2016

A Data Driven Approach to Quantify the Impact of Crashes

Obaidur Rahman Kazi
University of Kentucky, obaidur.kazi.uky@gmail.com
Digital Object Identifier: https://doi.org/10.13023/ETD.2016.516

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Obaidur Rahman Kazi, Student

Dr. Mei Chen, Major Professor

Dr. Yi-Tin Wang, Director of Graduate Studies

A DATA DRIVEN APPROACH TO QUANTIFY THE IMPACT OF CRASHES

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in Civil Engineering
in the College of Engineering
at University of Kentucky

By

Obaidur Rahman Kazi

Lexington, Kentucky

Director: Dr. Mei Chen,

Associate Professor of Civil Engineering

Lexington Kentucky

2016

ABSTRACT OF THESIS

A DATA DRIVEN APPROACH TO QUANTIFY THE IMPACT OF CRASHES

The growth of data has begun to transform the transportation research and policy, and open a new window for analyzing the impact of crashes. Currently for the crash impact analysis, researchers tend to rely on reported incident duration, which may not always be accurate. Further, impact of the crashes could linger a much longer time at upstream, even if the records are correct for the crash spot and it is a challenge to quantify the impact of a crash from the complex dynamics of the recurrent and non-recurrent congested condition. Therefore, a difference-in-speed approach is developed in this research to estimate the true crash impact duration using stationary sensor data and incident logs. The proposed method used the Kalman filter algorithm to establish traveler's anticipated travel speed under incident-free condition and then employ the difference-in-speed approach to quantify the temporal and spatial extent of the crash. Moreover, potential applications such as statistical models for predicting the impact duration and total delay were developed in this research. Later, an analysis on distribution of travel rate was performed to describe and numerically show to what extent crashes influenced travel rates compared with the normal conditions at different periods of the day and by the crash types. This study can help to shape incident management policies for different types of crashes at different periods and illustrates the usages of data to improve the understanding of crashes, their impact, and their distribution in a spatial-temporal domain.

KEYWORDS: Heat map, crash analysis, spatial and temporal impact, reliability, Big Data.

Obaidur Rahman Kazi

November 28, 2016

A DATA DRIVEN APPROACH TO QUANTIFY THE IMPACT OF CRASHES

By

Obaidur Rahman Kazi

Dr. Mei Chen

Director of Thesis

Dr. Yi-Tin Wang

Director of Graduate Studies

November 28, 2016

Date

Acknowledgements

I would like to express my deepest appreciation and gratitude to my advisor, Dr. Mei Chen for her continuous guidance, assistance, and support throughout my M.S. thesis process. Almost every day she allocated time for me from hers busy schedule to discuss any technical problem I faced. Also, I am grateful as she involved me in different research activities and projects. I could not have imagined having any better advisor and mentor for my graduate study.

Next, I wish to thank my entire Thesis Committee including Dr. Mei Chen, Dr. Reginald R. Souleyrette, and Dr. Nick Stamatiadis. Thank you for reviewing my thesis, and providing valuable insights about the research.

I would also thank my co-worker Xu Zhang, for being there to discuss my confusions. I want to cordially thank Tate, who was always available to help me. Thanks to Adam and Galen for their support.

Finally, I would like to thank my mother for all the support and inspiration she provides me every day from thousands of miles away, and my sisters Resma and Nipa, for inspiring me all the time. I will always be thankful to my other family members and friends from Bangladesh for their understanding and wholehearted support towards my work.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

The growth of data has begun to transform research and policy and transportation is no exception. As detection and communications technologies grow more sophisticated, transportation sensors continue to generate more data. Abundant data is available on infrastructure condition, operating characteristics, and traveler behavior across temporal and spatial domains. This data presents an opportunity to better understand the interactions between travelers and the traveling environment. Also, this data opens a new window for comprehensive analysis of traffic incidents and its spatiotemporal impact as congestion on normal traffic flow.

A traffic incident, such as a crash, is one of the major causes of congestion and unreliable travel time on roadways, which seriously affects travel experience and causes significant economic and environmental losses. Congestion on roadways is one of the major problems in the USA from both travelers and an operational point of view. This is due to the growth of population and increasing traffic demand. This congestion affects the nation's economy by causing thousands of unproductive hours, drastically alters the transportation schedule, causes distress to drivers and passengers, and increases atmospheric pollution.

Traffic congestion is the result of many different factors and mainly occurs when traffic demand exceeds the physical capacity of the roadway and traffic influencing events, such as an incident or bad weather, took place. That's why congestion could be divided into two major groups: recurrent and non-recurrent congestion. Recurrent congestion occurs almost every day at specific times of the day (peak period) when traffic flow exceeds capacity. On the other hand, non-recurrent congestion occurs due to unpredictable changes from time to time or day to day. Examples of non-recurrent congestion include unexpected events such as incidents, work zones, special events and weather, where peak demands are higher than normal [1].

1.2 Motivation

The strategies to address the congestion problem can be categorized as follows:

- Expanding the capacity of the existing transportation system
- Increasing operational efficiency of the existing infrastructure capacity

The experience over the years in the transportation world has shown the most efficient, intelligent and economical solution to minimize the congestion problem is to increase the operational efficiency of the existing transportation infrastructure. The strategies falling under this category are called Intelligent Transportation System (ITS). ITS can be defined as those techniques which ensure safe, fast, smooth and economical transportation of people and goods using the help of modern technology.

One of the important strategies of ITS is "Traffic Incident Management (TIM)," which is defined as the planned, systematic and coordinated use of human, institutional and technical resources to reduce the duration and impact of incidents, and improve the safety of motorists and crash victims [2].

The main purpose of traffic incident management is to reduce the influence of traffic incidents, which requires a timely and precise estimation of traffic incident duration. By performing a reliable prediction of incident duration, traffic management could deploy appropriate measures around the traffic incidents and provide travelers with real-time traffic information to reduce incident related traffic congestion.

Several state's Department of Transportation (DOTs) have adopted traffic incident management programs to reduce the incident duration and incident induced traffic congestion. These efforts have proven beneficial in terms of the return on capital investment. According to the 2012 Urban Mobility Report, incident management treatments have saved about 337 million hours of delay, which is equivalent to 7.2 billion dollars saved on congestion costs.

To help transportation agencies in developing a more efficient TIM programs, the need to examine traffic incident's impact on traffic flow drew more attention of researchers. Currently for the crash impact analysis, agency tends to rely on incident log. But, there are some problems in the reported time frame; first of all it is not always accurate, most of the cases reported end time is not reflective of the true end of the crash impact. Sometimes the end time was missing in the indent log. Because in the event of a crash, responding officers fill out a crash report along with other duties such as making arrangements for emergency service vehicles to arrive at the scene of the crash, clearing and securing the roadway, and collecting information surrounding the crash. Generally, all of this is done under time pressure and understandably, filling out a crash report cannot always be the officer's first priority. Sometimes, reports are completed once the officer is back at the station several hours or days after the crash occurred which leads to an unreliable data entry in the crash report. Even if the reported end time is accurate for the crash location, crash impact may linger for a much longer time at upstream of the roadway. Thus, there is a need to identify the true end-time/impact of the crash.

Therefore, the chief motivation of this study is to capture the "true" impact of the crash by analyzing the stationary sensor data and develop a framework to automatically identify the spatiotemporal extent of congestion due to the crash.

1.3 Research Objectives

Within the context of the events and needs presented earlier, the objectives of the research are-

- Developing a method that can capture the spatiotemporal extent of congestion due to a crash by combining incident data with stationary sensor data-set. The step by step process provides a practical approach to develop a crash-related congestion quantification method.

- Exploring the possible applications of the methodology which includes:
  - Developing a statistical model to predict the duration of crash impact. The model takes into consideration the factors affecting the duration of impact and uses regression technique to develop a framework for predicting the duration
  - Developing a statistical model to predict total delay of the crash. The model identifies the factors that affect the total delay due to the crash and uses regression technique to develop a framework for predicting the delay.
  - Analyzing the impact of crash on reliability

Effective management of congestion induced by the crashes largely depends on timely response with appropriate measures such as deployment of tow trucks, inform travelers in advance, route diversion etc. An accurate forecast of the potential extent of the impact of crashes in terms of duration and delay would help to determine appropriate management strategies.

1.4 Organization of the Research

The research consists of six chapters. Chapter 1 shows the background, motivation and objectives of the research. Chapter 2 presents a brief discussion of the previous literature and critical analysis of previous research works. Chapter 3 provides the detailed description of the study area and data-sets used in the research. Chapter 4 presents the methodology to quantify the impact of crash and discussion of the case study. Chapter 5 reveals the detailed description of the possible applications of the proposed methodology. Finally, Chapter 6 summarizes the findings of this research and discusses its limitations and the scope for the future work.

This chapter provides a general outline of the full research. Objectives and background of the research are described here, which will help to understand the total framework of the research. The next chapter will show a critical review of previous research.

Chapter 2 Literature Review

2.1 Overview

Traffic congestion is a global problem. Traffic congestion related studies conducted in US reveals that building more and more highways is not a sustainable solution to minimize congestion problem. Increasing roadway capacity as a measure to reduce traffic congestion is not a sustainable measure, because it requires an enormous quantity of resources such as land, money, fuel and labor. On the other hand, alternative approaches such as land use change, improvement of roadway operations, travel demand management, incident management and work zone management have been suggested as effective and sustainable measures to minimize the congestion. Among these alternative options, incident management is the most promising short term action to alleviate congestion problems on freeways and urban arterials. That's why traffic incident management measures have received considerable attention from the traffic management agencies and plenty of research has been conducted to find out the most effective and optimum way to manage incident. The widespread success of such program inspired more cities to adopt incident management as a viable step forward in improving the city's transportation reliability and safety.

Lots of research has been conducted to better understand the incident and its different characteristics. A brief review of such study is presented in the following:

2.2 Incident Duration Studies

A large number of studies focused on examining incident duration. Incident duration is defined as the elapsed time from the moment an incident is detected until the cause is removed from the scene [3]. Over the last few years, various methodologies and techniques have been used to model and analyze the incident duration. These models mainly establish relationship between incident duration and different influencing factors. A set of variables significantly affecting incident duration have been identified. They are named as incident characteristics and listed as follows:

- incident type and severity,
- the number and type of vehicles involved,
- geometric characteristics,
- temporal characteristics,
- environmental effects, and
- operational factors.

The most representative approaches for incident duration models are described in the following:

Linear regression analyses: Garib et al. developed linear regression models to estimate magnitude and duration of freeway incident delays [3]. The author developed a multiple linear regression model based on 205 incidents over a two-month period from Oakland, California, to predict incident duration as a function of six significant variables: number of lanes affected (X_1), number of vehicles involved (X_2), truck involvement as binary variable (X_3), natural logarithm of the police response time (X_4), time of day as binary variable (X_5) and weather conditions as binary variable (X_6). Then the final log-based regression model is given by:

$$\text{Log (Duration)} = 0.87 + 0.027\, X_1\, X_2 + 0.2\, X_3 + 0.68\, X_4 - 0.17\, X_5 - 0.24\, X_6$$

Non-parametric regression methods: The basis of nonparametric regression is to make current decisions based on past, similar experience. Smith and Smith used non-parametric regression model to predict incident duration; however the performance of the model was unsatisfactory, with an average error more than 20 min [4].

Time sequential methods: Khattak et al. developed time sequential model by identifying ten distinct stages of incident duration based on the availability of information [5]. Each stage has a separate truncated regression model and the model progressively add more variables. The purpose of the study is to demonstrate the methodology rather than show its performance.

Conditional probability analyses: Developing conditional probability is another use of probability in incident duration. Traffic Management agencies might be interested to know the probability of an incident lasting 30 minutes given that it has been already been active for 15 minutes, or similar case. To give answer for such situation, Nam and Mannering, used hazard based models (using conditional probabilities to find the likelihood that an incident will end next short time period given its continuing duration) to develop incident duration model [6].

Moreover, incident duration models based on probabilistic distribution analyses ([7], [8]), support vector regression [9], discrete choice models [10], Fuzzy logic models [11], Bayesian classifier [12], artificial neural networks[13] are also frequently explored by the researchers. It should be noted that all of the developed models are often site/facility specific and calibration is required for their use at other location/facilities.

2.3 Incident Delay Studies

Set of research has primarily focused on examining incident induced delay. An incident induced delay can be defined as additional delay produced by the incident. Garib et al. developed linear regression models to estimate cumulative incident delay as a function of incident duration, traffic demand, and capacity reduction represented by number of lanes affected and number of vehicles involved [3]. Again to evaluate the effectiveness and performance of the freeway, Skabardonis et al. developed a methodology to estimate the incident induced delay by comparing the travel time(calculated from loop detector's speed) under incident and incident free condition [14]. The methodology was developed based on the assumption that incident will affect the transportation system by increasing the travel time of the road users. Additionally, a large number of studies focused on examining delay based on queuing theory and shock wave analysis[15]. However, there are some limitations of such queuing and shockwave based studies such as queuing models require identification of capacity reduction, demand change to calculate the extent of delay which is difficult to measure due to stochastic nature of the incidents [14]. Again shockwave models estimate delay based on wave speed, but inaccuracy in estimating wave speed might cause serious misinterpretation of the incident induced delay [16].

2.4 Incident and Simulation

There are few studies which conduct extensive examination to determine the potential of using traffic simulation model for the analysis of the effects of traffic incident and corresponding incident management strategies. Cragg and Demestsky developed a CORSIM simulation model to assess incident impact and traffic diversion strategies on freeway [17]. Zhang et al. used TSIS simulation to predict delays due to incident on freeway [18]. Recently a legislation has been passed in South Carolina regarding quick clearance criteria in an incident site. Fries et al. assessed the impact of quick clearance criteria deployed in South Carolina using Paramics based simulation [19]. Kabit et al. developed a VISSIM simulation model to quantify the impacts of major traffic incidents and estimate their associated cost [20]. These simulation based studies demonstrate promising results of using simulation based approaches to determine the impact of incident. But the needs for detailed traffic and incident data, calibration of simulation model limited their uses for large scale analysis.

2.5 Incident and Reliability

Transportation researchers have recently turned their attention toward travel time reliability. Using traffic crash and empirical traffic flow data collected from the Netherlands, Tu et al. presented an empirical travel time reliability analysis [21]. One limitation in their research was that the duration and severity of each accident were unknown, so they assumed each accident had a duration of three hours. Yu et al. used reliability analysis to assess freeway crash risks and to evaluate hazardous freeway segments [22]. Reliability analysis accomplishes this by integrating traffic flow parameters and real-time crash occurrence risk at the disaggregate level with weather parameters. Yu et al. found this method provided more accurate crash predictions than logistic regression [22]. Zhong et al. used data on rural roads in Wyoming to model and predict crashes [23]. The data they used included accident records, traffic volume, speed, and other factors, from 36 roads over a 10-year period. Negative binomial regression and Poisson regression were used to examine the causes of rural crashes. Multiple regression approaches have attempted to analyze the relationship between crash rates and geometric roadway features. However, multiple studies have found linear regressions are unsuitable ([24], [25]), [23]) demonstrated that roads with higher speeds and traffic volumes elevated crash rates at certain higher risk locations. Wright et al. showed that incidents produce higher values in all reliability measures [26]. They also examined how incidents affect the probability of traffic congestion on freeway segments. Compared to the normal condition, they found that shoulder incidents significantly increased the probability of freeway segment traffic breakdown, while incidents spread across multiple lanes resulted in the most significant increases in travel time variability and in the buffer index.

2.6 Incident Impact Studies

With the advancement of technology, the traffic monitoring system is also equipped with modern sensory devices which generate and archive large amount of data each day. This data is continuous and capture the dynamics of traffic at monitored segments of the highway network. Now-a-days many studies have begun to investigate those archived traffic sensor data sets and incident logs to quantify the impact of incident. Chung and Recker utilized the loop detector data to identify incident induced congestion [27]. They applied the integer programming technique to identify temporal and spatial extent of the region of congestion caused by accident and estimates associated delay. Pan et al. also investigated archived traffic sensor data and incident log to estimate the impact of incidents [28]. With the availability of massive sensor data sets and incident logs, more data driven and location specific approaches should be developed to identify the spatiotemporal extent of traffic incident and incident characteristics, which significantly influence the incident induced congestion.

2.7 Other Studies

Few studies have looked at the interactive effects of traffic and weather factors and roadway geometry on different crash types. Among them, Yu et al. attempted to explore the use of microscopic traffic and weather indicators to differentiate between crash types and to analyze the crash type propensity at the micro-level for three major crash types — rear-end, sideswipe, and single-vehicle crashes [29]. Ahmed et al. investigated the effect of the interaction between roadway geometric features and real-time weather and traffic data on the occurrence of crashes on a mountainous freeway [30]. They found that geometric factors were significant in all seasons. Crash likelihood could double during the snowy season due to slick pavement conditions and steep grades, and when combined, produced a hazardous road surface. On the other hand, Hojati et al. presented a framework to exhaustively mine traffic-incident data and directed subsequent analysis toward an incident delay and travel-time reliability model [31].

Though there are several proposed models that are highly efficient, they cannot be applied to other cases because different studies call for the use of different variables. As such, results may not be transferable across different locations. Data collection and reporting process have also been incommensurate. While the findings of previous studies will not reduce the number of crashes/incidents, they will reduce their effects and guide the traffic management center to take adequate measure to minimize the congestion.

In this study we developed a data driven approach to quantify the impact of crash, not only relying on the historical trends of the traffic, rather considering both pre-crash on-going traffic condition and historical trends. For this purpose, we introduced the Kalman filter algorithm to combine current traffic and historical trends to formulate crash free normal traffic pattern which will open a new window to capture the dynamics of the impact of the crash during both recurrent and non-recurrent congested condition. Furthermore, this method captures the actual impact of crash and real end time of the crash impact, which is more accurate than the end time reported in the incident log.

In this chapter, various past studies have been discussed briefly. Also, the limitations of many existing studies are presented. In the next chapter, a detailed description of the study area and data sets used in our research will be provided.

Chapter 3 Data Collection

Stationary sensor data is used to develop the methodology to capture the crash impact. This chapter provides a brief description of the study area and different data sets.

3.1 Study Area

This study examines the impacts of the crashes along northbound and southbound I-65 in the Louisville metropolitan area. The study segment is 5.6 miles long in the northbound direction which starts from MP-131 and ends at MP-136.6. Again, the segment is 5 miles long in the southbound direction which starts from MP-136 and ends at MP-131. The speed limit along both segments is 55 mph. There are a total of 26 stationary sensors located in the study area which continuously collect the speed, volume and occupancy information at each sensor location. Figure 1 shows the spatial extent of the study corridor.

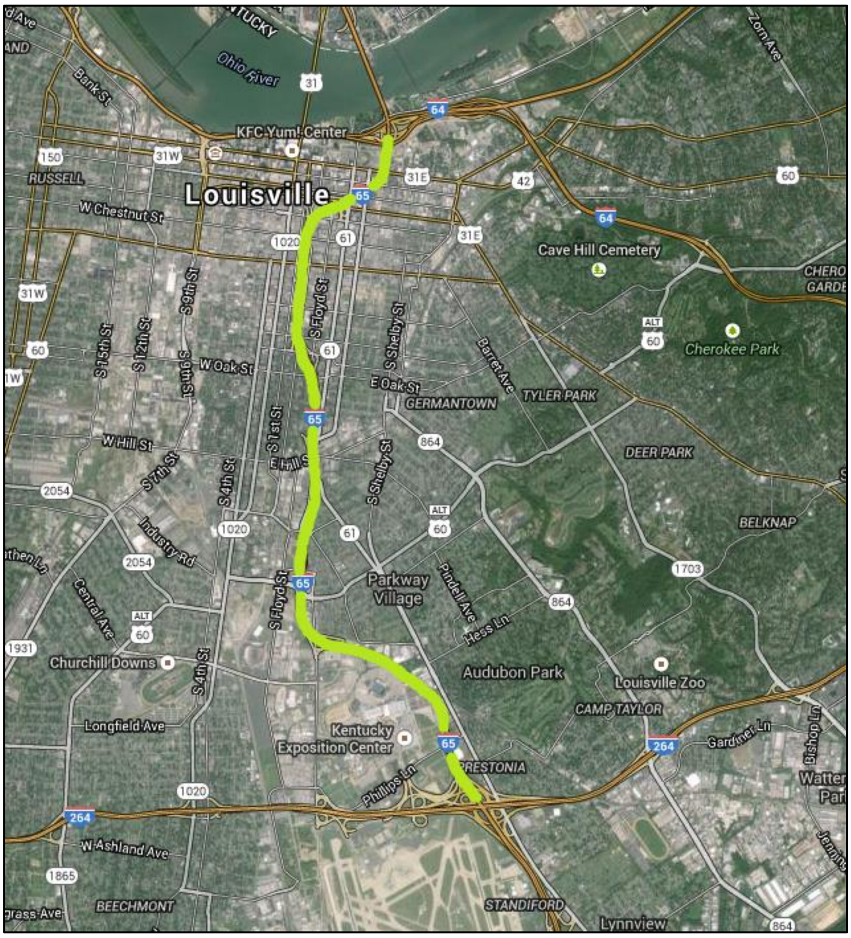


Figure 1: Study Corridor of I-65

3.2 Stationary Sensor Data

Stationary sensor data is provided by TRIMARC. TRIMARC is a regional traffic management center designed to improve the performance of the freeway system in metropolitan Louisville, which extends into Southern Indiana. The TRIMARC data environment contains time, speed, volume, and lane occupancy data for each day. This information was recorded at 15-minute intervals along each detector section. The sensor data was originally recorded in 30-second slots. The TRIMARC server aggregated them every 15 minutes. There are 15 TRIMARC sensors located on I-65 N and 11 sensors on I-65 S. The average spacing between two sensors is approximately 0.4 mile. Figure *2* maps the sensor locations on I-65 N. In this study, the data is collected from January 2011 to December 2013.

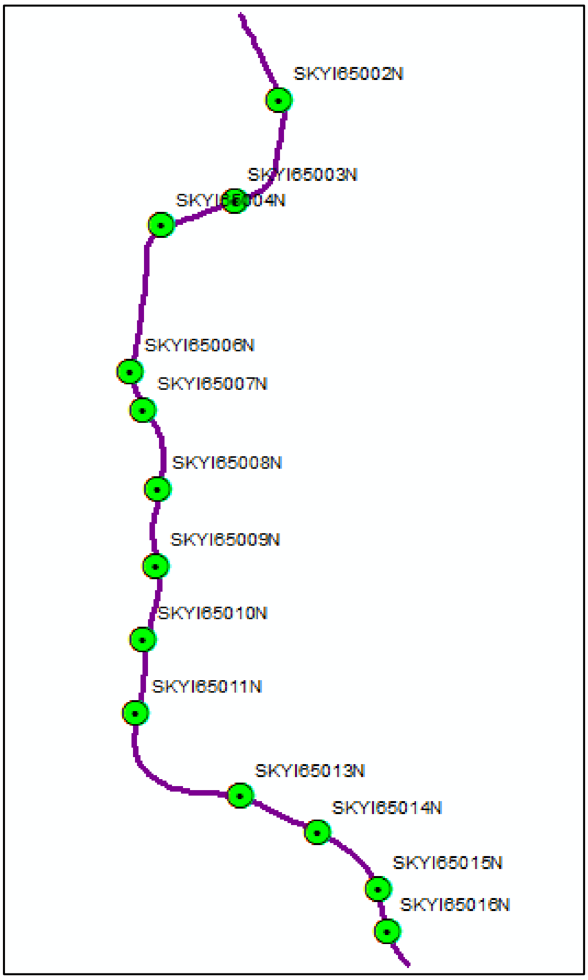


Figure 2: Location of Stationary Sensors on I-65N

3.3 Incident Data

The TRIMARC data environment also contains incident logs. An incident log is comprised of the incident description, spatial, temporal and environmental characteristics, and a short descriptive information about the incident. A detailed description of all records in the incident data is provided in the Appendix A. In this study, the incident data is collected from January 2011 to December 2013. A generic description of the incident records considered in this study is given in the following:

- Incident Description: This section contains general characteristic features of an incident like the type of incident, the numbers of vehicle involved, injury or not, number of lane blocked etc.
- Spatial Characteristics: This section consists of approximate location of the incident, which is documented by recording the name of the route, direction and the nearest mile-marker.
- Temporal Characteristics: This section contains the time at which the incident was notified to the TMC, the time at which the incident site was cleared and brought back to normal condition and total duration of the incident
- Environmental Characteristics: This section contains weather conditions (such as rainy or sunny, snowy or not etc) at the time of the incident.

3.4 Weather Data

Historical weather records are downloaded from https://www.wunderground.com/, which collects the data from weather sensors at Louisville International Airport, which is about 2 miles away from the study corridor. The data includes weather information such as temperature, wind speed, direction, visibility, weather condition, precipitation etc. In our study we have collected and processed the rain information that occurred in 2011, 2012 and 2013 in the study area. Later, this rain information is matched with the incident data to get the complete picture of the incident.

This chapter provides a brief overview about the study area and a brief description of the data sets that have been used in the research. In the next chapter, research methodology will be discussed elaborately.

Chapter 4 Research Method

4.1 Overview

Traffic incident is one of the major causes of non-recurrent congestion on roadways which seriously affects travel experience and causes economic and environmental losses. If a crash occurs during peak period, it exacerbates the congestion by lowering the traffic speed. The need to examine the crash impact on traffic flow is very important to develop an efficient crash management program; also, the end time reported in the incident log is not always reflective to the true end of the crash impact. Sometimes impact of the crash on traffic flow continued beyond the reported end time. Thus, there is a need to identify the actual end time and real impact of the crash on traffic. In order to identify the impact of the crash, a data driven approach is proposed in this research by analyzing the stationary sensor data under both crash-free and crash conditions. This methodology can capture crash impact on traffic flow for both recurring and non-recurring congested conditions and identify both crash scenario and spatiotemporal impacts of the crash.

4.2 Identification of Crash Impact Zone

In order to automatically identify the impact of the crash on the traffic flow, this study proposes a data-driven approach to analyze the stationary sensor data along with the incident log under the crash condition. The proposed methodology to identify the crash impact zone contains five major steps:

1. Obtaining the current speed profile under crash condition
2. Identifying the crash scenario
3. Determining the background speed profile
4. Identifying the crash impact from the difference-in-speed profile
5. Determining the impacted region

4.2.1 Obtaining the Current Speed Profile

When a crash occurred during the uncongested condition, it is very easy and straightforward to isolate the crash impact from normal traffic conditions. However, when a crash occurred during congested condition, it is very difficult to separate the crash impact from the congested traffic condition. In order to know if there is any impact of crash, first a current speed profile is obtained for the day when crash occurred. A current speed profile provides the ground truth measure of real time traffic conditions under the impact of both crash-induced and recurrent congested condition. To represent the current speed profile visually, using all stationary sensors data, a current speed contour map has been developed, which is described as follows:

At first, we assume each sensor measurement represents the segment traffic condition from that sensor to the adjacent upstream stationary sensor. And the current traffic speed V of the j^{th} segment at i^{th} time slice could be denoted as V(i,j), where i= 1,2,3….96 (as ninety six 15 minute time slice is equal to 1 day) and j= 1,2,3…….s (s is the total number of sensor). Later the current traffic speed measurements are coded as a continuous color (red to green represents low speed to high speed) to build a contour map. This contour map can also be called space time velocity map or heat map. The heat map (Figure 3) increases the visual understanding of the spatiotemporal change of speed and also highlights the congested area.

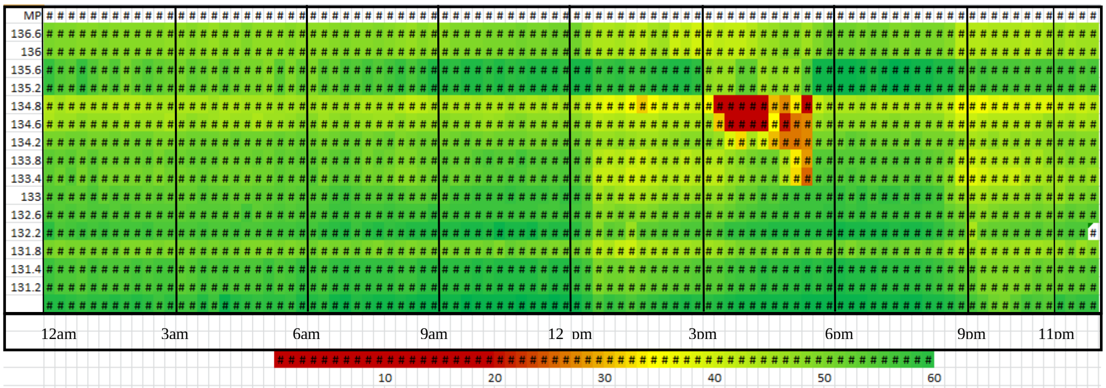


Figure 3: Heat Map

4.2.2 Identifying the Crash Scenario

In this step, the methodology will identify the crash scenarios, which can be divided into three categories:

- Type-1: Crash Induced Congestion
- Type-2: Crash without Congestion
- Type-3: Congestion Induced Crash

Crash induced congestion could be defined as the scenario when congestion occurred as a result of the crash, which is defined in this study as Type-1 crash.

Crash without congestion is the scenario when there was no congestion after the crash, which is defined in this study as Type-2 crash.

Finally, congestion induced crash is the scenario when the roadway was congested before the crash, which is defined in this study as Type-3 crash.

A search technique has been developed to identify these three types of crash scenarios. The technique searches whether there is an existence of congestion before and after the crash occurrence, knowing the start time and location of the crash from incident log. The searching process has two windows:

- Time window
- Space window

The time window consists of four time slices (15 min each); if the accident start time is T, then the four-time slice will be T-1, T, T+1, T+2. Again, the space window consists of two immediate upstream sensors (u_1, u_2) of the crash location. Each time-slice & sensor is considered as one cell. Thus, real time speed of the total eight cells will be checked to see if there is any congestion or not. The congestion will be determined by the following rule.

$$\text{Congested if } \frac{Speed\ limit - Current\ Speed}{Speed\ limit} > 25\%$$

At first all the time slices at u_1 will be analyzed, if there was no congestion, then all time slices at u_2 will be analyzed

- If congestion is found at $(T\text{-}1)^{th}$ time period, that would be considered a Type-3 crash
- If congestion is found at T^{th} or $(T+1)^{th}$ or $(T+2)^{th}$ time period, that would be considered a Type-1 crash
- If congestion is not found at any time-slice at any location, that would be considered a Type-2 crash.

Figure 4 shows the different types of crash scenarios. Star mark represents the crash start time and the actual location of the crash. Part (a), (b) & (c) represent the Type-1 crash, where three different scenarios present three different starting point of the impact which are at the crash moment, after small time lag and at the upstream segment respectively. Part (d) represents the Type-2 crash, where is no congestion after the crash. Part (e) & (f) represent a Type-3 crash, where crash occurred in the middle of recurrent congestion. In this way, the step will identify the start time and the location of the starting point of the crash impact.

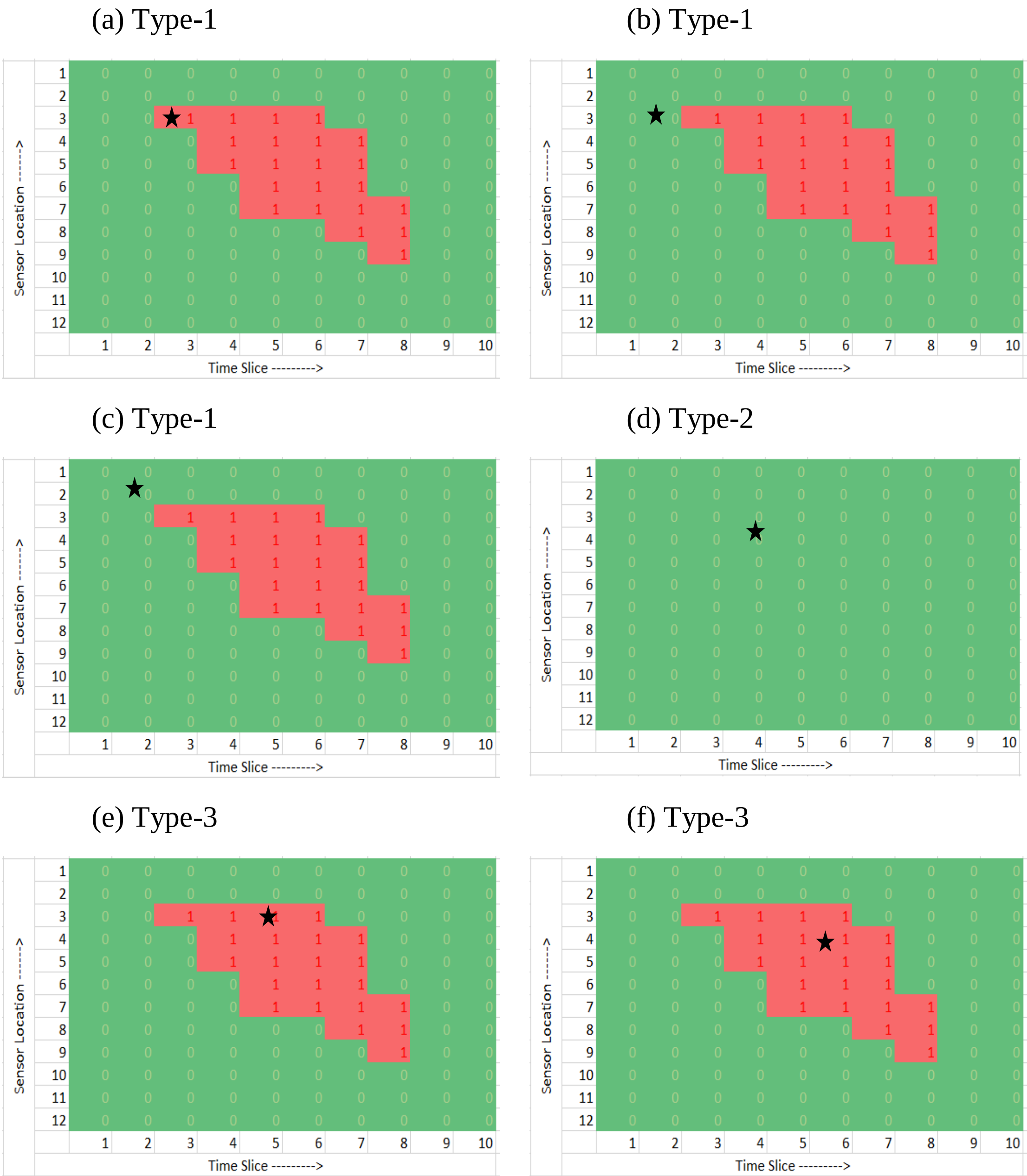


Figure 4: Crash Scenarios

4.2.3 Determining the Background Speed Profile

A background speed profile is a profile which reflects the expected traffic condition throughout the day for a specific location. Before identifying the impact of the crash, it is very important to know the normal traffic/speed condition of that location. As we know, the recurrent congestion is the traveler's expected traffic condition as long as it occurs periodically. On the other hand, non-recurrent congestion is often caused by unexpected incidents or inclement weather condition.

Therefore, the background speed profile should reflect both recurrent congestion and free flow traffic condition as expected by the daily road users. That is why a background speed profile should be constructed based on the traffic data that shares a similar traffic pattern (same period of time, weekend vs weekday etc)

Instead of using the fixed background speed profile, in this research a dynamic profile has been introduced because the pre-crash condition of a specific day may be different from an "average" day. To capture this variation, the dynamic background profile has been created by combining both the ongoing traffic condition and the historical trends. Kalman Filter Algorithm is used to predict this dynamic background speed profile. A Kalman filter is a powerful mathematical tool that can estimate the future states of the variables even without knowing the precise nature of the system modeled [32]. It is a recursive procedure that corrects its estimates whenever new observations become available, with the objective of minimizing the estimated error covariance. Kalman filter has been used widely in various fields of transportation such as forecasting traffic parameters, predicting bus arrival time[33]. As our main interest of this study is to identify the individual impact of each crash, the filter starts creating a background speed profile instantly when the crash occurs.

The filter procedure developed in this study is designed to predict the crash-free normal speed profile based on both historical profile and the pre-crash condition. The historical average speed and the variance of each sensor location for every 15 min interval are used as the inputs of the state predictor in Kalman filter. The filter procedure developed in this study is designed to predict speed of the next time slice knowing the speed of the previous time slice. The whole process of generating the background speed profile is explained in the following:

Now assume, k denotes the pre-crash time slice, k+1 denotes the time slice at crash moment. The term x_k is the historical average speed for the particular sensor, A_k is the ratio of historical average speed of k+1th and kth time slice. r_k is the real time speed

State Prediction : $\hat{X}_{k+1} = A_k x_k + w_k$
Observed Speed : $Z_k = \beta_k r_k + v_k$

Where,

A_k = state transition model which is applied to the previous state
β_k= 1
w_k = white noise associated with the transition process which is assumed to have zero mean and variances of Q_k
v_k = observation noise which is assumed to have zero mean and variances of R_k

The overall filtering process is the recursive prediction process. At the moment crash occurs, formulation of background speed profile will be started using the following process:

- Step 1 : Initialize
  Set k = (T-1); T= Accident start time
- Step 2 : Initialize Observed Speed, Z_k
- Step 3 : Initialize Covariance P_k
- Step 4 : Extrapolate state variable.

$\hat{X}_{k+1} = A_k x_k$

- Step 5 : Extrapolate Covariance
  $\hat{P}_{k+1} = A_k P_k A_k^T + Q_k$
- Step 6 : Compute Kalman Gain
  $K_{k+1} = \hat{P}_{k+1} * \beta_k (\beta_k \hat{P}_{k+1} \beta_k^T + R_k)^{-1}$
- Step 7 : Update State variable
  $X_{k+1} = \hat{X}_{k+1} + K_{k+1}(Z_k - \beta_k \hat{X}_{k+1})$
  Stop if k+1=96, otherwise go to step 8
- Step 8 : Update Covariance
  $P_{k+1} = (1 - \beta_k K_{k+1}) \hat{P}_{k+1}$
- Step 9 : Update Observed Speed
  $Z_{k+1} = \beta_{k+1} X_{k+1}$
- Step 10 : Update Time slice
  k=k+1
  go to Step 2.

This Kalman filter algorithm starts with taking the pre-crash speed as an input to predict the speed of the next time period and this process will continue for all of the upstream sensors and create a background speed profile at each sensor location of the whole corridor. We have used predicted speed of one time slice as the pseudo-observed speed (Step-9) for the prediction of the next time slice. The main motivation to propose this strategy is the fact that our key objective is to get the normal traffic speed pattern to capture the special events (such as crash) from daily traffic. The method captures the normal traffic speed pattern effectively.
Figure 5 represents how ongoing traffic profile and historical trends are combined in generating the background speed profile. The first part shows on a specific day, the pre-crash speed was higher than the historical average speed, so background speed profile has been started above the historical trend. On the other hand, the second part presents that on a particular day the pre-crash speed was lower than the historical average speed, so the background speed profile has been initiated from below the historical trend.

During the calculation of the historical average speed at different time slices, we have found that weekday and weekend settings show a significantly different traffic/speed pattern (Figure 6). So the average speed is calculated separately for weekdays and weekend. If the crash occurred on a weekday, the average speed of the weekday at that location will be considered as the input of background speed profile. For weekend crashes, the average speed of the weekend will be used as the input for the Kalman filter based background speed profile.

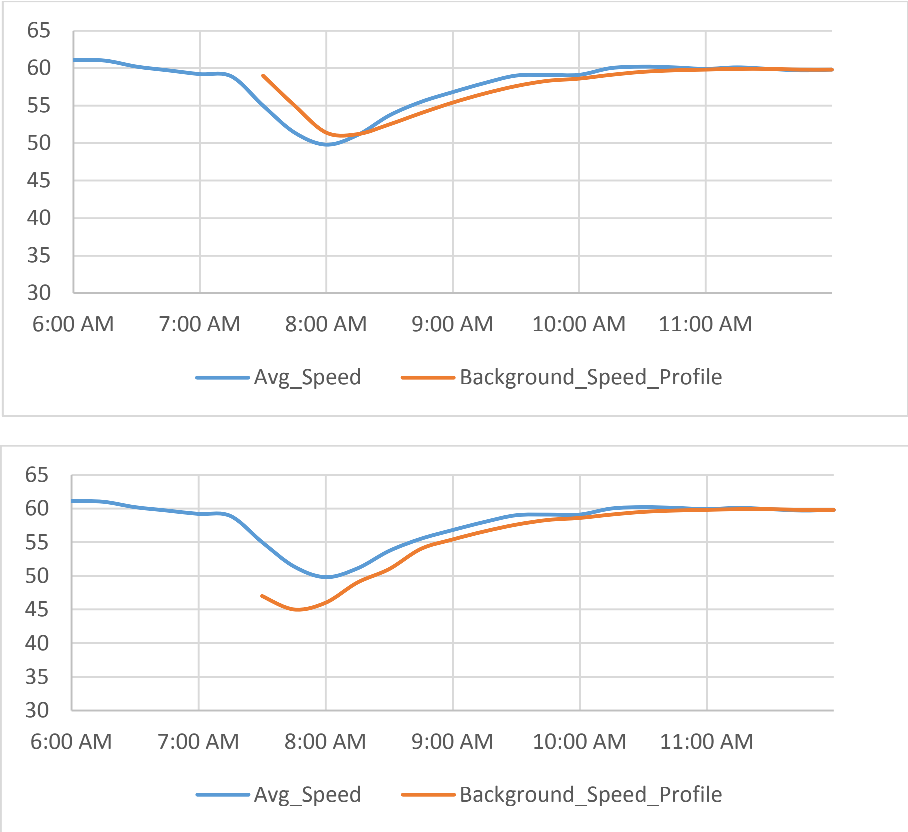


Figure 5: Capturing Dynamics of Traffic by the Background Speed Profile

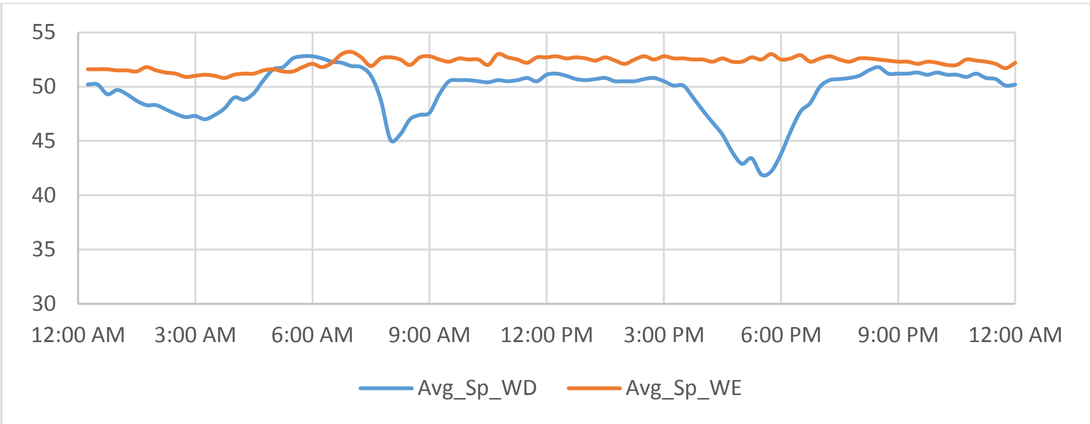


Figure 6: Average Speed Comparison- Weekday vs Weekend

4.2.4 Identifying the Crash Impact from the Difference-in-Speed Profile

By superimposing the current speed profile over the background speed profile, the difference in speed profile can be established and used as the basis for estimating the crash impact. Figure 7 shows the process of separating the crash impact from the background speed profile. The start and end time of the crash impact could be identified visually from the difference in speed profile. In order to automate the process the following formulations are used.

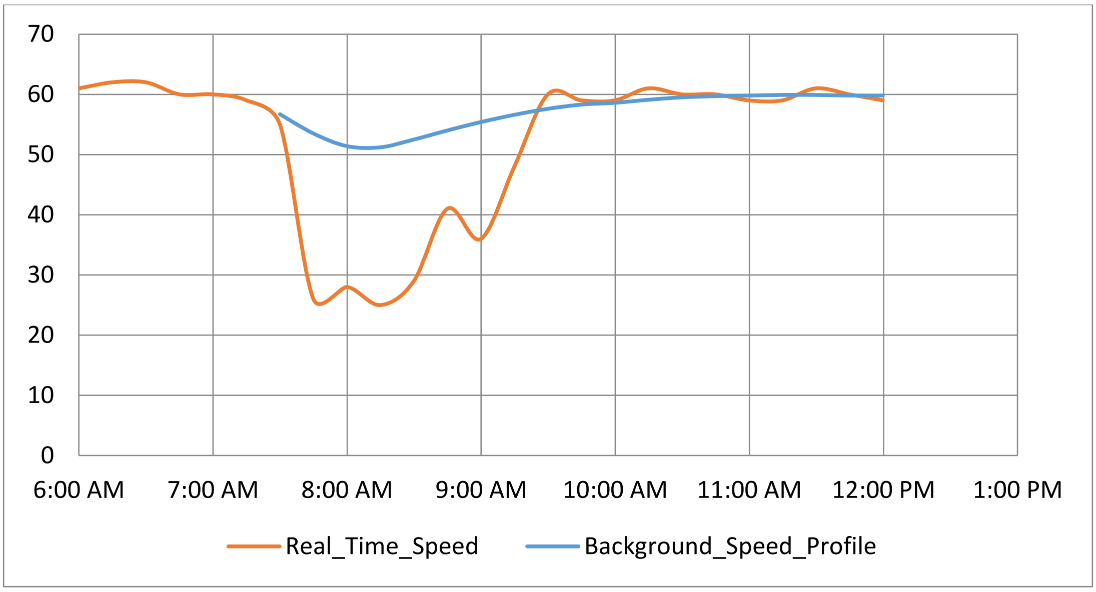


Figure 7: Real Time vs Background Speed Profile

If BSP<CSP,

DS= 0

If BSP>CSP and CSP< (75%*SL)

DS = BSP-CSP

If BSP>CSP and (75%*SL) < CSP < SL

DS = BSP-CSP, when DS $\geq$5 otherwise DS=0

If BSP>CSP and SL< CSP

DS = BSP-CSP, when DS $\geq$10, otherwise DS=0

Where,

BSP = Background Speed Profile (mph)

CSP = Current Speed Profile (mph)

DS = Difference in Speed (mph)

SL = Speed Limit (mph)

This process will continue from accident start time to the remaining portion of the day for all upstream sensors from the location of crash. In this study, all DS value were not taken as the crash impact. Instead, a filtering process has been introduced to filter the noise. In this way, the process adds an empirical tolerance value that specifies a least DS value at different level of current speed which must be achieved by DS to be considered as an impact of the crash.

To visually represent this impact, a contour map has been created, where Y-axis represents the location of different sensors and X-axis represents the different time period. Now this map will show whether the segment is congested during the crash time. If DS is greater than zero, than it is assumed that the corresponding segment is congested. That is why the selection of DS is very important and a filtering process is used to control the noise.

Figure 8 shows an example of the contour map showing the difference in speed profile. This map identifies the area where the crash shockwave propagates. It also captures other events of congestion which are not caused by the crash. Those external events will be excluded in the next step.

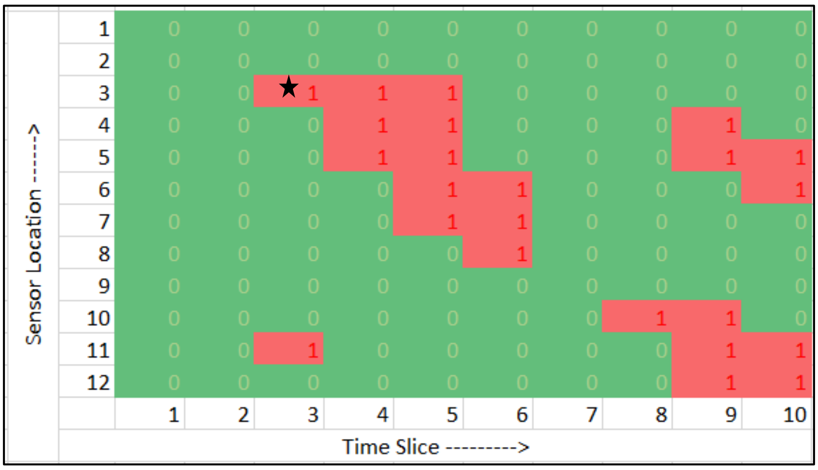


Figure 8: Contour Map of Difference in Speed Profile

4.2.5 Determining the Impacted Region

The final step is to identify the boundary of the spatiotemporal impact of the crash. The following conditions are considered to define the final boundary.

- After the crash, the spatiotemporal progression of the shockwave must be uninterrupted.
- The spatiotemporal boundary of the crash shockwave progression must be at upstream.
- Entire boundary of the impacted region must be contiguous

Fulfilling these conditions, the final boundary is determined. Figure 9 shows some examples of the impossible shape.

- Figure 9 (a) represents the irregular progression of the crash shockwave which is clear violation of our assumption. According to our consideration the spatiotemporal progression of the shockwave must be uninterrupted and it should advance at upstream in a cascading format. So the marked irregular portions should be ignored in the final region.
- There should not be any hole [Figure 9 (b)] in the impact region which is the violation of uninterrupted progression concept. These holes are created when current speed profile crosses the threshold of the background speed profile towards no-congested stage for few moments, then again returns to the congested condition. This occurs due to the highly stochastic nature of the traffic. So when such hole is found which is surrounded by the crash impacted

region, then the hole is considered as a part of the final impacted region. This inclusion will not affect the calculation of total delay due to crash.

- Figure 9 (c) & (d) are the violation of the assumption that the entire boundary of the impacted region must be contiguous. Here we observe the presence of some external events which are not related to the crash, those events are excluded from the final impact region.

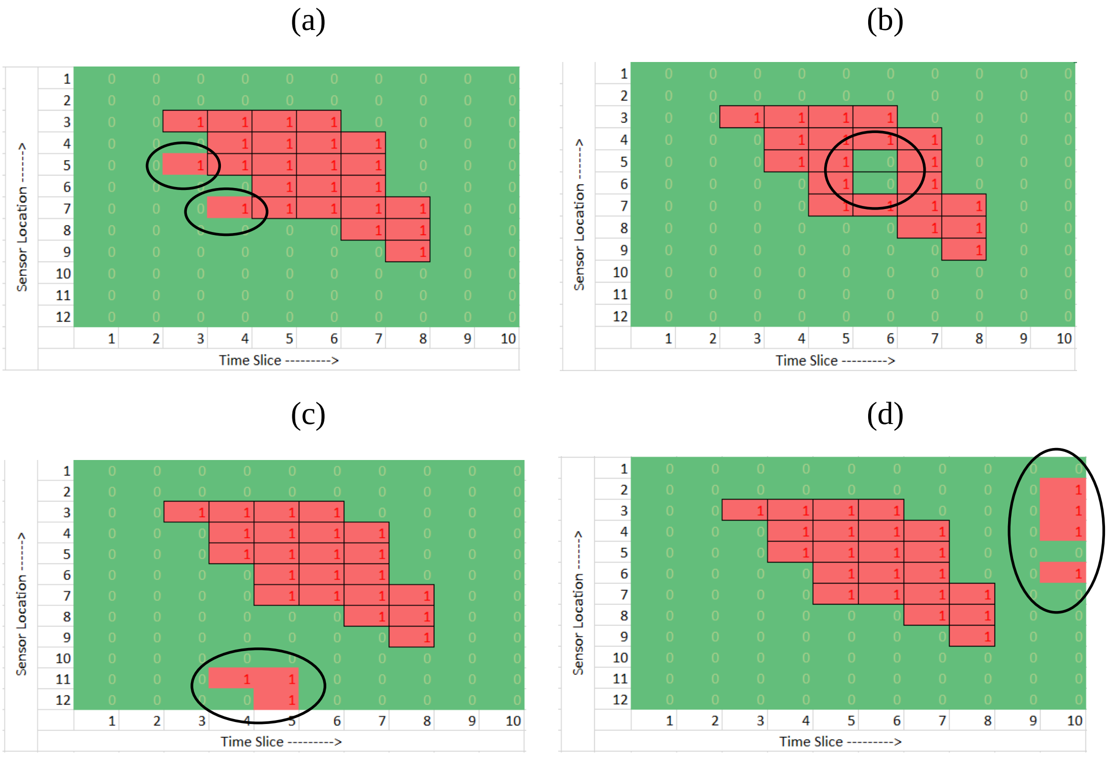


Figure 9: Impossible Shapes

Resolving the impossible shape, we have found the final boundary of the crash impact zone (Figure 10). From the final boundary, temporal and spatial length of the crash impact can be determined knowing the horizontal and vertical length of the boundary respectively. Spatial length represents how far a crash shockwave spreads and temporal length represents how much time a crash disturbs the normal flow of traffic. As each cell represents 15 minutes of time period, the calculated impact duration would be multiples of 15 minutes.

Since spatial aspects of the crash impact are an important issue, that issue has also been addressed using the proposed method. Since spatial length shows the number of upstream sensors affected by crash impact and we know the distance between two sensors is approximately 0.4 miles, the spatial extent of crash can be calculated from this information.

The success in identifying the crash impacts within the dynamic of the traffic environment depends on an accurate representation of background (crash-free) traffic condition at the time of the crash. The proposed method handles this issue by developing the expected normal condition using the Kalman filter algorithm that traveler would anticipate.

Instead of using fixed background profile, here we have used dynamic background profile to capture the dynamic of everyday traffic. That is why filter procedure is designed to predict the speed at the moment crash occurs, based on the input of the pre-crash speed. Thus, a new approach is introduced in this study for estimating background profile by combining both the ongoing traffic condition and the historical trends.

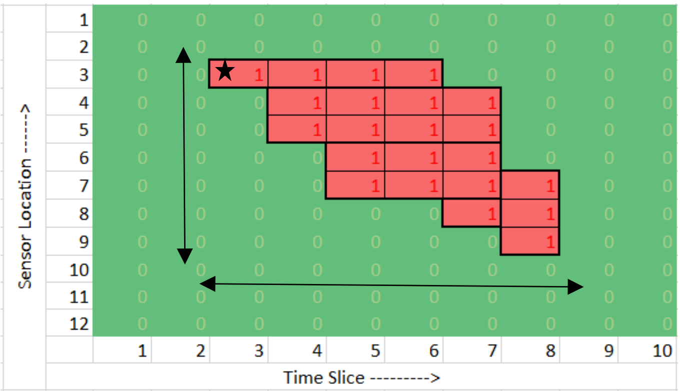


Figure 10: Final Impact Zone

A background profile may not always represent the true crash-free traffic condition at the time of crash. But the main goal is to have a reasonable approach to achieve a crash free normal speed profile that the traveler would anticipate during the time of the crash. As we know, traffic patterns seem to repeat themselves at specific times of the day and days of the week. Kalman filter based background profile could capture those trends in an effective way.

The proposed method assumed that crash will induce congestion and congestion will be identified from the reduction of speed; based on this assumption impact zone is captured. In reality a crash may not immediately induce congestion, so the proposed method easily adjusted this time lag problem by expanding the searching window; instead of limiting the searching time window at the crash occurrence time (T), the method will expand the time window up to the $(T+2)^{th}$ period. In this way, a small time-lag scenario will be included and a large time lag scenario will be excluded.

4.3 Visualizing Impact of the Crash: Case Study

To test the performance of the methodology and to show it visually, a visualization tool was developed in the spreadsheet. The tool creates a space time velocity map (also known as a heat map), where the horizontal axis represents the time of the day and the vertical axis represents the distance, or the location of the sensors/length of the segment. The heat map recorded average speeds on those segments at different times of the day (Described at Step-1).

Using this heat map, the user can input the date and find the traffic speed at different time of day. Since our interest is to identify the impact of the crash, a link between the incident data and the stationary sensor data has been developed. By inputting an incident identifier (Incident ID), incident information and corresponding traffic speed can be generated for that day. A space-time velocity map visualizes traffic data on a defined space-time window.

The visualization tool has three panels:

- 1st panel shows the incident information — when and where the incident happened, how long it lasted, the type of incident, condition, lane blocked information, and other information.
- 2nd panel shows the space time velocity map or heat map for the entire day.
- 3rd panel shows the heat map of the impacted region of the crash, which is identified using the above described methodology.

A narrative description of Figure *11* would read as follows: *On February 20, 2012, at 1:08 pm, an accident occurred at milepoint 134.8 along I-65N. The crash blocked one lane of the freeway, and the accident zone was cleared at 2:19 pm. Traffic was interrupted for 71 minutes. (Panel-1)*

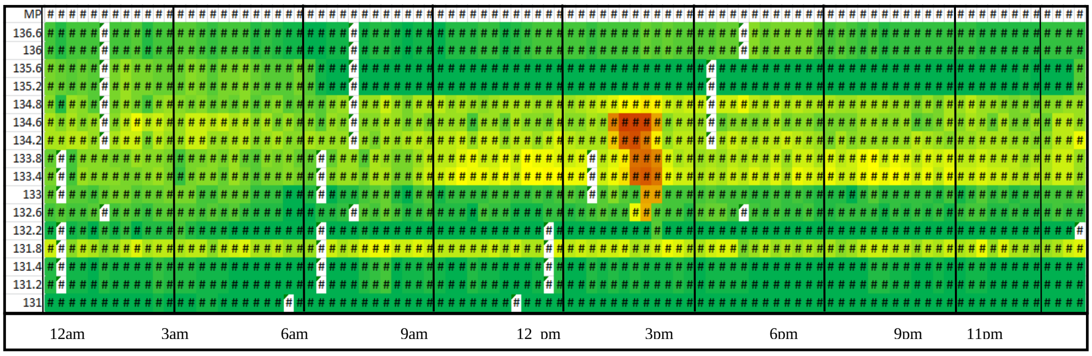
The 2nd panel shows the space-time velocity map of that day. We can see the real time speed of the traffic and significant speed drop during the crash period

The 3rd panel quantifies the spatiotemporal impact of the crash using the five steps crash impact identification method. It represents how long the crash affects the normal traffic flow and how far the impact propagates along the upstream segment. From the Figure *11*, we can see six horizontal cells (each cell equals 15 minutes) are affected by the crash, meaning crash's impact lasted approximately 90 minutes. It also affected six immediate upstream sensors meaning crash affect about (5*0.4) =2 mile upstream segment. In this region, a significant decline in vehicle speeds occurred due to the crash. Furthermore, this is an example of Type-1 crash; because there was no congestion before the crash but the congestion started after the crash occurred.

Crash-induced congestion is one of the major causes of traffic delays. The proposed methodology integrated incident data with the traffic sensor data to provide a data driven approach to quantify the crash induced congestion. Quantifying the crash induced congestion helps to assess various congestion mitigation measures and to monitor road performance. It will provide a greater insight about the crash impacts and guide the traffic management agencies toward improvements in the operation of the road networks.

The visualization tool/heat map makes it very easy to visually identify and understand the spatial temporal extent of the crash events. The case study demonstrated the performance of the methodology by screening all traffic data and identifying the impact zone of the crash.

24206	Incident ID	
2/20/2012	Start Date	#
Monday	Week day	M
	Start Time	1:08 PM
	End Time	2:19 PM
	Total (min)	71
	HWY	I-65
	MP	134.8
	Est. Clear	1 Hours
	Incident Type	Accident
	Lanes Blocked	1 Lane(s) Blocked
	Notes	The left lane is blocked. May be viewed on TRIMARC camera 3.
	Condition	Dry Pavement, Sunny, Possible Injury, Rear-End Collision, Vehicle Damage, Car, Van, SUV



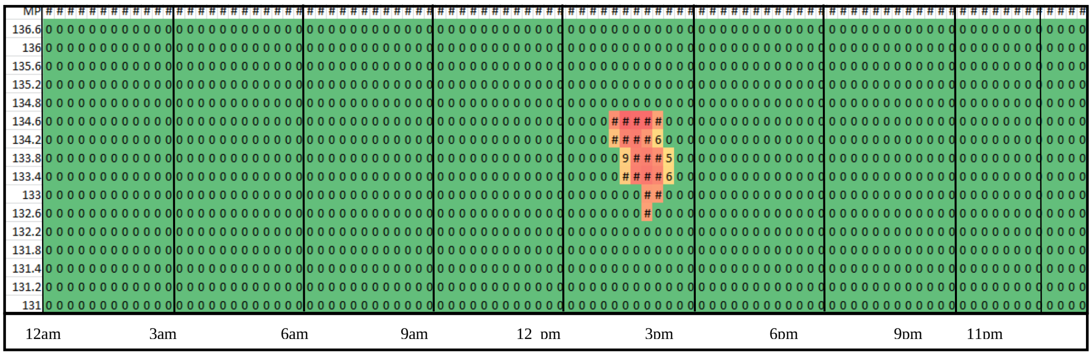


Figure 11: Visualization Tool (Heat Map)

Chapter 5 Applications

5.1 Overview

Identification of the crash impact zone opens a new window for in depth analysis of the actual effect of the crash on traffic flow. In this study temporal impact of the crash (impact duration) and impact delay are analyzed. Temporal impact is the duration of impact which starts since the normal traffic flow has been disrupted after the crash and ends when the normal traffic condition has been resumed. Possible applications of impact duration that are proposed in this study are enhanced incident duration modeling and analysis of the impact of crash on reliability.

Additionally, this study defines impact delay as the additional delay produced by the crash which is calculated by the reduction in segment speed from the background speed profile. Regression analysis has been utilized to build the statistical model for both impact duration and total delay.

Most of the available commercial statistical softwares are able to compute the regression analysis. All of the statistical computation for this research has been done using JMP 10 and R-software package.

A set of variables have been explored to build the statistical model. They are:

Information Types	**Independent Variables**		**State**
Accident Characteristics	Lane Blocking		(Single/Multi/Shoulder)
	Injury		(Yes/No)
	Vehicle Damage		(Yes/No)
	Collision Type		(Rear-end/Side-swipe/Crossing)
	Impact Duration (For Delay Model)		minute
Traffic Characteristics	Average Speed	Before (1,2,3 & 4 u/s segment)	mph
		At (1,2,3 & 4 u/s segment)	mph
		After (1,2,3 & 4 u/s segment)	mph
	Volume	Before crash	V_{phpl} (vehicle per hour per lane)
		At crash	V_{phpl} (vehicle per hour per lane)
		After crash	V_{phpl} (vehicle per hour per lane)
Weather Condition	Rain		(Yes/No)
Time of Day	Time		(peak/off-peak)

5.2 Impact Duration

One of the main objectives of the TIM program is to reduce the impact of the incident/crash on normal traffic flow. The most effective way to do this is to clear the incident scene as quickly as possible. The ability of quickly estimate the impact duration can help highway authorities effectively allocate their emergency resources to minimize the negative effect of incident. Additionally, documenting the impact end time and understanding their properties will allow better crash management strategies in the future. That is why in this study an investigation was conducted to understand the properties of the impact duration and factors affecting them.

The impact duration of a crash can be defined as the time elapsed between the beginning of disruption of normal traffic condition since crash and when the normal traffic condition is resumed. Total impact duration can be divided into following subdivision:

- Detection time (time required to detect the presence of a crash)
- Response time (time between notification of a crash to the incident response team and their arrival to the crash site after being informed)
- Clearance time (time required to clear the crash site)
- Traffic recovery time (time required to resume the normal traffic condition after being cleared the incident)

In our study impact duration are only calculated for Type-1 and Type-3 crashes. Type-2 crashes are not included in the impact duration model because they have no impact on traffic. There is a wide range of methods that could be used to predict the impact duration. In this study three methods have been explored to predict the impact duration

- Multiple linear Regression
- Logistics Regression
- Quantile Regression

5.3 Multiple Linear Regression

Multiple linear regression is a useful method for prediction, variable screening, parameter estimation and system explanation. In this section the regression analysis procedure is briefly described to predict the impact duration. The typical multiple linear regression model can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \ldots\ldots + \beta_n X_n$$

Where, Y is the dependent variable, X_1, X_2 …..X_n are the independent variables and β_0, β_1, β_2 …… β_n are the coefficients.

The unknown coefficients are determined using the least square method. Stepwise selection techniques were used to select the number of independent variables for the regression model. Several regression models were developed and investigated to identify the contributing factors and its contribution to predict the impact duration. JMP statistical software package was used to run the stepwise regression process for the selection of best multiple linear regression models. The criteria used in this study required that all variables added to the regression were statistically significant to a level of 0.05.

However, to gain additional insight in what variables are consistently used in the best regression model, an alternate approach is investigated. This alternate approach selects the best possible subsets of regression model based on the highest r-squared value. This approach did not consider the significance of each variable, rather it checked if the addition of the new variable increased the r-squared value or not. If the r-squared value is increased, then the variable is added, otherwise the variable is ignored. Again it should be noted that this procedure does not necessarily select the best model, rather provides an additional insight into what variables are consistently used in the best models.

5.3.1 Impact Duration Prediction Model

A statistical model for predicting the impact duration has been developed using the above described multiple linear regression procedure and applied to the same incident data-set described previously. The analysis produced the following impact duration prediction model:

$$\ln(Y) = 4.928 - 0.029X_1 - 0.119X_2$$

Alternatively, the model can be written as:

$$Y = e^{(4.928 - 0.029X_1 - 0.119X_2)}$$

Where,

Y = Impact duration in minute;

X_1= Post crash space mean speed in mph of 1st four upstream segment;

X_2= Binary variable for weather condition such as rain (Yes =0, No=1).

The model can predict 32% (R-squared) variation of the impact duration in a natural logarithmic format as a function of two independent variables (post-crash speed and weather condition). No other variables tested either individually or jointly were found to be significant. It is worth mentioning that the p-value for the corresponding F-statistics of overall model is found to be less than 0.05 (F-statistics = 90.37) and the p-value for the parameter of each independent variable is less than 0.05. So there is enough evidence of linear relationship between natural logarithm of impact duration and each independent variable which confirm the model is an adequate predictor for impact duration. Detailed results are provided in Appendix-B.

In the model the dependent variable has been transformed into natural logarithmic format because normal probability plot indicated that impact duration does not follow a normal distribution. An effort has been made by transforming the duration into different functions, but none of them satisfy the normality. However, transformation of impact duration into natural log format shows a better result than the non-transformed format. Since the model will be used as point estimators for the impact duration and not for determining confidence intervals, the normality problem can be ignored [34].

It can be observed that the coefficients of the independent variables of the model are negative. This indicates that shorter impact duration is expected with higher post-crash speed. The result also indicates that the duration of the impact of the crash is expected to be less severe in a no rain condition than in a rainy condition.

Two independent variables (speed and weather condition) are found to be significant in predicting impact duration for the dataset we used. Lane blocking, injury and other independent variables were not found significant for this model. One issue with the incident log is that it does not always keep consistent record of lane-blocking, injury, and other information. An effort has been made to see if this incomplete data set could give any additional insight or not, but it has been found that the incomplete data-set is insignificant in predicting the impact duration for the crash.

A more complete data-set might give additional insight for predicting the crash impact. However current model can predict the impact duration just knowing the post-crash speed at the upstream segment and the weather condition (rainy or not), which can be determined easily from real-time sensor data and weather information.

This is a new way to predict impact duration for the crash; instead of relying on the given incident duration, we have detected the actual start and end time of the impact of the crash and modeled this duration using significant independent variables. This model will help the freeway management authority to predict the actual impact duration of the crash and provide them an idea for what control strategies should be implemented to minimize the traffic congestion and thus improve the freeway performance.

The R-squared value (0.32) for this multiple linear regression model is not very high. One possible reason would be that the impact duration is measured as a multiple of 15 minutes, but predicted duration is in continuous number format which increases the difference between predicted and observed values.

5.4 Logistics Regression

In our analysis a number of independent variables are categorical in nature. If statistical modeling using linear regression would involve a lot of categorical independent variables, then the model would become bulky and inconvenient for further usage. On the other hand, logistic regression regresses the probability of a categorical outcome and is robust with both categorical and continuous independent variables. An attempt has been made to estimate the probability of different classes of impact duration using ordinal logistics regression.

In this section logistics regression methodology is introduced to develop an impact duration model. The general form of logistic regression model that is used in this study is presented in the following section.

5.4.1 Introduction

Logistics regression is a regression model where the dependent variable is categorical and independent variables can be categorical, continuous, or both. Depending on the nature of the response variable, logistics regression could be divided into following subdivision.

1. Binary Logistics Regression: The binary logistic regression model is used where response variable is binary or dichotomous. Here the response variable is taking only two values, occurrence or non-occurrence of a specific event, and is usually coded as Yes or No; 1 or 0.
2. Ordinal Logistics Regression: The ordinal logistics regression model is used where response variable is polytomous and ordered. Here dependent variable is coded as three or more ordered categories, for example low, medium, high.
3. Nominal Logistics Regression: The nominal logistics regression model is used where response variable is polytomous and unordered. Here dependent variable is coded as three or more categories, for example sunny, rainy, cloudy.

In this study impact duration is divided into seven categories. It should be noted that, impact duration is calculated as multiple of 15 minutes after the five-step process because our data comes in 15 minutes interval. Since impact duration is calculated as a multiple of 15 minutes, duration is categorized as 15 minutes or as increments of 15-minutes. The ranges of impact duration are listed in Table 1.

Table 1: Categories of Impact Duration

Categories	Range (min)
A	0-15
B	16-30
C	31-60
D	61-90
E	91-120
F	121-180
G	>180

The model investigates the effect of various factors influencing the impact duration using an ordinal logistics regression approach. Ordinal logistics regression is used when the dependent variable is in a categorical form and has three or more levels with a natural ordering (such as shortest, short, medium, large and largest). In the case of the impact duration model, the response variable also has the category from low to high as described in Table 1.

5.4.2 General Form

The general form of an ordinal logistics regression with K distinct category could be written as:

$$P(y \le k) = \frac{e^{\theta_k + \beta x\prime}}{1 + e^{\theta_k + \beta x\prime}}$$

Where,

K = The number of distinct categories

k = The category number, taking values 1, 2,…., K-1

$P(y \le k)$ = The probability of falling the response into category k or below

θ_k = The constant associated with the k^{th} response category

β = The vector of coefficients associated with the predictor variable

x' = The vector of predictor variables

The regression constants and coefficients are then calculated using a logit link function and by linking the probabilities to a linear combination of the predictor variables which shown below:

$$LOGIT[P(y \le k)] = log_e\left(\frac{P(y \le k)}{1 - P(y \le k)}\right) = \theta_k + \beta x'$$

Using a method equivalent to the maximum likelihood estimation procedure, the coefficients are estimated. Once the coefficients are evaluated, the cumulative probabilities and individual response probabilities can be calculated as follows:

Cumulative probability of the first response category : $P(y \le 1) = \frac{e^{\theta_1 + \beta x\prime}}{1 + e^{\theta_1 + \beta x\prime}}$

Cumulative probability of the second response category : $P(y \le 2) = \frac{e^{\theta_2 + \beta x\prime}}{1 + e^{\theta_2 + \beta x\prime}}$

And so on…

For the last response category, Cumulative probability: $P(y \le K) = 1.0$

Then individual probability of each response category can be calculated from the equation which shown below:

Probability of the first response category: $P(y = 1) = \frac{e^{\theta_1 + \beta x\prime}}{1 + e^{\theta_1 + \beta x\prime}}$

Probability of the second response category: $P(y = 2) = \frac{e^{\theta_2 + \beta x\prime}}{1 + e^{\theta_2 + \beta x\prime}} - \frac{e^{\theta_1 + \beta x\prime}}{1 + e^{\theta_1 + \beta x\prime}}$

The probabilities calculated using above equations depends on the predictor variable pattern x'. Therefore, probabilities of the responses could be calculated by changing the value of predictor variable x'. This helps to find the individual effect of different predictor variables.

5.4.3 Results of the Logistics Regression

At first each independent variable fits with the response variable to check the significance of the independent variable. It is found that the post-crash space mean speed of the first four upstream segments and the weather condition are significant for predicting incident duration, which is concordance with the result of multiple linear regression.

The results of the logistics regression are summarized in the Table 2, which consists of the regression constants and the regression coefficients of the predictor variables (Post-crash speed, weather condition). The regression constants corresponding to the different response categories (θ_1, θ_2, θ_3, $\theta_{4,}$ θ_5, θ_6) are calculated by assuming default factor levels for all the predictor variables. JMP also computes the associated p-values for the regression constants and coefficients. For ascertaining the significance of the predictor variable factors, a 5% significance level is assumed throughout this research. Hence a particular predictor variable factor would be considered statistically significant, if the corresponding p-value was found less than 0.05.

Table 2: Logistic Regression Results

Predictor Variable	Factors	Regression-Coefficient(β)	p-value
Const(A) -- θ_1		-5.084	<0.0001*
Const(B) -- θ_2		-3.814	<0.0001*
Const(C) -- θ_3		-2.26	<0.0001*
Const(D) -- θ_4		-1.21	<0.0001*
Const(E) -- θ_5		-0.484	0.0232*
Const(F) -- θ_6		0.605	0.0129*
Post-crash Speed		0.0749	<0.0001*
Weather Condition	Default: No Rain		n/a
	Rain (Y)	-0.3363	0.0011*

Based on the regression constants and the coefficients obtained by the regression analysis, the probabilities of different responses (impact duration class) can be calculated for any predictor variable scenario. For any particular predictor variable situation, the response probabilities can be calculated by using the corresponding regression constant (θ, depending on the response probability) and appropriate regression coefficients (β, depending on the predictor variable).

The effect of each predictor variable on the responses can be studied by changing the regression coefficients of that variable only and holding all the other coefficients unchanged. A positive regression coefficient indicates reduction in the impact duration time due to the corresponding factor. On the other hand, a negative regression coefficient indicates an increment in the impact duration time.

5.4.4 Case Study and Discussion

An example of calculating the probability of each response of the impact duration based on logistics regression results is presented in the following section. According to logistics regression result, we have found two significant independent variables. Between them, one variable is categorical (weather condition) and the other is continuous (post-crash speed) in nature.

Post-crash speed has a positive coefficient which means with the increase of speed, impact duration will be decreased. On the other hand, weather condition has negative coefficient which indicates impact duration will be increased compare to the default weather condition (No rain)

Since a continuous variable is present in the model, there is no absolute base condition for this model. Absolute base condition is achieved when all independent variables become categorical. Therefore, the response probabilities are calculated assuming a random speed value and weather condition (Rain or No rain).

Now sample calculation has been shown for Speed = 10 mph and weather = No rain condition. Based on the logistics regression result, response probabilities can be calculated as follows:

Probability of impact duration being "A_(0-15 min)":

$$= \frac{e^{\theta_1+\beta x\prime}}{1+e^{\theta_1+\beta x\prime}} = \frac{e^{-5.084+0.0749*10}}{1+e^{-5.084+0.0749*10}} = 0.01$$

Probability of impact duration being "B_(16-30min)":

$$= \frac{e^{\theta_2+\beta x\prime}}{1+e^{\theta_2+\beta x\prime}} - \frac{e^{\theta_1+\beta x'}}{1+e^{\theta_1+\beta x'}} = \frac{e^{-3.814+0.0749*10}}{1+e^{-3.814+0.0749*10}} - \frac{e^{-5.084+0.0749*10}}{1+e^{-5.084+0.0749*10}}$$
$$= 0.03$$

Similarly

Probability of impact duration being "C_(31-60 min)" = 0.14

Probability of impact duration being "D_(61-90 min)" = 0.21

Probability of impact duration being "E_(91-120 min)" = 0.18

Probability of impact duration being "F_(121-180 min)" = 0.23

Probability of impact duration being "G_(>180 min)" = 0.20

Again response probabilities for the Speed = 10 mph and Weather = Rainy condition can be calculated as follows:

Probability of impact duration being "A_(0-15 min)":

$$= \frac{e^{\theta_1+\beta x\prime}}{1+e^{\theta_1+\beta x\prime}} = \frac{e^{-5.084+0.0749*10-0.3363}}{1+e^{-5.084+0.0749*10-0.3363}} = 0.009$$

Probability of impact duration being “B_(16-30 min)”:

$$= \frac{e^{\theta_2+\beta x'}}{1+e^{\theta_2+\beta x'}} - \frac{e^{\theta_1+\beta x'}}{1+e^{\theta_1+\beta x'}}$$
$$= \frac{e^{-3.814+0.0749*10-0.3363}}{1+e^{-3.814+0.0749*10-0.3363}} - \frac{e^{-5.084+0.0749*10-0.3363}}{1+e^{-5.084+0.0749*10-0.3363}} = 0.02$$

Similarly

Probability of impact duration being “C_(31-60 min)” = 0.10

Probability of impact duration being “D_(61-90 min)” = 0.17

Probability of impact duration being “E_(91-120 min)” = 0.17

Probability of impact duration being “F_(121-180 min)” = 0.25

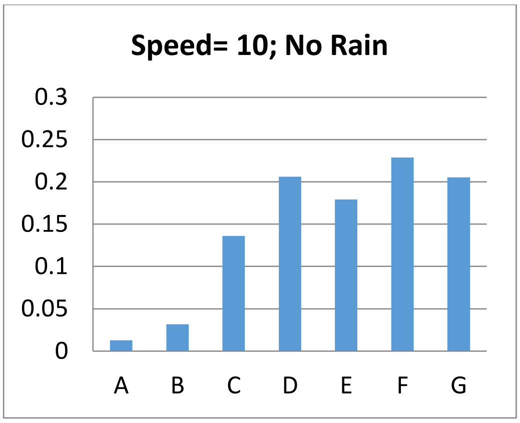
Probability of impact duration being “G_(>181 min)” = 0.26

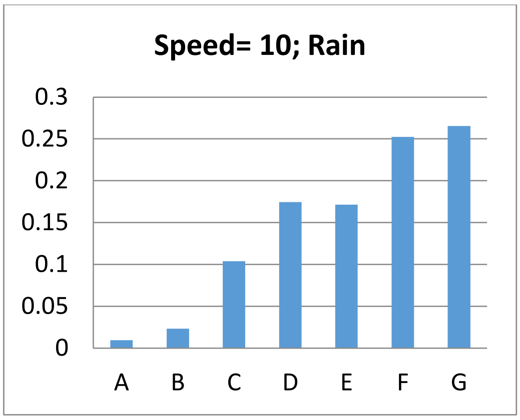
Examining the above probability values, we can say that at 10 mph post-crash speed and with the no rain condition, the impact duration has the highest probability to lie in the “F” class which means the crash impact would have lasted 120-180 minutes. On the other hand, at the same speed level with the rain condition, the impact duration has the highest probability to lie in the “G” class which means the impact would last more than 180 minutes. It is observable that during the rain condition, probability of higher impact duration increased compared to the no rain condition.

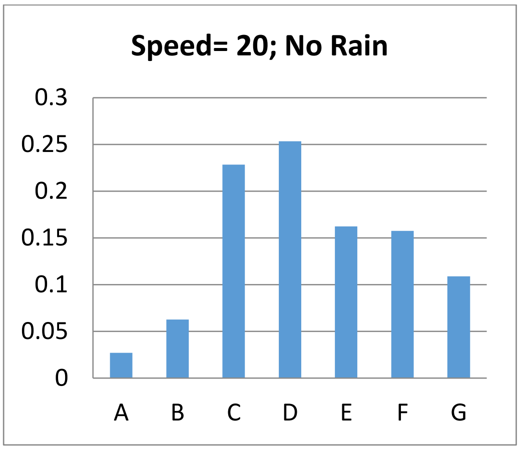
To further investigate this phenomena, individual probability of responses is calculated at different post-crash speed level (10,20,30,40,50,60 mph) for both the rain and the no rain condition and presents this result in Figure 12.

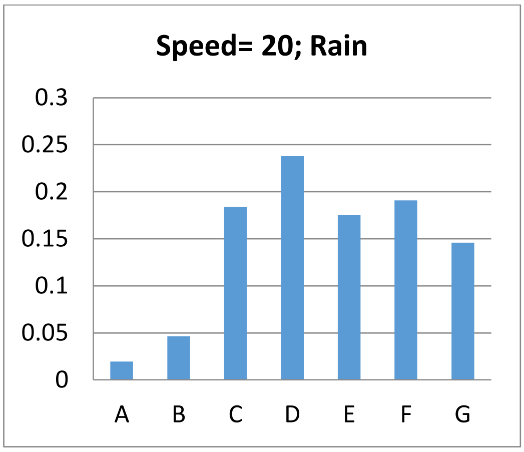
From Figure 12, it is observable that, with the increase of speed, probability of lower impact duration increased for both the no rain and the rain condition. There is sufficient evidence from the data showing this behavior as indicated by the fact that the coefficients are significant at 5% confidence level.

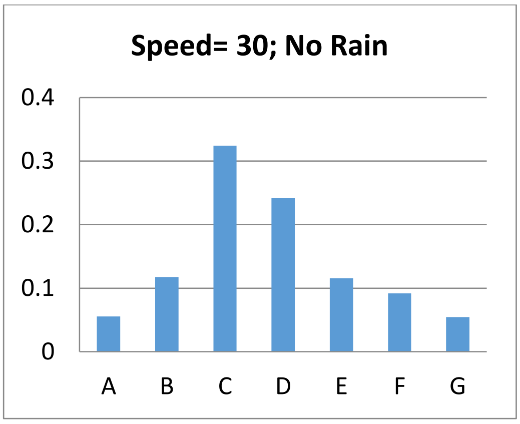
It should be noted that the proposed model provides a framework to make reasonable judgement about the impact duration due to crash. Moreover, the model can predict the impact duration by using the post-crash speed at the upstream segment and the weather condition (rainy or not) which can be determined easily from real-time sensor data and weather information. This information would be very helpful for improving freeway incident management and decision making.

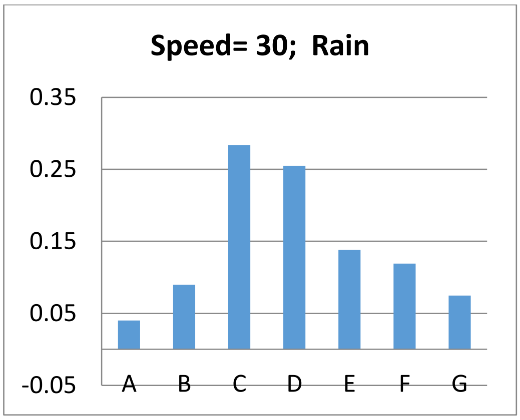

Speed= 10; No Rain
0.3
0.25
0.2
0.15
0.1
0.05
0
A B C D E F G

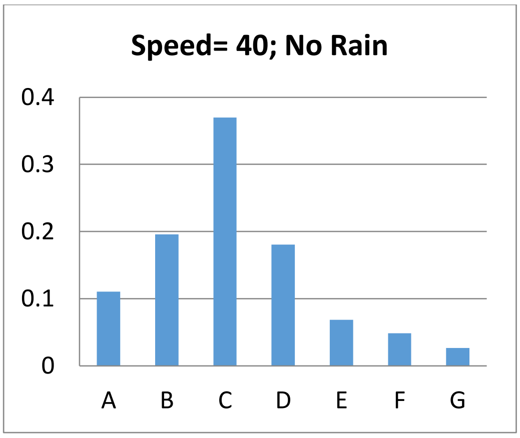

Speed= 10; Rain
0.3
0.25
0.2
0.15
0.1
0.05
0
A B C D E F G

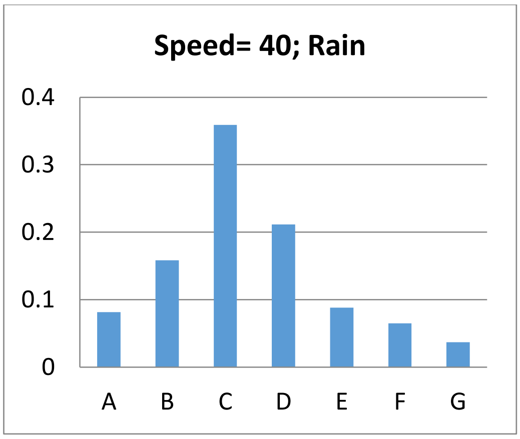

Speed= 20; No Rain
0.3
0.25
0.2
0.15
0.1
0.05
0
A B C D E F G


Speed= 20; Rain
0.3
0.25
0.2
0.15
0.1
0.05
0
A B C D E F G


Speed= 30; No Rain
0.4
0.3
0.2
0.1
0
A B C D E F G


Speed= 30; Rain
0.35
0.25
0.15
0.05
-0.05
A B C D E F G


Speed= 40; No Rain
0.4
0.3
0.2
0.1
0
A B C D E F G


Speed= 40; Rain
0.4
0.3
0.2
0.1
0
A B C D E F G

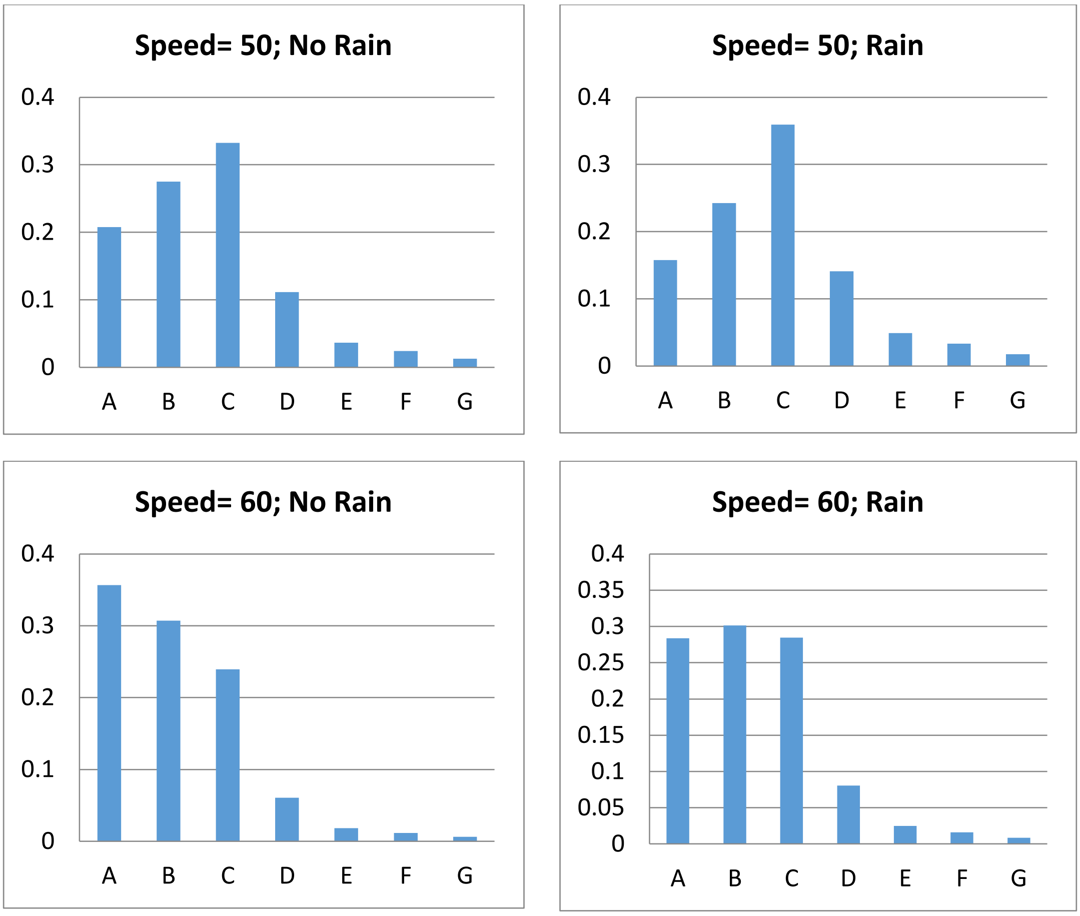


Figure 12: Response Probabilities for Impact Duration at Different Speed and Weather Condition

5.5 Quantile Regression

OLS model only represents the average relationship between the response and the explanatory variables. But quantile regression method enables us to explore potential effects on the shape of the distribution. In this study, one of our objectives is to understand the properties of the impact duration and the factors affecting them. It is found that the duration does not follow the normal distribution and the distribution is skewed to the right, which represents heterogeneity. Therefore, OLS model for impact duration would only convey a partial picture on the whole distribution and also violates the basic homoscedasticity assumption. On the other hand, quantile regression relaxes such assumptions and can be applied to estimate the relationship between any part of the distribution of the response and the explanatory variable. That's why the quantile regression model is tested more suitable to quantify the effects of the explanatory variables and how the effects are different across the distribution.

In this section, quantile regression methodology is introduced to develop an impact duration model. First, a brief description of quantile regression is presented.

5.5.1 General Form of Quantile Function

Quantile regression is a way of estimating functional relationships between the variables for all portions of a probability distribution [35]. It estimates the conditional quantiles of a dependent variable distribution in the linear model that provides a more complete picture of causal relationships between variables. The advantage of quantile regression compared to the ordinary least square (OLS) regression is that quantile regression estimates are more robust against the outliers.

Now suppose for a random variable Y, the cumulative distribution function is F(y), where $(y) = P(Y \leq y)$, then the τ^{th} quantile or percentile function of Y would be,

$Q(\tau) = F^{-1}(\tau)$, where $0 \leq \tau \leq 1$

$Q(\tau)$ can be mathematically formulated as

$$Q(\tau) = \beta_0(\tau) + X_1\beta_1(\tau) + X_2\beta_2(\tau) \ldots. + X_n\beta_n(\tau)$$

where $\beta_0(\tau)$ is the intercept and $\beta_1(\tau)$, $\beta_2(\tau)$...... $\beta_n(\tau)$ represent the coefficients of the explanatory variable at the τ^{th} quantile or percentile.

The above equation can be estimated by solving the following minimization problem correspondingly.

$$\arg min \sum_{i=1}^{m} \rho_\tau \{y_i - (\beta_0(\tau) + X_1\beta_1(\tau) + X_2\beta_2(\tau) \ldots. + X_n\beta_n(\tau))\} \ldots\ldots\ldots\ldots(1)$$

Where ρ_τ is the loss function and can be defined by

$$\rho_\tau(\alpha) = \begin{cases} (\tau - 1).\alpha\,; & \alpha < 0 \\ \tau.\alpha\,; & \alpha \geq 0 \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots\ldots..(2)$$

Where, $\alpha = y_i - (\beta_0(\tau) + X_1\beta_1(\tau) + X_2\beta_2(\tau) \ldots. + X_n\beta_n(\tau))$

Equation (1) and (2) can be reformulated into a standard linear programming problem, which can be easily solved with the simplex method.

5.5.2 Quantile Regression Model for Impact Duration

Using the above described methodology, quantile regression model was developed to predict the duration of the impact of a crash. From the experience of OLS method, it has been recognized that the resulting estimates of various effects on the conditional mean of impact duration were not indicative of the size and nature of these effects on the upper tail of the impact duration distribution. A more complete picture of the variable effects can be presented by estimating a family of quantile regression functions.

The impact of different significant independent variables on the whole distribution of the duration is shown in Figure 13. The x-axis represents the percentiles of interest, ranging from the 5th to the 95th percentile. The y-axis represents the independent variable effect in minutes. The solid red line represents the conditional mean outputted by the OLS method, while the dashed red lines show the conventional 95 percent confidence interval of the mean. Meanwhile, the dash dotted black line represents the percentile values and the shaded gray area shows a 95 percent pointwise confidence band for the quantile regression estimates.

It is found that three independent variables (space mean speed of two upstream station at crash moment, injury and weather condition) significantly affect the impact duration. Figure 13 shows the summary of the quantile regression results where we have three independent variables and an intercept. For each of the four coefficients, we plot 19 distinct quartile regression estimates for τ ranging from the 5th to the 95th percentile. For each variable, these point estimates can be interpreted as the impact of one unit change of the variable on duration, when the other variables remain unchanged.

The intercept of the model is the estimated value when all independent variables are zero; it may be interpreted as the estimated quantile function for the duration distribution of a crash event with no injury, no rain, and zero space mean speed of two upstream stations at the crash moment. For example, at the 50th percentile the intercept was 104.73 minutes, again at the 75th percentile the intercept became 160.88 minutes which represents that if a crash occurred with no injury, no rain, and the space mean speed of two upstream stations was zero, then the 75th percentile impact duration would be 160.88 minutes.

Now we will discuss the effect of the independent variables on the impact duration and tell a story focused on insights, not just the data. At any chosen quantile, we might ask how does change in speed affect the duration? The second panel (Figure 13) answers the question. For example, according to the 50th percentile model, per unit (mph) increase of speed will decrease 1.85 minutes duration time which means 10mph speed gain will more likely reduce the duration by 18.5 minutes; on the other hand, 10mph speed reduction will add 18.5 more minutes in the 50th percentile impact duration if other variables remain unchanged. Similarly, according to the 75th percentile model, 10mph speed gain will deduct the 75th percentile duration by about 28 minutes and vice versa.

The other two variables are rain and injury, an interesting observation was found during quantile analysis with these variables. It is found that, at lower quartile, rain and injury are not significant variables to predict duration. Rain became a significant independent variable at the 40th percentile and remained significant at the upper tail of the duration. But injury was found as a significant variable only at upper quantile (75th percentile and higher quantile) and positive coefficient reveals that crash with injury significantly affects the impact duration and yield higher duration. Similarly, presence of rain also increases the crash impact and yields higher duration.

At the higher quantile both variables (injury and rain) are significant, which reveals the insight that crashes with higher duration are more likely to occur with the presence of rain and injury. Based on the location of OLS line and quantile regression lines, it is observed that OLS overestimates the effect of rain and injury at lower quantile (above gray area) while underestimates the effect of those variables at higher quantile (Figure 13), because OLS model only addresses the average relationship between the variables not the influence of the variables at different percentile. Moreover, OLS model can only address the question, “is rain/injury important in predicting impact duration?” But it cannot answer the question “does rain/injury influence the impact duration differently for crashes with higher duration than those with average duration?” A more comprehensive picture of the effect of the predictors is explained by the quantile regression; for example, the 90th percentile duration of crashes with rain is 36.37 minutes higher than crashes with no rain. According to OLS model, it is 17.54 minutes higher, thus OLS model underestimates the effect of rain at longer duration.

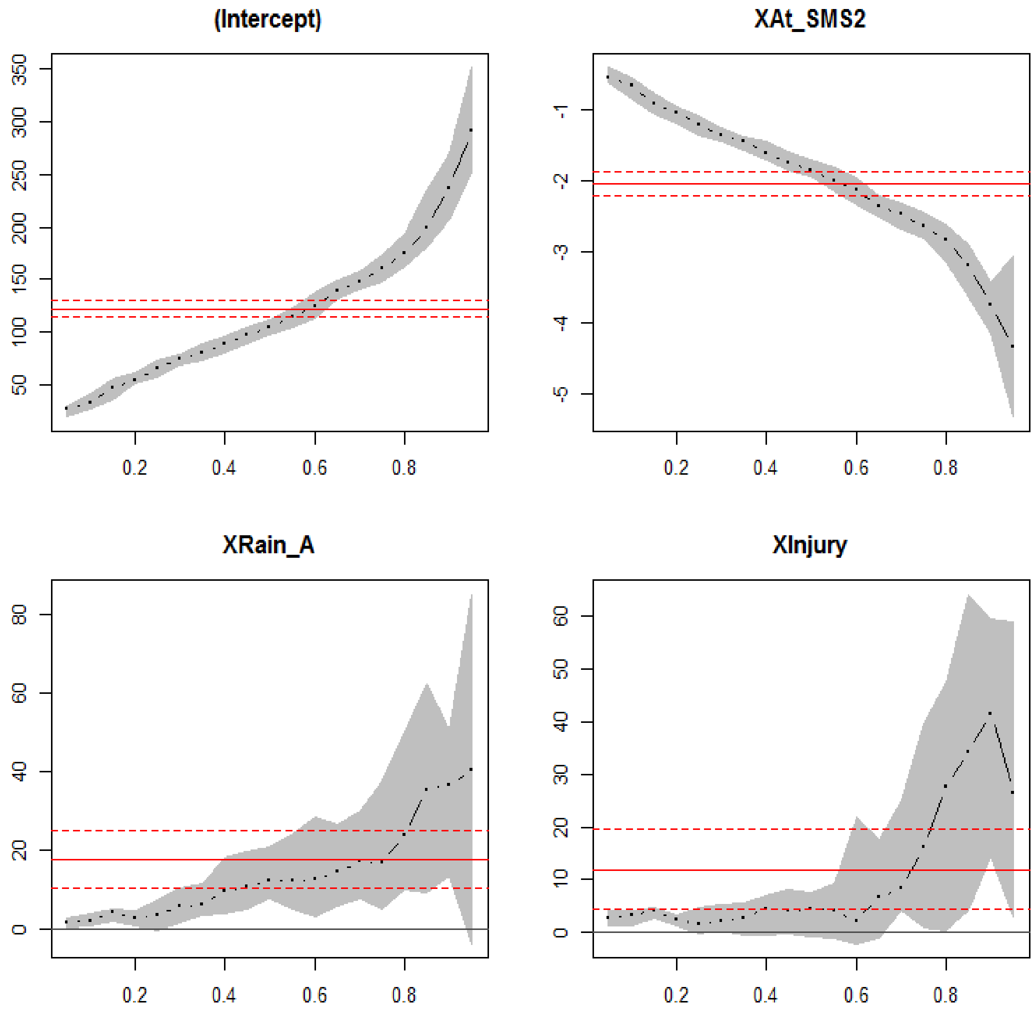


Figure 13: Quantile Regression Result

In this way, quantile regression provides a detailed picture of the duration of impact. Instead of giving a single output, it provides a range of values for best possible to worse possible scenario of a crash in a form of duration. Quantile regression also provides insights about the different variable’s effect at the different portions of the duration distribution. It identifies the variables which trigger longer impact duration. This information will help the agency to determine the work plan to minimize the impact of the crash. In our study we have found a crash with injury would have added 4 minutes (50th per.) to 26 more minutes (95th per.) in total impact duration if other variables remain unchanged.

In this way, analysis of different variables at different percentile would provide additional insights about the crash and a better solution could be made to reduce larger impact duration.

Table 3 provides the quantile regression results for predicting the impact duration. It shows OLS, 25^{th} percentile, 50^{th} percentile, 75^{th} percentile and 95^{th} percentile regression results in a tabular format and for different values of explanatory variables it shows resulting impact duration. The table contains four sections; 1^{st} section shows the resulting duration for a crash at different quantile when space mean speed of two

Table 3: Impact Duration Prediction

Variable	X	OLS(Mean)		25th		Median		75th		95th	
		β	β*X	β	β*X	β	β*X	β	β*X	β	β*X
constant		122.58	122.58	65.54	65.54	104.73	104.73	160.88	160.88	290.77	290.77
At_SMS2	30	-2.04	-61.2	-1.22	-36.6	-1.85	-55.5	-2.81	-84.3	-4.35	-130.5
Rain	0	17.54	0	3.16	0	12.05	0	16.59	0	40.4	0
Injury	0	11.95	0	1.46	0	4.39	0	16.06	0	26.41	0
Total Duration			61.38		28.94		49.23		76.58		160.27

Variable	X	OLS(Mean)		25th		Median		75th		95th	
		β	β*X	β	β*X	β	β*X	β	β*X	β	β*X
constant		122.58	122.58	65.54	65.54	104.73	104.73	160.88	160.88	290.77	290.77
At_SMS2	30	-2.04	-61.2	-1.22	-36.6	-1.85	-55.5	-2.81	-84.3	-4.35	-130.5
Rain	1	17.54	17.54	3.16	3.16	12.05	12.05	16.59	16.59	40.4	40.4
Injury	0	11.95	0	1.46	0	4.39	0	16.06	0	26.41	0
Total Duration			78.92		32.1		61.28		93.17		200.67

Variable	X	OLS(Mean)		25th		Median		75th		95th	
		β	β*X	β	β*X	β	β*X	β	β*X	β	β*X
constant		122.58	122.58	65.54	65.54	104.73	104.73	160.88	160.88	290.77	290.77
At_SMS2	30	-2.04	-61.2	-1.22	-36.6	-1.85	-55.5	-2.81	-84.3	-4.35	-130.5
Rain	0	17.54	0	3.16	0	12.05	0	16.59	0	40.4	0
Injury	1	11.95	11.95	1.46	1.46	4.39	4.39	16.06	16.06	26.41	26.41
Total Duration			73.33		30.4		53.62		92.64		186.68

Variable	X	OLS(Mean)		25th		Median		75th		95th	
		β	β*X	β	β*X	β	β*X	β	β*X	β	β*X
constant		122.58	122.58	65.54	65.54	104.73	104.73	160.88	160.88	290.77	290.77
At_SMS2	30	-2.04	-61.2	-1.22	-36.6	-1.85	-55.5	-2.81	-84.3	-4.35	-130.5
Rain	1	17.54	17.54	3.16	3.16	12.05	12.05	16.59	16.59	40.4	40.4
Injury	1	11.95	11.95	1.46	1.46	4.39	4.39	16.06	16.06	26.41	26.41
Total Duration			90.87		33.56		65.67		109.23		227.08

upstream stations at crash moment was 30 mph without injury and rain, 2nd section shows the resulting duration for same speed, without injury but rainy condition; 3rd section shows the result for same speed, without rain but a crash with injury and 4th section shows resulting duration with same speed with injury and rainy condition. This table demonstrates how the resulting duration changes at different quantiles with the change of explanatory variables.

Table 4 provides a chart of prediction error (absolute difference between predicted and measured duration) at different quantile. As our measured impact duration is calculated at 15 minute increments, a prediction error less than 15 minutes was selected to compare the performance of different quantile models. It has been observed that, at the 25th percentile, about 55.5% of crashes have been predicted with less than 15 minutes prediction error; at the 50th and the 75th percentile about 49% and 39% of crashes have been predicted with less than 15 minutes prediction error respectively. So the 25th percentile model could be used to calculate least projected duration and the 75th percentile model could be used to determine maximum projected duration. The 95th percentile model might overestimate the actual impact, so the 75th percentile model is proposed to project the maximum impact duration.

From the incident management point of view, crashes with longer duration are more critical. Quantile regression analysis enables the agency to identify the causes and measure the effect of contributing variables for such longer duration. This information will help in making plans to reduce the longer duration. Moreover, accurate prediction of impact duration of crashes will help traffic management centers to avoid crash induced congestion by implementing several strategies such as route diversion, informing travelers in advance about crashes and it's impact duration etc.

Table 4: Comparison of Percentage of Samples at Different Prediction Tolerances

Prediction Error (min)	OLS	25th	50th	75th	95th
<=5	13.1	39	29.8	20.4	1.9
<=10	25.1	49.9	40.1	32	2.9
<=15	**39.8**	**55.5**	**49.2**	**39.2**	**3.9**
<=30	64.5	66.4	70.4	59.3	11.3
<=60	85.8	82.9	86.7	80.5	37.4

In this section, three regression methods have been implemented to predict the impact duration. Among them quantile regression provides a detailed picture of the crash in form of duration, instead of providing a single/mean value. The result from the quantile regression shows that effect of rain and injury at higher percentiles are more significant and average relationship underestimates such effect. This result can be conveyed to the travelers to assist their trip planning during crash situations and effective response strategies could be implemented by the transportation agencies such as route diversion, quickly clearing the crash site etc. to minimize the impact of crashes on the traffic flow.

5.6 Impact Delay Model

In this section, delay due to crash will be investigated. The procedures developed in this study will be useful for the performance evaluation of accident management systems by quantifying congestion due to crash in terms of the total delay to evaluate the benefit of accident management systems. This study defines impact delay as the additional delay produced by the crash, which is calculated by the reduction in segment speed from background speed profile. The crash impact zone is found by completing the five step procedure. Then the delay is calculated for the whole crash impact zone which is negatively affected by the crash using the following equation:

$$TD = \sum_{\forall\, i,j \in\, impacted-cells} max\{L_j(\frac{1}{S_{ij}} - \frac{1}{\hat{S}_{ij}})V_{ij}\,, 0\}$$

Where
TD = Total delay due to crash impact (veh-hr)
L_j = Length of freeway segment *j* (in mile);
V_{ij} = Volume of traffic at j^{th} segment during time slice *i*.
S_{ij} = Speed affected by crash at j^{th} segment during time slice *i* (in mph).
$\hat{S}_{ij}$ = Predicted Speed using Kalman Filter at j^{th} segment during time slice *i* (in mph).

Then a statistical prediction model for predicting the total impact delay has been developed using the previously described multiple linear regression procedure and applied to the same incident data-set. The analysis produced the following impact delay prediction model:

$$\ln(Y_1) = 4.97 - 0.06X_1 + 0.013X_2$$

Alternatively, it can be written as:

$$Y_1 = e^{(4.97 - 0.06X_1 + 0.013X_2)}$$

Where,

Y_1 = Total impact delay (in veh-hr)

X_1= Post-crash space mean speed (in mph) of 1st four upstream segment

X_2= Impact duration in minute

The model can predict 74% (R-squared) variation of total delay in a natural logarithmic format as a function of two independent variables (post-crash speed and impact duration). No other variables either tested individually or jointly were found to be significant. It is worth mentioning that p-value for the corresponding F-statistics of overall model is found to be less than 0.05 (F-statistics = 2.13) and p-value for the parameter of each independent variable is less than 0.05. There is enough evidence of the linear relationship between natural logarithm of total delay and each independent variable which confirm the model is an adequate predictor for total delay due to the crash impact.

Post-crash speed and impact duration are found significant in predicting impact delay for the dataset we used. It can be observed that the coefficient of the post-crash speed of the model is negative. This indicates that the total delay will decrease with the increase of post-crash speed. On the other hand, coefficient of the impact duration is positive which indicates that total delay of the crash will be increased with the increase of impact duration.

In the model, the dependent variable has been transformed into natural logarithmic format because the normal probability plot indicated that total delay does not follow a normal distribution. Although transformation of dependent variable does not show pure normality, but it yields better model than non-transformed form. Since the model will be used as point estimators for the total delay, not for determining confidence intervals, the normality problem can be ignored. [34]

The current model can predict the total delay just knowing the post-crash speed at the upstream segment and the impact duration. The post-crash speed can be determined easily from real-time sensor data and impact duration can be achieved from the impact duration prediction model. If impact duration was not available, then we can assume a different impact duration value and input it in the model, and we will get an idea about the trend of total delay. This model will help the freeway management authority to predict total delay of the crash and give them an idea of what control strategies should be implemented to minimize traffic congestion and delay which will improve freeway performance.

5.7 Impact on Reliability

Impact on reliability analysis enables us to understand how different factors affect the travel experience of the users. In this study, travel rate (seconds/mile) was treated as a measure of effectiveness to understand how crashes affect the travel rate. Travel rate can be defined as the time required for traveling per unit distance (ex.-mile). This analysis can guide the agencies toward improvements in the operation of road networks. For example, when an agency experiences unreliable travel times because of incidents (crashes), the agency may increase its spending on incident management systems and safety improvements. Conducting the impact on reliability analysis required the following steps:

- Select the region or facilities of interest and study period
- Compile the travel rate data for each facility
- Identify what types of nonrecurring events (peak/off-peak crash, different types of crash etc.) are present in the data
- Develop cumulative distribution functions (CDFs) of the travel rate (TR) for each combination of nonrecurring events

In this study, we analyzed the impact of crashes on I-65 NB and SB corridor. First, we calculated the daily travel rate (sec/mile) (based on TRIMARC speed data of 15-minutes intervals) of the study segment for the year 2011 to 2013. Travel rates are then separated into two groups: peak period and off-peak period travel rate. The morning and afternoon peak periods occurred between 6 AM and 9 AM and 4 PM and 7 PM, respectively.

Our analysis was based on the crashes in the study segment. The impact duration which was determined after the five step process has been used to select the time slices impacted by crash. Peak travel rates were grouped into two classes: one containing the 15- minute travel rates that were influenced by crashes (termed “peak, crash”) and the other travel rates during periods that were unaffected by crashes (“peak, no crash”). Off-peak travel rates were also separated into two groups: “off-peak, no crash” and “off-peak, crash”.

As Figure 14 indicates, during a crash, the travel rate increases significantly over to the non-crash condition. Crashes during peak periods increase the travel rate (which means a higher delay) for both directions on the interstate. Figure 15 and Figure 16 depict a CDF of travel rate, which tell a better story about route performance.

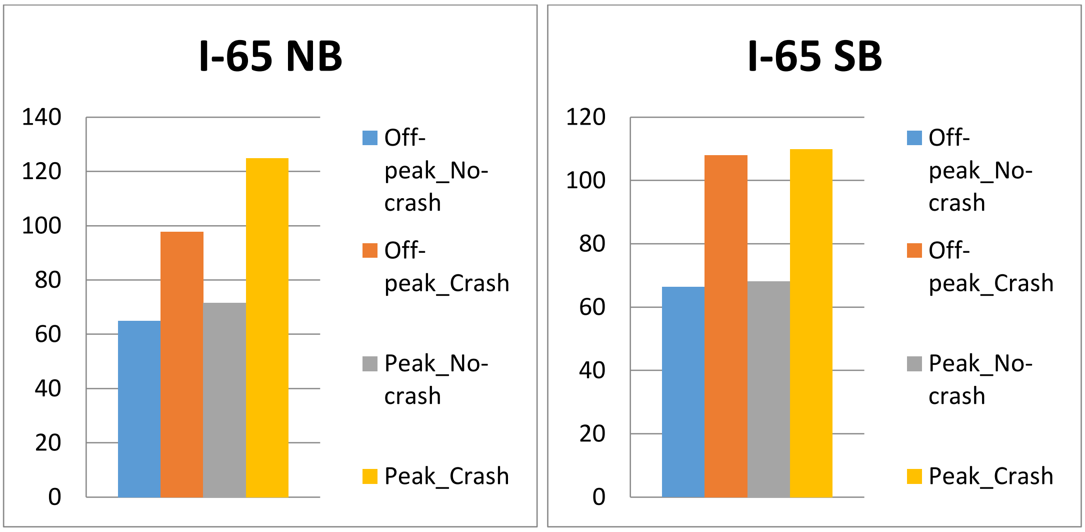


Figure 14: Average Travel Rate (sec/mile)

To explain those figures, let’s consider a travel rate of 80 sec/mile. According to I-65N (Figure 15), 96 percent of vehicles could travel at 80 sec/mile or less during an off-peak hour in non-crash situations, while 84 percent of vehicles travel at this rate during peak hours when there are no crashes. However, if a crash occurred during an off-peak period, only 41 percent of vehicles could travel at 80 sec/mile or less. If a crash happened during the peak period, the situation worsened. Only 23 percent of vehicles could achieve that travel rate. Similar trends were observed for I-65 S (Figure 16).

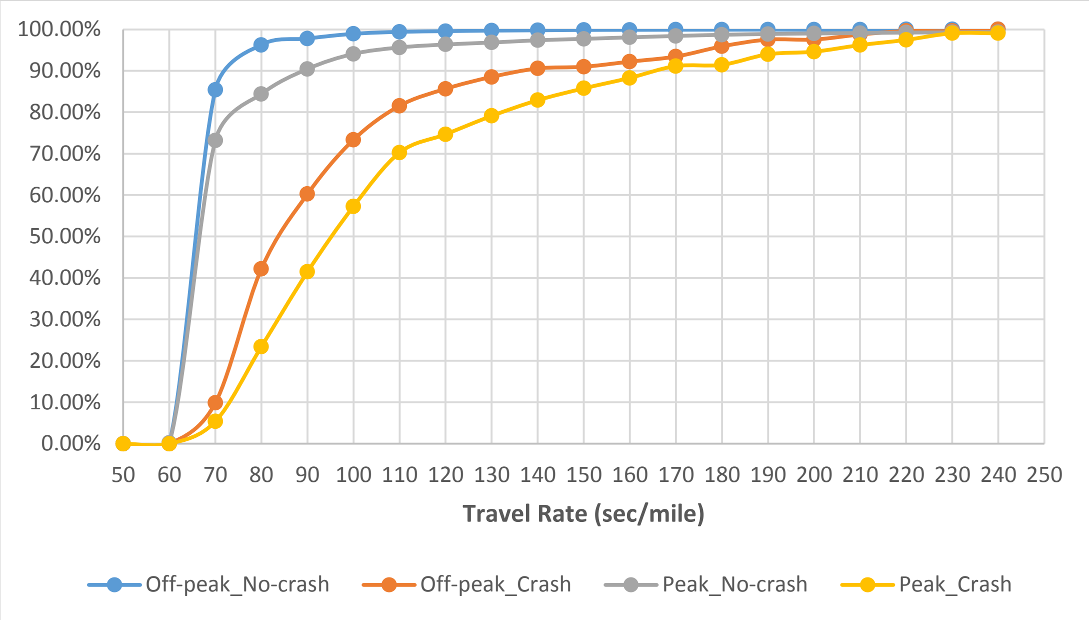


Figure 15: CDF of Travel Rates for I-65N

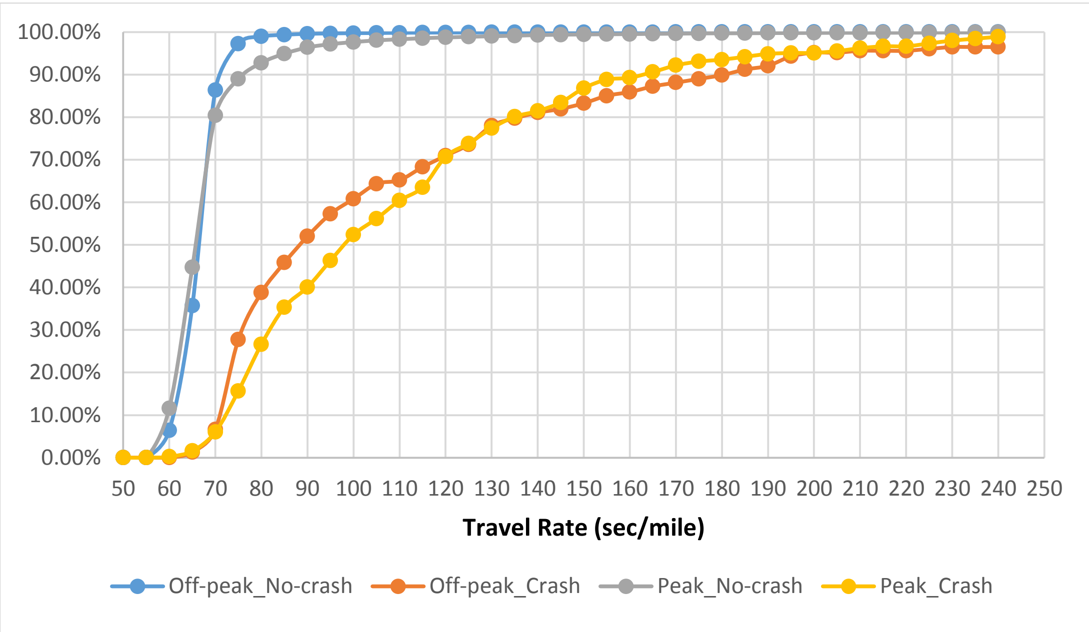


Figure 16: CDF of Travel Rates for I-65S

The benefit of using the actual impact duration over the reported incident duration by incident log is that it provides a more accurate estimation of CDF of travel rate and shows the true situation of the traveler experience. The same CDF of travel rate analysis was also conducted using the reported incident duration to show the difference between two estimations and the graphical representations of the CDF of travel rate are provided in the appendix-C. Figure 17 demonstrates the comparison between reported and measured duration in calculating CDF of travel rate at I-65N. From Figure 17 (where R=Reported, M=Measured, P=Peak, Cr=Crash), it is clear that CDF of travel rate using reported incident duration underestimates the true effect of the crash. For example, considering the measured impact duration along I-65N during off-peak crash, only 42% of vehicles could travel at 80 sec/mile or less, but using the reported incident duration, it is found that 65% of vehicles could travel at that rate which is very high. Similar trends were observed during peak-crash condition; while using the measured duration only 22% of vehicles could travel at 80 sec/mile or less, but considering the reported duration, it is found that 41.7%. However, during non-crash situations, the two estimations are almost same. Our main objective is to identify the true impact of the crash on the travel rate, and that is why actual impact duration that has been found after the five step crash identification process is suggested to use for estimating the CDF of travel rate which provides a more accurate estimation of the impact of the crash on the travel experience.

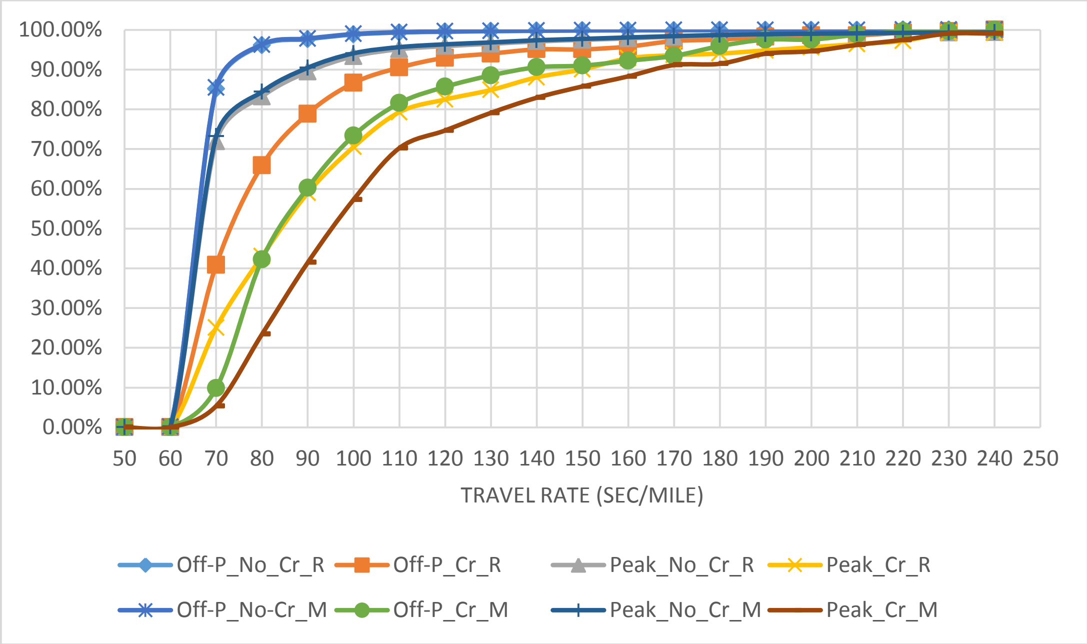


Figure 17: CDF of TR at I-65N for Reported vs Measured Duration

Next, we analyzed travel rate information for each route under four scenarios to determine the impacts of the crash on route travel time. These four conditions were:

1. No crashes: normal condition with no crashes
2. Crash with single lane blocked: crashes with a single lane blocked scenario with and without the shoulder blocked
3. Crash with multiple lanes blocked: crashes with multiple lanes blocked scenario with and without the shoulder blocked
4. Crash with only the shoulder blocked: crashes that block only the shoulder.

We drew a CDF of travel rate for the four scenarios to analyze the effect of different crash types. Taking a travel rate of 90 sec/mile as a baseline, along I-65N (Figure *18*), about 97% of vehicles could travel at 90 sec/mile or less during when there were no crashes. However, when a crash occurred, just 40% of vehicles maintained this travel rate when a single lane was blocked. Only 51% of vehicles moved at this rate under scenarios when just the shoulder was blocked. When multiple lanes were blocked due to a crash, the situation worsened: only 34% of vehicles traveled at 90 sec/mile or less travel rate. Similar trends were evident on I-65 S (Figure 19).

However, there was an exception on I-65 S (Figure 19). In this data set, all three types of the crash had a significant impact on travel rate but all types of the crash show a similar pattern. Although the multi-lane block crashes show a worse situation compared to the others, the differences are very small for the same travel rate. Again single lane and shoulder blocking crash graph almost overlap each other which indicates similar pattern between them.

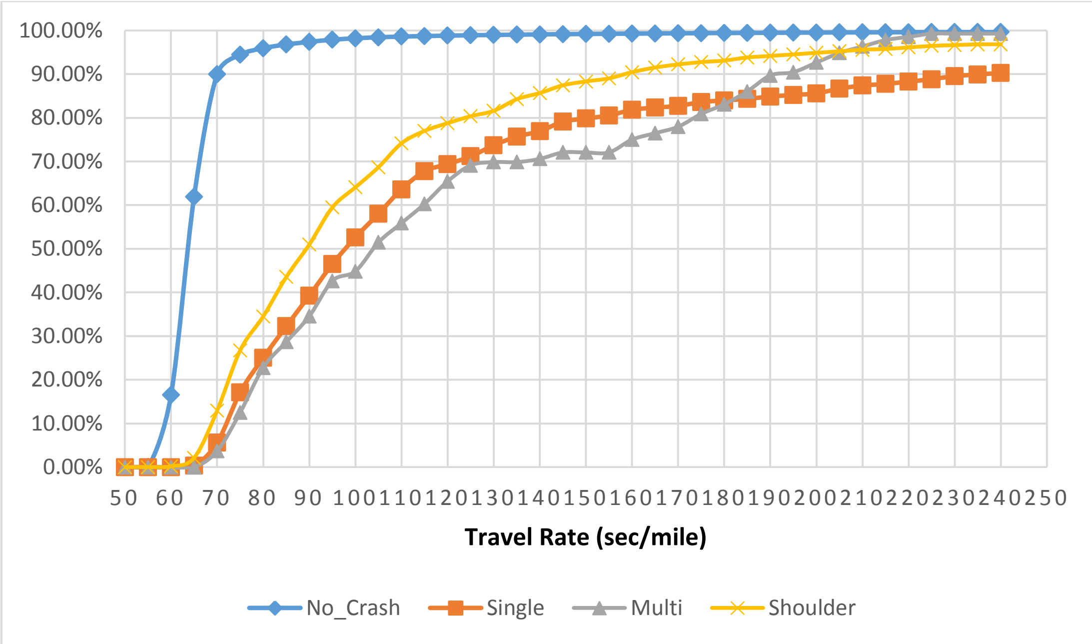


Figure 18: CDF of Travel Rate for Crash Types at I-65N

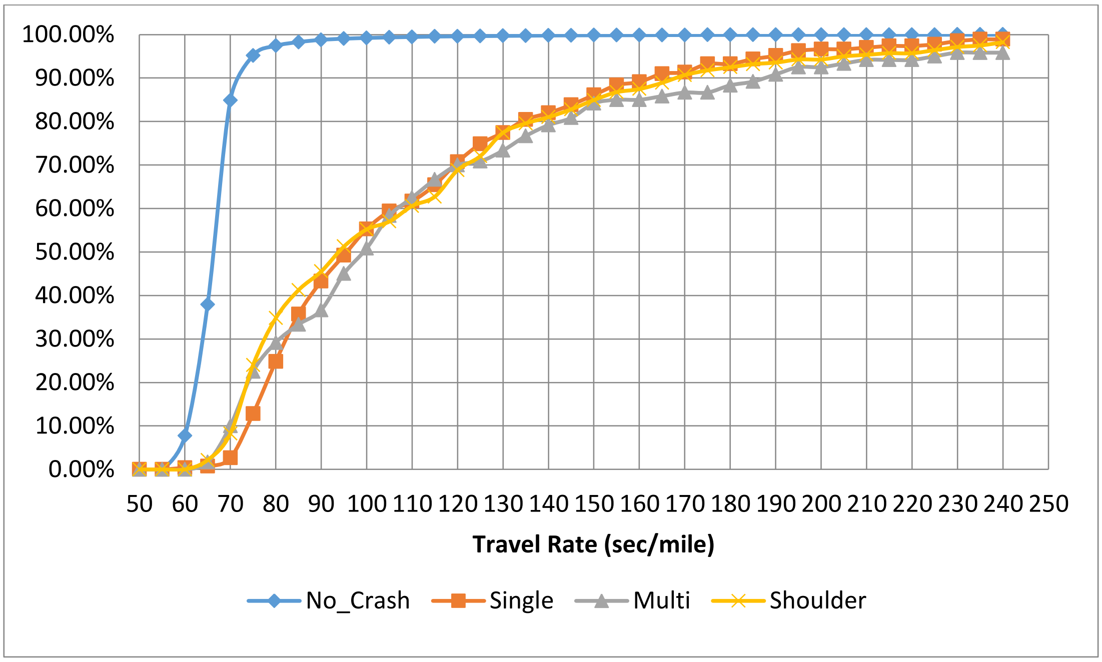


Figure 19: CDF of Travel Rate for Crash Types at I-65S

Using impact on reliability analysis, our main objective was to build a graphical framework to show the effect of crashes on travel experience. That objective was fulfilled with satisfaction. CDF of travel rate creates a better understanding about the effect of crashes on travel rate and quantify that effect in a graphical format.

This CDF of travel rate analysis provides additional insight about the effect of the crashes at different situations. This provides guidance to the traffic management agency about which actions might be taken to mitigate the impact of the crash.

5.8 Benefits of the Applications

The main objectives of the traffic management are to

- reduce the impact of congestion
- improve the level of service and performance of the freeway
- quick response to the incidents and clear the incident site as quick as possible
- dispatch appropriate message in advance to the users about current condition of the freeway

Traffic management could convey the message about the recurrent congestion knowing the historical trends of traffic. But in case of non-recurrent events such as incidents, inclement weather condition or other special events, the operator might panic because they do not have specific guidelines about how to measure the intensity of such events.

Among those non-recurrent events, crashes reduce the traffic flow at a great extent creating a temporary bottleneck on the roadway. In this study, we quantify the actual impact of crashes by the five step process and later use this impact duration value to build regression models. We propose three prediction models for the impact duration and one prediction model for the total delay.

Two impact duration models can predict the duration of crash impact as a function of two variables and quantile regression can predict as a function of three variables. These variables could be determined immediately after a crash occurred. Total delay model also can predict cumulative delay for each crash as a function of two variables. Among them, one variable (impact duration) cannot be determined until the normal traffic flow is reinstated. To use this model as an online tool for delay prediction, we can use the output of the impact duration model as an input of the delay model. Thus, together these two models provide an approximate picture of a crash scenario which would be very effective for the traffic management center in decision making.

This decision making includes what control strategy should be included to minimize the traffic congestion and improve the freeway performance. One control strategy would be to provide up-to-date information on freeway conditions to the users so that they can make relevant decisions before using that road which is known as Traveler Information System (TIS). The most commonly available traveler information systems are variable message signs and roadside or commercial radio broadcast.

When traffic operators detect an accident, they can use the proposed model to approximately calculate the impact duration and total delay for that particular accident. As previously discussed, the impact duration model would provide an approximate duration time in minutes and the delay model would provide the users an approximate delay caused by a specific accident in terms of vehicle-hour. Then the operator could utilize the average traffic flow information to estimate the average delay for each vehicle. For example, if the total delay for a specific crash was found to be 200 vehicle-hour in a freeway section with a traffic flow of 3000 vehicle per hour, then estimated average delay for each vehicle would be 4 minutes (200*60/3000). In this case traffic management can transmit two pieces of information (impact duration and delay per vehicle) to users via variable message signs or other suitable media. Now-a-days almost everyone uses smartphones and different types of apps for navigation purposes; Traffic management could transmit this information via smartphone/apps to the travelers in advance. This would help travelers to plan their trips and avoid the crash segment. A simple framework to convey this information is presented in Figure 20.

Based on the predicted duration of the crash impact and total predicted delay on the freeway and current conditions on local arterials streets, the traffic management center can implement appropriate diversion strategies in the network.

Moreover, reliability analysis (CDF of travel rate) provides additional insights about the impact of crashes; it showed how crashes affected travel rates during peak and off-peak periods and how different types of crashes produce significantly different outcomes.

Additionally, the CDF analysis could be used to conduct before and after study. For example, after implementing a certain incident management strategy, if the agency wants to know how much the new strategy improves the performance of the roadway, they can conduct this CDF of travel rate analysis before and after the implementation and could identify the change of performance of the roadway.

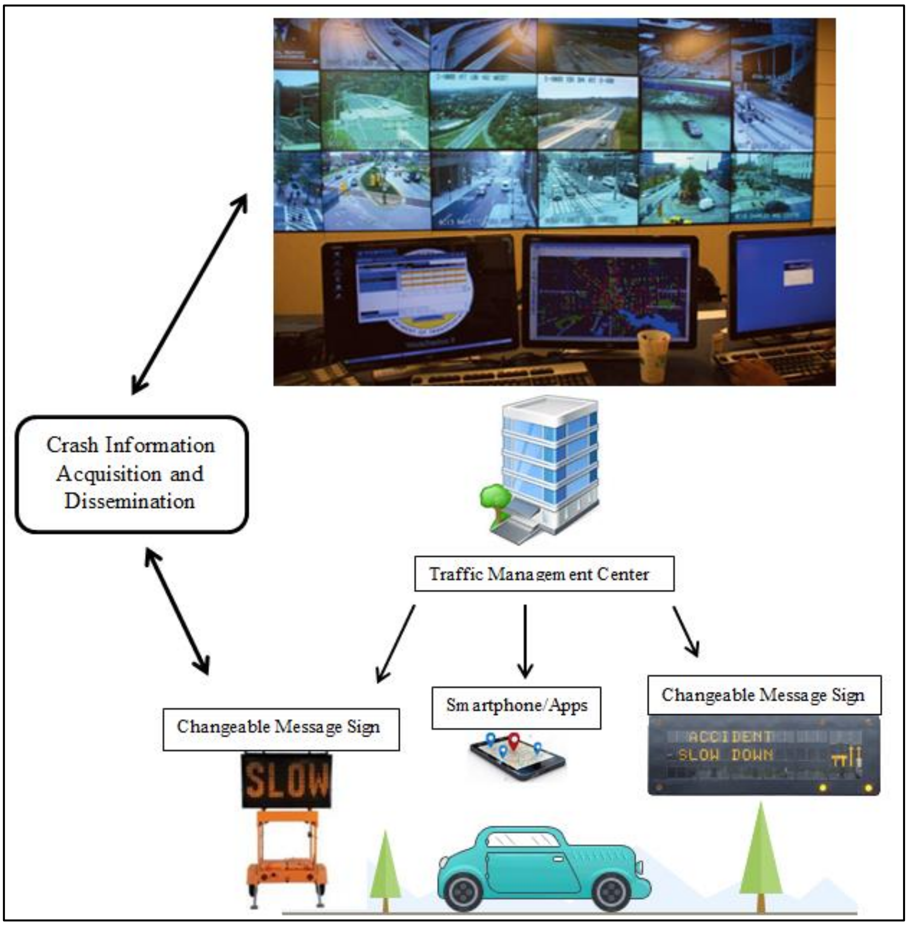


Figure 20: Advance Traveler Information System

This chapter provides the examples of possible applications. This chapter also provides detailed analysis and insights of each application such as impact duration model, total delay model and CDF of travel rate analysis. In the next chapter, findings of the total study will be discussed along with future works.

Chapter 6 Conclusions

6.1 Overview

Crash induced congestion is one of the major causes of the traffic delay. This study proposed a methodology for identification of spatiotemporal impact of each crash based on the stationary sensor data and presented possible applications of this methodology.

Identification of the temporal and spatial extent of the congested region due to the crash opens a new window to determine delay and duration of that crash. Quantifying crash induced delay helps monitoring performance of the roadways and assessing various congestion mitigation measures.

The methodology defines crash impacts as the reduction of traffic speed experienced by the traveler under crash conditions and the speed reduction is determined with respect to the expected traffic speed based on traveler's past experience. The method involves the development of background speed profile by using Kalman Filter algorithm and then superimposing the current traffic speed under crash conditions onto it. The resulting difference between two profiles shows the reduction of speed due to the crash and indicates the crash induced congestion. In this way, we can capture the dynamics of the impact of the crash during recurrent and non-recurrent congested conditions.

The accuracy of this method depends on the availability and the accuracy of the stationary sensor data and the incident log in the database. This in turn depends on the accuracy and consistency of the speed and incident logging procedure of the traffic management center because the incident start time and current speed data are directly used in the proposed methodology to identify the impact of crash.

6.2 Findings

In this research, a methodology to identify the impact of crash was proposed and examined in depth. Additionally, possible applications were investigated for the proposed methodology. Based on the analysis and modeling results, some of the important findings and conclusions are given below:

Integration of the incident data with the stationary sensor data provides the traffic management center a data driven approach to measure the crash induced congestion and its temporal and spatial extent. The case study demonstrated the performance of the methodology to automatically identify the impact of each crash. The use of a simple, yet informative heat map enhances our visual understanding to identify the spatiotemporal impact of crashes.

This study presented three models for predicting impact duration of crashes. One model can predict 32% variation of impact duration based on two independent variables (post-crash speed and weather condition). The second model can predict the probability of impact duration for lying at different response categories based on the same two independent variables. Moreover, the quantile regression provides a detailed picture of the duration of impact. Instead of giving a single output, it provides a range of values for the best possible to the worst possible scenario of a crash in a form of duration. It is worth mentioning that these models can predict the impact duration just knowing the post-crash speed at the upstream segment and the weather condition (rainy or not) which can be determined easily from real-time sensor data and weather information.

Furthermore, the paper presented a regression model for predicting delay for the crash impact. The impact delay model showed that 74% variation of the total delay can be predicted as a function of two independent variables (post-crash speed and impact duration). The proposed statistical models (impact duration and delay) are the possible applications of the five-step crash identification process, which can be used for informing the road users about freeway conditions so that they can make relevant decisions before using that road. Moreover, based on the predicted duration of the crash impact, total predicted delay on the freeway and current conditions on local arterials streets, the traffic management center can implement appropriate diversion strategies in the network to minimize congestion.

On the other hand, CDF of travel rates tells a better story about the impacts of crashes — it shows to what extent crashes influenced the travel rate during peak and off-peak periods. CDF of travel rate also showed how different types of crashes produce significantly different outcomes. Among the different crash types, crashes that blocked multiple lanes induce the most significant negative impacts on travel rate. Compared to normal conditions, when 97% to 98% of vehicles could travel at a specified travel rate, only 34% to 36% of vehicles could travel at that rate during multi-lane crashes. Crashes that blocked an individual lane or just the shoulder produced similar impacts on travel rates in both directions. The findings in this study can help to shape crash management policies for different types of crashes at different periods.

6.3 Future Research

The critical philosophy of examining the crash impact within the dynamic nature of traffic is to derive an accurate representation of background speed profile at the time of crash which reflects the traveler's expected normal condition. In this study, we combine both historical data and on-going traffic conditions before a crash to construct background profile, but it may not always be reflective of the normal traffic condition of that day. Several factors, such as weather conditions and seasonality, may influence the background speed profile. These factors should be further investigated for constructing the background speed profile.

In this study, we focused on developing a practical approach to construct a background profile to identify the impact of crash rather than investigating all possible factors that might influence background speed profile. In future research, ideal approach to construct background speed profile would be some sort of predictive model that will consider all essential factors that might have influence and derive an unbiased estimation of the normal traffic condition.

Besides, we have tested our methodology using stationary sensors data. In the case of roadways with few or no sensors, third party data (such as probe data) could be used to make this kind of analysis.

One issue with the incident data was that it did not always keep the consistent record of lane-blocking, injury and other information. A more complete data-set might provide additional insight for predicting the crash impact and total delay. Future research should focus on: (1) developing the regression model (duration/delay) with enrich data-set, (2) calibrating the models using the data from other sites, and (3) potential variables (lane blocking, injury etc) that are omitted should be examined for other locations.

Operation efficiency and traffic safety are considered as the most important elements among highway system performance measurement. Traffic congestion serves as a proxy for efficiency, and crash analysis can be used to evaluate highway safety. With the advances in big data, improving operations and safety in real-time is now possible. However, to fully realize the power of this data, we need to develop more applications for this data. This research has illustrated how data sets can be analyzed innovatively to improve our understanding of crashes, their impacts, and ultimately their distribution in spatial temporal domain.

In this chapter, findings of the total research is described briefly. It also provides recommendations for future research. In the next section appendices and references are provided.

Appendices

Appendix A: Incident Log Records

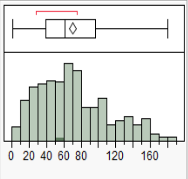
Field Name	Description
ID	Incident Identification Number
Type	Type of Incident
Start Date	Start date of incident
Start Time	Start time of incident
End Date	End date of incident
End Time	End time of incident
Total	Duration of Incident
State	Name of State
Hwy	Name of the route
Direction	Direction of the route
MP	Closest mile marker
Conditions	Pavement Condition
Est. Clear	Estimated clearance time
Lat	Latitude of the location
Long	Longitude of the location
Lanes Blocked	Number of lanes closed
Notes	Descriptive information of the incident

Appendix B: Results of the Statistical Model

Results of the Impact Duration Model

Distributions

Impact_Duration



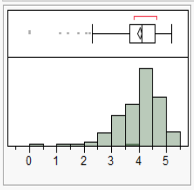
Quantiles

100.0%	maximum	180
99.5%		180
97.5%		159
90.0%		135
75.0%	quartile	97
50.0%	median	61
25.0%	quartile	39
10.0%		22
2.5%		11.4
0.5%		1
0.0%	minimum	1

Summary Statistics

Mean	70.007752
Std Dev	40.515812
Std Err Mean	2.0595345
Upper 95% Mean	74.057062
Lower 95% Mean	65.958442
N	387

ln(duration)



Quantiles

100.0%	maximum	5.19296
99.5%		5.19296
97.5%		5.0689
90.0%		4.90527
75.0%	quartile	4.57471
50.0%	median	4.11087
25.0%	quartile	3.66356
10.0%		3.09104
2.5%		2.43021
0.5%		0
0.0%	minimum	0

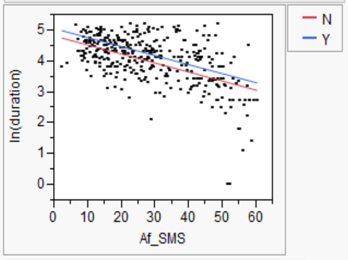
Summary Statistics

Mean	4.0337493
Std Dev	0.7504932
Std Err Mean	0.0381497
Upper 95% Mean	4.1087566
Lower 95% Mean	3.9587421
N	387

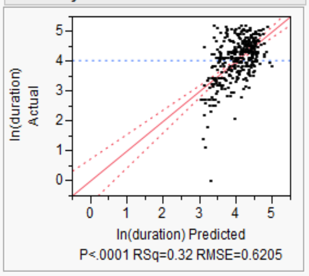
Response ln(duration)

Whole Model

Regression Plot



Actual by Predicted Plot



Summary of Fit

RSquare	0.320052
RSquare Adj	0.316511
Root Mean Square Error	0.620458
Mean of Response	4.033749
Observations (or Sum Wgts)	387

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	69.58281	34.7914	90.3747
Error	384	147.82784	0.3850	Prob > F
C. Total	386	217.41065		<.0001*

Lack Of Fit

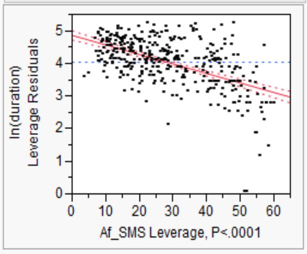
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	294	118.21110	0.402079	1.2218
Pure Error	90	29.61674	0.329075	Prob > F
Total Error	384	147.82784		0.1307
				Max RSq
				0.8638

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.9281452	0.073673	66.89	<.0001*
Af_SMS	-0.029141	0.002256	-12.91	<.0001*
Rain_A[N]	-0.119369	0.037971	-3.14	0.0018*

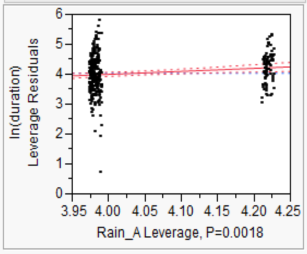
Af_SMS

Leverage Plot



Rain_A

Leverage Plot



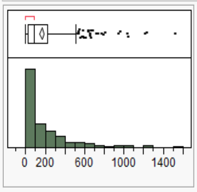
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
N	3.9806964	0.03577088	3.97077
Y	4.2194347	0.06695983	4.25419

Results of the Total Delay Model

Distributions

Delay(Vh-Hr)



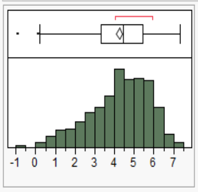
Quantiles

100.0%	maximum	1524.9
99.5%		1230.2
97.5%		752.115
90.0%		411.78
75.0%	quartile	225.15
50.0%	median	86.3
25.0%	quartile	27.3
10.0%		7.84
2.5%		2.455
0.5%		1.111
0.0%	minimum	0.4

Summary Statistics

Mean	164.57102
Std Dev	210.48782
Std Err Mean	10.258549
Upper 95% Mean	184.73552
Lower 95% Mean	144.40653
N	421

Ln(Delay)



Quantiles

100.0%	maximum	7.32968
99.5%		7.11493
97.5%		6.62249
90.0%		6.02047
75.0%	quartile	5.41675
50.0%	median	4.45783
25.0%	quartile	3.30689
10.0%		2.05919
2.5%		0.89792
0.5%		0.10488
0.0%	minimum	-0.9163

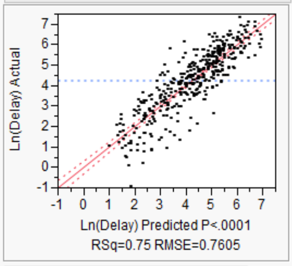
Summary Statistics

Mean	4.2510338
Std Dev	1.5169344
Std Err Mean	0.0739309
Upper 95% Mean	4.3963544
Lower 95% Mean	4.1057132
N	421

Response Ln(Delay)

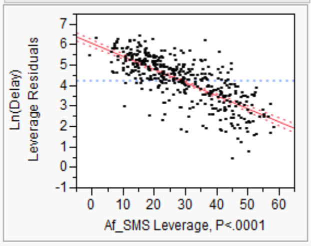
Whole Model

Actual by Predicted Plot



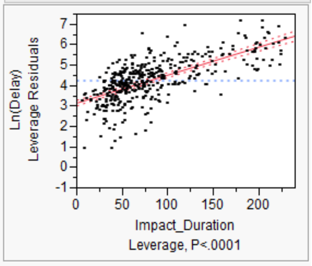
Af_SMS

Leverage Plot



Impact_Duration

Leverage Plot



Summary of Fit

RSquare	0.74986
RSquare Adj	0.748663
Root Mean Square Error	0.760492
Mean of Response	4.251034
Observations (or Sum Wgts)	421

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	724.70814	362.354	626.5323
Error	418	241.74970	0.578	Prob > F
C. Total	420	966.45783		<.0001*

Lack Of Fit

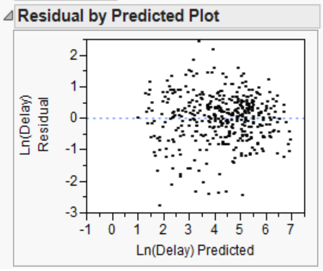
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	408	239.00277	0.585791	2.1325
Pure Error	10	2.74692	0.274692	Prob > F
Total Error	418	241.74970		0.0902
				Max RSq
				0.9972

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.9766428	0.121106	41.09	<.0001*
Af_SMS	-0.06393	0.002842	-22.50	<.0001*
Impact_Duration	0.0136707	0.000743	18.40	<.0001*

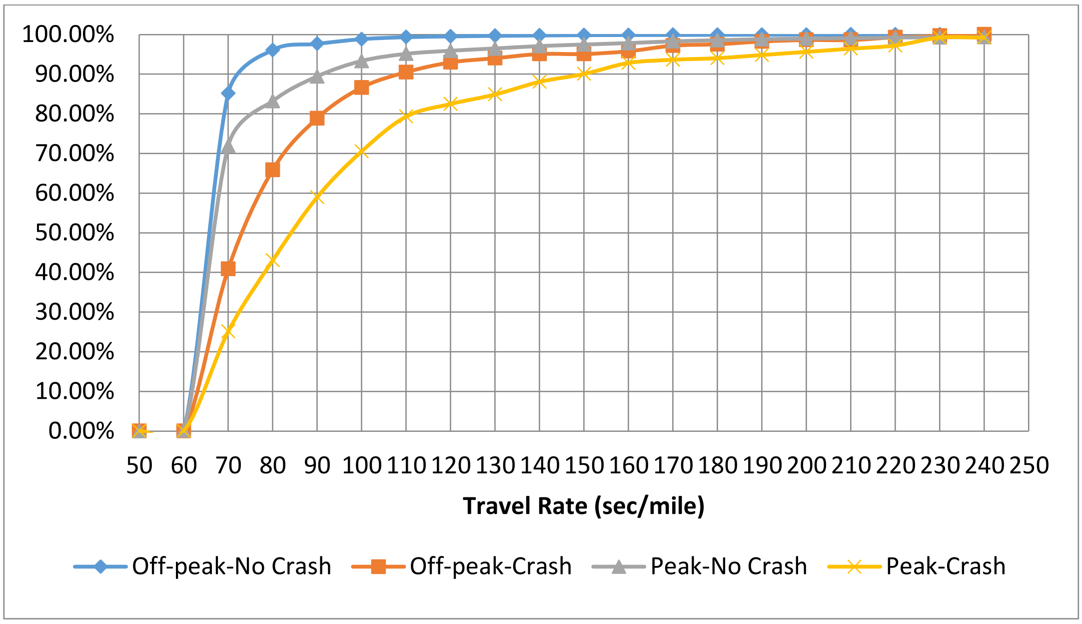
Effect Tests

Residual by Predicted Plot

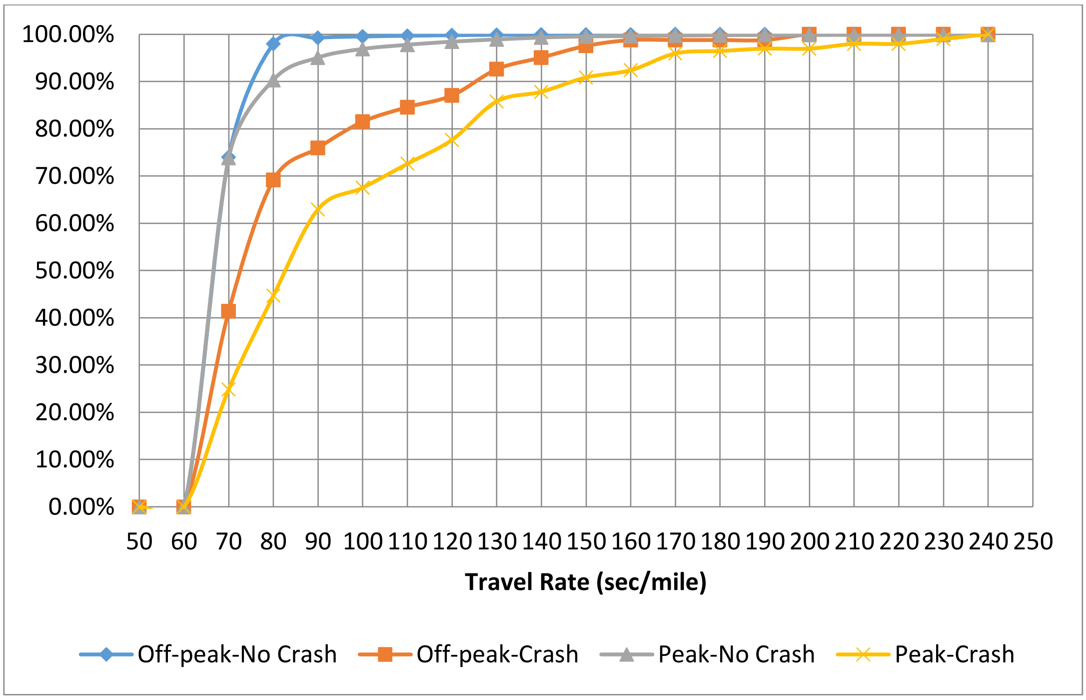


Appendix C: CDF of Travel Rate

Cumulative Distribution of Travel Rate at I-65N using reported incident duration



Cumulative Distribution of Travel Rate at I-65S using reported incident duration



References

1. Lomax, T., D. Schrank, S. Turner, and R. Margiotta., *Selecting travel reliability measures*. 2013: Texas Transportation Institute.
2. Neudorff, L. G., Randall, J. E., Reiss, and R. Gordon, *Freeway management and operations handbook*. 2003.
3. Garib, A., A. Radwan, and H. Al-Deek, *Estimating magnitude and duration of incident delays.* Journal of Transportation Engineering, 1997. **123**(6): p. 459-466.
4. Smith, K.W. and B.L. Smith, *Forecasting the clearance time of freeway accidents*. 2001: Research Report, STL-2001-01.
5. Khattak, A.J., J.L. Schofer, and M.-h. Wang, *A simple time sequential procedure for predicting freeway incident duration.* Journal of Intelligent Transportation Systems, 1995. **2**(2): p. 113-138.
6. Nam, D. and F. Mannering, *An exploratory hazard-based analysis of highway incident duration.* Transportation Research Part A: Policy and Practice, 2000. **34**(2): p. 85-102.
7. Giuliano, G., *Incident characteristics, frequency, and duration on a high volume urban freeway.* Transportation Research Part A: General, 1989. **23**(5): p. 387-396.
8. Golob, T.F., W.W. Recker, and J.D. Leonard, *An analysis of the severity and incident duration of truck-involved freeway accidents.* Accident Analysis & Prevention, 1987. **19**(5): p. 375-395.
9. Wu, W., S.Y. Chen, and C.J. Zheng, *Traffic incident duration prediction based on support vector regression.* Proceedings of the ICCTP, 2011. **2011**: p. 2412-2421.
10. Lin, P.W., N. Zou, and G.L. Chang. *Integration of a discrete choice model and a rule-based system for estimation of incident duration: a case study in Maryland*. in *CD-ROM of Proceedings of the 83rd TRB Annual Meeting, Washington, DC*. 2004.
11. KIM, H.J. and C. Hoi-Kyun, *A comparative analysis of incident service time on urban freeways.* IATSS research, 2001. **25**(1): p. 62-72.
12. Boyles, S., D. Fajardo, and S.T. Waller. *A naive Bayesian classifier for incident duration prediction*. in *86th Annual Meeting of the Transportation Research Board, Washington, DC*. 2007.
13. Wang, W., H. Chen, and M.C. Bell, *Vehicle breakdown duration modelling.* Journal of Transportation and Statistics, 2005. **8**(1): p. 75-84.
14. Skabardonis, A., K. Petty, H. Noeimi, D. Rydzewski, and P. Varaiya, *I-880 field experiment: Data-base development and incident delay estimation procedures.* Transportation Research Record: Journal of the Transportation Research Board, 1996(1554): p. 204-212.
15. Wang, Y., M.E. Hallenbeck, and P. Cheevarunothai, *Quantifying incident-induced travel delays on freeways using traffic sensor data*. 2008, Transportation Northwest, University of Washington.
16. Messer, C.J., C.L. Dudek, and J.D. Friebele, *Method for predicting travel time and other operational measures in real-time during freeway incident conditions.* Highway Research Record, 1973(461).
17. Cragg, C.A. and M.J. Demetsky, *Simulation Analysis of Route Diversion Strategies for Freeway Incident Management*. 1995.

18. Zhang, H., F. Ni, and S. Yang. *Prediction of freeway traffic incident delay based on simulation*. *15th International IEEE Conference on Intelligent Transportation Systems*. 2012. IEEE.
19. Fries, R., C. Hamlin, M. Chowdhury, Y. Ma, and K. Ozbay, *Operational impacts of incident quick clearance legislation: a simulation analysis.* Journal of advanced transportation, 2012. **46**(1): p. 1-11.
20. Kabit, M.R.B., P. Charles, L. Ferreira, and I. Kim , *Modelling major traffic incident impacts and estimation of their associated costs.* Transportation planning and technology, 2014. **37**(4): p. 373-390.
21. Tu, H., H. Li, H. Van Lint, and H. van Zuylen, *Modeling travel time reliability of freeways using risk assessment techniques.* Transportation Research Part A: Policy and Practice, 2012. **46**(10): p. 1528-1540.
22. Yu, R., Q. Shi, and M. Abdel-Aty, *Feasibility of incorporating reliability analysis in traffic safety investigation.* Transportation Research Record: Journal of the Transportation Research Board, 2013(2386): p. 35-41.
23. Zhong, C., V.Sisiopiku, K. Ksaibati, and T. Zhong, *Crash prediction on rural roads*. in *Proceeding of the 3rd International Conference on Road Safety and Simulation, Indianapolis, Indiana*. 2011.
24. Miaou, S.P. and H. Lum, *Modeling vehicle accidents and highway geometric design relationships.* Accident Analysis & Prevention, 1993. **25**(6): p. 689-709.
25. Okamoto, H. and M. Koshi, *A method to cope with the random errors of observed accident rates in regression analysis.* Accident Analysis & Prevention, 1989. **21**(4): p. 317-332.
26. Wright, B., Y. Zou, and Y. Wang, *Impact of Traffic Incidents on Reliability of Freeway Travel Times.* Transportation Research Record: Journal of the Transportation Research Board, 2015(2484): p. 90-98.
27. Chung, Y. and W.W. Recker, *A methodological approach for estimating temporal and spatial extent of delays caused by freeway accidents.* IEEE Transactions on Intelligent Transportation Systems, 2012. **13**(3): p. 1454-1461.
28. Pan, B., U. Demiryurek, C. Gupta, and C. Shahabi, *Forecasting spatiotemporal impact of traffic incidents for next-generation navigation systems.* Knowledge and Information Systems, 2015. **45**(1): p. 75-104.
29. Yu, R., M.A. Abdel-Aty, M.M. Ahmed, and X. Wang, *Crash-type propensity analysis with bayesian models using microscopic traffic and weather data*. in *Transportation Research Board 92nd Annual Meeting*. 2013.
30. Ahmed, M., M. Abdel-Aty, and R. Yu, *Assessment of interaction of crash occurrence, mountainous freeway geometry, real-time weather, and traffic data.* Transportation research record: Journal of the Transportation Research Board, 2012(2280): p. 51-59.
31. Hojati, A.T., L. Ferreira, P. Charles, and M.R. bin Kabit, M.R., *Analysing freeway traffic-incident duration using an Australian data set.* Road & Transport Research: A Journal of Australian and New Zealand Research and Practice, 2012. **21**(2): p. 19.
32. Kalman, R.E., *A new approach to linear filtering and prediction problems.* Journal of basic Engineering, 1960. **82**(1): p. 35-45.
33. Chen, M., X. Liu, and J. Xia, *Dynamic prediction method with schedule recovery impact for bus arrival time.* Transportation Research Record: Journal of the Transportation Research Board, 2005(1923): p. 208-217.

34. Mason, R.L., R.F. Gunst, and J.L. Hess, *Statistical design and analysis of experiments: with applications to engineering and science*. Vol. 474. 2003: John Wiley & Sons.
35. Koenker, R., *Quantile regression*. 2005: Cambridge University Press.

VITA

- Obaidur Rahman Kazi
- Place of Birth: Kishoreganj, Bangladesh
- Educational Institutions Attended:
    - Bangladesh University of Engineering and Technology: Bachelors of Science in Civil Engineering
- Professional Positions Held:
    - University of Kentucky: Graduate Research Assistant
    - University of Kentucky: Graduate Teaching Assistant
    - Presidency University: Lecturer
    - University of Kentucky Chapter of ITE: Student Member
- Scholastic Honors:
    - University of Kentucky
        - Graduate Research Assistantship
    - Bangladesh University of Engineering and Technology
        - University Merit Scholarship
    - Dhaka Education Board
        - Education Board Merit Scholarship
- Academic Publications and Presentations:
    - Kazi, O.R., and Chen, M., *A Data-Driven Approach to Analyze the Impact of Crash*. Poster presented at University Transportation Center (UTC) Conference for the Southeastern Region, Knoxville, 2016.
    - Chen, M., Kazi, O.R., and Fields, T., *Develop a Highway Inventory Data Quality Control Scheme (Pl-26)*. Final Report, Kentucky Transportation Cabinet, 2016.
    - Zhang, X., Kazi, O.R., and Chen, M., *An Effective Data Fusion Algorithm for the Travel Time Reliability Application*. Poster presented at the 56th Annual Meeting of Transportation Research Forum, Atlanta, 2015.
    - Chen, M., Kazi, O.R., Zhang, X., and Gibson, B., *Developing an Integrated Framework for Safety and Mobility Analysis*. Final Report, Southeastern Transportation Center, 2015.