2014

Urban Expressway Safety and Efficiency Evaluation and Improvement using Big Data

Qi Shi
University of Central Florida

URBAN EXPRESSWAY SAFETY AND EFFICIENCY EVALUATION AND IMPROVEMENT USING BIG DATA

by

QI SHI

B.Sc Tongji University, China, 2011
M.S. University of Central Florida, 2013

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at University of Central Florida
Orlando, Florida

Fall Term
2014

Major Professor: Mohamed Abdel-Aty

ABSTRACT

In an age of data explosion, almost every aspect of social activities is impacted by the abundance of information. The information, characterized by alarming volume, velocity and variety, is often referred to as "Big Data". As one fundamental elements of human life, transportation also confronts the promises and challenges brought about by the Big Data era. Big Data in the transportation arena, enabled by the rapid popularization of Intelligent Transportation Systems (ITS) in the past few decades, are often collected continuously from different sources over vast geographical scale. Huge in size and rich in information, the seemingly disorganized data could considerably enhance experts' understanding of their system. In addition, proactive traffic management for better system performance is made possible due to the real-time nature of the Big Data in transportation.

Operation efficiency and traffic safety have long been deemed as priorities among highway system performance measurement. While efficiency could be evaluated in terms of traffic congestion, safety is studied through crash analysis. Extensive works have been conducted to identify the contributing factors and remedies of traffic congestion and crashes. These studies lead to gathering consensus that operation and safety have played as two sides of a coin, ameliorating either would have a positive effect on the other. With the advancement of Big Data, monitoring and improvement of both operation and safety proactively in real-time have become an urgent call.

In this study, the urban expressway network operated by Central Florida Expressway Authority's (CFX) traffic safety and efficiency was investigated. The expressway system is equipped with multiple Intelligent Transportation Systems (ITS). CFX utilizes Automatic Vehicle Identification (AVI) system for Electronic Toll Collection (ETC) as well as for the provision of real-time information. Recently, the authority introduced Microwave Vehicle Detection System (MVDS) on their expressways for more precise traffic monitoring. These traffic detection systems collect different types of traffic data continuously on the 109-mile expressway network, making them one of the sources of Big Data. In addition, multiple Dynamic Message Signs are currently in use to communicate between CFX and motorists. Due to their dynamic nature, they serve as an ideal tool for efficiency and safety improvement.

Careful examination of the Big Data from the ITS traffic detection systems was carried out. Based on the characteristics of the data, three types of congestion measures based on the AVI and MVDS system were proposed for efficiency evaluation. MVDS-based congestion measures were found to be better at capturing the subtle changes in congestion in real-time compared with the AVI-based congestion measure. Moreover, considering the high deployment density of the MVDS system, the whole expressway network is well covered. Thus congestion could be evaluated at the microscopic level in both spatial and temporal dimensions. According to the proposed congestion measurement, both mainline congested segments and ramps experiencing congestion were identified. For congestion alleviation, the existing DMS that could be utilized for queue warning were located. In case of no existing DMS available upstream to the congestion area, the potential area where future DMS could be considered was suggested.

Substantial efforts have also been dedicated to Big Data applications in safety evaluation and improvement. Both aggregate crash frequency modeling and disaggregate real-time crash prediction were constructed to explore the use of ITS detection data for urban expressway safety analyses. The safety analyses placed an emphasis on the congestion's effects on the Expressway traffic safety. In the aggregate analysis the three congestion measures developed in this research were tested in the context of safety modeling and their performances compared. Multi-level Bayesian ridge regression was utilized to deal with the multicollinearity issue in the modeling process. While all of the congestion measures indicated congestion was a contributing factor to crash occurrence in the peak hours, they suggested that off-peak hour crashes might be caused by factors other than congestion. Geometric elements such as the horizontal curves and existence of auxiliary lanes were also identified to significantly affect the crash frequencies on the studied expressways.

In the disaggregate analysis, rear-end crashes were specifically studied since their occurrence was believed to be significantly related to the traffic flow conditions. The analysis was conducted in Bayesian logistic regression framework. The framework achieved relatively good classifier performance. Conclusions confirmed the significant effects of peak hour congestion on crash likelihood. Moreover, a further step was taken to incorporate reliability analysis into the safety evaluation. With the developed logistic model as a system function indicating the safety states under specific traffic conditions, this method has the advantage that could quantitatively determine the traffic states appropriate to trigger safety warning to motorists. Results from reliability analysis also demonstrate the peak hours as high risk time for rear-end crashes. Again, DMS would be an essential tool to carry the messages to drivers for potential safety benefits.

In existing safety studies, the ITS traffic data were normally used in aggregated format or only the pre-crash traffic data were used for real-time prediction. However, to fully realize their applications, this research also explored their use from a post-crash perspective. The real-time traffic states immediately before and after crash occurrence were extracted to identify whether the crash caused traffic deterioration. Elements regarding spatial, temporal, weather and crash characteristics from individual crash reports were adopted to analyze under what conditions a crash could significantly worsen traffic conditions on urban expressways. Multinomial logit model and two separate binomial models were adopted to identify each element's effects. Expected contribution of this work is to shorten the reaction and clearance time to those crashes that might cause delay on expressways, thus reducing congestion and probability of secondary crashes simultaneously.

Finally, potential relevant applications beyond the scope of this research but worth investigation in the future were proposed.

ACKNOWLEDGMENT

The author would like to thank his advisor, Dr. Mohamed Abdel-Aty, for his invaluable guidance, advice and support and encouragement toward successful completion of his doctoral course. The author wishes to acknowledge the support of his committee members, Dr. Essam Radwan, Dr. Boo Hyun Nam, Dr. Naveen Eluru and Dr. Chris Lee.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AI – Artificial Intelligent

ATIS -- Advanced Traveler Information System

ATM – Active Traffic Management

AVI -- Automatic Vehicle Identification

BCI – Bayesian Credible Interval

Caltrans -- California Department of Transportation

CFX -- Central Florida Expressway Authority

CI -- Congestion Index

DelDOT -- Delaware Department of Transportation

DIC – Deviance Information Criteria

DMB -- Digital Multimedia Broadcasting

DMS -- Dynamic Message Signs

DUI -- Driving Under the Influence

ETC -- Electronic Toll Collection

FDOT -- Florida Department of Transportation

FHWA -- Federal Highway Administration

FITM -- Freeway Incident Traffic Management

GIS -- Geographic Information

GPS – Global Positioning System

HFST -- High friction surface treatment

HOT -- High Occupancy Lane

ICS -- Incident Command System

IIA -- Independence from Irrelevant Alternatives

ITS -- Intelligent Transportation System

KEC -- Korean Expressway Corporation

LCS -- Lane Control System

LOS – Level of Service

M&E -- Monitoring and Evaluation

MDTA -- Maryland Transportation Authority

MDXWay -- Miami-Dade Expressway Authority

MNL -- Multinomial Logit

MP -- Milepost

MVDS -- Microwave Vehicle Detection System

MnDOT -- Minnesota Department of Transportation

MOT -- Maintenance of Traffic

MUTCD -- Manual on Uniform Traffic Control Devices

MVMT -- Million Vehicle Miles Traveled

NB – Negative Binomial

NH Turnpike Bureau -- New Hampshire Department of Transportation Bureau of Turnpikes

ORT -- Open Road Tolling

PDO --Property Damage Only

PLN – Poisson-lognormal

PRT -- Perception-Response Time

PSV -- Polished Stone Value

RCI -- Road Characteristics Inventory

RISC -- Rapid Incident Scene Clearance

RTMS -- Remote Traffic Management Sensors

MVDS -- Microwave Vehicle Detector Sensor

SOP -- Standard Operating Procedure

SQL -- Structured Query Language

STARR -- Specialty Towing and Roadside Repair

TMC -- Traffic Management Center

TTI -- Travel Time Index

UCF -- University of Central Florida

WV Parkways Authority -- West Virginia Parkways Authority

CHAPTER 1: INTRODUCTION

1.1 Overview

In an age of data explosion, almost every aspect of social activities is impacted by the abundance of information. The information, characterized by alarming volume, velocity and variety, is often referred to as "Big Data" (Beyer and Laney 2012). As one of the fundamental elements of human life, transportation also confronts the promises and challenges brought about by the Big Data era. Big Data in the transportation arena, enabled by the rapid popularization of Intelligent Transportation Systems (ITS) in the past few decades, are often collected continuously from different sources over vast geographical scale. Huge in size and rich in information, the seemingly disorganized data could considerably enhance experts' understanding of their system. In addition, proactive traffic management for better system performance is made possible due to the real-time nature of the Big Data in transportation.

Operation efficiency and traffic safety have long been deemed as priorities among highway system performance measurements. While efficiency could be evaluated in terms of traffic congestion, safety is studied through crash analysis. Extensive works have been conducted to identify contributing factors and remedies of traffic congestion and crashes. These studies lead to gathering consensus that operation and safety are two sides of a coin, ameliorating either would have a positive effect on the other. With the advancement of Big Data, monitoring and improvement of both operation and safety proactively in real-time have become an urgent call.

The Central Florida Expressway Authority (CFX) System utilizes Automatic Vehicle Identification (AVI) system for Electronic Toll Collection (ETC) as well as for the provision of real time information to motorists within the Advanced Traveler Information System (ATIS). Recently CFX also introduced Microwave Vehicle Detection System (MVDS) for more precise traffic detection. Despite that the detection technologies for each traffic detection system can be distinct, they share several common features. These systems monitor the traffic flow continuously and archive the traffic data on short time interval (e.g., 30 seconds, 1 minute). In addition, they are often installed with relatively short spacing on the managed freeways and expressways especially in urban areas. Thus they have the advantage of reflecting the traffic states along the roadways in real-time. The availability of real-time traffic data has transformed the outlook of numerous aspects of traffic operation and safety, both in research and practice. They allow operators to evaluate the traffic conditions at extremely microscopic level (i.e., specific locations at specific time). Researchers use the data to restore traffic conditions prior to individual crashes and summarize the common patterns leading to unsafe traffic conditions. As a result, proactive traffic management strategies can be developed to improve the overall performance of roadway networks.

To realize the envisioned improvement, the authority needs a medium to communicate with motorists on their system. Dynamic Message Signs (DMS) serve as an ideal tool since they can convey the required message to drivers in a timely manner. Nevertheless, only proper emplacement of them ensures their effectiveness.

1.2 Objectives

The main objective of this research is to identify potential applications of Big Data utilizing the ITS infrastructure. The applications are focused on real-time traffic operation efficiency (congestion) and safety evaluation and improvement. The applications depend on efficient use of the AVI, MVDS and DMS systems. Consequently, the viability of using the AVI and MVDS for high quality data and DMS for timely warning are investigated. To achieve the objective, several questions have to be answered as shown below:

1. How is the data quality provided by the ITS detection systems? Are the Big Data eligible for real-time traffic operation and safety management?
2. Based on the ITS infrastructure, how to evaluate the current operation efficiency of expressways using Big Data? How can DMS be used for congestion management?
3. How are the current traffic safety conditions on the expressways? How can the ITS systems help improve safety management in real-time?
4. What potential extensions of the Big Data can be considered to improve the ITS applications in the future?

To answer each question, several tasks were carried out. The first question addresses the data quality and information contained within the Big Data achieved by the following tasks:

a) Evaluation of the current deployment of the ITS systems on the studied urban expressway network.
b) Assessment of the qualities of the Big Data.
c) Identification of useful information for traffic operation and safety evaluation and improvement.

The second question addresses the operation efficiency on urban expressways and Big Data's application in efficiency evaluation and improvement. The question is answered by the following tasks:

d) Development of congestion measures based on the Big Data and real-time efficiency evaluation using the developed congestion measures.

e) Congestion improvement through DMS.

The third question tries to explore the applications of the Big Data in traffic safety evaluation and improvement. The question is answered by the following tasks:

f) Aggregate traffic safety evaluation of the urban expressways.

g) Real-time traffic safety evaluation and improvement.

h) Assessment of relationship between safety and operation efficiency.

The final question is about potential extensions of the Big Data applications in the transportation arena. The future applications will be proposed but not discussed in detail.

1.3 Dissertation Organization

The organization of the dissertation is as follows: following this chapter, existing applications and studies of ITS data on operation efficiency and safety evaluation are reviewed and summarized in Chapter 2. Chapter 3 provides an overview of the expressway system studied in this research. It covers the current deployment of ITS systems from which the Big Data are generated. Data qualities are examined and the useful information for traffic efficiency and safety evaluation would be identified and prepared. Chapters 4 and 5 are about traffic efficiency evaluation and improvement on urban expressways. Chapter 4 focuses on Big Data applications in development of congestion measures and congestion evaluation. Chapter 5 discusses the use of DMS in operation efficiency improvement. Chapters 6 to 9 give a comprehensive analysis about the traffic safety on the expressways. Chapter 6 introduced the uncapped AVI data in aggregate safety evaluation. Chapter 7 offers comprehensive analysis of expressway safety conditions and contributing factors that could lead to crashes on expressways. Chapter 8 proposes real-time traffic safety evaluation and improvement. Chapter 9 explores the relationship between safety and operation efficiency from a post-crash perspective. Finally, Chapter 10 summarizes the dissertation and raises potential improvement for future applications of Big Data in the transportation arena.

CHAPTER 2: LITERATURE REVIEW

2.1 General

This chapter reviews the existing studies and applications of ITS traffic data on traffic operation and safety evaluation and improvement. Moreover, since DMS plays an important role in potential operation and safety improvement. The deployment of DMS system is also investigated. The review consists of four parts. In the first part, the common ITS traffic detection technologies and applications of ITS traffic data in traffic management will be reviewed will be introduced. In the second part, existing crash frequency studies and their conclusions on crash contributing factors will be synthesized. In the third part, current work on real-time traffic safety evaluation will be summarized. And in the last part of the literature review, DMS implementation will be reviewed.

2.2 ITS Traffic Detection Technologies and Applications

In recent decades a wide variety of traffic detection technologies have been tested and applied for improved performance of automatic traffic monitoring. According to Martin *et al.* (2003), the state-of-art detection technologies fall into three categories: in-roadway detectors, over-roadway detectors and off-roadway technologies. The terms of "in-roadway" and "over-roadway" are used in the Traffic Detector Handbook (Klein *et al.,* 2006). The in-roadway detectors are also known as intrusive detectors and over-roadway detectors are commonly known as non-intrusive detectors. In the meantime, probe vehicles and remote sensing are emerging techniques that are becoming more popular in transportation field.

The in-roadway sensor is one that is embedded in the pavement of the roadway, embedded in the subgrade of the roadway, or taped or otherwise attached to the surface of the roadway (Mimbela and Klein, 2000). Most common in-roadway detector technologies include inductive loop detectors, weigh-in-motion sensors, magnetometers, tape switches, microloops, pneumatic road tubes, and piezoelectric cables. The in-roadway detector technologies have been implemented since the early stage of automatic traffic surveillance thus they are applications of relatively mature technologies. However, they have several drawbacks such as disruption of traffic for installation and repair. Also, they have high failure rates in certain conditions (Martin *et al.,* 2003) such as poor road surface conditions and adverse weather.

An over-roadway sensor is one that is mounted above the roadway itself or alongside the roadway, offset from the nearest traffic lane by some distance. Existing over-roadway sensors include video image processors, microwave radar sensors, ultrasonic, passive infrared and laser radar sensors, and passive acoustic sensors (Mimbela and Klein, 2000). Compared with the in-roadway sensors, the over-roadway sensors have the significant advantage that they minimize the disruption of traffic during installation and maintenance.

Besides the intrusive and non-intrusive traffic detection technologies, off-roadway technologies are developing fast. Probe vehicle and remote sensing are currently two new sensing technologies. Probe vehicles require in-vehicle devices. Remote sensing technology applies arterial or satellite images to analyze and extract traffic information (Martin *et al.,* 2003). Probe vehicles have opened new fields for traffic researchers since they track individual vehicle movement. Remote sensing is performed by aircraft or satellites. However, for real-time traffic monitoring its utilization is quite limited.

Probe vehicle technologies include Global Positioning System (GPS), cellular phones, Automatic Vehicle Identification (AVI) and Automatic Vehicle Location (AVL). They collect real-time traffic data for operation monitoring, incident detection and route guidance. Although the probe vehicle systems require high implementation cost and fixed infrastructure, they have the advantages of continuous data collection automated data collection and no disruption of traffic (Martin *et al.,* 2003).

Of all the detector systems introduced above, two systems that served as traffic data sources in this research will be discussed in detail.

The microwave radars are normally mounted adjacent to the roadways. They are typically insensitive to inclement weather. In addition, they offer direction measurement of speed and multiple lane operation. Two types of microwave detectors exist: Doppler Microwave Detectors and Frequency-modulated Continuous Wave (FMCW) Detectors. FMCW detectors are also referred to as true-presence microwave detectors. Both types of the detectors could detect volume, occupancy, classification and speed. However, the Doppler microwave detectors could only recognize vehicles above a minimum speed. The true-presence vehicle detectors, on the other hand, could detect stopped vehicles.

Automatic Vehicle Identification (AVI) refers to various components and processes that allow for the identification of vehicles for the purposes of charging the toll and providing data for various traffic-management strategies. Currently, two major types of AVI technology are under deployment: laser and Radio Frequency (RF). Laser systems utilize a bar-coded sticker attached to the vehicle. However, the detection is sensitive to weather and dirt. RF system uses transponder attached to vehicles and tag readers to identify the unique tag ID for ETC. Compared with other intrusive or non-intrusive detector systems, the AVI system has the ability to provide space mean speed information and travel time information (Riley, 1999). Nevertheless, the data collection capability of AVI system depends on the coverage area of AVI infrastructures (Martin *et al.,* 2003).

With the fast development of traffic detection technologies, their applications in traffic management become a significant topic for traffic operators. Considering the functionality of each type of traffic detection technology, each system can be exploited by certain types of traffic management applications (Antoniou *et al.,* 2011). Based on the potential applications of these detection systems, they can be broken into point, point-to-point and area-wide.

The inductive loop detectors, microwave detectors and video image detectors all belong to the point-based sensors since they reflect the traffic conditions at the installed locations. AVI system is point-to-point based. The same vehicle is identified at various locations based on which the travel time and space mean speed are calculated. GPS and cellular phones could be deemed as area-wide sensors. They collect travel information of individual vehicles continuously over large area as long as the vehicles are equipped with these devices. These data collection technologies create new opportunities in dynamic traffic management as well as other aspects of traffic simulation and prediction (Antoniou *et al.,* 2011).

The benefits of traffic detection technologies include direct and indirect applications. Direct applications could be congestion reduction, automatic incident detection (AID), travel time estimation. Indirect applications are carried out through enhancement of traffic modeling in the model development, calibration and validation processes.

2.3 Crash Frequency Studies

The use of ITS traffic data not only leads to more precise and timely traffic management for operators, it also introduces new perspectives in traffic safety research. The following part first summarizes current usage of the ITS traffic data in safety analysis. Then methodologies used to conduct traffic frequency studies will be explained. In the end, some conclusions of existing studies are provided.

2.3.1 ITS traffic data in traffic safety studies

Traffic safety can be affected by various factors such as driver characteristics, roadway geometry, traffic condition and weather (Huang and Abdel-Aty, 2010; Yu and Abdel-Aty, 2014). Among these factors, traffic flow parameters are most extensively studied. The effects of traffic variables (speed, volume, occupancy) on crash occurrences have been discussed in numerous studies. Before the introduction of ITS traffic data, only vague traffic indicators such as Average Annual Daily Traffic (AADT), Level of Service (LOS), etc. were available. Since the researchers began to use ITS traffic data in their work, traffic data at more microscopic level were applied. Daily traffic volume by weekday and weekend, even hourly volume are available. Moreover, the traffic detection data enable accurate measurement of speed, occupancy which is a surrogate measure of traffic density and vehicle classification. As a result, the quality of traffic data used in safety modeling has been greatly enhanced.

Of all the detection systems, loop detectors are first introduced and the most widely used traffic data sources for traffic safety analysis (Lee *et al.,* 2006; Golob *et al.,* 2008; Lee and Abdel-Aty, 2008; Xu *et al.,* 2013a). Loop detectors could provide information about speed, volume and occupancy based on one-minute or half-minute interval. However, the weakness of the loop detectors is that during inclement weather conditions, such as heavy rain or ice on the pavement, the accuracy of their detected data could be greatly reduced. In addition, the maintenance of the devices in case of malfunction could be very complicated. Recent development of non-intrusive traffic detection systems such as Microwave Vehicle Detection System (MVDS) (Akin *et al.,* 2011; Ahmed and Abdel-Aty, 2013b; Yu *et al.,* 2013b; Yu and Abdel-Aty, 2014), Video Detectors (Versavel, 1999; Zhang *et al.,* 2007; Laureshyn *et al.,* 2010), Automatic Vehicle Identification (AVI) sensors (Abdel-Aty *et al.,* 2012b; Ahmed *et al.,* 2012b) to be used in traffic data analysis. MVDS traffic data provide similar information as the loop detectors. Vehicle types can be also derived from MVDS data. The advantage of MVDS compared with loop detectors lies in that their performance is not greatly affected by weather conditions and they are installed along the roadside which make them easy to maintain. While the loop detectors, MVDS, and Video Detectors measure time mean speed at specific locations, the AVI sensors calculate space mean speed for a segment where AVI detectors are installed at both ends of the segment. However, vehicle count is not available from the AVI data.

2.3.2 Statistical Modeling Techniques

To study the crash mechanisms analytically, extensive research has been conducted at both the macro- and micro-levels. Micro-level traffic safety studies investigate motor-vehicle crashes with relatively small number of occurrences within a specific locations (e.g., roadway segment, intersection), but use information in a great amount of detail. Crash frequency studies at the micro-level are the cornerstone before proceeding to more detailed safety evaluation.

Poisson (Jovanis and Chang, 1986; Joshua and Garber, 1990; Miaou 1994) and Negative Binomial (NB) (Hauer and Hakkert, 1988; Poch and Mannering, 1996; Miaou and Lord, 2003) distributions are commonly used to represent the crash frequency distribution. Thus, in most studies more advanced safety models are eventually extensions from Poisson or NB models. Lord and Mannering (2010) summarized the advantages and disadvantages of the existing frequency models. In this proposal, the main statistical technique adopted is random-parameters (also known as multi-level or hierarchical) approach and Bayesian inference techniques.

Random-parameters approach is valued for its flexibility in accounting for unobserved heterogeneity which the fixed-parameters approach fails to capture. When the data structure exhibits hierarchical characteristics, multi-level models should be the suited modeling approach. In the multi-level model structure, parameters of explanatory variables vary at more than one level (Gelman and Hill, 2006). Several studies have benefited from the strength of random-parameter models (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009; Anastasopoulos and Mannering, 2011; Dinu and Veeraragavan, 2011; El-Basyouny and Sayed, 2011; Venkataraman *et al.,* 2011b; Venkataraman *et al.,* 2013; Xie *et al.,* 2013).

Anastasopoulos and Mannering (2009) used random-parameters NB model to study the crash mechanisms on 322 rural interstate highway segments in Indiana. The effects of pavement characteristics, roadway geometric characteristics, and truck percentages on crash frequencies were examined. The authors concluded that random-parameters negative binomial models better than standard negative binomial model.

El-Basyouny and Sayed (2009) employed the Poisson-lognormal (PLN) model with random parameters for 392 urban road segments belonging to 58 corridors in Vancouver, British Columbia. Corridor variations were modeled by fitting a different regression curve for each corridor. Urban road geometric characteristics contributing to crash occurrence were identified.

Venkataraman *et al.* (2011b, 2013) analyzed 1,153 directional segments of the seven Washington State interstates with nine-year's crash data. In their study, several geometric variables such as curvature effects, roadway lighting, left and right side shoulder width were studied. They also implemented a random parameter approach to estimate crash frequency by severity, number of vehicles involved and location types.

Xie *et al.* (2013) conducted a corridor-level signalized intersection safety analysis for 195 intersections of 22 corridors in Shanghai, China. Corridor-level data and intersection-level data were employed to construct a multi-level NB model.

The Bayesian inference is a statistical method appropriate for the goal to estimating the effects of variables at different levels properly as it treats the group-level as "prior information" in estimating the individual-level coefficients (Gelman and Hill, 2006). Recent trends see more and more researchers use Bayesian inference to predict crash frequencies.

Huang and Abdel-Aty (2010) discussed the application of multi-level Bayesian analysis in traffic safety. They pointed out the commonly observed hierarchical data structure in traffic safety analysis. In response, methods to solve multi-level problems were discussed. Artificial intelligent (AI) models are criticized of being black boxes incapable of generating interpretable results. Generalized estimating equations (GEE) provide estimates with acceptable properties only for the fixed parameters in the model. Bayesian hierarchical approach was recommended for explicitly address multi-level data structure while rendering interpretable results.

Guo *et al.* (2010) proposed to use Bayesian inference in their intersection safety studies. They argued that the signalized intersections located along the same corridor share similar traffic flow, geometric design, land use and signal control timing. And the assumptions for the generalized linear models (GLM) about independent observations are violated. The Bayesian approach could evaluate the effects of variables at their own levels. The author also underlined that the key to a Bayesian model is the selection of appropriate likelihood function and prior distributions.

Ahmed *et al.* (2011a) and Yu and Abdel-Aty (2013b), Yu *et al.* (2013b) used Bayesian hierarchical models to account for the seasonal and spatial correlations for a mountainous segment of Interstate 70 in Colorado. The authors found significant seasonal effect on crash occurrence on the roadway segment of interest. To account for the spatial correlations and evaluate the effects of candidate variables during each season, the authors used random effect model with Bayesian techniques. And they concluded that random effect model outperformed the others.

In most of the studies involving Bayesian inference techniques, the open source software WinBugs are used for model construction. To evaluate the model performance, the deviance information criteria (DIC) introduced by Spiegelhalter *et al.* (2003b) was used as a Bayesian measure of model complexity and fit. DIC can be viewed as the combination of measure of model fitting ($\bar{\mathrm{D}}$) and penalization of number of effective variables (p_D) in the model.

$$\mathrm{DIC} = \bar{\mathrm{D}} + p_D \tag{2-1}$$

When the difference in DIC between two models is more than 10, it is assured that the model with smaller DIC is better. The difference of DIC between 5 and 10 is considered substantial. Otherwise it is hard to make a definitive judgment (Spiegelhalter *et al.,* 2003a). Bayesian Credible Interval (BCI) was used for variable estimation; and the 95% BCI was measured to test the significance of variables. If the BCI does not contain 0, then the effect of variable is significant.

2.3.3 Crash contributing factors

A variety of factors could contribute to crash occurrence considering the highly random nature of crashes. However, to uncover the factors that significantly affect crashes and come up with strategic countermeasures for safety improvement, traffic characteristics and roadway geometric elements have been systematically studied during the past several decades.

2.3.3.1 Traffic characteristics

Regarding the effects of the traffic variables on traffic safety, numerous studies have explored how they would affect the crash occurrence. Traffic characteristics can often be classified as speed, density, flow, and congestion. Relationship between speed, density and flow:

$$q = k\bar{v} \tag{2-2}$$

Where q is flow (vehicles per unit time); k is density (vehicles per length of road); and $\bar{v}$ is mean speed.

For the effects of speed, it is reasonably safe to assume that increased speed would mean that the accidents that have occurred would be more severe (O'Donnell and Connor, 1996; Shankar and Mannering, 1996; Kockelman and Kweon, 2002; Hauer, 2009a, b; Kweon *et al.,* 2009; Xu *et al.,* 2013b). It is less straight forward for the relationship between speed and the possibility of accidents occurring. Several study results confirmed that higher speed will lead to higher number of crashes or higher accident rates. Taylor *et al.* (2002) implemented cross-sectional analysis on rural road segment in Britain, and also confirmed that positive relationship between accident frequency and average speed. Nilsson (2004) in his study found positive relationship between changes in speed and the number of accidents. The magnitude, however, depends on crash types. Elvik *et al.* (2004) concluded the causal relationship between changes in speed and changes in road accidents. They found the number of accidents will go down if speed goes down and vice versa. Aarts and van Schagen (2006) did a review on the relationship between driving speed and the risk of road crashes. They concluded that higher speed would lead to higher crash rate. In contrast, in several other studies, the effect of speed on crash occurrence is the opposite. Baruya (1998a, b), Ahmed *et al.* (2011b), and Yu *et al.* (2013a) found decreased average speed 5 – 10 min prior to crash occurrence is associated with higher likelihood of crashes. Generally speaking, speed has mixed effects on road safety.

Speed variance, on the other hand, is argued to cause safety problems instead of speed itself. In an early study about the effect of speed variance on crash occurrence, Garber and Gadiraju (1989) concluded that accident rated do not necessarily increase with an increase in average speed but do increase with an increase in speed variance. Golob and Recker (2003a) further concluded that left-lane collisions on urban freeways are more likely induced by volume effects, right-lane collisions are more closely tied to speed variances in adjacent lanes. As for crash severity, Lave (1985) found that fatality rate was strongly associated with speed variance rather than average speed.

Relationship between traffic density and accidents has been investigated less in the literature due to the dearth of relevant data. Some research used Volume over Capacity (V/C) to represent density (Ivan *et al.,* 2000; Lord *et al.,* 2005). Recently researchers used occupancy as a measure of traffic density (Lord *et al.,* 2005). In the research by Lord *et al.* (2005), they incorporated traffic density in their safety performance function and concluded that for single-vehicle crashes, density is negatively related with the crash frequencies; however, for multi-vehicle crashes, density is positively related with crash frequencies. This result reminded researchers that when analyzing the effects of traffic density on traffic safety, crash types should be taken into consideration.

The impact of traffic congestion on crash occurrence has also been explored by researchers. Baruya (1998) in their study concluded that the degree of congestion has a negative effect on crash frequency. Noland and Karlaftis (2005) investigated congestion and safety in London using an area-wide spatial analysis approach. Their study showed that there is little effect of traffic congestion on road safety. However, they suggested this might be due to that congestion are highly localized and time-of-day specific, therefore a more precise congestion measurement should be used. A recent study by Kononov *et al.* (2008) found that total as well as fatal and injury crash rats increase with the increase in traffic congestion. Wang *et al.* (2013b) found that increased traffic congestion is associated with more fatal and serious injury accidents and has little impact on slight injury accidents. The authors explained this might be due to the higher speed variance and worse driving behavior in the presence of congestion.

Annual Average Daily traffic has been widely utilized as a traffic flow indicator. Most safety studies have incorporated this information as an exposure factor. Overall the traffic flow is positively related with crash frequencies in the existing literature (Hauer *et al.,* 2002; Chin and Quddus, 2003; Wang *et al.,* 2009; Zhang *et al.,* 2012). Recent works begins to explore the potential that whether a surrogate measure of disaggregate volume is worth investigation and could be used as an alternative to aggregate volume indicators such as AADT. By using volume information aggregated at smaller time interval, researches now have the capability to study the effect of traffic flow during specific condition or time interval. Yu and Abdel-Aty (2013a) used traffic volume on weekdays and weekend to reveal and compare the features for weekday and weekend crashes. Hossain and Muromachi (2013) used high-resolution traffic data and aggregated the traffic volume into 5-minute interval for both total vehicle count and heavy vehicle count.

2.3.3.2 Roadway geometric characteristics

Roadway geometric design has long been claimed to play important role in traffic safety. Good designs should expect where drivers tend to make mistakes and leave some leeway for driving errors. Many studies have taken geometric element into account when conducting traffic safety evaluation. In most cases, the geometric information could be gathered in two ways; 1) fixed-length sections (Shankar *et al.,* 1995), that the roadway section is divided into the fixed length. 2) homogeneous sections (Milton and Mannering, 1998), that a new section would be generated by changes of any geometric variables. Some deficiencies exist for both segmentation methods. Homogeneous sections can result in sections with segment length too short for safety analysis. Crash locations are reported to nearest milepost, resulting in some misplacement if the segment is too short. Segment with fixed length could have the values of geometric characteristics changed within one segment. Recently, researchers have used the segmentation method that keeps the homogenous properties of roadway segment but combines the segment that is too short with adjacent sections as illustrated by Ahmed *et al.* (2011a).

The effects of various roadway geometrics on road accident frequency have been discussed:

Shankar *et al.* (1995) found the number of horizontal curves and maximum grade had a positive relationship with accident frequency in the Seattle area during the period 1988 -1993.

Milton and Mannering (1998) found narrow lanes (less than 3.5m) and sharp horizontal curves tend to decrease accident frequency in Eastern Washington.

Abdel-Aty and Radwan (2000) modeled traffic accident occurrence and involvement on SR 50 in Central Florida. They reached the conclusion that narrow lane width, larger number of lanes, narrow shoulder width and reduced median width increased the likelihood for accident involvement. Horizontal curvature was positively related to crash occurrence in the study.

Noland (2003a, b) used county-level highway accident data during 1987-1990 in the state of Illinois concluded that an increase in the number of lanes and lanes width was associated with increased fatalities; and an increase in the outside shoulder width was found to be associated with reduced accidents.

Haynes *et al.* (2007, 2008) suggested that more curved roads in an area resulted in less road accidents; further, they demonstrated that road curvature has an inverse relationship with fatal accidents in urban settings. Curvature was found to be a protective factor.

In Kononov *et al.*'s (2008) study about urban freeways. They reached the conclusion that an increased number of lanes increased the number of accidents.

Park and Lord (2009) in their studies about the effects of freeway design on safety stated that for urban freeways, number of lanes and curvatures were positively associated with crash frequencies. The median width is negatively related with crash occurrence.

Ahmed *et al.* (2011a) and Yu *et al.* (2013a) studied the hazardous factors on a mountainous freeway segment in Colorado. They found the vertical curves played important part in crash occurrence. They found that wider median width, larger number of lanes and higher degree of curvature all negatively affected the crash likelihood.

For several roadway geometric elements, consistent conclusions about their effects on traffic safety were reached, such as the number of lanes, the median width, lane width and shoulder width. The effect of horizontal curves obviously has a mixture of effects based on existing studies. To interpret the mixed effects, both engineering and driver behavior aspects should be taken into account. Wang *et al.* (2013a) pointed out that curvature might be risky considering its engineering effect. On the other hand, drivers might drive more slowly and cautiously on curved roads.

2.4 Real-time crash prediction

Exactly because of the real-time data collection by traffic detection systems on the roadways, traffic data generated on short time intervals (1-minute, 30-second) enable researchers to look at the crash precursors on individual crash levels. In most of the cases, real-time information about traffic, weather just before the crash occurrence would be incorporated in predicting crashes. The real-time crash predictions could unveil one important component in traffic safety that aggregated analysis sometimes failed to explore: the traffic dynamics – sudden formation of disrupted traffic condition (Hossain and Muromachi, 2012).

Common predictors in the real-time crash predictions include average speed, standard deviation of speed, and coefficient of variation of speed, traffic volume and occupancy aggregated at upstream and downstream detector locations. Recent research has seen real-time information incorporated and their effects found significant, especially in areas where inclement weather conditions are common.

Two general analysis approaches are employed in real-time traffic safety studies, statistical methods and data mining based methods. Statistical methods include logistic regression (Hourdos *et al.,* 2006), matched case-control logistic regression (Abdel-Aty *et al.,* 2004a; Abdel-Aty *et al.,* 2005; Zheng *et al.,* 2010), Bayesian statistics (Abdel-Aty *et al.,* 2012a; Ahmed *et al.,* 2012a; Yu and Abdel-Aty, 2013b); data mining based methods could be neural networks (Pande and Abdel-Aty, 2006), classification trees (Pande and Abdel-Aty, 2006), random forests (Pande *et al.,* 2011; Ahmed and Abdel-Aty, 2012; Hossain and Muromachi, 2012) and support vector machines (Yu and Abdel-Aty, 2013c), etc. Statistical methods presented the effects of variables in an interpretable way; however they assumed a linear relationship between the dependent and independent variables. While data mining methods often make predictions with very high accuracy, they are questioned because the analyzing process is like a black box.

Hourdos *et al.* (2006) employed video data to identify crash-prone conditions at freeway high-crash locations in Minnesota. Logistic regression model was used to find indicators of crash-prone conditions. Average speed, speed variation, wet pavement and reduced visibility were found to be significant factors increasing crash likelihood. The model achieved 58% accurate detection rate. One weakness of logistic regression is it has no control over other potential factors beside the ones being tested.

Abdel-Aty *et al.* (2004a, 2005) utilized matched case-control logistic regression method to predict crashes in real-time. The matched case-control analysis was employed to explore the effects of traffic flow variables while controlling for the effects of other confounding variables through the design of the study. The authors noticed that multi-vehicle crashes on freeways under high- and low-speed traffic conditions differed in severity and mechanisms. Two separate models were evaluated in the matched-case control framework. They concluded that the low-speed crashes occurred mostly in persisting congestions; however, the queues resulted from the crash dissipated quickly. In contrast, high-speed crashes often occurred under smooth traffic conditions; disruptive traffic conditions originating from downstream could cause driving errors.

Bayesian matched-case logistic regressions have been employed in the study about visibility related crashes by Abdel-Aty *et al.* (2012a). The advantages of using the Bayesian approaches, as the authors pointed out, include (1) it provides a natural and principled way of combining prior information with the data to yield a posterior belief, (2) it presents full distributional profile of parameters rather than single coefficient estimates to fully account for the uncertainty associated with single parameter estimates in classical statistics, and (3) small sample inference proceeds in the same manner of a large sample. Both loop detector data and AVI data were used in this study, it was found that the model estimated from loop detector data indicated the average speed observed at the nearest downstream station along with the coefficient of variation in speed observed at the nearest upstream station, all at 5-10 min prior to the crash time, were significant to visibility related crashes. The AVI data suggested only coefficient of variation in speed was significant.

Yu and Abdel-Aty (2013b) employed Bayesian multi-level logistic regression to study the single- and multi-vehicle crash mechanisms on a mountainous freeway segment. The modeling approach enabled the authors to account for the seasonal variations, crash-unit-level diversity and segment-level random effects on crash occurrence. The author stated by Bayesian inference techniques, more unobserved heterogeneity could be captured and the better classification ability the model would have.

Pande and Abdel-Aty (2006) explored the use of data mining techniques in predicting traffic crashes in real time. The authors employed classification trees, multi-layer perceptron neural networks. The authors suggested that data mining typically involves analysis where objectives of the data analysis have no bearing on the data collection strategy. The authors explained the reason why they used these data mining techniques as, the data mining process has two key components, variable selection procedure based on classification tree and neural network based modeling procedure with parameters identified through the preceding classification tree as inputs. Results of the model showed that 71% non-crashes and 57% crashes were correctly identified. Average speed, difference in occupancy on adjacent lanes and variations in volume and speed were found to be significantly associated with the crash occurrence.

Random forests method is an ensemble classifier that consists of many decision trees. Compared with traditional classification trees, random forest could obtain unbiased error estimates with no need for a separate cross validation test data set. Ahmed and Abdel-Aty (2013a) implemented the random forest technique to identify the significant traffic factors affecting the crash occurrence using AVI data on CFX's system. The authors concluded that AVI data were promising in providing a measure of crash risk in real time. However, they suggested it is useful when AVI segments are within 1.5 miles on average.

Yu and Abdel-Aty (2013c) employed the support vector machine in real time crash risk evaluation. Support vector machine was originally designed based on statistical learning theory and the structural risk minimization. The authors listed some previous studies using support vector machine and demonstrated that compared with multi-layer feed forward neural network and probabilistic neural network models, support vector machine models provided lower misclassification rate, higher correct detection rate. The study object was a mountainous freeway segment. The real-time risk prediction result showed that support vector machine will work well with small sample size; variable selection is needed before model estimation using support vector machine. Finally, the results confirmed with logistic regression models.

2.5 Dynamic Message Signs Implementation

2.5.1 Overview of DMS

Dynamic Message Signs (DMS), also known as Variable Message Signs (VMS) or Changeable Message Signs (CMS), are one of the primary Intelligent Transportation Systems (ITS) components. Their deployment is intended to provide motorists en-route with real-time traffic information in order to improve operations, reduce crashes, and help drivers make more educated route choice decisions, especially under adverse traffic conditions.

According to Dudek (2004), the real-time information not only benefits individual drivers and the responsible transportation agency, but also the general public. Motorists are interested in reaching their destinations as safely as possible without undue delays. The transportation agency is interested in utilizing the available highway capacity of the corridor or network and to improve safety and mobility.

The information displayed on the DMS could be classified into three major types (Dudek, 2004): 1) early warning messages; early warning messages give motorists advance notice of slow traffic and queuing ahead and are effective in reducing secondary crashes. 2) advisory messages; advisory messages provide motorists with useful information about a specific problem along their route. This information allows motorists to change their speed or path in advance of the problem area, or may encourage them to voluntarily take an alternative route to their destination. 3) alternative route messages; alternative route messages influence motorists to travel to their chosen destination by using different routes than originally intended. The alternative route is one designated by the transportation agency. In cases when the freeway is physically closed as a result of construction, crash, or natural disaster, the motorists are notified that an alternative route must be used.

Two types of DMS exist, namely permanently mounted DMS and portable DMS. Since portable DMS are used at a specific location temporarily where events or engineering work are occasional, the implementation of DMS and the effects on driver behavior in this study is about permanently mounted DMS.

In Florida, the Guidelines for the Use of Dynamic Message Signs on the Florida State Highway System (2008) state that the application of DMS as important traffic control devices involves several different situations, such as emergency (severe crashes, evacuations), incident scenes (traffic crashes, disabled vehicles, spilled cargo, roadway construction or maintenance), adverse weather conditions (icy roadway, fog, high winds), traffic delay (traffic congestion) and even normal conditions (peak and off-peak periods in the absence of incidents).

Several projects to evaluate the effectiveness of using variable message signs in European cities (Chatterjee and Hadi, 2013) have been conducted. Table 2-1 showed the applications of DMS in these studies. Four different types of traffic information were tested. Drivers' behavior reflected that for incident messages, it is not only the severity of the incident, but also other factors such as the specific location mentioned and the availability of viable alternative routes to avoid the problem location are what they expected. For route guidance information, when the route advice differs from that given normally, substantial diversions occur. For continuous information describing the traffic state on a major route, information increases the use of the major route and reduces use of alternative routes if there are no traffic problems reported on the major route. Travel time information was well regarded by drivers and found to be effective in inducing route changes. The study concluded that the deployments of DMS to inform drivers of traffic conditions have proved successful in terms of improving network travel times and reducing environmental impacts.

In the United States, while DMS systems are widely used, not many studies have yet explored how the DMS could affect drivers' behavior. A few studies involving the effects of DMS were conducted through questionnaires, simulations. Benson (1996) explored the motorist attitudes about DMS content in Washington, D.C. area. The research was done through a survey of more than 500 motorists. Results showed that exact location of accidents, time-tagging traffic information received high levels of support. Failing in these regards were met with less interest, including delay time estimates, safety messages, and posting of alternative routes.

Peeta *et al.* (2000) conducted an on-site stated preference user survey to evaluate the drivers willingness to divert based on DMS content in case of incidents. Results showed besides what has been mentioned in the above study by Benson (1996), other significant factors also included socio-economic characteristics, network spatial knowledge, and confidence in the displayed information.

Table 2-1: Applications of Dynamic Message Signs (DMS)

Project test site	Number of signs and type of road network	Type of information	VMS control strategy	Typcical message content
Bristol, UK	five VMS adjacent to the Park and Ride site on the arterial corridor	Phase 1: air quality information; Phase 2: comparative travel times for bus (Park and Ride) and car	automatic message selection based on air quality monitoring, SCOOT and AVL data	'AIR POLLUTION HIGH–USE P + RIDE' or 'TO CITY CENTRE P + R 15 MIN CAR 24MIN'
London, UK	30 VMS in Greater London, many facing in-bound drivers on arterial routes	incident information (advance warning or immediate warning)	set manually by the police based on information from the traffic congestion log	'MON 9 NOV 10AM–1 PM WHITEHALL AVOID AREA' or 'STAMFORD HILL CLOSED LONG DELAYS'
Lyon, France	five VMS (three at the south-east entrance and two at the northwest entrance to Croix Rousse tunnel)	congestion or incident information	automatic message selection based on travel time predictions and the observation of incidents	'TRAFIC FLUIDE', 'QUAI DE RHONE RG TRAFIC DENSE' or 'RUE GRENETTE FERMEE'
Paris, France	150 VMS on the SIRIUS motorway network to the east of Paris	travel time or congestion information (including cause of the problem, if relevant), depending on the traffic situation and sign location	automatic message selection based on an expert system that processes data from loop detectors and video cameras	'A104 → A4: 21MN A3 →A1: 3MN ' or 'A1: A 2KM BOUCHON = 3KM'

Applications of Dynamic Message Signs (DMS)

Project test site	Number of signs and type of road network	Type of information	VMS control strategy	Typcical message content
Piraeus, Greece	five VMS for in- and out-bound passenger port traffic	route guidance indicating the advised direction to a specific destination	automatic selection of a rerouting strategy based on traffic estimates	'GATE C ' or 'ATHENS →'
Southampton, UK	26 VMS on arterials, 17 VMS on motorway skirting Southampton	incident information (advance or immediate warnings)	set manually or VMS plan selected from an integrated strategy library	'BOAT SHOW PLEASE USE SIGNED CAR PARKS' or 'ROADWORKS BEVOIS VALLEY ROAD CLOSED'
Toulouse, France	10 VMS on approach roads to the urban ring road	route guidance indicating the advised direction to a specific destination	global routing strategy computes the controls for all signs	'VERS JEANNE D'ARC GAUCHE RECOMMANDE'
Turin, Italy	26 VMS on a network controlled by the Integrated Town Control Architecture of the Turin 5T system	route guidance indicating the advised direction to a specific destinations, plus a reason for the diversion when advice differs from normal	VMS control system calculates sign settings to meet target flows and turning percentages	conventional direction board but with indicator lights and rotating prism elements to advise diversions; variable text panel to display reason for the diversion, e.g. 'TRAFFICO INTENSO'
Valencia, Spain	32 traffic information VMS	congestion information	automatic message selection based on link flow/occupancy data	'FLUID', 'DENS' or 'CONGESTIO' description for each route

2.5.2 Queue Warning via DMS

One of the most important functions of DMS is queue warning, such that the traffic information at downstream locations could be sent to upstream drivers. Queue warning is a strategy to warn drivers of upcoming congestion and allow drivers enough time in advance to make decisions whether to detour or stay on the route. From a safety perspective it also allows drivers enough decision and reaction time to slow down.

The potential benefits offered by the queue warning system include prevention of primary and secondary crashes (also the severity is expected to be reduced), delay of the onset of congestion, and travel time improvement. The basic principle behind queue warning is alerting the drivers to congestion conditions, drivers' caution will rise, and more smooth and uniform traffic flow be achieved.

In Germany, the queue warning system involves displaying a congestion pictograph on each side of the speed harmonization gantry or DMS indicating congestion ahead. The gantries are generally spaced 1 km apart, and the system typically begins reducing speeds between three and four gantries before an incident. It was also noted that users are interested in knowing the location of the queue and what route they should take to avoid it (Mirshahi *et al.,* 2007).

The Netherlands alerts travelers to congestion and queues by flashing lights and speed signs on variable speed limit signs. The system, generally located every 500 m, provides queue tail warning and protection in known bottleneck locations. The Dutch have seen definite benefits from their congestion warning system. The throughput increased and incidents decreased (Mirshahi *et al.,* 2007).

2.5.3 Placement of Dynamic Message Signs

Since DMS are commonly used in queue warning, the placement of DMS therefore is discussed together with the guidelines for locating queue warning signs. Exact guidelines about the placement of DMS in the United States were not found. General principles have been discussed, which is presented in this section.

Currently the Manual on Uniform Traffic Control Devices (MUTCD) (2009) provides general guidance on the locations of DMSs as follows:

- DMS should be located sufficiently upstream of known bottlenecks and high crash locations to enable road users to select an alternative route or take other appropriate action in response to a recurring condition.
- DMS should be located sufficiently upstream of major diversion decision points, such as interchanges, to provide adequate distance over which road users can change lanes to reach one destination or the other.
- DMS should not be located within an interchange except for toll plazas or managed lanes.
- DMS should not be positioned at locations where the information load on drivers is already high because of guide signs and other types of information.

- DMS should not be located in areas where drivers frequently perform lane-changing maneuvers in response to static guide sign information, or because of merging or weaving conditions.

Although DMS system acts as multi-function media to motorists and could not simply be defined as warning signs, displaying warning messages is still crucial to the system. The standard for locating warning signs could be referred to for DMS locations. The MUTCD (Traffic) (2009) addresses the placement of warning signs as in Table 2-2. The warning signs should leave the drivers with adequate Perception-Response Time (PRT) to make corresponding adjustment. However, it is stated that the distance between the sign and the warned location should not be too far, such that drivers might tend to forget the warnings, especially in urban areas.

In California Department of Transportation's Traffic Manual Chapter 4 about signs, the placement of warning signs is also specified. They state that in rural areas, the warning signs should normally be placed about 150m (0.09mi) in advance of the conditions. On high-speed roads, particularly on freeways, the advance warning distance may have to be as great as 450m (0.28mi) or more (2006). These standards are similar to those provided by MUTCD.

Besides these general principles, researchers also conduct a variety of studies on the deployment of DMS. Among the current researches, some of them made an endeavor to find the optimal locations for these DMS.

Table 2-2: Guidelines for Placement of Warning Signs (adapted from MUTCD 2009)

Posted or 85th-Percentile Speed	Advance Placement Distance[1]								
	Condition A: Speed reduction and lane changing in heavy traffic[2]	Condition B: Deceleration to the listed advisory speed (mph) for the condition							
		0[3]	10[4]	20[4]	30[4]	40[4]	50[4]	60[4]	70[4]
20 mph	225 ft	100 ft[6]	N/A[5]	--	--	--	--	--	--
25 mph	325 ft	100 ft[6]	N/A[5]	N/A[5]	--	--	--	--	--
30 mph	460 ft	100 ft[6]	N/A[5]	N/A[5]	--	--	--	--	--
35 mph	565 ft	100 ft[6]	N/A[5]	N/A[5]	N/A[5]	--	--	--	--
40 mph	670 ft	125 ft	100 ft[6]	100 ft[6]	N/A[5]	--	--	--	--
45 mph	775 ft	175 ft	125 ft	100 ft[6]	100 ft[6]	N/A[5]	--	--	--
50 mph	885 ft	250 ft	200 ft	175 ft	125 ft	100 ft[6]	--	--	--
55 mph	990 ft	325 ft	275 ft	225 ft	200 ft	125 ft	N/A[5]	--	--
60 mph	1,100 ft	400 ft	350 ft	325 ft	275 ft	200 ft	100 ft[6]	--	--
65 mph	1,200 ft	475 ft	450 ft	400 ft	350 ft	275 ft	200 ft	100 ft[6]	--
70 mph	1,250 ft	550 ft	525 ft	500 ft	450 ft	375 ft	275 ft	150 ft	--
75 mph	1,350 ft	650 ft	625 ft	600 ft	550 ft	475 ft	375 ft	250 ft	100 ft[6]

[1] The distances are adjusted for a sign legibility distance of 180 feet for Condition A. The distances for Condition B have been adjusted for a sign legibility distance of 250 feet, which is appropriate for an alignment warning symbol sign. For Conditions A and B, warning signs with less than 6-inch legend or more than four words, a minimum of 100 feet should be added to the advance placement distance to provide adequate legibility of the warning sign.
[2] Typical conditions are locations where the road user must use extra time to adjust speed and change lanes in heavy traffic because of a complex driving situation. Typical signs are Merge and Right Lane Ends. The distances are determined by providing the driver a PRT of 14.0 to 14.5 seconds for vehicle maneuvers (2005 AASHTO Policy, Exhibit 3-3, Decision Sight Distance, Avoidance Maneuver E) minus the legibility distance of 180 feet for the appropriate sign.
[3] Typical condition is the warning of a potential stop situation. Typical signs are Stop Ahead, Yield Ahead, Signal Ahead, and Intersection Warning signs. The distances are based on the 2005 AASHTO Policy, Exhibit 3-1, Stopping Sight Distance, providing a PRT of 2.5 seconds, a deceleration rate of 11.2 feet/second^2, minus the sign legibility distance of 180 feet.
[4] Typical conditions are locations where the road user must decrease speed to maneuver through the warned condition. Typical signs are Turn, Curve, Reverse Turn, or Reverse Curve. The distance is determined by providing a 2.5 second PRT, a vehicle deceleration rate of 10 feet/second^2, minus the sign legibility distance of 250 feet.
[5] No suggested distances are provided for these speeds, as the placement location is dependent on site conditions and other signing. An alignment warning sign may be placed anywhere from the point of curvature up to 100 feet in advance of the curve. However, the alignment warning sign should be installed in advance of the curve and at least 100 feet from any other signs.
[6] The minimum advance placement distance is listed as 100 feet to provide adequate spacing between signs.

Abbas and McCoy (1999) suggested a genetic algorithm to identify the best DMS locations where they would maximize the potential reduction in vehicle delay as a result of traffic diversion when incident occurs. Fu *et al.* (2007) extended the work of Abbas and McCoy. They introduced the time-of-day variation in travel demand and inherent variations in incident characteristics across links and over time. Chiu and Huynh (2007) proposed a two-stage stochastic program to determine the optimal DMS locations with the objective of reducing the life cycle cost of the system. Shang and Huang (2007) introduced cell transmission model to optimize the location of a DMS board, aiming to minimize the system's total travel time. Jianjun *et al.* (2010) discussed the best location for DMS based on Grey situation decision.

To improve the traffic operation is the major concern of these studies. As for road safety, which although has been stated as one of the benefits that DMS will bring, is rarely considered in the DMS allocation procedure. The effects of DMS on road safety are mostly evaluated after they have been installed. Rämä and Kulmala (2000) and Luoma *et al.* (2000) studied the effects of variable message signs for slippery road conditions on reported driver behavior and found VMS increased traffic safety by decreasing mean speeds and increasing the following distances. Erke *et al.* (2007) concluded in their study that large speed reductions were observed when drivers approached the DMS. And safety problems may result from the attention overload or distraction due to the information on the DMS. Tay and De Barros (2010) conducted a study in which both questionnaire and on-road test were implemented. They concluded that the road safety messages on DMS had only relatively small albeit beneficial effect on driver's attitudes and on-road traffic speed.

2.6 Conclusions

Considerable progress has been made in traffic detection technologies in the past few decades. While the introduction of new systems is intended to offer more precise readings and higher system reliability, applications of these systems in traffic efficiency and safety are always worth exploration.

A large body of literature has investigated the traffic safety issues and crash contributing factors. Many of these studies at both aggregate and disaggregate levels have explored the effects of traffic flow parameters on crash occurrence. However, based on current findings, no consensus has yet been reached on how traffic congestion affects traffic safety. In addition, congestion as a typical traffic phenomenon on urban expressways not only exerts its impact on safety, but also serves as a direct measure of operation efficiency. As a result, traffic safety and efficiency are closely interrelated. There is a need to examine both and to explore the potential for improvement of both simultaneously, especially with the available traffic detection data. More specifically, given the ITS traffic detection data, how researchers could quantify the urban expressway efficiency? Are there definitive effects of congestion on safety and in turn how does traffic safety reacts to operation efficiency? If the relationship between efficiency and safety acts like two sides of the same coin, will there be methods to improve both of them?

All of these questions are explored in this dissertation for a more comprehensive understanding of the urban expressway efficiency, safety, their interrelationship and potential improvement using the ITS systems on the expressways. Considering the deployment of ITS detection systems on the expressways, Bayesian multi-level framework is extensively used in safety analysis, combined with data mining technologies to facilitate the modeling construction process (random forest, K-means clustering). Also, since there is an explicit interest in the relationship between efficiency and safety, correlation between variables need to be accounted using several statistical methods (correlation test, ridge regression). For the improvement of safety and efficiency simultaneously, feasibility of incorporating methodology in other principles (reliability analysis) is explored.

CHAPTER 3: EXPRESSWAY SYSTEM OVERVIEW

3.1 Introduction

To explore the potential of Big Data generated by the ITS systems in efficiency and safety evaluation and improvement for urban expressways, the expressway network in Central Florida area was selected as the data source and study location. The expressway system is located in the densely populated area of Central Florida, serving residents and tourists. Multiple ITS infrastructures are currently deployed on the network. Other pertinent information needed for analyses is also available. This chapter will take a glimpse of the individual expressways constituting the network. Then it will briefly describe the traffic detection systems which serve as the source of Big Data. The DMS will be a crucial tool for both efficiency and safety improvement, thus an introduction about the DMS system on the expressways will be given. Moreover, the roadway geometric information and crash data used in this research will also be covered in this chapter.

3.2 Expressway System

The Central Florida Expressway Authority (CFX) operates and maintains the region's 109-miles of expressway networks of interest as shown in Figure 3-1 (2014a). Currently, there are five toll roads under or partly under the management of CFX. Although future extensions have been planned, this research only focuses on existing CFX system. The five expressways connect Orlando and neighboring areas, serving both residents and visitors.

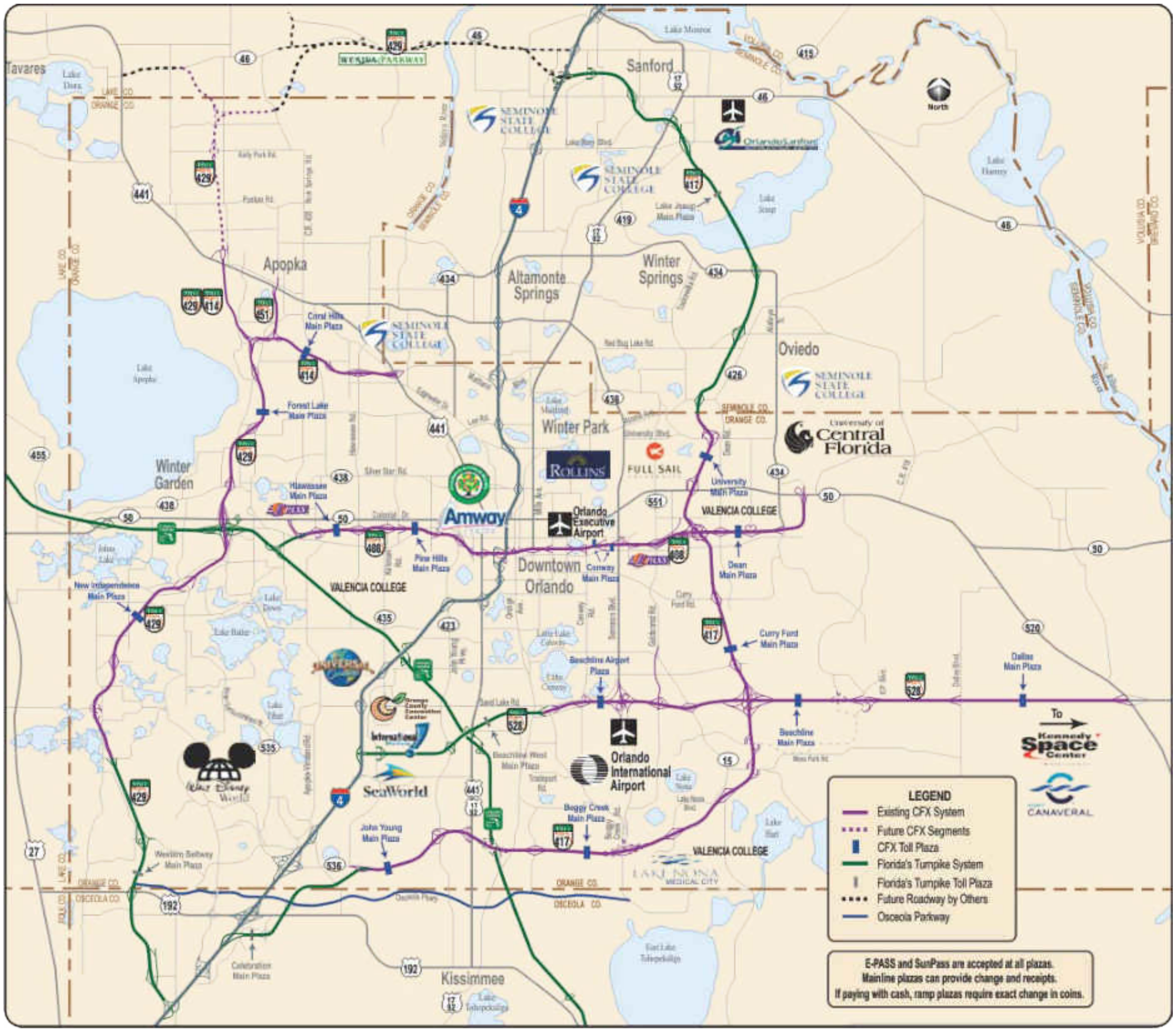


Figure 3-1: Expressways under CFX Management (2014d)

According to CFX, State Road 408 (Spessard L. Holland East-West Expressway) is the backbone of the Expressway Authority's 109-mile network. Except for the westernmost mile, the expressway is owned and operated by CFX. An estimated 125,000 - 135,000 vehicles a day travel the 21-mile expressway through downtown Orlando. Land uses along the expressway include residential, commercial and services, transportation (airport), academic institutions (2014b).

State Road 414 (John Land Apopka Expressway) is a 9-mile east-west corridor and is relatively new in the system. It improves access to SR 429 which is another expressway in CFX system, Interstate 4 and many local roads in the greater Apopka area (2014c).

A segment of 33 miles of the State Road 417 located in Orange County is under CFX management. This segment is also known as the GreeneWay. It provides the suburban areas near Orlando with convenient access if motorists need to travel between Sanford, Oviedo, the University of Central Florida, East Orlando or Kissimmee. SR 417 was the first in the system to have all mainline toll plazas converted to Express Lanes, which keep traffic moving by allowing customers to pay tolls at the posted highway speed (2014a).

State Road 429 (Daniel Webster Western Beltway) was developed in partnership between the Expressway Authority and the Florida's Turnpike Enterprise. CFX operates 23 miles of the expressway. The function of the SR 429 is to provide West Orange and Osceola counties with an alternate north-south route to Interstate 4 (2014b).

State Road 528 (Martin B. Andersen Beachline Expressway) provides a crucial connection for residents and tourists between the attractions area, the Orlando International Airport and the East

Coast beaches and Cape Canaveral. The Expressway Authority operates the 23 miles of the expressway (2014d).

For a thorough evaluation of current operation and safety performance of the network, comprehensive efforts have been made to collect data from different sources. Traffic data from two different detection systems, DMS message information, roadway geometric characteristics data, and crash data have been collected.

3.3 Traffic Detection Systems on the Expressways

When CFX converted mainline toll plazas to open tolling express lanes, they adopted the Automatic Vehicle Identification (AVI) system for Electronic Toll Collection (ETC). If vehicles traveling on CFX's expressways are equipped with E-PASS or SunPass, they don't have to stop to pay the tolls. The AVI detectors will keep records of the vehicle information and calculate the tolls according to the distance that the vehicles traveled. Although AVI detectors can archive traffic information, they are not designed for this objective. Since 2012, CFX has introduced Microwave Vehicle Detection System (MVDS) to their expressway network. These detectors are specifically installed for traffic monitoring. The two systems exhibit substantial difference between them, however, both of them could be leveraged to provide traffic professionals with valuable traffic information. In this study, both data were collected based on their availability.

3.3.1 AVI Traffic Data

AVI detectors are installed at toll plazas for Electronic Toll Collection (ETC) and at other locations for travel time estimation. The deployment of AVI system since 2005 and system updates afterwards were provided by CFX. The AVI traffic data were collected from September 2012 to July 2014. Based on the length between adjacent AVI detectors and the timestamps that the vehicles are detected at these locations, the traveling speed for a segment can be calculated. The segmentation of expressways based on the AVI sensors is shown in Appendix A. Table 3-1 summarizes the number of AVI segments per direction and basic statistics on each of the five expressways.

Table 3-1: AVI Segments on CFX Expressway System

Route ID	Direction	No. of Segments	Segment Length			
			Mean	Std Dev	Min	Max
SR 408	EB	23	0.926	0.479	0.290	1.853
	WB	23	0.977	0.524	0.332	2.287
SR 414	EB	3	1.529	0.555	0.928	2.022
	WB	4	2.445	2.948	0.350	6.811
SR 417	NB	16	1.959	0.829	0.751	3.848
	SB	20	1.567	0.770	0.378	3.098
SR 429	NB	10	1.895	1.220	0.704	4.271
	SB	9	2.136	1.534	0.614	4.536
SR 528	EB	8	2.740	2.25149	0.329	7.058
	WB	9	2.578	2.129	0.861	7.597

Figure 3-2 illustrates the deployment of AVI sensors on the expressway network. SR 408 has the smallest AVI segment length of the five expressways. SR 528 has relatively short AVI segments near the international airport and west to SR 417. However, on the suburban segments leading to the coast area, the distance of adjacent AVI sensors could be above 7 miles. The distance of adjacent AVI tag readers are determined on two basic criteria: 1) the need for toll collection; 2) the need for travel time estimation. In urban areas, the accessibility of the expressway has to accommodate the travelers' demand of entering and exiting the expressways. This makes the toll collection for a relatively short spacing necessary on SR 408.

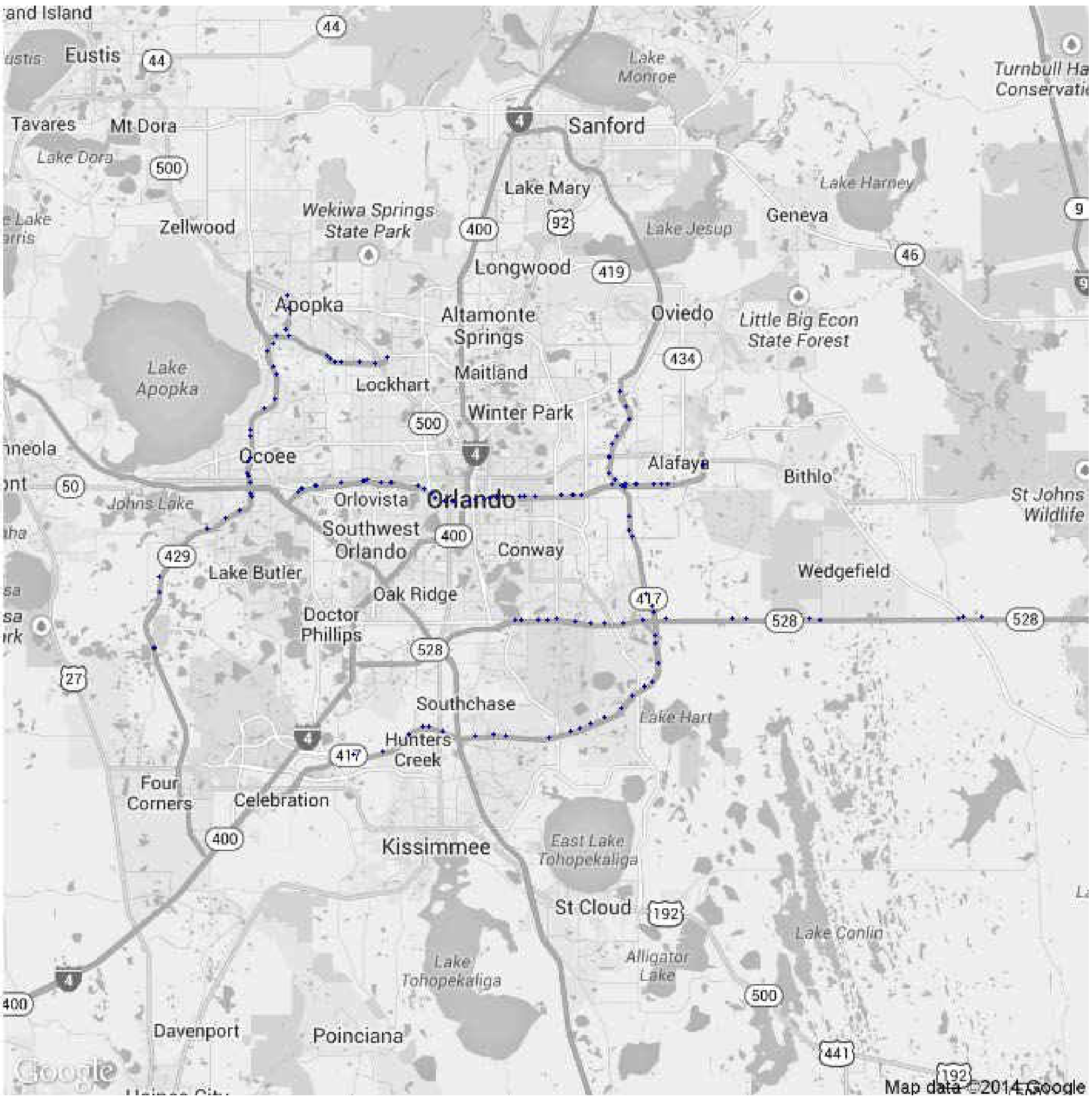


Figure 3-2: Deployment of AVI Sensors on Expressway Network

Traffic data generated from the AVI system can be categorized in two types, one is the more traditional capped AVI data and the other is the uncapped AVI data collected since the beginning of the project. The capped AVI data contain space mean speed information for each detection segment on one minute interval basis. The speed is capped at speed limit. Therefore, the AVI data during this time period are referred to as capped AVI data. While providing traffic information in real-time, the capped AVI data is not able to reflect the real situation as shown in Figure 3-3.

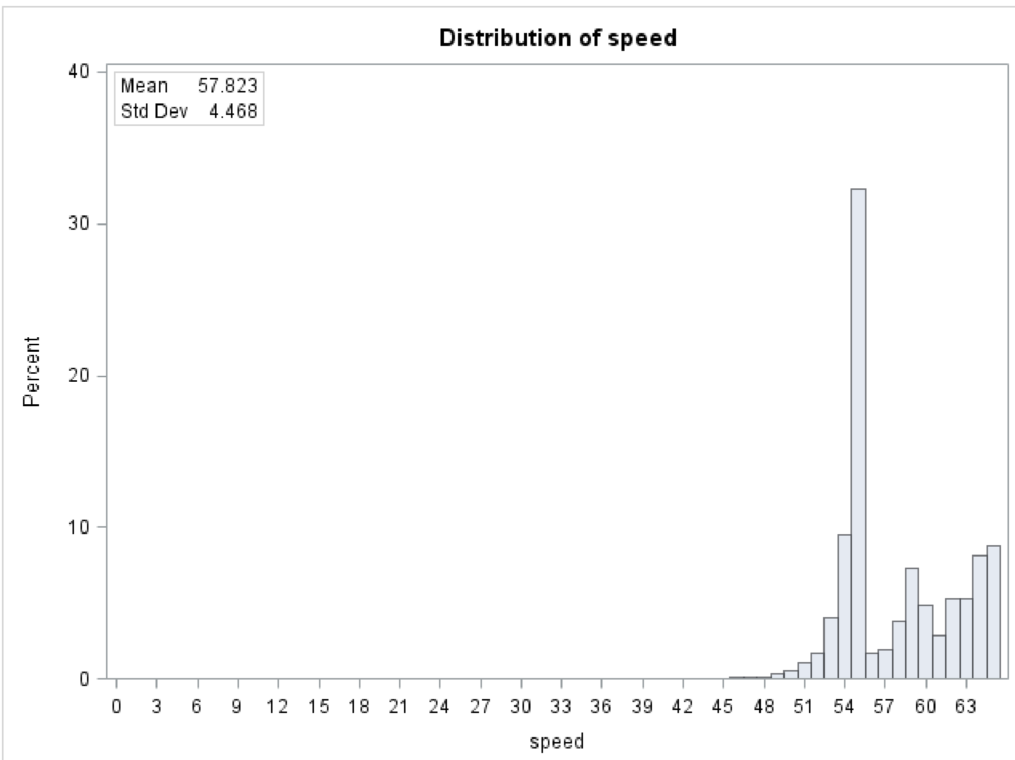


Figure 3-3: SR 408 Eastbound Capped AVI Data (Aug, 2013)

Since September 2012, CFX and Atkins have archived the raw readings by the AVI detectors. The raw readings contain the encrypted vehicle ID information and the timestamps of detection at each AVI detector location. Based on the information, the speed of individual vehicles can be derived using equation (3-1).

$$\text{speed} = \frac{|\text{milepost}_{\text{upstream}} - \text{milepost}_{\text{downstream}}|}{\text{timestamp}_{\text{downstream}} - \text{timestamp}_{\text{upstream}}} \quad (3\text{-}1)$$

Consequently, traffic information from the raw readings is referred to as the uncapped AVI data. Compared with capped AVI data, the uncapped data is not trimmed at speed limit, thus reflecting the speed of the segment at a specific time closer to the real traffic condition. In addition, the uncapped AVI data is based on individual vehicles, therefore providing some insights about the traffic volume on the expressways. Although the traffic count by AVI sensors is not complete traffic volume, they still offered precious information about real-time traffic volume before the introduction of MVDS sensors. Figure 3-4 is the distribution of speed for the same expressway during the same time period. The figure shows that the uncapped AVI data has its speed more normally distributed. Based on the comparison results, it is suggested to use the uncapped AVI data for more precise operation evaluation.

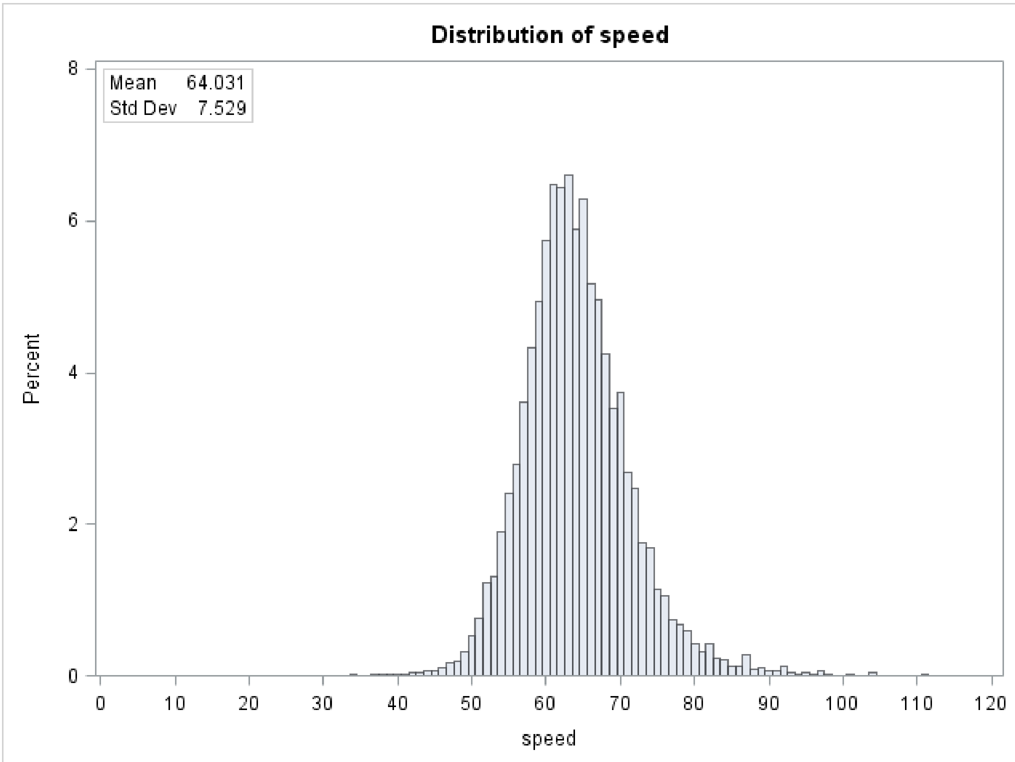


Figure 3-4: SR 408 Eastbound Uncapped AVI Data (Aug, 2013)

3.3.2 MVDS Traffic Data

MVDS was initially introduced to CFX's expressways since 2012. In 2013, the whole network operated by CFX was covered by MVDS as displayed in Figure 3-5. The system is specifically designed for traffic monitoring.

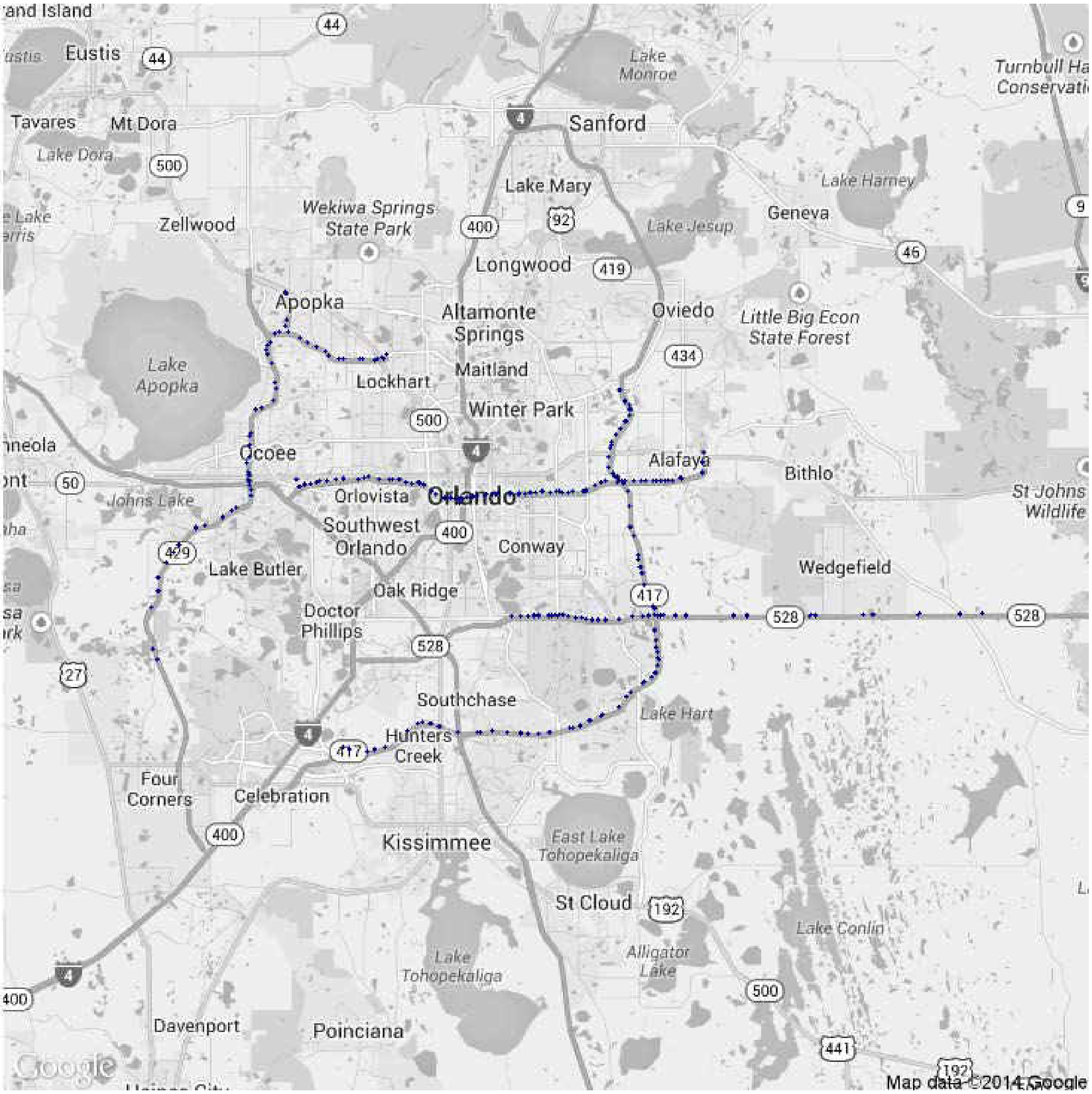


Figure 3-5: Deployment of MVDS Sensors on the Expressway Network

For the purpose of this project, MVDS data have been collected since July, 2013. MVDS does not identify individual vehicles. They return aggregated traffic flow parameters for each lane of the cross-section where the MVDS detector is installed at one minute interval basis. The traffic parameters include traffic volume, time mean speed, lane occupancy and traffic volume by vehicle length. Four types of vehicles were defined by their lengths:

- Type 1: vehicles 0 to 10 feet in length
- Type 2: vehicles 10 to 24 feet in length
- Type 3: vehicles 24 to 54 feet in length
- Type 4: vehicles over 54 feet in length

Additional information from MVDS traffic data includes the timestamp when the sensor is polled. It has been mentioned above that the sensors are polled every one minute. Also, unique sensor identifier and lane identifier are contained within the data. The sensor identifier consists of the roadway (i.e., SR 408, SR414, SR 417, SR 429 and SR 528), milepost and direction. The lanes are counted from the roadway medium to the outside lane. The lanes fall into four categories, which are Mainline, Ramp, Mainline TP Express and Mainline TP Cash. Mainline TP Expressway indicates express lanes at mainline toll plazas; vehicles equipped with tags do not need to slow down on these lanes when they pass the toll plazas. Mainline TP Cash means toll booth at mainline toll plazas; vehicles need to stop and pay tolls. On the expressways, these two types of lanes are physically separated. The types of lanes and number of lanes at each MVDS detection location can be seen in Appendix B.

Compared with the AVI traffic data, MVDS data reflect the traffic states at their installed locations instead of a segment. They also have several advantages over AVI data. The first is the scale of MVDS system. As shown in Table 3-2, the MVDS sensors significantly outnumber the AVI sensors and average distance between adjacent detectors of MVDS is much smaller than that of the AVI system. The higher deployment density means traffic information from more locations is gathered and more detailed knowledge about the expressway system is available. The second advantage is traffic data for different types of lanes from MVDS system. Given that MVDS sensors monitor traffic conditions on each traveling lane, traffic data at toll plazas and on ramps can be collected. AVI data only provide traffic information of a cross-section on the mainline. However, to have a general understanding about the expressway performance, analysis of toll plazas and ramps are necessary as well. The third advantage is the richness of traffic information from MVDS data. Capped AVI data only has capped speed information which does not reflect real-world traffic conditions while uncapped AVI data has more realistic speed data and part of the traffic volume information. The traffic count from uncapped AVI data is not the complete traffic volume. MVDS data include speed, complete traffic volume, and lane occupancy as a surrogate measure of traffic density and the volume by vehicle lengths.

Table 3-2: MVDS on the CFX Expressway System

Route	Length (mi)	Direction	MVDS Detectors							
			Total	Mainline (including TP Express)	TP Cash	Ramp	Distance between adjacent detectors			
							Mean	Std Dev	Min	Max
SR 408	21.4	EB	57	55	8	39	0.38	0.18	0.1	1
		WB	56	55	8	39	0.39	0.18	0.1	1
SR 414	9.5	EB	14	14	2	8	0.44	0.17	0.2	0.7
		WB	13	12	2	7	0.46	0.23	0.1	0.9
SR 417	31.5	NB	56	55	7	31	0.58	0.28	0.2	1.3
		SB	56	55	7	32	0.58	0.28	0.2	1.2
SR 429	22	NB	29	28	4	17	0.68	0.54	0.2	2.8
		SB	29	27	4	16	0.68	0.59	0.1	3.1
SR 528	22.4	EB	29	26	4	19	0.84	0.79	0.1	3
		WB	29	29	4	18	0.84	0.82	0.1	3.1

In conclusion, MVDS on CFX expressway network is more suitable for traffic monitoring. Nevertheless, AVI data will also be used in the report. It is expected that by using the two types of traffic data, better understanding about the expressways will be reached.

3.4 DMS Systems on Expressways

CFX installed numerous Dynamic Message Signs (DMS) on their expressways. The DMS are electronic signs on roadways to give motorists real-time information. As shown in Figure 3-6 and Table C-1 in Appendix C, in total 37 DMS are currently in use on CFX expressways and 35 of them are located on the five expressways involved in this study. The other two DMS are located on SR 451 and SR 520, respectively. The DMS data from September 2012 to September 2013 were collected. The DMS data record the DMS identifier information, the messages displayed on the boards, and the displayed time and duration of each message. SR 408 and SR 417 have the most DMS on the mainline according to Table 3-3. And SR 414, SR 429 and SR 528 have relatively fewer DMS installed. The differences are mainly affected by the traffic demand and the length of the segment. SR 408 has the highest traffic load among the five expressways. SR 417 on the other hand is the longest expressway in the system. However, the average distance between DMS on SR 408 is much smaller than the other roadways.

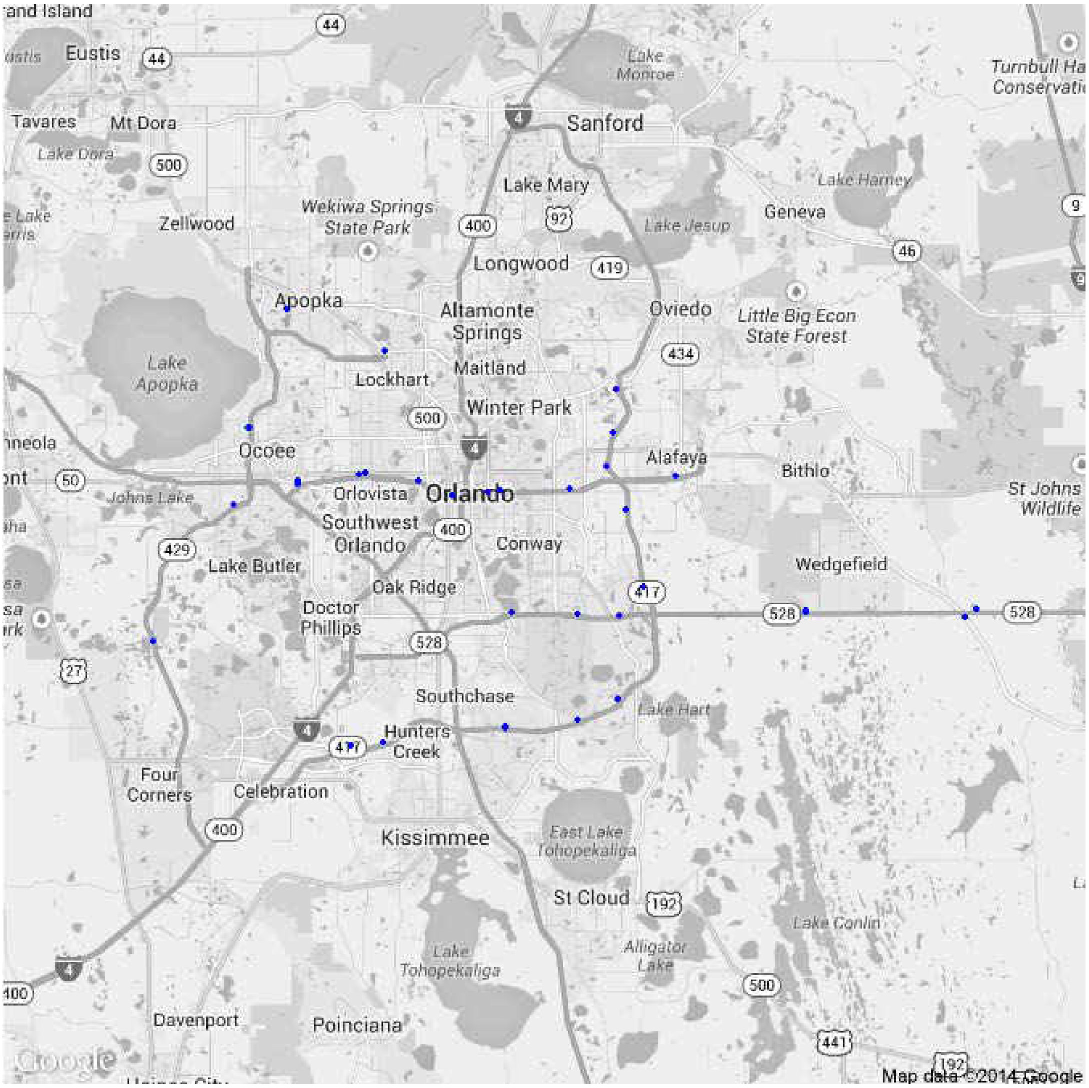


Figure 3-6: Deployment of DMS on the Expressway Network

Table 3-3: DMS on CFX Expressway System

Route	Direction	Total Number	On Mainline	Average Distance between DMS (mi)	On Ramp
SR 408	EB	7	6	3.84	1
	WB	5	5	3.925	0
SR 414	EB	0	0	--	0
	WB	1	1	--	0
SR 417	NB	5	5	6.725	0
	SB	6	6	5.94	0
SR 429	NB	3	3	6.6	0
	SB	2	2	4.3	0
SR 528	EB	3	3	7.3	0
	WB	3	3	8.85	0

DMS is used to convey the messages from traffic operators to motorists on road. According to the messages displayed on the DMS board in the one year period, several types of messages could be summarized. Typical messages include:

- travel time estimation message (the most common type)
- alert message: silver alert, child abduction, LEO alert, traffic alert
- law enforcement message: safety belt use, moving over for emergency vehicle
- adverse weather warning message: fog, smoke, brush fire, low visibility
- congestion warning message: congestion, heavy congestion, expected congestion
- lane close/open message: lane block, all lanes block, ramp close, ramp open, planned close/open of lanes/ramps
- unexpected events warning message: incident, crash, debris, disabled vehicle
- planned events message
- road work warning message: construction, road work, planned road work, work zone, rolling road block

Among the listed types of messages above, several types of messages are not expected to alter motorists' behaviors, such as messages related to travel time estimation, alert, law enforcement and planned events. Other messages are displayed to heightened travelers' awareness of the traffic conditions, weather, incidents, etc. These types of messages are expected to reduce risks on the expressways and improve the traffic flow. Besides these commonly displayed messages, blank signs are also observed during the study period. In other cases signs under test will show "TEST" message. The DMS data will be used to evaluate current adequacy of DMS.

3.5 Road Geometric Characteristics Data for Expressways

Roadway geometry has been verified in previous research (Shankar *et al.,* 1995; Milton and Mannering, 1998; Park and Lord, 2009; Ahmed *et al.,* 2011a; Christoforou *et al.,* 2011; Venkataraman *et al.,* 2011a; Hossain and Muromachi, 2012; Le and Porter, 2012; Yu and Abdel-Aty, 2013b) to have significant impact on traffic operation and safety. In this project, we first collected the geometric characteristics data in 2012 and updated the data in 2013. Florida Department of Transportation (FDOT) maintained the Road Characteristics Inventory (RCI) database that has the complete roadway geometry and other relevant information. The RCI database has hundreds of variables. Only the most relevant variables were chosen for the data preparation. In sum 14 variables have been selected, including pavement condition, number of lanes, auxiliary lane type, shoulder type and width, median type and width, inside shoulder type and width, horizontal degree of curvature, speed limit, section AADT, D factor, K factor and truck percentage. The expressways are divided into homogeneous segments. If one of the geometric characteristics' variables changes, a new segment will be generated. For the convenience of study, the smallest segment length is specified to be 0.1 mile. Segments smaller than 0.1 mile will be combined with adjacent segment which shares higher geometric similarity.

3.6 Crash Data on Expressways

In Florida, crashes are recorded in two formats of crash reports, namely the short form and long form. Long form crash reports are designed to keep records of more severe crashes, especially those involving injuries or fatalities. Short form crashes are mostly used for property damage only crashes. Two databases served as the crash data source. One is FDOT Crash Analysis Reporting (CAR) system and the other is Signal 4 Analytics (S4A) system. CAR database has longer history. However, they only archive the long form crashes. In contrast, S4A is newly developed and has both short and long form crashes. The issue with S4A database is that for the crashes occurred in early years (e.g., early 2000s), the short form crashes were not complete. After June 2012, S4A has the complete crash data from both types of reports for whole Florida. This current research covers two-year period since September 2012, thus having no problem with the crash data. For insurance, the research team still made a comparison between the CAR and S4A system using crash data from July 2012 to December 2013. All crashes recorded to occur on the expressways from CAR system were extracted as shown in Table 3-4. It should be noted that the crashes happening on the expressways are not necessarily on the segments operated by CFX. The number of crashes for each roadway that can be matched in S4A system was also retrieved. The results confirmed that S4A can replace CAR database in our research. As a result, in the safety analysis, the crash data were collected from S4A system.

Table 3-4: Crash in CAR and S4A

Expressways	Crash Count in CAR	Crash Matched in S4A
SR 408	731	730
SR 414	270	264
SR 417	401	401
SR 429	110	110
SR 528	592	590

3.7 Crash Data Preparation

The crashes contained in S4A are geocoded data with longitude and latitude information. Nevertheless, crash direction, and roadway milepost are not available. To locate these crashes and assign the direction and milepost information to these crashes, a Geographic Information System (GIS) network specifically for the expressways is created using ArcGIS. The original GIS data is downloaded from FDOT website. The research team made adjustment to keep only the expressways as shown in Figure 3-7.

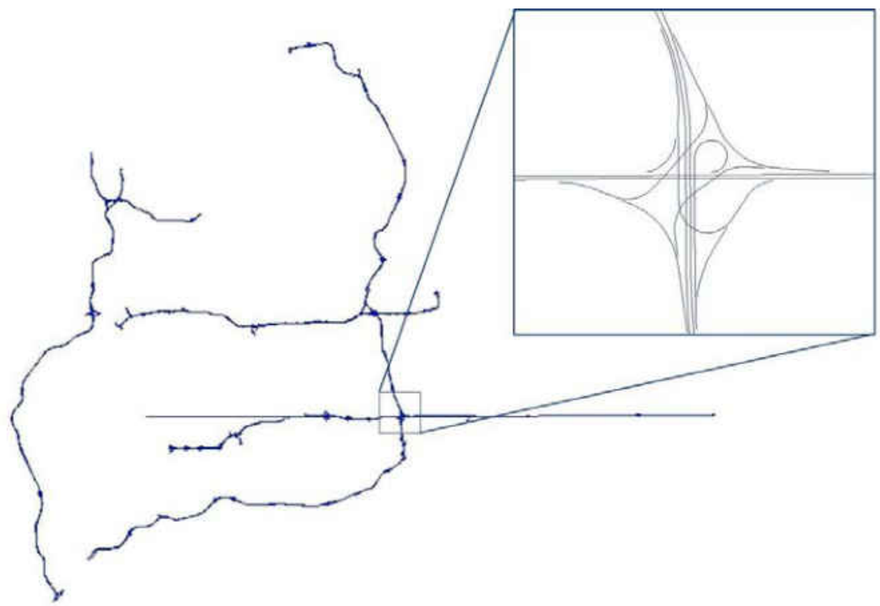


Figure 3-7: Expressway Network in GIS

As seen in Table 3-5, a small portion of crashes during each year were unmapped and could not be located. The research team referred to the original crash report for indications of crash location and found several issues that might cause this issues. First of all, these unmapped crashes lacked the coordinate information and were assigned with the location (0, 0). Second, the description of these crashes was unclear therefore the location could not be identified. For instance, one crash was described as "near Holland Toll Plaza" for which the toll plaza could not be found on SR 408. Third, some crashes have conflicting descriptions about the crash site such as that the crash street is recorded as SR 408. However, in the narrative part the police officer described that the crash was on SR 414. In conclusion, these crashes were not used for the safety analysis.

Table 3-5: Crash Data Preparation for the CFX Expressway System

Year	Orange County Crashes	Expressway Mapped Crashes	Expressway Unmapped Crashes	SR 408 Crashes	SR 414 Crashes	SR 417 Crashes	SR 429 Crashes	SR 528 Crashes
2012	35847	1375	88	626	25	369	78	277
2013	40476	1484	79	700	40	355	76	313
2014 (Jan -- Jun)	21322	859	35	361	35	218	67	178

To locate the crashes on the expressways, all of the crashes occurring within Orange County during a specific year were selected in the first place as shown in Figure 3-8. Then an initial selection of crashes on the expressways was conducted. In the crash report, there is one column indicating the crash street based on which crashes on expressways could be collected. However, the naming of the expressways is not consistent. As a solution, several key words that can be used for the same expressway were extracted using the Structured Query Language (SQL) technique as shown in Table 3-6. The "%" means any string of zero or more characters and "_" means any single character within the string in SQL. Using these criteria, the initial selection was made as displayed in Figure 3-9.

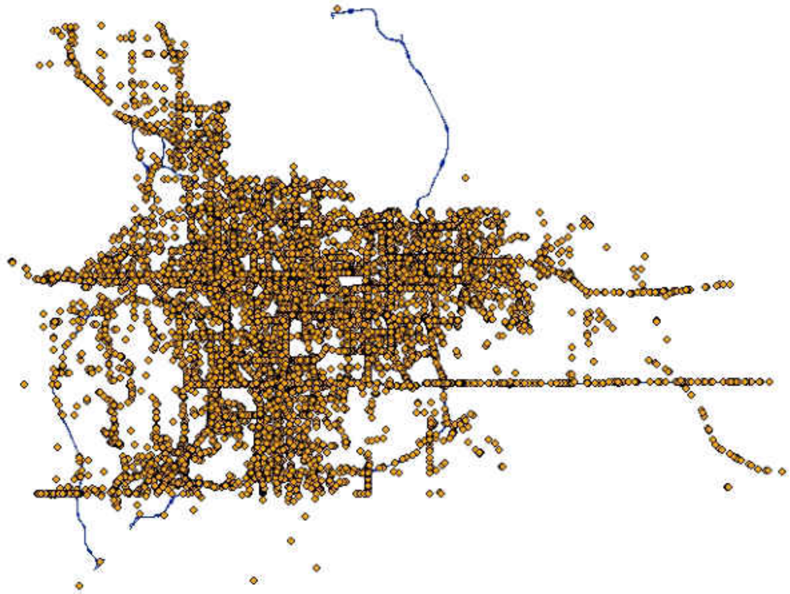


Figure 3-8: Total Crashes of Orange County in 2013

Table 3-6: Key Words Used for Expressway Crash Selection

Expressway	Key Words
SR 408	"%408%", "%E-W%", "%E/W%", "%EAST_WEST%", "EW %", "%EASTWEST%"
SR 414	"%414%", "%APOPKA EXPY%", "%JOHN LAND%", "%MAITLAND BLVD%"
SR 417	"%417%", "%CENTRAL_FL%", "GREENEWAY"
SR 429	"%429%", "%BELTWAY%", "%WEBSTER%"
SR 528	"%528%", "%BEELINE%", "BEACHLINE"

Figure 3-9: Initial Selection of Expressway Crashes

As can be seen in Figure 3-9, the majority of the crashes after the initial selection are located on the expressways. A few of the crashes that are not related to the expressways were also included because they share the same key words that are used to filter the expressway crashes. In addition, some crashes on the expressways occurred on the segments that are not operated by CFX. These crashes would also not be included in further analysis. Therefore a final selection based on the roadway segments that are operated by CFX was executed. The roadway information for each expressway was archived in FDOT Interchange Report (2012). It should be noted that this report is last updated in 2012 therefore a few sections that underwent changes might not be reflected in Table 3-7. For northern segment used to be part of SR 429 is now SR 451. And currently SR 414 and SR 429 share one segment together. Therefore SR 414 now extends for 9.62 miles. The interchange of SR 408 and SR 417 was changed during these time period, thus the segment from

MP 16.649 to MP 17.090 no longer exists. For the other parts of the system, most segments remain the same.

Table 3-7: Expressway Segment and Operation Authority

Route	County	Roadway ID	Local Milepost	Cumulative Milepost	Authority
SR 528	Orange	75471000	00.000 – 08.421	00.000 – 08.421	Florida's Turnpike
		75002000	07.944 – 30.341	08.421 – 30.818	CFX
		75005000	00.000 – 04.957	30.818 – 35.775	Florida's Turnpike
	Brevard	70007000	00.000 – 09.956	35.775 – 45.731	Florida's Turnpike
		70070000	05.200 – 12.968	45.731 – 53.499	Florida's Turnpike
SR 408	Orange	75474000	00.000 – 00.759	00.000 – 00.759	Florida's Turnpike
		75008170	01.417 – 05.132	00.759 – 04.474	CFX
		75008000	00.382 – 11.852	04.474 – 15.944	CFX
		75008160	00.000 – 06.260	15.944 – 22.204	CFX
SR 414	Orange	75340000	00.000 – 09.620	00.000 – 09.620	CFX
SR 417	Osceola	92472000	00.000 – 02.906	00.000 – 02.906	Florida's Turnpike
	Orange	75472000	00.000 – 02.192	02.906 – 05.098	Florida's Turnpike
		75301000	00.000 – 20.017	05.098 – 25.115	CFX
		75300000	00.000 – 11.501	25.115 – 36.616	CFX
		77470000	00.000 – 17.445	36.616 – 54.061	Florida's Turnpike
SR 429	Osceola	92473000	00.000 – 04.528	00.000 – 04.528	Florida's Turnpike
	Orange	75473000	00.000 – 05.325	04.528 – 09.853	Florida's Turnpike
		75320000	18.000 – 40.000	09.853 – 31.853	CFX

The results of the final selection are illustrated in Figure 3-10. In this figure, the crashes not related to expressways and those crashes not occurring on segments operated by CFX have been excluded. In the final crash data, crashes on the mainline, ramps and toll plaza cash lanes on the segments managed by CFX are selected. Figure 3-11 shows the detail about how these crashes are assigned. Both crash direction and mileposts are assigned to the crashes using ArcGIS. For each year, the same process was repeated to extract the expressway crash data from 2012 to June, 2014.

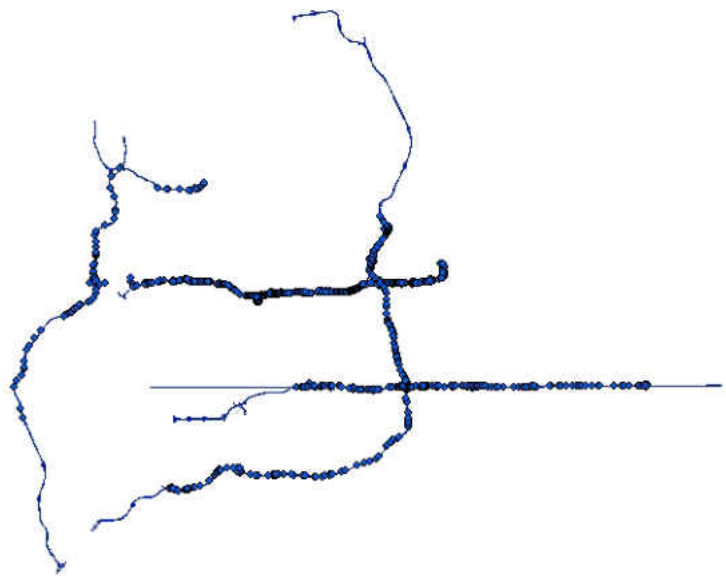


Figure 3-10: Final Selection of Expressway Crashes in 2013

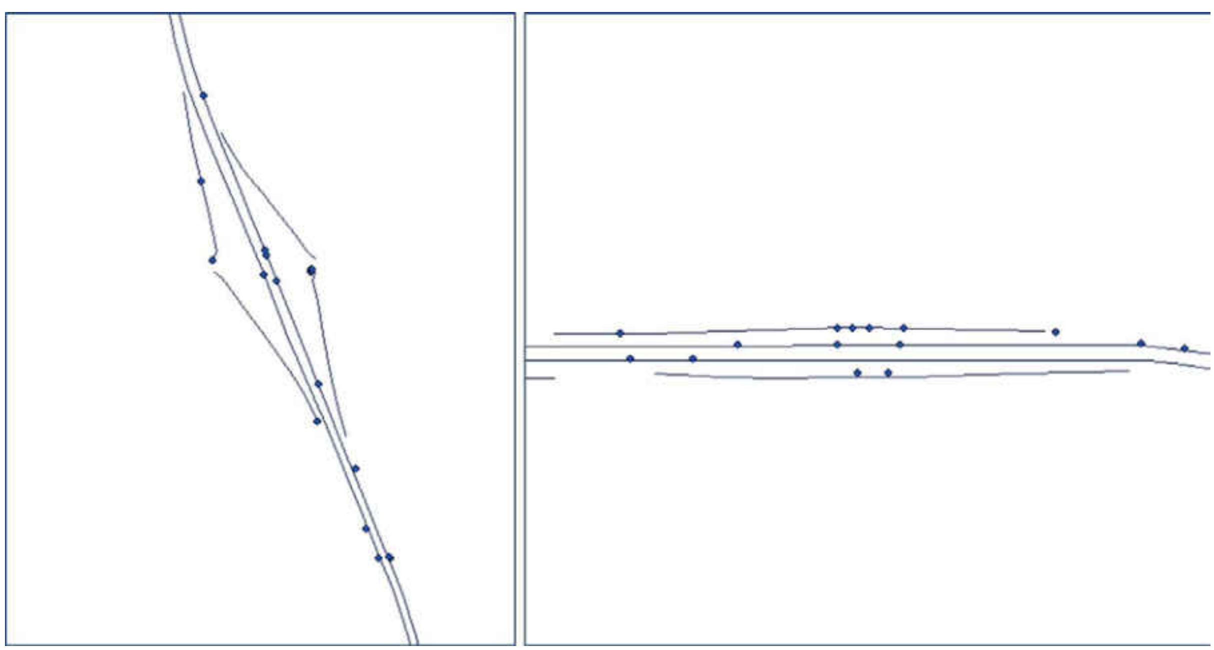


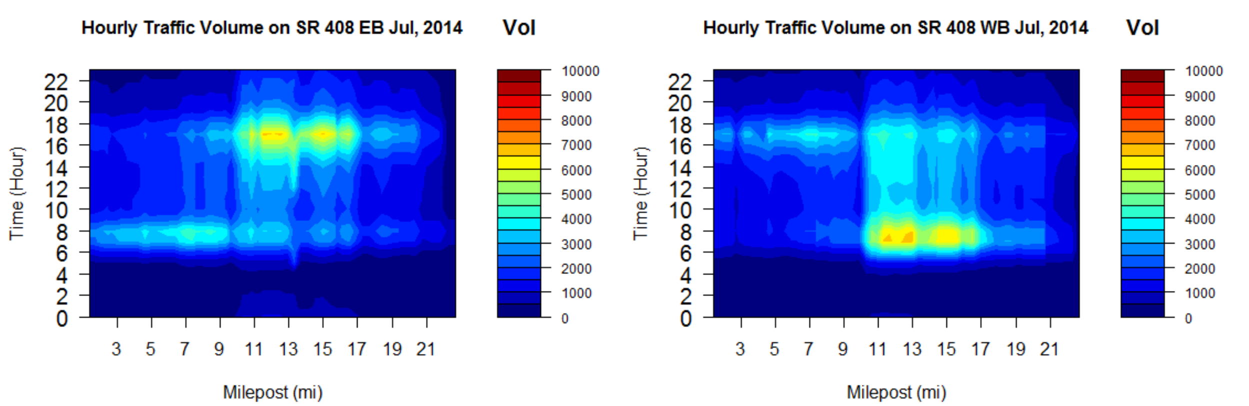
Figure 3-11: Crash Match on Mainline, Ramp and Toll Plaza Cash Lanes

The crash count for each expressway in each year is shown in Table 3-5 at the beginning of this section. SR 408 has the most crashes and SR 414 has the lowest crash count during each year. SR 417 has slight more crashes than SR 528. Several factors can contribute to the crash pattern on the expressway system. SR 408 is the spine of the system and carries the most traffic. SR 417 and SR 528 have relatively long segment length. However, the eastern part of SR 528 is located in suburban area, thus lighter traffic on this segment lead to fewer crashes compared with SR 417. SR 414 is the shortest of the five expressways and most crashes on SR 414concentrate near the eastern end of the roadway segment.

CHAPTER 4: TRAFFIC EFFICIENCY EVALUATION

4.1 Overview

As stated in the system overview in section 3.2, the five expressways are located in Central Florida area. SR 408, SR 528 and SR 414 travel along east-west direction; SR 417 and SR 429 travel along north-south direction. SR 408 carries the most traffic in the expressway system, especially commuting traffic. By taking the merit of MVDS traffic data, traffic operation on the expressways can be examined at more microscopic level. MVDS data from July, 2014 was selected to represent the most recent traffic operation states on the expressways. As illustrated in Figure 4-1 taking SR 408 as an example, the spatial-temporal characteristics of weekday average hourly traffic volume on mainline could be easily captured.



(a) (b)

Figure 4-1: Spatial-Temporal Hourly Volume Distribution on SR 408 (a) Eastbound and (b) Westbound

For SR 408, Eastbound experiences significant high demand during evening rush hours while the traffic reaches its peak on Westbound during morning rush hours. Hourly traffic volume on SR 408 during peak hours can rise to above 7000 vehicles per hour. The high demand exists around 6:00 to 9:00 AM in the morning and 16:00 to 19:00 PM in the afternoon. The segments that experience the high volume extend from around Milepost (MP) 11 to MP 17. For other segments and other time period, it can be seen that the traffic volumes are relatively stable and mostly below 3000 vehicles per hour. This preliminary review of SR 408 suggests when and where the congestion is likely to occur. Future congestion evaluation should focus on these segments during peak hours. By viewing the traffic demand at both spatial and temporal dimension, it has also been confirmed about how dynamic the traffic flow can be. Use of ITS traffic detection data enables operators and researchers to have precise and detailed knowledge about the performance of their roadways.

In the same way, it is found that SR 414 has relatively low hourly volume even during the morning (for Eastbound) and evening (for Westbound) peak hours. In non-peak hours, the hourly volume is generally below 600 vehicles per hour. During peak hours, there will be moderate increase in traffic volume. However, the peak hour traffic volume on SR 414 is still below 2000 vehicles per hour, which is similar to non-peak hourly volume on SR 408. Moreover, since SR 414 has the shortest segment length of the five expressways, traffic increases on the whole roadway segment during the peak hours.

SR 417 shows the similar pattern during the peak hours. Nevertheless, from around MP 7 to MP 27, on the 20-mile segment, hourly traffic volume only increases mildly in the morning and evening peak hours. In contrast, from MP 27 to MP 37, the hourly volume is significantly higher during the same time period. Especially for Northbound, hourly traffic can reach to about 5000 vehicles per hour around 17:00 PM.

SR 429 Northbound experiences mild traffic increase in the evening and Southbound in the morning. Peak hour volume is relatively low compared with SR 408, SR 417 and SR 528.

SR 528 accommodates most of its peak hour traffic on the segment from MP 8 to MP 13. Evening peak hours exert on SR 528 Eastbound while morning peak hours appear on Westbound. During peak hours, the hourly traffic volume can reach to 4000 vehicles per hour.

Synthesizing the mainline operation on the five expressways, it can be seen that SR 408 and SR 528 share similar traffic patterns and SR 417 and SR 429 share the similar traffic pattern. Considering the expressway locations in Figure 3-1, it can be seen that in the morning, traveling directions that experience higher demand is towards downtown Orlando while in the evening the direction is opposite. This pattern is straightforward considering the function of downtown area and should be taken into account for further analysis. Table 4-1 summarizes the findings in the preliminary analysis based on hourly traffic volume.

Table 4-1: Expressway System Operation Overview

Expressway	Traveling Direction	Direction of Morning Peak Hour	Segments with High Hourly Volume	Direction of Evening Peak Hour	Segments with High Hourly volume
SR 408	EB – WB	WB	MP 11 -- MP 17	EB	MP 11 -- MP 17
SR 528	EB -- WB	WB	MP 09 -- MP 13	EB	MP 09 -- MP 13
SR 414	EB -- WB	EB	MP 04 -- MP 09	WB	MP 04 -- MP 09
SR 417	NB -- SB	SB	MP 27 -- MP 37	NB	MP 27-- MP 37
SR 429	NB -- SB	SB	MP 24-- MP 30	NB	MP 24 -- MP 30

Besides mainline traffic conditions, ramps were also examined. In most cases, one MVDS sensor will be installed at diverging area near off-ramp and merging area near on-ramp. As shown in Figure 4-2, if a ramp is on-ramp, then its volume will be represented as a positive value. If a ramp is off-ramp, the volume will be negative indicating the traffic leaves the expressway system. Ramps on SR 408 Eastbound have been used as an illustration. The daily volume is calculated as the average volume at each detection location for each weekday in July, 2014. SR 408 has several ramps with daily volume above 25000 vehicles per day on weekdays. On other expressways, high volume on ramps usually ranges from 10000 to 20000 vehicles per day. Considering the milepost of the ramps with high traffic volume, it can be found that most of these ramps are located near interchanges with other arterials that also have high traffic demand on weekdays.

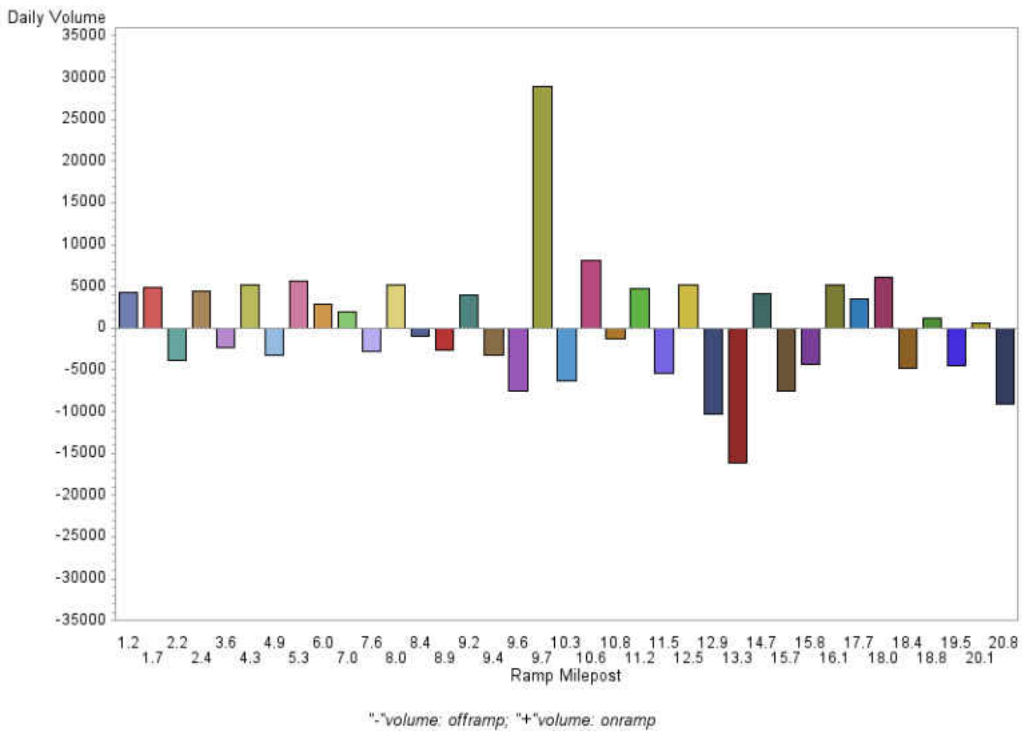


Figure 4-2: SR 408 Eastbound Weekday Ramp Traffic Volume

By reviewing the traffic on expressway mainline and ramps, a general impression of current expressway operation can be gained. For further evaluation, mainline operation will be thoroughly investigated. Ramps with daily volume higher than 10000 vehicles per day will also be studied.

4.2 Congestion Measurement

Traffic operation on expressways focuses on providing motorists with efficient movement to their destinations. To achieve this goal, reducing congestion is the most important task. Measuring congestion accurately is a prerequisite in congestion management. Traditionally, volume-to-capacity (V/C) ratios and level of service (LOS) are implemented by transportation authorities as indicators of congestion intensity (Grant *et al.,* 2011). Nevertheless, traffic demand can vary considerably in both temporal and spatial dimensions and roadway capacity can be reduced by incidents as discussed above. In such cases, V/C ratios and LOS lack the capability to capture the variability of congestion. With the fast development of ITS technology, real-time congestion measurement is becoming an urgent call. On the expressways, AVI and MVDS traffic detection systems are employed. Both of these systems archive the traffic data in real-time manner. In this project, multiple congestion measures were introduced and compared based on these two systems.

4.2.1 AVI-based Congestion Measurement

Congestion can be measured from three aspects, namely density, travel time and travel speed. AVI system is able to calculate the travel time of vehicles on a segment. Therefore, congestion measured by travel-time was introduced for the AVI system.

Travel time index (TTI) is the commonly accepted measure used to evaluate traffic congestion. It is defined as the ratio of actual travel time to an ideal (free-flow) travel time (Systematics 2005) as shown in Equation (4-1)

$$\mathrm{TTI} = \frac{\text{actual travel time}}{\text{free flow travel time}} \tag{4-1}$$

It indicates the additional time spent on a trip made during peak traffic hours compared to an ideal trip on the same corridor. On CFX expressway system, free flow travel time for each segment is offered in AVI traffic data. Free flow travel time is calculated based on segment length and average speed limit. Average speed limit of a segment accounts for that speed limits may vary within the segment. From the Enhancing Expressway Operations Using Travel Time Performance Data (Griffin, 2011), the levels of congestion and the corresponding travel time index for the studied expressways are listed in Table 4-2:

Table 4-2: Travel Time Index and Congestion Levels

Functional Class	Travel Time Index for different Congestion Levels		
	Below congestion threshold	Moderate Congestion	High Congestion
Freeway	Less than 1.25	1.25 – 1.99	Higher than 2.00

4.2.2 MVDS-based Congestion Measurement

Different from AVI system, MVDS sensors reflect the traffic conditions at the installed points rather than segments. Speed, volume and lane occupancy will be archived on one-minute interval basis.

Multiple congestion measures can be developed from the MVDS traffic data. Occupancy is defined as the percent of time a point on the road is occupied by vehicles (Hall 1996). Gerlough and Huber (1975) referred to occupancy as a surrogate for density. Compared with traditional V/C Ratio or LOS, occupancy has the advantage that it could be monitored in real-time.

The speed detected by MVDS detector is spot speed and the rate of reduction in speed caused by congestion from the free flow speed condition is adopted as congestion index (Hamad and Kikuchi, 2002; Hossain and Muromachi, 2012). The congestion index (CI) is expressed as

$$\mathrm{CI} = \frac{\text{free flow speed} - \text{actual speed}}{\text{free flow speed}} \text{ when } \mathrm{CI} > 0; \qquad (4\text{-}2)$$

$$= 0 \text{ when } \mathrm{CI} \leq 0$$

The CI is a continuous congestion indicator ranging from 0 to 1. The free flow speed is the 85th percentile speed at the corresponding detection point. From equation (4-2) above it can be seen that when the actual speed is above free flow speed, CI will be recorded as 0. When CI increases, the congestion becomes more severe.

Currently, for the congestion measures calculated from MVDS data, no specific relationship between occupancy/CI and level of congestion is available. However, the TTI value of 1.25 and 2 are approximately equivalent to CI value of 0.2 and 0.5. And According to the congestion plots, when CI reaches 0.2 and 0.5, the corresponding occupancy (%) is about 15 and 25. Therefore, the research team set up the following congestion levels defined by occupancy and CI as displayed in Table 4-3. Nevertheless, further refinement of these thresholds might be possible.

Table 4-3: MVDS-Based Congestion Measures and Congestion Levels

Congestion Measure	Expressway Congestion Levels		
	Below congestion threshold	Moderate Congestion	High Congestion
Occupancy (%)	≤ 15	15 – 24.99	≥ 25
CI	≤ 0.2	0.2 – 0.499	≥ 0.5

In conclusion, expressway mainline congestion will be evaluated using the three congestion measures using traffic data from the two traffic detection systems. For ramps, only occupancy and CI could be used since no AVI segment is available on ramps.

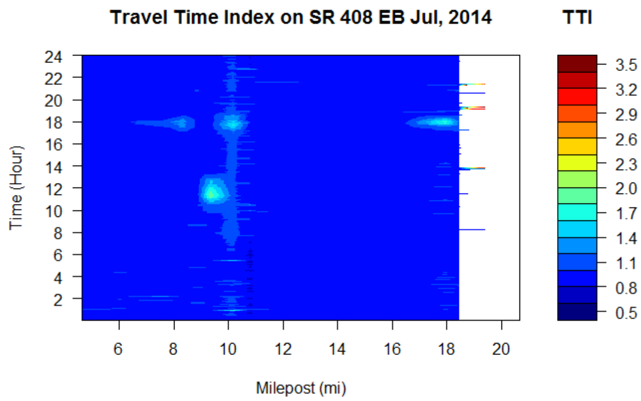
4.3 Expressway Congestion Evaluation

4.3.1 Mainline Congestion

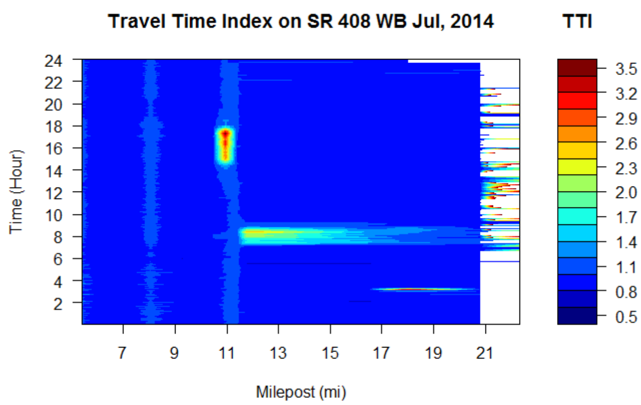
Two major efforts have been made to evaluate the congestion for the expressways. One is the evaluation of spatial-temporal distribution of current congestion on the expressways. The other effort is to identify the trend of congestion during the past one year to determine whether congestion is worsened or alleviated on the expressways. For this longitudinal comparison, five months during the past year was selected considering the availability of traffic data.

To measure current expressway congestion conditions, the traffic data were aggregated at five-minute interval and averaged by the weekdays for each month. Contour plots were generated to illustrate the spatial-temporal property of the congestion.

The TTI congestion plot shown in Figure 4-3 show a proportion of data near MP 3.0 and MP 20 were missing for both directions in July, 2014. As mentioned in the section about AVI system, the expressway system undertook major update in April, 2014. As a result, in the recent three months (May, 2014 – July, 2014) the AVI traffic data have the similar issues. For specific segments, no records were available during specific time period. Despite the incompleteness of the AVI data, some patterns could still be found from Figure 4-3, on SR 408 Eastbound, congestion is found near MP 9.0 and MP 18 in the evening peak hours. On SR 408 Westbound, morning congestion is observed from MP 11 to MP 15. These congestion patterns could also be found in Figure 4-4 and Figure 4-5, indicating that AVI data could reflect congestion to certain extent. However, it is still important to have the complete data to evaluate the performance of AVI-based congestion measure. In the following analysis, TTI won't be used to evaluate the current congestion on the system due to this data completeness issue.



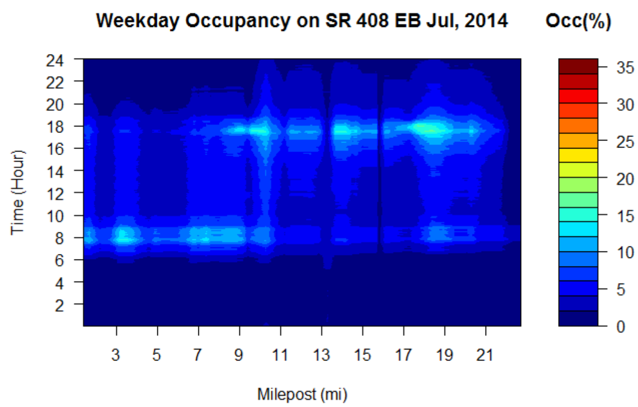
(a)



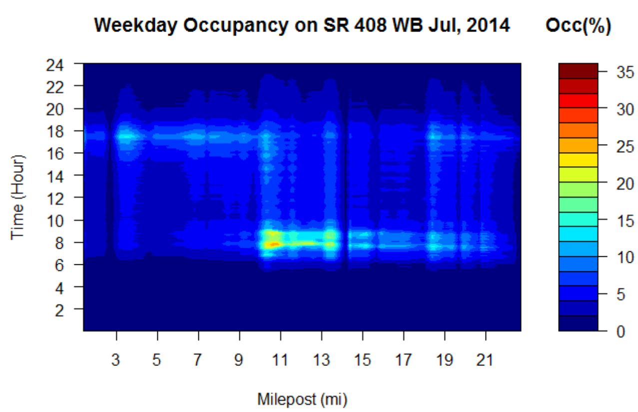
(b)

Figure 4-3: Mainline Weekday Travel Time Index of SR 408 (a) Eastbound and (b) Westbound

The congestion plots derived from occupancy and CI (Figure 4-4, Appendix D, Figure 4-5 and Appendix E) exhibit comparable congestion patterns for the expressways. As mentioned above, the number of MVDS sensors installed along the expressways is significantly more than that of the AVI sensors. In addition, the MVDS system is stable in terms of active sensors during the study time period. Therefore, the MVDS data is relatively complete and stable.

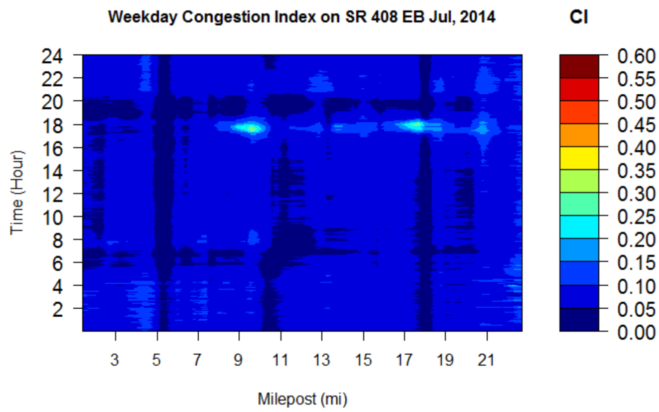


(a)

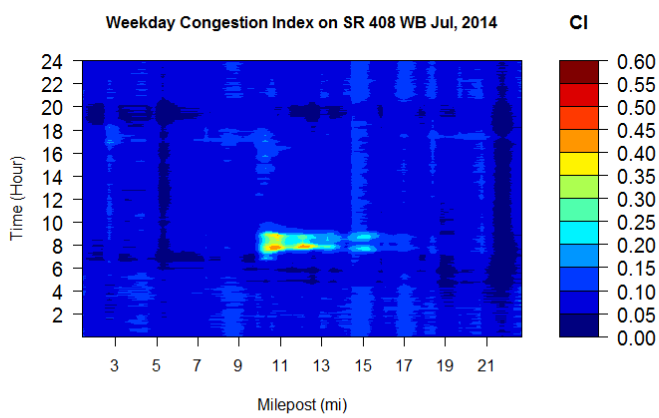


(b)

Figure 4-4: Mainline Weekday Occupancy of SR 408 (a) Eastbound and (b) Westbound



(a)



(b)

Figure 4-5: Mainline Weekday Congestion Index of SR 408 (a) Eastbound and (b) Westbound

Based on occupancy and CI, congestion conditions on the five expressways can be summarized. SR 408 experiences moderate congestion on Eastbound in morning peak hours and high congestion on Westbound in the evening peak hours. However, it should be noticed that the congestion intensity changes with time. When it is approaching peak hours, the congestion intensity gradually increases. Once the peak time is passed, the congestion becomes less severe.

The congested area for SR 408 is approximately from MP 17 to MP 19 on Eastbound and from MP 10 to MP 13 on Westbound. For SR 414, only the Eastbound experiences moderate congestion during morning peak hours, the congested segment is near MP 9.3. For SR 417, both directions experience moderate congestion during evening peak hours, the congested segments are short, only near the interchanges with University Blvd which leads to University of Central Florida. No congestion was detected on SR 429. On SR 528, congestion is detected on Eastbound in the evening and on Westbound in the morning.

In conclusion, based on the most recent MVDS data, the five expressways show distinct congestion patterns. SR 408 is affected by congestion most significantly. SR 528 also experience high congestion on the segment near MP 10 to MP 12. SR 417 and SR 528 only expect moderate congestion on specific locations for short time intervals. SR 429 has no congestion either in the morning or evening peak hours.

In the above congestion evaluation, the three candidate congestion measures were all applied. For a more detailed assessment of their performance, one segment on SR 408 Westbound with traffic data from February 2014 was extracted for the comparison. The selected AVI segment runs from MP 16.488 to MP 15.245. Within the AVI segment, five MVDS sensors are installed at the locations as shown in Figure 4-6.

The Figure 4-6 indicates that in the morning peak hours, both MVDS and AVI detect congestion on the segment. However, there is about one hour lag for the peak between the two types of data.

The time period for congestion by the CI is considered more reliable since congestion was mostly found from 7:00 AM to 9:00 AM. On the other hand, congestion by TTI occurred from 8:30 AM to 9:30 AM. Furthermore, CIs from MVDS sensors within one AVI segment differ from each other. Downstream (MP 15.2) has the highest congestion intensity and upstream (16.5) has the lowest congestion intensity. However, the AVI data could not reflect this detail. The high deployment density of the MVDS system ensures better reflection of the congestion of at different locations within a queue.

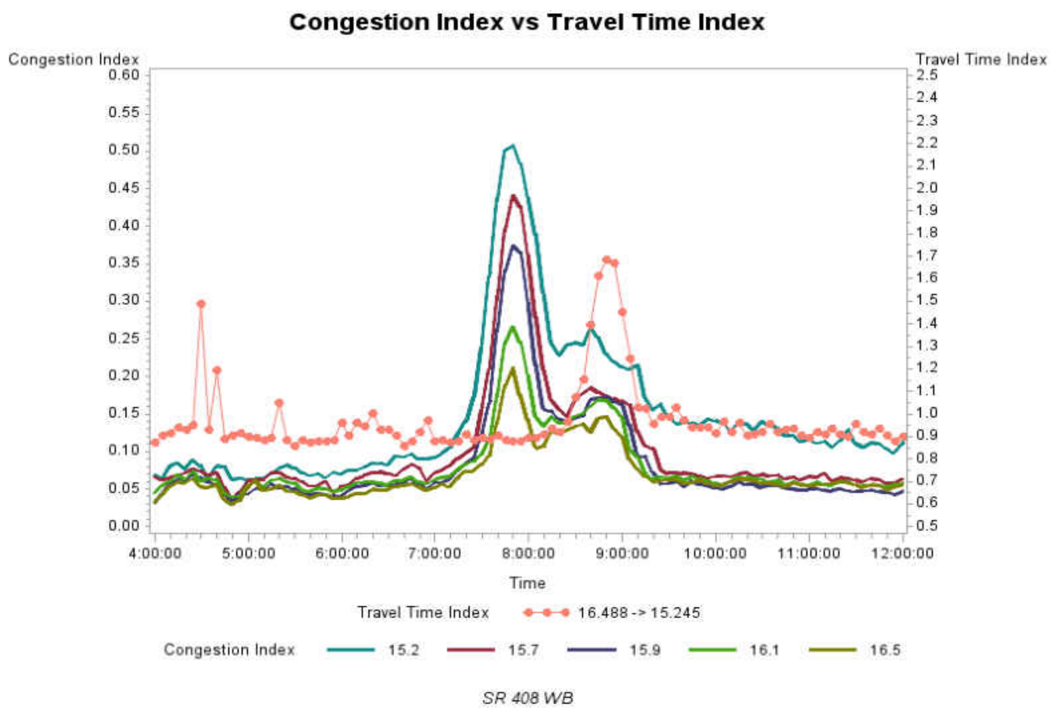


Figure 4-6: AVI-based TTI Profile vs MVDS-based Congestion Index Profile

Occupancy and CI are both derived from MVDS system. Therefore their performances in congestion detection are similar (Figure 4-7) according to several detectors. These findings imply that MVDS traffic data is more appropriate for congestion monitoring for several reasons:

1) MVDS system works in a more reliable manner, thus resulting in more complete data; 2) MVDS system has much more sensors than AVI system, consequently generating more detailed information about the expressways.

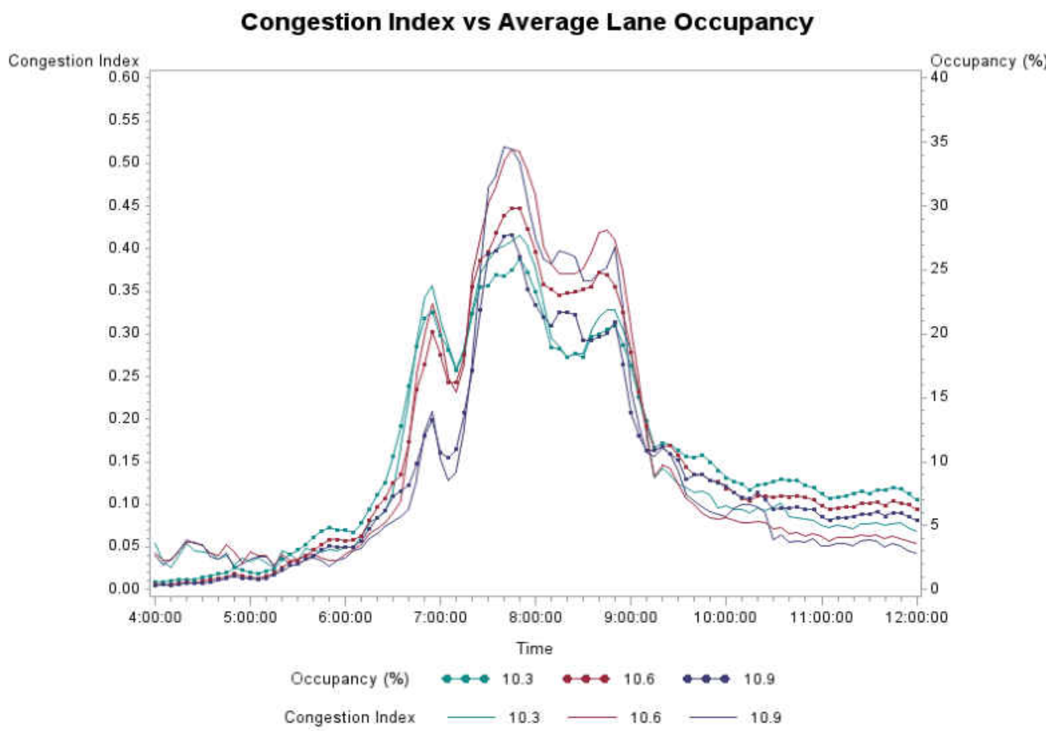


Figure 4-7: MVDS-based Congestion Index and Occupancy

In addition to the analysis of current congestion conditions on the expressways, the longitudinal trend of congestion during the past one year was also examined using the MVDS data. Five months, namely July 2014, May 2014, February 2014, November 2013 and August 2013 were selected. To see the trend of congestion, the system occupancy and CI were defined and calculated. The system occupancy and CI are the average occupancy and CI from all the MVDS sensors. They are used to represent the general congestion condition on the expressway (Figure

4-8, Figure 4-10). After the system congestion for each expressway is generated, the peak congested time can be identified. Then the detailed congestion information at different locations of the expressways at the time when the system congestion intensity is the highest can be evaluated (Figure 4-9, Figure 4-11). Overall, the longitudinal congestion evaluation consists of two parts, 1) whether the congestion has been alleviated during the last year for the whole expressway; 2) what are the changes for each specific location regarding to congestion intensity. The evaluation was conducted using both occupancy and CI.

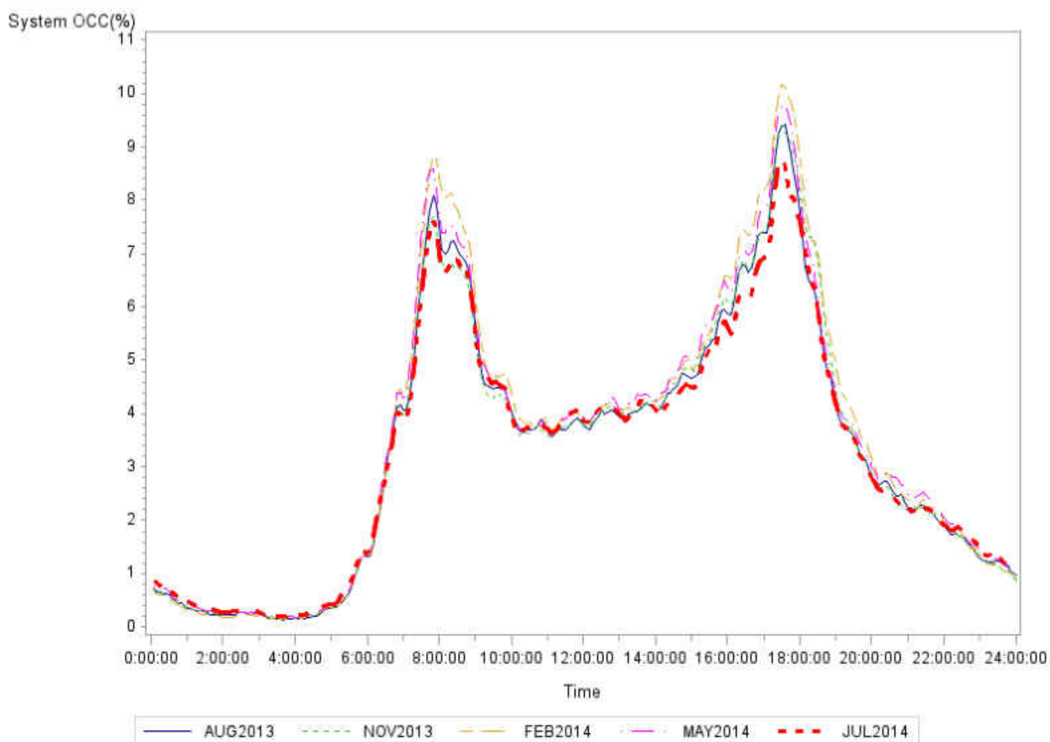


Figure 4-8: SR 408 Eastbound System Occupancy and Trend of Congestion

From a system point of view, Figure 4-8, Figure 4-10, etc. all confirmed that on SR 408, congestion is alleviated significantly in the recent month. Of the five months, both SR 408

Eastbound and SR 408 Westbound have the lowest system CI and occupancy. SR 408 is known as the spine of the expressway network since it carries the heaviest traffic load among the five expressways. Congestion on SR 408 is also the most severe compared with other expressways within the system. As a consequence, reduction in congestion intensity proves significant improvement in traffic operation. The same conclusion can be reached for SR 414. For SR 417 and SR 429, the improvement is not that significant yet congestion is not worsened. For SR 528, the conclusions from CI and occupancy diverge a little. According to the system occupancy, congestion conditions in the two months of 2013 were better than those in the three months of 2014. On the contrary, the system CI indicates that the three months in 2014 have less severe congestion than the two months in 2013. Nevertheless, at the system level, it cannot be judged that congestion on SR 528 has been improved significantly.

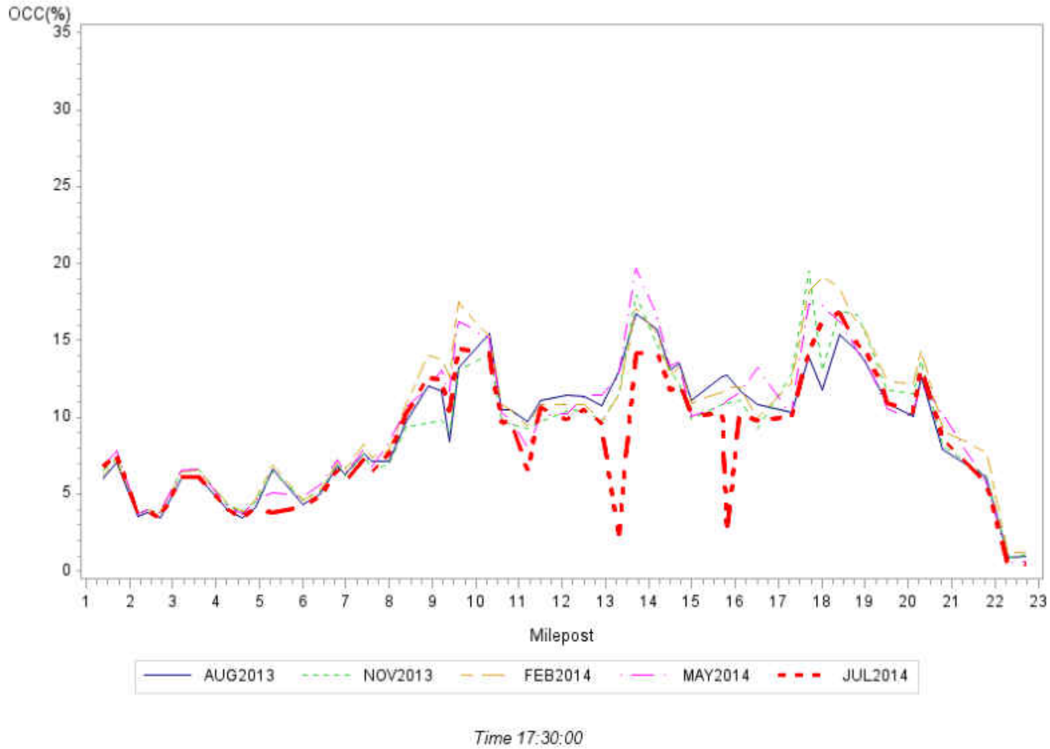


Figure 4-9: SR 408 Eastbound Peak Hour Occupancy and Trend of Congestion

Based on the system congestion evaluation, the exact time when the system reaches the peak congestion intensity can be identified. Congestion conditions at each detection point at this peak time were extracted for SR 408, SR 417 and SR 528. SR 414 and SR 429 were excluded because for SR 414 only one detection point has congestion and for SR 429 there is no congestion detected. This effort provides some insights into the trend of congestion at different locations along the road.

For SR 408 Eastbound (Figure 4-9 and Figure 4-11), congestion intensity measured by occupancy and CI implies that the congested segments experience relatively less severe congestion compared with February and May 2014, but not necessarily better than 2013. For Westbound, the alleviation of congestion at congested segments is significant.

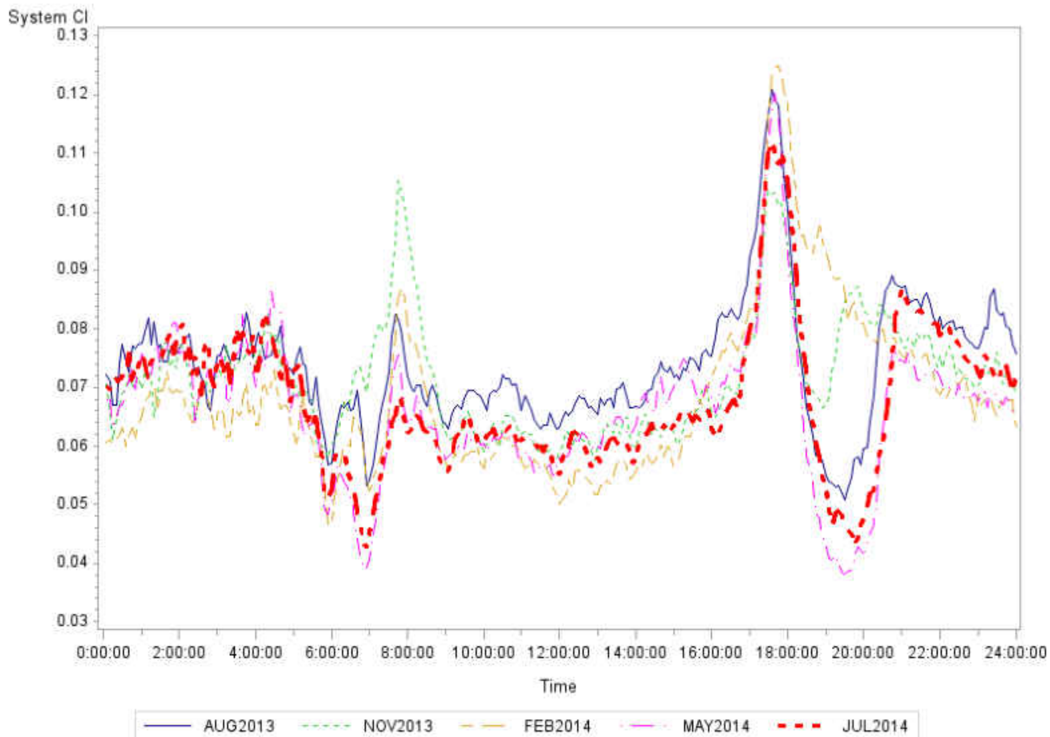


Figure 4-10: SR 408 Eastbound System Congestion Index and Trend of Congestion

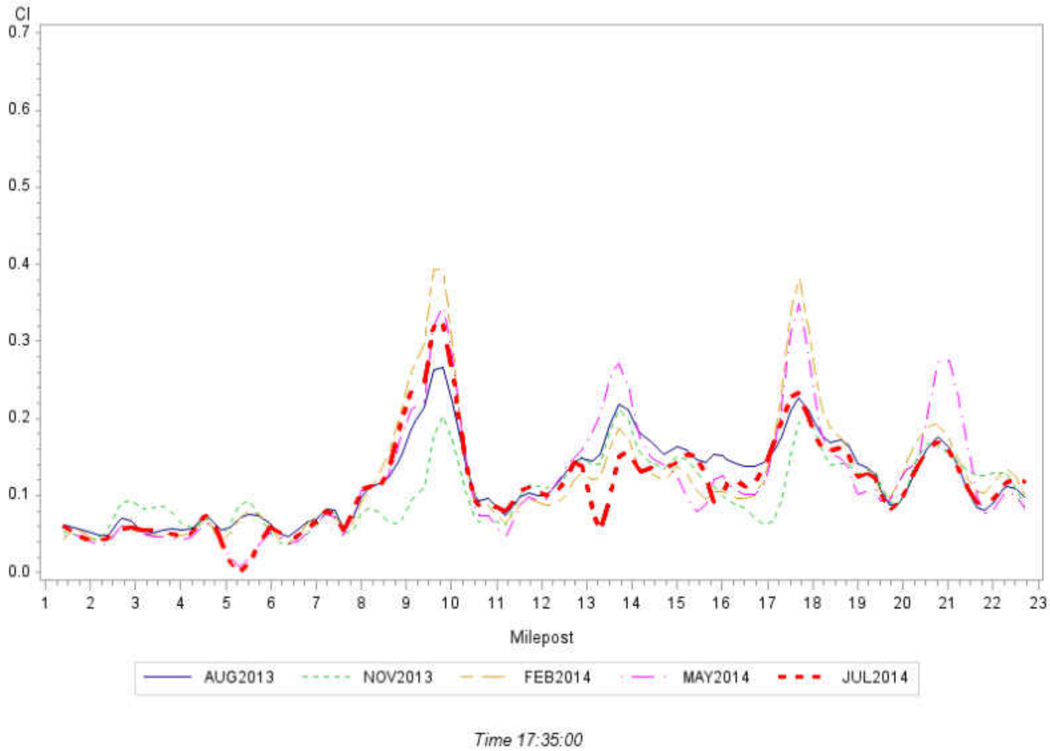


Figure 4-11: SR 408 Eastbound Peak Hour Congestion Index and Trend of Congestion

For SR 417, congestion intensity remains stable during the past one year according to occupancy. Based on CI, congestion increased near MP 31 for both Northbound and Southbound increased in a small amount. In conclusion, the congestion conditions on SR 417 did not change significantly.

For SR 528, congestion concentrated on the segment near MP 10 for both Eastbound and Westbound. Comparing the congestion at this segment in the five months during the past one year, it is suggested by both CI and occupancy that significant reduction of congestion intensity has been achieved.

As a result, the longitudinal analysis confirms operation improvement on SR 408, SR 414 and congested segments on SR 528. SR 417 and SR 429 remain stable during the past one year. Considering SR 408 and SR 528 are the expressways that experience most congestion, significant improvement on these two expressways indicate the successful management by the CFX.

4.3.2 Ramp Congestion

Ramps on the expressways lead the vehicles into the mainline and divert the vehicles out of the system. When congestion occurs on off-ramps, queues can build up on the mainline and affect the mainline congestion and safety. When congestion occurs on on-ramps, traveling speed on mainline and ramps are likely to differ significantly. The variation of speed near the merging area can greatly affect the mainline speed as well as probability of sideswipe crashes. Identifying ramps experiencing congestion will help improve both operation and safety on the expressways.

SR 408 has 37 ramps on Eastbound and 37 ramps on Westbound. SR 414 Eastbound has 8 ramps and SR 414 Westbound has 7 ramps. There are 31 ramps located on SR 417 Northbound and 32 ramps on SR 417 Southbound. SR 429 contains 17 ramps on Northbound and 16 ramps on Southbound. For SR 528, 19 ramps are on the Eastbound and 18 ramps on the Westbound. To identify the ramps that experience congestion, only the ramps having more than 10000 vehicles per day on weekdays were selected. If the ramps have less volume, they are unlikely to have congestion. Both occupancy and CI were used to evaluate the congestion conditions on the ramps.

As seen in Figure 4-12 and Figure 4-13, several ramps experience congestion during the peak hours on SR 408. For SR 414, SR 417 and SR 528, the ramps on the three expressways were found to have no congestion on them. The on-ramp at MP 9.7 on SR 408 Eastbound has moderate congestion during the evening peak hours according to the profile of CI. This on-ramp has relative high occupancy for most time of the day. The location of this on ramp is at the interchange of SR408 and I-4 in downtown Orlando. On SR 408 Westbound, congestion occurs on two off-ramps in the morning peak hours, at MP 9.9 and MP 10.3 respectively. Off-ramp at MP 9.9 takes vehicles off the expressway to I-4 Westbound and off-ramp at MP 10.3 takes vehicles from SR 408 to I-4 Eastbound. The congestion on SR 408 ramps indicate that in the morning, more traffic is traveling from I-4 to SR 408 while in the evening more vehicles are traveling from SR 408 to I-4.Except the ramps on SR 408, only one ramp on SR 429 was found to have congestion. The off-ramp is at MP 19.8 of SR 429 Southbound. During the evening peak hours, both CI and occupancy increased significantly and indicate moderate congestion. This congested ramp is located next to Winter Garden Village, a big shopping center and residential area. The land use property of this region explains why congestion is found on this particular ramp.

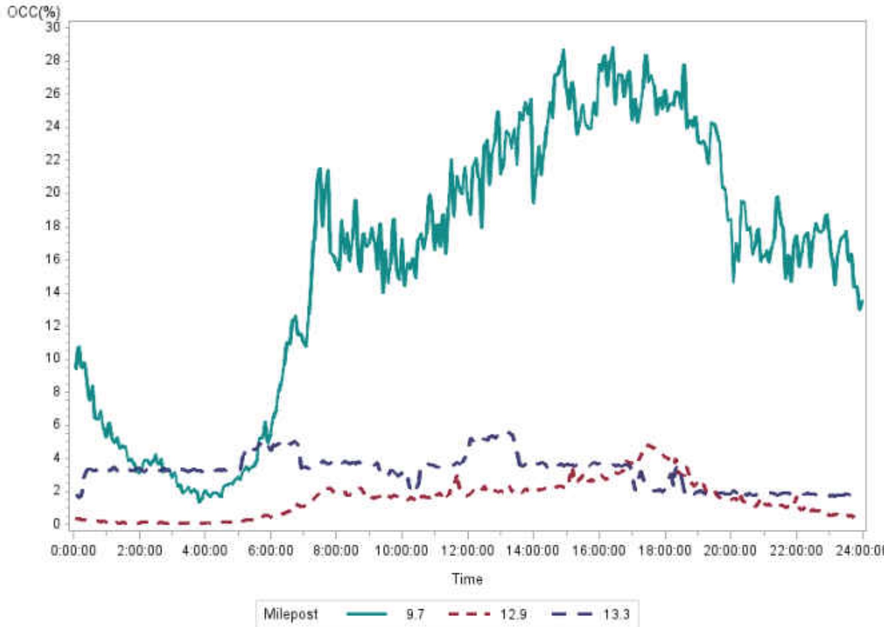


Figure 4-12: SR 408 Eastbound Ramp Occupancy Profile

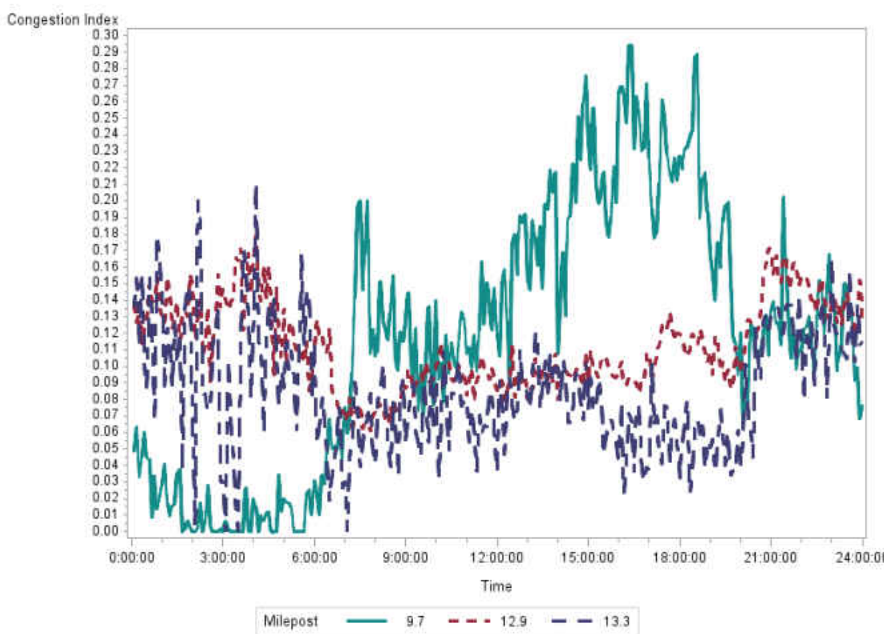


Figure 4-13: SR 408 Eastbound Ramp Congestion Index Profile

Results of the ramp congestion evaluation identify four ramps on the expressways that experience congestion. Three of the ramps are on SR 408 and one on SR 429. In addition, three of the four ramps are off-ramps and one is on ramp. Future operation improvement should also take these ramps into consideration.

CHAPTER 5: DMS APPLICATION IN CONGESTION MANAGEMENT

5.1 DMS Application in Queue Warning

Traffic congestion results in travel delay, excess fuel consumption and congestion costs. In 2010, Orlando ranked the 3rd in Florida in terms of the above three standards (38,260,000 hours delay, 11,883,000 gallons excess fuel, and 811 million dollars congestion costs) (Schrank *et al.* 2012).

Traditionally, increasing road capacity has been a common countermeasure to alleviate congestion. Due to the limited land resources, public investment and increasing travel demand, only adding more capacity to the system is not a long-term solution. Public transportation, road pricing and other measures to reduce traffic are also encouraged or implemented to ease traffic congestion. However, these steps are not able to change the current supply and demand greatly and quickly enough. Consequently, traffic management is crucial in order to enhance the efficiency of the road and lessen the intensity of congestion.

Recent years saw the rise of active traffic management (ATM). It is a combination of congestion management techniques. The goal of the ATM strategies is to better the roadway system's performance on road segments experiencing frequent congestion, or susceptible to crashes, bottlenecks and adverse weather conditions.

Queue warning is a strategy to warn drivers of upcoming congestion and allow drivers enough time in advance to make decisions whether to detour or stay on the route. From a safety perspective it also allows drivers enough decision and reaction time to slow down. Dynamic message signs (DMS) are the most common media to convey the traffic information to upstream drivers. Variable speed limits and lane control signals often work together with queue warning signs.

The potential benefits offered by the queue warning system include prevention of primary and secondary crashes (also the severity is expected to be reduced), delay of the onset of congestion, and travel time improvement. The basic principle behind queue warning is alerting the drivers to congestion conditions, drivers' caution will rise, and more smooth and uniform traffic flow be achieved.

5.2 Expressway Congestion Area and Suggested DMS Locations

On the expressways, the congestion areas have been identified in previous chapter. Based on the identified congested segments or locations, the DMS upstream can be used for congestion warning. Furthermore, for the congested ramps, DMS can also be used to remind motorists about the traffic condition on downstream ramps. In case that no DMS exists upstream to the congested segments, future DMS can be considered at these places.

Table 5-1 shows the congested segment on SR 408 Eastbound is located around MP 18. It can be seen from the table that the length of the queue changes over time. In total the congestion lasts for about 40 minutes from 17:30 PM to 18:10 PM. The maximum length of the queue was extracted to represent the congestion condition. According to the end of the queue when it is most congested, the upstream DMS is located. Figure 5-1 visually displays the congested segments and nearest upstream DMS. The beginning of queue (MP 18.4) is near the interchange with Dean Road and the end of the queue (MP 17.7) is at the interchange with SR 417. The nearest upstream DMS is at MP 15.2. The distance between the DMS and end of queue is 2.5 miles, which can warn drivers when they approach the queue.

Table 5-1: SR 408 Eastbound Congestion Area

Hour	17						18	
Minute	30	35	40	45	50	55	0	5
Beginning of Queue	18.4	18.4	18.4	18.4	18.4	18	18	18
End of Queue	18	18	17.7	17.7	17.7	17.7	17.7	17.7

SR 408 Westbound as shown in Table 5-2, Figure 5-2 and Figure 5-3 has two queues in the morning peak hours. The first queue is near Orlando Executive Airport. The congestion exists from Conway Road Toll Plaza to the interchange with Semoran Blvd. One DMS is located at MP 15.2. The second queue is from the interchange with I-4 to interchange with Crystal Lake Drive, about 2.3 miles long. Although one DMS is found at MP 11.8 within the congested segment, no DMS upstream can be used for queue warning. Therefore, one DMS is suggested near MP 13.6 to MP 14.6 for potential implementation in the future for queue warning in Figure 5-3.

Congestion on SR 414, SR 417 and SR 528 and the existing or suggested DMS could be identified in the same way. On specific expressways, only one detector reports congestion. For these conditions, the place with congestion will be specified as congestion location. SR 414 Eastbound experiences congestion near the end of the roadway segment. Only one MVDS sensor at the interchange with Orange Blossom Trail detects congestion for around 15 minutes from 7:50 AM to 8:05 AM. However, there is currently no DMS upstream to this congestion location. Correspondingly, potential location for new DMS is suggested to be MP 7.3 to MP 8.3.

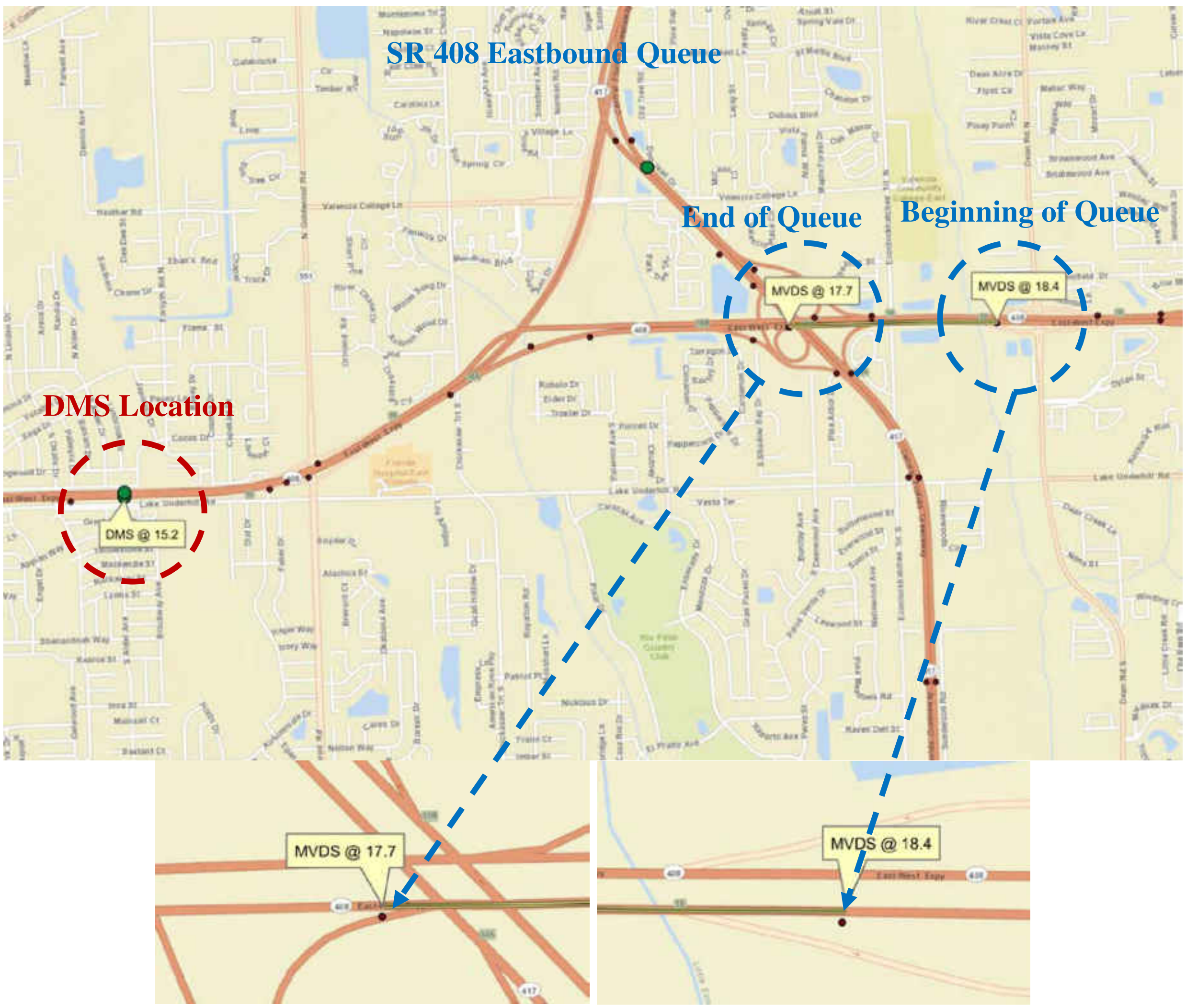


Figure 5-1: SR 408 Eastbound Congestion Segment and Upstream DMS Location

Table 5-2: SR 408 Westbound Congestion Area

Hour	7								8												9	
Minute	20	25	30	35	40	45	50	55	0	5	10	15	20	25	30	35	40	45	50	55	0	5
Beginning of Queue 1	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3		
End of Queue 1	13.6	13.6	13.6	13.6	13.6	14.4	14.4	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6	13.6		
Beginning of Queue 2	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3	10.3
End of Queue 2	10.6	10.6	10.6	11.6	12.1	12.6	12.6	12.6	12.6	12.1	11.6	11.6	11.6	11.6	11.6	11.6	11.6	11.6	11.6	10.9	10.6	10.6

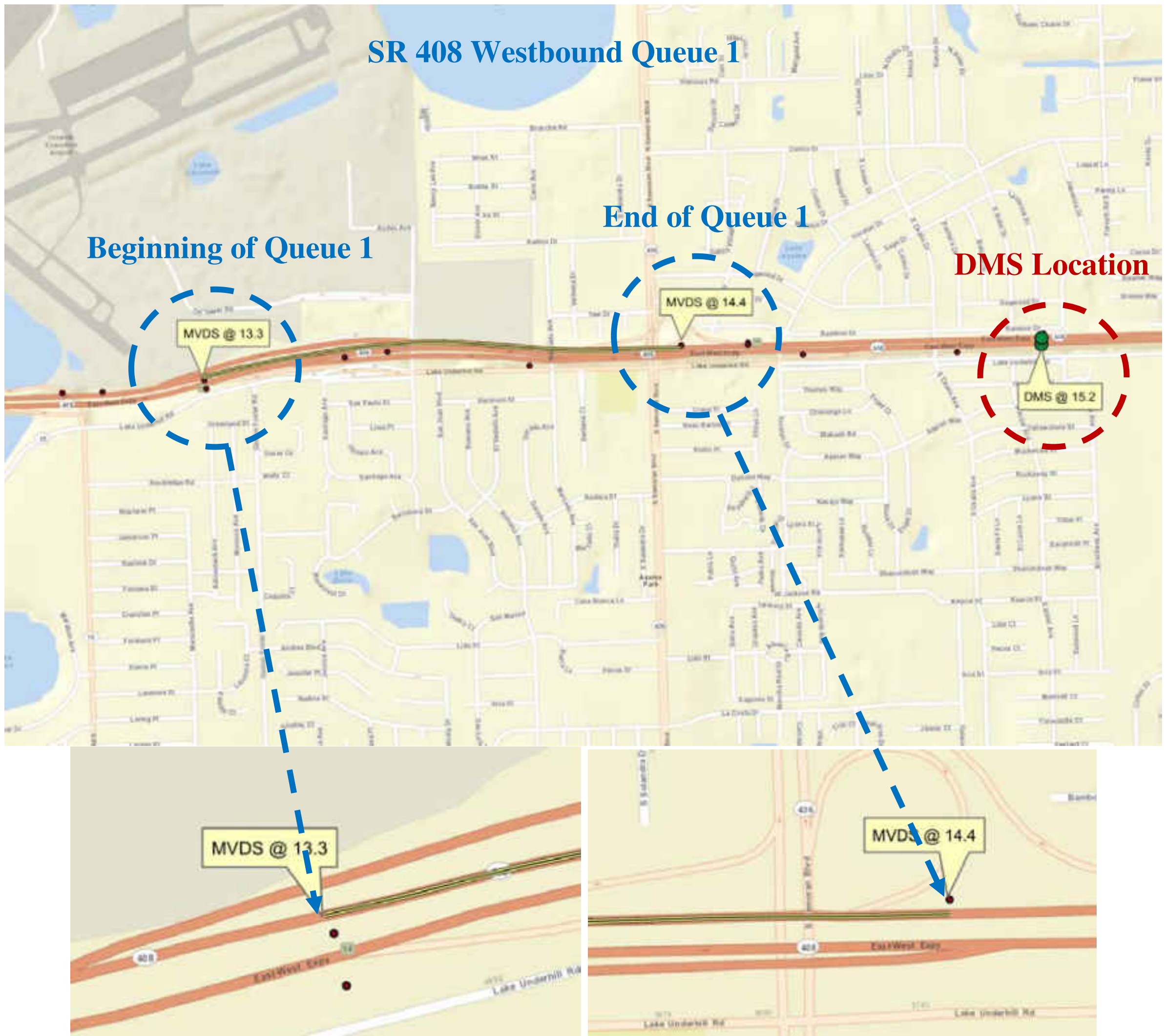


Figure 5-2: SR 408 Westbound Congestion Segment 1 and Upstream DMS Location

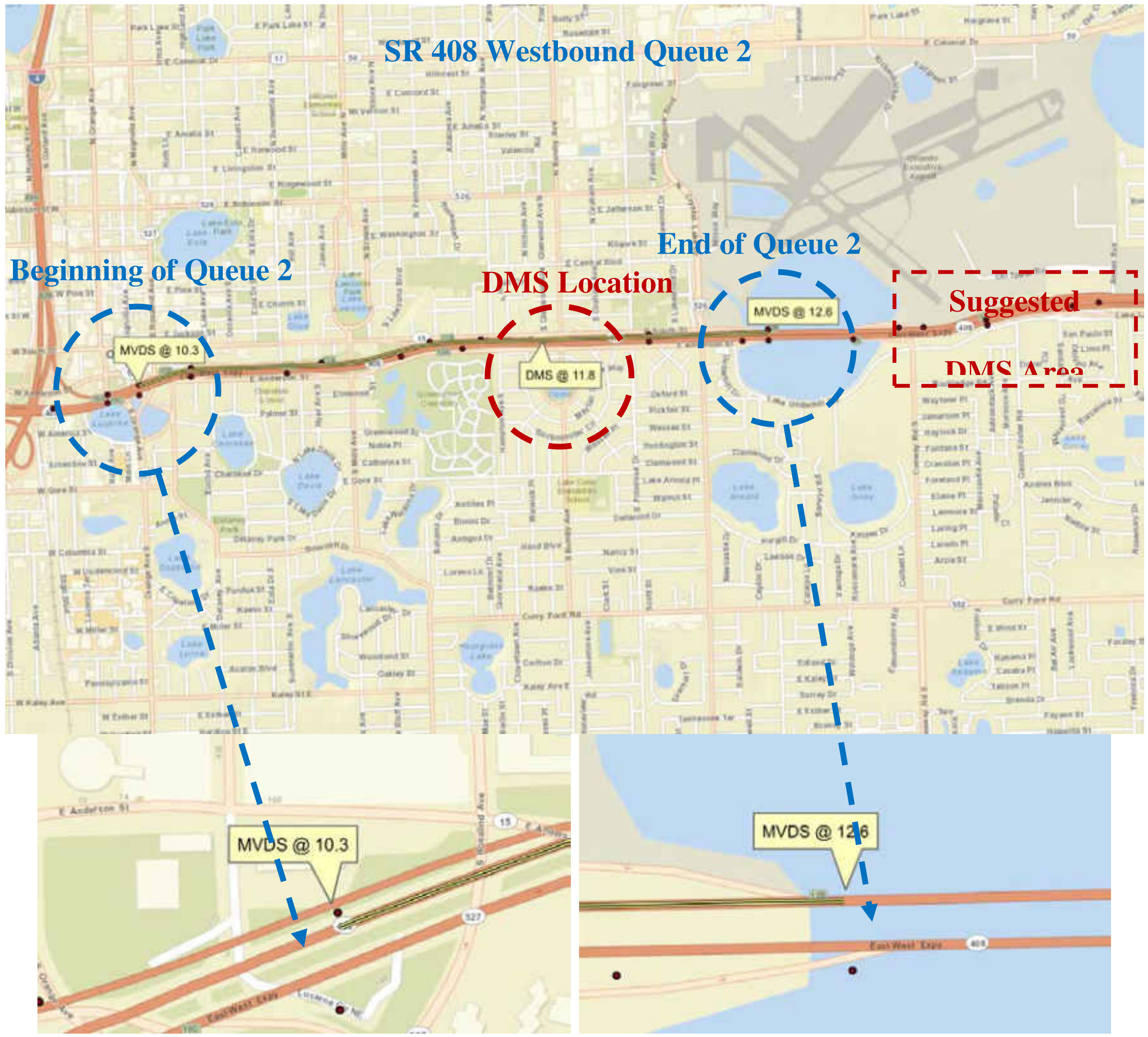


Figure 5-3: SR 408 Westbound Congestion Segment 2 and Suggested DMS Area

SR 417 Northbound experiences congestion on one segment and at one detection location. The congested segment is located between the interchange with East Colonial Dr and the interchange with University Blvd. The congested segment is relatively short and one upstream DMS is located at MP 33.4 which can be used for queue warning. The congested location at 37.7 on SR 417 Northbound is at the boarder of Orange and Seminole County with no upstream DMS nearby. It is suggested that a DMS at MP 35.7 to MP 36.7 might be considered in the future. The detection location at Southbound MP31.9 near interchange with SR 408 is also identified as a congested location. Congestion at this location might be due to the merging traffic from SR 417 Southbound and SR 408 Eastbound. Currently no DMS is located upstream to this congestion location, thus potential DMS can be installed in the segment from MP 32.9 to MP 33.9.

On SR 528 Eastbound has evening congesting lasting for about one hour. The congested segment is near interchange with Tradeport Dr and west to Beachline Airport Mainline Toll Plaza. When vehicles approach the toll plaza, there is only one express lane on the cross-section. And 35 mph speed limit is imposed on this lane. Therefore during the peak hours, queue can build up on this lane. One DMS is to this segment at MP 8.6, about 0.4 mile upstream to the end of the queue. On SR 528 Westbound at the interchange with SR 417, morning congestion is observed. The morning congestion exists during from 7:45 AM to 8:10 AM and from 8:45 AM to 8:55 AM. No existing DMS is located upstream to the SR 528 Westbound congestion area. As a response, to warn the motorists of oncoming congestion, a DMS at MP 12.0 to MP 13.0 can be considered.

Table 5-3 summarizes the congestion segment on each direction of the five expressways. The congested time period is also given. For the DMS application for queue warning, existing locations of DMS signs upstream to the identified congested area have been checked. In case a DMS is found, its milepost is given. If no DMS is available at present, the suggested DMS area is listed for consideration. In total, there are four locations that already have DMS which could be used for queue warning and five locations where future implementation of DMS could be considered for the purpose of congestion management.

Table 5-3: Mainline Congestion Segment and Location Identification and DMS Application for Congestion Management

Expressway	Direction	Congested Segment / Location	Congested Time Period	Existing Upstream DMS	DMS Location / Suggested DMS Area
SR 408	EB	MP 18.4 (Interchange with Dean Rd) → MP 17.7 (Interchange with SR417)	17:30--18:10	Yes	15.2
	WB	MP 13.3 (Close to Conway Rd) → MP 14.4 (Interchange with Semoran Blvd)	07:20--09:00	Yes	15.2
	WB	MP 10.3 (Interchange with I4) → MP 12.6 (Interchange with Crystal Lake Dr)	07:20--09:10	No	13.6--14.6
SR 414	EB	MP 9.3 (Interchange with Orange Blossom Trail)	07:50--08:05	No	7.3--8.3
SR 417	NB	MP 35.5→ MP 35.2 (between East Colonial Dr and University Blvd)	17:20--18:00	Yes	33.4
	NB	MP 37.7 (near boarder between Orange and Seminole County)	17:20--18:00	No	35.7--36.7
	SB	MP 31.9 (south to Interchange with SR408)	17:20--18:00	No	32.9--33.9
SR 528	EB	MP 9.8 → MP 9.0(Interchange with Tradeport Dr, west to Beachline Airport Mainline Toll Plaza)	17:10--18:15	Yes	8.6
	WB	MP 10.3 → MP11.0 (Interchange with Semoran Blvd)	07:45--08:10 08:45--08:55	No	12.0--13.0

In addition to mainline congestion management using DMS, congestion conditions on ramps could also be informed to motorists via DMS. On the expressway system, four ramps are found to experience congestion during the morning or evening peak hours as illustrated in Figure 5-4 to

Figure 5-7. Three ramps are located on SR 408, and the fourth on SR 429. Upstream DMS have been found for each of the ramp. Consequently, they can serve as queue warning signs. On SR 408 Eastbound, the congested ramp is from I-4 Westbound to SR 408 Eastbound located at MP 9.7. During 15:00 to 19:40, congestion exists on this ramp. The DMS at MP 7.7 can be used to warn motorists on the mainline of potential congestion at the merging area near the ramp. On SR 408 Westbound, two off-ramps, one connected with I-4 Eastbound and the other one connected with I-4 Westbound are identified to be congested during in the morning peak hours. When queues are building up on these ramps and affecting mainline traffic, the DMS at MP 11.8 can be considered for queue warning. The congested off-ramp on SR 429 Southbound is at the interchange with Daniels Road. The region nearby is large shopping center and residential area. Congestion is found from 17:00 PM to 18:20 PM. The nearest upstream DMS is located at MP 20.7, and can be used to warn the motorists about the congestion on this ramp in advance. Table 5-4 summarizes the information about ramp congestion and DMS application for queue warning on ramps.

Table 5-4: Ramp Congestion Identification and DMS Application for Congestion Management

Expressway	Direction	Ramp Location	Congested Time Period	Upstream DMS Location
SR 408	EB	MP 9.7 (I4 WB → SR 408 EB)	15:00 -- 19:40	7.7
SR 408	WB	MP 9.9 (SR 408 WB → I-4 WB)	07:25 -- 08:00	11.8
SR 408	WB	MP 10.3 (SR 408 WB → I-4 EB)	07:35 -- 09:05	11.8
SR 429	SB	MP 19.8 (Interchange with Daniels Rd)	17:00 -- 18:20	20.7

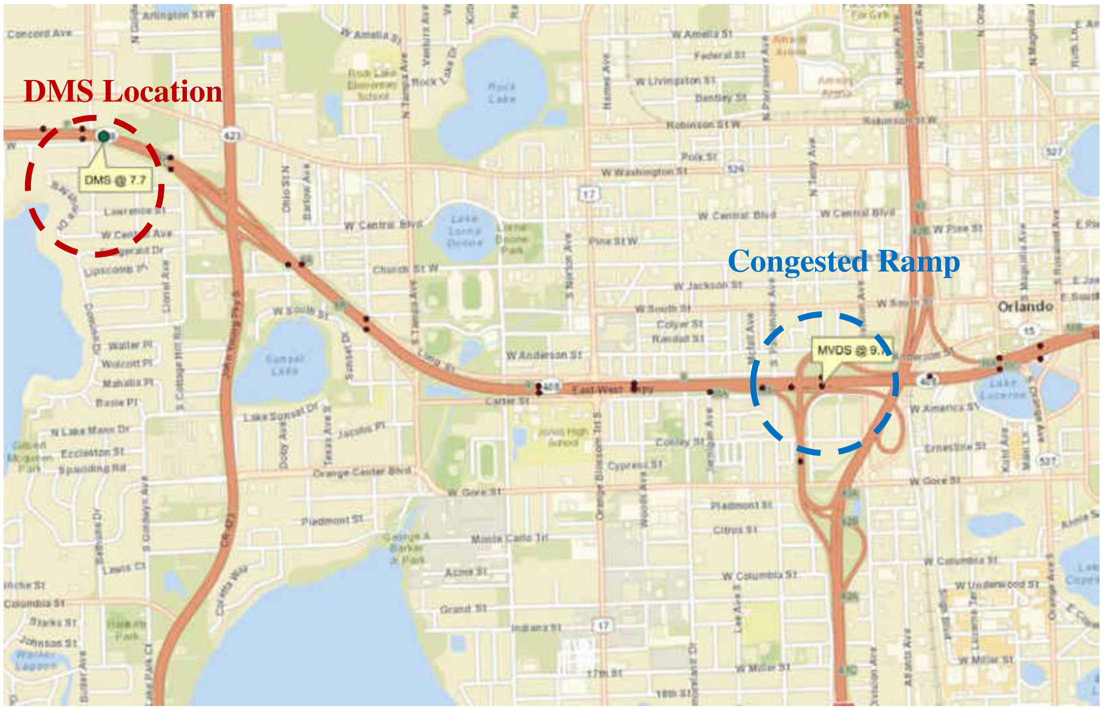


Figure 5-4: SR 408 Eastbound Congested Ramp and Upstream DMS Location

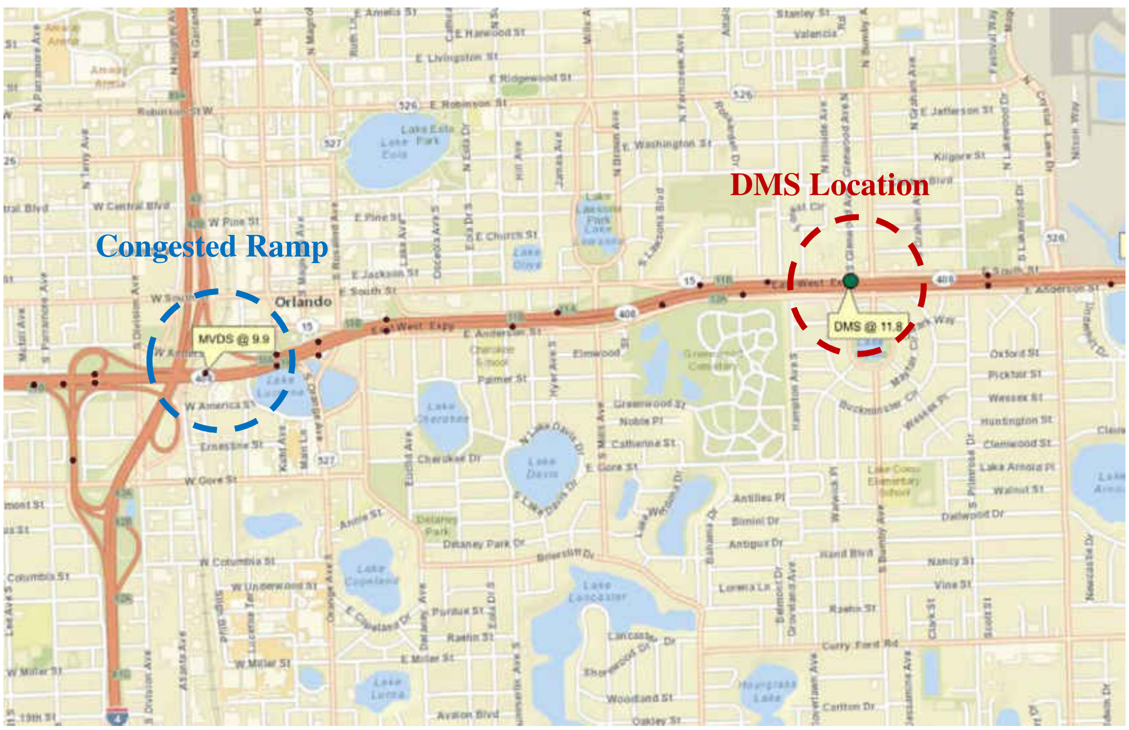


Figure 5-5: SR 408 Westbound Congested Ramp 1 and Upstream DMS Location

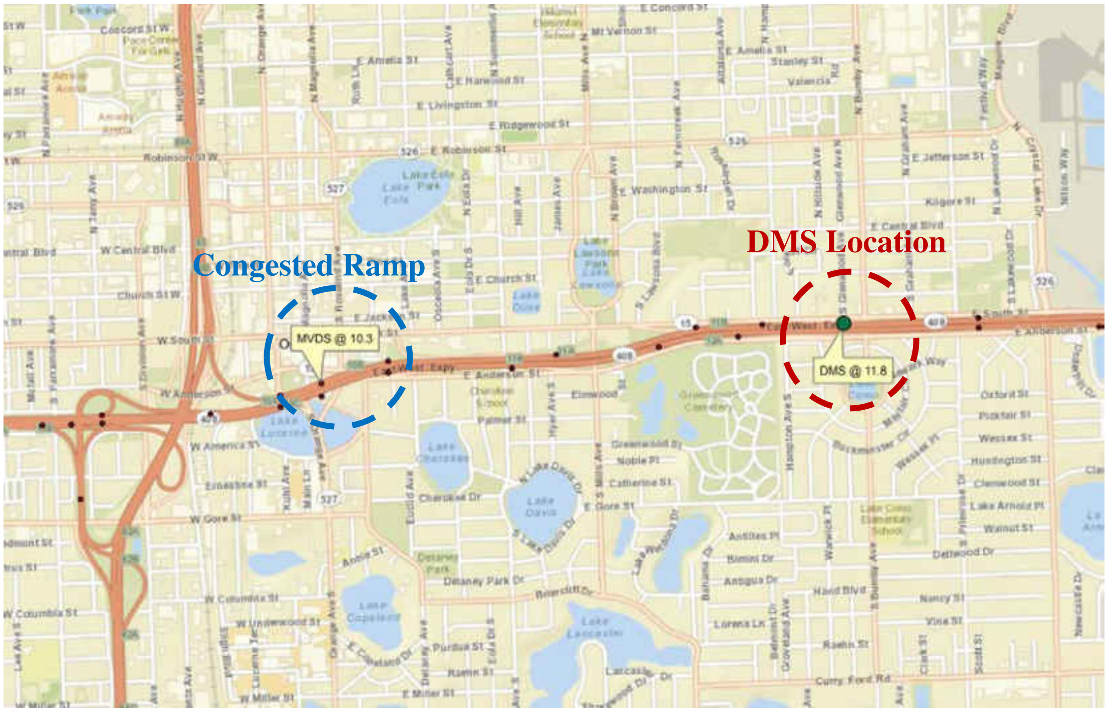


Figure 5-6: SR 408 Westbound Congested Ramp and Upstream DMS Location

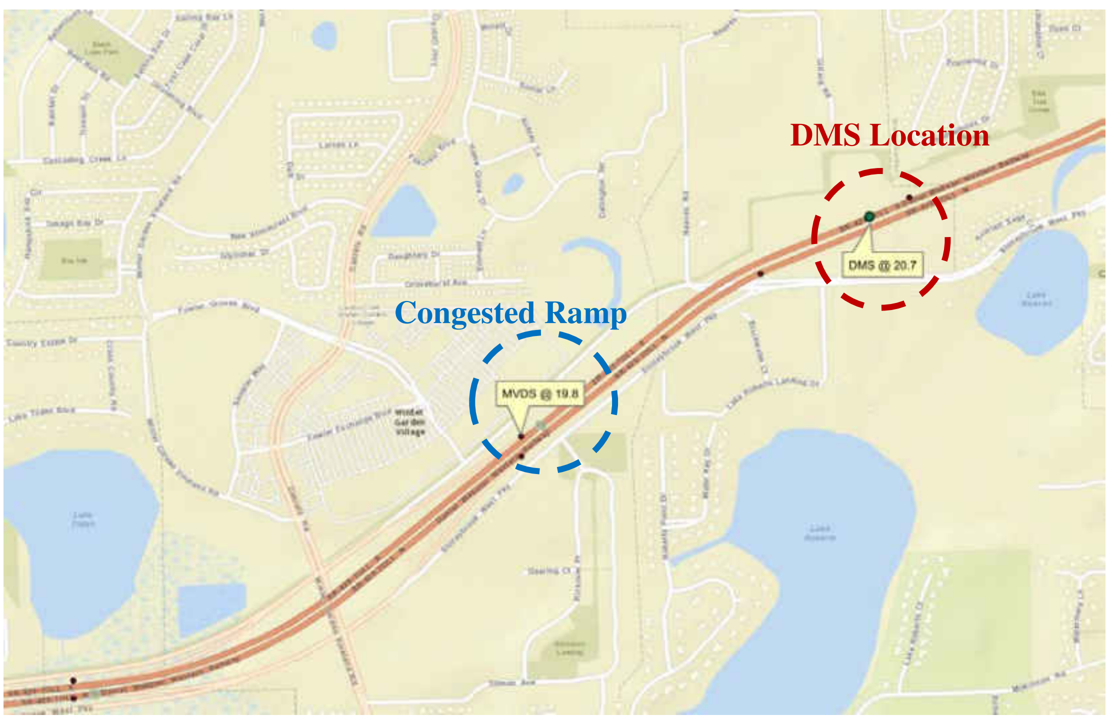


Figure 5-7: SR 429 Southbound Congested Ramp and Upstream DMS Location

CHAPTER 6: APPLICATION OF UNCAPPED AVI DATA IN AGGREGATE SAFETY EVALUATION

6.1 Introduction

Road safety is a major concern of transportation professionals. To gain better insights of crash occurrence mechanisms, extensive research has been conducted at both the macro- and micro-levels. Most existing studies at micro-level can be generalized as aggregate analysis (e.g., crash frequency analysis, crash rate analysis) and disaggregate analysis (e.g., real-time crash prediction). For aggregate analysis, identification of risk factors of total crash frequency is of prime importance before taking crash type and severity into consideration.

This study explored the contributing factors of traffic crashes on the 21-mile urban expressway segment of State Road 408 (SR 408). The toll road extends through downtown Orlando and is heavily used with CFX reporting 125,000 – 135,000 vehicles travel along the expressway daily. Due to large volume of commuter traffic, the expressway of interest has significantly higher volume on weekdays than weekend as illustrated in Figure 6-1 using traffic count by the AVI detector on SR 408 eastbound near the interchange with Interstate-4. Moreover, crash data indicated that around 80% of all crashes occurred on weekdays. The large traffic volume and crash concentration on weekdays necessitate the safety evaluation on the expressway.

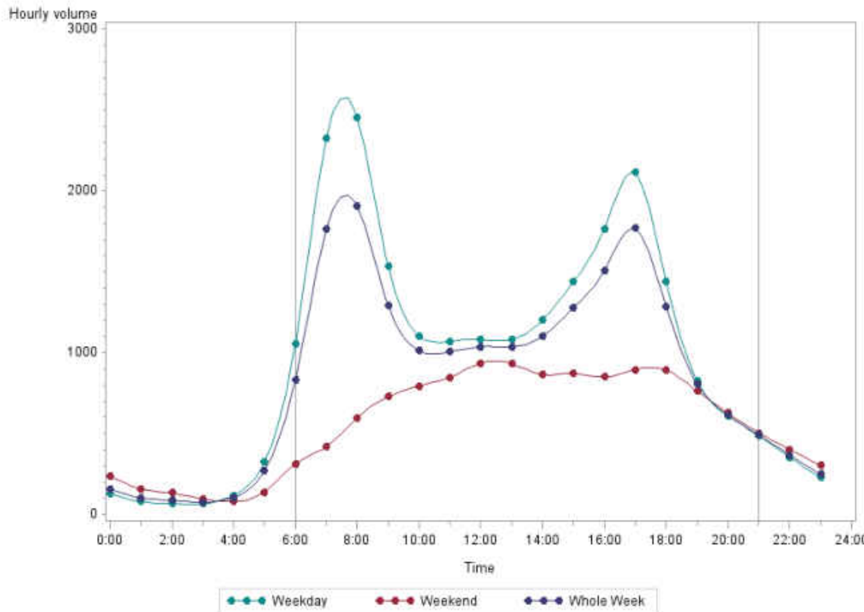


Figure 6-1: Hourly Volume by AVI detector on SR 408 Eastbound near the Interchange with Interstate-4

On the studied expressway, an Automatic Vehicle Identification (AVI) system is installed for Electronic Toll Collection (ETC) and anonymous travel time estimation. Previous studies (Abdel-Aty *et al.,* 2012; Ahmed and Abdel-Aty, 2012) on this road have discussed the possibility of using the capped AVI data for real-time crash predictions. One notable disadvantage of the capped AVI data is that the speed is capped at speed limit during the processing. Therefore the distribution of speed from capped AVI data would be misleading. The use of capped and uncapped AVI traffic data in traffic safety analysis was compared for crashes occurring on weekdays to take into account the traffic characteristics and crash occurrence on SR 408.

While the uncapped AVI data is rarely available in existing studies, its advantages against capped AVI data are obvious. By measuring its performance in safety evaluation, there will be hope that uncapped AVI data could be archived routinely in the future and could extend the usage of AVI system in traffic operation and safety.

In this study, the uncapped AVI traffic data of which speed is not capped were collected for crash frequency analysis along with roadway geometric information. The combination of traffic (43 segments) and roadway geometric characteristics (196 segments) data meant that multiple segments based on geometric characteristics would be nested within one AVI segment, thus exhibiting hierarchical data structure. The data structure indicated that multi-level model framework should be considered for traffic safety evaluation. Two multi-level models, with random parameters at only the traffic data level and at both the traffic and geometry data level were constructed and compared with the classic Negative Binomial model. The three models were evaluated with the Bayesian inference framework to determine the most efficient model in terms of model fitting and identifying the contributing factors to crashes and hence suggesting the corresponding safety countermeasures.

6.2 Data Description and Preparation

Three sets of data were prepared for the crash frequency analysis, 1) one-year capped and uncapped weekday AVI data for SR 408 from September 2012 to August 2013 provided by CFX, 2) crash data on the same roadway for the corresponding time period, 3) roadway geometric characteristics data downloaded from Florida Department of Transportation (FDOT) Roadway Characteristics Inventory (RCI) database. A descriptive statistics summary of variables from the three data sources is provided in Table 6-1.

Table 6-1: Statistic Summary of Variables for the Final Model

Variables	Description	Number	Mean	Std. dev.	Min	Max
Dependent variable						
Crash.ct	Crash counts per segment on weekdays	196	2.08	3.04	0	19
Independent variables						
RCI segment level variables						
Loglength	Logarithmic of segment length	196	-1.65	0.47	-2.60	-0.27
Isldwdth	Inside shoulder width	196	6.94	3.48	0	12
Auxlane	1 if segment has auxiliary lane; 0 if not	196	0.34	0.48	0	1
Hrzdgcrv	Horizontal degree of curvature	196	0.48	0.89	0	5.25
AVI segment level variables						
LogAADT	Logarithmic of directional AADT	43	10.43	0.40	9.18	10.99
Pavi_avgspd	Average speed based on capped AVI data on weekdays	43	57.91	4.26	52.25	64.60
Pavi_stdspd	Standard deviation of speed based on capped AVI data on weekdays	43	3.31	1.79	1.01	8.85
Logvol	Logarithmic of median traffic count per segment based on uncapped AVI data on weekdays	43	9.34	0.53	8.11	10.58
Uavi_avgspd	Average speed based on uncapped AVI data on weekdays	43	64.41	5.14	53.33	64.60
Uavi_stdspd	Standard deviation of speed based on uncapped AVI data on weekdays	43	6.48	2.00	4.13	11.27

Although environmental effects on crash occurrence have been examined in some previous studies (Shankar *et al.,* 1995; Ahmed *et al.,* 2011), weather information was not available in this study. In addition, the urban expressway of interest is located in Orlando, Central Florida area where climate is mild all year round. Thunderstorms or fog, while they can pose threats to drivers, exert their effects only during short time intervals and can be random, making them more appropriate in real-time rather than frequency models. In this crash frequency study, risk factors of traffic condition and roadway geometry are explored.

As far as we know, this is the first time uncapped AVI data has been used for traffic safety analysis. The uncapped AVI data contains the original time stamp of vehicle tag readings. The encrypted transponder (tag) ID of each vehicle, the date and time when the vehicle is detected, the AVI detector ID and the location (milepost) of the detectors are available. Providing data in terms of encrypted transponder IDs protects the privacy of CFX customers. The authors developed an algorithm to calculate the average speed per AVI segment. The encrypted vehicle information was used to filter out the time stamps when the same vehicle was detected at different AVI detectors. Based on the mileposts of two adjacent AVI detectors and the time interval between successive time stamps, the average speed of the vehicle at this segment can be calculated. Extreme speed records were removed according to the definition of abnormal data offered by CFX. As noted, not all vehicles travelling on the toll road are equipped with transponders. Based on the uncapped AVI data, the E-Pass usage is lower than the reported value as shown in Table 6-2. One contributing factor would be that the information of a portion of vehicles was not archived in the first place, for which the reason is unknown. Other possible factors are discussed in Table 6-2.

Nevertheless, the sample size is still sufficient enough to make the speed information from AVI recordings reliable to reflect the real traffic condition on the road. The capped AVI speed distribution indicates that speed limit of 55 and 65 mph are enforced on the expressway segment. However, uncapped AVI speed suggests that in reality the speed is more normally distributed. There are in total 43 AVI segments on SR 408, of which 23 are Eastbound (EB) and 20 are the Westbound (WB).

As stated, SR 408 is a toll road and mostly takes commuter traffic. The time period when the road is heavily used rather than when it is lightly traveled is assumed to uncover the relationship between crash and traffic flow more accurately. Accordingly, average and standard deviation of speed per segment on SR 408 for weekday traffic are calculated.

Crash data on weekdays of the same twelve months (Sep. 2012 – Aug. 2013) on SR 408 were prepared. In Florida, two types of crash report are used by the above agencies; namely long form and short form crash reports. To have the complete crash data, both long and short form crash data were incorporated in this research. To investigate in the contributing factors affecting traffic safety on the mainline of expressway, crashes occurring on the ramps and at manual or coin toll plazas were excluded. In total 500 crashes were identified occurring on the mainline of SR 408 in the one year period, of which 405 crashes occurred on weekdays (81% of total).

Roadway geometric characteristics data were collected from FDOT RCI database. The roadway geometric information collected included the number of lanes, type and width of auxiliary lane, left-/right-shoulder, median, horizontal degree of curvature, speed limit, pavement condition, sectional AADT, average directional factor of AADT (D factor), etc. The roadway is divided into homogeneous segments based on these geometric characteristics. When one of the values of these geometric factors changes, a new RCI segment will be generated. If a RCI segment is shorter than 0.1 mile, it is combined with adjacent RCI segments with higher similarity. The motivation and an example of combining these small segments were addressed in a previous study (Ahmed *et al.*, 2011a). A total of 196 RCI segments for SR 408 were generated (98 segments for each direction). The AADT data stored in the FDOT RCI database is the total AADT for both directions. The directional D factor must be used to derive AADT for each direction. In RCI database, a constant D factor of 52.7% was assigned to represent the percentage of AADT of the bound with higher volume on SR 408. On the other hand, the uncapped AVI data keeps the information at each vehicle level, making it possible to calculate traffic volume during specific time interval. A preliminary comparison between these two types of volume was conducted as shown in Table 6-2.

Table 6-2: Comparison between AADT and Uncapped AVI Volume

Data	AADT	Uncapped AVI volume
Strength	• reflect the total traffic volume • AADT is widely used in traffic analysis	• reflect the traffic volume pattern for each AVI segment • give detailed volume pattern for specific time interval of interest
Weakness	• AADT is the combined traffic for both directions • volume for each direction is assigned by a fixed directional factor • could not reflect volume pattern at more detailed level	• the volume captured by AVI detector is only for the vehicles equipped with toll tags • the true percentage of the AVI volume to the total volume is hard to be determined
Comparison	The volume at each AVI detector accounts for 40 – 60 % of AADT volume for that segment	

The comparison implied that the volume by AVI detector was not equivalent to AADT. The ratio between AVI traffic volume and AADT could only partially reveal the relationship between the tag-equipped traffic and the total traffic since even within the segment of a constant AADT, the traffic volumes vary across locations. The comparison here is intended to show the potential that a surrogate measure of disaggregate volume is worth investigation and could be used as an alternative to aggregate volume indicators such as AADT. To see their performance in safety analysis, both AADT and AVI traffic volume were incorporated in the modeling process. Considering that there is a percent of vehicles that are not detected by the AVI sensors which could lead to lower values of traffic counts, the median value of AVI traffic volumes was chosen as the traffic volume indicator at each AVI detector locations.

When the three data sets were combined together, the data showed significant multi-level structure features. Crashes were aggregated at the RCI segment level. The 196 RCI segments corresponded to 43 AVI segments. As a result, multiple RCI segments were covered by one AVI segment, and shared the same traffic data. Besides, the traffic counts and speed from uncapped AVI data were used in traffic safety analysis for the first time. Evaluation of their performance was needed against the conventionally used AADT and capped AVI speed data.

6.3 Methodology

In this study, multi-level NB regression is implemented. Two types of multi-level models, 1) random parameters at only AVI segment level variables; 2) random parameters at both AVI and RCI segment level variables; and the classic NB model were compared. In addition, since uncapped AVI data was introduced in safety analysis for the first time, comparisons between the use of AADT, capped AVI speed and uncapped AVI speed and volume (AVI transactions) were conducted in type (1) multi-level model framework.

The basic structure of the multi-level model can be shown as:

$$Y_{ij} \sim \mathrm{Negbin}(\lambda_{ij}, r) \tag{6-1}$$

Where Y_i is the crash count at RCI segment i $(i = 1, \ldots, 196)$ belonging to AVI segment j $(j = 1, \ldots, 43)$. λ_{ij} is the expected crash frequency. r is the dispersion parameter.

For multi-level model type (1) with random parameters at only AVI segment level variables, the expectation λ_{ij} for NB distribution could be expressed as

$$\lambda_{ij} = e_{ij}^{\delta}\exp(\alpha_{j[i]} + \mathbf{X}_{\boldsymbol{i}}\boldsymbol{\beta}) \tag{6-2}$$

where e_{ij}^{δ} is the exposure variable. In crash-frequency studies, vehicle miles traveled (VMT) ($\text{VMT} = \text{Traffic volume} \times \text{Segment length}$) is often used as an exposure factor. In case of this study, the segment length was RCI level exposure variable and traffic volume AVI level variable. To account for their effect properly, they were used separately at the two different levels.

At RCI segment level:

$$\log(\lambda_i) = \delta_l \mathrm{Log}length_i + \alpha_{j[i]} + \mathbf{X}_{\boldsymbol{i}}\boldsymbol{\beta} \tag{6-3}$$

At AVI segment level:

$$\alpha_j = \delta_v \mathrm{Log}volume_j + \mathbf{U}_{\boldsymbol{j}}\boldsymbol{\gamma} \tag{6-4}$$

$\mathbf{X}_{\boldsymbol{i}}$ stands for roadway geometric related risk factors at the RCI segment level, $\mathbf{U}_{\boldsymbol{j}}$ is travel flow related risk factors at AVI segment j $(j = 1, \ldots, 43)$. $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of regression coefficients at the RCI segment level and AVI segment level, respectively. δ_l and δ_v stand for coefficients of exposure variables Loglength and Logvolume.

For multi-level model type (2) with random parameters at both AVI and RCI segment level variables, the model specification should be updated accordingly,

$$\lambda_{ij} = e_{ij}^{\delta}\exp(\alpha_{j[i]} + \mathbf{X}_{\boldsymbol{i}}\boldsymbol{\beta}) \tag{6-5}$$

And at RCI segment level:

$$\log(\lambda_i) = \delta_{l_i}\text{log}length_i + \alpha_{j[i]} + \mathbf{X_i\beta} \tag{6-6}$$

At AVI segment level:

$$\alpha_j = \delta_v \text{log}volume_j + \mathbf{U_j\gamma} \tag{6-7}$$

δ_{l_i} is the RCI segment level coefficient of Loglength that varies across each observation. Other parameters have the same meaning as in the model type (1). For $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, δ_l and δ_v non-informative normal priors with mean at 0 and variance at 1000 are set. α_j follows a normal distribution with mean determined by Equation (6-4) and variance $1/\sigma^2$, $\sigma \sim uniform(0{,}100)$ as suggested by Gelman (2007). δ_{l_i} is also set to follow normal distribution $N(\mu[i]{,}1/\sigma^2)$, $\mu[i] \sim N(0{,}1000)$ and $\sigma \sim uniform(0{,}100)$.

Full Bayesian inference using Markov Chain Monte Carlo (MCMC) algorithm was implemented to construct the multi-level models. The Bayesian model's structure is inherently hierarchical. The prior distribution $f(\theta|a)$ of the model parameter θ with prior parameters a can be considered one level of hierarchy, with the likelihood $f(y|\theta)$ as the final stage of a Bayesian model. The posterior distribution is $f(\theta|y) \propto f(y|\theta)f(\theta|a)$ via the Bayes theorem (Ntzoufras 2011). The model calibrations were done in WinBUGS software. Three chains of 20,000 iterations were run for each model, of which 5,000 iterations were fed as burn-in.

The deviance information criteria (DIC) introduced by Spiegelhalter *et al.* (2003a) was used as a Bayesian measure of model complexity and fit. DIC can be viewed as the combination of measure of model fitting ($\overline{\mathrm{D}}$) and penalization of number of effective variables (p_D) in the model. When the difference in DIC between two models is more than 10, it is assured that the model with smaller DIC is better. The difference of DIC between 5 and 10 is considered substantial. Otherwise it is hard to make a definitive judgment (Spiegelhalter *et al.* 2003a). Bayesian Credible Interval (BCI) was used for variable estimation; and the 95% BCI was measured to test the significance of variables. If the BCI does not contain 0, then the effect of variable is significant.

6.4 Modeling Results and Discussion

6.4.1 Capped AVI Data vs. Uncapped AVI Data

Traffic variables have been confirmed as important risk factors for crash occurrence. AADT, capped AVI data, and uncapped AVI data were all available in this study. Concerning these traffic variables, several data issues arise: 1) directional AADT is calculated based on segment AADT for both directions and a constant D factor; 2) traffic counts based on uncapped AVI data can reflect the traffic pattern during the time span of interest (weekday) but only the vehicles with tags are captured; 3) speed derived from capped AVI data is capped at the speed limit while uncapped AVI data does not suffer from this problem. As a result, whether the adoption of uncapped AVI data will improve the estimation and our understanding of the risk factors of crashes and generate better model fitting was evaluated.

Three combinations of traffic data, 1) AADT and capped AVI speed, 2) AADT and uncapped AVI speed, 3) uncapped AVI traffic counts and speed were used and compared. The three combinations were evaluated in the multi-level framework in which only AVI segment level traffic variables were treated with random parameters. The parameter estimation and modeling results are shown in Table 6-3. Variables at the AVI segment level and RCI segment level were estimated simultaneously.

Table 6-3: Parameter Estimation and Model Fitting for Different Traffic Data

Multi-level model with random parameter for AVI segment level variables						
Variable	AADT and capped AVI speed		AADT and uncapped AVI speed		Uncapped AVI	
	Mean	95% BCI	Mean	95% BCI	Mean	95% BCI
AVI segment level variables						
LogAADT	0.001	(-0.317, 0.326)	0.131	(-0.205, 0.453)	-	-
Logvol	-	-	-	-	0.614	(0.297, 0.932)
Pavi_avgspd	0.038	(-0.023, 0.098)	-	-	-	-
Pavi_stdspd	-0.021	(-0.210, 0.162)	-	-	-	-
Uavi_avgspd	-	-	0.023	(-0.036, 0.084)	-0.067	(-0.113, -0.024)
Uavi_stdspd	-	-	-0.114	(-0.312, 0.080)	0.147	(0.024, 0.285)
RCI segment level variables						
Loglength	0.966	(0.602, 1.295)	0.959	(0.582, 1.317)	0.917	(0.600, 1.241)
Isldwdth	-0.029	(-0.110, 0.047)	-0.030	(-0.104, 0.038)	-0.064	(-0.134, 0.007)*
Hrzdgcrv	-0.284	(-0.505, -0.077)	-0.282	(-0.480, -0.084)	-0.281	(-0.499, -0.095)
Auxlane	0.375	(0.042, 0.719)	0.339	(0.004, 0.704)	0.265	(-0.049, 0.589)*
Model Evaluation						
$\overline{D}$	655.565		654.424		650.265	
p_D	36.160		36.876		32.421	
DIC	691.725		691.300		682.686	

* variable insignificant at 95% BCI but significant at 90% BCI

The effects of RCI segment variables of the three models agreed with each other. However, the significance of the candidate variables was different. The larger the logarithm of segment length, the more crashes would occur on the segment. This explained the impact of exposure variable on crashes. RCI segments with wider inside shoulder width were associated with fewer crashes. On the urban expressway, wider inside shoulder could allow drivers on the inner lane some leeway for driving errors. The existence of auxiliary lanes increased the chance of crashes on through travel lanes. The effect of auxiliary lanes as illustrated in Figure 6-2 has not been adequately discussed before. Auxiliary lanes are required to provide turning movements and speed changes at critical locations such as ramps. On the segment with auxiliary lanes, more lane-changing behaviors are expected in close proximity; speed difference might also be significant among through travel lanes. The turbulence of traffic flow on mainline segment with auxiliary lanes could then result in more crashes. The effect of horizontal degree of curvature should be reviewed carefully. Previous research has found mixture effect of horizontal curves. Hauer (1999), Abdel-Aty and Radwan (2000), Bonneson *et al.* (2007) suggested larger curvature would increase the crash likelihood. On the other hand, Milton and Mannering (1998), Haynes et al. (*2007*), Ahmed et al. (2011a) found negative associations between horizontal curves and crash frequencies. It should be interpreted as that curvature might be risky considering its engineering effect, but drivers might drive more slowly and cautiously on curves. In case of this study, the horizontal degree of curvature was negatively related with crash frequency. On the one hand, the maximum horizontal degree of curvature on the expressway is 5.25° per 100 feet, which is not too sharp. On the other hand, commuter traffic takes large portion of the total traffic, indicating the drivers are familiar with the roadway geometry. As a consequence, drivers are more likely to adapt their behavior according to the road geometric characteristics and also be more careful when driving on curves. The combined effects of road geometric features and driver behavior could result in the negative effect of horizontal curvature on crash occurrences.

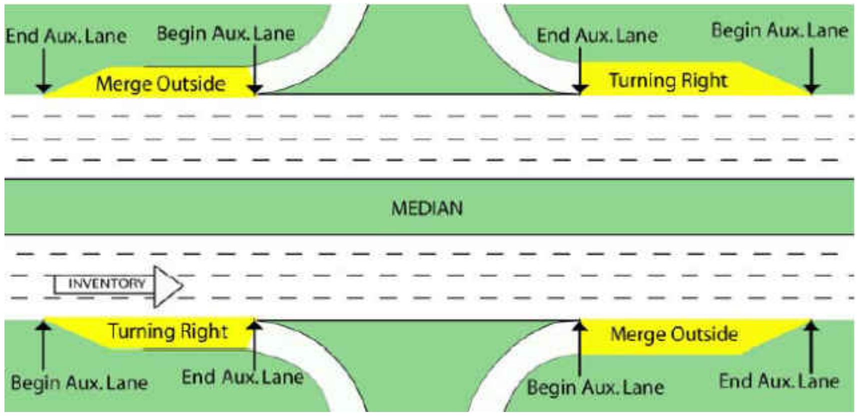


Figure 6-2: Auxiliary Lanes on Expressway from RCI Handbook

The effects of AVI segment level variables were not significant in the first two models incorporating AADT and capped AVI speed information. This could be due to the data issues that were mentioned at the beginning of this section. A constant D factor lacks the capability of determining the direction with higher traffic volume for each segment and assigning the volume for each direction. Censored speed from capped AVI data may not reflect the true effect of speed on crash occurrence. Traffic counts and speed (average and standard deviation) from uncapped AVI data were proved significant for weekday crashes. Their effects were easy to interpret; higher volume, congestion (lower traveling speed) and turbulence (large standard deviation of speed) would increase the crash likelihood. The goodness-of-fit of the three models were also compared. Models involving AADT and capped AVI speed data had relatively high DIC values (691.725 and 691.300, respectively). The model with traffic counts and speed from uncapped AVI data had the lowest DIC value (682.686). The difference of DIC was more than 5, proving that the use of uncapped AVI data was superior in describing crash mechanisms for the urban expressway segment in this study.

6.4.2 Negative Binomial model vs. multi-level models comparison

Multi-level model with random parameters for AVI segment level traffic variables were used to compare the performance of different traffic data in the section above. While the modeling approach accommodated the hierarchical data structure, its performance against classic NB model and multi-level model with random parameters for both AVI and RCI segment level variables were of interest to investigate. In some of the previous random-parameters studies, the parameters vary across observations (Anastasopoulos and Mannering, 2009; Dinu and Veeraragavan, 2011; Venkataraman *et al.,* 2013). In El-Basyouny and Sayed's (2009) study, 392 urban road segments belong to 58 corridors and the parameters vary across corridor groups. In this section, three models, namely the NB model and two multi-level models were compared using uncapped AVI data. The segment of SR 408 in this study was divided into 43 AVI segments and 196 RCI segments. However, the segmentation of the RCI segments was not directly related to the AVI segments. The assumption that RCI segments within one AVI segment share higher similarity while being distinct between different AVI segments could not hold. Besides accounting for AVI segment level effects, unobserved heterogeneity across RCI segments is worth consideration by introducing random parameters at this level. These issues are the motivation for the second multi-level model. Table 6-4 summarizes the modeling results, in which crashes occurred on weekdays.

Table 6-4: Parameter Estimation and Model Comparison for NB and Multi-level Models

Crash frequency on weekdays						
Variable	NB model		Multi-level models			
			Fixed Parameters		Random parameters at RCI segment level	
	mean	95% BCI	mean	95% BCI	mean	95% BCI
AVI segment level variables						
Logvol	0.674	(0.460, 0.894)	0.614	(0.297, 0.932)	0.633	(0.302, 0.984)
Uavi_avgspd	-0.076	(-0.105, -0.047)	-0.067	(-0.113, -0.024)	-0.070	(-0.117, -0.025)
Uavi_stdspd	0.118	(0.021, 0.218)	0.147	(0.024, 0.285)	0.157	(0.018, 0.290)
RCI segment level variables						
Loglength	0.744	(0.372, 1.092)	0.917	(0.600, 1.241)	0.998	(0.721, 1.309)
SD of Loglength distribution	-	-	-	-	0.333	(0.165, 0.489)
Isldwdth	-0.029	(-0.084, 0.026)	-0.064	(-0.134, 0.007)*	-0.068	(-0,141, -0.002)
Hrzdgcrv	-0.225	(-0.449, 0.007)*	-0.281	(-0.499, -0.095)	-0.296	(-0.504, -0.091)
Auxlane	0.087	(-0.281, 0.437)	0.265	(-0.049, 0.589)*	0.320	(-0.030, 0.687)*
Model evaluation						
$\overline{D}$	719.649		650.265		583.312	
p_D	7.898		32.421		78.945	
DIC	727.547		682.686		662.257	

* variable insignificant at 95% BCI but significant at 90% BCI

The parameter estimations of the traffic and road characteristics variables were consistent for the three models. However, the multi-level model identified significant geometric factors while NB model could not. It can be seen that when the coefficient of Loglength varies across observations, the effects of other variables become more significant. When random parameters existed at both AVI and RCI segment levels, the effect of inside shoulder width was significant at 95% BCI rather than 90% BCI with random parameters only at AVI level. The result confirmed that by taking account of the heterogeneity across RCI segments, improvement in parameter could be further achieved. Comparing the DIC values of the models, the multi-level models showed better fitting than NB model. By assigning random parameters to Loglength, the multi-level model further improved. Nevertheless, assigning variables with random parameters also caused increase of the effective number of variables incorporated in the model. Consequently, when modeling crash occurrence with hierarchical structured data, multi-level is more suitable than the NB model but the tradeoff of model complexity should be considered at the same time.

6.5 Summary and Conclusions

The traffic detection systems are the foundation of the Intelligent Transportation System (ITS). Development of traffic detection technology also brings new and more precise traffic data that benefits traffic safety analysis. Previous studies by Abdel-Aty *et al.* (2012), Ahmed and Abdel-Aty (2012) have used the speed data generated from Automatic Vehicle Detection (AVI) systems for real-time crash prediction on the same road. Constrained by the availability of data, the speed data in these studies were capped at 1-min interval and capped at the speed limits. In this study, uncapped AVI data was introduced, which archives the original time stamps of the transponder readings and the speed is not capped. The number of vehicles equipped with tags passing an AVI detector can also be calculated as an indicator of traffic volume.

How well the uncapped AVI data performs in traffic safety analysis has never been evaluated before. Comparisons between AADT and AVI traffic counts, speed data from capped AVI and uncapped AVI data were conducted using multi-level crash frequency models on SR 408. The models were evaluated for crashes occurring on weekdays. With some issues with capped AVI speed data pointed out, the modeling results confirmed that it was more appropriate to use uncapped AVI speed data. The strength and weakness of AADT and AVI traffic volume were listed and their performances in safety analysis were tested in the models. The results showed promising use of disaggregate volume as an alternative to aggregate volume indicators such as

AADT. At current stage, uncapped AVI data are seldom archived by transportation authorities. However, the results from this study encourage the uncapped AVI data to be stored and their use to be extended beyond ETC and travel time estimation.

When incorporating the roadway geometric information along with traffic data in crash frequency models, it is necessary to select the models based on the data structure. In case of this study, multi-level models approach was proved better. While the multi-level models were more complex than NB model to estimate, they offered significant improvement in model fitting. In addition, by estimating the variables at different levels separately, the multi-level model provided more accurate parameter estimation.

Higher volume, lower speed and larger variation of speed were identified as risk factors of crash occurrence. For roadway geometric characteristics, the longer segment length, auxiliary lanes were positively associated with crash occurrence. Horizontal degree of curvature and wider inside shoulder width were significant with a negative sign, indicating that sharper curves and wider inside shoulder width reduced crash likelihood. Previous studies have found mixture effects of horizontal degree of curvature on crashes, which should be understood from both engineering effects and human behavior. For the expressway segment in this study, the sharpest degree of curvature is 5.25° per 100 feet which is not too sharp. Commuter traffic indicates that drivers are likely to be familiar with the roadway geometry and prepared for the curves. The effect of auxiliary lanes on mainline crash occurrence was generally overlooked. However, relatively small distance between ramps on urban expressways makes segments with auxiliary lanes deserve careful safety investigation. Lane-changing and speed changes at these segments can greatly affect the possibility of driving errors.

In light of opportunities to further improve safety on urban expressway, countermeasures should be taken from aspects of both the traffic and roadway geometric characteristics. SR 408 has Dynamic Message Signs (DMS) installed to inform travelers of the estimated time to a specific point. In addition to travel time estimation, the DMS could be used to display traffic warnings to alert drivers of traffic conditions at the downstream segment. On segment with auxiliary lanes, more lane-changing behaviors and speed changes are expected. Close attention should be paid to the geometric design of segments auxiliary lanes, providing markings and signs to guide the drivers and forgive potential driving errors.

CHAPTER 7: AGGREGATE ANALYSIS OF CONGESTION'S IMPACT ON URBAN EXPRESSWAY SAFETY

7.1 Introduction

Providing motorists with efficient and safe traffic system has long been considered a priority of traffic professionals. With the growth in traffic demand outpacing construction of road infrastructure, congestion and safety concerns arise. In urban areas, many traffic authorities have turned to toll/turnpike facilities and efficient use of Intelligent Transportation Systems (ITS) techniques as remedies for congestion and to improve safety. Whether ameliorating congestion or safety would have positive effects on the other depends on the relationship between congestion and safety. It is widely acknowledged that crashes lead to non-recurrent congestion while the effects of congestion on crashes seem inconclusive.

This study aims at identifying the relationship between congestion and crashes for urban expressway State Road (SR) 408 in Central Florida based on the existing ITS facilities. The 21.4-mile expressway of interest is managed by Central Florida Expressway Authority (CFX). The toll expressway accommodates large commuter traffic and travels through the downtown Orlando area. Consequently, recurrent congestion during peak hours is observed on the expressway. For more efficient operation, CFX has installed Automatic Vehicle Identification (AVI) system for Electronic Toll Collection (ETC) and travel time estimation. In 2013, Microwave Vehicle Detection System (MVDS) was introduced for traffic monitoring on the expressways. Based on the real-time data derived from the detection systems, CFX provides motorists with travel time and other incident via the Dynamic Message Signs (DMS). The multiple ITS instruments deployed on the expressway give the authority a favorable margin for better understanding of the traffic conditions and potential for improvement. In this study, traffic parameters from the MVDS and AVI systems along with roadway geometric characteristics were collected to identify their relationship with crash occurrence. Compared with previous crash frequency studies exploring congestion's effects on safety, the current work applied ridge regression within Bayesian framework to avoid the problem of multicollinearity. Through this method, it is expected that more accurate parameter estimates will be achieved.

7.2 Statement of Problem

Traffic congestion especially in urban areas has been receiving increasing attention during the past few decades. Congestion is not only costly in terms of time losses and fuel waste, but also causes unstable traffic flow because of the stop-and-go traffic pattern. Whether congestion has an impact on traffic safety has been researched in many previous studies. One early study by Shefer and Rietveld (1997) stated that congestion leads to lower numbers of fatalities because of lower speed in congestion. Baruya (1998) investigated crash frequencies in four European countries and declared that congestion is positively associated with crash frequency. Golob and Recker (2003b) in their study of Southern California freeways found rear-end crashes are generally associated with high variations in relatively low speeds which are commonly observed in congested traffic. Kononov et al. (2008) developed safety performance functions (SPFs) using Colorado, California, and Texas data. Their conclusion suggested that safety deteriorates with the degradation of congestion expressed through level of service. Wang et al. (2009) explored the relationship between crash and congestion for an orbital motorway around London using a spatial analysis approach. They found that congestion imposes little or no impact on the frequency of road accidents. In a recent review of previous works about the congestion's effect on road safety, Wang et al. (2013c) concluded that existing research has not yet reached an agreement on the impact of traffic congestion. To understand the complex phenomenon, several factors should be taken into account; 1) how is congestion measured. With the development of traffic detection systems, new congestion measurements become available. Whether these detection systems reflect the congestion accurately is crucial in the congestion-safety studies. 2) Congestion could be time specific in urban area. Freeways/expressways in urban areas carry large commuter traffic in morning and evening peak hours. The necessity to consider peak and non-peak traffic hours is worth investigation. 3) The complex nature of crashes often requires multiple explanatory variables included in safety analysis. The traffic speed, volume, and density could be correlated which could introduce multicollinearity. Multicollinearity could affect accurate interpretation of the variables' effects (Berry and Feldman, 1985). Yet existing studies haven't addressed the problem adequately. The conflicting results in previous congestion-safety studies might be resulting from this issue.

7.2.1 Congestion Performance Measures

Congestion is easy to be observed and interpreted. Nevertheless, how to measure congestion could be a challenging task. Whether congestion estimates are accurate and useful largely depends on the available data. For a long time, Volume-to-Capacity Ratio (V/C Ratio) and the Level of Service (LOS) have been the primary congestion indicators since continuous data collection was not feasible (National Research Council, 2010). However, they lack the flexibility to capture the dynamics of congestion. With the development of ITS equipment, more direct measurement of speed, travel time, and volume becomes available. New approaches to measure congestion are proposed (Lomax *et al.,* 1997; Schrank and Lomax, 2007). Current congestion measures can be broken down into three general categories: density-based, travel-time-based and travel-speed-based. Traffic occupancy has become more available to reflect the traffic density on roadways thanks to the deployment of ITS facilities. The numerous types of ITS traffic detectors also promote the creation of other congestion measures. Travel Time Index (TTI) is a widely used travel-time-based congestion measure. It is defined as the ratio of the travel time during the peak period to the time required to make the same trip at free-flow speeds (Schrank and Lomax, 2007). TTI can be measured by the AVI system which identifies vehicles at different detectors and calculates the travel time. In the meantime, speed-based congestion measures are also applied. Christidis and Rivas (2012) in their report of European road congestion defined the intensity of congestion as the ratio of the average to free flow speeds. Hossain and Muromachi (2012) used the rate of reduction in speed caused by congestion from the free-flow speed condition as congestion index. In this study, the expressway has AVI and MVDS systems. Multiple congestion measures were tested to see their effects on traffic safety.

7.2.2 Time-specific Nature of Congestion

Another issue related to the congestion in crash frequency studies is the problem of averaging. Mensah and Hauer (1998) discussed the potential problems resulting from averaging traffic flows over a period of time. They argued that the cause-effect relationship is between crashes and the flows prevailing near the time of crash occurrence and there are two (daytime and nighttime) and perhaps many more cause-effect functions in the course of the study period. Wang et al. (2009) in their study suggested that in reality, the level of congestion varies over time and averaging may have an impact on the effect of congestion on crashes. In this paper, as it is still a crash frequency study, aggregating the traffic flow data is unavoidable. We examined whether it is necessary to take into account the peak and non-peak hours when exploring the congestion's effect on traffic safety.

7.2.3 Multicollinearity

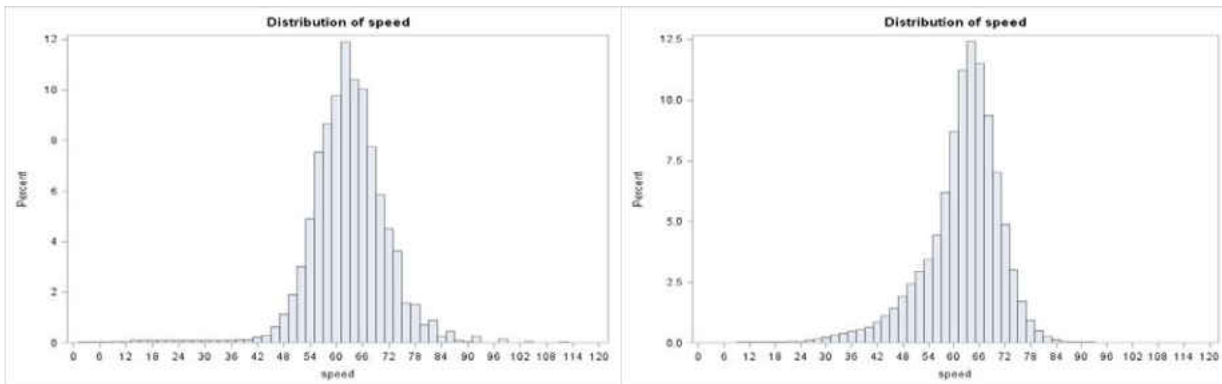
Multicollinearity occurs when the explanatory variables are correlated among each other. The inherent collinearity won't affect the overall model fitting (Burnham and Anderson 2002); however, it can result in parameter estimates with incorrect signs and implausible magnitudes (Mela and Kopalle 2002). In crash frequency studies, the traffic data are observational data in most cases and not controlled. Correlations exist among speed, density and volume. Nevertheless, existing literature shows that this issue is not adequately addressed. This might be a reason why researchers haven't reached a conclusive statement about the congestion's effect on traffic safety. Solutions to the multicollinearity problem were generalized by Congdon (2007) which include: 1) the introduction of extra information; 2) the multivariate reduction to a smaller set of uncorrelated predictors; 3) ridge regression. In this study, both roadway geometric characteristics and traffic flow parameters are incorporated and no extra information is available at this stage. Multicollinearity is present mainly among the traffic flow parameters. Despite that the focus of this research is congestion's effect, we still kept volume as it is an important exposure variable.

As a result, ridge regression is adopted to deal with the multicollinearity issue and to provide accurate parameter estimation.

7.3 Data Preparation

Three types of data were prepared to uncover the effects of congestion on traffic safety: 1) traffic flow data from AVI and MVDS systems on the expressway provided by CFX; 2) roadway geometric characteristics data downloaded from Florida Department of Transportation (FDOT) Roadway Characteristics Inventory (RCI) database; 3) crash data on the expressway of interest from Signal Four Analytics database.

The three congestion measures used in congestion measurement were also introduced in the aggregate safety analysis to determine whether they could reach consistent conclusions on the effects of congestion on traffic safety. To prepare the Travel Time Index, Congestion Index and occupancy information, both AVI and MVDS data were collected. There were in total sixteen months of raw (uncapped) AVI traffic data collected for this research from September 2012 to December 2013. The MVDS traffic data were collected since July 2013 and at the time of conducting this research, six months data were available. A preliminary comparison was conducted to check how comparable the six months MVDS data is to the sixteen-month AVI data. Figure 7-1 shows that the speed distributions are similar for the two types of data. MVDS average daily traffic volumes in the six months were compared with expressway AADT for 2012 and no significant changes in traffic volume were found in the two years. Therefore we believe that the six-month MVDS traffic data could reflect the traffic conditions of the studied time period as well.



(a) (b)

Figure 7-1: (a) AVI Speed Distribution; (b) MVDS Speed Distribution

Consequently, traffic data were collected from the two ITS detection systems. Three congestion measures (i.e., TTI, occupancy, CI) were developed and fed to the crash frequency models. The congestion measures were aggregated at 5-minute interval each day and then averaged during the study time period. Speed information from the two sources was also prepared. As for the traffic volume, only the MVDS daily traffic volumes were used. Although AVI detectors can capture vehicles equipped with tags, they are not able to report the total volume on the expressway. As an alternate measure of traffic volume for the AVI segments, the average volume at the MVDS detection points within an AVI is calculated. In addition, allowing for the peak-hour traffic effects, the traffic data were aggregated into two groups: peak hours (7:00 – 9:00 and 17:00 – 19:00) and non-peak hours (the rest time period during 6:00 – 21:00). The time period of 21:00 – 6:00 was excluded considering the light traffic during this period. It is more likely that crashes occurred during the discarded time period are affected by other factors rather than traffic flow parameters.

FDOT RCI online database stores comprehensive roadway characteristics information. When multiple geometric characteristic variables are selected, the roadway will be automatically divided into homogeneous segments by the system. The segments are specified as RCI segments. To avoid the segments with length too short for analysis, it was decided in this study that the smallest segment length is 0.1 mile. If a RCI segment length is below this threshold, it will be combined with adjacent segment with higher geometric similarity. The motivation and an example of combining these small segments were addressed in a previous study (Ahmed *et al.,* 2011a). The roadway geometric information included the number of lanes, existence of auxiliary lanes, horizontal degree of curvature, speed limit, etc. In total, there are 75 segments generated for EB and 76 segments for WB.

The crashes in Florida are archived in two formats, namely short and long form crash reports. Short form crash reports are created to record less severe crashes (i.e., property damage only crashes) while long form crash reports are for more severe crashes (i.e., injury and fatality involved crashes). The current work includes both types to guarantee that complete crash data are collected. Given the sixteen months traffic data available, the crash data during the same period was collected. To reveal the relationship between safety and congestion, two factors were considered: 1) the data should reflect traffic conditions for the days when recurrent congestion occurs; 2) the crashes should be more likely to be influenced by traffic flow. Taking the two points into account, we retained the mainline crashes occurring from 6:00 to 21:00 on weekdays for further analysis. Based on these criteria, 472 crashes were identified.

After the three data sets were prepared, they were merged together. The crashes were aggregated into the RCI segments on which they took place during peak and non-peak hours. The total number of observations is $(75 + 76) \times 2 = 302$. Then traffic data were assigned to the RCI segments. Since the AVI data reflect traffic conditions for segments on the expressway, each RCI segment is matched with the traffic information of the AVI segment that includes the RCI segment. On SR 408, there are 22 AVI detectors on the Eastbound (EB) and 20 AVI detectors on the Westbound (WB), resulting in 21 segments for EB and 19 for WB. In contrast, the MVDS detectors keep record of traffic flow parameters at the installed points. In this study, 110 MVDS detectors (55 for EB and 55 for WB) were utilized on the expressway. For each RCI segment, if it has no MVDS detector within it, then its traffic states will be represented by the nearest downstream MVDS detector. If the segment contains one MVDS detector, it will use the traffic data from that detector. In cases that one segment has multiple MVDS detectors, the traffic flow parameters will be averaged. Table 7-1 summarizes the information of significant variables involved in the final model.

Table 7-1: Statistics Summary of Variables

Variables	Description	Number	Mean	Std. dev.	Min	Max
Crash	Crash frequency counts	151 × 2	1.563	2.471	0	19
Length	RCI segment length	151	0.283	0.225	0.1	1.847
Auxiliary lane	1 if segment has auxiliary lane; 0 if not	151	0.358	0.481	0	1
AVI volume	Average daily volume of AVI segment during peak or non-peak hours	40 × 2	19104	10366	5142	46830
TTI	Average TTI of AVI segment	40 × 2	0.969	0.096	0.767	1.257
MVDS volume	Average daily volume at MVDS detector during peak or non-peak hours	96 × 2	19201	10505	4830	54422
Occupancy	Average occupancy (%) at MVDS detector	96 × 2	5.613	2.283	1.989	14.269
CI	Average CI (%) at MVDS detector	96 × 2	3.920	2.470	1.517	17.283

7.4 Methodology

7.4.1 Diagnostics of Multicollineairity

Total independence between explanatory variables is rare in observational data analysis. However, when the multicollinearity problem is severe, issues with coefficient estimates will arise. Therefore, detection of multicollinearity is a necessary step especially when highly correlations among explanatory variables are suspected. In crash frequency studies analysis, congestion could be related with traffic volume and speed. To examine the presence of multicollinearity, several procedures could be applied: correlation test, comparing the coefficients of determination, calculating tolerance (TOL) and the Variance Inflation Factor (VIF) (Kalatzis *et al.,* 2011).

In this study, correlations between variables were tested using the Pearson's correlation test. Klein (1962) suggested that if the determination coefficients of auxiliary regressions are greater than that of the general regression, multicollinearity is statistically significant. The coefficient of determination of a linear regression model is calculated as

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \tag{7-1}$$

In crash frequency modeling, the most widely adopted model framework is Poisson regression. Thus the determination coefficient of general regression calculated in this study is the pseudo R^2. The determination coefficient of auxiliary regression is expressed as R_k^2 , is for regression of the kth explanatory variable on all the other explanatory variables (Greene, 2003). TOL and VIF are two measures that assess multicollinearity. TOL of the kth variable is equal to $1 - R_k^2$. TOL ranges from 0 to 1 and higher TOL value indicates less severe multicollineairty. VIF is the reciprocal of TOL, which shows how multicollinearity has increased the instability of the coefficient estimates (Freund *et al.,* 2000). As a rule of thumb, Neter et al. (1996) thought that multicollinearity is only severe at $VIFs > 10$.

7.5 Bayesian Ridge Regression

Full Bayesian inference using Markov Chain Monte Carlo (MCMC) algorithm was used to construct the models. Bayesian modeling framework has been praised for its capability to deal with hierarchical data structure (Huang and Abdel-Aty, 2010). In this study, the hierarchical data structure exist as multiple RCI segments are nested within one AVI or MVDS segment and sharing the same traffic data. In addition, for each segment, crashes were aggregated into peak hours and non-peak hours, hence introducing another level. The hierarchical crash frequency model can then be set up as:

$$Y_{ijt} \sim Poisson(\lambda_{ijt}) \tag{7-2}$$

Y_{ijt} is the observed crash frequency at RCI segment i ($i = 1,2,\ldots,151$) belonging to AVI or MVDS segment j ($j = 1,2,\ldots 40$ for AVI segments and $j = 1,2,\ldots 96$ for MVDS segments) during peak and non-peak hour t ($t = 1,2$). It follows Poisson distribution with parameter λ_{ijt}. However, one limitation of Poisson model that has been discussed in numerous previous studies is the lack of ability to handle over-dispersion (Lord and Mannering, 2010). Two types of random effect terms were included to cope with the issue as shown below:

$$log(\lambda_{ijt}) = a_{jt[i]} + \mathbf{X_i\beta} + \delta_1\epsilon_{it} + \delta_2\epsilon_i \tag{7-3}$$

where $\mathbf{X_i}$ are geometric characteristics variables at RCI level and $\boldsymbol{\beta}$ is the vector of coefficients. ϵ_{it} is segment-peak-hour uncorrelated random effects and ϵ_i is the segment correlated random effects. The random effects follow normal distribution $\epsilon_{it} \sim N(0,1/\tau_{it})$ and $\epsilon_i \sim N(0,1/\tau_i)$ respectively. $a_{jt[i]}$ is the AVI or MVDS level effects. In normal regression, it is defined as $a_{jt} = \mathbf{U_{jt}}\boldsymbol{\gamma}_\mathbf{t}$. $\mathbf{U_{jt}}$ are traffic parameters for peak and non-peak hours. $\boldsymbol{\gamma}_\mathbf{t}$ are vectors of coefficients. In this paper, ridge regression was implemented to deal with multicollinearity. Early applications of ridge regression trace back to Hoerl's work (Hoerl, 1962; Hoerl and Kennard, 1970). When we have a function with the form like $a_{jt} = \mathbf{U_{jt}}\boldsymbol{\gamma}_\mathbf{t}$ and multicollinearity among the $\mathbf{U_{jt}}$, traditional analysis method such as OLS would lead to incorrect inference of variables. By ridge regression, small bias will be induced, however, more precise regression parameter estimates will be yielded (Congdon, 2007). The ridge regression approach can be viewed as a version of standard posterior Bayes regression with exchangeable prior distribution on the elements of the regression vector (Lindley and Smith, 1972). Nevertheless, to make this assumption plausible, a preliminary standardization is needed (Congdon, 2007). Following the suggestions of Ntzoufras and

Crainiceanu (Ntzoufras, 2002; Crainiceanu, 2004), the variables and their coefficients are standardized as in equation (6-4),

$$\begin{cases} z_{jt} = \frac{(u_{jt} - \bar{u}_{jt})}{\mathrm{sd}(u_{jt})} \\ \gamma_t = \frac{b_t}{\mathrm{sd}(u_{jt})} \end{cases} \tag{7-4}$$

And a_{jt} becomes $a_{jt} = \mathbf{Z_{jt}b_t}$.

The models were calibrated in the WinBUGS software (Lunn *et al.,* 2000). $\boldsymbol{\beta}$ were assigned with non-informative prior normal distribution $N(0, 10^6)$. Ridge regression further restricts $\mathbf{b_t}$ to be exchangeable with $b_t \sim N(0, \phi^{-1})$ and $\phi \sim gamma(0.001, 0.0001)$.The random effects terms τ_{it} and τ_i follow $gamma(0.001,0.001)$. When $(\delta_1, \delta_2) = (1,0)$, the model is with uncorrelated random effect; when $(\delta_1, \delta_2) = (0,1)$, it is model with correlated random effect. In total three types of models were tested: 1) uncorrelated random effects model without considering peak and non-peak hours; 2) uncorrelated random effects model with peak/non-peak hours random parameters; 3) correlated random effects model with peak/non-peak hours random parameters. Each model was simulated with three chains of 20000 iterations. The first 5000 iterations were discarded as burn-in period. Model convergence was visually checked according to that the trace plots of the three chains overlap one another (Spiegelhalter *et al.,* 2003a). Models were compared using Deviance Information Criterion (DIC). DIC consists of two elements, the model fitting part $\bar{D}$ and the number of effective variables p_D. Smaller DIC indicates better model. Parameter significance was investigated using Bayesian Credible Interval (BCI). If 0 is not contained within the 95% BCI, the effect of the variable is significant. With the selected best model, the performances of different congestion measures were compared in the same model setting. In the end, traditional crash frequency model was conducted to illustrate how the parameter estimates changed because of the ridge regression method.

7.6 Modeling Results and Discussion

7.6.1 Presence of Multicollinearity

To examine whether multicollinearity exists among the traffic data, correlation coefficients, coefficients of determination, and TOL/VIF values were calculated. Among the three candidate congestion measures, CI from MVDS data was used to represent the congestion condition on the expressway. Average traffic volume (Avg_mvdsvol), average speed (Avg_spd) and average CI (Avg_ci) were used to test multicollinearity. Table 7-2 shows that during peak hours, the traffic parameters have much higher correlations compared with non-peak hours. As expected, traffic volume is positively related with congestion and negatively related with speed. The VIFs for both peak and non-peak hours are less than 10, indicating that the multicollinearity problem is not severe. However, based on the other standard, the coefficients of determination of the auxiliary regressions are higher than that of general regression, which means the presence of multicollinearity. Therefore, when exploring the effects of congestion on traffic safety, multicollinearity should still be taken into account.

Table 7-2: Summary of Multicollinearity Test

Peak Hours	Correlation between variables			Coefficient of Determination	TOL	VIF
	MVDS volume	Speed	CI			
MVDS volume	1.000	-0.315	0.468	0.744	0.256	3.906
Speed	-0.315	1.000	-0.669	0.826	0.174	5.736
CI	0.468	-0.669	1.000	0.847	0.153	6.544
VIF mean						5.395
R^2 general				0.458		
Non-peak Hours	**Correlation between variables**			**Coefficient of Determination**	**TOL**	**VIF**
	MVDS volume	Speed	CI			
MVDS volume	1.000	-0.172	0.009	0.693	0.307	3.253
Speed	-0.172	1.000	-0.404	0.743	0.257	3.887
CI	0.009	-0.404	1.000	0.738	0.262	3.819
VIF mean						3.653
R^2 general				0.383		

7.6.2 Model Comparison

Three types of ridge regression models were compared. Table 7-3 illustrates the parameter estimates for each model and the overall performances of the models. RCI level geometric variables and MVDS level traffic variables were evaluated in the hierarchical model framework. The mean, standard deviation (SD) and the 95% BCI of the coefficients of estimated variables are given.

The effects of geometric variables in the three models are comparable. Logarithmic segment length affects the crash frequency with a positive sign, which implies that crash frequency is likely to increase under larger exposure. Another geometric variable identified to significantly affect the crash count is the existence of auxiliary lanes. The auxiliary lanes provide space for turning movements and speed changes on the mainline near ramps. Vehicles merging in or diverging from the mainline can cause disturbance in traffic flow and increase crash likelihood. In model type (3), the variable of auxiliary lane is significant at 90% BCI.

Table 7-3: Parameter Estimates and Model Fitting

Variables	(1) Uncorrelated random effects			(2) Random parameter with uncorrelated random effects			(3) Random parameter with correlated random effects		
	Mean	SD	95% BCI	Mean	SD	95% BCI	Mean	SD	95% BCI
RCI level									
Intercept	0.873	0.196	(0.466, 1.250)	0.970	0.187	(0.601, 1.344)	0.947	0.241	(0.487, 1.406)
log(Length)	0.769	0.137	(0.495, 0.1.042)	0.801	0.134	(0.541, 1.082)	0.821	0.179	(0.490, 1.180)
Auxiliary lane	0.268	0.168	(-0.061, 0.595)	0.211	0.167	(-0.120, 0.535)	0.343	0.194	(-0.012,0.749)*
MVDS level									
log(MVDS volume)	0.646	0.151	(0.349, 0.948)	0.580[1] 0.149[2]	0.352 0.037	(-0.074,1.269) (0.077,0.225)	0.482[1] 0.128[2]	0.313 0.028	(-0.097,1.120) (0.078,0.183)
CI	0.162	0.026	(0.112, 0.215)	0.791[1] 0.209[2]	0.312 0.145	(0.151, 1.384) (-0.067,0.478)	0.671[1] 0.092[2]	0.278 0.127	(0.129,1.211) (-0.138,0.336)
Model performance									
$\bar{D}$	755.196			763.963			749.718		
p_D	124.000			113.099			90.698		
DIC	879.195			877.062			840.416		

*significant at 90% BCI
[1] peak hours; [2] non-peak hours

For the MVDS level traffic variables, the random parameter models considering peak and non-peak hours outperform the type (1) uncorrelated random effects model and reveal the sophisticated effects of traffic parameters on safety. In the type (1) model, both logarithmic volume and CI have significant positive effects on crash likelihood. However, the random parameter models indicate that the effects of the variables vary for different time periods. Logarithmic volume is only significant during the non-peak hours. During peak hours, although it is still positively related with crashes, the impact is not significant. Congestion, on the other hand, only has significant influence on safety during the peak hours. During the non-peak hours, congestion rarely occurs on the expressway and the safety problems are highly likely to be caused by other factors such as weather conditions or driving errors. The results generated by the three models illustrate that in order to have more accurate understanding of the congestion's impact on safety, we should take into account the time-specific nature of congestion in urban areas. The effects of congestion on expressway safety differ between peak and non-peak traffic hours.

Regarding the performances of the three models, DIC values were compared. The random parameter model with correlated random effects has the lowest DIC. For each segment, the crash counts were observed during peak and non-peak hours, thus could lead to repeated observations per segment. As a result, the observations are correlated. Several studies have proved that by introducing the correlated random effects, models for repeated measures would provide better model fitting (smaller DIC) (Ntzoufras, 2011; Yu *et al.,* 2013b). In this study, it is confirmed that by applying correlated random effects, the model fitting ($\overline{D}$) improved and the model complexity (p_D) reduced. In addition, the SD of traffic variables in model type (3) decreased compared with model type (2), making the parameter estimation of traffic parameters more reliable.

7.6.3 Congestion Measure Selection

According to the model comparison from the above section, the random parameter model with correlated random effects was adopted to evaluate the performances of the three candidate congestion measures. Of the three congestion measures, occupancy and CI were derived from MVDS data while TTI was calculated from AVI data.

Comparing the performances of the three measures as shown in Table 7-4, it can be seen that the parameter estimates in the model with occupancy and CI share higher similarity than those with the TTI. The effects of geometric variables in the three models agree with each other. Larger logarithmic volume will always increase crash occurrence regardless of peak or non-peak hours. Yet the effect of traffic volume should not be confused with the effect of congestion. Although volume and congestion are positively related, higher volume does not necessarily mean congestion. Under the same congestion condition, the segment with more lanes and higher traveling speed will have higher traffic volume. And the effects of volume on crash frequency should be interpreted as an exposure variable. The three congestion measures have similar effects on crash frequency. Congestion is likely to lead to more crashes during peak hours, whereas the effects are insignificant during non-peak hours. The impact of congestion during peak hours is easy to understand. The insignificant relationship between congestion during non-peak hours and crash occurrence might imply that other factors (i.e. driving errors, weather conditions) serve as more important factors contributing to crashes under the light traffic condition.

To explore the congestion measure that gives the best modeling performance, DIC values for the models in Table 7-4 were compared. Spiegelhalter et al. (Spiegelhalter *et al.,* 2003a) suggested that when the difference of DIC between two models is greater than 10, the model with smaller DIC is assured to be better; the difference between 5 and 10 is considered substantial; and if the difference is below 5, it is not significant. In this study, the DIC of the model using CI as congestion measure has the lowest DIC, and the difference is between 5 and 10, thus making a substantial improvement in model performance.

Table 7-4: Parameter Estimates and Model Fitting

Congestion Measures	Occupancy			Congestion Index (%)			Travel Time Index		
Variables	Mean	SD	95% BCI	Mean	SD	95% BCI	Mean	SD	95% BCI
RCI level									
Intercept	0.964	0.256	(0.479, 1.472)	0.947	0.241	(0.487, 1.406)	0.993	0.204	(0.580, 1.372)
log(Length)	0.842	0.180	(0.483, 1.192)	0.821	0.179	(0.490, 1.180)	0.852	0.146	(0.563, 1.143)
Auxiliary lane	0.411	0.233	(-0.059, 0.786)*	0.343	0.194	(-0.012,0.749)*	0.303	0.175	(-0.012, 0.656)*
AVI/MVDS level									
log(MVDS volume)	0.572[1]	0.342	(-0.045, 1.228)	0.482[1]	0.313	(-0.097,1.120)	--	--	--
	0.157[2]	0.045	(0.067, 0.245)*	0.128[2]	0.028	(0.078,0.183)	--	--	--
log(AVI volume)	--	--	--	--	--	--	1.015[1]	0.522	(0.025, 1.987)
	--	--	--	--	--	--	0.869[2]	1.350	(-1.742, 3.602)
Occupancy	0.572[1]	0.334	(0.024, 1.216)	--	--	--	--	--	--
	-0.031[2]	0.081	(-0.146, 0.187)	--	--	--	--	--	--
CI	--	--	--	0.671[1]	0.278	(0.129,1.211)	--	--	--
	--	--	--	0.092[2]	0.127	(-0.138,0.336)	--	--	--
TTI	--	--	--	--	--	--	0.634[1]	0.355	(0.026, 1.352)
	--	--	--	--	--	--	-1.232[2]	1.295	(-3.818, 1.216)
Model performance									
$\overline{D}$	751.326			749.718			754.224		
p_D	94.147			90.698			93.212		
DIC	845.473			840.416			847.436		

*significant at 90% BCI
[1]peak hours; [2]non-peak hours

7.6.4 Model Fitting with/without Ridge Regression

After we constructed the crash frequency model using Bayesian ridge regression, one issue remains: how the crash frequency model with ridge regression differs from the model without ridge regression. Consequently, the model using the same hierarchical structure but without ridge regression was investigated. Table 7-5 summarizes the models with and without ridge regression. The DIC values for the two models are comparable with difference less than 5. This is expected as the multicollinearity won't affect model fitting. The parameter estimates for the two models are similar for the geometric variables and non-peak traffic parameters. However, the estimated effects of the traffic variables during peak hours vary considerably between the two models. The coefficients of traffic volume and congestion without ridge regression during peak hours are much smaller than those with ridge regression. In the multicollinearity test as shown in Table 7-2, it has been confirmed that high correlations exist among the traffic parameters during peak hours. If this issue is not handled properly in model construction, the estimates of the variables' impact will be affected. As a result, ridge regression can be a solution when multicollinerity is detected and the correlated variables are necessary to be retained in the model.

Table 7-5: Parameter Estimates and Model Fitting

	Without ridge regression			With ridge regression		
Variables	Mean	SD	95% BCI	Mean	SD	95% BCI
RCI level						
Intercept	-2.587	1.666	(-5.932, 0.151)	0.947	0.241	(0.487, 1.406)
log(Length)	0.834	0.172	(0.469, 1.170)	0.821	0.179	(0.490, 1.180)
Auxiliary lane	0.386	0.201	(0.001, 0.776)	0.343	0.194	(-0.012,0.749)*
MVDS level						
log(MVDS volume)	0.304[1]	0.178	(0.009, 0.657)	0.482[1]	0.313	(-0.097,1.120)
	0.131[2]	0.028	(0.082, 0.192)	0.128[2]	0.028	(0.078,0.183)
CI	0.329[1]	0.165	(0.059, 0.657)	0.671[1]	0.278	(0.129,1.211)
	0.085[2]	0.141	(-0.174,0.367)	0.092[2]	0.127	(-0.138,0.336)
Model performance						
$\bar{D}$	750.780			749.718		
p_D	92.878			90.698		
DIC	843.658			840.416		

*significant at 90% BCI
[1] peak hours; [2] non-peak hours

7.7 Conclusions

Traffic congestion and safety are the priorities of traffic professionals. In urban areas, the growing traffic has raised an urgent call for traffic authorities to provide both efficient and safe transportation system to the motorists. With the rapid development of ITS techniques, more proactive traffic management can be achieved. To improve the congestion and safety conditions simultaneously, a better understanding of their relationship is required. In this study, the effects of congestion on crash frequency based on an urban expressway in Central Florida were investigated. The 21.4-mile expressway of interest is equipped with 42 Automatic Vehicle Identification (AVI) detectors and 90 Microwave Vehicle Detection System (MVDS) traffic detectors. Comprehensive traffic information was extracted from the ITS systems.

To properly evaluate the effects of congestion on crash frequency, the presence of multicollinearity was checked. Multicollinearity is caused by the correlation among independent variables and will interfere with parameter estimates in regression. In traffic safety analysis, the traffic parameters are related. It was found in this study that during peak traffic hours (7:00 – 9:00 and 17:00 – 19:00), high correlations were verified. Since the coefficients of determination of auxiliary regressions are greater than the value of general regression, multicollinearity exists and should be taken into consideration.

Three models were constructed using Bayesian ridge regression. The longer segment length indicates higher exposure and increases the crash frequency. On the segment with auxiliary lanes, more frequent merging and diverging behaviors are expected. These movements can raise the likelihood of crashes. Regarding the effects of traffic parameters, their influences on crash occurrences were better revealed by assigning them with random parameters taking into account the peak and non-peak hours. In the Poisson model with uncorrelated random effects, volume and congestion were found to be positively related to crash frequency. Nevertheless, in the random parameter models, traffic volume was only significant during non-peak hours and congestion was only significant during peak hours. These conclusions suggest the necessity to distinguish the time of day when investigating how congestion is related to traffic safety. Moreover, the correlated random effects model provides better model fitting and smaller standard variance of coefficients for traffic parameters.

Previous studies have made effort to discover the effect of congestion on traffic safety but failed to reach a consistent conclusion. One of the possible explanations is the selection of congestion measures. In this study, Travel Time Index (TTI) from AVI data, occupancy and Congestion Index (CI) from MVDS data were developed and tested in the crash models. The multiple congestion measures were adopted in hope of more definite conclusion. It turned out that the three congestion measures have similar effects on crash frequency. Higher congestion intensity could increase the crash count during peak hours while they have no significant impact on crash frequency during the non-peak hours. Congestion Index was found to outperform the other two congestion measures based on the model fitting performance.

Ridge regression was introduced in this paper to deal with multicollinearity. This method was expected to generate more accurate parameter estimates since multicollinearity was detected in the data. In response, a model under the same modeling structure but without ridge regression was built to compare against the ridge regression model. It was found that the two models have highly comparable DIC values since correlations among explanatory variables won't affect model fitting. However, parameter estimates vary between the two regressions. These findings shed some light on how the effects of congestion on safety should be properly interpreted. The presence of multicollinearity can cause changes in the magnitudes of coefficients, and even the significance and signs of parameters in more extreme cases. In the presence of this issue, ridge regression serves as a practical tool to gain more accurate understanding about the effects of the variables-of-interest.

While congestion was confirmed by multiple measures to significantly increase crash frequency in this study, further improvement can still be made. In future work, it is expected to include other expressways rather than focusing on one expressway to reach more conclusive statement. Whether crash types should be taken into consideration is another argument worth discussion. Also, it is desirable to evaluate the congestion-crash relationship using traffic data prevailing near the crash time for more direct cause-effect. Real-time crash prediction models could serve as an alternative approach to evaluate the relationship.

CHAPTER 8: BIG DATA APPLICATIONS IN REAL-TIME SAFETY EVALUATION AND IMPROVEMENT

8.1 Introduction

In an age of data explosion, almost every aspect of social activities is impacted by the abundance of information. The information, characterized by alarming volume, velocity and variety, is often referred to as “Big Data” (Beyer and Laney, 2012). As one fundamental elements of human life, transportation also confronts the promises and challenges brought about by the Big Data era. Big Data in transportation arena, enabled by the rapid popularization of Intelligent Transportation System (ITS) in the past few decades, is often collected continuously from different sources over vast geographic scale. Huge in size and rich in information, the seemingly disorganized data could considerably enhance experts’ understanding of their system. In addition, proactive traffic management for better system performance is made possible due to the real-time nature of the Big Data in transportation.

Operation efficiency and traffic safety have long been deemed as priorities among highway system performance measurement. While efficiency could be evaluated in terms of traffic congestion, safety is studied through crash analysis. Extensive works have been conducted to identify contributing factors and remedies of traffic congestion and crashes respectively. These studies lead to gathering consensus that operation and safety have played as two sides of a coin, ameliorating either would have a positive effect on the other. With the advancement of Big Data, monitoring and improvement of both operation and safety proactively in real-time have become an urgent call.

This study focuses on Central Florida Expressway Authority (CFX)'s system in Central Florida area. Three expressways are located in the densely populated urban area. The toll expressways communicate downtown area, airport and other attraction areas in Orlando, serving both commuters and tourists. Multiple ITS systems are equipped on the system for electronic toll collection and traveler information. In 2013, the authority introduced Microwave Vehicle Detection System (MVDS) to monitor traffic conditions across different sections of the expressways. A total of 275 MVDS detectors are densely allocated along the 75-mile expressway network, with average spacing less than 1 mile. Comprehensive traffic flow parameters are archived on one-minute interval basis without interruption. As a result, the large geographic scale of deployment and continuous data collection provide a full view of network performance and serve as the source of Big Data. In this work, real-time operation and safety analyses based on MVDS data are carried out in the hope to shed some light on the Big Data applications in real-time safety monitoring and improvement.

8.2 Background

Effective strategies to improve traffic operation and safety simultaneously require profound understanding about their features and relationship. In the age of information, these objectives could be efficiently realized through Big Data applications. Traffic congestion can be viewed as a product of the interaction between demand and capacity. Periodic high demand at specific bottlenecks during peak hours can result in recurrent congestion while incidents, especially crashes, reducing roadway capacity temporarily lead to non-recurrent congestion. To catch this dynamic process, Big Data generated from the ITS detection system could be leveraged to develop congestion measurement in real-time. In the meantime, crash occurrence is often regarded as random events affected by human behavior, roadway design, traffic flow and weather conditions. Big Data applications also introduce new perspectives in safety analysis. Thanks to the advantages brought by Big Data, researchers are able to restore the traffic condition for each crash case and draw general conclusions using individual crash data. As a result, Big Data applications in the current work will focus on developing congestion measurement and uncovering the relationship between safety and congestion, both in real-time.

8.2.1 Real-time crash prediction

Currently real-time crash analysis emphasizing congestion has not been adequately explored. Some existing research include Christoforou *et al.* (2011) and Hossain and Muromachi (2012). Nevertheless, considering the analytical methods enabled by Big Data, the interrelation of congestion and traffic safety, and the vision to improve them together, this topic worth thorough investigation. In this study, we adopt real-time safety evaluation incorporating congestion. In the hope to propose an integrated improvement strategy for congestion and safety, we further introduce the first order reliability method (FORM) widely used in structural reliability analysis to determine when it is appropriate to trigger the warning to the expressway system.

8.2.2 Rear-end crash

It has been widely accepted that the same traffic state could impose distinct impact on traffic safety regarding to crash type and severity. Congestion's effect on traffic safety, to be properly analyzed, should also follow the same reasoning. So far, the relationship between congestion and crash severity gained relatively more research attention (Shefer and Rietveld, 1997; Quddus *et al.,* 2009; Wang *et al.,* 2009). Crash types under congestion, on the other hand, haven't been adequately addressed.

Common sense informs us that on freeways/expressways single vehicle crashes are more likely to occur with driving errors under free flow condition; sideswipe crashes might be caused by inappropriate lane-change behaviors and speed variation between lanes; and rear-end crashes often involve multiple vehicles with the leading vehicle suddenly decelerating. Under congestion, vehicles approaching the queue end have to slow down in advance otherwise the likelihood of rear-end crashes could be increased considerably. Lee *et al.* (2006) stated that rear-end crashes took a large proportion of total freeway crashes. Abdel-Aty *et al.* (2007) further confirmed that rear-end crashes are highly related with congestion, particularly if speed variation and average occupancy are elevated. Golob and Recker's (2003b) research conclusions were in line with the above study, associating rear-end collisions with high variations in relatively low speed. Christoforou *et al.* (2011) implemented multivariate probit model and indicated that the rear-end crashes were more probable under congestion while sideswipes more probable under "intermediate" density traffic regimes. In all, to achieve a more conclusive statement about the relationship between crash occurrence and crash mitigation under congestion, the crash types should be taken into consideration.

By synthesizing the discussion above, this study underlines the Big Data applications in congestion and safety analyses and improvement. Speed-based congestion index is applied to continuously monitor the traffic congestion in spatial-temporal dimensions. The relationship between congestion and rear-end crashes is explored within real-time modeling framework and FORM analysis. Strategies established on MVDS traffic data are proposed to improve operation and safety in real-time on the urban expressways.

8.3 Data Preparation

In this study, only three expressways in the system were selected for analysis, namely SR 408, SR 417 and SR 528. The reason is that only a small portion of crashes on CFX's network occurred on SR 414 and SR 429. In addition, almost no congestion was detected on the mainline of these two expressways except for the segment at the end of SR 414 Eastbound. On the 75-mile system of SR 408, SR 417 and SR 528, a total of 275 MVDS detectors have been installed. The system is well covered with the average distance between adjacent detectors less than 1 mile. The MVDS data have been collected since July 2013. At the time of this research, eight-month (July, 2013 to February 2014) traffic data have been collected. Congestion Index was selected to measure congestion in real-time.

During the studied time period, a total of 581 crashes occurred on the three expressways of which 243 are rear-end crashes (Table 8-1). SR 408 has both the highest crash count and rear-end crash rate among the three expressways. This phenomenon is credited to the large traffic volume and commuter traffic on SR 408 compared with the other two expressways. For each crash, traffic data of 5 to 10 minutes prior to the crash from two upstream and two downstream MVDS detectors (Figure 8-1) closest to the crash location are collected and aggregated.

Table 8-1: Crash Occurrence On Expressways

Expressway	Rear-end Crashes		
	No	Yes	Total
SR 408	133(45.86%)	157(54.14%)	290
SR 417	106(76.26%)	33(23.74%)	139
SR 528	99(65.13%)	53(34.87%)	152
Total	338(58.18%)	243(41.82%)	581

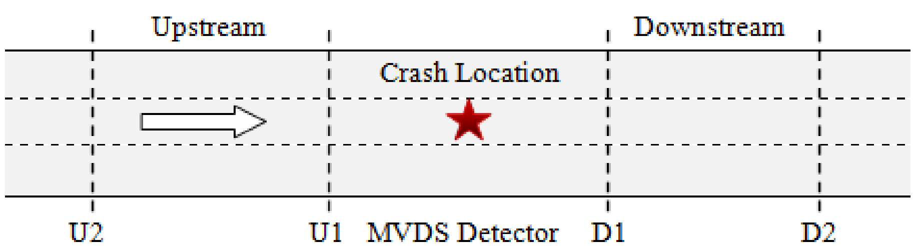


Figure 8-1: Crash Location And MVDS Detector Assignment

For the 243 rear-end crashes, 962 non-crash cases are matched. In previous matched case-control studies, control cases were selected for the same location at the same time and same weekday but in different weeks (Abdel-Aty *et al.,* 2004b; Ahmed and Abdel-Aty, 2012; Xu *et al.,* 2012). The motivation behind this design is to control the variability caused by the time of day, season and roadway geometric characteristics. On the urban expressways, congestion tends to be recurrent especially when large commuter traffic is expected. Therefore following the traditional procedure by extracting traffic parameters at the same time as a crash case for non-crash cases might conceal the true effect of congestion on safety which is exactly what we are interested in. Taking this into account, we matched non-crash cases by extracting traffic conditions at the same location three hours and six hours before and after crash during which time no crash was observed, making the ratio between study and control group approximately 1:4. The total 1205 observations are then divided into training (70%) and validation (30%) data sets. When making the data partition, we made sure that the crash case and its matched non-crash cases are assigned to the same data set. For each MVDS detector selected (xx: u1, u2, d1, d2), logarithm of volume (log_xx_vol), truck percentage (xx_trkpct), average, standard deviation and logarithm of coefficient of variation of the speed (xx_avgspd, xx_stdspd, and log_xx_cvspd), speed difference between inner and outer lanes (xx_spddiff), number of lanes at the detection location (xx_lanes) and congestion index (xx_ci) were extracted.

Besides traffic parameters, the spatial-temporal related properties of crashes were also prepared. Whether the crash occurred during peak hours (peak) defined as 7:00 ~ 9:00 am and 17:00 ~ 19:00 pm were set up as binary variable. The posted speed limit (maxspeed), the horizontal curvature and existence of auxiliary lanes near ramps at the crash locations were all incorporated from FDOT Roadway Characteristics Inventory (RCI) database.

8.4 Methodology

8.4.1 Real-time Congestion monitoring

The speed-based congestion index is adopted in this study to measure the congestion intensity on both spatial and temporal scales. The free flow speed is the 85th percentile speed at the detection location. CI is a continuous variable with the range between 0 and 1. The increase in CI value indicates higher congestion level. The CI has been aggregated into 5-minute intervals at each detection location. Filled contour plots are created to visualize the congestion distribution.

8.4.2 Random forest

Random forest is an ensemble classifier using many decision tree models to vote for the most popular class (Breiman, 2001). A single decision tree suffers from high variance or bias. In contrast, random forest offers unbiased estimates of the classification error as trees are added to the forest. Also, strong law of large numbers guarantees random forest is robust against over-fitting.

One of the basic practices of random forest in real-time traffic safety evaluation is to estimate variable importance (Abdel-Aty and Haleem, 2011; Yu and Abdel-Aty, 2014). The random forest algorithm estimates the importance of a variable by looking at how much prediction error increases (or decrease in accuracy) when OOB (out-of-bag) data for that single variable is permuted (Liaw and Wiener, 2002). Another measure is the total decrease in node impurities denoted by Gini coefficient from splitting on the variable, averaged over all trees (Breiman, 2006). The former measure has the drawback that it overestimates the variable importance of highly correlated variables (Strobl *et al.,* 2008). The latter, on the other hand, does not perform fairly with predictors of many categories (Strobl *et al.,* 2007). In this study, both measures along with Pearson's correlation test were employed to select the important variables for real-time safety evaluation.

8.4.3 Bayesian logit model

To predict real-time crash likelihood, logistic regression models under Bayesian framework were evaluated. The target variable is the binary indicator of crash occurrence, with probability p for crash case ($y = 1$) and $1 - p$ for non-crash case ($y = 0$). Three types of logistic models were constructed and their performances compared: (1) the matched case-control logit model; (2) fixed effects logit model; (3) random parameters logit model differentiating peak and non-peak hours. The specifications of the logit models are illustrated below:

$$y_i \sim Binomial(p_i, 1) \tag{8-1}$$

For model type (1),

$$logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{X_i\beta} + \varepsilon_{j[i]} \tag{8-2}$$

For model type (2),

$$logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \boldsymbol{X_i\beta} \tag{8-3}$$

For model type (3),

$$logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_{0[t]} + \boldsymbol{X_i\beta_t} \tag{8-4}$$

Where β_0 is constant term, $\boldsymbol{\beta}$ is the vector of explanatory variable coefficients. $\beta_{0[t]}$ and $\boldsymbol{\beta_t}$ bear the same meaning except that they are for peak ($t = 1$) and non-peak ($t = 0$) traffic hours. $\varepsilon_{j[i]}$ denotes the contribution to the logit of all terms constant within the j^{th} group (Hosmer and Lemeshow, 2004). In Bayesian inference, prior distributions for parameters have to be justified first. Here non-informative priors were assigned. For $\beta_0/\beta_{0[t]}$ and each element in $\boldsymbol{\beta}/\boldsymbol{\beta_t}$, they were assigned to follow $normal(0, 10^6)$. ε_j has normal distribution $normal(0, 1/\tau)$ where $\tau \sim gamma(0.001, 0.001)$.

The models were calibrated in the software WinBUGS. Three chains were simulated with the first 5000 iterations fed as burn-in out of the 15000 iterations. Parameter convergence was assured by checking if the trace plots of the three chains appear to be overlapping one another (Spiegelhalter *et al.,* 2003a). The Deviance Information Criterion (DIC) was used as a Bayesian measure of model complexity and fit. Smaller DIC indicates better model. Bayesian Credible Interval (BCI) was used for parameter estimation. If the 95% BCI does not contain 0, then the effect of the variable is significant.

8.4.4 First Order Reliability Analysis

Reliability and risk analysis is essential concern in structural engineering. The combined actions of the elements of the structure will lead to system failure when specific conditions are met. As a result, reliability analysis is used to determine the critical point that distinguishes safe and unsafe conditions. In traffic safety, we can also interpret each crash as a failure of the expressway system. Similarly, we can derive the combination of traffic parameters indicating potential hazards on the system.

Reliability is denoted by the probability of limit state function (LSF) $g(\boldsymbol{X}) > 0$ where $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$. When $g(\boldsymbol{X})$ is less than zero, it means that the system is in failure region. Then reliability could be numerically expressed as

$$R = P\{g(\boldsymbol{X}) > 0\} = \int_{g(x)>0} f_x(\boldsymbol{X})dx \tag{8-5}$$

Where $f_x(\boldsymbol{X})$ is the joint distribution of $\boldsymbol{X}$. In the logit model of real-time crash prediction, each case is corresponded with *p*, if p is higher than a cutoff point (c), then the case will be classified as crash, otherwise non-crash. Based on inverse logit function, we can apply the reliability analysis in real-time safety evaluation as follows:

$$p_i = \frac{\exp(\beta_0 + \boldsymbol{X_i\beta})}{1+\exp(\beta_0 + \boldsymbol{X_i\beta})} < c \tag{8-6}$$

$$g(\boldsymbol{X}) = -\boldsymbol{X_i\beta} - \beta_0 + \log(\frac{c}{1-c}) > 0 \tag{8-7}$$

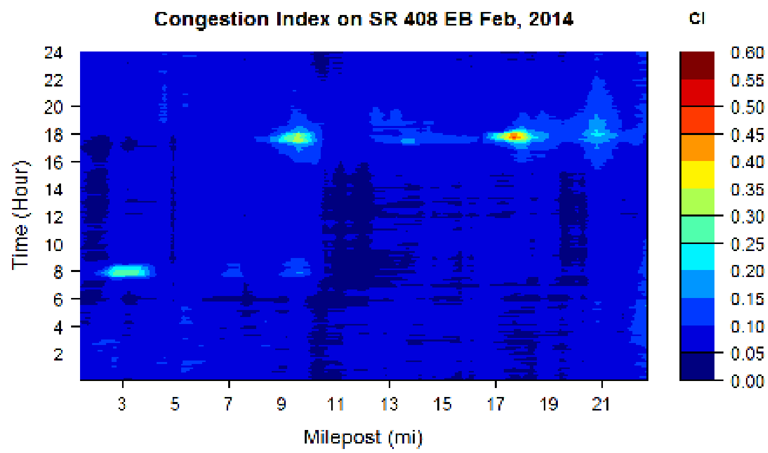
To get the joint distribution $f_x(\boldsymbol{X})$, the distribution of each significant variable and the correlations between them were determined first. SAS procedure SEVERITY was used to test candidate distributions and -2 Log Likelihood chosen as the comparison criterion. Then the joint probability density function (PDF) is calculated and transformed from the variables' original $\boldsymbol{X}$ space to the standard normal $\boldsymbol{U}$ space using Nataf transformation (Li *et al.,* 2008).

When the transformed joint PDF $f_x(\boldsymbol{U})$ with the corresponding LSF $g(\boldsymbol{U})$ are available, First Order Reliability Method (FORM) is applied to calculate the critical point. The term "first order" means that the method approximates the LSF by taking the first order Taylor expansion. The critical point is achieved by maximization of the transformed joint PDF at the limit state $g(\boldsymbol{U}) = 0$. Then it is transformed back into $\boldsymbol{X}$ space. A more detailed description of FORM can be found in Yu *et al*. (2013c) .

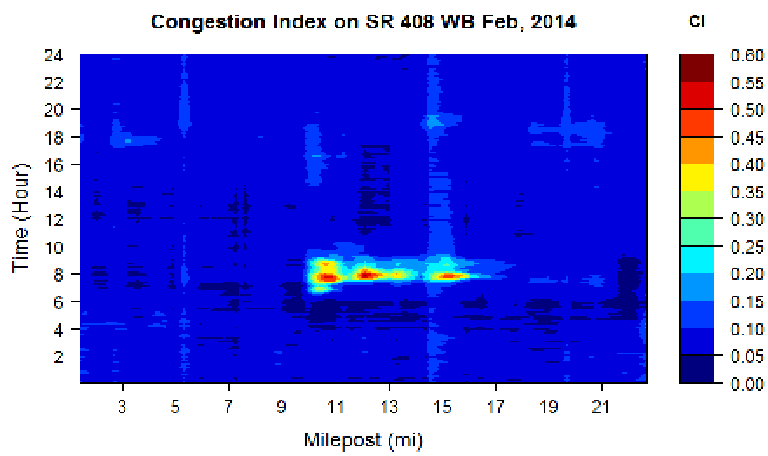
8.5 Congestion Monitoring and Safety Modeling Results

8.5.1 Congestion evaluation

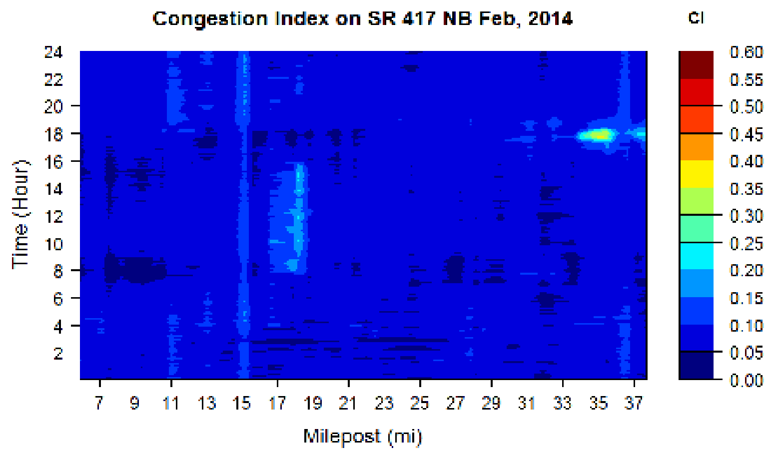
Of the eight-month period, the most recent month (February, 2014) was selected to reflect the congestion conditions on the three expressways. As mentioned, the CI values were aggregated at five-minute intervals for each station. To achieve more stable conclusions about congestion segments and time duration, they were averaged by the weekdays in February. Currently, CFX applies TTI as their congestion measurement. TTI of 1.25 and 2.0 are defined as thresholds for moderate and high congestion. The ratio between actual travel speed and free flow speed given the two TTI congestion thresholds are namely 4:5 and 2:1, which are equivalent to CI of 0.2 and 0.5 respectively. As a result, CI values of 0.2 and 0.5 are set up as the moderate and high congestion thresholds. From the Figure 8-2 (a) – (f) below, it can be concluded that the level of congestion is highly localized and time specific. For the same expressway, morning and evening peak hours are identified on multiple detecting points. For the same location, the congestion levels vary significantly across different time of day. Therefore, to achieve more accurate congestion detection, continuous monitoring is necessary. For traffic safety studies, the use of real-time congestion measurement should also be encouraged to reveal the true effect of congestion on crash occurrence.



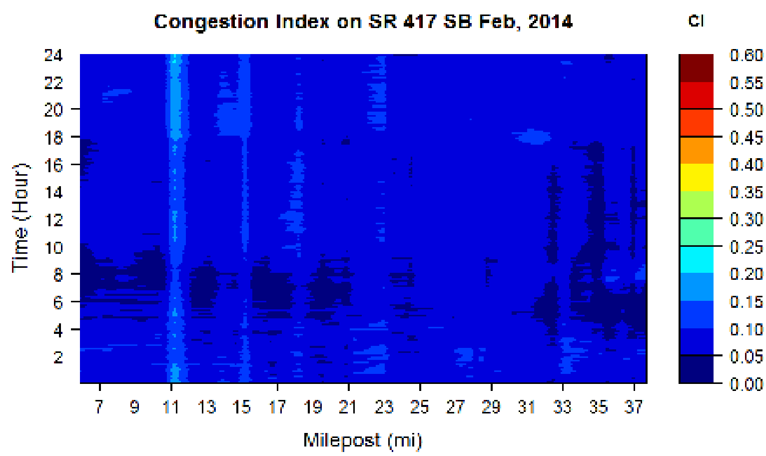
(a)



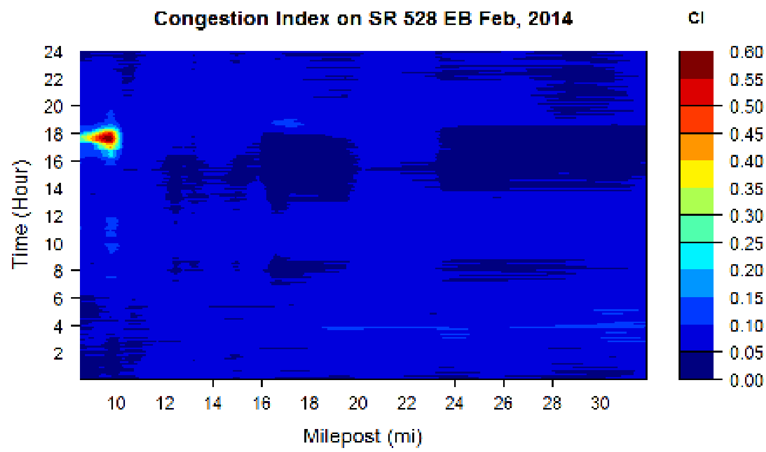
(b)


Congestion Index on SR 417 NB Feb, 2014
CI
Time (Hour)
24
22
20
18
16
14
12
10
8
6
4
2
7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37
Milepost (mi)
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

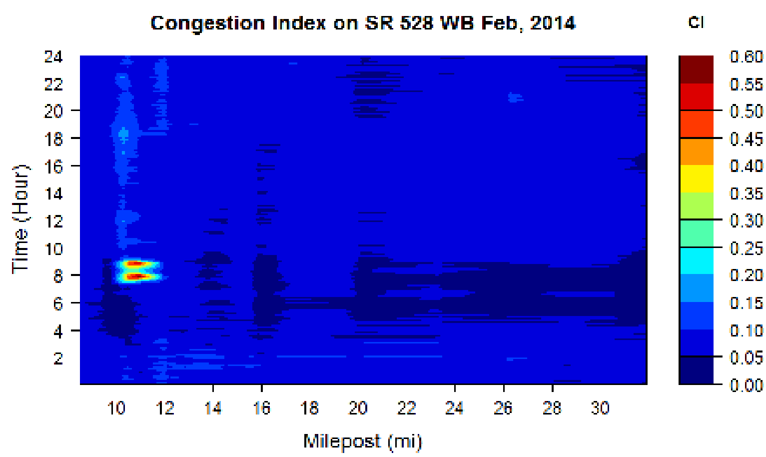
(c)


Congestion Index on SR 417 SB Feb, 2014
CI
Time (Hour)
24
22
20
18
16
14
12
10
8
6
4
2
7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37
Milepost (mi)
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

(d)



(e)



(f)

Figure 8-2: Congestion Evaluation for the Expressway System (a)-(f)

8.5.2 Variable selection

The random forest model for variable selection was constructed in R-based data mining tool Rattle (Williams 2011). In the model specification, 4 variables were randomly sampled at each split. As suggested by Breiman (2006), the number of trees to grow in random forest should not be too small, to ensure that every observation gets predicted at least a few times, we grew 500 trees. The variable importance rankings only listed the most important 20 variables out of the total 37 candidate variables; other candidate variables deemed unimportant by the algorithm were omitted. Figure 8-3 depicts that while there are minor differences between the variable rankings by the two importance ranking methods, the most important types of variables are basically the same. Logarithm of volume, peak hour, average speed and congestion index are the crucial variables. However, the correlation between variables should be investigated before identifying the variables to be incorporated in the final model. To cope with the issue, Pearson's correlation test in Table 8-2 and a simple logistic regression were run to keep the significant variables but controlling for correlations.

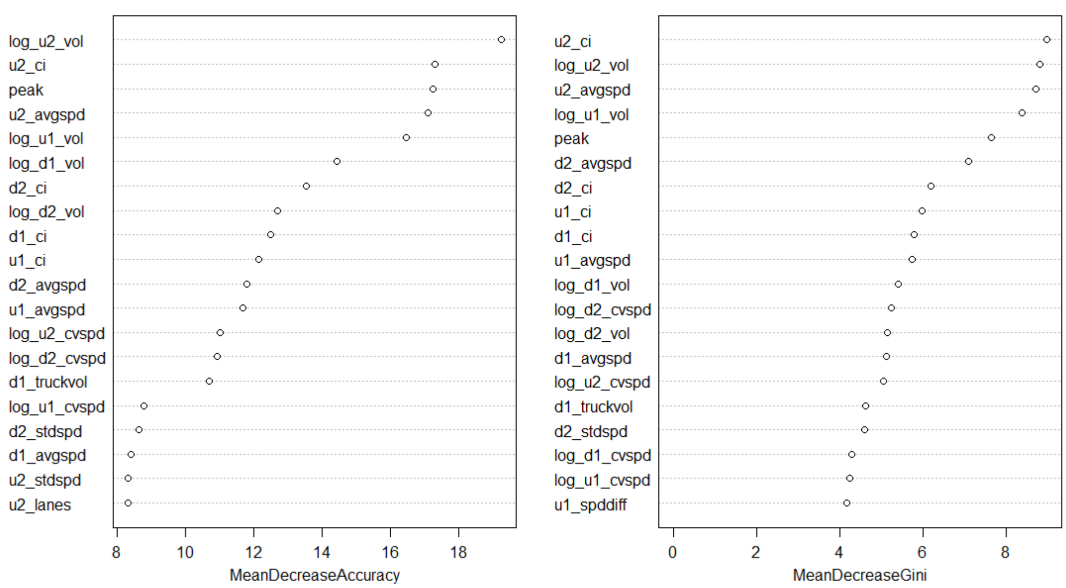


Figure 8-3: Variable Importance Based on Random Forest

Table 8-2: Pearson's Correlation Test for Variables in the Final Model

Pearson's correlation	peak	log_u2_vol	u2_avgspd	d1_ci
Peak	1.0000	0.3437	-0.2767	0.3289
log_u2_vol	0.3437	1.0000	-0.2426	0.1331
u2_avgspd	-0.2767	-0.2426	1.0000	-0.3705
d1_ci	0.3289	0.1331	-0.3705	1.0000

Synthesizing the results from random forest, correlation test and preliminary logistic regression, four variables were selected for the real-time crash prediction model: the peak hour indicator, logarithmic volume and average speed at U2 station and the CI at D1 station. Descriptive statistics of these variables are provided in Table 8-3.

Table 8-3: Summary of Descriptive Statistics

	Description	Mean	Std Dev	Min	Max
crash	Rear-end crash: non-crash=0; crash=1	0.201	0.401	0.000	1.000
peak	Peak hours: nonpeak=0; peak=1	0.161	0.368	0.000	1.000
log_u2_vol	Log volume of U2 station	4.611	1.118	0.693	6.762
u2_avgspd	Average speed of U2 station	62.138	8.735	2.500	98.000
d1_ci	Congestion Index of D1 station	0.058	0.103	0.000	0.909

8.5.3 Real-time logit models

Three types of real-time logit models based on the training data were tested and compared. All the four variables involved in the final model stage appear to be significant at 95% confidence interval (Table 8-4). Peak hour is proved to significantly increase crash likelihood. Logarithmic volume at U2 station is positively related with crash occurrence while the average speed at the same location is negatively associated with rear-end crash. The congestion index at D1 station is also found to contribute to crashes. While Congestion Index is a direct measure of congestion intensity, higher traffic volume and lower speed are also regarded as indirect indicators of congestion. In this study, the effects of traffic parameters at both upstream and downstream locations all converged to the same statement: rear-end crashes are significantly affected by traffic congestion on urban expressways.

Table 8-4: Variable Effects and model comparison

	Random Effect		Fixed Effect		Random Parameter	
	Mean	95% BCI	Mean	95% BCI	Mean	95% BCI
Intercept	--	--	-1.505	(-3.487, 0.541)	-1.031[1] -3.315[2]	(-4.176, 2.589) (-5.268, -0.092)
peak	1.905	(1.435,2.419)	1.857	(1.373, 2.363)	-- --	-- --
log_u2_vol	0.275	(0.116, 0.435)	0.382	(0.171, 0.596)	0.338[1] 0.823[2]	(0.117, 0.554) (0.291, 1.374)
u2_avgspd	-0.057	(-0.070, -0.045)	-0.042	(-0.066, -0.016)	-0.032[1] -0.048[2]	(-0.059, -0.002) (-0.087, -0.017)
d1_ci	6.053	(3.253, 9.546)	6.809	(3.658, 10.920)	7.288[1] 6.190[2]	(3.428, 12.160) (6.200, 10.630)
Model Estimation						
$\bar{D}$	634.211		632.975		629.562	
p_D	4.879		5.171		6.652	
DIC	639.090		638.146		636.214	

[1] non-peak hours; [2] peak hours.

As for the model selection, both model performance and prediction power were compared. The DIC values for the three models are comparable with random parameter model having slightly lower DIC. To evaluate the prediction power of the model, the cutoff point that maximizes both sensitivity and specificity has to be determined first. A graphical method is to select the point where sensitivity and specificity curves cross (Hosmer Jr and Lemeshow, 2004). The optimal cutoff point in this study is 0.14 as illustrated in Figure 8-4. The prediction outcome, sensitivity, specificity, overall accuracy rate and area under the ROC (AUC) of both training and validation data sets are calculated for each model (Table 8-5). The performances of the three models based on training data are similar. In the validation data set, the random parameter outperforms others in specificity and overall accuracy rate.

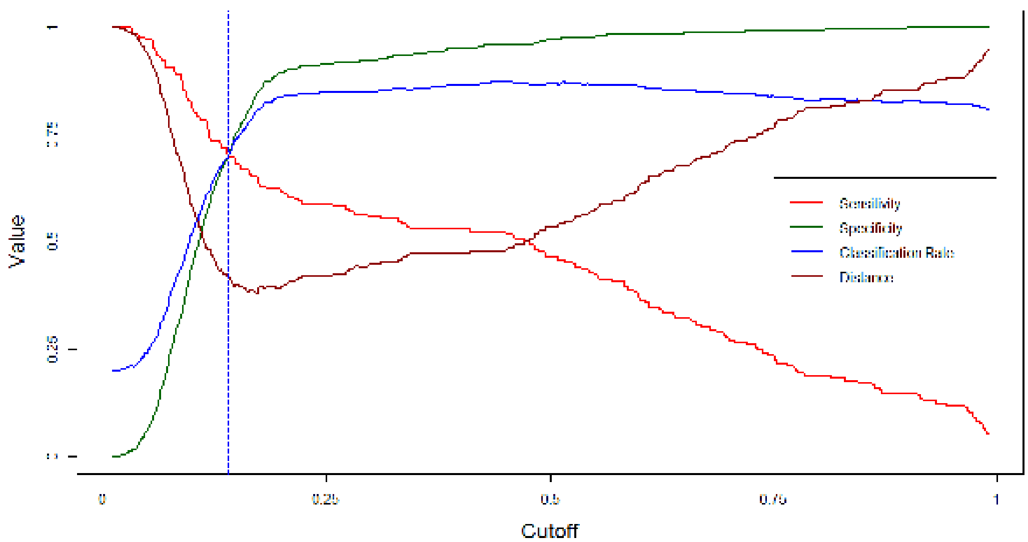


Figure 8-4: Cutoff Point Determination

Table 8-5: Model Classification Results

		Random Effect		Fixed Effect		Random Parameter	
Training Data							
	Predicted	0	1	0	1	0	1
Observed	0	493	182	474	201	491	184
	1	52	118	50	120	52	118
	Sensitivity	0.706		0.706		0.694	
	Specificity	0.702		0.702		0.727	
	Accuracy	0.703		0.703		0.721	
	AUC	0.774		0.779		0.781	
Validation Data							
	Predicted	0	1	0	1	0	1
Observed	0	205	82	202	85	210	77
	1	23	50	23	50	23	50
	Sensitivity	0.685		0.685		0.685	
	Specificity	0.714		0.704		0.732	
	Accuracy	0.708		0.700		0.722	
	AUC	0.755		0.755		0.755	

Based on the model fitting and prediction power of the training and validation data, we prefer the random parameter model as the input for reliability analysis. Another practical reason we choose the random parameter model is that FORM analysis could not deal with categorical variables.

8.5.4 Reliability Analysis

In reliability analysis, since we are more interested in the congestion's effects on crashes, we focus on the peak hour conditions during which time period congestion is more predictable. As discussed earlier, when the combined effects of individual elements in a system reach certain state, the system has high probability of failure and reliability analysis is used to determine this critical point in advance. As the application of reliability analysis in traffic safety evaluation, the critical point at which the crash contributing factors would most likely cause unsafe conditions on the expressway system is desired. Safety interventions will then be triggered if critical point is reached. On the basis of random parameter logistic regression results, we defined the LSF for peak hour as

$$g(\boldsymbol{X}) = -0.823\log(\text{u2\_vol}) + 0.048(\text{u2\_avgspd}) - 6.190(\text{d1\_ci}) + 1.50 \qquad \text{(8-8)}$$

Then the distribution for each variable was fitted as displayed in Table 8-6. Five types of candidate distributions were tested. For each distribution fitting, convergence and goodness of fit were recorded. The smaller -2 Log likelihood value, the better the candidate distribution fits. The log_u2_vol follows normal distribution. Weibull distribution has the lowest -2 Log likelihood values for u2_avgspd and d1_ci.

Table 8-6: Distribution Fitting and Selection

Distribution	log_u2_vol			u2_avgspd			d1_ci		
	Converged	-2 LL	Selected	Converged	-2 LL	Selected	Converged	-2 LL	Selected
Normal	Yes	496	Yes	Yes	1663	No	Yes	-76	No
Lognormal	Yes	1053	No	Yes	1757	No	Yes	-443	No
Exponential	Yes	1098	No	Yes	2058	No	Yes	-389	No
Weibull	Yes	555	No	Yes	1645	Yes	Yes	-458	Yes
Gamma	Yes	769	No	Yes	1719	No	Yes	-452	No

*-2LL: -2 Log Likelihood

Given the distribution of each variable, the FORM analysis was conducted using OpenSees software (Mazzoni *et al.,* 2006). In FORM analysis, it is the whole expressway system instead of individual locations that is under inspection. As input for the FORM model, basic statistics (i.e. mean and standard deviation) and the correlation matrix of the three variables during peak hours are required. The results of the FORM critical point are shown in Table 8-7. According to the critical point, when the system Congestion Index reaches 0.075, there would be high probability of a crash on the expressway.

Table 8-7: FORM Analysis for Peak Hour

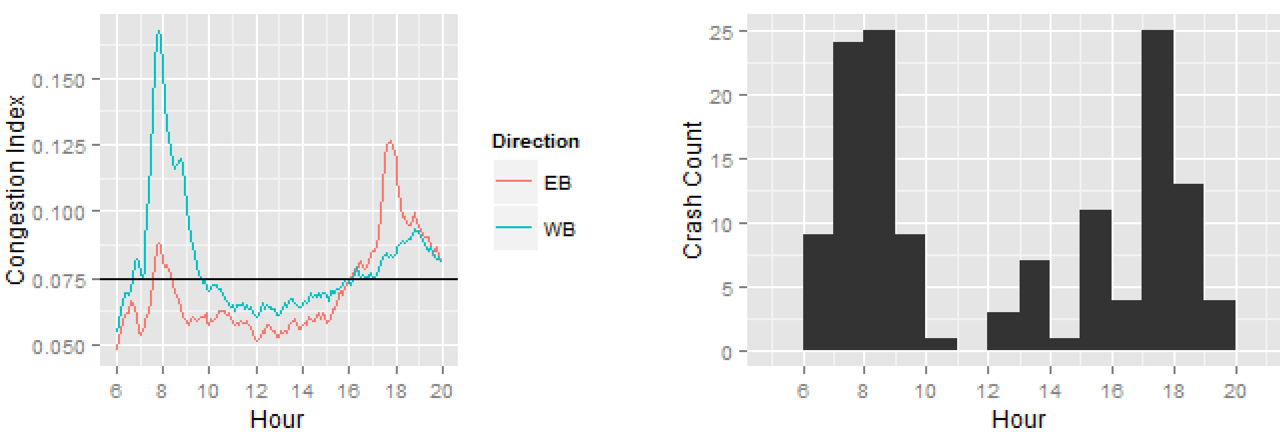
	Basic Statistics		Pearson Correlations			FORM Critical Point
	Mean	Std. Dev	log_u2_vol	u2_avgspd	d1_ci	
log_u2_vol	5.007	0.561	1	-0.366	0.259	5.17
u2_avgspd	67.448	4.785	-0.366	1	-0.371	67.0
d1_ci	0.071	0.034	0.259	-0.371	1	0.075

Figures 8-5 to Figure 8-7 depict the profiles of (a) average CI along time for each traveling direction on the expressway and (b) the crash frequencies along time. The system CI profiles are in accordance with plots in Figure 8-2. Figures 8-5 to Figure 8-7 also show that the time period when the system CI is above 0.075 is mostly morning or evening peak hours during which more rear-end crashes are observed. These figures visually prove that the critical point from FORM analysis is able to determine the system's safety condition based on congestion intensity.

Regarding to the d1_ci for each crash case, the average CI during peak hours is 0.2 (Table 8-8), the same as the moderate congestion threshold. Combining the conclusions of FORM analysis for the whole system and individual crash cases, real-time safety monitoring could be carried out using CI as a safety indicator: when the system CI reaches 0.075, the CI at each detector location will be examined; if it exceeds 0.20, safety countermeasures should be triggered at the upstream segment.

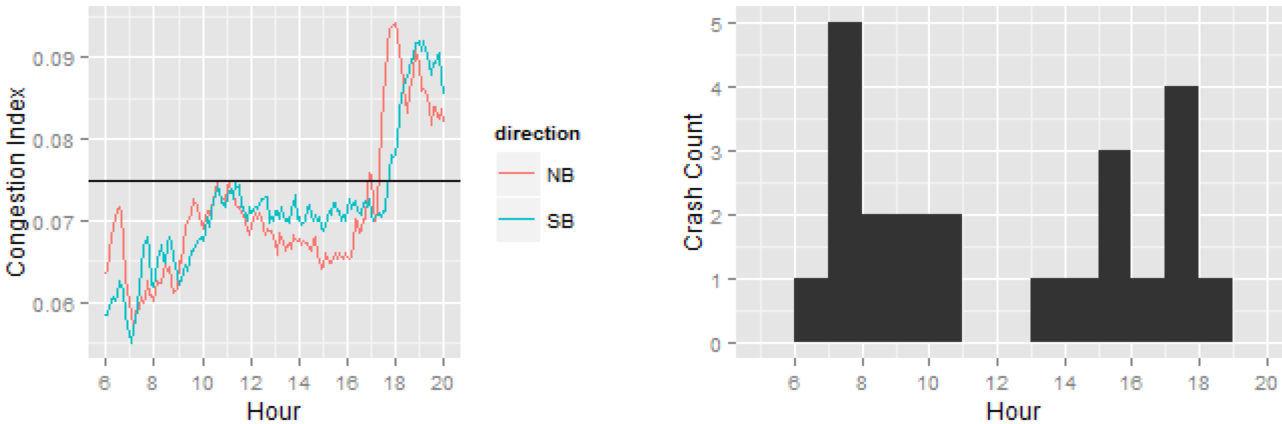
Table 8-8: Congestion Index for Crash Cases

Analysis Variable : d1_ci			
peak	Crash Count	Mean	Std Dev
0 (non-peak)	120	0.0683	0.1310
1 (peak)	123	0.2000	0.2349



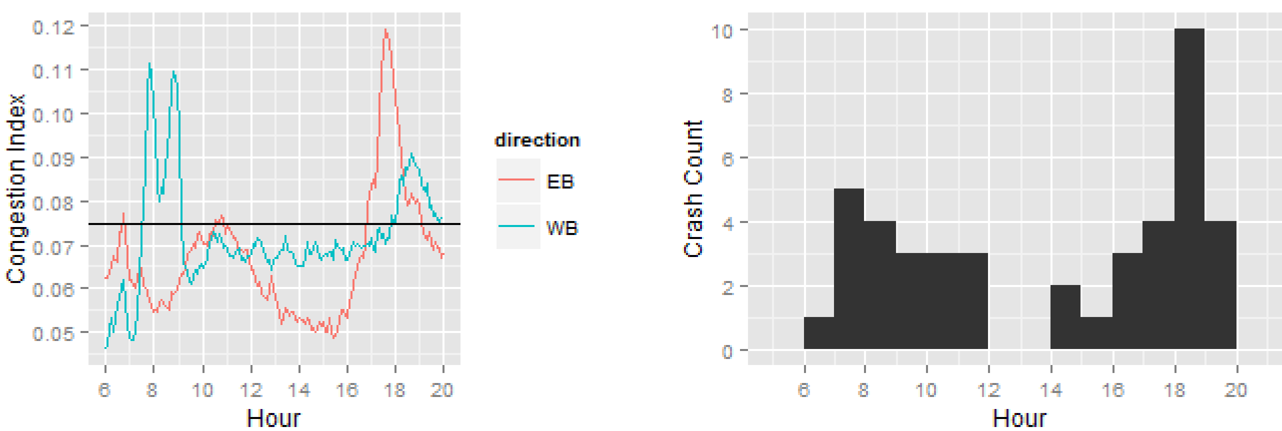
(a) (b)

Figure 8-5: Profiles of Congestion Index and Rear-end Crashes for SR 408



(a) (b)

Figure 8-6: Profiles of Congestion Index and Rear-end Crash for SR 417



(a) (b)

Figure 8-7: Profiles of Congestion Index and Rear-end Crash for SR 528

8.6 Conclusions

The rapid development of ITS systems in the past few decades has catalyzed the implementation of Big Data in the transportation arena. To harness the power of Big Data for better traffic system performance, it is vital to take full advantage of its real-time nature. In this study, the viability of monitoring and improving traffic operation and safety on urban expressways in Central Florida using real-time Microwave Vehicle Detection System (MVDS) data is researched. From perspectives of volume, velocity and variety, the MVDS should be regarded as a main source of Big Data. The detection system archives spot speed, volume, lane occupancy and volume by vehicle type per lane on minute basis. Congestion detection and the real-time safety analysis were developed for three expressways based on these data.

Traditional congestion measures lack the ability to capture the variability of congestion. Real-time congestion measurement based on Big Data is therefore more desirable to identify the congestion pattern in both the temporal and spatial dimensions. Congestion Index was introduced to measure congestion intensity and visualized via filled contour plot. It was found that congestion on the urban expressways is highly time and location specific. Recurrent congestion during morning and evening peak hours are observed for specific locations. Faced with the large traffic demand during peak hours, the traffic authorities could not always expand the system capacity as a solution. Currently, Dynamic Message Signs (DMS) have been widely applied on the CFX system for travel time estimation. However, they could also be used for congestion warning. Information of congestion locations and potential delay would leave drivers enough time to adjust their speed and raise their awareness of surrounding traffic. If smoother traffic flow is achieved, it is expected that congestion will be alleviated. In extreme cases of total shutdown of the expressway, traffic could diverge at nearby ramps to avoid deterioration of congestion by DMS suggestions. As a conclusion, application of Big Data for better operation should emphasize real-time monitoring of traffic condition and a quick response based on the retrieved data.

How congestion affects the crash occurrence has been discussed in some existing literature. In aggregate safety analysis, the issue related to averaging congestion intensity might be the cause of the insignificant effects of congestion found in many crash frequency studies. In case of this study, it was verified that the congestion was highly localized and time specific. As a result, to gain better understanding whether congestion leads to more crashes, it was deemed better to be evaluated under real-time modeling framework. Big Data enables the restoration of traffic for each crash case. Moreover, rear-end crashes were selected as the target since their connection with congestion could be more straightforward. Both data mining and Bayesian statistics techniques were adopted to identify the leading contributing factors to crashes in real-time. The results concluded that peak hour, higher volume and lower speed at upstream locations, and high congestion index (CI) at downstream detection point significantly increased crash likelihood. Thus, direct (CI) and indirect (volume, speed) congestion indicators all support the assumption that congestion has an impact on rear-end crashes.

Different from previous real-time traffic safety studies, the current work takes one step further by incorporating reliability analysis to determine the conditions at which it is appropriate to trigger safety warnings on the expressway. First-Order Reliability Method (FORM) model was constructed based on the real-time crash prediction model and the critical point of system CI was calculated. When the system reaches the critical point, it does not necessarily mean equal risk for each section. Accordingly, the CI value for each crash case was investigated. It was found that the average CI for peak hour crashes was equal to the congestion threshold, which suggested that when congestion is detected at a specific location, both congestion and safety warnings should be sent to motorists.

As a final effort of this study, we propose combined real-time monitoring of congestion and safety on urban expressways through the MVDS system. A strategy taking both system and local congestion into account is summarized in Table 8-9. The proposed framework is to highlight the association between congestion and safety. On urban expressways, improving either one would be beneficial to the other.

Table 8-9: Real-time monitoring of congestion and traffic safety

		Individual detector Congestion Index (CI)	
		< 0.2	≥ 0.2
System Congestion Index (CI)	< 0.075	Safe and no congestion	congestion warning
	≥ 0.075	Prepare for safety warning	congestion and safety Warning

Despite that this study conducted relatively comprehensive analyses on both congestion and safety, there are still plenty of room for further improvement: 1) the current work focuses on rear-end crashes only. However, Big Data applications could be easily extended to other crash types and severities. 2) multiple factors other than congestion could lead to crashes in real world. In this study, only traffic dynamics were included in the real-time models. However, to fully realize the power of Big Data, more data sources should be utilized. Real-time weather condition could be an important factor for expressway operation and safety as well. Future work should also make use of Big Data from other sources.

CHAPTER 9: BIG DATA APPLICATIONS IN POST-CRASH ANALYSIS FOR SAFETY AND EFFICIENCY IMPROVEMENT

9.1 Introduction

How to provide motorists with efficient and safe services is the principal concern for traffic engineers. Past decades have seen the development of high speed facilities and introduction of advanced Intelligent Transportation System (ITS) technologies to improve highway operation. In the meantime, safety campaigns including regulation, education, and scientific research have been carried out to bring down the losses associated with crashes. Although great efforts were made, issues regarding traffic safety and operation still remain hot topics for researchers. Extensive studies have been conducted to explore crash contributing factors and corresponding countermeasures to reduce crash occurrence. Conclusions from most existing literature have confirmed the relationship between traffic flow parameters and safety conditions. In the face of incidents, incident duration has also been examined by many researchers to reduce its impact on traffic operation.

In this study, one issue that is overlooked by both types of research discussed above is investigated. Different from analyses identifying factors leading to crash occurrence or factors affecting the incident duration, the objective of this study tries to answer the following questions: 1) do all the crashes cause congestion? 2) if not, what factors make the crashes' impact on congestion diverse? To achieve the goal, three urban expressways operated by Central Florida Expressway Authority (CFX) were evaluated. The expressways are toll roads connecting downtown Orlando and neighboring area, carrying both commuting and tourist traffic. For more accurate and effective traffic monitoring, the authority have installed Microwave Vehicle Detection System (MVDS) on the expressways. On the 75-mile network of interest, 275 MVDS detectors are deployed. These detectors monitor traffic flow continuously and archive the data at one-minute interval. Operational performance of the expressways then can be evaluated through the MVDS traffic data. In this study, real-time traffic data and the detailed information from crash reports were extracted for each crash case to identify the effects of crashes on traffic congestion. Both Bayesian binomial and multinomial logit models were utilized to identify the factors leading to those potential diverse effects.

These two types of studies aim at preventing traffic crash occurrence and limiting traffic incident duration. Yet there is still a need to provide a more comprehensive understanding of the safety-operation relationship. Traffic crashes pose much more hazard for motorists on the roadways and cause huge social-economic losses compared with other types of traffic incidents. The effects of crashes then should be examined in more details. Moreover, the effects of incidents on traffic flow could be distinct. In some cases, only traveling lanes or shoulders are blocked due to the incidents; in some cases both traveling lanes and shoulders are blocked during different phase of clearance; and in other cases neither is blocked. Therefore it is possible that the incident duration of a crash is different from the time duration it affects operation especially congestion. These issues serve as the motivation of this study. In this current study, real-time traffic information was introduced to illustrate the effects of crashes on traffic congestion. Individual crash reports were utilized to identify significant factors leading to these effects. Expected contributions from this paper are deeper insights into the operation-safety relationship and practical suggestions for allocation of the rescue resources.

9.2 Data Preparation

The three expressways under evaluation in this study are the same segments as in the real-time safety analysis. The MVDS data have been collected since July, 2013. Except for April, 2014 during which month the authority upgraded their system and did not archive the MVDS data, eleven months traffic data till June, 2014 were collected.

The crash data were downloaded from Signal Four Analytics database. For each crash case, the basic information (crash time, geocoded location, crash type, severity, vehicles involved, weather conditions, etc.) is incorporated in the crash report. During the studied time period, 838 crashes occurred on the mainline of the three expressways. The geocoded locations of crashes were used to match MVDS detectors to the crashes. As illustrated in Figure 9-1, to study the effects of crashes on congestion, the detector upstream to the crash location can reflect the traffic condition after crash occurrence. Consequently, the nearest upstream MVDS detector (U1) was assigned to each crash case. The traffic conditions 10 to 5 minutes prior to the reported time of crash and 0 to 5 minutes after the reported time were extracted. The 10 to 5 minutes instead of 5 to 0 minutes prior to crashes were selected to account for the possible delay between the real crash time and the time it is reported and recorded. Among the total 838 crashes, the real-time traffic data were successfully matched for 809 crashes and missing for the other 29 crashes. In the following analysis of the effects of crashes on congestion, the 809 crashes were used.

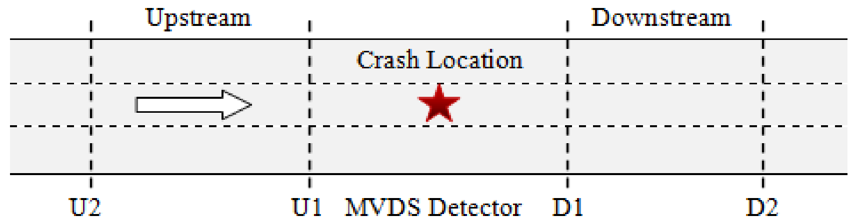
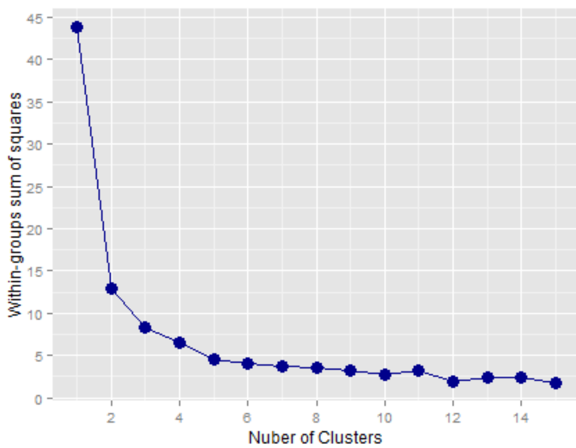
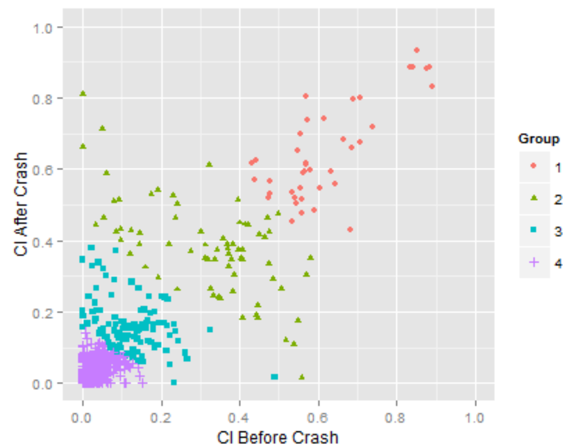


Figure 9-1: Crash Location and MVDS Detector Assignment

9.3 Crash Classification

How the crash alters the congestion status on the mainline can be identified by comparing the CIs before and after crashes. With the total 809 crashes, the patterns of before-after congestion conditions were evaluated first using clustering method. To partition the crashes into different clusters within which they share higher similarities, K-means clustering method was tested. K-means clustering method is a popular method for unsupervised classification. Several traffic safety studies have implemented this technique to group the crash data (Oltedal and Rundmo, 2007; Anderson, 2009; Xu *et al.,* 2012). In the K-means clustering analysis, the number of clusters has to be specified in advance. Therefore the selection of appropriate number of clusters will be crucial for the interpretation of the clustering results. The theoretical foundation of the method lies in that the sum of squares of the observations to their assigned cluster centers is a minimum (Hartigan and Wong, 1979). Figure 9-2 (a) was generated to show the total within-groups sum of squares under different number of clusters. The sharp decrease from 1 to 4 clusters and the relatively flat curve after 4 clusters suggest a 4-cluster solution. Figure 9-2 (b) shows that K-means clustering classified the crashes based on the congestion status before and after crashes. However, the clustering results do not differentiate the crash effects clearly. Within the same group, part of the crashes exhibit significant changes in before-after congestion intensity while others not. As a result, machine learning might be inappropriate for the purpose of this research and manual classification was applied.





(a) **(b)**

Figure 9-2: (a) K-mean Cluster Determination; (b) K-mean Cluster Results

In the scatter plot of Figure 9-2 (b), most of the dots representing the 809 crashes concentrate in the lower left corner. These dots indicate that the crashes occurred under non-congested conditions and did not lead to congestion afterwards. A proportion of the crashes were located along the 45-degree line in the higher part. Therefore they stand for those crashes happening under congestion. Nevertheless, their effects on traffic were very limited and did not worsen the congestion conditions. Another significant portion of crashes are in the upper left side of the figure. For these crashes, they occurred either under congested or non-congested conditions. The CIs after the crashes are much higher than CIs before crashes, which imply them as crashes that deteriorate the congestion on the mainline. One will also notice few crashes are located at the lower right side of the figure. Plain interpretation for these crashes should be that these crashes were observed under congestion conditions. Yet after the crash occurrences, the congestion intensity was relieved significantly. These cases will rarely exist in reality. Based on the above analysis, five clusters were manually created. Figure 9-3 illustrates how the Type 1 to Type 4 crashes were classified.

Considering crashes in the 5th cluster are illogical to common sense and they are only 46 cases out of the 809 crashes, investigation into this type of crashes wouldn't have practical meaning for improvement of emergency response. This 5th cluster was eliminated from further analysis. Figure 9-4 shows the manual clustering results. From the figure, we can answer our first question raised previously: crashes do not necessarily cause traffic congestions on the urban expressways. The contributing factors that make their impact diverse will be investigated using the information from individual crash reports in the following section.

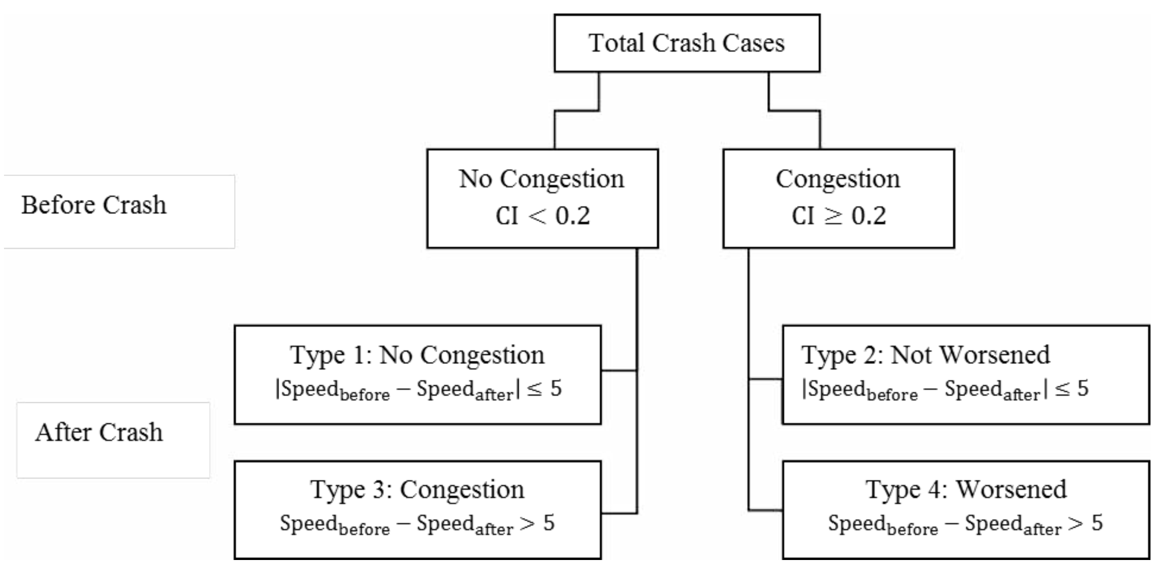


Figure 9-3: Crash Classification Procedure

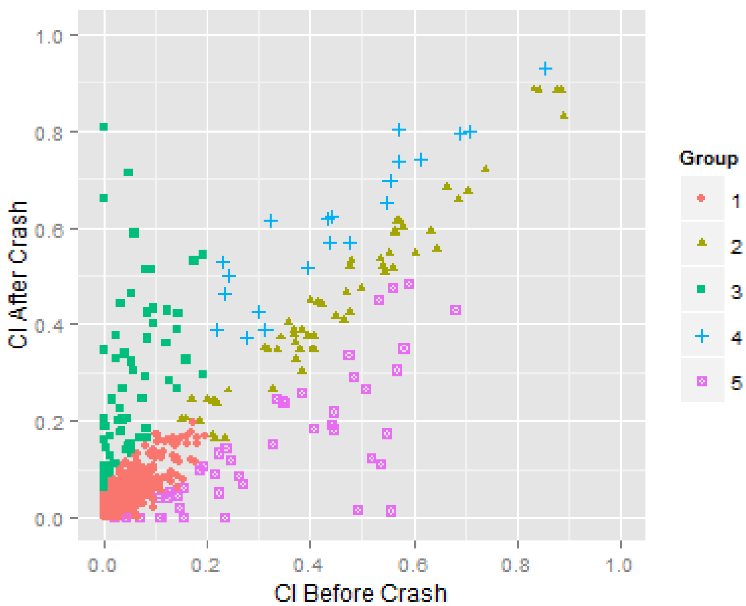


Figure 9-4: Crash Classification Based on the Effects of Crashes

9.4 Bayesian Logit Model

As discussed above, four types of crashes were classified. To statistically study their patterns and the contributing factors to each class, logit models were applied. Given that four types of crashes were involved, multinomial logit (MNL) models were considered. The MNL model was constructed in Bayesian framework. However, a concern raised regarding the independence from irrelevant alternatives (IIA) assumption in multinomial logit regression. IIA assumption means that the choice between two alternatives is unaffected by introduction of additional choices, which might not hold true in reality. To overcome the issue, nested logit models are suggested by researchers to relax the IIA assumption (Anderson *et al.,* 1992). Regarding the current work, Figure 9-3 shows the nested structure of the crash classification. However, traffic congestion conditions prior to crashes can be determined by the MVDS data and two separate binomial logit models conditional on prior crash conditions instead of nested logit model can be developed. The first binary model compares the crash effects under non-congested before crash condition (Type 1 vs. Type 3) while the second binary model compares their effects under congestion before condition (Type 2 vs. Type 4). The specifications of the models are generalized below:

Binary logit model

$$\pi(\mathrm{x}) = P(Y = 1|\mathbf{x}) = 1 - P(Y = 0|\mathbf{x}) \tag{8-2}$$

$$\log \frac{\pi(\mathrm{x})}{1-\pi(\mathrm{x})} = logit[\pi(\mathrm{x})] = \beta_0 + \boldsymbol{X\beta} \tag{8-3}$$

where Y is the binary response and $\boldsymbol{X}$ stands for the matrix of explanatory variables. β_0 and $\boldsymbol{\beta}$ are intercept and vector for parameter coefficient.

Multinomial logit model

$$\pi_j(\mathrm{x}) = P(Y = j|\mathbf{x}) \text{ with } \sum_j \pi_j(\mathbf{x}) = 1 \tag{8-4}$$

$$\log \frac{\pi_j(\mathrm{x})}{\pi_1(\mathrm{x})} = \beta_0 + \boldsymbol{X\beta},\ j = 2,3, \dots J \tag{8-5}$$

where Y is a categorical response with J categories ($J = 4$ in this study). The probability of each category sums to one. $\boldsymbol{X}$, β_0, and $\boldsymbol{\beta}$ bear the same meaning as in the binary model. In Bayesian inference, prior distributions for the parameters are required. In both the binary and multinomial logit models, non-informative priors $Normal(0, 10^3)$ are assigned to β_0 and $\boldsymbol{\beta}$. The models were built in WinBUGS software, 15000 iterations were run and the first 5000 were discarded as burn-in period. To ensure parameter convergence, three chains were simulated and the trace plots overlapped one another. The Deviance Information Criterion (DIC) was used as a Bayesian measure of model complexity and fit (Spiegelhalter *et al.,* 2003b). Bayesian Credible Interval (BCI) was used for parameter estimation. If the 95% BCI does not contain 0, then the effect of the variable is significant. Classifier performances of the logit models were evaluated using receiver operating characteristics (ROC).

9.5 Modeling Results and Discussion

9.5.1 Variable Description

To identify how the crashes affected the traffic congestion on the expressways. Information potentially pertinent to crash effects was extracted from the crash reports. The information could be broken down into four categories: spatial related factors, temporal related factors, crash related factors and weather related factors. Spatial related factors are expressway identifiers and number of lanes at the crash locations. Temporal related factors include peak-hour indicator, weekend indicator. Peak hours are defined as 7:00 to 9:00 and 17:00 to 19:00. Crash related factors are the number of vehicles involved in a crash and crash severity. Crashes involving four or more vehicles were rare and therefore they were combined with crashes involving three vehicles. On the three expressways, only two fatal crashes occurred during the study time period. In response, injury and fatal crashes were aggregated to become severe crashes against the property damage only (PDO) crashes. Weather related factors are the weather conditions recorded at the time of crash occurrences. Fog cases were few and they were combined with rainy conditions. All of the seven candidate variables are categorical variables as shown in Table 9-1. In Table 9-1, summary statistics of the crash frequencies for each category of the variables and the percentage of each type of crashes are provided.

Whether to incorporate all of the variables in Table 9-1 in the Bayesian logit models depends on the correlation between the variables. For categorical variables, the Pearson's Chi-square test was implemented. Of the seven candidate variables as shown in Table 9-2, peak hour indicator, crash severity, number of lanes and weather condition were identified to be independent from each other with $\mathrm{P-value} \geq 0.05$. They were included in the final logit models. Other three variables, namely expressway, weekend and vehicles involved were significantly correlated with some of the four variables above and did not outperform the four variables in the logit models. Therefore they were excluded.

Table 9-1: Statistics Summary of Variables

Crash Clusters	Type 1		Type 2		Type 3		Type 4		Total
Expressway									
SR 408	290	(76.7%)	49	(13.0%)	31	(8.2%)	8	(2.1%)	378
SR 417	169	(92.9%)	4	(2.2%)	7	(3.8%)	2	(1.1%)	182
SR 528	153	(75.4%)	9	(4.4%)	30	(14.8%)	11	(5.4%)	203
Peak Hour									
No	447	(88.9%)	16	(3.2%)	36	(7.2%)	4	(0.8%)	503
Yes	165	(63.5%)	46	(17.7%)	32	(12.3%)	17	(6.5%)	260
Weekend									
No	465	(77.8%)	58	(9.7%)	55	(9.2%)	20	(3.3%)	598
Yes	147	(89.1%)	4	(2.4%)	13	(7.9%)	1	(0.6%)	165
Number of Vehicles Involved									
1	178	(89.9%)	2	(1.0%)	17	(8.6%)	1	(0.5%)	198
2	367	(77.6%)	51	(10.8%)	41	(8.7%)	14	(3.0%)	473
3+	67	(72.8%)	9	(9.8%)	10	(10.9%)	6	(6.5%)	92
Number of Lanes									
2	333	(89.8%)	5	(1.3%)	29	(7.8%)	4	(1.1%)	371
3	174	(72.2%)	33	(13.7%)	25	(10.4%)	9	(3.7%)	241
4	65	(67.7%)	17	(17.7%)	7	(7.3%)	7	(7.3%)	96
5	40	(72.7%)	7	(12.7%)	7	(12.7%)	1	(1.8%)	55
Crash Severity									
PDO	421	(81.3%)	51	(9.8%)	35	(6.8%)	11	(2.1%)	518
Severe	191	(78.0%)	11	(4.5%)	33	(13.5%)	10	(4.1%)	245
Weather Condition									
Clear	398	(81.1%)	45	(9.2%)	37	(7.5%)	11	(2.2%)	491
Cloudy	111	(79.9%)	9	(6.5%)	13	(9.4%)	6	(4.3%)	139
Rain/Fog	103	(77.4%)	8	(6.0%)	18	(13.5%)	4	(3.0%)	133
Total	612	(80.2%)	62	(8.1%)	68	(8.9%)	21	(2.8%)	763

9.5.2 Multinomial Logit Model Results

Multinomial Bayesian logit model was evaluated first (as shown in Table 9-3). Type 1 crashes were set up as the baseline. Peak hours were found to significantly increase the logarithmic odds ratio of other types of crashes against Type 1 crashes. The effects of peak hours can be understood as during peak hours, recurrent congestion tends to occur. Therefore crashes under congested traffic flow are more likely to be observed during peak hours. In addition, even if no congestion exists prior to crashes, the higher traffic volume in peak hours will cause the impact of crashes to be more prominent.

The number of lanes at the crash location was also found positively related with the log odds ratio. From a demand-capacity point of view, number of lanes on expressway reflects the traffic demand on the segment. More lanes indicate higher traffic load and potential congestion. Thus the effects of number of lanes on Type 2 and Type 4 crashes are understandable. Type 3 crashes occurred under non-congested before crash conditions. If the cross section has only two lanes, single vehicle crashes are more likely to occur. The vehicles can move to shoulder or median after crashes and reduce their impact on the upstream traffic. In contrast, if the crashes occurred in the middle of a cross section with more than 2 lanes, the probability of multi-vehicle crashes would be higher and cause more severe delay before they could be moved out of the roadway.

Table 9-2: Pearson's Chi-square Correlation Test for Variables

Chi-square (P-value)	Expressway	Weekend	Peak Hour	Vehicles Involved	Crash Severity	Number of Lanes	Weather Condition
Expressway	--	4.041 (0.1326)	11.508 (0.0032)	42.238 (<.0001)	3.763 (0.1524)	148.933 (<.0001)	9.407 (0.0091)
Weekend		--	15.508 (<.0001)	14.556 (0.0001)	0.323 (0.5697)	0.123 (0.7262)	0.086 (0.7691)
Peak Hour			--	13.695 (0.0002)	1.927 (0.1650)	1.582 (0.2084)	2.820 (0.0931)
Vehicles Involved				--	0.018 (0.8936)	15.340 (<.0001)	0.110 (0.7397)
Crash Severity					--	0.1746 (0.6761)	1.686 (0.1942)
Number of Lanes						--	3.787 (0.0517)
Weather Condition							--

Table 9-3: Parameter Estimates and Model Fitting for Multinomial Logit Model

	$\log(P_{Type2}/P_{Type1})$			$\log(P_{Type3}/P_{Type1})$			$\log(P_{Type4}/P_{Type1})$		
	Mean	Std. Errors	95% BCI	Mean	Std. Errors	95% BCI	Mean	Std. Errors	95% BCI
Intercept	-5.379	0.564	(-6.497, -4.358)	-3.452	0.329	(-4.094, -2.823)	-7.026	0.854	(-8.932, -5.458)
Peak: Yes vs. No	2.252	0.332	(1.629, 2.956)	1.057	0.276	(0.484, 1.555)	2.755	0.618	(1.635, 4.069)
Lanes: 3 vs. 2	2.896	0.518	(1.971, 4.022)	0.631	0.308	(0.016, 1.241)	1.884	0.670	(0.735, 3.414)
Lanes: 4 vs. 2	2.970	0.559	(1.945, 4.134)	0.226	0.477	(-0.772, 1.106)	2.364	0.716	(1.081, 3.877)
Lanes: 5 vs. 2	2.609	0.673	(1.327, 3.977)	0.742	0.494	(-0.215, 1.672)	0.540	1.450	(-2.744, 3.027)
Weather: Cloudy vs Clear	-0.143	0.453	(-1.04, 0.700)	0.402	0.355	(-0.341, 1.067)	1.049	0.601	(-0.120, 2.234)*
Weather: Rain vs Clear	-0.246	0.445	(-1.145, 0.613)	0.664	0.327	(0.023, 1.282)	0.619	0.655	(-0.746, 1.791)
Severity: Severe vs PDO	-0.720	0.377	(-1.479, -0.004)	0.777	0.270	(0.269, 1.304)	0.839	0.500	(-0.193, 1.881)
Model Performance									
$\overline{D}$					907.893				
p_D					24.016				
DIC					931.909				
AUC					0.715				

*significant at 90% BCI

Weather conditions in the crash report have three categories. Compared with clear weather condition, cloudy weather won't significantly alter the log odds ratio of crashes. Rainy/fog conditions, nevertheless, will greatly increase the probability of congestion after crashes under non-congested conditions. Rain and fog can significantly impair drivers' visibility and the friction between pavement and tires. Once a crash occurs under these weather conditions, the severity level might be high and the adverse weather can extend the time needed to clear the scene. Under congestion, the adverse weather's effect is not significant since the speed of vehicles is expected to be low. Less severe crashes are expected under congested conditions. Thus the impact of adverse weather on congestion might be limited in this situation.

The crash severity also exhibits distinct effects on different crash types. If the traffic conditions before crashes are congested and the crashes do not worsen congestion, their severities would be much lower compared with Type 1 crashes. However, under non-congested conditions, if the crashes are severe, they have significant higher chance to result in congestion. One should understand that the speed under congested or non-congested traffic prior to crashes would mean quite different severity levels; and the severity levels will partially determine congestion status after crashes together with other factors.

9.5.3 Binomial Logit Model Results

The multinomial model in the above section provided relatively comprehensive and sound conclusions about the effects of crashes on congestion. However, the interpretation of the factors often involves differentiating congestion conditions prior to crashes first. To gain more clear understanding and relaxing the IIA assumption of the multinomial logit model, two separate binomial logit models based on the congestion status before crashes were constructed. Table 9-4 displays the modeling results.

Table 9-4: Parameter Estimates and Model Fitting for Separate Binomial Logit Model

	$\log(P_{Type3}/P_{Type1})$			$\log(P_{Type4}/P_{Type2})$		
	Mean	Std. Errors	95% BCI	Mean	Std. Errors	95% BCI
Intercept	-3.465	0.311	(-4.110, -2.901)	--	--	--
Peak: Yes vs. No	1.054	0.260	(0.558, 1.580)	--	--	--
Lanes: 3 vs. 2	0.634	0.315	(0.039, 1.261)	-2.013	0.517	(-3.087, -1.094)
Lanes: 4 vs. 2	0.160	0.470	(-0.864, 1.011)	-1.332	0.508	(-2.335, -0.403)
Lanes: 5 vs. 2	0.838	0.479	(-0.168, 1.755)*	-2.704	1.322	(-5.663, -0.645)
Weather: Cloudy vs Clear	0.340	0.363	(-0.394, 1.066)	--	--	--
Weather: Rain vs Clear	0.691	0.340	(0.037, 1.317)	--	--	--
Severity: Severe vs PDO	0.795	0.272	(0.272, 1.323)	1.587	0.601	(0.431, 2.781)
Model Performance						
$\bar{D}$	419.943			88.288		
p_D	8.101			3.938		
DIC	428.044			92.227		
AUC	0.686			0.728		

*significant at 90% BCI

Both Type 1 and Type 3 crashes had non-congested before crash conditions. The effects of the variables were the same as those found in the multinomial model. For Type 2 and Type 4 crashes which had the congested before crash conditions, the results shed some lights not revealed by the multinomial model. First of all, only the number of lanes and severity were found to significantly influence the probability of these two types of crashes. Since the congested traffic was mostly due to peak hour traffic before Type 2 and Type 4 crashes, the peak hour indicator would not play a crucial rule classifying these two crash types. For vehicles moving in congestion, speed has already been reduced. The effects of weather conditions on traffic flow parameters would be limited. The effects of number of lanes are worth elaboration. In contrast to the findings regarding Type 1 and Type 3 crashes, more lanes can efficiently reduce the impact of crashes under congested conditions. Crashes under congestion are more likely to involve multiple vehicles and block the traveling lanes. If the crash spot has more lanes, other vehicles can use adjacent lanes and avoid total shutdown of the mainline. The effects of crash severity do not differ for non-congested and congested conditions. The severe crashes will worsen the congestion conditions significantly in both cases.

Since the multinomial and binomial logit models employed different data, direct comparison via DIC is not appropriate. Area under the ROC Curve (AUC) was calculated to evaluate the performances of the models. The AUC values were all about 0.7, meaning the overall performances highly comparable. Both the multinomial and separate binomial models answered our second question regarding how the crashes could have distinct effects on congestion. According to the structure of crash classification, the separate binomial logit models generate results slightly easier for understanding and unveil distinct effects of the contributing factors on different types of crashes.

9.6 Conclusions

Traffic safety and operation are major indicators of highway performance. A large body of literature has investigated the operation-safety relationship. The current study focuses on one issue that was overlooked by previous studies: how crashes lay their impact on traffic congestion.

To answer the question, three expressways managed by CFX in Central Florida area were investigated. Detailed information from crash reports and real-time traffic data from MVDS system on the expressways were extracted.

The crashes were first clustered according to their effects on congestion. Traffic congestion status before and after each crash case were matched with the crash data. Machine Learning (K-means) method was tested to partition the data. However, the clustering results didn't offer reasonable insights into the crashes effects. As a solution, the crashes were manually classified. Four types of the effects of crashes were identified for further analysis. Based on the real-time traffic data, it was found that not all of the crashes would lead to congestion on the urban expressways. Both crashes occurring under congested or non-congested traffic flow could either increase the congestion intensity afterwards or exert insignificant influence on downstream traffic. To understand the distinct effects, information from crash reports were applied in statistical analysis.

Since the target of the statistical analysis is classification, logit models under Bayesian framework were constructed. Considering the structure used to cluster these crashes, both multinomial logit model and two separate binomial logit models were tested. Seven candidate variables that were possibly pertinent to crash effects were prepared. All of the candidate variables were categorical and Pearson' Chi-square test was conducted. Peak hour indicator, crash severity, number of lanes, and weather conditions were retained in the logit models. All of the four variables were found significant in the multinomial model and the binomial model for uncongested conditions. The separate binomial models generated results easier for explanation.

Under non-congested before crash conditions, the peak hours suggest higher traffic load. The crashes during peak hours would also pose more significant impact on the traffic congestion. If the roadway experiences no congestion prior to crashes, then the more lanes the segment has, the higher probability that crashes occurring on it would lead to congestion. It could be explained as when a crash occurs in the middle lanes of a cross section, it is possible to involve multiple vehicles and hence cause traffic congestion. With non-congested traffic flow before crashes, adverse weather could significantly increase the probability of congestion after crashes. On the contrary, in the binomial model for congested before crash conditions, only the number of lanes and crash severity were significant. The effects of crash severity are the same as that in uncongested conditions. However, the number of lanes shows different impact. Under congestion, in the face of crash occurrence, more lanes suggest less probability of congestion. Traveling speed will be greatly reduced by congestion, and therefore the crash manner. It is expected that single-vehicle crashes due to driving error or distraction would reduce under the congested conditions while the probability of multi-vehicle crashes greatly increases. In this case, more traveling lanes imply that motorists can use alternative lanes and avoid total shutdown of the mainline. As a result, traffic authorities should be careful when they interpret the crash effects on safety. The traffic state prior to crashes should be taken into account.

Potential improvement for future emergency response strategies can be raised based on the findings of this research. First of all, the real-time ITS traffic data should be incorporated in the response procedure. To estimate if the reported crashes would deteriorate congestion, current traffic condition at the crash site should be referred to. Second, the time of the crash, weather conditions at that time and the geometric characteristics of the crash location should all be considered. More police patrols might be helpful during peak hours or under adverse weather conditions. As for the segments with multiple lanes, their effects should be evaluated based on congestion levels. Last but not least, the necessity to report the potential severity of a crash has been confirmed. The quicker response for severe crashes can effectively diminish their effects on congestion.

CHAPTER 10: CONCLUSIONS

In the 21st century, data have become the foundation of scientific research and industrial operation in numerous fields. Intelligent Transportation Systems (ITS) traffic detection technologies on urban expressways have generated the Big Data that help the operators understand their system at the microscopic level. Researchers are also able to explore the efficient utilization of these data. In this dissertation, research aiming at traffic efficiency and safety evaluation and improvement using the Big Data for urban expressways were carried out. The following part of this chapter will summarize the efforts to achieve the objective, findings and implications based on the findings.

10.1 Summary of Research and Findings

The objective of this dissertation is to explore the Big Data applications in urban expressway efficiency and safety evaluation and improvement. The evaluation was conducted separately for efficiency and safety. Based on the evaluation results, corresponding improvement suggestions have been made. Nevertheless, considering efficiency and safety do not stand independently from each other, in the safety evaluation part, the relationship between these two elements was studied.

In the dissertation, the urban expressways operated and maintained by Central Florida Expressway Authority (CFX) were adopted as the data source and application. The expressway network consists of five expressways covering the downtown Orlando and neighboring areas, serving both commuting and leisure trips. In total the system reaches 109 miles. On the studied system, multiple ITS infrastructures can be found. The authority collects tolls using the combination of open tolling and cash lanes. Therefore the expressways have been equipped with Automatic Vehicle Identification (AVI) system for Electronic Toll Collection. This system could communicate with individual vehicles through the sensor and toll tags installed in the vehicle. Thus it has the ability to identify the unique ID of the vehicles using tags. Based on their readings at different locations, the AVI system could calculate the travel time on a specific segment and derive the space mean speed. However, it lacks the ability to reflect traffic volume on the road. Since 2012, the authority has begun to introduce Microwave Vehicle Detection System (MVDS). The system is specifically designed for traffic monitoring. They archive the detailed traffic flow parameters at installed locations and return the data on one-minute interval basis. Compared with the AVI system, the MVDS has much higher deployment density on the expressways and the traffic data is aggregated at lane level instead of a cross-section level. These two systems served as one of the sources of Big Data in this dissertation.

In addition, the expressways also have more than thirty Dynamic Message Signs (DMS) installed at different locations. These DMS function as a real-time medium to convey the most up-to-date traveling information to drivers. Most of the messages displayed on the boards are related to estimated travel time. However, in case of congestion, incidents and other events, they could also deliver relevant information to drivers. This real-time functionality makes DMS among the ideal tools for safety and efficiency improvement.

To fulfill all the objectives proposed, other data such as geometric characteristics, crashes and Geographic Information System (GIS) were collected and processed for analysis as well.

In the efficiency evaluation, congestion conditions on the expressways were examined. For this purpose, appropriate congestion measures that could implement the traffic detection data were first proposed. In the end, three types of congestion measures, time-based, speed-based and density-based congestion measures were introduced. Travel Time Index is a time-based congestion measure that could be derived using the AVI data. They measure the extra travel time needed to finish a trip on a specific segment during peak hours compared with that under free flow conditions. Congestion Index indicates the reduction in speed caused by congestion. Occupancy is a surrogate measure of traffic density, defined as the percentage of time a location is occupied by vehicles. Both Congestion Index and occupancy are derived from MVDS data. Comparisons were made between these congestion measures. It was found that all of the congestion measures could approximately reflect the same congestion conditions. Due to the recent update of AVI system on the CFX network, several segments suffer from data missing problem. Also, there were some discrepancies about the congestion time reflected by AVI and MVDS system. The MVDS system performs relatively more stable than the AVI system. Furthermore, since MVDS system is deployed with higher density, multiple MVDS sensors are normally installed within one AVI segment, therefore reflect the congestion condition with more detail.

Based on the MVDS congestion measures, both mainline and ramp congestion conditions were evaluated. Significant trend was found that in the morning peak hours, the traffic is moving from neighboring areas to Orlando and the opposite direction in the evening peak hours. Congestion segments on each expressway were identified. The length of queue changes as the congestion intensity varies over time. Based on the maximum queue length on the mainline, the DMS upstream were identified. The identified DMS could be used for queue warning in peak hours. It is expected that by warning the motorists in advance when they approach the congested area could help smooth traffic flow, reduce speed variation and finally alleviate congestion and crash risks. For congested areas where no existing DMS were found, potential locations for future DMS were suggested. Since no explicit guidelines were found about placement of DMS, the area one to two miles upstream to the end of queue were used. For ramps, the congested ramps were identified, most of which are off-ramps. Off-ramp congestion could lead to queues backing up on the mainline thus disrupting the mainline traffic. The same procedure was implemented to identify the DMS for ramp queue warning.

Expressway safety evaluation was carried out from different angles. First, the aggregate analysis which focuses on the crash frequency was conducted. Then to exploit the advantages of the Big Data, real-time crash prediction combined with reliability analysis were explored. Last but not least, the real-time traffic data after crash occurrence was also extracted to identify the effects of crashes on expressway operation. Relevant factors that lead to these effects were identified. These efforts were aimed at efficient utilization of the Big Data for safety evaluation, also bearing an emphasis on exploring the relationship between congestion and safety.

Since only the AVI data were available at the beginning of this research, their application in aggregate analysis was tested. In the aggregate analysis, the expressway SR 408 which carries the heaviest traffic and experiences the most crashes in the system was under investigation. Traditionally, only the AVI data capped at speed limit were used in safety analysis probably out of safety consideration. Compared with the capped AVI data, the uncapped AVI data had significant advantages in that they could reflect real-world traveling speed. In addition, although AVI should not be used for volume detection, they could still reflect the volume trend on the expressways before other traffic detection systems become available on the system. This study made an effort by introducing the uncapped AVI data to see if they could outperform the capped AVI data in safety evaluation. Multilevel Bayesian framework was adopted to fulfill the objective. Based on the results, uncapped AVI data are suggested to be used in future safety evaluation. Geometric elements such as inside shoulder width, horizontal curvature and existence of auxiliary lanes were all found to have an impact on the crash frequency. On the studied urban expressway, higher volume, lower speed and higher speed variation were the significant traffic flow parameters that influenced the crash occurrence. The traffic states satisfying these conditions were commonly observed under congested conditions, implying congestion's effects on safety. Based on these findings, DMS should be considered for both efficiency and safety improvement in the future.

Since the MVDS system was introduced to CFX expressway network, safety evaluation was expected to be improved by the new data. The new system made precise traffic volume available.

Moreover, new congestion measures were developed based on MVDS data. Thus direct evaluation of congestion's effects on safety became possible. To properly evaluate the effects of traffic parameters on crash frequency, the presence of multicollinearity was checked and taken into consideration. The presence of multicollinearity can cause changes in the magnitude of coefficients, and even the significance and signs of parameters in more extreme cases. Ridge regression was introduced in this study to deal with multicollinearity. Three models were constructed using Bayesian ridge regression. The longer segment length indicates higher exposure and increases the crash frequency. On the segment with auxiliary lanes, more frequent merging and diverging behaviors are expected. These movements can raise the likelihood of crashes. Regarding the effects of traffic parameters, their influences on crash occurrences were better revealed by assigning them with random parameters taking into account the peak and non-peak hours. In the Poisson model with uncorrelated random effects, volume and congestion were found to be positively related to crash frequency. Nevertheless, in the random parameter models, traffic volume was only significant during non-peak hours and congestion was only significant during peak hours. These conclusions suggest the necessity to distinguish the time of day when investigating how congestion is related to traffic safety. Moreover, the correlated random effects model provides better model fitting and smaller standard variance of coefficients for traffic parameters.

In this study, Travel Time Index (TTI) from AVI data, occupancy and Congestion Index (CI) from MVDS data were developed and tested in the crash models. The multiple congestion measures were adopted in hope of more definite conclusion. It turned out that the three congestion measures have similar effects on crash frequency. Higher congestion intensity could increase the crash count during peak hours while they have no significant impact on crash frequency during the non-peak hours. Congestion Index was found to outperform the other two congestion measures based on the model fitting performance.

Real-time crash prediction was carried out following the crash frequency study. The real power of the Big Data is also better exhibited in disaggregate analysis since they can restore the traffic conditions prior to crashes that the traditional aggregate analysis couldn't. In this study, it was first verified that the congestion was highly localized and time specific based on Congestion Index. As a result, to gain better understanding whether congestion leads to more crashes, it was deemed better to be evaluated under real-time modeling framework. Rear-end crashes were selected as the target since their connection with congestion could be more straightforward. Crashes on three of the five expressways were used in the real-time study. The exclusion of the other two expressways was because only a small proportion of crashes on the network occurred on these two expressways. Both data mining and Bayesian statistics techniques were adopted to identify the leading contributing factors to crashes in real-time. The results concluded that peak hour, higher volume and lower speed at upstream locations, and high congestion index (CI) at downstream detection point significantly increased crash likelihood. Thus, direct (CI) and indirect (volume, speed) congestion indicators all support the assumption that congestion has an impact on rear-end crashes.

This work also took one step further by incorporating reliability analysis to determine the conditions at which it is appropriate to trigger safety warnings on the expressway. First-Order Reliability Method (FORM) model was constructed based on the real-time crash prediction model and the critical point of system CI was calculated. When the system reaches the critical point, it does not necessarily mean equal risk for each section. Accordingly, the CI value for each crash case was investigated. It was found that the average CI for peak hour crashes was equal to the congestion threshold, which suggested that when congestion is detected at a specific location, both congestion and safety warnings should be sent to motorists.

Existing studies using real-time traffic data might underline the importance to proactively prevent or at least lower the crash likelihood. However, crashes are still highly random in nature, they can be reduced by effective traffic management, but cannot be eliminated. In case that a crash has already occurred, real-time traffic data also helps researchers get insights into their effects on operation. Traffic conditions immediately before and after crash occurrence were extracted to identify if these crashes on expressways led to congestion afterwards. First of all, unsupervised learning method K-means clustering was used to classify the crashes based on the before-after-crash traffic conditions, but the results did not seem logical. Therefore a self-defined classification was used. Based on the results, individual crash information was used to construct the logit models under Bayesian framework for classification to see what factors led to these distinct effects. How the factors would affect the operation after crashes were better explained by dividing them into congested and non-congested traffic before crashes.

Even under non-congested pre-crash conditions, the same segment is still expected to have higher traffic load in peak hours than non-peak hours. The crashes during peak hours would also pose more significant impact on the traffic congestion. If the roadway experiences no congestion prior to crashes, then the more lanes the segment has, the higher probability that crashes occurring on it would lead to congestion. It could be explained as when a crash occurs in the middle lanes of a cross section, it is possible to involve multiple vehicles and hence cause traffic congestion. With non-congested traffic flow before crashes, adverse weather could significantly increase the probability of congestion after crashes. On the contrary, in the binomial model for congested before crash conditions, only the number of lanes and crash severity were significant. The effects of crash severity are the same as that in uncongested conditions. However, the number of lanes shows different impact. Under congestion, in the face of crash occurrence, more lanes suggest less probability of congestion. Traveling speed will be greatly reduced by congestion, and therefore the crash type and number of vehicles involved could also be affected. It is expected that single-vehicle (e.g., off-road) crashes due to driving error or distraction would reduce under the congested conditions while the probability of multi-vehicle (e.g., rear-end, sideswipe, etc.) crashes greatly increases. In this case, more traveling lanes imply that motorists can use alternative lanes and avoid total shutdown of the mainline. As a result, traffic authorities should be careful when they interpret the crash effects on safety. The traffic state prior to crashes should be taken into account.

10.2 Implications

This dissertation has several findings regarding the quality of the Big Data, their use in efficiency evaluation, future improvement of traffic operation through DMS, expressway safety evaluation using Big Data and safety improvement.

As for the quality of traffic detection data, different systems can provide different traffic parameters that could be used for efficient traffic operation. For toll expressways, current trend is open tolling, thus the Automatic Vehicle Identification (AVI) system is indispensable in most of the Electronic Toll Collection (ETC) system. Although they could only provide limited traffic information, the travel time and mean space speed are still valuable traffic parameters that most intrusive or non-intrusive detectors fail to provide. These data are extremely useful for travel time estimation. Compared with the AVI system, the microwave radar system, in this study the Microwave Vehicle Detection System (MVDS) system could offer much more detailed traffic information. Unlike the AVI detectors that do not keep records of traffic volume or occupancy, the MVDS sensors can have these functions. In addition, since they have the ability to detect the traffic by lane, when they are installed near ramp or toll plazas, they can reflect the ramp and toll plaza cash lane traffic conditions. Furthermore, compared with intrusive traffic detection system, MVDS also has the advantage that it does not interrupt traffic during installation and maintenance. Based on the performance of AVI and MVDS systems on CFX network, it seems that MVDS is more stable. But a general conclusion about the comparison should still be based on larger sample size. Nevertheless, it is safe to state that to have precise understanding about the traffic conditions on the toll expressways, ITS systems specifically designed for traffic monitoring such as the MVDS system in addition to AVI system could be significantly beneficial.

Since the traffic detection systems return the traffic flow parameters in real-time manner, congestion measurement could be conducted at small time interval, for instance 5-minute or even 1-minute in case the precision and reliability of readings guaranteed. Not only congestion on mainline, but also on ramps could be identified. It should be noted that congestion is a quite dynamic process, the congestion intensity changes quickly in spatial and temporal dimensions. Real-time monitoring is therefore of significant importance if operators would like to be proactive in traffic management. DMS is an ideal tool for congestion improvement, since they have the capability to reflect what the traffic sensors detect on the expressways to the road users in real-time. Appropriate allocation of the DMS is crucial. On one hand, the DMS should leave motorists ample time to make decisions after they read the messages. On the other hand, the DMS shouldn't be too far from the locations about which the information is displayed so drivers won't forget the information. Currently no explicit guidelines exist. Therefore in case of this study, if no existing DMS sign is available for queue warning, one to two miles upstream the end of queue are proposed.

As for the safety evaluation, aggregate analysis indicated several geometric characteristics playing an important role in crash occurrence. The length of studied segment is an exposure variable. With other conditions held the same, the longer the segment length, the more crashes are likely to be observed. The existence of auxiliary lanes also affects the crash likelihood on expressways. On mainline segments with auxiliary lanes, more lane-changing maneuvers are expected. The turbulence in traffic flow can lead to speed variation thus increasing the likelihood of rear-end crashes and sideswipe crashes. Horizontal curvature was found to negatively affect crash risks. While the effects might be contradictory to people's perception of curve's effect on vehicles, it could be interpreted from the angle of drivers' reaction when the drive on curves. Drivers tend to be more cautious when they drive on curves compared with on straight sections. Wider inside shoulder tends to reduce crashes since they provide additional space for motorists. In cases that vehicles experience rotating, fishtail or skidding, the wider shoulders provide motorist spaces to regain the control of their vehicles or at least space for vehicles to rest on.

For the effects of traffic flow on safety, the use of uncapped AVI data is suggested to replace the capped AVI data. The uncapped AVI data reflect traveling speed on expressways closer to reality compared with capped AVI data. By different congestion measures, congestion was al found a significant factor affecting traffic safety during peak hours, while for off-peak time period other factors such as adverse weather conditions, human errors or debris might be responsible for crash occurrence.

The real-time crash analysis focuses on rear-end crashes. Congestion again was confirmed to be the influencing factors of crashes in real-time. For real-time safety improvement, the reliability analysis was introduced. They have the advantage to identify the states when the system would face failure, or crash in case of this study. Of course the ideal implementation of reliability analysis is to apply them at each detection location. However, since hundreds of detectors are deployed on the system, for this current work the calculation for each station is not impossible. Nevertheless, even applying this method at the system level shed some light into future improvement. When the system reaches certain states, safety conditions at individual locations that experience congestion should be scrutinized carefully. DMS upstream could be utilized to advise drivers to slow down or be prepared for slow traffic. The timely warning could potentially bring down the rear-end crash likelihood.

Besides real-time crash prediction, efficient clearance of crash site would also help improve expressway efficiency and safety. With limited resources, one priority is to target the crashes that have higher propensity to cause traffic delays on the expressways. Peak hour, number of lanes on the segment, crash severity, and weather conditions were found to significantly affect the crash effects, but their effects were also impacted by congestion conditions prior to crashes.

This dissertation discusses in detail the applications of urban expressway efficiency and safety evaluation and improvement using Big Data. However, the applications of Big Data are not limited to efficiency and safety. They could also be used for travel time estimation, especially considering multiple ITS systems are installed on many expressways. The fusion of data will be a key component in the process for better travel time algorithm. Also, the traffic detection data have great expectations in traffic simulation. Use of real-world traffic data could help simulation environment reflect reality as much as possible. Moreover, improvement could be first tested in the simulation at relatively small cost and be evaluated before they are put into practice.

In conclusion, the coming of Big Data era has begun to transform the outlook of traffic studies. While the topics discussed in this dissertation endeavor to show their great potentials, future work will always be needed to explore new sources of data and their applications for a better transportation system.

APPENDIX A: AVI SYSTEM SEGMENTATION

Table A-1: SR 408 Eastbound AVI System Segmentation

Link	Up_station	Down_station	Up_milepost	Down_milepost	Direction
1	AVI-0408E-WBoundary	AVI-0408E-GoodHomesRd	1.456	2.155	EB
2	AVI-0408E-GoodHomesRd	AVI-0408E-HiawasseeRd	2.155	3.468	EB
3	AVI-0408E-HiawasseeRd	AVI-0408E-KirkmanRd	3.468	4.632	EB
4	AVI-0408E-KirkmanRd	AVI-0408E-PineHillsRd	4.632	5.937	EB
5	AVI-0408E-PineHillsRd	AVI-0408E-JYP	5.937	7.619	EB
6	AVI-0408E-JYP	AVI-0408E-TampaAve	7.619	8.338	EB
7	AVI-0408E-TampaAve	AVI-0408E-OBT	8.338	8.859	EB
8	AVI-0408E-OBT	AVI-0408E-I4	8.859	9.347	EB
9	AVI-0408E-OBT	AVI-0408E-I4_Ramp	8.859	9.347	EB
10	AVI-0408E-I4	AVI-0408E-Orange_Ramp	9.347	10.191	EB
11	AVI-0408E-Orange_Ramp	AVI-0408E-MillsAve	10.191	10.809	EB
12	AVI-0408E-MillsAve	AVI-0408E-MillsAve_DMS	10.809	11.099	EB
13	AVI-0408E-MillsAve_DMS	AVI-0408E-BumbyAve	11.099	11.447	EB
14	AVI-0408E-BumbyAve	AVI-0408E-ConwayRd	11.447	12.841	EB
15	AVI-0408E-ConwayRd	AVI-0408E-SemoranBlvd	12.841	13.328	EB
16	AVI-0408E-SemoranBlvd	AVI-0408E-Goldnrd_DMS	13.328	15.181	EB
17	AVI-0408E-Goldnrd_DMS	AVI-0408E-GoldenrodRd	15.181	15.614	EB
18	AVI-0408E-GoldenrodRd	AVI-0408E-ChickasawTr	15.614	16.416	EB
19	AVI-0408E-ChickasawTr	AVI-0408E-EconTr	16.416	17.966	EB
20	AVI-0408E-EconTr	AVI-0408E-DeanRd	17.966	18.472	EB
21	AVI-0408E-DeanRd	AVI-0408E-RouseRd	18.472	19.424	EB
22	AVI-0408E-RouseRd	AVI-0408E-AlafayaTr	19.424	20.669	EB
23	AVI-0408E-AlafayaTr	AVI-0408E-EColonialDr	20.669	22.266	EB

Table A-2: SR 408 Westbound AVI System Segmentation

Link	Up_station	Down_station	Up_milepost	Down_milepost	Direction
1	AVI-0408W-GoodHomesRd	AVI-0408W-WColonl_Ramp	2.2362	1.0522	WB
2	AVI-0408W-GoodHomesRd	AVI-0408W-WBoundary	2.2362	1.5222	WB
3	AVI-0408W-HiawasseeRd	AVI-0408W-GoodHomesRd	4.5232	2.2362	WB
4	AVI-0408W-Kirkman_DMS	AVI-0408W-HiawasseeRd	4.8552	4.5232	WB
5	AVI-0408W-KirkmanRd	AVI-0408W-Kirkman_DMS	5.4092	4.8552	WB
6	AVI-0408W-PineHillsRd	AVI-0408W-KirkmanRd	6.008	5.4092	WB
7	AVI-0408W-FergusonRd	AVI-0408W-PineHillsRd	7.347	6.008	WB
8	AVI-0408W-JYP	AVI-0408W-FergusonRd	8.072	7.347	WB
9	AVI-0408W-OBT	AVI-0408W-JYP	9.298	8.072	WB
10	AVI-0408W-I4	AVI-0408W-OBT	10.488	9.298	WB
11	AVI-0408W-SummerlinAve	AVI-0408W-I4	10.964	10.488	WB
12	AVI-0408W-SummerlinAve	AVI-0408W-I4_Ramp	10.964	10.488	WB
13	AVI-0408W-MillsAve	AVI-0408W-SummerlinAve	11.399	10.964	WB
14	AVI-0408W-BumbyAve	AVI-0408W-MillsAve	11.806	11.399	WB
15	AVI-0408W-CrystalLkDr	AVI-0408W-BumbyAve	12.605	11.806	WB
16	AVI-0408W-SemoranBlvd	AVI-0408W-CrystalLkDr	14.563	12.605	WB
17	AVI-0408W-OxalisAv_DMS	AVI-0408W-SemoranBlvd	15.245	14.563	WB
18	AVI-0408W-GoldenrodRd	AVI-0408W-OxalisAv_DMS	16.488	15.245	WB
19	AVI-0408W-SR417	AVI-0408W-GoldenrodRd	18.033	16.488	WB
20	AVI-0408W-DeanRd	AVI-0408W-SR417	18.538	18.033	WB
21	AVI-0408W-RouseRd	AVI-0408W-DeanRd	19.706	18.538	WB
22	AVI-0408W-AlafayaTr	AVI-0408W-RouseRd	20.815	19.706	WB
23	AVI-0408W-EColonialDr	AVI-0408W-AlafayaTr	22.331	20.815	WB

Table A-3: SR 414 Eastbound AVI System Segmentation

Link	Up_station	Down_station	Up_milepost	Down_milepost	Direction
1	AVI-0414E-MardenRd	AVI-0414E-KeeneRd	4.313	6.335	EB
2	AVI-0414E-KeeneRd	AVI-0414E-HiawasseeRd	6.335	7.263	EB
3	AVI-0414E-HiawasseeRd	AVI-0414E-OBT	7.263	8.9	EB

Table A-4: SR 414 Westbound AVI System Segmentation

Link	Up_station	Down_station	Up_milepost	Down_milepost	Direction
1	AVI-0414W-KeeneRd	AVI-0429S-WestRd	6.62	-0.191	WB
2	AVI-0414W-KeeneDMS	AVI-0414W-KeeneRd	6.97	6.62	WB
3	AVI-0414W-HiawasseeRd	AVI-0414W-KeeneDMS	8.127	6.97	WB
4	AVI-0414W-OBT	AVI-0414W-HiawasseeRd	9.59	8.127	WB

Table A-5: SR 417 Northbound AVI System Segmentation

Link	Up_station	Down_station	Up_milepost	Down_milepost	Direction
1	AVI-0417N-IDrive_DMS	AVI-0417N-JYP	6.453	9.43	NB
2	AVI-0417N-JYP	AVI-0417N-OBT	9.43	10.638	NB
3	AVI-0417N-OBT	AVI-0417N-LandstarBlvd	10.638	13.004	NB
4	AVI-0417N-LandstarBlvd	AVI-0417N-BoggyCrk_DMS	13.004	14.653	NB
5	AVI-0417N-BoggyCrk_DMS	AVI-0417N-BoggyCreek	14.653	16.663	NB
6	AVI-0417N-BoggyCreek	AVI-0417N-LakeNona	16.663	18.82	NB
7	AVI-0417N-LakeNona	AVI-0417N-Narcooss_DMS	18.82	20.581	NB
8	AVI-0417N-Narcooss_DMS	AVI-0417N-Narcoossee	20.581	21.332	NB
9	AVI-0417N-Narcoossee	AVI-0417N-MossParkRd	21.332	22.519	NB
10	AVI-0417N-MossParkRd	AVI-0417N-InnovationWy	22.519	23.602	NB
11	AVI-0417N-InnovationWy	AVI-0417N-SR528	23.602	24.905	NB
12	AVI-0417N-SR528	AVI-0417N-LeeVistaBlvd	24.905	27.203	NB
13	AVI-0417N-LeeVistaBlvd	AVI-0417N-CurryFordRd	27.203	29.515	NB
14	AVI-0417N-CurryFordRd	AVI-0417N-SR408	29.515	32.495	NB
15	AVI-0417N-SR408	AVI-0417N-University	32.495	36.343	NB
16	AVI-0417N-University	AVI-0417N-Seminole_DMS	36.343	37.796	NB

Table A-6: SR 417 Southbound AVI System Segmentation

Link	Up_station	Down_station	Up_milepost	Down_milepost	Direction
1	AVI-0417S-JYP_DMS	AVI-0417S-IDrive_DMS	8.057	6.447	SB
2	AVI-0417S-JYP	AVI-0417S-JYP_DMS	10.287	8.057	SB
3	AVI-0417S-OBT	AVI-0417S-JYP	11.294	10.287	SB
4	AVI-0417S-LandstarBlvd	AVI-0417S-OBT	13.871	11.294	SB
5	AVI-0417S-Landstar_DMS	AVI-0417S-LandstarBlvd	14.648	13.871	SB
6	AVI-0417S-BoggyCreek	AVI-0417S-Landstar_DMS	17.746	14.648	SB
7	AVI-0417S-BoggyCrk_DMS	AVI-0417S-BoggyCreek	18.243	17.746	SB
8	AVI-0417S-LakeNona	AVI-0417S-BoggyCrk_DMS	19.564	18.243	SB
9	AVI-0417S-NarcoosseeRd	AVI-0417S-LakeNona	22.14	19.564	SB
10	AVI-0417S-MossParkRd	AVI-0417S-NarcoosseeRd	23.527	22.14	SB
11	AVI-0417S-InnovationWy	AVI-0417S-MossParkRd	24.461	23.527	SB
12	AVI-0417S-SR528	AVI-0417S-InnovationWy	26.321	24.461	SB
13	AVI-0417S-LeeVistaBlvd	AVI-0417S-SR528	27.924	26.321	SB
14	AVI-0417S-CurryFordRd	AVI-0417S-LeeVistaBlvd	30.241	27.924	SB
15	AVI-0417S-CuryFord_DMS	AVI-0417S-CurryFordRd	30.936	30.241	SB
16	AVI-0417S-SR408_E	AVI-0417S-CuryFord_DMS	32.709	30.936	SB
17	AVI-0417S-EColonial	AVI-0417S-SR408_E	34.899	32.709	SB
18	AVI-0417S-EColonl_DMS	AVI-0417S-EColonial	35.277	34.899	SB
19	AVI-0417S-University	AVI-0417S-EColonl_DMS	36.99	35.277	SB
20	AVI-0417S-Seminole_DMS	AVI-0417S-University	37.79	36.99	SB

Table A-7: SR 429 Northbound AVI System Segmentation

Link	Up_station	Down_station	Up_milepost	Down_milepost	Direction
10	AVI-0429N-SeidelRd_DMS	AVI-0429N-Independence	11.808	14.734	NB
11	AVI-0429N-Independence	AVI-0429N-CR535	14.734	19.005	NB
12	AVI-0429N-CR535	AVI-0429N-CR535_DMS	19.005	20.710	NB
13	AVI-0429N-CR535_DMS	AVI-0429N-Turnpike_S	20.710	21.792	NB
14	AVI-0429N-Turnpike_S	AVI-0429N-Turnpike_N	21.792	22.496	NB
15	AVI-0429N-Turnpike_N	AVI-0429N-SR438	22.496	23.668	NB
16	AVI-0429N-Turnpike_S	AVI-0408E-WBoundary	21.792	25.252	NB
17	AVI-0429N-Trnpike_Ramp	AVI-0429N-SR438	22.795	23.668	NB
18	AVI-0429N-SR438	AVI-0429N-PlantSt_DMS	23.668	24.972	NB
19	AVI-0429N-PlantSt_DMS	AVI-0429N-WestRd	24.972	26.421	NB

Table A-8: SR 429 Southbound AVI System Segmentation

Link	Up_station	Down_station	Up_milepost	Down_milepost	Direction
1	AVI-0429S-Independence	AVI-0429S-SeidelRd	15.454	11.808	SB
2	AVI-0429S-CR535	AVI-0429S-Independence	19.990	15.454	SB
3	AVI-0429S-CR535_DMS	AVI-0429S-CR535	20.710	19.990	SB
4	AVI-0429S-Turnpike_S	AVI-0429S-CR535_DMS	21.909	20.710	SB
5	AVI-0429S-Turnpike_N	AVI-0408E-WBoundary	22.927	18.971	SB
6	AVI-0429S-Turnpike_N	AVI-0429S-Turnpike_S	22.927	21.909	SB
7	AVI-0429S-WColonialDr	AVI-0429S-Turnpike_N	23.541	22.927	SB
8	AVI-0429S-SR438	AVI-0429S-WColonialDr	24.692	23.541	SB
9	AVI-0429S-WestRd	AVI-0429S-SR438	27.076	24.692	SB

Table A-9: SR 528 Eastbound AVI System Segmentation

Link	Up_station	Down_station	Up_milepost	Down_milepost	Direction
1	AVI-0528E-WBndry_DMS	AVI-0528E-TradeportDr	8.659	8.988	EB
2	AVI-0528E-TradeportDr	AVI-0528E-SR436	8.988	10.257	EB
3	AVI-0528E-SR436	AVI-0528E-GoldenrodRd	10.257	11.663	EB
4	AVI-0528E-GoldenrodRd	AVI-0528E-NarcoosseeRd	11.663	13.128	EB
5	AVI-0528E-NarcoosseeRd	AVI-0528E-SR417	13.128	15	EB
6	AVI-0528E-SR417	AVI-0528E-ICP	15	19.467	EB
7	AVI-0528E-ICP	AVI-0528E-DallasBlvd	19.467	23.519	EB
8	AVI-0528E-DallasBlvd	AVI-0528E-SR520	23.519	30.577	EB

Table A-10: SR 528 Westbound AVI System Segmentation

Link	Up_station	Down_station	Up_milepost	Down_milepost	Direction
1	AVI-0528W-TradeportDr	AVI-0528W-WBndry_DMS	9.82	8.501	WB
2	AVI-0528W-SR436_W	AVI-0528W-TradeportDr	10.806	9.82	WB
3	AVI-0528W-GoldenrodRd	AVI-0528W-SR436_W	12.475	10.806	WB
4	AVI-0528W-NarcoosseDMS	AVI-0528W-GoldenrodRd	14.019	12.475	WB
5	AVI-0528W-SR417	AVI-0528W-NarcoosseDMS	16.253	14.019	WB
6	AVI-0528W-ICP	AVI-0528W-SR417	20.139	16.253	WB
7	AVI-0528W-Dallas_DMS	AVI-0528W-ICP	23.246	20.139	WB
8	AVI-0528W-SR520	AVI-0528W-Dallas_DMS	30.843	23.246	WB
9	AVI-0528W-SR520_DMS	AVI-0528W-SR520	31.704	30.843	WB

APPENDIX B: MVDS SYSTEM AND LANE MANAGEMENT

Table B-1: SR 408 Eastbound MVDS System and Lane Management

Eastbound		Number of lanes			Eastbound		Number of lanes		
ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp	ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp
1	1.2			2	30	11.5	5		1
2	1.4	2			31	12.1	5		
3	1.7	2		2	32	12.5	5		1
4	2.2	3		1	33	12.9	5		2
5	2.4	3		1	34	13.3	5		2
6	2.7	3	2		35	13.7	3	3	
7	3.2	2	1		36	14.2	3	2	
8	3.6	2		1	37	14.5	4		
9	4.3	3		1	38	14.7	4		2
10	4.6	4			39	15	5		
11	4.9	3		1	40	15.7	4		2
12	5.3	3		1	41	15.8	4		1
13	6	3	2	1	42	16.1	4		1
14	6.4	3	1		43	16.5	5		
15	6.8	3			44	17.3	3		3
16	7	3		1	45	17.7	2		1
17	7.4	3			46	18	2		1
18	7.6	3		1	47	18.4	2		1
19	8	3		1	48	18.8	2		1
20	8.4	3		1	49	19	2	2	
21	8.9	3		1	50	19.4	2	1	
22	9.2	3		1	51	19.5	2		1
23	9.4	4		1	52	20.1	2		1
24	9.6	3		1	53	20.3	2		
25	9.7			1	54	20.8	2		1
26	10.3	3		1	55	21.8	2		
27	10.6	4		1	56	22.3	2		2
28	10.8	5		1	57	22.7	2		
29	11.2	5		1					

Table B-2: SR 408 Westbound MVDS System and Lane Management

Westbound		Number of lanes			Westbound		Number of lanes		
ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp	ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp
1	1.2			2	29	11.6	4		1
2	1.4	2			30	12.1	5		
3	1.6	3		2	31	12.6	5		2
4	2	3		1	32	13	5		1
5	2.4	3		1	33	13.3	3	2	
6	2.7	2	1		34	13.6	3	4	1
7	3.2	2	2		35	14.2	5		1
8	3.6	2		1	36	14.4	4		1
9	4.3	3		2	37	14.5	5		
10	4.6	4			38	15.2	5		
11	4.9	3		1	39	15.7	5		1
12	5.3	3		1	40	15.9	4		1
13	5.9	3	2	1	41	16.1	4		2
14	6.3	3	2		42	16.5	5		
15	6.8	3			43	17	3		2
16	7.3	3		1	44	17.8	3		1
17	7.4	4			45	18	3		1
18	7.6	3		1	46	18.4	2		1
19	8.1	3		1	47	18.8	2		1
20	8.4	3		1	48	19	2	1	
21	8.9	3		1	49	19.4	2	2	
22	9.2	3		1	50	19.7	3		1
23	9.7	3		1	51	19.9	2		1
24	9.9	2		2	52	20.7	3		
25	10.3	3		1	53	20.8	2		1
26	10.6	4			54	21.8	2		
27	10.9	4		2	55	22.3	2		1
28	11.3	5		1	56	22.7	2		1

Table B-3: SR 414 Eastbound MVDS System and Lane Management

Eastbound		Number of lanes		
ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp
1	3.6	2		3
2	3.8	2		2
3	4.3	3		1
4	4.9	3		
5	5.4	3	2	
6	5.8	3	1	
7	6.3	3		
8	6.6	3		1
9	7.2	3		
10	7.4	3		1
11	8.1	3		1
12	8.3	4		
13	8.9	3		2
14	9.3	2		1

Table B-4: SR 414 Westbound MVDS System and Lane Management

Westbound		Number of lanes		
ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp
1	3.8	2		3
2	4.3	3		1
3	4.9	3		
4	5.4	3	1	
5	5.8	3	2	
6	6.3	3		
7	6.6	3		1
8	7.2	3		
9	7.4	3		1
10	8.1	3		2
11	8.3	4		
12	9.2			2
13	9.3	3		1

Table B-5: SR 417 Northbound MVDS System and Lane Management

Northbound		Number of lanes			Northbound		Number of lanes		
ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp	ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp
1	5.9	2			29	23.9	2		1
2	6.2	2		2	30	24.5	2		
3	7.2	2	2		31	25	2		1
4	7.5	2	1		32	26.1	2		3
5	8.2	2			33	26.9	4		
6	9.4	2		1	34	27.3	3		2
7	10.1	2		1	35	27.9	3		1
8	10.6	2		1	36	28.1	3	2	
9	11	2		2	37	28.5	3	1	
10	12.2	2			38	28.7	4		
11	13.1	3		2	39	29.5	3		1
12	13.9	2		1	40	30.2	2		1
13	14.5	2			41	31.2	2		
14	15.2	4			42	31.9	2		
15	15.6	2	1		43	32.5	2		1
16	16.4	2			44	33			1
17	16.6	2		1	45	33.3	4		1
18	17.9	2		1	46	33.6	4		
19	18.2	2			47	34	3		2
20	18.8	2		1	48	34.6	3		1
21	19.3	2		1	49	35.2	3		
22	20.4	2			50	35.5	2	2	
23	20.9	2			51	36	3	2	
24	21.3	2		1	52	36.4	4		2
25	22	2		1	53	36.7	3		1
26	22.5	2		1	54	36.9	3		1
27	23	2		1	55	37.2	3		
28	23.6	2		1	56	37.7	2		

Table B-6: SR 417 Southbound MVDS System and Lane Management

Southbound		Number of lanes			Southbound		Number of lanes		
ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp	ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp
1	5.9	2			29	24.2	2		1
2	6.2	2		2	30	24.5	2		1
3	7.2	2	1		31	24.9	2		1
4	7.5	2	2		32	26.1	3		2
5	8.2	2			33	26.9	4		
6	9.4	2		1	34	27.3	3		1
7	10.3	2		2	35	27.9	4		1
8	10.7	2		1	36	28.1	3	1	
9	11.2	3		1	37	28.5	3	2	
10	12.2	2			38	28.7	4		
11	13.2	2		1	39	29.5	3		1
12	13.9	2		1	40	30.2	2		1
13	14.7	2			41	31.2	2		
14	15.2	3			42	31.9	2		
15	15.6	2	2		43	32.5	2		1
16	16.4	2			44	32.9			1
17	16.6	2		1	45	33.1	3		2
18	17.7	2		1	46	33.6	3		2
19	18.2	2			47	34.5	3		1
20	18.8	2		1	48	34.8	3		1
21	19.5	2		1	49	35.2	3		
22	20.4	2			50	35.5	2	1	
23	20.9	2			51	36	2	2	
24	21.3	2		1	52	36.4	3		1
25	22.2	2		1	53	36.7	2		1
26	23	3			54	37	2		1
27	23.2	2		1	55	37.2	2		
28	23.5	2		1	56	37.7	2		

Table B-7: SR 429 Northbound MVDS System and Lane Management

Northbound		Number of lanes			Northbound		Number of lanes		
ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp	ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp
1	10.9	2			16	22.8	2		1
2	13.7	2			17	23.2	2		2
3	14.6	2		1	18	23.6	3		2
4	15.3	2		1	19	24.5	2		1
5	16.1	2			20	24.7	2		
6	16.7	2	1		21	26	2		
7	17.2	2	1		22	26.3	2		1
8	18.4	2			23	26.8	2		1
9	18.9	2		1	24	27.7	2	2	
10	19.8	2		1	25	27.9	2	1	
11	20.4	2			26	28.9	2		1
12	21.7	2		1	27	29.4	2		1
13	21.9	2		2	28	29.6	3		
14	22.4	2		1	29	30	2		2
15	22.6			3					

Table B-8: SR 429 Southbound MVDS System and Lane Management

Southbound		Number of lanes			Southbound		Number of lanes		
ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp	ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp
1	10.9	2			16	22.7	2		2
2	11.5	2			17	23.2	2		1
3	14.6	2		1	18	24.2	2		1
4	15.3	2		1	19	24.5	3		
5	16.1	2			20	24.7	2		
6	16.7	2	1		21	26	2		
7	17.2	2	2		22	26.3	2		1
8	18.4	2			23	26.8	2		1
9	18.9	2		1	24	27.7	2	1	
10	19.8	2		1	25	27.9	2	2	
11	20.7	2			26	28.9	2		1
12	21.9	2		1	27	29.4	3		1
13	22.2			2	28	29.6	4		
14	22.4	2		1	29	29.8	2		2
15	22.5			2					

Table B-9: SR 528 Eastbound MVDS System and Lane Management

Eastbound		Number of lanes			Eastbound		Number of lanes		
ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp	ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp
1	8.5	3		1	16	15.3			3
2	9	3		1	17	15.7			2
3	9.5	3		1	18	15.9	2		2
4	9.8	3			19	16.6	2	2	
5	10.3	2		3	20	17.2	2	2	
6	10.7	2		1	21	19.5	2		1
7	10.8			1	22	20.2	2		1
8	11.1	2		1	23	23.2	2		
9	11.7	3		1	24	23.5	2		1
10	12.5	3		1	25	25.9	2	2	
11	12.8	4			26	26.3	2	1	
12	13.2	3		2	27	28.6	2		
13	13.8	2		1	28	30.6	2		1
14	14.5	2			29	31.9	2		
15	15	2		2					

Table B-10: SR 528 Westbound MVDS System and Lane Management

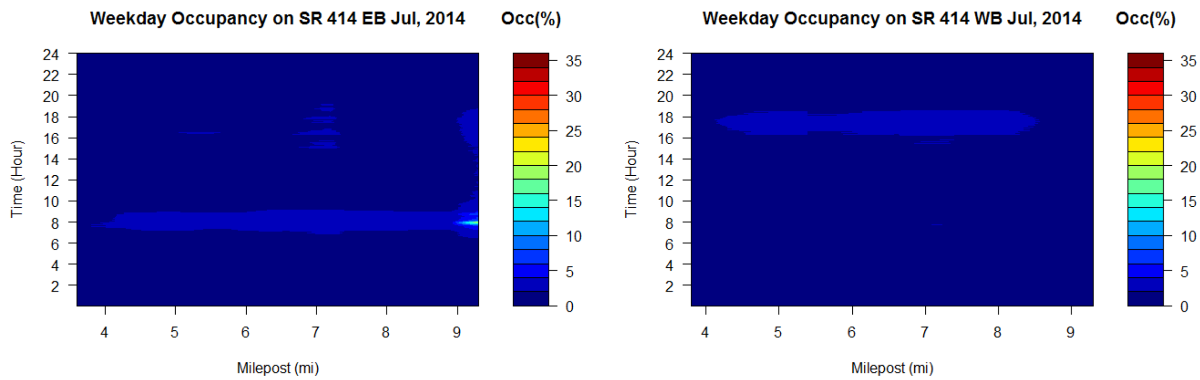
Westbound		Number of lanes			Westbound		Number of lanes		
ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp	ID	Milepost	Mainline (w/ TP Express)	TP Cash	Ramp
1	8.5	3		1	16	15.3	2		2
2	9.4	3		1	17	15.6	2		1
3	9.5	3		1	18	15.9	2		2
4	9.8	3			19	16.6	2	2	
5	10.3	4			20	17.2	2	2	
6	10.5	2		2	21	19.5	2		1
7	10.9	2		1	22	20.2	2		1
8	11	3		2	23	23.3	2		
9	12	4		1	24	23.5	2		1
10	12.2	3		1	25	25.9	2	1	
11	12.5	3		1	26	26.3	2	2	
12	12.8	3			27	28.6	2		
13	13.2	3		1	28	30.6	2		1
14	13.8	3		1	29	31.9	2		
15	14.5	2							

APPENDIX C: DMS LOCATIONS ON EXPRESSWAYS

Table C-1: DMS Locations on Expressway System

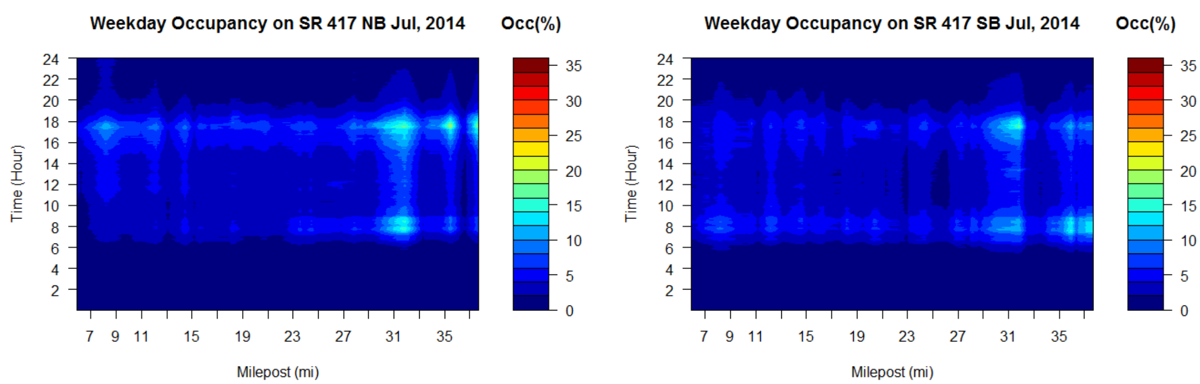
ID	DMS ID	Expressway	Direction	Milepost	Type
1	SR 408 EB @ MM 01.0 Ramp	SR 408	EB	1.0	Ramp
2	SR 408 EB @ MM 01.4	SR 408	EB	1.4	Mainline
3	SR 408 EB @ MM 04.4	SR 408	EB	4.4	Mainline
4	SR 408 EB @ MM 07.7	SR 408	EB	7.7	Mainline
5	SR 408 EB @ MM 11.1	SR 408	EB	11.1	Mainline
6	SR 408 EB @ MM 15.2	SR 408	EB	15.2	Mainline
7	SR 408 EB @ MM 20.6	SR 408	EB	20.6	Mainline
8	SR 408 WB @ MM 04.9	SR 408	WB	4.9	Mainline
9	SR 408 WB @ MM 09.3	SR 408	WB	9.3	Mainline
10	SR 408 WB @ MM 11.8	SR 408	WB	11.8	Mainline
11	SR 408 WB @ MM 15.2	SR 408	WB	15.2	Mainline
12	SR 408 WB @ MM 20.6	SR 408	WB	20.6	Mainline
13	SR 414 WB @ MM 09.6	SR 414	WB	9.6	Mainline
14	SR 417 NB @ MM 06.5	SR 417	NB	6.5	Mainline
15	SR 417 NB @ MM 14.7	SR 417	NB	14.7	Mainline
16	SR 417 NB @ MM 20.6	SR 417	NB	20.6	Mainline
17	SR 417 NB @ MM 27.0	SR 417	NB	27.0	Mainline
18	SR 417 NB @ MM 33.4	SR 417	NB	33.4	Mainline
19	SR 417 SB @ MM 08.1	SR 417	SB	8.1	Mainline
20	SR 417 SB @ MM 14.7	SR 417	SB	14.7	Mainline
21	SR 417 SB @ MM 18.2	SR 417	SB	18.2	Mainline
22	SR 417 SB @ MM 30.9	SR 417	SB	30.9	Mainline
23	SR 417 SB @ MM 35.3	SR 417	SB	35.3	Mainline
24	SR 417 SB @ MM 37.8	SR 417	SB	37.8	Mainline
25	SR 429 NB @ MM 11.8	SR 429	NB	11.8	Mainline
26	SR 429 NB @ MM 20.7	SR 429	NB	20.7	Mainline
27	SR 429 NB @ MM 25.0	SR 429	NB	25.0	Mainline
28	SR 429 SB @ MM 20.7	SR 429	SB	20.7	Mainline
29	SR 429 SB @ MM 25.0	SR 429	SB	25.0	Mainline
30	SR 451 SB @ MM 01.5	SR 451	SB	1.5	Mainline
31	SR 520 WB @ SR 528	SR 520	WB	--	Mainline
32	SR 528 EB @ MM 08.6	SR 528	EB	8.6	Mainline
33	SR 528 EB @ MM 11.8	SR 528	EB	11.8	Mainline
34	SR 528 EB @ MM 23.2	SR 528	EB	23.2	Mainline
35	SR 528 WB @ MM 14.0	SR 528	WB	14	Mainline
36	SR 528 WB @ MM 23.2	SR 528	WB	23.2	Mainline
37	SR 528 WB @ MM 31.7	SR 528	WB	31.7	Mainline

APPENDIX D: MAINLINE CONGESTION MEASUREMENT (1) OCCUPANCY



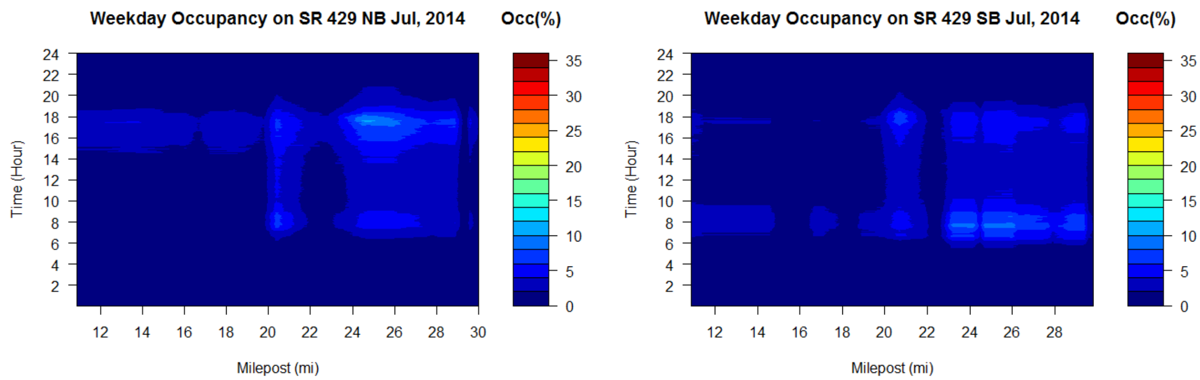
(a) (b)

Figure D-1: Mainline Weekday Occupancy of SR 414 (a) Eastbound and (b) Westbound



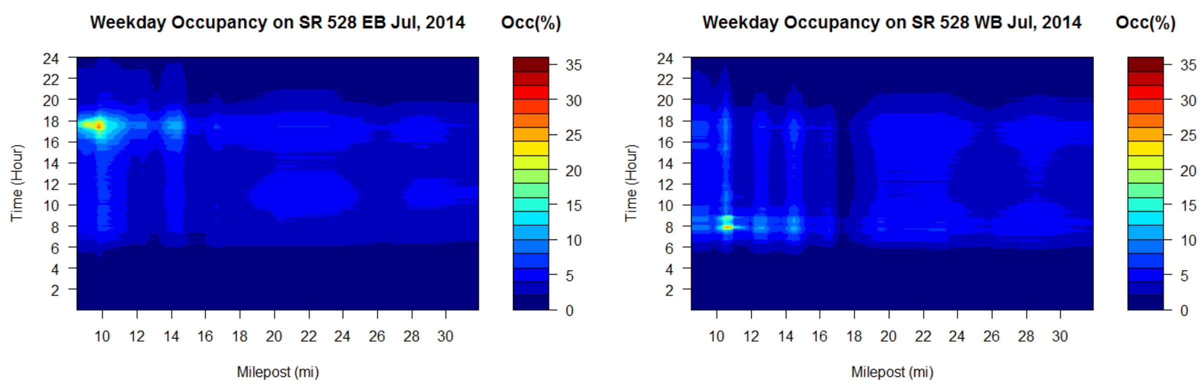
(a) (b)

Figure D-2: Mainline Weekday Occupancy of SR 417 (a) Northbound and (b) Southbound



(a) (b)

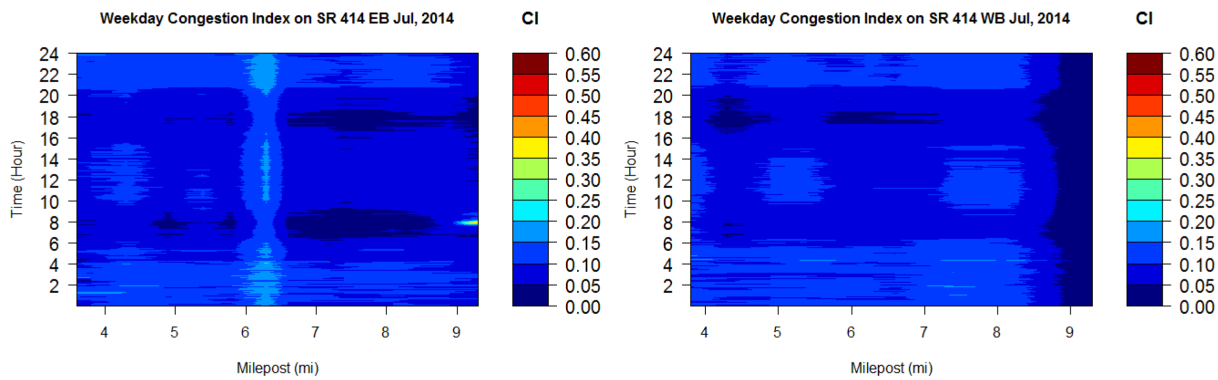
Figure D-3: Mainline Weekday Occupancy of SR 429 (a) Northbound and (b) Southbound



(a) (b)

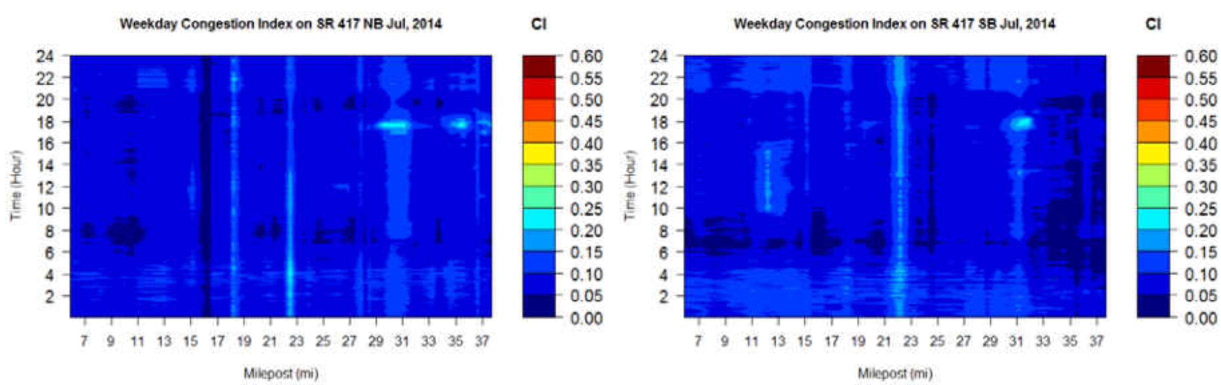
Figure D-4: Mainline Weekday Occupancy of SR 528 (a) Eastbound and (b) Westbound

APPENDIX E: MAINLINE CONGESTION MEASUREMENT (2) CONGESTION INDEX



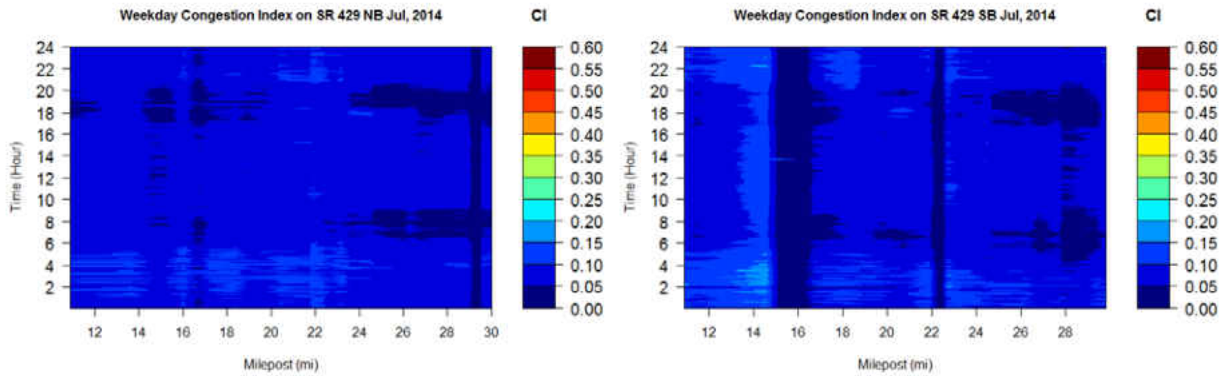
(a) (b)

Figure E-1: Mainline Weekday Congestion Index of SR 414 (a) Eastbound and (b) Westbound



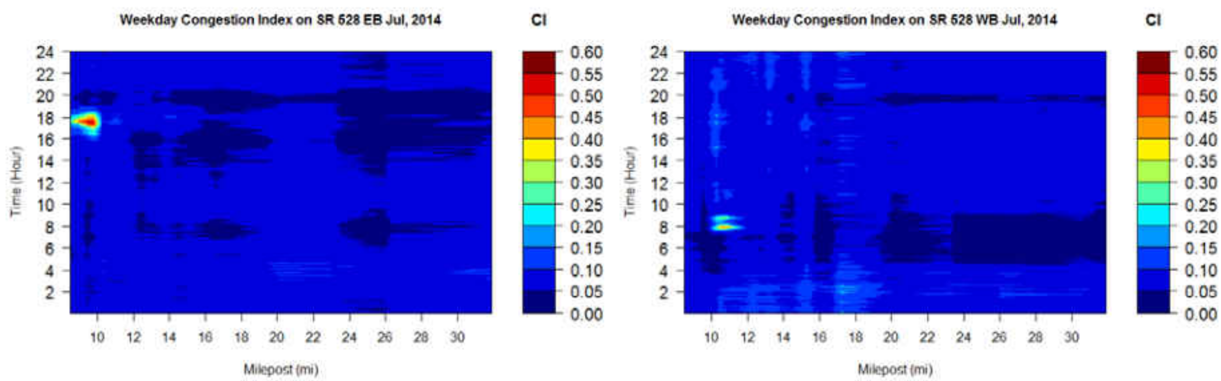
(a) (b)

Figure E-2: Mainline Weekday Congestion Index of SR 417 (a) Northbound and (b) Southbound



(a) (b)

Figure E-3: Mainline Weekday Congestion Index of SR 429 (a) Northbound and (b) Southbound



(a) (b)

Figure E-4: Mainline Weekday Congestion Index of SR 528 (a) Eastbound and (b) Westbound

LIST OF REFERENCES

Aarts, L., Van Schagen, I., 2006. Driving speed and the risk of road crashes: A review. Accident Analysis & Prevention 38 (2), 215-224.

Abbas, M., Mccoy, P., Year. Optimizing variable message sign locations on freeways using genetic algorithms. In: Proceedings of the 78th Annual Meeting of the Transportation Research Board, Washington, DC.

Abdel-Aty, M., Haleem, K., 2011. Analyzing angle crashes at unsignalized intersections using machine learning techniques. Accident Analysis & Prevention 43 (1), 461-470.

Abdel-Aty, M., Pande, A., Lee, C., Gayah, V., Cunningham, R., Dhindsa, A., Dilmore, J., 2007. Linking crash patterns to its-related archived data—phase ii, volume i: Real-time crash risk assessment models bd-550-5. Florida Dept. Transp., Tallahassee, FL.

Abdel-Aty, M., Uddin, N., Pande, A., 2005. Split models for predicting multivehicle crashes during high-speed and low-speed operating conditions on freeways. Transportation research record: journal of the transportation research board 1908 (1), 51-58.

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F., Hsia, L., 2004a. Predicting freeway crashes from loop detector data by matched case-control logistic regression. Transportation research record: journal of the transportation research board 1897 (-1), 88-95.

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F.M., Hsia, L., 2004b. Predicting freeway crashes from loop detector data by matched case-control logistic regression. Transportation Research Record: Journal of the Transportation Research Board 1897 (1), 88-95.

Abdel-Aty, M.A., Hassan, H.M., Ahmed, M., Al-Ghamdi, A.S., 2012a. Real-time prediction of visibility related crashes. Transportation Research Part C: Emerging Technologies 24 (0), 288-298.

Abdel-Aty, M.A., Hassan, H.M., Ahmed, M., Al-Ghamdi, A.S., 2012b. Real-time prediction of visibility related crashes. Transportation research part C: emerging technologies 24, 288-298.

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis & Prevention 32 (5), 633-642.

Ahmed, M., Abdel-Aty, M., 2013a. A data fusion framework for real-time risk assessment on freeways. Transportation Research Part C: Emerging Technologies 26 (0), 203-213.

Ahmed, M., Abdel-Aty, M., 2013b. A data fusion framework for real-time risk assessment on freeways. Transportation Research Part C: Emerging Technologies 26, 203-213.

Ahmed, M., Abdel-Aty, M., Yu, R., 2012a. Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data. Transportation research record: journal of the transportation research board 2280 (-1), 51-59.

Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011a. Exploring a bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accident Analysis & Prevention 43 (4), 1581-1589.

Ahmed, M., Yu, R., Abdel-Aty, M., Year. Safety applications of automatic vehicle identification and real-time weather data on freeways. In: Proceedings of the 18th ITS World Congress.

Ahmed, M.M., Abdel-Aty, M., Yu, R., 2012b. Bayesian updating approach for real-time safety evaluation with automatic vehicle identification data. Transportation Research Record: Journal of the Transportation Research Board 2280 (1), 60-67.

Ahmed, M.M., Abdel-Aty, M.A., 2012. The viability of using automatic vehicle identification data for real-time crash prediction. Intelligent Transportation Systems, IEEE Transactions on 13 (2), 459-468.

Akin, D., Sisiopiku, V.P., Skabardonis, A., 2011. Impacts of weather on traffic flow characteristics of urban freeways in istanbul. Procedia-Social and Behavioral Sciences 16, 89-99.

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis & Prevention 41 (1), 153-159.

Anastasopoulos, P.C., Mannering, F.L., 2011. An empirical assessment of fixed and random parameter logit models using crash-and non-crash-specific injury data. Accident Analysis & Prevention 43 (3), 1140-1147.

Anderson, S.P., De Palma, A., Thisse, J.F., 1992. Discrete choice theory of product differentiation MIT press.

Anderson, T.K., 2009. Kernel density estimation and k-means clustering to profile road accident hotspots. Accident Analysis & Prevention 41 (3), 359-364.

California manual on uniform traffic control devices, 2006.. In: Transportation, C.D.O. ed.

Guidelines for the use of dynamic message signs on the florida state highway system, 2008.. Florida Department of Transportation.

Interchange report, 2012.. Florida Department of Transportation.

Central florida greeneway, 2014a.. Central Florida Expressway Authority.

Daniel webster western beltway, 2014b.. Central Florida Expressway Authority.

John land apopka expressway, 2014c.. Central Florida Expressway Authority.

Martin b. Andersen beachline expressway, 2014d.. Central Florida Expressway Authority.

Antoniou, C., Balakrishna, R., Koutsopoulos, H.N., 2011. A synthesis of emerging data collection technologies and their impact on traffic management applications. European Transport Research Review 3 (3), 139-148.

Baruya, A., 1998. Speed-accident relationships on european roads. In: Proceedings of the 9th International Conference on Road Safety in Europe.

Benson, B.G., 1996. Motorist attitudes about content of variable-message signs. Transportation research record: journal of the transportation research board 1550 (1), 48-57.

Berry, W.D., Feldman, S., 1985. Multiple regression in practice Sage.

Beyer, M.A., Laney, D., 2012. The importance of ‘big data’: A definition. Stamford, CT: Gartner.

Breiman, L., 2006. Randomforest: Breiman and cutler’s random forests for classification and regression. URL http://stat-www. berkeley. edu/users/breiman/RandomForests, R package version.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: A practical information-theoretic approach Springer.

Chatterjee, S., Hadi, A.S., 2013. Regression analysis by example John Wiley & Sons.

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis & Prevention 35 (2), 253-259.

Chiu, Y.-C., Huynh, N., 2007. Location configuration design for dynamic message signs under stochastic incident and atis scenarios. Transportation Research Part C: Emerging Technologies 15 (1), 33-50.

Christidis, P., Rivas, J.N.I., 2012. Measuring road congestion. Institute for Prospective and Technological Studies, Joint Research Centre.

Christoforou, Z., Cohen, S., Karlaftis, M.G., 2011. Identifying crash type propensity using real-time traffic data on freeways. Journal of Safety Research 42 (1), 43-50.

Congdon, P., 2007. Bayesian statistical modelling John Wiley & Sons.

Crainiceanu, C., 2004. A short introduction to winbugs.

Dinu, R., Veeraragavan, A., 2011. Random parameter models for accident prediction on two-lane undivided highways in india. Journal of Safety Research 42 (1), 39-42.

Dudek, C.L., 2004. Changeable message sign operation and messaging handbook Federal Highway Administration, Operations Office of Travel Management.

El-Basyouny, K., Sayed, T., 2009. Accident prediction models with random corridor parameters. Accident Analysis & Prevention 41 (5), 1118-1123.

El-Basyouny, K., Sayed, T., 2011. A full bayes multivariate intervention model with random parameters among matched pairs for before–after safety evaluation. Accident Analysis & Prevention 43 (1), 87-94.

Elvik, R., Christensen, P., Amundsen, A., 2004. Speed and road accidents. An evaluation of the Power Model. TØI report 740, 2004.

Erke, A., Sagberg, F., Hagman, R., 2007. Effects of route guidance variable message signs (vms) on driver behaviour. Transportation Research Part F: Traffic Psychology and Behaviour 10 (6), 447-457.

Freund, R., Littell, R.J.F.R.C., Littell, R., 2000. Sas system for regression Sas Institute.

Fu, L., Henderson, J., Li, S., 2007. Locating changeable message signs for advanced traffic information and management systems. Canadian Journal of Civil Engineering 34 (5), 651-663.

Garber, N.J., Gadiraju, R., 1989. Factors affecting speed variance and its influence on accidents. Transportation Research Record (1213).

Gelman, A., Hill, J., 2006. Data analysis using regression and multilevel/hierarchical models Cambridge University Press.

Gerlough, D.L., Huber, M.J., 1975. Traffic flow theory: A monograph Transportation Research Board, National Research Council Washington, DC.

Golob, T., Recker, W., 2003a. Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. Journal of Transportation Engineering 129 (4), 342-353.

Golob, T.F., Recker, W., Pavlis, Y., 2008. Probabilistic models of freeway safety performance using traffic flow data as predictors. Safety Science 46 (9), 1306-1333.

Golob, T.F., Recker, W.W., 2003b. Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. Journal of Transportation Engineering 129 (4), 342-353.

Grant, M., Bowen, B., Day, M., Winick, R., Bauer, J., Chavis, A., 2011. Congestion management process: A guidebook. Washington, DC: United States Department of Transportation-Federal Highway Administration.

Greene, W.H., 2003. Econometric analysis Pearson Education India.

Griffin, L.A., 2011. Enhancing expressway operations using travel time performance data. In: Proceedings of the Facilities Management and Maintenance Workshop, Nashville, TN.

Guo, F., Wang, X., Abdel-Aty, M.A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. Accident Analysis & Prevention 42 (1), 84-92.

Hall, F.L., 1996. Traffic stream characteristics. Traffic Flow Theory. US Federal Highway Administration.

Hamad, K., Kikuchi, S., 2002. Developing a measure of traffic congestion: Fuzzy inference approach. Transportation Research Record: Journal of the Transportation Research Board 1802 (1), 77-85.

Hartigan, J.A., Wong, M.A., 1979. Algorithm as 136: A k-means clustering algorithm. Applied statistics, 100-108.

Hauer, E., 2009a. Speed and safety. Transportation research record: journal of the transportation research board 2103 (1), 10-17.

Hauer, E., 2009b. Speed and safety. Transportation Research Record: Journal of the Transportation Research Board 2103 (-1), 10-17.

Hauer, E., Hakkert, A., 1988. Extent and some implications of incomplete accident reporting. Transportation Research Record (1185).

Hauer, E., Harwood, D.W., Council, F.M., Griffith, M.S., 2002. Estimating safety by the empirical bayes method: A tutorial. Transportation research record: journal of the transportation research board 1784 (1), 126-131.

Haynes, R., Jones, A., Kennedy, V., Harvey, I., Jewell, T., 2007. District variations in road curvature in england and wales and their association with road-traffic crashes. Environment and Planning A 39 (5), 1222.

Haynes, R., Lake, I.R., Kingham, S., Sabel, C.E., Pearce, J., Barnett, R., 2008. The influence of road curvature on fatal crashes in new zealand. Accident Analysis & Prevention 40 (3), 843-850.

Hoerl, A.E., 1962. Application of ridge analysis to regression problems. Chemical Engineering Progress 58 (3), 54-59.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12 (1), 55-67.

Hosmer Jr, D.W., Lemeshow, S., 2004. Applied logistic regression John Wiley & Sons.

Hossain, M., Muromachi, Y., 2012. A bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. Accident Analysis & Prevention 45 (0), 373-381.

Hossain, M., Muromachi, Y., 2013. Understanding crash mechanism on urban expressways using high-resolution traffic data. Accident Analysis & Prevention 57 (0), 17-29.

Hourdos, J., Garg, V., Michalopoulos, P., Davis, G., 2006. Real-time detection of crash-prone conditions at freeway high-crash locations. Transportation research record: journal of the transportation research board 1968 (-1), 83-91.

Huang, H., Abdel-Aty, M., 2010. Multilevel data and bayesian analysis in traffic safety. Accident Analysis & Prevention 42 (6), 1556-1565.

Ivan, J.N., Wang, C., Bernardo, N.R., 2000. Explaining two-lane highway crash rates using land use and hourly exposure. Accident Analysis & Prevention 32 (6), 787-795.

Jianjun, W., Erxun, H.M.-T., Long, M., Ying, S., Ping, L., Year. Study on location of the variable message signs based on grey situation decision. In: Proceedings of the Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference on, pp. 523-526.

Joshua, S.C., Garber, N.J., 1990. Estimating truck accident rate and involvements using linear and poisson regression models. Transportation planning and Technology 15 (1), 41-58.

Jovanis, P.P., Chang, H.-L., 1986. Modeling the relationship of accidents to miles traveled. Transportation Research Record 1068, 42-51.

Kalatzis, A.E., Bassetto, C.F., Azzoni, C.R., 2011. Multicollinearity and financial constraint in investment decisions: A bayesian generalized ridge regression. Journal of Applied Statistics 38 (2), 287-299.

Klein, L.A., Mills, M.K., Gibson, D.R., 2006. Traffic detector handbook: -volume ii.

Klein, L.R., 1962. Introduction to econometrics.

Kockelman, K.M., Kweon, Y.-J., 2002. Driver injury severity: An application of ordered probit models. Accident Analysis & Prevention 34 (3), 313-321.

Kononov, J., Bailey, B., Allery, B.K., 2008. Relationships between safety and both congestion and number of lanes on urban freeways. Transportation research record: journal of the transportation research board 2083 (1), 26-39.

Kweon, Y.-J., Hartman, S.E., Lynn, C.W., 2009. Stop versus yield on pedestrian-involved fatal crashes in the united states. Accident Analysis & Prevention 41 (5), 1034-1039.

Laureshyn, A., Svensson, Å., Hydén, C., 2010. Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation. Accident Analysis & Prevention 42 (6), 1637-1646.

Lave, C.A., 1985. Speeding, coordination, and the 55 mph limit. The American Economic Review 75 (5), 1159-1164.

Le, T., Porter, R., 2012. Safety evaluation of geometric design criteria for spacing of entrance-exit ramp sequence and use of auxiliary lanes. Transportation research record: journal of the transportation research board 2309 (-1), 12-20.

Lee, C., Abdel-Aty, M., 2008. Two-level nested logit model to identify traffic flow parameters affecting crash occurrence on freeway ramps. Transportation Research Record: Journal of the Transportation Research Board 2083 (1), 145-152.

Lee, C., Abdel-Aty, M., Hsia, L., 2006. Potential real-time indicators of sideswipe crashes on freeways. Transportation Research Record: Journal of the Transportation Research Board 1953 (1), 41-49.

Li, H., Lü, Z., Yuan, X., 2008. Nataf transformation based point estimate method. Chinese Science Bulletin 53 (17), 2586-2592.

Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R news 2 (3), 18-22.

Lindley, D.V., Smith, A.F., 1972. Bayes estimates for the linear model. Journal of the Royal Statistical Society. Series B (Methodological), 1-41.

Lomax, T., Turner, S., Shunk, G., Levinson, H.S., Pratt, R.H., Bay, P.N., Douglas, G.B., 1997. Quantifying congestion. Volume 1: Final report.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice 44 (5), 291-305.

Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. Accident Analysis & Prevention 37 (1), 35-46.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. Winbugs-a bayesian modelling framework: Concepts, structure, and extensibility. Statistics and computing 10 (4), 325-337.

Luoma, J., Rämä, P., Penttinen, M., Anttila, V., 2000. Effects of variable message signs for slippery road conditions on reported driver behaviour. Transportation Research Part F: Traffic Psychology and Behaviour 3 (2), 75-84.

Martin, P.T., Feng, Y., Wang, X., 2003. Detector technology evaluation. Mountain-Plains Consortium.

Mazzoni, S., Mckenna, F., Scott, M.H., Fenves, G.L., 2006. Opensees command language manual. Pacific Earthquake Engineering Research (PEER) Center.

Mela, C.F., Kopalle, P.K., 2002. The impact of collinearity on regression analysis: The asymmetric effect of negative and positive correlations. Applied Economics 34 (6), 667-677.

Mensah, A., Hauer, E., 1998. Two problems of averaging arising in the estimation of the relationship between accidents and traffic flow. Transportation Research Record: Journal of the Transportation Research Board 1635 (1), 37-43.

Miaou, S.-P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Analysis & Prevention 26 (4), 471-482.

Miaou, S.-P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: Dispersion parameter, functional form, and bayes versus empirical bayes methods. Transportation Research Record: Journal of the Transportation Research Board 1840 (1), 31-40.

Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation 25 (4), 395-413.

Mimbela, L.E.Y., Klein, L.A., 2000. Summary of vehicle detection and surveillance technologies used in intelligent transportation systems.

Mirshahi, M., Obenberger, J.T., Fuhs, C.A., Howard, C.E., Krammes, R.A., Kuhn, B.T., Mayhew, R.M., Moore, M.A., Sahebjam, K., Stone, C.J., 2007. Active traffic management: The next step in congestion management.

National Research Council, T.R.B., 2010. Hcm 2010 : Highway capacity manual. Transportation Research Board, Washington, D.C.

Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W., 1996. Applied linear statistical models Irwin Chicago.

Nilsson, G., 2004. Traffic safety dimensions and the power model to describe the effect of speed on safety.

Noland, R.B., 2003a. Medical treatment and traffic fatality reductions in industrialized countries. Accident Analysis & Prevention 35 (6), 877-883.

Noland, R.B., 2003b. Traffic fatalities and injuries: The effect of changes in infrastructure and other trends. Accident Analysis & Prevention 35 (4), 599-611.

Noland, R.B., Karlaftis, M.G., 2005. Sensitivity of crash models to alternative specifications. Transportation Research Part E: Logistics and Transportation Review 41 (5), 439-458.

Ntzoufras, I., 2002. Gibbs variable selection using bugs. Journal of statistical software 7 (7), 1-19.

Ntzoufras, I., 2011. Bayesian modeling using winbugs John Wiley & Sons.

O'donnell, C.J., Connor, D.H., 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis & Prevention 28 (6), 739-753.

Oltedal, S., Rundmo, T., 2007. Using cluster analysis to test the cultural theory of risk perception. Transportation research part F: traffic psychology and behaviour 10 (3), 254-262.

Pande, A., Abdel-Aty, M., 2006. Assessment of freeway traffic parameters leading to lane-change related collisions. Accident Analysis & Prevention 38 (5), 936-948.

Pande, A., Das, A., Abdel-Aty, M., Hassan, H., 2011. Estimation of real-time crash risk. Transportation research record: journal of the transportation research board 2237 (-1), 60-66.

Park, B.-J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accident Analysis & Prevention 41 (4), 683-691.

Peeta, S., Ramos, J.L., Pasupathy, R., 2000. Content of variable message signs and on-line driver behavior. Transportation research record: journal of the transportation research board 1725 (1), 102-108.

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. Journal of Transportation Engineering 122 (2), 105-113.

Quddus, M.A., Wang, C., Ison, S.G., 2009. Road traffic congestion and crash severity: Econometric analysis using ordered response models. Journal of Transportation Engineering 136 (5), 424-435.

Rämä, P., Kulmala, R., 2000. Effects of variable message signs for slippery road conditions on driving speed and headways. Transportation Research Part F: Traffic Psychology and Behaviour 3 (2), 85-94.

Riley, J.D., 1999. Evaluation of travel time estimates derived from automatic vehicle identification tags in san antonio, tx. Virginia Polytechnic Institute and State University.

Schrank, D., Eisele, B., Lomax, T., 2012. Tti's 2012 urban mobility report. Texas A&M Transportation Institute. The Texas A&M University System.

Schrank, D.L., Lomax, T.J., 2007. The 2007 urban mobility report Texas Transportation Institute, Texas A & M University.

Shang, H.-Y., Huang, H.-J., Year. A cell transmission model and its application in optimizing the location of variable message signs. In: Proceedings of the Natural Computation, 2007. ICNC 2007. Third International Conference on, pp. 75-79.

Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. Journal of Safety Research 27 (3), 183-194.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis & Prevention 27 (3), 371-389.

Shefer, D., Rietveld, P., 1997. Congestion and safety on highways: Towards an analytical model. Urban Studies 34 (4), 679-692.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003a. Winbugs user manual. version.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003b. Winbugs user manual. Cambridge: MRC Biostatistics Unit.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC bioinformatics 9 (1), 307.

Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC bioinformatics 8 (1), 25.

Systematics, C., 2005. Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation Federal Highway Administration.

Tay, R., De Barros, A., 2010. Effectiveness of road safety messages on variable message signs. Journal of Transportation Systems Engineering and Information Technology 10 (3), 18-23.

Taylor, M., Baruya, A., Kennedy, J., Britain, G., 2002. The relationship between speed and accidents on rural single-carriageway roads.

Traffic, C., Manual on uniform traffic control devices.

Venkataraman, N., Ulfarsson, G., Shankar, V., Oh, J., Park, M., 2011a. Model of relationship between interstate crash occurrence and geometrics. Transportation research record: journal of the transportation research board 2236 (-1), 41-48.

Venkataraman, N., Ulfarsson, G.F., Shankar, V.N., 2013. Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type. Accident Analysis & Prevention 59, 309-318.

Venkataraman, N.S., Ulfarsson, G.F., Shankar, V., Oh, J., Park, M., 2011b. Model of relationship between interstate crash occurrence and geometrics. Transportation Research Record: Journal of the Transportation Research Board 2236 (1), 41-48.

Versavel, J., Year. Road safety through video detection. In: Proceedings of the Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on, pp. 753-757.

Wang, C., Quddus, M., Ison, S., 2013a. A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the uk. Transportmetrica A: Transport Science 9 (2), 124-148.

Wang, C., Quddus, M.A., Ison, S.G., 2009. Impact of traffic congestion on road accidents: A spatial analysis of the m25 motorway in england. Accident Analysis & Prevention 41 (4), 798-808.

Wang, C., Quddus, M.A., Ison, S.G., 2013b. The effect of traffic and road characteristics on road safety: A review and future research direction. Safety Science 57 (0), 264-275.

Wang, C., Quddus, M.A., Ison, S.G., 2013c. The effect of traffic and road characteristics on road safety: A review and future research direction. Safety science 57, 264-275.

Williams, G., 2011. Data mining with rattle and r: The art of excavating data for knowledge discovery Springer.

Xie, K., Wang, X., Huang, H., Chen, X., 2013. Corridor-level signalized intersection safety analysis in shanghai, china using bayesian hierarchical models. Accident Analysis & Prevention 50, 25-33.

Xu, C., Liu, P., Wang, W., Li, Z., 2012. Evaluation of the impacts of traffic states on crash risks on freeways. Accident Analysis & Prevention 47, 162-171.

Xu, C., Tarko, A.P., Wang, W., Liu, P., 2013a. Predicting crash likelihood and severity on freeways with real-time loop detector data. Accident Analysis & Prevention 57, 30-39.

Xu, C., Wang, W., Liu, P., 2013b. Identifying crash-prone traffic conditions under different weather on freeways. Journal of Safety Research 46 (0), 135-144.

Yu, R., Abdel-Aty, M., 2013a. Investigating the different characteristics of weekday and weekend crashes. Journal of Safety Research 46 (0), 91-97.

Yu, R., Abdel-Aty, M., 2013b. Multi-level bayesian analyses for single- and multi-vehicle freeway crashes. Accident Analysis & Prevention 58 (0), 97-105.

Yu, R., Abdel-Aty, M., 2013c. Utilizing support vector machine in real-time crash risk evaluation. Accident Analysis & Prevention 51 (0), 252-259.

Yu, R., Abdel-Aty, M., 2014. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. Safety Science 63, 50-56.

Yu, R., Abdel-Aty, M., Ahmed, M., 2013a. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis & Prevention 50 (0), 371-376.

Yu, R., Abdel-Aty, M., Ahmed, M., 2013b. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis & Prevention 50, 371-376.

Yu, R., Shi, Q., Abdel-Aty, M., 2013c. Feasibility of incorporating reliability analysis in traffic safety investigation. Transportation Research Record: Journal of the Transportation Research Board 2386 (-1), 35-41.

Zhang, G., Avery, R.P., Wang, Y., 2007. Video-based vehicle detection and classification system for real-time traffic data collection using uncalibrated video cameras. Transportation Research Record: Journal of the Transportation Research Board 1993 (1), 138-147.

Zhang, Y., Xie, Y., Li, L., 2012. Crash frequency analysis of different types of urban roadway segments using generalized additive model. Journal of Safety Research 43 (2), 107-114.

Zheng, Z., Ahn, S., Monsere, C.M., 2010. Impact of traffic oscillations on freeway crash occurrences. Accident Analysis & Prevention 42 (2), 626-636.