EXAMINING METHODS FOR IDENTIFYING THE OCCURRENCE OF SECONDARY CRASHES

By

Hadis Nouri

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of requirements
for the degree of

Civil Engineering—Doctor of Philosophy

2022

ABSTRACT

EXAMINING METHODS FOR IDENTIFYING THE OCCURRENCE OF SECONDARY CRASHES

By

Hadis Nouri

Traffic crashes are a particular concern in urban areas, where the occurrence of a collision heightens the risk of subsequent secondary crashes upstream, particularly under high levels of traffic congestion. There is considerable difficulty in estimating the number of such crashes, and in identifying roadway locations and circumstances where the risks of such crashes are most pronounced. In light of these concerns, there is significant value in advancing our understanding of these issues, including our ability to predict and mitigate the potential for secondary crashes on freeways. A significant challenge in this regard is the ability to effectively identify a secondary crash with respect to the both the spatial temporal thresholds within which secondary crashes occur. Contemporary approaches are often based on static spatiotemporal impact windows, or on dynamic approaches that consider traffic flow conditions. Both methods are subject to important limitations that are investigated as a part of this research. As a part of this study, crash data from the Michigan interstate system was used to identify secondary crashes. A detailed review of police crash reports is conducted to verify which crashes are secondary in nature by examining standard fields on the report form, as well as information from the narrative section completed by the investigating officer. The influence of spatiotemporal window sizing (relative to the time and location of the primary crash) is explored with respect to the sensitivity and specificity of secondary crash detection in order to determine thresholds that yield minimal error. A static approach based on a large number of predefined window sizes was used to compare the rate of secondary crash identification.

The static method was shown to consistently overestimate secondary crash occurrence and these results varied across thresholds sizes. Subsequent efforts used a dynamic approach, where the window size was varied based upon changes in speed profiles on the associated road segments. Real-time traffic and speed data were used to identify secondary crashes and the results vary considerably based upon the method employed. The research also identified contextual environments where the risks of secondary crashes are most pronounced through the estimation of a series of regression models, culminating in guidance to assist road agencies in effectively monitoring and clearing crashes and other incidents to minimize the potential for secondary crashes.

This thesis is dedicated to Mom and Dad. Thank you for always believing in me. This thesis work is dedicated to my husband, Roozbeh, who has been a constant source of support and encouragement during my journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW

Traffic incidents, such as crashes and vehicle breakdowns, cause significant congestion in urban areas and cause 30 to 40 percent of all congestion (Skabardonis *et al.*, 1995; Ozbay and Kachroo, 1999). The congestion caused by an incident can also increase the potential for upstream traffic crashes. Such events, generally referred to as secondary crashes, usually increase the time needed for traffic flow to return to normal (i.e., pre-incident) levels. Between 2 and 15 percent of the initial incidents can cause secondary crashes, leading to traffic operations complications (Moore, Giuliano and Cho, 2004; Hirunyanitiwattana, 2006).

Secondary crashes are one of the many undesirable consequences of crashes and other types of incidents. Such crashes are typically defined based upon the congested spatiotemporal boundaries impacted by primary crashes (Yang et al. 2018). Secondary crashes have increasingly been recognized as a significant problem in freeways that frequently affect both traffic operations and safety (Imprialou *et al.*, 2014). As reported by Owens et al. (2010), as many as 20 percent of all crashes and 18 percent of all fatalities on freeways result from secondary crashes. It has also been shown that the occurrence of an earlier crash could increase the risk of secondary crashes by more than six times (Tedesco *et al.*, 1994; Owens *et al.*, 2010). Karlaftis et al. (1999) found that if the clearance time of an initial incident increases by an additional minute, the likelihood of secondary crashes may rise by about 2.8 percent (Karlaftis *et al.*, 1999).

There has been considerable variability in estimates as to the proportion of all crashes that are secondary in nature. This is due to several factors, including differences in the contextual environment of these studies and challenges that are inherent in determining those crashes that are directly due to the occurrence of a prior crash (Sarker *et al.*, 2015). For example, Raub (1997) found that more than 15 percent of the crashes reported by police may be secondary in nature (Raub, 1997). The study by Karlaftis et al. (1999) examined primary crash characteristics and showed that more than 15 percent of all crashes might have resulted from an earlier incident (Karlaftis *et al.*, 1999). Moore et al. (2004) estimated secondary crash rates between 1.5 and 3.0 percent, significantly lower than previous studies suggested (Moore, Giuliano and Cho, 2004). Zhan et al. (2008) investigated incidents that resulted in lane blockages as potential causes of secondary crashes on Los Angeles freeways using crash records and traffic data from inductive loop detectors. The result showed that only 7.9 percent of all lane blockage incidents resulted in secondary crashes (Zhan et al. 2008).

Due to substantial economic and safety risks associated with secondary crashes, transportation agencies have taken various measures to minimize and mitigate the potential for and impacts of such crashes (Yang et al. 2018). One main challenge in investigating this issue is the inherent difficulty in effectively identifying which crashes are actually due to a prior crash or other incidents (Sarker *et al.*, 2015). Existing studies have made great efforts to explore the underlying mechanisms of secondary crashes, and relevant methodologies evolved regarding the identification, modeling, and prevention of these crashes. To date, there is significant variability in both the results and underlying methods used to identify secondary crashes (Yang et al. 2018).

1.1 Existing Methods for Identification of Secondary Crashes

Research has generally defined secondary crashes based on congested spatiotemporal boundaries impacted by primary crashes (Yang et al. 2018). The reliability of the spatial and temporal information of the prior incident is critical to the accuracy of secondary crash detection. Defining the impact area of an initial incident or crash is generally the first step in identifying these spatiotemporal boundaries. Various research studies have investigated different approaches to identify and analyze secondary crashes. These studies can be mainly classified into three types, including manual identification of crashes using real-time data (e.g., cameras from traffic management centers) or historical records (i.e., police crash reports), automatic identification using static spatiotemporal windows, and automatic identification using dynamic windows. In the latter two approaches, after identifying the impact area of the primary crash, the second step is to identify the secondary crashes that occur within the resultant spatiotemporal boundaries (Kitali, Alluri, Sando and Lentz, 2019; Kitali, Alluri, Sando and Wu, 2019). The following sections provide further descriptions of these three approaches to secondary crash identification.

1.1.1 Manual Method

Manual identification of secondary crashes can be done in either real-time or using historical data from police crash reports. Real-time identification requires visual verification of crashes through active monitoring. This is typically done by transportation agency personnel, such as staff from transportation management canters, incident responders, or law enforcement. Agencies have traditionally used this approach to identify and respond to events in near real-time. The process is simple and straightforward; however, manual identification is inefficient and can be unreliable and inconsistent for the purposes of large-scale identification (Kitali, Alluri, Sando and Lentz, 2019). This approach is also only viable in areas where there is continuous coverage of the roadway network through either closed-circuit cameras, courtesy patrol vehicles, or other resource-intensive approaches.

Large-scale manual identification of secondary crashes has been done in a limited number of studies using information from police crash reports, which are a very useful source of information for such purposes (Zhang et al. 2020). In a study by Zheng (2015), five years of crash data from Wisconsin were analyzed. A procedure was developed to automatically evaluate the narrative sections of police crash reports and detect potential secondary crashes if the narrative explicitly mentioned the crash was secondary in nature. Results found that the average distances from the primary crash to the upstream secondary crash were 0.29 miles. In addition, the observed average time-lapse was found to be 17 minutes between the primary and secondary crashes (Zheng *et al.*, 2015).

1.1.2 Static Method

The second approach, referred to as the static method, was first proposed in a study by Raub (1997). A fixed spatiotemporal threshold is used to identify potential secondary crashes in the static approach. Raub (1997) considered a spatial threshold of 1600 meters upstream of the primary crash and 15 minutes after the clearance of the crashes as a temporal threshold for identification purposes (Raub, 1997a). Several studies have investigated the spatiotemporal distribution of secondary crashes using various thresholds (Tedesco *et al.*, 1994; Raub, 1997a; Karlaftis *et al.*, 1999a; Chang and Steven, 2002; Moore, Giuliano and Cho, 2004; Kopitch and Saphores, 2011; Jalayer, Baratian-Ghorghi and Zhou, 2015; Tian, Chen and Truong, 2016). Table 1-1 summarizes the spatiotemporal windows that have been used in prior research that utilized a static approach.

Chung (2013) found an average time gap of 65.81 minutes and an average distance of 1.34 miles between primary and secondary crashes (Chung, 2013). Junhua et al. (2016) investigated the spatiotemporal gaps between crashes and found an average gap time of 74 minutes and a mean distance threshold of 4.52 miles. In addition, in 19.4 and 26.5 percent of the cases, gaps of less than one mile and 10 minutes were observed, respectively (Wang, Liu, *et al.*, 2016). Kitali et al. (2019) concluded that 90 percent of secondary crashes were detected within the spatial threshold of 5 miles and temporal threshold of 150 minutes. Based on this study, the distance gap was shown to vary greatly under different traffic conditions (Kitali, Alluri, Sando and Lentz, 2019).

Defining representative spatial and temporal thresholds play a critical role in the success of this method. There are also inherent trade-offs involved as considering large spatiotemporal windows leads to better sensitivity (i.e., identification of crashes that are actually secondary in nature), but at the expense of worse specificity (i.e., false identification of crashes that are not actually secondary) (Zheng *et al.*, 2015).

Moreover, considering a fixed spatiotemporal threshold may result in under or overestimating secondary crash frequencies for smaller or larger spatiotemporal thresholds. The static method is somewhat subjective and arbitrary and does not allow for consideration of the dynamic nature of traffic as the spatiotemporal thresholds vary based upon the level of traffic congestion and various other factors (Zhang, Green, and Chen 2019).

Table 1- 1: Summary of a spatiotemporal window in the static method

Author	**Spatial Boundaries**	**Temporal Boundaries**
Raub (1997)	1 mile	15 minutes
Karlaftis et al. (1999)	1 mile	15 minutes
Moore et al. (2004)	2 miles	120 minutes
Hirunyanitiwattana and Mattingly (2006)	2 miles	60 minutes
Pigman et al. (2011)	3.62 miles	42 minutes
Chung (2013)	1.34 miles	65.81 minutes
Wang et al. (2016)	4.518 miles	74 minutes
Kitali (2019)	5 miles	150 minutes
Chang et al. (2003)	2 miles	120 minutes
Zhan et al. (2008)	2 miles	15 minutes

1.1.3 Dynamic Method

Finally, the third approach is a dynamic method that establishes the spatiotemporal thresholds based on the primary incident's characteristics and concurrent traffic flow conditions. In order to overcome the static approach’s limitations, recent studies have investigated various dynamic approaches, such as queuing models, speed contours, shockwave theory, and vehicle probe data to identify secondary crashes (Junhua et al. 2016; Park and Haghani 2016b; Xu et al. 2016; Zhang, Cetin, and Khattak 2015)

1.1.3.1 Queuing Model

Dynamic approaches mainly use prevailing traffic flow conditions in order to identify secondary crashes and may facilitate better capture flow of the traffic and the queue formation process (Yang, Guo, and Xu 2019). Several studies developed queuing models to capture the progression of the region in which secondary crash occurs (Sun and Chilukuri 2010; Sun and Chilukuri 2007; Vlahogianni, Karlaftis, and Orfanou 2012; Chengjun Zhan, Gan, and Hadi 2009). Traffic arrival rate, departure rate, crash duration, lane capacity, and travel speed are some of the contributing factors that are used to capture the vehicle queue length (Yang et al. 2017).

1.1.3.2 Shockwave Theory

Shockwave theory is used to evaluate the dynamic traffic impact of a primary crash. In a study by Zheng et al. (2014), shockwave theory is used to model the dynamic impact area of primary crashes and identify secondary crashes occurring within these areas of large-scale transportation systems. The study utilized 2010 data from nearly 1,500 miles of freeways in Wisconsin. The result showed over 85 percent of secondary crashes were of three major crash types, including two-vehicle rear-end collisions, multiple-vehicle rear-end collisions, and sideswipes (Zheng *et al.*, 2014).

A total of 49,753 crashes from 2010 to 2012 on California interstate freeways, along with their corresponding upstream loop data, were analyzed by the shockwave boundary filtering method to identify secondary crashes. Based on the result, secondary accidents accounted for 1.08 percent, much lower than previous research estimates (Wang et al. 2016).

In another study, traffic shockwave speed and volume at the occurrence of a primary accident were considered in order to identify secondary crashes. In order to investigate contributing factors to secondary crash occurrence logistic regression model was developed. The study analyzed accident records from three years on California interstate freeways. Results show that primary crashes with long durations may expressively raise the possibility of secondary crashes. In addition, unsafe speed and weather are found to be factors contributing to the secondary crash occurrence (Wang, Xie, et al. 2016).

1.1.3.3 Vehicle Probe Data

Vehicle probe technology is used for real-time traffic estimation, and it is a common practice for data providers to report data on real-time traffic message signs. Studies attempted to explore the dynamics of traffic evolution during the primary crash using vehicle probe technology. This method proved to have a better result in identifying secondary crashes in comparison to the static method (Park and Haghani 2016a; Park, Haghani, and Hamedi 2013; Yang et al. 2017). In another study, using vehicle probe technology, a new data-driven analysis framework was developed to support the identification of secondary crashes that consists of three major components. At first, the impact area of a primary crash was detected. Then, the boundary of the impact area was estimated, and secondary crashes within the boundary were identified. The test results show that the proposed approach can best describe the impact area and identify up to 95 percent of the simulated crashes (Yang et al. 2017). However, this approach is limited to freeway segments which probe vehicle data is available.

1.1.3.4 Speed Contour

Wang and Jiang (2020) proposed an approach of influencing/leveraging the spatiotemporal evolution of shockwaves in speed contour plots in order to identify secondary crashes on freeways. It has been demonstrated that the defined region corresponding to a single primary crash is generally consistent with the spatiotemporal evolution of shockwaves (Wang and Jiang 2020). Speed contour plots were used in a study by Yang et al. (2014) to identify secondary crashes. Based on the results, 75 and 50 percent of all secondary crashes occur within two hours and two miles upstream of the primary crash, respectively. In addition, rear-end crashes were found to be the dominant secondary crash and improper lane changing, distracted driving as well as unsafe speed is considered to be significant contributing factors (Yang et al. 2014). Kitali et al. (2019) tried to identify the impact area of primary crashes using speed data. Based on the study, depending on the spatial and temporal influence area of the primary crash, the process of identifying secondary crashes varies. In this study, prevailing speed data in each section of the freeway was used to identify the impact range of the primary crash. Following all crashes within that impact area have been considered secondary crashes. The study's main objective was to determine the effect of traffic flow characteristics that change over space and time, such as speed, which has a significant impact on queue formation as a result of the primary crash. Results from the study showed that almost 8 percent of crashes are secondary crashes, and also more than 75 percent of secondary crashes were due to congested traffic conditions (Kitali et al. 2019).

Following the identification of secondary crashes, some previous studies have focused on investigating major factors contributing to the occurrence of secondary crashes. The study by Raub (1997) found that clearance time, peak hours, and weekdays are associated with more secondary crashes (Raub, 1997a). The study by Hirunyanitiwattana (2006) identifies secondary and primary crash characteristics in the California Highway System. The study revealed secondary crash rates increases in the region with high traffic volumes during morning and evening peak hours (Hirunyanitiwattana, 2006). Karlaftis et al. (1999) applied a logistic regression model to examine what primary crash characteristics are associated with the likelihood of a secondary crash. They suggested that the type of vehicle involved, the clearance time, season, and lateral location of the primary crash are significant factors (Karlaftis et al. 1999). More studies investigated contributing factors that affect the secondary crash occurrence, as shown in Table 1-2. The majority of studies used logistic regression models, and some used probit models to evaluate the existence of a significant difference between primary and secondary crashes (Khattak, Wang, and Zhang 2010; Khattak, Wang, and Zhang 2009; Vlahogianni et al. 2010; Vlahogianni, Karlaftis, and Orfanou 2012; Yang et al. 2014; Yang, Bartin, and Ozbay 2013; Zhan et al. 2008; Chengjun Zhan, Gan, and Hadi 2009).

Table 1- 2: Modeling approaches and contributing factors that affect secondary crashes

Author	**Method**	**Test variables**
Karlaftis et al. (1999)	Logistic regression	Clearance time, vehicle type, vehicle location, season, day of week
Hirunyanitiwattana and Mattingly (2006)	Proportional test	Time of day, roadway classification, primary crash, severity level, crash type
Zhan et al. (2008)	Logistic regression	Incident duration, time, environmental condition, incident type, location and traffic condition, lane closure, injuries, vehicle type

Table 1-2 (Cont'd)

Zhan et al. (2009)	Logistic regression	Incident duration, time, environmental condition, incident type, location, traffic condition, lane closure, injury condition, vehicle type
Khattak et al. (2009)	Binary probit regression models	Detection source, crash type, response vehicles, AADT, whether left shoulder affected, whether during peak hours, vehicle involved
Zhang and Khattak (2010)	Ordinal regression	Incident duration, whether truck involved, number of vehicles, lane blockage, segment length, number of lanes, curve, AADT
Vlahogianni et al. (2010)	Bayesian network	Time, number of vehicles, distance, duration, type of vehicle, location, maximum queue length, duration of queue observed upstream
Zhang and Khattak (2011)	Ordinary least squares (OLS) regression	The characteristics of primary crashes, road geometry, traffic
Vlahogianni et al. (2012)	Probit models	Duration, crash type, number of lanes, number of vehicles, heavy vehicle, travel speed, hourly volume, rainfall, downstream geometry, upstream geometry
Yang et al. (2013a)	Logistic regression	Time period, rear end, severity, duration, work zone, weekend, winter, lane closure, truck involved
Yang et al. (2013b, 2014a,b)	Probit model	The frequency of secondary crashes, spatiotemporal distributions, clearance time, crash type, severity

1.2 Summary and Research Objectives

Secondary crashes affect traffic operations and safety. These crashes are a performance measure in evaluating traffic incident management programs. Several approaches have been introduced to identify secondary crashes. Static and dynamic methods are mainly used in order to identify secondary crashes. Several thresholds have been suggested for defining the primary crash impact area and secondary crashes. However, there are some important limitations with these existing methods. For example, the static threshold method does not consider the dynamic nature of traffic conditions, introducing an implicit assumption that crashes occur at uniform rates irrespective of traffic flow conditions. Further, many studies focused on understanding the reliability of one window size have not included extensive validation with a detailed review of police-reported crash data. As such, the static approaches generally result in an overestimation of actual secondary crashes.

Dynamic approaches address this limitation by determining the spatiotemporal thresholds of primary crashes based on real-time traffic flow characteristics such as speed and density. However, dynamic models heavily rely on real-time traffic data, which are costly and only available in limited locations. For instance, approaches proposed based on queue length estimations require detailed queuing information, which may not be available at every location.

The goal of this research is to advance our understanding of the nature of secondary crashes, including the circumstances under which such crashes are most likely to occur. To address this goal, this study aims to:

1. Conduct a detailed investigation of police crash reports in order to identify the actual number and rate of secondary crashes on the Michigan interstate network;
2. Evaluate various spatial and temporal thresholds in terms of the precision and accuracy in identifying potential secondary crashes;
3. Compare scenarios under which various static and dynamic methods present advantages or disadvantages in identifying secondary crashes;
4. Assess the frequency of secondary crashes as a function of roadway characteristics.

As a part of these investigation, the research provides important insights into key areas, such as the trade-off between the sensitivity and specificity of static and dynamic models, particularly as it relates to the effect of window sizing or spatiotemporal thresholds on data reliability. This includes understanding the effect of the size of the static window in a large dataset and the correlation between static window predictions of secondary crash and actual number of secondary crashes.

This research also advances our understanding of dynamic secondary crash identification by estimating the impact range of primary crashes on upstream traffic using speed data and identifying secondary crashes that occur within this range. This method helps to better capture the effects of changes in traffic flow characteristics that occur over space and time and affect issues such as queue formation due to primary crashes. Compared to the previous spatiotemporal thresholds, the proposed approach provides an accurate, feasible impact area for secondary crash identification. The research also presents a sensitivity analysis of different spatial and temporal thresholds of primary crashes on the detection of secondary crashes.

Lastly, following the identification of secondary crashes through both the static and dynamic method, this research involves the development of a series of regression models in order to identify the interrelationships between secondary crash occurrence and various roadway and traffic characteristics of interest.

The remainder of this dissertation is organized as follows:

- Chapter 2 presents the results of the application of static methods for secondary crash identification. This includes the development of a crash-pairing algorithm developed to select spatially and temporally nearby crash pairs. Further, enhancements to the static methods are introduced by optimizing the trade-off between sensitivity and specificity to find the effect of window sizing or spatiotemporal thresholds on the reliability of data. In addition, the manual approach is used to define the control set, which is used to validate the accuracy of static methods used in order to identify secondary crashes. Furthermore, following the identification of secondary crashes, logistic regression and a negative binomial model were developed in order to investigate major factors contributing to the occurrence of secondary crashes.

- Chapter 3 presents a dynamic method in order to identify secondary crashes. Crash data and speed data in the Detroit freeway area were used to identify the impact area of the primary crash and secondary crash identification, respectively. In addition, the manual approach is used to define the control set, which is used to validate the accuracy of dynamic methods used in order to identify secondary crashes.

CHAPTER 2. MANUAL METHOD AND STATIC WINDOW SIZING

The static and dynamic methods were used to identify secondary crashes. In the static method, a fixed spatial and temporal threshold is used for secondary crash identification. In the dynamic method, depending on queue length and traffic flow characteristics impact area of a primary crash varies. Therefore, the actual representation of traffic flow is not considered in the static method. One of the most important aspects of this research is determining whether a crash is actually secondary in nature. This determination is ultimately based upon information from the police crash report forms. To this end, in order to identify secondary crashes, manual approach is used to identify secondary crashes from police crash reports. The result will be used to validate the accuracy of the static method used in the identification of secondary crashes. In the manual approach, narratives from the police crash reports were checked manually. Whereas in the static process, fixed spatiotemporal thresholds were considered to identify secondary crashes.

Data used in this study are drawn from police-reported crash data from the Michigan Traffic Crash Facts (MTCF) data query tool, which is maintained by the Michigan State Police (MSP) Office of Highway Safety Planning (OHSP). This tool allows users to have free access to query all crash reports from Michigan law enforcement agencies dating back to 2004. Detailed information is available from each crash, including PDF copies of the police crash reports. With respect to this study, these reports include essential details, such as the date, time, and location of the crash, and a crash narrative section, which provides details of the circumstances of the crash as determined by the investigating officer.

The study area includes the Michigan interstate mainline system. The study area includes the entire Michigan interstate mainline system. In 2018, a total of 312,798 crashes occurred throughout Michigan and, based on the Highway Class filter on MTCF, 35,123 crashes were indicated to have occurred on the interstate system. Next, crashes occurring on either an interstate exit or entrance ramp were removed using a roadway inventory file provided by the Michigan Department of Transportation (MDOT). Based on this filter 7,359 crashes that occur on-ramps were excluded. Subsequently, 363 crashes were removed where the crash report was either missing or incomplete. The final sample included 26,679 crashes. The individual crash report forms were all subsequently downloaded from MTCF, along with pertinent summary information (e.g., crash-ID, date, time, location, crash narrative) in spreadsheet format. Table 2-1 provides information about the crashes included in the analysis.

Table 2- 1: Crashes used in the analysis

Criteria	**Number**
Total crashes in interstate mainline Michigan	34,437
Missing or incomplete crash reports	363
Crashes on ramps	7,395
Total crashes included in the analysis	26,679

2.1 Keyword-Searching Approach/ Checking Narratives

One of the most important aspects of this research is determining whether a crash is actually secondary in nature. This determination is ultimately based upon information from the police crash report forms. To this end, in order to identify secondary crashes, information from two primary fields in the crash report form was utilized. This included a series of standard fields that are used to designate various subsets of crashes, as well as a keyword search or manual approach that was used to review the narrative section from police crash reports. After identifying those crashes that were secondary in nature, the accuracy of the static window method was used to assess the efficacy of various fixed spatiotemporal time and distance thresholds in identifying secondary crashes.

At the onset of the study, reports for all crashes occurring on the Michigan interstate system in 2018 were obtained from the MTCF database. Police crash reports are critical to identifying secondary crashes as the investigating officers generally have either first- or second-hand information regarding the cause of a crash and various precipitating factors. However, the reporting accuracy depends on officers' training, their understanding of how such crashes are defined, and related knowledge that a primary crash has occurred (Zhang et al. 2020). On the Michigan UD-10 crash report form, the contributing circumstances field indicates those factors that precipitated the occurrence of a crash, see Table 2-2. This field is also useful for explicitly identifying the occurrence of a secondary crash.

Table 2- 2: Contributing circumstance codes on Michigan UD-10 crash report

None	Other
Backup - Other Incident	Glare
Backup - Reg. Congestion	Shoulders
Prior Crash	Traffic Control Device
Unknown	

For each crash report, the unique crash identification number (crash-ID) was determined, along with data from a contributing circumstances field and the officer's crash narrative. The contributing circumstances field provides a list of common factors that are found to precipitate the occurrence of a crash. This field includes three primary codes that may be indicative that a secondary crash has occurred: (1) prior crash; and (2) backup due to other incident. However, prior experience has shown there is often some variability in terms of how different officers complete this and other related fields on the crash report form.

Consequently, as a first step, all narratives for crashes where one of the secondary crash related contributing circumstances were indicted were manually reviewed in order to assess whether the crash was truly secondary in nature. There are two conditions to determine secondary crashes for this method;

1) The prior crash contributing circumstance was selected, and there was no conflicting information in the narrative section (see example Figure 2-1 and 2-2); or
2) The narrative section explicitly indicated the occurrence of a prior crash, though one of the other (i.e., not "prior crash") contributing circumstances was selected.

Based on the crash code, 1,896 crashes were coded as being due, at least in part, to a prior crash under the contributing circumstance field. For all those crashes, crash narratives have been reviewed manually, and the result showed that 277 crashes (14.6 percent) were found to be not meet the conditions and therefore not related to prior crashes. For such crashes, another reason other than a prior crash was mentioned in the narrative as the cause of a crash occurrence, see Table 2-3 for example of miscoded crashes. Also, in case that crash narrative section was blank, crash considered a secondary crash.

Table 2- 3: Example of crashes with secondary crash code that not meet the secondary crash identification conditions

Crash-ID	Crash Code	Narrative
1253395	Backup Due to Other Incident	Unit 1 was traveling E/B on I-96 when she lost control, ran off the roadway to the left, and struck the cable barrier.
1253356	Backup Due to Other Incident	Vehicle 1 spun out after losing control and was struck by Vehicle 2.
1256253	Prior Crash	Driver 1 lost control after hitting a patch of ice. She was adamant that she was not going to fast and that the crash was caused by ice. She left the road and struck the cable barrier.

In the second step, the keyword-searching approach was used to identify additional target crashes based on the crash narratives. The keywords that were used in this method were *previous crash, another crash, prior crash, previous accident, another accident, and prior accident,* which are keywords that are used in the narratives by the officer to describe a secondary crash. These keywords were chosen after a manual review of secondary crash narratives that were identified in the previous step. Based on this method, an additional 249 secondary crashes were identified. The finding from this method also shows that law enforcement typically coded the contributing circumstances as backup due to regular congestion or other incidents instead of prior crashes.

In total, 1,872 secondary crashes were identified based on the crash code and word searching approach. There were 882 cases where the contributing circumstance was noted as a prior crash. Among these, 155 were found to have been due to some other (i.e., non-crash) event, such as a vehicle breakdown. Similarly, 892 of 1,014 crashes where the contributing circumstance was due to backup caused by another incident appeared to have been due to another prior crash. Table 2-4 shows the summary of secondary crash results in a manual approach.

Table 2- 4: Secondary crash results in manual method

Contributing circumstances (crash code)	**Total Nr. of crashes**	**Confirmed Nr. of secondary crashes**	**Other (non-secondary) crashes**
Prior Crash	882	731	155
Backup - Other Incident	1014	892	122
Backup - Reg. Congestion (Identified from Narrative)	5,221	62	5,159
Other (Identified from Narrative)	19,562	187	19,373
Total	26,679	1,872	24,807

Table 2-5 shows the final result from the manual approach. Based on the result from the current method, almost 7.02 percent of the interstate mainline crashes are considered secondary crashes.

Table 2- 5: Summary of manual approach result

TYPE OF CRASH	**NUMBER OF CRASH**	**PERCENTAGE**
Secondary crash	1,872	7.02
Crashes not due to congestion or another crash	24804	92.98
Total	26,679	100

Authority: 1949 PA 300, Sec.257.622
Compliance: Required MSP UD-10E
Penalty: $100 and/or 90 days (Rev 01/2016)

External # 0426203 | Crash ID 1253399

Page 01 of 01
File Class 93001

STATE OF MICHIGAN TRAFFIC CRASH REPORT

Incident # 617718	ORI MI 4106100	Department Name MSP Rockford	Reviewer BRETT SOVA

Crash Date	Crash Time	No. of Units	Crash Type	Special Circumstances	Special Checks
01/03/2018	10:00	01	Single Motor Vehicle	● None ○ Fleeing Police ○ Hit and Run ○ Unknown ○ School Bus ○ Animal	○ Fatal ○ Non-Traffic Area ○ ORV/Snowmobile

County	Traffic Control	Relation to Roadway	Weather	Area
41 - Kent	None	On the Road	Snow	FRWY Other Freeway Areas

City/Twsp	Contributing Circumstances	Light	Road Surface Condition	Total Lanes	Speed Limit	Posted
11 - Grand Rapids Twp	1st Prior Crash 2nd	Daylight	Snow	02	70	Yes

Work Zone (if applicable) Type | Workers Present | Activity | Location

LOCATION

Prefix	Primary Road Name	Road Type	Suffix	Divided Roadway
	I-96	FWY		

Distance / Direction 528 Feet W | Trafficway Divided Hwy with Barrier

Prefix	Intersecting Road Name	Road Type	Suffix	Divided Roadway
	FULTON	AVE		

UNIT / DRIVER

Unit Number	Unit Known	State	Driver License Number	Date of Birth (Age)	License Type	Endorsements	Sex	Total Occupants	Hazardous Action
01	Yes	MI	#############	##/##/#### (34)	● Operator ○ Chauffeur ○ Moped	○ Cycle ○ Farm ○ Recreation	F	01	Speed Too Fast

Unit Type	Driver Information	Driver is Owner	Injury	Position	Restraint
MV	################################ ################################ GRAND RAPIDS, MI 49525 (###) ###-####	Yes	O	Front - Left	Shoulder and Lap Belt

Driver Condition at Time of Crash	Driver Distracted By	Ejected	Trapped	Airbag Deployed
1st Appeared Normal 2nd	Not Distracted			Not Deployed

Hospital NONE | Ambulance NONE

Alcohol Suspected	Contributing Factor	Alcohol Test Type	Alcohol Test Results	Interlock Device
No	No	○ Breath ○ Blood ○ Urine ○ Field ○ PBT ○ Refused ● Not Offered	○ Pending Test Results:	No

Drug Suspected	Contributing Factor	Drug Test Type	Drug Test Results	Citation Issued
No	No	○ Blood ○ Urine ○ Field ○ Refused ● Not Offered	○ Pending Test Results:	○ Hazardous ○ Other

Vehicle Registration	State	Vehicle Description Year	Make	Model	Color
CFY8802	MI	2010	FORD	LGT CONVTNL 'F'	BLK

VIN	Vehicle Type	Special Vehicles	Private Trailer Type	Vehicle Defect
1FTEW1E87AFC65327	Pickup Truck	Not Applicable		

Automation System(s) in Vehicle | Automation System Level in Vehicle | Automation System Level Engaged at Time of Crash

Insurance Company	Insurance Policy #	Towed By	Towed To
################################	################################	DRIVEN/DRIVER	

Location of Greatest Damage	First Impact	Extent of Damage (Power Unit and/or Trailers)	Vehicle Direction	Vehicle Use	Action Prior
08	08	Functional Damage	E	Private	Avoiding Veh Front/Back

Sequence of Events (● indicates MOST harmful event) | First 01 - Loss of Control | Second 03 - Ran Off Roadway-Left | Third ● 25 - Guardrail Face | Fourth

PASSENGERS

Passenger Information | Date of Birth (Age) | Sex | Position | Restraint | Injury | Ejected | Trapped | Airbag Deployed
Hospital | Ambulance
Passenger Information | Date of Birth (Age) | Sex | Position | Restraint | Injury | Ejected | Trapped | Airbag Deployed
Hospital | Ambulance

TRUCK/BUS

Carrier Information | USDOT | MC | MPSC
Driver's CDL Type | Endorsements ○H ○P ○T ○N ○S ○X | CDL Exempt ○ Farm ○ Other
GVWR/GCWR ○ 10,000 lbs. or Less ○ 10,001 - 26,000 lbs. ○ Greater than 26,000 lbs. | Vehicle Configuration | Cargo Body Type | Medical Card | Hazardous Material ○ Placard ○ Cargo Spill | ID # | Class #

OWNERS

Owner Information | Owner Information

Damaged Property | Public | Owner & Phone

Figure 2- 1: Example crash report and narrative indicating a secondary crash

UNIT / DRIVER

Unit Number | Unit Known | State | Driver License Number | Date of Birth (Age) | License Type: O Operator, O Chauffeur, O Moped | Endorsements: O Cycle, O Farm, O Recreation | Sex | Total Occupants | Hazardous Action

Unit Type | Driver Information | Driver is Owner | Injury | Position | Restraint

Driver Condition at Time of Crash: 1st, 2nd | Driver Distracted By | Ejected | Trapped | Airbag Deployed

Hospital | Ambulance

Alcohol Suspected | Contributing Factor | Alcohol Test Type: O Breath, O Blood, O Urine, O Field, O PBT, O Refused, O Not Offered | Alcohol Test Results: O Pending, Test Results: | Interlock Device

Drug Suspected | Contributing Factor | Drug Test Type: O Blood, O Urine, O Field, O Refused, O Not Offered | Drug Test Results: O Pending, Test Results: | Citation Issued: O Hazardous, O Other

Vehicle Registration | State | Vehicle Description: Year, Make, Model, Color

VIN | Vehicle Type | Special Vehicles | Private Trailer Type | Vehicle Defect

Automation System(s) in Vehicle | Automation System Level in Vehicle | Automation System Level Engaged at Time of Crash

Insurance Company | Insurance Policy # | Towed By | Towed To

Location of Greatest Damage | First Impact | Extent of Damage (Power Unit and/or Trailers) | Vehicle Direction | Vehicle Use | Action Prior

Sequence of Events (● indicates MOST harmful event): First, Second, Third, Fourth

PASSENGERS

Passenger Information | Date of Birth (Age) | Sex | Position | Restraint | Injury | Ejected | Trapped | Airbag Deployed

Hospital | Ambulance

Passenger Information | Date of Birth (Age) | Sex | Position | Restraint | Injury | Ejected | Trapped | Airbag Deployed

Hospital | Ambulance

TRUCK/BUS

Carrier Information | USDOT | MC | MPSC

Driver's CDL Type | Endorsements: O H, O P, O T, O N, O S, O X | CDL Exempt: O Farm, O Other

GVWR/GCWR: O 10,000 lbs. or Less, O 10,001 - 26,000 lbs., O Greater than 26,000 lbs. | Vehicle Configuration | Cargo Body Type | Medical Card | Hazardous Material: O Placard, O Cargo Spill | ID # | Class #

OWNERS

Owner Information | Owner Information

WITNESS

Witness Information | Witness Information

Investigated at Scene	Reported Date (Time)	1st Investigator Name (Badge)	2nd Investigator Name (Badge)	Photos
Yes	01/03/2018 (10:19)	JARED J ORBAN (231)		No

Narrative

Unit 1 was traveling E/B on I-96 when she tried to avoid another crash in front of her, lost control, ran off the roadway to the left, and struck the guardrail.

Diagram

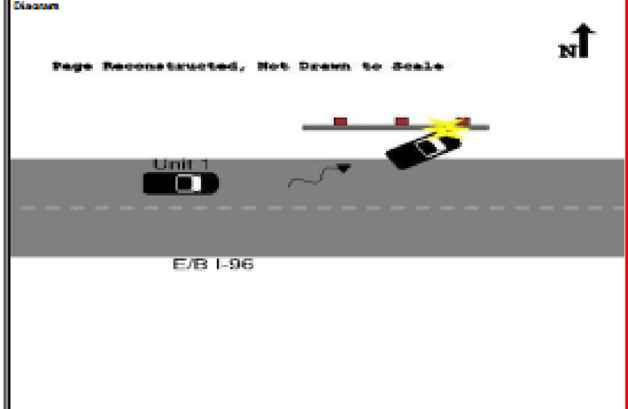


Figure 2- 2: Example crash report and narrative indicating a secondary crash

2.2 Static Sizing: Spatiotemporal Window

Under the static approach, a crash is classified as secondary in nature if it falls within a predefined time-space window originating from another (prior) crash. In order to identify potential secondary crashes, each crash is associated with its corresponding interstate road number, the geospatial location on the road, and the associated date and time. Using linear referencing in ArcGIS, the exact locations for each crash along a particular highway were determined based upon a route-specific identification number and a mile marker. Consecutive crashes on each road segment were identified based on date and time.

The distance between two consecutive crashes was calculated from the difference between corresponding mile points. It is essential to mention that each direction has been considered separately, and only crashes that are happening in the same direction and upstream of the primary crash have been considered. For each crash, a spatiotemporal window was assigned, and then the events in the window were recorded as a secondary crash. In view of the large size of the database, nearest neighboring methods were coded in MAPLE, which is a math software to enable global identification of the nearest event in the crash database[1].

The problem with the spatiotemporal window can be best summarized in Figure 2-3, which shows increasing the size of the window will increase true positives but will also increase false positives. Accordingly, there is an inherent trade-off between sensitivity and specificity of the given method, which can be tweaked to achieve a comprehensible result. Here the sensitivity defines as the probability of correctly identifying a secondary crash versus specificity is a

[1] Maplesoft, a division of Waterloo Maple Inc.. (2019). *Maple*. Waterloo, Ontario. Retrieved from https://hadoop.apache.org

probability of correctly identifying a non-secondary crash. In the following section, this trade-off will be explored.

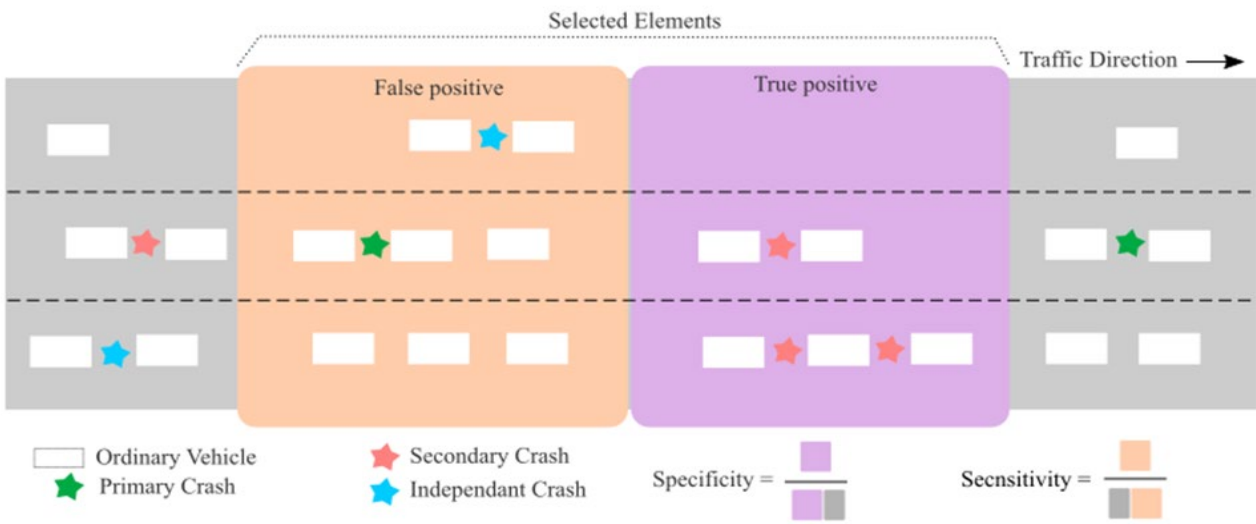


Figure 2- 3: Trade-off between sensitivity and specificity

2.3 Analysis and Results

Sizing of ST window: After determining the spatiotemporal thresholds between consecutive crash events within the 2018 interstate crash dataset, different time and distance intervals were used to define different sizes of spatiotemporal windows. Figure 2-4 (a) shows the probability density function for all crashes happening within 15 minutes of another crash and within different distance gaps from 1/2 to 5 miles. The inset shows further details within a one-mile radius. As can be seen, most of the crashes that are potentially secondary in nature occur within the first 0.2 miles and 15 minutes from the primary crash. Figure 2. 4 shows the probability density function for all crashes happening within the first mile gap and different time gaps from 0 to 120 minutes. As shown in Figure 2-4, within the first-mile gap, most crashes occur in the first 7-minute period after the primary crash, with gradual and persistent decreases in subsequent thresholds.

Considering the average frequency of occurring incidents to be a constant throughout the search space, the elevation in density near the peak shows the high specificity which is probability of correctly identifying a non-secondary crash of the static window to crashes that occurred in that region. As expected, the specificity fades as the window size increases while the sensitivity, probability of correctly identifying a secondary crash, increases.

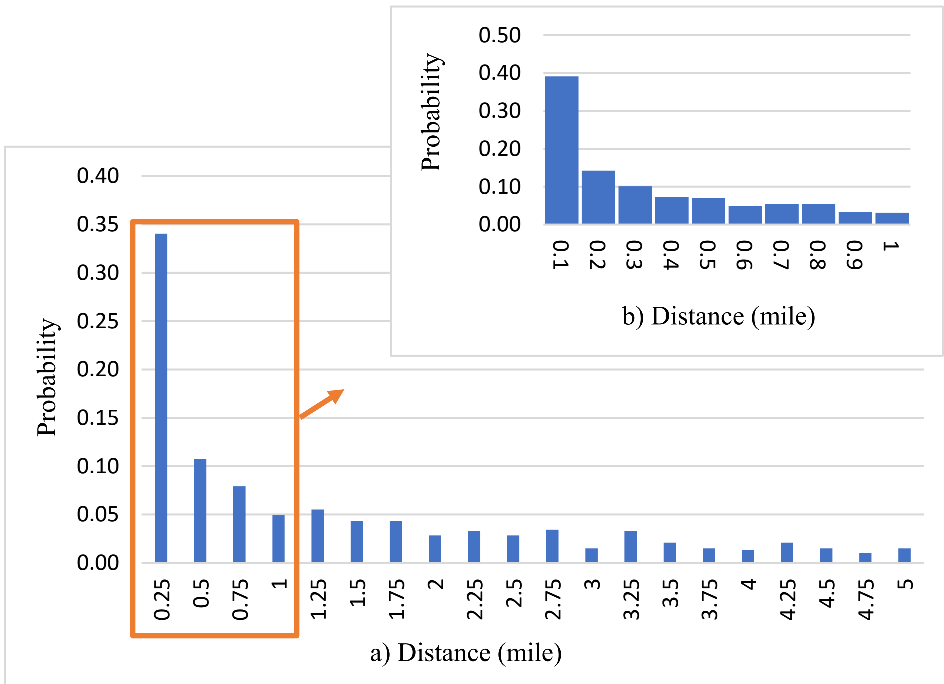


Figure 2- 4: (a) Density function for fixed time grid=15 minutes and various distance grid, (b) Higher resolution inset of distribution in shorter distance (1 mile)

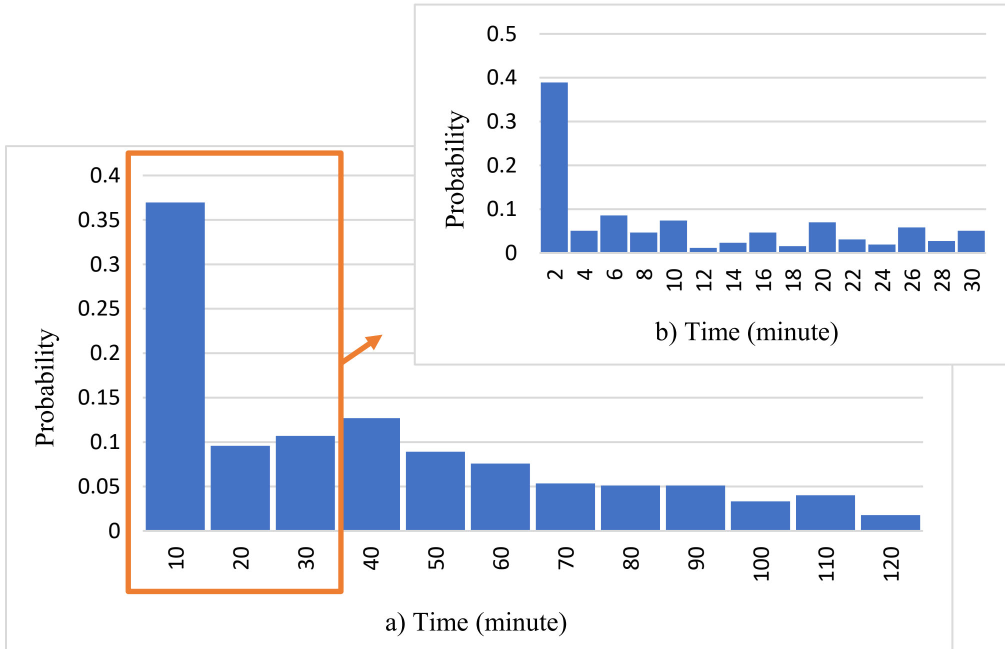


Figure 2- 5: (a) Density function for fixed distance grid =0.05 and various time grid (b) Higher resolution inset of distribution in shorter time gap (30 minutes)

In order to understand the effect of window size on the accuracy of the predictions, one can plot the predictions obtained with a spatiotemporal approach against the actual confirmed events as determined by the crash code, and manual approach described previously. To this end, Figure 2 – 6 (a) was plotted with respect to the following four parameters

- $N_{M[L,T]}$: Number of confirmed secondary crashes identified by the manual approach which fall within a specific spatiotemporal window from the first crash
- $N_{M-Total}$: Total number of confirmed secondary crashes identified by manual approach in the largest window
- $N_{S[L,T]}$: Number of crashes that exists within a specific spatiotemporal window from the first crash
- $N_{S-Total}$: Number of crashes in the largest window

Figure 2-6. (a) demonstrates the normalized plot of the secondary crashes occurring within spatiotemporal windows of different distances with increments of one mile (a fixed time gap of 15 min is assumed) against the total number of secondary crashes identified by manual word searching within the largest window. Here the largest window is 15 minutes and 6 miles. The total crashes within this time gap window is 977 and from those 171 confirmed secondary crashes. The red line shows the normalized plot of the manually identified secondary crash with respect to different window sizes. As expected, as the window size increases, all secondary crashes identified by the manual approach will be covered by the spatiotemporal window. To describe what percentage of the crashes that fall within the spatiotemporal window are secondary crashes, Figure 2-6. (b) was developed where the ratio of the secondary crash to the total number of crashes for different sizes of windows was plotted. Similarly, as the spatiotemporal window grows, the sensitivity of the static method fades due to the large number of non-secondary crashes that are included (false positives).

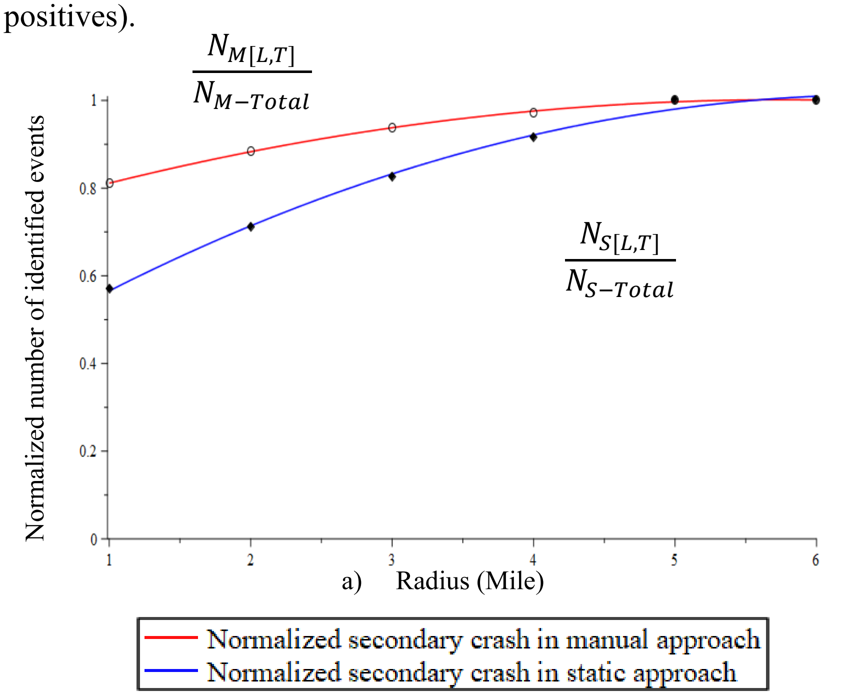
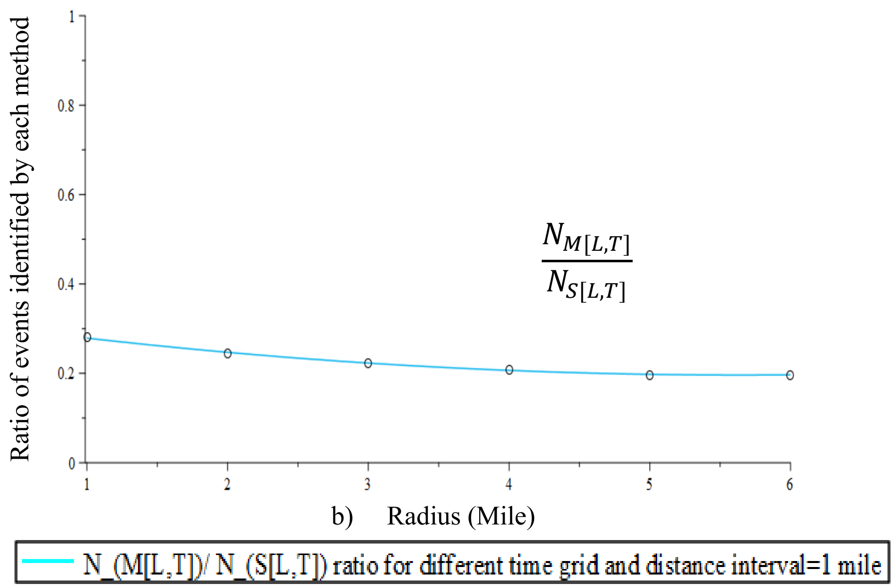


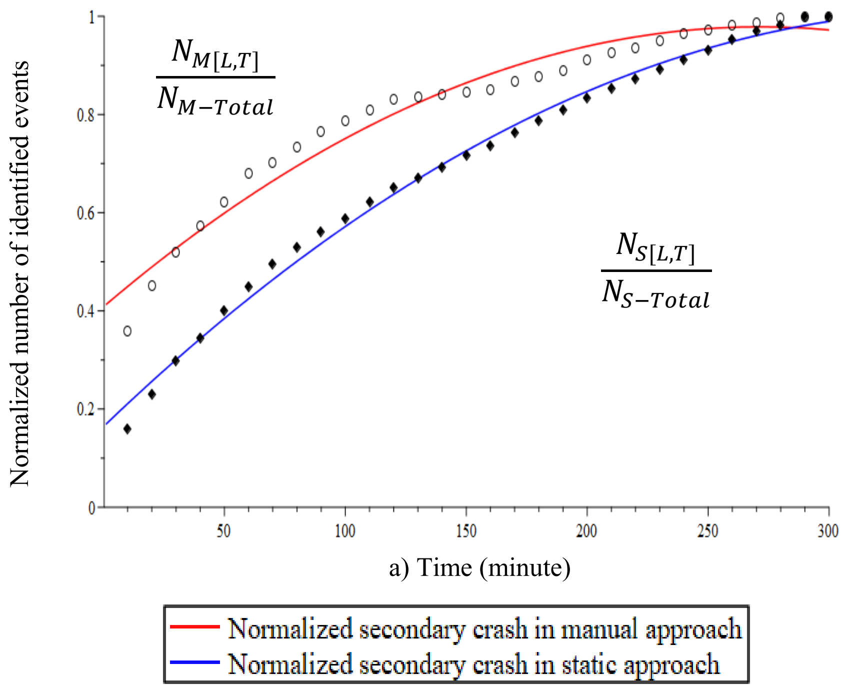
Figure 2- 6: (a) Normalized plot of accumulated events registered by manual and static methods in windows with a time size of 15 minutes and various distance gaps

Figure 2- 6 (Cont'd)



(b) Accuracy of the static method shown by the number of confirmed secondary crashes captured vs. those captured by static method for with time size of 15 minutes and various distance gap

Figure 2 - 7 shows that a similar statement can be made when the windows are growing in the time dimension, as well. While the offset and the slope of the normalized static method and manual approach curves may be different (compare Figure. 2-6 (a) and Figure. 2-7 (a), the blue line shows the ratio of crashes within the designated spatiotemporal window to the total number of crashes in the static method. This fact is better shown in Figures 2-6. (b) and 2-7. (b), where the ratio of confirmed secondary crashes that were identified in the manual approach (red curve) against crashes within the spatiotemporal window (blue curve) is plotted. It should be noted that the total number of crashes in the largest window here within a 6-mile distance gap (Figure 2-6) and 300 minutes time gap (Figure 2-7) is different therefore, the percentages vary in Figure 2-6. (a), and Figure 2-7. (a), accordingly.



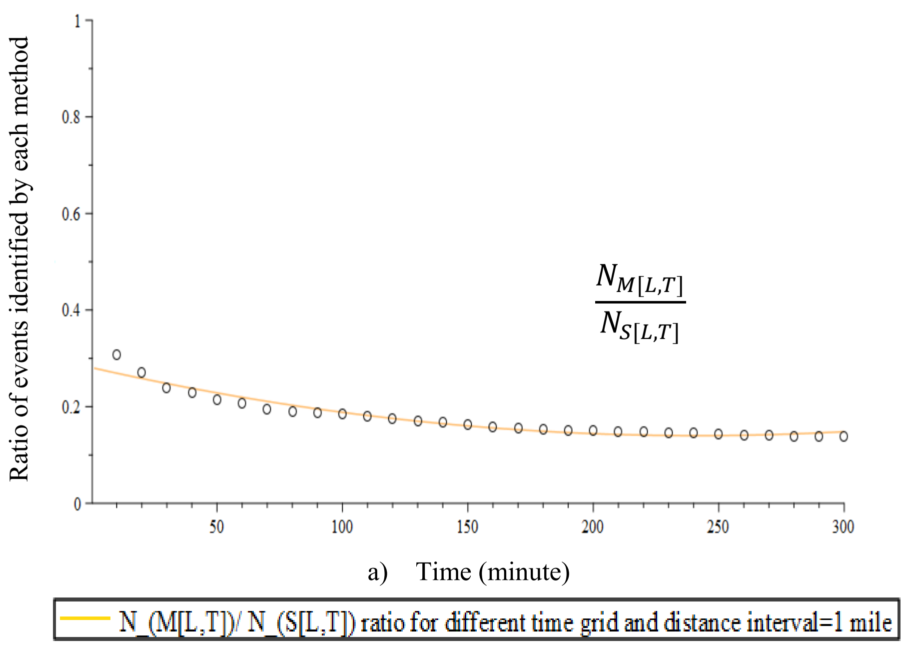


Figure 2- 7: (a) Normalized plot of accumulated events registered by manual approach and static methods in windows with a gap size of 1 mile and various time gaps, (b) Accuracy of the static method shown by the number of confirmed secondary crashes captured vs. those captured by a static method.

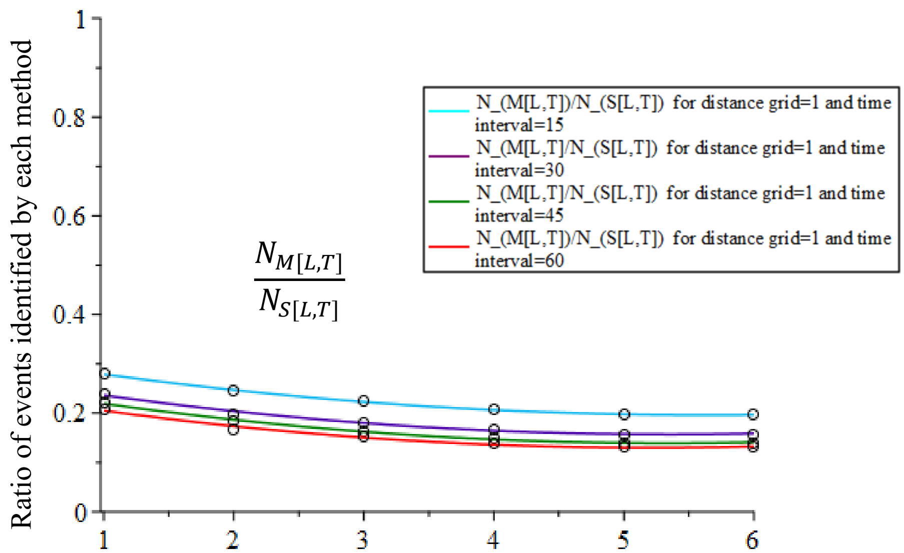
In order to illustrate the loss of accuracy by decreasing sensitivity, the ratio of all verified secondary crashes to the estimated number of secondary crashes under the static approach is evaluated by plotting $N_{M[L,T]} / N_{S[L,T]}$ for different sizing of spatiotemporal windows. This plot shows that the sensitivity (i.e., probability of correctly identifying a secondary crash) of the static method is highest at the smallest time and distance windows. The proportion of secondary crashes that are correctly identified, α, is illustrated on windows with different spatial and temporal sizes (see Figure 2-8. (a) and Figure 2-8. (b). In general, Figure 2-8 suggests that the static method performs poorly at larger time and distance thresholds. The general trend here implies that the rate of crashes identified by the static method stabilizes at distances of approximately 3 miles and time periods of approximately 60 minutes in these scenarios. The same general pattern is observed in the analysis of different geographic regions, within the same regions during different seasons, and across different highway segments. In other words, one can say Equation 1,

$$\boldsymbol{N_{M[L,T]}} = \boldsymbol{\alpha}\, \boldsymbol{N_{S[L,T]}} \quad \text{where} \quad \boldsymbol{\alpha} \cong [0.27 - 0.09] \qquad (1)$$

- $N_{M[L,T]}$: Confirmed secondary crash events in the manual approach
- $N_{S[L,T]}$: Number of crashes that exists within a specific spatiotemporal window from the first crash
- α: Convergence limit (Sensitivity)

Therefore, α is the sensitivity of the of secondary crashes identified by the static window approach. It can be seen that within the aforementioned spatiotemporal window, as the window grows, the sensitivity drops to reach the line which has a constant declining rate which is correlated with the linear expansion of window size. The declining rate of the line can be considered almost constant since after certain window size based on the literature, secondary crashes rarely occur beyond a time and distance thresholds, here for windows larger than 6 distance and 300 minutes time.

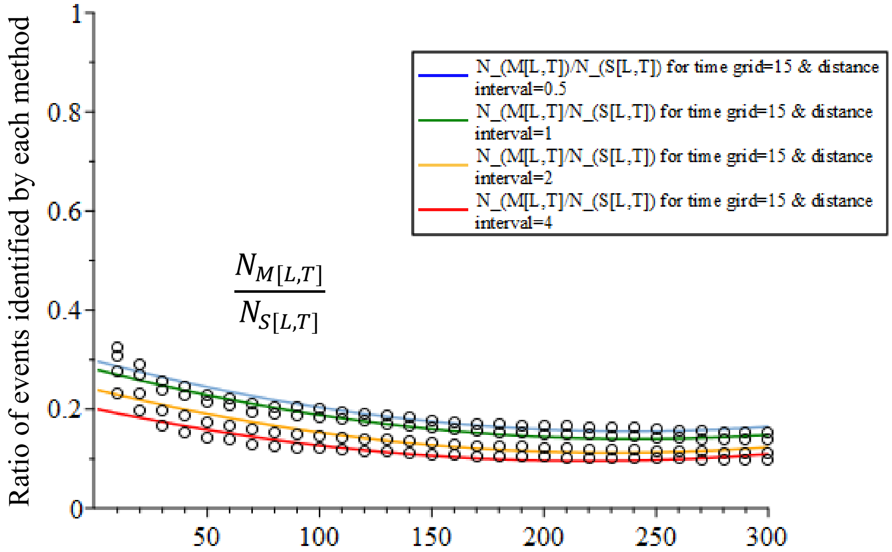
The maximum drop in sensitivity occurs right before merging of the α to the line Therefore, a spatiotemporal window can be used to estimate the number of confirmed secondary crashes identified by the manual approach.



a) Radius (mile)

Figure 2- 8: (a) The ratio of confirmed secondary crash events in the manual approach to the total predicted events in static approach within a gap size of 1 mile and various time intervals. (b) The ratio of confirmed secondary crashes in the manual approach to the total predicted events in the static approach within a gap size of 15 minutes and various distance gaps.

Figure 2- 8 (Cont'd)



b) Time (minute)

Table 2-6 shows the number of crashes within each spatiotemporal window (projected positive) and the number of confirmed secondary crashes and the ratio within each spatiotemporal window (true positives). The table shows the specificity and sensitivity.

Table 2- 6: Secondary crash distribution for interstate roads in Michigan based on static and manual approach

Distance grid (Mile)	Time gap (Min)	Number of crashes in spatiotemporal window N_S[L,T]	Number of verified secondary crashes within spatiotemporal window N_M[l,T]	Specificity (within 300min, 6mile)	Sensitivity
1	15	509	142	93%	27.70%
	30	773	185	93%	24.00%
	60	1155	254	94%	21.80%
	300	2605	362	94%	13.80%
3	15	740	166	93%	22.40%
	30	1207	220	93%	18.30%
	60	1929	318	94%	16.50%
	300	5151	526	95%	10.20%

Table 2- 6 (Cont'd)

6	15	977	171	93%	17.50%
	30	1611	235	93%	14.60%
	60	2764	354	94%	13.40%
	300	7431	638	unknown	8.90%

A similar analysis has been done for each interstate roadway in Michigan in order to identify secondary crashes in each freeway. The goal was to determine which road is more critical and concerned in the possibility of secondary crashes occurrence. As previously mentioned, each direction has been considered separately. For each primary crash, crashes that occur in the same direction and upstream of a primary crash, are considered, and their spatiotemporal gap has been recorded. Table 2-7 shows percentages of secondary crashes in each of thirteen interstate roadways in Michigan. The results are based on the 6 miles and 300 minutes space-time window.

Table 2- 7: Secondary crashes for interstate roads in Michigan based on static and manual approach

Freeway	**Number of Crashes**	**Number of confirmed secondary crashes in manual approach**	**Percentages of confirmed secondary crashes in manual approach**	**Number of crashes in spatiotemporal window**	**Number of confirmed secondary crashes in spatiotemporal window**	**Percentages of secondary crashes in spatiotemporal window**
I-69	1835	125	6.8	199	33	16.6
I-75	7041	423	6.0	3016	182	6.03
I-94	7564	605	8.0	1627	180	11.1
I-96	5021	336	6.7	1072	116	10.8
I-194	52	2	3.8	12	0	0.0
I-196	1439	112	7.8	309	42	13.6
I-296	239	23	9.2	110	10	9.1
I-375	95	3	3.1	21	3	14.3
I-475	286	15	5.2	166	6	3.6
I-496	400	42	10.5	102	11	10.8
I-675	90	4	4.4	47	4	8.5
I-696	1984	152	7.7	533	50	9.4
I-275	633	32	5.1	366	14	3.8
Total	26,679	1,872	7.0	7586	651	8.6

A similar correlation factor has been observed in this set of results. The number of secondary crashes identified by static methods in each road is higher than the number of secondary crashes identified by the manual approach. Based on the result, interstate roads I-496 and I-375 are assumed to have the highest and the lowest rate of secondary crashes by 10.5 percent and 3.1 percent consecutively.

Figure 2-9 shows the comparison of the spatiotemporal distribution of crashes in static window versus distribution of confirmed secondary crashes in relation to previous crash temporally, Figure 2-9 (a) and spatially, Figure 2-9 (b). Both figures show the frequency of crashes are higher in shorter time and distace interval. In addition, the crash frequency drops with increase in time and distance gap.

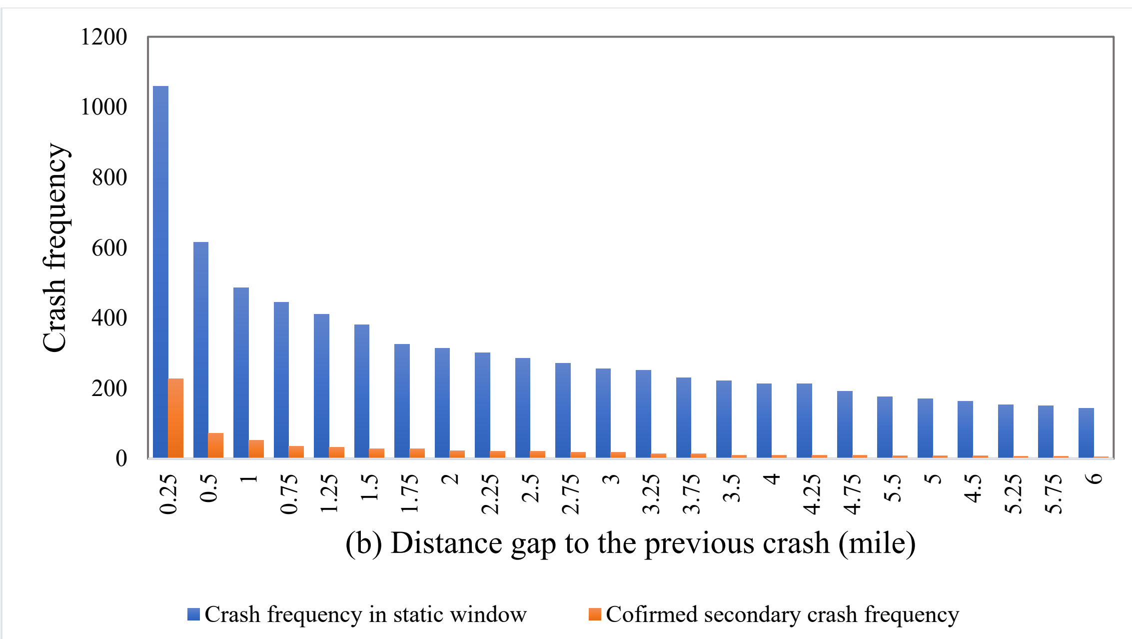


Figure 2- 9: Comparison of the spatiotemporal distribution of crashes in static window versus distribution of confirmed secondary crashes in relation to previous crash (a) Temporal distribution (b) Spatial distribution.

Figure 2-9 (Cont'd)

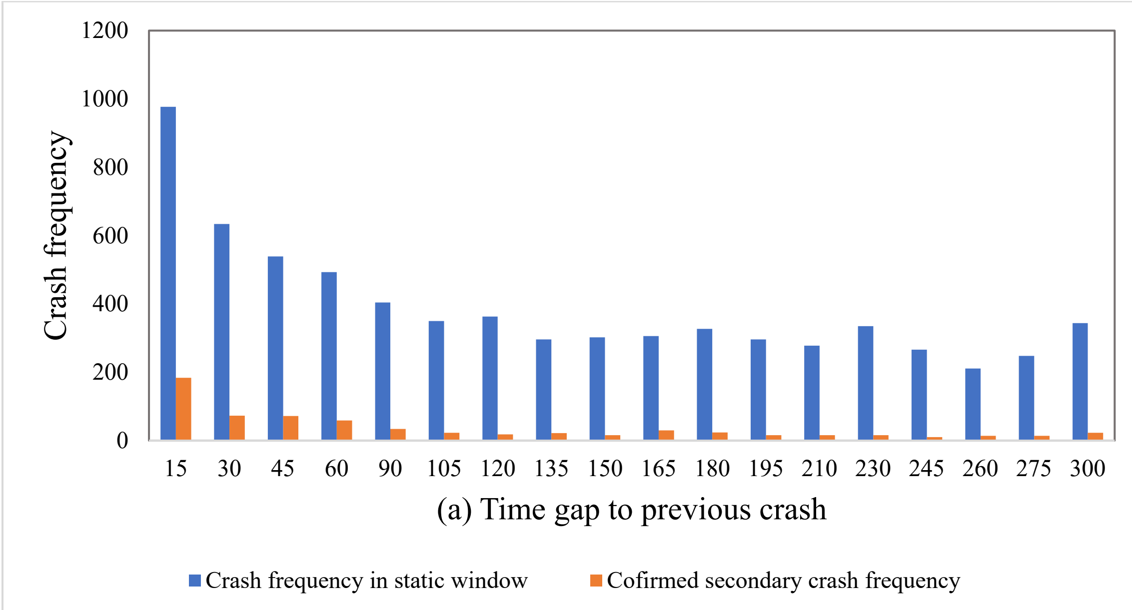


Figure 2-10 shows the temporal and spatial distribution and characteristics of the actual confirmed secondary crashes within each static window. Temporally, approximately 65 percent of the secondary crashes were found to occur within 90 minutes time gap from the previous crash. Spatially, about 80 percent of the secondary crashes occurred within a 2.5-mile distance gap from the previous crash. Generally, about 60 percent of secondary crashes occurred within 75 minutes of the time gap of the previous crash and within one mile upstream of the previous crash. In other words, about 40 percent of secondary crashes occurred beyond the most commonly used one mile and 75 spatiotemporal thresholds.

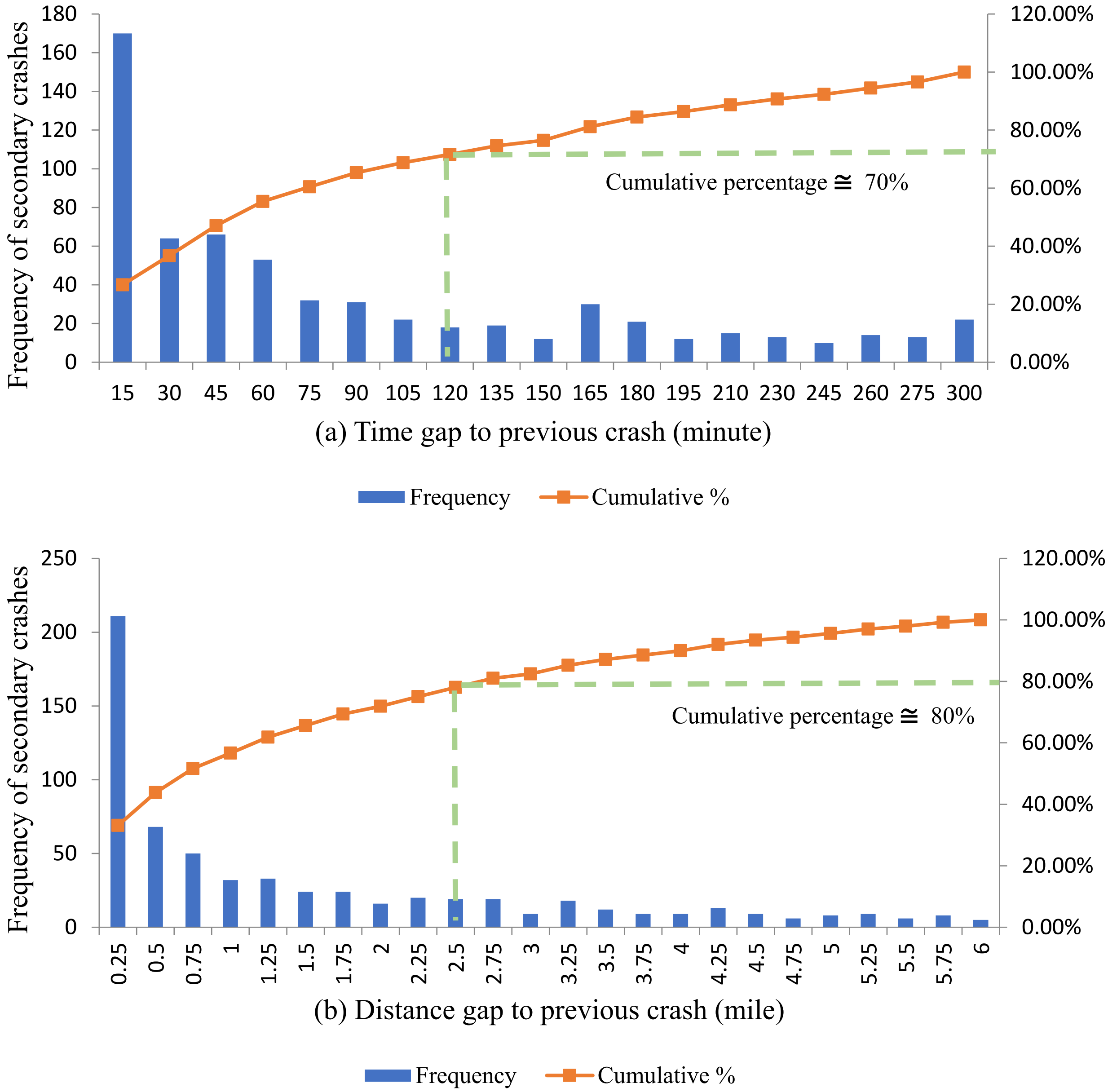


Figure 2- 10: Spatiotemporal distribution of secondary crashes in relation to previous crash (a) Temporal distribution (b) Spatial distribution.

2.4 Discussion and Conclusion

Crashes constitute a significant source of delays, system unreliability, and inefficiency on freeways. The congestion caused by primary crashes often exposes the subsequent vehicle to the risk of secondary crashes. While secondary crashes are relatively infrequent, they pose a significant safety risk in freeways and highly affect traffic operations and flow. Despite substantive research efforts, there is still considerable uncertainty as to the magnitude and nature of secondary crashes. The spatial and temporal influence of primary crashes on road users are closely related to occurrences of secondary crashes. Some studies, mostly based on static methods, have defined secondary incidents based on fixed spatial and temporal thresholds. In this approach, a fixed spatiotemporal window is assumed around the primary crash. In addition, the static approach considers the same window for all types of primary crashes regardless of the upstream traffic flow, density and speed.

In this work, by leveraging a huge database of all events on Michigan Interstate roads in 2018, a keyword-searching/manual approach has been performed to define the control set of a secondary crash based on police reports. Results from manual approach are then used to validate the accuracy of the static method in order to identify secondary crashes. Based on manual results, about 7 percent of interstate crashes were recorded by police officers as secondary crashes. In addition, a large set of static window sizes was explored, and it was found that while predicting secondary crashes with fixed-size windows yield a significant overestimate, window sizes can be used to derive values that are linearly correlated with the confirmed number of secondary crashes regardless of the window size, traffic flow, density, and speed.

By benchmarking secondary crash densities identified using different static thresholds with confirmed secondary crash density obtained by the manual approach, it has been shown that the static method consistently overestimates secondary crash rates, this can be seen in Figure 2-11. Table 2-8 shows the result from some of the previous studies which applied a static approach to identify secondary crash rates, and Figure 2-11 demonstrate the comparison of the result from the previous studies with the result from the current study considering different spatiotemporal thresholds.

Table 2- 8: Summary of secondary crash rates in literate

Study	**Secondary Crash Rate**	**Spaciotemporal Threshold**
Raub (1997)	15%	15 min and 1 mile
Karlaftis et al. (1999)	35%	15 min and 1 mile
Moore et al. (2004)	1.5% to 3%	2 hours and 2 miles upstream in both directions
Kopitch and Saphores (2011)	5.53%	2 hours and 2 miles upstream in both directions
Green et al. (2012)	0.10% to 0.15%	80 min and 1,000 ft
Zhan et al. (2008)	7.90%	Clearance time + 15min and 2 miles

Figure 2-11 shows the comparison of the secondary crash rates from previous studies with the secondary crash rates within the current study. The blue dots in Figure 2-11 shows the secondary crash rate in different studies considering the static method and designated spatiotemporal thresholds. The orange color dots show the secondary crash rates within the current research regardless of the spatiotemporal thresholds.

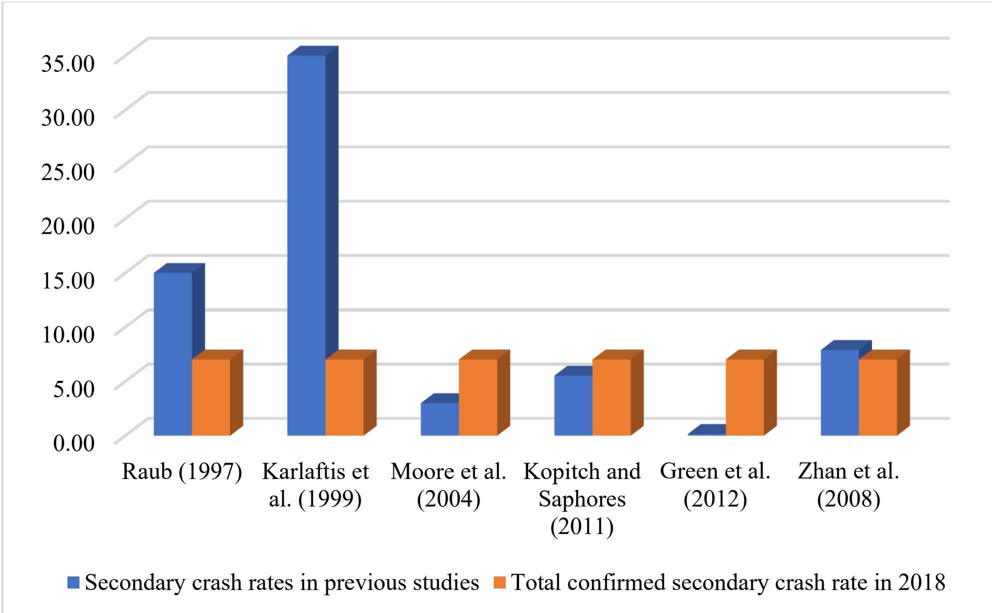


Figure 2- 11: Comparison of secondary crash rates in previous studies which applied static approach with the current study

It should be noted that secondary crashes occur within the spatiotemporal impact area of the primary crash therefore, shorter spatiotemporal windows have been considered. It was found that with the increase in spatiotemporal window sizing, the specificity fades as the sensitivity increases. Identifying the factors that lead to secondary crashes is the first step toward preventing the occurrence of secondary crashes. Existing studies have used several statistical models to analyze the risk of secondary crash occurrence. The current research has adopted logistic regression and negative binomial models to identify characteristics that distinguish secondary crashes from primary crashes. This study's proposed methodological approach and research findings provided insights into the effects of traffic conditions, geometric characteristics, weather conditions, and primary crash characteristics on the probability of multiple secondary crashes on freeways.

The logistic regression model suggests that the number of lanes, weather conditions, posted speed limit, crash severity, which involves fatal injury, number of units involved in the crash, and crashes with emergency medical service involved are among key variables that affect secondary crash occurrence. The negative binomial model suggests that annual average daily traffic (AADT), large urbanized areas (with a population of more than 200,000), and median with concrete barriers are among the key variables that affect secondary crash occurrence. This result is expected to provide useful information in developing policies and strategies to prevent the occurrence of secondary crashes. Moreover, the developed model can also be incorporated into advanced traffic control systems on freeways to avoid the occurrence of secondary crashes.

Secondary crashes caused by other non-crash incidents and also the effect of crashes in the opposite traffic direction deserve more investigation. In summary, the static method may fail to capture the impact area of primary crashes and often overestimate the secondary crash by considering all the nearby events as the secondary crash. On the other hand, dynamic approaches address this limitation by determining the spatiotemporal thresholds of primary crashes based on real-time traffic flow characteristics such as speed and density. Further investigation and dynamic method are recommended for future study.

CHAPTER 3. SECONDARY CRASH IDENTIFICATION BASED ON SPEED DATA

Secondary crashes occur within the impact area of a prior incident and can lead to an increase in traffic flow, fluctuation, and risk of subsequent crashes. In order to mitigate the safety impact and congestion associated with secondary crashes, strategies should be developed to reduce the potential for such crashes. As described in the previous section, the static method identifies secondary crashes based on pre-specified spatiotemporal parameters. It has serious limitations as it fails to capture the actual impact range of primary crashes.

Dynamic methods address the limitations associated with static methods. Despite their widespread application, static studies generally run into concerns as to their reliability due to their one-size-fits-all approach to the problem. Many prior studies using static methods have also assessed sensitivity of the results without explicitly validating secondary crash estimates with ground truth data as to the actual number of crashes in a large pool of data. Such approaches generally result in an overestimation of actual secondary crashes. The static threshold method also generally does not consider the actual representation of traffic conditions. The influence area, from both a temporal and spatial perspective, is expected to vary based upon real-time traffic flow characteristics (e.g., speed, density) and other factors. Compared to the static approach, the dynamic method is more advanced and reliable by limiting the search space based on traffic flow characteristics rather than assigning a static spatiotemporal window. However, the implementation of the dynamic approach depends on the availability of real-time traffic data. While traffic sensors for real-time traffic flow measurements are only available on limited access facilities, the use of the dynamic method is limited to the locations with available sensor data. Moreover, this method is resource-hungry and data-intensive. In this thesis, a dynamic secondary crash identification method is proposed, which focuses on estimating the impact range of the primary crash using speed data. The proposed approach aims to use the data from traffic flow characteristics, such as speed, which change over space and time to describe the queue formation as a result of a primary crash.

The contributions of this research are summarized as follows:

- Identify secondary crashes from the integration of the speed contour plot and the spatiotemporal evolution of the primary crash impact area.
- The current method can determine impact areas associated with multiple incidents and confirm that each impact area is consistent with the spatiotemporal evolution of shockwaves.
- The proposed approach should lead to reducing the misidentification of secondary crashes compared to the static approach that considers fixed spatiotemporal thresholds.
- Lastly, this research aims to identify those contextual environments where the risks of secondary crashes are most pronounced, culminating in guidance to assist road agencies in effectively monitoring and clearing crashes and other incidents to minimize the potential for secondary crashes

3.1 Data Acquisition

Data used in this study are drawn from police-reported crash data from the Michigan Traffic Crash Facts (MTCF) data query tool, which is maintained by the Michigan State Police (MSP) Office of Highway Safety Planning (OHSP). This tool allows users free access to query all crash reports from Michigan law enforcement agencies dating back to 2004. Detailed information is available from each crash, including PDF copies of the police crash reports. With respect to this study, these reports include important details, such as the date, time, and location of the crash, as well as a crash narrative section, which provides details of the circumstances of the crash as determined by the investigating officer.

In 2018, a total of 312,798 crashes occurred throughout Michigan, and based on the Highway Class filter on MTCF, 34,437 crashes were indicated to have occurred on the interstate system. Next, crashes occurring on either an interstate exit or entrance ramp were removed using a roadway inventory file provided by the Michigan Department of Transportation (MDOT). Based on this filter, 7,359 crashes that occur on-ramps were excluded. Subsequently, 363 crashes were removed where the crash report was either missing or incomplete. Given the resources required for this dynamic analysis, the study area was constrained to include only the Detroit metro area interstate mainline system. Interstate in Detroit area includes all roads that are located in Macomb, Oakland, and Wayne county. The final dataset included a total of 13,392 crashes in the Detroit area. The individual crash report forms were all subsequently downloaded from MTCF, along with pertinent summary information (e.g., crash-ID, date, time, location, crash narrative) in spreadsheet format.

In addition, real-time traffic data and speed from the Regional Integrated Transportation Information System (RITIS) website were used in this study. "RITIS is an automated data sharing, dissemination, and archiving system that includes many performance measures, dashboard, and visual analytics tools that help agencies to gain situational awareness, measure performance, and communicate information between agencies and to the public"[2]. Real-time speed data for every 15-minute interval for every interstate segment was downloaded from RITIS. In order to acquire stable traffic flow rates, literature recommended utilizing a minimum of 15 minutes measurement

[2] https://ritis.org/

intervals (Smith and Ulmer, 2003). It should be noted that natural traffic flow data at shorter time intervals may contain a large amount of noise (Guo *et al.*, 2017).

Michigan roadways consist of different PR-Numbers, and each PR-Numbers consists of different XD-segments with different mile points. PR-Number is the physical road number of the segment, as imported from the Michigan Geographic Framework and XD-segment stands for extreme definition segment. Based on the definition, "XD-segments are segments that cover more miles of road than TMC segments, generally with greater granularity, and with the ability to adapt more quickly to changes in the road network and the addition of new roads and new markets" (*Glossary - INRIX*, no date). In total, there are 967 segments and 32 PR-number within Detroit area interstate roadways, see Figure 3-1.

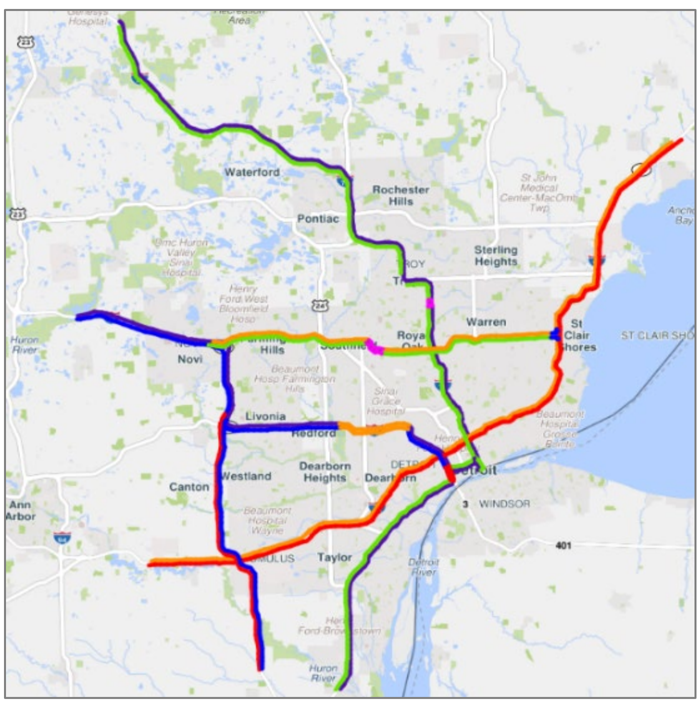


Figure 3- 1: Interstate roadways in the Detroit area

From speed data downloaded from RITIS, speed data were missing for 83 segments (Figure 3-2). For those segments, the speed will be interpolated based on speed data from adjacent segments. Missing data replaced by the average speed of the segment below and above that missing segment.

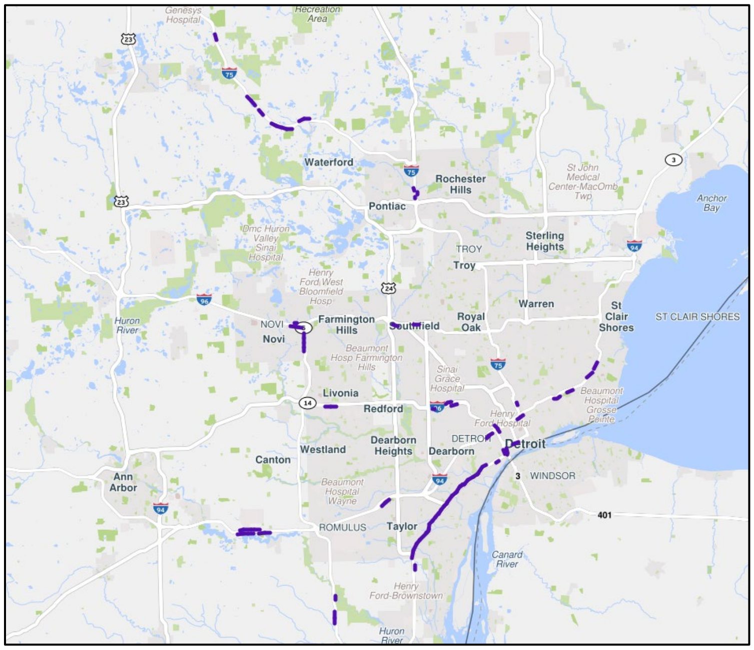


Figure 3- 2: Segments that speed data is missing

ArcGIS was used to create a new linear reference system based on the Detroit crash data and prepared linear referencing files for XD-segments, PR-numbers, and crashes in the Detroit area.

3.2 Determination of Spatiotemporal Speed Matrix

Literature suggests that the evolution of travel speed in a link can be visualized by a speed contour plot (Park, Gao and Haghani, 2017; Wang, Qi and Jiang, 2018; Wang and Jiang, 2020). To construct a speed contour plot, a road section is segmented into i sections and these sections labeled 1 to i from upstream to downstream. The time period is discretized into T intervals labeled 1 to T. Here $T = 96$, as the time interval is 15-minutes, so the time period is discretized from 1 to 96 for a 24-hour time period. The combination of a specific time period and a particular road segments defines a cell in the speed contour matrix, see Figure 3-3. $S_{t,i}$ is defined as a travel speed in segment i within time interval t. Figure 3-4 demonstrate average speed contour matrix using yearly speed data observation on each day of the week. $\bar{S}_{t,i}$, is the average speed on segment i during time interval t with the standard deviation of $\sigma_{t,i}$. It should be noted that separate yearly average speed profile for each of the seven days of the week was calculated.

Time

Segments / Traffic Flow Direction

	1	…	t	…	96
1	S 1.1	…	S t.1	…	S 96.1
2	S 1.2	…	S t.2	…	S 96.2
⋮	…	…	⋮	…	⋮
i	S 1.i	…	S t.i	…	S 96.i
⋮	…	…	⋮	…	⋮
I	S 1.I	…	S t.I	…	S 96.I

Figure 3- 3: Speed contour matrix. $S_{t,i}$, is the speed on segment i during time interval t

Time

Segments / Traffic Flow Direction

	1	…	t	…	96
1	$\bar{S}_{1.1}, \sigma_{1,i}$	…	$\bar{S}_{t.1}, \sigma_{t,1}$	…	$\bar{S}_{96.1}\ \sigma_{96,1}$
2	$\bar{S}_{1.2}, \sigma_{1,2}$	…	$\bar{S}_{t.2}, \sigma_{t,2}$	…	$\bar{S}_{96.2}, \sigma_{96,2}$
⋮	⋮	…	⋮	…	⋮
i	$\bar{S}_{1.i}, \sigma_{1,i}$	…	$\bar{S}_{t.i}, \sigma_{t,i}$	…	$\bar{S}_{96.i}, \sigma_{96,i}$
⋮	…	…	⋮	…	⋮
I	$\bar{S}_{1.I}, \sigma_{1,I}$	…	$\bar{S}_{t.I}, \sigma_{t,I}$	…	$\bar{S}_{96.I}, \sigma_{96,I}$

Figure 3- 4: Average speed contour matrix. $\bar{S}_{t,i}, \sigma_{t,i}$, is the speed on segment i during time interval t with the standard deviation of $\sigma_{t,i}$

3.3 Determination of Impact Area

The main goal is to compare the yearly speed matrix, where $\bar{S}_{t,i}$ is defined as yearly average travel speed in segment i within time interval t, for each day of the week, with a daily speed contour matrix, $S_{t,i}$ and assign a threshold to determine the spatiotemporal range and whether the daily speed is noticeably smaller than the yearly average travel speed for each day of the week. In the current study, $\bar{S}_{t,i}$ was calculated for each day of the week separately, as the yearly average speed varies for each day of the week. Therefore, a separate speed profile for each of the seven days of the week was calculated. The value of the cut-off deviation has a significant influence on describing the affected zones after a crash. Decreasing the cut-off thresholds reduces the affected zone upstream. Different scenarios have been considered to determine the crash impact area. In the current study, various cut-off deviation, such as 5 mph and 10 mph cut-off-speed, and standard deviation (STD), 1.65STD, 2STD, 3STD has been considered, and secondary crashes were identified based on each scenario.

$$Q_{t,i} = \begin{cases} 1 & if\ S_{t,i} - \bar{S}_{t,i} \leq \sigma_{t,i}\bar{S}_{t,i} \\ 0 & Otherwise \end{cases} \qquad (1)$$

$Q_{t,i}$: Discriminant binary indicator

$\sigma_{t,i}$: Standard deviation

i: Segment number

t: Time step

$S_{t,i}$: Speed on segment i in time step t

$\bar{S}_{t,i}$: Average yearly speed of the day of the week on segment i during time step t

To be specific if $Q_{t,i} = 1$, the matrix cell is considered a congested area. As a result, the discriminant binary indicator $Q_{t,i}$ can be used to indicate whether the vehicle speed in segment i during time interval t is substantially lower than the yearly average speed within each day of the week. If there is an existing crash in cell i, t, the speed reduction is assumed to be due to the crash occurrence. Figure 3-5 shows an example of a speed contour plot for day 107 (04/17/2018) within PR-number 639107 (a segment in I-96 WB) in the Detroit area. Here $T = 96$ and $I = 8$. The time interval is 15-minutes, so the time period is discretized from 1 to 96 for a 24-hour time period. Based on the direction of traffic flow, segment 1 is considered to be upstream of segment 8. Based on the definition, for cells that speed is below the yearly average speed, the color changes from white to red.

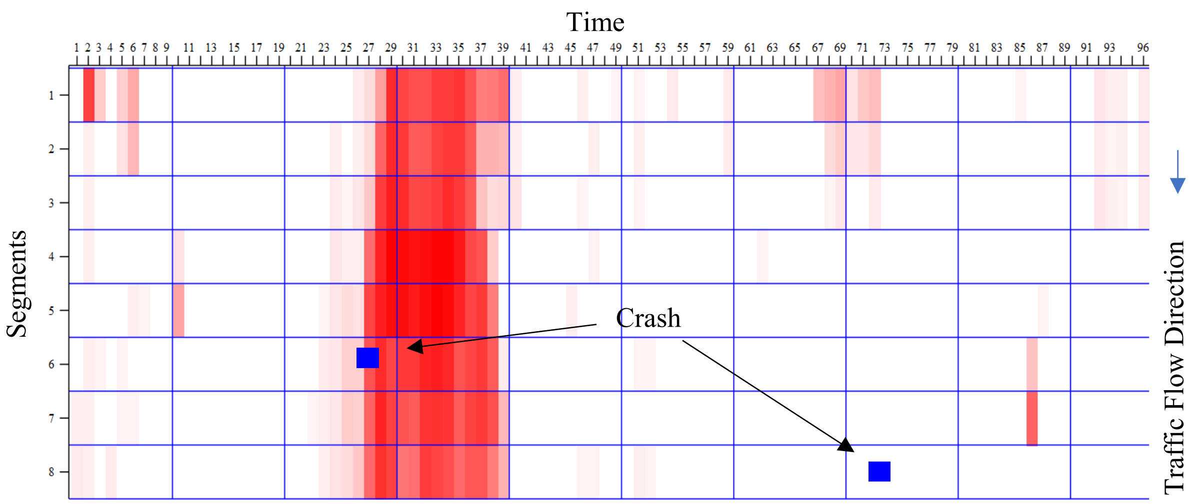


Figure 3- 5: Example of Speed contour plot at day 107 within PR-number 639107

3.4 Secondary Crash Identification Approach

As mentioned in the previous section, each crash is matched to a specific location along the roadway segment based on geographic coordinates using ArcGIS. In addition, roadways consist of different PR-Numbers, and each PR-Number consists of different XD-segments with different mile points. The following steps were performed in order to identify secondary crashes:

- Speed trend plotted based on yearly average speed data in 2018 for each day of the week and each segment.
- Average speed trend at each section, with respect to the day of the week
- Estimating crash impact duration and secondary crash identification

3.4.1 Speed trend plotted based on average speed data on each day of the week and each segment

Recurrent speed trends for each XD-segments were plotted based on average speed data for the year 2018 in each day and each segment. The process will be demonstrated for PR-number 639107 (I-96 WB), see Figure 3-6. This PR-number consists of 8 XD-segments located on I-96 westbound, see Table 3-1.

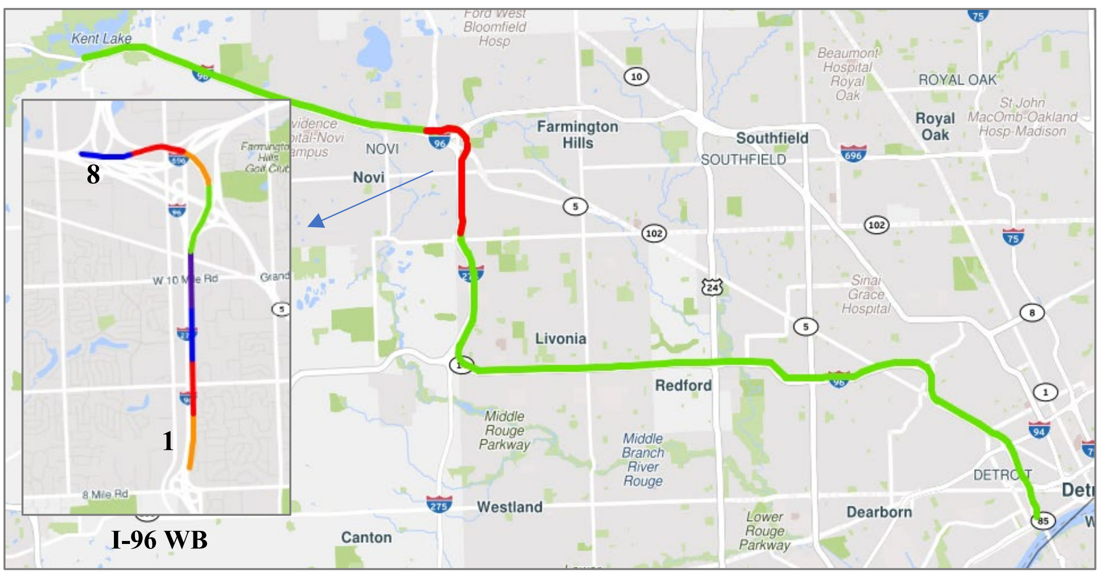
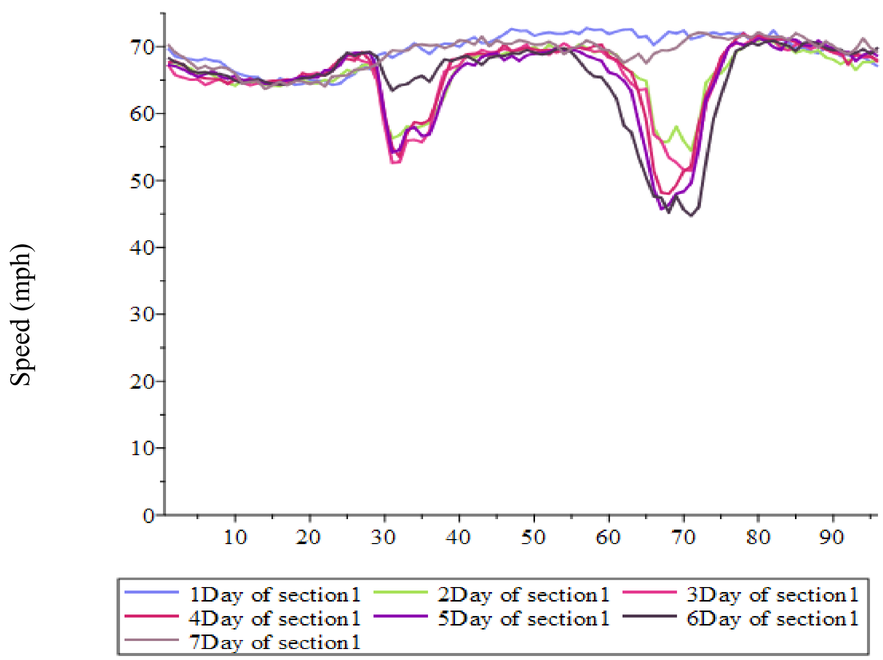


Figure 3- 6: Demonstration of PR-number 639107 (I-96 WB)

Table 3- 1: Segments and mile points within PR-number 639107

PR-Number	XD-segment	Mile point	Segment number
639107	1346346161	0.513	1
639107	1346346122	0.5127	2
639107	1346346133	0.5115	3
639107	1346452489	0.5245	4
639107	1346452504	0.6611	5
639107	1346453321	0.3862	6
639107	1346453331	0.5254	7
639107	1346453345	0.2391	8

Figure 3-7 shows the average 15-minute speed plot for 24 hours in the first segment within PR-number 639107 (I-96 WB). It can be seen from the diagram that the average speed in section one varies between 65 to 70 miles per hour. In addition, the speed drops during the morning peak hour, from 7:30-10:30 am, and evening peak hour, from 3:30-6:30 pm. As expected, such peak hour effects are generally observed on weekdays.



a) Time

Figure 3- 7: a) Speed Trend in section 1 within PR-number 639107 (Sunday=1, Monday=2, Tuesday=3, Wednesday=4, Thursday=5, Friday=6, Saturday=7) b) PR-number 639107 (I-96 WB) with 8 XD-segments

Figure 3-8 demonstrates the average speed for all 8 segments within 639107 PR-Number aggregated over the year. Colors show a different speed range, orange, the highest, and blue, the lowest speed within the segment. The same trend can be observed that the speed significantly drops within morning and evening peak hours. Moreover, speed is considerably lower in the last four segments (segment 5-8). The reason could be the location of those segments that are located at the system interchange.

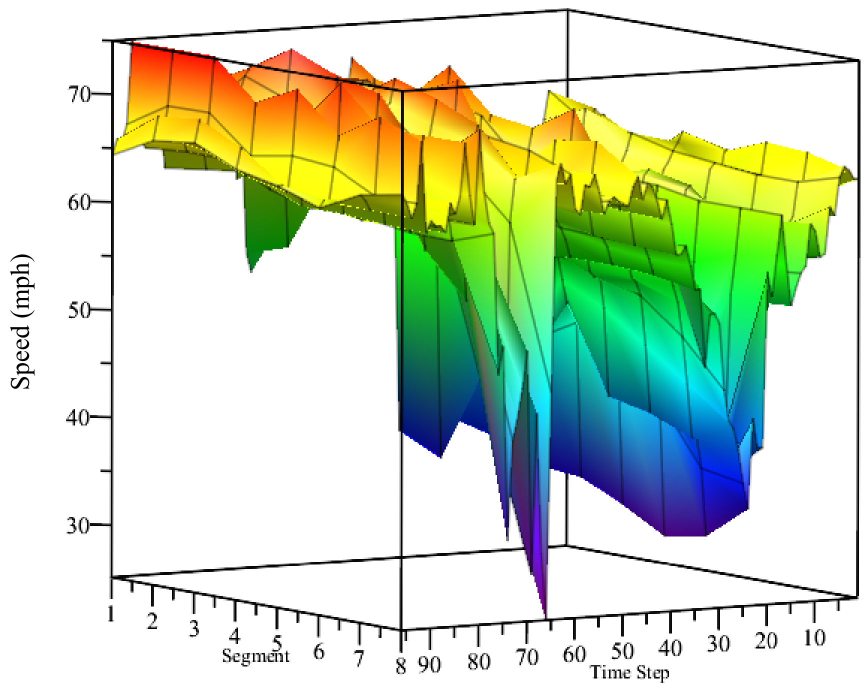


Figure 3- 8: Yearly Speed Average for all 8 segments within PR-number 639107

Figure 3-9 demonstrates the yearly average speed for each time slot during a day (96 Time slot) in various segments. Figure 3-9, shows that average speed varies in different segments, approximately from 75 to 65 mph. Furthermore, it also illustrates that the speed drops in the last four segments. As mentioned, lower speed at the last 4 segments may be induced by their locations, as they are located at a curve. Each line shows the yearly average speed evolution per time slot in all 8 segments. No significant difference in yearly average speed in various time slots during a day was observed.

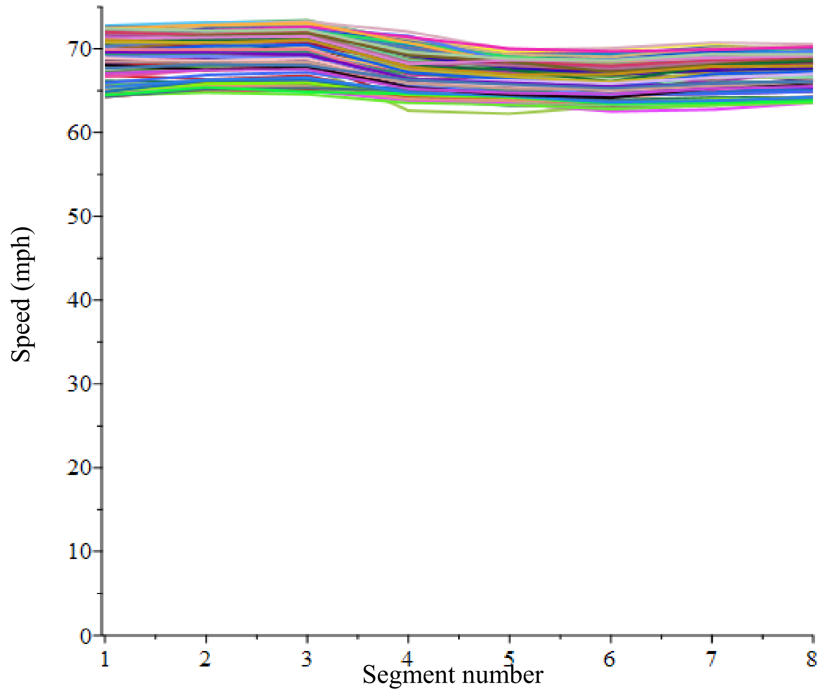


Figure 3- 9: Yearly speed average within each time slot (PR-number 639107)

3.4.2 Average speed trend at each section, with respect to day of the week

The speed data for the same time and location were collected from all days in 2018, and the yearly average speed at each XD-segment, with respect to the day of the week, will be calculated. Subsequently, the result from daily speed compared with the annual average speed. The result will be demonstrated in the heat map, see Figure 3-10.

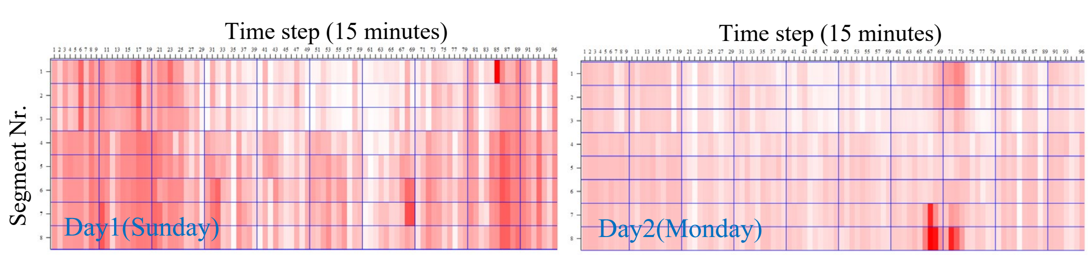


Figure 3- 10: Average speed profile for each day of the week within PR-number-639107

Figure 3- 10 (Cont'd)

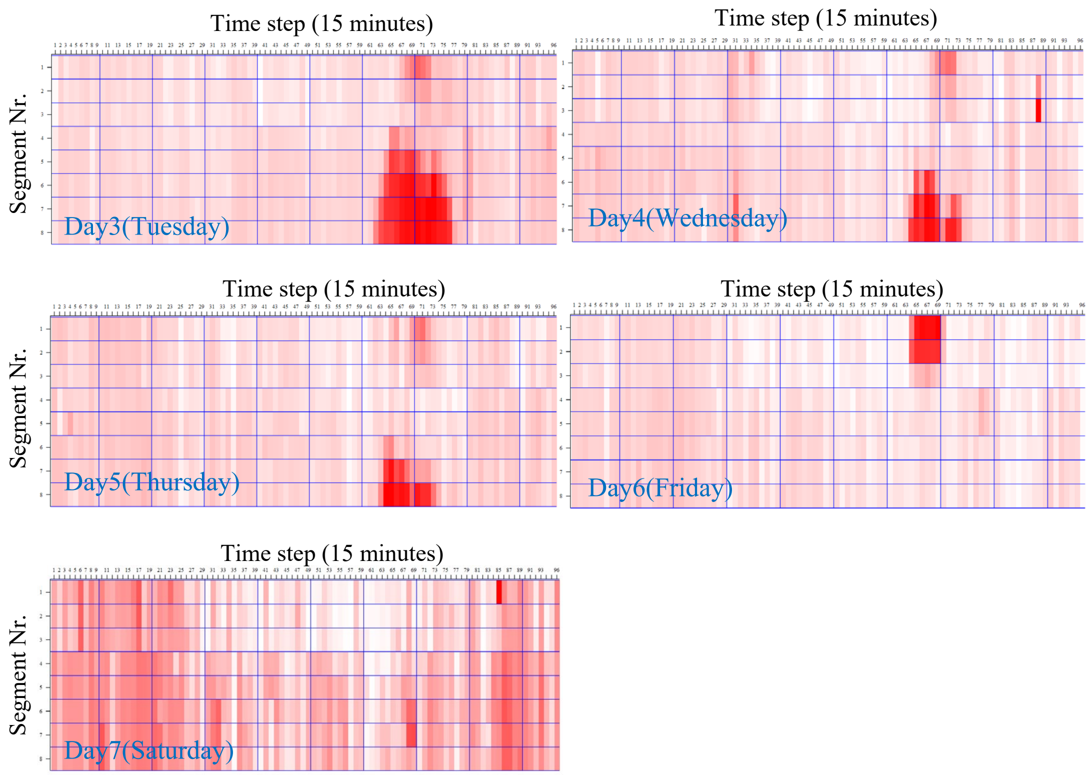


Figure 3- 11 demonstrates the heat map of the relative speed of the traffic in a 24-hour period on October 19th 2018, along I-96 WB in 15-min speed intervals with the yearly average speed in all 8 segments within PR-number 639107.

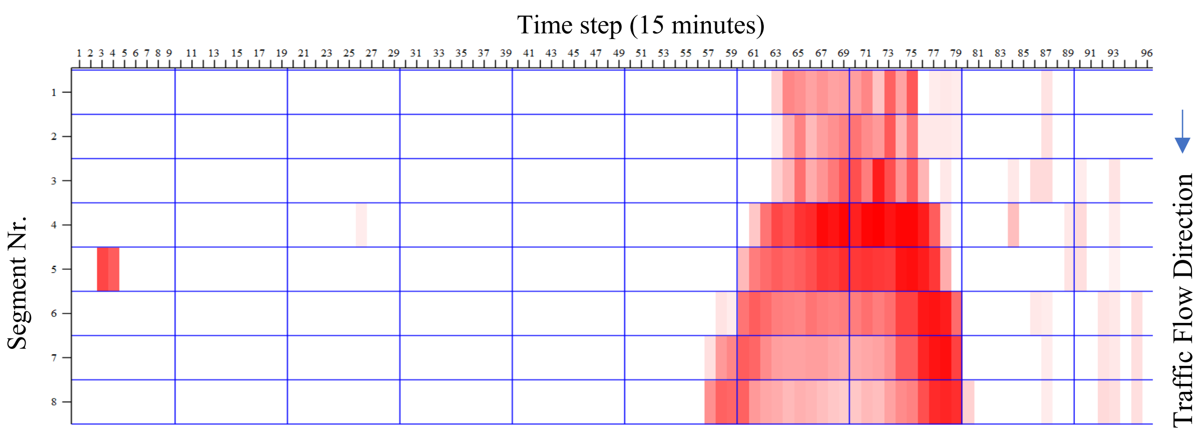


Figure 3- 11: Difference between daily and yearly average speed (October 19th 2018)

Accordingly, one heat map can be generated for each day of a section. If the speed is lower than the yearly average speed, the color changes from white to red. As the difference increases, the color will be intensified. Note that the speed increase has not been considered. Red zones describe the time and location of significant speed drops from the yearly average speed. In this corridor, significant congestion occurred, and the speed drop started in segment eight and continued till segment one, which is upstream of traffic, see Figure 3-12.

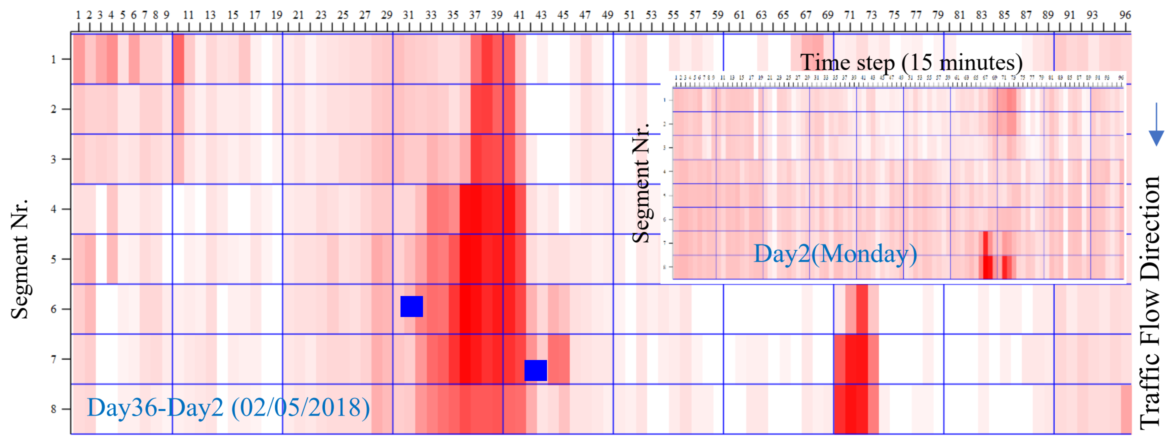


Figure 3- 12: Different average speed profiles for the day Monday, February 5th (02/05/2018) of the week within PR-number-639107

3.4.3 Estimating crash impact duration and secondary crash identification

Next, the crashes within each PR-number are extracted from the interstate crash database and implemented in the heat map. It should be noted that most segments do not experience even two crashes on the same day and thus can be automatically eliminated from the search space. Plotting the distribution of events over the year, Figure 3-13 is created, which describes the density of daily crashes in 2018 within PR-number 639107. The total number of crashes in 2018 in that PR-Number is 138. Using a colored gradient contour of white to red, Figure 3-13, can be used to quickly demonstrate the days with no crashes or one crash. Excluding those days, the dynamic method search space can be quickly constrained to 28 days with more than one crash.

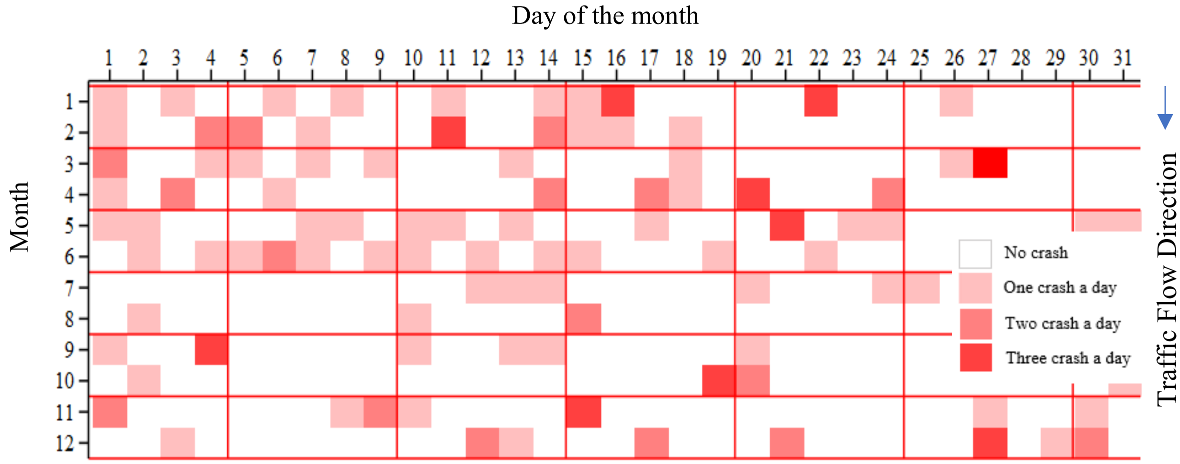


Figure 3- 13: Contour plot of the density of crashes in 2018 within PR-number 639107

Further, the speed data at the time of each crash, $S_{t,i}$ has been compared to the average yearly speed trend within that segment, $\bar{S}_{t,i}$. Speed data at the time of crashes were used to establish a recurrent speed profile of the section under normal traffic conditions. Speed plot trends of crashes plotted to identify the incident impact duration time. The incident impact duration is defined as the duration between the time that incident was detected and the time that speed returned to the normal trend, which is the yearly average speed for each day of the week.

It was hypothesized that when the speeds from the incident reporting times are lower than the defined boundary of average speeds, the speed drop is assumed to be affected by the occurrence of an incident. In this case, the speed profile for each XD-segment is assumed to be affected by the occurrence of an incident when the speed at the incident times is substantially lower than the defined average speed. The speed drop in each road segment was compared spatially and temporally with the average annual speed in that segment to identify secondary crashes. In the case that the speed drops near the incident location, for every crash, the time and the distance in the upstream direction of the traffic are recorded till the speed gets back to the annual average speed.

Once the incident impact area for all crashes is identified, the model will search for other incidents occurring within the affected spatiotemporal window. Any crash within the impact area and upstream of a primary crash will be categorized as a secondary crash.

$$C_i = SC\left(C_j\right) \;\; if \begin{cases} Sg_i < Sg_j \;\; upstream\ condition \\ t_i > t_j\ time\ condition \\ t_i \in [temporal\ impact\ area\ of\ C_j] \\ Sg_i \;\in [spatial\ impact\ area\ of\ C_j] \end{cases} \;, \; S_{t,i} < \bar{S}_{t,i}^{dow} \quad (2)$$

C_i: Crash i

C_j: Crash j

SC: Secondary crash

Sg: Segment of the crash occurrence

t: Time of the crash occurrence

$S_{t,i}$: Speed at the time of each crash

$\bar{S}_{t,i}^{dow}$: Average yearly speed within on segment i in time step t

If multiple crashes were detected within the affected spatiotemporal window, all of them would be categorized as secondary crashes. In the example depicted in Figure 3-14 a showcase of crashes that occurred on Friday, October 19th, 2018, along I-96 WB is provided. On this particular day, three crashes occurred along the study corridor, resulting in significant congestion, the average speed, $\bar{S}_{t,i}$, dropped below the recurring speeds along this corridor, $S_{t,i}$. Two of these three crashes were considered as secondary crashes.

Table 3- 2: Crashes on Friday, October 19th, 2018, along I-96 WB

Crash ID	Date	Day of the week	Time	PR-Nr.
1514269	10/19/2018	6	15:35	639107

Table 3- 2 (Cont'd)

1514280	10/19/2018	6	16:30	639107
1509026	10/19/2018	6	16:50	639107

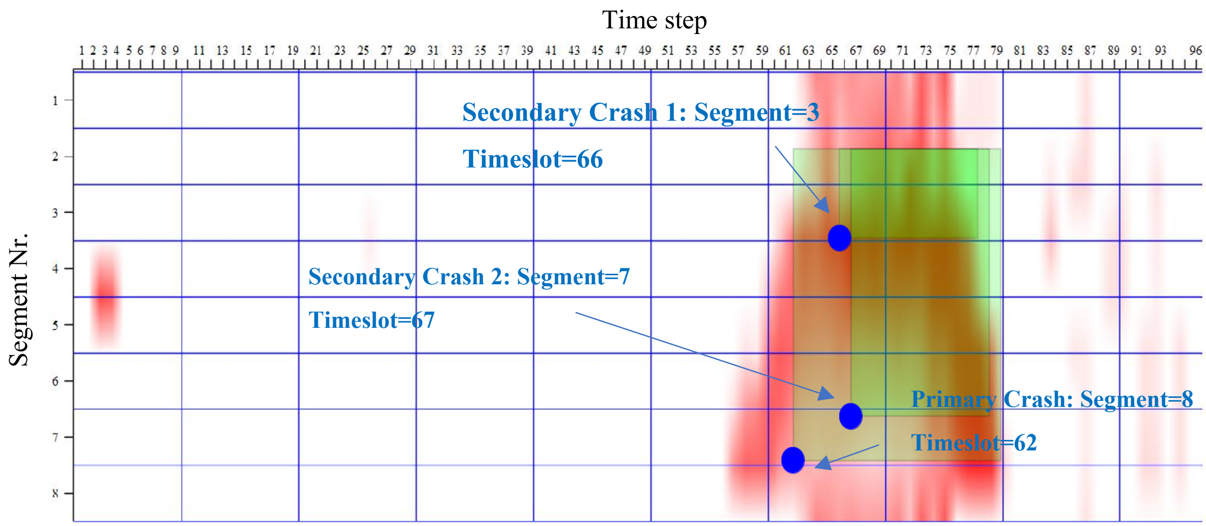


Figure 3- 14: Detection of secondary crashes using speed data

Crash 1 occurred at 15:35 pm (time slot 62) on segment 8 and affected eight segments in the upstream direction (from 8 to 1).

$$Crash_1: [\, t_1 = 62, i_1 = 8] \quad (3)$$
$$S_{t,i} < \bar{S}_{t,i}^{dow}$$
$$\forall\, i = [1 \ldots ,8]$$
$$\forall\, t = [62 \ldots ,80]$$

$Crash_1$: First crash with crash Id: 1514269

i: Segment

t: Time step

$\bar{S}_{t,i}^{dow}$: Average yearly speed on segment i in time step t

$S_{t,i}$: Speed on segment i in time step t

The speed drop continues from time slot 62 to 80. It is worth noting that crash occurrence is considered to be the source of the congestion and speed drop, however, it may be possible that the speed reduction is not due to only the crash occurrence. From the Figure 3-14, it can be clearly observed that congestions and queue formations occur after the primary crash. However, less information has been obtainable in the Figure 3-14 about whether the queue formations resulted from recurrent congestion or another crash in the previous road segment. In order to eliminate the effects of recurrent congestions, the spatial and temporal influencing range of the prior crash should be determined.

As a result of congestion caused by the primary crash and significant speed reduction, another crash occurred at 16:30 (time step 66) on segment three. This crash occurred 55 minutes later and upstream of the primary crash on segment 3. The crash resulted in a drop-in speed from time slot 66 to 78 and from segment 3 to 1.

$$Crash_2: [\, t_2 = 66, i_2 = 3] \quad (4)$$
$$S_{t,i} < \bar{S}_{t,i}^{dow}$$
$$\forall\, i = [1 \ldots, 3]$$
$$\forall\, t = [66 \ldots, 78]$$

$Crash_2$: Second crash with crash Id: 1514280

i: Segment

t: Time step

$\bar{S}_{t,i}^{dow}$: Average speed on segment i in time step t

$S_{t,i}$: Speed on segment i in time step t

Following those crashes, another crash occurred at 16:50 (time slot 67) on segment 7. The speed drop continues from time slot 67 to 79 and from segment 7 to 1.

$$Crash_3: [\ t_3 = 62, i_3 = 8]\ \ (5)$$
$$S_{t,i} < \bar{S}_{t,i}^{dow}$$
$$\forall\ i = [1 \dots, 7]\ ,$$
$$\forall\ t = [67 \dots, 79]$$

$Crash_2$: Third crash with crash Id: 1509026

i: Segment

t: Time step

$\bar{S}_{t,i}^{dow}$: Average speed on segment i in time step t

$S_{t,i}$: Speed on segment i in time step t

In the showcase, the first crash is considered a primary crash, and other crashes are considered as secondary crashes as there are located in the primary crash impact area. The same analysis was done for days with multiple crashes. In some cases, a secondary crash could be a primary crash and leads to additional crashes.

3.5 Results and Discussion

3.5.1 Secondary Crashes Identified by Manual Method Within Detroit Area

Secondary crashes were identified in the Detroit area using the manual method. In this approach, police crash reports were used to identify secondary crashes. This method was used to evaluate the sensitivity of spatiotemporal thresholds and also to determine the extent of under or overestimation of secondary crashes when compared with the dynamic method. Each crash report includes detailed information about the crash, such as date, time, location, and a crash narrative and crash code. In the manual approach, narratives from the police crash reports were checked manually. In total, there were 13,392 crash reports in the Detroit region, and the information from these reports was converted to a spreadsheet format for review.

Based on the crash code, 859 crashes were identified as being due, at least in part, to a prior crash under the contributing circumstance field. For all those crashes, crash narratives have been reviewed manually, and the result showed that about 82 percent, or 707 of them, were associated with a previous crash and secondary in nature. The rest of the crash reports were assumed not to be related to prior crashes. For such crashes, other reasons other than prior crashes were mentioned in the narrative as the cause of a crash occurrence, which means crash code and narratives were not correlated. As mentioned in the previous chapter, the manual approach is used for the rest of the crash reports. Due to this approach additional 122 secondary crashes were identified. The result is demonstrated in Table 3-3. The result shows that almost 6.2 percent of the crashes were considered secondary crashes within Detroit interstate mainline system.

Table 3- 3: Result for reviewing the crash reports with secondary related crash code

Contributing circumstances (crash code)	**Total number of crashes**	**Confirmed number of secondary crashes**	**Other (non-secondary) crashes**
Backup - Other Incident	455	382	73
Prior Crash	404	325	79
Other (Identified from narrative)	12,533	122	12,411
Total	13,392	829	12,563

3.5.2 Secondary Crashes Identified Using the Dynamic Method in Detroit Area

The proposed approach used Detroit crash data (13,392 crashes) from MTCF database and real-time speed data from RITIS. Various scenarios have been considered as cut-off deviations, such as 5 mph and 10 mph cut-off-speed and STD, 1.65STD, 2STD, 3STD. Secondary crashes have been identified based on different scenarios. The result shows that the identified secondary crashes accounted for 3 to 10 percent of the Detroit crashes based on different scenarios, see Table 3-4. It can be observed that the scenario with 5-mph cut-off deviation has the highest, and 3STD has the lowest number of identified secondary crashes.

Table 3- 4: Secondary crash results from the dynamic approach for various cut-off scenarios

Dynamic Method Scenario	**Nr of Secondary Crash**	**Percentage of Secondary Crash**
5 mph cut off	1301	9.72
10 mph cut off	828	6.18
(Standard Deviation) STD	1102	8.23
1.65STD	762	5.69
2STD	623	4.65
	414	3.09
Total number of crashes	13,392	

Further, the result from different scenarios in the dynamic method was compared with the result from the manual approach, see Table 3-5. From those crashes classified in the dynamic method as secondary crashes, some have been identified in the manual approach as well as the secondary crash. Here crashes identified in the manual approach are confirmed as actual secondary crashes. The percentages of actual secondary crashes have been calculated from the ratio of the number of secondary crashes identified in the manual method to those determined by the dynamic method considering various scenarios. The percentages of the actual secondary crash identified in the dynamic method are the highest in 3STD scenario by about 37 percent and the lowest in 5-mph scenario by 20 percent, respectively.

Table 3- 5: Comparison of secondary crashes identified by dynamic and manual method

Dynamic method Scenarios	Nr. of Secondary Crashes identified in dynamic method	Nr. of confirmed secondary crashes (manual method)	Percentage of confirmed secondary crashes
5 mph cut off	1301	259	19.9
10 mph cut off	828	207	25.0
(Standard deviation) STD	1102	249	22.6
1.65STD	762	209	27.4
2STD	623	195	31.3
3STD	414	155	37.4

- 829 (≅ 6.2%) total number of actual secondary crashes in the Detroit area (based on manual method)

3.5.3 Static Sizing: Spatiotemporal Window in Detroit area

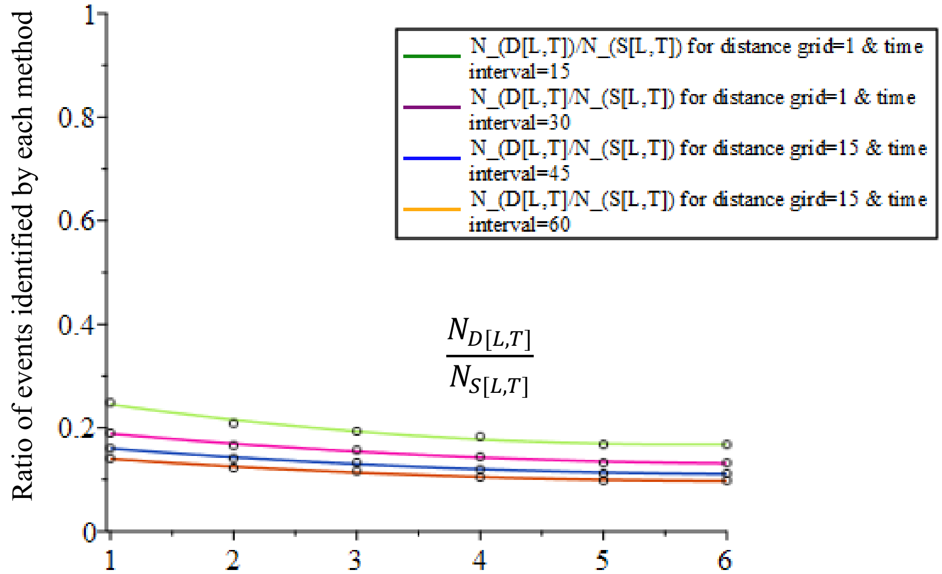
In order to compare the result from dynamic approach to the static approach similar process employed see previous chapter (section 2.2). In this section the number of secondary crashes that has been identified in dynamic approach within each spatiotemporal window determined. Each crash is associated with interstate road number, location on the road, date, and time. Using linear referencing in ArcGIS, the exact locations (mile points) for each crash along the interstate road were determined. In the first step, consecutive crashes on each road segment were identified based on date and time. From the difference between corresponding mile points, the distance between two consecutive crashes was calculated. After determining the spatiotemporal thresholds between consecutive crash events within the 2018 interstate crash dataset in the Detroit area, different time and distance intervals (the distance interval varies from 1 to 6 miles and the time interval from 0 to 300 minutes) were used to define different sizes of spatiotemporal windows based on the result from the different scenarios in the dynamic method. This approach has been explained in detail in the previous chapter (section 2.2).

The same analysis has been done for the dynamic process after determining the spatiotemporal thresholds between consecutive crash events. In order to illustrate the loss of accuracy by increasing sensitivity, the ratio of all verified secondary crashes in the dynamic approach to the total predicted events in a static approach (spatiotemporal window) is demonstrated by plotting $N_{D[L,T]}/\,N_{S[L,T]}$ for different sizing of spatiotemporal windows, see Equation 6,

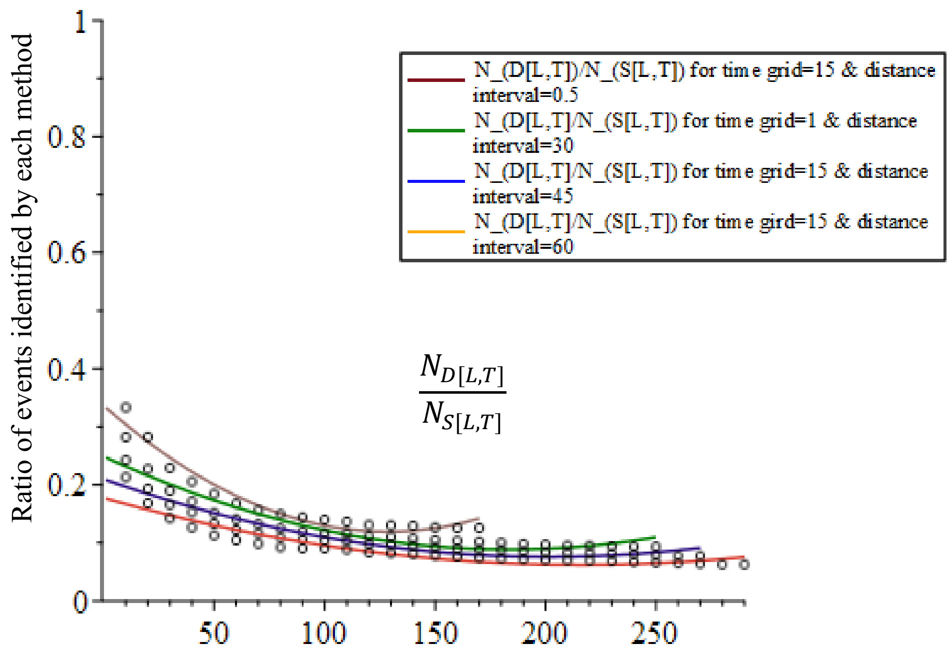
$$\boldsymbol{N_{D[L,T]}} = \boldsymbol{\alpha N_{S[L,T]}} \quad \text{where} \quad \boldsymbol{\alpha} \cong [0.16 - 0.22] \quad (6)$$

- $N_{D[L,T]}$ = Number of secondary crash events identified in a dynamic method
- $N_{S[L,T]}$ = Number of crashes that exists within a specific spatiotemporal window from the first crash
- α: Convergence limit (Sensitivity)

The convergence limit α was observed on windows with different spatial and temporal sizes, see Figure 3-15.



a) Radius (Mile)



b) Time (minute)

Figure 3- 15: The ratio of actual confirmed events in the dynamic method to the total predicted events in the static approach within a gap size of 1 mile and various time intervals. b) The ratio of actual confirmed events in the dynamic method to the total predicted events in the static approach within a gap size of 15 minutes and various distance gaps

In other words, it can be seen that as the window grows, the accuracy decreases, and limit α can be considered as sensitivity which is the probability of correctly identifying a secondary crash. Also, the specificity of dynamic approach which is the probability of correctly identifying a non-secondary crash calculated, see Table 3-6. Here the crash data within each window compared with the crash data within the largest window (spatiotemporal window of 6 mile and 300 minutes).

Table 3- 6: Secondary crash distribution for interstate roads in Detroit area based on static and dynamic approach

Distance grid (Mile)	**Time gap (Min)**	**Number of crashes in spatiotemporal window N_S[L, T]**	**Number of verified secondary crashes in dynamic approach within spatiotemporal window N_D[L,T]**	**Specificity (300min, 6mile)**	**Sensitivity**
1	15	204	77	86%	38%
	30	315	109	87%	35%
	60	482	151	88%	31%
	300	1076	250	89%	23%
3	15	299	95	87%	32%
	30	496	139	87%	28%
	60	814	199	88%	24%
	300	2171	377	90%	17%
6	15	394	108	87%	27%
	30	669	165	87%	25%
	60	1119	235	88%	21%
	300	3170	480	unknown	15%

The temporal and spatial characteristics of secondary crashes within each static window can be observed in Figure 3-16. Temporally, approximately 75 percent of the secondary crashes were found to occur within 100 minutes time gap from the previous crash. Spatially, about 80 percent of the secondary crashes were found to occur within 2.5-mile distance gap from the previous crash.

Generally, about 68 percent of secondary crashes occurred within 75 minutes of the time gap of the previous crash and within 1.5 miles upstream of the previous crash. In other words, about 32% of secondary crashes occurred beyond the most commonly used 1.75 miles and 75 spatiotemporal thresholds. These statistics confirm that the proposed dynamic approach identified more secondary crashes than the traditional manual method and less than the static method, which means that the static method overestimates the number of secondary crashes.

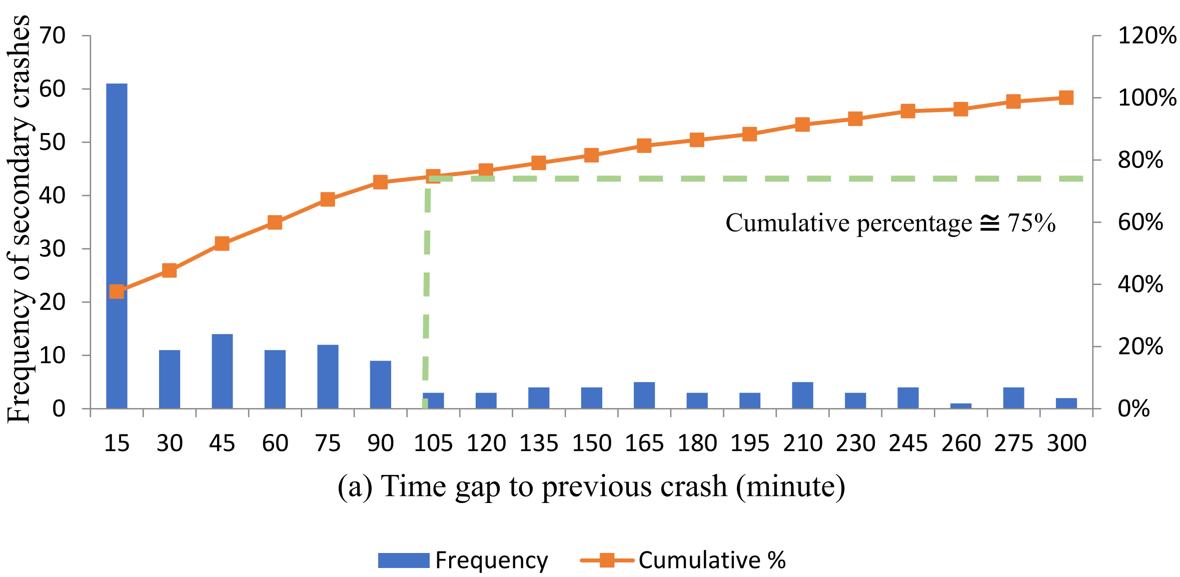
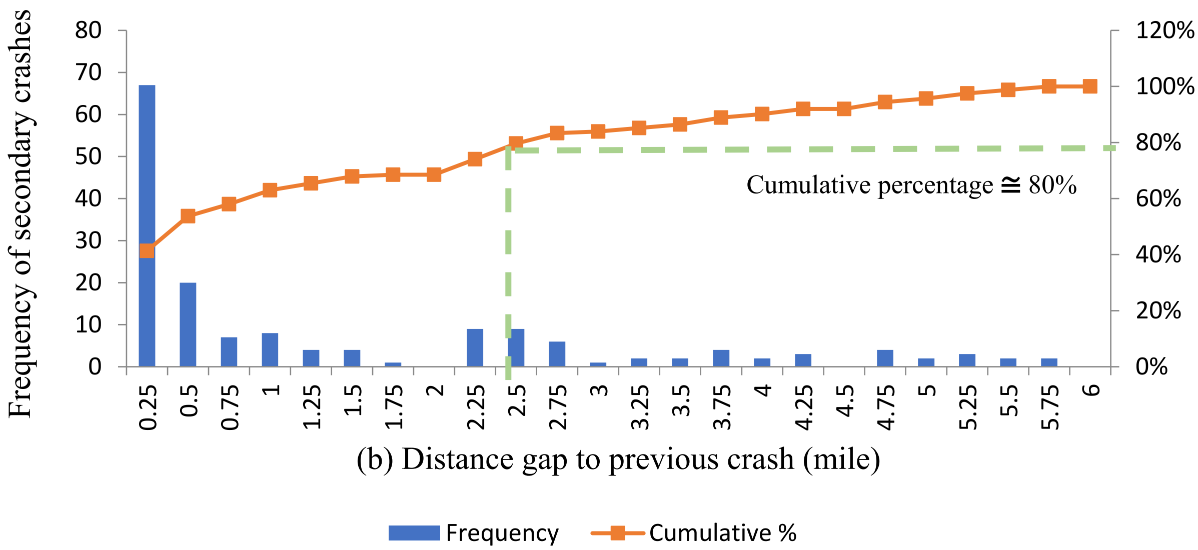


Figure 3- 16: Spatiotemporal distribution of secondary crashes in relation to previous crash (a) Temporal distribution (b) Spatial distribution

Figure 3- 16 (Cont'd)



3.6 Discussion and Conclusions

Crashes are a major source of delays, system unreliability, and inefficiency on freeways. Congestion caused by a crash may increase the potential of subsequent vehicles to the risk of secondary crashes. Such crashes have been identified as a major problem in freeways that frequently affect both traffic operations and safety. Therefore, transportation agencies have taken various measures to minimize and mitigate the potential for and impacts of such crashes. Identifying secondary crashes is not a straightforward procedure as the definition is subjective. Past studies have proposed manual, static, and dynamic approaches to identify secondary crashes. Static methods have defined secondary crashes based on a fixed spatial and temporal threshold. In this approach, a fixed spatiotemporal window is assumed around the primary crash, which often overestimates the secondary crash by considering all the nearby events as the secondary crash. Furthermore, the static approach considers the same window for all types of primary crashes regardless of the upstream traffic flow, density and speed. The dynamic approach identifies a dynamic spatiotemporal impact area for each primary crash, in contrast to the static method, which considers a predefined threshold for the primary crash.

This research proposes a secondary crash identification method on freeways by tracking the spatiotemporal evolution of traffic flow. In this work, by leveraging a huge database of all events in Michigan Detroit interstate roads in 2018, a secondary crash identification approach from the integration of speed contour plot and the spatiotemporal evolution of primary crash impact area was proposed. Real-time travel speed data for every 15 minutes time interval was downloaded from RITIS and used in the method. In order to identify the crash impact area, the daily speed has been compared with the yearly average yearly speed within each day of the week. For each primary crash, a spatiotemporal speed matrix and corresponding speed contour plot within every segment are constructed. The area is considered congested when the daily speed is lower than the average speed. If there is an existing crash in the section, the speed reduction is assumed due to the crash occurrence. Further, if another crash occurs within the primary crash impact area, it is considered a secondary crash. It has been demonstrated that the static method consistently overestimates and with the increase in spatiotemporal window seizing, the specificity fades as the sensitivity increases.

In addition, the number of secondary crashes identified by the dynamic method is highly dependent on the cut of speed. Based on the dynamic method, the total number of secondary crashes identified in the Detroit area varies from 3 to 10 percent, considering different scenarios. Different scenarios have been considered as cut-off deviations such as 5 mph and 10 mph cut off-speed as well as STD, 1.65STD, 2STD, 3STD. So, the 5-mph cut-off point scenario was considered to have the least sensitivity and 3STD the highest sensitivity consecutively.

Logistic regression and negative binomial model were applied in order to identify factors that affect secondary crashes is the first step toward preventing the occurrence of secondary crashes. The result from the logistic regression model suggests that weather conditions, posted speed limit, and crash severity, which involves minor injury, are among the key variables that affect secondary crash occurrence. The result from the negative binomial model suggests that annual average daily traffic (AADT), median with a concrete barrier, and a number of lanes and right shoulder width are among the key variables that affect secondary crash occurrence. This result is expected to provide useful information in developing policies and strategies to prevent the occurrence of secondary crashes. Moreover, the developed model can also be incorporated in advanced traffic control systems on freeways to prevent the occurrence of secondary crashes.

With the comparison of the proposed approach to static and dynamic methods, it is expected that the proposed approach will lead to a reduction in the misidentification of secondary crashes. In addition, results may help to perform necessary strategies to mitigate secondary crashes, including improved traffic management policies and the implementation of advanced intelligent transportation warning systems. While this study only examined 2018 data on interstate roads in the Detroit area, it may not be a comprehensive representation of the whole state. Furthermore, secondary crashes caused by other non-crash incidents and also the effect of crashes in the opposite traffic direction deserve more investigation.

CHAPTER 4. MODELING AND PREDICTING SECONDARY CRASH RISK

4.1 Logistic Regression Analysis

Existing studies have used several statistical models to analyze the risk of secondary crash occurrence. Among these studies, a number of studies e.g. (Karlaftis *et al.*, 1999; Zhan *et al.*, 2008) have adopted logistic regression models to identify those characteristics that distinguish secondary crashes from primary crashes. The results of such analyses can help to discern those scenarios where secondary crashes are most likely to occur, providing agencies with important insights to help with incident response and management activities.

In the logistic regression framework, each crash can be characterized into one of two dichotomous outcomes, either the crash was secondary in nature (i.e., due to the occurrence of a previous, downstream crash) or it was not. The general form of this relationship is as follows,

$$Y_i = logit(P_i) = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} \quad (7)$$

Where the response variable Y_i is the logistic transformation of the probability of a crash being secondary in nature (P_i). The variables X_{i1} to X_{ik} are factors assumed to be related to the occurrence of a secondary crash, β_0 is an intercept, and β_1 to β_k are estimated regression parameters for each independent variable. These regression parameters are positive for those variables that are positively correlated with secondary crashes (i.e., secondary crashes are more likely as these variables are increased). Negative parameters are reflective of those variables that are underrepresented (i.e., less likely) among secondary crashes.

4.1.1 Data Description and Summary

The initial dataset included a total of 26,679 crashes that occurred on mainline interstates in Michigan in the calendar year 2018. These data have been filtered out to consider only those crashes that occurred on roads with between two and five lanes and with speed limits from 55 mph to 75 mph. This reduced the final data set to 25,366 crashes. Table 4-1 shows the descriptive statistics corresponding to these data.

Table 4 - 1: Descriptive statistics for analysis dataset

Variables	**Mean**	**Standard Deviation**
Interstate highway where the crash occurred		
I-69 (1 if yes; 0 if no)	0.066	0.249
I-75 (1 if yes; 0 if no)	0.259	0.438
I-94 (1 if yes; 0 if no)	0.283	0.450
I-96 (1 if yes; 0 if no)	0.195	0.396
I-196 (1 if yes; 0 if no)	0.056	0.229
I-275 (1 if yes; 0 if no)	0.024	0.153
I-296 (1 if yes; 0 if no)	0.009	0.096
I-475 (1 if yes; 0 if no)	0.011	0.105
I-496 (1 if yes; 0 if no)	0.015	0.124
I-194, I-375, I-675 (1 if yes; 0 if no)	0.007	0.085
I-696 (1 if yes; 0 if no)	0.074	0.262
Emergency medical services involved (1 if yes; 0 if no)	0.007	0.083
Total number of lanes at the site of the crash		
Two (1 if yes; 0 if no)	0.322	0.467
Three (1 if yes; 0 if no)	0.409	0.492
Four (1 if yes; 0 if no)	0.234	0.423
Five (1 if yes; 0 if no)	0.035	0.184
Urban area type		
Rural (1 if yes; 0 if no)	0.182	0.386
Small Urban and Small Urbanized (1 if yes; 0 if no)	0.117	0.322
Large Urbanized (1 if yes; 0 if no)	0.701	0.458
Time at which crash occurred		
Morning Peak hour (6:00 - 9:00) (1 if yes; 0 if no)	0.186	0.389
Evening Peak hour (15:00 - 19:00) (1 if yes; 0 if no)	0.275	0.446

Table 4 -1 (Cont'd)

Off-Peak hour (1 if yes; 0 if no)	0.526	0.499
Day of week on which crash occurred		
Weekdays (1 if yes; 0 if no)	0.776	0.417
Weekend (1 if yes; 0 if no)	0.220	0.420
Number of units involved in the crash		
One (1 if yes; 0 if no)	0.421	0.494
Two (1 if yes; 0 if no)	0.490	0.500
More than two (1 if yes; 0 if no)	0.089	0.285
Relationship of crash to the roadway		
On the Road (1 if yes; 0 if no)	0.828	0.377
Median (1 if yes; 0 if no)	0.045	0.207
Shoulder (1 if yes; 0 if no)	0.064	0.245
Outside of Shoulder/Curb (1 if yes; 0 if no)	0.057	0.232
Gore/On-Street Parking/Off Roadway/Sidewalk/Bicycle Lane (1 if yes; 0 if no)	0.006	0.077
Weather Conditions		
Clear and cloudy (1 if yes; 0 if no)	0.699	0.459
Rain (1 if yes; 0 if no)	0.124	0.330
Snow (1 if yes; 0 if no)	0.157	0.364
Other (Fog, Severe Crosswinds, etc.) (1 if yes; 0 if no)	0.020	0.140
Crash Severity		
Fatal injury (1 if yes; 0 if no)	0.003	0.056
Suspected Serious Injury (1 if yes; 0 if no)	0.014	0.117
Suspected Minor Injury (1 if yes; 0 if no)	0.045	0.207
Possible injury (1 if yes; 0 if no)	0.125	0.331
No injury (1 if yes; 0 if no)	0.813	0.390
Posted Speed Limit		
55 mph (1 if yes; 0 if no)	0.122	0.327
60-65 mph (1 if yes; 0 if no)	0.034	0.181
70 mph (1 if yes; 0 if no)	0.777	0.416
75 mph (1 if yes; 0 if no)	0.067	0.250

Crashes occurring on a total of thirteen interstate highways were included in the sample. Of these, the majority of crashes (54.2%) occurred on I-75 and I-94. Crashes were least frequent on bypass routes, such as I-194 and I-375. The Michigan UD-10 police crash report form classifies roads into four area type categories: 1) Rural (population is less than 5,000); 2) Small Urban (urban cluster population is 5,000 - 49,999); 3) Small Urbanized (population is 50,000 - 199,999); and 4) Large Urbanized (population is 200,000 or more). Approximately 70 percent of crashes occurred in a large urbanized area. Approximately 78 percent of the crashes occurred on weekdays, and almost 70 percent happened during clear or cloudy weather conditions.

From all crashes, only one percent involved more than two vehicles. In this study, based on occurrence time, crashes were categorized into two groups, namely, those that occurred during peak hours (06:00 to 10:00 and 15:00 to 19:00) and those that occurred during off-peak hours (9:01 am - 14:59 pm; 19:01 pm – 5:59 am). The information about peak hours has been determined based on the MDOT Freeway Congestion & Reliability Report in 2019. Based on the data, almost 53 percent of crashes are happening during off-peak hours. Overall, the summary statistics showed that almost 77 percent of crashes occurred on roads with a 70-mph posted speed limit.

4.1.2 Analysis and Result of Logistic Regression Model

Estimation results for the logistic regression model for secondary crashes are shown in Table 4-2. All of the factors listed in the previous section were included in the initial model. The model was then tested to determine the significant variables. All of the identified variables are significant at the 0.05 level.

Table 4 - 2 : Logistic regression model results for secondary crash likelihood

Variables	**Estimate**	**SE**	**P-value**
Intercept	-3.850	0.141	< 0.001
I-94 **(baseline)**			
I-69	-0.047	0.120	0.697
I-75	-0.151	0.075	0.045
I-96	-0.183	0.076	0.017
I-196	-0.191	0.120	0.111
I-275	-0.491	0.193	0.011
I-296	0.057	0.230	0.806

Table 4-2 (Cont'd)

I-475	-0.161	0.271	0.551
I-496	0.137	0.177	0.438
I-194, I-375, I-675	-0.682	0.402	0.090
I-696	-0.026	0.111	0.813
Urban areas – Rural **(baseline)**			
Urban areas - Small Urban and Small Urbanized	0.023	0.102	0.823
Urban areas - Large Urbanized	-0.102	0.088	0.246
Emergency medical services involved	1.122	0.201	< 0.001
Off-Peak hour **(baseline)**			
Morning Peak hour	-0.098	0.067	0.142
Evening Peak hour	-0.459	0.063	< 0.001
Weekend **(baseline)**			
Weekdays	0.026	0.065	0.690
Number of units - 1 **(baseline)**			
Number of units - 2	1.677	0.081	< 0.001
Number of units - more than 2	2.206	0.100	< 0.001
Crash Severity - No injury **(baseline)**			
Crash Severity - Fatal injury	0.739	0.306	0.016
Crash Severity - Suspected Serious Injury	0.171	0.192	0.373
Crash Severity - Suspected Minor Injury	0.042	0.121	0.726
Crash Severity - Possible injury	0.008	0.074	0.911
Weather Condition - Clear and cloudy **(baseline)**			
Weather Condition - Rain	0.227	0.079	0.004
Weather Condition - Snow	0.738	0.066	< 0.001
Weather Condition - other	0.103	0.203	0.612
Number of lanes- 2 **(baseline)**			
Number of lanes- 3	-0.365	0.072	< 0.001
Number of lanes- 4	-0.465	0.082	< 0.001
Number of lanes- 5	-0.792	0.082	< 0.001
Relationship of the crash to the roadway- On the Road **(baseline)**			
Relationship of the crash to the roadway - Median	0.123	0.166	0.459
Relationship of the crash to the roadway - Shoulder	0.314	0.118	0.008
Relationship of the crash to the roadway - Outside of Shoulder/Curb	-0.131	0.169	0.440
Relationship of the crash to the roadway - Other	-0.527	0.464	0.256

Table 4-2 (Cont'd)

Speed Limit - 55 mph **(baseline)**			
Speed Limit - 60_65 mph	0.679	0.149	< 0.001

The result shows that the probability of secondary crash occurrence is lower in peak hours in comparison to non-peak hours. In addition, the likelihood of secondary crash occurrence is higher within the morning peak hour (6:00 AM to 9:00 AM) than the evening peak hour (15:00 PM to 19:00 PM). This result is consistent with the findings of the study by Vlahogianni et al. (2010), which found that during peak periods, crash influence is most likely increasing both temporally and especially in upstream traffic direction. Moreover, by expanding the crash duration, an extended response and clearance time may induce a significant likelihood of a secondary crash (Vlahogianni *et al.*, 2010). However, a few other studies found peak hours as an insignificant factor in increasing the possibility of secondary crash occurrence (Khattak, Wang and Zhang, 2009; Xu *et al.*, 2016; Sarker *et al.*, 2017). One reason could be the speed drop in peak hour.

Based on the result from Table 4-2 if all other factors are fixed, secondary crashes are more likely to occur when there are two and more than two vehicle units involved in the crash. Previous studies show mixed findings. This result is consistent with the findings from the study by Zhan et al. (2008) and Kopitch and Saphores (2011), where the number of vehicles is a significant factor in the likelihood of secondary crashes (Zhan *et al.*, 2008; Kopitch and Saphores, 2011). Khattak et al. (2009) proposed three binary probit models to examine the interdependence between primary crash duration and secondary crash occurrence. Their findings showed that primary crash duration, AADT, and the number of involved vehicles positively affect the likelihood of secondary crashes (Khattak, Wang and Zhang, 2009). However, few other studies do not support this finding (Vlahogianni, Karlaftis and Orfanou, 2012; Park and Haghani, 2016a; Park, Gao and Haghani, 2017). The result shows that secondary crashes are more associated with crash injuries. Also, the likelihood of secondary crash occurrence is higher when primary crash results in fatality. One of the possible reasons could be that a fatal crash is likely to lead higher effect on traffic flow on freeways, leading to a higher likelihood of multiple secondary crashes.

Based on the result, secondary crash likelihood is higher during the week and decreased on weekends. This result is inconsistent with the finding of the previous study (Xu *et al.*, 2016). Also, the likelihood of secondary crashes increases within rainy and snowy weather conditions, which is consistent with the previous study (Khattak, Wang and Zhang, 2011; Mishra *et al.*, 2016; Wang, Liu, *et al.*, 2016). In particular, the possibility of the secondary crash occurrence is higher in snowy weather. One reason could be that bad weather reduces visibility and friction between pavement and tires. Therefore, drivers have less time and space to take crash avoidance maneuvers.

The chance of secondary crash occurrence is the highest on the roads with two lanes. The result shows that the probability of secondary crash occurrence decreases as the number of lanes increases. One possible reason is that with increasing the number of lanes vehicles could prevent secondary crashes by changing the lanes. This result is consistent with the findings of the study by Sarker et al. (2017) and Zhan et al. (2008), where the number of lanes was a factor that was found to be one of the key variables affecting secondary crash likelihood, whereas in the study by Park and Haghani (2016) and Park et al. (2017) the number of lanes was found to be negatively related to secondary crash occurrences (Zhan *et al.*, 2008; Park and Haghani, 2016a; Park, Gao and Haghani, 2017; Sarker *et al.*, 2017).

The result in Table 4-2 shows that secondary crashes are more likely to occur in the median and shoulder of the road. The likelihood of secondary crash occurrence is higher on roads with 60 mph and 65 mph speed limit. This could be because by increasing the speed limit at the crash location, flowing vehicles do not have enough time to break and prevent secondary crashes. This finding is consistent with the results of a previous study that speed is a significant factor affecting secondary crash likelihood. The study found that segments with higher posted speed limit (>55 mph) incur more secondary crashes compared with lower speed limit roads (Sarker *et al.*, 2017).

4.2 Negative Binomial Model

In addition to distinguishing between those factors associated with secondary (as compared to primary) crashes, further insights can be obtained by examining how frequently secondary crashes occur on individual road segments. As crash frequencies on a given road segment are composed of non-negative integers, count data models such as the negative binomial represent an appropriate analysis framework. Within the context of this study, the probability of the number of secondary crashes, y, occurring on interstate segment i, during a specific year of the analysis period is given as shown in Equation 3,

$$P(y_i) = \frac{e^{-\lambda}\lambda_i^{y_i}}{y_i!} \qquad (8)$$

Where, λ_i is the average number of secondary crashes for segment i. λ_i is a function of various site-specific characteristics as shown in Equation 4,

$$\lambda_i = EXP(\beta_0 + \beta_1 X_i + \beta_2 X_i + \cdots + \beta_k X_k + \varepsilon_i) \qquad (9)$$

where X_1 to X_k are a series of independent variables (e.g., traffic volumes, geometric characteristics, number lanes), β_1 to β_k are a series of parameters estimated from the regression model, and EXP(ε_i) is a gamma-distributed error term with mean equal to one and variance of α.

4.2.1 Data Summary

The data used in the analysis was interstate road segments in Michigan. The data was excluded from the sufficiency file provided by the Michigan Department of Transportation

(MDOT). The National Functional Crash (NFC) code was used to filter the interstate road segments. NFC code classifies each street and highway based upon its primary function. The sample size contains 1,557 rows, each row has the information of unique segment number and mile point information. Table 4-3 provides descriptive statistics for the segments included in the final database. The curve length percentage demonstrates the geometric characteristic of the road. The curve percentage has been calculated from the length of the curve within the segment divided by the total segment length.

AADT values ranged from 1,830 to approximately 103,000 vehicles per day (vpd), with an average of 30,768 vpd. The curve length percentage shows the geometric characteristics of the segment. Based on the data from 1,557 segments 460, about 30 percent of the segments contain curves. The right shoulder width parameter is the predominant width, to the nearest foot, of the improved shoulder on the right side of the roadway for divided segments or both sides of the roadway for undivided segments. The pavement edge or painted edge line is used as a reference point to determine the shoulder's width. The left shoulder width is the predominant width, to the nearest foot, of the improved shoulder on the left side of the roadway for divided segments. More than half of the segments are located in a large urbanized area with a population 200,000 or more, which comprised almost 52 percent of the sample. Table 4-3 also includes details about the frequency of the number of crashes within each road segment. Almost half of the crashes happen within interstate road segments on I-75 and I-94. Table 4-3 also provides details of the speed limit on segments where crashes are observed. The data shows that approximately 79 percent of crashes occurred on segments with a 70 mph speed limit.

Table 4 - 3: Descriptive statistics of pertinent variables

Parameter	**Min.**	**Max.**	**Mean**	**Std. Dev.**
Curve Length Percentage	0	100	10.492	21.125
Road number on which the crash occurred				
I-69 (1 if yes; 0 if no)	0	1	0.120	0.325
I-75 (1 if yes; 0 if no)	0	1	0.252	0.434
I-96 (1 if yes; 0 if no)	0	1	0.173	0.379
I-94 (1 if yes; 0 if no)	0	1	0.268	0.379
I-196 (1 if yes; 0 if no)	0	1	0.061	0.239
I-275 (1 if yes; 0 if no)	0	1	0.025	0.156
I-296 (1 if yes; 0 if no)	0	1	0.005	0.072
I-475 (1 if yes; 0 if no)	0	1	0.021	0.142
I-496 (1 if yes; 0 if no)	0	1	0.019	0.135
I-194, I-375, I-675 (1 if yes; 0 if no)	0	1	0.016	0.126
I-696 (1 if yes; 0 if no)	0	1	0.040	0.196
Speed limit				
55 – 65 mph (1 if yes; 0 if no)	0	1	0.087	0.281
70 mph (1 if yes; 0 if no)	0	1	0.789	0.408
75 mph (1 if yes; 0 if no)	0	1	0.125	0.330
width of the shoulder on the right side of the roadway				
Right Shoulder Width - 0 to 10 ft (1 if yes; 0 if no)	0	1	0.690	0.462
Right Shoulder Width - 11 to 14 (1 if yes; 0 if no)	0	1	0.310	0.462
width of the shoulder on the left side of the roadway				
Left shoulder width - 0 to 8 ft (1 if yes; 0 if no)	0	1	0.640	0.360
Left shoulder width 9 to 17 ft (1 if yes; 0 if no)	0	1	0.360	0.480
predominant type of median for divided segments				
Concrete barrier (1 if yes; 0 if no)	0	1	0.347	0.476
Guardrail, graded with ditch (1 if yes; 0 if no)	0	1	0.653	0.476
Urban areas designated through FHWA				
Rural (population is less than 5,000) (1 if yes; 0 if no)	0	1	0.274	0.446
Small Urban (urban cluster population is 5,000 - 49,999) (1 if yes; 0 if no)	0	1	0.072	0.258
Small Urbanized (population is 50,000 - 199,999) (1 if yes; 0 if no)	0	1	0.138	0.345
Large Urbanized (population is 200,000 or more) (1 if yes; 0 if no)	0	1	0.516	0.500

Table 4 -3 (Cont'd)

The total number of lanes at the site of the crash				
Two (1 if yes; 0 if no)	0	1	0.550	0.498
Three (1 if yes; 0 if no)	0	1	0.336	0.472
Four (1 if yes; 0 if no)	0	1	0.114	0.318
Annual Average Daily Traffic (AADT)	1,830	103,100	30,768.214	21,924.061

4.2.2 Analysis and Result of Negative Binomial

This section presents the results of negative binomial models that were estimated to investigate the relationship between secondary crash frequency within each interstate road segment. Parameter estimates are presented for the model, along with the standard errors, t-statistic, and p-value. The model includes a variable that specifies the percentage of the curve within each road segment and AADT, median type, speed limit, number of lanes, and shoulder widths. When interpreting the results from the model, a positive parameter estimate indicates that secondary crashes increase as the independent variable is increased, and the converse is true for negative parameter estimates. Table 4-4 presents the results for total secondary crashes with respect to interstate road segments.

Table 4 - 4: Model results for total secondary crashes

Variables	Estimate	SE	z-value	P-value
Intercept	-1.543	1.085	-14.228	< 0.001
I-94 **(baseline)**				
I-69	-0.097	0.149	-0.651	0.515
I-75	-0.174	0.109	-1.594	0.111
I-96	-0.102	0.103	-0.987	0.324
I-196	0.233	0.159	1.461	0.144
I-275	-0.835	0.261	-3.201	0.001
I-296	0.478	0.364	1.314	0.189
I-475	-0.235	0.317	-0.740	0.459
I-496	0.494	0.246	2.005	0.045

Table 4 – 4 (Cont'd)

I-194, I-375, I-675	-0.178	0.447	-0.399	0.690
I-696	-0.194	0.177	-1.095	0.273
Urban areas -Rural **(baseline)**				
Urban areas - Small Urban	0.310	0.175	1.773	0.076
Urban areas - Small Urbanized	0.151	0.139	1.091	0.275
Urban areas - Large Urbanized	0.324	0.124	2.625	0.009
Speed Limit - 55 – 65 mph **(baseline)**				
Speed Limit - 70 mph	0.002	0.137	0.011	0.991
Speed Limit - 75 mph	0.015	0.239	0.061	0.951
number of lanes- 2 **(baseline)**				
number of lanes- 3	-0.378	0.117	-3.230	0.001
number of lanes- 4	-0.645	0.167	-3.871	< 0.001
Guardrail, graded with ditch **(baseline)**				
Median - Concrete barrier	0.325	0.100	3.247	0.001
Right Shoulder Width - 0 to 10 ft **(baseline)**				
Right Shoulder Width - 11 to 14 ft	-0.080	0.080	-1.002	0.316
Left shoulder width - 0 to 8 ft **(baseline)**				
Left shoulder width 9 to 17 ft	-0.144	0.090	-1.601	0.109
Curve Length Percentage	0.000	0.002	-0.284	0.777
log (AADT)	1.495	0.108	13.823	< 0.001

The results from Table 4-4 show that some of the independent variables, such as the curve length percentage, and shoulder width, did not exhibit a clear relationship with the total number of secondary crashes. This finding is inconsistent with the results from the previous study that show curve segments lead to an increased risk of secondary crashes (Zhan *et al.*, 2008). Also, in the study by Sarker et al. (2017), results show that roads with broad right shoulders (width >14 ft) have fewer secondary crashes compared to roads with narrow right shoulders. This is because sufficient right shoulder allows the traffic incident management agencies to manage the incident more effectively without significantly compromising the roadway's capacity (Sarker *et al.*, 2017). Based on the results in Table 4-4, the frequency of secondary crashes has no relationship with the speed limit of the road segment. This finding is inconsistent with the results from a previous study, which was one of the key variables in affecting secondary crash likelihood (Karlaftis *et al.*, 1999a; Hirunyanitiwattana, 2006; Sarker *et al.*, 2017).

However, several independent variables were shown to strongly correlate with secondary crash frequency. Secondary crash frequency increased at the road segments with concrete barrier median, consistent with the previous study's finding (Sarker *et al.*, 2017). The study considered two types of median type, raised median and no raised median type. The result shows that roads with a raised median have more secondary crashes than roads without a raised median.

The secondary crash frequency decreases in the segments with three and four lanes compared to the segments with two lanes, which is consistent with the previous study's result where the number of lanes is among key variables that affect secondary crash occurrence (Sarker *et al.*, 2017).

Based on Table 4-4, the coefficient of the variable Urban areas - Large Urbanized is positive, indicating that the number of secondary crashes increases in the large urbanized areas with more than 200,000 population. The reason could be an increase in population leads to higher traffic volume, which increases the number of crashes and, consequently, the number of secondary crashes. The study by Sarker et al. (2017) analyzed the effect of land use on secondary crash occurrences and found that land use is among the key variables that affect secondary crash occurrences. The study considered suburban and urban areas, and the result shows that the number of secondary crashes is higher in urban areas (Sarker *et al.*, 2017).

The results from Table 4-4 show that annual average daily traffic (AADT) is statistically significant, and with the increase in the AADT the number of secondary crashes increased. One of the possible explanations is that higher traffic volume represents lower time headway between vehicles which leaves drivers less time for taking crash avoidance maneuvers when meeting hazardous satiations. This may lead to an increase in the risks of a secondary crash. This result is consistent with the finding from previous studies that crash risks increase with an increase in traffic volume (Khattak, Wang and Zhang, 2009, 2011; Zhang and Khattak, 2011; Mishra *et al.*, 2016; Sarker *et al.*, 2017).

CHAPTER 5. CONCLUSION

Crashes constitute a significant source of traffic congestion, in addition to reducing transportation system reliability, and efficiency, particularly on limited-access freeways. The congestion caused by primary crashes often exposes the following upstream vehicles to a heightened risk of secondary crashes. Therefore, transportation agencies have taken various measures to minimize and mitigate the potential for such crashes' and their resultant impacts. Although secondary crashes are relatively infrequent, they constitute a considerable safety concern and significantly impact traffic operations. Despite substantive research efforts, there is still significant uncertainty about the magnitude and nature of secondary crashes. The spatial and temporal impact of primary crashes on the road is closely related to occurrences of secondary crashes.

Past studies have proposed manual, static, and dynamic approaches to identify secondary crashes. Static methods have defined secondary crashes based on fixed spatial and temporal thresholds. In this approach, a fixed spatiotemporal window is assumed with respect to the time and location of the primary crash. However, this approach often overestimates the rate of secondary crashes by classifying all events within these windows as secondary in nature. Furthermore, the static approach considers the same window sizes for all types of primary crashes regardless of the upstream traffic flow, density, and speed. In contrast, the dynamic approach identifies a spatiotemporal impact area for each primary crash that varies based upon traffic flow characteristics. In general, more severe crashes result in greater speed reductions and have impacts that extend further spatially and over longer durations temporally.

In this work, by leveraging a vast database of all crashes occurring on Michigan Interstate roads in 2018, an extensive manual review has been performed to identify actual secondary crashes and define this control set of secondary crashes based on information from police crash reports. The manual approach results are then used to assess the accuracy of the static method in identifying secondary crashes. Based on the manual approach, about seven percent of all interstate crashes were recorded by police officers as being secondary in nature. In addition, the role of static window sizes was explored. This study suggests that while predicting secondary crashes with fixed-size windows yield a significant overestimate; window sizes can be used to derive linearly correlated values with the confirmed number of secondary crashes regardless of the window size, traffic flow, density, and speed.

This research further proposed a secondary crash identification method on freeways by tracking the spatiotemporal evolution of traffic flow. In this work, by leveraging a vast database of all crashes on interstate roads in Detroit, Michigan, a secondary crash identification approach was proposed from the integration of a speed contour plot and the spatiotemporal evolution of the primary crash impact area. Real-time travel speed data for every 15-minute time interval were collected from the Regional Integrated Transportation Information System (RITIS). To identify the crash impact area, the daily speed has been compared with the yearly average speed within each corresponding day of the week. For each primary crash, a spatiotemporal speed matrix and corresponding speed contour plot within every segment are constructed. The area is considered congested when the daily speed is lower than the average speed. If there is an existing crash in the section, the speed reduction is assumed due to the crash occurrence. Further, if another crash occurs within the primary crash impact area, it is considered a secondary crash.

In addition, the number of secondary crashes identified by the dynamic method is highly dependent on the cut-off speed that is used to identify periods during which the primary crash introduced non-recurrent congestion. Different scenarios have been considered in terms of these threshold values, such as 5 mph and 10 mph cut-off-speeds, as well as reductions of 1, 1.65, 2, and 3 standard deviations below the long-term average speeds for each day-of-week/time-of-day combination. The dynamic approach results show that the total number of secondary crashes identified in the Detroit area varies from 3 to 10 percent, considering different scenarios. So, the 5-mph cut-off point scenario was considered the least sensitivity and 3STD the highest sensitivity consecutively.

Identifying the factors that lead to secondary crashes is the first step toward preventing the occurrence of secondary crashes. Existing studies have used several statistical models to analyze the risk of secondary crash occurrence. The current research has adopted logistic regression and negative binomial models to identify characteristics distinguishing between secondary and primary crashes. This study's proposed methodological approach and research findings provided insights into the effects of traffic conditions, geometric characteristics, weather conditions, and primary crash characteristics on the probability of multiple secondary crashes on freeways.

The logistic regression model suggests that the number of lanes, weather conditions, posted speed limit, crash severity (particularly those resulting in fatal injury), number of units involved in the crash, and crashes with emergency medical service involved are among the key variables that are associated with the secondary crash occurrence. The negative binomial model suggests that annual average daily traffic (AADT), large urbanized areas (with a population of more than 200,000), and segments where median concrete barriers are present are among the key variables that are associated with the secondary crash occurrence. These results provide helpful information in developing policies and strategies to prevent the occurrence of secondary crashes. Moreover, the developed model can also be incorporated into advanced traffic control systems on freeways to help mitigate the risk of secondary crashes and allow agencies to be prepared for circumstances under which the risks of secondary crashes are elevated.

With the comparison of the proposed approach to static and dynamic methods, it is expected that the proposed approach will reduce the misidentification of secondary crashes. In addition, results may help to perform necessary strategies to mitigate secondary crashes, including improved traffic management policies and advanced intelligent transportation warning systems. While this study only examined 2018 data on interstate roads in the Detroit area, it may not be a comprehensive representation of the whole state. As such, additional research is warranted to understand differences that may exist on freeways with different traffic and geometric characteristics.

The static and dynamic windows provide a fundamental tool to quantify how the occurrence of a secondary crash is influenced by primary crash severity. The tool could also help understand how quickly information should be transferred about the occurrence and location of traffic incidents to the upstream drivers to prevent secondary crashes. A dynamic approach could be used for locating critical time/zones in order to adopt proper strategies to prevent the risk of secondary crash occurrence based on the average speed profile per year and identifying high-risk zones. In addition, identifying zones with the likelihood of secondary crash occurrence will allow pre-emptive deployment of responding agencies such as highway patrols, emergency medical services, towing agencies, etc.

Both static and dynamic methods, the two most common approaches used to define the impact area of the primary crash, have limitations that restrict their practical applications. Although the dynamic method is proven to yield more accurate results, applying it requires real-time traffic data, which is only available in limited locations. On the other hand, the static method, which considers predefined and fixed spatiotemporal thresholds, does not yield reliable results.

Secondary crashes caused by other non-crash incidents and the effect of crashes in the opposite traffic direction deserve more investigation. In summary, the static method may fail to capture the impact area of primary crashes and often overestimate the secondary crash by considering all the nearby events as the secondary crash. On the other hand, dynamic approaches address this limitation by determining the spatiotemporal thresholds of primary crashes based on real-time traffic flow characteristics such as speed and density. Further investigation and dynamic methods are recommended for future study.

A complete understanding of secondary crash characteristics, contributing factors with respect to traffic, geometric conditions, and crash details can simplify and accelerate the identification of secondary crashes without analyzing individual reports. While most automatic identification methods of the secondary crash remain limited to the spatiotemporal boundary analysis, it has been demonstrated that the dynamic method is substantially more relevant in locations where the traffic flow is monitored and recorded.

Ultimately, this research provides important insights that can aid road agencies in more proactive management of traffic crashes and other incident clearance activities. With that being said, there are some practical limitations, and the following research tasks are recommended as the next steps building upon the results of this research,

- Investigating the role of prevailing traffic characteristics on secondary crashes should be considered in greater detail. This study shows that speed reductions have pronounced impacts on secondary crash occurrence. However, additional information, such as traffic volume levels and other measures may help to further our understanding of these relationships. In general, many secondary crashes occur during congested traffic conditions, primarily using varying spatiotemporal thresholds depending on the prevailing traffic conditions.

- Conducting additional case studies and varying spatiotemporal thresholds depending on the prevailing traffic conditions is expected to improve the accuracy of the thresholds used in the static model.
- In a dynamic approach, the effect of special events and holidays, road maintenance and its effects on average speed, percentage of lane closure, shoulder blocked should also be investigated.
- In addition, the role of attributes such as work zones, design features, vehicle technology, and pavement conditions in secondary crash occurrence should be investigated as these factors could affect the average speed in a segment.

BIBLIOGRAPHY

BIBLIOGRAPHY

Chang, G.-L. and Steven, R. (2002) ‘Performance Evaluation of CHART (Coordinated Highways Action Response Team) Year 2002 (Final Report) Performance Evaluation of CHART’, *University of Maryland, College Park and Maryland State Highway Administration*.

Chung, Y. (2013) ‘Identifying primary and secondary crashes from spatiotemporal crash impact analysis’, Transportation Research Record, (2386), pp. 62–71.

Guo, J. *et al.* (2017) ‘Short-term traffic flow prediction using fuzzy information granulation approach under different time intervals; Short-term traffic flow prediction using fuzzy information granulation approach under different time intervals’, *IET Intelligent Transport Systems*. Institution of Engineering and Technology, 12(2), pp. 143–150.

Hirunyanitiwattana, W. S. P. M. (2006) ‘Identifying secondary crash characteristics for California highway system’.

Imprialou, M. I. M. *et al.* (2014) ‘Methods for defining spatiotemporal influence areas and secondary incident detection in Freeways’, *Journal of Transportation Engineering*, 140(1), pp. 70–80.

Jalayer, M., Baratian-Ghorghi, F. and Zhou, H. (2015) ‘Identifying and characterizing secondary crashes on the Alabama state highway systems’, *Advances in Transportation Studies*, (37), pp. 129–140.

Karlaftis, M. G. *et al.* (1999a) ‘ITS impacts on safety and traffic management: an investigation of secondary crash causes’, *Journal of Intelligent Transportation Systems 5*, no. 1, pp. 39–52.

Karlaftis, M. G. *et al.* (1999b) ‘ITS Impacts on Safety and Traffic Management: An Investigation of Secondary Crash Causes’, *ITS Journal*, 5(1), pp. 39–52.

Khattak, A. J., Wang, X. and Zhang, H. (2010) ‘Spatial analysis and modeling of traffic incidents for proactive incident management and strategic planning’, *Transportation Research Record*, (2178), pp. 128–137.

Khattak, A. J., Wang, X. and Zhang, H. (2011) ‘iMiT: A Tool for Dynamically Predicting Incident Durations, Secondary Incident Occurrence, and Incident Delays’, *TRB 90th Annual Meeting Compendium of Papers DVD*, (January), pp. 1–17.

Khattak, A., Wang, X. and Zhang, H. (2009) ‘Are incident durations and secondary incidents interdependent?’, *Transportation Research Record*, (2099), pp. 39–49.

Kitali, A. E., Alluri, P., Sando, T. and Wu, W. (2019) ‘Identification of Secondary Crash Risk Factors using Penalized Logistic Regression Model’, *Transportation Research Record*, 2673(11), pp. 901–914.

Kitali, A. E., Alluri, P., Sando, T. and Lentz, R. (2019) ‘Impact of Primary Incident Spatiotemporal Influence Thresholds on the Detection of Secondary Crashes’, *Transportation Research Record*, 2673(10), pp. 271–283.

Kopitch, L. and Saphores, J.-D. M. (2011) ‘Assessing Effectiveness of Changeable Message Signs on Secondary Crashes’, No. 11-427.

Mishra, S. *et al.* (2016) ‘Effect of primary and secondary crashes: Identification, visualization, and prediction’, p. No. CFIRE 09-05.

Moore, J. E., Giuliano, G. and Cho, S. (2004) ‘Secondary accident rates on Los Angeles freeways’, *Journal of Transportation Engineering*, 130(3), pp. 280–285.

Owens, N. *et al.* (2010) ‘Traffic incident management handbook’, *Washington, DC: Federal Highway Administration, Office of Transportation Operations*, (Report No. Vol. 9. FHWA-HOP-10-013).

Ozbay, K. and Kachroo, P. (1999) ‘Incident management in intelligent transportation systems’, *MA: Artech House Publishers*.

Park, H., Gao, S. and Haghani, A. (2017) ‘Sequential interpretation and prediction of secondary incident probability in real time’, No. 17-062.

Park, H. and Haghani, A. (2016a) ‘Real-time prediction of secondary incident occurrences using vehicle probe data’, *Transportation Research Part C: Emerging Technologies*. Elsevier Ltd, 70, pp. 69–85.

Park, H. and Haghani, A. (2016b) ‘Stochastic Capacity Adjustment Considering Secondary Incidents’. IEEE, 17(10), pp. 2843–2853.

Park, H., Haghani, A. and Hamedi, M. (2013) ‘Quantifying non-recurring congestion impact on secondary incidents using probe vehicle data’, *54th Annual Transportation Research Forum, TRF 2013*, (March 2013), pp. 6–17.

Raub, R. A. (1997a) ‘Occurrence of secondary crashes on urban arterial roadways’, *Transportation Research Record*, (1581), pp. 53–58.

Raub, R. A. (1997b) ‘Secondary crashes: An important component of roadway incident management’, *Transportation Quarterly*, 51(3), pp. 93–104.

Sarker, A. A. *et al.* (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*. Elsevier Ltd, 60, pp. 142–160.

Sarker, A. A. *et al.* (2017) ‘Prediction of secondary crash frequency on highway networks’, *Accident Analysis and Prevention*. Elsevier Ltd, 98, pp. 108–117.

Skabardonis, A. *et al.* (1995) *Freeway Service Patrol Evaluation. Berkeley,CA: California PATH Research Report*. California.

Smith, B. L. and Ulmer, J. M. (2003) 'Freeway Traffic Flow Rate Measurement: Investigation into Impact of Measurement Time Interval', *Journal of Transportation Engineering, 129(3)*, pp. 223–229.

Sun, C. C. and Chilukuri, V. (2010) 'Dynamic incident progression curve for classifying secondary traffic crashes', *Journal of Transportation Engineering*, 136(12), pp. 1153–1158.

Sun, C. and Chilukuri, V. (2007) 'Secondary Accident Data Fusion for Assessing Long-Term Performance of Transportation Systems', (MTC Project 2005-04), pp. 1–35.

Tedesco, S. *et al.* (1994) 'Development of a model to assess the safety impacts of implementing IVHS user services', *IVHS America Annual Meeting. 2 Volumes*.

Tian, Y., Chen, H. and 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.

Vlahogianni, E. I. *et al.* (2010) 'Freeway operations, spatiotemporal-incident characteristics, and secondary-crash occurrence', *Transportation Research Record*, (2178), pp. 1–9.

Vlahogianni, E. I., Karlaftis, M. G. and Orfanou, F. P. (2012) 'Modeling the effects of weather and traffic on the risk of secondary incidents', *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 16(3), pp. 109–117.

Wang, J., Xie, W., *et al.* (2016) 'Identification of freeway secondary accidents with traffic shock wave detected by loop detectors', *Safety Science*. Elsevier Ltd, 87, pp. 195–201.

Wang, J., Liu, B., *et al.* (2016) 'Modeling secondary accidents identified by traffic shock waves', *Accident Analysis and Prevention*. Elsevier Ltd, 87, pp. 141–147.

Wang, Z. and Jiang, H. (2020) 'Identifying Secondary Crashes on Freeways by Leveraging the Spatiotemporal Evolution of Shockwaves in the Speed Contour Plot', *Journal of Transportation Engineering, Part A: Systems, 146(2), 04019072*. American Society of Civil Engineers (ASCE), 146(2).

Wang, Z., Qi, X. and Jiang, H. (2018) 'Estimating the spatiotemporal impact of traffic incidents: An integer programming approach consistent with the propagation of shockwaves', *Transportation Research Part B: Methodological, 111*, pp. 356–369.

Xu, C. *et al.* (2016) 'Real-time estimation of secondary crash likelihood on freeways using high-resolution loop detector data', *Transportation Research Part C: Emerging Technologies*. Elsevier Ltd, 71, pp. 406–418.

Yang, B., Guo, Y. and Xu, C. (2019) 'Analysis of Freeway Secondary Crashes with a Two-Step Method by Loop Detector Data', *IEEE Access*. IEEE, 7, pp. 22884–22890.

Yang, H. *et al.* (2014) 'Development of online scalable approach for identifying secondary crashes', *Transportation Research Record*, 2470(26), pp. 24–33.

Yang, H. *et al.* (2017) 'Use of ubiquitous probe vehicle data for identifying secondary crashes', *Elsevier*, pp. 138–160.

Yang, H. *et al.* (2018) 'Methodological evolution and frontiers of identifying, modeling and preventing secondary crashes on highways', *Accident Analysis and Prevention*. Elsevier Ltd, 117, pp. 40–54.

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

Zhan, C. *et al.* (2008) 'Understanding the characteristics of secondary crashes on freeways', *87th Annual Meeting of the Transportation Research Board, TRB, No. 08-1835*.

Zhan, C., Gan, A. and Hadi, M. (2009) 'Identifying secondary crashes and their contributing factors', *Transportation Research Record*, (2102), pp. 68–75.

Zhang, H., Cetin, M. and Khattak, A. J. (2015) 'Joint analysis of queuing delays associated with secondary incidents', *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 19(2), pp. 192–204.

Zhang, H. and Khattak, A. (2011) 'Spatiotemporal patterns of primary and secondary incidents on urban freeways', *Transportation Research Record*, (2229), pp. 19–27.

Zhang, X. *et al.* (2020) 'Identifying secondary crashes using text mining techniques', *Journal of Transportation Safety and Security*. Taylor & Francis, 12 (10)(0), pp. 1338–1358.

Zhang, X., Green, E. and Chen, M. (2019) 'Impact of Primary Incident Spatiotemporal Influence Thresholds on the Detection of Secondary Crashes', *Transportation Research Record*. Elsevier Ltd, 2673(3), pp. 1–16.

Zheng, D. *et al.* (2014) 'Identification of Secondary Crashes on a Large-Scale Highway System', *Transportation Research Record: Journal of the Transportation Research Board*, 2432(1), pp. 82–90.

Zheng, D. *et al.* (2015) 'Analyses of multiyear statewide secondary crash data and automatic crash report reviewing', *Transportation Research Record*, 2514, pp. 117–128.