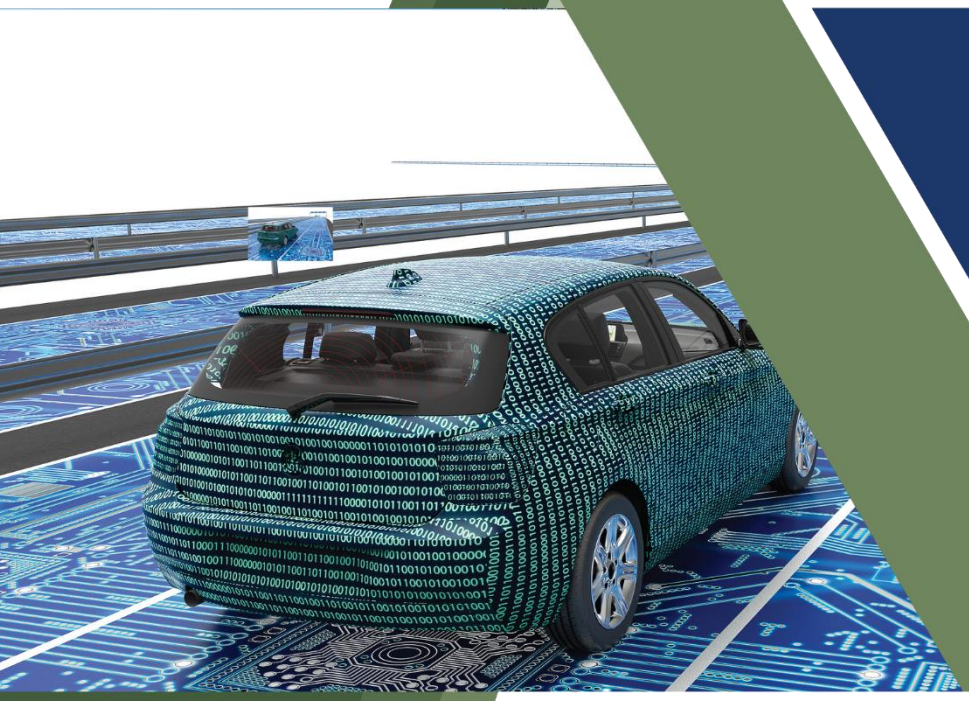


FINAL REPORT

Project C

July 2021

Performance Measurement & Management using Connected & Automated Vehicle Data

Dr. Mohammed Hadi | Florida International University
Dr. Virginia Sisiopiku | University of Alabama at Birmingham
Dr. Siva Srinivasan | University of Florida
Dr. Ossama E. Ramandan | CDM Smith
Md. Ahsanul Islam | Florida International University
Dr. Leila Azizi | Florida International University
Siddhartha Gulhare | University of Florida
Dr. Tao Wang | Florida International University
Dr. Shahadat Iqbal | Florida International University

STRIDE | Southeastern Transportation Research, Innovation, Development and Education Center

UF | Transportation Institute UNIVERSITY of FLORIDA

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. Project C	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Performance Measurement & Management using Connected & Automated Vehicle Data		**5. Report Date** 7/14/2021 **6. Performing Organization Code**
7. Author(s) Dr. Mohammed Hadi, Florida International University Dr. Virginia Sisiopiku, University of Alabama at Birmingham Dr. Siva Srinivasan, University of Florida		**8. Performing Organization Report No.** STRIDE Project C
9. Performing Organization Name and Address **Florida International University**, Civil & Environmental Engineering,10555 W. Flagler Street, EC 3680, Miami, FL 33174 **University of Florida**, College of Design, Construction & Planning, PO Box 115701, Gainesville, FL 32611		**10. Work Unit No.** **11. Contract or Grant No.** Funding Agreement Number 69A3551747104
12. Sponsoring Agency Name and Address University of Florida Transportation Institute Southeastern Transportation Research, Innovation, Development and Education Center (STRIDE) 365 Weil Hall, P.O. Box 116580 Gainesville, FL 32611 U.S Department of Transportation/Office of Research, Development & Tech 1200 New Jersey Avenue, SE Washington, DC 20590 United States		**13. Type of Report and Period Covered** 5/1/2017 to 7/14/2021 **14. Sponsoring Agency Code**

15. Supplementary Notes

16. Abstract

The availability of connected vehicle (CV) data, even at lower market penetrations, can be sufficient to support critical transportation performance measurement and management functions. This study developed a framework, methods, and algorithms for using CV data to estimate measures to support agency processes. As such, the study investigated the use of CV data to estimate metrics that can be currently estimated using existing data sources including those related to mobility, reliability, and environmental impacts. In addition, the study investigated the estimation and utilization of additional mobility and safety metrics that cannot be estimated based on existing sources of data. The developed framework and methods to estimate performance measures can be used by a system operator, a planner, or an automated system to support decisions associated with the agency business processes. The methods can be also used in the real-time operations of traffic management centers (TMCs) to determine the traffic states. In addition, machine learning models were developed for use by the TMCs for short-term prediction of traffic conditions to support proactive activation of operational plans to mitigate potential deterioration in mobility and safety performance.

17. Key Words connected vehicles, performance measurement, transportation management and operations, data analytics	**18. Distribution Statement** No restrictions		
19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages** 227 Pages	**22. Price**

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

ACKNOWLEDGEMENT OF SPONSORSHIP AND STAKEHOLDERS

This work was sponsored by a contract from the Southeastern Transportation Research, Innovation, Development and Education Center (STRIDE), a Regional University Transportation Center sponsored by a grant from the U.S. Department of Transportation's University Transportation Centers Program.

LIST OF AUTHORS

Lead PI:

Mohammed Hadi, Ph.D.
Professor, Department of Civil & Environmental Engineering, Florida International University
(305) 348-0092 |hadim@fiu.edu
ORCID 0000-0003-2233-8283

Co-PI:

Virginia P. Sisiopiku, Ph.D.
Professor, Department of Civil, Construction and Environmental Engineering, University of Alabama at Birmingham
(205) 934-9912 | vsisiopi@uab.edu
ORCID 0000-0003-4262-8990

Sivaramakrishnan (Siva) Srinivasan, Ph.D.
Associate Professor, Department of Civil and Coastal Engineering, University of Florida
352-294-7807 |siva@ce.ufl.edu
ORCID 0000-0002-9008-0192

Additional Researchers:

Ossama E. Ramadan, Ph.D.
Postdoctoral Fellow, Department of Civil, Construction, and Environmental Engineering, University of Alabama at Birmingham
(205) 934-1574 | oramadan@uab.edu

Md. Ahsanul Islam
Graduate Research Assistant, Department of Civil, Construction, and Environmental Engineering, University of Alabama at Birmingham
(205) 934-1574 | ahsan08@uab.edu

Leila Azizi, Ph.D.
Former Graduate Research Assistant, Department of Civil and Environmental Engineering, Florida International University, Miami, FL.
(305) 348-1393 | laziz001@fiu.edu

Siddhartha Gulhare
Graduate Research Assistant, Department of Civil and Coastal Engineering, University of Florida | gulhare.s@ufl.edu

Tao Wang, Ph.D.
Research Associate, Department of Civil and Environmental Engineering, Florida International University, Miami, FL.
(305) 348-1393 |tao5431@gmail.com

Shahadat Iqbal, Ph.D.
Research Associate, Department of Civil and Environmental Engineering, Florida International University, Miami, FL.
(305) 348-1393 | miqba005@fiu.edu

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AASHTO	American Association of State Highway and Transportation Officials
AADT	Average Annual Daily Traffic
AHTD	Annual Hours of Truck Delay
ALDOT	Alabama Department of Transportation
AMCD	Advanced Messaging Concept Development
ANN	Artificial Neural Networks
ANOVA	Analysis of variance
BI	Buffer Index
BSM	Basic Safety Message
CALTRANS	California Department of Transportation
CCTV	Closed-circuit Television
CD	Control Delay
CHTS	California Household Travel Survey
CMAQ	Congestion Mitigation and Air Quality
CMEM	Comprehensive Modal Emissions Model
CMS	Changeable Message Sign
CV	Connected Vehicles
CV	Connected Automated Vehicles
DMS	Dynamic Message Signs
DRAC	Deceleration Rate to Avoid the Crash
DSRC	Dedicated Short-Range Communication
DSS	Decision Support System
EF	Emission Factor
EIA	U.S. Energy Information Agency
FFS	Free Flow Speed
FFTT	Free Flow Travel Time
FHWA	Federal Highway Administration
FN	False Negative
FP	False Positive
FR	Flowrate
GHG	Greenhouse Gas Emitter
GMM	Gaussian Mixture Models
GPS	Global Positioning System
GUI	Graphical User Interface
HAR	Highway Advisory Radio
HCM	Highway Capacity Manual
HPMS	Highway Performance Monitoring System
INFLO	Intelligent Network Flow Optimization
ITS	Intelligent Transportation Systems
JPO	Joint Program Office
LDT	Light Duty Truck

LDV	Light Duty Vehicle
LOS	Levels of Service
LOTTR	Level of Travel Time Reliability
MAPE	Mean Absolute Percent Error
MMITSS	Multi-Modal Intelligent Traffic Signal Systems
MOVES	Motor Vehicle Emission Simulator
NCHRP	National Cooperative Highway Research Program
NGSIM	Next Generation SIMulation
NHS	National Highway System
NHTS	National Household Travel Survey
NO	Number of Oscillations
NPMRDS	National Performance Management Research Data Set
NREL	National Renewable Energy Laboratory
NVP	Number of Vehicles in the Platoon
OBD	On-board Diagnostic
OOB	Out Of Bag
ORNL	Oak Ridge National Laboratory
PCA	Principal Component Analysis
PET	Post Encroachment Time
PF	Project Adjustment Factor
PHED	Peak Hour Excessive Delay
PIP	Percentage of Vehicles in the Platoon
PL	Procedural Language
PLS	Partial Least Square
PRT	Perception Reaction Time
PTI	Planning Time Index
RCI	Rear-end Crash Index
RCI	Roadway Congestion Index
RF	Random Forest
RPCGB	Regional Planning Commission of Greater Birmingham
RSU	Roadside Units
SAE	Society of Automotive Engineers
SD	Standard Deviation
SDF	Stopping Distance of the Following vehicle
SDI	Stopping Distance
SDL	Stopping Distance of the Leading vehicle
SDPE	Standard Deviation of Percentage Error
SEMI-ODE	Southeast Michigan Operational Data Environment
SND	Speed Normal Deviate
SPaT	Signal and Phasing Timing
SPMD	Safety Pilot Model Deployment
SQL	Structured Query Language
STRIDE	Southeastern Transportation Research, Innovation, Development, and Education Center

SVM	Support Vector Machine
TCA	Trajectory Conversion Algorithm
TCA	Trajectory Conversion Algorithm
TET	Time Exposed Time–to–collision
TETI	The Index of Time Exposed to Time to Collision
TIT	Time Integrated Time-to-collision
TMC	Traffic Management Centers
TN	True Negative
TNM	Traffic Noise Model
TP	True Positive
TRI	Travel Rate Index
TSMO	Transportation System Management and Operations
TTC	Time–to–collision
TTMI	Travel Time Misery Index
TTPV	Travel Time Percent Variation
TTR	Travel Time Reliability
TTTR	Truck Travel Time Reliability
VIIPD	Vehicle Infrastructure Integration Probe Data
VISSIM	Verkehr In Städten SIMulationsmodell
VMS	Variable Message Signs
VMT	Vehicle Miles of Travel
VSP	Vehicle Specific Power
VT-Micro	Virginia Tech Microscopic Energy and Emission Model
XGB	Extreme Gradient Boosting

ABSTRACT

The availability of connected vehicle (CV) data, even at lower market penetrations, can be sufficient to support critical transportation performance measurement and management functions. This study developed a framework, methods, and algorithms for using CV data to estimate measures to support agency processes. As such, the study investigated the use of CV data to estimate metrics that can be currently estimated using existing data sources including those related to mobility, reliability, and environmental impacts. In addition, the study investigated the estimation and utilization of additional mobility and safety metrics that cannot be estimated based on existing sources of data. The developed framework and methods to estimate performance measures can be used by a system operator, a planner, or an automated system to support decisions associated with the agency business processes. The methods can be also used in the real-time operations of traffic management centers (TMCs) to determine the traffic states. In addition, machine learning models were developed for use by the TMCs for short-term prediction of traffic conditions to support proactive activation of operational plans to mitigate potential deterioration in mobility and safety performance.

Keywords: Connected Vehicles, Performance Measurement, Transportation Management and Operations, Data Analytics

EXECUTIVE SUMMARY

This study investigated the use of CV data to estimate metrics that can be currently estimated using existing data sources including those related to mobility, reliability, and environmental impacts. In addition, the study investigated the estimation and utilization of additional mobility and safety metrics that cannot be estimated based on existing sources of data.

To explore the potential for using new measures in assessing system performance based on CV data, the study examined the use of microscopic measures in combination with the usually used macroscopic measures for traffic congestion evaluation, traffic state categorization, traffic flow breakdown prediction, and estimation of traffic safety. The macroscopic measures are the mean speed, traffic flow rate, and occupancy. The investigated microscopic measures for the stated purpose are the location and speed of individual vehicles, acceleration/deceleration of individual vehicles, standard deviations of the speeds of individual vehicles, standard deviations of speed between vehicles, and two disturbance measures. The disturbance measures to capture the stop-and-go operations are the number of oscillations and a measure of disturbance durations in terms of the time exposed time–to–collision (TET), which has been used in other studies as a safety surrogate measure. However, this measure of disturbance duration requires the locations and speeds of both the leading and following vehicles and therefore cannot be measured accurately with low sample sizes of CV. Thus, this study derived models to estimate this measure based on speed parameters. In addition, machine learning models were developed for use in real-time for short-term prediction of traffic conditions to support proactive activation of operational plans to mitigate potential deterioration in performance.

A second part of the study developed a detailed methodological framework to describe the process of determining performance measures based on CV data. The proposed methodological framework has four parts, namely: (i) physical data flow diagram, (ii) processes and process groups hierarchical diagram, (iii) individual process designs, and (iv) logical data flow diagram. The proposed methodological framework addressed the data aggregation issue and introduced data aggregation algorithms in the field to estimate performance measures. In order to validate the proposed framework, the study utilized a simulation model of a 14-mile long section of interstate I-65 between exit 247 and 261A in Birmingham, AL.

A third part of the study investigated the application of limited connected vehicle data to estimate real time pollutant emission and fuel consumption of the transportation network. This study applied emission models to a large trajectory dataset for different vehicle categories and soak time.

1. INTRODUCTION

This study investigated the use of CV data to estimate metrics that can be currently estimated using existing data sources including those related to mobility, reliability, and environmental impacts. In addition, the study investigated the estimation and utilization of additional mobility and safety metrics that cannot be estimated based on existing sources of data.

1.1 BACKGROUND

The estimation of performance measures is critical to transportation system planning, planning for operations, operations, and management process. The performance measures are usually estimated based on existing technologies include travel time, travel time reliability, volume, density/occupancy, vehicle classification, and incident occurrence. Queue length, back of queue, and emission have also been derived based on the above measures. These measures have been estimated based on existing detection technologies such as point detectors and vehicle matching technologies. Signalized intersection movement measures have also been derived based on high resolution controller data. Data from third party vendors collected using probe vehicles have also been widely used, particularly travel time and origin-destination measures.

The availability of connected vehicle (CV) data, even at lower market penetrations, can be sufficient to support critical transportation performance measurement and management functions. Connected vehicle (CV) technologies promise to allow the estimation of measures currently provided by other technologies, as well as measures that cannot be collected by existing sensor technologies. Examples of the additional measures include stops, accelerations, and decelerations, shockwave speed, detailed signalized intersection movement-level measures, potential for crashes, weather impacts, and emissions. A relatively low market penetration of CV may be required for estimating some of the measures, while other measures will require high market penetrations to produce accurate results.

This study developed a framework, methods, and algorithms for using CV data to estimate measures to support agency processes. As such, the study investigated the use of CV data to estimate metrics that can be currently estimated using existing data sources including those related to mobility, reliability, and environmental impacts. In addition, the study investigated the estimation and utilization of additional mobility and safety metrics that cannot be estimated based on existing sources of data. The developed framework and methods to estimate performance measures can be used by a system operator, a planner, or an automated system to support decisions associated with the agency business processes. The methods can be also used in the real-time operations of traffic management centers (TMCs) to determine the traffic states. In addition, machine learning models were developed for use by the TMCs for short-term prediction of traffic conditions to support proactive activation of operational plans to mitigate potential deterioration in mobility and safety performance.

1.2 OBJECTIVES

The goal of this study is to investigate the use of data collected using connected vehicle, combined with data from other sources, to support the performance measurement of transportation system for planning and operation purposes. The specific objectives are:

- Develop a methodological framework to describe the process of determining performance measures based on CV data
- Examine the use of microscopic traffic measures in combination with the usually used macroscopic traffic measures for traffic congestion evaluation, traffic state categorization, traffic flow breakdown prediction, environmental impacts estimation, and traffic safety assessment
- Demonstrate the accuracy of estimating different performance measures using CV data, possibly in combination with other data sources.

1.3 OVERVIEW AND REPORT ORGANIZATION

To explore the potential for using new measures in assessing system performance based on CV data, the study examined the use of microscopic measures in combination with the usually used macroscopic measures for traffic congestion evaluation, traffic state categorization, traffic flow breakdown prediction, and estimation of traffic safety. The macroscopic measures are the mean speed, traffic flow rate, and occupancy. The investigated microscopic measures for the stated purpose are the location and speed of individual vehicles, acceleration/deceleration of individual vehicles, standard deviations of the speeds of individual vehicles, standard deviations of speed between vehicles, and two disturbance measures. The disturbance measures to capture the stop-and-go operations are the number of oscillations and a measure of disturbance durations in terms of the time exposed time–to–collision (TET), which has been used in other studies as a safety surrogate measure. However, this measure of disturbance duration requires the locations and speeds of both the leading and following vehicles and therefore cannot be measured accurately with low sample sizes of CV. Thus, this study derived models to estimate this measure based on speed parameters. In addition, machine learning models were developed for use in real-time for short-term prediction of traffic conditions to support proactive activation of operational plans to mitigate potential deterioration in performance.

A second part of the study developed a detailed methodological framework to describe the process of determining performance measures based on CV data. The proposed methodological framework has four parts, namely: (i) physical data flow diagram, (ii) processes and process groups hierarchical diagram, (iii) individual process designs, and (iv) logical data flow diagram. The proposed methodological framework addressed the data aggregation issue and introduced data aggregation algorithms in the field to estimate performance measures. In order to validate the proposed framework, the study utilized a simulation model of a 14-mile long section of interstate I-65 between exit 247 and 261A in Birmingham, AL.

A third part of the study investigated the application of limited connected vehicle data to estimate real time pollutant emission and fuel consumption of the transportation network. This study applied emission models to a large trajectory dataset for different vehicle categories and soak time.

This report is organized in four chapters. The first chapter is the introduction which describes the project background and objectives. The second chapter describes a methodology to determine freeway performance based on combinations of macroscopic and microscopic measures such as platooning measures, safety measures, traffic state identification, traffic safety, and traffic flow breakdown prediction considering connected vehicle environment. Chapter 3 describes the framework and proof of concept for the utilization of CV data for the performance measurement to support transportation planning and operation applications. The last chapter (Chapter 4) describes a methodology to measure emission-based performance measurement using speed/acceleration data collected from CV's.

2. FREEWAY PERFORMANCE MEASUREMENT USING COMBINATION OF MACROSCOPIC AND MICROSCOPIC MEASURES

2.1 INTRODUCTION

Transportation system performance is a key component in congestion management, traffic safety management, setting agency priorities, and making policy decisions. Emerging connected and automated vehicle technologies (CV/CAV), shared autonomy, and shared mobility will significantly affect the demand and supply of the transportation network. They will also increase data quantity and quality, allowing the use of better performance measures and better estimation of existing measures.

The increase in the market penetration of connected vehicles in the coming years will provide an important source of data for planning, planning for operations, and operations and management of transportation systems. The improved quality, quantity, details, and types of data provided by CV will allow for a better estimation of system performance, and the development and application of more effective strategies based on this estimation. Transportation System Management and Operations (TSMO) agencies are currently collecting data using point detectors, automatic vehicle identification technologies such as Bluetooth and Wi-Fi readers, video analytics, and private sector vendor data. The parameters currently obtained and used based on these data sources usually include volume, speed, occupancy/density, and travel time measurements. The National CV Field Infrastructure Footprint Analysis document, produced by the American Association of State Highway and Transportation Officials (AASHTO) (Wright, James, J. Kyle Garrett, Christopher J. Hill, Gregory D. Krueger, Julie H. Evans, Scott Andrews, Christopher K. Wilson, Rajat Rajbhandari, 2014) recommends that public agencies should assess and trade-off the opportunities to use connected vehicle probe data aggregation and processing versus the continued deployment, operations, and maintenance of traditional ITS vehicle detection versus purchasing private sector data.

Due to the limited data availability, the traffic mobility and safety performance estimations in real-time operations have been mainly based on the three fundamental macroscopic measures (speed, occupancy, and volume). The introduction of connected vehicles, connected automated vehicles, and advanced infrastructure sensors will allow the collection of microscopic measures that can be used in combination with the macroscopic measures for better estimation of traffic mobility and safety than what can be done with the macroscopic measures by themselves. There are a number of studies that have investigated using CV data for performance measurement. However, these studies have mainly focused on measures that can be estimated by existing technology. Also, there have been limited efforts to investigate the potential of taking advantage of the more detailed data obtained from CV in deriving additional microscopic measures for use in assessing system operations. The availability of CV data will allow for the collection of parameters such as the distributions of the time headways between vehicles, variations of the speed between vehicles, variations of the speed of each vehicle, and variations

of acceleration of each vehicle. A relatively low market penetration of CV may be required for estimating some of the detailed measures, while other measures will require high market penetrations to produce accurate results. Measures such as time headway, volume, and density will not be available based on data from low market penetrations of CV. Thus, other surrogate measures are needed for the estimation of performance measures at lower CV market penetrations.

Based on NCHREP report 551(NCHRP, 2006), the most commonly used criteria by agencies to select performance measurements for utilization are: easy to understand, well defined and quantifiable, describing existing conditions, predictability, accuracy of precision, variability by transportation alternatives, and consistently interpretability. The newly derived measures will provide agencies with additional capabilities to estimate system performance.

1.1 Objectives

The goal of this task is to investigate the use of microscopic measures that can be estimated in real-time operations based on CV/CAV data in combination with the commonly used macroscopic measures to estimate and predict system performance. For this purpose, this study defines additional microscopic measures to quantify traffic disturbances. The disturbance measures are utilized to capture the stop-and-go operations and include the number of oscillations and a measure of disturbance durations in terms of the time exposed time-to-collisions, which has been used in other studies as a safety surrogate measure. These microscopic measures are used in congestion evaluation based on platooning, traffic state recognition, traffic breakdown prediction, and safety assessment using trajectory data at low market penetrations of CV data. The findings from this research will improve agency decision-making and allow optimized operations and better outcome performance.

1.2 Scope

The specific scope of this task are:

- Assessing the level of traffic platooning as an indication of traffic congestion using surrogate measures that can be estimated at low market penetrations of CV data
- Developing disturbance measures based on the time exposed time–to–collision (TET) index and the number of oscillations as indicators of traffic flow instability and unsafe conditions
- Developing models to estimate the TET index based on other measures at low market penetrations of CV data
- Developing methods for traffic state identification utilizing combinations of microscopic and macroscopic measures
- Developing methods for real-time prediction of traffic state based on combinations of microscopic and macroscopic measures

- Developing methods for real-time traffic safety estimation utilizing combinations of microscopic and macroscopic measures

2.2 LITERATURE REVIEW

This section first provides information about CV data elements, available real word CV data, and its applications to transportation system. It then presents a literature review related to study's purposes including: (a) platooning formation (b) traffic state classification, (c) traffic flow state prediction and (d) safety assessment.

2.2.1 Connected Vehicle Data Elements

Connected vehicle data are generated from vehicles and communicated to either the roadside units (RSU) or central facilities for processing and use. These data are useful for mobility, safety, and environmental applications. Obtaining some of the useful data elements requires connection to the vehicles on-board diagnostic port (OBD-II). The data is transmitted using connected vehicle messages utilizing dedicated short-range communication (DSRC) or other communication technologies such as cellular communications. The connected vehicle message types and components are specified in the Society of Automotive Engineers (SAE) J2735 standards (SAE International, 2009).

The basic safety message (BSM), specified in J2735, contains vehicle safety-related information broadcasted to surrounding vehicles, but can be also sent and/or captured by the infrastructure. The BSM, as defined in the J2735 standards, consists of two parts. Part 1 is sent in every BSM message broadcasted ten times per second. It contains core data elements, including vehicle position, heading, speed, acceleration, steering wheel angle, and vehicle size. Part 2 consists of a large set of optional elements such as precipitation, air temperature, wiper status, light status, road coefficient of friction, antilock brake system activation, traction control system activation, and vehicle type. However, not all of these parameters are currently available from every vehicle.

CV data communication involves communications between vehicles, infrastructure and road users to talk each other. This is normally categorized as Vehicle to Vehicle (V2V), Vehicle to Infrastructure (V2I), and Vehicle to Everything (V2X) such as to pedestrians and bicycles. V2I communication provides significant opportunity to collect CV data for use in measuring system performance.

A key challenge is the collection, storage and analysis of data from connected vehicles. The amount of this data is considerably larger than traditional transportation data collected by existing sensors such as speed, volume and, and occupancy measurements. CV offers the opportunity to collect detailed microscopic data from each vehicle ten times per second.

2.2.1.1 Data from Real-World Deployments of Connected Vehicles

One of the important considerations of the real-world deployments of the US Department of Transportation's intelligent transportation systems, Joint Program Office (JPO) is to share the collected ITS including CV data with the users. To date, the available connected vehicles data on freeways is summarized in Table 2-1 (ITS DataHub, 2019).

TABLE 2-1 AVAILABLE CV DATA ON FREEWAYS FROM US FUNDED DEPLOYMENTS

CV Data	Short Description	Location
Safety Pilot Model Deployment (SPMD)	This data provides BSM, vehicle trajectories, and various driver-vehicle interaction data	Ann Arbor, Michigan
Wyoming CV Pilot BSM Sample	This is a live running log of sanitized BSM	Wyoming
Tampa CV Pilot BSM Sample	Generates data from the interaction between vehicles and between vehicles and infrastructure	Tampa, FL

2.2.1.2 Mapping Performance Measures to System Applications

The Wyoming CV pilot demonstration team has identified the applications of performance measures in nine categories focusing on improvement of traffic safety and mobility (Hartman et al., 2016). The measure applications were categorized in: (1) improving performance during bad weather condition, (2) improving ability of traffic management centers (TMC) to generate alerts and advisories, (3) efficiently disseminate traveler information, (4) effectively disseminating and receiving I2V and V2I alerts from the TMC, (5) improving information to fleet manager, (6) effectively transmitting and receiving V2V messages, (7) automatic emergency notification of crashes, (8) reducing speed variation, and (9) reducing crashes.

2.2.2 Platooning Measures

This section first defines platoon and then reviews existing studies that developed and utilized methods to estimate platooning in congestion evaluation. These studies mainly focus on the estimation of the percentage of vehicles in the platoon and the platoon size distribution based on macroscopic and microscopic traffic flow parameters.

As the flow rate increases, two or more vehicles will form platoons. Platooning is important as it has serious implications on traffic operations. The Highway Capacity Manual (HCM) (2016) defines platoon as “a group of vehicles traveling together either voluntarily or involuntarily because of signal control, geometric or other factors.” A vehicle with a higher desired speed catches up with a slow vehicle and is forced to follow it, reducing its speed to maintain its desired following distance.

Traffic flow breakdown on a freeway and its stochastic nature are strongly related to the platooning and platoon size (Shiomi et al., 2011). Shiomi et al. (2011) defined the traffic breakdown probability as a function of platoon size and its lead vehicle’s speed, given the traffic

flow rate. They also utilized a probability density function of the appearance of a platoon of size of "x," in which "x" is the number of vehicles in the platoon. Thus, the likelihood of a platoons of a given size can be estimated utilizing this function. In addition to their use in determining the probability of breakdown, the platooning characteristics can be used as criteria to determine the level of service of a facility. When drivers are forced to adjust their speed and follow the leading vehicle, they perceive lower serviceability. Platooning also indicates lower maneuverability, in terms of the ability to change lanes, once again indicating a lower level of service. Furthermore, there is an increase in conflicts with the presence of platoons, indicating an impact on safety. A previous study showed that the number of platooned vehicles over one mile can be used as a better serviceability criterion in multi-lane highways, compared to density (Chatterjee et al., 2017). Thus, measuring or estimating platooning attributes, as part of traffic operations and management, will allow TSMO agencies to make better decisions to implement active traffic management strategies to reduce the probability of breakdown and improve the level of service. The estimation of the platooning characteristics on a freeway segment is important for both planning and operation analysis and can be used to assess the level of operations and service of traffic, as well as to potentially predict the safety of the traffic stream.

The relationship between platooning and traffic flow stability has been a subject of research in traffic flow theory and applications. Platooning characteristics have been associated with three types of instability that represent increasing levels of instability: local instability, platoon (or string) instability, and traffic flow instability. Determining the distribution of platoon size and the distribution of the inter-platoon gaps can be used in assessing the level of instability. Traffic flow is considered unstable if it contains enough long platoons and short inter-platoon gaps resulting in the instability within one platoon transferring to the next platoon and the disturbance continuing to grow in amplitude (Pueboobpaphan & Arem, 2011).

The review of literature shows that the time headway has been widely used for platoon identification. Different thresholds of time headway, ranging from 3.0 seconds to 6.0 seconds, have been used in the past to identify platooning vehicles. Lay (1998) suggested three groups for time headway. When the headway is less than 2.5 seconds, the traffic is recognized to be following (in the platoon). When it is between 2.5 seconds and 9.0 seconds, the traffic is considered to be either following or free (not platooning), and when it is more than 9.0 seconds, the traffic is considered free.

Shimoi et al. (2011) defined platoons as vehicles whose speeds are strongly correlated with the speed of the lead vehicle. They studied the relationship between the time headway and the correlation coefficient in the speed of two successive vehicles and determined that the critical headway is 4.0 seconds. Rahman et al. (2012) used a time gap threshold of 4.0 seconds instead of time headway to define platooned vehicles. Vogel (2002) found that the 6.0-second time headway is optimal for identifying non-following vehicles. Al-Kaisy and Durbin (2009) identified vehicles as being in a platoon based on plotting the mean speed of the vehicles and their time headways. The time headway that indicates platooning was identified to be 3.0 seconds. The HCM points to 3.0 seconds as the threshold, although it recommends using the percentage of

time spent following to determine the platooned vehicles. Gattis, et al. (1997) used 5.0 seconds as a time headway criterion to define platoons.

Yang et al.(2015) conducted empirical analysis on platoons from free flow to congestion flow at on-ramp bottlenecks in I-405 Santa Monica, California and I-95 Backlick, North Virginia. They found that the platoon time headway fits a normal distribution, with a mean of 1.42 seconds and standard deviation of 0.42 seconds in Backlick. In Santa Monica, the lognormal distribution had a mean of 1.66 seconds and standard deviation of 0.84 seconds. They also found that the non-platoon vehicle time headway fits shifted negative exponential distribution with a mean of 3.01 seconds and standard deviation of 1.12 seconds.
Researchers also linked the position of a vehicle in a platoon with the standard deviation of speed. Jiang et al. (2015) conducted an experimental study of car-following behavior in a 25-car platoon using GPS data and found that the standard deviation of speed increases in a concave manner with the position of the vehicle in the platoon. Tian et al. (2015) investigated the growth pattern of traffic oscillation using US-101 trajectories data collected by the Next Generation SIMulation (NGSIM) program and produced the following relationship between the location of the vehicle in a platoon and its standard deviation of speed:

$$SDv = -10.4\ E\text{^}((-NVP/94.29)\ )+10.56 \qquad 2\text{-}1$$

where SDv is the standard deviation of speed of each vehicle (m/s) and NVP is number of vehicles in the platoon.

Researchers also investigated the variation in the percentage of vehicles in a platoon with traffic volume. Sun et al.(2005) examined the relationship between traffic volume and the percentage of vehicles platooning on a highway, with two lanes in each direction and short-term and long-term work zones in Illinois. They showed that as the volume increases, the percentage of platooning vehicles increases. The following Equation was developed based on the collected data:

$$FR = -1.377 + \ln(PIP) \qquad 2\text{-}2$$

where FR is the hourly flow rate (veh/hr/lanes) ranging between 400 and 1,400 veh/hr/lane, and PIP is the percentage of vehicles in the platoon.

2.2.3 Safety Measures

Rear-end collisions are a main safety concern on freeways, mainly caused by slow or stopped traffic. Because collisions are rare events, crash data for at least three years is required to have a sufficient sample size to assess traffic safety (Rahman & Abdel-Aty, 2018). The safety assessment to support TSMO, particularly in real-time operations, does not have access to sufficient crash data to assess safety. To assess safety for a shorter period of time or when such data is not available, traffic conflicts have been used as a technique to assess the safety at a location (Li et al., 2017; Rahman & Abdel-Aty, 2018), with the assumption that the conflict

statistics is correlated with the risk of actual collisions (Dijkstra et al., 2010; Lu et al., 2011). A conflict is a scenario where two drivers will likely collide without evasive action.

In order to evaluate rear-end crashes, surrogate measures of safety have been proposed to allow the development of relationships between the likelihood of crashes and traffic stream flow parameters (Kuang et al., 2015; Z. Li et al., 2014). These measures have been widely used as indicators to evaluate rear-end crash risk and to quantify the number of conflicts. Examples are the time–to–collision (TTC), time exposed time–to–collision (TET), time integrated time-to-collision (TIT), post encroachment time (PET) and deceleration rate to avoid the crash (DRAC) (Abdel-Aty & Pande, 2005; Guido et al., 2012; Li et al., 2017; Peng et al., 2017; Rahman & Abdel-Aty, 2018). Moreover, since a rear-end crash may occur due to insufficient safety distance between the leading and the following vehicle, Oh et al. (2006) proposed a rear-end crash index (RCI) based on the safe stopping distance in car following to indicate dangerous conditions. The macroscopic traffic flow parameters (flow, speed, and density) have been used in safety performance estimation (Chang & Xiang, 2003). However, these measures do not adequately capture the interactions among individual vehicles. A study by Zheng (2012) showed that crashes happen in congested conditions and in the transition conditions about six times and two times, compared to free flow conditions, respectively. There are limited studies that utilize microscopic traffic flow data in estimating the safety performance. Zheng (2012) reported that a combination of speed, speed variance and flow as a good indicator of traffic's chaos which has adverse impact on traffic safety. A study by Zheng et al. (2010) also showed that traffic oscillations are related to the standard deviation of speed and crash rate tends to increase as the standard deviation of speed increases. Other studies also found a direct relationship between the standard deviation of speed and safety (Abdel-Aty & Pande, 2005; Kamrani et al., 2018; Lee et al., 2002). Arvin et al. (2019) used lateral and longitude accelerations of individual vehicles that were estimated based on CV data collected as part of the Safety Pilot Model Development (SPMD) as a measure of driver's volatility that impacts crash frequencies at intersections.

The above review of the literature indicates that although several studies have investigated using surrogate measures for safety performance assessment, there is limited work on utilizing microscopic parameters at the individual vehicle level, that will potentially become available from small market penetrations of emerging vehicle technology, to assess safety.

2.2.4 Traffic State Identification

The categorization and recognition of the traffic state, particularly the occurrence of breakdown, is critical to traffic flow analysis and effective traffic management and operations. Due to the data availability, the identification of the traffic states has been mainly based on the three macroscopic measures (speed, occupancy, and volume). The introduction of connected vehicles, connected automated vehicles, and advanced infrastructure sensors will allow the collection of microscopic measures that can be used in combination with the macroscopic measures for better recognition of the traffic state.

Understanding traffic flow breakdown mechanism and the probability of its occurrence is important in traffic analysis, management and operations. The causes of traffic breakdowns include high traffic volume that exceeds the maximum allowable throughput and/or bottlenecks where the capacity drops due to disturbance caused by individual drivers such as lane changing and abrupt braking. The Highway Capacity Manual (HCM) procedure for basic freeway segments categorize the traffic states into six levels of service (LOS) and analysts have generally assumed that the breakdown occurs in the threshold between LOS E and F, where the demand exceeds capacity of the freeway segment. The traffic flow rate at breakdown can be lower than the estimated average roadway capacity, as defined in the HCM. Research efforts have confirmed that breakdown can happen stochastically and not at a deterministic value of capacity (Dong & Mahmassani, 2012; Elefteriadou et al., 1995).

For modeling the probability of traffic breakdown, the classification of traffic conditions into congested and uncongested conditions is important. Such classification, if estimated and eventually predicted, will allow traffic management agencies to activate operational plans to address the adverse impacts of congestion. Traffic flow breakdown is usually defined as a speed drop of a certain amount when the traffic demand exceeds capacity. The HCM defines breakdown on freeways as a condition when the speed drops below a certain threshold (e.g., 40 mph) and/or by a certain amount (e.g., 10 mph) (Elefteriadou, 2017) and it is sustained at least for three time-intervals (e.g., 15 minutes totally).

Researchers have used visual observations of the traffic flow-speed, flow-occupancy and speed-occupancy diagrams to evaluate breakdown using threshold values of speed, flow, density/occupancy, or combinations of these variables (Dehman, 2014; Kondyli et al., 2013; Laflamme & Ossenbruggen, 2017; Yeon et al., 2009). The change-point regression modeling was also used to identify the critical value to estimate the breakdown based on the speed-occupancy diagram (Kidando et al., 2019). Clustering models, which are unsupervised machine learning algorithms, such as the Gaussian mixture model (Kidando et al., 2018; Ko & Guensler, 2005) and K-Means clustering (Elfar et al.,2018; Jingxin et al., 2013) have also been used to identify the traffic state based on speed and density. The speed threshold used to identify breakdown in previous studies widely varied from 25 mph to 50 mph, showing that there is no agreed-on definition of breakdown (Kidando et al., 2019).

Some researchers also defined traffic regimes based on oscillations and categorized traffic conditions as none-oscillatory, damped oscillatory, and oscillatory regimes. The breakdown is associated with the oscillatory regime that indicate unstable conditions (Herman et L, 1959; Swaroop & Rajagopal, 1999). A traffic condition is considered stable if the flow is able to handle disruptions without breaking down. Treiber and Kesting (2013) divided the traffic flow states into stable and unstable and recognized key factors of congested traffic instability such as disruption and propagation probability. It was found that there is always a growth of perturbation in the congestion regime (Treiber & Kesting, 2013). Treiber and Kesting (2013) defined five stability classes based on density, with two classes are unconditionally stable or unstable but the remaining three classes can be stable, unstable or metastable and there is a probability of traffic breakdown associated with these states. However, an experimental study

by Jiang et al. (2018) showed that traffic instability is related more to speed rather than density. They also found that for the same average traffic speed, some experiments showed stable traffic flow while others showed unstable traffic flow. This indicates that speed by itself may not be a sufficient indicator of breakdown and other parameters are needed to indicate traffic instability. Some researchers assumed traffic flow to be stable if the slope of the flow-speed fundamental diagram is positive and is unstable if the slope is negative. However, it was shown that under the negative slope, the traffic flow is not necessarily unstable and depends on driving behaviors (Pueboobpaphan & Arem, 2011). Chatterjee et al. (2017) also used platoon characteristics to determine three different traffic states: free flow, stable flow and constraint flow.

The majority of previous studies used either visual observation or clustering methods using macroscopic parameters (speed, flow, and density) to identify traffic state. However, the review of literature presented in this section indicates that other factors should be considered for accurate traffic state identification and prediction.

2.2.5 Traffic Flow Breakdown Prediction

The prediction of the traffic flow breakdown is critical to effective traffic management and operations. Active traffic management applications have generally used macroscopic traffic metrics in their decisions to activate operation plans. The traffic breakdown has been identified mainly based on thresholds of speed and/or occupancy/density. The critical speed or density at capacity, as identified in the highway capacity manual (HCM) can be used for this purpose. In terms of traffic flow, however, research efforts have confirmed that breakdown can happen stochastically and not at a deterministic value of capacity (Dong & Mahmassani, 2012; Elefteriadou et al., 1995). As mentioned before, the introduction of connected vehicles, connected automated vehicles, and advanced infrastructure sensors will allow the estimation of microscopic traffic metrics that can be used in combination with the macroscopic measures commonly used for better identification of the traffic breakdown in terms of both mobility and safety.

Traffic flow breakdown is usually defined as a speed drop of a certain amount. There are numerous studies that investigated the traffic breakdown phenomena. The first probabilistic breakdown model was proposed in 1972 (Bullen, 1972) identifying the probability of breakdown as an increasing function of the flow rate. Wang et al. (2010) developed a model for predicting the breakdown probability based on the expected future density, utilizing Markov transition. Son et al. (2004) developed a probabilistic model of traffic breakdown, triggered by merging vehicles, by applying a wave propagation model with random disturbance. Elefteriadou et al. (1995) defined the breakdown mechanism as resulting from a large vehicle groups entering freeway through an on-ramp causing speed drops. Shiomi et al. (2011) defined the traffic breakdown probability as a function of platoon size and its lead vehicle's speed given the traffic flow rate. Ahn et al. (2017) studied a stochastic modeling of traffic breakdown for freeway merge bottlenecks with consideration of headway distribution. Chen et al. (2014) proposed a model for traffic breakdown, caused by perturbations of on-ramp merging vehicles,

based on queuing theory. Dong et al. (2012) studied traffic breakdown, focusing on disturbance caused by speed changes of the lead vehicle and the propagation by followers. Kondyli (2009) also studied breakdown probability at freeway ramp merge based on the stochasticity of driver behaviors in accepting gaps and making decisions.

With the rapid advancement of machine learning, it is possible to utilize large amount of data including microscopic traffic data that includes detailed trajectory data to improve the accuracy of traffic state identification and prediction. The machine learning methods are mainly categorized in three categories: naïve, parametric, and non-parametric methods (Van &Van, 2012). There are numerous studies in transportation engineering on using various machine learning methods in a wide variety of applications. However, there is a limited research on using machine learning methods in conjunction with high-resolution data in real-time traffic state estimation. Elfar et al. (2018) used trajectory data to predict congestion utilizing speed and standard deviation of speed between individual vehicles. There are also existing studies that used deep learning methods in traffic congestion and trajectory prediction (Khajeh Hosseini & Talebpour, 2019; Zhang et al., 2019). However, these existing studies (Elfar et al., 2018; Khajeh Hosseini & Talebpour, 2019) focused on the classification of congestion for short periods of time in the future (10 sec/ 20 sec) and did not combine mobility and safety metrics in the prediction. In addition, these studies used data that were collected only at congested conditions and do not covered all traffic conditions (Elfar et al., 2018; Zhang et al., 2019).

2.3 METHODOLOGY

This chapter presents the surrogate measures, utilized data, and methods that were used to achieve the research task.

2.3.1 Surrogate Measures

Before providing more details on the utilized methodology and data, this section describes the microscopic traffic parameters investigated for use in the listed study' purposes and how they are calculated. A Procedural Language extension to Structured Query Language (PL/SQL) program was developed with Oracle SQL Developer to calculate these measures utilizing the trajectory datasets used in this study.

2.3.1.1 Standard Deviations of Speeds

The standard deviation of speed is related to the shockwave and platoon formation, which preludes traffic breakdown (Elfar et al., 2018). Some studies also showed that the increase in standard deviation of speed will result in traffic breakdown (Krauss, 1997; Kühne, 1984). The standard deviations of speeds of individual vehicles (SDv) and Standard deviations of the speeds between vehicles (SDt) are calculated as follows:

$$SD_T = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(V_i - \bar{V})^2} \qquad 2\text{-}3$$

$$SDv=\sqrt{\frac{1}{n_1-1}\sum_{j=1}^{n_1}(S_j-\bar{S})^2} \quad 2\text{-}4$$

where, n is the number of vehicles, V_i is the speed of vehicle i, $\bar{V}$ is the mean speed of all vehicles, $n_1$is the number of speed data collected for a specific vehicle at 1/10th of a second time interval, S_j is the speed of the vehicle at interval j, and $\bar{S}$ is the mean of speed data for each vehicle.

2.3.1.2 The Index of Time Exposed to Time to Collision (TETIndex)

Time-to-Collision (TTC) is a primary conflict measure, introduced by Hayward (Hayward, 1972), and has been widely used as a surrogate safety measure for rear-end crashes. TTC is defined as the expected time for two vehicles, following each other, to collide if they remain at their present speed and on the same path (Hayward, 1972). It has been reported that lower TTC is a good indicator of the probability of collision but cannot be directly linked to the severity of the collision. Balas et al. (Balas & Balas, 2007) used the inverse of the TTC as an indicator of safety. Another type of TTC referred to as TTC_{Break} (Peng et al., 2017) was proposed to consider situations, in which the leading vehicle stops suddenly. This parameter is calculated as follows:

$$TTC_{Break}(t)=\frac{x_{i-1}(t)-x_i(t)-\bar{L}}{v_i(t)} \quad \text{if } v_i(t)>v_{i-1}(t) \quad 2\text{-}5$$

where TTC_{Break} (t) is the time to collision value of vehicle I at time t, xi (t) is the position of vehicle I at time t, vi(t) is the speed of vehicle I at time t, and $\bar{L}$ is the average length of vehicles. An aggregate indicator, the TET, was introduced to assess the safety performance of monitored segment in space and time. The TET reflects the total time spent under dangerous traffic conditions, characterized by TTC values below a threshold value TTC*. The TET is calculated as follows.

$$TET(T)=\sum_{i=1}^{n}\delta_t\Delta t \quad 2\text{-}6$$

$$TET=\sum_{t=1}^{T}TET(t) \quad 2\text{-}7$$

$$\text{WHERE} \quad \delta_t=\begin{cases}1, & 0<TTC_{break}(t)\le TTC^*\\ 0, & \text{else}\end{cases}$$

where, Δt is the time step of the trajectory data collection, which is usually set at 0.1 second in simulation, TTC* is the threshold, n is the number of vehicles, and T is time.
The TTC* threshold, referenced above is used to differentiate the unsafe car following conditions from the ones that are considered safe. According to the past research (Abdel-Aty & Pande, 2005), the threshold values are usually set between 0.5 second to 3.5 second. Fan et al. (Fan et al., 2013) recommended using a TTC* threshold of 2 seconds in freeway merge areas to identify the conflicts. The sensitivity analysis in previous research also showed that changing the TTC* value within the investigated range does not have a significant effect on the results of the safety assessment (Li et al., 2017). Based on the above review, this study uses TET with a threshold (TTC*) of 2 seconds, as the surrogate safety measure to be estimated based on CV trajectory data.

In this study, TET is calculated for each 33subsegment and normalized using Equation 2-8 to determine a TET index, which is proposed in this study as an index of disturbance duration.

$$TETINDEX=((TET/N)/T) \qquad 0 <= TETINDEX <= 1 \qquad 2\text{-}8$$

where TET, n and T are as in Equation 2-7.

2.3.1.4 Number of Oscillations (NO)

In addition to the duration of disturbance, reflected by the TETIndex described in Equation 2-8, the number of oscillations (NO), reflecting the stop/go maneuvers, was also defined in this study. An oscillation is defined as a deceleration phase followed by an acceleration phase. Stop and go traffic is the mechanism of traffic state transition to congestion and is related to traffic breakdown and instability (Zheng et al., 2010). An oscillation occurs when the speed of the follower vehicle is changing while the leader's speed does not. This study measured NO directly from individual's vehicles acceleration/deceleration. The percentage of vehicles having oscillation in each 300 ft segment is calculated and used as another disturbance indicator.

2.3.1.5 Rear-End Crash Index (RCI)

Rear-end crash Index (RCI) is another surrogate safety measure proposed based on the safe stopping distance in car following situations (Oh et al., 2006; Ozbay et al., 2007). Safe stopping distance is defined as that, at which the follower vehicle can safely reduce speed to avoid colliding with the leading vehicle when the leading vehicle reduces its speed or stops. Using the RCI, the judgment of whether a conflict occurs is based on the trajectories parameters of two consecutive vehicles including the relative speed, distance, and acceleration between the leading and following vehicles. RCI was defined as follows:

$$RCI = \left(\sum_{t=1}^{T} RC(t)\right)/n \qquad 2\text{-}9$$

$$RC(\tau) = \sum_{i=1}^{n} SDI * \Delta t \qquad 2\text{-}10$$

$$\text{WHERE } SDI = \begin{cases} 0\ (\text{safe}), & \text{if } SDF < SDL \\ 1\ (\text{unsafe}), & \text{else} \end{cases}$$

where, RC(t) is the number of conflicts at time t, Δt, t, T and n are as defined before and the stopping distance (SDI) is calculated as function of the reaction distance and braking distance as follows:

$$SDF = v_f * PRT + \frac{{v_f}^2}{2*a_f} \qquad 2\text{-}11$$

$$SDL = v_l * PRT + \frac{{v_l}^2}{2*a_l} + \bar{l} \qquad 2\text{-}12$$

$$SDL_{Break} = \bar{l} \qquad 2\text{-}13$$

$$\text{WHERE } \bar{l} = x_{i-1}(t) - x_i(t) - L \qquad 2\text{-}14$$

where SDF is the stopping distance of the following vehicle, SDL is the stopping distance of the leading vehicle, PRT is the perception reaction time, set to 1.5 seconds (AASHTO, 2011), a is

deceleration rate, v is the speed and $\bar{l}$ is relative distance. x_i (t) and $\bar{L}$ are as defined before. Please note that since this study uses $\mathrm{TTC}_{\mathrm{Break}}$, SDL_{Break} was defined in this study to extend the original concept of SDL to reflect conditions, at which the leading vehicle stops suddenly causing a conflict with the lagging vehicle.

2.3.2 Utilized Data

A complete understanding of the traffic conditions will benefit from collecting the data at the individual vehicle level and generating time-space diagrams showing vehicle interactions, disturbances in traffic flow, congestion formation, and propagation that are important in assessing traffic safety and mobility. As stated earlier, the recent advancement of emerging technologies such as CV and CAV promise to provide such data in real-time. However, CV and CAV data are not yet available from traffic streams with high market penetrations of these technologies to allow its use in this study. The real-world trajectory data collected as part of the Federal Highway Administration (FHWA) by Next Generation SIMulation (NGSIM) program was also collected only for congested conditions, and thus does not fully satisfy the data requirements of this study. Therefore, simulation modeling was used to produce trajectory data for different congestion levels for use in the development of the models in this study. However, real-world vehicle level data in addition to macroscopic traffic data were used in calibrating the simulation model. In addition, vehicle-level real-world data from other locations were used in validating and evaluating the developed models, as described later in this chapter. Thus, the utilized data includes the following three groups:

- Simulated vehicle trajectory from microscopic simulation
- Real-world vehicle trajectory from NGSIM data
- CV data from real-world deployments of connected vehicles (CV)

The trajectory data from microscopic simulation, NGSIM data, and CV data include information about vehicles' speeds, accelerations/decelerations, and locations at the 1/10th of a second resolution. These trajectory data were associated with subsegments per lane, each measuring 300 feet long. The segment performance was assessed for each 5-minute interval, unless otherwise stated.

2.3.2.1 Microscopic Simulation

Simulation modeling of a freeway segment was performed utilizing the PTV's Verkehr In Städten SIMulationsmodell (VISSIM) microscopic simulation tool. The simulated segment includes an on-ramp and off-ramp with three main lanes throughout the section. The merge of the on-ramp traffic creates a bottleneck with different levels of congestion at the upstream section, depending on the traffic level. The segment was simulated and calibrated to reflect real-world conditions on the eastbound segment of the I-580 in California, between Strobridge Ave and Redwood Rd. The calibration process utilized data and results from a FHWA study, in which the time headways of all vehicles were collected utilizing a drone video recording and analyzed using an image processing software (Hale et al., 2019). The FHWA study utilized a

previously developed method (Rakha & Gao, 2010) to calibrate the Wiedemann model based on the collected headway data. Please refer to Hale et al. (2019) and Rakha & Gao (2010) for the utilized VISSIM calibration processes and its validation.

A one-mile segment in the simulation was divided into sub-segments per lane, each is 300 feet long, to allow a detailed spatial analysis of the segment for each 5-minute time intervals for a total period of 15-minutes. The first 30 minutes of VISSIM simulation was set as a warm-up period and thus was excluded from the period of performance assessment. Figure 2-1 shows the coded freeway section.

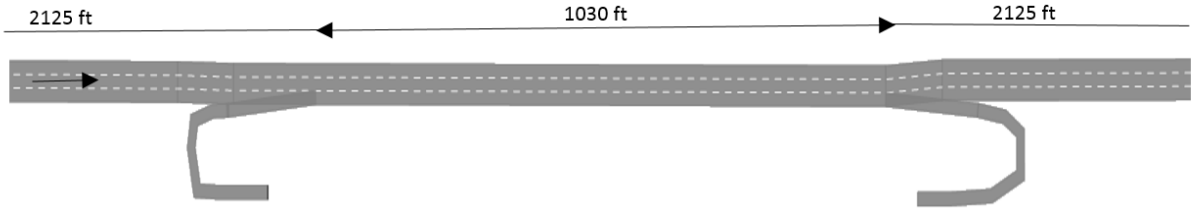
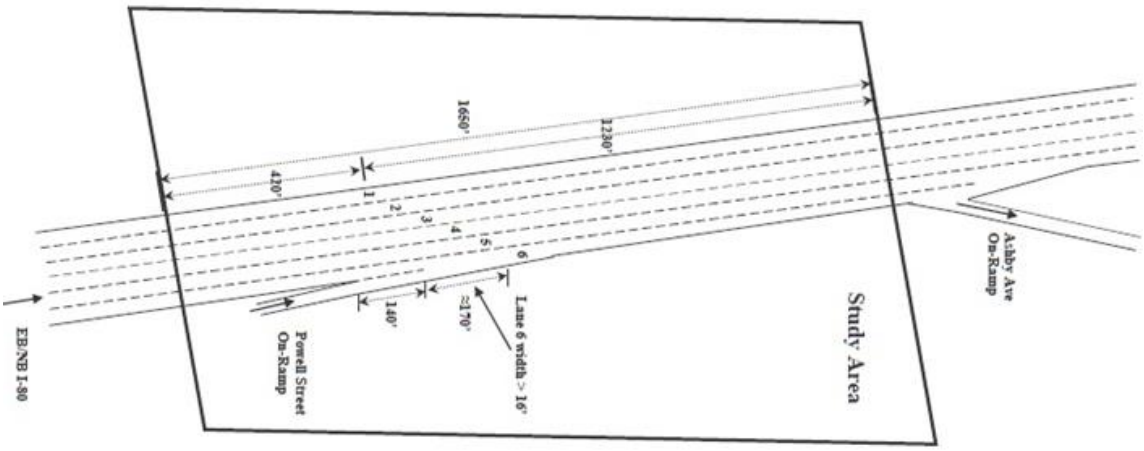


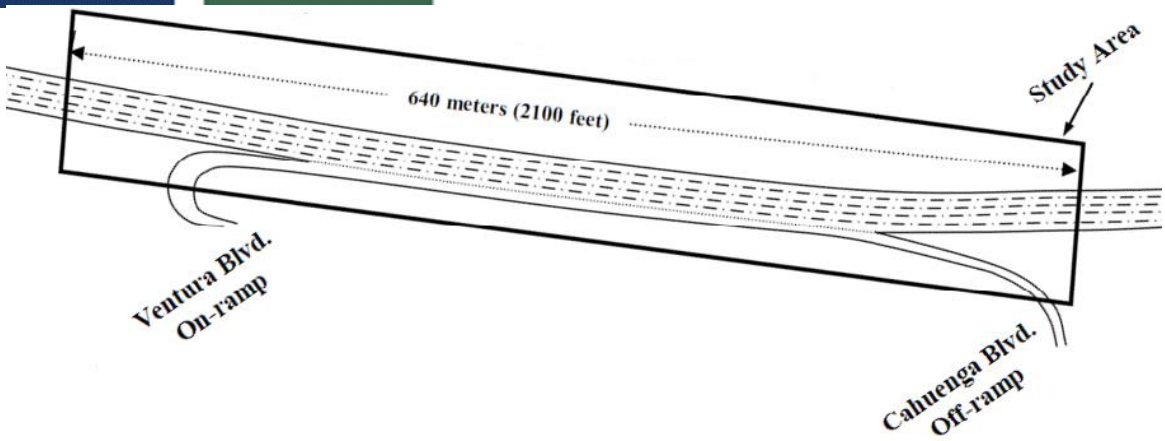
FIGURE 2-1: CODED FREEWAY SECTION

2. 3.2.2 NGSIM Data

The data from the simulation effort described above was used to derive a model to estimate safety and mobility performance. To validate the transferability of the model in producing acceptable results in other locations not used in the calibration of the simulation model, this study used real-world trajectories data collected from both I-80 and US 101 segment in California as part of the NGSIM program (FHWA, 2007) for the two locations. The segment of US101 is about 2,100 feet long and has five mainline lanes, one on-ramp and one off-ramp. The data were collected for the period between 7:50 A.M. to 8:35 A.M, for a total period of 45 minutes. The segment of I-80 is about 1,650 feet long and has five mainline lanes, one on-ramp, and one off-ramp. The I-80 data were collected for the period between 4:00 P.M. and 4:15 P.M. Figure 2-2 shows the study areas of I-80 and US101.



(a)



(b)

FIGURE 2-2: NGSIM PROGRAM STUDY AREAS (A) I-80 AND (B) US101

2.3.2.3 Connected Vehicle Data

Further evaluation of the application of the developed models for safety and breakdown prediction was done using data collected based on BSM from connected vehicles as part of the SPMD project in Ann Arbor, Michigan. The data was collected for 3,600 feet of the eastbound direction of I-94. This study used this data to assess the performance of the developed method when the data is obtained from low market penetrations of CV. The SPMD included about 2,800 instrumented vehicles and about 70 miles of roads instrumented with road side units (RSU) (Henclewood, et al., 2014). The CV and RSU communicated via DSRC communication. The selected segment has two lanes and located between ramps.

2.3.2.4 Descriptive Statistics of Utilized Data

The summary statistics for utilized data including the simulated dataset, CV dataset, and NGSIM dataset are reported in Table 2-2. The features included in Table 2-2 are microscopic and macroscopic measures explained earlier in section 2.3.1.

TABLE 2-2: DESCRIPTIVE STATISTICS OF FEATURES

Features	Mean	Standard Deviation	Minimum	Maximum
Real-World CV Data and Detector Data from the SPMD Safety Pilot Project				
Mean Speed (ft/s)	93.88	15.05	45.18	120.41
Average SDv (ft/s)	0.88	0.74	0.04	4.08
SDt (ft/s)	6.66	5.05	0.1	22.3
TETIndex	0.03	0.02	0.01	0.15
NO (%)	17.15	36.63	0	100
Flow (vphl)	693	228.1	104	1814
Occupancy (%)	13.25	10.23	1.5	56.6

Simulated Data				
Mean Speed (ft/s)	73.83	25.60	12.20	106.34
Average SDv (ft/s)	1.59	1.38	0.09	5.02
SDt (ft/s)	5.95	3.53	1.91	17.87
TETIndex	0.035	0.032	0.0013	0.17
NO (%)	16.28	24.23	0	100
Flow (vphl)	1772	322	693	2488
Occupancy (%)	28.98	16.34	7.29	95.86
NGSIM: I-80 (4:00 P.M. and 4:15 P.M.)				
Mean Speed (ft/s)	25.76	4.39	15.69	37.20
Average SDv (ft/s)	3.86	0.73	2.35	5.75
SDt (ft/s)	6.74	1.62	4.05	12.90
TETIndex	0.09	0.03	0.05	0.19
NO (%)	96.28	5.51	76.47	100
Flow (vphl)	1293	210.9	824	1941
Occupancy (%)	40.84	9.44	25.08	81.53
NGSIM: US101 (7:50 A.M. to 8:35 A.M.)				
Mean Speed (ft/s)	34.25	5.08	24.08	45.15
Average SDv (ft/s)	3.62	0.49	2.42	4.46
SDt (ft/s)	9.88	1.72	6.70	14.28
TETIndex	0.13	0.025	0.07	0.19
NO (%)	76.26	21.14	18.75	100
Flow (vphl)	1677	134.2	1392	2004
Occupancy (%)	41.38	4.65	31	55.21

2.3.3 Platooning Measures

One of the objectives of this task is to estimate platooning measures at low market penetration of CV data. Several studies used time headway and its distribution as criteria to identify platooned vehicles. Although the headway distribution is a good measure to use in estimating the platooning characteristics, the data required for headway estimation will not be available based on data from low market penetrations of CV data, as described earlier. The headway measurement requires the location and speed of both the leading and lagging vehicles, and these vehicles will have to be equipped with CV technologies to provide this information. The headway data can be also obtained from connected automated vehicles if information from the vehicle sensors installed in these vehicles such as image-based sensors and microwave sensors are made available for performance measurement. However, this is not expected in the near future. Thus, measures other than time headway is needed to estimate platooning at lower CV market penetrations.

This section identifies methods for the estimation of the percentage of vehicles in the platoon and the distribution of the platoon size based on surrogate measures that can be assessed using CV data at a relatively low market penetration of connected vehicles. This study utilizes two surrogate measures for this purpose, the standard deviation of speed between vehicles (SDt), and the average of the standard deviations of the speeds of individual vehicles (SDv) as were described before.

To investigate the relationship between the surrogate measures and the platooning characteristics, this study utilized real-world trajectory data collected as part of the Federal Highway Administration (FHWA) NGSIM program. Since this data was collected for congested conditions with high levels of platooning, trajectory data generated from simulation analysis were used to supplement the real-world data, which provided the coverage of additional platooning levels. The mean speed, SDv and SDt were calculated based on the real-world and simulated trajectories. A moving average method was used to remove the noise in the vehicle speed profile. All vehicles were sorted by their entrance times to the 300 ft subsegment and were assigned vehicle identifications for use in the determination of the platooning of vehicles. Two criteria were used to decide whether a vehicle is platooned or not. The first criterion is the time headway. A time headway of 4.0 seconds was selected as a threshold to identify the platoon based on data analysis that involved plotting vehicle headway vs. speed, sensitivity analysis that involves varying the headway threshold between 3.0 seconds to 5.0 seconds, and findings based on the literature review. Since the standard deviation of speed of the leading vehicles should be small, another criterion was used to supplement the time headway criterion in the identification of a non-platooned vehicle, or the leading vehicle of a platoon based on the standard deviation of individual vehicle speed. Based on examining the data used in this study and the review of literature, the standard deviation of speed below 1.0 ft/sec was used as an indication of a non-platooned vehicle or the leading vehicle in a platoon, if the vehicle also meets the headway criterion. The reason for using 1.0 ft/sec is because in Equation 2-1, if the specified number of vehicles in a platoon is 1, then the SDv is 0.26 m/s (0.88 ft/sec), which is close to 1 ft/s.

The above procedure allows the estimation of the percentage of vehicles in a platoon based on all available trajectories. In addition, the distribution of the size of the platoon was estimated by estimating the position of each vehicle in the platoon based on a model developed by Tian et al. (2017) that links the standard deviation of the speed of a vehicle to its position in the platoon. This model was previously presented in Equation 2-1. This relationship was validated based on the data collected in this study, as discussed later in this document.

The above discussion summarizes the estimation of two platooning measures for use in this study: the percentage of vehicles in a platoon and the distribution of the size of the platoon. Since the purpose of this study is to link the two surrogate measures (SDv and SDt) to the platooning measures discussed above, the relationships between the surrogate measures and the platooning measures were then developed based on all available trajectories since the platooning measures can only be measured if all trajectories are used. The trajectories of all vehicles cannot be obtained unless the market penetration of CV is 100%. Thus, the developed

relationship was derived for use to estimate the platooning measures based on the surrogate measures for lower market penetrations.

As stated earlier, the trajectories utilized in this study were obtained from two sources: real-world trajectories collected from I-80 in Emeryville, California by the NGSIM program, and simulated trajectories for the same corridor. For both the real-world and simulation, the segment was divided into five subsegments per lane, each measuring 300 feet long. The platooning measures and surrogate measures were estimated for each subsegment per 15-minute time interval.

2.3.4 Safety Measures

Another objective of this task is to use trajectory data in estimating metrics of the safety of freeway segments using data collected from low market penetrations of CV. The safety performance was assessed based on indicators of the interactions between pairs of vehicles in the traffic stream. These interactions at the individual's vehicle level creates perturbation in the traffic flow that can potentially lead to traffic breakdown or collisions.

This study investigated methods for the utilization of detailed trajectory data to measure disturbances metrics as indictors of the perturbation of traffic flow to identify unsafe conditions. The utilized disturbance metrics are the number of oscillations (NO) and a measure of disturbance durations index based on time exposed time-to-collisions (TET), referred to as TETIndex. This study also measured NO directly from individual vehicles' acceleration/deceleration in transition to congested conditions. TET has been widely used as a safety surrogate measure, however, to the best of author's knowledge, there is no study on the identification of its thresholds to justify activating plans to mitigate unsafe conditions in real-time operations. TET also has not been sufficiently investigated as a real-time indicator of safety under different conditions and with smaller sample sizes of trajectory data. The TETIndex estimation requires the location and speed of both the leading and following vehicles and therefore cannot be measured accurately at low sample sizes of vehicle trajectories. Thus, this study derived regression models to estimate the TETIndex based on speed parameters for use in cases such as low market penetrations of CV. The considered speed parameters are the mean speed, the average of the standard deviations of the speeds of individual vehicles (SDv), and standard deviations of speed between vehicles (SDt). The developed model was tested using real-world trajectory data from two locations that were not used in the development of the model. The model was also used to estimate the TETIndex at low market penetrations of real-world vehicle trajectories based on CV data and the results were related to crash data at the site.

The disturbance metrics were also related to an additional surrogate measure of safety referred to the rear end crash index (RCI) to confirm the critical values of the disturbance metrics that indicate unsafe conditions, as described later. The parameters used in this research are described in the section 2.3.1. When calculating these parameters, a moving average method was used to remove the noise in the vehicle speed profile. The estimation was done for each

300 ft sub-segment of the case study facility, per lane and for each 5-minute time intervals. Figure 2-3 presents an overview of the utilized data sources, the developed method, and the associated evaluation.

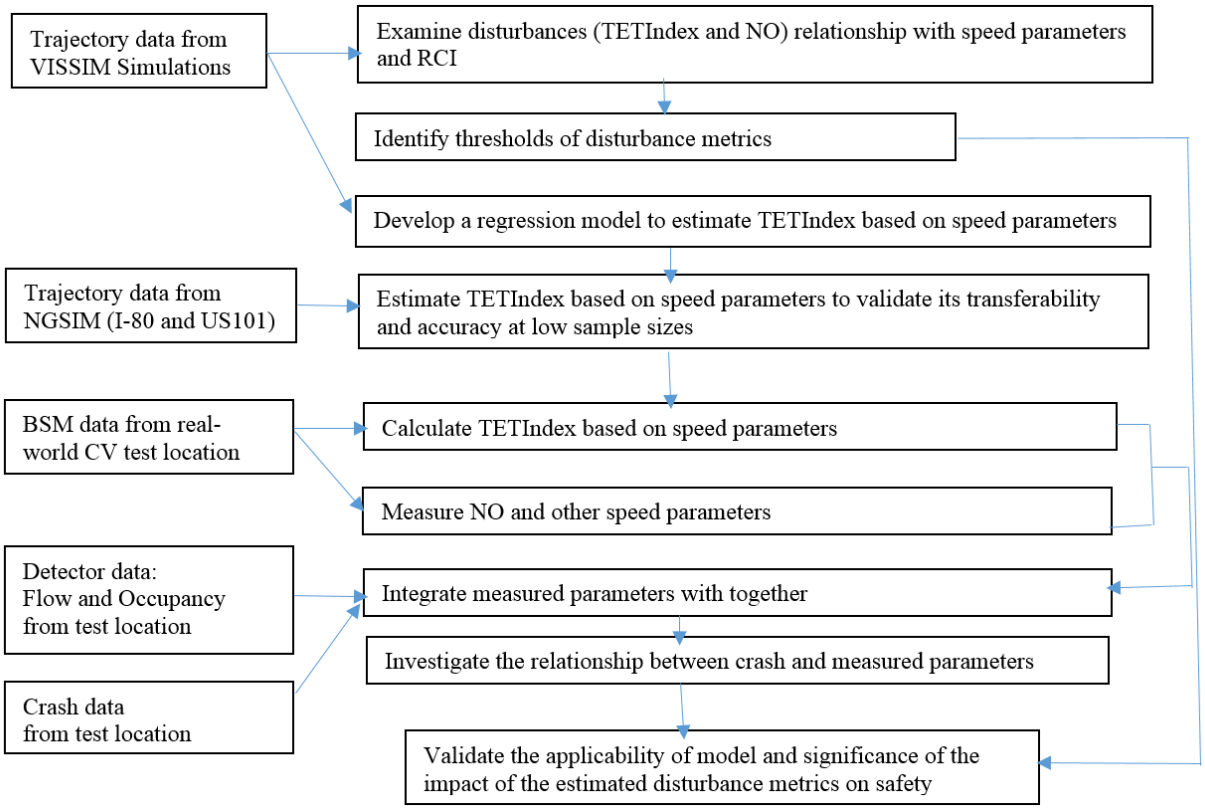


FIGURE 2-3: METHODOLOGY FRAMEWORK OF THE ASSESSMENT OF SAFETY

2.3.5 Traffic State Identification

This study investigated the use of microscopic features that contribute to traffic flow perturbations as indicators of breakdown, in addition to the commonly used macroscopic measures. Three methods were used and compared in identifying the breakdown state. The first and second methods use clustering technique and the change-point regression, respectively, based on macroscopic measures, as have been done in previous studies and discussed in the literature review. The third method, as a contribution of this study, uses a combination of macroscopic and microscopic features in the clustering analysis to identify the traffic state. The first step in the analysis was to identify the conditions, under which there is uncertainty in whether a breakdown has occurred when using the macroscopic measures alone. This was determined by using clustering based on the macroscopic measures and identifying the uncertain traffic conditions with high probabilities of belonging to two clusters, one cluster indicating a "breakdown" and one indicating "non-breakdown". The Gaussian Mixture Models (GMM) with the three macroscopic features of traffic (flow rate, speed and density) was used to identify the uncertain phase of traffic conditions. The reason for using the GMM clustering for this purpose instead of using other more commonly used clustering methods is that the

GMM clustering is capable of providing the probability of belonging to the identified clusters (Bishop, 2006).

The identified uncertain traffic condition phase from the GMM analysis was then further analyzed using the K-Means clustering, with consideration of both macroscopic and microscopic features. The microscopic features include the standard deviation of individual's vehicle speed, number of oscillation (stop/go), and TET, which are indicators of disturbances. The results from using these additional microscopic features in the clustering are then evaluated and compared to those using only macroscopic measures. In addition, the performance of the method was compared with the results obtained from using a deterministic value of speed as an indicator of breakdown as identified using the change-point regression analysis procedure based on the speed-occupancy relationship. Finally, the method was tested using NGSIM dataset.

2.3.5.1 Clustering Analysis

Clustering is the grouping of a set of data into clusters where the data in the same cluster are similar in some sense. There are several clustering methods that have been proposed and utilized in the literature. The two approaches mentioned earlier, the GMM and K-Means clustering methods have their own strengths and weakness. GMM is a probabilistic clustering that fits a set of number of Gaussians (number of components) to the data and estimates the Gaussian distribution parameters (the mean and variance) for each cluster and the size of a cluster. It then calculates the probability of the data points belonging to each cluster (Neal, 2007). The K-Means method clusters the data points based on the average squared distance between the points such that the distance in the same cluster is minimized (Arthur & Vassilvitskii, 2007). The K-Means methods, which is the most widely used clustering method, clusters the data deterministically while in reality there might be some overlapping between the clusters. The GMM addresses this issue by providing the probability of a data point belonging to a specific cluster. However, the K-Means clustering was found to performs better with high dimensional data (Neal, 2007).

As mentioned earlier, there is uncertainty in the data belonging to clusters around the breakdown point. Thus, in this study a hybrid clustering approach that combines GMM and K-Means was used to better categorize the data around the breakdown point. The clustering algorithm and analysis were implemented using the Scikit-Learn library in Python. First, the GMM was used to identify the traffic phase with uncertainty where the probability of being in more than one cluster is significantly higher than zero. Then, the K-Means clustering was used to classify that data points in the phase of uncertain traffic conditions into "breakdown" and "non-breakdown" conditions based on different combinations of macroscopic and microscopic traffic features. Note that clustering the uncertain phase with GMM was also tried. It was found that the K-Means clustering outperformed the GMM with regard to the Silhouette Coefficient (SC), which is a measure used for clustering assessment, explained in the next paragraph. Recognizing the difference in scale between the used features, the values of each feature were standardized before clustering by subtracting the mean and then dividing by the standard deviation of the feature values (Hale, 2018). The results of the clustering algorithm were assessed using the SC as a performance measure of clustering. The SC value for each point is a

measure of how that point is similar to the points in its own cluster, compared to the points in another cluster. The value is between -1 and 1. Values close to one indicate data that the data is very well clustered (Liu et al., 2016). In addition to the similarity of the points in the clusters, this value also increases with the decrease in the number of features. Thus, this should be considered when comparing the performance of different clustering alternatives using the SC.

2.3.6 Traffic Flow Breakdown Prediction

This study investigated the use of combinations of macroscopic and microscopic features to identify the mobility and safety state of freeway traffic. The investigated microscopic features included the standard deviation of individual vehicle's speed and the standard deviation of speed between vehicles, in addition to the two-disturbance metrics of traffic flow, mentioned earlier. The two-disturbance metrics are the number of oscillations (stop/go) and a metric that has been widely used as a surrogate metric to safety, referred to as the time exposed time–to–collision (TET). These disturbance metrics are expected to be good indicators of the perturbation in traffic flow and also safety, thus allowing more accurate alerts to traffic management agencies regarding entering a nonacceptable traffic state. The developed and evaluated models predict the breakdown state in term of combined mobility and safety metrics in the next 5-minute interval in real-time operations. Three different machine learning approaches were used and evaluated for developing state prediction models using data from simulation modeling and data from a real-world CV deployment, as part of the SPMD.

2.3.6.1 Utilized Machine Learning Approaches

Machine learning methods/algorithms are classified as supervised, unsupervised, and reinforcement learning. Supervised learning requires training data that include feeding paired inputs and outputs to the model. Examples of supervised learning are statistical regressions, K-nearest neighbors, support vector machine (SVM), artificial neural networks (ANN), decision trees, and tree ensembles. With unsupervised learning, the input data is not associate with outputs. Examples are clustering and association rules. Reinforcement learning can observe the conditions and select the best actions for a given situation (Bishop, 2006). In this study, both supervised and unsupervised learning techniques are used to support prediction in real-time operations. Clustering, which is an unsupervised learning approach, is used in off-line operations to better partition the data into breakdown and no-breakdown patterns based on the macroscopic metrics and the microscopic disturbance metrics. Then, the clustered data is used off-line to train supervised machine-learning classification approaches for short-term real-time prediction of the breakdown in real-time operations,

For the real-time classification of traffic state, as explained in previous sections, this study first categorized the historical traffic data using unsupervised clustering technique that separates the traffic states into "breakdown" and "non-breakdown", based on combinations of microscopic and macroscopic features, that reflect both mobility and safety. The identified traffic state was then used as a binary response label, to the data that is used in the training and testing of the machine learning classification methods used in this study.

The top ten classifiers that are widely used in research and practice were attempted using the Scikit Learn library in Python. Each of these classifiers has its own weakness and strengths Based on the initial evaluation results, three of the classifiers were selected for use in this study. These selected classifiers are the SVM and two tree ensemble techniques, which are the Random Forest (RF) and Extreme Gradient Boosting (XGB). RF is one of the simplest, popular and accurate learning algorithms and act excellent in prediction performance. RF trains each tree individually using a random sample of the data. XGB, known as regularized boosting, is an implementation of gradient boosted decision trees designed for speed and performance. XGB builds each new tree correcting errors made by the previous trees. SVM is a linear/non-linear separator that transform the data and found the optimum boundary to separate the data into classes (Bishop, 2006).

As stated earlier, several macroscopic and microscopic features were utilized as inputs to the investigated models. Recognizing the difference in scale between the used features, the values of each feature were standardized before clustering by subtracting the mean and then dividing by the standard deviation of the feature values. Since most features are correlated with each other, first, the Principal Component Analysis (PCA) was utilized to reduce the dimensionality of data and handle the collinearity between variables. The utilized number of components was selected based on the number of input features and projected variance (Bishop, 2006).

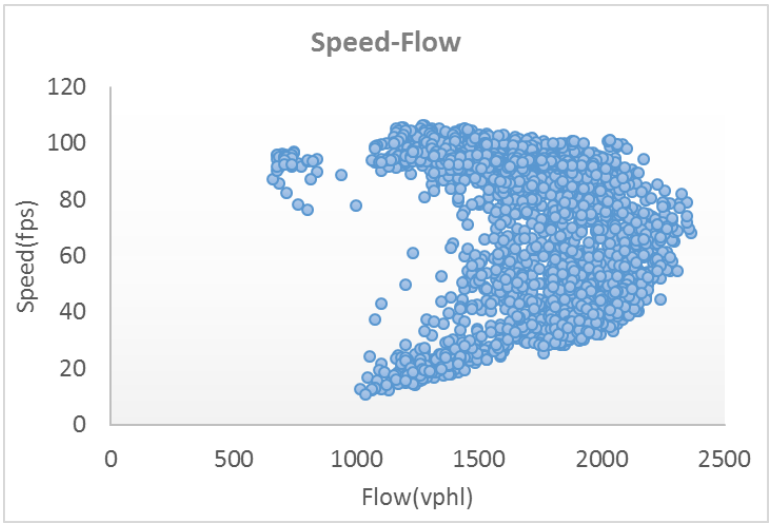
2.4 ANALYSIS AND RESULTS

This section presents the results and discussion of the development, validation, and evaluation of the models and methods mentioned in Section 2.2.3.

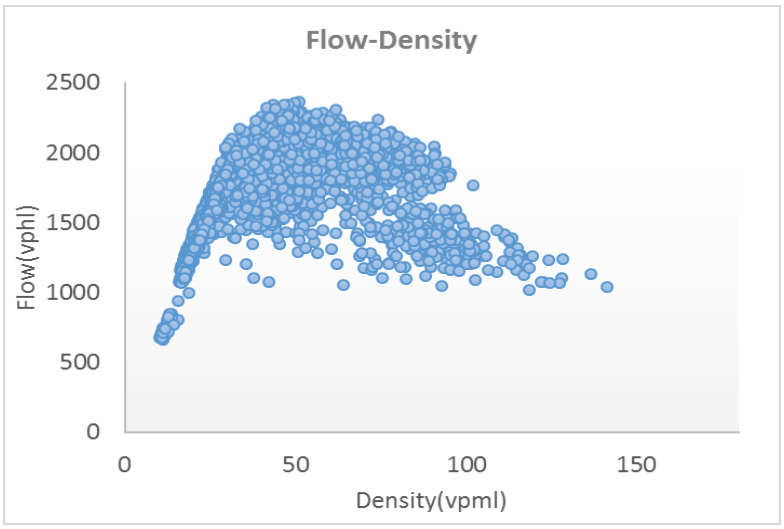
2.4.1 Relationships between Parameters

First, the study performed an initial exploration of the relationship between the three fundamental macroscopic variables (flow, density, and speed) for the simulated data as presented by the fundamental diagrams and shown in Figure 2-4. A Change-Point Regression with Gaussian Mixture analysis was conducted based on the data from the calibrated simulation model and the critical speed at capacity was found to be around 65 ft/second. This value was used in further analysis of the results, as described later.

(a)



(b)



(c)

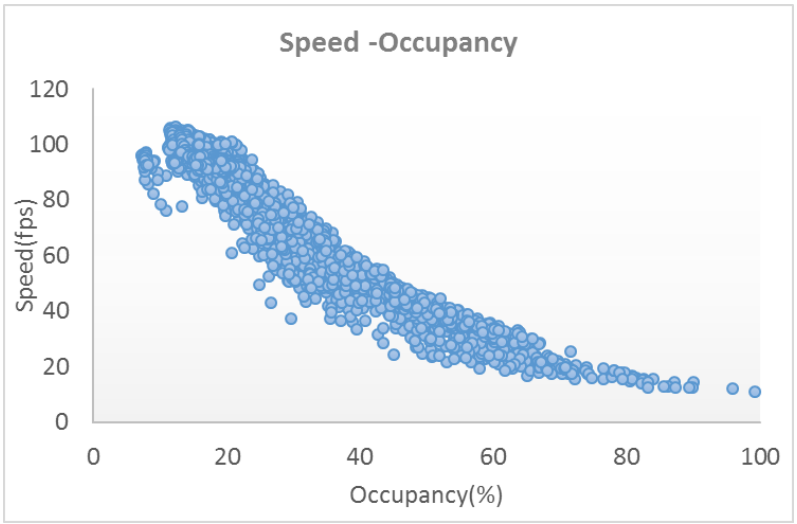


FIGURE 2-4: FUNDAMENTAL DIAGRAM OF THE MACROSCOPIC TRAFFIC FEATURES

Then, this study performed a visual inspection of the relationship between disturbance metrics (TETIndex and NO) with speed parameters (S, SDv and SDt), as described next.

2.4.1.1 Disturbance Metrics: TETIndex and NO Analysis

The relationship between the speed parameters are shown in Figure 2-5. Figure 2-5 (a) shows that as the speed decreases, the SDv, increases reaching a maximum of about 4 ft/s at speed slightly below the critical speed at capacity (about 55 ft/s which is lower than the speed at capacity, which is 65 ft/s). However, the SDv does not increase further when the speed decreased below 65 ft/s since the traffic is already in a congested regime. Figure 2-5 (b) shows that as the speed decreases, the SDt, increases reaching a maximum of about 8 ft/s around the breakdown point. As the traffic becomes more congested and the speed drops, the SDt decreases due to the constraint on the desired speeds of vehicle by their leaders.

(a)

(b)

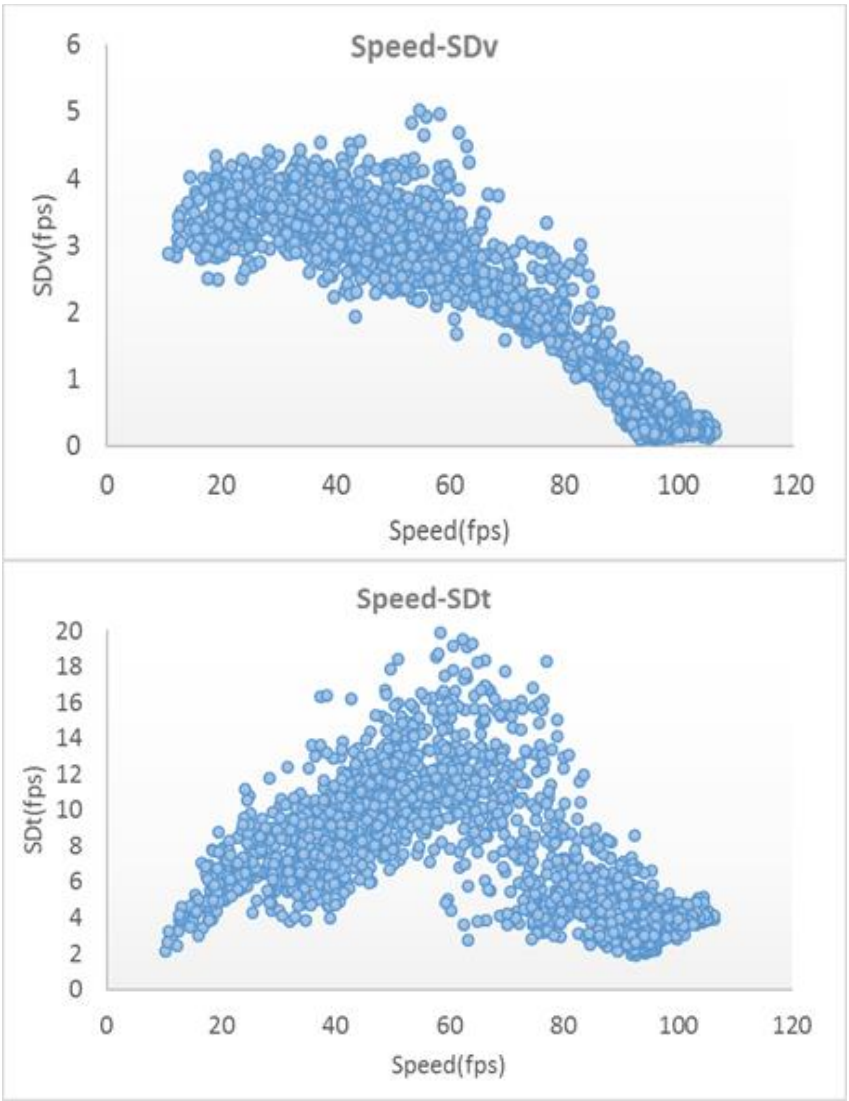


FIGURE 2-5: THE RELATIONSHIP BETWEEN (A) SDV AND THE SPACE MEAN SPEED, (B) SDT, AND THE SPACE MEAN SPEED

Visual inspection of the relationships of the two disturbance metrics (TETIndex and NO) and the speed parameters are shown in Figure 2-6. It can be seen that the TETIndex and NO increase with the decrease in the mean speed. They also increase with increase in SDv. In addition, the TETIndex and NO increase sharply as the speed drops below 65 ft/sec and SDv increases above 2.0-2.5 ft/sec. The values of the TETIndex and NO at the break point, beyond which the two values of the two variables increase sharply, are about 0.03 and 10%, respectively. Figure 2-6 shows that the values of the TETIndex and NO with the maximum SDt are also around 0.03 and 10%, respectively. These values seem to indicate the start of the transition phase of traffic conditions, at which the perturbation of traffic flow is more likely to grow. Figure 2-6 shows higher values of the disturbance metrics including a TETindex of 0.05 and a NO of 20% are clearly in the congestion region of the curves.

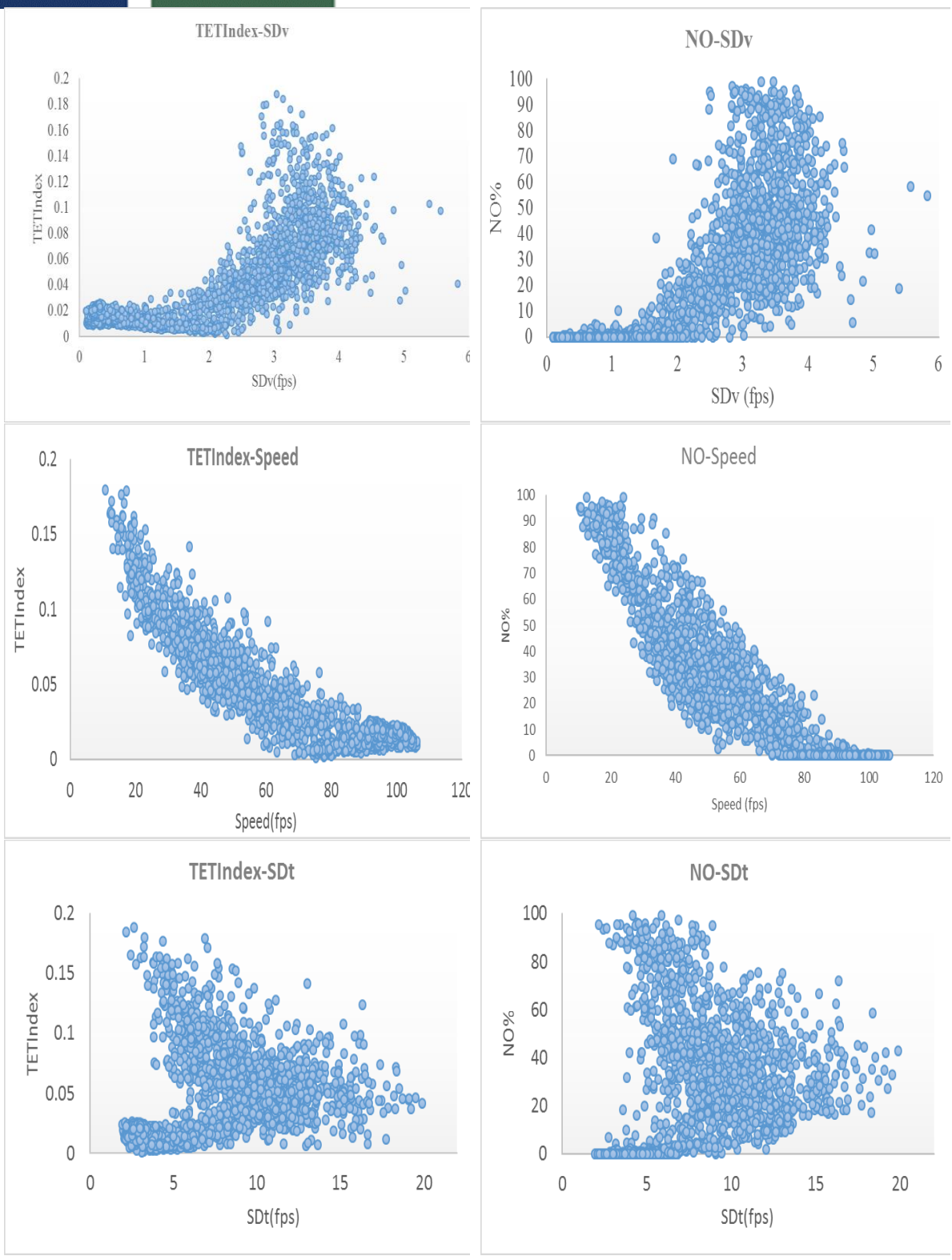


FIGURE 2-6: THE RELATIONSHIP BETWEEN DISTURBANCES WITH MEAN SPEED SDV AND SDT

2.4.2 Platooning Measures

Next, this study investigated the CV use of data at a relatively low market penetration of CV for platooning measure estimation including the percentage of vehicles in the platoon and the distribution of the platoon size based on surrogate measures. This study utilized two surrogate measures for this purpose, the standard deviation of speed between vehicles, and the average of the standard deviations of the speeds of individual vehicles.

2.4.2.1 Platoon Percentage Determination

As described earlier, the number of vehicles in the platoon was calculated in previous studies by identifying whether or not a vehicle is platooned based on the measured time headways and the standard deviation of individual vehicle speeds based on vehicle trajectories. The result was used to calculate the percentage of the vehicles in the platoon. This calculation is possible only if 100% of the vehicle trajectories are available. Thus, the standard deviation of speed between vehicles (SDt), which can be measured at low CV market penetrations, was also measured based on the trajectories for potential use to estimate the percentage of vehicles in the platoon at low market penetrations. A regression analysis was conducted to derive the relationship between the percentage of vehicles in platoon for each subsegment and time interval and the corresponding SDt.

The analysis of the results shows that at low volumes, when the segment is operating at Level of Service C to D or better, according to the HCM procedure, the SDt does not change with the increase in demand and the percentage in platoons, as shown in Figure 2-4. However, an increase in the SDt is observed with a decrease in speed as traffic approaches breakdown, reaching a maximum close to traffic breakdown, as shown in Figure 2-5. Figure 2-5 shows that the maximum observed SDt is at a speed around 59 ft/sec (about 41 mph), which is beyond the critical speed at capacity estimated, which according to the HCM procedure, is around 77 ft/sec (52 mph). Around that point, which corresponds to about 84% of vehicles in the platoon, a significant increase in the SDt is observed, as shown in Figure 2-7. This can be explained by the fact that some vehicles may still be traveling at a relatively higher speed when they are not in platoons, while others are constrained due the following of slower vehicles in the platoons Beyond the point mentioned above (84% platoon percent and 40 mph speed), as the speed decreases further and the platooning percentage increases, the relationship between the percent of vehicles in the platoon and SDt follows a negative log linear relationship, as shown in Figure 2-7. The decrease in the SDt with the increase in platoon percentage, and thus congestion, is expected since the increase in platooning after breakdown will further reduce the ability of vehicles to travel at their desired speeds. As shown in Figure 2-7 and Figure 2-8, SDt decreases from 12 ft/sec when 84% of the vehicles were in the platoon and the average speed was 40 mph to 6 ft/sec at a speed of 20 ft/sec (13 mph) and 95% or more of the vehicles in platoon. It should also be mentioned that a direct relationship has also been found between the SDt and safety in previous work. Thus, the discussed relationship may be used to further assess the impacts of different levels of platooning on safety, in addition to mobility.

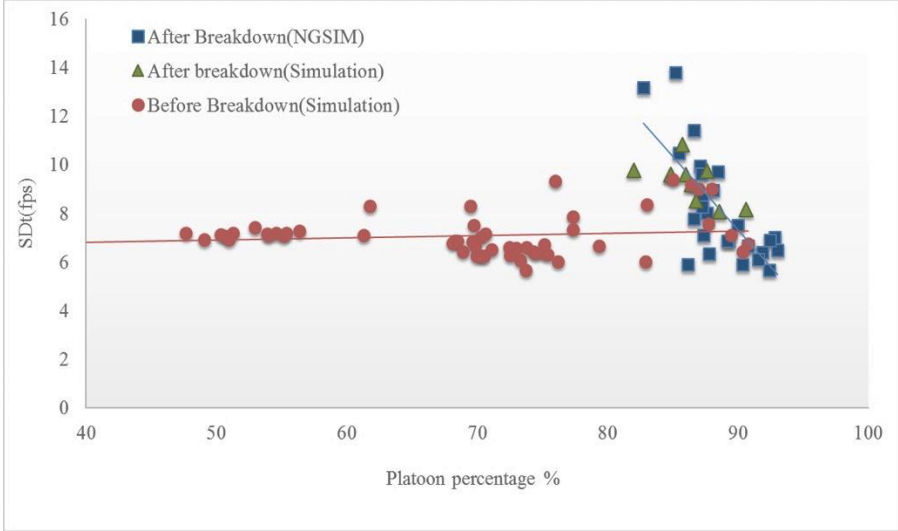


FIGURE 2-7: THE RELATIONSHIP BETWEEN SDT AND THE PERCENTAGE OF VEHICLES IN THE PLATOON

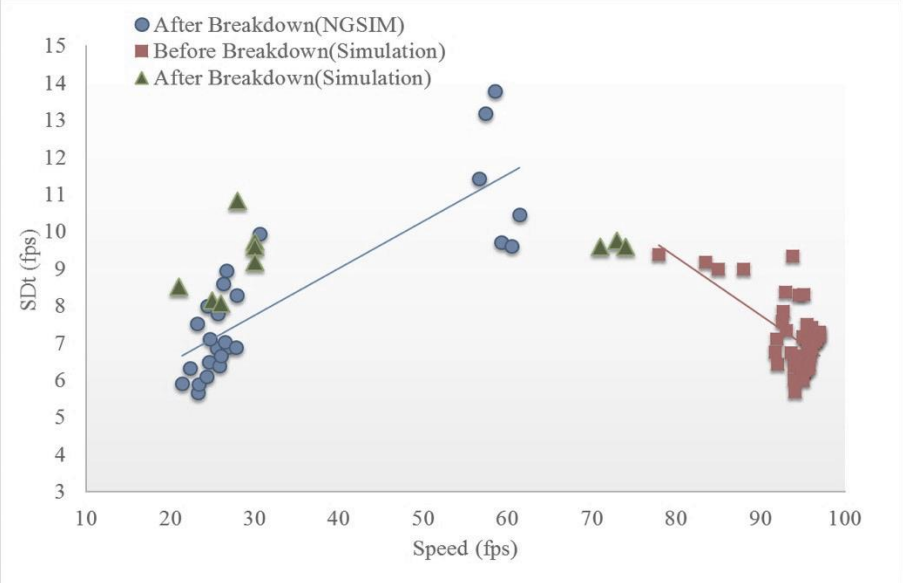


FIGURE 2-8: THE RELATIONSHIP BETWEEN SDT AND THE SPACE MEAN SPEED

The fitted function between the percent of vehicles in the platoon and SDt for "after traffic breakdown" is provided in Equation 2-15.

$$P=\alpha+\beta\log(SD_T) \qquad \textbf{2-15}$$

Where P is the Percentage in the Platoon in the "After Breakdown" conditions, SDt is the standard deviation of speed between vehicles (ft/s), and α and β are coefficients. The statistical

software R was used to fit the regression equations and produce the statistics required to assess the significance of the relationships. The regression analysis results are presented in Table 2-3. Various transition forms were investigated, and the best form was selected. In addition to the statistical test results presented in Table 2-3, the model was also validated using the residual plot and the quantile-quantile (q-q) plot.

TABLE 2-3: DEVELOPED EQUATIONS TO ESTIMATE THE PERCENT IN PLATOON

Equation	α	β	R-squared value	Adjusted R-squared	t value of α (Pr(>|t|))	t value of β (Pr(>|t|))	Residual (P-value)
After Breakdown	111.09	-10.67	0.6354	0.625	38.565 (2e-16)	-7.811 (3.55e-09)	3.547e-09

The equation in Table 2-3 can be used to estimate the percent of vehicles in platoons during breakdown conditions that appear to occur when 84% of the vehicles or more are in the platoon and the speed is below the estimated speed at capacity. The equation can be used to estimate the percent of vehicles in the platoon as a function of SDt when these estimates are made based on data from a low market penetration of connected vehicles. This is because the platooning characteristics cannot be measured directly at low CV market penetrations. In this study, the accuracy of this estimation was assessed at market penetrations of 5%, 10%, 50% and 80%. For each market penetration, the percent of vehicles in the platoon was calculated based on the measured standard deviation of speed between vehicles. The assessment was performed utilizing the NGSIM data for the I-80 segment described previously. The vehicles with connected vehicle equipment for each CV penetration were selected randomly from all vehicles in the traffic stream. However, the accuracy of the estimation of the speed standard deviation is expected to depend on this selection when there is a high variation in the speed characteristics of the vehicles in the link, particularly at low market penetrations. Thus, a Monte Carlo analysis was used in this study to account for this stochasticity by randomly selecting different CV vehicles from the traffic stream for each Monte Carlo run. Twenty Monte Carlo runs were conducted, and the estimates of the percent of vehicles in the platoon were estimated using the equation in Table 2-3, based on the SDt generated from these runs, and were then compared with those actually measured using vehicle trajectories from a 100% market penetration of CV. The following equations provide expressions of the measures used to assess the quality of the estimation:

$$\text{MEAN ABSOLUTE PERCENT ERROR (MAPE)} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y}{y_i}\right| \quad \text{2-16}$$

$$\text{STANDARD DEVIATION OF PERCENTAGE ERROR (SDPE)} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(w_i{}^2 - n\bar{w}^2)} \quad \text{2-17}$$

where $w_i = \frac{y_i - y}{y_i}$, y_i is the estimated value of the ith run, y is the value at 100% MP, n is the total number of runs (n=20), and $\bar{w}$ is the average of all the w_i. Different quality measurements have different significance. SDPE is a measure of reliability of the estimates. MAPE is the average error of all runs. The results of the quality measure calculations are shown in Table 4-2. As can be seen from the results in Tables 2-4, the error in the estimation of the platoon percentage is low. Thus, it can be concluded that the platoon percentage can be estimated at the low market penetration of CV data accurately based on the methodology presented in this study.

TABLE 2-4: THE QUALITY OF THE ESTIMATION OF THE PERCENTAGE OF VEHICLES IN PLATOON WITH DIFFERENT MARKET PENETRATIONS

Accuracy Measure	Market Penetration			
	5%	10%	50%	80%
MAPE	1.9010	1.0423	0.2968	0.1882
SDPE	1.2912	0.8532	0.2910	0.1448

2.4.2.1.1 SENSITIVITY ANALYSIS ON THE UTILIZED TIME HEADWAY THRESHOLD

An analysis was conducted to determine the sensitivity of the platoon percentage in this study for the utilized time headway threshold. Figure 2-9 shows how the estimated percentage of vehicles in a platoon in the NGSIM data varies as the standard deviation of speed between vehicles changes with different time headway thresholds (5-second, 4-second and 3-second time headways).

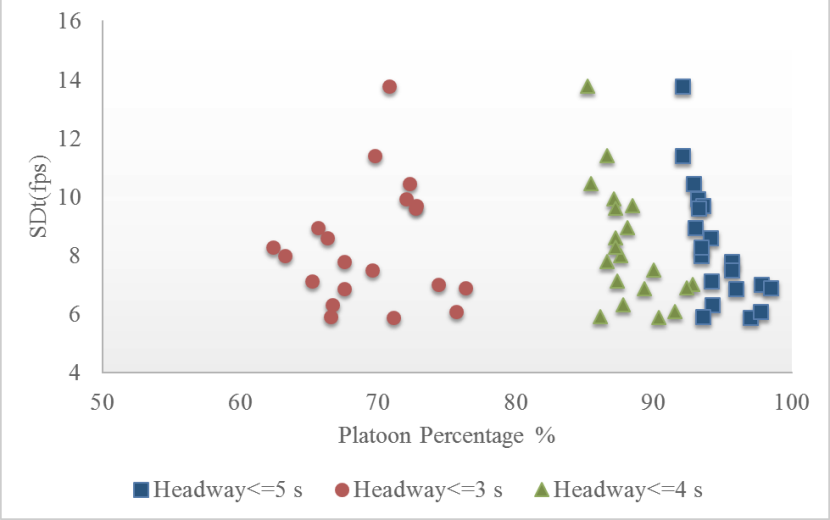


FIGURE 2-9: PLATOON PERCENTAGES ESTIMATED USING DIFFERENT HEADWAY THRESHOLDS

As can be seen from Figure 2-9, the use of a 3-second headway as a threshold results in a lower percentage of vehicles in platoon and the use of a 5-second headway results in a higher percentage of vehicles in platoon than the 4-second headway. During a traffic congestion period, only 63% to 76% of the vehicles were identified as platooning when using the 3-second

headway criterion. This indicates that using the 3-second headway underestimate the percentage of vehicles in platoon. When the 5-second headway is used as the threshold, 100% of vehicles were identified as platooned. However, examining the standard deviation of the speed of individual vehicles shows that there are non-platooned vehicles, indicating that the 5-second threshold possibly overestimate the platooned vehicles percentage. Thus, the 4-second headway is used as the threshold in this study.

2.4.2.2 Determination of Platoon Size Distribution

In addition to the percent of vehicles in the platoon discussed in the previous section, this study also investigates estimating the platoon size distribution based on the standard deviation of each vehicle speed (SDv) at different market penetrations. The estimation is conducted for the "after breakdown." conditions Each non-platooned vehicle is considered a special vehicle platoon with platoon size equal to one. Tian et al. (20) developed an equation to estimate the platoon size based on the standard deviation of individual vehicle speed using NGSIM data for the U.S. 101 Freeway in California. In this study, we obtained the standard deviations of speed for several vehicles with different positions in the platoons utilizing NGSIM I-80 data. It was determined that the equation developed by Tian et al. (2017) produces reasonable results for the purpose of this research. This equation, as presented in the review of literature in Equation 2-1, was utilized in this study and rearranged to calculate the size of the platoon as a function of the SDv, as follows:

$$\text{NVP} = -94.29(\ln(10.56 - 0.3048\text{SD}_\text{v}) - \ln(10.4)) \qquad \textbf{2-18}$$

where SDv is the standard deviation of the speed of each vehicle (ft/sec), and NPV is the number of vehicles in the platoon. The position of each vehicle in the platoon was estimated based on Equation 2-18 and with different CV market penetrations. The estimated positions were categorized into four groups, as shown in Table 2-5. The results in Table 2-5 are based on the full set of trajectories of the I-80 NGSIM data in a peak period from 4:00 p.m. to 4:15 p.m.

TABLE 2-5: PLATOON SIZE PERCENTAGE BASED ON ALL TRAJECTORIES

Subsegment	%. of Vehicles in Platoon Size Group			
	1<=x<=2	2<x<=5	5<x<=15	x>15
1	0.75	13.43	57.09	28.73
2	3.14	28.63	60.39	7.84
3	3.82	32.44	56.87	6.87
4	2.27	25.38	57.58	14.77
5	4.60	24.52	53.64	17.24

The percentage of vehicles in each platoon size group was estimated utilizing data that assumed different market penetrations of CV. The accuracy of this estimation was assessed at market penetrations of 5%, 10%, 20%, 50% and 80%, compared to the base value of the comparison, which is the estimation when utilizing data from CVs at 100% market penetration. Twenty Monte Carlo runs were conducted to account for the stochasticity due to the random selection of different CV vehicles from the traffic stream, as discussed earlier. Each of these

runs represents a different day of operations with different vehicles selected to be equipped with CV technology for each day.

The t-test of the difference in the mean and the Chi-square ($\chi 2$) on the difference in the frequency distribution tests (McShane, 2011) were conducted to assess the accuracy of the estimated platoon size mean and distribution, respectively. For example, Table 2-6 shows the t-test and $\chi 2$ results for different market penetrations for one of the test's subsegments. The results of the t-test on the difference in the mean and the $\chi 2$ test of the platoon size distribution based on the mean of the 20 runs can be used to assess the adequacy of the estimation for planning purposes. The data used are average values based on data from 20 days, represented by the 20 runs. The results of the $\chi 2$ for each of the 20 days (runs) represent the adequacy of the estimation for use in the real-time operation for that single day.

TABLE 2-6: T-TEST AND X2 TEST FOR DIFFERENT MARKET PENETRATION LEVELS

Hypothesis test	Chi-square test on Frequency Distribution		t-test on the Mean	
Market Penetration	Is the Average of the Runs passing Test (Rejecting Null Hypothesis)	No. of Individual Runs Passing (Rejecting Null Hypothesis)	t-value	Pass (Rejecting Null Hypothesis)
5%	Y	10 out of 20	1.013	Y
10%	Y	14 out of 20	0.5239	Y
20%	Y	18 out of 20	0.6665	Y
50%	Y	20 0ut of 20	1.3209	Y

As can be seen from Table 2-6, when comparing the results from using the data at the 5% and 10% market penetration levels and using the data at the 100% level, the null hypothesis of no difference between the estimation of the means according to the t-test and no difference in the frequency distribution of the platoon size based on the average of the runs could not be rejected at the 95% confidence level. However, when considering the platoon size distribution for individual days, the null hypothesis can be rejected for 10 of the 20 days with a 5% market penetration, and 6 of the 20 days with a 10% market penetration at the 95% confidence level, indicating that the estimate is not adequate for operations at the 5% and 10% market penetration levels

At a higher market penetration of 20%, the null hypothesis of the $\chi 2$ test can be rejected for most of the individual runs (days) at the 95% confidence level. With this CV market penetration (20%), the estimates can be used for both planning (based on the average of the runs) and operations (based on individual runs). Table 2-7 shows a comparison of the platoon size distribution using the data from each of the runs of the Monte Carlo simulation with a 20% market penetration, the average of the runs with a 20% market penetration, and the

distribution based on a 100% market penetration of CV. Inspection of the data in this table confirms that most Monte Carlo runs with the 20% CV market penetration produces results that are comparable to those obtained with the 100% market penetration.

TABLE 2-7: COMPARISON OF THE PLATOON SIZE DISTRIBUTION ESTIMATED WITH THE 20% AND 100% CV MARKET PENETRATION LEVELS

Market Penetration	Run	% of Vehicles in Platoon Size Group				Mean of value
		1<=x<=2	2<x<=5	5<x<=15	15<x<=30	
20%	1	0.00	3.77	69.81	26.42	13.06
	2	3.77	13.21	64.15	18.87	11.18
	3	1.92	7.69	65.38	25.00	12.46
	4	0.00	13.21	47.17	39.62	14.09
	5	1.89	15.09	56.60	26.42	12.16
	6	0.00	9.43	62.26	28.30	12.92
	7	1.89	9.43	58.49	30.19	13.00
	8	0.00	13.21	60.38	26.42	12.44
	9	1.89	15.09	52.83	30.19	12.63
	10	0.00	13.21	66.04	20.75	11.74
	11	0.00	18.87	49.06	32.08	12.78
	12	0.00	16.98	50.94	32.08	12.91
	13	1.89	11.32	54.72	32.08	13.11
	14	1.89	11.32	56.60	30.19	12.88
	15	1.89	16.98	50.94	30.19	12.51
	16	0.00	13.21	56.60	30.19	12.92
	17	0.00	20.75	50.94	28.30	12.19
	18	0.00	13.21	58.49	28.30	12.68
	19	0.00	7.55	54.72	37.74	14.23
	20	0.00	9.43	60.38	30.19	13.16
	Mean	0.85	12.65	57.33	29.17	-
100%	-	0.75	13.43	57.09	28.73	12.65

2.4.2.2.1 RELATIONSHIP BETWEEN PLATOONING AND DISTURBANCE METRICS

This section discusses the relationships between platooning measures derived as expressed above and the disturbances metrics utilized in this study. The number of vehicles in platoon (NVP) was calculated based on Equation 2-18. The relationships between NVP and disturbance metrics were then explored as displayed in Figure 2-10.

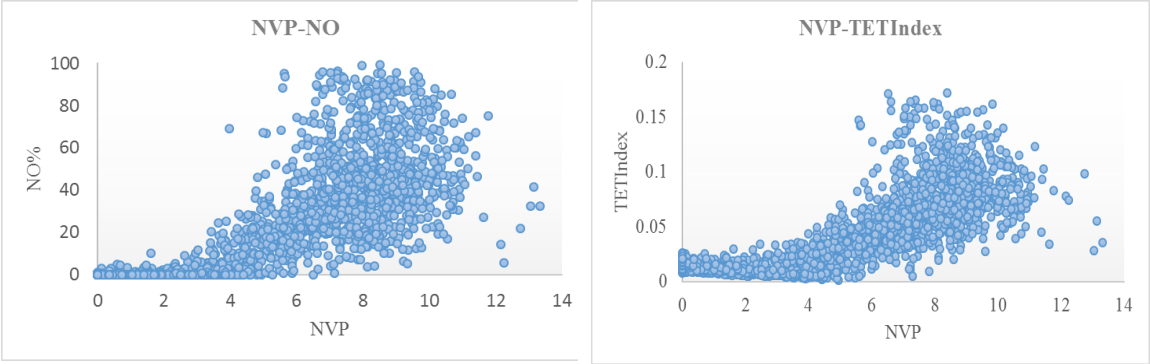


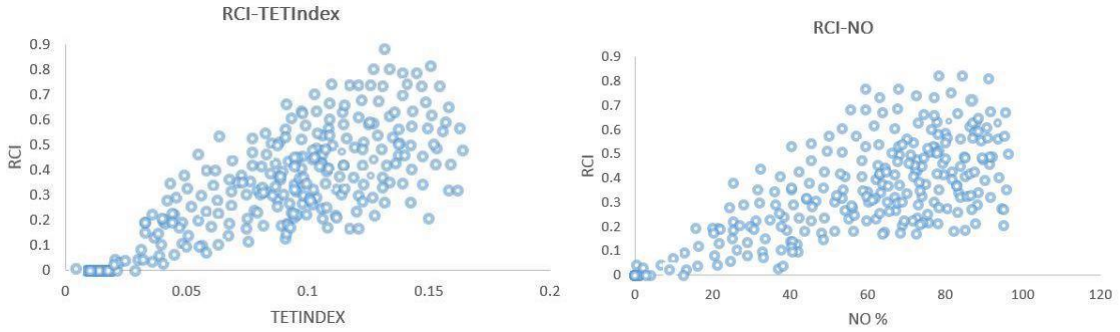
FIGURE 2-10: RELATIONSHIP BETWEEN DISTURBANCE METRICS WITH NVP

As can be seen from Figure 2-10, the NVP increases after NO of 10% and TETIndex of 0.03. As stated earlier, the values of the TETIndex and NO also start increasing sharply at NO greater than 10% and TETIndex higher than 0.03. The NPV is greater than 15 when the NO exceeds 20% and TETIndex exceeds 0.05, which are the values previously reported where the traffic condition is in congested region. These results are consistent with founding in section 2.4.1.1.

2.4.3 Safety Measures

Real-time safety assessment can be used as an important input to traffic management and operations. It is envisioned that when a threshold value of the estimated TETIndex and NO as a surrogate measure to safety are exceeded, a central decision support system (DSS) will recommend the activation of operational plans on the freeway such as metering and/or dynamic speed limit to reduce the probability of crashes. Thus, there is a need to identify the thresholds for these parameters that justify the activation of the plans. For this identification, this study uses a visual inspection of the graphical views of the relationships between the disturbance metrics and the speed parameters and RCI.

The relationship between the speed parameters are shown in Figure 2-5. Visual inspection of the relationships of the two disturbance metrics (TETIndex and NO) and the speed parameters are shown in Figure 2-6. As it was explained in the section of 2.4.1, The values of the TETIndex and NO at the break point beyond which the two values of the two variables increase sharply are about 0.03 and 10%, respectively. Figure 2-7 shows that higher values of the disturbance metrics including a TETindex of 0.05 and a NO of 20% are clearly in the congestion region of the curves. Further inspection was done by examining the relationship between the TETIndex and NO and a third surrogate measure to safety, the RCI plotted in Figure 2-11. According to Oh et al (Oh et al., 2006), RCI values less than 0.2 can be considered as acceptable from rear-end crash risk point of view. Utilizing RCI value of 0.2 as reference, Figure 2-11 shows that this value is associated with TETIndex less than 0.03 and NO less than 10% and thus has an acceptable rear-end crash risk. Thus, based on the above, it was decided to use TETIndex of 0.03 and NO of 10% in this study, as the thresholds to determine unsafe conditions and potentially activate strategies to mitigate these conditions.



(a) (b)

FIGURE 2-11: THE RELATIONSHIP BETWEEN RCI AND (A) TETINDEX AND (B) NO

2.4.3.1 Developing Model to Estimate TETindex at Low Market Penetration of CV Data

As mentioned earlier, the TETIndex is not obtainable from small samples of data. Thus, regression analyses were conducted to derive the relationship between the TETIndex as the dependent variable and speed, SDv, and SDt, as the independent variables based on the full set of simulation results. The first step was to examine any collinearity in the data by producing the correlation matrix among the three variables.

In addition, factor analysis was used to investigate how the variable values cluster to eliminate the impacts of independent variables correlating strongly with each other without significantly deteriorating the model fit of the data. Based on this analysis, it was determined that due to the relatively high correlation, the SDv variable can be either eliminated or respecified by utilizing a new variable that is a function of SDv (Johnson & Wichern, 2007). The respecification was done by dividing SDv by speed. The resulting new variable was then used as an independent variable. This new variable is actually the coefficient of variation of the speed of individual vehicles. Several linear and non- linear functions between the TETIndex, Mean Speed (S), SDv and SDt were tested using multivariate regression analyses using the statistical software R. However, the best fit was found when using a machine-learning regression method referred to as Partial Least Square (PLS) regression, which is a robust method in prediction and can handle data, which are strongly collinear and noisy (Johnson & Wichern, 2007; Wang, et al., 2008). The resulting equation is provided below:

$$TETINDEX = \beta_0 + \beta_1 SDT + \beta_2 (SDV / S) \qquad 2\text{-}19$$

where, β_i are the regression coefficients, and the other variables are as defined before. The unit of speed parameters is ft/s. The PLS regression result is in Table 2-8. The quality of the model shows an acceptable R-squared (the coefficient of determination) and Q-squared (the cross

validated value which is calculated on the basis of the cross-validation). The error of the model based on the comparison of the estimated and observed TETIndex in the simulation is 14%. Substituting the previously identified critical values of SDv , SDt and mean speed of 2 ft/s, 8 ft/s and 65 ft/s respectively, in Equation 2-19, result in TETIndex of 0.03, which is the critical value identified for this parameter earlier, further confirming that this model produce reasonable value.

TABLE 2-8: COEFFICIENTS AND QUALITY OF THE TETINDEX ESTIMATION BASED ON PLS REGRESSION

Variable	Coefficient	Std. deviation	Lower bound (95%)	Upper bound (95%)
β_0	0.010621	0.001	0.009	0.012
β_1	0.000241	0.0001	0.00007	0.001
β_2	0.637769	0.008	0.621	0.655
The quality of the model:				
Statistic	**Comp1**	**Comp2**		
Q^2 cum	0.79	0.88		
R^2 cum	0.79	0.88		
MAPE= 14%				

To validate the performance of the regression model presented in Equation 2-19, the model was tested for two locations with real-world data collected as part of the NGSIM program (I-80 and US101). This testing was conducted to assess the transferability of the regression model developed based on simulation data to other locations not used in the calibration of the simulation model. The model was tested under different sample size of data. The vehicles that were assumed to be equipped with technology that provides their trajectories for use in the model application were selected randomly from all vehicles in the traffic stream. Since the accuracy of the estimation is expected to depend on the random selection, a Monte Carlo analysis was used to account for the stochasticity by randomly selecting different vehicles from the traffic stream for each Monte Carlo run. Twenty Monte Carlo runs were conducted for each investigated market penetration of data, and the TETIndex was estimated based on speed parameters using Equation 2-19. Statistical tests indicated that 20 runs are sufficient. The results obtained with each market penetration were compared with the base value, which is the TETIndex calculated using Equation 2-8 based on the full NGSIM data. The MAPE and SDPE (Equations 2-16 and 2-17) used to assess the quality of the estimation.

The error of the model based on the MAPE and SDPE are reported in Table 2-9 for the two NGSIM datasets. As can be seen from Table 2-9, the developed regression model based on simulation was able to predict the TETIndex at an accuracy of 15% to 20% for locations that are different from the location used in the simulation.

TABLE 2-9: THE QUALITY OF ESTIMATIONS OF THE DEVELOPED REGRESSION MODEL FOR DIFFERENT SAMPLE BASED ON NGSIM DATA

Accuracy measure (Mean Value)		Sample Size			
		5%	10%	20%	100%
I-80	MAPE	20%	19.8%	19.6%	19%
	SDPE	0.78	0.51	0.45	0.29
US101	MAPE	16%	15.5%	15.8%	15%
	SDPE	0.55	0.35	0.28	0.18

2.4.3.2 Real World CV Data Analysis

As mentioned earlier, CV data was extracted to evaluate the performance of the application of the model for a freeway segment in Michigan. The data indicates that the speed on the segment varied between 45 ft/s and 120 ft/s, and the peak period was between 3:00 P.M. and 5:00 P.M. The data was aggregated for each 5-minute interval. Then, the mean speed, average SDv, SDt and NO parameters were calculated based on vehicle trajectories for each interval. Since the TETIndex cannot be measured at low market penetrations of CV, Equation 2-19 was used to calculate the TETIndex based on speed parameters. Descriptive statistics of the measured traffic parameters for the segment were shown in Table 2-10.

Test location crash data for the year 2013 was also obtained and integrated with the traffic data for the time interval prior to crash occurrence. Crash was used as a binary dependent variable. A total of 35 crashes occurred in the test location in 2013. It is interesting that all crashes happened at TETindex values greater than 0.025. More sever crashes happened at TETIndex higher than 0.03 and more property damage crashes happened at TETIndex higher than 0.05, where the traffic conditions become congested.

Before applying statistical model to the crash, the Random Forest (RF) was used to rank the important variables on crash occurrences. RF is a non-parametric statistical method that is based on decision trees (Bishop, 2006). The R package “randomforest’ (Liaw, 2002) was used to select the important variables. A higher accuracy represents a higher variable importance. The results are shown in Figure 2-12. As can be seen, the most important variables in predicting the occurrence of crashes are the TETIndex and NO.

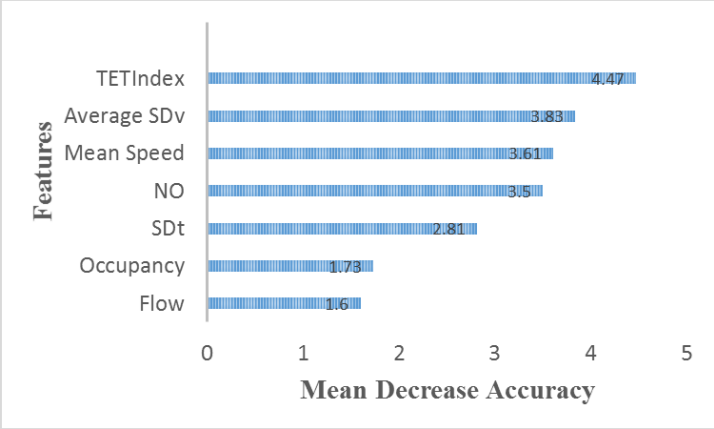


FIGURE 2-12: VARIABLE IMPORTANCE AS IDENTIFIED BY THE RANDOM FOREST

Since some important variables identified by RF technique are correlated with each other, first, the Principal Component Analysis (PCA) was utilized to reduce the dimensionality of data and handle the collinearity between variables. The utilized number of components was selected based on the number of input features and projected variance (Krauss, 1997).

The Random Forest technique was used to fit a crash model to the data using the Scikit Learn library in Python. The values of each input feature were standardized before use. As data samples are small, all data were used as a training dataset and k-fold cross validation was used to estimate the accuracy of the model with 5-fold. Fine-tuning of model parameters was done using the Grid Search in the utilized tool, to tune the model by searching for the best hyper parameters and keeping the classifier with the highest accuracy. The overall accuracy with different utilized variables along with the tuned parameters is reported in Table 2-10. As can be seen, the developed model of the crash frequency for the Michigan test segment corresponds more with adding disturbance metrics to input variables than excluding them. This indicates that the utilized disturbances metrics are good indicators of traffic safety and they can use as inputs to predict crash in real-time operations.

TABLE 2-10: SUMMARY OF TUNED PARAMETERS AND MODEL ACCURACY WITH DIFFERENT INPUT VARIABLES

Evaluation		All variables	All variables without disturbance metrics
Overall accuracy (SD)		0.82 (+/- 0.15)	0.78 (+/- 0.21)
Tuned parameters	max_depth	3	3
	max_features	1	1
	n_estimators	100	100

**SD=Standard Deviation*

** n_estimators are the number of trees used in the ensemble,*

max_depth controls the depth of each tree, and

max_feature is the size of the random subsets of features to consider when splitting a node

2.4.4 Traffic State Identification

This section describes the method developed for traffic state recognition (identification) from traffic operation point of view. Prior to conducting clustering for the purpose of traffic state recognition, the study performed an initial exploration of the three fundamental macroscopic variables (flow, density, and speed), for the simulated data as presented by the fundamental diagrams and shown in Figure 2-4. As mentioned in section 2.4.1, a Change-Point Regression with Gaussian Mixture Analysis was conducted using speed and occupancy data and the critical speed at capacity was found to be 65 ft/second. However, the fundamental diagram, shown in Figure 2-4 indicates that the relationships between the three fundamental macroscopic traffic features are very scattered and that the critical speed is somewhere between 50 ft/second and 70 ft/second and the critical density is between 40 vehicle/mile/lane to 60 vehicle/mile/lane. This further indicates the need to use clustering and consider the microscopic features to identify breakdown.

As mentioned earlier, the GMM clustering was used to identify the uncertain traffic condition phase when the three macroscopic features (traffic flow, speed and density) were used in the clustering. Different numbers of components based on LOS (two to six) were used in the GMM clustering, and the SC was checked to determine the best number of components. It was found that the GMM with two components is the best investigated option. When examining the resulting GMM clusters, it was found that 365 datasets out of the 2997 are overlapping between two clusters and have probability of greater than zero to being on another cluster. Interestingly, these datasets are all in the mean speed range between 50 ft/seconds to 70 ft/seconds. This confirms the stochasticity observed in the fundamental diagrams in Figure 2-4.

The next step was to cluster the data in the uncertain traffic condition phase based on different combinations of macroscopic and microscopic features using the K-Means clustering algorithm. The results were then examined to determine the best combinations of features in separating the traffic state into two clusters of "breakdown" and "non-breakdown". Please note that the goal of the study was defining traffic sate regarding breakdown, however, different number of clusters (k) was also attempted and the optimum number of clusters was found to be two. The percentages of the data in the "breakdown" and "non-breakdown" clusters obtained using different combinations of features in the K-Means clustering and also those separated by the deterministic value of speed at capacity are shown in Table 2-11 based on whole dataset. As shown in Table 2-11, the percentage of the data in the breakdown cluster ranges from 32% to 42% depending on the utilized features.

TABLE 2-11: PERCENTAGE OF TRAFFIC STATES IN THE BREAKDOWN CLUSTER (GROUP1) AND NON-BREAKDOWN CLUSTER (GROUP2) WHEN USING DIFFERENT COMBINATIONS OF FEATURES

Clusters		Features Utilized in K-Means Clustering															
	S< 65 ft/s	TET-S-Oc-SDv	NO-S-Oc-SDv	S-Oc-SDv	TET-S-Oc	TET-S-Oc-NO	S-Oc-NO	S-Oc	TET	SDv	NO	S-TET-Flow	S-SDv-Flow	S-NO-Flow	D-Flow-SDv	D-Flow-NO	D-Flow-TET
(%) (Group 1)	35.2	35.8	36.1	37.4	31.8	32.0	33.8	34.1	23.0	42.0	25.9	36.2	41.7	37.3	41.0	36.3	33.9
(%) (Group 2)	64.7	64.2	63.8	62.5	68.1	68.0	66.1	65.8	74.0	58.0	74.1	63.7	58.3	62.7	59.0	63.7	66.1

**S=Space Mean Speed, Oc=Occupancy, D=Density, SDv=Average of Standard deviation of individual's vehicles, NO=Number of oscillations, TET=TETIndex*

The next step is to assess the results from the K-Means clustering presented in Table 2-11. First, the results were examined to determine the ability of different combinations to represent certain state correctly. Certain states were defined as speed lower than 50 ft/sec and greater than 70 ft/sec as the congested and uncongested conditions respectively. The percentage errors on certain conditions were then reported and are shown in Table 2-12. It can be seen that only five of the sixteen investigated combinations of the features produced zero errors in certain conditions. Note that additional combinations with flow and density, speed and flow and adding the SDt to the combinations were tried but were not reported in Table 2-11 and Table 2-12, since they did not produce improvements to the analysis.

Further examination was done by visually inspecting examples of trajectory data to determine the ability of the above options to isolate stop and go conditions from other conditions. Two random examples were selected to illustrate this visual inspection (Figure 2-13). Examples (a) and (b) have speed less than 65 ft/s and should be categorized as breakdown according to the change-point regression based on the speed-occupancy relationship. However, Example (a) is stable while Example (b) is unstable (have stop and go conditions).

TABLE 2-12: EVALUATIONS OF THE DERIVED CLUSTERS

	Features Utilized in K-Means Clustering																
	S< 65 ft/s	**TET-S-Oc-SDv**	**NO-S-Oc-SDv**	**S-Oc-SDv**	**TET-S-Oc**	**TET-S-Oc-NO**	**S-Oc-NO**	**S-Oc**	**TET**	**SDv**	**NO**	**S-TET-Flow**	**S-SDv-Flow**	**S-NO-Flow**	**D-Flow-SDv**	**D-Flow-NO**	**D-Flow-TET**
Percent Error in certain congested	0	0	0	0	0.1	0	1.2	0	10.6	0	12.2	0.1	0	0.1	0.4	0.1	0.6
Percent Error in certain uncongested	0	0.1	0.1	0.48	0	0	0	0	0	7.3	0	0.8	6.29	1.29	5.9	1.7	0.64
Breakdown evaluation based on the three examples (corresponding with Figure 2)																	
Example (a)	N	N	N	N	Y	Y	Y	N	Y	N	Y	Y	N	N	N	Y	Y
Example (b)	N	N	N	N	N	N	N	N	Y	N	Y	Y	N	N	Y	N	N
Average Silhouette	-	0.67	0.7	0.69	0.7	0.69	0.68	0.71	0.73	0.77	0.76	0.53	0.56	0.54	0.53	0.52	0.51

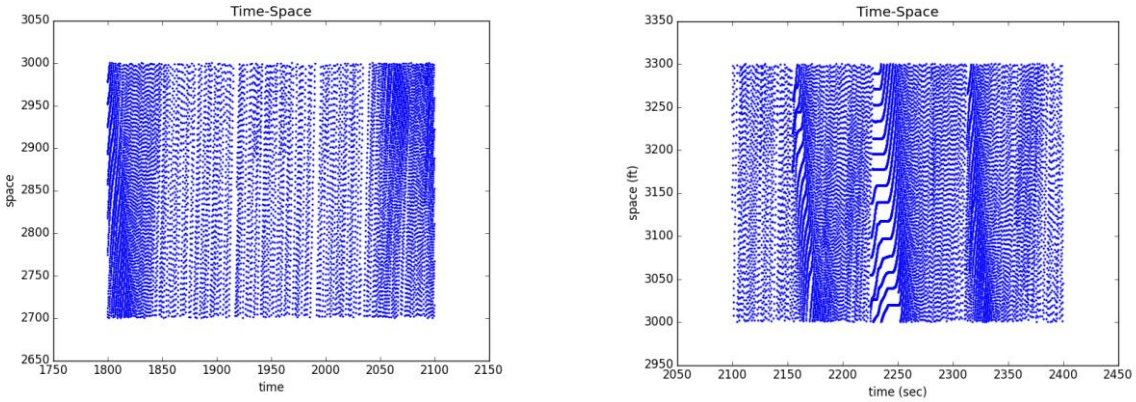
***S=Space Mean Speed, Oc=Occupancy, D=Density, SDv=Average Standard deviation of individual's vehicles, NO=Number of oscillations, TET=TETIndex*

*** Bold colors shows clustering options that passed the evaluation correctly*

**** Y=stable condition (non- breakdown) N=unstable condition (breakdown)*

***** Certain congested is defined condition of speed lower than 50 ft/s and certain uncongested is defined conditions of speed greater than 70 ft/s*

Page left blank.



(a) Mean Speed=56.6 ft/s, TETIndex= 0.03, NO= 9%

(b) Mean Speed=51.8 ft/s, TETIndex=0.05, NO=20%

FIGURE 2-13:TIME-SPACE FOR THE TWO EXAMPLES OF THE VISUAL INSPECTION

Based on the results presented above regarding the error in assigning data to the congested /uncongested conditions in relation to speed, visual inspection, and the average SC; as shown in Table 2-12, it was determined that clustering with the "TETIndex, NO, Mean Speed and Occupancy" features, is the best clustering option among the investigated ones. The performance of utilizing the clustering based on these features (TETIndex, NO, Mean Speed and Occupancy), referred to as the "selected clustering" in the rest of this document, to identify "breakdown" and "non- breakdown" conditions was further compared to the utilization of clustering based on speed and occupancy only and also to using the 65 ft/sec deterministic threshold determined according to the change-point regression.

The results are reported in Table 2-13 and show that the clustering based on the four macroscopic and microscopic features can better account for the number of disturbance and disturbance duration that reflect the slow and go operations compared to the other two options. As can be seen from Table 2-13, the clustering with the "TETIndex, NO, Mean Speed and Occupancy", captures more disturbances in the breakdown cluster compared to using the deterministic value of speed at capacity and the clustering based on speed and occupancy. This selected clustering also reports the highest percentage of "non-breakdown" conditions for the whole dataset. It also clusters 56.1% of the uncertain phase (the measurements with speed between 50 ft/sec and 70 ft/sec) as breakdown. Using the deterministic value of speed suggests 79.2% of the uncertain phase is in breakdown. It can be seen that the total improvement of the selected clustering compared to the clustering with the deterministic speed at capacity and with clustering with speed and occupancy are about 30% and 20%, respectively. This means that 30% of the data in the uncertain phase considered as congested condition when using the deterministic value of a speed of 65 ft/sec are actually stable.

TABLE 2-13: COMPARISON OF THE SELECTED CLUSTERING WITH A DETERMINISTIC VALUE OF SPEED AND WITH CLUSTERING BASED ON SPEED-OCCUPANCY IN CAPTURING THE AMOUNT OF DISTURBANCES IN THE UNCERTAIN PHASE (50 FT/SEC<MEAN SPEED<70 FT/SEC)

Condition in the Uncertain Phase (385 dataset out of 2997)	Mean Speed < 65 ft/s	Mean Speed and Occupancy	TET, NO, Mean Speed, and Occupancy
Breakdown Condition (%)	79.2	70.6	56.1
% of Reading NO>20% in cluster of breakdown	71.8	69.1	82.8
% of Reading TET>0.05 in cluster of breakdown	33.4	32.35	44.9

Moreover, it was found that in the breakdown cluster, the value of TETIndex is greater than 0.05 and the NO is greater than 20%. These two microscopic features could be considered together to decide if traffic flow perturbation could grow leading to breakdown or not. As described earlier. Table 2-14 shows the percentage of the captured disturbances from the selected clustering. As can be seen, the TETIndex of higher than 0.05 and NO higher than 20% can capture disturbances fully in breakdown conditions, while using the TETIndex and NO separately fails to identify the instability fully.

TABLE 2-14: THE RESULTS OF SELECTED CLUSTERING IN CAPTURING DISTURBANCES IN CLUSTER OF BREAKDOWN

Defined traffic conditions	**% of Reading TETIndex>0.05**	**% of Reading NO>20%**	**% of Reading TETIndex>0.05 And NO>20%**
Certain congested	91.7	95.2	100%
Breakdown (Uncertain Phase)	44.9	82.8	100%
Non-breakdown (Uncertain Phase)	7.05	38.2	0%
Certain uncongested	0.1	0.85	0%

**From whole dataset 13% are in uncertain phase and 25% and 62% are in congested and uncongested conditions, respectively*

The TETIndex of 0.05 and NO of 20% are also identified as values representing breakdown conditions based on the visualization from Figure 2-14. Figure 2-14 shows the relationship between the TETIndex and NO with the SDv, based on simulation results and NGSIM data (from US 101 in California). As can be seen, a TETIndex of 0.05 and NO of 20% are the change points where the diagrams bend and start a sharp increase further indicating that these values are good indicators of breakdown. Please note that although Figure 2-14 shows dataset from both VISSIM and NGSIM, only VISSIM data was used to develop the method and the NGSIM data was used to test the method.

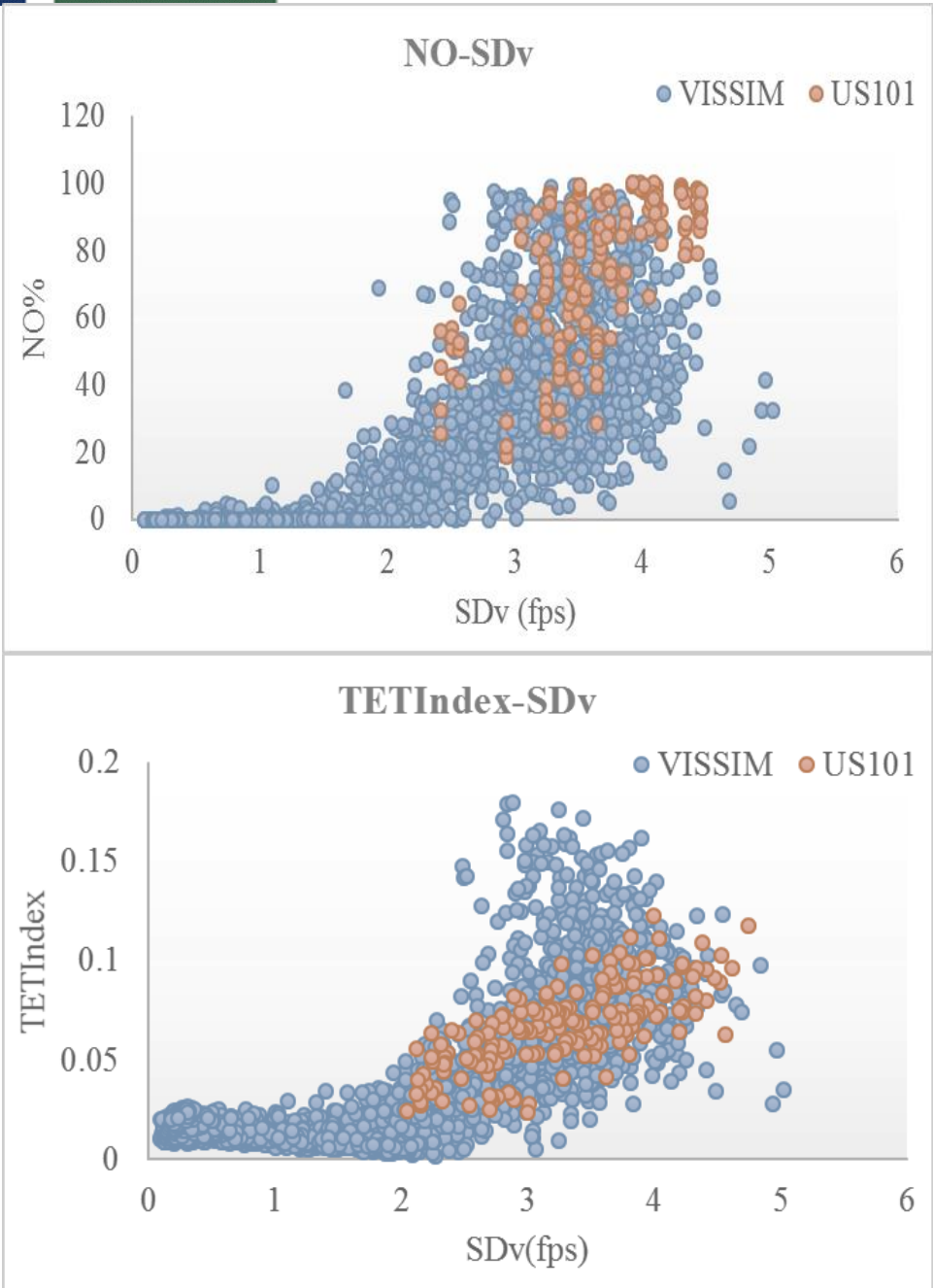


FIGURE 2-14: THE RELATIONSHIP BETWEEN DISTURBANCES WITH SDV

2.4.4.1Testing the Method with Trajectory Data of NGSIM US 101

The NGSIM dataset of US 101 in California was also used to test the proposed method. Although this dataset does not cover the different traffic states, the study investigated the inclusion of capturing the traffic disturbances in the traffic state identification. To be consistent with study by Lu et al. (2009) on NGSIM US 101 dataset about analyzing fundamental diagram, the same segment and same time window of one second was selected. The test segment is located in the upstream section of US101 in lane one which is about 550 feet long and for time period of 7:50 ~ 8:05 A.M. Lu et al(Lu et al., 2009) defined the critical density of around 95

vpml. As described in proposed method, first, the GMM clustering with macroscopic features was used and uncertain phase was identified. This uncertain phase located between density of 85 vpml and 100 vpml. The identified uncertain traffic condition phase from the GMM analysis is then further analyzed using the K-Means clustering based on "TETIndex, NO, Mean Speed and Occupancy" to classify the uncertain phase into two clusters. The results are shown in Figure 2-15. The examination of Cluster 1 and Cluster 2 show that the TETIndex is between 0.0\6 to 0.099 in cluster 1 and is between 0.095 to 0.125 in Cluster 2. This indicates that this method not only can be used to identify breakdown from non-breakdown conditions, but also can be used to identify the level of congestion in the breakdown phase.

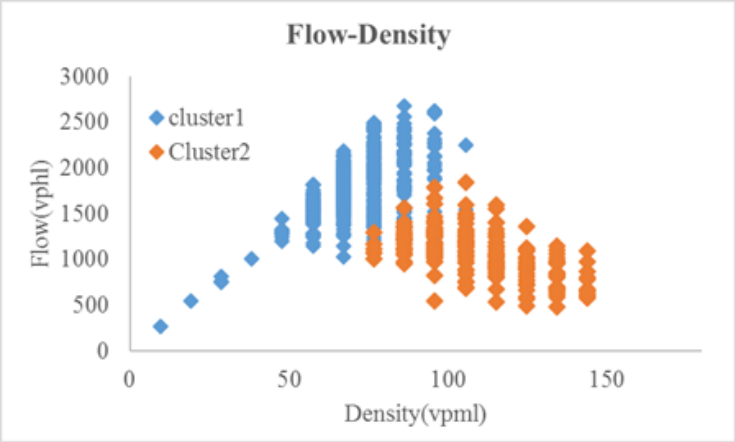
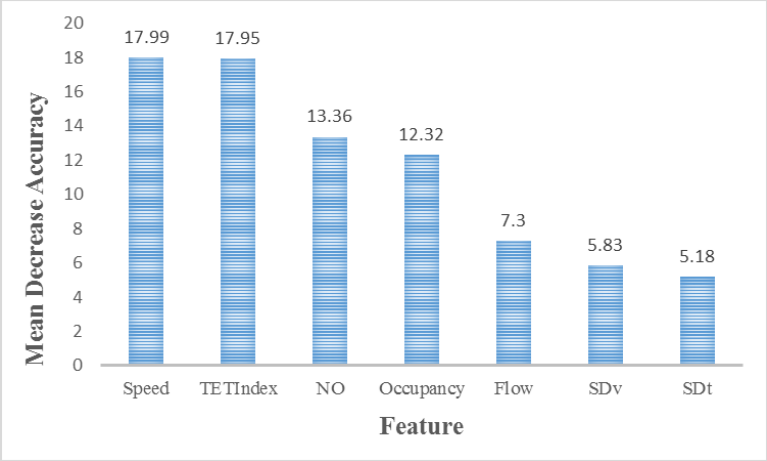


FIGURE 2-15: DATA CLUSTERS IN THE TEST SEGMENT OF THE NGSIM US101 DATASET

2.4.5 Traffic Flow Breakdown Prediction

As stated earlier, the results of clustering discussed in the previous section, was used as a binary label to build the breakdown prediction model utilizing three machine learning methods that can act as classifiers of traffic patterns in real-time operations. Before applying a model to predict breakdown, the RF approach was used to rank the importance of the variables. RF is a non-parametric method that is based on decision trees (Liu et al., 2016). The R package "randomforest' was used to identify the importance of the variables. The RF was grown by building 200 decision trees and by randomly selecting two predictor variables at each split since this number of variables results in the minimum Out Of Bag (OOB) error. The assessment of the results was made using a metric referred to as Mean Decrease Accuracy. The results are shown in Figure 2-16. A higher accuracy value in the figure represents a higher variable importance. As can be seen in Figure 4-16 (a), the four most important variables in breakdown prediction are the mean speed, the TETIndex, NO, and occupancy. The RF was also applied to the real-world CV dataset. The results shown in Figure 4-16 (b) show that the four most important features to predict breakdown are the same four features identified for the simulation data.

(a)



(b)

FIGURE 2-16: VARIABLE IMPORTANCE RANKING USING RANDOM FOREST TECHNIQUE (A) THE SIMULATION DATASET (B) THE REAL-WORLD CV DATASET

The next step is to use the three selected machine learning methods (the SVM, RF, and XGB) to develop models for real-time applications. The simulated data were split to training and testing datasets in 8:2 ratio with a random selection algorithm to make sure each dataset represent the maximum variance of data and to minimize having a biased dataset. For the real-world CV data, considering that the sample size is small, all data were used in the training and the k-fold cross validation was used to estimate the accuracy of the model with 5-folds.
The features were temporally lagged for two 5-minute time intervals to allow the prediction of the traffic state in the following 5-minute interval. This means that the feature values estimated for the past two five minutes are used as inputs to the machine learning. The output is the predicted binary label of breakdown/non-breakdown for the next five minutes. To assess the effect of utilizing the defined disturbance and safety metrics in clustering and classification machine learning approaches on the accuracy of state prediction, two scenarios were studied as follows.

Scenario (A) involved the use of macroscopic and microscopic features as inputs to the machine learning. The utilized metrics are the flow, mean speed, occupancy, SDv, SDt, NO and TETIndex. In addition, the binary label obtained from clustering using the microscopic and macroscopic features, as discussed earlier, was used as a label that was also used in the training.

Scenario (B) involved all features in Scenario (A) without the two disturbance and safety metrics. Thus, the utilized metrics are the flow, mean speed, occupancy, SDv, and SDt. In addition, instead of using the binary label resulting from clustering, the binary label Fine-tuning of the model parameters was done for each scenario and each method using the Grid Search, to tune the models by searching for the best hyper parameters and keeping the classifier with the highest accuracy. The final selected model parameters for each machine learning approach and each of the two scenarios (A and B) are reported in Table 2-15. In Table 2-15, the parameter C is the cost parameter of the error and shows the strength of the regularization. Gamma is a parameter for non-linear SVM. n_estimators is the number of trees used in each of the two ensembles (RF and XGB). Learning _rate controls is used in fixing the error from the previous iteration. max_depth controls the depth of each tree. max_feature is the size of the random subsets of features to consider when splitting a node.

TABLE 2-15: SUMMARY OF TUNED PARAMETERS WITH THREE SCENARIOS FOR EACH CLASSIFIER WITH TWO DATASET

Model	Scenarios	C	gamma	kernel	learning_rate	max_depth	max_features	n_estimators
Simulated Data								
SVM	A	1	NA	Linear	NA	NA	NA	NA
	B	10	0.1	RBF	NA	NA	NA	NA
RF	A	NA	NA	NA	NA	7	3	50
	B	NA	NA	NA	NA	7	2	150
XGB	A	NA	NA	NA	0.1	5	1	200
	B	NA	NA	NA	0.01	5	1	150
Real-World CV Data								
SVM	A	0.1	NA	Linear	NA	NA	NA	NA
	B	0.1	NA	Linear	NA	NA	NA	NA
RF	A	NA	NA	NA	NA	3	2	20
	B	NA	NA	NA	NA	3	1	20
XGB	A	NA	NA	NA	0.01	3	2	200
	B	NA	NA	NA	0.5	5	1	150

** RBF= Radial Basis Function kernel SVM*

The performance of each of the classifiers in the prediction in each scenario were assessed in terms of the overall accuracy, recall, precision, balanced accuracy, F1 Score, and confusion matrix, as reported in Table 2-16. In the confusion matrix, the rows are the predicted state and the columns are the actual state. The higher are the numbers on the diagonal of the matrix, the

more accurate is the estimation. The other goodness of the models used in this study are defined below (Bishop, 2006).

The accuracy score is the number of the correct predictions made by the model. It indicates the overall performance of the model showing the fraction of the correct predictions over $n_{samples}$ and is computed as:

$$Accuracy\,(y,\hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1(\hat{y}_i = y_i) \quad \textbf{2-20}$$

Precision, computed as below, is a measure of how accurate the positive predictions are. A high precision index indicates that most of the exampled labeled as positive are actually positive.

$$Precision = \frac{TP}{TP+FP} \quad \textbf{2-21}$$

Recall refers to the coverage of the actual positive sample. In other words, a high recall index indicates that a class is correctly recognized.

$$Recall = \frac{TP}{TP+FN} \quad \textbf{2-22}$$

In the above equations, TP are the true positives, TN are the true negatives, FP are the false positives (Type I error), and FN are the false negatives (Type II error).

The F1 score is a hybrid metric useful for unbalanced classes. The F1 score is computed as the harmonic mean of the precision and recall indices. It complements the precision index and is especially useful when uneven class distribution is present.

$$F1\ score = 2*((\text{PRECISION}*\text{RECALL})/(\text{PRECISION}+\text{RECALL})) \quad \textbf{2-23}$$

When comparing the results of the evaluations in Table 2-16, it can be seen that Scenario (A) produced better results than Scenario (B) when used with all machine learning methods, confirming that the state estimation based on disturbance metrics combined with macroscopic metrics produce better results. All three investigated machine learning methods produced good accuracy with the RF approach producing somewhat better results than the other two methods.

The above results show the accuracy metrics for the overall prediction accuracy. However, this study also assessed the accuracy for a particular condition that is of specific importance to traffic management. This measure assesses the accuracy of the prediction of the occurrence of breakdown in the next five minutes when the states of the two previous five minutes are non-breakdown. This is very important to allow the activation of new management plans to address the breakdown before it occurs. Table 2-16 shows the accuracy of this prediction as the "% Error in Predicting Transition to Congestion." Again, Scenario (A) produced better results with this metric compared to Scenario (B) for all investigated conditions. Also, the RF method appears from the results in Table 2-16 to produce somewhat better results than the other methods.

TABLE 2-16: PERFORMANCE OF THE THREE CLASSIFIERS IN STATE PREDICTION BASED ON SIMULATION DATA

Model	Scen-arios	Overall Accuracy	Pre-cision	Recall	balanced accuracy	F1	Confusion Matrix	% Error in Predicting Transition to Congestion
SVM	A	0.939	0.939	0.939	0.922	0.939	[130 4] [8 56]	7
	B	0.924	0.924	0.924	0.906	0.924	[126 5] [10 57]	14
RF	A	**0.954**	0.954	0.954	0.941	0.954	[131 3] [6 58]	**5**
	B	0.934	0.935	0.934	0.913	0.940	[128 3] [10 57]	9
XGB	A	0.949	0.950	0.949	0.930	0.949	[132 2] [8 56]	7
	B	0.924	0.924	0.924	0.901	0.924	[125 3] [12 58]	19

**Note that the whole simulated dataset was 2968 data points where 20% was taken as test set with random selection. As the study considered 3 sequences of 3 time-interval of each 5 minutes, so the predicted size is about 198.*

The method and scenarios tested using simulation trajectories as described earlier were also tested using CV-data from a low market penetration deployment. As mentioned earlier, the evaluation in this study also utilizes CV data that was extracted to evaluate the performance of the application of the model for a freeway segment. The data indicates that the speed on the segment varied between 45 ft/s and 120 ft/s, and the peak period was between 3:00 P.M. and 5:00 P.M. The results from applying the three machine learning techniques with Scenario (A) and Scenario (B) were evaluated using the 5-Fold cross validation. Table 2-17 shows the results from the evaluation. As with the simulated data, Table 2-17 shows that Scenario "A" had higher accuracy compared to Scenario 'B" indicating the benefit of using the microscopic metrics. However, the machine learning algorithm that performed the best in this case was the XGB model.

TABLE 2-17: PERFORMANCE OF THE THREE CLASSIFIERS IN STATE PREDICTION BASED ON REAL-WORLD CV DATA

Model	Scenarios	Overall Accuracy (SD)
SVM	A	0.94 (+/- 0.15)
	B	0.91(+/- 0.15)
RF	A	0.91 (+/- 0.14)
	B	0.91(+/- 0.15)
XGB	A	**0.95 (+/- 0.13)**
	B	0.90(+/- 0.12)

**SD=Standard deviation*

2.5 CONCLUSION

This study proposed a methodology for the estimation and prediction of new measures based on CV data for potential use in off-line planning and in real-time management of traffic operations and safety. These measures include platooning measures, traffic disturbance measures, and safety surrogate measures. The percentage of platooned vehicles and the distribution of the platoon size were estimated based on surrogate measures that can be assessed using CV data at relatively low market penetrations of connected vehicles. The utilized measures are SDv and SDt. Relationships between the surrogate measures and the platooning measures were identified and utilized based on available trajectories data for different market penetrations of CV. The results show that the percentage of vehicles in the platoon can be accurately and reliably estimated at relatively low CV market penetrations. For the platoon size distribution estimation, a low market penetration of 5% is adequate when using the data for planning purposes based on multiple days. However, a minimum of 20% market penetration of CV is needed to estimate the platoon distribution for individual day operations.

This study also defined disturbance metrics and examined utilizing them as traffic safety and stability indicators for potential use in real-time operations. These disturbances metrics are the number of oscillations (NO) and a measure of disturbance durations index in terms of the time exposed time-to-collisions (TETIndex). The TET parameter has been used in the past as a safety surrogate measure. This study introduced its use for the first time as an indicator of traffic breakdown analysis and traffic safety analysis. Since TETIndex estimation cannot be measured at low market penetration of CV data, a regression model was derived based on speed parameters to estimate this parameter. Statistical testing of the model and associated parameters indicate that the model is significant and has a mean absolute percent error of 14%. Then, the developed regression model was further validated using real-world trajectory data collected by the NGSIM program from two locations that were not used in the calibration of the simulation model. The results showed that the TETIndex can be estimated with low samples of trajectory data (e.g., data from low market penetration of trajectory connected vehicles) based on speed parameters with an error of around 15%-20%.

The application of the model to estimate safety risk utilizing trajectory data from a real-world deployment with low market penetration of CV data showed that including the utilized disturbance metrics allow better recognition of the crash risk. The study also confirmed that a TETIndex of 0.03 with NO of 10% can be used as thresholds above which the probability of perturbation growth and crash occurrence increase. This study found that that the investigated disturbance metrics can be use as indicators of unsafe conditions as part of decision support tools that include the activation of transportation management strategies to reduce the probability of unsafe traffic and ease traffic disturbances that have adverse impacts on traffic safety.

This study also utilized the defined microscopic disturbance and surrogate safety measures and examined the benefit of utilizing them in traffic state classification and prediction in combination with macroscopic traffic parameters. The combined macroscopic and microscopic measures are the TETIndex, NO, SD, SDt, mean speed, traffic flow rate, and occupancy. The measures were used as inputs to a hybrid unsupervised clustering and three different supervised classifiers (SVM, RF, and XGB). The results indicate that the utilization of macroscopic measures by themselves in the traffic state estimation creates uncertainty with regard to traffic performance based on microscopic characteristics. This uncertainty covers in a relatively wide range of speed around the transition from the uncongested to the congested traffic conditions. The results of the evaluation performed in this study indicate that the combination of features that produced the best categorization of traffic state using clustering are the NO, TETIndex, average speed, and occupancy. The clustering results were compared to those obtained using a deterministic value of speed at capacity, derived using change-point regression and the results from clustering based only on speed and occupancy. It was also found that a TETIndex greater than 0.05 And NO greater than 20% can be used as criteria in the breakdown identification. The method was tested utilizing real-world NGSIM dataset. It was concluded that the proposed method of using the traffic disturbance parameters can also be used to categorize different levels of congestion.

The utilization of TETIndex and NO as disturbance metrics in combination with other metrics also increases the accuracy of traffic state prediction based on the results from the application of three supervised machine learning classifiers from both simulated trajectory dataset and real-world CV dataset. All three machine learning approaches investigated in this study (the SVM and two tree ensembles) performed well with slight variations in performance, depending on the specific case study data used in the investigation. It can be concluded that the investigated disturbance and surrogate metrics can be used as inputs to machine learning to predict traffic flow breakdown in terms of mobility and safety in real-time traffic operations. Such use is recommended as part of decision support tools that recommend the activation of transportation management strategies to reduce the probability of traffic breakdown and ease traffic disturbances.

3. FRAMEWORK FOR UTILIZING CV DATA FOR PLANNING AND OPERATION APPLICATIONS

3.1 INTRODUCTION

Performance measurement can play a vital role in decision making at both federal and state levels. Proper transportation measurement and management process helps to enhance transportation system planning and operations. Estimated performance measures can be used by a system operator or planner in order to support decisions associated with these processes. Such measurements can also be used to derive information for dissemination to travelers, third-party data aggregators, traveler information service providers, and other agencies. Performance measures can be either quantitative or qualitative. Examples of quantitative performance measures include volume, density, travel time, speed, queue length, and emissions. Qualitative performance measures include user satisfaction, driver compliance, and driver frustration. Performance measures can be estimated based on existing technologies such as traffic surveillance involving closed-circuit television (CCTV), machine vision equipment, and sensors including subsurface induction loop, acoustic, and radio frequency (RF).

Connected vehicle (CV) technologies promise to allow the estimation of performance measures currently provided by other technologies, as well as measures that cannot be collected by existing sensor technologies. Examples of additional performance measures that can be estimated from data obtained through CVs include stops, accelerations, and decelerations, shockwave speed, detailed signalized intersection movement-level measures, and the potential for crashes, to name a few.

Studies confirm that a relatively low market penetration of CV may be required for estimating some performance measures, while other measures will require high market penetrations to produce accurate results (Iqbal et al., 2018; Khan et al., 2017; Khazraeian et al., 2017). The availability of CV data, even at small percentages, may be sufficient to support critical transportation management functions. For example, such data can be beneficial in identifying abnormalities in data detection and processing issues associated with existing technologies.

3.1.1 Performance Measurement Definition

Performance measurement can be defined in various ways. Generally, it is an approach to evaluate the efficiency and effectiveness of the system. A comprehensive definition of performance measurement, based on a national performance review, is offered by the US Federal Highway Administration (FHWA). Accordingly “performance measurement is a process of assessing progress toward achieving predetermined goals, including information on the efficiency with which resources are transformed into goods and services (outputs), the quality of those outputs (how well they are delivered to clients and the extent to which clients are satisfied) and outcomes (the results of a program activity compared to its intended purpose), and the effectiveness of government operations in terms of their specific contributions to program objectives.” (NCHRP, 2003). Moreover, FHWA defines Transportation Performance

Management (TPM) as a "strategic approach that uses system information to make investment and policy decisions to achieve national performance goals." (FHWA, 2017a).

Performance measurement is an important element of congestion management. In 2014, congestion on the US transportation system resulted in 6.9 billion hours delay with an extra 3.1 billion gallons of fuel at a total cost of $160 billion (Schrank et al., 2015). Continuous monitoring of congestion presence and associated impacts using performance measures can provide valuable information about the location, type, severity, and extent of congestion over space and time, which will help to take necessary actions in order to address these problems.

3.1.2 Connected Vehicles (CVs)

CVs are one of the latest innovations in the era of road transportation systems and an indispensable part of Intelligent Transportation Systems (ITS). CVs use sensors and various communication technologies such as vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications to interact with each other. Such technologies allow information exchange, cooperative localization and map updating, and facilitate cooperative maneuvers between vehicles within a short range of each other.

CVs are proactive, well cooperative and coordinated and can provide a 360-degree awareness that informs a driver of hazards and situations that they cannot see as well as reduce or eliminate crashes through a combination of driver advisories, driver warnings, and vehicle controls. For example, the Intersection Movement Assist application warns drivers when it is unsafe to enter an intersection. The Do Not Pass application warns drivers when it is not safe to pass a slower moving vehicle while the Emergency Electronic Brake Light application notifies the driver when an out-of-sight vehicle located downstream is breaking.

The major benefit of vehicle connectivity is that information can be shared in an efficient and reliable manner and can be used to support drivers' decision-making process. This, in turn, can improve traffic safety, operational efficiency, road utilization, and user satisfaction. Moreover, the autonomous vehicle capability is increasingly built into vehicles in the form of self-parking, adaptive cruise control, and assisted-braking with a potential for full automation. In addition to supporting driver decision making, CVs generate the full set of vehicle data that can be utilized to support and enhance performance monitoring. CV data include brake status, turn signal status, vehicle length, vehicle width, and bumper height, as well as time, heading angle, lateral acceleration, longitudinal acceleration, yaw rate, throttle position, steering angle, headlight status, wiper status, external temperature, and vehicle mass.

3.1.3 Problem Statement

It has been already established that performance measurement is an indispensable part of effective transportation systems management. Thus, performance measurement is a topic of huge interest both internationally and in the US and can play a vital role in decision making at both federal and state levels. Proper transportation measurement and management processes

help to enhance transportation systems internal operations. The recent introduction of CV technologies created new opportunities for use of CV data for performance measurements. A review of the literature confirms that existing studies are limited in scope as they mainly focused on the generation of specific measures such as travel time, density and queue length estimation by using CV data from different sources including the Next Generation Simulation (NGSIM) data set (Argote et al., 2012; Hao et al., 2014; Nam et al., 2017; Qiu et al., 2010) and safety pilot model deployment (SPMD) data set (Khattak et al., 2017; Jun Liu et al., 2016; Mousa et al., 2017; Zheng et al., 2017).

Moreover, existing studies did not develop a list of performance measures that can be calculated based on emerging CV data nor have they considered data that can be collected in accordance with the Society of Automotive Engineer (SAE) message sets and the expected availability of the data. It should be mentioned that, according to CV data standards, no permanent vehicle identifications are assigned to any vehicles. In addition, the earlier studies also have not addressed new performance measurements related to national highway system performance, freight movement on the interstate, Congestion Mitigation and Air Quality (CMAQ) program – traffic congestion, and CMAQ – on-road mobile source emissions that were recently established by FHWA in the Moving Ahead for Progress in the 21st Century (MAP-21) in 2012 (FHWA 2017b).

3.1.4 Objectives

The objective of this study is to investigate the use of data collected from CVs, alone or in combination with data from other sources, to support transportation system performance measurement for transportation planning and operation purposes. Novel performance measures are also developed, considering the availability of emerging vehicle technologies through Dedicated Short Range Communications (DSRC) and/or wide area cellular technology. To achieve the study objectives, the following tasks are performed:

- Development of a methodological framework to estimate system performance measurements using CV data in support of transportation operations, management, and planning. The framework describes data requirements, performance metrics, and performance measures as well as which types of performance measures can be calculated using CV data.
- Proof of concept as part of framework validation where performance measurements are compared using traditional and CV data. To obtain the necessary CV data, the study developed a simulation model of a study segment of an interstate in the Birmingham region using the microscopic simulation software VISSIM. The simulation test bed is used to generate BSM data using the trajectory conversion analysis (TCA) tool. These, in turn, are used to generate performance measurements in accordance with the proposed framework.

3.2 LITERATURE REVIEW

The primary purpose of this chapter is to summarize past studies related to transportation performance measurement using CV data, discuss the data standard of CVs, identify traditional and CV-related data sources for performance measurement, and discuss existing performance measurement techniques.

3.2.1 Existing Peer-Reviewed Studies

The literature review identified a number of earlier studies that used CV data (obtained from field tests or generated via simulation) to determine travel time, density, volume, queue length, and emissions. However, there has not been a good synthesis of these studies. To address this issue, existing studies are summarized below, and contributions and limitations are highlighted where possible.

There are several studies in the literature focusing on travel time estimation using CV data. Mousa et al. (2017) estimated travel time based on basic safety message (BSM) info and found that the mean absolute errors are 13 and 20 seconds for 5- and 20-minute horizons respectively. These estimates were based on eXtreme Gradient Boosting (XGB) algorithm and the BSM data came from a Safety Pilot Model Deployment conducted in Ann Arbor, Michigan. However, this study did not account for different market penetration rates, which is an important consideration. Zou et al. (2010) estimated travel time based on Vehicle Infrastructure Integration Probe Data (VIIPD) messages according to J2735 standards and found average travel time error percentages of 27.6%, 12.5%, and 8.2% for 1%, 5%, and 10% market penetrations, respectively. These estimates were based on traffic simulations of a hypothetical network. Izadpanah et al. (2011) conducted a study to determine travel time using vehicle trajectory data from GPS data loggers on a freeway segment. The results showed that the measured and ground truth travel time had no significant difference. In another study, Argote et al. (2012) estimated some common arterial measures of effectiveness including average speed, average delay per unit distance, average number of stops, average acceleration noise, and queue length based on CV data obtained from Next Generation Simulation (NGSIM) data. A drawback of this study is that it uses the vehicle ID but does not consider the change of vehicle ID during its course of travel, as specified in the J2735 standards.

Studies focusing on density estimation based on CV data report that high market penetration of CVs is required in order to get accurate results. A study by Khan (2015) based on simulation modeling, showed that the use of CV data as input into an advanced estimation algorithm can provide an accuracy of at least 85% when the CV penetration level was 50% or more, with the estimation accuracy increasing with the increase in the market penetration. The same study reported that density estimations that used an algorithm based on point detector data resulted in an accuracy rate between 42.5% and 62.2%. An incremental benefit-cost analysis indicated that the use of CV provides a higher return on investment, compared to the use of loop detectors. However, the study did not assess the accuracy of CV data utilization for market penetrations below the 50% market penetration level. In another study, Khan et al. (2017)

assessed the accuracy of CV data utilization for market penetrations below the 50% market penetration level and found that 20% or more CV penetration level can provide 85% accuracy. Nam et al. (2017) also conducted a study to estimate density using probe vehicle data of Next Generation and Simulation (NGSIM) dataset and found that estimated densities reflect ground truth density and accuracy of density increases with the increase of penetration rates. A number of studies examined the potential of using a low sample size of probe vehicles in combination with point detector data to improve density estimation accuracy. Al-Sobky et al. (2016) conducted a study to determine traffic density using two smartphones inside two vehicles and an observer to obtain count data. The results showed that measured density is close to the actual density at the 5% significance level. The error of the density estimated using this method ranges from 1.3% to 15%, with an average of 8%. However, this proposed system is not applicable for a high percentage of heavy vehicles and uninterrupted flow condition. In another study, Qiu et al. (2010) combined detector data with probe data to estimate density and found that the relative error for the given periods can be improved from 30% based on point sensor data, and to 4% to 6% based on point sensor data plus probe vehicle data. They used two loop detectors placed 1,000 feet apart, with two probe vehicles driven five round trips along the section. Once again, this indicates the potential of using CV data in combination with point detection to estimate density at low market penetrations of CV.

Zheng et al. (2017) estimated traffic volumes and found mean absolute percentage error (MAPE) of traffic volumes in the rage of 9-12%. This estimate was based on low market penetration rates ranging from 3 to 12%. In their study they used two sources of connected vehicle (CV) data, namely the Safety Pilot Model Deployment (SPMD) project in the city of Ann Arbor, MI and vehicle trajectory data in China.

Queue length estimation using CV data also requires high market penetration of CVs, as reported in the literature. J.-Q. Li et al. (2013) combined probe trajectory and signal timing data to estimate the queue length and found that the mean absolute percentage error decreased with the increase in the market penetration. This estimate was based on microscopic simulation data. Osman et al. (2016) investigated cycle-by-cycle queue length using Basic Safety Messages (BSMs) based on shockwave analysis and found estimation errors to be between 0 and 33%. However, a study conducted by Khazraeian et al. (2017) indicated that a relatively low market penetration (around 3% to 6%) for a congested freeway is sufficient for accurate and reliable estimation of the queue length. Even at 3% market penetration, the CV-based estimation of the back of queue identification is significantly more accurate than that based on detector measurements. It was also found that CV data allows faster detection of the bottleneck and queue formation.

Recent studies indicate that incident detection and collision warning can be estimated using CV data. Wolfgram et al. (2018) detected the occurrence of incidents quickly and reliably using CV data and found that availability of CV data can reduce the detection time, from minutes to just seconds. This estimate was based on empirical and simulation data. Tajalli et al. (2018) investigated the vehicle collision problem at a signalized intersection using CV data and technologies based on simulation data under three sets of scenarios including various volumes

of vehicles, compliance rates, and CV penetration rates. Analysis of results showed that the number of V2V conflicts decreased from 24 to 16, to 5 and vehicle to pedestrian (V2P) conflicts decreases from 56 to 33, to 0 for increasing market penetration rates from 0% to 50%, to 100% respectively.

Work zone safety can also benefit from CV technologies, even at a low market penetration rate. A study conducted by Genders et al. (2016) showed that market penetration rates lower than 40% increase the safety of the traffic network, meanwhile, MPR more than 40% decreases the safety of the network. Authors also mentioned that work zone information through CV technologies helps to modify driving behavior and decay travel time.

Several studies examined eco-driving such as maintaining speed, acceleration, checking proper tire pressure etc. and concluded that it is a great option for reducing CO2 in the environment. Barth et al. (2009) measured the effectiveness of eco-driving based on simulation and real-world experiments and found that 10-20% CO2 emission and fuel consumption can be easily reduced. They used three types of traffic data sets, namely speed, flow, and density. The data were collected from the California PeMS system under actual traffic conditions. In another study, Rakha et al. (2011) estimated fuel consumption and instantaneous speeds for eco-driving based on SPaT information from CV communication technologies and found that most fuel optimal speed profiles can be easily identified. These estimates were based on two models, namely a VT-Micro model and a Vehicle Dynamics model.

3.2.2 Connected Vehicle (CV) Data Standard

CVs transmit information through various wireless communication technologies, sensors and other in-vehicle technologies. Dedicated short range communication (DSRC) is a reliable technology that is used frequently to transmit information between vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I). The Society of Automotive Engineers (SAE) established Dedicated Short Range Communication (DSRC) Message set Dictionary standard. Data attributes of different types of DSRC messages according to SAE (2016) standard J2735 are summarized in the following paragraphs.

3.2.2.1 Basic Safety Message (BSM)

Basic safety message is used to exchange information of safety data related to vehicle movement. The data transmission rate is 10 times per second. It consists of two parts of data, namely Part-I (BSM core data) and Part-II (optional) (SAE, 2016). Some data attributes of Part-I are vehicle temporary ID, latitude, longitude, elevation, positional accuracy, transmission state, speed, heading, steering wheel angle, acceleration set 4way, brake system status, and vehicle size. Data attributes of Part-II include vehicle safety extensions such as vehicle events flags, path history, path direction, exterior lights; special vehicle extensions such as emergency details, event description, trailer data, trailer unit description list; supplemental vehicle extensions such as vehicle type classification data, various V2V probe data, detected obstacle data etc. The detailed description of data attributes of BSM data Part-I and Part-II are provided in the Appendix A and Appendix B, respectively.

3.2.2.2 Signal Phase and Timing (SPAT) Message

The SPAT message transmits the current signal status, remaining phase time, and the next phase status of the intersection to the vehicle. Data attributes of SPAT message include time stamp (minute of the year); intersection state list such as intersection reference ID, intersection status object, lane ID, movement list, maneuver assist list etc. The detailed description of data attributes of SPAT message is available in the Appendix C.

3.2.2.3. Map Data (MAP) Message

MAP message represents geometric information of intersections as well as segments of roadway. Some data attributes of MAP message are time stamp (minute of the year), layer type, layer ID, intersection geometry list, road segment list, data parameters, restriction class list. The detailed description of data attributes of MAP message is provided in the Appendix D.

3.2.3 Data Sources

Traditional and CV data sources currently exist that can provide reliable data for estimation of various performance measures. The most commonly used data sources are summarized in the following subsections.

3.2.3.1 Traditional Data Sources

Traditional databases include the Highway Performance Monitoring System (HPMS), the National Performance Management Research Data Set (NPMRDS), the CMAQ Public Access System, the U.S. Energy Information Agency (EIA), and the National Household Travel Survey (NHTS). A brief description of existing data sources and types of data that they provide for performance measurement are summarized in Table 3-1.

TABLE 3-1: BRIEF DESCRIPTION OF EXISTING DATA SET

Name of data set	Types of data	Frequency
Highway Performance Monitoring System (HPMS)	Distance, Geographic Location, Jurisdiction, Types of areas (Rural, Small Urban, Urbanized), Average Annual Daily Traffic (AADT), Through lane numbers, Serviceability ratings, and others	Annually
National Performance Management Research Data Set (NPMRDS)	TMC code, measurement_tstamp, Speed, Average speed, Reference speed, Travel time, Data density, Direction, Geographic location, etc.	Monthly
CMAQ Public Access System	On-road mobile source emissions data	Annually
U.S. Energy Information Agency (EIA)	Emission conversion factors	
National Household Travel Survey (NHTS)	Number of Households, number of vehicles, number of people, daily travel information for all modes of transportation, vehicle miles of travel (VMT), average vehicle occupancy.	Periodically

3.2.3.2. Connected Vehicle (CV) Data Sources

The Intelligent Transportation Systems (ITS) Joint Program Office (JPO) was established by U.S. Department of Transportation (USDOT) to provide data from different sources to support researchers, and developers. Currently, the ITS JPO website (USDOT (U.S. Department of Transportation), 2018) includes the following field data sets of CV, mentioned in Table 3-2.

TABLE 3-2: BRIEF DESCRIPTION OF CV FIELD DATA SET (USDOT 2018)

Name of Data set	Types of Message	Time period	Location	Sample size
Advanced Messaging Concept Development (AMCD)	BSM		Virginia Smart Road in Blacksburg, Virginia; Fairfax County, Virginia	12 vehicles
Multi-Modal Intelligent Traffic Signal Systems (MMITSS)	BSM, MAP, SPAT	03/02/2015 to 03/04/2015	Anthem, Arizona	16 vehicles
Wyoming Connected Vehicle (CV) Pilot project	BSM		I-80 in Wyoming's southern border	400 vehicles
Safety Pilot Model Deployment (SPMD)	BSM	10/012012 to 04/30/2013	Ann Arbor, Michigan	Approximately 3,000 vehicles
Intelligent Network Flow Optimization (INFLO) data	BSM	01/12/2015 to 01/16/2015	I-5 in Seattle, WA	21 vehicles
2014 ITS World Congress Connected Vehicle Test	BSM, MAP, SPAT	09/082014 to 09/10/2014	Detroit, MI	9 vehicles
Southeast Michigan Operational Data Environment (SEMI-ODE)	BSM, MAP, SPAT	04/05/2016 to 04/07/2016	Southeast Michigan	
National Center for Atmospheric Research (2009)	BSM	04/06/2009 to 04/22/2009	Michigan Test Bed	27 vehicles

3.2.4 Performance Measurement Techniques

The Transportation Performance measurement techniques can be classified into two groups, namely established performance measurement and emerging performance measurement. Details for each are offered in the subsections below.

3.2.4.1 Established Performance Measurements

The FHWA established four types of new performance measurements in the Moving Ahead for Progress in the 21st Century (MAP-21). According to (FHWA 2017), these performance measurements are:

I. national highway system performance,
II. freight movement on the interstate,
III. CMAQ program - traffic congestion, and
IV. CMAQ - on-road mobile source emissions.

Figure 3-1 presents data requirements, performance metrics, and performance measures for each type in a systematic way for easy reference.

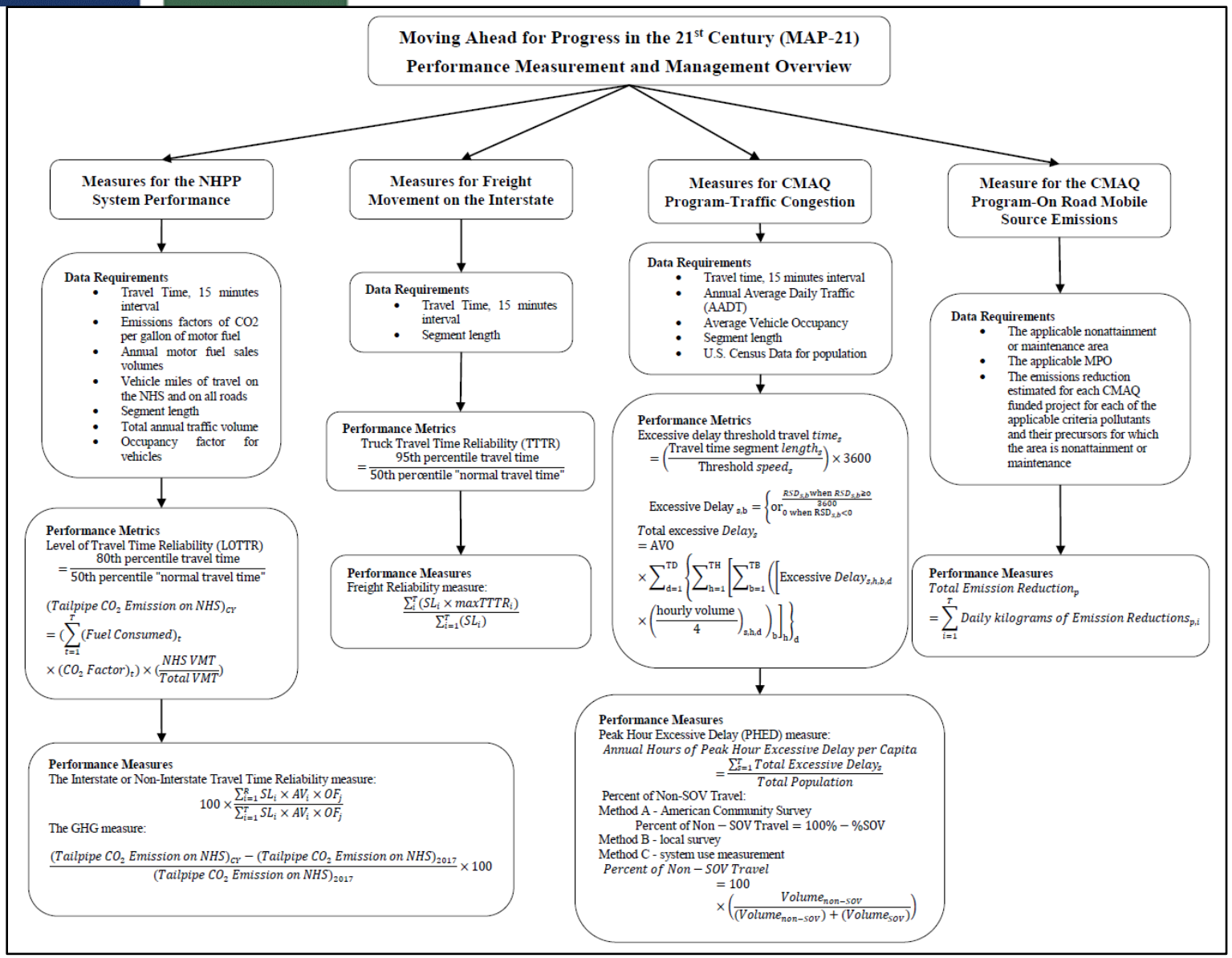


FIGURE 3-1: OVERVIEW OF MAP-21 PERFORMANCE MEASUREMENT AND MANAGEMENT

3.2.4.2 Emerging Performance Measurements

Emerging performance measurement systems can be classified into two categories: i. Operational performance measurement, and ii. Planning performance measurement. Operational and planning performance measurements are discussed in more detail in the following paragraphs.

3.2.4.2.1 OPERATIONAL PERFORMANCE MEASUREMENTS

According to NCHRP (2003) report, commonly used performance measures for operational effectiveness of highway systems and segments are summarized in Table 3-3.

TABLE 3-3: OPERATIONAL PERFORMANCE MEASURES (NCHRP, 2003)

Performance Measures	Typical Definition
Commercial vehicle safety violations	Number of violations issued by law enforcement based on vehicle weight, size, or safety
Delay caused by incidents	Increase in travel time caused by incidents
Density	Passenger cars per hour per lane
Duration of congestion	Period of congestion
Evacuation clearance time	Reaction and travel time for evacuees to leave area at risk
Incidents	Traffic interruption caused by a crash or another unscheduled event
Rail crossing incidents	Traffic crashes that occur at highway-rail grade crossings
Recurring delay	Travel time increases from congestion but does not consider incidents
Response time to weather-related incidents	Period required for an incident to be identified and verified and for an appropriate action to alleviate the interruption to traffic to arrive at the scene
Roadway congestion index	Cars per road space
Security for highway and transit	Number of violations issued by law enforcement for acts of violence against travelers
Speed	Distance divided by travel time
Toll revenue	Dollars generated from tolls
Traffic volume	Annual average daily traffic, peak-hour traffic, or peak-period traffic
Travel costs	Value of driver's time during a trip and any expenses incurred during the trip (vehicle ownership and operating expenses, tolls, or tariffs)
Travel time	Distance divided by speed
Vehicle occupancy	Persons per vehicle
Weather-related traffic incidents	Traffic interruptions caused by inclement weather

3.2.4.2.2 PLANNING PERFORMANCE MEASUREMENTS

Planning performance measures based on NCHRP (2003), NCHRP (2008) and AASHTO (2012) reports are summarized and presented below in Table 3-4.

TABLE 3-4: PLANNING PERFORMANCE MEASURES

Measure	Calculation Method
Roadway congestion index	$\dfrac{\frac{VMT_{Freeway}}{Lane-mile_{Freeway}} \times VMT_{Freeway} + \frac{VMT_{Arterial}}{Lane-mile_{Arterial}} \times VMT_{Arterial}}{13{,}000 \times VMT_{Freeway} + 5{,}000 \times VMT_{Arterial}}$
Travel rate index	$\dfrac{\frac{60}{Speed_{Freeway}}}{\frac{60}{Freeflowspeed_{Freeway}}} \times VMT_{Freeway} + \dfrac{\frac{60}{Speed_{Arterial}}}{\frac{60}{Freeflowspeed_{Arterial}}} \times VMT_{Arterial}}{VMT_{Freeway} + VMT_{Arterial}}$
Delay per eligible driver	Total delay (includes recurring and incident delay) per eligible driver
Delay per capita	Total delay (includes recurring and incident delay) per person
Wasted fuel per eligible driver	Difference between fuel consumption in existing conditions and fuel consumption based on free flow speeds per driver
Wasted fuel per capita	Difference between fuel consumption in existing conditions and fuel consumption based on free flow speeds per driver
Congestion cost per eligible driver	Costs in dollars of congestion based on comparison of existing conditions and free-flow conditions per eligible driver
Congestion cost per capita	Costs of congestion based on comparison of existing conditions and free-flow conditions per eligible driver
Annual person-hours of delay	$Daily\ vehicle\ hours\ of\ delay \times 250\ working\ days\ per\ year \times 1.25\ persons\ per\ vehicle$
Percent congested travel	$\dfrac{VMT\ under\ congested\ conditions}{Total\ VMT\ for\ the\ area}$
Travel rate index	$\dfrac{\text{Travel time under congested conditions}}{\text{Travel time under uncongested conditions}}$
Travel time percent variation	$\dfrac{\text{Standard deviation}}{\text{Average travel time}} \times 100\%$

Measure	Calculation Method
Travel time buffer index	$\frac{95\%\ confidence\ travel\ rate_{(in\ minutes\ per\ mile)} - Average\ travel\ rate_{(in\ minutes\ per\ mile)}}{Average\ travel\ rate_{(in\ minutes\ per\ mile)}} \times 100\%$
Travel time misery index	Average of the travel rates for the longest 20% of the trips − Average travel rates for all trips
Planning Time Index	$\frac{95th\ percentile\ travel\ time\ (minutes)}{FFS\ or\ PSL\ travel\ time\ (minutes)}$
Travel Time Index	$\frac{Actual\ travel\ rate\ (minutes\ per\ mile)}{FFS\ or\ PSL\ travel\ rate\ (minutes\ per\ mile)}$
Congested Roadway (miles)	$\sum Congested\ segment\ lengths\ (miles)$
Annual Hours of Truck Delay (AHTD)	$\sum \left(\frac{Freight\ VMT}{Travel\ speed} - \frac{Freight\ VMT}{Agency\ specified\ threshold\ speed}\right) \times \times 7 \times 52$
Freight Reliability Index	$\frac{80th\ percentile\ travel\ time}{Agency\ travel\ time}$

3.3 TRANSPORTATION PERFORMANCE MEASUREMENT FRAMEWORK DEVELOPMENT

Development of a framework for transportation performance measurement system can play a vital role toward the improvement of the effectiveness and efficiency of the performance measurement analysis. The framework can be either conceptual, theoretical or methodological. These frameworks provide a structured hierarchy of procedures and processes to guide transportation authorities, engineers, planners, and agencies.

3.3.1 Existing Peer-Reviewed Studies

Most of the existing studies mainly focused on three types of the frameworks, namely conceptual, theoretical, and methodological. Sometimes, these frameworks can be used interchangeably (Tamene, 2016). The conceptual framework is something new that one can develop based on own concepts, whereas the theoretical and methodological frameworks are based on an existing theory or set of theories (Tamene, 2016).

3.3.1.1 Conceptual Framework

This framework is very common in several fields of studies, including those that are related to system performance measurement and management. Chowdhury et al. (2015) developed a forensic framework to evaluate severity condition of pavement distress based on pavement condition, historical data, and field core samples (Chowdhury et al., 2015). Griffis et al. (2013) proposed a performance-based framework for wind engineering (Griffis et al., 2013) where the authors considered the inelastic behavior of buildings under wind loading. In another study, Nasution et al. (2017) developed a conceptual framework to support policymakers to create public policy based on two approaches, big data and system dynamics (Nasution et al., 2017). A comprehensive conceptual framework to analyze construction performance was proposed by

Maloney (1990) and is shown in Figure 3-2. Moreover, several studies also addressed conceptual framework. Table 3-5 summarizes some of the studies of the various fields as a reference and to provide insights about the use of conceptual frameworks in engineering and other applications.

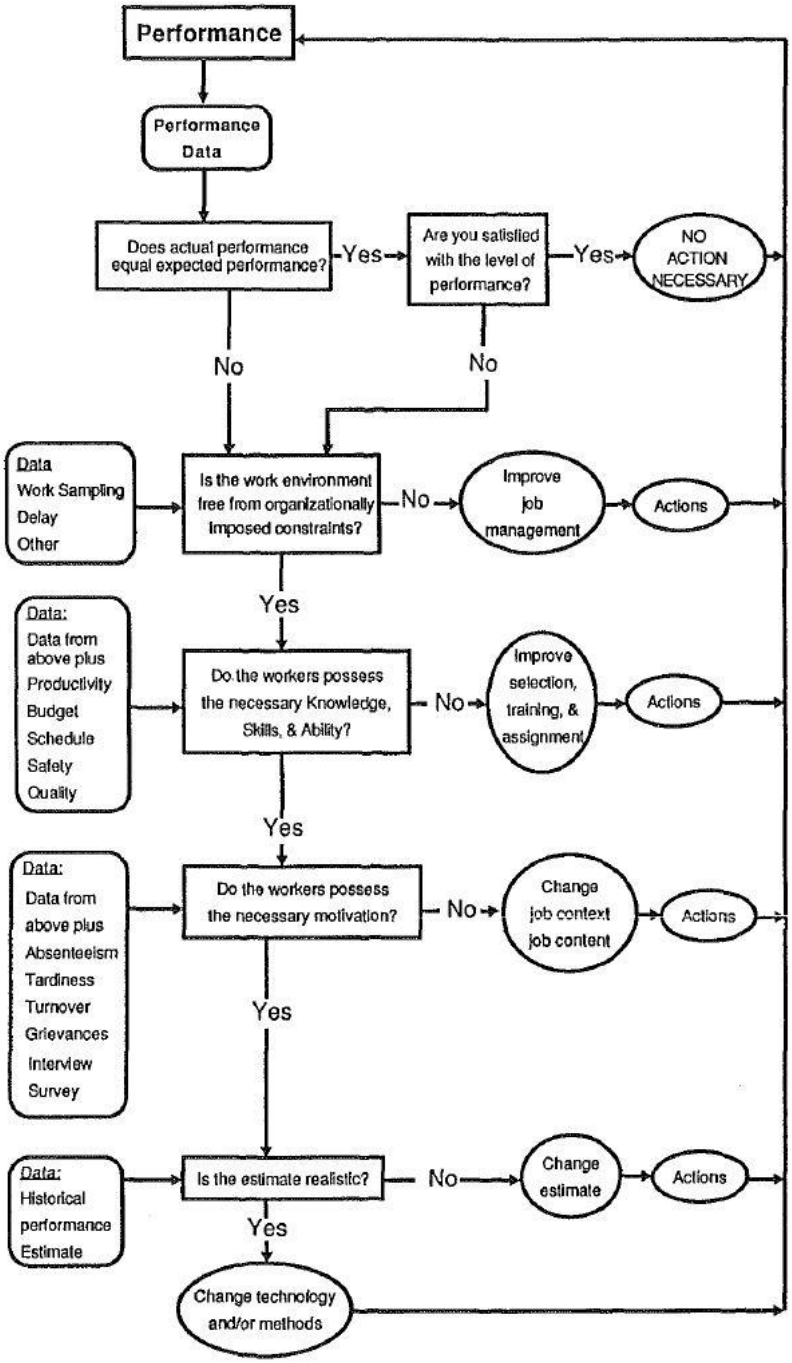


FIGURE 3-2: A FRAMEWORK OF CONSTRUCTION PERFORMANCE ANALYSIS. ADAPTED FROM (MALONEY, 1990)

TABLE 3-5: APPLICATION OF THE CONCEPTUAL FRAMEWORK

Authors	Application
Fwa et al. (1997)	Pavement friction management of airport runways
Gökalp et al. (2016)	Handling big data for industry
Gong et al. (2012)	Assessing climate-related heat effects on craft time utilization in the construction industry
A. G. Hobeika et al. (1993)	Real-time traffic diversion model
J. Y. Liu et al. (2013)	Assessing the impact of green practices on collaborative work in China's construction industry
Ock et al. (2016)	Smart Building Energy Management Systems (BEMS) simulation
H. Wang et al. (2014)	Evaluating the effectiveness of emergency response system for oil spill
M'baya et al. (2017)	Assessing the modernization of Legacy Systems
Al-Ruithe et al. (2017)	Cloud data governance-driven decision making
Junxiao Liu et al. (2014)	Performance measurement of public-private partnerships

3.3.1.2 Theoretical Framework

There are several studies in the literature that present theoretical frameworks. Papadimitriou et al. (2010) proposed a theoretical framework considering crossing behavior of pedestrians along a trip. In another study, Arif et al. (2012) developed a theoretical framework to improve the operational effectiveness of transportation infrastructure asset management. An example of a theoretical framework is shown in the Figure 3-3, based on the work of (Sopian et al., 2017). In addition, Table 3-6 summarizes some of the recent and past studies related to theoretical framework development.

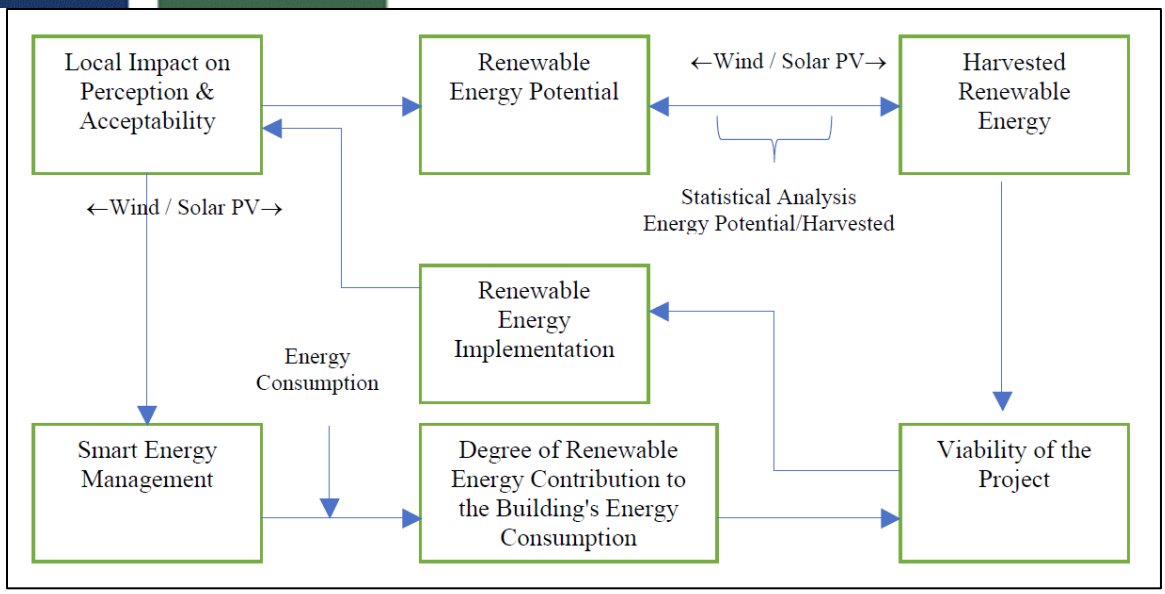


FIGURE 3-3: THEORETICAL FRAMEWORK OF SMART ENERGY MANAGEMENT, ADOPTED FROM (SOPIAN ET AL., 2017)

TABLE 3-6: APPLICATION OF THE THEORETICAL FRAMEWORK

Authors	Application
Xia et al. (2011)	Intelligent Transportation System (ITS) data processing
Green et al. (2005)	Estimation of the number of equivalent strain cycles
Z. Wang et al. (2013)	Selection of construction project delivery method
Fontana et al. (2018)	Controlling pressure-reducing valves in water distribution networks
Al-Shamisi et al. (2016)	Developing an active cyber situational awareness model

3.3.1.3 Methodological Framework

A methodological framework is a set of principles which helps to achieve a particular goal in order to plan and execute any projects. Vos et al. (2012) proposed a methodological framework to evaluate the effectiveness of software testing tools and techniques to help practitioners compare results among similar designs. To optimize planning, design, and operations of airports, Karlaftis et al. (1996) developed a methodological framework to forecast air travel demand. As an example, Figure 3-4 illustrates a comprehensive methodological framework based on four dimensions, namely, benefit, cost, value, and risk (BCVR) that was developed by F. Li et al. (2015) to help decision-makers to resolve various types of decision problems in order to improve performance management systems.

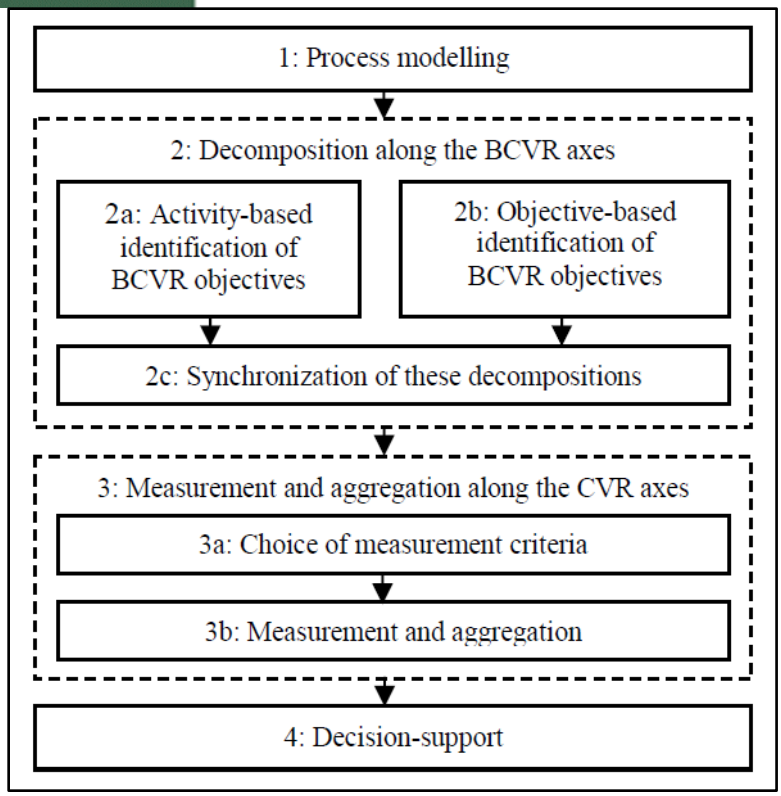


FIGURE 3-4: METHODOLOGICAL FRAMEWORK FOR INDUSTRIAL PERFORMANCE MANAGEMENT, ADAPTED FROM (F. LI ET AL., 2015)

In addition, Table 7 summarizes some of the recent and past studies related to methodological framework development.

TABLE 3-7: APPLICATION OF THE METHODOLOGICAL FRAMEWORK

Authors	Application
Morandi et al. (2013)	Patient safety measures in robotic surgery
Pacheco et al. (2014)	Processing data in a methodological context
Marques Almeida et al. (2015)	Estimation of truck factor
Gómez et al. (2015)	Forecasting vessel performance

The literature review of framework options revealed that a methodological framework is the sequence of several techniques or methods (Karlaftis et al., 1996; F. Li et al., 2015) and thus it is the best framework to use for the purposes of this study. As stated earlier, this study will evolve several techniques of performance measurements based on existing concepts and theories in order to meet its aforementioned objectives. Hence, the methodological framework has been selected for this study and is explored in the following paragraphs.

3.3.2 Development of a Methodological Framework

A team of researchers from the University of Alabama at Birmingham in collaboration with Florida International University developed a methodological framework design for

transportation performance measurement consisting of four parts, namely: (i) physical data flow diagram, (ii) processes and process groups hierarchical diagram, (iii) individual process designs, and (iv) logical data flow diagram. The framework design, represented by these four diagrams is described in the following subsections.

3.3.2.1. Physical data flow diagram

The physical data flow diagram illustrates the physical entities of the proposed system and the data flows between such system entities, including:

(i) vehicles;
(ii) communication technology such as DSRC, LTE, fiber optics, on-board units (OBUs) etc.;
(iii) road-side units (RSUs);
(iv) traffic control devices such as changeable message sign (CMS), variable message signs (VMSs), dynamic message signs (DMSs), highway advisory radio (HAR) etc.;
(v) traffic management center (TMC); and
(vi) drivers or travelers.

This component of the framework is illustrated in Figure 3-5 whereas Figure 3-6 presents the system architecture & physical data flow diagram.

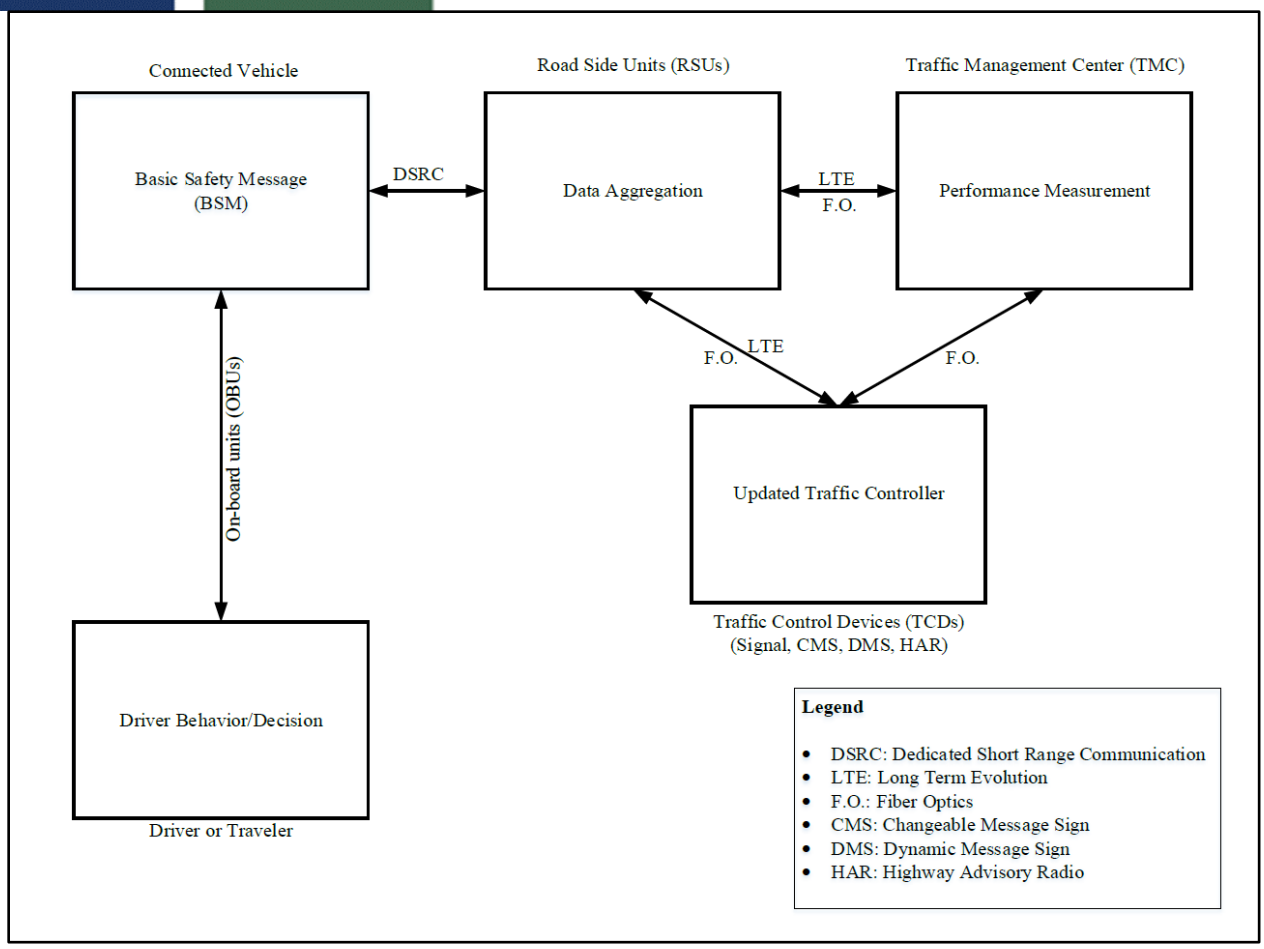


FIGURE 3-5: PHYSICAL DATA FLOW DIAGRAM

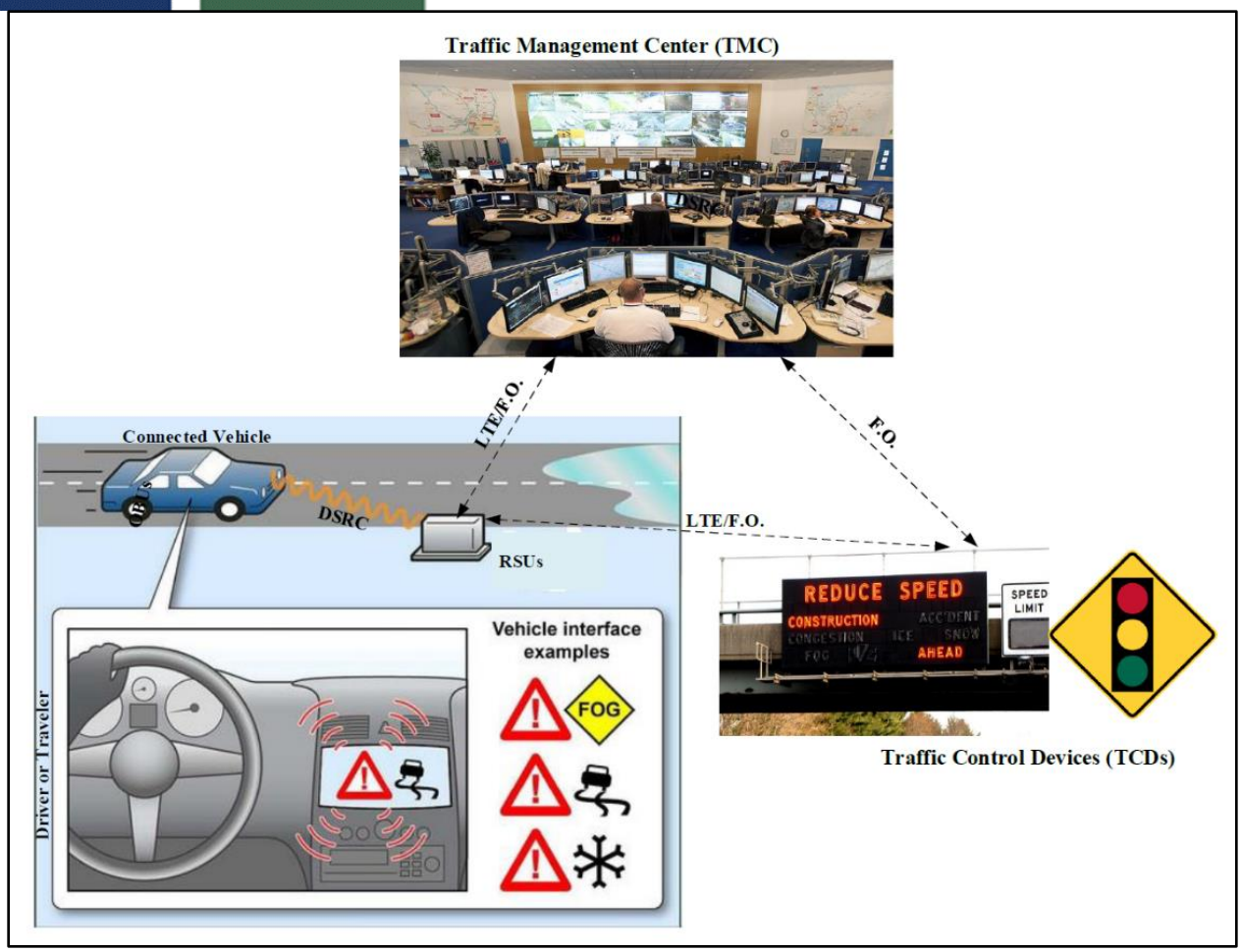


FIGURE 3-6: SYSTEM ARCHITECTURE & PHYSICAL DATA FLOW DIAGRAM

In this framework, vehicles transmit basic safety messages (BSMs) to the road side units (RSUs) through DSRC. The RSUs are equipped with special purpose computers and CCTV cameras. Data aggregation and preliminary processing are performed by the special purpose computers at the RSUs, and the output is transmitted to TMCs through LTE or fiber optic communications. Computers at TMCs implement performance measurement processes and the output is transmitted back to the drivers through CMSs or on-board displays. Additionally, output is transmitted through computer displays to traffic control operators.

The proposed physical architecture of RSUs is illustrated in Figure 3-7. Main entities of the physical architecture of RSUs are FLIR Thermi Cam V2X, DSRC transceiver, Larson Davis sound level meter-831, special purpose computer, LTE, fiber optics, and traffic management center (TMC).

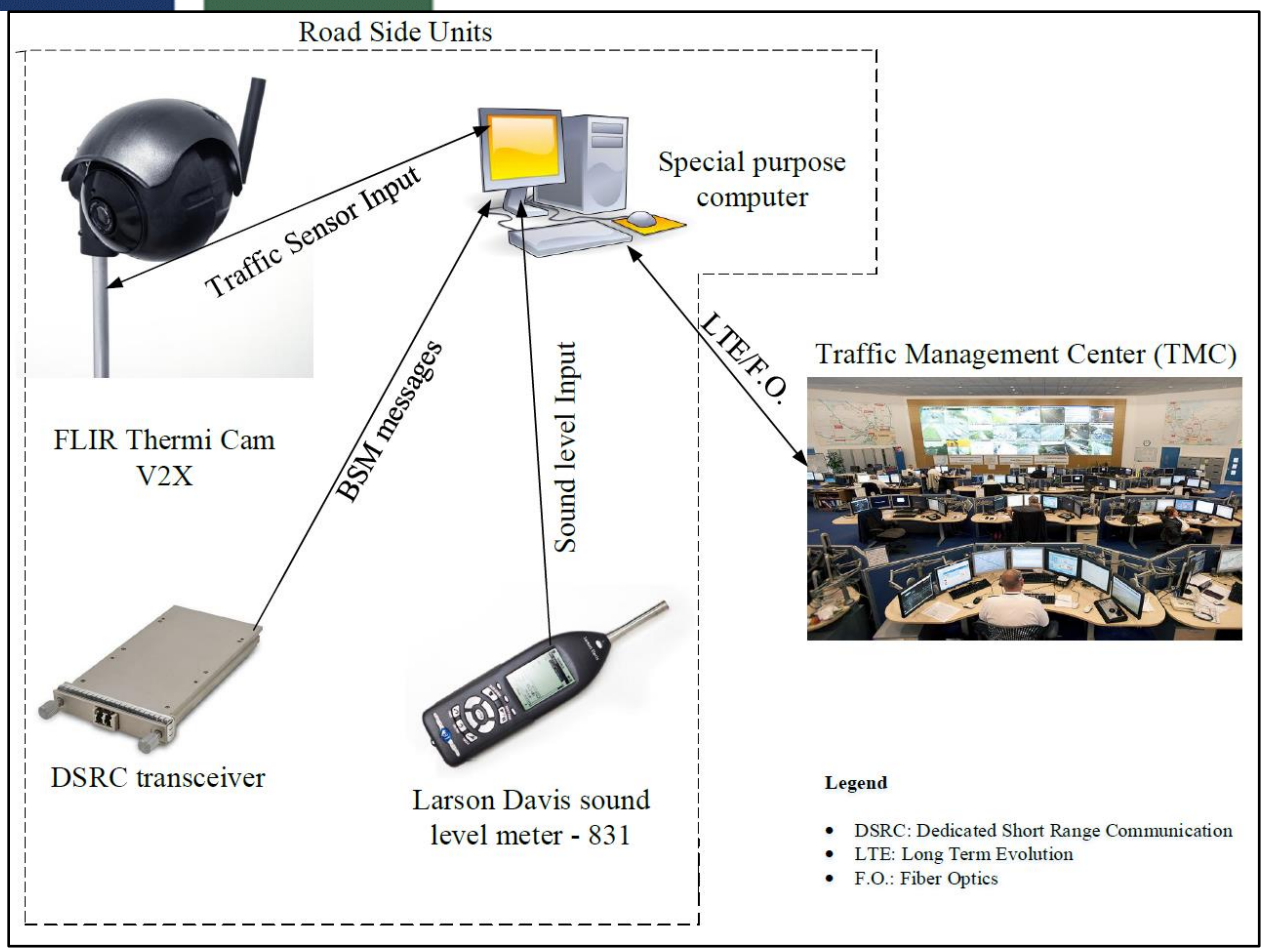


FIGURE 3-7: PHYSICAL ARCHITECTURE OF ROAD SIDE UNITS

The specific functions of entities are described below:

- FLIR Thermi Cam V2X collects vehicle counts, occupancy, classification, speeds, headway, gap etc., DSRC transceiver collects BSM messages, and Larson Davis sound level meter-831 collects sound level data.
- All collected information is transferred to a special purpose computer.
- The special purpose computer processes data to calculate performance measures. Processed data are transferred to TMC through LTE or fiber optics.

3.3.2.2 Processes and process groups hierarchical diagram

There are two sets of process groups in this framework. The first set, illustrated in Figure 3-8, describes data aggregation process groups at RSUs, while the second set, illustrated in Figure 3-9, describes performance measurement process groups at the TMC. While not all inclusive, the hierarchical diagrams presented in Figure 3-8 and 3-9 provide a good picture of the process groups and processes involved in data aggregation at RSUs and the process groups data flow diagram at TMC, respectively.

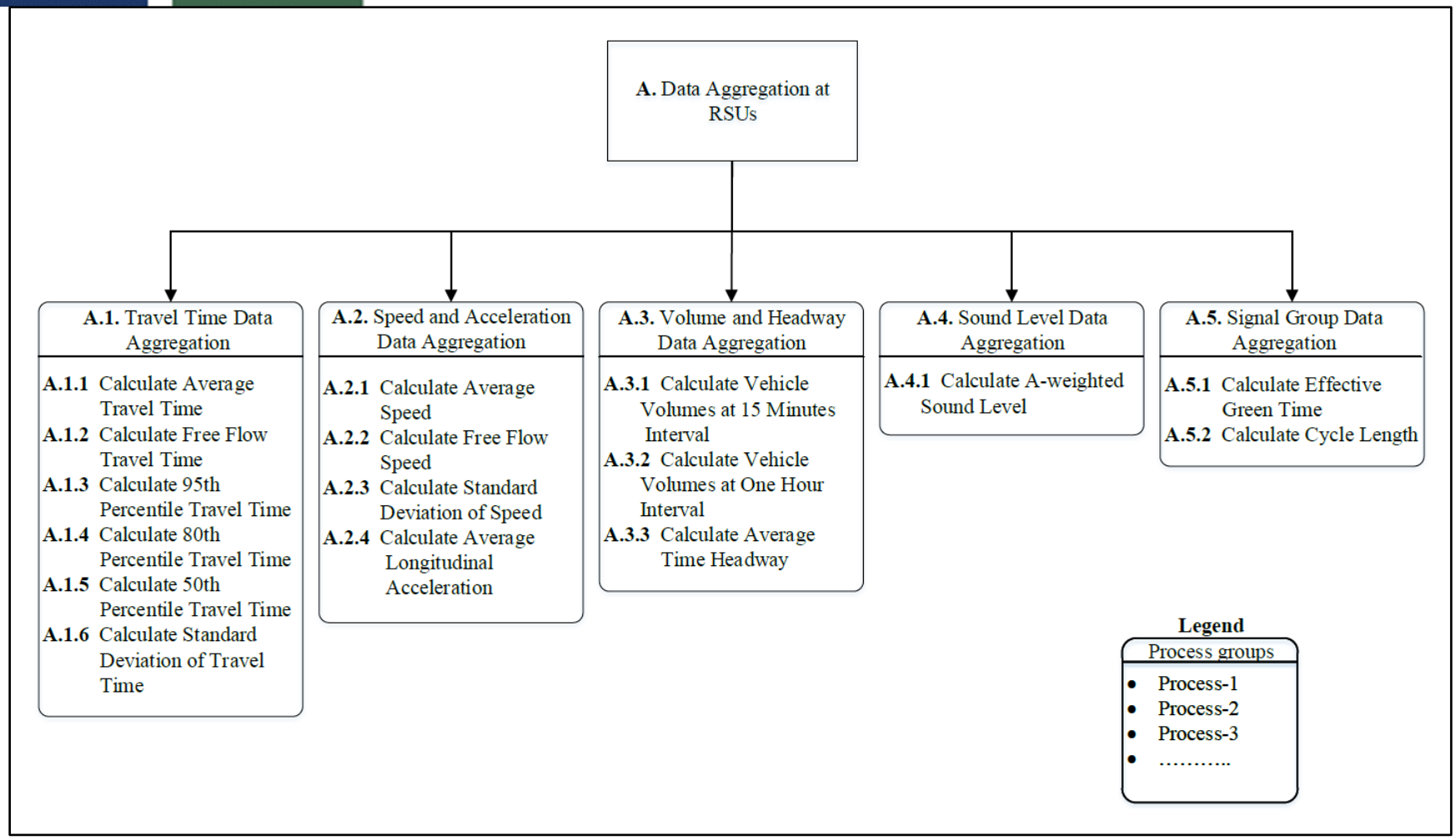


FIGURE 3-8: PROCESS GROUPS AND PROCESSES HIERARCHAL DIAGRAM FOR DATA AGGREGATION AT RSUS

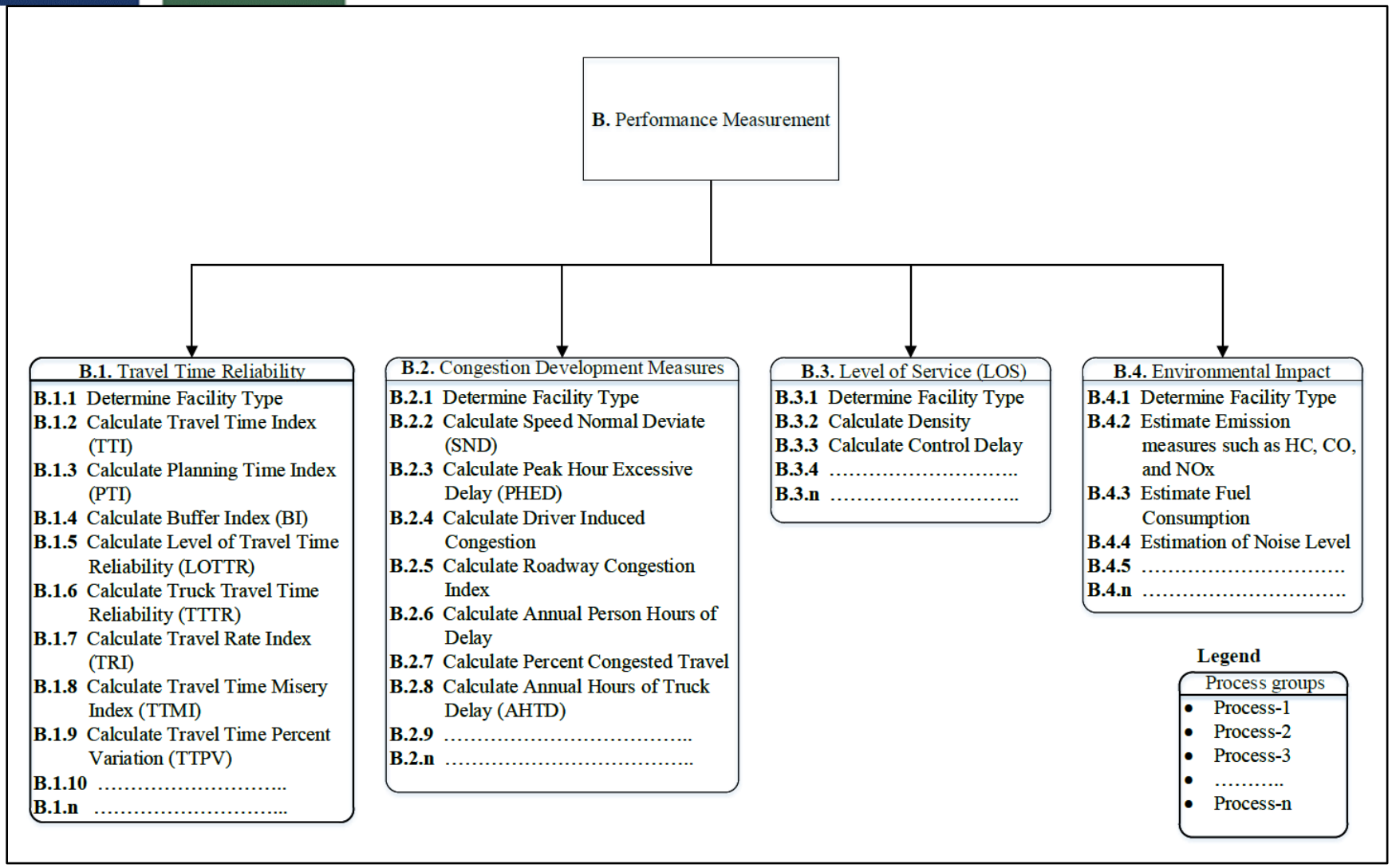


FIGURE 3-9: PROCESS GROUPS DATA FLOW DIAGRAM AT TMC

3.3.2.3 Individual Process Designs

As presented earlier, performance measurement and management is performed through sets of processes. Each individual process can be represented by its inputs, tools and techniques, and outputs. The following subsections present individual process designs, grouped by process groups as illustrated in the previous Figure 3-8 and Figure 3-9.

3.3.2.3.1 PROCESS GROUP A.1 – TRAVEL TIME DATA AGGREGATION

Process Group A.1 "Travel time data aggregation" contains six processes to aggregate discrete measurements that are harvested from BSM and traffic sensors, to generate segment-representative 15-min and daily travel time measurements. The sought aggregate per-segment travel times are 15-minute average travel time; segment free flow travel time; segment 95th, 80th, and 50th percentile travel times; and the standard deviation of the 15-minute average travel time measurements. The following subsections illustrate the process designs details for these six processes.

3.3.2.3.1A PROCESS A.1.1 – CALCULATE AVERAGE TRAVEL TIME

Performance measurement calculations for a given segment depend on 15-minute average travel times. BSMs have time-stamped instantaneous speeds which cannot be directly used in such calculations. This framework proposes Process A.1.1 "Calculate average travel time," as illustrated in Figure 3-10, to calculate the 15-minute average travel time. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the 15-minute average travel time are:

- Speed: This speed refers to instantaneous speed.
- BSM Timestamp: It indicates the specific time at which the data was recorded.
- Vehicle ID: It is temporary and assigned to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Basic Freeway: Indicates which types of facility is being used for analysis.
- Segment Length: Distance covered by each Road Side Unit.
- RSU ID: Identification of Road Side Units.
- Vehicle Count: Number of the vehicle within a specific segment at a particular time interval.
- Traffic Sensor Timestamp: Specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm developed in this study and illustrated in Figure 3-11 to average the input speed over 15-minute intervals and divide the segment length by the average speed to calculate the 15-minute average travel time. Thus, the process output includes 15-minute average travel time for the specific segment, and the RSU ID which identifies that segment.

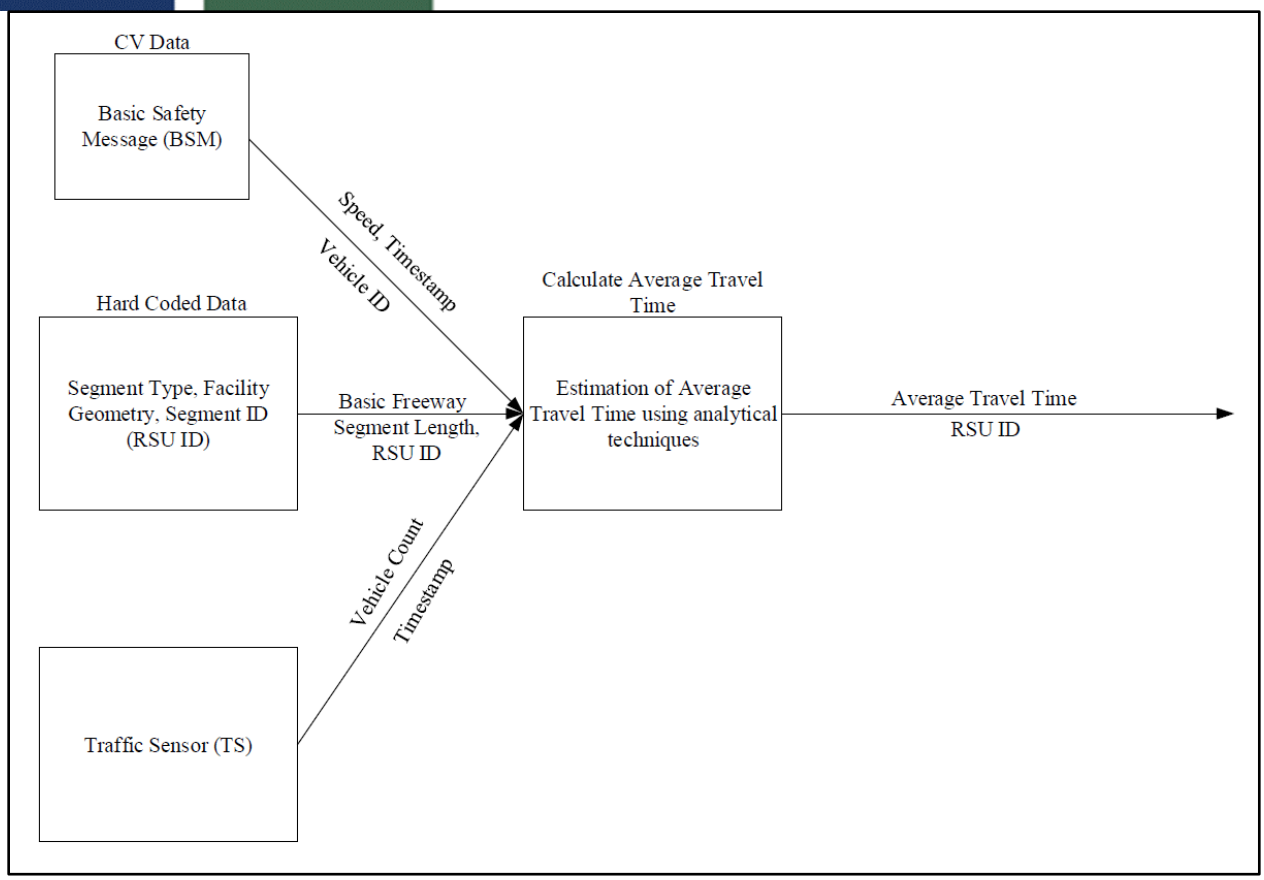


FIGURE 3-10: SCHEMATIC DESIGN OF PROCESS A.1.1 "CALCULATE AVERAGE TRAVEL TIME"

CV_Data = (Timestamp as time, Vehicle ID as integer, Speed as double)

CV_S15 = (15-minutes interval ID as integer, Average speed as double)

CV_TT15 = (15-minutes interval ID as integer, Average travel time as double)

Segment length as double

RSU ID as an integer

Facility type as short text

Average travel time as double

Global Interval (i as an integer shown in Appendix E (Table 8-1), Start time as time, End time as time)

For i = 0 to 95

Average_Speed = Average (SELECT Speed From CV_Data

WHERE Interval (i).Start ≤ CV_Data.Timestamp

≤ Interval (i).End)

CV_S15 (i) = Average_Speed

Average_Travel Time = Segment Length/CV_S15 (i).Average_Speed

CV_TT15 (i) = Average_Travel Time

End

Average travel time = CV_TT15.Average_Travel Time

RSU_ID = RSU ID

FIGURE 3-11: ALGORITHM FOR CALCULATING AVERAGE TRAVEL TIME

3.3.2.3.1B. PROCESS A.1.2 – CALCULATE FREE FLOW TRAVEL TIME

Free flow travel time is an important parameter to estimate performance metrics such as travel time index, planning time index, and annual person-hours of delay. These performance metrics require free flow travel time; however, BSMs have time-stamped instantaneous speed at 0.1-second intervals which cannot be directly used in such calculations. This framework proposes Process A.1.2 "Calculate free flow travel time," as illustrated in Figure 3-12, to calculate the free flow travel time. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the free flow travel time are:

- Speed: It refers to instantaneous speed.
- BSM Timestamp: It indicates the specific time at which the data was recorded.

- Vehicle ID: It is temporary and assigns to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Facility Type: Indicates which types of facility are being used for analysis; i.e., basic freeway segment, freeway merge segment, urban street segment, etc.
- Number of lane: Number of the lane within a specific segment.
- Segment Length: Represented by the distance covered by each Road Side Unit.
- RSU ID: Identification of Road Side Units that is also used to identify segments.
- Vehicle Count: Number of the vehicle within specific segment at a particular interval.
- Traffic Sensor Timestamp: It also means specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm illustrated in Figure 3-13 to estimate free flow travel time by averaging the input speed for those intervals where the traffic flow meets free flow conditions. To illustrate the logic of this process, the case of a basic freeway segment is being considered. Accordingly, free flow conditions are met when traffic flow is under 1,000 veh/hr. Then the segment length is divided by the calculated average speed to estimate the free flow travel time. Thus, the process output includes the free flow travel time for the specific segment, and the RSU ID which identifies that segment.

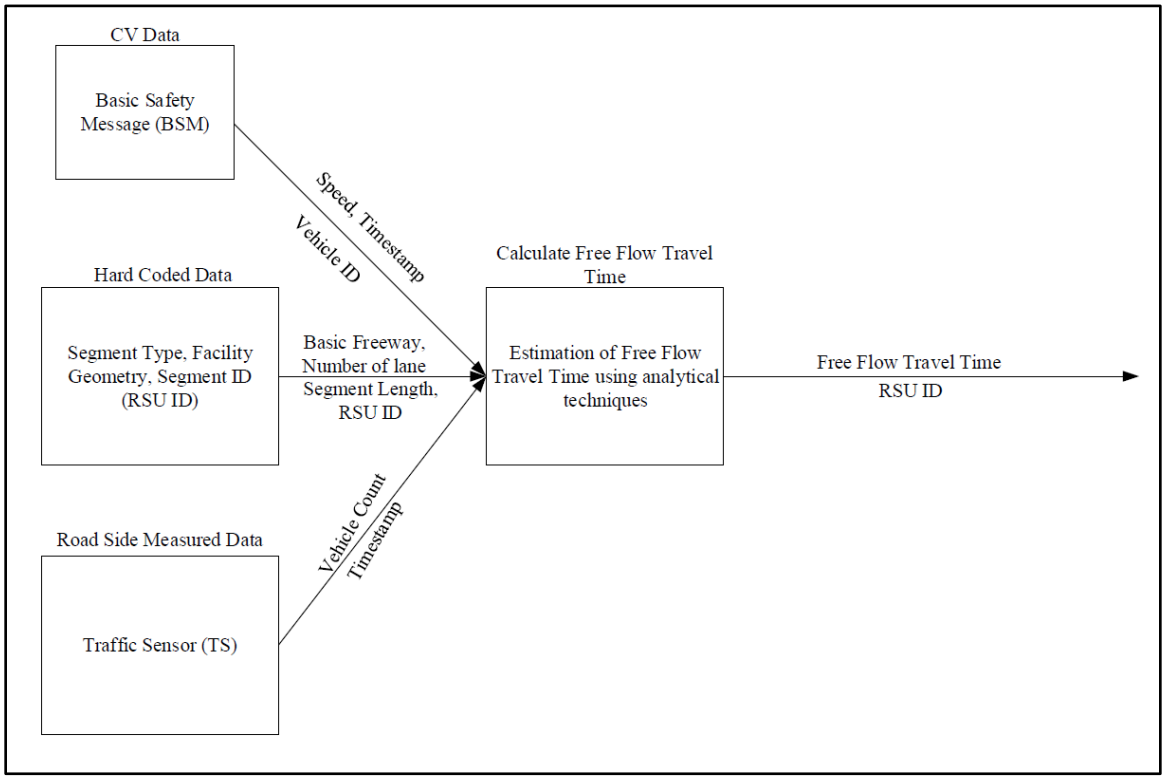


FIGURE 3-12: SCHEMATIC DESIGN OF PROCESS A.1.2 "CALCULATE FREE FLOW TRAVEL TIME"

CV_Data = (Timestamp as time, Vehicle ID as integer, Speed as double)

TS_Data = (Timestamp as time, Vehicle count as integer)

CV_S15 = (15-minutes interval ID as integer, Average speed as double)

CV_TT15 = (15-minutes interval ID as integer, Average travel time as double)

TS_V15 = (15-minutes interval ID as integer, Vehicle count/lane as integer)

Number of lanes as an integer

Segment length as double

RSU ID as an integer

Facility type as short text

FFTT as double

Global Interval (i as an integer, Start time as time, End time as time)

For i = 0 to 95

Average_Speed = Average (SELECT Speed From CV_Data

WHERE Interval (i).Start ≤ CV_Data.Timestamp

≤ Interval (i).End)

CV_S15 (i) = Average_Speed

Average_Travel Time = Segment Length/CV_S15 (i).Average_Speed

Vehicle_count = TS_V15 (i)

If Vehicle_count ≤ 250

then CV_TT15 (i) = Average_Travel Time

End

FFTT = CV_TT15.Average_Travel time

RSU_ID = RSU ID

FIGURE 3-13: ALGORITHM FOR CALCULATING FREE FLOW TRAVEL TIME

3.3.2.3.1C. PROCESS A.1.3 – CALCULATE 95TH PERCENTILE TRAVEL TIME

Travel time reliability measures of a given segment depend on 15-minute 95th percentile travel time. But the primary challenge is to generate 15-minute 95th percentile travel time for a particular segment using BSMs which contain stamped instantaneous speed at the 0.1-second interval. This framework proposes Process A.1.3 "Calculate 95th percentile travel time," as illustrated in Figure 3-14, to calculate the 15-minute 95th percentile travel time. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the 15-minute 95th percentile travel time are:

- Speed: It means instantaneous speed.
- BSM Timestamp: It indicates the specific time at which the data was recorded.
- Vehicle ID: It is temporary and assigns to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- Segment Length: Distance covered by each Road Side Unit.
- RSU ID: Identification of Road Side Units.
- Vehicle Count: Number of the vehicle within specific segment at a particular interval.
- Traffic Sensor Timestamp: It also means specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm illustrated in Figure 3-15 to estimate the 15-minute 95th percentile travel time. The algorithm calculates the 95th percentile of the input speed over 15-minute intervals and divides the segment length by the 95th percentile speed. So, the process output includes 15-minute 95th percentile travel time for the specific segment, and the RSU ID which identifies that segment.

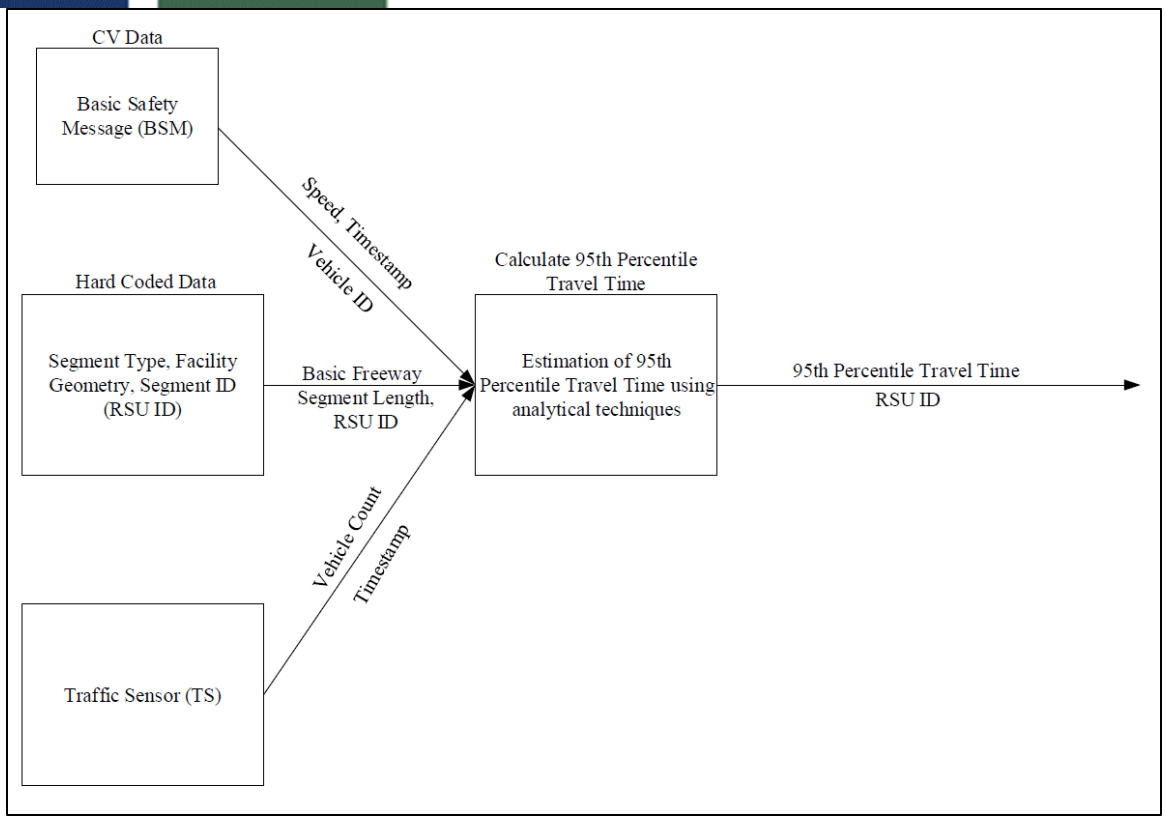


FIGURE 3-14: SCHEMATIC DESIGN OF PROCESS A.1.3 "CALCULATE 95TH PERCENTILE TRAVEL TIME"

CV_Data = (Timestamp as time, Vehicle ID as integer, Speed as double)

CV_S15 = (15-minutes interval ID as integer, 95th percentile speed as double)

CV_TT15 = (15-minutes interval ID as integer, 95th percentile travel time as double)

Segment length as double

RSU ID as integer

Facility type as short text

95th percentile travel time as double

Global Interval (i as integer, Start time as time, End time as time)

For i = 0 to 95

95th percentile_Speed = Pct.95 (SELECT Speed From CV_Data

WHERE Interval (i).Start ≤ CV_Data.Timestamp

≤ Interval (i).End)

CV_S15 (i) = 95th percentile_Speed

95th percentile_travel time = Segment Length/CV_S15 (i). 95th percentile_Speed

CV_TT15 (i) = 95th percentile_travel time

End

95th percentile travel time = CV_TT15. 95th percentile_travel time

RSU_ID = RSU ID

FIGURE 3-15: ALGORITHM FOR CALCULATING 95TH PERCENTILE TRAVEL TIME

3.3.2.3.1D. PROCESS A.1.4 – CALCULATE 80TH PERCENTILE TRAVEL TIME

The level of travel time reliability estimation (a MAP-21 performance metric) depends on 15-minute 80th percentile travel times. BSMs contain time-stamped instantaneous speed at 0.1-second intervals which cannot be directly used in such calculations. This framework proposes Process A.1.4 "Calculate 80th percentile travel time," as illustrated in Figure 3-16, to calculate the 15-minute 80th percentile travel time. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the 15-minute 80th percentile travel time are:

- Speed: It refers to instantaneous speed.
- BSM Timestamp: It indicates the specific time at which the data was recorded.

- Vehicle ID: It is temporary and assigns to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- Segment Length: Distance covered by each Road Side Unit.
- RSU ID: Identification of Road Side Units.
- Vehicle Count: Number of the vehicle within specific segment at a particular interval.
- Traffic Sensor Timestamp: It also means specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm illustrated in Figure 3-17 to estimate the 15-minute 80th percentile travel time by calculating the 80th percentile of the input speed over 15-minute intervals and dividing the segment length by the 80th percentile speed. Thus, the process output includes 15-minute 80th percentile travel time for the specific segment, and the RSU ID which identifies that segment.

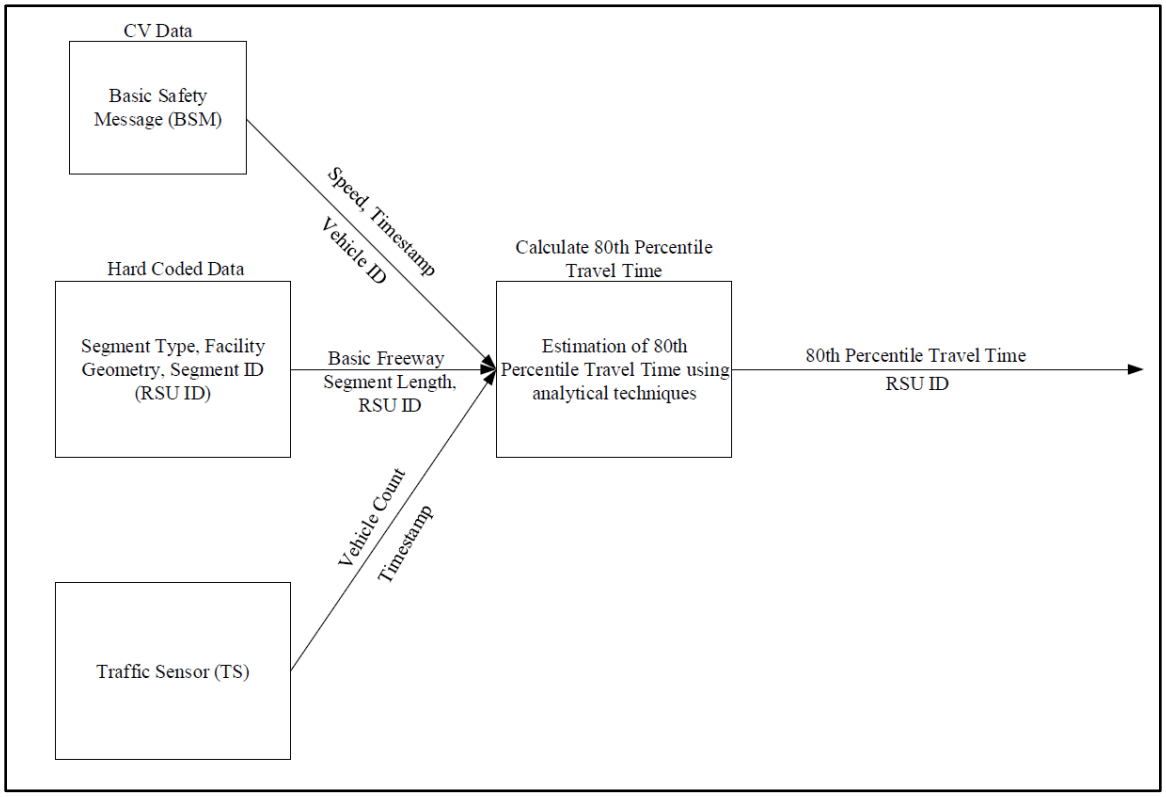


FIGURE 3-16: SCHEMATIC DESIGN OF PROCESS A.1.4 "CALCULATE 80TH PERCENTILE TRAVEL TIME"

CV_Data = (Timestamp as time, Vehicle ID as integer, Speed as double)

CV_S15 = (15-minutes interval ID as integer, 80th percentile speed as double)

CV_TT15 = (15-minutes interval ID as integer, 80th percentile travel time as double)

Segment length as double

RSU ID as integer

Facility type as short text

80th percentile travel time as double

Global Interval (i as integer, Start time as time, End time as time)

For i = 0 to 95

80th percentile_Speed = Pct.80 (SELECT Speed From CV_Data

WHERE Interval (i).Start ≤ CV_Data.Timestamp

≤ Interval (i).End)

CV_S15 (i) = 80th percentile_Speed

80th percentile_travel time = Segment Length/CV_S15 (i).80th percentile_Speed

CV_TT15 (i) = 80th percentile_travel time

End

80th percentile travel time = CV_TT15. 80th percentile_travel time

RSU_ID = RSU ID

FIGURE 3-17: ALGORITHM FOR CALCULATING 80TH PERCENTILE TRAVEL TIME

3.3.2.3.1E. PROCESS A.1.5 – CALCULATE 50TH PERCENTILE TRAVEL TIME

CVs transmits BSMs which contain time-stamped instantaneous speed at 0.1-second intervals. Such raw data cannot be directly used for estimating performance measurements. Performance measurements require 15-minute 50th percentile travel times at each segment. This framework proposes Process A.1.5 "Calculate 50th percentile travel time," as illustrated in Figure 3-18, to calculate the 15-minute 50th percentile travel time. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the 15-minute 50th percentile travel time are:

- Speed: It refers to instantaneous speed.
- BSM Timestamp: It indicates the specific time at which the data was recorded.

- Vehicle ID: It is temporary and assigns to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- Segment Length: Distance covered by each Road Side Unit.
- RSU ID: Identification of Road Side Units.
- Vehicle Count: Number of the vehicle within specific segment at a particular interval.
- Traffic Sensor Timestamp: It also means specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm illustrated in Figure 3-19 to estimate the 15-minute 50th percentile travel time. The algorithm calculates the 50th percentile of the input speed over 15-minute intervals and divides the segment length by the 50th percentile speed. So, the process output includes 15-minute 50th percentile travel time for the specific segment, and the RSU ID which identifies that segment.

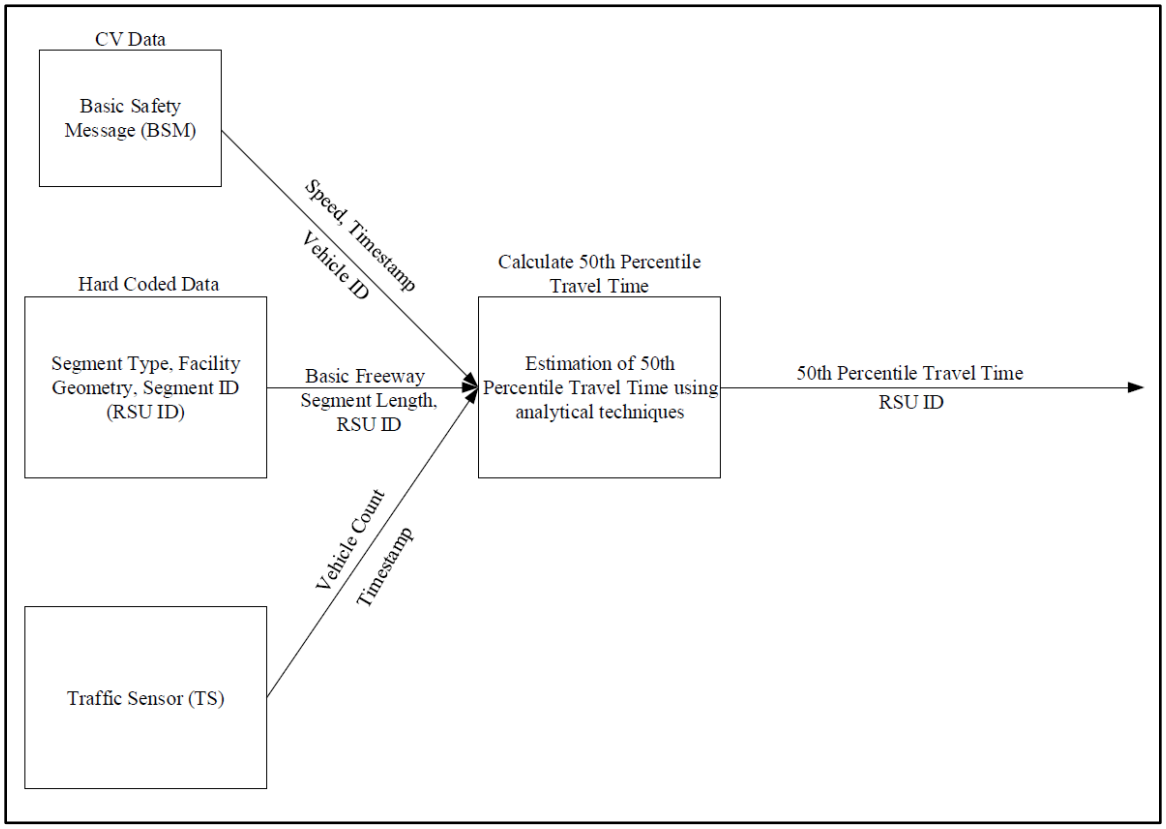


FIGURE 3-18:SCHEMATIC DESIGN OF PROCESS A.1.5 "CALCULATE 50TH PERCENTILE TRAVEL TIME"

CV_Data = (Timestamp as time, Vehicle ID as integer, Speed as double)

CV_S15 = (15-minutes interval ID as integer, 50th percentile speed as double)

CV_TT15 = (15-minutes interval ID as integer, 50th percentile travel time as double)

Segment length as double

RSU ID as integer

Facility type as short text

50th percentile travel time as double

Global Interval (i as integer, Start time as time, End time as time)

For i = 0 to 95

50th percentile_Speed = Pct.50 (SELECT Speed From CV_Data

WHERE Interval (i).Start ≤ CV_Data.Timestamp

≤ Interval (i).End)

CV_S15 (i) = 50th percentile_Speed

50th percentile_travel time = Segment Length/CV_S15 (i).50th percentile_Speed

CV_TT15 (i) = 50th percentile_travel time

End

50th percentile travel time = CV_TT15. 50th percentile_travel time

RSU_ID = RSU ID

FIGURE 3-19: ALGORITHM FOR CALCULATING 50TH PERCENTILE TRAVEL TIME

3.3.2.3.1F. PROCESS A.1.6 – CALCULATE STANDARD DEVIATION OF TRAVEL TIME

Performance measurement calculations for a given segment also depend on 15-minute standard deviation of travel times. BSMs have time-stamped instantaneous speed at 0.1-second intervals which cannot be directly used in such calculations. This framework proposes Process A.1.6 "Calculate standard deviation of travel time," as illustrated in Figure 3-20, to calculate the 15-minute standard deviation of travel time. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the 15-minute standard deviation of travel time are:

- Speed: It refers to instantaneous speed.
- BSM Timestamp: It indicates the specific time at which the data was recorded.

- Vehicle ID: It is temporary and assigns to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- Segment Length: Distance covered by each Road Side Unit.
- RSU ID: Identification of Road Side Units.
- Vehicle Count: Number of the vehicle within specific segment at a particular interval.
- Traffic Sensor Timestamp: It also means specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm illustrated in Figure 3-21 to calculate the 15-minute standard deviation of travel time by calculating the standard deviation of the input speed over 15-minute intervals and divides the segment length by the standard deviation of speed. Thus, the process output includes 15-minute standard deviation of travel time for the specific segment, and the RSU ID which identifies that segment.

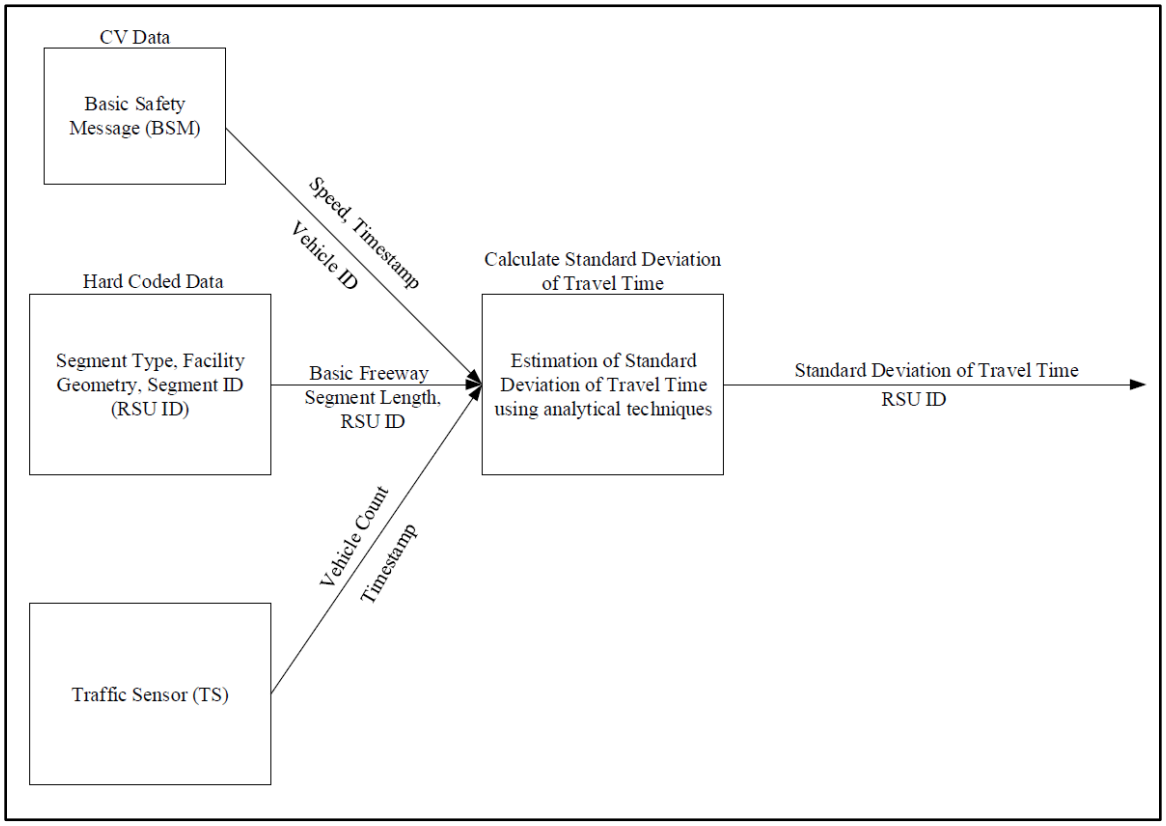


FIGURE 3-20: SCHEMATIC DESIGN OF PROCESS A.1.6 "CALCULATE STANDARD DEVIATION OF TRAVEL TIME"

CV_Data = (Timestamp as time, Vehicle ID as integer, Speed as double)

CV_S15 = (15-minutes interval ID as integer, Standard deviation of speed as double)

CV_TT15 = (15-minutes interval ID as integer, Standard deviation of travel time as double)

Segment length as double

RSU ID as integer

Facility type as short text

Standard deviation of travel time as double

Global Interval (i as integer, Start time as time, End time as time)

For i = 0 to 95

Standard deviation_Speed = Std. (SELECT Speed From CV_Data

WHERE Interval (i).Start ≤ CV_Data.Timestamp

≤ Interval (i).End)

CV_S15 (i) = Standard deviation_Speed

Standard deviation_travel time = Segment Length/CV_S15 (i).Standard deviation_Speed

CV_TT15 (i) = Standard deviation_travel time

End

Standard deviation_travel time = CV_TT15. Standard deviation_travel time

RSU_ID = RSU ID

FIGURE 3-21: ALGORITHM FOR CALCULATING STANDARD DEVIATION OF TRAVEL TIME

3.3.2.3.2 PROCESS GROUP A.2 – SPEED AND ACCELERATION DATA AGGREGATION

The speed and acceleration of vehicle are an excellent indicator to evaluate transportation performance measurements. Process Group A.2 "Speed and acceleration data aggregation" comprises four processes to aggregate BSM and traffic sensors data at 15-minute intervals for each segment. The processed speed and acceleration data at each segment are 15-minute average speed; segment free flow speed; standard deviation of the 15-minute average speed; and 15-minute average longitudinal acceleration. The following subsections explain the process designs for these four processes in detail.

3.3.2.3.2A PROCESS A.2.1 – CALCULATE AVERAGE SPEED

Average speed is a necessary parameter in the calculation of transportation performance measurements. Estimation of 15-minutes average speed is needed, however, the real-time

traffic information data like BSMs provide second by second speed data with the timestamp. Thus, the raw data needs to be processed to estimate 15-minute average speed for a given segment. This framework proposes Process A.2.1 "Calculate average speed," as illustrated in Figure 3-22, to calculate the 15-minute average speed. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the 15-minute average speed are:

- Speed: It refers to instantaneous speed.
- BSM Timestamp It indicates the specific time at which the data was recorded.
- Vehicle ID: It is temporary and assigns to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- RSU ID: Identification of Road Side Units.
- Vehicle Count: Number of the vehicle within specific segment at a particular interval.
- Traffic Sensor Timestamp: It also means specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm illustrated in Figure 3-23 to average the input speed over 15-minute intervals in order to calculate the 15-minute average speed. Thus, the process output includes 15-minute average speed for the specific segment, and the RSU ID which identifies that segment.

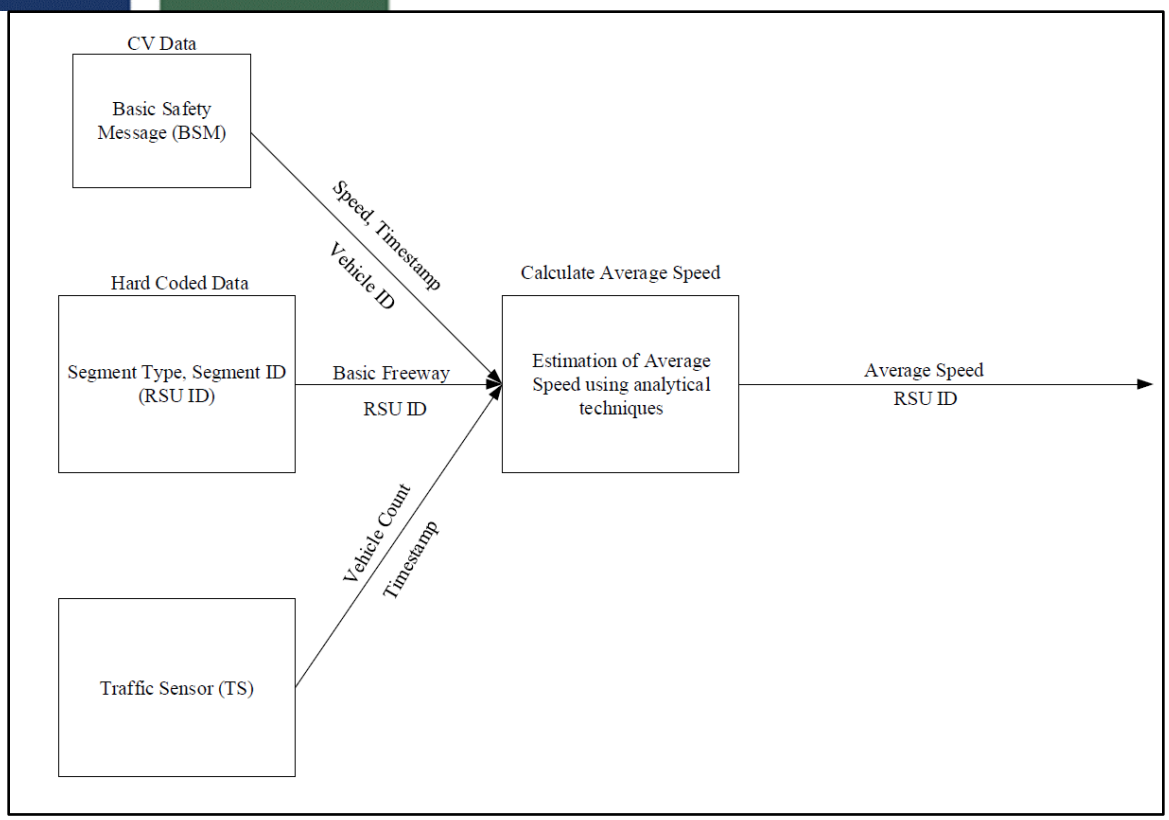


FIGURE 3-22: SCHEMATIC DESIGN OF PROCESS A.2.1 "CALCULATE AVERAGE SPEED"

```
CV_Data = (Timestamp as time, Vehicle ID as integer, Speed as double)

CV_S15 = (15-minutes interval ID as integer, Average speed as double)

RSU ID as integer

Facility type as short text

Average speed as double

Global Interval (i as integer, Start time as time, End time as time)

For i = 0 to 95

Average_Speed = Average (SELECT Speed From CV_Data
                                WHERE Interval (i).Start ≤ CV_Data.Timestamp
                                ≤ Interval (i).End)

CV_S15 (i) = Average_Speed

End

Average Speed = CV_S15.Average_Speed

RSU_ID = RSU ID
```

FIGURE 3-23: ALGORITHM FOR CALCULATING AVERAGE SPEED

3.3.2.3.2B. PROCESS A.2.2 – CALCULATE FREE FLOW SPEED

Free flow speed plays a vital role in the performance evaluation of transportation systems. Generally, performance measurements require free flow speed at 15-minute intervals for a given segment. This framework proposes Process A.2.2 "Calculate free flow speed," as illustrated in Figure 3-24, to calculate the 15-minute free flow speed by aggregating instantaneous speed of BSMs. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the 15-minute free flow speed are:

- Speed: It refers to instantaneous speed.
- BSM Timestamp: It indicates the specific time at which the data was recorded.
- Vehicle ID: It is temporary and assigns to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- Number of lane: Number of lane within a specific segment.
- RSU ID: Identification of Road Side Units.
- Vehicle Count: Number of the vehicle within specific segment at a particular interval.

- Traffic Sensor Timestamp: It also means specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm illustrated in Figure 3-25 to estimate 15-minute free flow speed by averaging the input speed over 15-minute intervals when the vehicle volume is less than or equal 250 vehicles per hour. Thus, the process output includes 15-minute free flow speed for a specific segment, and the RSU ID which identifies that segment.

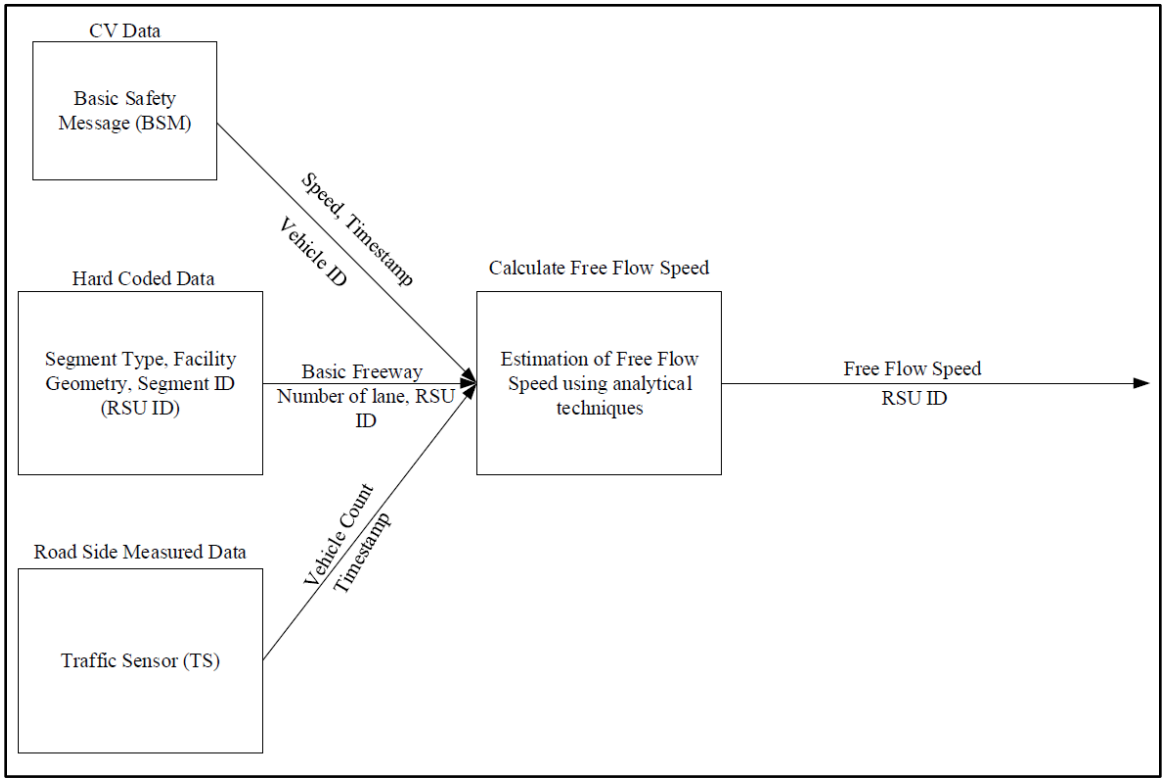


FIGURE 3-24: SCHEMATIC DESIGN OF PROCESS A.2.2 "CALCULATE FREE FLOW SPEED"

CV_Data = (Timestamp as time, Vehicle ID as integer, Speed as double)

TS_Data = (Timestamp as time, Vehicle count as integer)

CV_S15 = (15-minutes interval ID as integer, Average speed as double)

TS_V15 = (15-minutes interval ID as integer, Vehicle count/lane as integer)

Number of lanes as an integer

RSU ID as an integer

Facility type as short text

FFS as double

Global Interval (i as an integer, Start time as time, End time as time)

For i = 0 to 95

Average_Speed = Average (SELECT Speed From CV_Data

WHERE Interval (i).Start ≤ CV_Data.Timestamp

≤ Interval (i).End)

Vehicle_count = TS_V15 (i)

If Vehicle_count ≤ 250

then CV_S15 (i) = Average_Speed

End

FFS = CV_S15.Average_Speed

RSU_ID = RSU ID

FIGURE 3-25: ALGORITHM FOR CALCULATING FREE FLOW SPEED

3.3.2.3.2C PROCESS A.2.3 – CALCULATE STANDARD DEVIATION OF SPEED

Speed normal deviate (SND) calculations for a given segment depend on 15-minute standard deviation of speed. BSMs have time-stamped instantaneous speed at 0.1-second intervals which cannot be directly used in such calculations. This framework proposes Process A.2.3 "Calculate standard deviation of speed," as illustrated in Figure 3-26, to calculate the 15-minute standard deviation of speed. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the 15-minute standard deviation of speed are:

- Speed: It refers to instantaneous speed.
- BSM Timestamp: It indicates the specific time at which the data was recorded.

- Vehicle ID: It is temporary and assigns to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- RSU ID: Identification of Road Side Units.
- Vehicle Count: Number of the vehicle within specific segment at a particular interval.
- Traffic Sensor Timestamp: It also means specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm illustrated in Figure 3-27 to calculate the 15-minute standard deviation of speed by calculating the standard deviation of the input speed over 15-minute intervals. Thus, the process output includes 15-minute standard deviation of speed for the specific segment, and the RSU ID which identifies that segment.

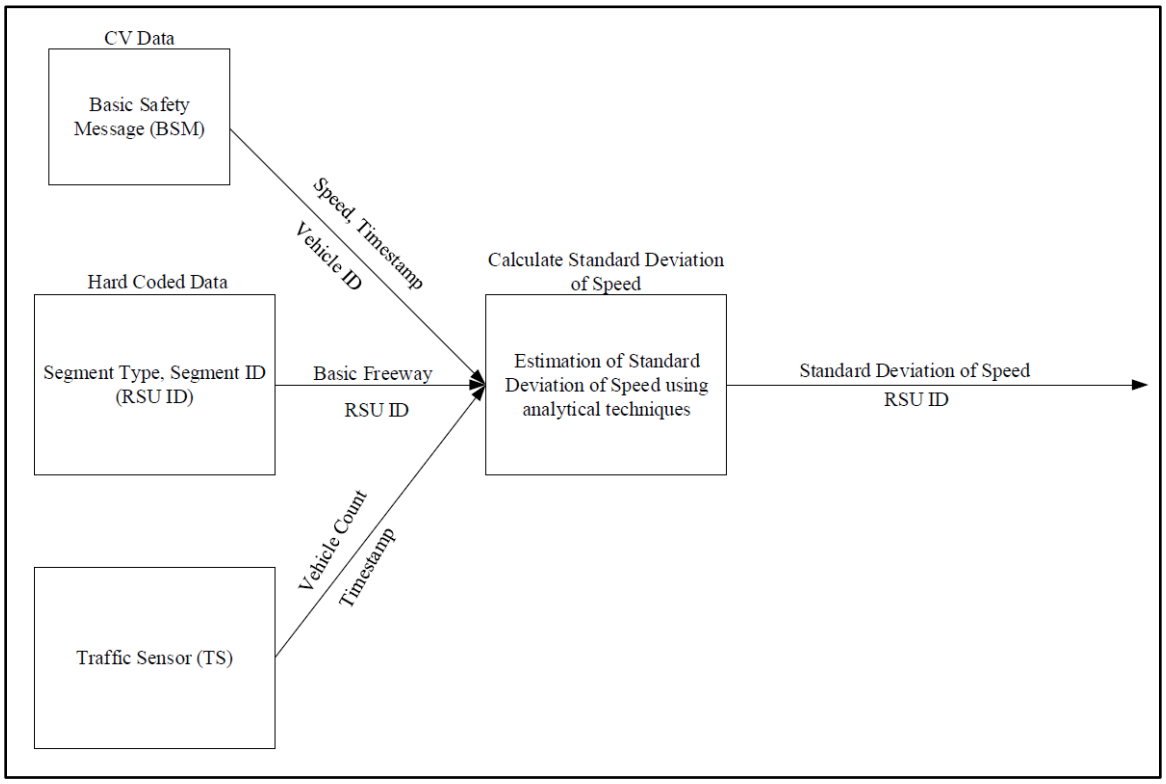


FIGURE 3-26: SCHEMATIC DESIGN OF PROCESS A.2.3 "CALCULATE STANDARD DEVIATION OF SPEED"

CV_Data = (Timestamp as time, Vehicle ID as integer, Speed as double)

CV_S15 = (15-minutes interval ID as integer, Standard deviation of speed as double)

RSU ID as integer

Facility type as short text

Standard deviation of speed as double

Global Interval (i as integer, Start time as time, End time as time)

For i = 0 to 95

Standard deviation_Speed = Std. (SELECT Speed From CV_Data

WHERE Interval (i).Start ≤ CV_Data.Timestamp

≤ Interval (i).End)

CV_S15 (i) = Standard deviation_Speed

End

The standard deviation of speed = CV_S15. Standard deviation_Speed

RSU_ID = RSU ID

FIGURE 3-27: ALGORITHM FOR CALCULATING STANDARD DEVIATION OF SPEED

3.3.2.3.2D PROCESS A.2.4 – CALCULATE AVERAGE LONGITUDINAL ACCELERATION

Emission rate and fuel consumption calculations for a given segment depend on 15-minute average longitudinal acceleration. BSMs have time-stamped instantaneous longitudinal acceleration at 0.1-second intervals which cannot be directly used in such calculations. This framework proposes Process A.2.4 "Calculate average longitudinal acceleration," as illustrated in Figure 3-28, to calculate the 15-minute average longitudinal acceleration. This process requires input from two data sources, namely BSM, and hard coded data in the RSU special purpose computer. Specifically, the inputs required to calculate the 15-minute average longitudinal acceleration are:

- Longitudinal acceleration: It is an instantaneous acceleration along travel direction.
- BSM Timestamp: It indicates the specific time at which the data was recorded.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- RSU ID: Identification of Road Side Units.

This process uses the algorithm illustrated in Figure 3-29 to average the input longitudinal acceleration over 15-minute intervals in order to calculate the 15-minute average longitudinal acceleration. Thus, the process output includes 15-minute average longitudinal acceleration for the specific segment, and the RSU ID which identifies that segment.

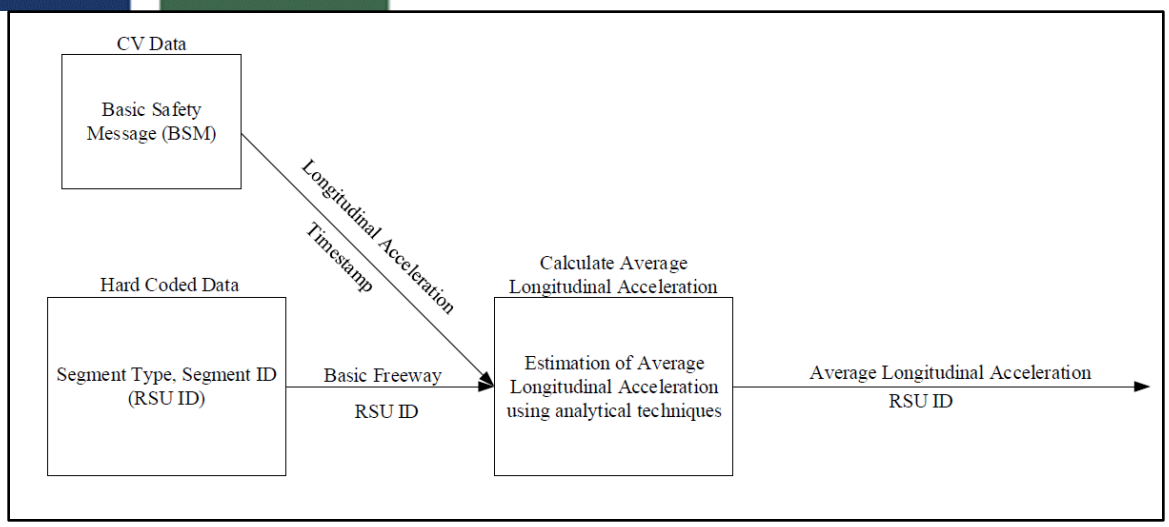


FIGURE 3-28: SCHEMATIC DESIGN OF PROCESS A.2.4 "CALCULATE AVERAGE LONGITUDINAL ACCELERATION"

CV_Data = (Timestamp as time, longitudinal acceleration as double)

CV_A15 = (15-minutes interval ID as integer, Average longitudinal acceleration as double)

RSU ID as integer

Facility type as short text

Average longitudinal acceleration as double

Global Interval (i as integer, Start time as time, End time as time)

For i = 0 to 95

Average longitudinal_Acceleration = Average (SELECT positive value of longitudinal acceleration

From CV_Data

WHERE Interval (i).Start ≤ CV_Data.Timestamp ≤ Interval (i).End)

CV_A15 (i) = Average longitudinal_Acceleration

End

Average longitudinal acceleration = CV_A15. Average longitudinal_Acceleration

RSU_ID = RSU ID

FIGURE 3-29: ALGORITHM FOR CALCULATING AVERAGE LONGITUDINAL ACCELERATION

3.3.2.3.3. PROCESS GROUP A.3 – VOLUME AND HEADWAY DATA AGGREGATION

Volume and headway data are required to estimate some vital traffic flow parameters such as flow rate, peak hour factor, and density. Process Group A.3 "Volume and headway data

aggregation" contains three processes to aggregate BSM and traffic sensors data, to generate segment representative 15-minute volume and headway data. The aggregated volume and headway data at each segment are (i) total vehicle volumes at 15-minute intervals; (ii) total vehicle volumes at one-hour intervals; and (iii) average time headway. The following subsections 3.3.2.3.3A through C illustrate the process designs for these three processes.

3.3.2.3.3A PROCESS A.3.1 – CALCULATE VEHICLE VOLUMES AT 15-MINUTES INTERVAL

Estimation of vehicle miles traveled (VMT) for a given segment depend on vehicle volumes at 15-minute intervals. This framework proposes Process A.3.1 "Calculate vehicle volumes at 15-minutes interval," as illustrated in Figure 3-30, to calculate the vehicle volumes at 15-minute intervals. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the vehicle volumes at 15-minute intervals are:

- BSM Timestamp: It indicates the specific time at which the data was recorded.
- Vehicle ID: It is temporary and assigns to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- RSU ID: Identification of Road Side Units.
- Vehicle Count: Number of the vehicle within specific segment at a particular interval.
- Traffic Sensor Timestamp: It also means specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm illustrated in Figure 3-31 to sum the input vehicle over 15-minute intervals in order to calculate the 15-minute vehicle volumes. Thus, the process output includes vehicle volumes at 15-minute intervals for the specific segment, and the RSU ID which identifies that segment.

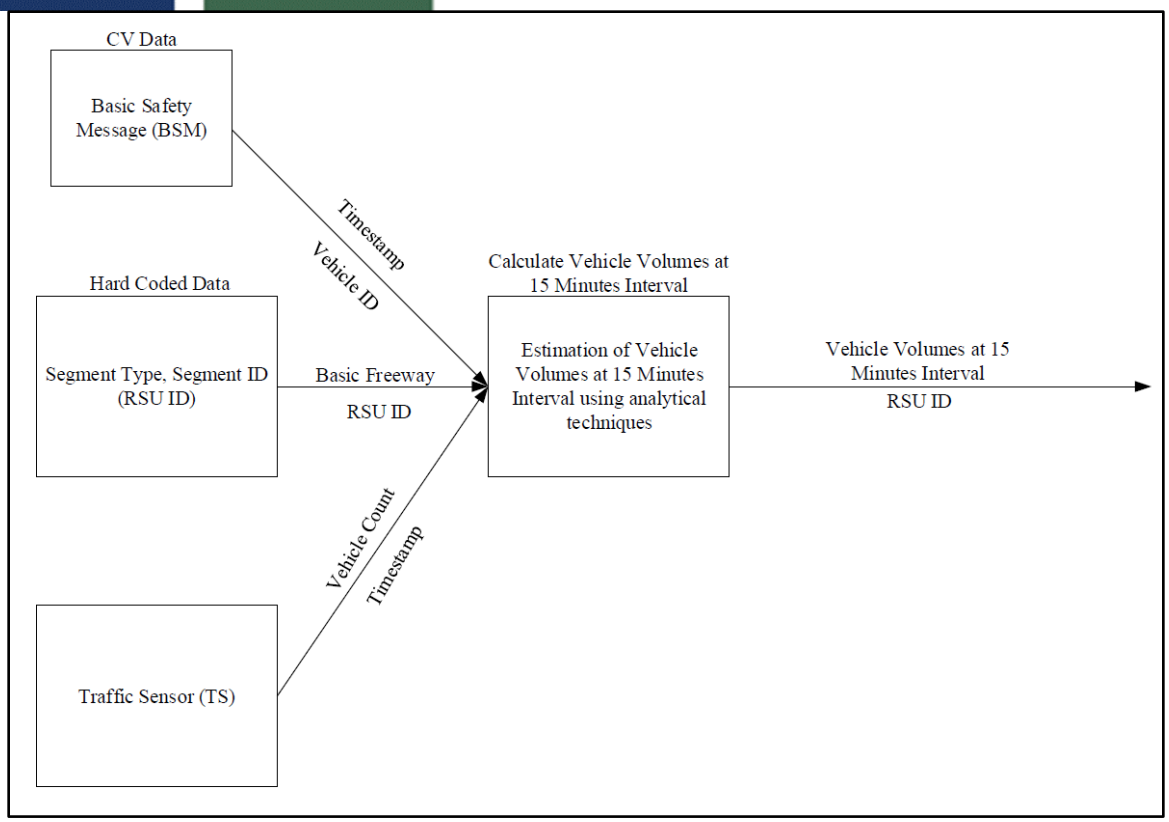


FIGURE 3-30: SCHEMATIC DESIGN OF PROCESS A.3.1 "CALCULATE VEHICLE VOLUMES AT 15-MINUTE INTERVALS"

TS_Data = (Timestamp as time, Vehicle count as integer)

TS_V15 = (15-minutes interval ID as integer, Vehicle count as integer)

RSU ID as integer

Facility type as short text

Vehicle count as double

Global Interval (i as integer, Start time as time, End time as time)

For i = 0 to 95

Vehicle_count = TS_V15 (i)

End

Vehicle volumes at 15-minutes interval = Vehicle_count. TS_V15

RSU_ID = RSU ID

FIGURE 3-31: ALGORITHM FOR CALCULATING VEHICLE VOLUMES AT 15-MINUTES INTERVAL

3.3.2.3.3B PROCESS A.3.2 – VEHICLE VOLUMES AT THE ONE-HOUR INTERVAL

Performance measurement calculations for a given segment also depend on vehicle volumes at one-hour interval. This framework proposes Process A.3.2 "Calculate vehicle volumes at one-hour interval," as illustrated in Figure 3-32, to calculate the vehicle volumes at one-hour interval. This process requires input from three data sources, namely BSM, hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the vehicle volumes at one-hour intervals are:

- BSM Timestamp: It indicates the specific time at which the data was recorded.
- Vehicle ID: It is temporary and assigns to all connected vehicles SAE (2016). Moreover, vehicle ID will change about 5 minutes later.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- RSU ID: Identification of Road Side Units.
- Vehicle Count: Number of the vehicle within specific segment at a particular interval.
- Traffic Sensor Timestamp: It also means specific time at which the data was recorded. However, Traffic Sensor is programmed to get data for every 15-minutes interval.

This process uses the algorithm illustrated in Figure 3-33 to sum the input vehicle over one-hour interval in order to calculate the one-hour vehicle volumes. Thus, the process output includes vehicle volumes at one-hour interval for the specific segment, and the RSU ID which identifies that segment.

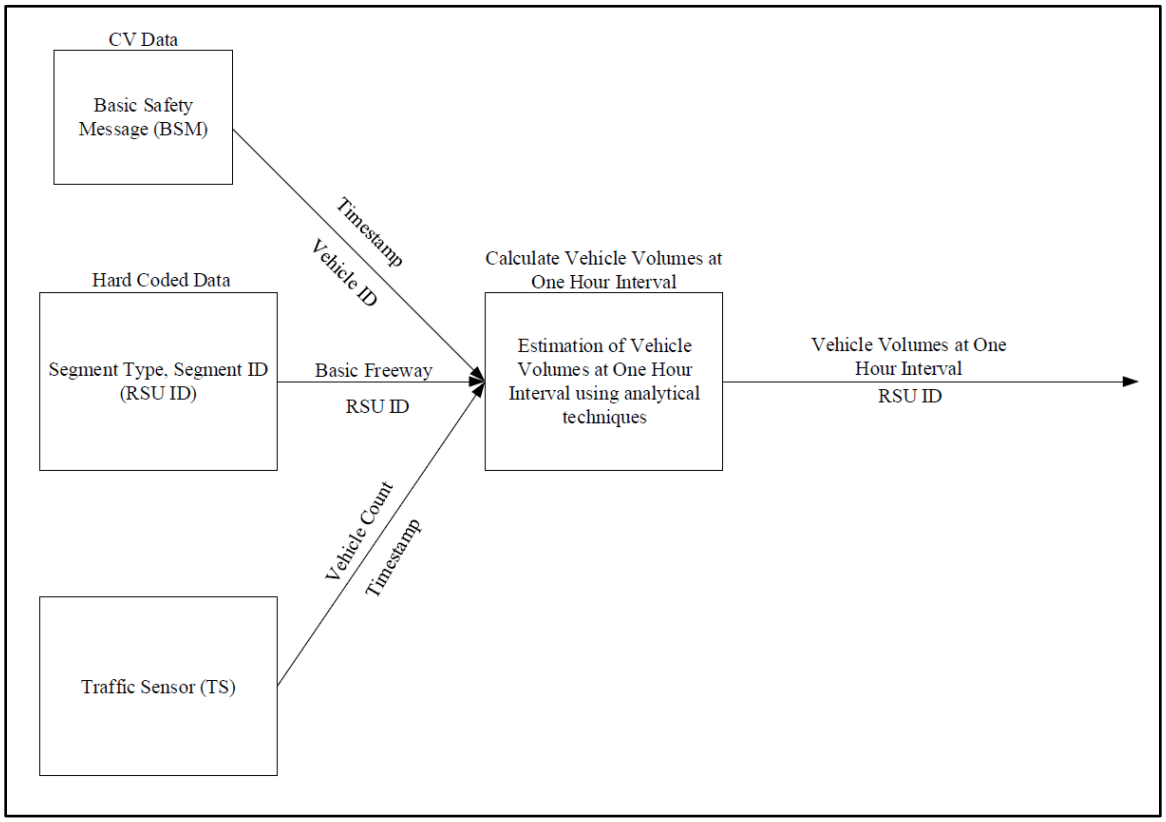


FIGURE 3-32: SCHEMATIC DESIGN OF PROCESS A.3.2 "CALCULATE VEHICLE VOLUMES AT ONE-HOUR INTERVAL"

TS_Data = (Timestamp as time, Vehicle count as integer)

TS_V60 = (60 minutes interval ID as integer, Vehicle count as integer)

RSU ID as integer

Facility type as short text

Vehicle count as double

Global Interval (j as integer shown in Appendix E (Table 8-2), Start time as time, End time as time)

For j = 0 to 23

Vehicle_count = Sum (SELECT Vehicle count From TS_Data

WHERE Interval (j).Start ≤ TS_Data.Timestamp

≤ Interval (j).End)

TS_V60 (j) = Vehicle_count

End

Vehicle Volumes at one-hour interval = TS_V60. Vehicle_count

RSU_ID = RSU ID

FIGURE 3-33: ALGORITHM FOR CALCULATING VEHICLE VOLUMES AT ONE HOUR INTERVAL

3.3.2.3.3C PROCESS A.3.3 – CALCULATE AVERAGE TIME HEADWAY

Flow rate and density calculations for a given segment depend on 15-minute average time headways. This framework proposes Process A.3.3 "Calculate average time headway," as illustrated in Figure 3-34, to calculate the 15-minute average time headway. This process requires input from two data sources, namely hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the 15-minute average time headway are:

- Time Headway: Time between successive vehicles on a lane or roadway. However, Traffic Sensor is programmed to get data for every 15-minutes interval.
- Traffic Sensor Timestamp: It indicates the specific time at which the data was recorded.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- RSU ID: Identification of Road Side Units.

This process uses the algorithm illustrated in Figure 3-35 to average the input time headway over 15-minute intervals in order to calculate the 15-minute average time headway. Thus, the process output includes 15-minute average time headway for the specific segment, and the RSU ID which identifies that segment.

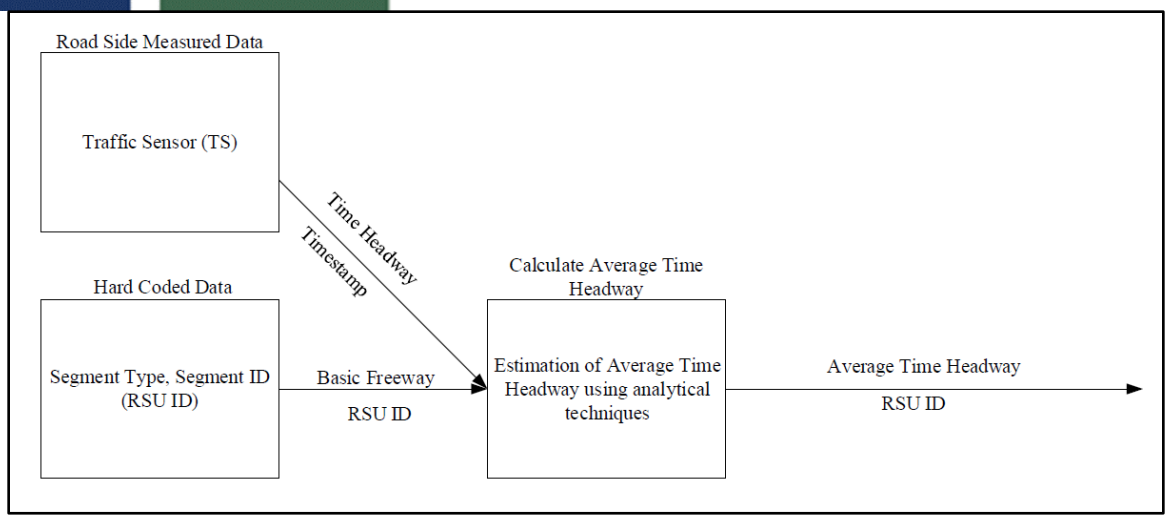


FIGURE 3-34: SCHEMATIC DESIGN OF PROCESS A.3.3 "CALCULATE AVERAGE TIME HEADWAY"

TS_Data = (Timestamp as time, Time Headway as double)

TS_H15 = (15-minutes interval ID as integer, Average time headway as double)

RSU ID as an integer

Facility type as short text

Average time headway as double

Global Interval (i as an integer, Start time as time, End time as time)

For i = 0 to 95

Average Time_Headway = TS_H15 (i)

End

Average time headway = Average Time_Headway.TS_H15

RSU_ID = RSU ID

FIGURE 3-35: ALGORITHM FOR CALCULATING AVERAGE TIME HEADWAY

3.3.2.3.4 PROCESS GROUP A.4 – SOUND LEVEL DATA AGGREGATION

The intensity of sound has an impact on human health. Process Group A.4 "Sound level data aggregation" contains one process to aggregate sound level data that is collected from the sound level meter, to process data at 15-minute intervals for each segment. The sought aggregate per segment sound level is 15-minute A-weighted sound level. The following subsection illustrates the proposed process design for this process.

3.3.2.3.4A PROCESS A.4.1 – CALCULATE A-WEIGHTED SOUND LEVEL

The intensity of sound level calculations for a given segment depend on 15-minute A-weighted sound level. This framework proposes Process A.4.1 "Calculate A-weighted sound level," as illustrated in Figure 3-36, to calculate the 15-minute A-weighted sound level. This process requires input from two data sources, namely hard coded data in the RSU special purpose computer, and the RSU-mounted traffic sensor. Specifically, the inputs required to calculate the 15-minute A-weighted sound level are:

- Sound level: It means the intensity of noise. However, Sound Level Meter (SLM) is programmed to get data for every 15-minutes interval.
- SLM Timestamp: It indicates the specific time at which the data was recorded.
- Basic Freeway: Indicates which types of facility are being used for analysis.
- RSU ID: Identification of Road Side Units.

This process uses the algorithm illustrated in Figure 3-37 to average the input sound level over 15-minute intervals in order to calculate the 15-minute A-weighted sound level. Thus, the process output includes 15-minute A-weighted sound level for the specific segment, and the RSU ID which identifies that segment.

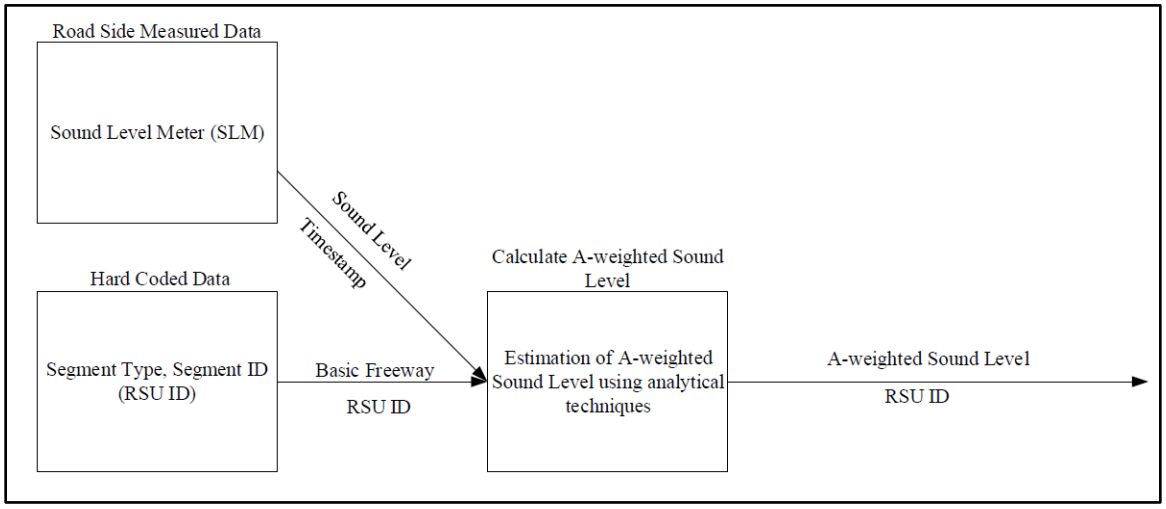


FIGURE 3-36: SCHEMATIC DESIGN OF PROCESS A.4.1 "CALCULATE A-WEIGHTED SOUND LEVEL"

SLM_Data = (Timestamp as time, Sound level as double)

SLM_SL15 = (15-minutes interval ID as integer, A-weighted sound level as double)

RSU ID as an integer

Facility type as short text

A-weighted sound level as double

Global Interval (i as an integer, Start time as time, End time as time)

For i = 0 to 95

A-weighted sound_level = SLM_SL15 (i)

End

A-weighted sound level = A-weighted sound_level.SLM_SL15

RSU_ID = RSU ID

FIGURE 3-37: ALGORITHM FOR CALCULATING A-WEIGHTED SOUND LEVEL

3.3.2.3.5. PROCESS GROUP A.5 – SIGNAL GROUP DATA AGGREGATION

Signal group data aggregation help to estimate control delay as well as Level of Service (LOS) of the signalized intersection. Process Group A.5 "Signal group data aggregation" comprises two processes to aggregate SPaT messages, to generate total effective green time and cycle length at 15-minute intervals for each segment. The following subsections a through b illustrate the process designs for these two processes.

3.3.2.3.5A PROCESS A.5.1 – CALCULATE EFFECTIVE GREEN TIME

LOS calculations for a given intersection depend on 15-minute total effective green. SPaT messages have time-stamped likely times which cannot be directly used in such calculations. This framework proposes Process A.5.1 "Calculate effective green time," as illustrated in Figure 3-38, to calculate the 15-minute total effective green time. This process requires input from two data sources, namely SPaT, and hard coded data in the RSU special purpose computer. Specifically, the inputs required to calculate the 15-minute total effective green time are:

- Likely Time: It refers to effective green time (Ibrahim et al., 2017).
- Intersection Reference ID: Identification of intersection.
- Timestamp: It indicates the specific time at which the data was recorded.
- Signalized Intersection: Indicates which types of facility are being used for analysis.
- RSU ID: Identification of Road Side Units.

This process uses the algorithm illustrated in Figure 3-39 to sum the input likely time over 15-minute intervals in order to calculate the 15-minute total effective green time. Thus, the process output includes 15-minute total effective green time for the specific segment, and the RSU ID which identifies that segment.

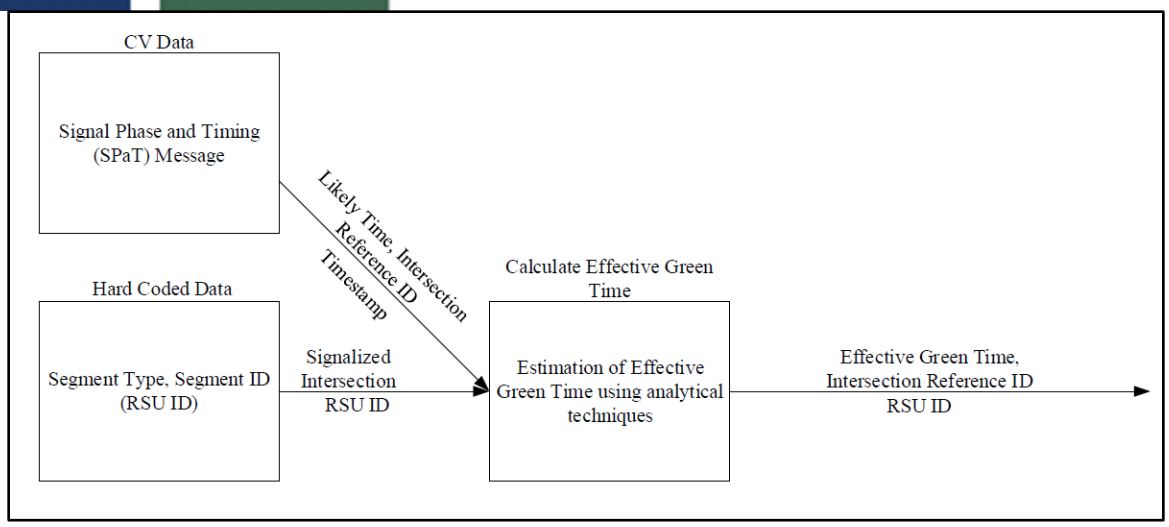


FIGURE 3-38: SCHEMATIC DESIGN OF PROCESS A.5.1 "CALCULATE EFFECTIVE GREEN TIME"

SPaT_Data = (Timestamp as time, Intersection Reference ID as integer, Likely Time as double)

SPaT_GT15 = (15-minutes interval ID as integer, Intersection Reference ID as integer, Total effective green time as double)

RSU ID as an integer

Facility type as short text

Total effective green time as double

Global Interval (i as an integer, Start time as time, End time as time)

For i = 0 to 95

Total Effective Green_Time = Sum (SELECT Likely Time From SPaT_Data

WHERE Interval (i).Start ≤ SPaT_Data.Timestamp ≤ Interval (i).End)

SPaT_GT15 (i) = Total Effective Green_Time

Intersection Reference_ID = SELECT Intersection Reference ID From SPaT_Data

WHERE Interval (i).Start ≤ SPaT_Data.Timestamp ≤ Interval (i).End

SPaT_GT15 (i) = Intersection Reference_ID

End

Total Effective green time = SPaT_GT15.Total Effective Green_Time

Intersection reference ID = SPaT_GT15.Intersection Reference_ID

RSU_ID = RSU ID

FIGURE 3-39: ALGORITHM FOR CALCULATING EFFECTIVE GREEN TIME

3.3.2.3.5B PROCESS A.5.2 – CALCULATE CYCLE LENGTH

To identify the LOS of a signalized intersection, 15-minute total cycle length is required. SPaT messages have time-stamped next times which cannot be directly used in such calculations. This framework proposes Process A.5.2 "Calculate cycle length," as illustrated in Figure 3-40, to calculate the 15-minute total cycle length. This process requires input from two data sources, namely SPaT, and hard coded data in the RSU special purpose computer. Specifically, the inputs required to calculate the 15-minute total cycle length are:

- Next Time: It means cycle length (Ibrahim et al., 2017).
- Intersection Reference ID: Identification of intersection.
- Timestamp: It indicates the specific time at which the data was recorded.
- Signalized Intersection: Indicates which types of facility are being used for analysis.
- RSU ID: Identification of Road Side Units.

This process uses the algorithm illustrated in Figure 3-41 to sum the input next time over 15-minute intervals in order to calculate the 15-minute total cycle length. Thus, the process output includes 15-minute total cycle length for the specific segment, and the RSU ID which identifies that segment.

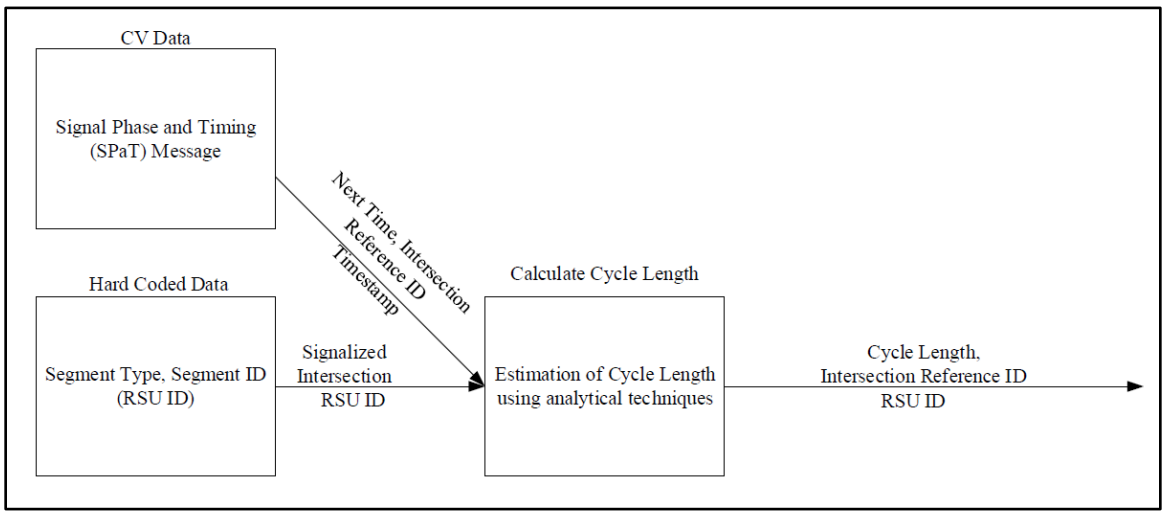


FIGURE 3-40: SCHEMATIC DESIGN OF PROCESS A.5.2 "CALCULATE CYCLE LENGTH"

SPaT_Data = (Timestamp as time, Intersection Reference ID as integer, Next Time as double)

SPaT_CL15 = (15-minutes interval ID as integer, Intersection Reference ID as integer, Total cycle length as double)

RSU ID as an integer

Facility type as short text

Total cycle length as double

Global Interval (i as an integer, Start time as time, End time as time)

For i = 0 to 95

Total Cycle_length = Sum (SELECT Next Time From SPaT_Data

WHERE Interval (i).Start ≤ SPaT_Data.Timestamp ≤ Interval (i).End)

SPaT_CL15 (i) = Total Cycle_length

Intersection Reference_ID = SELECT Intersection Reference ID From SPaT_Data

WHERE Interval (i).Start ≤ SPaT_Data.Timestamp ≤ Interval (i).End

SPaT_CL15 (i) = Intersection Reference_ID

End

Total cycle length = SPaT_CL15.Total Cycle_length

Intersection reference ID = SPaT_CL15.Intersection Reference_ID

RSU_ID = RSU ID

FIGURE 3-41: ALGORITHM FOR CALCULATING CYCLE LENGTH

3.3.2.3.6 PROCESS GROUP B.1 – TRAVEL TIME RELIABILITY

Travel time reliability (TTR) is a measure used to identify the extent of unexpected delay experienced by travelers. Travel time reliability measures better represents a user's travel experience than a simple average travel time. Eight measures of travel time reliability have been identified in this process group B.1 "Travel time reliability." These measures are travel time index, planning time index, buffer index, level of travel time reliability, truck travel time reliability, travel rate index, travel time misery index, and travel time percent variation. The following subsections explain the process designs for these eight measures.

3.3.2.3.6A PROCESS B.1.2 – CALCULATE TRAVEL TIME INDEX (TTI)

Travel time index is an important parameter to measure the average condition of congestion. It compares travel conditions in the peak period to travel conditions during free-flow or posted

speed limit conditions. This framework proposes Process B.1.2 "Calculate travel time index (TTI)" as illustrated in Figure 42, to calculate the 15-minute travel time index. This process requires three types of data from RSU data aggregation, namely 15-minute average travel time, 15-minute free flow travel time, and RSU ID, and facility type from Process B.1.1 "Determine facility type."

For a specific segment and time period, TTI is calculated as follows:

$$TTI = \frac{Average\ travel\ time}{FFTT} \quad \textbf{3-1}$$

Using equation (3-1), the output from this process is 15-minute TTR TTI, Facility for the specific segment.

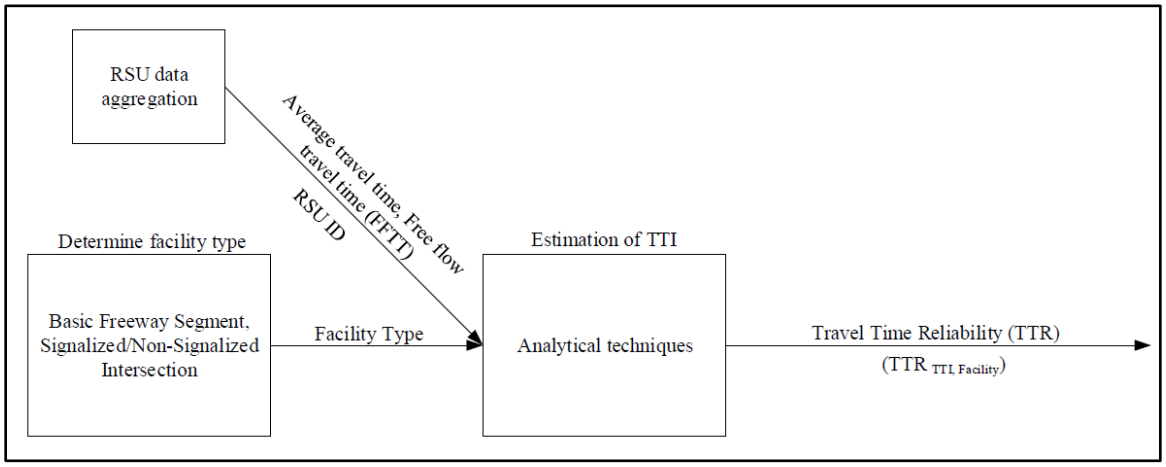


FIGURE 3-42: SCHEMATIC DESIGN OF PROCESS B.1.2 "CALCULATE TRAVEL TIME INDEX (TTI)"

3.3.2.3.6B PROCESS B.1.3 – CALCULATE PLANNING TIME INDEX (PTI)

Planning time index is the ratio of 95th percentile travel time and free flow travel time. It represents how much total time a traveler should allow to ensure on-time arrival. This framework proposes Process B.1.3 "Calculate planning time index (PTI)" as illustrated in Figure 43, to estimate the 15-minute planning time index at segment level. This process uses three types of data such as 15-minute 95th percentile travel time, segment free flow travel time at 15-minute interval, and RSU ID from RSU data, and facility type from Process B.1.1 "Determine facility type."

Equation (3-2) is used to calculate 15-minute PTI for the specific segment based on the 95th percentile travel time and the free flow travel time (FFTT). Hence, the process output include 15-minute $TTR_{PTI,\ Facility}$.

$$PTI = \frac{95th\ percentile\ travel\ time}{FFTT} \quad \textbf{3-2}$$

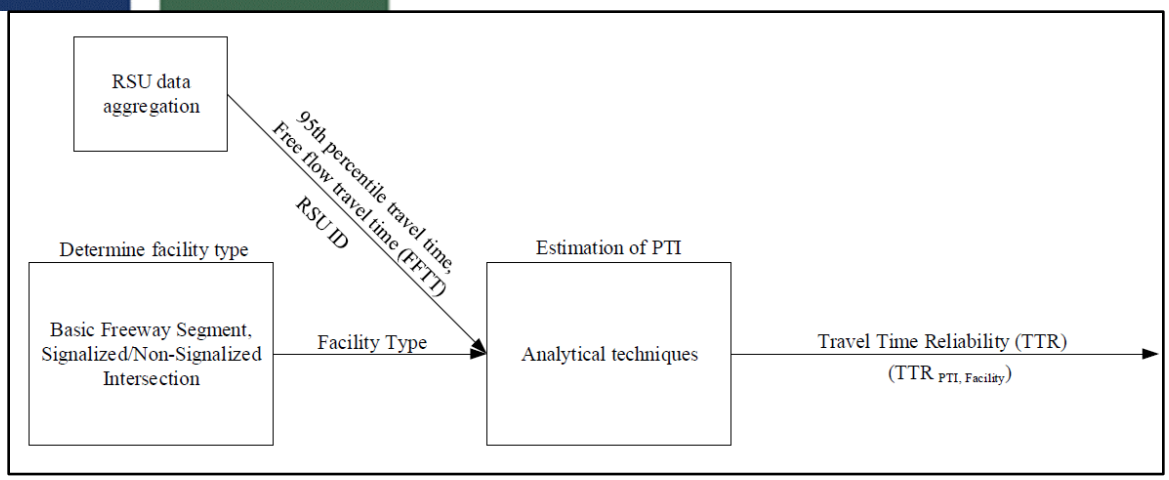


FIGURE 3-43: SCHEMATIC DESIGN OF PROCESS B.1.3 "CALCULATE PLANNING TIME INDEX (PTI)"

3.3.2.3.6C PROCESS B.1.4 – CALCULATE BUFFER INDEX (BI)

Buffer index also is an important parameter to measure travel time reliability. It represents the extra time that travelers must add to their average travel time when planning trips to ensure on-time arrival. This framework proposes Process B.1.4 "Calculate buffer index (BI)" to estimate 15-minute buffer index which is shown below in **Error! Reference source not found.**. This p rocess utilizes data from RSU data aggregation and Process B.1.1 "Determine facility type." Specifically, the inputs required to calculate the 15-minute buffer index are: 95th percentile of the average travel time, 15-minute average travel time, RSU ID, and facility type.

This process uses equation (3-3) to estimate 15-minute BI. Hence, the output is 15-minute $TTR_{BI, Facility}$ for the specific segment.

$$BI = \frac{95th\ percentile\ travel\ time - Average\ travel\ time}{Average\ travel\ time} \qquad 3\text{-}3$$

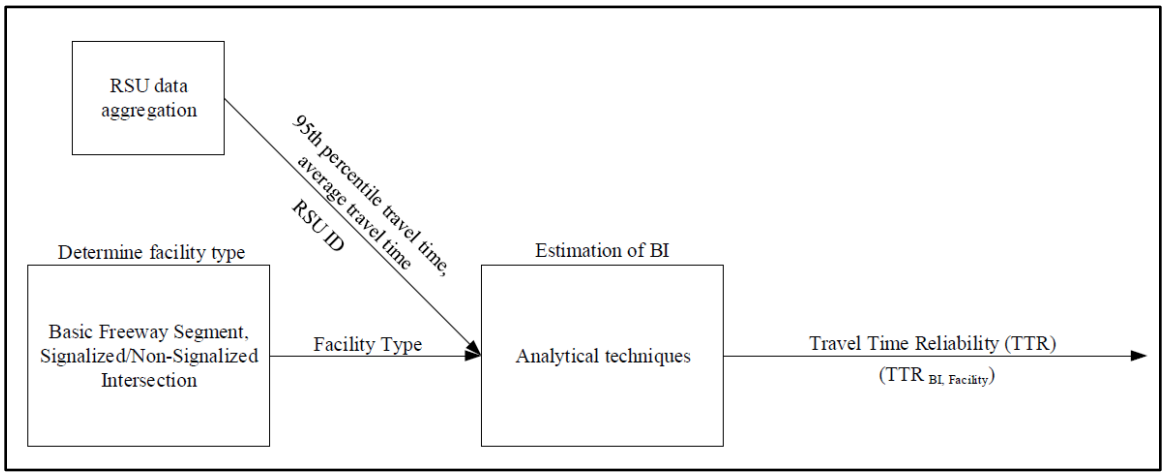


FIGURE 3-44: SCHEMATIC DESIGN OF PROCESS B.1.4 "CALCULATE BUFFER INDEX (BI)"

3.3.2.3.6D PROCESS B.1.5 – CALCULATE LEVEL OF TRAVEL TIME RELIABILITY (LOTTR)

According to MAP-21, level of travel time reliability is an important parameter to measure travel time reliability for passenger car vehicles. It is a new performance metric for travel time reliability. This framework proposes Process B.1.5 "Calculate level of travel time reliability (LOTTR)" as shown below in Figure 3-45 , to calculate 15-minute level of travel time reliability. Four types of data, namely 15-minute 80th percentile travel time, 15-minute 50th percentile travel time, RSU ID, and facility type are required in this process.

LOTTR is expressed as a ratio of the 80th percentile travel time of a reporting segment to the ''normal'' (50th percentile) travel time of a reporting segment occurring throughout a full calendar year. For a specific segment and time period, this is calculated as follows:

$$LOTTR = \frac{\text{80th percentile travel time}}{\text{50th percentile "normal travel time"}} \quad \textbf{3-4}$$

Thus, the process output is the 15-minute TTR $_{LOTTR,\ facility}$.

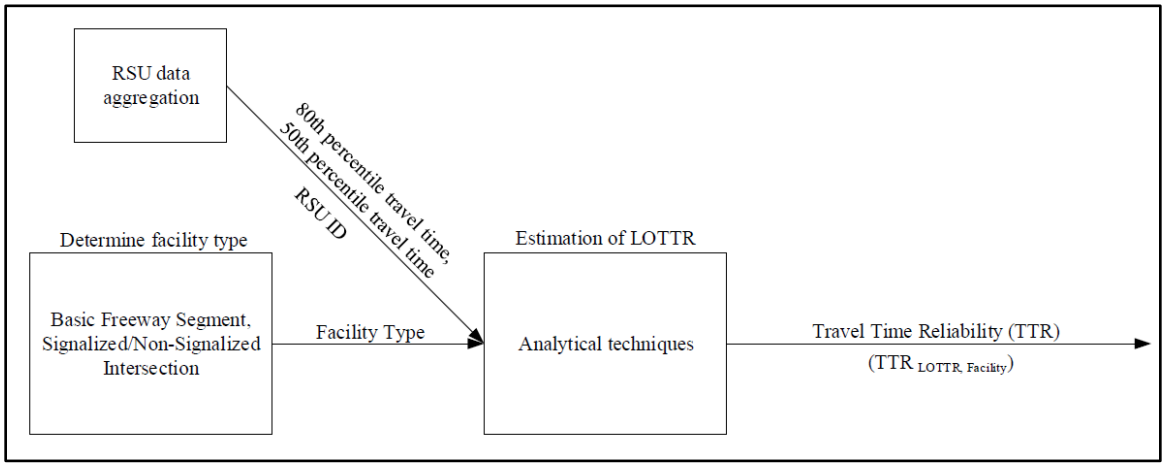


FIGURE 3-45: SCHEMATIC DESIGN OF PROCESS B.1.5 "CALCULATE LEVEL OF TRAVEL TIME RELIABILITY (LOTTR)"

3.3.2.3.6E PROCESS B.1.6 – CALCULATE TRUCK TRAVEL TIME RELIABILITY (TTTR)

According to MAP-21, TTTR is an important parameter for measuring travel time reliability for truck vehicles. It depends on 15-minute 95th percentile travel time, 15-minute 50th percentile travel time, and the facility type. This framework proposes Process B.1.6 "Calculate truck travel time reliability (TTTR)" as illustrated in Figure 3-46, to estimate 15-minute TTTR.

TTTR is expressed as the ratio of the 95th percentile travel time divided by the 50th percentile travel time for each segment and each time period. For a specific segment and time period, this is calculated as follows:

$$TTTR = \frac{\text{95th percentile travel time}}{\text{50th percentile "normal travel time"}} \quad \textbf{3-5}$$

Hence, using equation (3-5), the process output is TTR $_{TTTR,\ Facility}$ for every 15-minute at a specific segment.

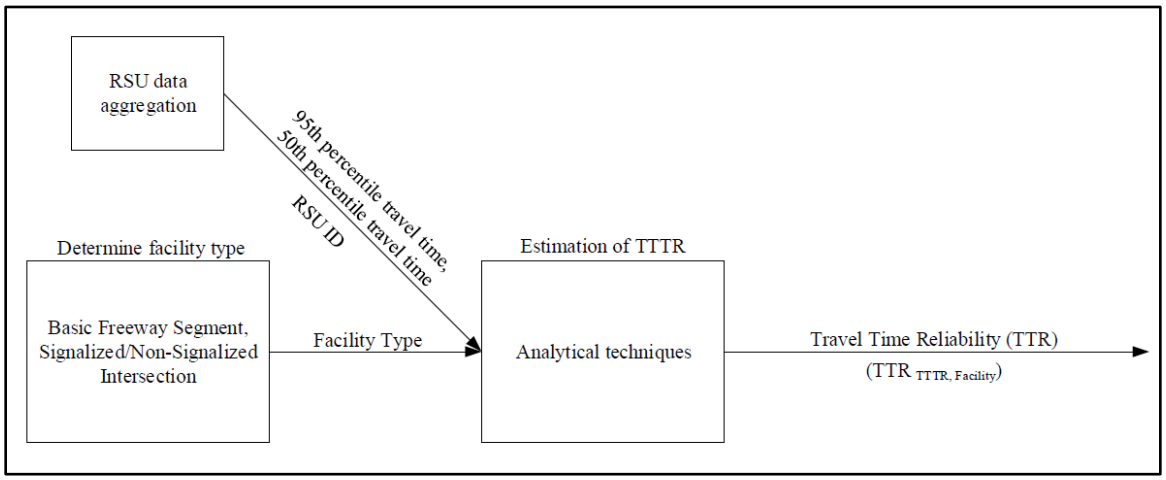


FIGURE 3-46: SCHEMATIC DESIGN OF PROCESS B.1.6 "CALCULATE TRUCK TRAVEL TIME RELIABILITY (TTTR)"

3.3.2.3.6F PROCESS B.1.7 – CALCULATE TRAVEL RATE INDEX (TRI)

According to Texas Transportation Institute's congestion measures in the urban mobility report, TRI also is an important parameter to measure travel time reliability (NCHRP, 2003). It requires several types of data inputs such as 15-minute average speed of freeway, 15-minute average speed of arterial, 15-minute segment free flow speed of freeway, 15-minute segment free flow speed of arterial, 15-minute vehicle volume of freeway, 15-minute vehicle volume of arterial, segment length of freeway, segment length of arterial, and RSU ID. This framework proposes Process B.1.7 "Calculate travel rate index (TRI)" to estimate 15-minute TRI as shown below in Figure 3-47.Generally, TRI indicates the amount of extra travel time. This process uses equation (3-6) to calculate 15-minute TRI for the specific segment. So, the process output is 15-minute TTR $_{TRI,\ Facility}$ for the specific segment.

$$TRI = \frac{\frac{\frac{60}{Speed_{Freeway}}}{\frac{60}{Freeflowspeed_{Freeway}}} \times VMT_{Freeway} + \frac{\frac{60}{Speed_{Arterial}}}{\frac{60}{Freeflowspeed_{Arterial}}} \times VMT_{Arterial}}{VMT_{Freeway} + VMT_{Arterial}} \quad \text{3-6}$$

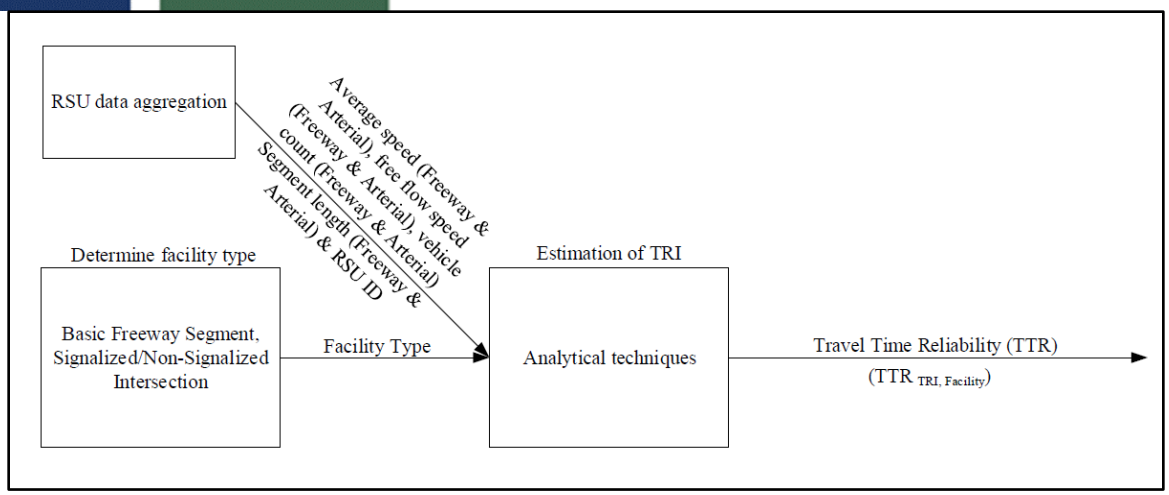


FIGURE 3-47: SCHEMATIC DESIGN OF PROCESS B.1.7 "CALCULATE TRAVEL RATE INDEX (TRI)"

3.3.2.3.6G PROCESS B.1.8 – CALCULATE TRAVEL TIME MISERY INDEX (TTMI)

TTMI measures the length of delay of only the worst trips. This framework proposes Process B.1.8 "Calculate travel time misery index (TTMI)" as shown below in Figure 3-48, to estimate 15-minute TTMI using data from RSU data aggregation and facility type from Process B.1.1 "Determine facility type." Specifically, the inputs required to estimate the 15-minute TTMI are: 15-minute average speed, RSU ID, and facility type.

In this process, the performance metric TTMI can be calculated by using equation (3-7). Hence, the process output is TTR $_{TTMI,\ Facility}$ for every 15-minute at a specific segment.

$$TTMI = \text{Average of the travel rates for the longest } 20\% \text{ of the trips} - \text{Average travel rates for all trips} \quad \textbf{3-7}$$

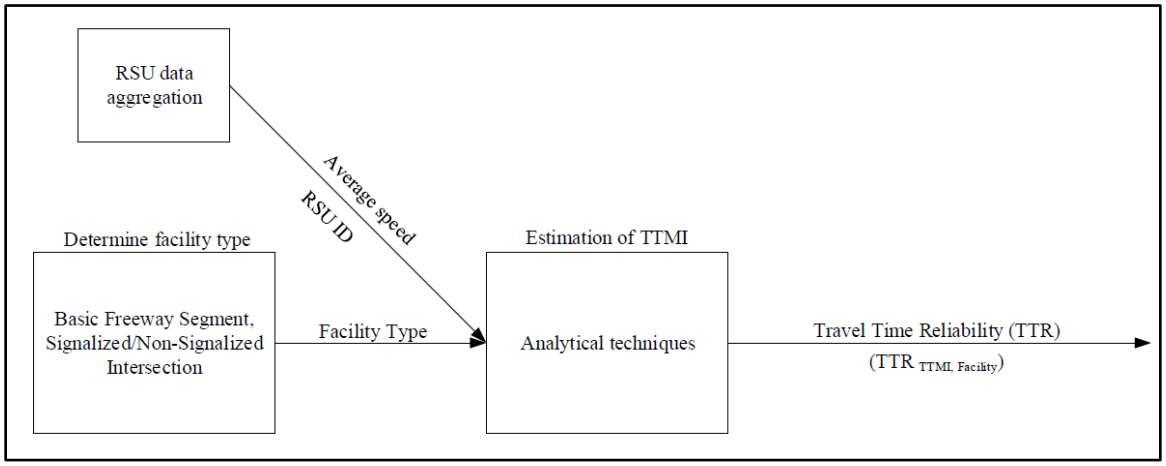


FIGURE 3-48: SCHEMATIC DESIGN OF PROCESS B.1.8 "CALCULATE TRAVEL TIME MISERY INDEX (TTMI)"

3.3.2.3.6H PROCESS B.1.9 – CALCULATE TRAVEL TIME PERCENT VARIATION (TTPV)

The travel time percent variation (TTPV) is another important parameter to measure travel time reliability. It is used to get a clear picture of the trend of travel time. Higher values of percent variation indicate less reliable travel time. This framework proposes Process B.1.9 "Calculate travel time percent variation (TTPV)" as illustrated in Figure 3-49, to calculate the 15-minute TTPV. This process requires four types of data: 15-minute average travel time, 15-minute standard deviation of travel time, RSU ID and facility type.

TTPV is expressed as the ratio of standard deviation of travel time and average travel time. This process uses equation (3-8) to estimate 15-minute TTPV for the specific segment. Thus, the process output is TTR $_{TTMI,\ Facility}$ for every 15-minute at a specific segment.

$$TTPV = \frac{\text{Standard deviation of travel time}}{\text{Average travel time}} \times 100\% \qquad \text{3-8}$$

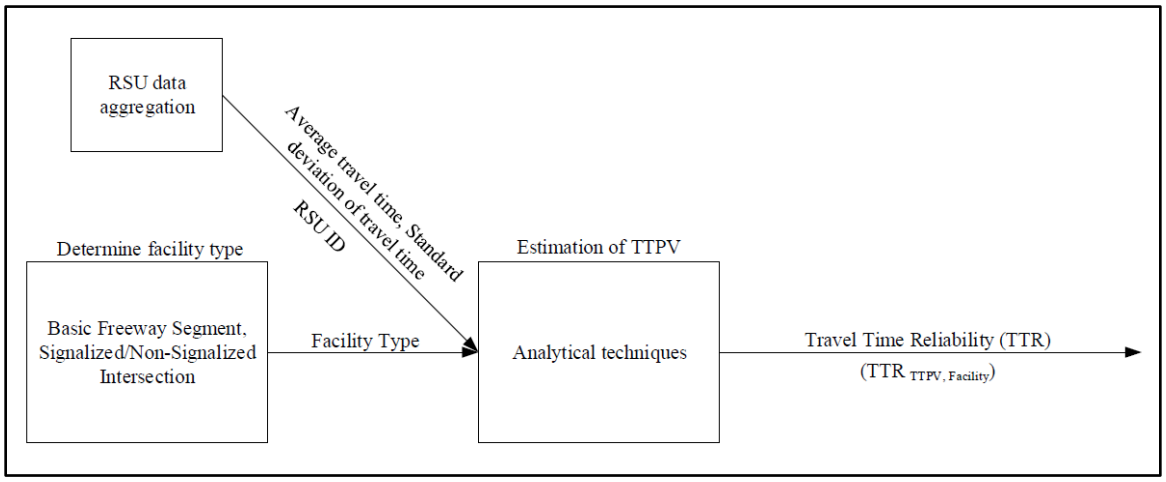


FIGURE 3-49: SCHEMATIC DESIGN OF PROCESS B.1.9 "CALCULATE TRAVEL TIME PERCENT VARIATION (TTPV)"

3.3.2.3.7 PROCESS GROUP B.2 – CONGESTION DEVELOPMENT MEASURES

Process Group B.2 "Congestion development measures" comprises 7 processes to estimate performance measures of the transportation system using aggregated data from RSUs data aggregation, to support advanced transportation management. These processes are speed normal deviate (SND); peak hour excessive delay (PHED); driver induced congestion; roadway congestion index; annual person hours of delay; percent congested travel; annual hours of truck delay (AHTD). The following subsections 3.3.2.3.7A through G illustrate the process designs for these eight processes.

3.3.2.3.7A PROCESS B.2.2 – CALCULATE SPEED NORMAL DEVIATE (SND)

The speed normal deviate (SND) is used to classify recurrent and non-recurrent congestion. This framework proposes Process B.2.2 "Calculate speed normal deviate (SND)" as shown below in Figure 3-50, to estimate 15-minute SND. The inputs of this process are: 15-minute average

speed, monthly average speed, 15-minute standard deviation of speed, RSU ID, and facility type.

SND represents the difference of the speed from its mean of the given data set, divided by the standard deviation of the speed of the data set. For a specific segment and time period, this is calculated as shown in equation (3-9):

$$SND = \frac{Speed - Average\ speed}{Standard\ deviation} \quad \text{3-9}$$

Using equation (3-9), this process output is CDMs $_{SND,\ Facility}$ for every 15-minute at a specific segment.

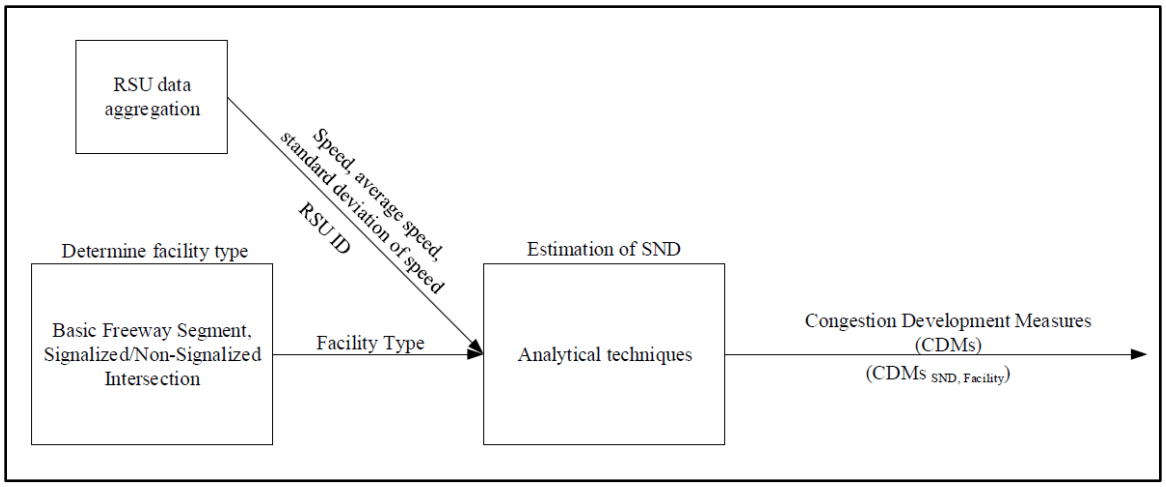


FIGURE 3-50: SCHEMATIC DESIGN OF PROCESS B.2.2 "CALCULATE SPEED NORMAL DEVIATE (SND)"

3.3.2.3.7B PROCESS B.2.3 – CALCULATE PEAK HOUR EXCESSIVE DELAY (PHED)

According to MAP-21, peak hour excessive delay (PHED) is used to assess traffic congestion to support the CMAQ program. PHED focuses on excessive delay experienced during peak hours in applicable urbanized areas. This framework proposes Process B.2.3 "Calculate peak hour excessive delay (PHED)" as illustrated in Figure 3-51, to estimate annual hours of peak hour excessive delay per capita. Several types of data are needed: segment length, hourly traffic volume, total population, AVO, RSU ID, and facility type.

This process uses the following steps to estimate PHED for the specific segment:

Step-01: Compute excessive delay threshold travel time, which can be calculated as shown in equation (3-10):

$$\textbf{Excessive delay threshold travel time}_p == \left(\frac{\textbf{Travel time segment } length_p}{\textbf{Threshold speed}}\right) \times 3600 \quad \text{3-10}$$

where, Travel time segment length = segment length for travel time; threshold speed = 35 mph for interstate and 15 mph for principal and other NHS roads; and p = segment number.

Step-02: Compute excessive delay, which can be estimated as follows:

$$\textbf{Excessive Delay}_{\text{p,a}} = \begin{cases} \text{or} \dfrac{RSD_{p,a} \text{ when } RSD_{p,a} \geq 0}{3600} \\ 0 \text{ when } \text{RSD}_{p,a} < 0 \end{cases} \quad \textbf{3-11}$$

Where, RSD (Reporting Segment Delay) = excessive delay threshold travel time$_p$ and RSD ≤ 900 seconds; and a = time interval, 15-minutes.

Step-03: Compute total excessive delay:

$$\textbf{Total excessive } \boldsymbol{Delay}_{p=} = \textbf{AVO} \times \sum_{\text{d=1}}^{\text{TD}} \left\{ \sum_{\text{h=1}}^{\text{TH}} \left[\sum_{\text{a=1}}^{\text{TB}} \left(\left[\textbf{Excessive } \boldsymbol{Delay}_{p,h,a,d} \times \left(\frac{\textbf{hourly volume}}{4} \right)_{\text{p,h,d}} \right)_{\text{a}} \right]_{\text{h}} \right\}_{\text{d}} \quad \textbf{3-12}$$

Where, TD = total number of days; d = day of the year; TH = total number of hour intervals; h = single hour interval; TB = total number of 15-minutes interval.

Step-04: Compute Peak Hour Excessive Delay (PHED) measure:

$$\boldsymbol{Annual\ Hours\ of\ Peak\ Hour\ Excessive\ Delay\ per\ Capita} = \frac{\sum_{p=1}^{T} \boldsymbol{Total\ Excessive\ Delay}_p}{\boldsymbol{Total\ Population}} \quad \textbf{3-13}$$

Hence, using equation (3-13), the process output is CDMs $_{\text{PHED, Facility}}$ for each year at a specific segment.

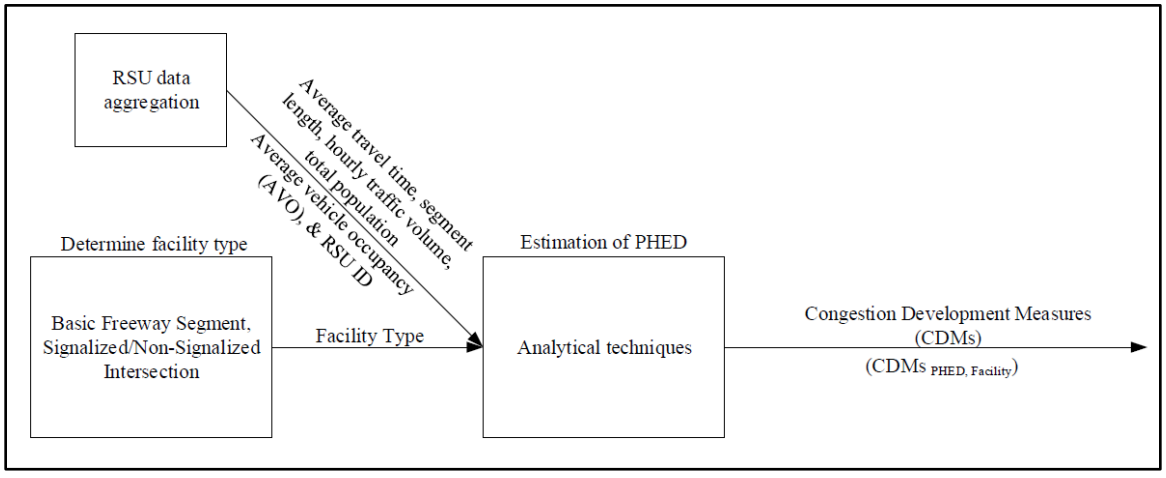


FIGURE 3-51: SCHEMATIC DESIGN OF PROCESS B.2.3 "CALCULATE PEAK HOUR EXCESSIVE DELAY (PHED)"

3.3.2.3.7C PROCESS B.2.4 – CALCULATE DRIVER INDUCED CONGESTION

Nowadays driving behavior is considered in the estimation of traffic congestion. This framework proposes Process B.2.4 "Calculate driver induced congestion" as shown below in Figure 3-52, to estimate 15-minute driver induced congestion. In doing so, this process utilizes 15-minute average time headway, and RSU ID from RSUs data aggregation, and facility type from Process B.1.1 "Determine facility type."

This process assumes that, if the time headway between successive vehicles is less than 3 sec/veh, then driver induced congestion can happen. So, an alarm message will display through CMS to warn all drivers not to follow closely.

Hence, the process output indicates, $CDMs_{Driver\ induced\ congestion,\ Facility}$ or no driver induced congestion for every 15-minute at a specific segment.

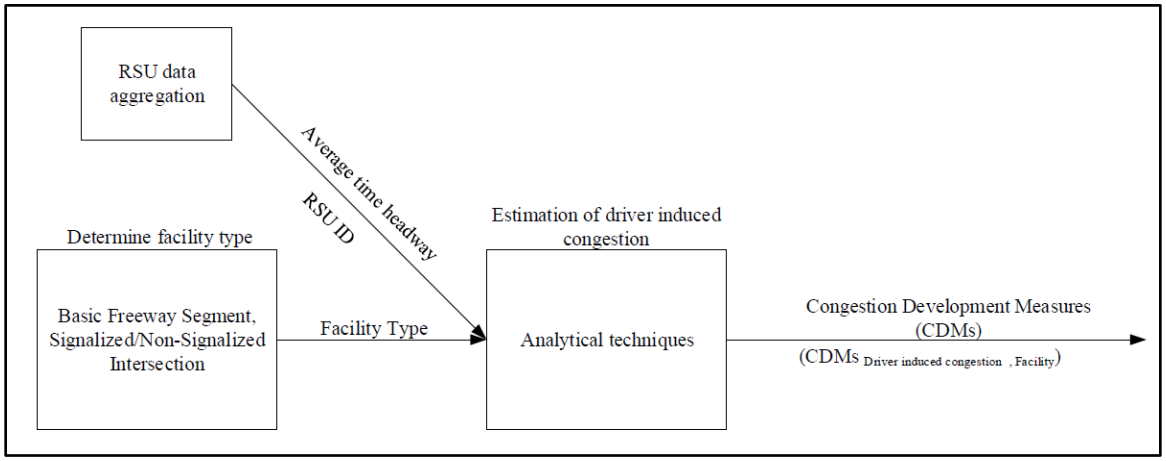


FIGURE 3-52: SCHEMATIC DESIGN OF PROCESS B.2.4 "CALCULATE DRIVER INDUCED CONGESTION"

3.3.2.3.7D PROCESS B.2.5 – CALCULATE ROADWAY CONGESTION INDEX (RCI)

Roadway congestion index is used to quantify congestion by focusing on daily vehicle miles traveled on both freeway and arterial roads. This Process B.2.5 "Calculate roadway congestion index" proposes a framework to estimate 15-minute roadway congestion index, which is shown in Figure 3-53. The input of this process are: 15-minute vehicle count of freeway, 15-minute vehicle count of arterial, segment length of freeway, segment length of arterial, lane width for both freeway and arterial, and RSU ID.

Roadway congestion index is expressed as cars per road space. For a specific segment and time period, this is calculated as follows:

$$Roadway\ congestion\ index\ (RCI) == \frac{\frac{VMT_{Freeway}}{Lane-mile_{Freeway}} \times VMT_{Freeway} + \frac{VMT_{Arterial}}{Lane-mile_{Arterial}} \times VMT_{Arterial}}{13{,}000 \times VMT_{Freeway} + 5{,}000 \times VMT_{Arterial}} \quad \text{3-14}$$

Thus, using equation (3-14), the process output is CDMs $_{\text{Roadway congestion index, Facility}}$ for every 15-minute at a specific segment.

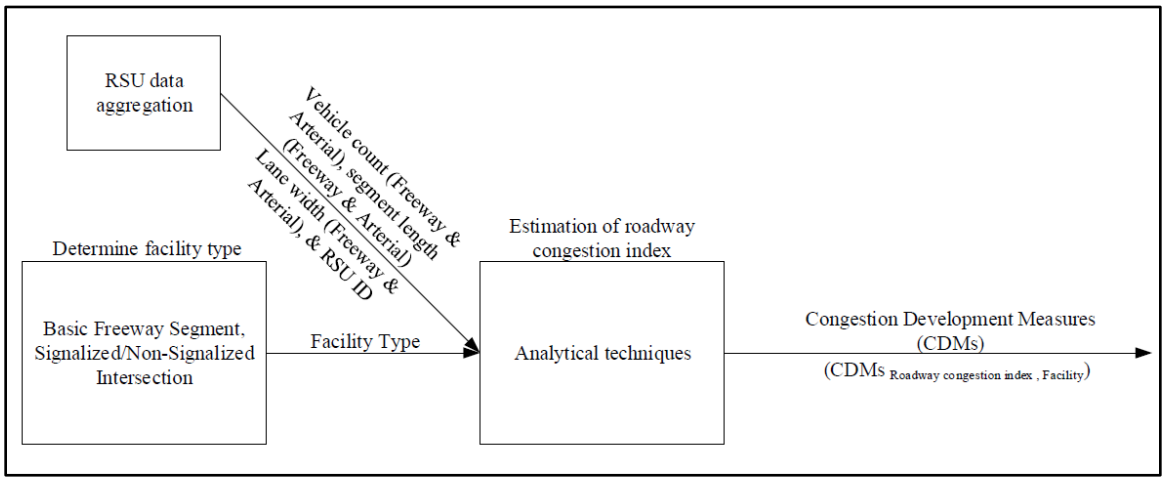


FIGURE 3-53: SCHEMATIC DESIGN OF PROCESS B.2.5 "CALCULATE ROADWAY CONGESTION INDEX"

3.3.2.3.7E PROCESS B.2.6 – CALCULATE ANNUAL PERSON HOURS OF DELAY

Annual person hours of delay is an indicator of mobility performance of a roadway. This framework proposes Process B.2.6 "Calculate annual person hours of delay" as illustrated in Figure 3-54, to estimate annual person hours of delay. Input data of this process are: 15-minute average travel time, 15-minute segment free flow travel time, RSU ID, and facility type.

For a specific segment and time period, annual person hours of delay is calculated as follows:

$$\boldsymbol{Annual\ person\ hours\ of\ delay} = = \boldsymbol{Daily\ vehicle\ hours\ of\ delay} \times \boldsymbol{250\ working\ days\ per\ year} \times \boldsymbol{1.25\ persons\ per\ vehicle} \quad \textbf{3-15}$$

The process output using equation (3-15) is CDMs $_{\text{Annual person hours of delay, Facility}}$ for each year at a specific segment.

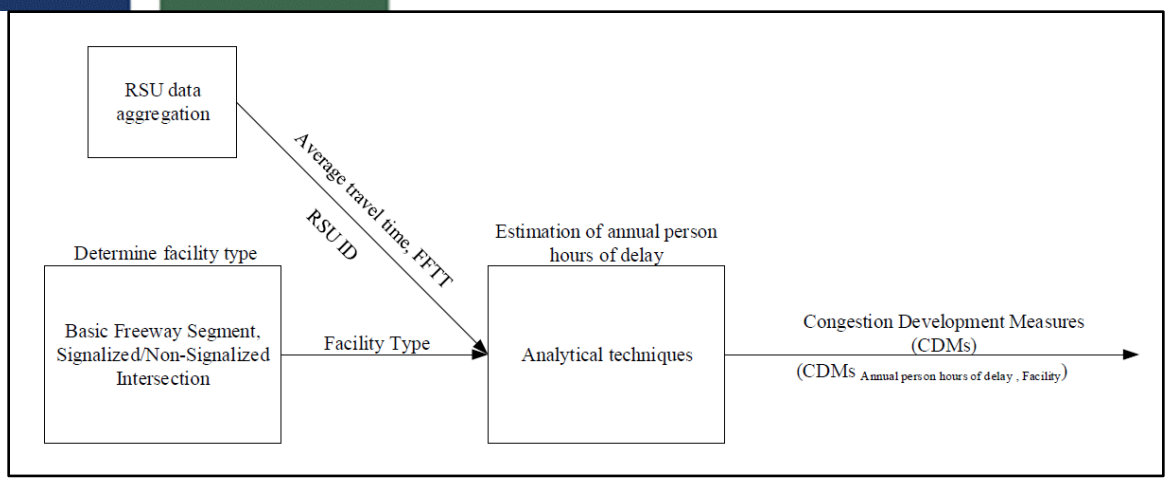


FIGURE 3-54: SCHEMATIC DESIGN OF PROCESS B.2.6 "CALCULATE ANNUAL PERSON HOURS OF DELAY"

3.3.2.3.7F PROCESS B.2.7 – CALCULATE PERCENT CONGESTED TRAVEL

The percent of congested travel measures the extent of congestion. This framework proposes Process B.2.7 "Calculate percent congested travel" to estimate 15-minute percent congested travel, which is shown below in Figure 3-55. The input data of this process are 15-minute vehicle count, segment length, RSU ID, and facility type. This process uses equation (3-16) to calculate 15-minute percent congested travel for the specific segment. Hence, the output of this process is CDMs $_{Percent\ congested\ travel,\ Facility}$ for every 15-minutes at a specific segment.

$$Percent\ congested\ travel = \frac{VMT\ under\ congested\ conditions}{Total\ VMT\ for\ the\ segment} \qquad 3\text{-}16$$

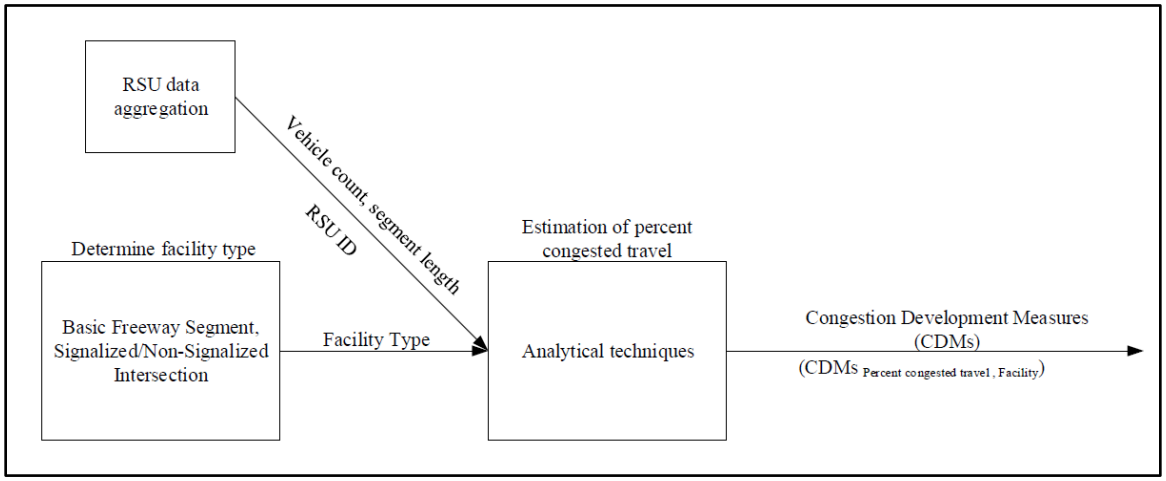


FIGURE 3-55: SCHEMATIC DESIGN OF PROCESS B.2.7 "CALCULATE PERCENT CONGESTED TRAVEL"

3.3.2.3.7G PROCESS B.2.8 – CALCULATE ANNUAL HOURS OF TRUCK DELAY (AHTD)

Annual hours of truck delay (AHTD) is the amount of extra travel time spent by each truck due to congestion. Several types of data are needed to estimate AHTD as shown in Figure 3-56.

These data are 15-minute average speed, 15-minute truck volumes, segment length, agency specified threshold speed or PSL, RSU ID, and facility type.

For a specific segment, AHTD is calculated as follows:

$$AHTD = \sum \left(\frac{Freight\,VMT}{Travel\,speed} - \frac{Freight\,VMT}{Agency\,specified\,threshold\,speed} \right) \times 7 \times \times$$

52 3-17

Hence, using equation (3-17), the output is CDMs $_{AHTD,\,Facility}$ for each year at a specific segment.

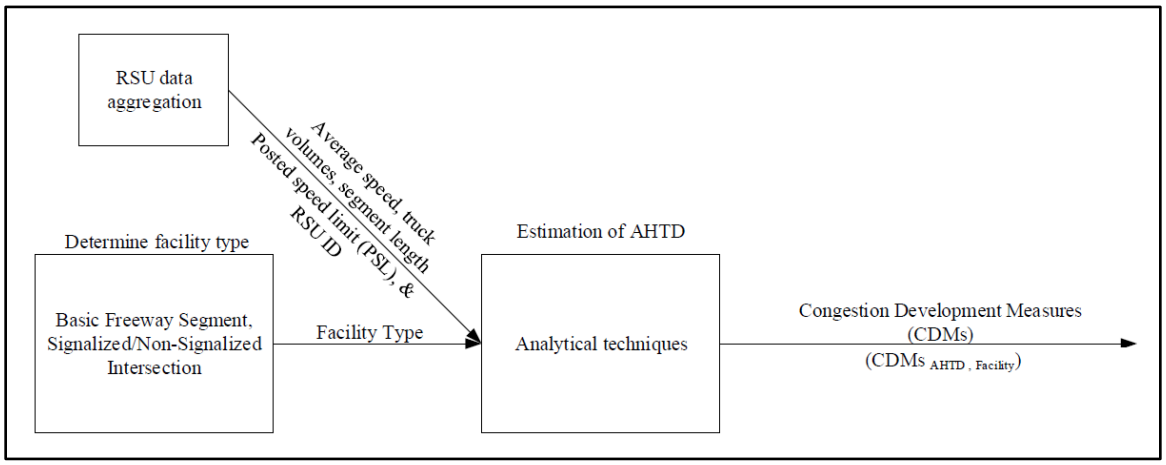


FIGURE 3-56: SCHEMATIC DESIGN OF PROCESS B.2.8 "CALCULATE ANNUAL HOURS OF TRUCK DELAY (AHTD)"

3.3.2.3.8 PROCESS GROUP B.3 – LEVEL OF SERVICE (LOS)

Level of Service (LOS) is a qualitative measure of traffic describing the operating conditions of a segment or facility. According to the Highway Capacity Manual (HCM, 2016), there are six levels of LOS ranging from level A to level F. Level A represents the best quality of traffic whereas level F represents the worst quality of traffic. This Process Group B.3 "Level of service" contains two processes to estimate LOS of the specific segment on the basis of density and control delay. The following subsections explain the process designs for these two processes.

3.3.2.3.8A PROCESS B.3.2 – CALCULATE DENSITY

Density is an important parameter to quantify LOS of the freeway segments. The density of any segment for freeways is the number of vehicles within the segment. This framework proposes Process B.3.2 "Calculate density" as shown in Figure 3-57, to identify LOS. The input data of this process are 15-minute average speed, 15-minute average time headway, number of lanes, RSU ID, and facility type. For a specific segment and time period, density is calculated based on flow rate (see equation (3-18)), average speed and number of lanes as shown in equation (3-19):

$$Flow\,rate = \frac{3600\ (s/h)}{Average\ time\ headway\ (s/veh)} \qquad 3\text{-}18$$

$$Density = \frac{\text{Flow rate (veh/h)}}{\text{Average speed (mile/h)} \times \text{Number of Lane}} \quad 3\text{-}19$$

The above-calculated density by using equation (3-19) can be used to identify LOS of the basic freeway segment using the criteria illustrated in Exhibit 12-15 of the HCM (HCM, 2016). Hence, the output of this process is LOS $_{Facility}$ for the specific segment.

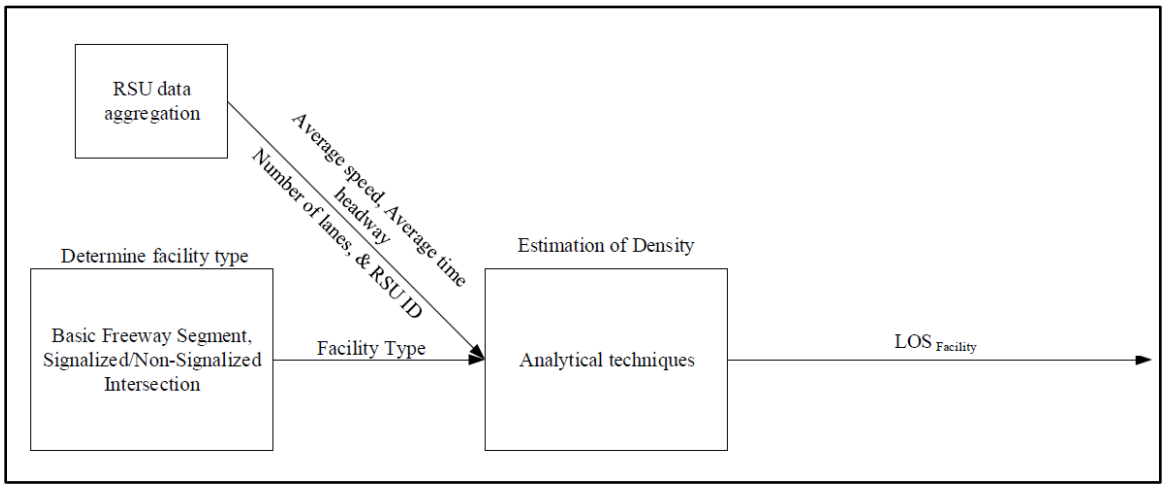


FIGURE 3-57: SCHEMATIC DESIGN OF PROCESS B.3.2 "CALCULATE DENSITY"

3.3.2.3.8B PROCESS B.3.3 – CALCULATE CONTROL DELAY (CD)

Control delay (CD) is an important parameter to measure the performance and identify LOS for the signalized intersection. This framework proposes Process B.3.3 "Calculate control delay" as illustrated in Figure 3-58, to calculate LOS at signalized intersections. The input data of this process are 15-minute effective green time, 15-minute cycle length, hourly traffic counts, RSU ID, and facility type. According to HCM (2016), control delay (CD) can be calculated as follows:

$$CD = d_1 \times (PF) + d_2 \quad 3\text{-}20$$

$$PF = f_{PA}\left(1 - \left[R_p \times \frac{g}{C}\right]\right) / \left(1 - \frac{g}{C}\right) \quad 3\text{-}21$$

$$If\ 0.50 < R_p \leq 0.85, then\ f_{PA} = 0.93 \quad 3\text{-}22$$

$$If\ 1.15 < R_p \leq 1.50, then\ f_{PA} = 1.15 \quad 3\text{-}23$$

$$Else\ f_{PA} = 1.00 \quad 3\text{-}24$$

$$d_1 = \frac{0.5C\left(1-\frac{g}{C}\right)^2}{1-[min(1,X)g/C]} \quad 3\text{-}25$$

$$d_2 = 900T\left[(X-1) + \sqrt{(X-1)^2 + \frac{4IX}{cT}}\right] \quad 3\text{-}26$$

$$I = 1.0 - 0.91\,X_u^{2.68} \geq 0.090 \quad 3\text{-}27$$

Where, d_1 = uniform delay (s/veh), d_2 = incremental delay (s/veh), PF is the progression adjustment factor, f_{PA} = adjustment factor for platoons arriving during green, R_p = platoon ratio, g = effective green time (s), C = cycle length (s), X = volume to capacity ratio, c = capacity (veh/h), T = analysis period duration (h), I = upstream filtering adjustment factor, and X_u = volume to capacity ratio of the upstream movements.

The above-calculated CD can be used to identify LOS of the signalized intersection using LOS criteria that are illustrated in Exhibit 19-8 of the HCM (HCM, 2016). Thus the process output is $LOS_{Facility}$ for every 15-minutes at a specific intersection.

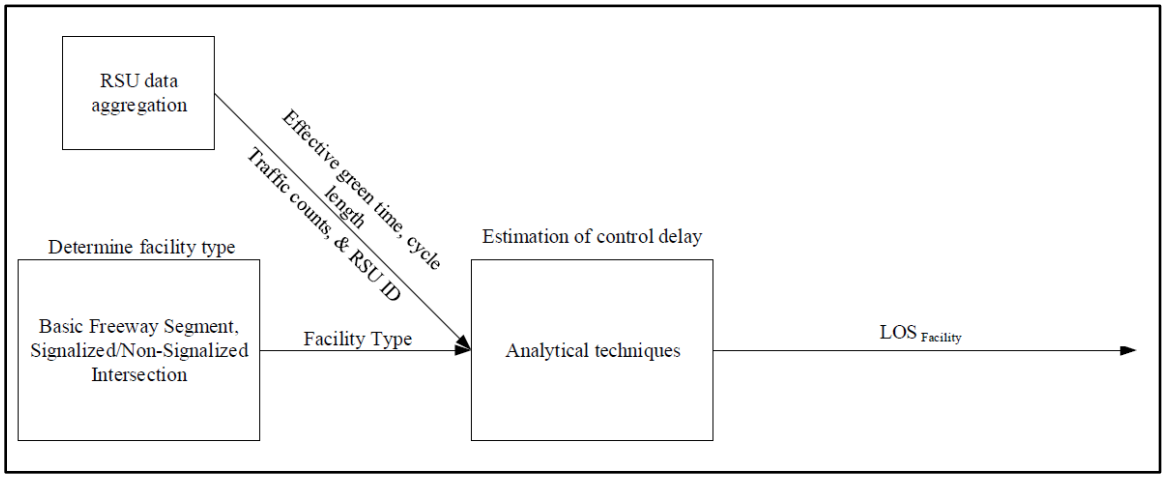


FIGURE 3-58: SCHEMATIC DESIGN OF PROCESS B.3.3 "CALCULATE CONTROL DELAY"

3.3.2.3.9 Process Group B.4 – Environmental Impact

Surface transportation has considerable impact on environment. This process group contains three processes to calculate some vital parameters of the environment in order reduce the impact on environment. These are rate of emission measures such as HC, CO, and NOx, rate of fuel consumption, and intensity of noise. The following subsections a through c illustrate the process designs for these three processes.

3.3.2.3.9A PROCESS B.4.2 – ESTIMATE EMISSION MEASURES

Vehicle emissions have considerable impacts on the environment. Research confirms that vehicle emissions are responsible for 45% of the pollutants in the environment in the United States (Ahn et al., 2002). This framework proposes Process B.4.2 "Estimate emission measures" as illustrated in Figure 3-59, to calculate 15-minute emission measures. The inputs required to calculate emission measures are segment length, 15-minute vehicle count, 15-minute average speed, 15-minute average longitudinal acceleration, RSU ID, and facility type.

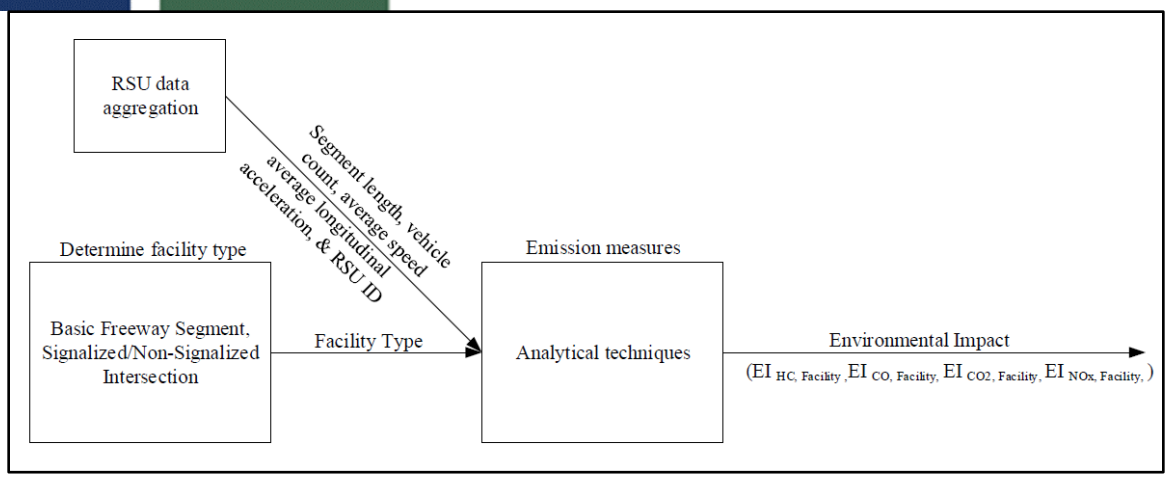


FIGURE 3-59: SCHEMATIC DESIGN OF PROCESS B.4.2 "ESTIMATE EMISSION MEASURES SUCH AS HC, CO, AND NOX"

There are several techniques to estimate rate of vehicle emissions, including the vehicle miles of travel (VMT) approach, vehicle miles of travel at specified speeds or speed ranges (VMT-S), driving mode approach, vehicle specific power (VSP), and Virginia Tech Microscopic energy and emission model (VT-Micro) (Frey et al., 2002; Rakha et al., 2003; Rouphail, 2013). Some details are provided below.

VMT approach. According to Rouphail (2013),

$$Emission\ rate, E = A \times E_f \qquad 3\text{-}28$$

Where, A = vehicle activity expressed in vehicle miles traveled by a vehicle; and E_f = emission factor. Emission factors (E_f) of hydrocarbon (HC), carbon monoxide (CO), and nitrogen oxides (NOx) are 1.25, 12.57, and 0.92g/mile respectively (Rouphail, 2013).

VMT-S approach. The rate of vehicle emission can be calculated using equation (3-27). However here, A = annual mileage; and the emission factor (E_f) is based on vehicle speed. An example of calculating emission factor of CO emissions is shown in Table 3-8.

TABLE 3-8: EMISSION FACTORS OF CO EMISSIONS, (SOURCE: ROUPHAIL (2013))

Speed range (mph)	≤10	10 – 30	30 – 50	>50
Corresponding E_f (g/mile)	18	9	10	11

Driving mode approach. Equation (3-27) can also be used to estimate emission rate for four types of driving modes such as acceleration, deceleration, idle, and cruise specified in the Frey study (Rouphail, 2013). Here, vehicle activity (A) equals to travel time spent in each of the four modes. Emission factor (E_f) can be computed from Figure 3-60.

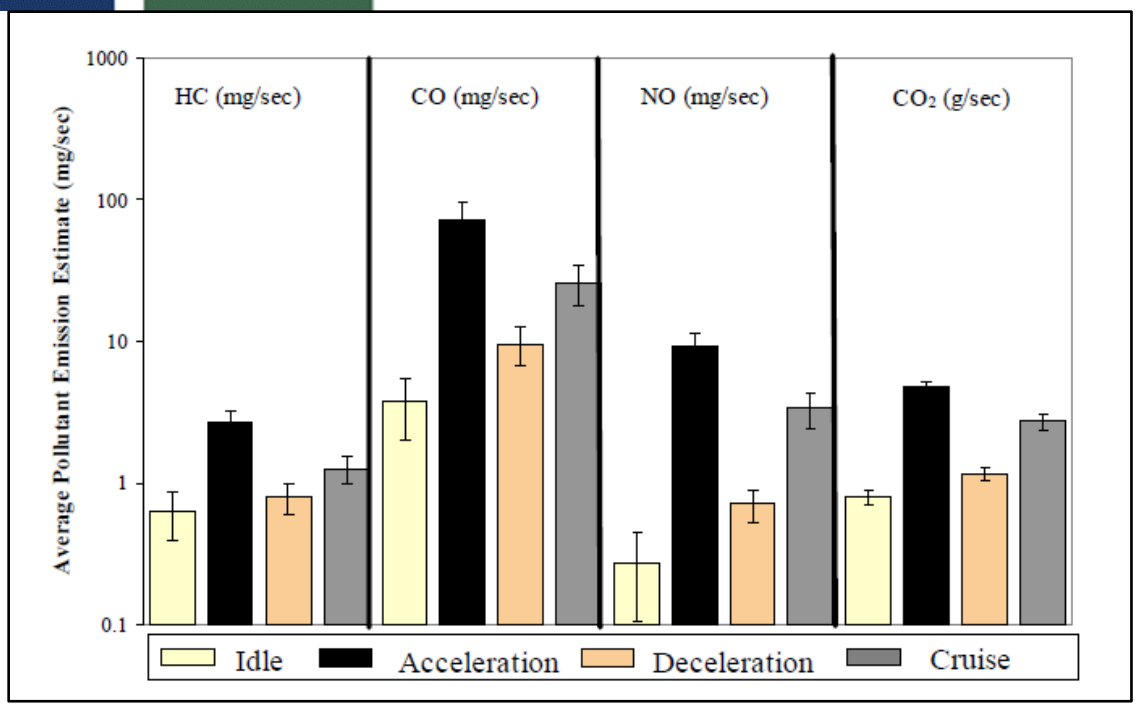


FIGURE 3-60: EMISSION FACTORS BY DRIVING MODE, ADAPTED FROM (FREY ET AL., 2002)

VSP approach. Vehicle Specific Power (VSP) distribution is a good technique to estimate emission measures. According to (Frey et al., 2002), VSP for a light-duty vehicle can be estimated using an equation given by EPA, which is:

$$VSP = v[1.1a + 9.81(\text{a tan}(\sin(grade))) + 0.132] + 0.000302v^3 \quad \textbf{3-29}$$

Ignoring the grade of the road, the equation of VSP for light duty vehicle becomes:

$$VSP = v(1.1a + 0.132) + 0.000302v^3 \quad \textbf{3-30}$$

where, VSP = vehicle specific power (kW/ton); v = instantaneous speed (m/s); and a = longitudinal acceleration (m/s^2).

After estimating VSP using equation (3-29), the VSP mode of each vehicle can be computed using Figure 3-61.

VSP Mode	Bin Range	VSP Mode	Bin Range
1	$VSP < -2$	2	$-2 \leq VSP < 0$
3	$0 \leq VSP < 1$	4	$1 \leq VSP < 4$
5	$4 \leq VSP < 7$	6	$7 \leq VSP < 10$
7	$10 \leq VSP < 13$	8	$13 \leq VSP < 16$
9	$16 \leq VSP < 19$	10	$19 \leq VSP < 23$
11	$23 \leq VSP < 28$	12	$28 \leq VSP < 33$
13	$33 \leq VSP < 39$	14	$VSP \geq 39$

FIGURE 3-61: VSP MODAL DEFINITIONS, ADAPTED FROM (ROUPHAIL, 2013)

Mathematically, the emission estimation method is expressed as follows:

$$E_{j,k,N} = \sum_{t=1}^{T_j} B_{itj} \times EF_i \quad \textbf{3-31}$$

Where, $E_{j,k,N}$ = emissions (k) of vehicle ID, N traveling on segment (j), in g/vehicle; B_{itj}= 1 if VSP mode (i) occurs in time step (t) on segment (j), zero otherwise;

EF_i = emission factor associated with VSP mode (i) which can be calculated from Figure 3-62;

T_j = average travel time on segment j (computed as segment (j) length over average segment travel speed);

i = 1, 2, …, 14;

j = 1, 2, 3, …, j;

N = vehicle ID; and

k = CO, HC, CO2, NOx.

Emissions for a segment can be calculated from equation (3-31), if the length of segment and traffic count are known:

$$EF_{segment,k,Facility} = \frac{\sum E_{j,k,N} * segment\ length\ (D_j)}{segment\ length\ (D_j)} * Traffic\ Count \quad \textbf{3-32}$$

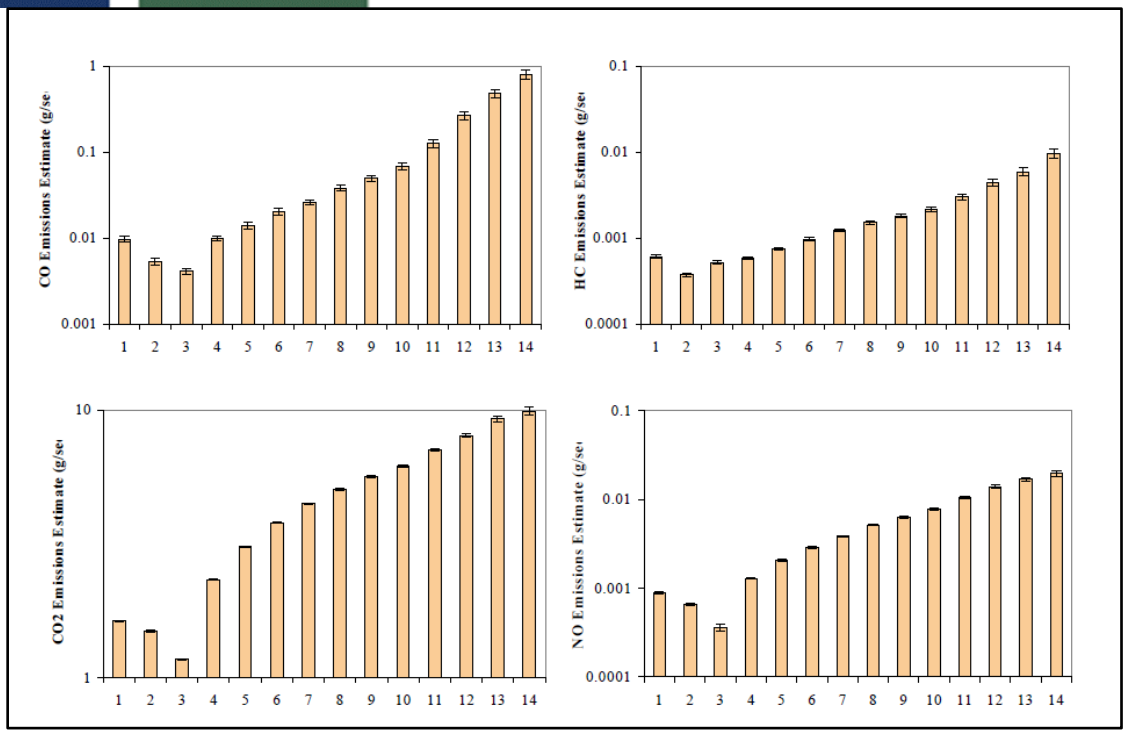


FIGURE 3-62: AVERAGE MODAL EMISSION RATES (G/SEC) FOR VSP BINS FOR CO, HC, CO_2, AND NOX, ADAPTED FROM (FREY ET AL., 2002)

VT-Micro model. This model comprised speed and acceleration data collected by the Oak Ridge National Laboratory (ORNL) to estimate emission rate (Antoine G. Hobeika et al., 2015; Rakha et al., 2003). Mathematically, emission rate is expressed as shown in equation (3-32) (Rakha et al., 2003):

$$MOE_e = e^{\sum_{i=0}^{3}\sum_{j=0}^{3}\left(K_{i,j}^{e}\times v^{i}\times a^{j}\right)} \quad \textbf{3-33}$$

Where, MOE_e is the instantaneous fuel consumption or emission rate (ml/s or mg/s); $K^e_{i,j}$ is model regression coefficient for MOE ''*e*'' at speed power ''*i*'' and acceleration power ''*j*''; v is the instantaneous speed (km/h); a is the instantaneous acceleration (km/h/s); I equals to the power of the speed (i.e., v, v^2, v^3); and *j* equals to the power to acceleration (i.e., *a*, a^2, a^3).

Most of the past and recent studies mentioned that second-by-second data such as vehicle speed, and acceleration helps to estimate emission rate accurately. Among the above mentioned approaches, only VSP approach and VT-Micro model comprises both data sets. However, VT-Micro model is not applicable beyond the boundaries of vehicle speed and acceleration that were used in calibration. Hence, the VSP approach has been selected for this study to estimate emission rate.

Using Eq. (31), the process outputs are $EI_{HC,\ Facility}$, $EI_{CO,\ Facility}$, $EI_{CO2,\ Facility}$, $EI_{Nox,\ Facility}$ for every 15-minutes at a specific segment.

3.3.2.3.9B PROCESS B.4.3 – ESTIMATE FUEL CONSUMPTION

Fuel consumption rate is related to emission rates. It has also considerable impacts on environments. In order to calculate fuel consumption rate, this framework proposes process B.4.3 "Estimate fuel consumption" which is shown below in Figure 3-63. The required data of this process are segment length, 15-minute vehicle count, 15-minute average speed, 15-minute average longitudinal acceleration, RSU ID, and facility type.

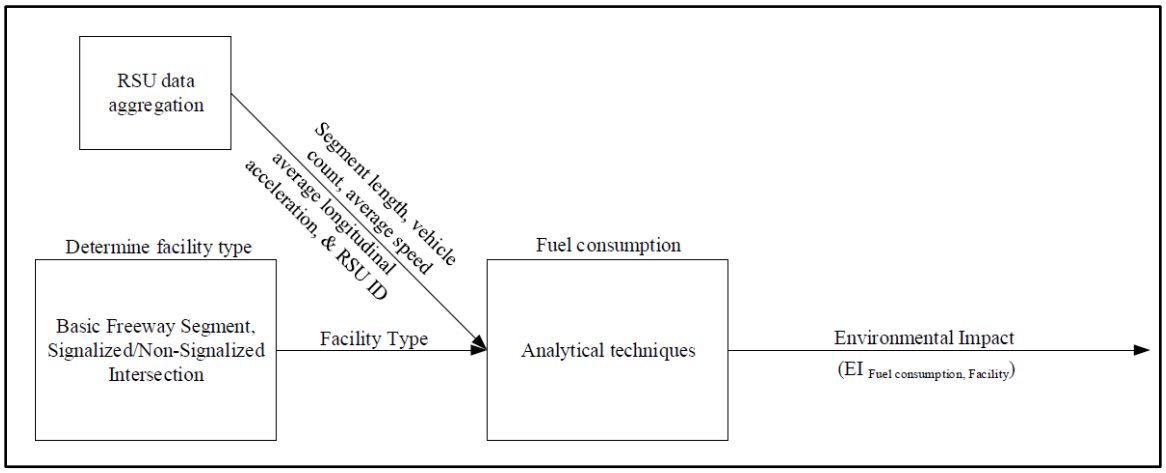


FIGURE 3-63: SCHEMATIC DESIGN OF PROCESS B.4.3 "ESTIMATE FUEL CONSUMPTION"

For a specific segment, fuel consumption can be calculated by using the following two steps:

Step-01: Compute emission rate of CO_2 (EI_{CO2}), CO (EI_{CO}), and HC (EI_{HC}) using equation (3-31) which is mentioned in process B.4.2, and

Step-02: Estimate fuel consumption using carbon balance method (Song et al., 2009) as shown in equation (3-33)

$$\boldsymbol{Fuel\ consumption} = \left(\boldsymbol{EI_{CO2}} \times \frac{\mathbf{12}}{\mathbf{44}} + \boldsymbol{EI_{CO}} \times \frac{\mathbf{12}}{\mathbf{28}} + \boldsymbol{EI_{HC}} \times \frac{\mathbf{12}}{\mathbf{13}}\right) \times \frac{\mathbf{1}}{\%\boldsymbol{C}} \qquad \textbf{3-34}$$

Where, %C equals to the percentage of gasoline by weight. The typical value of %C is 86.4% (Song et al., 2009).

The final output of this process by using Eq. (33) is $EI_{\text{Fuel consumption, Facility}}$ for every 15-minute at a specific segment.

3.3.2.3.9C PROCESS B.4.4 – ESTIMATION OF NOISE LEVEL

Surface transportation is one of the major sources of noise pollution. It is very important to reduce the impact of noise pollution using noise mapping. In order to calculate the automobile related noise level, process B.4.4 proposes a framework which is shown in Figure 3-64. This process uses 15-minute A-weighted sound level, RSU ID, and facility type in the estimation of noise level.

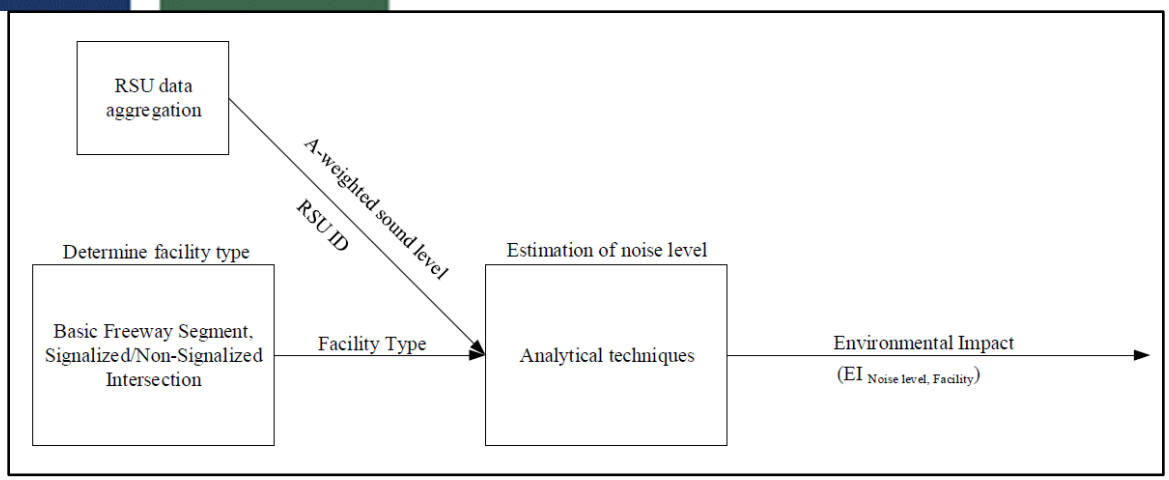


FIGURE 3-64: SCHEMATIC DESIGN OF PROCESS B.4.4 "ESTIMATION OF NOISE LEVEL"

There are several techniques regarding the estimation of the noise level. FHWA developed the Traffic Noise Model (TNM) to model for highway traffic noise (Yi et al., 2013). However, the intensity of noise can also be calculated based on field data. To quantify road traffic noise, four traffic noise metrics are required (de Kluijver et al., 2003). For a specific segment and time period, these are calculated as follows:

- Day average sound level,

$$L_d = \frac{1}{N}\sum_{T=i}^{j} L_{Aeq} \quad \textbf{3-35}$$

Where Ld = day average sound level; N = number of data, T = time period from i (07:00 AM local time) to j (07:00 PM local time), LAeq = A-weighted sound level.

- Evening average sound level,

$$L_e = \frac{1}{N}\sum_{T=k}^{h} L_{Aeq} \quad \textbf{3-36}$$

Where L_e = evening average sound level; N = number of data, T = time period from k (07:00 PM local time) to h (11:00 PM local time), L_{Aeq} = A-weighted sound level.

- Night average sound level,

$$L_n = \frac{1}{N}\sum_{T=r}^{s} L_{Aeq} \quad \textbf{3-37}$$

Where L_n = night average sound level; N = number of data, T = time period from r (11:00 PM local time) to s (07:00 AM local time), L_{Aeq} = A-weighted sound level.

- Day-evening-night average sound level,

$$L_{den} = 10log\frac{1}{24}\left(12 \times 10^{0.1 \times L_D} + 4 \times 10^{0.1 \times (L_e+5)} + 8 \times 10^{0.1 \times (L_n+10)}\right) \quad \textbf{3-38}$$

Hence, the process output by using equation (3-37) is $EI_{\text{Noise level, Facility}}$ for everyday at a specific segment.

3.3.3 Logical data flow diagram

This process is the combination of processes and process groups hierarchical diagram and individual process designs. These are illustrated by Figure 3-65 for a quick and convenient reference.

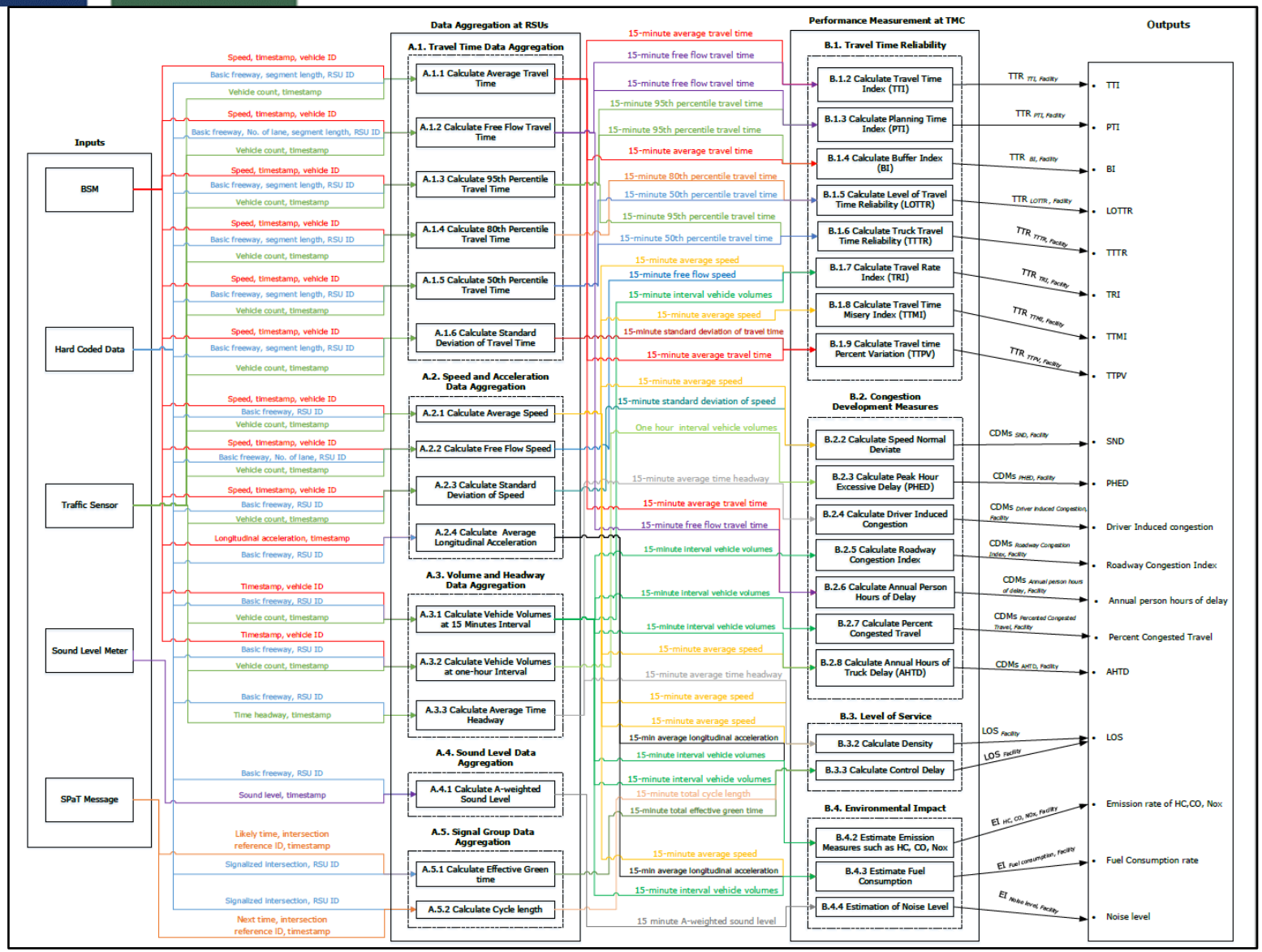


FIGURE 3-65: LOGICAL DATA FLOW DIAGRAM

3.4 PROOF OF CONCEPT STUDY

3.4.1 Introduction

This chapter presents a description of a case study, which is used as a proof of concept for this study. The case study uses microsimulation modeling to generate CV data for a study corridor and then applies appropriate procedures described in the methodological framework to calculate performance measures. A comparison of the performance measures generated from the simulation model to those obtained using data from traditional data sources is performed to validate the proposed framework.

More specifically, the following sections provide background information about the study corridor and the VISSIM simulation model used in the case study. Moreover, this chapter discusses VISSIM output data calibration using traditional data source, and the VISSIM output trajectory data conversion to BSMs using the Trajectory Conversion Algorithm (TCA) tool. A summary of results of the selected performance measures using CV data and traditional data sources is provided along with a comparison, which serves as a proof of concept of the proposed framework.

3.4.2 Study Corridor Location

A section of I-65 in the Birmingham, AL region is chosen as the study corridor for the purpose of this study. The study corridor is almost 14.40 miles long, extending from exit 247 to exit 261A as shown in the **Error! Reference source not found.**.

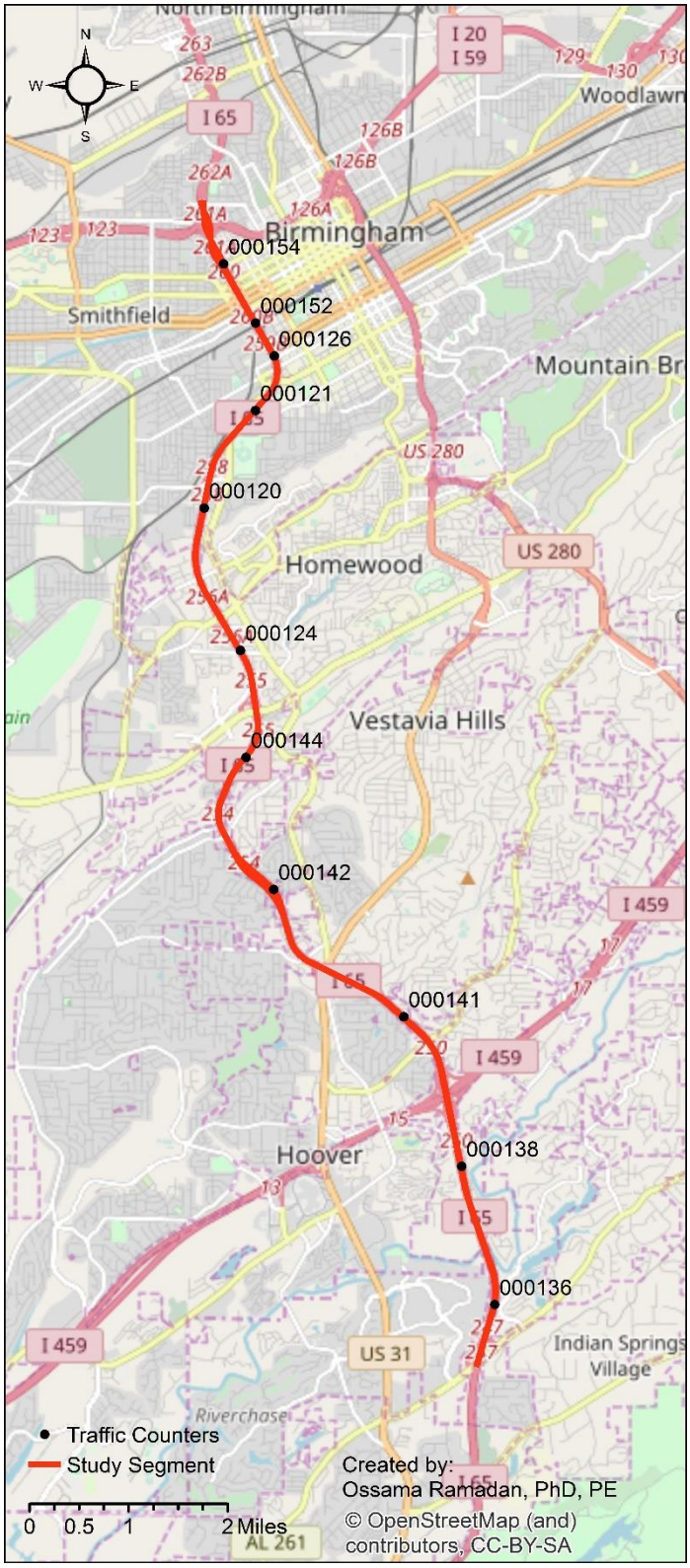


FIGURE 3-66: STUDY CORRIDOR LOCATION

The study corridor contains 53 Traffic Message Channels (TMCs) and was divided into four major segments. The main attributes of those segments are illustrated in Table 3-9. The study segment typically has three 12-ft lanes per mainline direction, with auxiliary lanes added at ramps locations. The posted speed limit on this study interstate is generally 60 mph with an advisory speed limit of 45 mph on the ramps.

TABLE 3-9: STUDY CORRIDOR ATTRIBUTES

Segment Name	Road Number	Travel Direction	TMC Count	Length (miles)
I20/59 to I459	I-65	Southbound	22	11.11
I459 to Valleydale Road		Southbound	4	3.24
I459 to I20/59		Northbound	23	11.23
Valleydale Road to I459		Northbound	4	3.21

3.4.3 Development of a simulation model

The microscopic simulation platform VISSIM 10.00 is used to build a simulation model along the study corridor ("PTV VISSIM 10," 2018). VISSIM is the microscopic stochastic traffic simulator that is mostly used as a tool for the design of urban public transportation systems, but has shown capabilities of reproducing freeway traffic behaviors as well (Gomes et al., 2004). VISSIM is a time-step and behavior-based simulation model, developed to model urban traffic by “Planung Transport Verkehr AG” of Karlsruhe, Germany based on the work of Wiedemann (Wiedemann, 1974, 1991). Along with CORSIM, VISSIM is one of the most popular traffic simulation software that is widely accepted by transportation professional around the globe and appropriate to use for the purposes of this study. In this specific application, the model of a study corridor is run under normal traffic conditions to get vehicle record data that are then used as an input to generate BSMs using Trajectory Conversion Algorithm (TCA) tool. This simulation model of the study corridor contains a total of 132 links. Description of each link is illustrated in Appendix F. The model was run for one hour from 8:00 AM to 9:00 AM using traffic volumes obtained from the Alabama Department of Transportation. The output data is collected from vehicle record output, which is a .fzp file containing vehicle speed, vehicle number, link number, lane index, acceleration, simulation second, and time of the day. The study run the simulation model three times and averaged the output data to calibrate and estimate the performance measures.

3.4.3.1 VISSIM Output Data Calibration

The VISSIM output data is calibrated using field measurement data set available through National Performance Management Research Data Set (NPMRDS) database. The study used two control limits to calibrate the VISSIM speed data outputs, namely the upper control limit (UCL) which is 15% more than the field measurement value, and lower control limit (LCL) which is less than the 15% of field measurement value. Analysis of results showed that VISSIM output speed data is within the upper control and lower control limit for all study segments as illustrated in Figure 3-67, Figure 3-68, Figure 3-69, Figure 3-70 respectively.

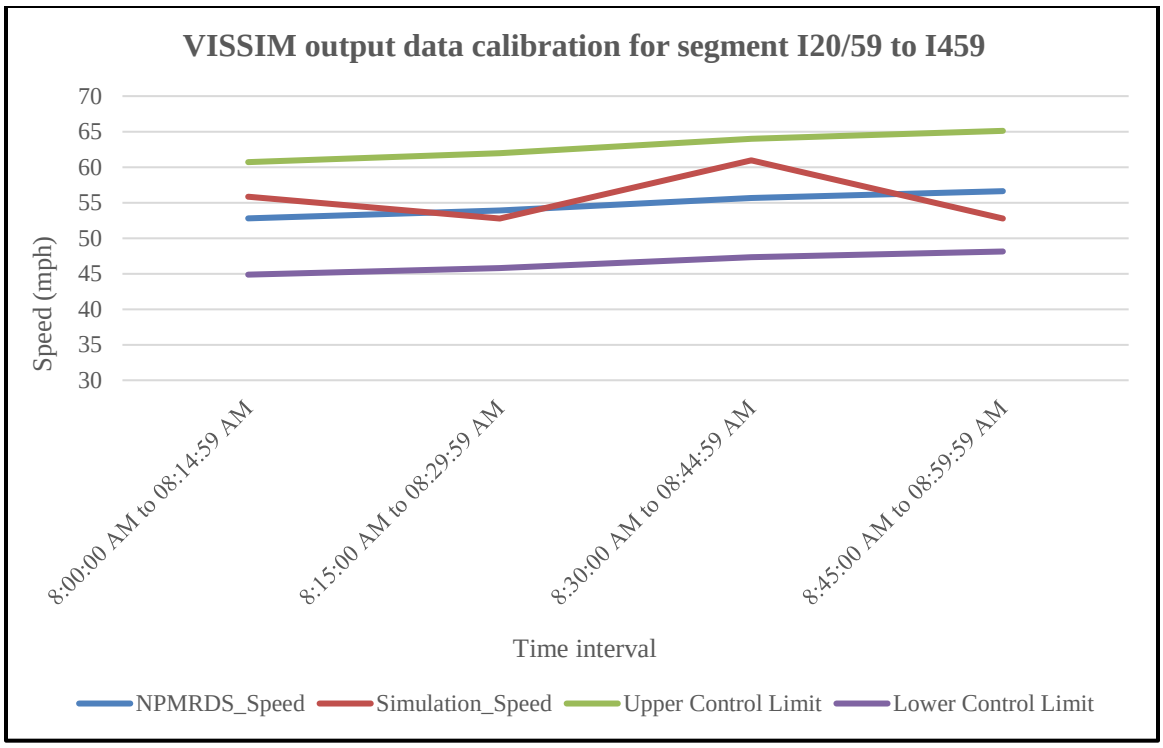


FIGURE 3-67. VISSIM OUTPUT DATA CALIBRATION FOR SEGMENT I20/59 TO I459

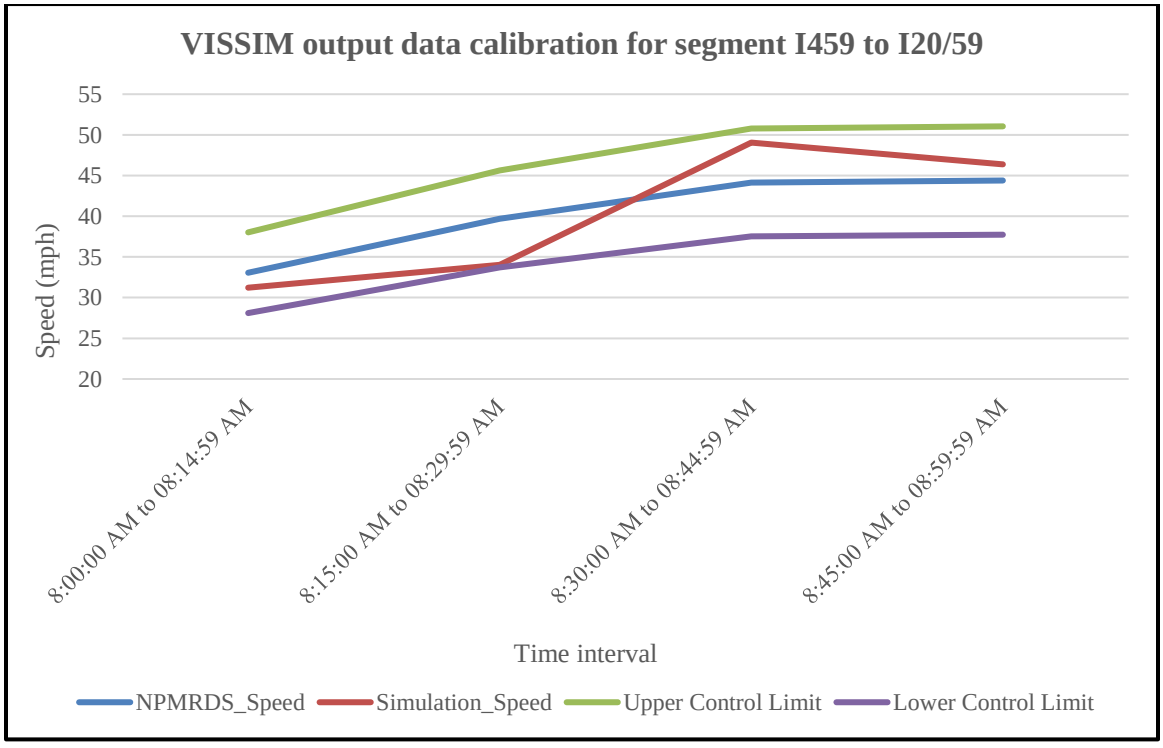


FIGURE 3-68. VISSIM OUTPUT DATA CALIBRATION FOR SEGMENT I459 TO I20/59

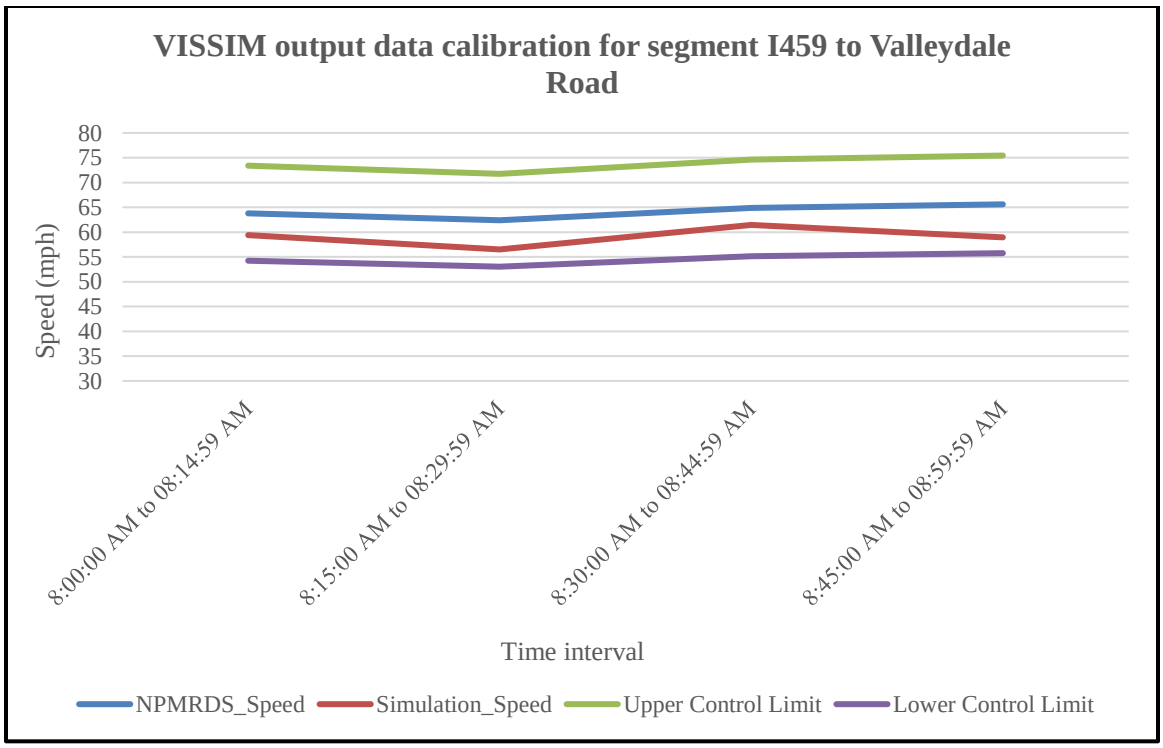


FIGURE 3-69. VISSIM OUTPUT DATA CALIBRATION FOR SEGMENT I459 TO VALLEYDALE ROAD

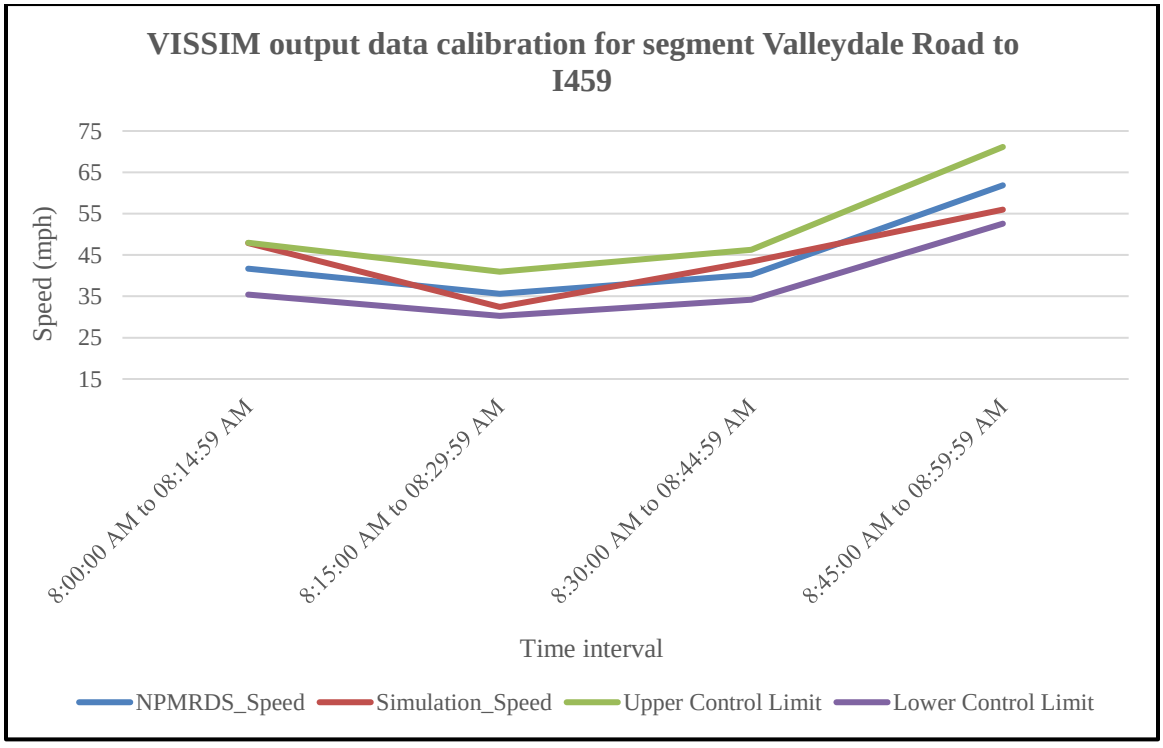


FIGURE 3-70. VISSIM OUTPUT DATA CALIBRATION FOR SEGMENT VALLEYDALE ROAD TO I459

Then, an ANOVA: Single Factor statistical F-test was performed to confirm the calibration through statistical analysis. To evaluate the upper critical value of the F distribution, the study considered a 5% significance level. The analysis of the F-test showed that the F value is less than the critical value of F, which is illustrated in Figure 3-71. Hence, there is no significant difference between VISSIM output data and filed measurement data, which is a desirable outcome.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
NPMRDS_Speed	16	816.26	51.01625	124.1573583
Simulation_Speed	16	798.91	49.93188	101.7695096

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	9.406953	1	9.406953	0.08327432	0.774893	4.170876786
Within Groups	3388.903	30	112.9634			
Total	3398.31	31				

FIGURE 3-71. STATISTICAL TEST TO VALIDATE VISSIM OUTPUT DATA

3.4.4 Trajectory Conversion Algorithm (TCA) tool

To generate BSMs, the study used the latest version of Trajectory Conversion Algorithm (TCA) tool [Version 2.3.3] (OSADP 2015), developed by the Federal Highway Administration (FHWA). The TCA Software is designed to test different strategies for producing, transmitting, and storing Connected Vehicle information (OSADP, 2015).

The TCA tool requires vehicle trajectory file, either collected from the real world or generated from a software such as VISSIM. The trajectory file from real world data should contain vehicle ID, time, speed, vehicle position (x, y coordinates), acceleration, and vehicle type of the vehicle. Also, vehicle record from VISSIM output file should contain vehicle number, speed, acceleration, vehicle type, simulation time, and vehicle position. TCA generates BSMs based on SAE J2735 standards using these trajectory files as an input file. For example, TCA tool changes the vehicle ID every 5-minutes. Moreover, TCA software is an open source software, and the user can select market penetration rate, transmission loss, the communication type (DSRC or Cellular), and also specify the roadside unit location.

3.4.4.1 Assumptions for generating BSMs using TCA tool

The calibrated VISSIM output data were used as an input file for TCA tool to generate BSMs Part-I. BMSs Part-I data attributes were introduced in Chapter 2 of this document and details are available in Appendix B. The study assumed 100% market penetration rate, data transmission loss 10%, and DSRC technology as the communication type only. A total of 14 roadside units (RSUs) were placed along the study corridor at 1-mile interval, which are shown in Figure 3-72.

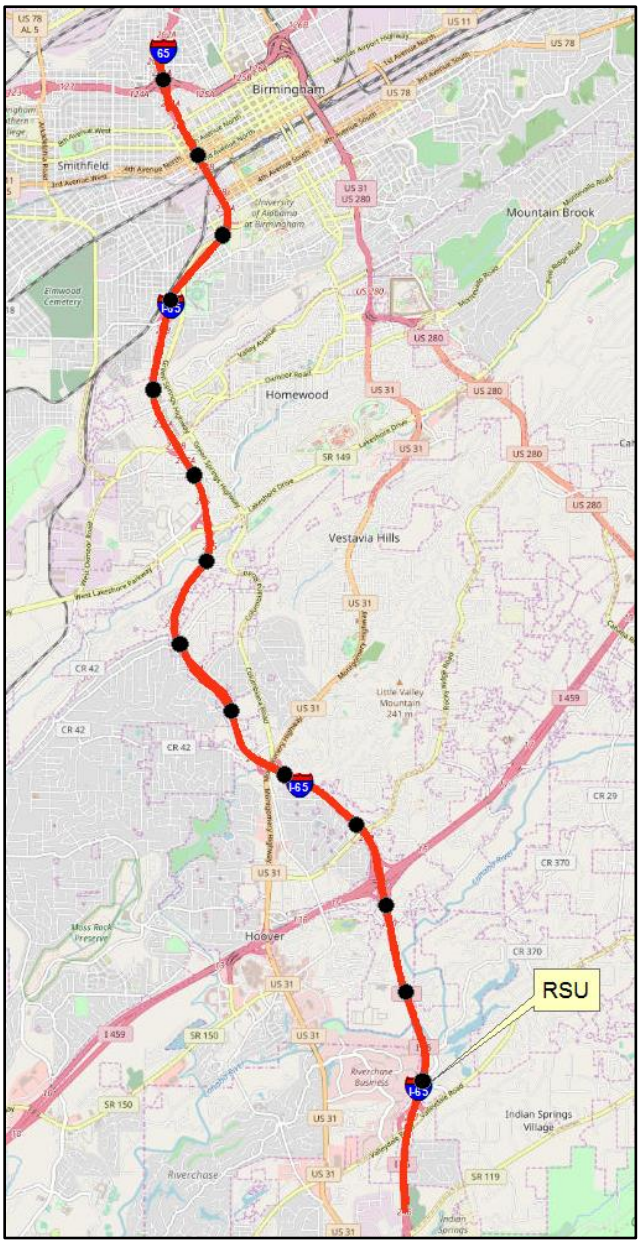


FIGURE 3-72. RSU LOCATION ALONG THE STUDY CORRIDOR

3.4.5 Application of framework to calculate the selected performance measures

The proposed methodological framework as described in Chapter 3 of this document contains twenty performance measures. For illustration purposes, this study selected three performance measures, namely TTI, PTI, and SND to establish the proposed framework as a proof of concept. Analysis procedures and results of these performance measures using the proposed methods in the framework are discussed below in the following subsections.

3.4.5.1 Data collection

As mentioned earlier, BSM Part-I data has been generated along the study corridor using TCA tool. The sample data contains DSRC_MessageID, temporary vehicle ID, transtime, location_x, location_y, speed, heading, instantaneous acceleration, brake pressure, brake status, and transmission received time. A sample of BSM data Part-I for the I-65 study corridor is shown in Figure 3-73.

DSRC_MessageID	Vehicle_ID	BSM_tmp_ID	transtime	X	Y	Speed	Instant_Acceleration	Heading	brakeStatus	brakePressure	hardBraking	transTo	transmission_received_time
02	1	2051799291	0	-4494	19153	68.69	0.33	-9999	0000	0	0	1	0
02	1	2051799291	1	-4499	19107	69.28	0.58	186.5	0000	0	0	1	1
02	5	3234925705	1	-2458	6432	77.38	0	-9999	0000	0	0	10	1
02	6	2220986051	1	-3224	16266	61.22	0.67	-9999	0000	0	0	3	1
02	7	2459590290	1	114	483	77.09	0.45	-9999	0000	0	0	14	1
02	8	3336490039	1	-3229	16275	80.79	0.86	-9999	0000	0	0	3	1
02	9	3974256673	1	-725	5096	65.02	1.02	-9999	0000	0	0	11	1
02	10	2845872919	1	-357	3696	64.75	0.58	-9999	0000	0	0	12	1
02	11	837701584	1	-3231	16261	63.59	0.32	-9999	0000	0	0	3	1
02	12	3877035076	1	-4278	14736	63.45	0.58	-9999	0000	0	0	4	1
02	13	3058786577	1	-4446	12699	66	0.73	-9999	0000	0	0	5	1
02	15	1999292319	1	-360	3693	68.76	1.09	-9999	0000	0	0	12	1
02	16	245065966	1	-3226	16245	66.11	0.44	-9999	0000	0	0	3	1
02	17	4236809084	1	122	499	62.76	0.56	-9999	0000	0	0	14	1
02	18	2467933529	1	-4188	8981	71.15	0.9	-9999	0000	0	0	8	1
02	20	2958436573	1	-4217	8909	68.37	0.83	-9999	0000	0	0	8	1
02	22	1636061193	1	-4271	14744	66	0.44	-9999	0000	0	0	4	1
02	23	962646063	1	-4268	14740	72.65	0.49	-9999	0000	0	0	4	1
02	24	1306578549	1	-4445	12687	70.76	0.35	-9999	0000	0	0	5	1
02	26	3813417771	1	-730	5112	66.65	0.79	-9999	0000	0	0	11	1
02	28	1593315671	1	-361	3540	62.49	0.35	-9999	0000	0	0	12	1
02	30	3500010566	1	-4447	12648	55.54	1.2	-9999	0000	0	0	5	1

FIGURE 3-73. SAMPLE OF BSM DATA PART-I, I-65 STUDY CORRIDOR

3.4.5.2 Calculation of the selected performance measures using CV data

As the main objective of this study is to validate the proposed framework, the study used the proposed algorithm to calculate average travel time, free flow travel time, and 95th percentile travel time, average speed, and standard deviation of speed using connected vehicle BSMs. Then, these calculated values are used to estimate TTI, PTI, and SND. Results of these performance measures are illustrated in the following section.

3.4.5.3 Results

According to the proposed algorithm, free flow speed is calculated based on hourly traffic flow rate (1000 pc/hr/lane). The study utilized speed data from NPMRDS data set and monthly traffic volume report from Alabama Department of Transportation (ALDOT) to identify the segment free flow speed. From monthly traffic volume report, it is found that traffic flow rate is less than 1000 pc/hr/lane only from 12:00 AM to 5:00 AM. To estimate free flow speed, each

link of the study corridor was matched with the NPMRDS TMCs. Then, average speed was calculated from 12:00 AM to 5:00 AM for each link along the study corridor, and this average speed is illustrated in Table 3-10, identified as a free flow speed to estimate free flow travel time.

TABLE 3-10: FREE FLOW SPEED OF EACH LINK ALONG THE STUDY CORRIDOR

Link number	Road Name	Direction	FFS (mph)
1	I-65	Northbound	73
3	I-65	Northbound	74
5	I-65	Northbound	74
7	I-65	Northbound	74
9	I-65	Northbound	74
12	I-65	Northbound	71
17	I-65	Northbound	73
19	I-65	Northbound	73
22	I-65	Northbound	73
25	I-65	Northbound	71
28	I-65	Northbound	70
31	I-65	Northbound	67
37	I-65	Southbound	60
42	I-65	Southbound	68
46	I-65	Southbound	70
48	I-65	Southbound	70
51	I-65	Southbound	73
55	I-65	Southbound	70
58	I-65	Southbound	72
62	I-65	Southbound	72
64	I-65	Southbound	75
66	I-65	Southbound	74
68	I-65	Southbound	74
69	I-65	Southbound	74

The results of the selected performance measures are summarized and discussed in the following subsections.

3.4.5.3.1 TRAVEL TIME INDEX (TTI)

The average travel time index values for the selected study corridor are summarized in Table 3-11. TTI value varies from 1.12 to 2.67. A TTI values close to 1 indicates free flow travel time condition.

TABLE 3-11: AVERAGE TTI ALONG THE STUDY CORRIDOR USING BSMS

Segment Name	Road	Direction	Time interval	Average
I20/59 to I459	I-65	Southbound	08:00:00 to 08:14:59	1.21
I20/59 to I459	I-65	Southbound	08:15:00 to 08:29:59	1.59
I20/59 to I459	I-65	Southbound	08:30:00 to 08:44:59	1.12
I20/59 to I459	I-65	Southbound	08:45:00 to 08:59:59	1.58
I459 to I20/59	I-65	Northbound	08:00:00 to 08:14:59	2.67
I459 to I20/59	I-65	Northbound	08:15:00 to 08:29:59	1.99
I459 to I20/59	I-65	Northbound	08:30:00 to 08:44:59	1.65
I459 to I20/59	I-65	Northbound	08:45:00 to 08:59:59	1.81
I459 to Valleydale Road	I-65	Southbound	08:00:00 to 08:14:59	1.21
I459 to Valleydale Road	I-65	Southbound	08:15:00 to 08:29:59	2.10
I459 to Valleydale Road	I-65	Southbound	08:30:00 to 08:44:59	1.18
I459 to Valleydale Road	I-65	Southbound	08:45:00 to 08:59:59	1.38
Valleydale Road to I459	I-65	Northbound	08:00:00 to 08:14:59	2.32
Valleydale Road to I459	I-65	Northbound	08:15:00 to 08:29:59	2.13
Valleydale Road to I459	I-65	Northbound	08:30:00 to 08:44:59	2.39
Valleydale Road to I459	I-65	Northbound	08:45:00 to 08:59:59	1.68

3.4.5.3.2 PLANNING TIME INDEX (PTI)

The average planning time index values for the study corridor locations are summarized in **Error! Reference source not found.**. A PTI values close to 1 indicates that traveler can travel the s egments with free flow speed.

TABLE 3-12: AVERAGE PTI ALONG THE STUDY CORRIDOR USING BSMS

Segment Name	Road Name	Direction	Time interval	Average PTI
I20/59 to I459	I-65	Southbound	08:00:00 to 08:14:59	0.89
I20/59 to I459	I-65	Southbound	08:15:00 to 08:29:59	0.92
I20/59 to I459	I-65	Southbound	08:30:00 to 08:44:59	1.01
I20/59 to I459	I-65	Southbound	08:45:00 to 08:59:59	1.04
I459 to I20/59	I-65	Northbound	08:00:00 to 08:14:59	1.07
I459 to I20/59	I-65	Northbound	08:15:00 to 08:29:59	1.10
I459 to I20/59	I-65	Northbound	08:30:00 to 08:44:59	1.16
I459 to I20/59	I-65	Northbound	08:45:00 to 08:59:59	1.19
I459 to Valleydale Road	I-65	Southbound	08:00:00 to 08:14:59	1.26
I459 to Valleydale Road	I-65	Southbound	08:15:00 to 08:29:59	1.32
I459 to Valleydale Road	I-65	Southbound	08:30:00 to 08:44:59	1.32
I459 to Valleydale Road	I-65	Southbound	08:45:00 to 08:59:59	1.33
Valleydale Road to I459	I-65	Northbound	08:00:00 to 08:14:59	1.77
Valleydale Road to I459	I-65	Northbound	08:15:00 to 08:29:59	1.79
Valleydale Road to I459	I-65	Northbound	08:30:00 to 08:44:59	1.88
Valleydale Road to I459	I-65	Northbound	08:45:00 to 08:59:59	1.90

3.4.5.3.3 SPEED NORMAL DEVIATE (SND)

The Speed Normal Deviate (SND) is a good indicator used to classify recurrent and non-recurrent congestion. In the case study, SND values were calculated for each 15-minute time interval for the selected study corridor. Table 3-13 summarizes the calculated maximum segment SND values. Higher SND values indicate higher speed drops.

TABLE 3-13: MAX SND VALUE ALONG THE STUDY CORRIDOR USING BSMS

Segment Name	Road Name	Direction	Time interval	Max SND
I20/59 to I459	I-65	Southbound	08:00:00 to 08:14:59	0.7
I20/59 to I459	I-65	Southbound	08:15:00 to 08:29:59	-10.5
I20/59 to I459	I-65	Southbound	08:30:00 to 08:44:59	4.0
I20/59 to I459	I-65	Southbound	08:45:00 to 08:59:59	-9.9
I459 to I20/59	I-65	Northbound	08:00:00 to 08:14:59	-0.4
I459 to I20/59	I-65	Northbound	08:15:00 to 08:29:59	-1.9
I459 to I20/59	I-65	Northbound	08:30:00 to 08:44:59	0.4
I459 to I20/59	I-65	Northbound	08:45:00 to 08:59:59	1.2
I459 to Valleydale Road	I-65	Southbound	08:00:00 to 08:14:59	4.1
I459 to Valleydale Road	I-65	Southbound	08:15:00 to 08:29:59	-14.1
I459 to Valleydale Road	I-65	Southbound	08:30:00 to 08:44:59	4.9
I459 to Valleydale Road	I-65	Southbound	08:45:00 to 08:59:59	-17.0
Valleydale Road to I459	I-65	Northbound	08:00:00 to 08:14:59	0.7
Valleydale Road to I459	I-65	Northbound	08:15:00 to 08:29:59	-9.0
Valleydale Road to I459	I-65	Northbound	08:30:00 to 08:44:59	-0.4
Valleydale Road to I459	I-65	Northbound	08:45:00 to 08:59:59	-7.8

3.4.6 Calculation of the selected performance measures using NPMRDS data

The selected performance measures were also calculated using data from the field measurement data set NPMRDS. Analysis procedures and results of these performance measures using the conventional data collection method are illustrated in the following subsections.

3.4.6.1 Data collection

This study used one hour data for August 31, 2017 from 8:00 AM to 9:00 AM of the National Performance Management Research Data Set (NPMRDS) obtained from Federal Highway Administration (FHWA) with the help of the Regional Planning Commission of Greater Birmingham (RPCGB). The data set contains travel time, speed, historic average speed, reference speed, and data density based on probe vehicles for every 5 minutes, 10 minutes, 15 minutes, and 1 hour, seven days per week and covers the entire National Highway System (NHS) containing all interstates and US highways. Now, this data are collected by INRIX. Sample travel time data for the study corridor are shown in Table 3-14.

.

TABLE 3-14: TRAVEL TIME SAMPLE DATA

tmc_code	measurement_tstamp	travel_time_seconds
101N04375	8/31/2017 8:00:00 AM	31.31
101N04375	8/31/2017 8:05:00 AM	31.31
101N04375	8/31/2017 8:10:00 AM	40.16
101N04375	8/31/2017 8:15:00 AM	33.58
101N04375	8/31/2017 8:20:00 AM	32.99
101N04375	8/31/2017 8:25:00 AM	31.31
101N04375	8/31/2017 8:30:00 AM	29.79
101N04375	8/31/2017 8:35:00 AM	31.31

3.4.6.2 Method of calculation

Selected performance measures such as TTI, PTI, and SND were calculated using the methods applied by (Sullivan et al., 2017) and are discussed in the following section.

3.4.6.3 Results

TTI, PTI, and SND have been analyzed using NPMRDS travel time data set. Results of these performance measures are discussed in the following subsections.

3.4.6.4.1 TRAVEL TIME INDEX (TTI)

TTI values were calculated for the selected study corridor from 8:00 AM to 9:00 AM. Average segment TTI values are summarized in Table 3-15.

TABLE 3-15: AVERAGE SEGMENT TTI ALONG THE STUDY CORRIDOR USING NPMRDS DATA SET

Segment Name	Road	Direction	Time interval	Average
I20/59 to I459	I-65	Southbound	08:00:00 to 08:14:59	1.11
I20/59 to I459	I-65	Southbound	08:15:00 to 08:29:59	1.08
I20/59 to I459	I-65	Southbound	08:30:00 to 08:44:59	1.05
I20/59 to I459	I-65	Southbound	08:45:00 to 08:59:59	1.03
I459 to I20/59	I-65	Northbound	08:00:00 to 08:14:59	2.69
I459 to I20/59	I-65	Northbound	08:15:00 to 08:29:59	1.94
I459 to I20/59	I-65	Northbound	08:30:00 to 08:44:59	1.46
I459 to I20/59	I-65	Northbound	08:45:00 to 08:59:59	1.31
I459 to Valleydale Road	I-65	Southbound	08:00:00 to 08:14:59	1.10
I459 to Valleydale Road	I-65	Southbound	08:15:00 to 08:29:59	1.12
I459 to Valleydale Road	I-65	Southbound	08:30:00 to 08:44:59	1.08
I459 to Valleydale Road	I-65	Southbound	08:45:00 to 08:59:59	1.07
Valleydale Road to I459	I-65	Northbound	08:00:00 to 08:14:59	1.67
Valleydale Road to I459	I-65	Northbound	08:15:00 to 08:29:59	2.15
Valleydale Road to I459	I-65	Northbound	08:30:00 to 08:44:59	2.00
Valleydale Road to I459	I-65	Northbound	08:45:00 to 08:59:59	1.10

3.4.6.4.2 PLANNING TIME INDEX (PTI)

Based on the analysis of NPMRDS data set for the study corridor, PTI values were calculated and summarized in Table 3-16.

TABLE 3-16: AVERAGE PTI ALONG THE STUDY CORRIDOR USING NPMRDS DATA SET

Segment Name	Road	Direction	Time interval	Average
I20/59 to I459	I-65	Southbound	08:00:00 to 08:14:59	1.82
I20/59 to I459	I-65	Southbound	08:15:00 to 08:29:59	1.22
I20/59 to I459	I-65	Southbound	08:30:00 to 08:44:59	1.14
I20/59 to I459	I-65	Southbound	08:45:00 to 08:59:59	1.15
I459 to I20/59	I-65	Northbound	08:00:00 to 08:14:59	1.14
I459 to I20/59	I-65	Northbound	08:15:00 to 08:29:59	3.92
I459 to I20/59	I-65	Northbound	08:30:00 to 08:44:59	1.95
I459 to I20/59	I-65	Northbound	08:45:00 to 08:59:59	2.89
I459 to Valleydale Road	I-65	Southbound	08:00:00 to 08:14:59	2.62
I459 to Valleydale Road	I-65	Southbound	08:15:00 to 08:29:59	1.56
I459 to Valleydale Road	I-65	Southbound	08:30:00 to 08:44:59	1.11
I459 to Valleydale Road	I-65	Southbound	08:45:00 to 08:59:59	1.22
Valleydale Road to I459	I-65	Northbound	08:00:00 to 08:14:59	2.77
Valleydale Road to I459	I-65	Northbound	08:15:00 to 08:29:59	1.31
Valleydale Road to I459	I-65	Northbound	08:30:00 to 08:44:59	1.21
Valleydale Road to I459	I-65	Northbound	08:45:00 to 08:59:59	1.13

3.4.6.4.3 SPEED NORMAL DEVIATE (SND)

SND values were also calculated for the study corridor at 15-minute interval. Maximum segment SND values are summarized in the Table 3-17.

.

TABLE 3-17: MAXIMUM SEGMENT SND ALONG THE STUDY CORRIDOR USING NPMRDS DATA SET

Segment Name	Road Name	Direction	Time Interval	Max SND
I20/59 to I459	I-65	Southbound	08:00:00 to 08:14:59	-4.04
I20/59 to I459	I-65	Southbound	08:15:00 to 08:29:59	-2.124
I20/59 to I459	I-65	Southbound	08:30:00 to 08:44:59	3.305
I20/59 to I459	I-65	Southbound	08:45:00 to 08:59:59	-4.598
I459 to I20/59	I-65	Northbound	08:00:00 to 08:14:59	-7.19
I459 to I20/59	I-65	Northbound	08:15:00 to 08:29:59	-11.449
I459 to I20/59	I-65	Northbound	08:30:00 to 08:44:59	10.632
I459 to I20/59	I-65	Northbound	08:45:00 to 08:59:59	5.719
I459 to Valleydale Road	I-65	Southbound	08:00:00 to 08:14:59	2.833
I459 to Valleydale Road	I-65	Southbound	08:15:00 to 08:29:59	-1.083
I459 to Valleydale Road	I-65	Southbound	08:30:00 to 08:44:59	0.721
I459 to Valleydale Road	I-65	Southbound	08:45:00 to 08:59:59	2.833
Valleydale Road to I459	I-65	Northbound	08:00:00 to 08:14:59	-8.715
Valleydale Road to I459	I-65	Northbound	08:15:00 to 08:29:59	-4.748
Valleydale Road to I459	I-65	Northbound	08:30:00 to 08:44:59	1.089
Valleydale Road to I459	I-65	Northbound	08:45:00 to 08:59:59	13.167

3.4.7 Comparison between performance measures using proposed framework and conventional method

3.4.7.1 Hypothesis

We hypothesized that there is no significant statistical difference between performance measurements calculated using the proposed framework and CV data and those that are calculated using conventional methods and data sources.

3.4.7.2 Comparison

To test the abovementioned hypothesis, an ANOVA: Single Factor statistical F-test was performed. The test was used to find if there is a significance difference between the calculated performance measures using BSMs and NPMRDS data set at the 5% significance level.

3.4.7.2.1 TRAVEL TIME INDEX (TTI)

From the ANOVA statistical test it is found that there is no significant difference between TTI values produced from the proposed framework using CV data and those produced from traditional data at the 5% significance level as F value is less than the critical value of F, as illustrated in Figure 3-74.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Avg_TTI_BSMs	16	27.98847222	1.749279514	0.227258
Avg_TTI_NPMRDS	16	22.95193676	1.434496047	0.257114

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.792709046	2	0.396354523	1.582019	0.222791	3.327654
Within Groups	7.265579457	29	0.250537223			
Total	8.058288503	31				

FIGURE 3-74: ANOVA: SINGLE FACTOR STATISTICAL TEST OF TTI VALUES

3.4.7.2.2 PLANNING TIME INDEX (PTI)

The ANOVA: Single Factor statistical F-test confirmed that there is a no significance difference in the calculated PTI values using BSMs and NPMRDS data set at the 5% significance level. The results of the F-test are illustrated in Figure 3-75.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Avg_PTI_NPMRDS	16	28.16796	1.760497777	0.722063909
Avg_PTI_BSMs	16	20.92597	1.307873264	0.117341437

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.638951597	1	1.638951597	3.905030161	0.057400714	4.170876786
Within Groups	12.59108019	30	0.419702673			
Total	14.23003179	31				

FIGURE 3-75: ANOVA: SINGLE FACTOR STATISTICAL TEST OF PTI VALUES

3.4.7.2.3 SPEED NORMAL DEVIATE (SND)

Similarly to the comparison of results for the TTI and PTI indices, the ANOVA statistical test of SND values confirmed that differences in SND values calculated from CV and field data are not statistically significant as the F value obtained is less than the critical value of F, as illustrated in Figure 3-76.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Max_SND_BSMs	16	-54.706	-3.419125	47.437384
Max_SND_NPMRDS	16	-3.648	-0.228	44.758866

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	81.46623	1	81.46623	1.7672352	0.1937459	4.1708768
Within Groups	1382.9437	30	46.098125			
Total	1464.41	31				

FIGURE 3-76. ANOVA: SINGLE FACTOR STATISTICAL TEST OF SND VALUES

Overall, the statistical analysis performed above confirms that there is a close agreement between the performance measures (i.e., TTI, PTI and SND) generated from a. the proposed methodological framework using CV data generated by the VISSIM simulation platform and the TCA tool and b. actual field data obtained from the NPMRDS travel time data set. Thus, the proof of concept for the proposed framework is successful.

3.5 CONCLUSION AND RECOMMENDATIONS

This chapter presents the research summary, study implications, and future directions based on the research conducted for this study.

3.5.1 Summary of Research

Performance measurement and management is of great importance toward achieving operational effectiveness of roadways. Field or simulated data can be used to determine performance measures. Emerging vehicle technologies, like connected vehicles (CVs) create new opportunities for collecting new types of transportation data that can improve the accuracy of transportation system performance measurement. Proliferation of CVs is also expected to increase data quantity and quality and enable the development of new performance measures.

In light of the rapid progress in the area of CV technologies, this study (a) developed a methodological framework to estimate system performance measurements using CV data and (b) provided a proof of concept to validate the proposed framework. In doing so, performance measurements were calculated using traditional field data and simulated CV data for a freeway

study segment in Birmingham, AL and compared. A close agreement between the findings serves as a proof of concept for the proposed framework.

First, the study conducted a detail literature review to identify the attributes of CV data and available data sources. A literature search was also employed to identify and document performance measures that can be calculated using emerging CV data.

The study then developed a detailed methodological framework to describe the process of determining performance measures from CV data. The proposed methodological framework contains four parts, namely: (i) physical data flow diagram, (ii) processes and process groups hierarchical diagram, (iii) individual process designs, and (iv) logical data flow diagram. A primary challenge was to aggregate connected vehicle data at RSUs. The proposed methodological framework addressed the data aggregation issue and introduced data aggregation algorithms at RSUs to estimate performance measures.

In order to validate the proposed framework, the study developed a VISSIM simulation model of a 14-mile long section of interstate I-65 between exit 247 and 261A in Birmingham, AL. Data calibration was done using upper and lower control limits of the NPMRDS speed data (+/-15%). The speed data produced by the simulation model were within the upper and lower control limits of the NPMRDS data set, thus confirming the reliability of the simulation outputs.

The TCA tool was used to generate BSMs using VISSIM output as an input and considering market penetration rate 100%. The outputs from this process were then used to generate selected performance measures following the procedures presented in the methodological framework. The performance measures generated for demonstration purposes included Travel Time Index (TTI), Planning Time Index (PTI), and Speed Normal Deviate (SND). Field data from the NPMRDS data set were also used to calculate similar performance measures using the conventional method. A statistical comparison between the two sets of performance measures was performed using the ANOVA: Single Factor statistical F-test. The results from the statistical test showed that the F value were less than the critical values of F within 5% significance level for all performance measures tested. This finding indicate that there is no significant difference between performance measures generated using the VISSIM output data and the NPMRDS data. Hence, proposed framework is valid, and it can be used in practical applications.

3.5.2 Implications for Practice

This methodological framework proposed in this study serves as a reference to transportation agencies, researchers, and consultants involved in the assessment of transportation network performance using performance measures. Advances in transportation system performance measurement resulting from this study are expected to improve transportation policy decision

making, optimize planning and operations, and improve transportation system outcomes in the future.

3.5.3 Recommendations for Future Research

CV technology has the potential to address major issues affecting the US transportation system and introduce new exciting opportunities for data sharing in the transportation sector. CV-related research is a growing and exciting area. The market penetration rate of CV is expected to increase significantly within the next few years. The full potential benefits of CVs can only be achieved through a connected environment. So, the development of a framework that can used to estimate transportation system performance measures is of great importance.

It is important to note that the groundbreaking work performed in this study should be followed by additional research to expend the scope of the work in the near future. Some study limitations are listed below, along with recommendations for potential future work.

- The study used simulation data. Evaluation of the proposed framework using a real-world testbed data is also recommended.
- This study only considered a basic freeway section. Incorporation of other facilities such as merging, diverging, and weaving to evaluate the framework performance should be considered as an extension of the current work.
- The proof of concept study focused on a freeway network. Methods and algorithms to combine data collected from existing and emerging sources with enhanced models and optimization algorithms should be considered in future studies to optimize and manage signal operations.
- This study considered twenty performance measures in the methodological framework. However, the framework proposed in this study may need to be updated and expanded to incorporate emerging performance measures in the future.

Consideration of different market penetration rates is also recommended to evaluate the proposed framework in future studies.

4. EMISSION ESTIMATION BASED ON CV DATA

4.1 INTRODUCTION

Increased traffic in transportation networks does not only affects the mobility but also increases emissions and fuel consumption. The transportation sector has become one of the largest energy consumer and greenhouse gas emitter (GHG). In the U.S., vehicles are responsible for one-third of total CO_2 emission (EPA, 2008) and consume approximately 30% of total petroleum (Annual Energy Outlook, 2008). Several measures in terms of new policies/infrastructure have been taken or are being planned to curb vehicular emissions. In order to estimate the impact of these measures on environment, it is important to measure/quantify vehicular emissions. It is possible to monitor pollutants and therefore, real time air quality but they are generated from various sources such as energy sector, manufacturing, agriculture, air transportation, ground transportation, etc. It is very difficult to isolate and quantify the effect of ground transportation (vehicles) in total emission production. Several models have been developed to estimate vehicle emission and fuel consumption. Traditionally, estimating emission and fuel consumption is a two-stage procedure. The first stage involves estimation of travel and/or driving pattern, followed by second stage, which involves estimation of emission corresponding to travel/traffic data estimated in the first stage. These models can be broadly divided into macroscopic and microscopic. Macroscopic models require simpler, macro input such as average speed and then estimate emissions and fuel based on relationships between macro inputs and emission rates (usually developed using standard drive cycles). Microscopic models require finer inputs i.e. second-by-second speed and acceleration values and therefore measure emission at second by second level. These finer inputs (speed profiles) can appreciate small changes in driving patterns and therefore can distinguish emission and fuel consumption between various driving cycles with similar macroscopic characteristic such as average speed. Microscopic models can be used to measure change in emissions due to transportation operational projects such signal coordination. These models can also be used to monitor real time emission if speed and acceleration values of all the vehicles are known in the entire network/study area. The current traffic data collection techniques such as fixed data recording sources (loop detector, traffic cameras), probe vehicles methods (floating car method) can measure macroscopic traffic parameters like speed, density but cannot be used to extract finer traffic data i.e. speed profiles of each vehicle.
With the advent of connected vehicle technology, it should be possible to install an infrastructure of roadside units, which can receive second by second speed and acceleration of the connected vehicles (CVs) in the vicinity and estimate real time emission and fuel consumption (this infrastructure can also be used for other transportation applications such as signal coordination, safety, etc.). These roadside units can be installed strategically in the transportation network to calculate real time estimates of emission and fuel from the entire network. Such infrastructure will also provide an opportunity to study temporal and spatial variations in emissions and to assess the environmental impact of any change in transportation policy/infrastructure. However, due to slow deployment, CVs will very slowly penetrate into the

network and it will take decades to achieve complete connected vehicle infrastructure. The market share of CVs will gradually grow and, but it is possible to utilize CV data (after its market penetration crosses certain threshold) to augment the current transportation analysis.
The important question for today is how limited CV data, which will be available in near future, can be used to estimate real time vehicle emissions without any information about unconnected vehicles in the network. Nevertheless, it should be possible to estimate instantaneous speed and acceleration of unconnected vehicles within some confidence interval using traffic flow theories. The reconstruction of trajectories of unconnected vehicles is beyond the scope of this chapter. However, this chapter attempts to estimate the possible distribution of emission and fuel consumption in network with limited available CV data. For analysis, this study assumes that speed and acceleration of unconnected vehicles can be determined in multiples of fixed values, therefore is possible to approximate speed and acceleration to the correct predefined bins (which allow uncertainty in calculation). The study applied microscopic emission models (CMEM and VT-Micro, explained in next section) to large dataset of speed profiles of vehicles collected using global positioning system (GPS) for different vehicle categories and determined the normalized frequency (probability) distribution of emission and fuel consumption for each speed and acceleration bin. Later, emission and fuel consumption were estimated for NGSIM data assuming different market shares of connected vehicle. The study is limited to microscopic emissions of light duty vehicles and focuses on carbon dioxide emission and fuel consumption only.
The structure of this chapter is as follows. Section 0 reviews the literature of emission models and connected vehicles. The methodology and data is described in Section 0. Section 0 describes microscopic estimation procedure. The development of distributions are explained in Section 0. Application of the study is described in section 0. Section 0 summarizes and concludes the study.

4.2 LITERATURE REVIEW

This section reviews emission models and their data requirements. It also reviews connected vehicle technology, their applications with limited connected vehicle data and reconstruction of unconnected vehicle trajectories.

4.2.1 Emission Models

Several emission models have been developed to estimate vehicle emission and fuel consumption and to evaluate the environmental impact of transportation decision making. These models are can be broadly classified as macroscopic and microscopic models and were developed by establishing relationship between emission rates and transportation metrics using standardised driving cycle intended to represent driving patterns on roads. For example, macroscopic emission models require macro/coarser inputs such average speed, distance travelled, vehicle fleet composition etc., which can be obtained from macroscopic travel demand model, travel survey, traffic data collection, etc. Since the input variables are macro/coarser in nature, macroscopic emission models can be applied to transportation

planning with static approach to estimate total/average emission for large networks and to evaluate environmental impacts. Some examples of macroscopic models are MOVES, MOBILE6, COPERT III, etc. MOVES was developed by US Environmental Protection Agency (EPA, 2014). These models cannot appreciate the difference in emission rates from different driving cycles, which have similar aggregate representation such as average speed.

Microscopic models require instantaneous (second by second) speed and acceleration profile of all vehicles. These fine second by second inputs can measure small changes in driving pattern and therefore, can appreciate change in emission and fuel consumption due to operational transportation projects such as signal coordination, traffic calming measures, eco-routing, etc. These models, in combination traffic microsimulations, are being widely applied to various transportation problems. These are either physical based or statistical based models. The two commonly used microscopic models are CMEM and VT-Micro models, which have been applied in the study, are explained below.

4.2.1.1 CMEM Model

Comprehensive Modal Emission Model (Scora and Bartha, 2006) is power based emission model developed at University of California-Riverside by Bartha et al. 1996. It calculates the tractive power by incorporating change in kinetic energy, change in potential energy, rolling resistance, wind resistance, friction loss in engine and power utilized in accessories such as air conditioning. After calculating the required engine power, it estimates the tailpipe emission and fuel consumption rates.

CMEM models estimates fuel consumption and emission rates for a wide variety of light duty vehicles (LDVs) and light duty trucks (LDTs) as a function of their operating mode. The modelling process decomposes the emission generation into modules of corresponding physical phenomenon: engine speed, air to fuel ratio, engine power, catalyst past fraction, fuel use and engine-out emission (Barth el al., 1996).

The model has classified LDVs into 12 categories based on power to weight ratio, accumulated mileage, emission certification level, fuel and emission control technology and emitter level category (Scora and Bartha, 2006). Models requires vehicle category, road grade, second by second speed profile and calibrated parameters (cold start coefficient and engine friction factor) as input. This study is limited to light duty vehicles and assumes road grade to be zero.

4.2.1.2 VT-Micro Model

VT-Micro is a regression model developed by experimenting with different polynomial combinations of speed and acceleration to create a dual regime model. It was developed by using data collected at Oak Ridge National Laboratory (ORNL) and at US Environmental Protection Agency. It estimates emission and fuel consumption instantaneous (second by second) speed and acceleration as explanatory variables. The estimated emission and fuel consumption rates were accurate and consistent with observed values with coefficient of determination ranging from 0.92 to 0.99. It classifies LDVs into five categories based on engine size, vehicle model year and mileage (Rakha et al., 2004).

Both CMEM and VT-Micro model have been extensively used to evaluate environmental impacts of transportation projects including roundabouts (Ahn et al., 2009; Coelho el at., 2006),

eco-routing (Booribonsomsin et al., 2012), signal timing optimization (Kwak et al., 2012), connected vehicles (Abianeh et al., 2009), etc. Most of these studies estimate emissions and fuel consumption based on the speed profiles obtained from traffic microsimulations. Other required inputs such as vehicle categories, soak time, etc. were assumed in these simulation studies.

The micro emission models can also be applied to proposed framework for real time estimations. However, it is difficult to obtain the required inputs the models i.e. second by second position, speed, acceleration, time into the trip, vehicle category and other parameters for all the vehicles near roadside unit and detailed road grade profile. The connected vehicle technology, explained in the next section, provides this opportunity to transmit some of these vehicle related information to the nearby infrastructure.

4.2.2 Connected Vehicle

Connected vehicle technology allows vehicle and infrastructure to communicate wirelessly using global positioning systems, sensors, etc. It allows vehicle to continuously transmit their status such as speed, position, etc. to nearby infrastructure at regular (small) interval. Several applications have been developed which use continuous position data to improve mobility (signal optimization and coordination), to monitor transportation networks, to estimate travel time, speed, queue lengths, etc. However, due to time required to setup infrastructure, possible slow deployment, resistance among drivers to share their positions, etc., CVs will penetrate slowly into the transportation system. Therefore, only a proportion of vehicle will participate initially.

The analysis, which rely on vehicle's position (mainly through GPS) usually, will have no information about non-communicating vehicles (unconnected vehicles) in the network. However, even with the limited penetration of CVs, it is possible to estimate positions (therefore, speeds and accelerations) of non-communicating vehicles. A reasonable approximation of position of non-communicating vehicles will significantly improve the quantum of data and therefore, improve the analysis. Even at the limited penetration rates, CV technology has many application. It can be used to estimate traffic volumes for signalized intersection (Zhang and Liu, 2017), queue length (Tiaprasert et al., 2015; Gao et al., 2019), platoon recognition (Tiaprasert et al., 2019), etc.

However, estimation of emissions and fuel consumption using microscopic models require second by second speed profile and other vehicular information for vehicle categorization. The limited speed profiles from small share of connected vehicle without vehicular information has very limited application. In order to utilize this dataset. This study makes assumption regarding reconstruction of speed profiles of unconnected vehicles and proposes a methodology to overcome limitation of absence of vehicular information.

4.3 METHODOLOGY

The methodology used in the study comprises of three stages as explained below.

- In the first stage, we estimated the second by second CO_2 and fuel rate of large real trajectory dataset with real soak time values, using CMEM and VT-Micro model for different vehicle categories. The road grades were assumed to be zero.
- The study does not focus on reconstruction of unconnected vehicle trajectories. In the second stage, we made assumptions regarding the reconstruction of speed profiles of unconnected vehicles. We assumed that second by second speed and acceleration of unconnected vehicles can be determined in the multiples of 5 km/h and 1 km/h/s respectively. We also assumed that the connected vehicles communicate their status every second with uncertainty of 0.5 km/h in speed and 0.25 km/h/s in acceleration. Later, we created the distributions of CO_2 and fuel rates for each speed and acceleration bin. Similar distributions of estimates were created for connected vehicles for reduced bin size. The different values of CO_2 and fuel rates in each bin accounted for estimates corresponding to different speed and acceleration values in same bin and for soak time and different vehicle categories.
- In the third stage, we estimated the distribution of total CO_2 emission and fuel consumption for NGSIM trajectory dataset, collected at Peachtree Street, Atlanta, using distributions od estimates developed in stage two. We randomly assigned labels to trajectories as connected or unconnected based on market share of CV. The second by second speed and acceleration values of connected and unconnected vehicles were rounded off for binning. Later we chose random estimate values from the distributions corresponding to bins assigned and them to estimate total CO_2 and fuel consumption. We repeated the total estimation procedure 1000 times to create a distribution of total estimates. The procedure was also repeated for different market shares.

4.3.1 Data

The study utilizes two vehicle trajectory datasets. The first dataset (CALTRANS) consisted of large number of speed profiles, which was to estimate CO_2 and fuel consumption using CMEM and VT-Micro models for various vehicle categories. The CALTRANS dataset is described in detail in section 0. Later, the distribution of total of CO_2 and fuel were estimated for the second vehicle trajectory dataset (NGSIM) using emission distributions developed using CALTRANS data. NGSIM trajectory is described in section 0.

4.4 ESTIMATION OF EMISSION RATES

This section describes the trajectory dataset used and the procedure to estimate CO2 and fuel using CMEM and VT-Micro model.

4.4.1 CALTRANS dataset

The pre-defined or simulated drive cycles cannot represent all real driving conditions or real distribution of speed and accelerations, soak times and cannot capture the real emission patterns (Joumard et al., 2000). The proposed methodology required drive cycles that cover

real world driving patterns. Therefore, vehicle trajectory dataset collected in California Household Travel Survey (CHTS) between 2010 and 2012 was utilized in the study. The CHTS was then the largest such regional travel survey ever conducted in the United States. The detailed travel survey was conducted from 42,500 households via multiple methods including 2910 in-vehicle global positioning system (GPS) devices, which were to be used for seven days. Transportation Secure Data Center, which is maintained by National Renewable Energy Laboratory (NREL, 2019), currently hosts the GPS data.

The drive cycle data provided by NREL include second-by-second speed profile collected from global positioning system (GPS). The speed profile of data of each vehicle in drive cycle data is grouped by date and passed through NREL's GPS filtering procedure to filter erroneous speed values (for details see, Duran and Earleywine, 2013). The drive cycle data does not contain the latitude and longitude spatial data. Therefore, the term 'speed profile' and 'GPS trajectory' are used interchangeably in the remaining chapter.

The dataset contains speed profiles and trip information table corresponding to each 2910 in-vehicle GPS device. The speed profiles are available in comma separated (.csv) files, grouped by date of travel (24 hour period). Each row in csv file contains time stamp and speed. The zero-speed drift segments were removed from speed profiles during filtration procedure. The dataset also has a trip information file (.csv) corresponding to each vehicle, which has date, start and stop time for each trip. The dataset contains speed profiles of 42,618 trips.

For analysis, we used 14,444 trips (31.48 million seconds of trip information) with duration between 20 minutes and 120 minutes, which were obtained, from 2441 vehicles. The distribution of trip duration is shown in Figure 4-1; approximately 50% of the trips were between 20 to 30 minutes of duration. The distribution of instantaneous speed and acceleration is shown in Figure 4-2. Figure 4-3 shows the distribution soak time of trips. Each trip was simulated in CMEM version 3.01 and VT-Micro model to estimate second by second CO_2 and fuel rate.

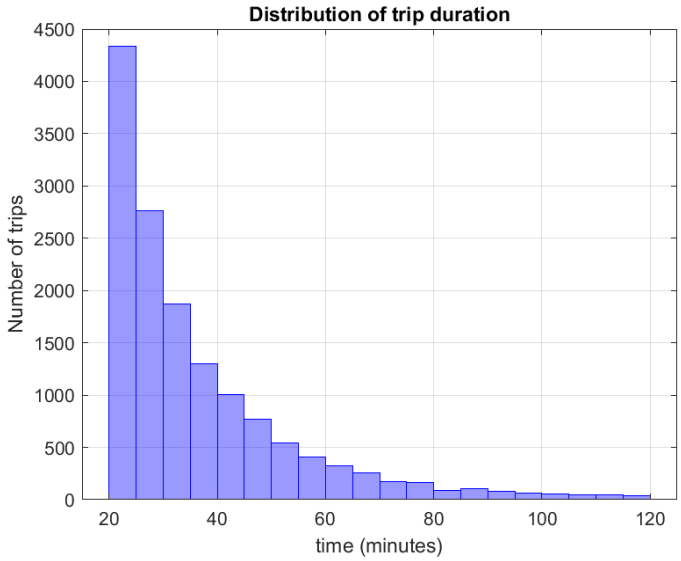


FIGURE 4-1 DISTRIBUTION OF TRIP DURATION IN CALTRANS DATASET

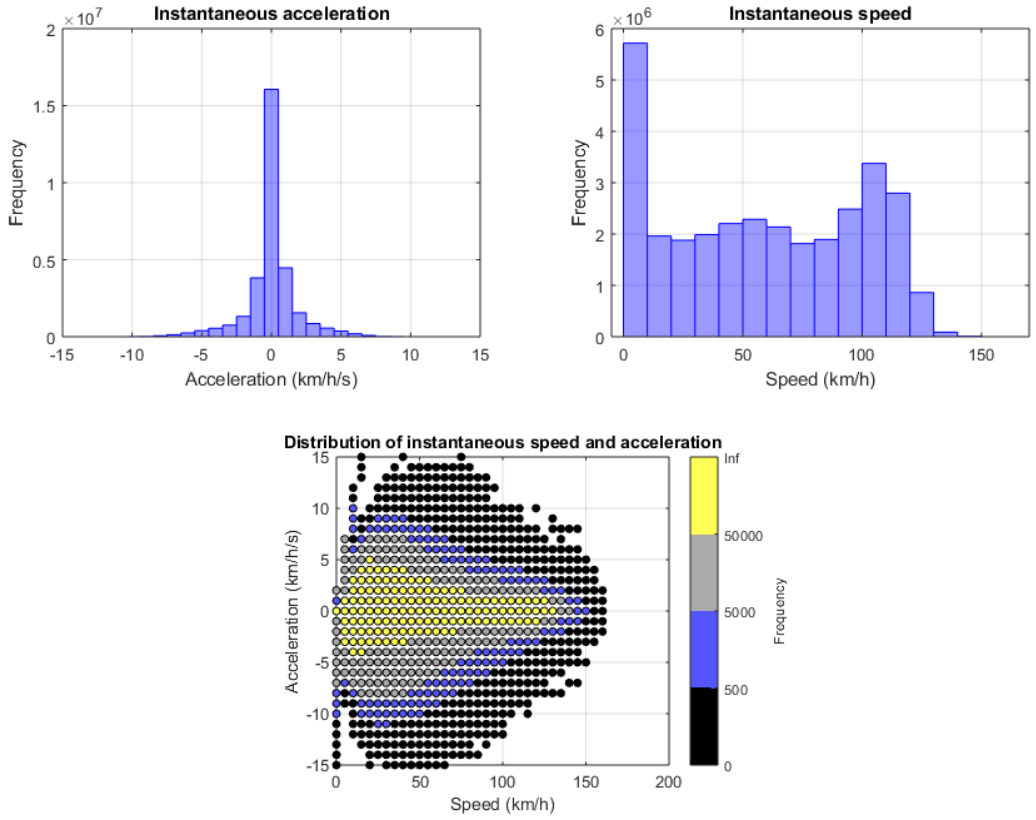


FIGURE 4-2 DISTRIBUTION OF INSTANTANEOUS SPEED AND ACCELERATION IN CALTRANS DATASET

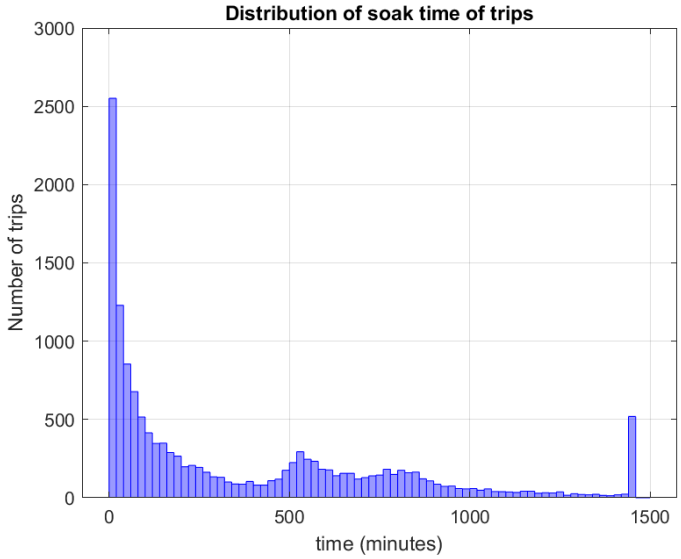


FIGURE 4-3 DISTRIBUTION OF SOAK TIME IN CALTRANS DATASET

4.4.2 Estimation of second by second CO2 and fuel consumption rate

Emission models group vehicles into various categories based on vehicle model year, mileage, weight to power ratio/engine size for estimation purpose. However, the detailed vehicle information have been determined to be potentially personally identifiable when paired with vehicle position information, therefore, it is unlikely for connected vehicles to share vehicle related information. CMEM model also utilizes model-calibrated parameters (such as cold start coefficient) as input, which require soak time information, which is impossible to determine. The relevant vehicular information available for CALTRAN dataset is shown Figure 4-4 in the form of histograms. Based on the available information about the vehicles, we have considered only light duty vehicles (LDV) for the study and then further limited our analysis to four categories (LDV 4-7) in CMEM and three categories (LDV 3-5) in VT-Micro model.

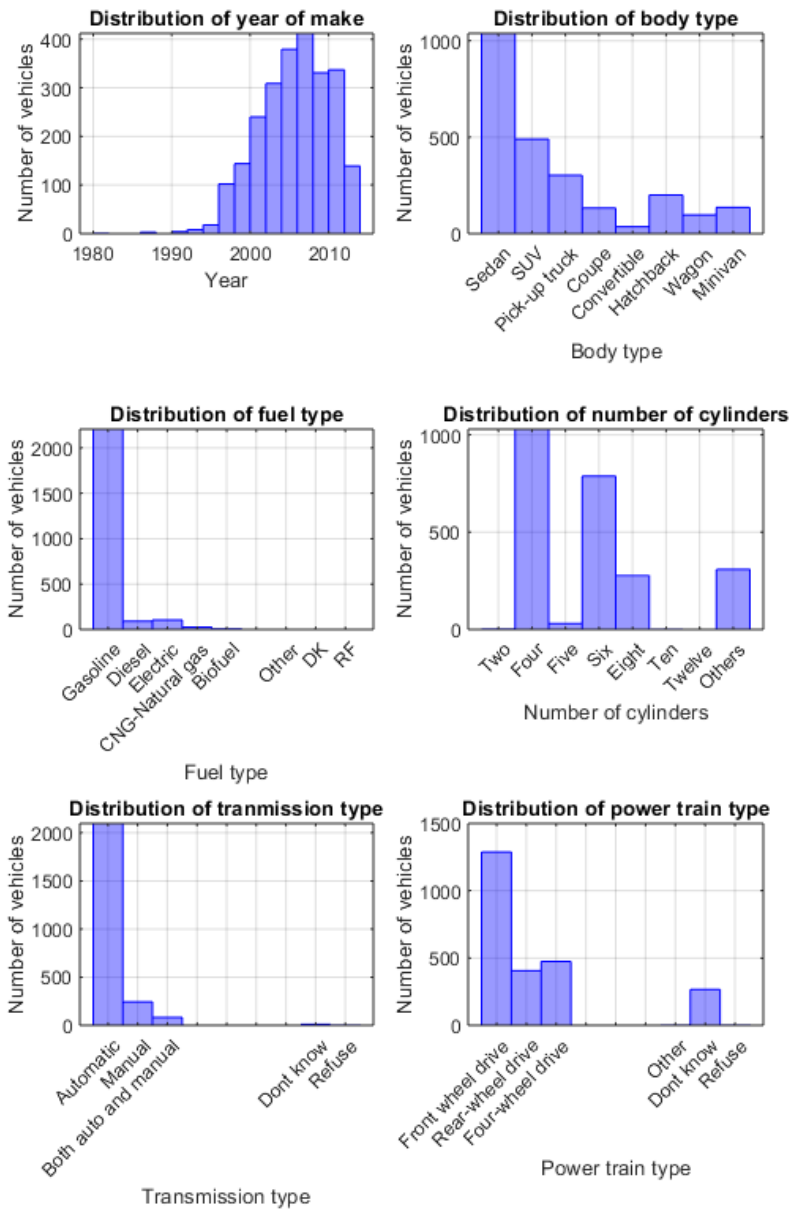


FIGURE 4-4 VEHICULAR RELATED INFORMATION

4.4.2.1 Estimation using CMEM model

CMEM model requires two input files: (a) Vehicle activity file, which contains second by second speed profile and (b) Vehicle control file, which contains information about vehicle category, soak time of trip, input/output units, etc. MATLAB script was written to read speed profiles and trip information files to create input files for CMEM model. It reads the trip information file of each vehicle and then reads csv file of corresponding date for speed profile. Speed profile for each trip was copied, imputed with zero speed values (rows with zero-speed drift were omitted

in the dataset) and speed was scaled to km/h unit. Another column of time into the trip (in seconds) was added and saved as vehicle activity file with extension .act.
Vehicle control file (.ctr) was created which contains vehicle category, soak time (in minutes), input unit of speed and desired output unit system (METRIC in this study) for output. We used default values for all other parameters in the model. Vehicle classification is an important step since vehicle categories have different emission rates. We limited our study to four categories of vehicle (LDV4 to LDV7) for analysis. CMEM model determines the operating condition (hot stabilized or cold start) based on the soak time, which specifies the time for which vehicle engine was off prior to the trip. The soak time of trip was determined by subtracting end time of the previous trip from the start time of the corresponding trip. The distribution of soak time was shown in Figure 4-3.
The screenshot of input files are shown in Figure 4-5. A batch script was written to automate the process of calculating emission and fuel consumption for each trip using CMEM. The estimation procedure was repeated for all the four categories.

```
# t(s), V(km/h)
1,0
2,3.9368
3,6.037
4,7.6362
5,8.7857
6,9.9556
7,11.464
8,13.449
9,15.975
10,19
11,22.358
12,25.817
13,29.152
14,32.18
15,34.735
```

(a)

```
# Soak time (Tsoak) is in minutes

IN_UNITS = METRIC
OUT_UNITS = METRIC
VEHICLE_CATEGORY = 4
Tsoak = 65
```

(b)

FIGURE 4-5 THE SCREENSHOT OF INPUT FILES FOR CMEM MODEL. (A) VEHICLE ACTIVITY FILE (B) CONTROL FILE.

CMEM model generates two output files for each trip: (a) Vehicle emission file, which contains second by second emission and fuel estimates in column format. Figure 4-6 shows the screenshot of the output data, which contains time (second), speed (km/h), acceleration (km/h/s), fuel, CO_2, CO, HC and NO_x, all in grams per second. However, this study focuses on carbon dioxide emission and fuel consumption only. (b) Summary file, which contains total distance travelled, total emissions, etc. The summary file was not used in the analysis.

1	0.00	0.953072	1.931560	0.368173	0.151664	0.003165
2	3.94	1.584260	3.415237	0.604779	0.195049	0.012868
3	6.04	1.509723	3.758078	0.553309	0.047790	0.011706
4	7.64	1.406854	3.497156	0.515608	0.045968	0.010132
5	8.79	1.310026	3.251561	0.480121	0.044252	0.008651
6	9.96	1.004083	2.475572	0.367994	0.038829	0.003970
7	11.46	1.092350	2.699449	0.400343	0.040394	0.005321
8	13.45	1.221129	3.026083	0.447540	0.042677	0.007291
9	15.98	1.377049	3.421560	0.504685	0.045440	0.009676
10	19.00	1.583150	3.944322	0.580220	0.049091	0.012830
11	22.36	1.839686	4.595016	0.674240	0.053633	0.016755
12	25.82	1.760354	4.393792	0.645165	0.052228	0.015541
13	29.15	1.785325	4.457131	0.654317	0.052670	0.015923
14	32.18	1.769278	4.416427	0.648435	0.052386	0.015678
15	34.74	1.710070	4.266247	0.626736	0.051338	0.014772

FIGURE 4-6 THE SCREENSHOT OF VEHICLE EMISSION FILE

4.4.2.2 Estimation using VT-Micro model

VT-Micro model requires csv file, which contains time (s), speed (km/h), and acceleration (km/h/s) for a trip for estimation. It has a graphical user interface (GUI) and has to be run in MATLAB. Since, the estimation of fuel and emission depends only on instantaneous speed and acceleration. A single csv file with speed profiles from all the trips was created and given as input for calculation. A column of acceleration was added next to speed's column before concatenating all trips. The screenshot of input file is shown in Figure 4-7, which shows few rows from the beginning and few rows from in-between. For the study, three categories of vehicle (LDV3 to LDV5) were chosen. We ran the input file in VT-Micro model for each of the three categories. The screenshot of VT-Micro GUI is also shown in Figure 4-8.

```
# t(s), V(km/h), a(km/h/s)
1,0,0
2,3.94,3.94
3,6.04,2.1
     .
     .
     .
95000,1.87,-1.05
95001,0,-1.87
95002,0,0
95003,2.47,2.47
     .
     .
     .
```

FIGURE 4-7 SCREENSHOT OF INPUT FILE FOR VT-MICRO

A single output file was generated corresponding to each vehicle category, which contains time (seconds), speed (km/h), acceleration (km/h/s), fuel (liters/sec), CO_2, CO, HC and NOx, all in grams per second, for all the trips. The unit of fuel was converted to grams/sec by multiplying with density (748.9 grams/liter).

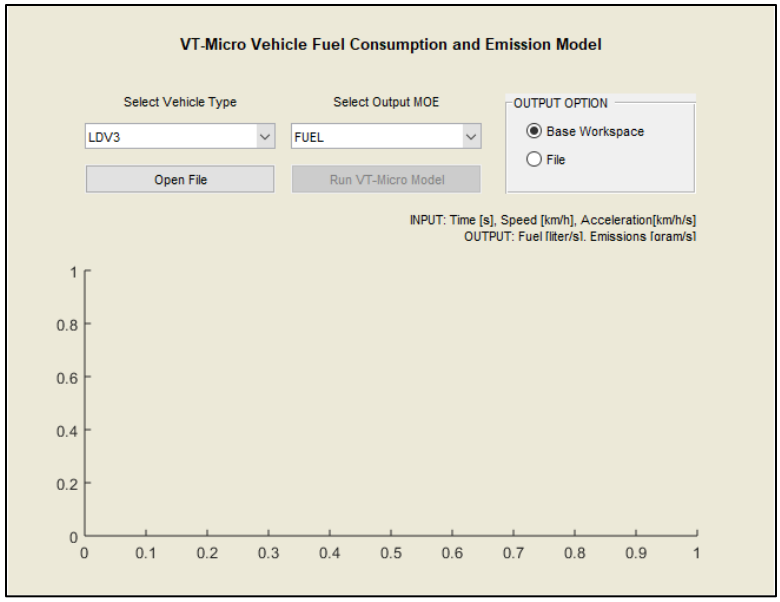


FIGURE 4-8 THE SCREENSHOT OF VT-MICRO GUI

4.4.3 Comparison of estimates from CMEM and VT-Micro

We had created a single table for both CMEM and VT-Micro, which contained trip ID, speed, acceleration, estimates corresponding to different vehicle categories. In order to understand the variation in estimates within categories of the same model and against the categories of another model. We calculated the range and mean of estimates for each second for each model for every trip. Figure 4-9 and Figure 4-10 shows the second by second mean (in dark lines) and range (in light bands) of estimates. Figure 4-9 corresponds to a trip with large soak time of 392 minutes while Figure 4-10 corresponds to trip with smaller soak time of 10 minutes.

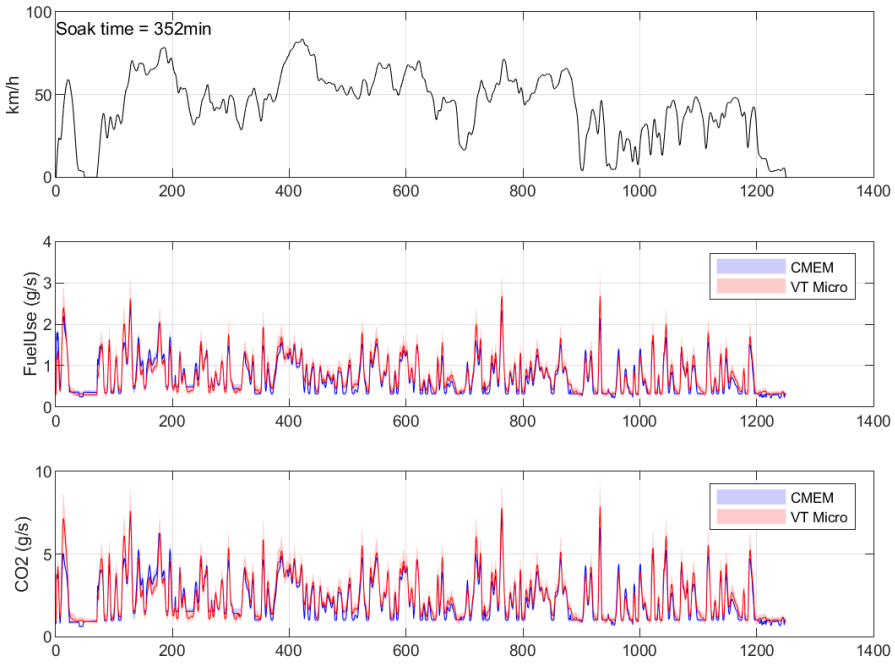


FIGURE 4-9 COMPARISON OF CMEM AND VT-MICRO SECOND BY SECOND ESTIMATES FOR A TRIP WITH HIGHER SOAK TIME

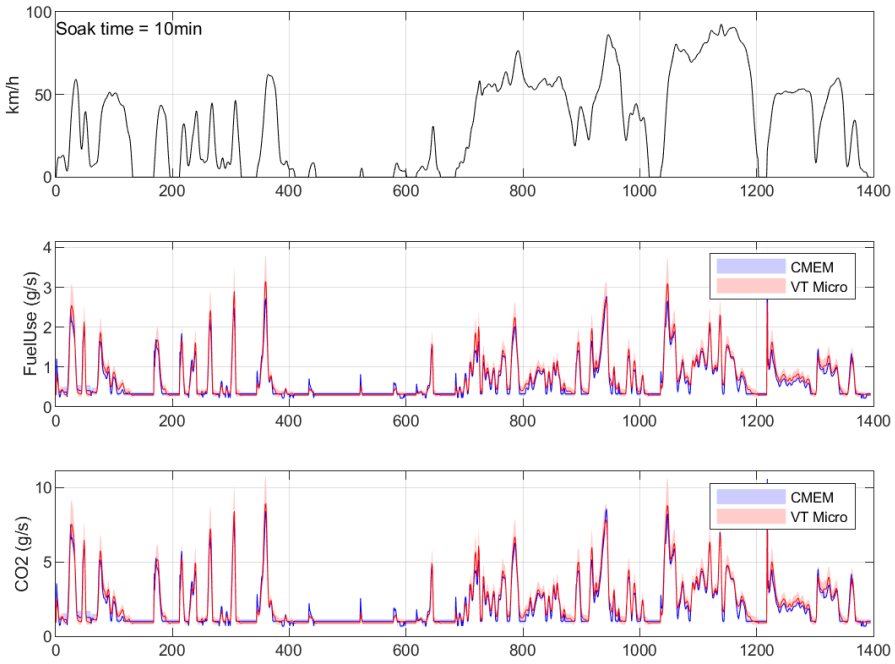


FIGURE 4-10 COMPARISON OF CMEM AND VT-MICRO SECOND BY SECOND ESTIMATES FOR A TRIP WITH LOWER SOAK TIME

4.5 PROBABILITY DISTRIBUTION OF ESTIMATES FOR EVERY BIN

For real time application, it may not be possible to run CMEM and VT-Micro models on speed profiles for different vehicle categories and soak times to estimate the mean and range of CO_2 and fuel. The market share of CV will also remain limited for a long period, which implies speed profiles will be available for only limited vehicles. It is important to reconstruct the second by second positions (therefore, speed and acceleration) of unconnected vehicles for estimating total CO_2 and fuel of the network. The speed profiles of unconnected vehicles can be reconstructed with uncertainties based on the market shares. However, the reconstruction of trajectories of unconnected vehicle is beyond the scope of this study. Here, we assumed that speed and accelerations can be estimated in multiples of 5 km/h and 1 km/h/s. We also assumed that roadside units receive second by second speed and acceleration of each CV in multiples of of 0.5 km/h and 0.25 km/h/s respectively. In addition, vehicle category and soak time cannot be determined. Hence, it is important to understand the variation in estimates of CO_2 and fuel rate within the defined bins, without knowledge of vehicle category and soak time. We first created distribution of estimates for unconnected vehicles for coarser bins (5 km/h and 1 km/h) and then same procedure was repeated for connected vehicles with finer bin size (0.5 km/h and 0.25 km/h/s). The procedure was same for both CMEM and VT-Micro model.
We rounded off speed and acceleration values to the multiples of five and one respectively and then developed normalized distributions (pdf) for each speed-acceleration bin for different

vehicle categories. The estimates from different vehicle categories were merged to create another distribution, which accounts for both vehicle category and soak time. Figure 4-11, Figure 4-12, Figure 4-13 and Figure 4-14 illustrates the distribution of fuel rates. Figure 4-11 shows the CMEM distribution of fuel rates for speed bin ranging from 27.5 km/h to 32.5 km/h and acceleration bin ranging from -0.5 km/h/s to 0.5 km/h/s. The second to the fifth plot were developed for different vehicle categories; the first distribution combines the distribution from different categories (assuming all categories are equally likely). These distributions are drawn from 152,013 seconds of data. Figure 4-12 shows the CMEM distribution for finer bins (29.75 to 30.25 km/h and -0.125 to 0.125 km/h/s), drawn from 3967 seconds of data. The x-axis range in Figure 4-11 and Figure 4-12 were kept same for better visual comparison. Similarly, Figure 4-13 and Figure 4-14 shows the distribution developed from VT-Micro model. There are few observations can be drawn from these distributions:

1. LDV5 category in VT-Micro has higher CO_2 and fuel consumption rate than other two categories (LDV3 and LDV4).
2. All four categories of CMEM model provides similar estimates of carbon dioxide and fuel.
3. VT-Micro model depends on instantaneous speed and acceleration and therefore, the distribution of estimates are very narrow for finer bins as shown in Figure 4-14.
4. CMEM model also depends upon several factors such as cold start. Therefore, distribution of estimates are wider for CMEM than VT-Micro for finer bins as shown in Figure 4-12 and Figure 4-14.

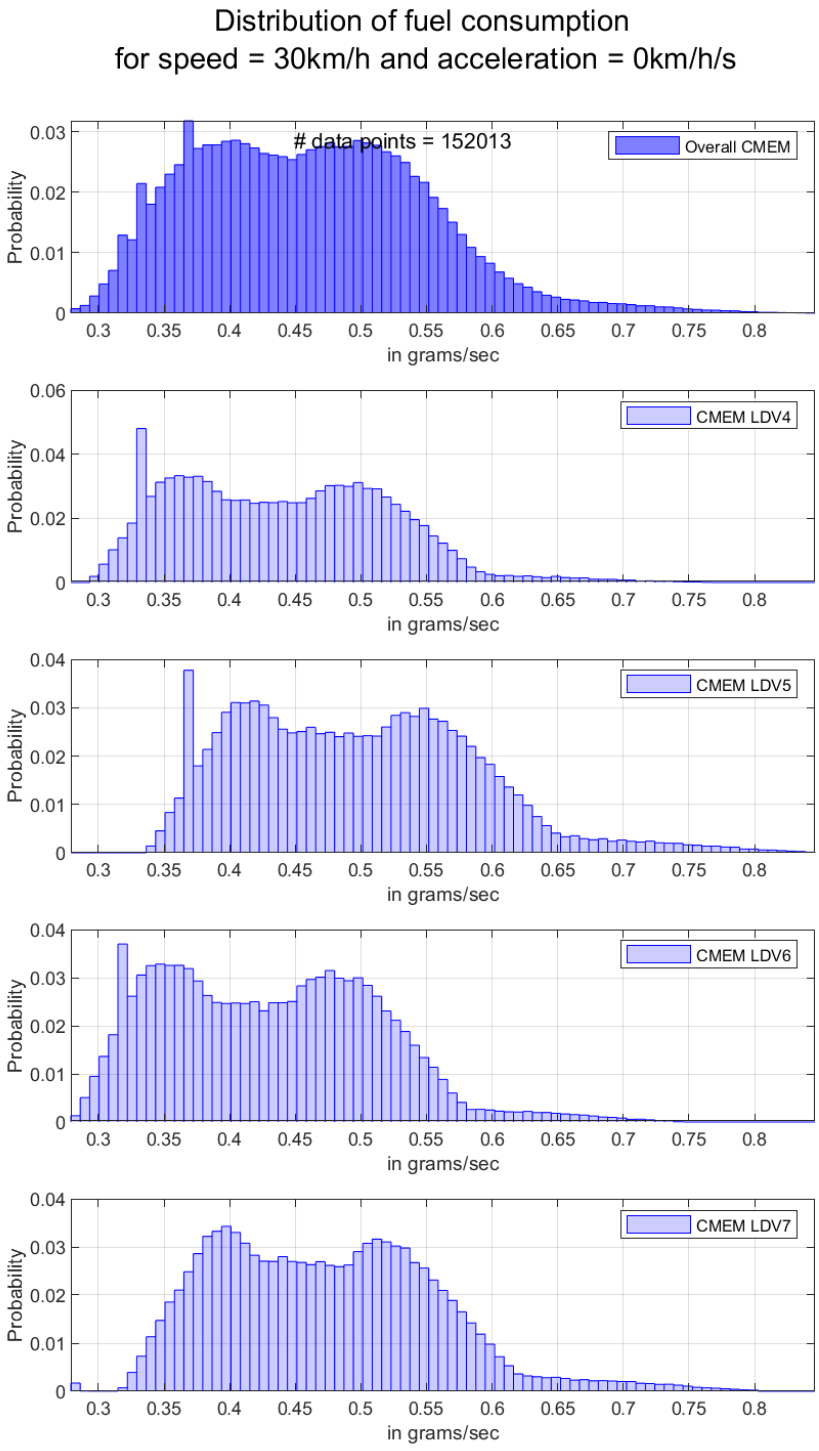


FIGURE 4-11 DISTRIBUTION OF CMEM FUEL RATE FOR 5 KM/H SPEED BIN AND 1 KM/H/S ACCELERATION BIN

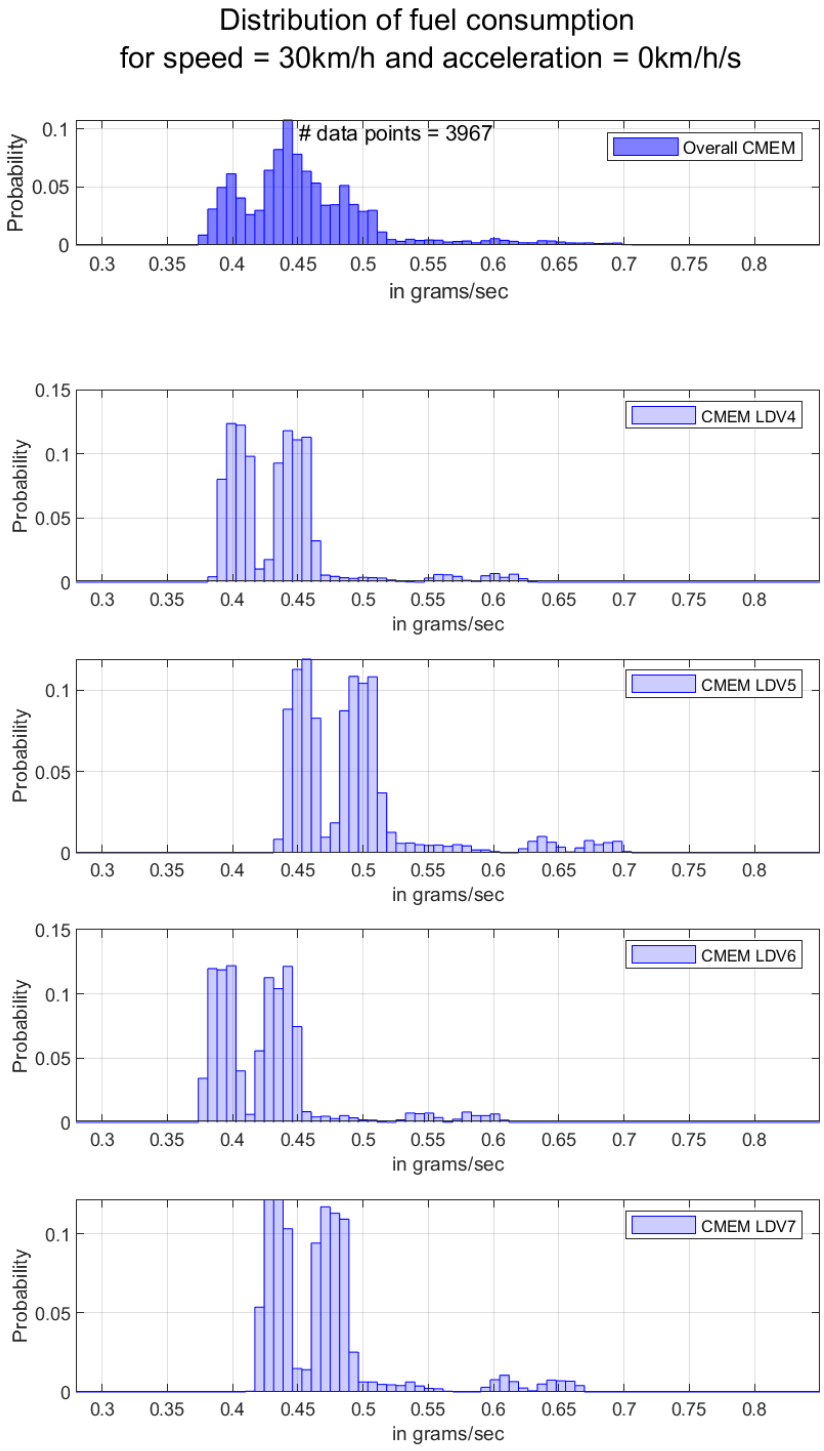


FIGURE 4-12 DISTRIBUTION OF CMEM FUEL RATE FOR 0.5 KM/H SPEED BIN AND 0.25 KM/H/S ACCELERATION BIN

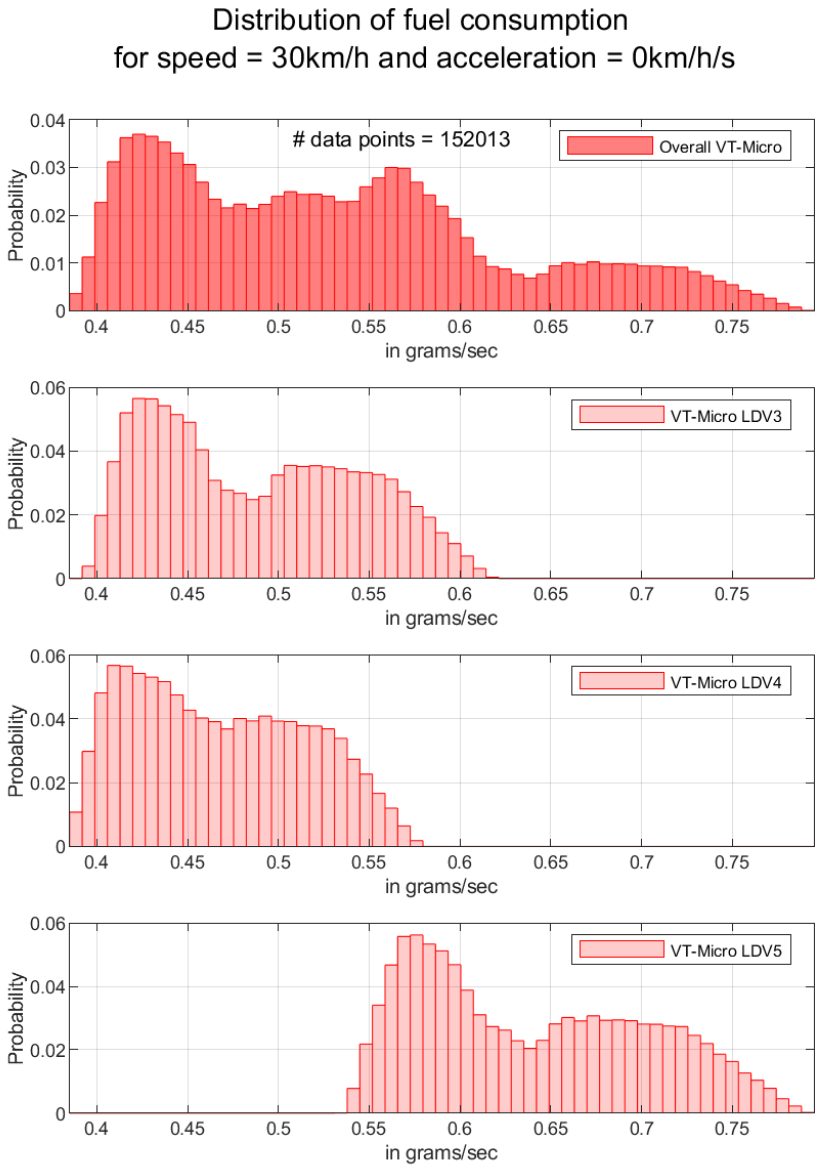


FIGURE 4-13 DISTRIBUTION OF VT-MICRO FUEL RATE FOR 5 KM/H SPEED BIN AND 1 KM/H/S ACCELERATION BIN

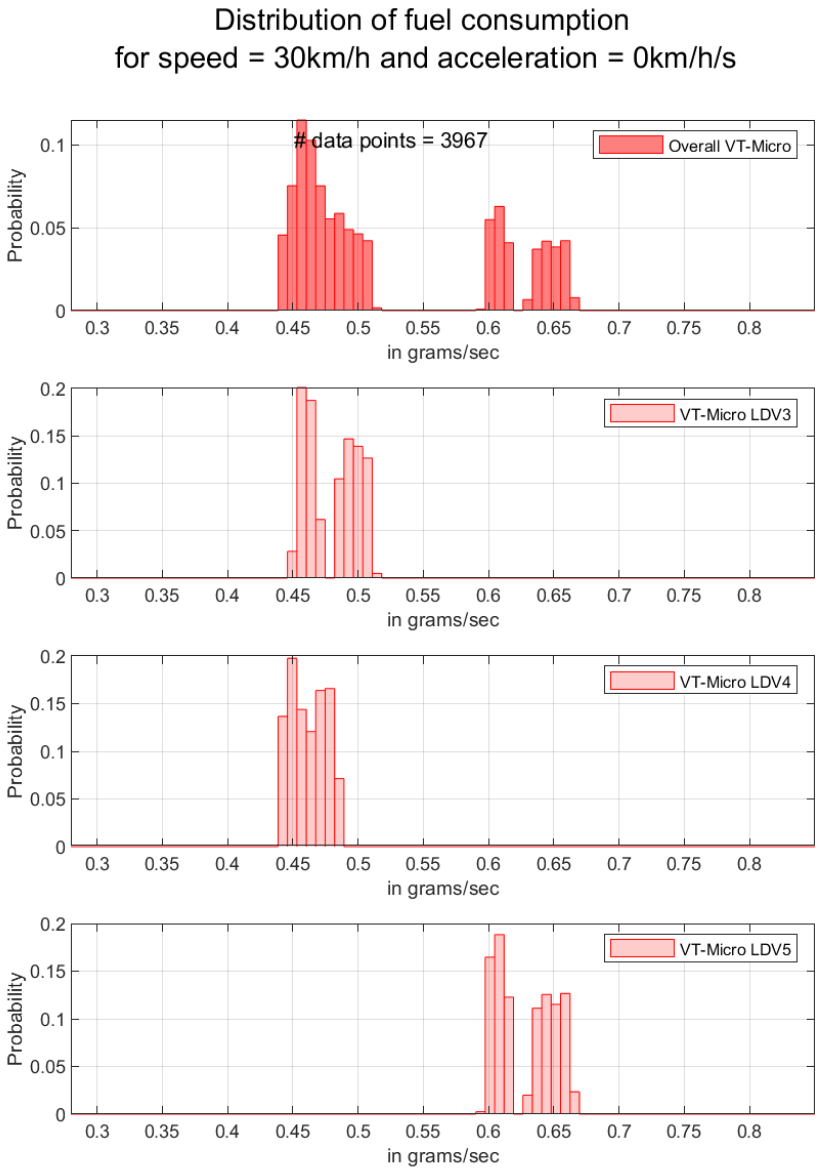


FIGURE 4-14 DISTRIBUTION OF VT-MICRO FUEL RATE FOR 0.5 KM/H SPEED BIN AND 0.25 KM/H/S ACCELERATION BIN

4.6 APPLICATION

We used another trajectory dataset (NGSIM) to demonstrate the application of this study in estimating total CO_2 and fuel consumption in 30 minutes for a section of network.

4.6.1 NGSIM dataset

We used the NGSIM vehicle trajectory dataset collected on a segment of Peachtree Street, Atlanta, Georgia with speed limit was 35 mph. The study section was approximately 0.64 km in length with two/three lanes in in each direction and had five intersections. A total of 30 minutes of trajectory data was processed from video collected on November 2008 for two 15 minutes interval – 12:45 PM to 1:00 PM and 4:15 PM to 4:30 PM. Vehicles were classified in three categories a) motorcycle 2) automobile and 3) bus and truck, 98% of observed vehicles were automobile.
In the dataset, vehicle trajectories are provided at resolution of 10 frames per second. Second by second speed and accelerations of all the vehicles were extracted. The data set consist of 87,394 seconds from 1,543 vehicles. The distributions of instantaneous speed and acceleration are shown in Figure 4-15.

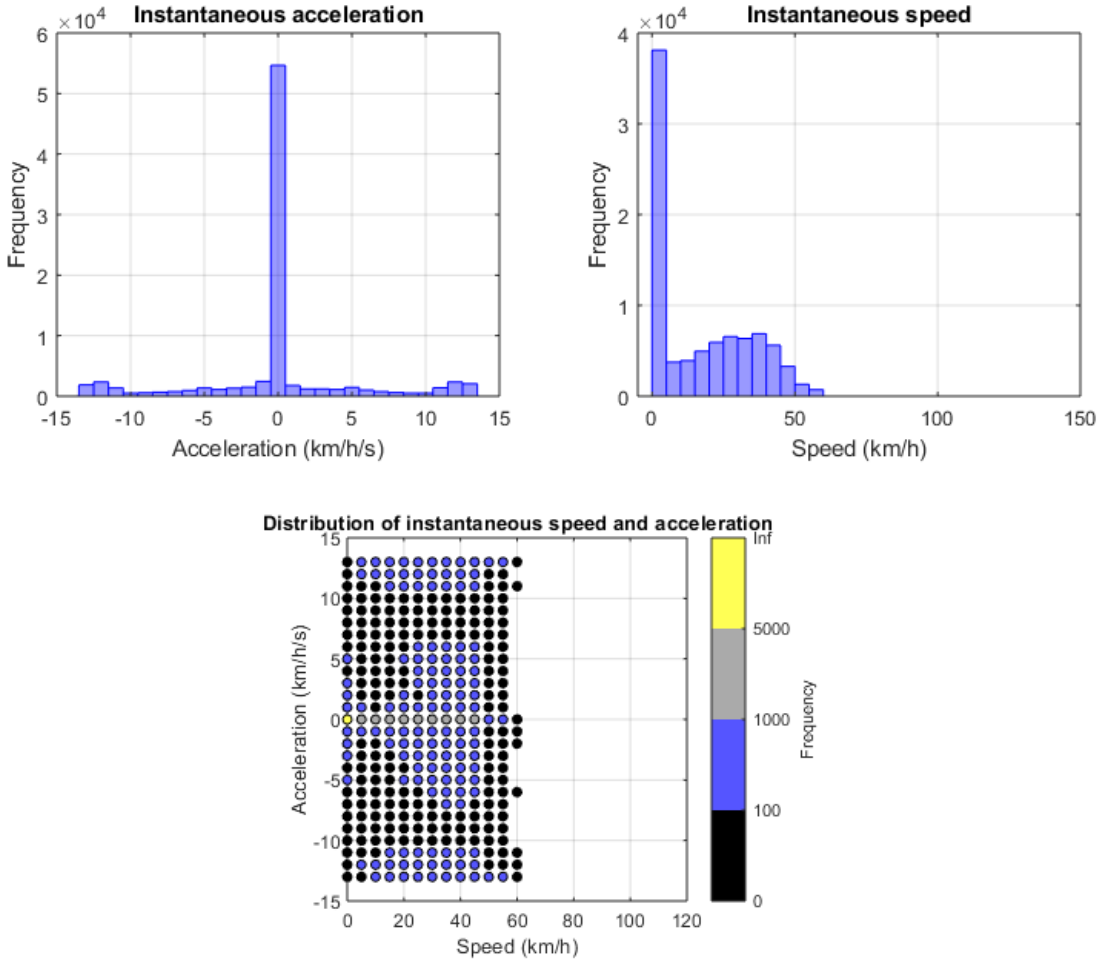


FIGURE 4-15 DISTRIBUTION OF SPEED AND ACCELERATION IN NGSIM TRAJECTORY DATASET

4.6.2 Estimation of total CO2 and fuel consumption for different market shares

We calculated the total CO_2 emission and fuel consumption for different market share (30%, 50%, 70% and 100%). The speed profiles were split into connected vehicles and unconnected vehicles based on the market share. The speed and acceleration are rounded off to 0.5 km/h and 0.25 km/h/s for connected vehicles and to 5 km/h and 1 km/h/s for unconnected vehicles. Later the distributions developed in previous section were used for estimation. For each timestamp, a random value is drawn from the distributions of CO_2 emission and fuel consumption of corresponding bin, which are then summed over all the trips for every second. We simulated estimation procedure 1000 times to derive distribution of total CO_2 and fuel for each market share. Figure 4-16, Figure 4-17, Figure 4-18 and Figure 4-19 shows the distributions of total estimates for different market shares. The following observations can be made from the distributions:

1. It is worth noting that the distributions of total estimates are very narrow, which suggest that sum of various random values of estimates drawn from the distributions of corresponding speed-acceleration bin is almost equivalent to number of points in the bin times the mean estimate value of corresponding bin.
2. The distribution of total CO_2 estimates of any market share of CMEM model is narrower/tighter than VT-Micro model as shown in Figure 4-18 and Figure 4-19. It can be attributed to the similar estimates obtained from different vehicle category in CMEM as compared VT-Micro.
3. It can also be observed that CMEM distributions for different market shares do not overlap as compared to VT-Micro model.

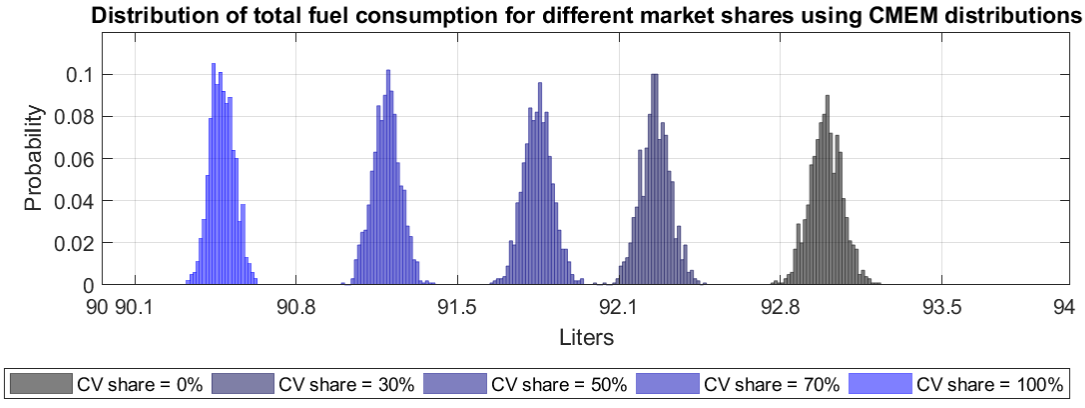


FIGURE 4-16 DISTRIBUTION OF TOTAL FUEL CONSUMPTION FOR DIFFERENT MARKET SHARE USING CMEM DISTRIBUTION

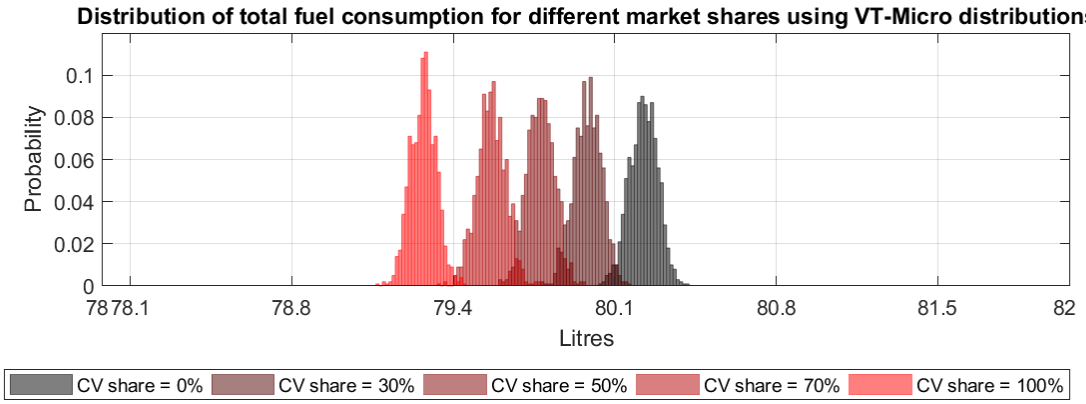


FIGURE 4-17 DISTRIBUTION OF TOTAL FUEL CONSUMPTION FOR DIFFERENT MARKET SHARE USING VT-MICRO DISTRIBUTION

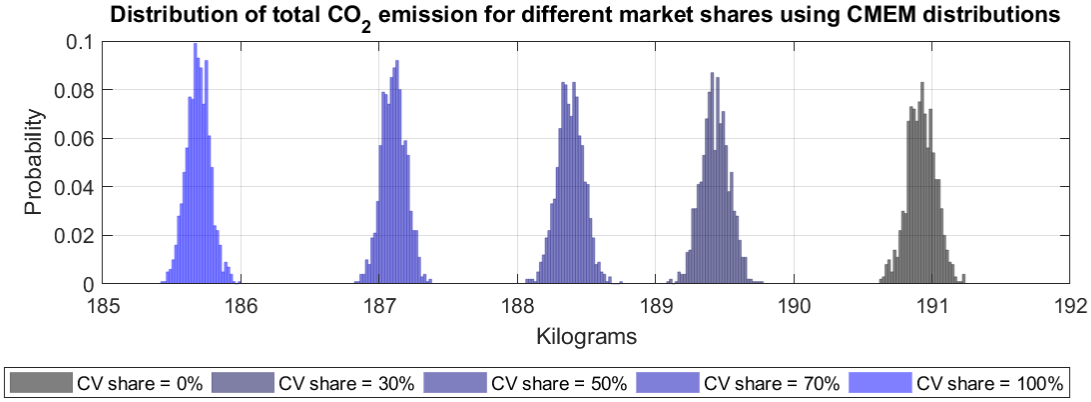


FIGURE 4-18 DISTRIBUTION OF TOTAL CO_2 EMISSION FOR DIFFERENT MARKET SHARE USING CMEM DISTRIBUTION

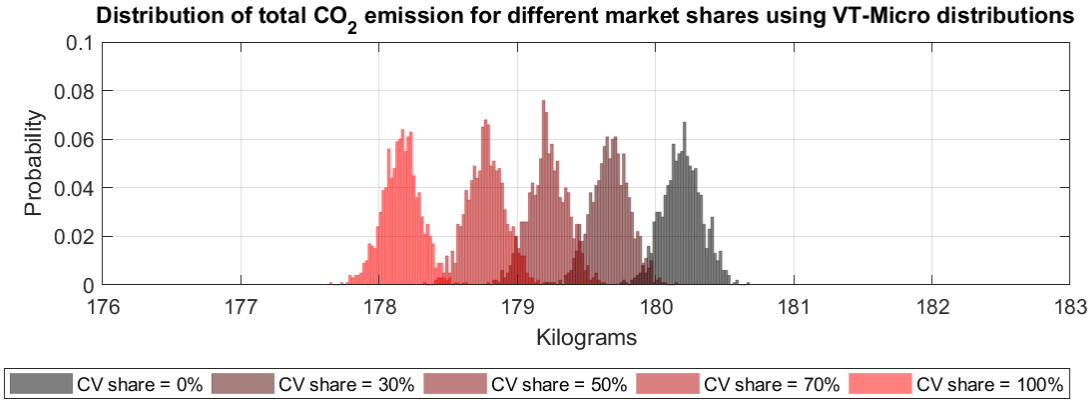


FIGURE 4-19 DISTRIBUTION OF TOTAL CO_2 EMISSION FOR DIFFERENT MARKET SHARE USING VT-MICRO DISTRIBUTION

The development of distributions of total CO_2 and fuel requires random draw from distribution of estimates for each speed and acceleration bin. Multiplying number of points in each bin with

mean estimate of the bin (as shown in Figure 4-20, Figure 4-21 and Table 4-1) can also estimate the total estimates. Table 4-2 shows the total estimates for 0% and 100% market share of CVs. This process is very fast and the results obtained confirms with the distribution of total estimates.

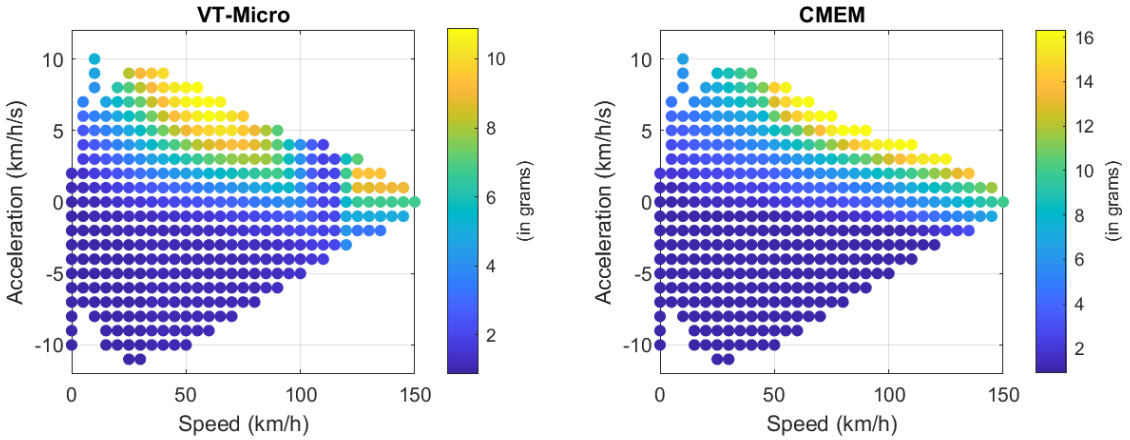


FIGURE 4-20 ILLUSTRATES THE MEAN VALUES OF DISTRIBUTION OF CO_2 ESTIMATES

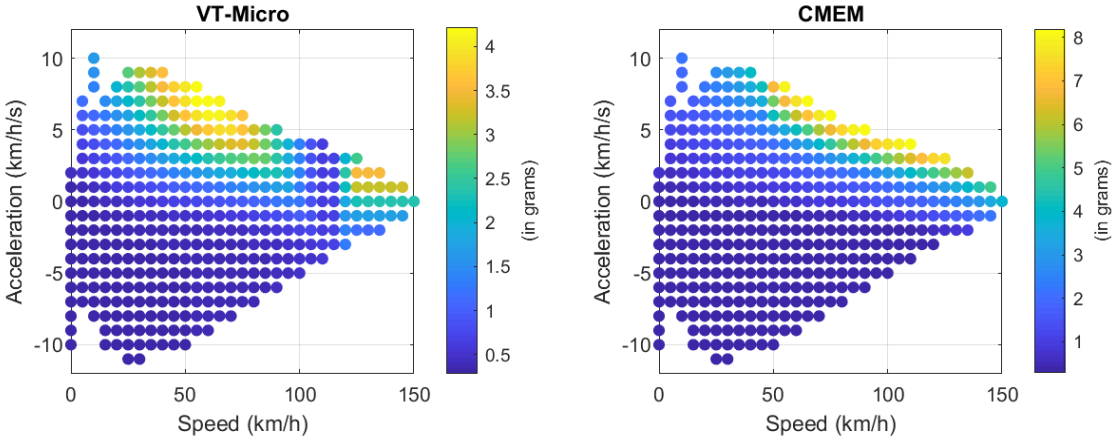


FIGURE 4-21 ILLUSTRATES THE MEAN VALUES OF DISTRIBUTION OF FUEL ESTIMATES

TABLE 4-1 MEAN OF ESTIMATES OF EMISSION AND FUEL CONSUMPTION FOR SELECTED BINS

Speed (km/h)	Acceleration (km/h/s)	CO_2 (g)		Fuel consumption (g)	
		CMEM	VT	CMEM	VT
0	0	1.06	0.93	0.34	0.29
15	-2	1.04	1.01	0.34	0.33
	0	1.15	1.26	0.37	0.41
	2	2.34	2.02	0.77	0.65

30	-2	1.05	1.09	0.34	0.36
	0	1.43	1.62	0.46	0.53
	2	2.90	3.11	0.94	1.00
45	-2	1.07	1.17	0.34	0.38
	0	1.62	2.01	0.52	0.66
	2	3.47	4.23	1.12	1.37

TABLE 4-2 TOTAL CO_2 AND FUEL ESTIMATED USING MEAN VALUES OF DISTRIBUTIONS FOR DIFFERENT MARKET SHARE OF CV

	0% CV share		100% CV share	
	Total Fuel consumption (liters)	Total CO_2 emission (kilograms)	Total Fuel consumption (liters)	Total CO_2 emission (kilograms)
CMEM	92.99	190.93	90.49	185.69
VT-Micro	80.24	180.20	79.33	178.16

4.7 SUMMARY AND CONCLUSION

The study investigates the application of limited connected vehicle data to estimate to real time CO_2 emission and fuel consumption of the transportation network. The microscopic emission models require second by second speed profile, vehicular characteristics, other parameters related engine cold start, etc. for estimation. CVs can provide transmit second by second speed and acceleration values, but it may not share other vehicular information. In addition, the share of connected vehicle is expected to be low.

This study applied CMEM and VT-Micro model to a large trajectory dataset for different vehicle categories and soak time. Later, it assumes that speed and acceleration can be determined in multiples of fixed values for both connected and unconnected vehicles. Therefore, we defined coarser for unconnected vehicles and finer bins connected vehicles for speed and acceleration and created distribution of emission estimates corresponding to each bin. Later we used these distributions to calculate total carbon dioxide and fuel estimates for another trajectory dataset. The study contributes in developing a methodological framework for estimating emission and fuel consumption with limited CV data. The current study has made following assumptions: 1) Road grades are assumed to be zero. 2) Speed and acceleration of unconnected vehicles are assumed to be determined in multiples of 5 km/h and 1 km/h/s, which should be reasonable for dense traffic but needs to be revised for low traffic volume network. 3) Limited vehicle categories are considered for the emission models. The estimation can be further improved by utilizing better technique to estimate speed profiles of unconnected vehicles and by incorporating road grade information in emission models.

5. REFERENCE LIST

AASHTO. (2011). *A Policy on Geometric Design of Highways and Streets, 6th Edition.*

AASHTO. (2012). SCOPM Task Force Findings on National-Level Performance Measures.

Abdel-Aty, M., & Pande, A. (2005). Identifying crash propensity using specific traffic speed conditions. *Journal of Safety Research*, *36*(1), 97–108.

Abianeh, A. S., Burris, M. W., Talebpour, A., & Sinha, K. C. (2009). The Impacts of Connected Vehicles on Fuel Consumption, and Traffic Operation under Recurring and Nonrecurring Congestion. *Journal of Transportation Engineering*.

Ahn, K., *Kronprasert*, N., & Rakha, H. (2009). Energy and environmental assessment of high-speed roundabouts. *Transportation Research Record*, *2123*(1), 54-65.

Ahn, K., Rakha, H., Trani, A., & Van Aerde, M. (2002). Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels. Journal of Transportation Engineering, 128(2), 182-190.

Al-Kaisy, A. & Durbin,C. (2009). Platooning on Two-Lane Two-Way Highways: An Empirical Investigation. *Journal of Advanced Transportation*, *43*, 71–88.

Al-Ruithe, M., & Benkhelifa, E. (2017). A Conceptual Framework for Cloud Data Governance-Driven Decision Making. IEEE.

Al-Shamisi, A., Louvieris, P., Al-Maulla, M., & Mihajlov, M. (2016). Towards a theoretical framework for an active cyber situational awareness model. IEEE.

Al-Sobky, A.-S. A., & Mousa, R. M. (2016). Traffic density determination and its applications using smartphone. Alexandria Engineering Journal, 55(1), 513-523. Traffic density determination and its applications using smartphone

Argote, J., Christofa, E., Xuan, Y., & Skabardonis, A. (2012). Estimation of Arterial Measures of Effectiveness with Connected Vehicle Data. 91st Annual Meeting Transportation Research Board.

Arif, F., & Bayraktar, M. E. (2012). Theoretical framework for transportation infrastructure asset management based on review of best practices. Construction Research Congress © ASCE.

Arthur, D., & Vassilvitskii, S. (2007). K-Means++: the Advantages of Careful Seeding. *Proc ACM-SIAM Symposium on Discrete Algorithms.*, *8*, 1027–1035.

Arvin, R., Kamrani, M., & Khattak, A. J. (2019). How instantaneous driving behavior contributes to crashes at intersections: Extracting useful information from connected vehicle message data. *Accident Analysis and Prevention*, *127*(January), 118–133.

Azizi, L., Iqbal, M. S., & Hadi, M. A. (2018). Estimation of Freeway Platooning Measures Using Surrogate Measures Based on Connected Vehicle Data. *97st Annual Meeting of Transportation Research Board,, Washington DC*.

Balas, V. E., & Balas, M. M. (2007). Driver assisting by inverse time to collision. *2006 World Automation Congress, WAC'06*, 1–6.

Barth, M., An, F., Norbeck, J., & Ross, M. (1996). Modal emissions modeling: A physical approach. *Transportation Research Record*, *1520*(1), 81-88.

Barth, M., & Boriboonsomsin, K. (2009). Energy and emissions impacts of a freeway-based dynamic eco-driving system. Transportation Research Part D: Transport and Environment, 14(6), 400-410. Energy and emissions impacts of a freeway-based dynamic eco-driving system.

Bishop, Ch. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Boriboonsomsin, K., Barth, M. J., Zhu, W., & Vu, A. (2012). Eco-routing navigation system based on multisource historical and real-time traffic information. *IEEE Transactions on Intelligent Transportation Systems*, *13*(4), 1694-1704.

Bullen, A. G. R. (1972). Strategies for practical expressway control. *Journal of the Transportation Engineering Division*, *98*(3), 599–605.

Chang, G. L., & Xiang, H. (2003). The relationship between congestion levels and accidents. *State Highway Adiministration*.

Chatterjee, S., Roy, D., Chakraborty, S., & Roy, S. K. (2017). Level of Service Criteria on Indian Multilane Highways based on Platoon Characteristics. *96st Annual Meeting of Transportation Research Board, Washington DC*.

Chen, X., Li,Z. Li,L., & Qixin. S. (2014). A Traffic Breakdown Model Based on Queuing Theory. Network and Spatial economics. *Springer*, *14*, 485–504.

Chowdhury, N., Hossain, Z., Yang, S., & Braham, A. (2015). A Framework for Premature Pavement Distress Evaluation. ASCE.

Coelho, M. C., Farias, T. L., & Rouphail, N. M. (2006). Effect of roundabout operations on pollutant

emissions. *Transportation Research Part D: Transport and Environment*, *11*(5), 333-343.

Dehman, A. (2014). Breakdown Maturity Phenomenon at Wisconsin Freeway Bottlenecks. *Transportation Research Record: Journal of the Transportation Research Board*, *2395*(1), 1–11.

De Kluijver, H., & Stoter, J. (2003). Noise mapping and GIS: optimising quality and efficiency of noise effect studies. Computers, Environment and Urban Systems, 27(1), 85-102.

Dijkstra, A., Marchesini, P., Bijleveld, F., Kars, V., Drolenga, H., & Maarseveen, M. (2010). Do Calculated Conflicts in Microsimulation Model Predict Number of Crashes? *Transportation Research Record: Journal of the Transportation Research Board*, *2147*(1), 105–112.

Dong, J., & Mahmassani, H. S. (2012). Stochastic modeling of traffic flow breakdown phenomenon: Application to predicting travel time reliability. *IEEE Transactions on Intelligent Transportation Systems*, *13*(4), 1803–1809.

Duran, A., & Earleywine, M. (2013). *GPS* data *filtration method for drive cycle analysis applications* (Vol. 1, No. NREL/CP-5400-53865). National Renewable Energy Lab. (NREL), Golden, CO (United States).

Elefteriadou, L. (2017). Freeway-Merging Operations: A Probabilistic Approach. *IFAC Proceedings Volumes*, *30*(8), 1283–1288.

Elefteriadou, L., Roger, P. R., & Mcshane, W. R. (1995). Probabilistic Nature of Breakdown at Freeway Merge Junctions. *Transportation Research Record: Journal of the Transportation Research Board*, *1484*, 80–89.

Elfar, A., Talebpour, A., & Mahmassani, H. S. (2018). Machine Learning Approach to Short-Term Traffic Congestion Prediction in a Connected Environment. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(45), 185–195.

Energy, U. S. D. (2008). Annual Energy Outlook 2008, With Projection to 2030. Energy Inf. Admin., Washington, DC, Rep. DOE/EIA-0383 (2008).

EPA, US. (2008). *Inventory of US Greenhouse Gas Emissions and Sinks: 1990-2006*. US Environmental Protection Agency.

EPA, US. (2014). Motor Vehicle Emission Simulator (MOVES) User Guide. *US Environmental Protection Agency*.

Fan, R., Wang, W., Liu, P., & Yu, H. (2013). Using VISSIM simulation model and Surrogate Safety Assessment Model for estimating field measured traffic conflicts at freeway merge areas.

IET Intelligent Transport Systems, *7*(1), 68–77.

FHWA. (2007). US Highway 101 Dataset.

FHWA. (2017a). *Transportation Performance Management.* Federal Highway Administration Transportation Performance Management

Fontana, N., Giugni, M., Glielmo, L., Marini, G., & Verrilli, F. (2018). Real-Time Control of a PRV in Water Distribution Networks for Pressure Regulation: Theoretical Framework and Laboratory Experiments. Journal of Water Resources Planning and Management, 144(1). Real-Time Control of a PRV in Water Distribution Networks for Pressure Regulation: Theoretical Framework and Laboratory Experiments. Journal of Water Resources Planning and Management

Frey, H., Unal, A., Chen, J., Li, S., & Xuan, C. (2002). Methodology for developing modal emission rates for EPA's multi-scale motor vehicle & equipment emission system.

Fwa, T. F., Chan, W. T., & Lim, C. T. (1997). Decision framework for pavement friction management of airport runways. Journal of Transportation Engineering, 123, 429-435.

Gao, K., Han, F., Dong, P., Xiong, N., & Du, R. (2019). Connected Vehicle as a Mobile Sensor for Real Time Queue Length at Signalized Intersections. *Sensors*, *19*(9), 2059.

Gattis, J., Alguire,M., Townsend, K. & Rao, S. (1997). Rural Two-Lane Passing Headways and Platooning. *Transportation Research Record*, *1579*, 27–34.

Genders, W., & Razavi, S. N. (2016). Impact of Connected Vehicle on Work Zone Network Safety through Dynamic Route Guidance. Journal of Computing in Civil Engineering, 30(2). Impact of Connected Vehicle on Work Zone Network Safety through Dynamic Route Guidance.

Gökalp, M. O., Kayabay, K., Akyol, M. A., Eren, P. E., & Koçyigit, A. (2016). Big data for industry 4.0: A conceptual framework. IEEE. Big data for industry 4.0: A conceptual framework

Gomes, G., May, A., & Horowitz, R. (2004). Calibration of VISSIM for a Congested Freeway (UCB-ITS-PRR-2004-4). UC Berkeley: California Partners for Advanced Transportation Technology: https://escholarship.org/uc/item/7bs9b2v3

Gómez, R., Camarero, A., & Molina, R. (2015). Development of a Vessel-Performance Forecasting System: Methodological Framework and Case Study. Journal of Waterway, Port, Coastal, and Ocean Engineering, 142(2), 04015016.

Gong, J., Gordon, C., & Azambuja, M. (2012). A Conceptual Framework for Assessing Climate-

Related Heat Effects on Craft Time Utilization in the Construction Industry. ASCE.

Green, R. A., & Lee, J. (2005). Computation of number of equivalent strain cycles: A theoretical framework. ASCE.

Griffis, L., Patel, V., Muthukumar, S., & Baldava, S. (2013). A Framework for Performance-based Wind Engineering. ASCE and ATC.

Guido, G., Vitale, A., Astarita, V., Saccomanno, F., Giofré, V. P., & Gallelli, V. (2012). Estimation of Safety Performance Measures from Smartphone Sensors. *Procedia - Social and Behavioral Sciences*, *54*, 1095–1103.

Hale, D. K., Button, L., Argote, J., Joshi, C., Zhang, X., Kondyli, A., & Bared, J. G. (2019). Calibration of Microsimulation Car-Following Models for Narrow Freeway Lanes. *98st Annual Meeting of Transportation Research Board, Washington DC*.

Han, Y., & Ahn, S. (2017). Stochastic Modeling of Breakdown at Freeway Merge Bottleneck. *96th Annual Meeting Transportation Research Board, Washington DC*.

Hao, P., Ban, X., Guo, D., & Ji, Q. (2014). Cycle-by-cycle intersection queue length distribution estimation using sample travel times. Transportation Research Part B: Methodological, 68, 185-204. Cycle-by-cycle intersection queue length distribution estimation using sample travel times

Hartman,K., Goplakrishna,D. & Kitchener, F. M. (2016). Performance Measurement and Evaluation Support Plan ICF/Wyoming CV Pilot.

Hayward, J. C. (1972). Near-miss determination through use of a scale of danger. *Highway Res. Rec.*, *384*(384), 22–34.

Hale. J. (2018). Scale, Standardize, or Normalize with Scikit-Learn – Towards Data Science. Accessed from https://towardsdatascience.com/scale-standardize-or-normalize-with-scikit-learn-6ccc7d176a02

Henclewood, D., Abramovich, M., Yelchuru, B. (2014). Safety Pilot Model Deployment-one Day Sample Data Environment Data Handbook. *USDOT Research and Technology Innovation Administrations*.

Herman, R., Montroll, E. W., Potts, R. B., & Rothery, R. W. (1959). Traffic Dynamics: Analysis of Stability in Car Following. *Operations Research*, *7*(1), 86–106.

HCM (2016). *Highway Capacity Manual, Sixth Edition: A Guide for Multimodal Mobility Analysis.* Transportation Research Board, Washington DC, USA.

Hobeika, A. G., Jung, H., & Bae, S. (2015). Contribution of Aggressive Drivers to Automobile Tailpipe Emissions under Acceleration and Braking Conditions. Journal of Transportation Engineering, 141(2). Contribution of Aggressive Drivers to Automobile Tailpipe Emissions under Acceleration and Braking Conditions

Hobeika, A. G., Sivanandan, R., Subramaniam, S., Ozbay, K., & Zhang, Y. (1993). Real-time traffic diversion model: Conceptual approach. Journal of Transportation Engineering, 119, 515-534.

Ibrahim, S., Kalathil, D., Sanchez, R. O., & Varaiya, P. (2017). Estimating Phase Duration for SPaT Messages. arXiv preprint arXiv:1710.05394.

Iqbal, M. S., Hadi, M., & Xiao, Y. (2018). Effect of Link-Level Variations of Connected Vehicles (CV) Proportions on the Accuracy and Reliability of Travel Time Estimation. IEEE Transactions on Intelligent Transportation Systems, 1-10. Effect of Link-Level Variations of Connected Vehicles (CV) Proportions on the Accuracy and Reliability of Travel Time Estimation

ITS DataHub. (2019). ITS DataHub, Intelligent Transportation System, US Department of Transportation. Accesed from: ITS DataHub, Intelligent Transportation System

Izadpanah, P., Hellinga, B., & Fu, L. (2011). Real-Time Freeway Travel Time Prediction Using Vehicle Trajectory Data. 90th Annual Meeting Transportation Research Board.

Jiang, R., Hu, M.B., Zhang, H.M., Gao, Z.Y., & Jia,B.(2015). On Some Experimental Features of Car-Following Behavior and How to Model Them. *Transportation Research Part B*, *80*, 338–354.

Jiang, R., Jin, C. J., Zhang, H. M., Huang, Y. X., Tian, J. F., Wang, W., & Jia, B. (2018). Experimental and empirical investigations of traffic flow instability. *Transportation Research Part C: Emerging Technologies*, *94*, 83–98.

Jingxin, X., & Mei, C. (2007). Defining traffic flow phases using intelligent transportation systems-generated data. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *11*(1), 15–24.

Johnson, R. A., & Wichern, D. W. (2007). *Appllied Multivariate Statistics Analysis*.

Joumard, R., André, M., Vidon, R., Tassel, P., & Pruvost, C. (2000). Influence of driving cycles on unit emissions from passenger cars. *Atmospheric environment*, *34*(27), 4621-4628.

Kamrani, M., Arvin, R., & Khattak, A. J. (2018). Extracting Useful Information from Basic Safety Message Data: An Empirical Study of Driving Volatility Measures and Crash Frequency at

Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(38), 290–301.

Karlaftis, M. G., Zografos, K. G., Papastavrou, J. D., & Charnes, J. M. (1996). Methodological framework for air-travel demand forecasting. Journal of Transportation Engineering, Part A: Systems, ASCE, 122, 96-104.

Khajeh Hosseini, M., & Talebpour, A. (2019). Traffic Prediction using Time-Space Diagram: A Convolutional Neural Network Approach. *Transportation Research Record*.

Khan, S. M. (2015). Real-Time Traffic Condition Assessment With Connected Vehicles. (Master of Science in Civil Engineering), Clemson University.

Khan, S. M., Dey, K. C., & Chowdhury, M. (2017). Real-Time Traffic State Estimation With Connected Vehicles. IEEE Transactions on Intelligent Transportation Systems, 18(7), 1687-1699. Real-Time Traffic State Estimation With Connected Vehicles

Khattak, A. J., & Wali, B. (2017). Analysis of volatility in driving regimes extracted from basic safety messages transmitted between connected vehicles. Transportation Research Part C: Emerging Technologies, 84, 48-73. Analysis of volatility in driving regimes extracted from basic safety messages transmitted between connected vehicles

Khazraeian, S., Hadi, M., & Xiao, Y. (2017). Assessment of the Benefits of Queue Warning in a Connected Vehicle Environment based on Surrogate Safety Measures. TRB 2017 Annual Meeting.

Kianfar, J., & Edara, P. (2013). A Data Mining Approach to Creating Fundamental Traffic Flow Diagram. *Procedia - Social and Behavioral Sciences*, *104*, 430–439.

Kidando, E., Moses, R., Kitali, A. E., Kwigizile, V., Lyimo, S. M., Chimba, D., & Sando, T. (2018). Exploring the influence of rainfall on a stochastic evolution of traffic conditions. *97th Annual Meeting Transportation Research Board, Washington DC*, 1–17.

Kidando, E., Moses, R., & Sando, T. (2019). A Statistical Approach for Estimating Speed Threshold for Traffic Breakdown Event Identification: a Model Accounting for Data Variations. *98th Annual Meeting Transportation Research Board,Washington DC*.

Ko, J., & Guensler, R. (2005). Characterization of Congestion Based on Speed Distribution: A Statistical Approach Using Gaussian Mixture Model. *84st Annual Meeting of Transportation Research Board, Washington DC*.

Kondyli, A. (2009). Breakdown Probability Model At Freeway-Ramp Merges Based On Driver Behavior. *University of Florida*.

Kondyli, Alexandra, Elefteriadou, L., Brilon, W., Hall, F. L., Persaud, B., & Washburn, S. (2013). Development and Evaluation of Methods for Constructing Breakdown Probability Models. *Journal of Transportation Engineering*, *139*(9), 931–940.

Krauss, S. , Wanger,P. &Gawron C. (1997). Metastable States in a Microscopic Model of Traffic Flow. *Physical Review E*, *55*(5), 5597.

Kuang, Y., Qu, X., & Wang, S. (2015). A tree-structured crash surrogate measure for freeways. *Accident Analysis & Prevention*, *77*, 137–148.

Kühne, R. (1984). Macroscopic Freeway Model for Dense Traffic Stop-Start Waves and Incident Detection. *9 Th International Symposium on Transportation and Traffic Theory*, 21–42.

Kwak, J., Park, B., & Lee, J. (2012). Evaluating the impacts of urban corridor traffic signal optimization on vehicle emissions and fuel consumption. *Transportation Planning and Technology*, *35*(2), 145-160.

Laflamme, E. M., & Ossenbruggen, P. J. (2017). Effect of time-of-day and day-of-the-week on congestion duration and breakdown: A case study at a bottleneck in Salem, NH. *Journal of Traffic and Transportation Engineering (English Edition)*, *4*(1), 31–40.

Lay, M. G. (1998). Handbook of Road Technology. Volume 2: Traffic and Transport. Gordon and Breach Science Publishers, London (England).

Lee, C., Saccomanno, F., & Hellinga, B. (2002). Analysis of Crash Precursors on Instrumented Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, *1784*(1), 1–8.

Li, F., Etienne, A., Siadat, A., & Vernadat, F. (2015). BCVR: A methodological framework for industrial performance management and decision-support, 1258-1267.

Li, J.-Q., Zhou, K., Shladover, S. E., & Skabardonis, A. (2013). Estimating Queue Length under the Connected Vehicle Technology: Using Probe Vehicle, Loop Detector, and Fused Data. TRB 2013 Annual Meeting.

Liu, J., & Khattak, A. J. (2016). Delivering improved alerts, warnings, and control assistance using basic safety messages transmitted between connected vehicles. Transportation Research Part C: Emerging Technologies, 68, 83-100. Delivering improved alerts, warnings, and control assistance using basic safety messages transmitted between connected vehicles

Liu, J., Love, P. E. D., Davis, P. R., Smith, J., & Regan, M. (2014). Conceptual Framework for the Performance Measurement of Public-Private Partnerships. ASCE. doi:10.1061/(ASCE)IS

Liu, J. Y., Low, S. P., & Yang, J. (2013). Conceptual Framework for Assessing the Impact of Green Practices on Collaborative Work in China's Construction Industry. Journal of Professional Issues in Engineering Education and Practice, 139(3), 248-255. Conceptual Framework for Assessing the Impact of Green Practices on Collaborative Work in China's Construction Industry

Li, Y., Wang, H., Wang, W., Xing, L., Liu, S., & Wei, X. (2017). Evaluation of the impacts of cooperative adaptive cruise control on reducing rear-end collision risks on freeways. *Accident Analysis and Prevention*, *98*, 87–95.

Li, Z., Ahn, S., Chung, K., Ragland, D. R., Wang, W., & Whon Yu, J. (2014). Surrogate safety measure for evaluating rear-end collision risk related to kinematic waves near freeway recurrent bottlenecks. *Accident Analysis and Prevention*, *64*, 52–61.

Liaw,A., &Wiener,M. (2002). Classification and Regression by RandomForest. *R News*, *2*, 18–22.

Liu, S., McGree, J., Ge, Z., & Xie, Y. (2016). Finding groups in data. *Computational and Statistical Methods for Analysing Big Data with Applications*, 29–55.

Lu, G., Cheng, B., Kuzumaki, S., & Mei, B. (2011). Relationship Between Road Traffic Accidents and Conflicts Recorded by Drive Recorders. *Traffic Injury Prevention*, *12*(4), 320–326.

Lu, X. Y., Varaiya, P., & Horowitz, R. (2009). Fundamental Diagram modelling and analysis based NGSIM data. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *42*(15), 367–374.

Maloney, W. F. (1990). Framework for analysis of performance. Journal of Construction Engineering and Management,, 116, 399-415.

Marques Almeida, A. M., & Santos, L. G. d. P. (2015). Methodological Framework for Truck-Factor Estimation Considering Vehicle–Pavement Interaction. Journal of Transportation Engineering, 141(2). Methodological Framework for Truck-Factor Estimation Considering Vehicle–Pavement Interaction

M'baya, A., Laval, J., & Moalla, N. (2017). An assessment conceptual Framework for the modernization of Legacy Systems. European Union.

McShane. (2011). Traffic engineering. *Pearson*.

Morandi, A., Verga, M., Oleari, E., Gasperotti, L., & Fiorini, P. (2013). A methodological framework for the definition of patient safety measures in robotic surgery: the experience of SAFROS Project. In Frontiers of Intelligent Autonomous Systems (pp. 381-390): Springer.

Mousa, S. R., & Ishak, S. (2017). An Extreme Gradient Boosting Algorithm for Freeway Short-Term Travel Time Prediction Using Basic Safety Messages of Connected Vehicles. TRB 96th Annual Meeting Compendium of Papers, Paper No. 17-04695.

Nam, D., Lavanya, R., Yang, I., Jeon, W. H., & Jayakrishnan, R. (2017). Traffic Density Estimation Using Radar Sensors Data from Probe Vehicles. ITS World Congress 2017 Montreal.

Nasution, F. B. B., Bazin, N. E. N., & Hasanuddin. (2017). Conceptual Framework for Public Policymaking based on System Dynamics and Big Data. IEEE, 19-21.

NationalResearchCouncil. (2016). Highway Capacity Manual: A Guide for Multimodal Mobility Analysis: Transportation research board.

NCHRP. (2003). Performance Measures of Operational Effectiveness for Highway Segments and Systems.

NCHRP, report 551. (2006). *Performance Measures and Targets for Transportation Asset Management*.

NCHRP. (2008). Cost-Effective Performance Measures for Travel Time Delay, Variation, and Reliability (978-0-309-27946-8).

Neal, R. M. (2007). Pattern Recognition and Machine Learning. *Technometrics*, *49*(3), 366–366.

NREL. (2019). Transportation Secure Data Center, National Renewable Energy Laboratory. Accessed Jan. 15, 2019: www.nrel.gov/tsdc.

Ock, J., Issa, R. R. A., & Flood, I. (2016). Smart Building Energy Management Systems (BEMS) Simulation Conceptual Framework. IEEE.

Oh, C., Park, S., & Ritchie, S. G. (2006). A method for identifying rear-end collision risks using inductive loop detectors. *Accident Analysis and Prevention*, *38*(2), 295–301.

Osman, O. A., Bakhit, P. R., & Ishak, S. (2016). Queue Estimation at Signalized Intersections Using Basic Safety Messages in Connected Vehicle Environments. 95th Transportation Research Board Annual Meeting.

Ozbay, K., Associate Professor, P. D., Yang, H., Research Assistant, G., Bartin, B., Research Associate, P. D., & Mudigonda, S. (2007). *Derivation and Validation of a New Simulation-based Surrogate Safety Measure*.

Pacheco, F., Rangel, C., Aguilar, J., Cerrada, M., & Altamiranda, J. (2014). Methodological framework for data processing based on the Data Science paradigm, 1-12.

Papadimitriou, E., Yannis, G., & Golias, J. (2010). Theoretical Framework for Modeling Pedestrians' Crossing Behavior along a Trip. Journal of Transportation Engineering ©ASCE, 136, 914-924. doi:10.1061/共ASCE壳TE.1943-5436.0000163

PTV VISSIM 10. (2018).

Peng, Y., Abdel-Aty, M., Shi, Q., & Yu, R. (2017). Assessing the impact of reduced visibility on traffic crash risk using microscopic data and surrogate safety measures. *Transportation Research Part C: Emerging Technologies*, *74*, 295–305.

Pueboobpaphan, R., & Arem, B. (2011). Driver and Vehicle Characteristics and Platoon and Traffic Flow Stability. *Transportation Research Record: Journal of the Transportation Research Board*, *2189*(1), 89–97.

Qiu, T., Lu, X.-Y., Chow, A., & Shladover, S. (2010). Estimation of Freeway Traffic Density with Loop Detector and Probe Vehicle Data. Transportation Research Record: Journal of the Transportation Research Board, 2178, 21-29. Estimation of Freeway Traffic Density with Loop Detector and Probe Vehicle Data

Rahman, A. and Lowenns,N. (2012). Analysis of Rainfall Impacts on Platooned Vehicle Spacing and Speed,. *Transportation Research Part F*, *15*, 395–403.

Rahman, M. S., & Abdel-Aty, M. (2018). Longitudinal safety evaluation of connected vehicles' platooning on expressways. *Accident Analysis and Prevention*, *117*, 381–391.

Rakha, H., Ahn, K., & Trani, A. (2004). Development of VT-Micro model for estimating hot stabilized light duty vehicle and truck emissions. *Transportation Research Part D: Transport and Environment*, *9*(1), 49-74.

Rakha, H., & Ding, Y. (2003). Impact of stops on vehicle fuel consumption and emissions. Journal of Transportation Engineering, 129(1), 23-32.

Rakha,H., and Gao,Y. (2010). Calibration of Steady-State Car-Following Models Using Macroscopic Loop Detector Data. *VT-2008-01 DTRS 99-G-003,Sponsored by the Virginia Department of Transportation and the U . S . Department of Transportation*.

Rakha, H., & Kamalanathsharma, R. K. (2011). Eco-driving at signalized intersections using V2I communication. 2011 14th International IEEE Conference on Intelligent Transportation Systems.

Rouphail, N. M. (2013). Traffic Congestion Management (Vol. 2): John Wiley & Sons, Inc.

SAE International. (2009). *J2735: Dedicated Short Range Communications (DSRC) Message Set Dictionary*.

SAE. (2016). Dedicated Short Range Communications (DSRC) Message Set Dictionary.

Schrank, D., Eisele, B., Lomax, T., & Bak, J. (2015). 2015 urban mobility scorecard.

Scora, G., & Barth, M. (*2006*). Comprehensive modal emissions model (cmem), version 3.01. User guide. Centre for Environmental Research and Technology. University of California, Riverside, 1070.

Shiomi, Y, Toshio,Y. & Ryuichi, K. (2011). Platoon-Based Traffic Flow Model for Estimating Breakdown Probability at Single-Lane Expressway Bottlenecks. *Transportation Research Part B*, *45*, 1314–1330.

Son, B, Taewan, K., Kim, H. J.& Lee,S. (2004). Probabilistic model of traffic breakdown with random propagation of disturbance for ITS application. *Lecture Notes in Artificial Intelligence*, 45–51.

Song, G., Yu, L., & Wang, Z. (2009). Aggregate fuel consumption model of light-duty vehicles for evaluating effectiveness of traffic management strategies on fuels. Journal of Transportation Engineering, 135(9), 611-618.

Sopian, K., Zaharim, A., Mat, S., Fazlizan, A., Maseran, N., Safari, M. A. M., . . . Rahim, A. A. (2017). The Theoretical Framework of Smart Energy Management System for Rural Area in Mersing Malaysia. IEEE.

Sullivan, A. J., Ramadan, O. E., & Islam, M. A. (2017). Regional Data Clearinghouse Activities - Transportation Data Center, Phase-II. Birmingham, AL

Sun. D & R. F. Benekohal. (2005). Analysis of Work Zone Gaps and Rear-End Collision Probability. *Journal of Transportation and Statistics*.

Swaroop, D., & Rajagopal, K. R. (1999). Intelligent cruise control systems and traffic flow behavior. *American Society of Mechanical Engineers, Dynamic Systems and Control Division (Publication) DSC*, *67*, 373–380.

Tajalli, M., & Hajbabaie, A. (2018). Collision mitigation at signalized intersection using connected vehicles data and technologies. TRB 2018 Annual Meeting.

Tamene, E. H. (2016). Theorizing Conceptual Framework. Asian Journal of Educational Research Vol, 4(2).

Tian, J., Jiang, R., Jia, B., Gao, Z., & Ma, S. (2015). Empirical analysis and simulation of the evolution concavity of traffic oscillations.

Tiaprasert, K., Zhang, Y., Wang, X. B., & Zeng, X. (2015). Queue length estimation using connected vehicle technology for adaptive signal control. *IEEE Transactions on Intelligent Transportation Systems*, *16*(4), 2129-2140.

Tiaprasert, K., Zhang, Y., & Ye, X. (2019). Platoon recognition using connected vehicle technology. *Journal of Intelligent Transportation Systems*, *23*(1), 12-27.

Treiber, M., & Kesting, A. (2013). Traffic flow dynamics: Data, models and simulation. In *Traffic Flow Dynamics: Data, Models and Simulation*.

USDOT (Federal Highway Administration). (2015, 20 November 2015). https://www.itsforge.net/78-news/140-tca-2-3-release

USDOT (U.S. Department of Transportation). (2018). Department of Transportation ITS JPO Data. https://www.its.dot.gov/data/

Van Lint, &Van,C. (2012). Short-Term Traffic and Travel Time Prediction Models. *Artificial Intelligence Applications to Critical Transportation Issues*, *22*, 22–41.

Vasudevan M., A. Jacobi, G. McHale, D. Thompson (RITA ITS JPO), K. Sakai, R. Watanabe, and Y. Tanaka, US-Japan Collaborative Research on Probe Data: Assessment Report. Final Report, FHWA-JPO-13-091, Washington D.C., November 2013

Vogel, K. (2002). What Characterizes A '"Free Vehicle"' In an Urban Area? *Transportation Research Part F: Traffic Psychology and Behavior*, *5*, 15–29.

Vos, T. E. J., Marin, B., Escalona, M. J., & Marchetto, A. (2012). A Methodological Framework for Evaluating Software Testing Techniques and Tools. Paper presented at the 2012 12th International Conference on Quality Software.

Wang, H., Ren, J., Wang, J., & Yang, J. (2014). Developing a Conceptual Framework to Evaluate the Effectiveness of Emergency Response System for Oil Spill. Safe, Smart, and Sustainable Multimodal Transportation Systems © ASCE.

Wang, H., Rudy, K. , Jia, L. & Daiheng, N. (2010). *Calculation of Traffic Breakdown Probability to Optimize Link Throughput. Applied Mathematical Modelling*. 3376–3389.

Wang, W., Chen, S., & Qu, G. (2008). Incident detection algorithm based on partial least squares regression. *Transportation Research Part C: Emerging Technologies*, *16*(1), 54–70.

Wang, Z., Wang, D., Yang, G., & Ding, J. (2013). Selection of Construction Project Delivery Method Based on Value-Added Analysis: A Theoretical Framework. ASCE.

Wiedemann, R. (1974). Simulation Des straßenverkehrsflusses. Institut für Verkehrswesen der Universität Karlsruhe.

Wiedemann, R. (1991). Modelling of RTI-Elements on multi-lane roads, 2.

Wolfgram, J., Huang, P. X., Zhao, Y., Christofa, E., & Xiao, L. (2018). A quick and reliable traffic incident detection methodology using connected vehicle data. TRB 2018 Annual Meeting.

Wright, J., Kyle Garrett,J., Hill,J., Krueger,G., Evans,J., Andrews,S., Wilson,C.K., & Rajbhandari,R. (2014). National connected vehicle field infrastructure footprint analysis. *No. FHWA-JPO-14-125. United States. Department of Transportation. Intelligent Transportation Systems Joint Program Office*.

Xia, Y., Shan, Z., Kuang, L., & Shi, X. (2011). A theoretical approach for its data analyses using cyber infrastructure. ICTIS © ASCE.

Yang, X., J., Zheng, j.,& Sun,J. (2015). Empirical Analysis of Platoon Characteristics: Evidence from Trajectory Data at Two Freeway On-Ramp Bottlenecks. *CICTP, ASCE*.

Yeon, J., Hernandez, S., & Elefteriadou, L. (2009). Differences in Freeway Capacity by Day of the Week, Time of Day, and Segment Type. *Journal of Transportation Engineering*, *135*(7), 416–426.

Yi, P., & Paparaju, S. (2013). Field Investigation of Traffic Noise by Pickup Trucks and Sports Utility Vehicles. Procedia-Social and Behavioral Sciences, 96, 2939-2944.

Zhang, Z., Ding, F., Zhou, Y., Ahn, S., & Ran, B. (2019). Deep Long Short-Term Memory Network Based Long-Term Vehicle Trajectory Prediction. *98st Annual Meeting of Transportation Research Board, Washington DC*, *1*(4800), 1–18.

Zheng, J., & Liu, H. X. (2017). Estimating traffic volumes for signalized intersections using connected vehicle data. Transportation Research Part C: Emerging Technologies, 79, 347-362. Estimating traffic volumes for signalized intersections using connected vehicle data

Zheng, Z. (2012). Empirical Analysis on Relationship between Traffic Conditions and Crash Occurrences. *Procedia - Social and Behavioral Sciences*, *43*, 302–312.

Zheng, Z., Ahn, S., & Monsere, C. M. (2010). Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis and Prevention*, *42*(2), 626–636.

Zou, Z., LI, M., & Bu, F. (2010). Link Travel Time Estimation Based on Vehicle Infrastructure Integration Probe Data. ICCTP 2010: Integrated Transportation Systems—Green•Intelligent•Reliable. ASCE, 2266-2276.

6. APPENDICES

6.1 APPENDIX A – BSM DATA ELEMENTS (PART-I)

- Msgcount
- Vehicle temporary ID
- Dsecond
- Latitude
- Longitude
- Elevation
- Positional accuracy
    1. Semi-major axis accuracy
    2. Semi-minor axis accuracy
    3. Semi-major axis orientation
- Transmission state
- Speed
- Heading
- Steering wheel angle
- Acceleration set 4way
    1. Longitudinal acceleration
    2. Lateral acceleration
    3. Vertical acceleration
    4. Yaw rate
- Brake system status
    1. Brake applied status
    2. Traction control status
    3. Antilock brake status
    4. Stability control status
    5. Brake boost applied
    6. Auxiliary brake status
- Vehicle Size
    1. Vehicle width
    2. Vehicle length

6.2 APPENDIX B – BSM DATA ELEMENTS (PART-II, OPTIONAL)

- Vehicle safety extensions
    1. Vehicle events flags optional
    2. Path history optional
        i. Full position vector
            A. Ddatetime
            B. Longitude
            C. Latitude
            D. Elevation
            E. Heading
            F. Transmission and speed
            G. Positional accuracy
            H. Time confidence
            I. Position confidence set
            J. Speed and heading and throttle confidence
        ii. GNSS status
        iii. Path history point list
            A. Lat offset
            B. Lon offset
            C. Elevation offset
            D. Time offset
            E. Speed
            F. Positional accuracy
            G. Coarse heading
    3. Path direction
        i. Radius of curvature
        ii. Confidence
    4. Exterior lights
- Special vehicle extensions
    1. Emergency details (optional)
        i. SSP index
        ii. Siren in use
        iii. Light bar in use
        iv. Multi-vehicle response
        v. Priviliged events optional
        vi. Response type optional
    2. Event description (optional)
        i. ITIS codes
        ii. Priority optional
        iii. Heading slice optional
        iv. Extent optional

3. Trailer data optional
    i. SSP index
    ii. Pivot point description
        A. Pivot Offset
        B. Pivot angle
4. Trailer unit description list
    i. Vehicle width
    ii. Vehicle length
    iii. Vehicle height optional
    iv. Trailer mass optional
    v. Bumper heights optional
    vi. Center of gravity
    vii. Front pivot description optional
    viii. Rear pivot description optional
    ix. Rear wheel offset optional
    x. Position offset
    xi. Elevation offset

- Supplemental vehicle extensions
    1. Vehicle type classification data
    2. Various V2V probe data
    3. Detected obstacle data
    4. Disabled vehicle report
    5. Oncoming lane speed reporting
    6. Raw GNSS measurements

6.3 APPENDIX C – DATA ATTRIBUTES OF SPAT MESSAGE

- Time stamp such as minute of the year
- Intersection state list
    1. Descriptive name optional
    2. Intersection reference ID
    3. Msgcount
    4. Intersection status object
    5. MOY such as minute of the year
    6. Time stamp such as dsecond optional
    7. Enabled lane list such as lane ID optional
    8. Movement list
        i. Descriptive name optional
        ii. Signal group
        iii. Movement event list
            A. Movement phase state
            B. Timing change details optional
                a. Start time
                b. Min end time
                c. Max end time
                d. Likely time
                e. Confidence
                f. Next time
            C. Advisory speed list optional
                a. Advisory speed type
                b. Speed advice optional
                c. Speed confidence optional
                d. Zone length optional
                e. Restriction class ID optional
    9. Maneuver assist list optional
        i. Lane connection ID
        ii. Queue length optional
        iii. Available storage length
        iv. Wait on stop optional
        v. Pedestrian bicycle detect optiona

6.4 APPENDIX D – DATA ATTRIBUTES OF MAP MESSAGE

- Time stamp such as minute of the year optional
- Msgcount
- Layer type optional
- Layer ID optional
- Intersection geometry list optional
    1. Descriptive name optional
    2. Intersection reference ID
    3. MsgCount
    4. Position 3D
    5. Lane width optional
    6. Speed limit list optional
        i. Speed limit type
        ii. Velocity
    7. Lane list
        i. Lane ID
        ii. Descriptive name optional
        iii. Ingress approach ID optional
        iv. Egress approach ID optional
        v. Lane attributes
            A. Lane direction
            B. Width of lane sharing
            C. Lane type attributes
                a. Lane attributes-vehicle
                b. Lane attributes-cross walk
                c. Lane attributes-bike
                d. Lane attributes-sidewalk
                e. Lane attributes-barrier
                f. Lane attributes-striping
                g. Lane attributes-tracked vehicles such as trains and trolleys
                h. Lane attributes-parking
        vi. Allowed maneuvers optional
        vii. Node list XY
            A. Node set xy
            B. Computed lane
                a. Lane ID
                b. Lane offset in x and y direction
                c. Lane rotation
                d. Lane path scale
        viii. Connects to list optional
        ix. Overlay lane list optional

8. Preempt priority list optional

- Road Segment List optional
    1. Descriptive name optional
    2. Road segment reference ID
    3. Position 3D
    4. Lane width optional
    5. Speed limit list optional
    6. Road lane set list
- Data parameters optional
    1. Process method - IA5 string optional
    2. Process agency - IA5 string optional
    3. Last checked date - IA5 string optional
    4. GeoId used - IA5 string optional
- Restriction class list optional
    1. Restriction class ID
    2. Restriction user type list

6.5 APPENDIX E – GLOBAL INTERVAL

TABLE 6-1: 15-MINUTES INTERVAL ID

15-minutes interval ID (i)	Start time	End time
0	00:00:00	00:14:59
1	00:15:00	00:29:59
2	00:30:00	00:44:59
3	00:45:00	00:59:59
4	01:00:00	01:14:59
5	01:15:00	01:29:59
6	01:30:00	01:44:59
7	01:45:00	01:59:59
8	02:00:00	02:14:59
9	02:15:00	02:29:59
10	02:30:00	02:44:59
11	02:45:00	02:59:59
12	03:00:00	03:14:59
13	03:15:00	03:29:59
14	03:30:00	03:44:59
15	03:45:00	03:59:59
16	04:00:00	04:14:59
17	04:15:00	04:29:59
18	04:30:00	04:44:59
19	04:45:00	04:59:59
20	05:00:00	05:14:59
21	05:15:00	05:29:59
22	05:30:00	05:44:59
23	05:45:00	05:59:59
24	06:00:00	06:14:59
25	06:15:00	06:29:59
26	06:30:00	06:44:59
27	06:45:00	06:59:59
28	07:00:00	07:14:59
29	07:15:00	07:29:59
30	07:30:00	07:44:59
31	07:45:00	07:59:59
32	08:00:00	08:14:59
33	08:15:00	08:29:59
34	08:30:00	08:44:59

15-minutes interval ID (i)	Start time	End time
35	08:45:00	08:59:59
36	09:00:00	09:14:59
37	09:15:00	09:29:59
38	09:30:00	09:44:59
39	09:45:00	09:59:59
40	10:00:00	10:14:59
41	10:15:00	10:29:59
42	10:30:00	10:44:59
43	10:45:00	10:59:59
44	11:00:00	11:14:59
45	11:15:00	11:29:59
46	11:30:00	11:44:59
47	11:45:00	11:59:59
48	12:00:00	12:14:59
49	12:15:00	12:29:59
50	12:30:00	12:44:59
51	12:45:00	12:59:59
52	13:00:00	13:14:59
53	13:15:00	13:29:59
54	13:30:00	13:44:59
55	13:45:00	13:59:59
56	14:00:00	14:14:59
57	14:15:00	14:29:59
58	14:30:00	14:44:59
59	14:45:00	14:59:59
60	15:00:00	15:14:59
61	15:15:00	15:29:59
62	15:30:00	15:44:59
63	15:45:00	15:59:59
64	16:00:00	16:14:59
65	16:15:00	16:29:59
66	16:30:00	16:44:59
67	16:45:00	16:59:59
68	17:00:00	17:14:59
69	17:15:00	17:29:59
70	17:30:00	17:44:59
71	17:45:00	17:59:59
72	18:00:00	18:14:59
73	18:15:00	18:29:59

15-minutes interval ID (i)	Start time	End time
74	18:30:00	18:44:59
75	18:45:00	18:59:59
76	19:00:00	19:14:59
77	19:15:00	19:29:59
78	19:30:00	19:44:59
79	19:45:00	19:59:59
80	20:00:00	20:14:59
81	20:15:00	20:29:59
82	20:30:00	20:44:59
83	20:45:00	20:59:59
84	21:00:00	21:14:59
85	21:15:00	21:29:59
86	21:30:00	21:44:59
87	21:45:00	21:59:59
88	22:00:00	22:14:59
89	22:15:00	22:29:59
90	22:30:00	22:44:59
91	22:45:00	22:59:59
92	23:00:00	23:14:59
93	23:15:00	23:29:59
94	23:30:00	23:44:59
95	23:45:00	23:59:59

TABLE 6-2: 60 MINUTES INTERVAL ID

60 minutes interval ID (j)	Start time	End time
0	00:00:00	00:59:59
1	01:00:00	01:59:59
2	02:00:00	02:59:59
3	03:00:00	03:59:59
4	04:00:00	04:59:59
5	05:00:00	05:59:59
6	06:00:00	06:59:59
7	07:00:00	07:59:59
8	08:00:00	08:59:59
9	09:00:00	09:59:59
10	10:00:00	10:59:59
11	11:00:00	11:59:59
12	12:00:00	12:59:59

60 minutes interval ID (j)	Start time	End time
13	13:00:00	13:59:59
14	14:00:00	14:59:59
15	15:00:00	15:59:59
16	16:00:00	16:59:59
17	17:00:00	17:59:59
18	18:00:00	18:59:59
19	19:00:00	19:59:59
20	20:00:00	20:59:59
21	21:00:00	21:59:59
22	22:00:00	22:59:59
23	23:00:00	23:59:59

6.6 APPENDIX F – ATTRIBUTES OF VISSIM SIMULATION MODEL

TABLE 6-3: ATTRIBUTES OF EACH LINK

LINK NO	NAME	LINKBEHAVTYPE	DISPLAYTYPE
1	valleydale_to_247_NB	1	1
2	RTlane_from_247	1	1
3	247_NB_OffToOnRamp	1	1
4	247_NB_OnRamp	1	1
5	247_to_250_NB	1	1
6	250_off_ramp	1	1
7	from_250_towards_I65	1	1
8	250_onramp	1	1
9	250_to_252	1	1
10	offramp_from_252	1	1
11	on_ramp_252_to_i65	1	1
12	252_to_254	1	1
13	on_Ramp_254	1	1
14	off_Ramp_254	1	1
16	255_offR	1	1
17	254_to_255	1	1
18	255_onR	1	1
19	255_256A	1	1
20	256A_offR	1	1
21	256A_onR	1	1
22	256A_258	1	1
23	258_offR	1	1
24	258_onR	1	1
25	258_259	1	1
26	259_offR	1	1
27	259_onR	1	1
28	259_260B	1	1
29	259B_onR	1	1
30	260B_offR	1	1
31	260_261	1	1
32	260B_onR	1	1
33	OffR_261A	1	1
34	OffR_261B	1	1
35	onR_261A	1	1
36	onR_261B	1	1
37	SB_260_261	1	1
38	SB_261B_offR	1	1

39	SB_261A_offR	1	1
40	SB_261A_onR	1	1
41	SB_261A_offR	1	1
42	259_260B	1	1
43	260B_offR	1	1
44	259B_onR	1	1
45	259_onR	1	1
46	258_259	1	1
47	259_offR	1	1
48	256A_258	1	1
49	258_onR	1	1
50	SB_OnR	1	1
51	255_256A	1	1
52	256A_onR	1	1
53	256A_offR	1	1
54	SB_OffR_255	1	1
55	254_to_255	1	1
56	255_onR	1	1
57	on_Ramp_254	1	1
58	252_to_254	1	1
59	off_Ramp_254	1	1
60	SB_offR	1	1
61	on_ramp_252_to_i65	1	1
62	250_to_252	1	1
63	250_onramp	1	1
64	from_250_towards_I65	1	1
65	250_off_ramp	1	1
66	247_to_250_NB	1	1
67	247_NB_OnRamp	1	1
68	247_NB_OffToOnRamp	1	1
69	valleydale_to_247_NB	1	1
10000	I65_to_247_connector	1	1
10001	C_onRamp_I65_247	1	1
10002	exit_250_to_off_ramp	1	1
10003	connect_250_onramp_to_I65	1	1
10004	C_from_offramp_252	1	1
10005	connect_onramp_252	1	1
10006	C_onR_254	1	1
10007	C_offR_254	1	1
10009	C_onR_255	1	1
10010	C_offR_256A	1	1
10011	C_onR_256A	1	1

10012	C_offR_258	1	1
10013	C_onR_258	1	1
10014	C_offR_259	1	1
10015	C_onR_259	1	1
10016	C_onR_259B	1	1
10017	C_offR_260B	1	1
10018	C_onR_260B	1	1
10019	C_OffR_261A	1	1
10020	C_OffR_261B	1	1
10021	C_onR_261A	1	1
10022	C_OnR_261B	1	1
10023	SB_Connector	1	1
10024	SB_C_261A	1	1
10025	SB_Co	1	1
10026	SB_C	1	1
10027	SB_C_260B	1	1
10028	Sb_C_259B	1	1
10029	SB_C	1	1
10030	SB_C	1	1
10031	SB_C	1	1
10032	SB_C	1	1
10033	SB_C_256A	1	1
10034	SB_C_256A	1	1
10035	SB_OffR_255	1	1
10036	SB_C_onR_255	1	1
10037	SB_C_254	1	1
10038	SB_C_254_On	1	1
10039	SB_C_252	1	1
10040	SB_C_252_onR	1	1
10041	SB_C_OffR_250	1	1
10042	Sb_C_250_onR	1	1
10043	SB_C_OffR_247	1	1
10044	C_247	1	1
10045	Sb_c_247	1	1
10046	Sb_C	1	1
10047	C_247	1	1
10048	Sb_C_250	1	1
10049	C_250	1	1
10050	SB_250C	1	1
10051	C_250	1	1
10052	SB_C	1	1
10053	C_nearUS31	1	1

10054	Sb_C	1	1
10056	Sb_C	1	1
10058	C	1	1
10059	Sb_C	1	1
10060	C	1	1
10061	Sb_C	1	1
10062	C	1	1
10063	sb_c	1	1
10064	C_259	1	1
10065	SB_C	1	1
10066	C_260	1	1