

**NDOT Research Report**

**Report No. 224-14-803**

**Task Order 10**

---

**Correlation Analysis of Freeway Traffic Status  
and Crashes with Nevada Data**

---

**November 2017**

**Nevada Department of Transportation  
1263 South Stewart Street  
Carson City, NV 89712**



## TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 224-14-803 TO 10	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Correlation Analysis of Freeway Traffic Status and Crashes with Nevada Data		5. Report Date November 11, 2017	
		6. Performing Organization Code University of Nevada, Reno	
7. Author(s) Hao Xu, Jianqing Wu and Zong Tian		8. Performing Organization Report No.	
9. Performing Organization Name and Address Center for Advanced Transportation Education and Research University of Nevada, Reno 1664 N. Virginia St, Mail Stop 258, Reno, Nevada, 89557		10. Work Unit No.	
		11. Contract or Grant No. NDOT agreement 224-14-803 Task Order 10	
12. Sponsoring Agency Name and Address Nevada Department of Transportation 1263 South Stewart Street Carson City, NV 89712		13. Type of Report and Period Covered Final Report	
		14. Sponsoring Agency Code NDOT	
15. Supplementary Notes This project was sponsored by the UNR SOLARIS University Transportation Center and NDOT			
16. Abstract This project is to study the correlation between freeway traffic status and crash risks with the historical freeway ITS data and related crash data in Nevada. With the comprehensive review of previous research results, the Center for Advanced Transportation Education and Research (CATER) at the University of Nevada, Reno (UNR) analyzed the interactive relationship between the freeway traffic status and the crashes. The research developed a new method to correct the crash time that was not accurately documented in the crash report. A crash risk prediction model has been developed to estimate the influence of traffic status on collision risk. The impact of crash severity and crash type on after-crash traffic flows was also analyzed. The real-time crash risk prediction models can be used to estimate the collision possibilities based on the real-time ITS data. The models and analysis were based on the Nevada data so that they can be directly applied on Nevada freeways without additional calibration effort. The Nevada Data Exchange (NDEX) data warehouse developed and managed by the Nevada Department of Transportation (NDOT) has been reviewed for its capability of feeding real-time freeway ITS data to the models. The connection to the NDEX warehouse has been tested. This project prepared Nevada for the future implementation of real-time crash risk prediction along freeways.			
17. Key Words Freeway risk prediction, real time, intelligent transportation system data, statistical analysis, traffic status		18. Distribution Statement No restrictions. This document is available through the: National Technical Information Service Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 31	22. Price

## Disclaimer

This work was sponsored by the Nevada Department of Transportation. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of Nevada at the time of publication. This report does not constitute a standard, specification, or regulation.



## **Final Report**

November 2017

# **Correlation Analysis of Freeway Traffic Status and Crashes with Nevada Data**

SOLARIS Consortium, Tier 1 University Transportation Center  
Center for Advanced Transportation Education and Research  
Department of Civil and Environmental Engineering  
University of Nevada, Reno  
Reno, NV 89557

---

Hao Xu, Jianqing Wu and Zong Tian  
Center for Advanced Transportation Education and Research  
University of Nevada, Reno

**DISCLAIMER:**

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

---

## EXECUTIVE SUMMARY

---

The conventional traffic safety study has been predicting crash rate or frequency by considering Annual Average Daily Traffic (AADT) and road properties, such as curve radius, roughness, superelevation et al. The conventional models help identification of high-crash sites and selection/development of traffic safety countermeasures. When various real-time traffic data are available from sensors widely deployed in today's intelligent transportation systems (ITS), a new direction to improve highway safety is real-time crash-risk prediction with the real-time traffic information. While previous research has started to look at the possible solutions, some questions were still not well answered. On the other side, when different regions have different ITS data, the models developed in other countries or states may not serve Nevada highways well.

This project is to study the correlation between freeway traffic status and crash risks with the historical freeway ITS data and related crash data in Nevada. With the comprehensive review of previous research results, the Center for Advanced Transportation Education and Research (CATER) at the University of Nevada, Reno (UNR) analyzed the interactive relationship between the freeway traffic status and the crashes. The research developed a new method to correct the crash time that was not accurately documented in the crash report. Crash and traffic flow data from one hour before and one hour after the accident were extracted to support the research. The negative binomial analysis is selected and applied to study the relationship between traffic flow and crash characteristics. A crash risk prediction model has been developed to estimate the influence of traffic status on collision risk. The impact of crash severity and crash type on after-crash traffic flows was also analyzed. The real-time crash risk prediction models can be used to estimate the collision possibilities based on the real-time ITS data. The models and analysis were based on the Nevada data so that they can be directly applied on Nevada freeways without additional calibration effort. The Nevada Data Exchange (NDEX) data warehouse developed and managed by the Nevada Department of Transportation (NDOT) has been reviewed for its capability of feeding real-time freeway ITS data to the models. The connection to the NDEX warehouse has been tested.

This project prepared Nevada for the future implementation of real-time crash risk prediction along freeways. The prediction methods and the available real-time data input have been well prepared. The project results can lead to the future crash warning system to warn drivers, traffic operators, and incident responders, which will help reduce crashes on freeways and minimize influence on traffic if an accident happens.

## TABLE OF CONTENTS

---

Executive Summary .....	i
Table of Contents .....	ii
List of Tables.....	iii
List of Figures .....	iv
Abbreviation .....	v
1 Introduction .....	1
2 Data and Data Processing .....	4
2.1 Data Source .....	4
2.2 Data Processing.....	5
2.2.1 Spatially relating sensor data and crashes.....	5
2.2.2 Temporally relating sensor data and crashes .....	7
3 Model Development and Analysis.....	9
3.1 Available Methodologies for Crash Risk Prediction.....	9
3.2 Selection of Methodology for Crash Risk Prediction .....	9
3.3 Influence of Traffic Status on Crash Risk .....	11
3.3.1 Scenario 1: speed lower than the normal speed.....	11
3.3.2 Scenario 2: speed higher than normal speed.....	12
3.4 Influence of Crashes on Traffic Flow.....	13
3.5 Influence of Difference between Average Speed and Speed Limit.....	15
4 NDOT NDEX .....	19
4.1 NDEX Web Services .....	19
4.1.1 WSDL .....	19
4.1.2 XSD.....	20
4.1.3 XML.....	20
4.2 NDEX Connection Guidance.....	20
5 CONCLUSION AND DISCUSSION .....	22
REFERENCES .....	24

## LIST OF TABLES

---

Table 2-1. Speeds of No-Crash Scenarios and Crash Scenarios .....	8
Table 3-1. Scenarios for crash prediction model .....	11
Table 3-2. Regression Analysis Results of Scenario 1 .....	12
Table 3-3. Regression Analysis results of Scenario 2 .....	12

## LIST OF FIGURES

---

Figure 2-1. Inductive loop detector locations of received ITS data.....	4
Figure 2-2. Flowchart of crash data processing.....	6
Figure 2-3. Estimation of the actual crash time.....	8
Figure 3-1 Relationship of after-crash speed and distance to the crash location.....	10
Figure 3-2 Relationship between crash severity and after-crash speed.....	14
Figure 3-3 Relationship between crash severity and after-crash traffic volume.....	14
Figure 3-4 Relationship between crash type and after-crash speed.....	15
Figure 3-5 Relationship between crash type and after-crash traffic volume.....	15
Figure 3-6 Frequency distribution of crashes under different speed deviations.....	16
Figure 3-7 Frequency distribution of different speed deviations.....	17
Figure 3-8 Relationship of speed deviation and crash risk.....	18

## ABBREVIATION

---

<b>AADT</b>	Annual Average Daily Traffic
<b>ADT</b>	Average Daily Traffic
<b>API</b>	Application Programming Interface
<b>CATER</b>	Center for Advanced Transportation Research and Education
<b>DMS</b>	Dynamic Message Signs
<b>FAST</b>	Freeway & Arterial System of Transportation
<b>GIS</b>	Geographic Information System
<b>HAR</b>	Highway Advisory Radio
<b>IC</b>	Injury Crashes
<b>ITE</b>	Institute of Transportation Engineers
<b>ITS</b>	Intelligent Transportation System
<b>LRS</b>	Linear Referencing System
<b>NCATS</b>	Nevada Citation and Accident Tracking System
<b>NDEX</b>	Nevada Data Exchange
<b>NDOT</b>	Nevada Department of Transportation
<b>NLCCA</b>	Nonlinear Canonical Correlation Analysis
<b>PDO</b>	Property Damage Only
<b>RTC</b>	Regional Transportation Commission
<b>RWIS</b>	Roadway Weather Information Systems
<b>SHSP</b>	Nevada Strategic Highway Safety Plan
<b>SOLARIS</b>	Safety and Operations of Large-Area Rural/Urban Intermodal Systems
<b>TMDD</b>	Traffic Management Data Dictionary
<b>TSE</b>	Traffic Safety Engineering
<b>UNR</b>	University of Nevada, Reno
<b>UTC</b>	University Transportation Center
<b>WSDL</b>	Web Services Description Language
<b>XML</b>	Extensible Markup Language
<b>XSD</b>	XML Schema Definition

## 1 INTRODUCTION

---

With the Intelligent Transportation System (ITS) technologies expanded, both traffic safety and mobility benefit from the new ITS applications. Federal and states increasingly emphasize the need to reduce traffic fatalities and serious injuries by making use of ITS technologies, which was also included in the Nevada Strategic Highway Safety Plan (SHSP). ITS techniques have been applied in traffic operation for decades with traffic sensors widely deployed. When new sensor technologies are being developed and implemented, research on the correlation between real-time traffic status from existing sensors and crash risk can serve as a parallel approach to improving traffic safety. When different traffic sensors, such as inductive loops, cameras, and radars, have been employed in various regions for different purposes, the inductive loop detectors along freeways continuously collect freeway traffic data, normally including the traffic flow rate, average speed, occupancy, and other information. The widely deployed inductive loop detectors offer the opportunity to study the correlation between real-time traffic and crashes before and after collisions happen.

Freeways are designed to provide high speed for people and freight transport. The unhindered flow of traffic saves time and reduces fuel consumption. At the same time, higher speed reduces a driver's ability to react to an emergency. The high speeds of traffic on freeways cause serious threats to life when any traffic crash occurs. The high speed and high volume of freeway traffic also cause concern of second crashes and severe delay after a collision. When traffic safety engineers conventionally use Average Annual Daily Traffic (AADT) volumes to evaluate freeway safety performance and select countermeasures, the real-time ITS data allows a more accurate safety evaluation and prediction model compared to the crash estimation models using AADTs. Real-time crash risk evaluation based on the real-time inductive loop detector data can be used to warn drivers, develop proactive countermeasures, and operate traffic flow to reduce the risk. On the other hand, modeling traffic flow changes after a crash will help engineers to respond better to the congestion caused by freeway crashes and be prepared for different after-crash situations. Understanding how different types of crashes influence traffic flow can also be combined with real-time crash risk estimation, which will be used to reduce the risk of second crashes.

Traffic volume has been considered as one of the significant factors contributing to crashes on freeways (1). The role of traffic flow in crash causation has been discussed and researched for many years. Numerous studies have been established to statistically link freeway crash occurrence and traffic characteristics. Earlier studies usually used long-term historical data such as AADT and hourly traffic volume as input traffic flow for crash prediction models. Gwynn (2) examined the hourly accident rate on a four-lane divided highway in New Jersey and reported that average daily traffic (ADT) was directly correlated to accident occurrence. ADT was considered to be a good predictor of accident

rates. Persaud et al. (3) studied the relationship between crashes and AADT on rural two-lane roads. His study confirmed the non-linear relationship between them.

Because of a lack of real-time traffic flow data, these methods could not provide a real-time warning on crash risk to drivers/operators. A previous study has shown the likelihood of a crash or crash potential is significantly affected by the short-term variance of traffic flow (4). For this reason, the crash risk needs to be evaluated on a real-time basis by monitoring the real-time traffic condition. ITS sensors have been widely deployed to report real-time traffic status, especially along freeways in urban areas. For using the real-time ITS data to estimate the real-time crash risk, the correlation between them needs to be studied with the historical ITS data before, during and after recorded crashes. Researchers have invested effort in developing real-time crash prediction models by using real-time traffic flow characteristics. Golob et al. (5) used ITS data of 30-second observations from inductive loop detectors to study 9,341 crashes on urban freeways. 30 minutes sensor data at each loop detector station were extracted to establish a stable measurement of traffic conditions before each accident. Crash characteristics for six different traffic flow regimes were generated. The results showed that high variance in flow/occupancy could lead to high risk of a crash. However, the specific relationship applied only to conditions during 1988 in Orange County, California. The revealed association between traffic status and crash risk cannot be directly applied to today's regions without local and latest data validation. Newer ITS and crash data are needed to be used to update the result. Another study conducted by Thomas considered the relationships among urban freeway accidents, traffic flow, weather, and light conditions by using ITS data (6). The results indicate that the type of collision is strongly related to median traffic speed and temporal speed variations of the left and interior lanes. The results also show that types of collision are associated with different weather conditions. For example, hit-object collisions and multi-vehicle crashes are more likely to occur on wet roads; furthermore, rear-end accidents are more likely to happen on dry roads during daylight. Lee et al. (4) built a real-time crash prediction model in freeway traffic with incident logs and traffic flow data extracted from loop detectors along a 6.21-miles stretch of the Gardiner Expressway in Toronto, Canada. This research revealed that the speed abruptly drops at the detector station immediately upstream of the crash site. Therefore, the moment of speeds significantly drops can be considered a reasonable estimation of the actual time of the crash. Abdel-Aty (1) studied the relationship between crash frequency and ITS data in Orlando, Florida by using archived loop data and a total of 3,146 crashes from the years 1999 through 2002. The results showed that five minutes before an accident affects its occurrence most significantly.

In summary, these studies prepared good foundational work for this project. However, limitations such as lack of real-time crash prediction models (functions) for implementation still exists. Besides that, existing studies did not address the influence distance of crashes to traffic flow. Furthermore, these studies did not model the influence of accidents on traffic flow after the occurrence of accidents. The study of traffic flow after the crash can help to improve traffic operation and reduce secondary crashes.

Therefore, this research project modeled the interactive relationship between real-time traffic status and crashes considering crash severities and types.

The Center for Advanced Transportation Research and Education (CATER) at the University of Nevada, Reno (UNR) performed the correlation analysis in this project to study the traffic pattern before and after crashes considering various crash severities and crash types. For accurately modeling the correlation between crashes and real-time traffic status, the received sensor data and crash data were preprocessed and filtered with predefined criteria. The crash time is significant for correlation analysis of traffic status and crashes, but the documented crash time may not be accurate. The CATER team developed a method to validate the logged crash time information and correct it when errors are found.

The revealed relationship between traffic status and crashes will help safety engineers to identify the traffic flow factors influencing highway safety, which may lead to better countermeasure selection or development. The developed crash prediction model provides real-time evaluation of crash risk, so it will contribute to the real-time traffic safety management in Nevada in the future. Limited resources of traffic agencies can be used more efficiently for improvement of traffic mobility and safety. The proposed project is closely related to the Theme of SOLARIS (Safety and Operations of Large-Area Rural/Urban Intermodal Systems) Institute, Tier 1 UTC (University Transportation Center) at UNR, - Safety and Operations of Large Area Rural/urban Intermodal Systems, especially in Traffic safety data management and crash mitigation and Technologies for safe traffic operations and management. The knowledge will help traffic engineers for development of safety countermeasures and traffic incident response plans.

This report is organized as follows. The Introduction section briefly discusses the background, literature review, expected benefits and objectives of this research. The second section presents the data source and data processing procedure. The third section introduces the model development effort and results, including components of available technologies for analysis, methodology selection, and analysis of interrelationship between traffic status and crashes, and the prediction models predicting real-time crash risk. The fourth section documents the NDOT Nevada Data Exchange (NDEX) data warehouse that can be an excellent data resource feeding the real-time prediction models. The last section summarizes the major effort and findings in this project.

## 2 DATA AND DATA PROCESSING

---

### 2.1 DATA SOURCE

The CATER team received the 2012 inductive loop detector data (495 sensor stations as shown in Figure 2-1) along freeways I-15, I-215 and US-95 in the Las Vegas area. The data were provided by the Freeway & Arterial System of Transportation (FAST) who is one of the first truly integrated Intelligent Transportation System (ITS) organizations in the U.S. The Regional Transportation Commission of Southern Nevada (RTC) is the official administrator of FAST. NDOT and the RTC are full-fledged funding partners, contributing to the operations and management of FAST.



**Figure 2-1. Inductive loop detector locations of received ITS data**

The data of two urban freeways (IR 15 and US 95) in Las Vegas were used in this project to study the relationship between traffic status and crashes. The following traffic flow characteristics are available in the ITS database and were used in this analysis: speed, traffic volume, and occupancy. Occupancy is the percentage of time in which vehicles are over a vehicle presence detector, such as a loop detector. Lane occupancy is used to predict traffic density — a measure of congestion. ITS data from a total of 410 loop detector stations along 92.7 miles freeway were used in this study.

The 2012 crash data in Las Vegas were obtained from the Nevada Department of Transportation (NDOT) Traffic Safety Engineering (TSE) Division. The following crash properties were considered in this analysis: crash type, severity, location and time. The historical crash data were in the format of Geographic Information System (GIS); each crash was located by the linear referencing system (LRS). The ESRI GIS software package ArcGIS was applied to process the historical crash data. The linear referencing system allows geographic processing with ArcGIS Linear Reference Functions. The historical crash data had been extracted from the Nevada Citation and Accident Tracking System (NCATS) database and were geo-located into GIS by NDOT TSE.

## **2.2 DATA PROCESSING**

### **2.2.1 Spatially relating sensor data and crashes**

Crash locations were determined by route IDs and milepost values in the GIS files; while loop detectors were expressed by the geographic coordinates (longitudinal and latitudinal) in the FAST detector database. The crash data and ITS data could be shown on the same map in the ArcGIS, then can be connected based on their spatial locations. There is a typical influence range of a crash on freeway traffic flow. In Golob et al.'s research (5), 78% crashes were within 0.25 miles distance from a detector station. Thus, this project used 0.25 miles as the threshold for connecting each crash to the related sensor station. In other words, a crash was connected to a sensor station when its distance to the station equals or is shorter than 0.25 miles. A total of 25,019 crashes in 2012 were linked to the selected freeway sensor stations and used for the correlation analysis. The crash filtering process is demonstrated by the flowchart in Figure 2-2.

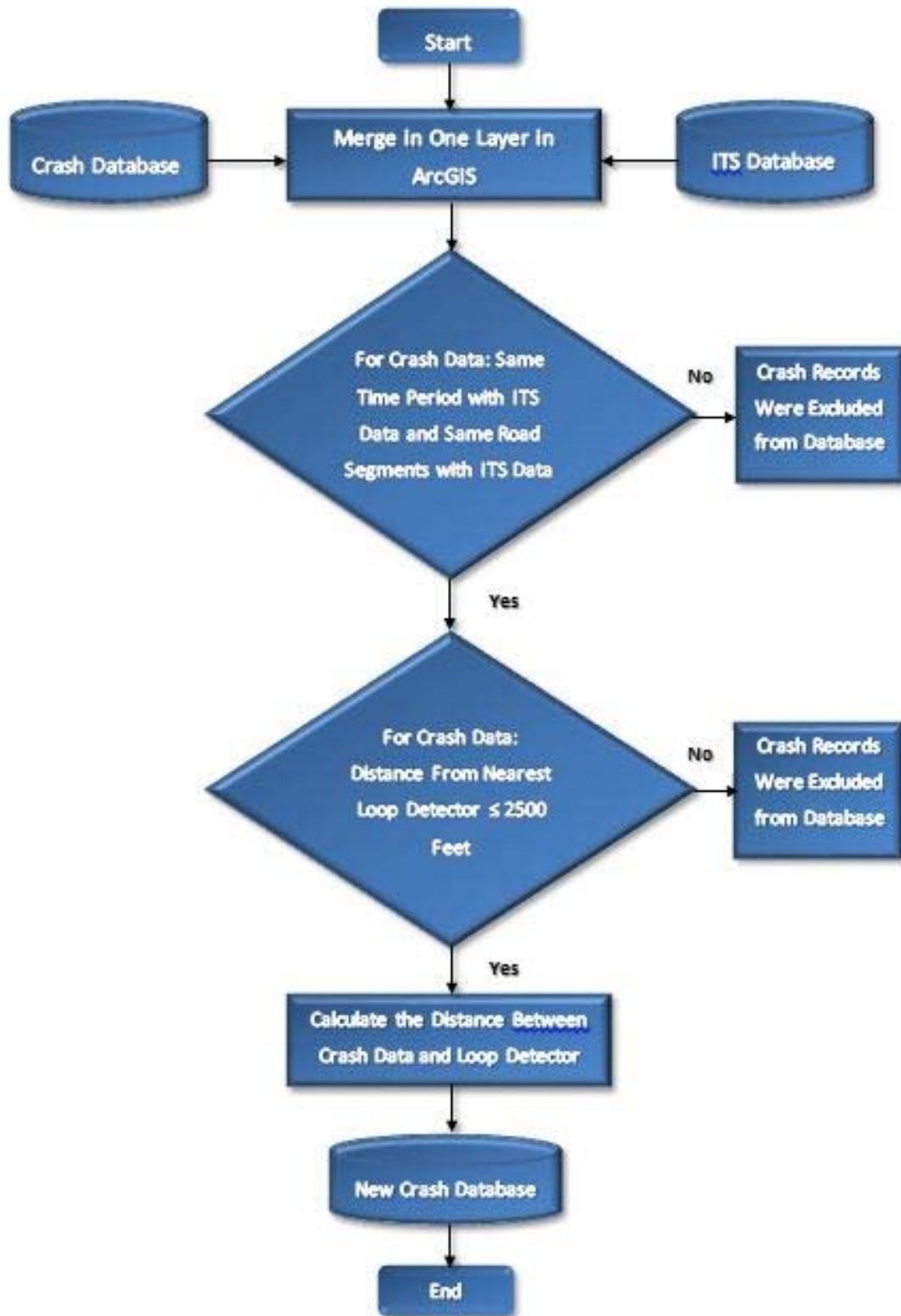


Figure 2-2. Flowchart of crash data processing

### 2.2.2 Temporally relating sensor data and crashes

The accuracy of crash time is significant to the analysis of the correlation between traffic status and crashes. Without accurate timestamp of a crash, it cannot be linked to the actual freeway traffic status, so the output of correlation analysis will not be trustable. The crash data were initially imported from the police accident reports to the NCATS database. The crash time documented in the police report may not be the actual incident moment, it is often an approximated time. The estimation may considerably deviate from the exact crash time, which will impact the accuracy of prediction models. The actual crash time needs to be identified before the model-development in this project. A method to determine the exact crash time was developed in Lee, et al.'s study (4). As introduced in the Introduction section, Lee et al. found that speed abruptly drops at the detector station immediately upstream of the crash site. The time at which the speed drops can be considered a reasonable estimation of the actual time of a crash. However, this method did not examine the influence of other factors on traffic flow. The speed drop may be caused by traffic fluctuation, traffic congestion without crashes or work zone changing road geometry. For eliminating other factors' influence, traffic flow data collected at the same loop detector at the same time-of-day and day-of-week were extracted. It is assumed that the crash time reported by the police has maximum one hour offset from the actual crash time (one hour before or one hour after the crash). One hour of traffic data around the recorded crash time was extracted from the ITS dataset to validate and correct the documented crash time. The threshold for confirming the crash time is shown in equation (1).

$$t_{\text{adjust}} = T_k, \text{ when } \begin{cases} v_{T_k} < v_{T_{k-1}} \\ \frac{v_{T_{k-1}} - v_{T_k}}{v_{T_{k-1}}} \gg 30\% \end{cases} \quad (1)$$

$t_{\text{adjust}}$  is the adjusted crash time.

$T_k$  is the  $k$ th time period (begin from the 1 hour before recorded crash time).

$v_{T_k}$  is the average speed observed in the  $k$ th time period

$v_{T_{k-1}}$  is the average speed in the  $(k-1)^{\text{th}}$  time period in no-crash case

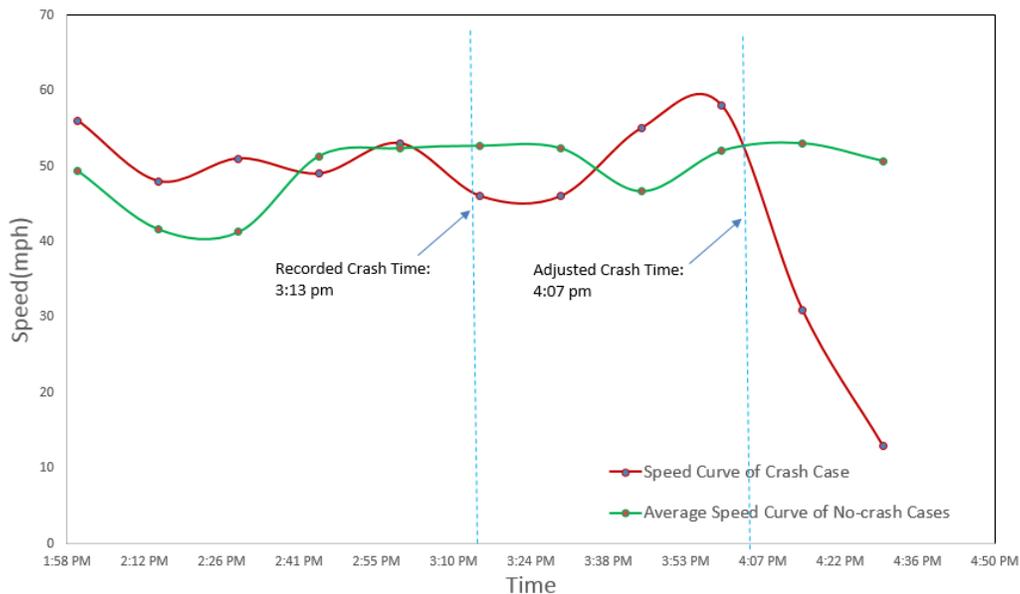
The interpretation of this equation is: if the speed drop after the documented crash time was more than 30% of the average speed (same time-of-day and day-of-week) without a crash, the crash time is considered accurate. An example is given to demonstrate the process. The example-crash happened on I-15 with the documented crash date and time as January 13, 2012 at 3:13 pm. The distance between the crash location and the upstream loop detector was 0.06 miles. The speeds from 2:00 pm to 4:30 pm were extracted from the ITS database. All speed records at the same period of a week without a crash were also retrieved. The detailed speed information is shown in Table 2-1. Figure 2-3 shows the difference between crash and no-crash cases. At 4:15 pm January 13, the speed had dropped by more than 30% compared to the speed at 4:00 pm in no-crash days. The crash time can now be estimated to occur before 4:15 pm with consideration the high

## Correlation Analysis of Freeway Traffic Status and Crashes with Nevada Data

influence to speed after the crash. The adjusted crash time was finally determined as 4:07 pm. This method is utilized to find the most accurate crash time. If no speed and time records meet equation 1, the crash time documented in the NCATS database was used as the accurate time.

**Table 2-1. Speeds of No-Crash Scenarios and Crash Scenarios**

Time	2:00 PM	2:15 PM	2:30 PM	2:45 PM	3:00 PM	3:15 PM	3:30 PM	3:45 PM	4:00 PM	4:15 PM	4:30 PM
<b>Speed of Crash Case(1/13/2012)</b>	56	48	51	49	53	46	46	55	58	31	13
<b>Speed of No-crash Case1(1/6/2016)</b>	62	49	55	60	54	46	52	47	55	45	39
<b>Speed of No-crash Case2(1/20/2016)</b>	24	31	34	45	59	61	54	60	61	59	46
<b>Speed of No-crash Case3(1/27/2016)</b>	62	45	35	49	44	51	51	33	40	55	67
<b>Average Speed of No-crash Cases</b>	49.33	41.67	41.33	51.33	52.33	52.67	52.33	46.67	52.00	53.00	50.67
$\frac{V_{T_{k-1}} - v_{T_k}}{V_{T_{k-1}}} \gg 30\%$	NO	YES	YES								



**Figure 2-3. Estimation of the actual crash time**

### 3 MODEL DEVELOPMENT AND ANALYSIS

---

#### 3.1 AVAILABLE METHODOLOGIES FOR CRASH RISK PREDICTION

Researchers have applied various multivariate statistical methods for correlation analysis between traffic status and crashes. The most commonly used methods include Poisson regression, negative binomial regression, and nonlinear canonical correlation analysis.

Poisson regression requires the response variable to be Poisson-distributed. Fridstrøm et al. (7) applied the Poisson regression model to measure the contribution of randomness exposure, weather, and daylight to the variation of road accident frequency. The research introduced the details of generalized Poisson regression. A disadvantage of Poisson regression is that overdispersion often exists in the results. Hypothesis test results become invalid because of overdispersion or extra variation in the Poisson model (8). Negative binomial regression is usually applied to deal with overdispersion in Poisson regression. Hauer et al. (9) used the negative binomial model to estimate crash frequency on an undivided four-lane urban road. Hadi et al. (8) used negative binomial regression to evaluate safety effects of cross-section design for various highway types. Nonlinear canonical correlation analysis (NLCCA) is a revised form of canonical correlation analysis that allows the variable sets to contain categorical variables. Golob et al. (6) successfully used NLCCA to build the correlation model between crash type, crash severity, and traffic status. In another study, Golob et al. (10) applied the same statistical method to evaluate the safety effects of changes in freeway traffic flow.

#### 3.2 SELECTION OF METHODOLOGY FOR CRASH RISK PREDICTION

Considering the practice in previous research and also the overdispersion in the data, the negative binomial model was selected to analyze the correlations between crash data and traffic flow characteristics. A negative binomial random variable  $Y$  (crash frequency) has a probability function (11) that can be written as equation (2).

$$P(Y = y) = \binom{y + k - 1}{y} \left(\frac{\mu}{\mu + k}\right)^y \left(\frac{k}{k + \mu}\right)^k \quad (2)$$

Where  $y = 0, 1, 2, \dots$ , and parameters  $\mu (= E(Y))$  and  $k$  are positive. The quantity  $1/k$  is known as the dispersion parameter, and the Poisson distribution is the limit as  $k$  approaches infinity. The negative binomial is commonly used to handle overdispersion with respect to the Poisson distribution because  $\text{Var}(Y) = \mu + \mu^2/k > \mu$ . R language was applied to build the negative binomial model and to perform diagnostics. The powerful graphical function of R allows clear depictions of the data and of the analytic results (12). The maximum likelihood estimation of model parameters was performed using “MASS” package developed in R.

The crash frequencies of each segment in 2012 were calculated and used as the base of the model. 2,500 feet upstream each loop detector was used in the calculations. Since the locations of the crashes vary from the nearest detector, the following process was

developed to eliminate the influence of the different distances. The threshold of distance from loop detector to crash location is less than or equal to 2,500 feet. It is assumed that the influence of traffic flow on crashes was highest at a distance of zero feet from the loop detector; the influence of traffic flow is lowest if the distance is 2500 feet.

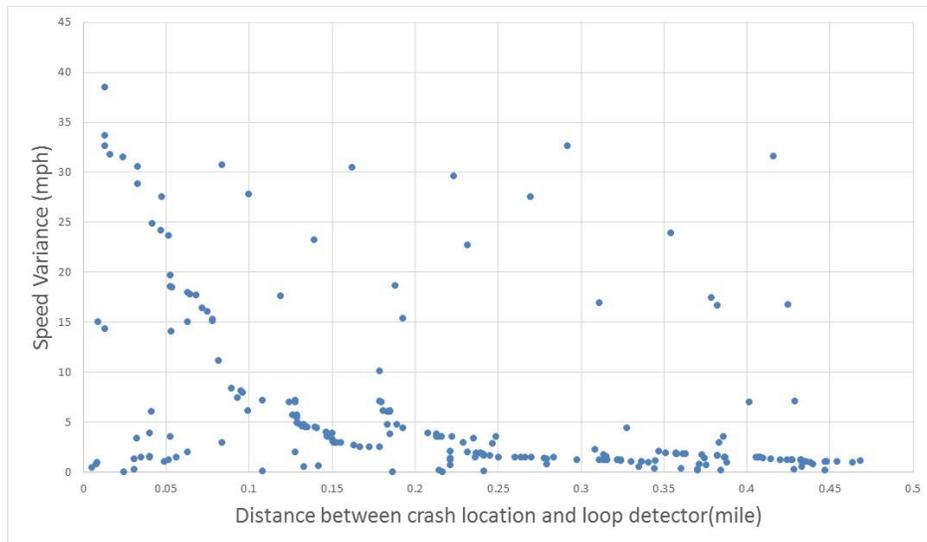
The crash frequencies were calculated (13) by using equation (3).

$$\text{Frequency} = \frac{C}{N} \tag{3}$$

C=Total number of crashes in the study period

N=Number of years of data

Another critical factor considered in the model is the distance between the crash location and the sensor. No current research was found to study the influence on crash risk differentiated by the distance to the sensors. This project analyzed the ITS data and crash data considering the different distance between crashes and their related sensors. The relationship between detected speeds and the distance from the crash location to the detectors is shown in Figure 3-1. The crash data were on January 2012 on the northbound IR15. After excluding all statistical outliers, Figure 3-1 shows that the speed variance decrease as the distance between the crash location and loop detector increases (Note: speed change is defined as the speed in a crash case minus the average speed in a no-crashcase). This mean smaller speed change at a far distance (upstream) can lead to same crash risk as the risk related to a higher speed change at a near location.



**Figure 3-1 Relationship of after-crash speed and distance to the crash location**

No equations were built to reflect this relationship because of dispersed data in Figure 3-1. Based on the finding in Kloeden et al.’s study in 2002 (14), a reserved coefficient was added into Rate to reflect the influence of the distance. The input (y) for the negative binomial was changed into equation (4).

$$y = \text{Frequency} * (1 + 0.3 * d/2500\text{ft}) \quad (4)$$

### 3.3 INFLUENCE OF TRAFFIC STATUS ON CRASH RISK

In order to predict crash risk more accurately, two different scenarios were analyzed separately as shown in Table 3-1.

**Table 3-1. Scenarios for crash prediction model**

Model	Scenarios	y	x1	x2	x3	x4	x5	x6
Negative Binomial	Speed before Crash<Normal Speed	Crash Frequency	Speed	Speed Change	Traffic Volume	Traffic Volume Change	Lane Occupancy	Lane Occupancy Change
	Speed before Crash>Normal Speed	Crash Frequency	Speed	Speed Change	Traffic Volume	Traffic Volume Change	Lane Occupancy	Lane Occupancy Change
Data Source		Calculated from equation 3	Captured by loop detectors					

#### 3.3.1 Scenario 1: speed lower than the normal speed

The summarized regression results of Scenario 1 are shown in Table 3-2. The output file of R software includes the estimate, standard error, z-scores and P value. The estimate of the negative binomial model shows the log relation between variables and crash frequency. Here is an example of the estimate of “volume.” The variable of “volume” has a coefficient of 0.000484. It means that for each one-unit increase in “volume”, the expected log value of crash frequency increases by 0.000484. The P value is the probability of finding the observed results when the null hypothesis of the study question is valid. If a factor has P <0.05, it means the influence of this element is statistically significant. P <0.001 indicates the influence of this factor is statistically highly significant.

The results indicate that traffic volume (with an estimate of 0.000484) in all the studied factors has the most significant influence on the crash frequency when the detected speed is lower than average (no-crash cases). In this scenario, the high traffic volume indicates high forecasted crash frequencies, which means a high crash risk. Although lower speeds may cause higher crash risk, the significant level of the influence is not apparent. No evidence supports that occupancy and occupancy variance may influence crash risk. The real-time crash-frequency prediction function has been developed for Scenario 1 as the following Equation (5):

$$\log(Y) = 5.60135 + 0.000484 * X \quad (5)$$

Where Y is crash rate,

X is traffic volume.

**Table 3-2. Regression Analysis Results of Scenario 1**

Scenario 1: Speed before Crash<Normal Speed					
	Estimate	Standard Error	Z value	Pr(> Z )	Significant Level
<b>(Intercept)</b>	5.60135	0.968688	5.782	7.36E-09	***
<b>speed</b>	-0.02526	0.014317	-1.765	0.07763	.
<b>speed change</b>	-0.01241	0.012175	-1.02	0.30797	
<b>volume</b>	0.000484	0.000156	3.101	1.93E-03	**
<b>volume change</b>	0.000466	0.0003	1.555	0.12	
<b>occupancy</b>	-0.01551	0.024259	-0.639	0.52271	
<b>occupancy change</b>	-0.0258	0.026014	-0.992	0.32126	
Significant Code: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

**3.3.2 Scenario 2: speed higher than normal speed**

Table 3-2 shows the regression results of the second scenario. The results show that speed, speed change and volume have the most significant influence on crash rate. Crash risk rises when speed, speed change or traffic volume become higher. The impact of occupancy on crash rate is not significant compared to the three elements introduced above. The real-time crash-frequency prediction function has been developed for Scenario 2 as the following Equation (6):

$$\log(Y) = 8.497172 + 0.073227 * X1 + 0.079802 * X2 + 0.001194 * X3 \quad (6)$$

Where Y is crash rate

- X1 is speed
- X2 is speed change
- X3 is traffic volume

**Table 3-3. Regression Analysis results of Scenario 2**

<b>Scenario 2: Crash Speed&gt;Normal Speed</b>
--

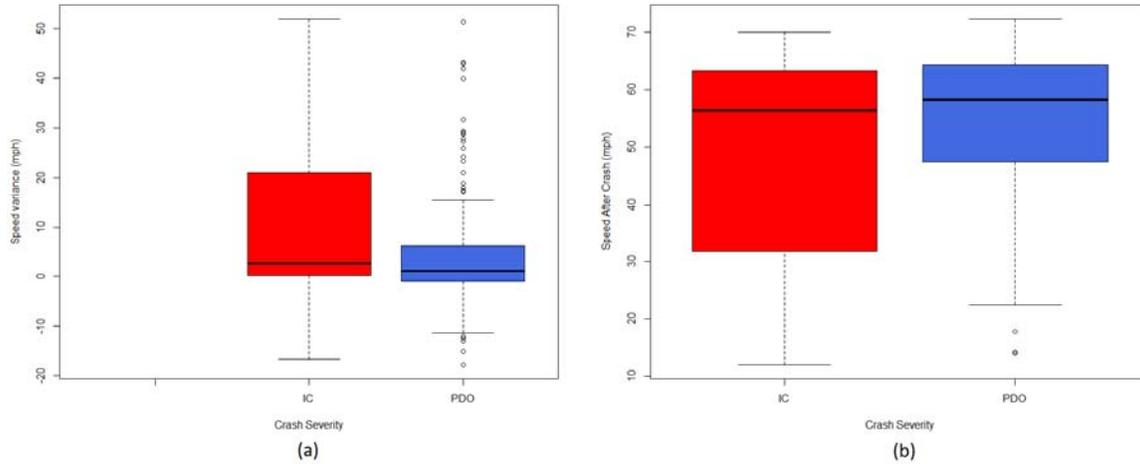
## Correlation Analysis of Freeway Traffic Status and Crashes with Nevada Data

	Estimate	Standard Error	Z value	Pr(> Z )	Significant Level
<b>Intercept</b>	8.497172	1.3865099	6.128	8.87E-10	***
<b>speed</b>	0.073227	0.0193261	3.789	0.000151	***
<b>speed change</b>	0.079802	0.0217197	3.674	0.000239	***
<b>volume</b>	0.001194	0.0002397	4.981	6.34E-07	***
<b>volume change</b>	0.001211	0.0006457	1.875	0.060727	.
<b>occupancy</b>	-0.10585	0.0455303	-2.325	0.020079	*
<b>occupancy change</b>	-0.10275	0.04703	-2.185	0.028901	*
Significant Code: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

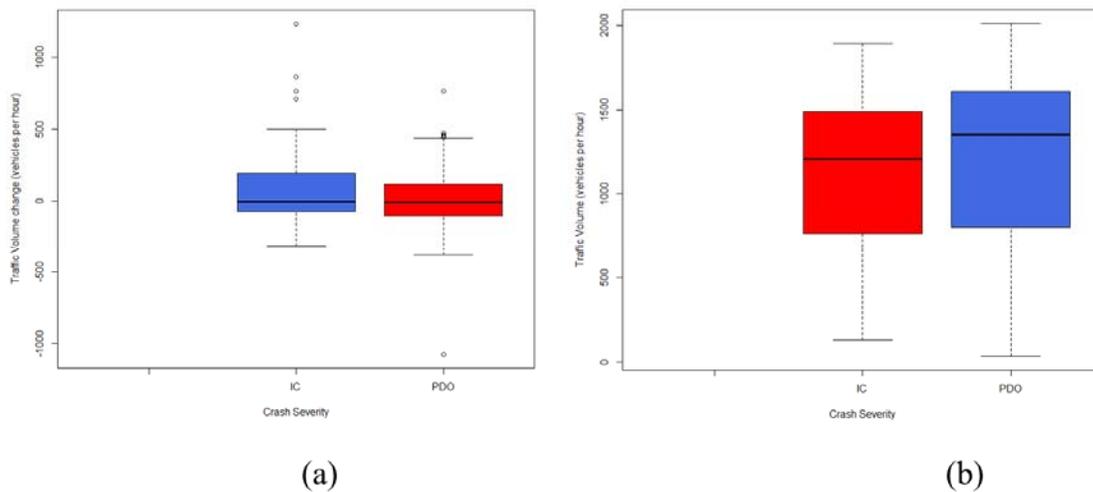
### 3.4 INFLUENCE OF CRASHES ON TRAFFIC FLOW

For analysis of crashes' influence on after-crash traffic status, crashes were classified into two groups with consideration of severities: injury crashes (IC) (including fatalities) and property-damage-only (PDO) crashes. Four freeway crash types were also considered in this analysis: non-collision, rear-end, angle and sideswipe. The average speed, speed change, traffic volume and volume change in one hour after crash occurrence were analyzed for understanding traffic patterns influenced by crashes. The influence of incident severity and crash type on traffic flow was plotted into box charts. Figure 3-2 and 3-3 show the relationship between crash severity and status of after-crash traffic flow. The box charts show the min, 25%, median, 75%, and max values. The circles in the box charts are outliers. After an accident, the traffic speed will drop compared with speed in no-crash cases (shown in Figure 3-2 a). After-crash speeds of PDOs had more outliers than ICs. The results of this component directly correlate to the scenario 1 of the crash rate prediction model introduced in Section 3.1. It means that the reduced after-crash speeds will bring higher crash risk to the traffic flow. In this analysis, after-crash speeds and volumes were studied with the scenario 1 crash prediction model to understand their influence on the risk of secondary crashes. Injury crashes lead to a much lower overall speed than the average speed (shown in Figure 3-2 b). This lower speed could potentially result in a higher risk of a second crash based on the developed crash rate prediction model. If the crash severity is high, it causes a sharp drop in speed, which is a more serious threat to traffic safety. Figure 3-3 shows the influence of PDO crashes and injury crashes on traffic volumes. The after-crash traffic volumes and volumes are influenced similarly by PDO crashes and injury crashes.

## Correlation Analysis of Freeway Traffic Status and Crashes with Nevada Data



**Figure 3-2 Relationship between crash severity and after-crash speed**



**Figure 3-3 Relationship between crash severity and after-crash traffic volume**

Figure 3-4 and 3-5 show how crashes of different types impact traffic flows (after-crash traffic). Angle and rear-end crashes caused sharp speed drop (Figure 3-4). The reduced speed leads to higher crash risk based on the crash prediction model developed for Scenario 1. Traffic engineers and incidence response teams should pay attention to the risk of the second crashes after angle or rear-end crashes. Non-collision and sideswipe crashes have low influence on traffic speed. Many outliers of speed change in non-collision and sideswipe crashes mean that no evident trend exists. There is no noticeable difference in traffic volume change between different crash types (Figure 3-5 a). The traffic volumes after angle and rear-end crashes were higher than traffic volumes after non-collision and sideswipe crashes (Figure 3-5 b). The higher traffic volumes also constitute the potential raised risk of a secondary accident.

## Correlation Analysis of Freeway Traffic Status and Crashes with Nevada Data

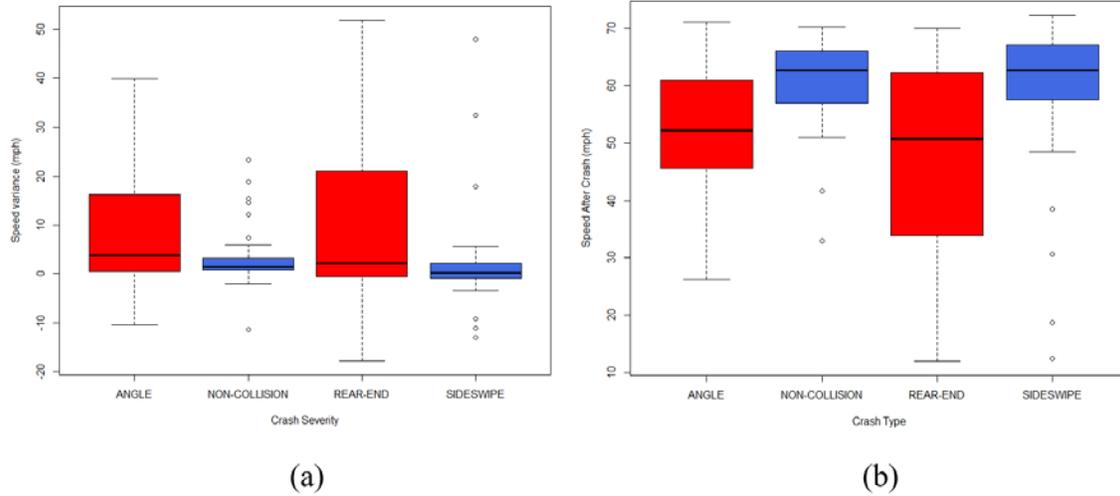


Figure 3-4 Relationship between crash type and after-crash speed

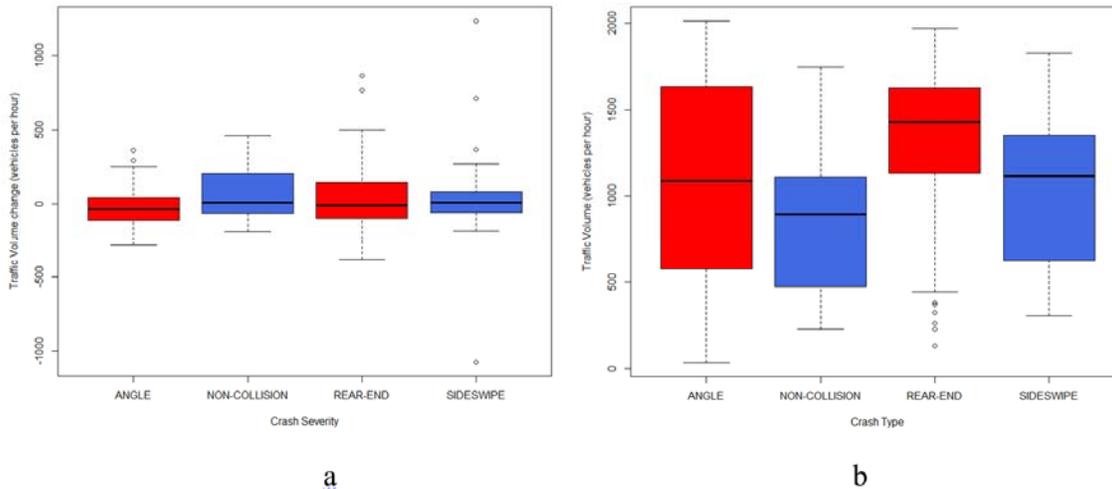


Figure 3-5 Relationship between crash type and after-crash traffic volume

### 3.5 INFLUENCE OF DIFFERENCE BETWEEN AVERAGE SPEED AND SPEED LIMIT

This research also analyzed how the difference between the average speed and the speed limit influences the crash risk. The average speeds along the studied road segments were recorded every 1 minute by the loop detectors. However, the actual record interval may be longer than 1 minute because of the instability of loop detectors or other human-made destruction. For a comprehensive and accurate understanding of the influence of speed difference from the speed limit, ten-minute data before each crash were checked to exclude the records with the interval longer than 1 minute. A total of 2,428 crash records with 1-minute-interval speed data were used for this analysis. The speed difference between the 10-minute average speed and the speed limit, named speed deviation in this study, was calculated with Equation (7). The range of speed deviation was from -55 mph to 10 mph.

$$V_d = V_a - V_l \tag{7}$$

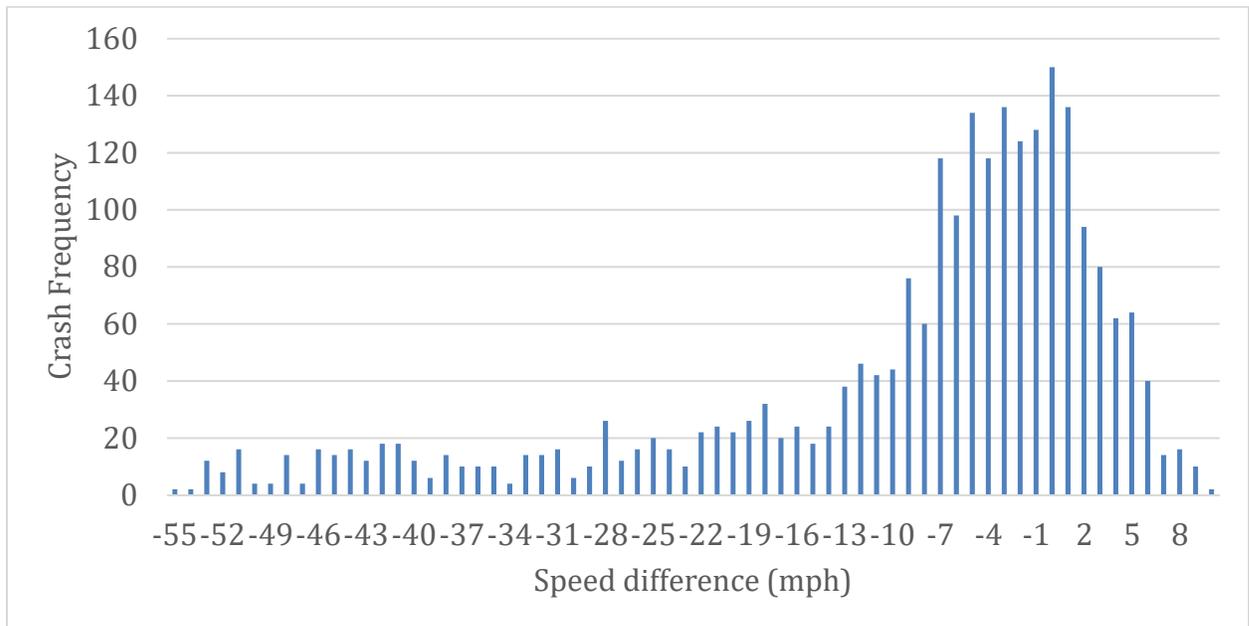
Where

$V_d$  is speed deviation

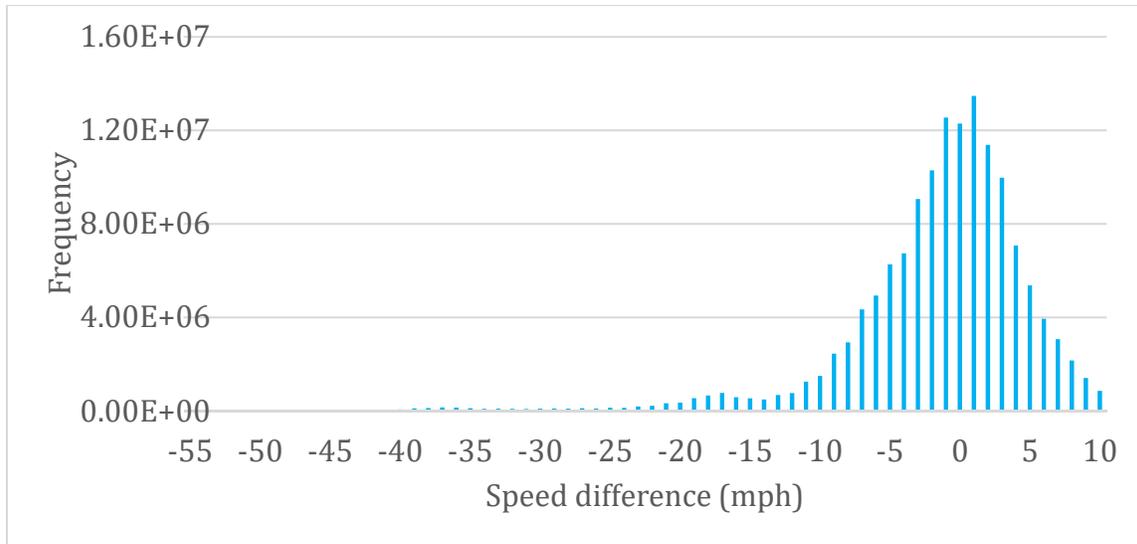
$V_l$  is speed limit, which is 65mph in this research,

$V_a$  is ten minutes average road section speed before crash

The relationship between crash frequencies and speed deviations was then studied, as shown in Figure 3-6. The frequency distribution chart shows that most crashes happened when speed deviation was from -10mph to 5mph. However, the crash frequency distribution itself cannot reasonably explain the influence of the speed deviation. The high crash frequency could be caused by the high frequency of the related speed range on the freeway segments. Therefore, Figure 3-7 summarized the frequency of speed deviation in the studied dataset.



**Figure 3-6 Frequency distribution of crashes under different speed deviations**



**Figure 3-7 Frequency distribution of different speed deviations**

Crash risk index, which can describe crash risk at different speed deviations, is estimated with Equation (8) to consider both crash frequency and speed range frequency.

$$y = F / F_{vd} \quad (8)$$

Where

y is crash risk index

F is the crash frequency

$F_{vd}$  is the speed deviation frequency

A higher index indicates that the crash risk of the speed difference between a 10-minute average speed and the speed limit. Figure 3-8 shows the scatter plot of crash risk index. An exponential equation (9) was generated, which fits the data well. The  $R^2$  of this equation is 0.8337 that means the relationship between speed deviation and crash risk index is strong.

$$y = 1E-05e^{-0.063x} \quad (9)$$

Where

x is speed deviation

y is crash risk index

When the average speed was higher than the speed limit, the influence of speed deviation on the crash risk is not significant. It should be noted that this conclusion is limited to the scenario that the speed deviation is lower than 10mph.

Correlation Analysis of Freeway Traffic Status and Crashes with Nevada Data

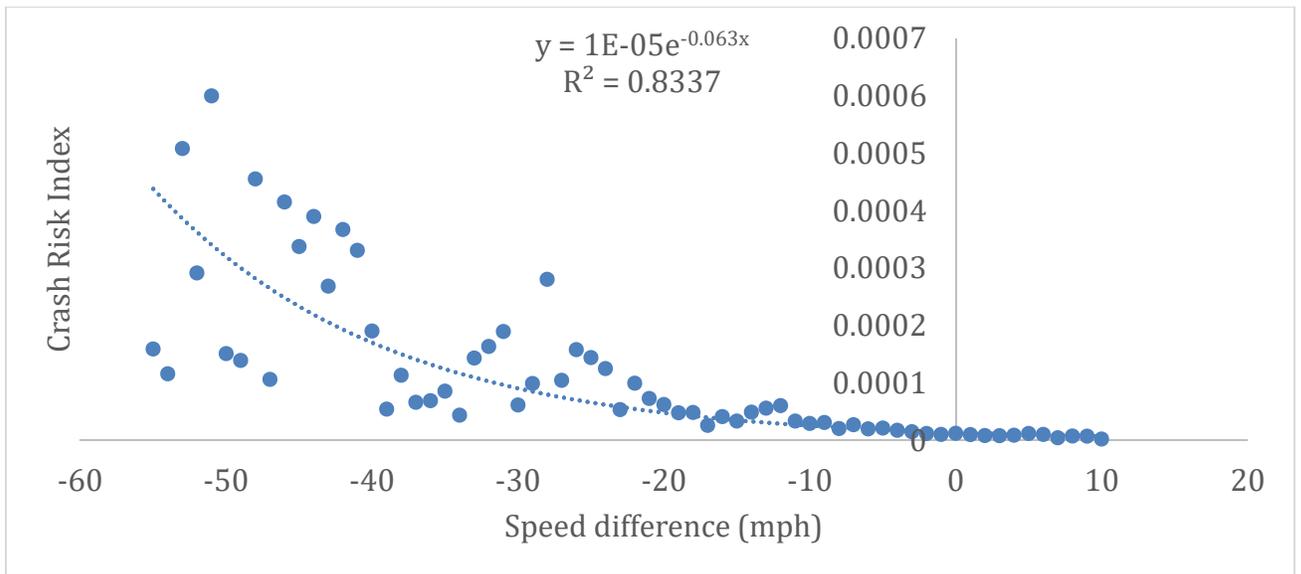


Figure 3-8 Relationship of speed deviation and crash risk

## 4 NDOT NDEX

---

NDOT has been collecting and managing different traffic data through various data platforms. The NDOT Traffic Operation Division has implemented a central data warehouse, NDEX, for accessing and sharing the traffic data. NDEX was expanded to accommodate different data sources and data formats. As the real-time ITS data is available through the NDOT NDEX system, the real-time models to evaluate crash risks can be combined with the real-time data to estimate the real-time crash risks along freeways in Nevada. With NDEX, crash evaluation and prediction could be made more efficient by pulling data and sharing results in real time.

A department of transportation has several different types of disparate management platforms that poll for the status of devices such as Roadway Weather Information Systems (RWIS), Highway Advisory Radio (HAR), traffic detectors, CCTV cameras, Dynamic Message Signs (DMS), etc. To facilitate information exchanges between DOT agencies and state's partners, the Institute of Transportation Engineers (ITE) has developed the Traffic Management Data Dictionary (TMDD) standard version 3.01 (<http://www.ite.org/standards/tmdd/>) for Traffic Management Center-to-Center Communications. The TMDD standard aims at providing standards-based, high-level definitions in a protocol-independent manner, which a system interface specification can be prepared and a data exchange system can be developed.

Based on the ITE TMDD, NDOT has designed, planned, and implemented the central data warehouse NDEX. Normal implementation of the Transportation data warehouse is to push data in a single direction into the data warehouse. The new data service improved communication by pushing data from the data warehouse into the data management platforms, so it allows the data management platforms to receive data from other Divisions. This allows any jurisdiction to pull all or some of the data being published by another jurisdiction. The NDEX work is complete and waiting for the development of data source management platforms.

### 4.1 NDEX WEB SERVICES

NDEX implemented the web services with the technologies of the Web Services Description Language (WSDL) (<https://www.w3.org/TR/2007/REC-wsdl20-20070626/>), XML Schema Definition (XSD) ([https://www.w3schools.com/xml/schema\\_intro.asp](https://www.w3schools.com/xml/schema_intro.asp)), and Extensible Markup Language (XML) (<https://en.wikipedia.org/wiki/XML>).

#### 4.1.1 WSDL

WSDL provides a model and an XML format for describing Web services. WSDL 2.0 enables one to separate the description of the abstract functionality offered by a service from concrete details of a service description such as “how” and “where” that functionality is offered.

WSDL 2.0 describes a Web service in two fundamental stages: one abstract and one concrete. Within each stage, the description uses a number of constructs to promote reusability of the description and to separate independent design concerns.

At an abstract level, WSDL 2.0 describes a Web service in terms of the messages it sends and receives; messages are described independent of a specific wire format using a type system, typically XML Schema. At a concrete level, a binding specifies transport and wire format details for one or more interfaces. An endpoint associates a network address with a binding. And finally, a service groups together endpoints that implement a common interface.

A WSDL 2.0 service description indicates how potential clients are intended to interact with the described service. It represents an assertion that the described service fully implements and conforms to what the WSDL 2.0 document describes. A WSDL 2.0 interface describes potential interactions with a Web service, not required interactions. The declaration of an operation in a WSDL 2.0 interface is not an assertion that the interaction described by the operation must occur. Rather it is an assertion that if such an interaction is (somehow) initiated, then the declared operation describes how that interaction is intended to occur.

### **4.1.2 XSD**

XSD (XML Schema Definition) is a World Wide Web Consortium (W3C) recommendation that specifies how to formally describe the elements in an Extensible Markup Language (XML) document. This description can be used to verify that each item of content in a document adheres to the description of the element in which the content is to be placed. XSD 1.1 became an approved W3C standard in April 2012.

### **4.1.3 XML**

In computing, Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. The W3C's XML 1.0 Specification and several other related specifications—all of them free open standards—define XML.

The design goals of XML emphasize simplicity, generality, and usability across the Internet. It is a textual data format with strong support via Unicode for different human languages. Although the design of XML focuses on documents, the language is widely used for the representation of arbitrary data structures such as those used in web services.

Several schema systems exist to aid in the definition of XML-based languages, while programmers have developed many application programming interfaces (APIs) to aid the processing of XML data.

## **4.2 NDEX CONNECTION GUIDANCE**

This section summarizes the procedure for connecting to the NDEX data warehouse based on the information offered by the NDOT Traffic Operation Division. The first step

is to obtain username and password from the NDOT Traffic Operation Division because NDEX is a restricted access based on authentication. Before programming the client to retrieve data or publish data on NDEX, SoapUI (<http://soapui.org>) is recommended for testing connections with NDEX, so the user can validate whether the authorized access is the needed one. The NDEX WSDL endpoint (<http://coloNDEXsrv.its.nv.gov/tmddws/TmddWS.svc>) can be accessed to download the WSDL file defining the web service of NDEX. In SoapUI, the user can create a project using the NDEX WSDL endpoint. Then SoapUI automatically generates EC and OC interfaces and dialogues administrator. The generated request messages (in the XML format) include the information of what data are requested, the authorized username and password, and other elements. After editing the request messages with the assigned username and password, the request message can be sent to the NDEX server. With the request message, the NDEX server will verify the username and password information, then determines whether to return the requested data. When the connection to NDEX and the received information being validated in SoapUI, the real-time data request and processing can be coded with different program languages for saving to a database or being processed in real time. Therefore, the real-time crash risk prediction along freeways can be applied to the real-time ITS data from NDEX, and the crash risk evaluation results can be shared through NDEX. The implementation of the real-time crash prediction system can be considered in the future projects.

NDEX currently includes the following data:

- Detector Station
- CCTV
- Dynamic Message Sign (DMS)
- Environmental Sensors (ESS)
- Highway Advisory Radio (HAR)
- Incidents/Events
- Ramp Meter
- Node, Link, Traffic Network

## 5 CONCLUSION AND DISCUSSION

---

This research project analyzed the correlation between freeway traffic status and crash risks with the freeway ITS data from the Las Vegas FAST. The research team identified the influence of different traffic flow characteristics (speed, speed change, volume, volume change, occupancy, and occupancy change) on the real-time crash risks. The UNR CATER team also developed negative binomial models to predict crash risk with traffic flow information from ITS sensors. This project studied the influence of crash characteristics on traffic flow after accidents, which determines the risk level of second crashes. The following findings are concluded based on the research results:

1. If observed speed is lower than average speed, there is a strong relationship between traffic volume and crash risk. The higher traffic volume, the higher the crash risk; lower speed also indicates a higher risk of a crash.
2. Speed, speed change and traffic volume have significant influence to crash risk if observed traffic speed is higher than normal speed. High speed, high variance and traffic volume will lead to high crash risk.
3. When crash severity increases, the speed change of following vehicles will increase which is harmful to traffic safety. Traffic volume will also decrease caused by low traffic capacity.
4. Angle and Rear-end crashes have significant influence to after-crash traffic flow when compared to other crash types. These two crashes are identified to the most dangerous crash types on freeways.
5. This research introduced a new method to better estimate the crash time. By comparing the speed change between crash case and no-crash case, the crash time can be better approximated. This new method will improve the accuracy of the regression model. A case control method was generated to estimate the actual crash time. The time point when the average speed in crash case has a deviation more than 10% of average speed in non-crash cases recorded in the same time of day and same location is considered as the real crash time.
6. Crash risk index is put forward to indicate the probability of potential crash. A high crash risk index means the probability of potential crash is high. The relationship between speed deviation and crash risk index fits the exponential distribution very well. The equation has a R<sup>2</sup> with 0.8337, which means there is strong relationship between speed deviation and crash risk index.

The ITS data and crash data used in this research is collected at 2012, new ITS data and crash data are expected to be used to improve the regression model. The sample data of average speed higher than speed limit is very limited in the available database. The conclusion that “the influence of speed deviation on crash risk index is not significant” does not suit the situation when average speed is 10mph higher than speed limit. More

## Correlation Analysis of Freeway Traffic Status and Crashes with Nevada Data

ITS data and crash data are still needed to explore the relationship between speed deviation and crash risk index for the situation that average speed is higher than speed limit. More ITS data will be obtained to improve upon the accuracy of the results of the regression model. The NDOT NDEX data warehouse is an excellent data source to feed the real-time prediction models. The data can also be used to further improve the accuracy of the developed models. Negative binomial model has been applied in analyzing the relationship between crash data and traffic flow, while NLCCA is considered to be another good analytic approach to study the correlation. Data mining may also be used to search the relationship.

## REFERENCES

---

1. Abdel-Aty, M, et al. "Predicting freeway crashes from loop detector data by matched case-control logistic regression." *Transportation Research Record: Journal of the Transportation Research Board* 1897 (2004): 88-95.
2. Gwynn, D.W. "Relationship of accident rates and accident involvements with hourly volumes." *Traffic Quarterly* 21.3 (1967).
3. Persaud, B.N., Retting, R.A. and Lyon, C.A. "Crash reduction following installation of centerline rumble strips on rural two-lane roads." *Accident Analysis & Prevention* 36.6 (2004): 1073-1079.
4. Lee, C., Hellinga, B. and Saccomanno, F. "Real-time crash prediction model for application to crash prevention in freeway traffic." *Transportation Research Record: Journal of the Transportation Research Board* 1840 (2003): 67-77.
5. Golob, T.F. and Recker, W.W. "A method for relating type of crash to traffic flow characteristics on urban freeways." *Transportation Research Part A: Policy and Practice* 38.1 (2004): 53-80.
6. Golob, T.F., and Recker, W.W. "Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions." *Journal of transportation engineering* 129.4 (2003): 342-353.
7. Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R. and Thomsen, L.K. "Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts." *Accident Analysis & Prevention* 27.1 (1995): 1-20.
8. Hadi, M.A., Aruldhas, J., Chow, L.F. and Wattleworth, J.A. "Estimating safety effects of cross-section design for various highway types using negative binomial regression." *Transportation Research Record* 1500 (1995): 169.
9. Hauer, E., Council, F. and Mohammedshah, Y. "Safety models for urban four-lane undivided road segments." *Transportation Research Record: Journal of the Transportation Research Board* 1897 (2004): 96-105.
10. Golob, Thomas F., Wilfred W. Recker, and Veronica M. Alvarez. "Tool to evaluate safety effects of changes in freeway traffic flow." *Journal of Transportation Engineering* 130.2 (2004): 222-230.
11. Yee, T.W. "The VGAM package." *R News* 8.2 (2008): 28-39.
12. Wu, J., Xu, H., Sun, Y. and Geng, X. "Effect of Road Characteristics and Driving Cycles on Accident Risk on Full-Access-Control Highways." *Transportation Research Board*, No. 17-00940 (2017).
13. Highway safety manual. Washington, DC: American Association of State Highway and Transportation Officials (2010).
14. Kloeden, C.N., McLean, J. and Glonek, G.F.V. "Reanalysis of travelling speed and the risk of crash involvement in Adelaide South Australia." *Australian Transport Safety Bureau*, 2002.



Nevada Department of Transportation  
Rudy Malfabon, P.E. Director  
Ken Chambers, Research Division Chief  
(775) 888-7220  
kchambers@dot.nv.gov  
1263 South Stewart Street  
Carson City, Nevada 89712