Florida International University

FIU Digital Commons

3-20-2020

Identification of Secondary Traffic Crashes and Recommended Countermeasures

Armana Sabiha Huq
Florida International University, ahuq002@fiu.edu

Recommended Citation

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

IDENTIFICATION OF SECONDARY TRAFFIC CRASHES AND RECOMMENDED

COUNTERMEASURES

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CIVIL ENGINEERING

by

Armana Sabiha Huq

2020

To: Dean John L. Volakis
     College of Engineering and Computing

This dissertation, written by Armana Sabiha Huq, and entitled Identification of Secondary Traffic Crashes and Recommended Countermeasures , having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Mohammed Hadi

_____
Atorod Azizinamini

_____
B M Golam Kibria

_____
Xia Jin , Major Professor

Date of Defense: March 20, 2020

The dissertation of Armana Sabiha Huq is approved.

_____
Dean John L. Volakis
College of Engineering and Computing

_____
Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2020

DEDICATION

To my parents Md Azizul Huq and Shahida Begum, my four sisters: Dr. Arifa,

Dr. Aysa, Tumpa, Aurin, my husband, Dr. Nazmul, my own and in-law relatives, my

friends, and colleagues. I am also dedicating this dissertation to all extremely supportive

faculties and administrative officers of the Florida International University, the

University Graduate School, the College of Engineering and Computing, the Department

of Civil and Environmental Engineering, the Office of International Student and Scholar

Services, the Division of Student Affairs office of the Vice President, especially Drs.

Lidia Kos, Anthony J. McGoron, Atorod Azizinamini, Nasar U. Ahmed, Alejandra Parra

and Mr. Carlton Ng, Ms. Claudia Biscardi . They are my source of strength and

inspiration. They have been motivating and helping me to reach my career goals.

ACKNOWLEDGMENTS

ABSTRACT OF THE DISSERTATION

IDENTIFICATION OF SECONDARY TRAFFIC CRASHES AND RECOMMENDED

COUNTERMEASURES

by

Armana Sabiha Huq

Florida International University, 2020

Miami, Florida

Professor Xia Jin, Major Professor

Secondary crashes (SCs) usually occur due to congestion or other prior incidents.

SCs are increasingly spotted as a significant issue in traffic operations, leading to reduced

capacity, extra traffic delays, increased fuel consumption, and additional emissions. SCs

have substantial impacts on traffic management resource allocation. One of the challenges

in the traffic safety area of the transportation industry is to determine an adequate method

for identifying SCs. The specific objectives of this study are: identification of SCs using

spatiotemporal criteria and exploring the contributing risk factors to the identified SCs.

Two different approaches were explored to identify SCs. The first approach is based

on a "static" method that employs a predefined 2 miles-2 hours fixed spatiotemporal

threshold. Four-year (2011 to 2014) crash and traffic data from the Crash Analysis

Reporting (CAR) system database were used. The linear referencing tool of Geographic

Information Systems (GIS) was applied to identify crashes that fell within the threshold.

About 1.49% of all crashes were identified as SCs. A Structural Equation Model (SEM)

was developed to investigate the contributing risk factors to the occurrence and severity

level of SCs. Model results revealed that a series of driver attributes contributed to the

occurrence of SCs, including the influence of alcohol or drug, inattentive driving, fatigue

or speeding. Other variables that might lead to higher probabilities of SCs include vehicle attributes (brake defects, motorcycles), roadway conditions (roadway surface, vision obstruction) and environmental factors (raining condition Given that about 40% of SCs were rear-end crashes, this study also examined contributing factors to severity levels of rear-end SCs. Results revealed that the presence of horizontal curves, presence of guardrail, and posted speed limit showed a significant influence on the severity level of SCs. Crash modification factors were also developed by considering the roadway and traffic characteristics.

In contrast to the static method, the dynamic approach identifies a dynamic spatiotemporal impact area for each primary incident using the Speed Contour Plot method. This analysis was explored using the Regional Integrated Transportation Information System (RITIS) and the SunGuide™ database for the year of 2014-2017. This study further analyzed contributing risk factors to SCs on I-95 and found that SCs were more likely to occur if primary incident clearance times were longer. It also revealed that SCs were more severe at night and on weekends. It implies that timely emergency responses would have a significant effect on mitigating SCs. These findings point to necessary strategies to mitigate SCs, including improved traffic management policies and implementation of advanced intelligent transportation warning systems.

One of the challenges in addressing SCs lies in the lack of quality databases (such as speed data and incident information) to appropriately identify and investigate SCs. Therefore, future efforts may focus on institute a framework that combines all levels of databases from multiple sources, which can help timely identification and investigation of SCs. This would lead to the development and implementation of efficient and effective countermeasures to mitigate SC and enhance safety.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND ACRONYMS

AADT	Annual Average Daily Traffic
AIC	Akaike Information Criterion
ASDA	Automatic Staudynamikanalyse
B/C	Benefit-cost
BIP	Binary Integer Programming
BLM	Binary Logistic Regression Model
BSCP	Binary Speed Contour Plot
CCTV	Closed Circuit Television Cameras
CMS	Changeable Message Signs
CV	Connected Vehicle
EMS	Emergency Medical Service
FDOT	Florida Department of Transportation
FHWA	Federal Highway Administration
FSP	Freeway Service Patrol
GIS	Geographic Information System
GORP	Generalized Ordered Response Probit
HCM	Highway Capacity Manual
HOV	High Occupancy Vehicle
INDOT	Indiana Department of Transportation
IPC	Incident Progression Curve
ITE	Institute of Transportation Engineers
ITS	Intelligent Transportation System

KDE	Kernel Density Estimation
MLSA	Multi-Layer Sensitivity Analysis
MLM	Multinomial Logit Model
MLP	Multi-Layer Perception
NB	Negative Binomial
NN	Neural Network
PC	Primary Crash
PDO	Property Damage Only
PI	Primary Incident
RCI	Roadway Characteristics Inventory
RSCP	Representative Speed Contour Plot
RTMC	Regional Traffic Management Center
SBSIF	Simulation Based Secondary Identification Filtering
SC	Secondary Crash
SCIA	Secondary Crash Identification Algorithm
SCM	Speed Contour Map
SCP	Speed Contour Plot
SEM	Structural Equation Modeling
STIA	Spatiotemporal Impact Area
SWBF	Shock Wave Boundary Filtering
SWT	Shockwave Theory
TDOT	Tennessee Department of Transportation
TIM	Traffic Incident Management
USDOT	United States Department of Transportation

CHAPTER 1

INTRODUCTION

1.1 Background

Traffic incidents are estimated to cause between 30-50% of the congestion problems on urban roadways (Skabardonis et al. 1995, Ozbay 1999, Kwon et al. 2006). Traffic incident or primary incident (PI) can be a crash or any type of incident such as disabled vehicles, debris on the roadway, emergency vehicles, police activity, vehicle fire, flooding, pedestrian, and so on. Traffic crashes are the most frequent incidents on highways and the ones with the most severe consequences. According to the National Highway Traffic Safety Administration (NHTSA), about 6.3 million highway crashes are reported annually in the United States, among which more than 32,000 are fatal crashes (NHTSA, 2016). These type of incidents often poses challenging problems in traffic operations and safety. Both transportation agencies and the general public are concerned about their notable direct and indirect impacts. It has been estimated that these highway crashes resulted in almost $1 trillion in economic loss and societal harm, dating back to 2010 (Blincoe et al., 2015).

The hazardous traffic conditions caused by any type of PI can cause additional traffic crashes, often referred to as secondary crashes (SCs). SCs are typically defined as crashes that occur within the spatial and temporal boundaries of the impact area of earlier PIs. Researchers have argued whether the principal cause of SCs is due to recurring or non-recurring congestions immediately after earlier PIs.

Moore et al. (2004) specified a comprehensive definition, "secondary crashes as those occurring in either direction of primary incident, within or at the boundary of the

queue formed by that incident". Per this definition, several studies including Raub (1997a, 1997b), Zhan et al. (2009), and Zhang & Khattak (2010b) assumed that potential SCs would be the result of PIs. However, Owens et al. (2009, 2010) defined SCs as those that occurred within the spatial and temporal boundaries of the impact areas that is formed due to earlier primary crashes (PCs).

The Federal Highway Administration (FHWA) Focus State Initiative defined SC as "unplanned incidents (starting at the time of detection) for which a response or intervention is taken, where a collision occurs either: a) within the incident scene, or b) within the queue (which could include the opposite direction) resulting from the original incident". Kentucky's highway incident management task force adopted a slightly different definition from FHWA, stating, "a secondary crash is a crash that has occurred due to non-recurring traffic congestion. The congestion should be a result of an earlier documented crash" (Pigmen et al., 2011). On the other hand, Xu et al. (2016) identified SCs accounting for recurrent congestions.

Although there is no standard definition of SCs, accurately defining those crashes is very critical (Park & Haghani, 2016). Several authors agreed that the reduction of SCs is a reliable performance measure for incident management systems (Sun & Chilukuri, 2010, Yang et al., 2018). That is why there is an urgent need to classify SCs with an appropriate methodological approach.

Previous studies determined that anywhere between 1-30% of total crashes were identified as SCs resulting from PC occurrence. (Karlaftis et al., 1999; Wang et al., 2016). Raub (1997a) found that 15% of all crashes on urban arterial roadways were secondary crashes caused by a prior incident. Chimba et al. (2014) reported that about 18% of all

freeway traffic fatalities were SCs. Xie et al. (2016) analyzed three years of crash records from May 2008 to April 2011 in Manhattan, finding that nearly 7.5% of crashes were SCs and 9.3% of those resulted in incapacitating and fatal injuries. Bryden & Fortuniewicz (1986) investigated a total of 3,302 barrier crashes in New York State and concluded that about 25% of total crashes were SCs, 90% resulting in fatalities. It has been reported that SCs can account for as high as 20% of all crashes and 18% of all fatalities on the United States' freeways (Owens, 2010). Considering the significant economic and social costs as well as the potential preventability, SC mitigation has become a priority for transportation agencies around the world.

1.2 Problem Statement

Researchers have found several issues and challenges in analyzing SCs, including:

- There are two major issues that could result in misclassification of SCs: (a) inadequate incident data; and (b) inconsistent approach to secondary crash identification (Yang et al., 2014b).

- Non-standard subjective selection of spatiotemporal thresholds inadequately identifies SCs (Yang et al., 2014b).

- It is difficult to determine if a SC was due to recurrent or non-recurrent congestion. The use of the category "accident ahead" underestimates SC records, while the category "congestion ahead" overestimates the SC frequencies. In general, crash records have insufficient data to identify SC (Sun & Chilukuri, 2010).

- Crashes are both a rare and random event, depending on various factors, including human, vehicle, roadway, and weather conditions. Since not all influencing factors are included in the statistical models, addressing unobserved heterogeneity is

critical (Sarker et al., 2017). Note that unobserved heterogeneity is an issue for any type of crash frequency analysis.

- It is difficult to evaluate the correlation between multiple incidents that occur within the spatiotemporal window of primary crashes (Haghani et al., 2006). The authors also mentioned that SC identification when congestion level information is unavailable might yield biased results.

- Intelligent Transportation Systems (ITS) strategies are becoming an essential components of Traffic Incident Management (TIM). On the other hand, effective TIM depends on reducing the risk of SCs, which is very challenging due to the stochastic nature of PC and SC. However, SC could be an effective TIM performance measure if properly recorded either in police crash reports or at the Traffic Management Center (TMC).

1.3 Research Objective

This research explores different approaches to identify the most suitable method of SC identification. Both static and dynamic approaches have been modeled to identify SCs. According to Sarker et al (2017), the dynamic approach is mostly applied on freeways over arterials, due to its discontinuous traffic flows and interrupted turning movements at intersections. Therefore, in this research work, the dynamic approach has been utilized only for the analysis of SCs of freeway facilities in Florida. And the static approach has been utilized for both freeway and arterial facilities in Florida. The findings of this research will be used by public and private sectors of the transportation industry in making operational strategies and management policies for improving traffic safety and mobility.

The specific objectives of this dissertation work are as follow:

1. Determining the most applicable method to identify SCs

2. Developing an enhanced model to quantify the major influential factors of SCs

1.4 Research Organization

This dissertation work includes a total of six chapters. Chapter 1 includes introduction, scope, and objective of this research work. Chapter 2 provides a comprehensive review of existing literature on SC identification, and methodological approaches to address the risk of potential SCs and its prevention policies. Chapter 3 presents the methodology that has been performed in this research to achieve the stated objectives. Chapter 4 describes the data sources and variables to perform the analysis. Chapter 5 includes the details of the results and findings. Chapter 6 summarizes the complete research and provides recommendations for future works.

CHAPTER 2

LITERATURE REVIEW

This chapter provides a review of existing literature on secondary crash occurrence on freeways and possible influential risk factors associated with secondary crashes. The different approaches adopted to identify secondary crashes are first discussed in detail. The models to predict the probability of secondary crash occurrences are then presented, followed by a discussion of the influential risk factors associated with secondary crashes. This section also presents several different challenges faced by researchers in analyzing secondary crashes. The summary of this chapter demonstrates in Appendix (Table A 2.1 – Table A 2.3).

2.1 State of the Art of SC Identification Methods

The identification and analysis of secondary crashes are difficult due to the lack of detailed information about prior incidents and relative traffic data (Zhan et al., 2008). As such, the most critical step in identifying secondary crashes is to determine the temporal and spatial boundaries of primary incidents (Khattak et al., 2007). The following approaches were found to be used frequently to identify secondary crashes (Sando et al., 2018). More specifically, static, dynamic, and the spatial analysis tools are discussed.

2.1.1 Fixed Spatiotemporal Thresholds Based

The static approach identifies secondary crashes using pre-specified spatial and temporal thresholds regardless of site and event-specific characteristics (e.g., Sarker et al., 2015; Mishra et al., 2016; etc.). The fixed spatiotemporal thresholds were predefined by the researcher's personal observation to identify those crashes.

Several researchers in 1990s used the static approach as this is the simplest approach of identifying SCs. Raub (1997a) retrieved past crash data and identified secondary crashes using fixed and predefined spatiotemporal boundaries. In Figure 1, the primary crash is at the point of (0,0) where distance and time are equal to zero. It is quite clear that crashes (a and b) at the same location and within a few minutes of the primary crash are related. However, it is very difficult to assume the distance (Ds) from a prior event, which has a probable influence of primary crashes for occurring secondary crashes. This influential distance is affected by traffic volume and travel times. During peak hours when traffic volume is high, this distance might be several kilometers.
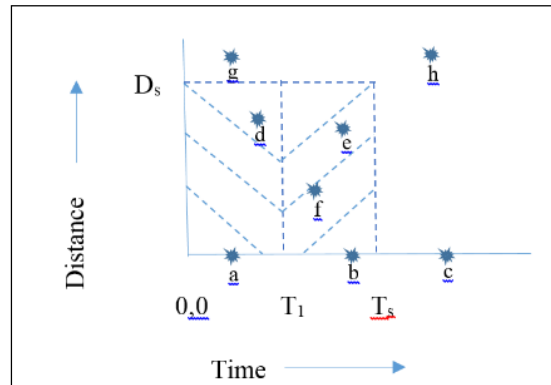
Figure 1: Defining Secondary Crashes Using Time-space Diagram (Raub, 1997a)

In Figure 1, it is assumed that crashes a, b, d, e, and f within the distance Ds and time Ts are linked to the prior crash. This assumption varies based on the researchers. Within the time T1, "a" and "d" crashes are more likely to be related than "e", "f", and "b". The rest of the crashes "c", "g" and "h" are assumed not to be influenced by the primary crash. According to Raub (1997a), distance Ds is less than 1 mile (1600 m), and time of effect Ts is 15 minutes longer than T1. The author used incident data for the 28-day period in January 1995 from the metropolitan region in Northern Chicago, Illinois. He found that the average time between primary crashes and secondary crashes was 36.4 minutes.

Several researchers, including Karlaftis et al. (1998, 1999), used similar spatiotemporal thresholds for identifying secondary crashes. Latoski et al. (1999) estimated the benefits of Hoosier Helper freeway service patrol in Northern Indiana during the daytime patrol and 24-hour patrol in 1996. In this study, the authors defined secondary crashes as those that occurred within 3 miles upstream and within 15 minutes after the clearance time of the primary crashes. The authors concluded that Hoosier Helper could reduce secondary crashes by 18.5% in winter and 36.3% in all other seasons per crash assisted.

Karlaftis et al. (1999) used a similar approach Karlaftis et al. (1998) to identify secondary crashes. Using four years of incident records from 1992-1995 on Borman Expressway, the authors determined 35% of all crashes to be secondary crashes. In this study, less than 0.8 km (1.5 km later on) in upstream and 15 minutes plus primary crash clearance time were considered as spatiotemporal thresholds to identify secondary crashes. Furthermore, the study also discovered that the assumed spatial distance was precisely accurate for higher traffic flow on Borman Expressway with nearly 140,000 vehicles per day. Zhan et al. (2008) utilized the Systems Management for Advanced Roadway Technologies (SMART) database from January 2005 to January 2007 in Florida to examine the likelihood of secondary crashes and their relation to primary incident characteristics. The crashes that occurred within 2 miles upstream in the same direction of the primary incident, and within the timeframe from the start time of primary incident to 15 minutes plus the clearance time were identified as secondary crashes. They found 413 secondary crashes that were linked with 352 primary incidents in 4,435 lane blockage incidents.

Moore et al. (2004) examined 84,684 crash records from 1998 to 1999 to identity secondary crash rates on Los Angeles freeways using a four-steps filtering method. In this study, the authors identified secondary crashes as those that occurred within 120 minutes and 2 miles of primary incidents. The authors concluded that secondary crashes per primary crash ranged from 1.5% to 3.0%, and secondary crashes per primary incident ranged from 0.7% to 1.3%. Hirunyanitiwattana & Mattingly (2006) used data from Highway Safety Information System (HSIS) to identify secondary crash characteristics in California. They compared primary and secondary crash characteristics using over 350,000 crashes in 1999 and 2000. The authors defined secondary crashes by a 2-mile-60-minute spatiotemporal window boundary. The authors used the approach that was similar to Moore et al. (2004) to exclude crashes in the opposite direction of primary crashes. They also found that the maximum queue length was between 2 miles and 5 miles. Per these spatio-temporal thresholds, about 4.4% of the total crashes were identified as secondary crashes, and this rate slightly increased from 1999 to 2000.

Sun & Chilukuri (2007) attempted to use the dynamic spatiotemporal threshold instead of a fixed boundary, as in the static approach to identify secondary crashes accurately. The study was based on 5,514 crashes in 2002 in Missouri. The authors used 2 miles-60 minutes static threshold window to define secondary accidents. As crash queues are continuously moving, the static threshold may have a significant risk of identifying secondary crashes with high positive and negative numbers (i.e., Type I and Type II errors, respectively). This limitation of the static method motivated the authors to apply dynamic spatiotemporal thresholds. The authors concluded that secondary crashes identified using the static and dynamic methods varied by more than 30%. Khattak et al. (2009) sought the

interdependence between primary incident duration and secondary incident occurrence using traffic incident and road inventory data in the Hampton Roads, Virginia in 2006. The study identified secondary incidents that occurred within 1 mile in the same direction and within the actual duration of primary incidents. However, actual duration plus 15 minutes threshold was used when the primary incident blocked lanes. By fixing the aforementioned threshold criteria, only 1.93% and 2.01% of 38,086 incidents were identified as primary and secondary incidents, respectively. Significant independent variables associated with longer duration were found to result in higher secondary incident occurrences include detection resources (Closed Circuit Television (CCTV) cameras, radio, and phone), accident (incident type), lane closure time, more vehicles involved, more response vehicles, higher AADT, left shoulder or ramp affected, and peak hours (Khattak et al., 2009).

Green et al. (2012) analyzed secondary crash occurrence on roadways in Kentucky from 2007 to 2010. The authors identified secondary crashes per the following definition: "a secondary crash is a crash that has occurred due to non-recurring traffic congestion. The congestion should be a result of an earlier documented crash". During the 18-month period from January 2009 to June 2010, a total of 9,330 crashes were coded as secondary crashes. The authors manually reviewed the police reports of these crashes and found that only about 3.88% (362 of 9,330) were secondary crashes. This inconsistency was due to two main reasons: the secondary crash had to occur as a result of a previous crash, and these must be secondary crashes and not secondary events. Moreover, this study further used an alternative algorithm to identify secondary crashes using the fixed spatiotemporal threshold window constituting 1.14 miles and within 80 minutes upstream of primary crashes. The authors also identified secondary crashes due to the "rubbernecking effect" by identifying

crashes within 0.2 miles of the primary crash. Overall, they found only a very small percentage of total crashes (0.10% to 0.15% of total annual crashes) to be secondary crashes.

Sarker et al. (2015) used both static and dynamic approaches to identify secondary crashes in Shelby County, Tennessee. The authors used three years (2010-2012) of crash data, freeway and arterial traffic data, incident management data, and roadway network data from the Tennessee Department of Transportation (TDOT). The authors used five different temporals (30, 60, 120, 180, and 300 minutes) and spatial thresholds (0.5, 1, 2, 3, and 5 miles) to identify secondary crashes using the static approach. The results revealed that the static approach inconsistently estimated secondary crash frequencies for various spatiotemporal thresholds. This study proved that dynamic approach is more accurate and reliable than static approach in identifying secondary crashes. A Secondary Crash Identification Algorithm (SCIA) was developed to identify secondary crashes resulting from primary crashes for the following possible five cases: (a) Case 1: Same Direction-Upstream; (b) Case 2: Opposite Direction-Upstream; (c) Case 3: Opposite Direction-Downstream; (d) Case 4: Opposite Direction-Upstream/Downstream; and (e) Case 5: Combination of first three cases.

Tian et al. (2016) conducted a pilot study in Florida to develop an efficient technique to identify secondary crashes. Sunguide® incident database, FDOT Crash Analysis Reporting (CAR) system database and geo-coded roadway geometric, and traffic data were analyzed for the year 2010. Similar to Sarker et al. (2015), Tian et al. (2016) utilized both static and dynamic approaches to identify secondary crashes. However, the static approach was used to create a new method using Geographic Information System

(GIS) to identify secondary crashes, and a dynamic approach was used to validate the feasibility of the proposed method. In this study, the spatiotemporal threshold window was fixed as 2 miles-2 hours in the same direction of primary crashes. This study also compared secondary crash frequencies and their characteristics using the following three spatiotemporal threshold groups: (a) Group 1: 2 miles, 2 hours; (b) Group 2: 2 miles, clearance time + 15 minutes; and (c) Group 3: 2 miles, clearance time + 30 minutes.

Khattak et al. (2010b) used Kernel density spatial analysis to observe the clustering of secondary incidents for five major Interstate freeways in Virginia. The Kernel density method approximates the concentration of secondary crashes per unit area or per unit length. Considering normal density or Gaussian function with mean 0 and variance 1, Kernel density function K calculates the shape of the bumps with kernel radius h. Equations 1 and 2 describe the Kernel density functions. The spatial distribution of the incidents for the entire network is estimated by these two equations. Compared to secondary incidents, as expected, the authors found that non-secondary incidents were widely distributed across roadway segments.

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left\{\frac{1}{h}(x - X_i)\right\} \tag{1}$$

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)} \tag{2}$$

where,

$\hat{f}(x)$ = Kernel density estimator at location, *x;*

n = observed number of events;

h = Kernel radius (bandwidth parameter);

X_i =   frequency of incidents observed on segment, $i;$ and

$K(x)$ =   symmetric probability Kernel density function.

Jalayer et al. (2015) identified secondary crashes based on 2 miles of queue length and 120 minutes of queue clearance time using 2010-2013 crash data from Alabama. The results revealed that only 5% of total crashes were identified as secondary crashes. Several studies concluded that the static approach is inaccurate in identifying secondary crashes due to its subjective assumption on fixed spatiotemporal thresholds (Sun & Chilukuri, 2006; Sun & Chilukuri, 2010; Khattak et al., 2010a; Chou et al., 2009). For example, a crash on an uncongested freeway that is far away from a primary crash should not be identified as a secondary crash (Park & Haghani, 2016).

On the other hand, overall static thresholds could be effective for a rough estimation of SCs in a specific study area due to their simplicity. However, they are inelastic and thus error-prone. Primary incident impact areas may vary significantly depending on weather, traffic conditions, or time of the day, so the likelihood of misidentification of SCs is significant. Moreover, Raub (1997) and Moore et al. (2004) concluded transferability of static thresholds is questionable for classifying SCs. To overcome the above mentioned limitations, several recent studies have used the dynamic approach, which uses flexible spatiotemporal thresholds.

2.1.2 Dynamic Spatiotemporal Thresholds Based

Principal idea of the dynamic approach is to identify secondary crashes using flexible spatiotemporal thresholds. This section includes previous studies related to the identification of secondary crashes using the dynamic approach.

Deterministic Queueing Model

Fundamentally, SCs identification utilizing this algorithm is based on maximum queue length. The estimated maximum queue length and dissipation time of primary incidents are used to identify SC spatiotemporal threshold. After that, PIs are linked to possible SCs through C+ programming environment. Traffic incidents and traffic count data are used for this kind of model to identify SCs. Queue length is calculated by the following deterministic queuing model with a single server (D/D/1). Based on total delay, t_i, the temporal boundary of secondary incidents can be estimated. Total remaining delay can be quantified by adding all the small trapeziums under the shaded area from t_i to t_e. Figure 2 illustrates the cumulative arrival and departure curve due to the incident bottleneck.

$$q(t_i) = q(t_{n-1}) + (t_i - t_{n-1})(\lambda_n - u^*) \qquad \text{for} \qquad t_{n-1}, t_i < T_c$$
$$q(t_i) = q(t_{n-1}) + (t_i - t_{n-1})(\lambda_n - u) \qquad \text{for} \qquad t_{n-1}, t_i > T_c$$

(3)

$$A_1 = \frac{1}{2}\big(q(t_n) - q(t_i)\big) \times (t_n - t_i)$$

$$A_2 = \frac{1}{2}\big(q(t_{n+1}) - q(t_n)\big) \times (t_{n+1} - t_n)$$

(4)

$$A_3 = \frac{1}{2}\big(q(t_{n+2}) - q(t_{n+1})\big) \times (t_{n+2} - t_{n+1})$$

………………

So, the remaining total delay at, $t_i = \sum_{k=1} A_k$.

The annotations of the above queueing model are as follows:

$q(t_i)$ =   Queue length for t_i;

u^* =   reduced capacity of the bottleneck;

u	=	returned capacity after primary incident clearance;
T_c	=	Primary incident clearance time;
t_{n-1}	=	$(n-1)^{th}$ time intervals from primary incident;
t_n	=	n^{th} time intervals from primary incident;
$A_c(t)$	=	cumulative arrival curve; and
$D_c(t)$	=	cumulative departure curve after bottleneck.
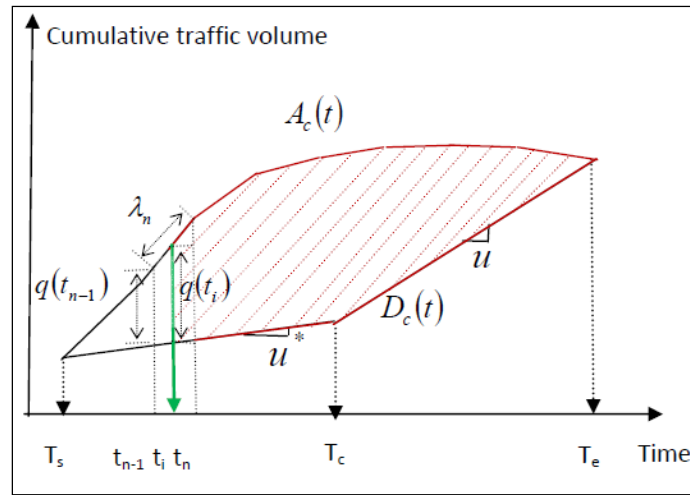
Figure 2: Queue Length and Remaining Total Delay at t_i (Khattak et al., 2010a)

Zhan et al. (2009) first introduced this method. The authors calculated incident recovery time from 33.34 minutes to 52.60 minutes and incident dissipation time from 0 to 21.76 minutes. By following Highway Capacity Manual (HCM) method, the study estimated maximum queue length from 1.09 miles to 1.49 miles. This study identified 225 secondary crashes that were linked to primary incidents. Khattak et al. (2010a) proposed a dynamic queue-based algorithm to develop STIA of primary incidents. The specific objective of this study was to develop an online prediction tool termed as DSD (Duration-Secondary Incident-Delay). After identifying dynamic spatiotemporal boundaries based on a simple Deterministic Queueing Model, the study identified secondary incidents (2% of

total recorded incidents) in both directions of primary incidents on Hampton Roads. However, Zhang & Khattak (2011) estimated average distance (1.2 miles) and time gap (34 minutes) between primary and secondary incidents based on 15 miles of maximum queue lengths. Vlahogianni et al. (2010) collected data via 220 CCTVs in Attica Tollway in Greece. The study utilized Bayesian Network and queueing model to identify the upstream STIA of primary crashes. The study revealed that 30% of the secondary crashes occurred approximately 1 km long and last for less than 1 hour.

Although queue-based approaches may offer a more accurate and dynamic representation of impact areas, they largely depend on the number and quality of available predictors. Considering that the factors affecting queue formation and dissipation may vary from case to case, it is likely that the impact areas predicted by these approaches might be inaccurate.

Speed-Based Matrix

This method is based on binary integer programming (BIP) to determine the actual spatiotemporal extent of delay using loop detector data. The basic idea of this approach is to develop congested cells caused by each primary incident and comparing other cells captured by the maximum extent of incident impact area with and without speeds. The detailed method includes three steps:

- *Step 1 - Construct the Spatiotemporal Speed Matrix:* Table 1 is the base condition matrix for the crash *free* scenario. In the matrix, if $s_j(t_m)$ is speed for a specific section j and for specific time interval t_m, then $s_{jn}(t_m)$ would be the speed distribution for n specific historical archived period. Table 1 includes a set of parameters $(\Omega_{j,m}=\Omega(\overline{s}_j(t_m),\sigma_{s_{j(t_m)}},\alpha_1,\alpha_2,.....,\alpha_p))$ corresponding to the *crash free*

speed distribution. Here, $\bar{s}_j(t_m)$, $\sigma_{s_{j(t_m)}}$, and α are mean speed, standard deviation, and level of confidence, respectively.

Table 1: Matrix of Crash-Free Speed $s_{jn}(t_m)$ (Chung 2013)

	Freeway Section (Traffic Flow Direction ⟵)					
Time	i	i-1	i-2	-	2	1
t_1	$\Omega_{i,1}$	$\Omega_{i-1,1}$	$\Omega_{i-2,1}$	-	$\Omega_{2,1}$	$\Omega_{1,1}$
t_2	$\Omega_{i,2}$	$\Omega_{i-1,2}$	$\Omega_{i-2,2}$	-	$\Omega_{2,2}$	$\Omega_{1,2}$
-	-	-	-	-	-	-
t_{M-1}	$\Omega_{i,M-1}$	$\Omega_{i-1,M-1}$	$\Omega_{i-2,M-1}$	-	$\Omega_{2,M-1}$	$\Omega_{1,M-1}$
t_M	$\Omega_{i,M}$	$\Omega_{i-1,M}$	$\Omega_{i-2,M}$	-	$\Omega_{2,M}$	$\Omega_{1,M}$

Table 2 shows the speed reduction profile due to crash occurrence in the section *i* at time t_1. This matrix can be schematically described by negative impact of crash (i.e., speed reduction) which follows shockwave theory, as illustrated in Figure 3(a).

Table 2: Matrix Observed Crash Speeds $\hat{s}_j(t_m)$ (Chung 2013)

	Freeway Section (Traffic flow direction ⟵)					
Time	i	i-1	i-2	-	2	1
t_1	$\hat{s}_i(t_1)$	$\hat{s}_{i-1}(t_1)$	$\hat{s}_{i-2}(t_1)$	-	$\hat{s}_2(t_1)$	$\hat{s}_1(t_1)$
t_2	$\hat{s}_i(t_2)$	$\hat{s}_{i-1}(t_2)$	$\hat{s}_{i-2}(t_2)$	-	$\hat{s}_2(t_2)$	$\hat{s}_1(t_2)$
-	-	-	-	-	-	-
t_{M-1}	$\hat{s}_i(t_{M-1})$	$\hat{s}_{i-1}(t_{M-1})$	$\hat{s}_{i-2}(t_{M-1})$	-	$\hat{s}_2(t_{M-1})$	$\hat{s}_1(t_{M-1})$
t_M	$\hat{s}_i(t_M)$	$\hat{s}_{i-1}(t_M)$	$\hat{s}_{i-2}(t_M)$	-	$\hat{s}_2(t_M)$	$\hat{s}_1(t_M)$

- *Step 2 - Determine Maximum Extent of Crash Influence.* The maximum extent of crash influence area due to the discontinuity of uncongested and congested traffic

flow is estimated. Due to a lack of crash scenario data (i.e., number of lane blockage, accident clearance time, etc.), this method assumed the worst-case scenario (i.e., total lane blockage) for the crash occurred in section, *i* at pre-specified time t_1. In this step, the real congested section specifically due to a crash, could be separated by a complete schematic speed distribution profile. Figure 3(b) displays the maximum set of freeway sections affected by a crash.
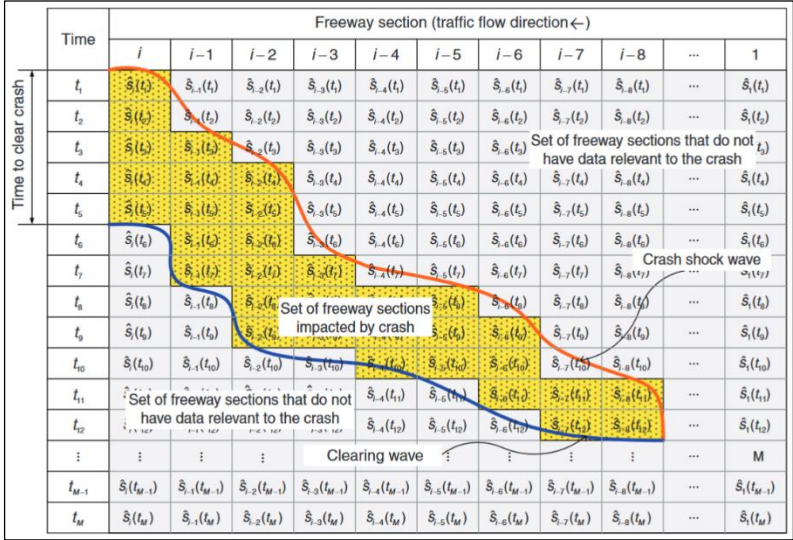
(a) Schematic of Spatiotemporal Freeway Sections Affected by a Crash
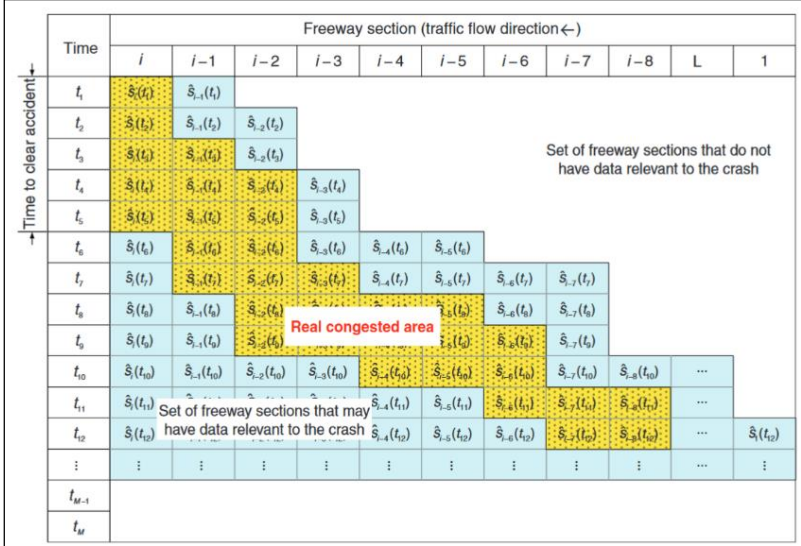
(b) Maximum Set of Freeway Sections Affected by a Crash

Figure 3: Estimation of Spatiotemporal Crash Impacts (Chung 2013)

- *Step 3 - Identify Spatiotemporal Crash Impact Region.* Using maximum extent of crash influence, the real congested crash region and other region are separated by comparing crash speed $\hat{s}_j(t_m)$ to crash-free speed $s_{jn}(t_m)$ at α level of confidence. It is assumed that any particular $\hat{s}_j(t_m)$ is not from the crash-free speed distribution $s_{jn}(t_m)$. Figure 4 represents the actual binary crash speed matrix where:

$$P_{jm} = \begin{cases} 0, & \hat{s}_j(t_m) \leq \bar{s}_j(t_m) - (\alpha \times \sigma_{s_{j(t_m)}}) \\ 1, & \hat{s}_j(t_m) > \bar{s}_j(t_m) - (\alpha \times \sigma_{s_{j(t_m)}}) \end{cases} \tag{5}$$

Note that the thick red line shows the actual boundary of the affected region.
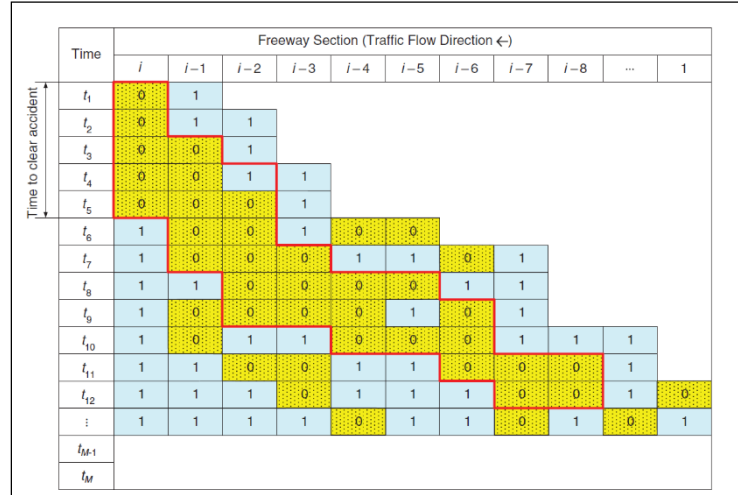
Figure 4: Binary Crash Speed Matrix (Chung, 2013)

The subset of cells for which the crash speeds are significantly different from the crash-free speeds, must follow certain topological properties. There are three problematic local shape configurations for the subset of cells that need to be addressed (Chung, 2013):

a. the spatiotemporal progression of the crash shock wave must be uninterrupted (i.e., an affected region does not contain holes),

b. the boundary of the spatiotemporal progression of the crash shock wave must be upstream [i.e., the vertical position *(t)* of any dot-shaded section *j* along the

boundary of the region must be either lower than or the same as the vertical position

of the neighboring dot-shaded section *j - n*], and

c. the entire boundary of the affected region must be contiguous.

These conditions and the problem of determining the "best" set of dot-shaded cells

can be formulated as the following binary integer programming (BIP) problem:

$$\min_{\delta_{jm}} Z = \sum_{\forall j,m} \left[P_{jm}.\delta_{jm} + \left(1 - P_{jm}\right).\left(1 - \delta_{jm}\right) \right] \tag{6}$$

subject to

$$\delta_{j+k,m} \leq \left[1 - (\delta_{j,m} - \delta_{j+1,m})\right].R \qquad\qquad \forall j, m; \ \forall k \leq J - j$$

$$\delta_{j,m+r} \leq \left[1 - (\delta_{j,m} - \delta_{j,m+1})\right].R \qquad\qquad \forall j, m; \ \forall r \leq M - m$$

$$\delta_{j,m+k} \leq \left[1 - (\delta_{j,m} - \delta_{j+1,m})\right].R \qquad\qquad \forall j, m; \ \forall k \leq M - m$$

$$\delta_{jm} = \begin{cases} 0 \\ 1 \end{cases}$$

where R is a large number, J is the maximum number of upstream sections, and M

is the maximum number of subinterval time periods that define maximum duration

assumed for congestion caused by a crash.

Chung (2012) proposed above mentioned "Speed-Based Matrix" for the first time

to quantify non-recurring congestion delay caused by crashes on urban freeways in Orange

County, California. Chung (2013) applied this method to identify secondary crashes using

one year of crash data from March 2001 to February 2002 in California. The author

assumed that crashes that occurred within the average time duration of the study and in the

vicinity of primary crashes would be secondary crashes. Using BIP, the author found 1,890

of 6,200 secondary crashes were linked to primary crashes. About 7.5% (141 of 1,890) of

them were in the same direction of the primary crashes, and only 3.8% (71 of 1,890) of them occurred due to the rubbernecking phenomenon. An average spatiotemporal boundary was found to be 1.34 miles-65.81 minutes for the same direction and 1.83 miles-81.49 minutes in the opposite direction. The author also identified cascading secondary crashes in both directions. In the same direction of primary crashes, about 6.1%, 1.1%, 0.2%, and 0.1% of the 1,890 crashes were first, second, third, and fourth secondary crashes. Similarly, 3.5% and 0.3% of the 1,890 crashes were the first and second secondary crashes in the opposite direction of primary crashes.

Speed Contour Plot

The main idea behind these approaches is to establish the speed contour (heat) map based on speed measurements from various sensor measurements. The time-space diagram is split into grid cells based on certain time intervals (e.g., 5 min, 15 min, etc.) and milepost of sensor stations. In general, each cell is determined to be congested or not based on the following condition:

$$V_{(t,s)}^b = \begin{cases} 1, if\ V_{(t,s)} < V_{(t,s)}^r \\ 0, \qquad\quad others \end{cases} \tag{7}$$

where $V_{(t,s)}$ represents its current speed; $V_{(t,s)}^b = 1$ means that the cell is congested and 0 means not congested; and $V_{(t,s)}^r$ depicts its reference speed obtained from historical data. Thus, the impact area of PI can be depicted using the congested cells following the occurrence of the PI. If another crash occurs within the congested cells, it has been identified as a SC. The key premise is the selection of the reference speed.

Yang et al. (2013, 2014a, 2014b, 2014c, 2014d), Xu et al. (2016), and Goodall (2017) utilized the "Speed Contour Plot" approach to identify SCs. Yang et al. (2014c)

proposed a new online-based approach by developing speed contour plot utilizing virtual sensors that collect data from private traffic information providers such as Bing Maps, Google Maps, and MapQuest, instead of traditional sensors (i.e., loop detector or roadside traffic sensors). The specific objective of this study was to develop a readily employable tool for large scale roadway network so that researchers could identify secondary crashes directly from the tool. The authors concluded that this algorithm is a more reliable method of identifying secondary crashes than the traditional static or dynamic queuing models. The authors also found that virtual sensors provide better results compared to the traditional sensors. Yang et al. (2013, 2014a) used the archived data from traditional traffic sensors and a similar approach using speed contour plots to identify secondary crashes. The study used traffic information (volume, speed, occupancy, and travel time) from traditional sensors for 27 miles on New Jersey Turnpike in 2011. The authors also compared the results using both the predefined fixed static approach and the aforementioned method. The authors concluded that the proposed algorithm captured more crashes compared to the static approach. The proposed method also reduced the incorrect classification of secondary crashes.

Goodall (2017) addressed two limitations, captured all the secondary crashes if at least 90% of the connecting line passes within the non-recurring congestion zone. Moreover, since secondary crashes can be associated at any time of the primary crash occurrence, the author categorized crashes that occurred at the tail or border of the queue as secondary crashes. Goodall (2017) applied this refined approach to identify secondary crashes on I-66 corridor using incident, speed, crash, and vehicle probe project surveillance for 2014. About 13.80% (340 of 2,466) of secondary crashes were found to be in correlation

with the upstream of primary crashes. The findings revealed that secondary crash occurred on an average once in each 24.8 incidents. Xu et al. (2016) followed the approach discussed by Yang et al. (2014c) to identify secondary crashes on I-880 freeway in CA using five years of crash and traffic data. Only 113 of 9,188 crashes (about 1.2%) were identified as secondary crashes based on the speed contour plot. This proportion was found to be consistent with several other studies, including Sarker et al. (2015), Wang et al. (2016), Park & Haghani (2016), and Mishra et al. (2016).

Simulation-Based Approach

A computationally efficient methodology, the Simulation Based Secondary Incident Filtering (SBSIF) method, is proposed for efficiently delineating the boundaries of the incident impact area in a time-space contour map of traffic speeds and employing the outcome in identifying secondary incidents.

In comparison to the static and dynamic approaches, this method accounts for the dynamic spatial and temporal properties of incident impact given prevailing traffic conditions. The SBSIF method is composed of two main tasks. The first task identifies the incident impact area that results from each primary incident, i.e., the portion of the time-space traffic speed contour map in which traffic speeds are impacted due to the incident (Figure 5). The second task employs the impact area to identify the secondary incidents from archived data (Figure 6).
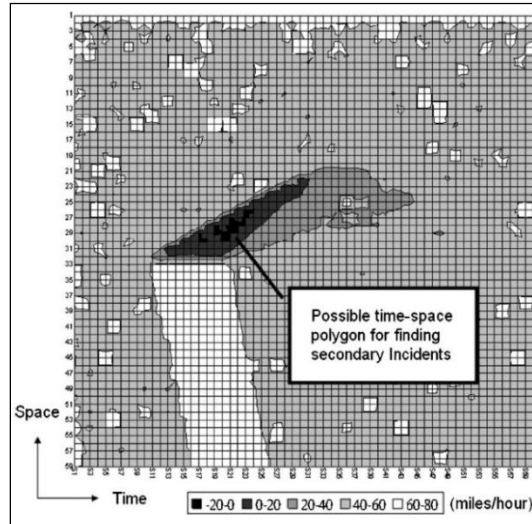
Figure 5: Traffic Speed Contour Map (Chou & Miller-Hooks 2009)
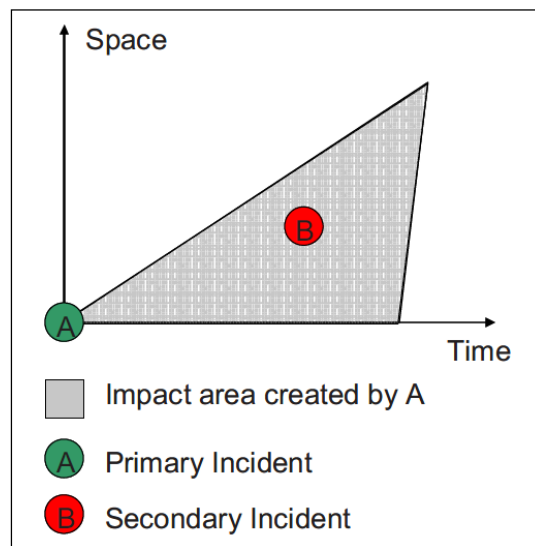
Figure 6: Secondary Incident Identification Through STIA (Chou & Miller-Hooks 2009)

The SBSIF method utilizes multi-regression models for the quick identification of the corner points of the incident impact areas. Each corner point is associated with two calibrated ordinal least-square regression models.

Figure 7 illustrates secondary incident identification for the specific average speed threshold under specific traffic conditions through the corner point identification approach. The x and y of each corner point of the polygon represent time and space value,

24

respectively. The side of the impact area is estimated using Equation 8 under the given

corner points (x_1, y_1) and (x_2, y_2). As such, any incident at time x in location y within the

polygon is classified as secondary incident resulting from primary incident.

$$y = y_1 + \left[\frac{(y_2 - y_1)}{(x_2 - x_1)}\right] \times (x - x_1) \tag{8}$$
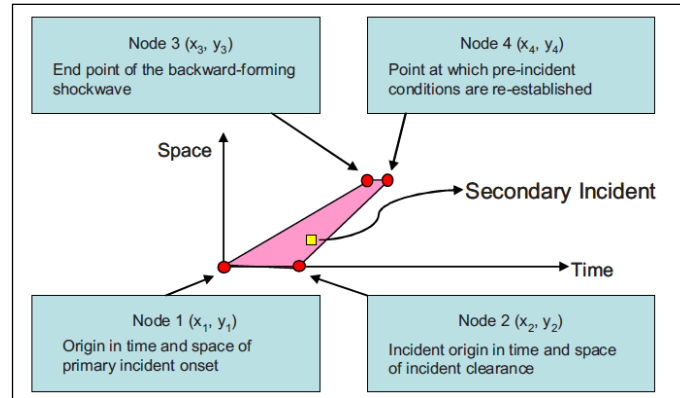
Figure 7: Corner Points Identification Approach (Chou & Miller-Hooks 2009)

Haghani et al. (2006) introduced a simulation-based approach to identify the

spatiotemporal impact area (STIA) of an isolated incident using a three-step process in

CORSIM micro-simulation software. The authors conducted a Multi-Layer Sensitivity

Analysis (MLSA) for the evaluation of Hudson Valley Highway Emergency Local Patrol

(H.E.L.P.) program. As part of MLSA, longitudinal location of secondary incidents was

defined with respect to dynamic queue formation based on shock wave theory. In this case,

queue formation on the part of I-287 corridor was explained by an analytical algorithm (Al-

Deek et al., 1995) considering mean occupancy rates of loop detector data. A set of time

intervals for a set of queue lengths were employed to determine the impact areas in both

the directions of primary incidents. Each iteration consisted of increasing the time with a

constant time interval to identify the impact areas.

Step 0: $S = \phi$. Select and remove p_{ik} from Q_p.

Step 1: Generate a corner point set, $C_{p_{ik}}$, of incident $p_{ik} \in Q_p$ via regression models with traffic conditions and incident properties.

Step 2: Let (x,y)-coordinates of $p_{ik} \in Q_p$ be the point of origin,$(0,0)$. Form an impact area, $IA_{p_{ik}}$, of incident $p_{ik} \in Q_p$ using line functions with the corner point set, $C_{p_{ik}}$.

Step 3: Compute (x,y)-coordinates, $(x,y)_{p_{ij}}$, for a potential secondary incident, $p_{ij} \in Q_s$, with data logs of mile marker and incident start timestamp with respect to $p_{ik} \in Q_p$.

Step 4: If $(x,y)_{p_{ij}} \in IA_{p_{ik}}$ and the traffic flow direction of p_{ij} is the same as p_{ik}, a primary-secondary pair is found, denoted (p_{ik}, p_{ij}).

Step 5: $S = S \cup \{(p_{ik}, p_{ij})\}$. Repeat Steps 3 and 4 for all $p_{ij} \subseteq Q_s$ and $p_{ij} \neq p_{ik}$.

Step 6: If $Q_p \neq \phi$, select and remove p_{ik} from Q_p and return to Step 1; otherwise, if $Q_p = \phi$, stop. The procedure terminates with the set of secondary incidents and the pairing of secondary incident to primary incidents.

Notation employed:

Q_p	:	Set of primary incidents consisting of Ω incident types, $Q_p = p_{ik}, \forall i \in (1,2,....n_k), \ \Omega \subseteq \Gamma$
p_{ik}	:	An incident, i, of incident type k, $\forall i = (1,2,...n_k), k \in (1,2,...K), \ p_{ik} \in Q_p$
Q_s	:	Set of potential secondary incidents consisting of Λ incident types, $Q_s = p_{ij}, \forall i \in (1,2,....n_j), \ \Lambda \subseteq \Gamma$
p_{ij}	:	An incident, i, of incident type j, $\forall i = (1,2,...n_j), j \in (1,2,...J), \ p_{ij} \in Q_s$
S	:	Set of primary-secondary incident pairs, $S = \{(p_{ik}, p_{ij})\}, \forall p_{ik} \in Q_p, p_{ij} \in Q_s$
(p_{ik}, p_{ij})	:	A primary-secondary incident pair, $\forall p_{ik} \in Q_p, p_{ij} \in Q_s$
$C_{p_{ik}}$	:	Set of corner points of incident $p_{ik}, \forall p_{ik} \in Q_p$
$IA_{p_{ik}}$	:	Impact area of incident $p_{ik}, \forall p_{ik} \in Q_p$
$(x,y)_{p_{ij}}$	:	(x,y)-coordinates of incident $p_{ij}, \forall p_{ij} \in Q_s$

Figure 8: Steps of SBSIF Method (Chou & Miller-Hooks 2009)

Figure 8 shows the detailed steps adopted in this method. This step was repeated until the traffic returned into the pre-stage of the primary incident condition. A secondary incident was considered within the impact areas for each time interval till the end interval. Based on this model, the secondary incident rate was found to be between 0.7% and 0.13%.

Chou & Miller-Hooks (2009) also adopted the Simulation Based Secondary Identification Filtering (SBSIF) method by considering the parameters (lane blockages, incident duration, speed, volume, etc.) used by Haghani et al. (2006). Chou & Miller-Hooks (2009) collected incident data for 10 miles on I-287 corridor in New York State from January to June 2006. The authors used one year of traffic data from 2007 to create the contour map where 693 primary incidents were simulated in CORSIM. About 3.80% and 4.30% of secondary incidents were identified in the same direction of primary incidents

using visual and regression applications of SBSIF method, respectively. The study also compared the results with the static filtering method for two spatiotemporal thresholds values adopted by Zhan et al. (2008) (Static threshold # 1: 2 miles-15 minutes) and Raub (1997a) (Static threshold #2: 1 mile-15 minutes). About 6.70% and 7.50% of secondary incidents were identified using static threshold #s 1 and 2, respectively. Only three incidents were erroneously identified by the SBSIF method, whereas the static approach based on 2 miles-15 minutes threshold had erroneously identified 23 secondary incidents.

Queuing Shockwave Based Algorithm

In this approach, a dynamic spatiotemporal impact area is outlined based on real time traffic conditions employing kinematic shockwave theory for bi-direction traffic flow, as shown in Figure 9. A crash is categorized as a secondary crash if it occurred within the estimated STIA.
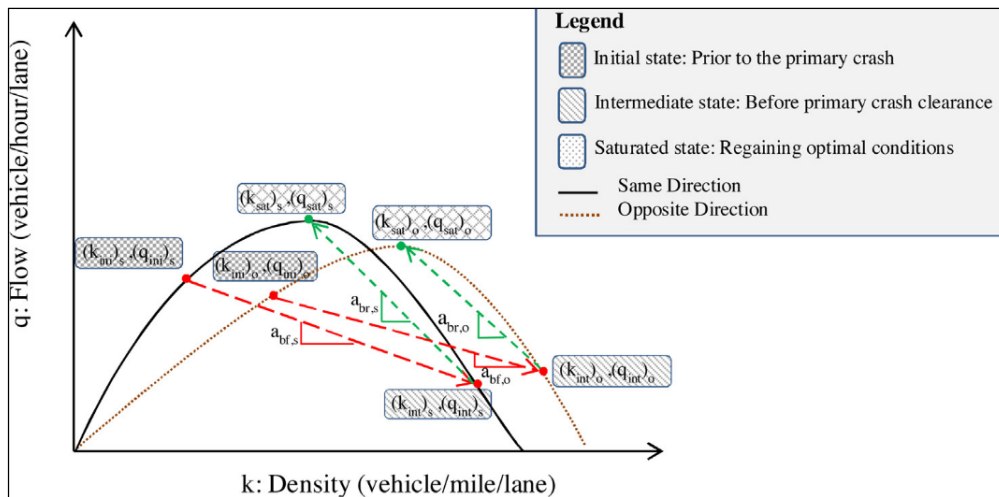
Figure 9: Shockwave Speed for Bi-Directional Traffic Flow (Sarker et al., 2017)

The Methodological framework consists of the following steps:

- *Step 1 - Developing Crash Pair Algorithm:* This step identifies candidate secondary crashes paired with primary crashes *(c_i, c_j)* based on pre-specified static thresholds. The algorithm consists of the following conditions:

  o *Condition 1:* $0 \leq t(c_j) - t(c_i) \leq T$

  o *Condition 2:* $d(c_i, c_j) \leq D$

  where,

  c_i =   former crash, *i*

  c_j =   later crash, *j*

  $t(c)$ =   *c* crash at *t* time since an early time origin in minutes

  $d(c_i, c_j)$ =   distance from crash c_i to c_j in miles

  T   =   temporal threshold in minutes

  D   =   spatial threshold in miles

  Two crashes are paired if they satisfy these two conditions. The spatial distance, *$d(c_i, c_j)$* between two crashes is estimated using Dijkstra's method. Dijkstra's method is an iterative approach that finds the shortest path from an origin to every node in a network. In addition, this distance needs to be calculated using the absolute difference in Beginning Log Mile (BLM). The position of the paired crashes is determined with in relation to each other, using their direction, BLM, and their respective spatial coordinates.

- *Step 2 - Identification of Secondary Crashes:* The two filters (ramp and impact area filters) are performed to identify secondary crashes from the paired crashes. The crash pairs need to be excluded where primary crashes occurred on highway ramps

since very few secondary crashes were found to have occurred on ramps. The impact area of a crash is defined between two simplified straight shockwave lines, one for the queuing shockwave and the other for the discharging shockwave.

- *Step 3 - Estimation of STIA:* In order to estimate STIA, following algorithm needs to be performed. Shockwave speeds have been calculated using the following equations. Moreover, clearance time (T_c) and time difference between primary crash and potential secondary crash can be directly collected from the database.

$$a_{br,s} \times (t - T_c) \leq d \leq a_{bf,s} \times t, \text{ when } t > T_c$$

$$0 \leq d \leq a_{bf,s} \times t, \text{ when } t < T_c$$

$$a_{bf,s} = \frac{(q_{ini})_s - (q_{int})_s}{(k_{ini})_s - (k_{int})_s} \tag{9}$$

$$a_{br,s} = \frac{(q_{int})_s - (q_{sat})_s}{(k_{int})_s - (k_{sat})_s} \tag{10}$$

Notations of the proposed algorithm are:

$a_{br,s}$ = backward-recovery shockwave speed in the same direction of primary crash;

$a_{bf,s}$ = backward-forming shockwave speed in the same direction of primary crash;

$(k_{ini})_s, (q_{ini})_s, (u_{ini})_s$ = density, flow, and speed in the same direction prior to primary crash;

$(k_{int})_s, (q_{int})_s, (u_{int})_s$ = density, flow, and speed in the same direction prior to primary crash clearance;

29

$(k_{sat})_s, (q_{sat})_s, (u_{sat})_s$ = density, flow, and speed in the same direction at optimal condition;

$a_{br,o}$ = backward-recovery shockwave speed in the opposite direction of primary crash;

$a_{bf,o}$ = backward-forming shockwave speed in the opposite direction of primary crash;

$(k_{ini})_o, (q_{ini})_o, (u_{ini})_o$ = density, flow, and speed in opposite direction prior to primary crash;

$(k_{int})_o, (q_{int})_o, (u_{int})_o$ = density, flow, and speed in opposite direction prior to primary crash clearance;

$(k_{sat})_o, (q_{sat})_o, (u_{sat})_o$ = density, flow, and speed in opposite direction at optimal condition;

t = time difference between primary crash and secondary crash; and

T_c = clearance time.

Algorithms based on queuing shockwaves can "better capture the effects of traffic characteristics (e.g., flow, speed, and density), that change over time and space, and affect both queue formation from a primary crash occurrence" (Sarker et al., 2017). Zheng et al. (2014, 2015) and Sarker et al. (2015, 2017) used queuing shockwave-based algorithms to identify secondary crashes. The methodological framework specifically conducted for large scale networks with limited available data. The *Crash Pair Algorithm* was found to be both reliable and faster compared to the ArcGIS program (Zheng et al., 2014). Zheng et

al. (2014) utilized State Trunk Network (STN) and the crash data for 1,500 miles of freeways in Wisconsin. The data was collected for the year 2010 from WisTransPortal Data Hub of the Traffic Operations and Safety (TOPS) Laboratory at the University of Wisconsin. The Two-Phase Automated procedure was applied on 7,034 crash records to identify secondary crashes on that large segment of freeways. In this study, only 79 secondary crashes were identified.

In contrast, Sarker et al. (2015, 2017) identified secondary crashes that occurred within STIA by employing shockwave theory. The likelihood of secondary crash occurrence was found to be much higher within 0.5 miles to 1 mile and within 30 minutes to 60 minutes of the prior crash occurrence, while the likelihood was found to be lower within 5 miles-300 minute threshold.

Incident Progression Curve

The temporal and spatial region of influence of a primary incident is delineated by the queue length resulting from a primary incident throughout the duration of an incident. Since a queue often persists after an incident has been cleared, the incident normalization time is required in addition to the incident clearance time. This temporal and spatial region of influence of a primary incident is bounded by a curve (Figure 10). Sun & Chilukuri (2010) named this curve as Incident Progression Curve (IPC).
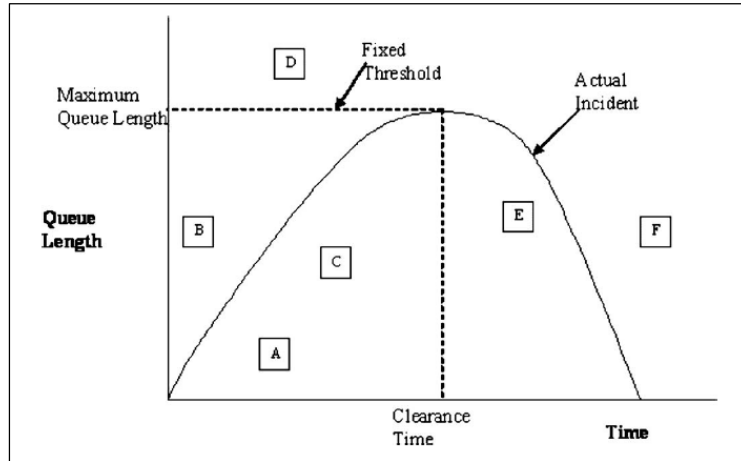
Figure 10: Static Threshold Versus Actual Incident Progression

The study methodology includes following four steps to identify secondary incidents:

- *Step 1 - Reviewing Intranet Incident Reports*: in this step, researchers need to download the pages of the reports at a regular interval through an automated computer script. Incident type, description and reporting time were extracted in a specific field via computer programming. All the information is then be integrated from different sources e.g., aircrafts, elevated traffic cameras, freeway service patrol, emergency management (fire, police, ambulance, and HAZMAT), and motorist calls.

- *Step 2 – Filling of Incomplete Incident Reports:* The final dataset needs to be constructed using all incidents based on second, third, and fourth-order polynomial models.

- *Step 3 - Master Incident Progression Curves:* In this step, a master IPC is developed by capturing the central propensity of incidents. To create a master IPC, all individual IPCs should be separated by an equal growth interval (Figure 11).

Median of the incident duration and queue length are estimated for each of the interval. The joined IPCs are sketched as an isolated IPC for a single incident.
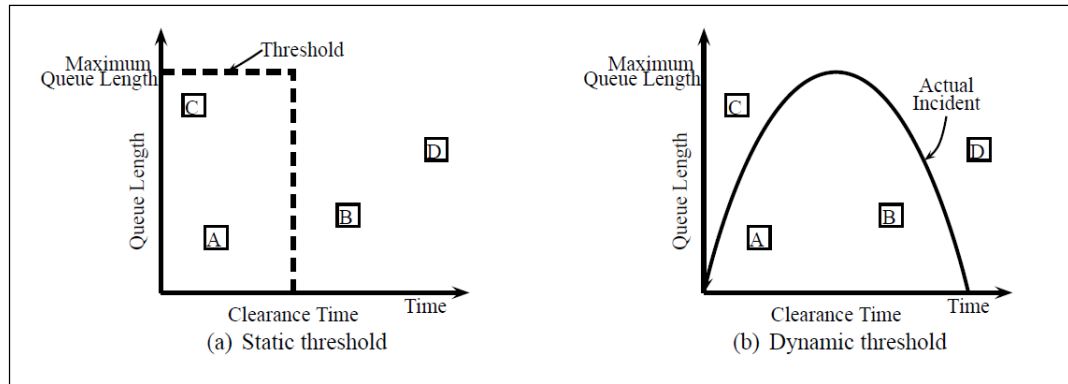
Figure 11: Static Versus Dynamic Incident Progression Curve (Shlayan et al., 2009)

- *Step 4 - Identification of Secondary Incidents Based on Dynamic IPC:* This step is to estimate the dynamic threshold based using the maximum queue length model using Equation 11. In this equation, Q is the length of the queue in miles, and t is the time duration from the primary incident occurrence in minutes. Sun & Chilukuri (2010) estimated values of the coefficients a_0, a_1, a_2, and a_3 (0.013873, 0.12652, -0.00094363, and -0.000007826).

$$Q = a_0 + a_1 t + a_1 t^2 + a_3 t^3 \tag{11}$$

Sun & Chilukuri (2010) developed a nonlinear regression incident progression curve (IPC) to identify secondary traffic crashes in the both directions of primary crashes. Other than previous studies, the authors attempted to improve static methodology by addressing the end of the dynamic queue through the entire curve. This study analyzed 480 intranet reports (5,514 crashes) for the freeway segments of I-70 and I-270 in St. Louis. Based on master IPC, a dynamic spatiotemporal threshold window was found as 3.09

miles-43.5 minutes. The results also showed that only 7.19% of total crashes were found as secondary crashes based on the dynamic threshold IPC.

Shlayan et al. (2009) also recommended a model for identifying secondary incidents employing the dynamic progression curve. The study considered both directions of primary incidents. The methodology of this research work demonstrated through a real-time case (a video example) and based on multiple VISSIM simulation models for 15 conditions in the Las Vegas area. It was found that the front of the queue was disseminated upstream.

The deployment of IPCs solely depends on detailed incident data. However, such data is difficult to obtain as incidents are random events, and most agencies do not track queue lengths in real time. That is why, identification of SCs using IPC approach is quite challenging to deploy.

Automatic Staudynamikanalyse Model

Automatic Staudynamikanalyse (ASDA), also known as automatic tracking of moving traffic jams, is a robust model for tracking and defining the spatiotemporal influence of traffic disturbance in freeways (Zheng et al., 2015). The ASDA method possibly be used to track a moving jam at all times (Kerner et al., 2004). Based on Kerner's three-phase traffic theory, Kerner et al. (2004) established the ASDA algorithm to capture freeway congested traffic flow patterns by employing spatiotemporal boundaries. Li & Bertini (2010) used this algorithm for capturing wide, moving jam. They concluded that ASDA might be a better model compared to the traditional speed threshold algorithms, especially because ASDA is based on not only traffic speed but also traffic flow.

This is a robust model to identify SCs on freeways. Automatic Staudynamikanalyse (ASDA) model algorithm considers not only traffic speed, but also the flow. Figure 12 schematically illustrates the ASDA model where Q_0 and Q_n are two successive detectors on a freeway road section. The model starts measuring the uninterrupted positions of the upstream front, $x_{up}^{(jam)}$ immediately after moving jam has been detected by Q_n at time t_0. Note that the upstream front of the moving jam is considered to have reached a detector at some time if the three following criteria need to be fulfilled:

- The speed at the detector is below the maximum speed threshold;

- The speed drop at the detector is greater than a certain threshold (estimated from real time data); and

- The difference between the speeds at the detector and the next downstream detector is greater than the speed differential
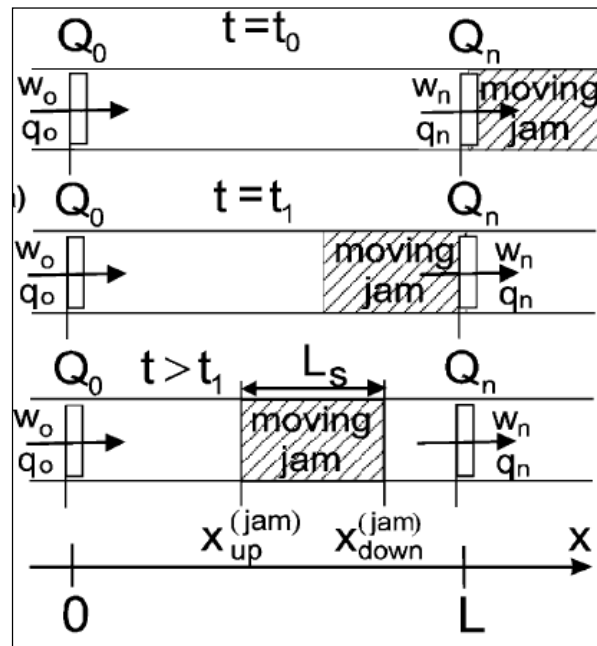
Figure 12: Schematic Illustration of ASDA Model (Kerner et al., 2004)

On the other hand, continuous positions of downstream front of traffic jam, $x_{down}^{(jam)}$ is measured by Q_n at time t_1. ASDA model estimates $x_{up}^{(jam)}$ and $x_{down}^{(jam)}$ resulting from primary incidents using Equations 12 and 13, where i and j are two detectors at t time values. These two indices increase in the same direction of traffic flow along x axis. L_{i+1}, L_j are coordinates of the corresponding detectors. $v_{up}^{(jam)}(t)$ and $v_{down}^{(jam)}(t)$ are upstream and downstream fronts of the moving jam velocities, respectively. $t_o^{(t+1)}$ is the time when the upstream front jam is calculated at $(i+1)$ detector and $t_1^{(j)}$ is the time when the downstream front jam is calculated at j detector. $w_o^{(i)}(t)$ and $q_o^{(i)}(t)$ are the average speed and traffic flow rate at detector i upstream of the jam. On the other hand, $w_{max}^{(j)}(t)$ and $q_{out}^{(j)(jam)}(t)$ are the averaged vehicle speed and the traffic flow rate at detectors j downstream of the wide moving jam. Whereas, q_{min} and ρ_{max} are the traffic flow rate and density inside the moving jam.

Using Equations 12 and 13, total jam width, L_s (Equation 14) at t, queue duration and maximum queue length resulting from primary incidents need to be calculated. So, spatiotemporal boundaries of the incident impact areas are easily defined. Consequently, secondary incidents could be identified if they occurred within the impact areas.

$$x_{up}^{(jam)}(t) = L_{i+1} + \int_{t_0^{(i+1)}}^{t} v_{up}^{(jam)}(t)dt$$

$$\approx L_{i+1} - \int_{t_0^{(i+1)}}^{t} \frac{q_0^{(i)}(t) - q_{min}}{\rho_{max} - \left(\frac{q_0^{(i)}(t)}{w_0^{(i)}(t)}\right)} dt, \ \ t \geq t_0^{(i+1)}, \ \ i = 1,2,\ldots \tag{12}$$

$$x_{down}^{(jam)}(t) = L_j + \int_{t_1^j}^{t} v_{down}^{(jam)}(t)dt \tag{13}$$

$$\approx \ L_j - \int_{t_1^{(j)}}^{t} \frac{q_{out}^{(j)(jam)}(t) - q_{min}}{\rho_{max} - \left(\dfrac{q_{out}^{(j)(jam)}(t)}{w_{max}^{(j)}(t)}\right)} dt, \ \ t \geq t_1^{(j)}, \ j = 1,2,\dots$$

$$L_s = \ x_{down}^{(jam)} - x_{up}^{(jam)} \tag{14}$$

The density parameter, ρ_{max} is estimated by Equation 15.

$$\rho_{max} = \frac{1000}{L_{PC}.A_{PC} + L_{HGV}.(1 - A_{PC})}\left[\frac{vehicles}{km}\right] \tag{15}$$

where,

L_{PC} = the expected average length of passenger cars (PC) including a (small) average

distance between vehicles inside the moving jam

A_{PC} = the fraction of PC, can be determined with local detector measurements

L_{PC} = the expected average length of heavy goods vehicles including a (small) average

distance between vehicles inside the moving jam

Imprialou et al. (2014) analyzed 1,287 incidents for the years 2007 and 2009 on a 31-mile freeway segment on Attica Tollway. The authors compared the results with the ASDA model and the cumulative plots method. The authors defined the spatiotemporal thresholds of the incident influence area to be 1.43 miles-70 minutes and 1.6 miles-80 minutes for the ASDA algorithm and cumulative plots method, respectively. For identifying secondary incidents, the authors initially constructed a second-order polynomial model that could link spatial distance from prior incident to the incident duration. The model was fitted based on Levenberg-Marquardt algorithm, a nonlinear least-squares curve-fitting procedure. Only 1.79% of secondary incidents were identified using this method. In another study conducted by Orfenou et al. (2011) using the same approach,

only 3.5% of the total accidents (30 of 856 accidents) were categorized as secondary accidents.

2.2 Influential Risk Factors and SCs Occurrence Model

Several researchers have used statistical models and tests to analyze the relationships between the primary crash characteristics, and the possibility of secondary crash occurrence. This section discusses the studies that adopted regression models, ordered probit models, logit models, multinomial logits models, Bayesian logit models, binomial logit models, proportional test, t-test, chi-square test, etc. to identify the risk factors of SCs.

2.2.1 Probability of SC Occurrence

Goodall (2017) developed the binary logistic regression model to predict the secondary crash occurrences. Latoski et al. (1999) also developed logistic regression models to predict the probability of a secondary crash as a function of clearance time in winter, clearance time during the rest of the year, vehicle type, weekday, and presence of a ramp or a median. In this research, the authors identified secondary crashes using the static approach: crashes that took place within 3 miles upstream and within the clearance time plus 15 min of a primary crash were identified as secondary crashes.

Kopitch & Saphores (2011) examined the probability of secondary crashes depending on various contributing factors using a simple logistic regression model. The authors focused on evaluating the impact of Changeable Message Signs (CMS) on reducing secondary crashes. They concluded that the effectiveness of a CMS in preventing secondary crashes increases between 2 miles and 11.15 miles, and then decreases between 11.15 mi and 22.3 miles, and these results are only slightly statistically significant.

The analysis was based on a combined database of weather data with geometric information and accident data in 2008 for 74 miles on I-5 from Mexico-US border to Orange County, CA. In this study, crashes that occurred within 2 miles and 2 hours in the both directions of primary crash occurrence were defined as secondary crashes. About 5.2% of all primary crashes were represented as secondary crashes. Moreover, findings also showed that CMS might be an effective incident management tool to mitigate potential secondary crashes.

Descriptive Statistics

A few studies used descriptive statistics to analyze secondary crashes. Yang et al. (2013, 2014a) conducted a descriptive analysis on the secondary crashes identified using speed contour plots. The findings are as follows:

- About 8% of all crashes were identified as secondary crashes. Every 11 non-secondary crashes were linked to one secondary crash.

- 50% of secondary crashes occurred within 2 miles upstream of primary crashes.

- 75% of secondary crashes were within 2 hours of primary crashes.

- Two or more vehicles were involved in secondary crashes. 75% of secondary crashes were rear-end crashes.

- "Following too closely" (54%) followed by "driver inattention" (12%), and "improper lane change" (18%) were the major contributing factors to the secondary crash occurrence.

- Clearance time for the identified secondary crashes was less than primary crashes. A high 90% of the secondary crashes were cleared within 90 minutes.

Tian et al. (2016) utilized both static and dynamic approaches to identify secondary crashes. The authors identified secondary crashes using GIS and a fixed spatiotemporal threshold window of 2 miles-2 hours in the same direction of primary crashes. They also conducted descriptive analysis on the identified secondary crashes. The findings include:

- "Careless driving" (50%) followed by "exceeded safe speed limit" (8.13%) and "no improper driving/action" (4.07%) were the major contributing factors of primary crashes that were linked to secondary crashes.

- A majority of secondary crashes were rear-end, followed by angle crashes.

- More than 50% of secondary crashes resulted in no injuries. Within the 2-hour temporal window, secondary crashes were found to result in the lowest percentage of fatal crashes (0.73%) and the highest percentage of PDO crashes (61.59%) were found.

Like Yang et al. (2013, 2014a) and Tian et al. (2016), Carrick et al. (2015) also conducted descriptive analysis using data from secondary crashes in Florida. The authors collected information about secondary crashes at the scene by police officers reporting traffic crashes. Some of the relevant findings include:

- Secondary crashes were found to be more likely to occur in cloudy or rainy conditions. Wet roadway surface condition was found to be involved in twice the number of SCs.

- The likelihood of secondary crashes was found to be high in hilly areas. Uphill and downhill grades combined were found to experience twice the number of crashes.

- Higher rates of secondary crashes were observed when they occurred "on the roadway" and involved three or more vehicles, including commercial vehicles.

2.2.2 Risk Factor Model

A study examined the underlying relationship between SC occurrences and different contributing factors. In general, both parametric and non-parametric models were sought to model the association of SCs with external conditions. With the modeled association, it is expected to provide transportation agencies with more insightful information when developing countermeasures to mitigate SC risks. A detailed of the modeling practices are provided below. Information regarding the authors, methods, considered variables, data used, and SC identification methods in each study are presented.

Logistic Regression Model

Junhua et al. (2016) used three years of loop detector data from California freeways to analyze secondary crashes. Based on 180 minutes-24.85 miles of spatiotemporal boundary, researchers identified 1,183 crash pairs for further analysis. They investigated major contributing factors involving primary crashes resulting in secondary crashes using a binary logistic regression model. The authors used the shock wave boundary filtering (SWBF) method based on queuing theory to identify secondary crashes. This method "provides real-time accident impact scope and is equipped with an automatic algorithm to conduct the filtering work circularly" (Junhua et al., 2016). "It includes three main steps: (1) calculate traveling speed of primary accident impact through flow and density information; (2) determine a feasible spatiotemporal district for secondary accidents by estimating the real time space-time scope of shock waves generated by every potential primary accident; and (3) match the primary accident with the corresponding loop data to calculate the spatiotemporal district for secondary accidents." (Junhua et al., 2016).

Junhua et al. (2016) found that crash processing duration significantly affects on the secondary crash occurrence. Several studies including Zhan et al. (2008, 2009); Karlaftis et al. (1999, 2009); Zhang & Khattak (2010a, 2010b, 2011); Yang et al. (2014b); Vlahogianni et al. (2010, 2012) made similar observations. However, tow away indicator, road surface condition, and other parties involved, were found to be insignificant. This study also identified the high significance of the following three types of shock waves generated by primary crashes: (a) Shock Wave 1: Generated at the time of the primary crash; (b) Shock Wave 2: Generated when rescue personnel or police arrive at the site to control traffic, and (c) Shock Wave 3: Dissipated when primary crash has been transacted, and the bottleneck is recovered. The study revealed that all these waves increased the likelihood of secondary crash occurrence. The authors concluded that "stepwise speed control is necessary in the downstream section of the bottleneck to slow down the traffic wave" (Junhua et al., 2016).

Mishra et al. (2016) used the above multinomial logit model to observe the correlation between secondary crashes with corresponding contributing risk factors. Increased number of vehicles involved in crashes, relatively high AADT, increased upstream flow was found to lead to a higher likelihood of secondary crash occurrence. Nonetheless, good weather condition decreased the likelihood of secondary crash occurrence. The study also found that the likelihood of secondary crashes increased if the primary crashes were rear-end compared to other crash types (i.e., angle, sideswipe, head-on, etc.). The authors found that more secondary crashes involved with higher upstream traffic flow. Primary crashes with rear-end collision type was found to be the predominant factor that contributed to secondary crash occurrences. The authors concluded that the

probability of a secondary and a tertiary crash are 37.18% and 4.05%, respectively. Note that the SCs that occurred in the upstream same direction of the primary crashes were considered in crash prediction modeling.

Zhan et al. (2008) used regression models, and identified the following five factors to have statistically significant effect on the likelihood of secondary incident occurrence: the number of vehicles involved in the primary incident, the number of lanes at the primary incident location, the primary incident duration, time-of-day of incident occurrence, and the probability of vehicle rollover occurring during the primary incident. The authors also concluded that incident visibility and the lane blockage durations of the primary incidents are significant contributing variables for determining the severity of secondary crashes.

Khattak et al. (2012) used a binary logit model to identify factors associated with the occurrence of secondary incidents resulting from primary incidents. Note that the other variables are self-explanatory. The authors found a significant positive correlation between secondary crash occurrences and longer primary incident duration, higher AADT, and primary incidents occurred during peak hours.

Neural Network Model

The following NN model is to identify SCs risk factors. This model is based on multi-layer perceptions (MLPs) and a logistic regression model with no interaction term.

$$y_p = \frac{1}{1 + e^{-net_j}} \tag{16}$$

where,

$$net_j = \sum_k \omega_{kj} h_k + \theta_j$$

y_p =   output of pth data

ω_{kj} =   connection weight between k^{th} neuron in the hidden layer and j^{th} neuron in the

         output layer with the bias term, θ_j

$h_k = \dfrac{1}{1+e^{-net_k}} =$ output of the hidden neuron

$$net_k = \sum_i \omega_{ik} x_i - \theta_i$$

ω_{ki} =   connection weight between k^{th} neuron in the hidden layer and i^{th} input variable

         with the bias term, θ_i

Neural networks (NNs) are efficient predictive models widely applied in function approximation and classification problems (Vlahogianni & Karlaftis, 2013). Vlahogianni et al. (2012) implemented logistic output activation function in their NN model to identify significant variables associated with secondary crashes on a 65.2 km urban motorway linked to the city center from the Athens International Airport.

The authors ranked the importance of independent variables in the modeling of the probability of having a secondary accident using two measures: mutual information, and the second is partial derivatives. The mutual information provides the overall significance of certain variables on the secondary accident likelihood under the prevailing roadway and traffic conditions; while the "partial derivatives-based ranking provides information on which variable to alter in order to affect immediate change to the secondary accident likelihood" (Vlahogianni et al., 2012).

Traffic speed, duration of the primary accident, hourly volume, rainfall intensity, and a number of vehicles involved in the primary accident were found to be the top five factors associated with secondary accident likelihood. The authors also found changes in

traffic speed and volume, blocked lanes, percentage of trucks, and upstream geometry to significantly influence the probability of secondary incident occurrence (Vlahogianni et al., 2012).

Park & Haghani (2016) used a Bayesian neural network for the sequential prediction of secondary incidents from the point of incident response to the clearance of primary incidents. The authors compared the results with Binary Logit model results. Probabilistic learning models were considered as conditional distribution of the dependent variables. The results revealed that the likelihood of secondary incidents was higher when the clearance time for primary incidents is between 10 minutes and 20 minutes, or more than 75 minutes.

Bayesian Random Effect Logit Model

Bayesian random effect logit model (Equation 17) is for predicting the secondary crashes due to primary crash characteristics, roadway geometric, and environmental conditions (Huq 2011). To avoid bias parameter estimation, this model consists of unobserved heterogeneous factors i.e., the work zones, design features, and pavement conditions. Markov Chain Monte Carlo (MCMC) simulations were used to quantify the random effect logit model.

$$y_n = \text{Bernoulli } (p_n) \tag{17}$$

$$\text{logit } (p_n) = \beta_0 + \theta_r + \beta_1 x_{1n} + \beta_2 x_{2n} + \ldots + \beta_i x_{in}; \theta_r \sim (0, \Sigma_\theta) \tag{18}$$

The elasticity of continuous independent variables such as traffic flow characteristics can be estimated using Equation 18. The pseudo-elasticity (see Equation 19) can be calculated to estimate the elasticity of non-continuous variables, such as indicator variables that take on values zero or one. The elasticity of a continuous independent

variable represents the percentage change in the dependent variable resulting from a 1% change in the independent variable (Washington et al., 2010). Similarly, pseudo-elasticity gives "the incremental change in frequency caused by changes in the indicator variables" (Washington et al., 2010).

$$E_i = \frac{\partial Y}{\partial x_i} \times \frac{x_i}{Y} = (1 - P)\beta_i x_i \tag{19}$$

$$E_i = \left\{ \frac{e^{\Delta(x'\beta)} \times (1 + e^{x_i \beta_i})}{e^{\Delta(x'\beta)} \times e^{x_i \beta_i} + 1} - 1 \right\} \times 100 \tag{20}$$

where,

y_n =   secondary crash indicator (i.e., 1 if a secondary crash is induced by a primary crash, and 0 otherwise) for the nth observation;

p_n =   P = probability of a secondary crash;

β_i =   coefficient of independent variable x_i;

x_{in} =   the value of variable i for sample n; and

θ_r =   random effect which captures the heterogeneity effects for freeway segment, r.

Using this mode, Xu et al. (2016) observed findings similar to Sarker et al. (2017); they found a positive correlation between AADT and secondary crash occurrences. The likelihood of secondary crashes was found to be much higher on weekends compared to weekdays. Compared to several different crash types, sideswipe primary crashes were found to be less likely to cause secondary crashes. Similar to Xie et al. (2016), Xu et al. (2016) indicated that the likelihood of secondary crashes decreased with increase in number of lanes. The authors also observed that the probability of secondary crashes was more likely to occur during morning peak periods. The authors also concluded that

including real-time traffic variables increased the prediction accuracy by 16.6%, and considering unobserved heterogeneity (consideration of random variables) effectively increased the prediction accuracy by 7.7%.

Generalized Ordered Response Probit Model

Sarker et al. (2017) developed a GORP model and confirmed that there was no significant evidence of unobserved heterogeneity in both expected count and propensity components. The study also compared the standard NB model with the final GORP model using log-likelihood ratio (LR) test statistic and found that the GORP model might be a superior statistical model. The study also found a significant effect of right shoulder width, speed limit, AADT, number of lanes, and presence of raised median on secondary crash occurrences. The findings revealed that about 10% increase in AADT increased secondary crash occurrences by 34.24%. Moreover, the authors also found that two-lane roads caused 73% more secondary crashes compared to road segments with three or more lanes, and locations with raised medians experienced more secondary crashes compared to undivided roads. Segments with right shoulder width less than 14 ft were found to be experienced more secondary crashes compared to roads with wider shoulders, and those segments with over 55 mph experienced more secondary crashes compared to segments with lower speed limits.

Structural Equation Models

Xie et al. (2016) proposed Structural Equation Modeling (SEM) that was used a combination of factor analysis and multiple regression analysis to analyze structural relations (Ullman, 2003). The authors investigated structural relationships between the contributing factors, presence of secondary collisions, and injury severity. The authors

conducted the chi-square difference test to estimate the goodness of fit of the three models. The authors concluded that, compared to SEMs, probit models overestimated the safety effects of confounding variables by mixing the direct and indirect effects (Xie et al., 2016).

The study found the following thirteen explanatory variables to contribute to the presence of secondary crashes: alcohol, drugs, inattention, inexperience, sleep, control disregarded, speeding, fatigue, defective brake, pedestrian involved, defective pavement, limited view, and rain. The sixteen variables were found to be increased in the risk of severe injuries presence of secondary crashes. They include alcohol, drugs, inattention, yield, illness, control disregarded, speeding, fatigue, cell phone, defective brake, motorcycle involved, bike involved, pedestrian involved, defective pavement, and at intersection. The likelihood of the occurrence of secondary crashes and severe injuries were found to be higher at nighttime compared to daytime conditions (Xie et al., 2016).

Gompit Model

Vlahogianni et al. (2012) used a Gompit model employing the Weibull distribution to identify secondary incident risk factors related to primary incident characteristics. The authors could not find a significant relation between collision type and secondary crash occurrence. However, they found that the likelihood of SCs is negatively correlated with the number of blocked lanes for primary incidents occurrence.

Probit Model

Vlahogianni et al. (2012) used probit model for estimating the probability of secondary incidents. The authors found that secondary crash occurrence was significantly affected by upstream geometry, rainfall intensity, speed, and traffic volume. The results

also revealed that lower speed and higher lane volume might increase secondary accidents. This output is consistent with Junhua et al. (2016).

Statistical Tests

This section focuses on the statistical tests applied in earlier discussed studies to analyze contributing risk factors associated with possible secondary crashes resulting from primary crashes.

Test for Proportions

Hirunyanitiwattana & Mattingly (2006) conducted the test for proportions to compare the differences in the characteristics of secondary crashes and primary crashes using crash data from 1999 and 2000 from California highway system. The authors identified secondary crashes using a 2-mile-60-minute spatiotemporal window boundary. Jalayer et al. (2015) also conducted the Test for Proportions to determine if there is any statistically significant difference between the primary and secondary crashes with regards to various characteristics including crash type, severity level, time of day, area type, average emergency response duration, and roadway classification. The null hypothesis is that the proportion of primary crashes by primary crash contributing factor is not statistically significantly different from the proportion of secondary crashes of the same classification. The alternate hypothesis is that the proportion of primary crashes by primary crash contributing factor is statistically significantly different from the proportion of secondary crashes of the same classification. The authors used static threshold of 2 miles-120 minutes to identify secondary crashes. The analysis was based on four years of data from 2010 through 2013 of Alabama State. These results were found to be consistent with

several other studies including Raub (1997a), Hirunyanitiwattana & Mattingly (2006), and Zhan et al. (2008). Some of the key findings are listed below (Jalayer et al., 2015):

- The probability of the secondary crash occurrence in urban areas was found to be higher than in rural areas.

- "Following too close" and "driving too fast" were found to be the main contributing factors of secondary crashes.

- Secondary crashes were found to result in no injury (i.e., PDO).

- Rear-end crashes were found to be overrepresented among secondary crashes.

- The proportion of secondary crashes with no emergency response required was found to be significantly higher compared to those in the primary crash.

- Compared to all crashes in Alabama, the secondary crashes were found to be overrepresented and underrepresented for interstate and state highways, respectively.

- Secondary crashes were found to be more frequent during morning and evening peak hours.

<u>Pearson's Chi-square Test</u>

Zheng et al. (2015) used Pearson's Chi-square (χ^2) test to test the independence between secondary crashes and general crashes (identified using the two-phase automatic identification process discussed in Section 2.2.5), and for the day of week, month of the year, and hour of the day. They did not observe statistically significant differences between secondary crashes and general crashes with respect to day of the week (χ^2=2.88). On the

contrary, secondary crashes were significantly different from primary crashes with respect to month of the year and hour of the day, the χ^2 values are 20.13 and 88.91, respectively.

t-Test

Tian et al. (2016) used t-test to determine if there is a statistically significant difference in crash rates, crash severity and other factors between secondary crashes identified. The authors found a very limited influence on temporal criteria of secondary crashes at 5% significance level. The results also indicated no significant differences between the temporal boundaries of clearance time plus 15 minutes and clearance time plus 30 minutes. However, rear-end crashes were found to increase when clearance time plus 30 minutes instead of 2 hours was considered.

2.3 Review on SC Prevention

Other than the identification and risk factor modeling of SCs, very few studies focused on the prevention of SC occurrences. The primary countermeasures explored in existing studies include the deployment of the active traffic management using changeable or variable message signs (CMS or VMS) variable speed limit control (VSL), and connected vehicles (CVs). For example, Kopitch and Saphores (2011) verified the effectiveness of 11 CMS that provided real-time traffic information about incidents, work zones, congestion, speed limits ahead, and alerts in reducing SC risk. It was found that the effectiveness of CMS increased between 2 and 11.15 miles and decreased between 11.15 and 22.3 miles. Li et al. (2014) introduced the strategy of implementing a variable speed limit with both weather and traffic flow information to mitigate SC risk. Two surrogate safety measures, including time exposed time-to-collision (TET) and time integrated time-to-collision (TIT), were found to be reduced by 40–50 percentage in a case study on I-880

in California during heavy rain conditions. Lately, Yang et al. (2017) examined the impact of connected vehicles on improving the situational awareness of drivers to mitigate SC occurrences. SC risk, measured by the number of simulated conflicts, was found to be significantly reduced if the market penetration rate of CVs on a highway was relatively high (e.g., 15%) in dense traffic conditions.

Other than the countermeasures, some studies also examined the benefits of service patrol programs in reducing SCs. For example, Karlaftis et al. (1999) examined the effect of the Hoosier Helper service patrol program on the Broman Expressway in Indiana. It was found that the program may help reduce SC likelihood by 18.5 percent in winter and by 36.3 percent in other seasons per crash assisted. The delay savings and crash cost savings from secondary crash reduction were $568,080 in 1995, which was 1.38 times the service patrol program cost. Although there was no quantitative assessment, some other studies also mentioned the use of service patrol programs as an effective countermeasure to reduce SC risk. For example, Khattak et al. (2012) suggested the improvement of coverage of service patrols and towing service on highway chokepoints that have higher SC occurrence probability.

Mitigation of post-crash impacts of SCs rather than the prevention of SCs has also been discussed by some researchers. Compared with previous studies that only used PC information, Park et al. (2016a) considered the evolution of PCs and SCs over time to determine an appropriate location for emergency response units. Linear programming approach with relaxed integrality constraint for integer variables was verified to be valid in reducing the expected total delay of crashes in a numerical study with data collected on an interstate highway.

2.4 Review of Study Constraints

This section includes several issues and constraints that have been faced by previous authors while doing their research on secondary crash analysis, are summarized below:

- As crash queues are continuously moving, the static threshold may have a significant risk of identifying secondary crashes with high positive and negative numbers (i.e., Type I and Type II errors, respectively).

- Two major issues resulted in misclassification of secondary crashes: (a) inadequate incident data; and (b) inconsistent approach to secondary crash identification (Yang et al., 2014).

- Nonstandard subjective selection of spatiotemporal thresholds cannot adequately identify secondary crashes (Yang et al., 2013).

- It is difficult to determine if a secondary crash was due to recurrent or non-recurrent congestion. The use of the category "accident ahead" underestimated the secondary crash records, while the category "congestion ahead" overestimated the secondary crash frequencies. In summary, crash records do not have sufficient information to identify secondary crashes (Sun et al., 2006).

- Crashes are rare and random events, which depend on various factors, including humans, vehicles, roadways, and weather conditions. Since not all influencing factors are included in the statistical models, addressing unobserved heterogeneity is critical (Sarker et al., 2017). Note that unobserved heterogeneity is an issue for any type of crash frequency analysis (Sarker et al., 2017). Moreover, the model

parameters can be constrained to be the same or allowed to vary between groups. So, it is difficult to identify parameter constraints (Xie et al., 2017).

- It is difficult to evaluate the correlation between multiple incidents that occur within the spatiotemporal window of primary crashes (Haghani et al., 2006). The authors also mentioned that secondary incident identification when congestion level information is unavailable might yield biased results. That is why, in this dissertation work, the dynamic approach was conducted by assuming that there were no multiple SCs occurred within the same congested spatiotemporal area. It eventually yields biased results.

- Though Connected Vehicle (CV) technology is the most advanced method of mitigating SCs, there are some significant limitations that still needs to be addressed (Yang et al., 2017).

CHAPTER 3

METHODOLOGY

3.1 Methodology for SC Identification

This chapter describes the methodology that has been applied in this research. The first two sections provide methods that have been performed to achieve the two specific objectives of this dissertation. A detailed description of the procedures involved in the methodology is presented in each subsequent section.

3.1.1 Static Approach

The static approach includes fixed spatiotemporal thresholds to identify SCs. According to Table A 2.1, the majority of previous studies considered 2 miles-2 hours threshold values to identify SCs. Therefore, this study used 2 miles-2 hours spatiotemporal thresholds while using the static approach. The ArcGIS, a mainstream Geographical Information System (GIS) software, was used to assign mileposts to all incidents. Based on previous studies, SCs can occur either in the upstream direction of the primary incidents (PIs) or in the opposite direction of the PIs. However, this study identified SCs only in the same direction and upstream of PIs. Figure 13 is the flow chart showing the static approach using GIS.
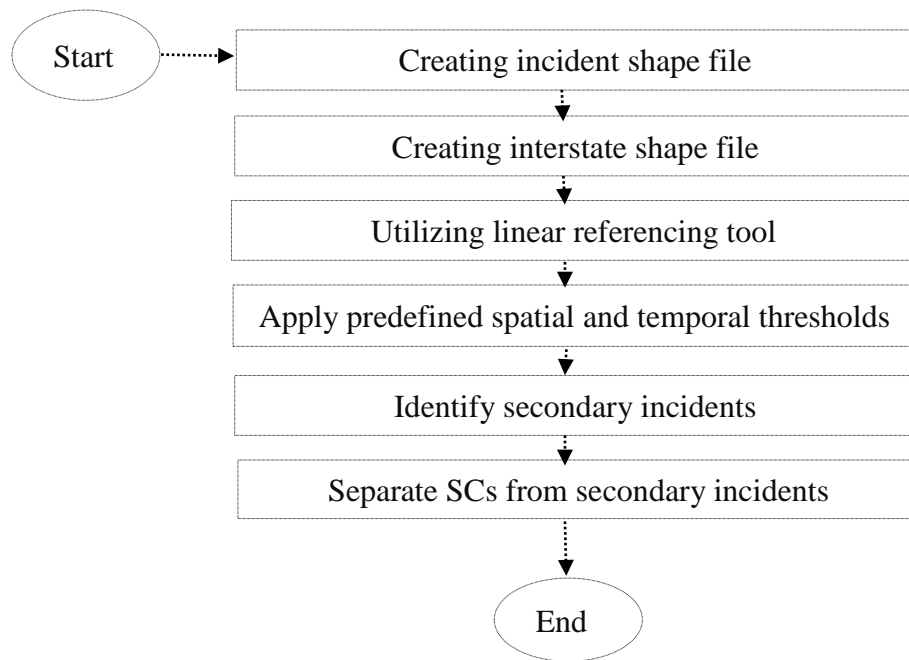
```
┌─────────┐            ┌──────────────────────────────────────────┐
│  Start  │┄┄┄┄┄┄►     │        Creating incident shape file      │
└─────────┘            └──────────────────────────────────────────┘
                                           ┆
                                           ▼
                       ┌──────────────────────────────────────────┐
                       │        Creating interstate shape file    │
                       └──────────────────────────────────────────┘
                                           ┆
                                           ▼
                       ┌──────────────────────────────────────────┐
                       │        Utilizing linear referencing tool │
                       └──────────────────────────────────────────┘
                                           ┆
                                           ▼
                       ┌──────────────────────────────────────────┐
                       │  Apply predefined spatial and temporal thresholds │
                       └──────────────────────────────────────────┘
                                           ┆
                                           ▼
                       ┌──────────────────────────────────────────┐
                       │        Identify secondary incidents      │
                       └──────────────────────────────────────────┘
                                           ┆
                                           ▼
                       ┌──────────────────────────────────────────┐
                       │    Separate SCs from secondary incidents │
                       └──────────────────────────────────────────┘
                                           ┆
                                           ┆
                                           ▼
                                      ┌─────────┐
                                      │   End   │
                                      └─────────┘
```

Figure 13: SCs Identification by Static Approach Using ArcGIS

Identification of Upstream Secondary Crashes

SCs occurred in the upstream direction of the Primary Incident (PI). The following steps were performed:

- *Step 1 - Assign Mileposts to Incidents*: Traffic incidents are mapped in GIS using the corresponding coordinates (latitude and longitude) in the dataset. Next, mileposts were assigned to each traffic incident using a linear referencing tool in ArcGIS.

- *Step 2 - Identify Potential Secondary Incidents that are Crashes*: While the primary event could be any incident and not necessarily a crash, this method focuses on identifying only SCs (and not secondary incidents). Thus, as one of the initial steps, the potential secondary incidents are checked to make sure that they are crashes .

56

- *Step 3 - Identify Upstream Potential SCs*: The occurrence of a PI is expected to result in a queue backup in the upstream direction. Therefore, SCs that occurred only in the same direction and upstream of the PI are identified by comparing the milepost of the PI with the milepost of the potential SCs.

- *Step 4 - Calculate Distance (Spatial Threshold)*: The mileposts of the PI and the potential SCs are used to compute the distance between the two.

- *Step 5 - Calculate Time Difference (Temporal Threshold)*: The time difference between the PI and the potential SC is calculated.

- *Step 6 - Extract PI-SC Pair*: Following the identification of the spatiotemporal relationship between the PI and the potential SC (in Steps 4 and 5), the respective pairs are extracted based on the set spatiotemporal criteria.

- *Step 7 - Store the Identified Secondary Crash*: The extracted SCs are stored, and the process is repeated for the rest of the incidents.

3.1.2 Dynamic Approach

In contrast to the static approach, this method utilizes the dynamic spatiotemporal impact areas (STIA) to identify secondary crashes (SCs). At first, the STIA of the primary incidents are deployed using speed contour plot (SCP). Later, incidents within or at the boundary of the STIA, have been recognized as SIs. In this research, both recurring and non-recurring traffic congestion formed by primary incidents are considered to identify SIs. Afterwards, SCs were extracted from SIs. The real time traffic data including speed information were extracted from RITIS and incident data were obtained from SunGuide™ incidents database from 2014 to 2017. A total of 21,589 incidents were extracted to identify

SCs in the same direction and upstream of PIs using SCP dynamic approach. The incident zone only includes Miami-Dade County areas.

Speed Contour Plot

This approach first determines STIA of the PI using real-time traffic flow data while accounting for the effects of recurring congestion. A secondary incident is then identified if it is within the STIA of the primary incident. Figure 14 shows flow chart of SCP method.

```
 ┌─────────┐      ┌──────────────────────────────────────────────────────────┐
 │  Start  │----▶ │ Preparing master database with incident, network, traffic │
 └─────────┘      │ operation data                                            │
                  └──────────────────────────────────────────────────────────┘
                                           │
                  ┌──────────────────────────────────────────────────────────┐
                  │ Assessing traffic flow characteristic using detector data │
                  └──────────────────────────────────────────────────────────┘
                                           │
                  ┌──────────────────────────────────────────────────────────┐
                  │ Organizing all data by date, route and direction of       │
                  │ traffic flow                                              │
                  └──────────────────────────────────────────────────────────┘
                                           │
                  ┌──────────────────────────────────────────────────────────┐
                  │ Extracting raw speed and average speed at detectors       │
                  └──────────────────────────────────────────────────────────┘
                                           │
                  ┌──────────────────────────────────────────────────────────┐
                  │ Developing an incident speed contour plot on time and     │
                  │ space values                                              │
                  └──────────────────────────────────────────────────────────┘
                                           │
                  ┌──────────────────────────────────────────────────────────┐
                  │ Locating speed drop area on speed contour plot            │
                  └──────────────────────────────────────────────────────────┘
                                           │
                  ┌──────────────────────────────────────────────────────────┐
                  │ Identifying secondary incidents within or at boundaries   │
                  │ of speed drop                                             │
                  └──────────────────────────────────────────────────────────┘
                                           │
                  ┌──────────────────────────────────────────────────────────┐
                  │ Separating SCs from total incidents                       │
                  └──────────────────────────────────────────────────────────┘
                                           │
                                      ┌─────────┐
                                      │   End   │
                                      └─────────┘
```
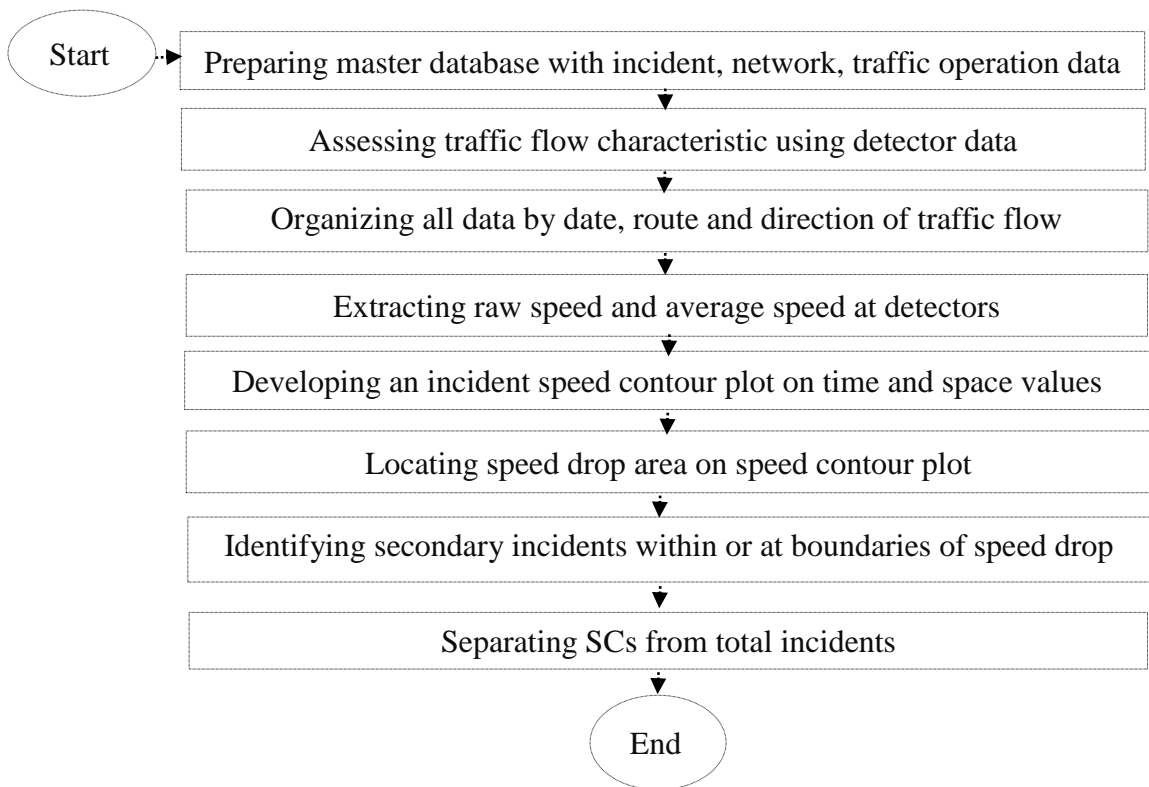
Figure 14: SCs Identification Using SCP Method

Detail explanation of the steps conducted in SCP algorithm are listed below:

- *Step 1* - The 5-min speed data were extracted from RITIS database to develop a speed contour plot for a prior incident. More specifically, the speed data were extracted for the time interval between 6 h before the prior incident and 6 h after

the prior incident from the loop detectors within 10 miles upstream the prior incident. Figure 15a illustrates an example of a speed contour plot for a prior incident, where congestions and queue formations were clearly observed after the prior incident. It is difficult, however, to determine whether the queue formations were caused by recurrent congestions or the prior incident. To account for effects of the recurrent congestions, the following two steps were further used to identify the spatial and temporal impact range of the prior incident.

- *Step 2* – 5-min speed data for the same time and location in the step 1 was extracted from incident-free days in one year. For example, the prior incident in Figure 15a occurred at the time of 12:30 am on June 2014 and the milepost of 8.65. Then the speed data were collected for the same time interval and location in Figure 15a from all incident-free days in 2014. Then the speed data for each time and location was averaged over all the incident-free days

- *Step 3* - To account for the potential effects of recurrent congestions, the average speed over incident-free days was subtracted in Step 2 from the speed data for each time and location in Step 1. The differences between speeds in Step 2 and Step 1 for various times and locations were then used to develop a new speed contour plot, which was used to identify the spatial and temporal impact range of the prior incident. Figure 15b illustrates the modified speed contour for identifying the spatial and temporal impact range of a prior incident.

- *Step 4* - The incidents that occurred in the spatial and temporal impact ranges of primary incidents were identified as secondary incidents.
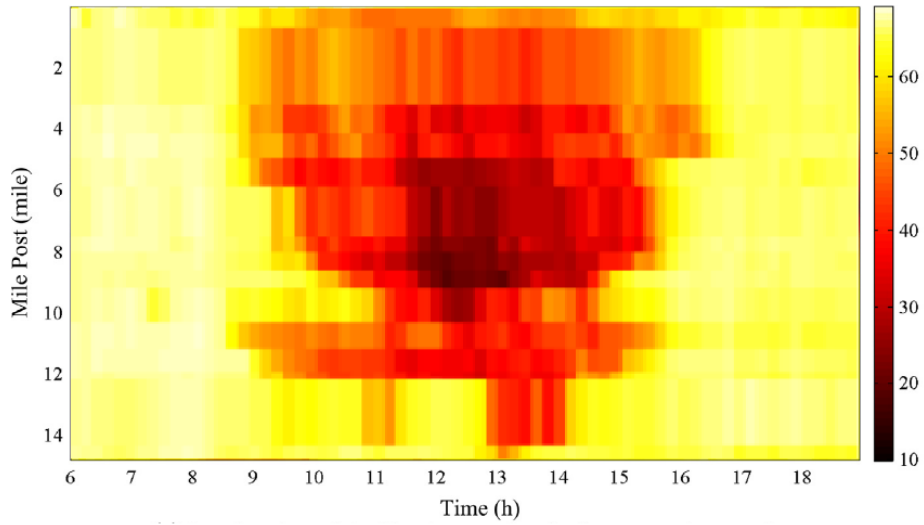
(a) With Recurrent Congestions
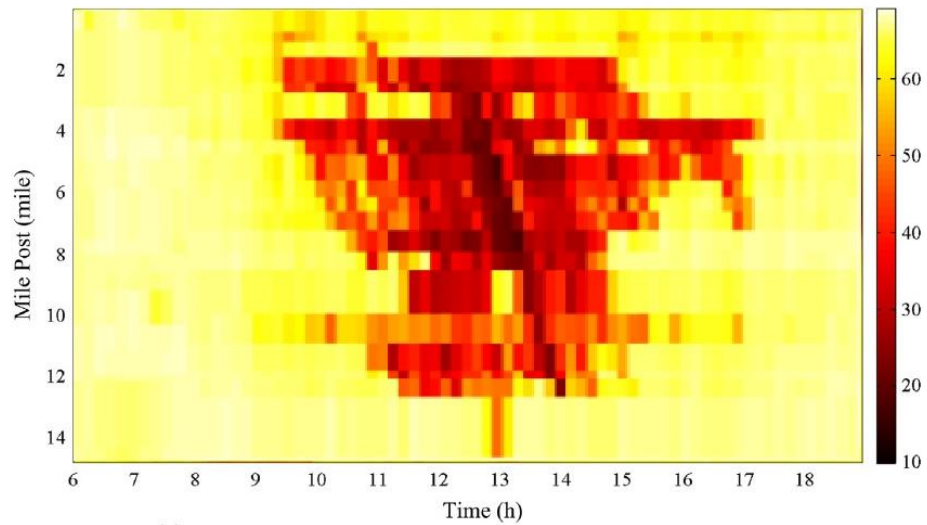
(b) Without Recurrent Congestions

Figure 15: Speed Counter Plot to Identify SCs

3.2 Modeling SC Occurrences and Risk Factors

As discussed in earlier sections, identifying secondary crashes is the most critical step of this research work. The next important step is calculating the probability of secondary crash occurrences and investigating the influential risk factors to SCs.

3.2.1 Statistical Hypothesis Test

A set of selected characteristics, including crash severity, collision type, driver actions, area type, were established to conduct the proportional test. In this test, accepting the null hypothesis statement that there is no statistically significant difference between the primary and the SCs in having a specific characteristic, then they are both the same. In other words, that characteristics have the same contributions to both primary and secondary crashes. The test results are discussed in more detail in the results section.

To determine if it is statistically significant based upon a pre-defined threshold probability (α), a proportional test examines the differences between the primary and the secondary crashes with regards to various characteristics. The fundamental equation to conduct the test (z test) is shown in Equation 21:

$$z = \frac{p_1^\wedge - p_2^\wedge}{\sqrt{P^\wedge(1 - P^\wedge)} * (\frac{1}{n_1} + \frac{1}{n_2})} \tag{21}$$

$$P^\wedge = \frac{x_1 + x_2}{n_1 + n_1} =$$

$$p_1^\wedge = \frac{x_1}{n_1}$$

$$p_2^\wedge = \frac{x_2}{n_2}$$

where

$P^\wedge$ = pooled sample proportion

$p_1^\wedge, p_2^\wedge$ = two compared population proportions

x_1, x_2 = number of successes for populations

n_1, n_2 = population sample sizes

The null hypothesis is when it states that there is no difference between the two population proportions. For a two-tailed test with 0.05 significance level, the null hypothesis is H_0: $p_1^{\wedge} = p_2^{\wedge}$ and the alternative hypothesis is defined as H_1: $p_1^{\wedge} \neq p_2^{\wedge}$ if $z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$.

3.2.2 Structural Equation Models

Structural Equation Modeling (SEM) is a multivariate statistical analysis technique that uses a combination of factor analysis and multiple regression analysis to analyze structural relations (Ullman, 2003). SEM framework (Figure 16) was developed to investigate structural relationships between the contributing factors, the presence of secondary collisions, and injury severity levels. The mean and variance adjusted weighted least squares (WLSMV) were used to estimate the parameters of the SEMs. The chi-square difference test has been conducted to estimate the goodness of fit of the three models. In this study, the SEM was developed for both nighttime and daytime crashes with equal regression coefficients.
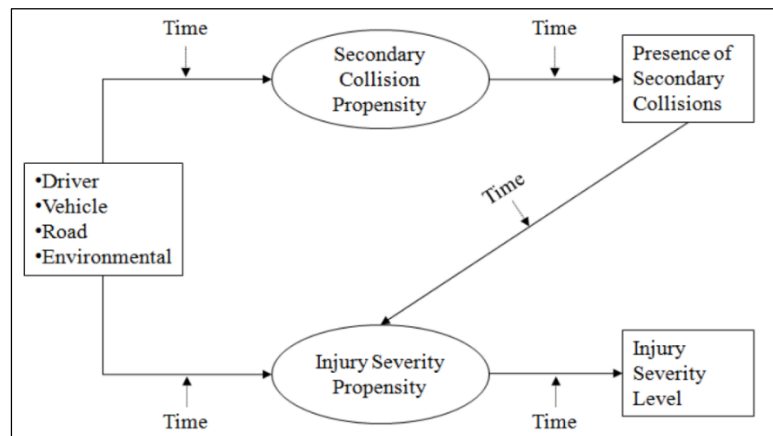
Figure 16: Structural Equation Model Framework (Xie et al., 2016)

The formulation for the proposed SEM is expressed using the following equations. Note that the contributing factors affecting secondary crashes have been identified using

Equation 22, and the effect of secondary collisions on injury severity has been analyzed using Equation 23.

$$sc_i^* = \boldsymbol{\alpha}'\mathbf{Z_i} + v_i; \ v_i \sim N(0,\sigma_v^2)$$

$$sc_i = 1, \text{ if } sc_i^* > \varphi, \ sc_i = 0, \text{ otherwise}$$

(22)

where

sc_i^* = a latent secondary collision propensity in crash, i ($i=1,2,.....,N$);

sc_i = secondary collision indicator (1 represents presence of secondary crash);

α' = vector of coefficients of exogenous variables;

Z_i = all exogenous variables; and

v_i = stochastic error for unobserved factors (independent and normally distributed).

$$y_i^* = \beta'X_i + \gamma sc_i + \varepsilon_i; \ \varepsilon_i \sim N(0,\sigma_\varepsilon^2)$$

$$y_i = k, \text{ if } \eta^{k-1} < y_i^* < \eta^k,$$

(23)

where,

y_i^* =   a latent injury severity propensity for crash, i ($i=1,2,.....,N$)

y_i =   observed injury severity level for crash i (1 = no injury, 2 = possible injury, 3 = non-incapacitating injury, 4 = incapacitating injury, and 5 = fatal)

β' =   vector of coefficients of exogenous variables

X_i =   all exogenous variables

γ =   the effect of the presence of secondary collisions on the injury severity

ε_i =   stochastic error for unobserved factors (independent and normally distributed)

k =   an index to represent injury severity outcome ($k = 1,2,....,K$)

η^k = upper threshold corresponding to the injury severity outcome k ($\eta^0 < \eta^1 .... <$

η^5, $\eta^0 = -\infty$, $\eta^5 = +\infty$)

3.2.3 Multinomial Logit Model

The multinomial logit (MNL) model was used in this research to understand how roadway geometric and non-geometric characteristics affect secondary crashes. MNL models are traditional discrete choice models that consider three or more outcomes; however, MNL models do not explicitly consider the ordering that may be present in these outcomes (Savolainen et al., 2011). Generally, in a multinomial logit (MNL) model, a crash is addressed in terms of injury severity outcomes in the sense that the propensity of crash i towards severity category k is represented by severity propensity function, T_{ki}, as shown in Equation 24 (Kim et al., 2008; Ye and Lord, 2014).

$$T_{ki} = \alpha_k + \beta_k X_{ki} + \varepsilon_{ki}, \tag{24}$$

where

α_k = a constant parameter for crash severity category k;

β_k = a vector of the estimable parameters for crash severity category k that follows KABCO; $k = 1,...,K$; ($K = 3$ in this research) representing no injury or property damage only/O ($k = 1$), possible injury/C and non-incapacitating injury/B ($k = 2$), and incapacitating injury/A and fatal injury/K ($k = 3$);

X_{ki} = a vector of explanatory variables affecting the crash severity of i at a severity category k (incident, roadway and traffic characteristics);

ε_{ki} = a random error term following the Type I generalized extreme value (i.e., Gumbel distribution) where, $i = 1, ..., n$; n is the total number of crash events

Equation 25 shows how to calculate the probability for each crash severity category. Let $P_i(k)$ be the probability of crash i ending in crash severity category k (Ye and Lord, 2014), such that

$$P_i(k) = \frac{\exp(\alpha_k + \beta_k X_{ki})}{\sum_{\forall k} \exp(\alpha_k + \beta_k X_{ki})}. \tag{25}$$

The coefficients, β_k, can be estimated by the maximum likelihood method. The MNL models are derived with the assumption that the unobserved factors are uncorrelated over the alternatives or outcomes which is known as the independence from irrelevant alternatives (IIA) (Train, 2009; Ye and Lord, 2014). This assumption limits MNL in the sense that it is very likely that the unobserved factors are shared by some outcomes. However, the IIA assumption makes MNL very convenient to use (Ye and Lord, 2014). Another assumption of MNL is worth to note that the error term is considered identical and independently distributed (IID) (Ben-Akiva and Lerman, 1985).

Underreported data is supposed to be biased when a crash prediction model is developed. But usually lower severity crashes (PDO) are likely to be underreported that may lead overestimation for higher severity and underestimation for lower severity crashes (Ye and Lord, 2011). Since estimates of exogenous variables remain unbiased in the MNL model, the log-odds ratio of outcomes is estimated. Only the difference in coefficients is identifiable by keeping the coefficients of one outcome are set to be the base category (zero value). The zero-value base category can solve this indeterminacy (Carson and Mannering, 2001). As estimated coefficients of the independent and dependent variables are not correlated in the MNL model, odds ratio (OR) is estimated relative to the base category.

3.2.3 Model Validation

There are three types of model goodness of fit were employed to assess the performance of above-mentioned models, SEMs in particular. They are named as the root mean square error (RMSEA), the comparative fit index (CFI), and Tucker Lewis index (TLI). The RMSEA is computed based on the chi-square statistic, but it considers the model complexity by including degrees of freedom. The RMSEA ranges from 0 to 1, with a smaller value indicating a better fit. Generally, a model with RMSEA less than 0.05 is favored (Markus, 2012).

On the other hand, CFI and TLI measure the relative improvement in the fit of the proposed model over that of a baseline model (null model with no explanatory variables). The CFI or TLI with the value higher than 0.9 indicates a good fit of model (Hu and Bentler, 1995). Equations (26, 27, 28) of RMSEA, CFI and TLI are mentioned below:

$$RMSEA = \sqrt{\frac{\chi_M^2 - df_M}{df_M(N-1)}} \tag{26}$$

$$CFI = 1 - \frac{\chi_M^2 - df_M}{\chi_B^2 - df_B} \tag{27}$$

$$TLI = \frac{\chi_B^2/df_B - \chi_M^2/df_M}{\frac{\chi_B^2}{df_B} - 1} \tag{28}$$

where

χ_M^2 = the chi-square statistic for proposed model, M

df_M = degrees of freedom for model M

χ_B^2 = the chi-square statistic for base model, B

df_B = degrees of freedom for model B; and

N   = sample size

3.3 Development of Crash Modification Factors

The cross-sectional analysis was used in this study to develop Crash Modification Factors (CMF) for rear-end secondary crashes (RSCs) on two-lane undivided roadways in District six areas in Florida. A CMF is a multiplicative factor, used to compute the expected number of crashes when a specific countermeasure is implemented at a specific site. A CMF of 0.9 indicates a 10% expected reduction in crashes, while a CMF of 1.3 indicates a 30% expected increase in crashes (Washington et al., 2010). Cross-sectional studies are recommended for CMF estimation when before-after studies cannot be conducted due to insufficient before and after crash data when a particular engineering countermeasure is implemented, or the date of the implemented treatment is unknown; or when it is difficult to distinguish the effect of a countermeasure from confounding factors.

In cross-sectional studies, crash experience at locations with and without a specific feature is studied; and then the difference in safety is attributed to that feature. To obtain reliable results from cross-sectional studies, it is critical that all locations are similar to each other in all other factors affecting crash risk. However, in practice, it is difficult to collect data for enough locations that are similar in all other factors affecting crash risk. Therefore, cross-sectional studies are often conducted through multiple variable regression models. The models are developed using crash data from sites both with and without the treatment. The change in crashes from a unit change in a specific variable can be estimated from regression model. The CMFs are then deduced from the model parameters (Washington et al., 2010). This research used the generalized linear model (GLM) approach with the negative binomial distribution (NB) to develop the relevant regression models. The models have crash frequency as the explanatory variable, and the roadway

characteristics as explanatory variables. Equation 29 illustrates the basic form of the regression model used in this study.

$$Y_i = exp\left(\beta_0 + \beta_1 \times ln\,AADT_i + \beta_2 \times CW_i + \beta_3 \times GR_i + \ldots + \beta_k \times X_{ik} + OFFSET\right) \tag{29}$$

where

Y_i	$=$	crash frequency on a road section i,
$AADT_i$	$=$	average annual daily traffic on a road section i (vehicle/day),
CW_i	$=$	width of a carriageway section i (ft),
GR_i	$=$	presence of guardrail along a road section i (0 if absent, 1 if present),
X_{ik}	$=$	roadway characteristic k (i.e., countermeasure) of road section i,
β_0	$=$	model intercept/constant,
$\beta_1, \beta_2,\ldots, \beta_k$	$=$	model coefficients, and
$OFFSET_i$	$=$	ln (3×(section length of road section i, i.e., SL_i)) for segments. Note

that the number *3* was used in the offset term because this study considered four years of crash data.

The regression coefficients and over-dispersion parameters were estimated using the glm.nb function of MASS package in the statistical software R (Kim et al., 2008). An offset term was added to the regression equation to predict the crash frequency in crashes per mile per year for segments, as shown in Equation 29. The CMFs can be inferred from the estimated model parameters, i.e., coefficients. As the model form is log-linear, the CMFs can be calculated using Equation 30.

$$CMF = exp\left(\beta_k\right) \tag{30}$$

CHAPTER 4

DATA PREPARATION

The research includes various and intensive databases to address the specific objectives, as mentioned in Section 1.3. The entire dissertation work is conducting the analysis of roadways within the District 6 of Florida, includes Miami-Dade and Monroe Counties (Figure 17). Crash Analysis Reporting (CAR) system database was explored for static approach to identify SCs. However, Regional Integrated Transportation Information System (RITIS) and SunGuide™ database were used to identify SCs using a dynamic approach. The following section explains the detail of databases and data preparation utilized in this research work.
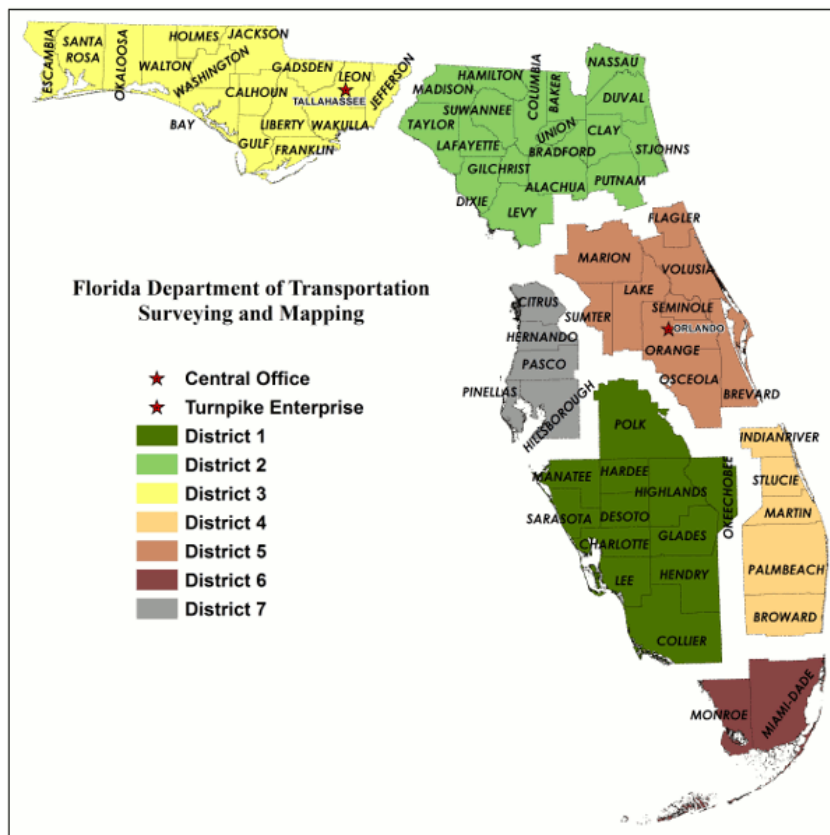
Figure 17: Study Area

4.1 Crash Data

Crash data were obtained from FDOT's Crash Analysis Reporting (CAR) repository for the years of 2011-2014. CAR database was utilized while using a static approach to identify SCs. The CAR database includes three levels of data files: crash level file, vehicle-driver-passenger level file, and non-motorist level file. The crash level file contains crash-related information such as crash number, roadway ID where the crash occurred, milepost of the crash location, crash severity, etc. The vehicle-driver-passenger file includes the road user-related information for each crash record, thus having information on the crash number, all vehicles involved in the crash, all drivers and passengers involved in the crash, etc. The non-motorist level data file includes information about each non-motorist involved in a crash such as crash number, type of non-motorist, non-motorist location, non-motorist injury severity, etc.

4.2 Roadway Characteristics Inventory Database

Detailed roadway characteristics information was extracted from the 2014 FDOT's Roadway Characteristics Inventory (RCI) database. The RCI is mainly used to identify the type of road configuration, the geometrics of roadway segments and intersections, e.g., overall surface lane width, number of lanes, shoulder type and width, median width, maximum speed limit, and other roadway and traffic characteristics. Figure 18 includes a screenshot of the RCI query list (Park et al., 2015).

Figure 18: A part of Roadway Characteristics Inventory Database Query List

Among more than 200 variables in the RCI database, the following variables were extracted for the Crash Modification Factors (CMFs) development of the RSCs on two-lane undivided roadways: pavement surface width (SURWIDTH), average annual daily traffic (SECTADT), shoulder type (SHLDTYPE), presence of guardrail (MLTRFSEP), max speed limit (MAXSPEED), presence of horizontal curve (CURCLASx), functional classification of roadways (FUNCLASS). Segmentation was conducted with only these variables. Segmentation was performed per the guidelines provided in the Highway Safety Manual (Celik et al., 2014). It ensured that all necessary segments have similar variables. According to AASHTO, segmentation is important to achieve homogenous characteristics in each of the segments in parallel to each variable selection to the analysis while developing CMF (AASHTO, 2010).

Crash database was merged with the roadway segments based on crash location and then linear referenced in ArcGIS. The shapefiles for roadways were collected from FDOT website. It ensured that all the segments had data for all the variables and segments with missing data for any variable were excluded from the analysis.

4.3 Traffic Incident Data from SunGuide Software

SunGuide™ is an Advanced Traffic Management System (ATMS) software uses for incident management to process and archive incident data on freeways. Traffic incident data is required to evaluate the incident verification and response durations. The District 6 SunGuide™ database was used to identify secondary crashes in freeways, I-95 corridor in particular. A total of 21,589 incidents were extracted to identify SCs using the dynamic approach. Figure 19 shows the selected four corridors (I-75, I-95, I-195 and I-395) in District 6 for SCs analysis on freeways using the dynamic approach. The incident zone only includes Miami-Dade County areas. For this study, the following information was retrieved from the SunGuide™ database for the years 2014 to 2017.

- Event ID

- Roadway

- Latitude and longitude of the event location

- Incident notification time

- Incident clearance duration

- Event type (i.e., crash, debris on roadway, disabled vehicle, emergency vehicles, flooding, pedestrian, police activity)

- Time of event

Figure 19: Selected Four Corridors in Miami-Dade County for SCs Analysis

4.4 Regional Integrated Transportation Information System Database

Regional Integrated Transportation Information System (RITIS) is situational awareness, data archiving, and analytics platform that has been used by transportation officials, first responders, planners, researchers, and more. RITIS fuses data from different agencies, systems, and even the private sector, enabling effective decision-making for incident response and planning. Within RITIS are a broad portfolio of analytical tools and features. Ultimately, RITIS allows a wide range of capabilities and insights, reduces the

cost of planning activities and conducting research, and breaks down the barriers within and between agencies for information sharing, collaboration, and coordination.

The RITIS database also includes many performance measures, dashboards, and visual analytics tools. These tools assist agencies in gaining situational awareness, measure performance, and communicate information between agencies and to the public. This database was explored while using the dynamic approach to identify SCs as like as SunGuide™ database. Particularly, raw speed and traffic data were retrieved from RITIS database for I-95 corridor within Miami-Dade County in Florida for the year 2014-2017.

CHAPTER 5

RESULTS AND ANALYSIS

5.1 Analysis of SCs Utilizing Static Approach

The static approach uses fixed spatiotemporal thresholds to identify SCs. In this research, 2 miles-2 hours threshold has been used to identify SCs in District 6 for the year of 2011-2014. A total of 4,966 SCs were identified, which is nearly 1.49% of total crashes within these four years. A set of selected characteristics, including crash severity, collision type, primary crash (PCs) contributing factor, driver actions, area type, highway classification, were established to conduct the proportional test and null hypothesis.

Furthermore, rear-end SCs (RSCs) were identified from the crash level data file using the first harmful event (FRST_HARM_EVNT_CD) description and manner of collision (IMPCT_TYP_CD). A total of 1,992 crashes were identified as RSCs crashes. Afterward, a total of 561 crashes were identified as RSCs in two-lane undivided roadways to develop Crash Modification Factors (CMFs).

5.1.1 Investigating the Primary and Secondary Crash Characteristics

As secondary crashes occur after PCs, it is also vital to a thorough investigation of the characteristics of PCs. To conduct this further observation, the proportional test, as described in Chapter 3, a null hypothesis (i.e., the proportion of primary crashes by primary crash contributing factor is not statistically significantly different from the proportion of SCs of the same classification) was considered. In this analysis, the 5% significance level (p-values <0.05) was considered to reject the null hypothesis. Table 3 includes only statistically significant variables. There are several variables have been found to be insignificant in this analysis e.g., careless driving, failed to yield right of way, improper

boarding and turning, traffic rules and signals violation, wring way driving, and other contributing action. According to obtained results, compared to the primary crashes, the secondary crashes are more likely to be associated with "following too closely" and "driving too fast" (Table 3) variables. Therefore, "following too close" and "driving too fast" are counted as the prominent type of contributing factors in the SCs, which is relevant to the previous studies (Tian et al. 2015, 2016). As for "driving under influence (DUI)" the p-value less than 0.05 indicates that the difference between the percentage of the primary and the secondary crashes caused by this factor is statistically significant. Since these percentages in the SCs are lower than those in the primary crashes, DUI cannot be considered as a major contributing factor in the SCs for further investigation.

Table 3: Driver Actions Using Proportional Test

	Year	Driving Too Fast	DUI	Following Too Close
Primary Crash (%)	2011	4.4	1.2	12.6
	2012	5.1	1.5	11.5
	2013	5.2	1.9	19.6
	2014	5.3	2.1	15.2
Secondary Crash (%)	2011	8.1	0.5	19.6
	2012	9.3	0.6	22.1
	2013	7.1	0.8	23.4
	2014	12.1	0.3	26.1
p-value	2011	0.000	0.000	0.000
	2012	0.000	0.000	0.000
	2013	0.000	0.000	0.000
	2014	0.000	0.000	0.000

Table 4 shows the distribution of PCs and SCs with respect to three different collision categories e.g., head-on, rear-end, side-swipe. The insignificant collision types include angle, front to rear, rear to side, and others as well. Table 4 clearly shows that the rear-end collision is the predominant type of crashes which is relevant to the previous studies (Mishra et al., 2016, Tian et al., 2015, 2016, Hirunyanitiwattana and Mattingly 2006, Zheng et al., 2014, 2015, Yang et al., 2014, Jalayer et al., 2015). The proportion of total primary crashes for rear-end collision types and their differences found to be statistically significant. Moreover, from Table 3, it was found that a higher likelihood of SCs was associated with "driving too fast" and "following too close" factors. The outcome is similar in rear-end crashes. Additionally, the percentages of other types of collisions (head-on and side-swipe) are higher in the PCs compared to the SCs, and their differences are statistically significant.

Table 4: Collision Type Using Proportional Test

	Year	Head-On	Rear-End	Side-Swipe
Primary Crash (%)	2011	7.1	50.3	22.1
	2012	8.9	51.1	23.4
	2013	9.1	55.2	25.9
	2014	10.1	59.1	26.1
Secondary Crash (%)	2011	4.4	66.1	24.2
	2012	5.1	68.2	25.3
	2013	6.3	66.1	26.1
	2014	7.9	67.1	24.5
p-value	2011	0.000	0.000	0.000
	2012	0.000	0.000	0.000
	2013	0.000	0.000	0.000
	2014	0.000	0.000	0.000

As shown in Table 5, unlike the fatal crashes, the percentage of the primary and the secondary crashes for all four years statistically differ from each other for injury and property damage only (PDO) severity types, meaning the secondary crashes were as fatal as the primary crashes in 2011 and 2012. The result reveals that the PDO crashes are the prevalent severity type for both primary and secondary crashes (Table 5). Hirunyanitiwattana and Mattingly 2006 and Jalayer et al. 2015 also found that PDO crash level is the prominent type of severity in their research works. Moreover, the proportion of SCs is significantly higher and lower than primary crashes for PDO and injury severity type, respectively.

Table 5: Crash Severity Using Proportional Test

	Year	Fatal	Injury	PDO
Primary Crash (%)	2011	0.6	31.5	60.1
	2012	0.7	35.6	58.7
	2013	0.9	35.7	57.2
	2014	1.0	37.1	55.4
Secondary Crash (%)	2011	0.4	29.1	65.4
	2012	0.4	25.2	61.7
	2013	0.5	30.1	59.2
	2014	0.6	29.5	57.3
p-value	2011	0.062	0.001	0.000
	2012	0.022	0.000	0.001
	2013	0.031	0.002	0.003
	2014	0.215	0.000	0.000

As per Table 6, most crashes occurred in the urban area with a higher proportion of the SCs similar to previous studies (Hirunyanitiwattana and Mattingly 2006, Zheng et al., 2014, Chimba et al., 2014). The hypothesis test results indicate that the differences between the primary and secondary crashes for both area types are significant, with a higher proportion of SCs in urban areas. Table 6 shows that nearly 30 percentage points more SCs occur on urban roadways compared to rural roads. It is also found that nearly 1.56% of total crashes is identified as SCs on urban arterial roads, whereas only 0.61% of total crashes occur on rural arterial roads.

Table 6: Crash Area Type Location Using Proportional Test

	Year	Urban	Rural
Primary Crash (%)	2011	51.4	40.1
	2012	52.1	41.1
	2013	53.9	41.1
	2014	56.5	42.5
Secondary Crash (%)	2011	60.9	32.1
	2012	62.8	32.5
	2013	64.3	33.1
	2014	65.1	34.2
p-value	2011	0.001	0.001
	2012	0.000	0.000
	2013	0.000	0.000
	2014	0.001	0.001

5.1.2 SEM Estimation Results

As mentioned in the methodology section, the SEM framework is used to model the presence of secondary crashes and injury severity levels (no injury, possible injury, non-incapacitating injury, incapacitating injury, and fatal). A total of 4,966 SCs were identified for the years of 2011 to 2014 from the FDOT's Crash Analysis Reporting (CAR) database for District 6 in Florida. The study explored all the three levels of CAR data files: crash level file, vehicle-driver-passenger level file, and non-motorist level file. The injury severity model was estimated for the four categories independent variables e.g., driver, vehicle, roadway and environmental features. Each of the categories includes multiple independent variables to investigate their effects on the injury severity of SCs. The "driver" category includes driver behavior and actions, the "vehicle" category includes vehicle condition and maneuvers, the "roadway" explores roadway geometric, and non-geometric features of roads and the "environment" category consider weather and lighting conditions.

The study considered different sets of regression coefficients for both daytime and nighttime crashes while developing the SEM model. The root mean square error (RMSEA) for this SEM is found to be less than 0.05 (0.032), and the comparative fit index (CFIs) and Tucker Lewis index (TLIs) are higher than 0.9 (0.912 and 0.907 respectively). Both the estimated values confirmed that the model was well fitted.

Table 7 reveals the estimation results. All the explanatory variables observed as statistically significant at 5% significance level (p-values <0.05). According to Table 7, nine explanatory variables found to contribute to the presence of secondary crashes. Drivers under the influence of alcohol are more likely to get involved with secondary collisions relative to those who are attentive drivers. Because the alcohol would affect the

80

judgment, reasoning, and reaction of drivers. It is also found that drinking alcohol may lead to aggressive driving, thus severe injuries, which is consistent with the earlier studies (Xie et al., 2016).

Table 7: SEM Model Estimation Results

	Daytime				Nighttime			
	Secondary Crashes		Injury Severity		Secondary Crashes		Injury Severity	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
Secondary Crashes	-	-	0.561	0.000	-	-	0.510	0.000
Driver								
Alcohol	0.812	0.000	0.419	0.000	0.493	0.000	0.490	0.000
Drug	0.795	0.000	0.852	0.000	0.594	0.041	0.478	0.030
Inattentive	0.312	0.000	0.111	0.000	0.123	0.000	0.091	0.000
Sleep	0.445	0.000	-	-	0.401	0.023	0.111	0.000
Speeding	0.811	0.000	0.488	0.000	0.721	0.000	0.410	0.000
Vehicle								
Defective Brake	0.605	0.000	0.591	0.000	0.791	0.000	0.461	0.000
Motorcycle Involved	-	-	1.325	0.000	-	-	1.512	0.000
Roadway								
Carriageway Surface	0.531	0.002	0.300	0.000	0.501	0.000	0.259	0.000
Obstructed Vision	0.321	0.032	-	-	0.259	0.011	-	-
Environment								
Rain	0.112	0.000	-	-	0.108	0.021	-	-
Threshold Values								
η^1	-	-	0.311	0.000	-	-	0.051	0.000
η^2	-	-	1.832	0.000	-	-	1.705	0.000
η^3	-	-	2.110	0.000	-	-	2.054	0.000
η^4	-	-	3.513	0.000	-	-	3.291	0.000
φ	1.435	0.000	-	-	1.222	0.000	-	-

For crashes caused by distracted driving, the likelihood of secondary collisions and severe injuries expected to be higher. Intuitively, drivers who fall asleep can lead to more chances of secondary crashes since they couldn't react in time even after the initial

collisions. However, sleeping causes more severe crashes at night (daytime data is missing in the table). Speeding also found to be associated with higher severe injury propensity, a result observed in the previous studies (Hirunyanitiwattana and Mattingly 2006).

Vehicles with brake defects tend to be exposed to secondary collisions and severe injuries since those vehicles couldn't stop fast enough. Khattak et al. (2007) found defective truck brakes were significantly associated with severe injuries.

The likelihood of SCs is more at the roadways with pavement defects and limited views. Pavement defects can also increase the risk of severe injuries. Furthermore, on rainy days, more SCs found to occur due to the slippery roadway and obstructed vision. This finding is also consistent with the previous studies (Vlahogianni et al., 2012, Xu et al., 2019 and Xie et al., 2016).

5.1.3 MNL Model Estimation Results for Rear-End SCs

From the above results, it is clear that rear-end is the most prominent collision type of SCs (Tables 3 and 4). Therefore, the Multinomial logit model (MNL) is developed to explore the most significant roadway geometric characteristics on the severities of rear-end SCs (RSCs). Rear-end crashes from 2011-2014 were identified from the crash level data file. From 2011-2014, a total of 1,992 crashes were identified as RSCs. KABCO (K=Fatal, A=Incapacitating Injury, B=Non-incapacitating Injury, C=Possible Injury, and O=Property Damage Only) severity levels were grouped into three categories K+A, B+C, and O for severity analysis.

Before fitting the model, the random forest technique was conducted to investigate the most important independent variables so that the model would well fitted appropriately and found that the out-of-bag (OOB) error across decision trees (500 prunes) is sufficient to

rank the importance of explanatory variables. The random forest technique is a popular learning tool was first proposed by Leo Breiman (Breiman, 2001). It generates a large number of decision trees to identify the random parameters in the model. Gini Index is utilized to identify most important variables before estimating the model parameters. The plot of out-of-bag (OOB) error across decision trees used to check whether the assumed pruned number of decision trees is sufficient to rank the importance of explanatory variables. The random forest tool eliminated the ten independent variables e.g., shoulder type, presence of roadside fixed objects (trees, utility poles, and others), bridge, culvert, curves, guardrail, median.

However, the estimated coefficients that are not found to be significant at the 5% level of significance are excluded from the Table 8. The goodness of fit of the model is measured by log-likelihood at convergence, log-likelihood at fitting the intercept, Likelihood Ratio Index, Akaike Information Criterion (AIC) and Schwarz Criterion (SC) values. The model is well fitted at 0.1 adjusted R^2 (Miaou et al., 1996). In addition, the model fitted the data properly with a large likelihood ratio (Chi-Square statistic = 951.3) and a very small P-value (<0.0001). Besides maximum likelihood estimates, the model output also includes the proportional odds ratio . The estimated results for the multinomial logit model are represented in Table 8.

Table 8: MNL Model Estimation Results

Predictor Variables[a]	Estimates	Odds Ratio (OR)
Constant [B+C]	0.789	2.20
Constant [K+A]	-1.985	0.14
Presence of Horizontal Curve [B+C]	0.086	1.09
Presence of Horizontal Curve [K+A]	0.139	1.15
Presence of Guardrail [B+C]	-0.152	0.85
Presence of Guardrail [K+A]	-0.174	0.84
Speed Limit (> 50 mph) [B+C][a]	0.166	1.18
Speed Limit (> 50 mph) [K+A][a]	0.191	1.21

Note: B+C stands for possible and non-incapacitating injury; K+A stands for fatal and severe injury; The base condition for injury severity is property damage only (O) crashes; [a]The base condition for speed limit is ≤ 30 mph.

The presence of horizontal curves is more likely to have more severe injuries from RSCs, as can be observed from Table 8. The coefficients for the presence of horizontal curve for possible and non-incapacitating (B+C) injuries and for fatal and severe (K+A) injuries found to be positive, and the corresponding ORs were estimated as 1.09 and 1.15 respectively. It indicates the probability of B+C and K+A crash probability over property damage only crashes as 1.09 and 1.15 times, respectively, where horizontal curves are present. The presence of guardrails reduces the K+A crash probability compared to property damage only crashes (OR = 0.84). Higher speed limits resulted in more severe crash risk ratios.

5.1.4 Cross-Sectional Study for CMFs of RSCs on Two-Lane Undivided Roadways

The rear-end secondary crash data from the previous section was then merged with the roadway segment database so that each site had the total number of RSCs that occurred during 2011-2014 to develop Crash Modification Factors (CMFs) for the two-lane undivided roadways. Table 9 presents the descriptive statistics of the variables for the two-lane undivided roadway segments mentioned in this study.

Table 9: Descriptive Statistics of the Roadway Characteristics Variables

Attribute	Attribute Category	Value
Total Roadway Segment Length (in miles)	--	431.1
Section AADT (vehicle/day)	Mean	10,549
	Standard Deviation	4,501
Carriageway Width (ft.)	Mean	27.60
	Standard Deviation	1.61
Shoulder Type (in miles)	Paved	366.6
	Lawn, Gravel/Marl	51.10
	Curb and Gutter	5.20
Presence of Horizontal Curve (in miles)	No	407.4
	Yes	23.70
Presence of Guardrail (in miles)	No	357.4
	Yes	73.70
Speed Limit (in miles)	≤ 30 mph	20.10
	30-50	50.20
	> 50 mph	355.8

At first, the negative binomial distribution models were developed by considering the following variables: AADT, Carriageway width, shoulder type, presence of curve, presence of guardrail, and speed limit. The variables (presence of curves, guardrails, carriageway width, speed limit) that found significant associated with the 5% level of significance, were further used to develop the final NB regression models. The CMFs were then estimated from the final models. Table 10 provides the CMFs. Considering the limited variability in the predictors within the dataset, the 95% confidence interval (CI) was considered acceptable while developing the NB regression models.

Table 10: RSCs CMFs for Two-lane Undivided Roadway Segments

Variables	Total Crashes	
	Coefficients	CMFs
Presence of Curve	0.3293	1.39
Presence of Guardrail	-0.0101	0.99
Carriageway Width	-0.1393	0.87
Speed Limit (> 50 mph)[a]	0.3436	1.41

Note: [a]The base condition for speed limit is $\leq$ 30 mph.

The presence of curves at roadway segments is found to have a coefficient of 0.3293 and a CMF of 1.39. It can be inferred from the CMF that the presence of curves increases RSCs crash probability by 39% per year per mile on two-lane undivided roadway segments in Florida. From Table 8, it is clear that the presence of curve increases the likelihood of RSCs occurrences which is also established in Table 10. The presence of guardrails is found to reduce RSCs occurrences by nearly 1%. An increase in carriageway width is found to have a positive impact on reducing the RSCs probabilities by 13% as they provide better

vehicle maneuvering options. Higher speeds found in high crash probabilities as perceived about 41%.

5.2 Analysis of SCs Utilizing Dynamic Approach

This section includes results of SCs on freeways utilizing the dynamic approach as per Chapter 3. A total of 21,589 incidents for the year of 2014-2017, were extracted from the SunGuide™ incidents database to identify SCs using speed contour plot (SCP) dynamic method. Table 11 shows the numbers of identified secondary, primary, and total incidents on the four corridors in Miami-Dade County in Florida (Figure 19). It is also found that the number of SCs is higher while using the static approach than the dynamic approach. It is also observed that using the static approach, the SCs rate is 4.6% of total crashes on freeways.

Table 11: Incident Distributions Using SCP Method Among Four Corridors

Freeway	Total Incidents	Number of Total Crashes	[a]Number of SCs Using Dynamic Approach	[b]Number of SCs Using Static Approach
I-95	18,160	8,141	105	363
I-195	1,038	534	6	23
I-395	1,211	517	5	29
I-75	1,180	586	9	38
Total	21,589	9,778	125	453

Note: [a]Utilizing SCP dynamic approach; [b]Utilizing static approach with 2 miles–2 hours fixed spatiotemporal thresholds

To validate, the result of secondary crash identification compared with those of the latest publications. Table 11 reveals that using dynamic method the ratio of identified

secondary crashes to the total crashes is within 1–1.5%, which is consistent with the findings of the latest publications (Park and Haghani, 2016; Sarker et al., 2015; Mishra et al., 2016). The following section includes the comprehensive SCs analysis for the I-95 corridor as it has higher crash rates than other corridors.

Figure 20: Secondary Crash Analysis on I-95 Corridor

Descriptive Statistics for I-95 Corridor

Figure 21 shows that 45% of total incidents on I-95 corridor are related to the vehicle to vehicle crash type. Among them, a total of 105 secondary crashes were identified using the dynamic approach for I-95 corridor for four years of incident data (2014 to 2017).

And 363 SCs by utilizing the static approach, which is nearly double than the identified

SCs by the dynamic approach. Among them, 70% SCs involve with PDO crashes, 25%,

and 5% are injury and fatal crashes, respectively.
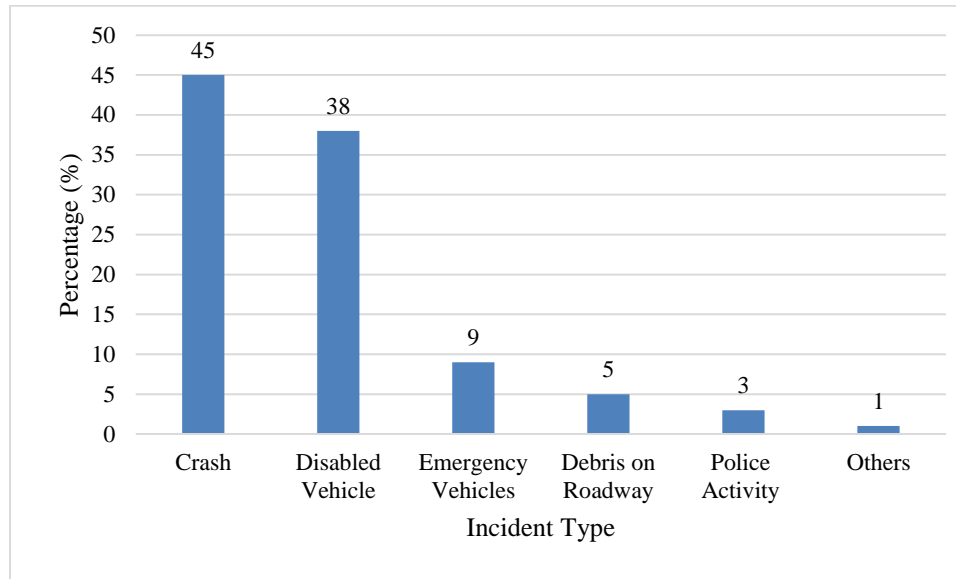
Figure 21: Percentage of Incidents by Incident Type on I-95

Figures 22 and 23 represent the distribution of SCs occurrences by month and day

of the week, respectively. SCs are more likely to occur in October and less in February.

The number of SCs is more likely to occur on Friday and less in Sunday. To be noted that

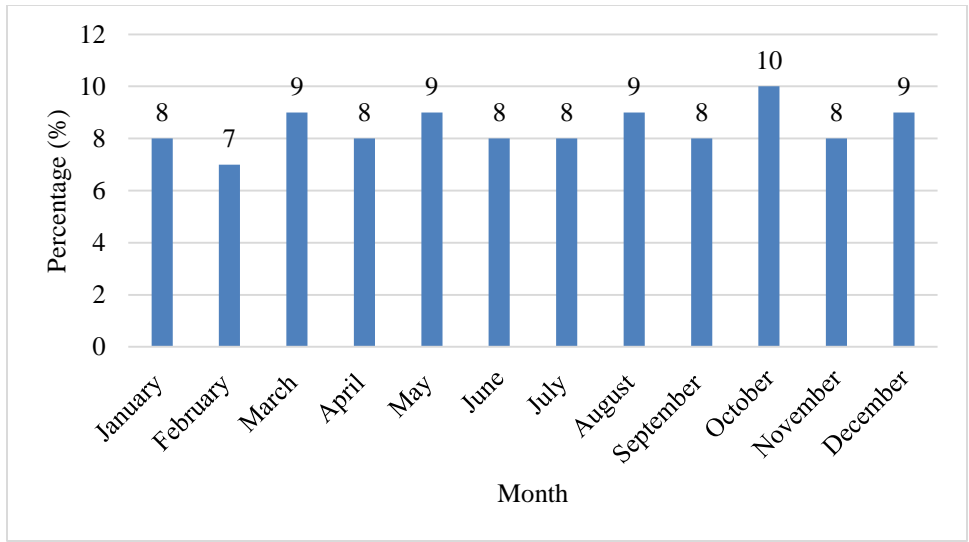the number of SCs is very close in every days of the week.
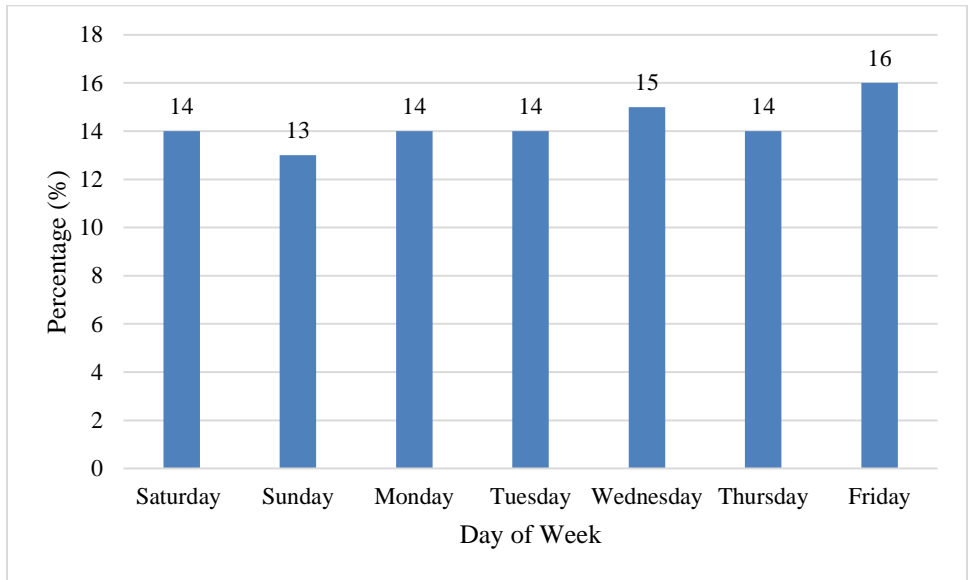
Figure 22: Percentage of SCs by Month

Figure 23: Percentage of SCs by Day of Week

Table 12 shows the incident characteristics on I-95 corridor. It reveals that SCs are more likely to occur if incident clearance times are longer.

Table 12: Incident Characteristics on I-95 Corridor

Incident Type	Number of Total Incidents	Percentage of SCs (%)	Average Incident Clearance Time (min)
Crash	8,141	97	69.1
Emergency Vehicles	1,600	2	41.98
Disabled Vehicles	6,832	1	36.96
Debris on Roadway	947	0	17.76

5.2.1 MNL Model Estimation Results for I-95 Corridor

Multinomial logit (MNL) model was developed for analyzing SCs risks on injury severity (no injury, possible injury and non-incapacitating injury, incapacitating injury and fatal) by the categorical covariates of temporal (crash date), environmental (lighting condition) and SCs occurrence time. Incident occurrence time defines as peak (0600 to 1000 and 1530 to 1830) and off-peak (other times of day).

The estimated results for the multinomial logit model are represented in Table 13. The estimated coefficients that are not found to be statistically significant at the 5% level of significance are excluded from the table. The other insignificant independent variable was the months of the year. The goodness of fit of the model is measured by the Akaike Information Criterion (AIC) and Schwarz Criterion values. The model is well fitted at 0.1 adjusted R^2. The model was also able to be fitted the data properly with a large likelihood ratio (Chi-Square statistic = 1,121.2) and a very small P-value (<.0001). Besides maximum likelihood estimates, the model output also includes standard errors, the Wald Chi-Square,

proportional odds ratio (the coefficients are exponentiated), and the 95% confidence

intervals for the proportional odds ratios (OR).

Table 13: MNL Model Results for I-95 Corridor

Predictor Variables[a]	Estimates[b]	OR[c]
Constant (I)	0.928	2.53
Constant (P)	-1.212	0.29
Day of Week		
Friday (I)	-0.511	0.6 (0.4, 0.8)
Friday (P)	-0.481	0.6 (0.4, 0.9)
Saturday (I)	-0.561	0.6 (0.4, 0.8)
Saturday (P)	-0.523	0.6 (0.4, 0.9)
Sunday (I)	-0.624	0.5 (0.3, 0.7)
Sunday (P)	-0.651	0.5 (0.3, 0.8)
SCs Occurrence Time		
Off-peak (I)	-0.842	0.4 (0.3, 0.6)
Off-peak (P)	-0.991	0.4 (0.3, 0.5)
Peak (I)	-1.771	0.2 (0.1, 0.2)
Peak (P)	-2.361	0.1 (0.1, 0.2)
Lighting Condition		
Daylight (I)	-0.990	0.4 (0.2, 0.7)
Daylight (P)	-1.123	0.3 (0.2, 0.7)
Night (I)	-1.031	0.4 (0.2, 0.5)
Night (P)	-0.978	0.4 (0.2, 0.7)

Likelihood Ratio Index (McFadden's Pseudo R2): 0.1
Akaike Information Criterion (AIC): 1213.1
Schwarz Criterion (SC): 1516.3
P-Value: <.0001

Note: [a]All predictor variables indicated as one if yes and zero if not. In parenthesis letter I and P

are for injury and PDO crashes where Fatal is the base case with coefficients restricted at zero.

[b]Standard errors are in parentheses;

[c]Lower, upper limits at the 95% confidence intervals (CI) are in parentheses

Temporal Factors

SCs are nearly two times (OR = 0.5) more likely to result in fatal when the crash occurs on Sunday. On the other hand, it is 1.7 times (OR = 0.6) more likely to have fatal crashes on both Friday and Saturday. So, weekdays are less likely to have fatal crashes than weekends. Moreover, regression coefficients of weekends are negative and significant, implying a significant decrease in injury and PDO crashes on weekends.

Injury severity tends consistently increases in peak hours. This group is twice as likely to result in fatal crashes compared to PDO crashes (OR = 0.1) than injury crashes (OR = 0.2). Off-peak hours are less likely to be involved in fatal crashes than peak hours.

Environmental Factors

From Table 13, it is evident that dark lighting condition has a significant effect on SCs. Of the other categories under the lighting condition variable, the dark condition is ten times more likely to result in fatal crashes in compared to PDO crashes (OR = 0.1) than in the daylight condition (OR = 0.4). Moreover, the negative value of the coefficient indicates that fatal crashes significantly increase in the dark light street conditions.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This study conducted an analysis focusing on secondary crashes with two objectives: identification of secondary crashes (SCs) using spatiotemporal criteria and exploring the contributing risk factors to the identified SCs. The study area includes the roadways within the District 6 of Florida, which includes Miami Dade and Monroe counties. Crash Analysis Reporting (CAR) system database was explored for the "static" approach to identify SCs. The Regional Integrated Transportation Information System (RITIS) and the SunGuide™ incidents database were explored to identify SCs using the "dynamic" approach.

For SC identification, both static and dynamic approaches were modeled. For the static approach, a 2 miles-2 hours fixed spatiotemporal threshold was used, which identified a total of 4,966 SCs, nearly 1.49% of total crashes in the district. Due to a lack of quality speed data, the dynamic method was only applied on freeway segments. For the dynamic approach, speed contour plots (SCP) were developed to identify SCs. For freeway corridors in the district, about 1.30% of total crashes were identified as SCs based on the dynamic approach compared to about 4.6% using the static approach. This indicates that the static approach significantly overestimates SCs, especially for freeway segments.

The analysis showed that 2.43 percentage point more SCs occurred on freeways compared to arterial roadways. This may be attributed to the uninterrupted flow with high speeds on freeways. It implies that the timely and effective traffic incident management (TIM) programs may help reduce SCs significantly. Looking at urban roadways and rural

94

roadways, it was found that nearly 30 percentage point more SCs occurred on urban roadways compared to rural roadways.

This study further developed severity and frequency models to investigate the contributing risk factors to the identified SCs. The analysis revealed that high speed and rear-end collisions were two prominent parameters for both primary and secondary crashes. Nearly 40% of SCs were related to rear-end SCs (RSCs). Based on a Structural Equation Model (SEM), several explanatory variables were identified as significant contributing factors to the occurrences of SCs, including drunk driving, driver fatigue, obstructed vision, surface condition of carriageways, rainy season (slippery roadways), defective brake and speeding. On the other hand, reckless driving (inattentiveness, speedy), defective vehicle conditions, poor roadway conditions, and the rainy season (slippery roadways) were associated with severe SCs.

Focusing on RSCs as a prominent type of collision in SCs analysis, a MNL model was developed to examine the impacts of roadway geometric and non-geometric characteristics on the severity of RSCs. In addition, Crash Modification Factors (CMFs) were developed for RSCs on two-lane undivided roadways based on results from a negative binomial regression model. Results revealed that the presence of horizontal curves, the presence of guardrail, carriageway width, and posted speed limit showed significant impacts on the occurrence of RSCs. Specifically, the presence of horizontal curves and speed limit above 50 mph would increase the probability of RSC by 39% and 41%, respectively. On the other hand, the presence of guardrail would reduce RSC by 1%.

In terms of SCs on freeway segments, model results revealed that weekends were more likely to have fatal crashes than weekdays, and dark conditions were ten times more

likely to result in fatal outcomes for SCs. These findings provided useful information for the development and identification of effective countermeasures to mitigate SCs.

6.2 Research Contributions

This research explored both "static" and "dynamic" approaches to identify SCs. The majority of previous studies only focused on freeway SCs analysis, while this study conducted a comprehensive analysis of SCs for all roadway types. The study also examined the impacts of driver characteristics, roadway conditions, and environmental factors on SC occurrence and severity level. This analysis would provide useful insights in developing countermeasures to reduce SCs and enhance safety, which has great implications in improving travel time reliability, reducing congestion and delay, reducing fuel consumption and emissions, and increasing quality of life.

6.3 Recommendations for Future Research

Future studies to extend this dissertation study could focus on the following aspects:

- The study only investigated the contributing risk factors of SCs. Future research can extend the analysis on examining the effects of primary incident characteristics on secondary crash characteristics.

- Appropriate countermeasure selection is quite important to analyze the reduction of SCs. Future research might need to include a more sophisticated method of developing crash modification factors and validating the study results.

- Future studies can focus on impact analysis of the advanced transportation systems e.g., connected and autonomous technology to mitigate SCs.

6.4 Limitation

The major limitation of this dissertation is that the analysis is limited to the data (District 6 in Florida). Therefore, this study could not make any definite statement from the output.

APPENDIX

Table A 2.1: Summary of Studies on Secondary Crash Identification Methods

Reference	Data (Type, Year, Location)	Spatiotemporal Criteria	SCs of Total Crashes (%)
(Raub 1997)[a]	Incident, 28 days in January 1995, Northern Chicago, Illinois	<1 mile, <Clearance time+15 minutes	15.5% (97/627)
(Karlaftis et al., 1999)[a]	Incident, 1992-1995, Borman Expressway, Northern Indiana	1.5 km, Clearance time+15 minutes,	34.7% (257/741)
(Latoski et al., 1999)[a]	Incident, 1996, Borman Expressway, Northern Indiana	No more than 3 miles upstream, Clearance time+15 minutes	7.7% (689/8,986)
(Zhan et al., 2008)[a]	Incident, 2005-2007, I-95, I-75, I-595, Fort Lauderdale, Florida	2 miles upstream, Clearance time+15 minutes	5.2% (413/7,903)
(Hirunyanitiwattana and Mattingly 2006)[a]	Crash, 1999-2000, California	2 miles, 2 hours	4.3% (15,442/354,854)
(Khattak et al., 2009)[a]	Incident, 2006, Hampton road, Virginia	1 mile upstream, actual duration	2% (764/38,086)
(Moore et al., 2004)[a]	Incident and Loop Detector data, March, May, July 1999 and last week of 1998, California	2 miles (both directions), 2 hours	0.2% (177/84,684)
(Kopitch and Saphore 2011)[a]	Incident, 2008, I-5 segments in Orange county, Southern California	2 miles upstream, 2 hours	5.5% (528/9,549)
(Chang and Rochon 2009)[a]	Incident, 2010, Maryland, Coordinated Highways Action Response Team (CHART)	2 miles, 2 hours; 0.5 mile, 0.5 hour (other direction)	3.6% (702/19,309)
(Tian et al., 2015, 2016)[a]	Crash, Incident data, 2010, Interstate, Florida	2 miles, 2 hours; 2 miles, clearance time + 15 minutes; 2 miles, clearance time + 30 minutes	326, 124 and 137
(Wang et al., 2016)[b]	Crash, Loop detector data, 2010-2012, Interstate, California	Spatiotemporal shockwave with 1 speed turning point	0.4% (209/49,753)

(Jalayer et al., 2015) [a]	Crash, 2010-2013, Alabama	2 miles, 2 hours	N/A
(Zhan et al., 2009)[b]	Incident, 2005-2007, I-95, I-75, I-595, Fort Lauderdale, Florida	Departure Traffic delay	3.2% (255/7,903)
(Sun and Chilukuri 2006, 2007, 2010) [b]	Incident, 2003, I-70, I-270 segments in Missouri	Incident progression curves	7.2% (397/5,514)
(Chou and Miller-Hooks 2009) [b]	Incident and simulated traffic data, 2007, New York	Determine impact area based on simulated speed contour map	3.9% (27/693)
(Vlahogianni et al., 2010, 2012) [b]	Incident, monitor and sensor data, 2007-2008, Attica Tollway, Greece	Identify influential area by ASDA model	16% (279/1,746)
(Yang et al., 2013, 2014) [b]	Crash and sensor data, 2011, New Jersey 2011	Determine spatiotemporal impact by speed contour map	8.9% (100/1,118)
(Imprialou et al., 2013) [b]	Incident, Detectors data, 2007-2009, Attica Tollway, Greece	Real influential area	5.3% (67/1,287)
(Park and Haghani 2016) [b]	Incident, Probe data, I-695	Determine spatiotemporal impact by speed contour map	1.1% (125/1,150)
(Sarker et al., 2015, 2017; Miahra et al., 2016) [b]	Crash, Sensor data, 2010-2012, Shelby county, Tennessee	Queuing shockwave-based	0.6% (570/91,325)
(Chung 2013) [b]	Crash and sensor data, 2001-2002, Orange county, California	Speed contour plot	3.4% (212/6,200)
(Goodle 2017) [b]	Incident, speed, crash, and vehicle probe data, 2014, I-66 corridor,	Speed contour plot	13.8% (340/2,466)
(Xu et al. 2016) [b]	Crash, Traffic Data, 2006-2010, I-880 freeway, California	Speed Contour Plot	1.2% (113/9,188)
(Haghani et al., 2006) [b]	Incident, Simulated traffic data, I-287 corridor	Simulation-based approach	0.7%-0.13%
(Zheng et al., 2014, 2015) [b]	Crash, State Trunk Network data, 2010, Wisconsin	Queuing shockwave-based	1.1% (79/7,034)
(Khattak et al., 2010) [b]	Incident, 2006, Hampton road, Virginia	Queue based method	2.2% (907/41,539)

	Incident, Loop detector data, January 2010 to December 2012, California	Queuing shockwave-based	17.7% (209/1183)
(Junhua et al., 2016) [b]	Incident, Loop detector data, January 2010 to December 2012, California	Queuing shockwave-based	17.7% (209/1183)
(Zhang and Khattak 2010, 2011) [b]	Incident, 2008, Hampton roads, Virginia	Determining queueing model	0.5% (317/61,455)
(Shlayan et al., 2009) [b]	Incident, Simulated traffic data, I-15, I-405, Seattle, WA	Incident progression curves	None
(Xu et al., 2019) [b]	Crash, Loop detector data, 2010 and 2015, I-5 N freeway, California	Speed contour-based	2.2% (214/9,828)

Note: [a]: Static Approach, [b]: Dynamic Approach

Table A 2.2: Summary of Secondary Crash Risk Factors

Models (Authors)	Explanatory Variables	Findings
LR (Junhua et al., 2016)	Crash severity, violation category, weather, tow away, road surface, lighting, traffic volume, duration, shock waves	- Crash processing duration significantly effect on SC occurrence - Tow away indicator, road surface condition, and other parties involved, were insignificant
LR (Zhan et al., 2008)	Number of vehicles involved, number of lanes, primary incident duration, rollover, midday (9:00 to 16:00), AM (6:00 to 9:00)	- The number of vehicles involved in the primary incident, the number of lanes at the primary incident location, the primary incident duration, time-of-day of incident occurrence, and if vehicle rollover occurs during the primary incident impact SC occurrence
LR (Khattak et al., 2010, 2012)	Time of day, weather, location, AADT, detection source, number of vehicles, incident type, lane closed, EMS, right and left shoulder, ramp, predicted incident duration	- A positive significant correlation between SC occurrence and longer primary incident duration, higher AADT, and primary incidents occurred during peak hours
NNs, GM, PM (Vlahogianni et al., 2012)	Duration, collision type, number of lanes, number of vehicles, heavy vehicle, travel speed, hourly volume, rainfall, alignment, upstream and downstream geometry	- Traffic speed, duration of PC, hourly volume, rainfall intensity, and number of vehicles involved in the primary accident were found to be the top five factors associated with SC likelihood - SC are negatively correlated with the number of blocked lanes for PC

		- Lower speed and higher lane volume might increase SCs.
BNN (Park and Haghani 2016)	Different stages of clearance time	-Likelihood of secondary incidents was higher at clearance time of primary incidents from 10 minutes to 20 minutes or more than 75 minutes
MLM (Mishra et al., 2016)	AADT, functional classification of roadway, number of vehicles involved, stream flow, incident type, weather indicator	- SC involved with higher upstream traffic flow - PC with rear end collision type was the predominant factor contributed to SC
BLM (Xu et al., 2016)	Severity, sideswipe, day of week, roadway surface, lane, average traffic volume, average speed, detector occupancy, difference in traffic volume between adjacent lanes	- Likelihood of SC is higher on weekend, roadway with lower number of lane and at morning peak - Sideswipe type primary crashes are less likely to occur SCs than other type of primary crashes -Difference in traffic volume between adjacent lanes also has significant risk effect on SC
GORP (Sarker et al., 2017)	Speed Limit, Number of Lanes, Land use, median type, ramp, High occupancy vehicle (HOV) indicator, AADT, Right shoulder, Segment length	- about 10% increase in AADT, increased SC occurrences by 34.24% - 2 lane roads caused 73% more SCs than roads with 3 or more lanes - Raised median have 267% more SCs compared to without raised median roads.
SEM (Xie et al., 2016)	Driver, vehicle, roadway characteristics, and environmental condition (rain)	- Drunk, inattentive and careless drivers who were reluctant to traffic control signals most likely to cause secondary collisions with higher level of injury severities. - Speeding, defective vehicle significantly increase likelihood of SC - SC with high injury level frequently occur at intersection than mid blocks
TFP (Hirunyanitiwattana and Mattingly 2006)	Area type, time of day, crash severity, collision type and factor, road classification	- Speeding is the major collision factor of SC than PC

		- Property damage only (PDO) crashes are more frequent in both PC and SC
PCT (Zheng et al., 2015)	Day of the week, month of the year, and hour of the day	- SC were significantly different from PC with respect to month of the year and hour of the day
t-test (Tian et al., 2015, 2016)	Crash severity, crash type, no improper action, careless driving	- Careless driving is the leading factor which accounts for more than 50% of the total primary incidents, followed by exceeding safety velocity limit (8.13%), and no improper driving/action (4.07%) in average - Rear-end collision type is predominant in SC
ZOPM (Xu et al., 2019)	Traffic volume, speed, weather condition, crash severity, crash type, road geometry	- Rainy weather and hit-run primary crash type are the significant factors to SC

Note: LR: Logistic Regression Model, NNs: Neural Networks, BNN: Bayesian Neural Network, GM: Gompit Model, PM: Probit Model, MLM: Multinomial Logit Model, BLM: Bayesian Random Effect Logit Model, GORP: Generalized Ordered Response Probit, SEM: Structural Equation Models, TFP: Test for Proportions, PCT: Pearson's Chi-square Test, ZOPM: Zero-inflated Ordered Probit Model

Table A 2.3: Summary of the Test for Proportions Study Results (Hirunyanitiwattana et al., 2006)

Variable	Null Hypothesis	Findings
Area type	The proportion of SCs in urban districts is not significantly different from the proportion of SCs in rural districts.	Urban districts have higher proportion of SCs than rural districts
Time of day	The proportion of PCs by time of day is not significantly different from the proportion of SCs of the same classification.	During the peak hour, the proportion of SCs is higher than that of PCs
Crash severity	The proportion of PCs by crash severity is not significantly different from the proportion of SCs of the same classification.	Probability of crash severities is greater for PCs than SCs
Collision type	The proportion of PCs by collision type is not significantly different from the proportion of SCs of the same classification.	Rear end proportion of SCs is higher than that of PCs
Primary collision factor	The proportion of PCs by primary collision factor is not significantly different from the proportion of SCs of the same classification.	Speeding is the major collision factor of SCs than PCs
Road classification	The proportion of PCs by road classification is not significantly different from the proportion of SCs of the same classification.	Proportion of SCs on urban freeway with larger than 4 lanes is higher than PCs.

Table A 2.4: Computation Equations for SCs Analysis

	Influential Risk Factor Models
LR (Zhan et al., 2008, Junhua et al., 2016, Khattak et al., 2010, 2012)	$P\left(y_i = \frac{1}{x_i}\right) = p_i = \frac{e^{(\beta x_i + \beta_0)}}{1 + e^{(\beta x_i + \beta_0)}}$; p_i = probability of i incident, β, β_0 = coefficients of variables x_i.
NNs (Vlahogianni et al., 2013)	$y_p = \frac{1}{1 + e^{-net_j}}$; $net_j = \sum_k \omega_{kj} h_k + \theta_j$, y_p = output of pth incident, ω_{kj} = connection weight between kth neuron in the hidden layer and j^{th} neuron in the output layer with the bias term θ_j, $h_k = \frac{1}{1 + e^{-net_k}}$ = output of the hidden neuron, $net_k = \sum_i \omega_{ki} x_i + \theta_i$, ω_{ki} = connection weight between k^{th} neuron in the hidden layer and i^{th} input variable with the bias term, θ_i
GM (Vlahogianni et al., 2013)	$p_i = 1 - \exp\left(-\exp\left(x_i\beta + \beta_0\right)\right)$; p_i = probability of incident i; β, β_0 = coefficients of variables x_i
PM (Vlahogianni et al., 2013)	$p_i = \varphi\left(x_i\beta + \beta_0\right) = \int_{-\infty}^{x_i\beta + \beta_0} \emptyset(t)dt$; p_i = probability of incident i, β, β_0 = coefficients of independent variables x_i; φ = cumulative standard distribution; and $\emptyset$ = density function.
MLM (Mishra et al., 2016)	$P(m) = \frac{e^{V_m}}{\Sigma_{k=1}^{M} e^{V^k}}$; $P(m)$ = probability of event m, V_j = utility of m, j = 1, 2,……., M.
BLM (Yang et al., 2014)	y_n = Bernoulli (p_n); logit $\left(p_n\right) = \beta_0 + \theta_r + \beta_1 x_{1n} + \beta_2 x_{2n} + .. + \beta_i x_{in}$; $\theta_r \sim (0, \Sigma_\theta)$; y_n = SCs indicator for nth observation; p_n = probability of SCs, β_i = coefficient of independent variable xi; x_{in} = the value of variable i for sample n; θ_r = random effect with the heterogeneity effects for freeway segment r.

Note: LR: Logistic Regression Model, NNs: Neural Networks, GM: Gompit Model, PM: Probit Model, MLM: Multinomial Logit Model, BLM: Bayesian Random Effect Logit Model

REFERENCES

American Association of State Highway and Transportation Officials (AASHTO). Highway Safety Manual. AASHTO, 2010.

Al-Deek, H., Garib, A., & Radwan, A. E. (1995). New Method for Estimating Freeway Incident Congestion. Transportation Research *Record*, 30-39.

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand (Vol. 9). MIT press.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Bryden, J. E., & Fortuniewicz, J. S. (1986). Traffic Barrier Performance Related to Vehicle Size and Type (No. HS-039 930).

Carrick, G., Srinivasan, S., & Bejleri, I. (2015). Descriptive Analysis and Characterization of Secondary Traffic Crashes in Florida. In Transportation Research Board 94th Annual Meeting (No. 15-0096).

Carson, J., and F. Mannering. (2001). The effect of ice warning signs on ice-accident frequencies and severities. Accident Analysis & Prevention, 33(1), 99-109.

Celik, A. K., and E. Oktay. A Multinomial Logit Analysis of Risk Factors Influencing Road Traffic Injury Severities in the Erzurum and Kars Provinces of Turkey. Accident Analysis and Prevention, Vol. 72, 2014, pp. 66-77.

Chang, G. L., & Rochon, S. (2009). Performance Evaluation and Benefit Analysis for CHART. Technical Rep., Maryland State Highway Administration, Hanover, MD.

Chou, C. S., & Miller-Hooks, E. (2009). Simulation-based Secondary Incident Filtering Method. Journal of Transportation Engineering, 136(8), 746-754.

Chimba, D., Kutela, B., Ogletree, G., Horne, F., & Tugwell, M. (2014). Impact of Abandoned and Disabled Vehicles on Freeway Incident Duration. Journal of Transportation Engineering, 140(3), 04013013.

Chung, Y., & Recker, W. W. (2012). A Methodological Approach for Estimating Temporal and Spatial Extent of Delays Caused by Freeway Accidents. IEEE Transactions on Intelligent Transportation Systems, 13(3), 1454-1461.

Chung, Y. (2013). Identifying Primary and Secondary Crashes from Spatiotemporal Crash Impact Analysis. Transportation Research Record: Journal of the Transportation Research Board, 2386, 62-71.

Ding, Z., Gong, W., Li, S., & Wu, Z. (2018). System dynamics versus agent-based modeling: A review of complexity simulation in construction waste management. *Sustainability*, *10*(7), 2484.

Goodall, N. J. (2017). Probability of Secondary Crash Occurrence on Freeways with the Use of Private-Sector Speed Data. Transportation Research Record: Journal of the Transportation Research Board, 2635, 11-18.

Haghani, A., Iliescu, D., Hamedi, M., & Yang, S. (2006). Methodology for Quantifying the Cost Effectiveness of Freeway Service Patrol Programs. University of Maryland.

Hirunyanitiwattana, W., & Mattingly, S. P. (2006). Identifying Secondary Crash Characteristics for California Highway System. In Transportation Research Board 85th Annual Meeting (No. 06-2900).

Hu, L. T., & Bentler, P. M. (1995). Evaluating Model Fit.

Huq, A. S. (2011). Identification of black spots on a selected segment of Dhaka-Chittagong highway: application of interactive highway safety design module (IHDSM). Australian Road Research Board. ISBN 187659263X

Imprialou, M. I. M., Orfanou, F. P., Vlahogianni, E. I., & Karlaftis, M. G. (2013). Methods for Defining Spatiotemporal Influence Areas and Secondary Incident Detection in Freeways. Journal of transportation engineering, 140(1), 70-80.

Junhua, W., Boya, L., Lanfang, Z., & Ragland, D. R. (2016). Modeling Secondary Accidents Identified by Traffic Shock Waves. Accident Analysis & Prevention, 87, 141-147.

Jalayer, M., Baratian-Ghorghi, F., & Zhou, H. (2015). Identifying and Characterizing Secondary Crashes on the Alabama State Highway Systems. Advances in Transportation Studies, 37, 129-140.

Karlaftis, M.G., Latoski, S.P., Richards N.J., & Sinha, K. C., (1998). An Empirical Analysis of Secondary Crash Causes. 77th Annual Meeting of the Transportation Research Board, CD-ROM, Washington, D.C. Paper 980017.

Karlaftis, M. G., Latoski, S. P., Richards, N. J., & Sinha, K. C. (1999). ITS Impacts on Safety and Traffic Management: An Investigation of Secondary Crash Causes. Journal of Intelligent Transportation Systems, 5(1), 39-52.

Khattak, A. J., Wang, X., Zhang, H., & Cetin, M. (2007). Primary and Secondary Incident Management: Predicting Durations in Real Time. Contract, 2009.

Khattak, A., Wang, X., & Zhang, H. (2009). Are Incident Durations and Secondary Incidents Interdependent? Transportation Research Record: Journal of the Transportation Research Board, 2099, 39-49.

Khattak, A. J., Wang, X., & Zhang, H., (2010a). A Tool for Dynamically Predicting Incident Durations, Secondary Incident Occurrence, and Incident Delays, In 12th World Conference on Transport Research Lisbon, Portugal.

Khattak, A., Wang, X., & Zhang, H. (2010b). Spatial Analysis and Modeling of Traffic Incidents for Proactive Incident Management and Strategic Planning. Transportation Research Record: Journal of the Transportation Research Board, 2178, 128-137.

Khattak, A., Wang, X., & Zhang, H. (2012). Incident Management Integration Tool: Dynamically Predicting Incident Durations, Secondary Incident Occurrence and Incident Delays. IET Intelligent Transport Systems, 6(2), 204-214.

Kim, J. K., Ulfarsson, G. F., Shankar, V. N., and Kim, S. (2008). Age and Pedestrian Injury Severity in Motor-vehicle Crashes: A Heteroskedastic Logit Analysis. Accident Analysis & Prevention, 40(5), 1695-1702.

Kopitch, L., & Saphores, J. D. M. (2011). Assessing Effectiveness of Changeable Message Signs on Secondary Crashes. In Transportation Research Board 90th Annual Meeting (No. 11-4270).

Kwon, J., Mauch, M., & Varaiya, P. (2006). Components of congestion: Delay from incidents, special events, lane closures, weather, potential ramp metering gain, and excess demand. Transportation Research Record, 1959(1), 84-91.

Latoski, S. P., Pal, R., & Sinha, K. C. (1999). Cost-effectiveness Evaluation of Hoosier Helper Freeway Service Patrol. Journal of Transportation Engineering, 125(5), 429-438.

Li, H., & Bertini, R. (2010). Comparison of Algorithms for Systematic Tracking of Patterns of Traffic Congestion on Freeways in Portland, Oregon. Transportation Research Record: Journal of the Transportation Research Board, 2178, 101-110.

Li, Z., Li, Y., Liu, P., Wang, W., & Xu, C. (2014). Development of a variable speed limit strategy to reduce secondary collision risks during inclement weathers. Accident Analysis & Prevention, 72, 134-145.

Markus, K. A. (2012). Principles and practice of structural equation modeling by Rex B. Kline. Structural Equation Modeling: A Multidisciplinary Journal 19(3), 509-512.

Miaou, S. P., Lu, A., & Lum, H. S. (1996). Pitfalls of using R^2 to evaluate goodness of fit of accident prediction models. Transportation Research Record, 1542(1), 6-13.

Mishra, S., Golias, M., Sarker, A., & Naimi, A. (2016). Effect of Primary and Secondary Crashes: Identification, Visualization, and Prediction (No. CFIRE 09-05).

Moore, J. E., Giuliano, G., & Cho, S. (2004). Secondary Accident Rates on Los Angeles Freeways. Journal of Transportation Engineering, 130(3), 280-285.

Moore, D. N., Schneider, W. H., Savolainen, P. T., and Farzaneh, M. (2011). Mixed Logit Analysis of Bicyclist Injury Severity Resulting from Motor Vehicle Crashes at Intersection and Non-intersection Locations. Accident Analysis & Prevention, 43(3), 621-630.

NHTSA, 2016. Quick Facts 2015, Accessed 18 August 2017, https://crashstats.nhtsa.dot.gov/.

Owens, N. D., Armstrong, A. H., Mitchell, C., & Brewster, R. (2009). Federal Highway Administration Focus States Initiative: Traffic Incident Management Performance Measures Final Report (No. FHWA-HOP-10-010).

Owens, N., Armstrong, A., Sullivan, P., Mitchell, C., Newton, D., Brewster, R., Trego, T. (2010). Traffic Incident Management Handbook. (Report No. FHWA-HOP-10-013). Washington, DC: Federal Highway Administration, Office of Transportation Operations.

Ozbay, K., & Kachroo, P. (1999). Incident management in intelligent transportation systems.

Park, H., & Haghani, A. (2016). Real-Time Prediction of Secondary Incident Occurrences Using Vehicle Probe Data. Transportation Research Part C: Emerging Technologies, 70, 69-85.

Park, H., Shafahi, A., & Haghani, A. (2016a). A stochastic emergency response location model considering secondary incidents on freeways. IEEE Transactions on Intelligent Transportation Systems, 17(9), 2528-2540.

Park, J., Abdel-Aty, M., Lee, J., & Lee, C. (2015). Developing crash modification functions to assess safety effects of adding bike lanes for urban arterials with different roadway and socio-economic characteristics. Accident Analysis & Prevention, 74, 179-191.

Raub, R. A. (1997a). Occurrence of Secondary Crashes on Urban Arterial Roadways. Transportation Research Record: Journal of the Transportation Research Board, 1581, 53-58.

Raub, R. A. (1997b). Secondary crashes: An Important Component of Roadway Incident Management. Transportation Quarterly, 51(3).

Sando, T., Alluri, P., Chuan, C., Haule, H., Kitali, A., Lentz, R., & Huq, A. (2018). Evaluation of Incident Response Improvements for Statewide Application: Learning from the New Regional Traffic Management Center in Jacksonville, Florida.

Sarker, A. A., Naimi, A., Mishra, S., Golias, M. M., & Freeze, P. B. (2015). Development of a Secondary Crash Identification Algorithm and Occurrence Pattern Determination in Large Scale Multi-Facility Transportation Network. Transportation Research Part C: Emerging Technologies, 60, 142-160.

Sarker, A. A., Paleti, R., Mishra, S., Golias, M. M., & Freeze, P. B. (2017). Prediction of Secondary Crash Frequency on Highway Networks. Accident Analysis & Prevention, 98, 108-117.

Savolainen, P. T., Mannering, F. L., Lord, D., and Quddus, M. A. (2011). The Statistical Analysis of Highway Crash-injury Severities: A Review and Assessment of Methodological Alternatives. Accident Analysis & Prevention, 43(5), 1666-1676.

Shlayan, N., Saddi, R. R., Kachroo, P., Nevada, D. O. T., & Ohene, F. (2009). The Moving Dynamic Nature of Progression Curves for Freeway Incident Related Congestion. IMWeb, University of Nevada, LA.

Skabardonis, A., Noeimi, H., Petty, K., Rydzewski, D., Varaiya, P., & Al-Deek, H. (1995). Freeway service patrol evaluation.

Sun, C., & Chilukuri, V. (2007). Secondary Accident Data Fusion for Assessing Long-Term Performance of Transportation Systems. Report No. MTC Project 2005-04, 2007.

Sun, C., & Chilukuri, V. (2010). Dynamic Incident Progression Curve for Classifying Secondary Traffic Crashes. Journal of Transportation Engineering, 136(12), 1153-1158.

Sun, C., & Chilukuri, V. (2006). Use of Dynamic Incident Progression Curve for Classifying Secondary Accidents. In Transportation Research Board 85th Annual Meeting (No. 06-1804).

Train, K. E. (2009). Discrete choice methods with simulation. Cambridge university press.

Tian, Y. (2015). Identifying Secondary Crashes by Using Geographic Information System (GIS) and Determining the Secondary Crashes Characteristics. Dissertations and Theses. Paper 253.

Tian, Y., Chen, H., & Truong, D. (2016). A Case Study to Identify Secondary Crashes on Interstate Highways in Florida by Using Geographic Information Systems (GIS). Advances in Transportation Studies, 2, 103-112.

Ullman, J. B., & Bentler, P. M. (2003). Structural Equation Modeling. John Wiley & Sons, Inc.

Vlahogianni, E., Karlaftis, M., Golias, J., & Halkias, B. (2010). Freeway Operations, Spatiotemporal-Incident Characteristics, and Secondary-Crash Occurrence. Transportation Research Record: Journal of the Transportation Research Board, 2178, 1-9.

Vlahogianni, E., Karlaftis, M., & Orfanou, F. P. (2012). Modeling the Effects of Weather and Traffic on the Risk of Secondary Incidents. Journal of Intelligent Transportation Systems, 16(3), 109-117.

Vlahogianni, E., & Karlaftis, M. (2013). Fuzzy-Entropy Neural Network Freeway Incident Duration Modeling with Single and Competing Uncertainties. Computer-Aided Civil and Infrastructure Engineering, 28(6), 420-433.

Wang, J., Xie, W., Liu, B., & Ragland, D. R. (2016). Identification of Freeway Secondary Accidents with Traffic Shock Wave Detected By Loop Detectors. Safety Science, 87, 195-201.

Washington, S. P., Karlaftis, M. G., & Mannering, F. (2010). Statistical and Econometric Methods for Transportation Data Analysis. CRC press.

Xu, C., Liu, P., Yang, B., & Wang, W. (2016). Real-Time Estimation of Secondary Crash Likelihood on Freeways Using High-resolution Loop Detector Data. Transportation Research Part C: Emerging Technologies, 71, 406-418.

Xie, K., Ozbay, K., & Yang, H. (2016). A Joint Analysis of Secondary Collisions and Injury Severity Levels Using Structural Equation Models. In Transportation Research Board 95th Annual Meeting (No. 16-0206).

Xie, K., Ozbay, K., & Yang, H. (2018). Secondary collisions and injury severity: A joint analysis using structural equation models. Traffic injury prevention, 19(2), 189-194.

Xu, C., S., Xu, C. Wang, J. Li. (2019). Investigating the factors affecting secondary crash frequency caused by one primary crash using zero-inflated ordered probit regression. Physica A 524, 121-129.

Yang, H., Bartin, B., & Ozbay, K. (2013). Investigating the Characteristics of Secondary Crashes on Freeways. In 92nd Annual Meeting of the Transportation Research Board, Washington, DC (Vol. 2).

Yang, H., Bartin, B., & Ozbay, K. (2014a). Mining the Characteristics of Secondary Crashes on Highways. Journal of Transportation Engineering, 140(4), 04013024.

Yang, H., Bartin, B., & Ozbay, K. (2014b). Use of Sensor Data to Identify Secondary Crashes on Freeways. Transportation Research Record: Journal of the Transportation Research Board, 2396, 82-92.

Yang, H., Ozbay, K., Morgul, E., Bartin, B., & Xie, K. (2014c). Development of Online Scalable Approach for Identifying Secondary Crashes. Transportation Research Record: Journal of the Transportation Research Board, 2470, 24-33.

Yang, H., Ozbay, K., & Xie, K. (2014d). Assessing the Risk of Secondary Crashes on Highways. Journal of Safety Research, 49, 143-149.

Yang, H., Wang, Z., & Xie, K. (2017). Impact of connected vehicles on mitigating secondary crash risk. International journal of transportation science and technology, 6(3), 196-207.

Yang, H., Wang, Z., Xie, K., Ozbay, K., & Imprialou, M. (2018). Methodological evolution and frontiers of identifying, modeling and preventing secondary crashes on highways. Accident Analysis & Prevention, 117, 40-54.

Ye, F., & Lord, D. (2014). Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models. Analytic methods in accident research, 1, 72-85.

Zhan, C., Gan, A., & Hadi, M. (2009). Identifying Secondary Crashes and Their Contributing Factors. Transportation Research Record: Journal of the Transportation Research Board, 2102, 68-75.

Zhan, C., Shen, L., Hadi, M. A., & Gan, A. (2008). Understanding the Characteristics of Secondary Crashes on Freeways. In Transportation Research Board 87th Annual Meeting (No. 08-1835).

Zhang, H., & Khattak, A. (2011). Spatiotemporal Patterns of Primary and Secondary Incidents on Urban Freeways. Transportation Research Record: Journal of the Transportation Research Board, 2229, 19-27.

Zhang, H., & Khattak, A. (2010a). Analysis of Cascading Incident Event Durations on Urban Freeways. Transportation Research Record: Journal of the Transportation Research Board, 2178, 30-39.

Zhang, H., & Khattak, A. (2010b). What is the Role of Multiple Secondary Incidents in Traffic Operations? Journal of Transportation Engineering, 136(11), 986-997.

Zheng, D., Chitturi, M. V., Bill, A. R., & Noyce, D. A. (2015). Analyses of Multiyear Statewide Secondary Crash Data and Automatic Crash Report Reviewing. Transportation Research Record: Journal of the Transportation Research Board, 2514, 117-128.

Zheng, D., Chitturi, M. V., Bill, A. R., & Noyce, D. A. (2014). Secondary Crash Identification on a Large-Scale Highway System. In Transportation Research Board 93rd Annual Meeting.

VITA

ARMANA SABIHA HUQ

2020	Ph.D., Civil Engineering, FIU, Miami, Florida; the Outstanding Student Life Awards, Scholar Category, FIU; the Third Place Winner of the Annual Book Scholarship, Gold Coast Chapter of ITE; Excellence in Civic Engagement and Leadership Medallion of Distinction, FIU
2019	Dissertation Year Fellowship, FIU, Miami, Florida
2019	Certified as a Road Safety Professional, Awarded the First Place in Florida Puerto Rico ITE Annual Meeting Research Competition in Puerto Rico; the Winner of the Women Engineers Local Tampa Collegiate Research Competition in Tampa; the National Lifesavers Conference Traffic Safety Scholar in Kentucky; the Most Valuable Academic of Leader in FIU; the Outstanding Student Life Awards, Scholar Category in FIU; Graduate and Professional Student Committee Conference Award at FIU
2018	Awarded the First Place in the Annual University Transportation Center Conference for the Southeastern Region Research Competition, SC; the Third Place Winner of Graduate Student Appreciation Week, Scholarly Forum at FIU
2017	Awarded the First Place Winner of Smart City Campaign in the United Nations Development Program, Bangladesh; Best President Award, Bangladeshi Student Organization at FIU; Council for Student Organizations Awards as the President of ITE Student Chapter at FIU; Graduate and Professional Student Committee Conference Award at FIU
2015-2019	Graduate Research and Teaching Assistant Department of Civil Engineering, FIU, Miami, Florida
2007-Present	Faculty, Accident Research Institute, BUET, Dhaka, Bangladesh
2009	M.Sc., Civil Engineering, BUET, Dhaka, Bangladesh
2006	B.Sc., Civil Engineering, BUET, Dhaka, Bangladesh

PUBLICATIONS AND PRESENTATIONS

Huq, A., Lee, M. and Jin, X. (2020), A Comprehensive Examination of Highway Secondary Crash Risk Factors, Transportation Research Board (TRB) 99th Annual Meeting, January 12-16, Washington, DC

Rabbi, S., Newaz, K., Huq, A. and Hoque, M. (2019), Development of Low-Volume Rural Road Safety Manual for Bangladesh, the 15th World Conference on Transport Research, May 26-31, Mumbai, India

Huq, A. (2019), Intelligent Transportation System Approach on Freeway Secondary Crash Identification and Mitigation, ITE Annual Meeting, Austin, TX, July 21-24

Huq, A. and Alluri, P. (2019), A Comprehensive Review of Secondary Crash Studies, TRB 98th Annual Meeting, January 13-17, Washington, DC

Huq, A. (2019), Analysis of Rear-end Secondary Crashes on Rural Two-lane Undivided Roadways, Florida Puerto Rico ITE Annual Meeting, Nov 6-8

Huq, A. (2019), Application of Connected Vehicle Technology to Mitigate Secondary Crashes, ASCE Annual Meeting, Orlando, FL, July 18-19

Huq, A. (2019), Modeling the Impact of Freeway Incident Characteristics on Secondary Crashes, Safe Systems Summit, Durham, NC, Apr 22-25

Huq, A. (2019), Application of Connected Vehicle Technology to Mitigate Secondary Crashes, Society of Women Engineers' WE Local Program Conference, Tampa, FL, Feb 15-16

Sando, T., Alluri, P., Chuan, C., Haule, H., Kitali, A., Lentz, R., and Huq, A. (2018), Evaluation of Incident Response Improvements for Statewide Application: Learning from the New Regional Traffic Management Center in Jacksonville, Funded Research Project Report, Florida Department of Transportation Research Center, Tallahassee, FL

Huq, A. (2018), Freeway Secondary Crash Identification and Mitigation, the 6th Annual UTC Conference for the Southeastern Region, Oct 24-25, Clemson University, SC

Alluri, P., Raihan, M. A., Saha, D., Wu, W., Huq, A., Nafis, S., and Gan, A. (2017). Statewide Analysis of Bicycle Crashes, Funded Research Project Report, Florida Department of Transportation Research Center, Tallahassee, FL

Huq, A. and Alluri, P. (2016), Comprehensive Study to Improve Bicycle Safety, the 10th University Transportation Center (UTC) Spotlight Conference: Bicycles and Pedestrians, Dec 1-2, Washington, DC