller Performance Measures" (2007). Civil Engineering Faculty Publications. 21.

Tang, Hua. (2013). Development of a Multiple-Camera Tracking System for Accurate Traffic Performance Measurements at Intersections. Intelligent Transportation Systems Institute, Center for Transportation Studies, University of Minnesota. Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/146949

Tang, Ruotian, Ryo Kanamori, and Toshiyuki Yamamoto. 2020. "Improving Coverage Rate for Urban Link Travel Time Prediction Using Probe Data in the Low Penetration Rate Environment" Sensors 20, no. 1: 265. https://doi.org/10.3390/s20010265

Waddell, J. M., Remias, S. M., Kirsch, J. N., & Young, S. E. (2020). Scalable and Actionable Performance Measures for Traffic Signal Systems using Probe Vehicle Trajectory Data. Transportation Research Record, 2674(11), 304–316. https://doi.org/10.1177/0361198120941847

Wai Wong, Shengyin Shen, Yan Zhao, Henry X. Liu, On the estimation of connected vehicle penetration rate based on single-source connected vehicle data, Transportation Research Part B: Methodological, Volume 126, 2019, Pages 169-191, ISSN 0191-2615, https://doi.org/10.1016/j.trb.2019.06.003.

Wang, X., Jerome, Z., Zhang, C., Shen, S., Kumar, V. V., & Liu, H. X. (2023). Trajectory Data Processing and Mobility Performance Evaluation for Urban Traffic Networks. Transportation Research Record, 2677(3), 355–370. https://doi.org/10.1177/03611981221115088

Z. Arbi, O. Belkahla Driss and M. K. Sbai, "A multi-agent system for monitoring and regulating road traffic in a smart city," 2017 International Conference on Smart, Monitored and Controlled Cities (SM2C), Sfax, Tunisia, 2017, pp. 17-22, doi: 10.1109/SM2C.2017.8071843.

Zeidler, Verena & Buck, Sebastian & Kautzsch, Lukas & Vortisch, Peter & Weyland, Claude. (2019). Simulation of Autonomous Vehicles Based on Wiedemann's Car Following Model in PTV Vissim.

Zhang, X., Hamedi, M., & Haghani, A. (2015). Arterial Travel Time Validation and Augmentation with Two Independent Data Sources. Transportation Research Record, 2526(1), 79–89. https://doi.org/10.3141/2526-09

Zhao, Y., Zheng, J., Wong, W., Wang, X., Meng, Y., & Liu, H. X. (2019). Estimation of Queue Lengths, Probe Vehicle Penetration Rates, and Traffic Volumes at Signalized Intersections using Probe Vehicle Trajectories. Transportation Research Record, 2673(11), 660–670. https://doi.org/10.1177/0361198119856340

Appendix A Code

The code provided outlines the procedures for importing raw simulated Connected Vehicle data and subsequently aggregating the trajectories based on specific spatial and temporal criteria. Additionally, it calculates both the average relative error and the Root Mean Square Error (absolute error).

```
import pandas as pd

# Load the trajectory data

data = pd.read_csv('trajectory_data.csv')

# Define the intersection geometry

intersection_geometry = {'lanes': ..., 'approaches': ..., 'movements': ...}

# For each vehicle, trace the trajectory

vehicles = data['vehicle_id'].unique()

vehicle_trajectories = {}

for vehicle in vehicles:

    vehicle_data = data[data['vehicle_id'] == vehicle]

    trajectory = list(zip(vehicle_data['x_coord'], vehicle_data['y_coord'], vehicle_data['timestamp']))

    vehicle_trajectories[vehicle] = trajectory

# Identify approach and movement for each vehicle

vehicle_approach_movement = {}

for vehicle, trajectory in vehicle_trajectories.items():

    # Analyze the trajectory to identify the approach and movement

    # This is a simplified placeholder - actual implementation can be complex

    approach = identify_approach(trajectory, intersection_geometry)

    movement = identify_movement(trajectory, intersection_geometry)

    vehicle_approach_movement[vehicle] = (approach, movement)

# Aggregate the data based on approach and movement
```

```
aggregated_data = {}
for vehicle, (approach, movement) in vehicle_approach_movement.items():
    if (approach, movement) not in aggregated_data:
        aggregated_data[(approach, movement)] = []
aggregated_data[(approach, movement)].append(vehicle_trajectories[vehicle])
import numpy as np
# Define the total number of vehicles
N_v = ...
# Define the desired CV market penetration rate
p = ...
# Generate a random boolean mask for selecting vehicles
mask = np.random.choice([True, False], size=N_v, p=[p, 1-p])
# Assume all_vehicle_data is a list of dictionaries, each representing a vehicle's data
all_vehicle_data = ...
# Select a subset of vehicles to be the connected vehicles
cv_vehicle_data = [data for data, m in zip(all_vehicle_data, mask) if m]
# Run the simulation and record trajectory data
for data in cv_vehicle_data:
    vehicle_id = data['id']
    # Run simulation for this vehicle and record data
# Repeat the procedure for a specified number of simulations
N = ...
all_D = []
all_D_c = []
```

```
for _ in range(N):
    # Run simulation and calculate D and D_c
    D = ...
    D_c = ...
    all_D.append(D)
    all_D_c.append(D_c)
# Calculate the mean values
mean_D = np.mean(all_D)
mean_D_c = np.mean(all_D_c)
import pandas as pd
import numpy as np
# Load data into DataFrame
data = pd.read_csv('your_data_file.csv')
# Calculate the delay for each vehicle
data['delay'] = data['entry_time'] - data['exit_time']
# Calculate the average overall delay
average_delay = data['delay'].mean()
# Calculate the 95th percentile queue length
queue_length_95 = np.percentile(data['queue_length'], 95)
# Identify stops based on velocity threshold (velocity should be in km/h)
velocity_threshold = 3
data['stopped'] = np.where((data['velocity'].shift(-1) <= velocity_threshold) & (data['velocity'] >
velocity_threshold), 1, 0)
# Calculate the percentage of stopped vehicles
```

```
percentage_stopped_vehicles = (data['stopped'].sum() / len(data)) * 100
print(f"Average overall delay: {average_delay}")
print(f"95th percentile queue length: {queue_length_95}")
print(f"Percentage of stopped vehicles: {percentage_stopped_vehicles}")

import numpy as np
def calculate_rmse(actual_values, estimated_values):
    n = len(actual_values)
    squared_errors = [(actual - estimated)**2 for actual, predicted in zip(actual_values,
estimated_values)]
    mean_squared_error = sum(squared_errors) / n
    rmse = np.sqrt(mean_squared_error)
    return rmse
```

Appendix B Signal Timing of Each Intersection

Figure B-1 The Signal Timing of Intersection No.1

Figure B-2 The Signal Timing of Intersection No.2

Figure B-3 The Signal Timing of Intersection No.3

Figure B-4 The Signal Timing of Intersection No.4

Figure B-5 The Signal Timing of Intersection No.5

Figure B-6 The Signal Timing of Intersection No.6

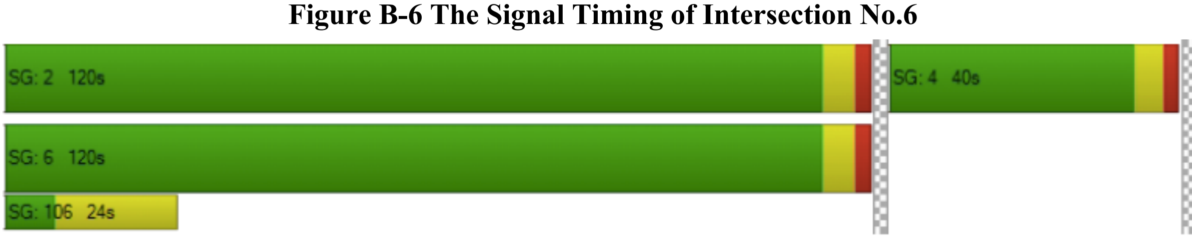


Figure B-7 The Signal Timing of Intersection No.7

Figure B-8 The Signal Timing of Intersection No.8

Figure B-9 The Signal Timing of Intersection No.9

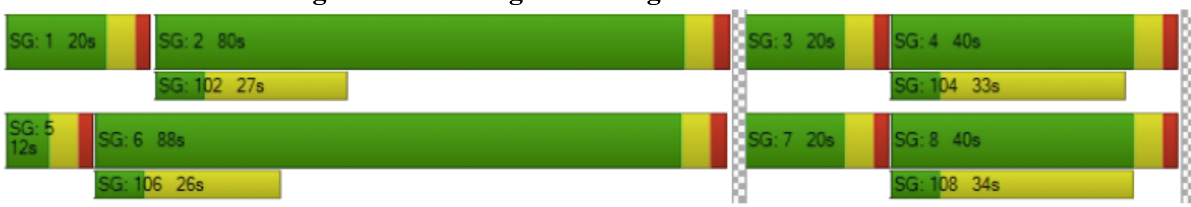


Figure B-10 The Signal Timing of Intersection No.10

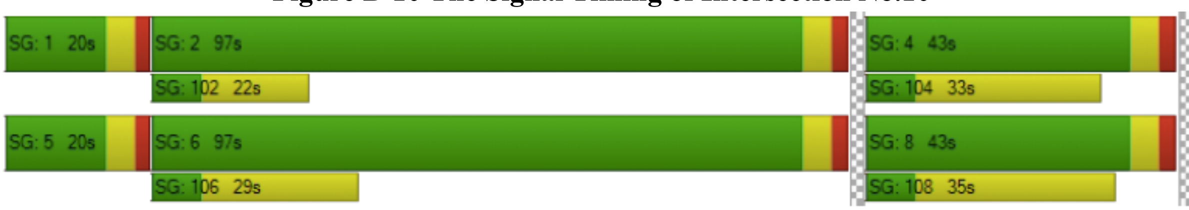


Figure B-11 The Signal Timing of Intersection No.11

Figure B-12 The Signal Timing of Intersection No.12

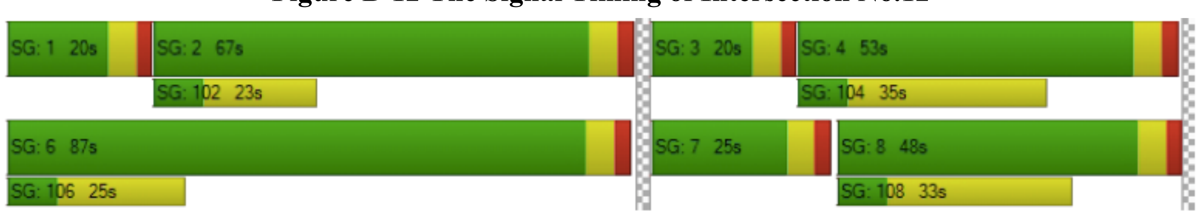


Figure B-13 The Signal Timing of Intersection No.13

Figure B-14 The Signal Timing of Intersection No.14

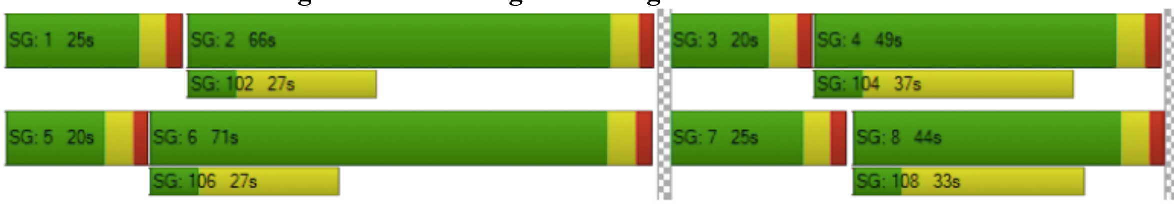


Figure B-15 The Signal Timing of Intersection No.15

Figure B-16 The Signal Timing of Intersection No.16

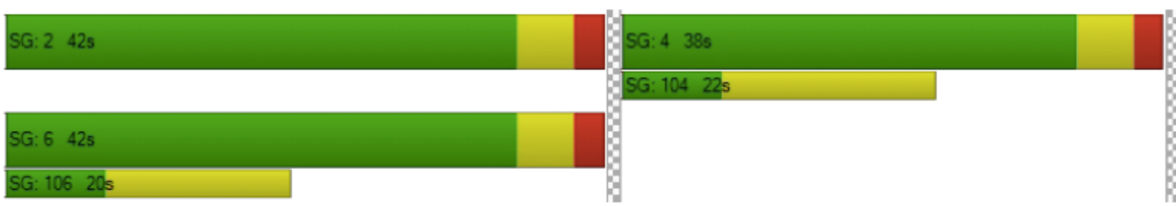


Figure B-17 The Signal Timing of Intersection No.17

Appendix C: Lane Configuration and Movements of the Individual Intersections

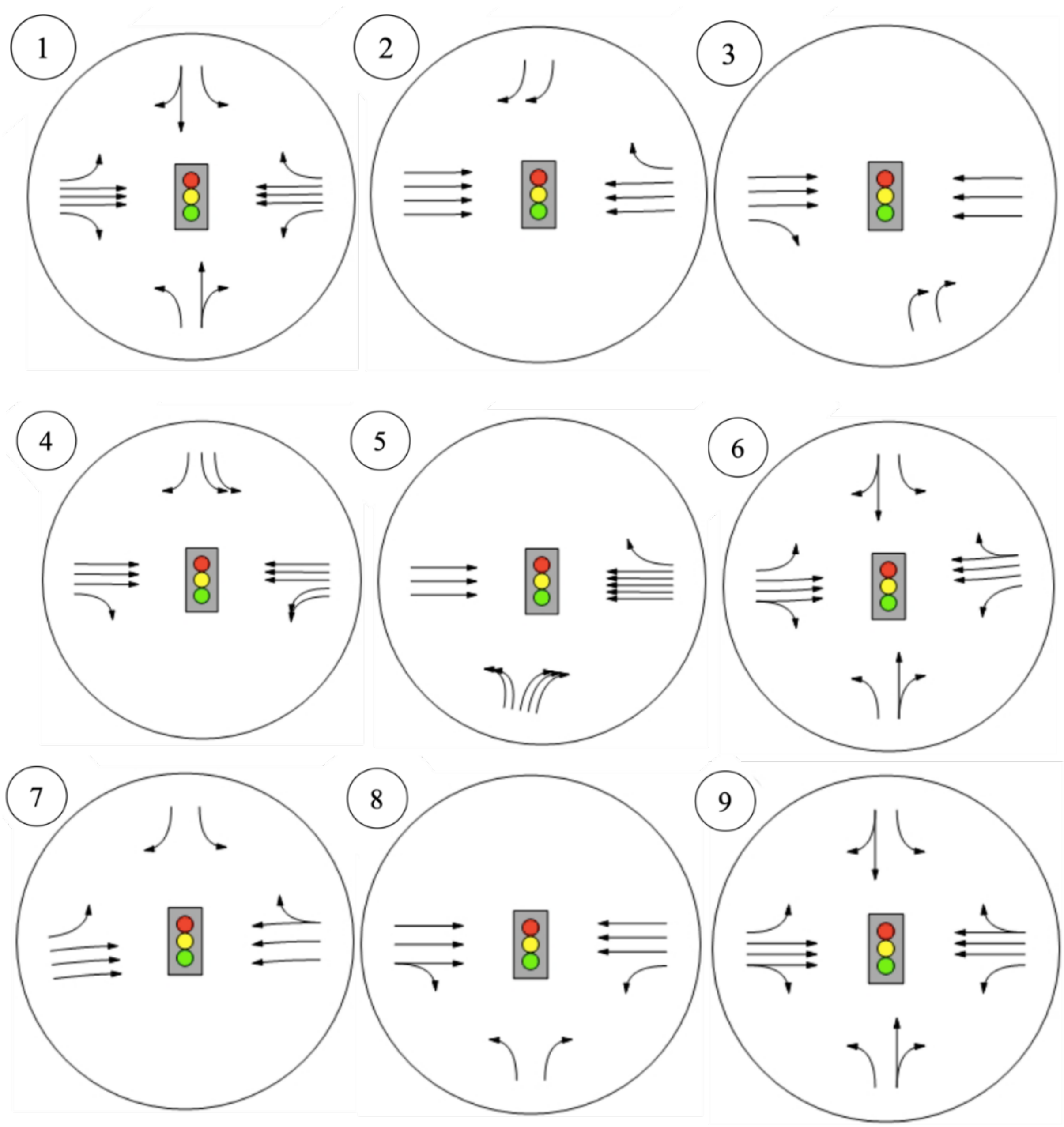


Figure C The Geometric Layout of Each Intersection along The Broward Blvd.

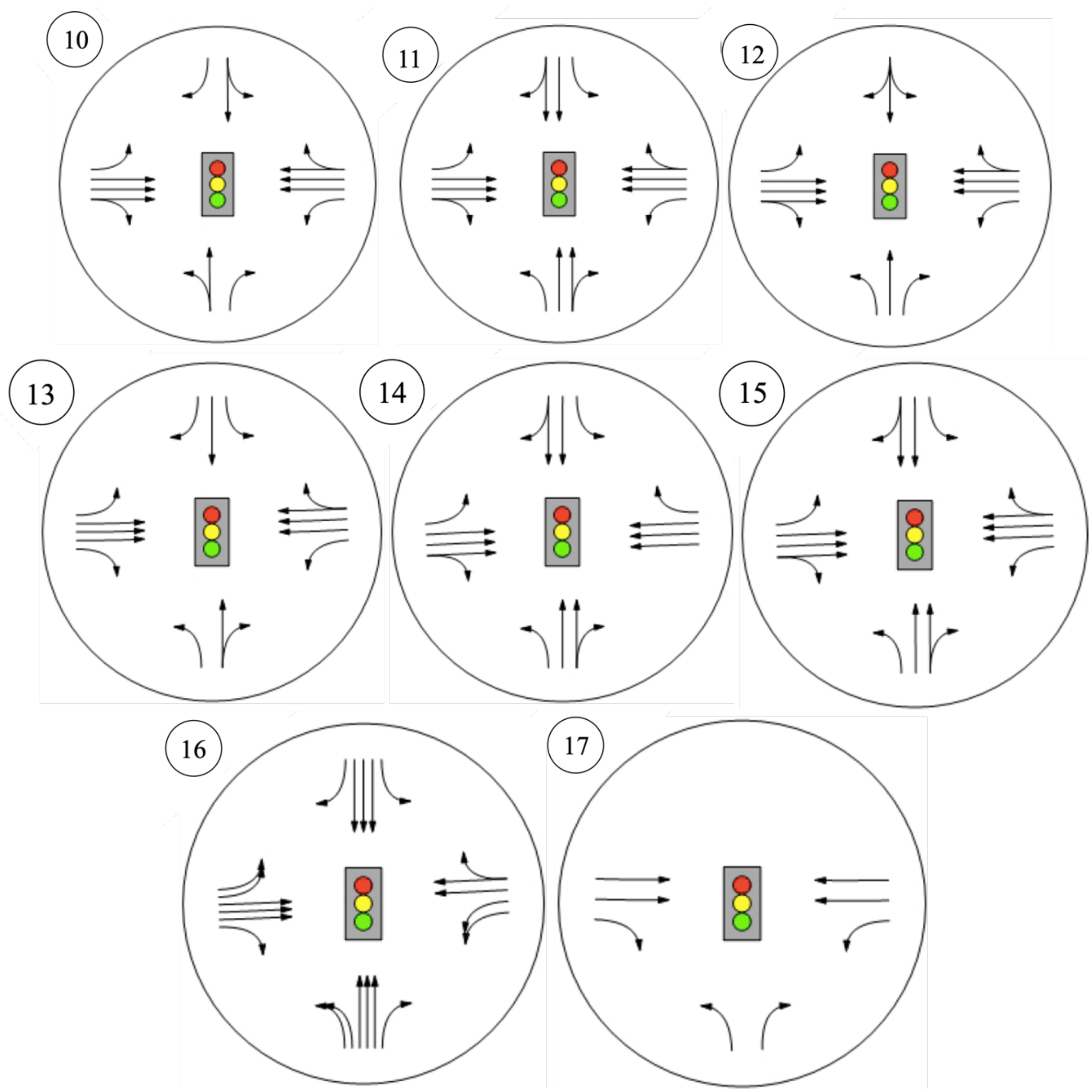


Figure C The Geometric Layout of Each Intersection along The Broward Blvd.