

EXPLORING RELATIONSHIPS BETWEEN HORIZONTAL
CURVE ROADWAY DEPARTURE CRASHES AND
GEOMETRIC DESIGN CONSISTENCY ON
RURAL, TWO-LANE HIGHWAYS

by

Mingde Lin

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Civil and Environmental Engineering

The University of Utah

December 2017

Copyright © Mingde Lin 2017

All Rights Reserved

ABSTRACT

In 2014, there were 17,791 fatalities as a result of roadway departure crashes in the U.S., representing 54% of all traffic fatalities in the U.S. Roadway departure crashes account for approximately 52% of traffic fatalities in the state of Utah. A significant number of roadway departure crashes occur on horizontal curves along rural, two-lane highways. Previous research has indicated that providing “consistent” designs that are compatible with driver expectations and capabilities can reduce the number of roadway departure crashes at these locations. Various measures of design consistency have been proposed to quantify the levels by which a road design meets driver expectations and capabilities, including speed differentials, alignment indices, and visual demand/work load estimates. Among them, alignment indices have been proven as direct design consistency measures to analyze crash frequency.

The objective of this research was to estimate relationships between the expected frequency of horizontal curve roadway departure crashes and geometric design consistency, characterized by using alignment indices along rural, two-lane highways in Utah. Negative binomial and zero-inflated negative binomial regression models were estimated which relate expected frequencies of roadway departure crashes to design and traffic characteristics of the rural, two-lane road segments. The dataset consists of 578 horizontal curves with corresponding design and traffic information, as well as characteristics of the upstream and downstream tangents and curves. Horizontal

alignment indices, curve lengths, average daily traffic volumes (ADTs), and general geometric variables were tested in the model specifications. To build the dataset for model estimation, roadway features were gathered along rural, two-lane state routes in Utah using the Utah Department of Transportation's LIDAR files. Crash data were also provided by the Utah Department of Transportation for these same routes and spanned the years 2008 through 2014. Eventually, the best two models were explored in this study. One model included the following parameters: the natural logarithm of average annual daily traffic, the changed radius rate, vertical curvature change rate, maximum change in degree of curvature, indicator variable for the presence of a vertical curve on a horizontal curve, and average grade. The other model had the same variables as the first model, but the ratio of average radius over radii replaced the changed radius rate and the average change in degree of curvature replaced the maximum change in degree of curvature.

TABLE OF CONTENTS

ABSTRACT	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
ACKNOWLEDGEMENTS	ix
Chapters	
1. INTRODUCTION	1
1.1 Problem Statement	1
1.2 Research Objective and Scope	3
2. LITERATURE REVIEW	5
2.1 Background of Design Consistency	5
2.2 Overview of Design Consistency Measures	7
2.3 Background of Alignment Indices	9
2.3.1 Horizontal Alignment Indices	10
2.3.2 Vertical Alignment Indices and Composite Alignment Indices	14
2.4 Count Models	16
2.4.1 Poisson Model and Poisson Lognormal Model	16
2.4.2 Negative Binomial Model	18
2.4.3 Zero-Inflated Model	19
2.5 Background of Data Collection Methods	20
3. RESEARCH METHODS	24
3.1 Negative Binomial Model	24
3.2 Zero-Inflated Negative Binomial Models	27
4. DATA COLLECTION	32
4.1 UDOT Data Files	32
4.1.1 Horizontal Curve Estimation and Validation	32
4.1.2 Visual Screening of Data in Google Earth	35
4.1.3 Final Horizontal Curve Segment Entity Database	39

4.1.4 AADT and Post Speed	40
4.2 Roadway Inventory Data File	41
4.3 Crash Data File	43
4.4 Variable Definitions and Descriptive Statistics	45
5. DATA ANALYSIS RESULTS	61
5.1 Relationship between Roadway Departure Crashes and Individual Design Consistency Measures.....	61
5.2 Relationship between Roadway Departure Crashes and All Design Consistency Measures.....	64
5.3 Exploring “Excessive” Zero Roadway Departure Crashes and All Design Consistency Measures.....	66
5.4 Model Selections.....	68
6. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS.....	78
6.1 Summary	78
6.2 Findings and Conclusions	79
6.2.1 Finding and Conclusion 1	80
6.2.2 Finding and Conclusion 2	81
6.2.3 Finding and Conclusion 3	81
6.2.4 Finding and Conclusion 4	82
6.3 Recommendations.....	83
REFERENCES	85

LIST OF TABLES

Table

1. The Califso et al. Design Consistency Evaluation on Alignment Indices	23
2. Raftery's (1995) BIC and Hilbe's (2011) AIC for Significant Levels	31
3. Minimum Design Tangent Length.....	50
4. Roadway Departure (RWD) Crash Descriptions.....	52
5. General and Crash Variables Descriptions	54
6. Horizontal Variables Descriptions.....	55
7. Vertical Variables Descriptions	57
8. Summary Descriptive Statistics for General and Crash Disaggregate Data.....	58
9. Summary Descriptive Statistics for Horizontal Disaggregate Data.....	59
10. Summary Descriptive Statistics for Vertical Disaggregate Data.....	60
11. Negative Binomial Models with Individual Design Consistency Measures	71
12. Model A with Design Consistency Measures.....	72
13. Model B with Design Consistency Measures	73
14. Zero-Inflated Negative Binomial Model A with Design Consistency Measures	75
15. Zero-Inflated Negative Binomial Model B with Design Consistency Measures	76
16. Comparisons between Zero-Inflated Negative Binomial and Negative Binomial in Model A and Model B (578 Observations).....	77

LIST OF FIGURES

Figure

1. Example of “broken” segments of a horizontal curve.	47
2. A screen capture of “Calculate Geometry” tool in ArcGIS for converting coordinates from degrees into meters.....	47
3. An example of curve with accurate information.....	48
4. An example of discrepancy between curve length and GPS coordinates.	48
5. An example of curve at or near intersection.	49
6. Example of winter closure information in Google Earth.....	49
7. An example of manual measurement in Google Earth.	50
8. Superelevation validation comparisons	50
9. Types of vertical alignments.....	51
10. Example of vertical alignment inside of horizontal curve.....	51
11. Example of video image from Web Navigator.	52
12. Alternative roadway departure (RWD) crash frequency comparisons among 7 years (2008-2014) in Utah.....	53
13. Roadway departure crashes distribution	74
14. Probability of roadway departure (RWD) crashes among different models.....	77

ACKNOWLEDGEMENTS

I would like to express my deep appreciation to my supervisor, Dr. Richard J. Porter, for his strong and continuous support of my research efforts and academic courses. I am very grateful to his excellent guidance, insight, and patience. I hope to use his professional advice to be thoughtful, articulate, and persistent in my future career.

I would also like to thank my committee members, Dr. Milan Zlatkovic and Dr. Xiaoyue Liu, for their kind support of and insights into my master's program. Dr. Milan Zlatkovic gave me strong support and very helpful guidance on my first research project in the first year of my Master's study. Dr. Xiaoyue Liu gave her thoughtful suggestions and career insights as well.

I thank the Federal Highway Administration (FHWA) and the University of Utah for funding this research. I also thank the Utah Department of Transportation for the shared research database which was captured by the Mandli Communications Consultant Company. I especially thank Thanh Le who handed over a significant polished dataset to me and gave me helpful directions as to my data manipulation. I also would like to thank Research Assistant Professor Juan Medina and Brendan Sean Duffy, who provided safety crash data for this research.

I am also grateful to my Utah traffic lab mates, Jeffrey Taylor, Yu Song, Anusha Musunuru, Zhuo Chen, Ivana Tasic, Kiavash Fayyaz, Jem Locquiao, and Michael Scott, for their companionship and selfless assistance. I would like to give my special thanks to

Jeffrey Taylor, Yu Song, and Anusha Musunuru, who always gave me friendly and kind suggestions on my research without any hesitation. In particular, Jeffrey Taylor helped me and provided excellent relevant studying resources on my validation of the database process and reviewed almost all of my work and Yu Song helped me to prepare for my defense presentation slides. In addition, I much appreciate all of my friends for being around me through my ups and downs. I could not have completed this research without your inspiration and encouragement.

Most importantly, I really want to thank my loving family. My parents and my grandparents have always given me unconditional love and hope. All of you encouraged and inspired me to keep moving forward without fear.

CHAPTER 1

INTRODUCTION

This chapter consists of two sections. First, the problem statement will provide an overview of traffic safety in the U.S. and define roadway departure crashes and design consistency. The second section defines the research objective and scope, and outlines the tasks required to accomplish the research objectives.

1.1 Problem Statement

Millions of people are killed or injured in highway crashes each year in the United States (U.S.). The National Highway Traffic Safety Administration (NHTSA) (2015) estimates the cost of motor vehicle crashes to be approximately \$871 billion per year. The societal cost of traffic injuries and fatalities includes personal harm and suffering, as well as economic losses. Millions of families have been mentally harmed as a result of losing their relatives in traffic crashes. Even though traffic fatalities in the U.S. have decreased in the last few years, more than 30,000 still occur each year. The American Association of State Highway and Transportation Officials (AASHTO) calls for a reduction of 1,000 fatalities per year to achieve their goal of a 50% reduction by 2030. AASHTO, as well as the Federal Highway Administration (FHWA) and state departments of transportation (DOTs), have also embraced the “Towards Zero Death”

traffic safety vision.

Roadway departure crashes have constituted a majority of highway fatalities in recent years. The FHWA defines a roadway departure crash as “a crash which occurs after a vehicle crosses an edge line or a center line, or otherwise leaves the traveled way” (FHWA, 2015). In 2014, 17,791 fatalities resulted from roadway departure crashes, which represented 54% of all traffic fatalities in the U.S. There were 54,036 motor vehicle crashes in Utah during 2014, which resulted in 23,364 injuries and 256 deaths. Failure to keep in the proper lane was identified as a crash cause in approximately 12% of all crashes and 20% of fatal crashes (Utah Department of Public Safety Highway Safety Office, 2014). Average annual roadway departure fatalities from 2007 to 2013 were approximately 52% of all fatalities in Utah (Jalayer, Mohammad, and Zhou, 2016). A significant number of roadway departure crashes occur on horizontal curves along rural, two-lane highways. (FHWA, 2016)

The AASHTO Highway Safety Manual (HSM) includes an abundance of analytical methodologies and techniques to estimate the expected number of all types of crashes on road segments or intersections (AASHTO, 2010). However, the current crash statistic methodologies in the HSM still need to be generated and updated with new information. Therefore, evaluation of design consistency measures provides potential effects on safety performance, by analyzing roadway design attributes with respect to driver expectancy. In other words, a “consistent” design is one that is compatible with driver expectation and capabilities. A consistent design along a rural highway has the potential to reduce crash severity and frequency. Design consistency has safety implications and is intuitively linked to roadway departure crashes. Ng and Sayed (2004)

and Wu et al. (2013) have attempted to explicitly link measures of design consistency to safety. These studies offer a starting point for additional analysis, but do not necessarily provide generalizable safety findings related to roadway departure crashes on horizontal curves along rural, two-lane roads in the U.S.

1.2 Research Objective and Scope

The objective of this research was to explore the relationship between the expected number of roadway departure crashes on horizontal curves and design consistency measures, focusing on alignment indices (geometric design characteristics) along rural, two-lane highways. Relationships were estimated using a cross sectional study design and a series of negative binomial regression models. Data were collected in Utah. The database for this research was built by leveraging the results of a large data collection effort conducted by the Utah DOT using mobile light detection and ranging (LiDAR). The research objective was accomplished through the following eight tasks:

- Review research literature on design consistency measures and the relationships between design consistency and safety.
- Review count model selections.
- Collect traffic and geometric design characteristics for rural, two-lane highways in Utah.
- Verify and refine geometric design characteristics and measurements provided by the Utah Data Portal and Mobile LiDAR data, including horizontal curve geometrics, cross slope, and vertical grade.
- Build horizontal curve segments for analysis.

- Define roadway departure crashes and merge crash counts to the defined horizontal curve segments.
- Estimate geometric design consistency measures in this study.
- Explore the relationship between the expected number of roadway departure crashes and design consistency through a series of negative binomial regression models, with design consistency defined in this study using a number of different alignment indices.

CHAPTER 2

LITERATURE REVIEW

This chapter includes an overview of design consistency and design consistency measures (i.e., operating speed, alignment indices, driver workload, and vehicle stability), especially for alignment indices measures, and different crash count models used in previous research. The first section provides background information on design consistency. The second section presents an overview on how different measurements are used to evaluate design consistency. Alignment indices, as important measurements for this study, will be discussed in more detail in this section. The third section introduces the background of crash count models and how these models are employed for safety and design consistency studies. The last section demonstrates the different data collection methods and the Mobile LiDAR method which is utilized in this study.

2.1 Background of Design Consistency

According to past research, design consistency has typically taken into account three considerations: driving performance, speed, and safety. Performance considerations address the impact of heavy driver workloads on a driver's readiness and understanding, which interrupt driver expectancy. Speed considerations address how different design elements impact the operating speed. Operating speed evaluates the design consistency

along different road elements. Safety addresses how geometric design measurements (e.g., alignment indices) impact highway safety from a transportation engineering perspective (Gibreel et al., 1999).

Design consistency has been evaluated and studied widely in the past century based on three considerations. In the middle of the 1960s, geometric design created the expectation and improved ability of the motorist to guide and control a vehicle in a safe driving manner (Glennon et al., 1978). In the early part of the 1980s, researchers found poor design consistency performance caused higher driver workload. The inconsistent design was summarized as “a geometric feature or combination of adjacent features that have such unexpectedly high driver workload that motorists may be surprised and possibly drive in an unsafe manner” (Messer, 1980). Later on, research revealed driving operation error is reduced more by geometric design variables that conform to drivers’ expectations than variables that violate their expectancies (Post et al., 1981). In terms of the design consistency concept, the definition recommended by Wooldridge et al. (2003) states that “Design Consistency is the conformance of a highway’s geometric and operational features with driver expectancy.”

This definition is the most applicable when considering multiple measures of effectiveness and different roadway environments. Wooldridge et al. (2003) created a survey for determining the definition of design consistency. Researchers provided five potential definitions to the U.S. state DOTs and transportation researchers. The phrase of “driver expectancy” was finally adapted instead of the terms of “similar roadway”, “section of highway”, or “driver workload”. The phrase “highway’s geometric features” from the design consistency definition referred to safety considerations which include

traffic accidents, vehicle stability, cross sections, horizontal alignment, vertical alignment, sight distance, and traffic volume. “Operating features” represents operating speed, design speed, and expected speed based on speed considerations.

From performance considerations, driver expectancy can be regarded as reasonable safety probabilities of driver behavior in a given environment. Alexander et al. (1986) indicates that “Expectancy relates to a driver’s readiness to respond to situations, events, and information in predictable and successful ways.”

Driver workload, driver anticipation, highway aesthetics, and interchange design are the factors which may interfere with driver expectancy. Even though many factors impact the design consistency evaluation, four quantitative measures were identified by past studies for directly or indirectly developing geometric design models to estimate the crash frequency based on safety considerations. The next subsection describes these four design consistency measures.

2.2 Overview of Design Consistency Measures

The four design consistency measures are categorized as speed differences, vehicle stability, alignment indices, and driver workload. The speed differences usually indicate the difference between operating speed and design speed ($V_{85} - V_d$) or the reduction in operating speeds between two successive elements (ΔV_{85}) (Lamm et al., 1999; Fitzpatrick and Collins, 2000). The meaning of V_{85} describes the 85th percentile operating speed, which is selected by the drivers under free flow conditions (Tarris et al., 1996). These speed equations are utilized to explain safety criteria. Past studies proved a larger speed difference caused a higher crash frequency (Anderson et al., 1999; Ng and

Sayed, 2004; Wu et al., 2013). However, Butsick et al. (2015) indicated that the speed differences only indirectly identify the reasons associated with the drop in speed, because speed differences act as surrogate measures of consistency. In addition, the estimation of the speed differences was limited by the field validation which ensures circumstantial applicability. Thus, Butsick et al. (2015) suggested utilizing geometric alignment data to measure design consistency with safety considerations that could be more practical.

Driver workload is the other significant measure for evaluating design consistency. Even though visual demand and available sight distance have been identified as two parameters to measure driver workload, almost all of the research utilized the visual demand of drivers to analyze design consistency. Visual demand is quantified by the amount of visual information the driver requires to maneuver the vehicle on the right track of the roadway (Fitzpatrick et al., 2000). Messer (1980) and Messer et al. (1981) developed two equations for drivers familiar and unfamiliar with the roadway as it relates visual demand to the horizontal curve radius. Ng and Sayed (2004) utilized the methodology developed by Messer et al. (1981) to indicate the positive relationship between crash frequency and lack of visual demand due to a longer distance of roadway caused by a larger curve radius. However, Andrew et al. (2015) also indicated that the measure of driver workload from performance considerations also serves as a surrogate measure of consistency.

Vehicle stability, another design consistency measure, is quantified by side friction. Side friction demand equations are formed by the 85th percentile operation speed, radius, and superelevation of the roadway (Lamm et al. 1999). Ng and Sayed (2004) utilized the equation developed by Lamm et al. (1999) and found the positive relationship

between crash frequency and higher changes in vehicle stability. Again, aforementioned measures of design consistency indirectly explore the relationship between geometric design consistency as surrogate variables and crash frequency. Thus, this study will focus on directly exploring the relationship between geometric alignment indices and crash frequency. More previous studies on alignment indices are introduced in the following section.

2.3 Background of Alignment Indices

Alignment indices are design consistency measures which directly focus on studying roadway geometric design parameters from a horizontal and vertical alignment perspective. Fitzpatrick et al. (2000) defined alignment indices as “quantitative measures of the general character of a roadway segment’s alignment.” When operating speeds are missing or have a poor prediction on long tangents, the changes in alignment indices are able to show where the geometric inconsistencies are located. In terms of increasing geometric inconsistencies, alignment indices change in relation to the high rate or large increase of different segments of the roadway. Fitzpatrick et al. (2000) mentioned three proposed indicators of geometric inconsistency which are described below:

- A large increase/decrease in the values of alignment indices for successive roadway segments.
- A high rate of change in alignment indices over some length of roadway.
- A large difference between the individual feature and the average value of the alignment index.

Anderson et al. (1999) summarized three major advantages of using alignment

indices in design consistency evaluations. First, alignment indices are easier to explain, utilize, or design for the practitioner in the transportation engineering field. Second, based on a system-wide perspective, alignment indices which consider horizontal and/or vertical alignment elements provide a quantitative mechanism for comparing successive geometric elements. This kind of mechanism is the basis of the definition of design consistency. Third, alignment indices are able to quantify the interaction between the horizontal and vertical alignments. Thus, alignment indices are classified as horizontal alignment indices, vertical alignment indices, and composite indices. In the following subsections, several alignment indices, which may be applied in this thesis, will be introduced based on horizontal alignment indices. The vertical alignment indices and combination indices play a subsidiary role.

2.3.1 Horizontal Alignment Indices

Much of the previous literature has suggested that the horizontal alignment indices are important, because these indices potentially exist in the relationship between curves, speeds, and crash rates. In this subsection, six alignment indices will be discussed:

- The curvature change rate (CCR),
- The degree of curvature (DC),
- The average radius of curvature (Avg. R),
- The changed radius rate (CRR),
- The ratio of average radius over radii (RRR), and
- The ratio of tangent length over radius (RTR).

Firstly, the curvature change rate (CCR) and the degree of curvature (DC) have

been used to evaluate geometric design consistency by Lamm et al. (1987), Morrall et al. (1994), and Faghri et al. (1999). Lamm et al. (1987) indicated both indices are equally important, while other studies selected the DC to assess consistency. CCR was recognized as an index with significant impacts on crash frequency by Castro et al. (2005). The CCR is defined as the ratio of the sum of deflection angles to the total length of the segment. The equation for CCR is shown in Eq. 2-1 below:

$$CCR = \sum(\Delta i/L) \text{ (degree/mile)} \quad (2-1)$$

Where:

$\sum \Delta i$ = deflection angle (degree);

L = length of segment (mile).

Castro et al. (2005) found a moderately good correlation between the crash rate variation and the increments of CCR. The DC is defined as the relation between the curve length and its radius. The equation for DC is shown in Eq. 2-2 below:

$$DC = 5730/R \text{ (degree/feet)} \quad (2-2)$$

Where:

DC = degree of curvature (degree);

R = curve radius (feet).

The equation for the degree of curvature of a segment is given as:

$$DC = \sum(DC_i/L) \text{ (Degree/Mile)} \quad (2-3)$$

Where:

DC_i = degree of curvature of each element of the segment (degree);

L = total segment length (mile).

In this study, the research team created an algorithm to determine the horizontal curve based on raw data from pieces of curves in the LiDAR database. The horizontal curve estimation processes are shown in the Data Collection section. Then, the degree of curvature can be determined later. The total segment length for each principle curve segment consists of the upstream tangent length, curve length, and downstream tangent length in this thesis. The DC along the tangent segments are zero. DC_i consists of DC in the middle of curve, and DCs in the upstream curve and downstream curve.

The average radius of curvature (Avg. R) is another important index which will be employed in this thesis. The definition of Avg. R is the average horizontal radius of curvature of the segment.

$$Avg R = \sum(R_i/N) \text{ (feet)} \quad (2-4)$$

Where:

R_i = radius of curve i (feet);

N = number of horizontal curves within the segment.

Fitzpatrick et al. (2000) and Anderson et al. (1999) have proven the sensitivity relationship between the average radius of curvature and crash frequency.

In addition, Anderson et al. (1999) showed a general comparison by applying a term of changed radius rate (CRR). The CRR is defined as the radius of the i th curve on the roadway section over the average radius for the whole test roadway section, as shown in Eq. 2-5.

$$CRR_i = R_i/R_{mean} \quad (2-5)$$

Where:

R_i = radius of curve i (feet);

R_{mean} = the average radius of horizontal curvature for a roadway segment (feet).

CRR is used to show inconsistency with the flat curves. If CRR is less than 1, it demonstrates a significant impact on design consistency in this roadway section. However, CRR does not categorize a curve as good, fair, or poor. Califso et al. (2009) developed and categorized alignment indices of design consistency criteria for evaluation. Their results identified two alignment indices based on a homogeneous sample of 15 subjects which were tested along a 6.8-mile length with four test sections. One of the alignment indices is the ratio of average radius over radii (RRR). It is formed as the ratio between the average radius of horizontal curvatures for a roadway segment and the radius of each individual horizontal curve. Eq. 2-6 describes the RRR as:

$$RRR = R_{mean}/R_i \quad (2-6)$$

Where:

R_{mean} = the average radius of horizontal curvatures for a roadway segment (feet);

R_i = the radius of horizontal curve i (feet).

In this study, the average radius of each segment has three horizontal curvatures. When the RRR is smaller, the geometric design is more consistent along the road. The other measure is the ratio of tangent length over radius (RTR), which represents the upstream tangent length over the bending radius of the horizontal curve. The equation for

RTR is shown below:

$$RTR = TL/R_i \quad (2-7)$$

Where:

TL = the tangent length (feet);

R_i = radius curve i (feet).

In this thesis, RTR consists of three measures: 1) the upstream tangent length over the radius of the middle curve, 2) the downstream tangent length over the radius of the middle curve, and 3) the average tangent length over the radius of the middle curve in each principle curve segment. Consistency ratings for the RRR and RTR are shown in Table 1.

2.3.2 Vertical Alignment Indices and Composite Alignment Indices

Due to the amount of mountainous terrain or hilliness along rural highways, these terrains affect the sensitivity relationship of the speed, crash frequency, and vertical alignment indices. The vertical curvature change rate (VCCR) and the average rate of vertical curvature (AVC) are two vertical alignment indices which have a significant influence on crash frequency. The definition of VCCR is the index of the gradient change per roadway segment. The VCCR is given as:

$$VCCR = \frac{\sum |A_i|}{\sum L} \left(\frac{\%}{mi} \right) \quad (2-8)$$

Where:

|A_i| = absolute gradient difference over vertical curve i (%);

L = length of segment (mile).

Castro (2005) found that the VCCR has been emphasized for analyzing crash rate variations. The difference of grade in each principle curve segment can be identified roughly by viewing LiDAR collection video. The length of the segment is the same as the length in the horizontal curve segment.

The average rate of vertical curvature (AVC) is a vertical alignment index which indicates the amount of change in the vertical alignment (Anderson et al., 1999). The equation for AVC is:

$$AVC = \frac{\sum(L_i/|A_i|)}{N} \quad (2-9)$$

Where:

L_i = length of the vertical curve i on the roadway section (feet);

$|A_i|$ = absolute gradient difference over vertical curve i (%);

N = number of vertical curves within the section.

Fitzpatrick et al. (2000) and Anderson et al. (1999) have found a relationship between the AVC and crash frequency. However, it may not be used in this study due to limitations in identifying the vertical curvature accurately.

Until now, only one composite alignment index has been applied by Castro et al. (2005). The composite alignment index (CCR combo) is defined by Fitzpatrick et al. (2000) as the sum of the horizontal curvature change rate and the vertical curvature change rate. The equation is shown as:

$$CCR\ combo = CCR + VCCR = \frac{\sum \Delta i}{L} + \frac{\sum |A_i|}{L} \quad (2-10)$$

Where:

Δi = deflection angle (degrees);

$|A_i|$ = absolute gradient difference over vertical curve i (%);

L = length of segment (feet).

Castro et al. (2005) indicated that the composite alignment index becomes a reasonably good tool to evaluate alignment consistency. The horizontal alignment part of the CCR combo has more effect on the model than vertical alignment indices.

2.4 Count Models

Different statistical models have been applied for modeling crash frequency in past decades, in order to explore potential methods for this study. Due to the non-negative integer characteristic of crash frequency data, the application of Poisson and negative binomial regression count models is the most appropriate choice for modeling crash frequencies. In the following subsections, this study will introduce definitions and applications of these count models, including the Poisson model, Poisson lognormal model, Poisson gamma / negative binomial (NB) model, and an extension of two previous models, specifically the zero-inflated Poisson (ZIP) and negative binomial models (ZINB).

2.4.1 Poisson Model and Poisson Lognormal Model

The Poisson model is the most basic crash count model, which assumes that the mean and variance are equal. The Poisson distribution usually has been used as an appropriate selection for crash frequency analysis. The characteristics of these crash

frequencies are relatively small non-negative integers. In the late 1980s and early 1990s, the Poisson regression approach for modeling crash frequencies was adopted and applied popularly for research studies (Jovanis and Chang, 1986; Jones et al., 1991; and Miaou and Lum, 1993).

However, crash frequencies sometimes follow a lognormal distribution which means a normal distribution falls on a logarithmic scale (Aguero-Valverde, 2013). Thus, the Poisson lognormal models have been employed for crash frequency studies since the late 1990s (Anderson et al., 1999; Miranda Moreno et al., 2005; Ma et al., 2007). Anderson et al. (1999) utilized the Poisson model and Poisson lognormal model to analyze the relationship between safety and geometric design consistency measures for rural, two-lane highways. In the section exploring the relationship between speed reduction and crashes occurring on horizontal curves, they collected 1,747 crashes for 5,287 curves, with a mean of 0.11 accidents per curve per year. They developed two approaches to treat exposure. In the first approach, the natural logarithm of AADT and curve length were used separately. In the second approach, an exposure variable known as million vehicle-kilometers of travel (MVKT) was applied in the Poisson model and provided a more significant model than the first approach. The authors decided to use the Poisson model instead of the negative binomial by assessing the results from goodness of fit criteria. Goodness of fit criteria will be discussed in greater detail in the Methodology section dealing with selecting better count models.

In the second section of this study, Anderson et al. (1999) employed the Poisson lognormal model to analyze the relationship between alignment indices and crash frequency, because the researchers found that their observed crash frequency distribution

in 3 years was more like the lognormal distribution than the Poisson distribution. They found that crash frequency was closely related with alignment indices measures. Most interestingly, they found the average radius of horizontal curvature and the average rate of vertical curvature had a greater effect on crash frequency than the ratio of maximum radius to minimum radius on a roadway section.

2.4.2 Negative Binomial Model

To overcome over-dispersion, the negative binomial regression model is developed using a gamma probability distribution. This model is frequently utilized by researchers in modeling crash frequency, and has been popular to estimate the average crash frequency from observed crash counts in transportation studies. Lord et al. (2005) reported that an abundance of previous research has found that the variance to mean ratios of crash data are greater than one (Abbess et al., 1981; Poch and Mannering, 1996; Hauer, 1997).

Saito et al. (2015) utilized a negative binomial model to predict crash frequency for horizontal curve segments of rural, two-lane highways in Utah. Crash sample periods were used either for a 3-year period from 2010 to 2012, or a 5-year period from 2008 to 2012. They also utilized the database from the Utah LiDAR collection, which was provided by UDOT's LiDAR asset management program. The database contained 1495 curved segments which were randomly selected in the state of Utah. The results showed that four significant variables impacted potential crash occurrence. These four variables include average annual daily traffic (AADT), segment length, total truck percentage, and horizontal curve radius. In this study, the curve radius will be transformed by different

combinations with other alignment indices or the total radius of the studied segment.

The characteristics of over-dispersion may include the preponderance of zeros, a condition which occurs when there is a greater-than-expected number of zero observations in the negative binomial process and the preponderance of large outcomes. To overcome this preponderance of zeros, a zero-inflated model will be presented and discussed in the following subsection.

2.4.3 Zero-Inflated Model

The zero-inflated count model is able to handle data with a preponderance of zeros. Essentially, the zero-inflated models are followed by a dual state process. Lord et al. (2005, 2007) indicated a dual state process includes a perfect state (zero state) and an imperfect state with a mean (non-zero state). In terms of highway safety, the perfect state represents the count of crashes per specific time period when there are zero accidents at an entity (intersection, road segment, etc.), and the imperfect state represents the count of crashes when there are more than zero accidents. However, a Poisson or negative binomial model cannot explain the “excess” zeros under this dual state process.

Lee and Mannering (2002) deployed a zero-inflated count model to analyze roadway run off (roadway departure) crashes on a 96.6 km (~ 60 miles) section of highway in Washington State. The total number of roadway run off crashes was 489 in a 3-year period. They found that posted speed limits above 85 km/h (~55mph) increased the crash frequencies in the negative binomial crash state (imperfect state) and decreased the frequencies in the zero state (perfect state). Increasing road shoulder width also increased the probability of roadway run off crashes in a perfect state. To prove the

implication of the estimation results, they also used the pseudo-elasticity to test the incremental change in the count of crashes by changes in their indicator variables.

Easa and You (2009) studied the relationship between crash frequency and relevant variables under five different alignment combinations, including horizontal curves combined with crest vertical curves, horizontal curves combined with sag vertical curves, and horizontal curves combined with multiple vertical curves, as well as these curves combined with grades of less than 5% and grades of more than 5%. They employed Poisson, negative binomial, zero-inflated Poisson and zero-inflated negative binomial models to explore each combination. They utilized ZIP and ZINB in the final estimated models. They came to two conclusions referred to in this study. First, the degree of curvature has the most significant impact on crash frequency. Second, the crash frequency on horizontal curves combined with sag vertical curves is greater than horizontal curves combined with crest vertical curves.

2.5 Background of Data Collection Methods

State DOTs have applied various methods on collecting roadside inventory data based on cost of time, equipment, and labor. Current methods used include integrated GPS/GIS mapping, field inventory, photo/video log, aerial imagery, satellite imagery, mobile LiDAR, Airborne LiDAR, terrestrial laser scanning, etc. A survey from the Highway Safety Manual (HSM) pointed out that air-based methods are less popular choices among state DOTs because of the difficulty in identifying small objects. Field inventory methods still require a heavy labor workload, provide less accurate data collection, and suffer from a lack of new supporting technology. Thus, this study focuses

on exploring data collected through integrated GPS/GIS mapping, photo/video log, terrestrial laser scanning, and mobile LiDAR.

Objective roadside inventory data collection methods have been studied in the past decade. Photo/video logs, as mobile collection methods, are able to automatically record photos/videos on roadway information after later processing. The advantage of this method is less exposure to traffic and short field data collection times. However, the drawback of this method is the inability to measure different feature dimensions, such as the coordinate of each tested milepost. Large data need to be reduced. Integrated GPS/GIS mapping systems used an integrated GPS/GIS field data logger to record and store inventory information. Outcomes of this method can be viewed in a mapping application. This method has low cost of equipment, easier data transferring, and low data reduction effort. The data reduction process involves inputting data into a computer aided design (CAD) software program, and importing the results into the drawing format which is easier to manipulate and analyze intuitively. But this method involves long field collection times and crew exposure to traffic. In addition, limitations could include a GPS outage, which could be caused by a tree or tall buildings.

In recent years, state DOTs have employed and updated collection technology, such as terrestrial laser scanning and Mobile LiDAR. The former uses direct 3D precision point information to acquire highway inventory data. The drawbacks of this method are long field data collection times, exposure to traffic, high initial cost, long data reduction time, and large data size. Mobile LiDAR has more mobility with an instrumented 3D precision point sensor and other sensors to capture geospatial data accurately and precisely. The remarkable advantage of Mobile LiDAR is the reduced amount of time for

collecting data – for a 20-mile segment of a highway, the time was reduced to 30 minutes from 10 days. It also improves the safety of the survey crew compared with other methods. The new system is able to measure at a rate of 50,000 to 500,000 points per second per scanner (Tang and Zakhor et al., 2011). Even though shortcomings of the Mobile LiDAR method include expensive equipment and the long data extraction and reduction time, this method is able to capture valuable data for DOT programs (Jalayer et al. 2014).

Data were made available through UDOT's online data portal, a central clearinghouse of all public UDOT data. The research team relied on a roadway inventory developed from LiDAR data and processed and calibrated by one or more data collection contractors. This resulted in direct and easy access to a significant number of roadway inventories not typically available in traditional datasets, including cross slope and vertical grade. However, the data were being processed in a way to support asset management, and the accuracy of certain data elements was at a level consistent with that need and inconsistent with safety analysis. Additional data processing will be presented in the data collection section.

Table 1. The Califso et al. Design Consistency Evaluation on Alignment Indices

RRR	Consistency Rating	RTR
<1.5	Good	<1
1.5 to 2	Fair	1 to 2
≥ 2	Poor	≥ 2

CHAPTER 3

RESEARCH METHODS

This chapter describes the research methods. A series of count models were estimated to explore the relationship between the expected number of roadway departure crashes and horizontal and vertical alignment indices. In the first section, the negative binomial (NB) model will be discussed. In the second section, zero-inflated negative binomial (ZINB) models are introduced to attempt to address the excessive zeros in the crash frequency database.

3.1 Negative Binomial Model

Negative binomial (NB) models have been estimated for over-dispersed crash frequency data, or data for which the variance is greater than the mean (Miaou et al. 1993; Shankar et al. 1995). A gamma-distributed error term in the NB model helps overcome erroneous coefficient estimates and erroneous inferences that result from ignoring the over-dispersion. Based on statistical road safety modeling (SRSM) (Hauer, 2004), the expected number of roadway departure crashes on segment i , μ_i is expressed by NB as:

$$\mu_i = E(Y_i) = \exp(\beta_0 + \sum_{j=1 \text{ to } k} \beta_j X_{ij} + \varepsilon_i) \quad (3-1)$$

Where:

$\mu_i = E(Y_i)$ = the expected number of roadway departure crashes on segment i ;

k = the number of independent variables;

X_{ij} = independent variable j on road segment i ;

β_0 = constant or intercept;

β_j = parameter that quantifies the magnitude and direction of the effect of independent variable j in X_{ij} on μ_i ;

ε_i = unknown or unmodeled effects on μ_i , represented as a disturbance term.

Alignment indices, a design consistency measure, were the primary explanatory variables of interest. Based on the previous literature review section, horizontal and vertical alignment indices were tested as potential right-hand-side variables in Eq. 3-1. In addition, other variables were also tested in model specifications to minimize omitted variable estimator bias. These variables included posted speed limit indicators, tangent length indicators, as well as the exposure measures LnAADT (natural logarithm of the annual average daily traffic of a roadway segment) and LnCL (natural logarithm of the curve length of a roadway segment). Omitted variable bias means over- or under-estimating the safety effect of design consistency variables due to missing unmeasured variables that are correlated with design consistency variables. The exposure variables LnAADT and LnCL are commonly specified in crash prediction modeling (Reurings et al., 2006). Various specifications of LnAADT and LnCL were tested, which will be discussed in the Data Analysis section.

$\text{Exp}(\varepsilon_i)$ is gamma distributed with mean 1 and variance α . This results in the mean-variance relationship being expressed as:

$$VAR(Y_i) = \mu_i + \alpha [\mu_i]^2 \quad (3-2)$$

Where:

$\mu_i = E(Y_i)$ = the expected number of roadway departure crashes on segment i ;

$VAR(Y_i)$ = variance of roadway departure crashes on segment i ;

α = over-dispersion parameter.

The over-dispersed data are represented by a value for α that is greater than 0. If α is less than 0, the data are under-dispersed. A larger estimate of α indicates greater over-dispersion. Eq. 3-2 indicates the variance is greater than the mean in most cases. The probability density function of the negative binomial distribution is defined as the following form (Miaou, 1994):

$$P(Y_i = y_i) = \frac{\Gamma(\frac{1}{\alpha} + y_i)}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \quad (3-3)$$

Where:

$P(Y_i = y_i)$ = the probability density function of the NB for roadway departure crashes on segment i ;

α = dispersion parameter;

μ_i = the expected number of roadway departure crashes on segment i ;

$\Gamma(\cdot)$ = a value of the gamma function.

The McFadden Pseudo R-Squared was used to evaluate goodness-of-fit of the negative binomial models. This measure is analogous to the R-squared term in linear regression, where values range from 0 to 1, but never approach 0 or 1, and higher values indicate better model fit. The McFadden Pseudo R-Squared is described in Eq. 3-4.

$$\rho^2 = 1 - \frac{L(full)}{L(0)} \quad (3-4)$$

Where:

ρ^2 = McFadden Pseudo R-Squared;

L(full) = log-likelihood of the model with explanatory variables;

L (0) = log-likelihood of the intercept-only model.

3.2 Zero-Inflated Negative Binomial Models

The zero-inflated count model was formally introduced by Lambert (1992) as a method of accounting for excessive zero counts. This model has been explored in traffic safety for the past two decades and mainly provides a method to handle study sites which have a preponderance of instances in which there are no crashes. This thesis will explore the zero-inflated negative binomial model (ZINB), in addition to a NB model, due to the excessive number of sites with zero roadway departure crashes. Generally, the ZINB uses a logit model to describe roadway departure crash frequencies in either a zero state or non-zero state, and the NB count model is used to describe the crash frequency of non-zero roadway departure crashes.

Mathematically, the probability density function of the zero-inflated negative binomial distribution with two states (zero and non-zero states) are represented as (Hosseinpour et al., 2014):

$$P(Y_i = y_i = 0) = P_i + (1 - P_i) \frac{1}{(1 + \alpha \mu_i)^{\frac{1}{\alpha}}} \quad (3-5)$$

$$P(Y_i = y_i > 0) = (1 - P_i) \frac{\Gamma\left(\frac{1}{\alpha} + y_i\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_i + 1)} \frac{(\alpha\mu_i)^{y_i}}{(1 + \alpha\mu_i)^{\left(y_i + \frac{1}{\alpha}\right)}} \quad (3-6)$$

Where:

y_i = the number of roadway departure crashes for segment i ;

P_i = the probability of segment i being in a zero crash state, which is fitted in a logistic regression model. The expression of P_i is shown as:

$$P_i = \frac{\exp(K_i\beta)}{1 + \exp(K_i\beta)} \quad (3-7)$$

Where:

K_i = the function of explanatory variables in logistic regression model;

β = the estimable coefficients.

In the ZINB model testing process, all geometric alignment indices will be tested by utilizing the NB model in the non-zero state; meanwhile, geometric variables and exposure variables will be tested by utilizing the logit binary model in the zero-state. In the zero-state logit binary model process, the positive signs of coefficients in the logit binary model implies the higher probability of being in the zero state. For example, if the coefficient for the indicator $RRR < 1.5$ (an indicator of poor design criteria) (Califso et al., 2009) had a positive sign in the logistic regression model and a statistically significant confidence interval, it would imply that $RRR < 1.5$ increases the probability of being in the zero state.

The Vuong test is commonly used to evaluate the appropriateness of using a zero-inflated count model, and it is used here to compare between the NB and ZINB models.

The test was first provided by Vuong (1989) and is shown below:

$$m_i = \ln \left(\frac{\sum_i P1(y_i|x_i)}{\sum_i P2(y_i|x_i)} \right) \quad (3-8)$$

$$V = m_{mean}(n)^{0.5}/SD(m) \quad (3-9)$$

Where:

$P1(y_i|x_i)$ = the predicted probability density function of the standard negative binomial;

$P2(y_i|x_i)$ = the predicted probability density function of the zero-inflated negative binomial;

m_{mean} = the mean of m_i ;

$SD(m)$ = the standard deviation of m_i ;

n = number of the observations;

V = the Vuong test for a standard normal distribution.

If V is greater than 1.96, it means the NB model is preferred over the ZINB model. If V is less than 1.96, then the ZINB model is preferred over the NB model. However, many previous studies reported the Vuong's statistic test did not apply a penalty for the complexity of model variables (Greene, 2000; Washington et al., 2003; Lord et al., 2007). Vuong (1989) even suggests applying Akaike's information criterion (AIC) and Bayesian information criterion (BIC) for correcting this test. These two goodness-of-fit measures are able to penalize the model and overcome the complexity. The equations are defined as follows:

$$AIC = -2LL + 2P \quad (3-10)$$

$$BIC = -2LL + P(\ln(n)) \quad (3-11)$$

Where:

LL = the logarithm of the maximum likelihood estimation for each model;

P = the number of model parameters;

n = the number of observations.

Simply speaking, the lowest value of AIC or BIC represents the preferable model. The statistically significant difference between two models was found by Raftery (1995) and Hilbe (2011). Table 2 presents the significant levels for AIC and BIC.

In this study, the number of observations will be more than 500, which means the ZINB will be more favored than the NB if the difference in the AIC is more than 2.5 and the ZINB has a lower AIC. However, if the two models did not show a significant difference in AIC, the difference of BIC would indicate the favored model which is indicated by the lower BIC model. STATA statistical software was utilized to estimate the NB and ZINB models. In addition, STATA will also be used to implement model selection tests, such as the AIC, BIC, and Vuong test. It provides and implements all equations shown above for negative binomial and zero-inflated negative binomial regression models.

Table 2 Raftery's (1995) BIC and Hilbe's (2011) AIC for Significant Levels

ΔAIC	Results if $A < B$	ΔBIC	Results if $A < B$
≤ 2.5	No difference	≤ 2	Weak difference
2.5 to 6	Prefer A if $n > 256$	2 to 6	Positive difference
6 to 9	Prefer A if $n > 64$	6 to 10	Strong difference
> 10	Prefer A	< 10	Very strong difference

CHAPTER 4

DATA COLLECTION

This chapter discusses the process of collecting different data resources for the final database. The final database consists of five main components of data files, including horizontal curvature data, traffic flow data, posted speed limit data, geometric roadway inventory data, and crash data. All data were obtained from the UDOT Data Portal and UDOT Traffic and Safety Division. The first section introduces the procedures of horizontal curve estimation and validation based on horizontal curvature data file. Also, the posted speed limit and annual average daily traffic (AADT) as traffic flow data will be subsequently described. Second, this section explains the validation procedures of vertical grade and cross slope, which are two major roadway inventory variables. The third section presents the definition and the collection of roadway departure crashes. The last section in this chapter shows the definition of all variables and descriptive statistic summary for the final data.

4.1 UDOT Data Files

4.1.1 Horizontal Curve Estimation and Validation

Initially, UDOT Data Portal provided thousands of “broken” pieces of horizontal curvature segments which were not used as the completed entire horizontal curve. The

main purpose of using this database was developing a method of combining “broken” curve segments into complete horizontal curves and estimating their key geometric characteristics. The key geometric characteristics include radius, deflection angle, curve degree, etc. The time of manipulation process was consumed at least 500 person-hours on this initial effort. The spent time were distributed on examining the data, testing various alternatives, developing the algorithms, and visually verifying all results in Google Earth. The major programming language was utilized by adopting VBA (Visual Basic for Application) in Microsoft Excel. The procedure of estimating horizontal curve was implemented in the following key steps:

Step 1. Imported the curve shape file into GIS software (ArcGIS) and computed the 2-dimensional Cartesian UTM coordinates from Longitudes and Latitudes in WGS84.

Step 2: Exported the attribute table to a CSV (Comma Separated Value) data file and imported it into Excel.

Step 3: Combined the short segments and estimated PC and PT locations, curve radii, deflection angles, and curve length.

Step 4: Examined and cleaned up the data. This step helped screen out abnormal and missing values. The research team examined the data and cleaned up the data using the following criteria:

- Curves with missing GPS coordinates.
- Curves with missing traffic volumes.
- Abnormally long (a few miles) and abnormally short (less than 0.05 mile) curves.
- Curves with at least one crash coded as intersection-related.

The following content of this section provide a brief description of details of the

algorithm applied in estimation of horizontal curve process. Figure 1 shows an example of a horizontal curve that is split into many short segments, which is shown at the end of this chapter. The cyan arc represented the entire horizontal curve while the yellow arc is one of the “broken” short segments.

In step 1, the “Calculate Geometry” tool was used to add X and Y fields into the attribute table in ArcGIS. The X and Y fields were then converted from Longitude and Latitude in degrees to X and Y position in meters. All calculations would later be done using these 2-dimensional X-Y coordinates in meters. Figure 2 is a screen shot showing the “Calculate Geometry” tool for converting GPS coordinates from degrees to meters in ArcGIS. This step of data conversion was done for the entire data file. The units of horizontal curve length in the final data file were converted from meters to miles.

In step 2, the data files in ArcGIS were exported into a CSV data file by using the data export tool in ArcGIS. Then, the data were imported into an Excel spreadsheet for further calculation. In step 3, the direction of all segments on the decreasing milepost direction was reversed. In the data file, UDOT noted “N” represented as the decreasing milepost direction and “P” represented as the increasing milepost direction. Related VBA codes were developed to identify all records in the data file. Those identified segments were systematically re-coded in the decreasing milepost direction. The new direction changed from “N” to “P”. In other words, the value of ending milepost was received in the new beginning milepost. Vice versa, the value of the beginning milepost became the new ending milepost.

In the last step, the order of data was sorted by route number and the mileposts after reversing the direction of segment. PC and PT locations were detected based on

various characteristics of each short segment (e.g., radius is very large for a tangent and within a reasonable range for a short segment on curve) by using VBA programming codes. The curve length was estimated from the mileposts of PC and PT. The deflection angle was calculated from the estimated curve length and the estimated curve radius. The equation is shown below.

$$DC = \frac{180}{\pi} * \frac{L}{2*R} \quad (4-1)$$

Where DC means deflection angle, L was curve length in feet, and R was radius in feet.

After this step, short “broken” curve segments were merged into complete horizontal curves with the estimated values for their key geometric features. After this process, the data file was reduced from about 6,500 curves down to 4,416 horizontal curves. The reduction in the number of curves can be explained by the following reasons. It was difficult for the developed algorithm to identify small deflection angles. The algorithm also had problems accurately detecting curves for winding stretches of roadway where no or very short tangents exist between curves. In addition, the data often seem to be inaccurate for road segments in mountainous terrain.

4.1.2 Visual Screening of Data in Google Earth

This subsection discusses how to visually check, verify, and recover a horizontal curve data file associated with all 4,416 curves in Google Earth. The procedures of validation and recovery were implemented in the following steps:

Step 1: Marked PC and PT locations in Google Earth.

Step 2: Checked and verified all curves in Google Earth.

Step 3: Recovered or aligned the new horizontal curves in Google Earth.

In the first step, GPS coordinates of PC and PT were created as place markers by using Keyhole Markup Language (KML) in Google Earth. PC and PT markers were coded with different colors for easy identification (Red color for PC and Green color for PT). Both PC and PT markers were attached with curve identifiers and key curve geometric characteristics (such as, radius, beginning and ending milepost, deflection angle, and curve length.). This key information of horizontal curves was prepared for validation when the curve was checked in Google Earth (discussed in Step 2). Figure 3 presents an example of PC and PT markers in Google Earth. The red color marker represented GPS location of PC and green color represented GPS location of PT. PC and PT markers have almost identical key information except milepost.

In the second step, all KML files of PC and PT markers of all horizontal curves were imported into Google Earth. Each individual horizontal curve was checked visually to verify the consistency between the key curve geometric characteristics attached to markers and locations of all markers. The distance measurement tool in Google Earth was also occasionally used for verifying the curve length. Figure 3 presented an accurate example of horizontal curve with relevant geometric characteristics. In this example, the horizontal curve has a deflection angle of 31.01 degrees with a 0.499-mile curve length. Judging the curve in Google Earth based on geometric design requirements, these numbers were reasonable to appear. A distance of 0.5-mile curve length was measured by the distance tool in Google Earth. Therefore, all information for this horizontal curve was consistent and the curve was tagged in the data file for analysis. Some inconsistent information was also found by using the same method to check. In Figure 4, it was an

example of a curve with inconsistent information. The marker labels indicated that the curve is 0.059-mile-long which means the location of PC and PT were almost at the same location. Thence, all curves with similarly inconsistent information as this example were removed from the dataset.

As Figure 4 illustrates, the GPS locations of PC and PT were inaccurate. The inaccurate GPS locations appeared more frequently in mountainous areas during the process of data validation. The potential reason for this inaccurate result was the poor GPS signal reception in the terrain of Utah's mountains. Thus, this part of the GPS database was recommended to be excluded without more accurate location data.

Through this process, the horizontal curves located at or near one or more intersections were screened out and identified. These cases were finally removed from the database for reducing crash analysis interference by intersections. In Figure 5, it represented an example of a horizontal curve located at an intersection. Traffic volume of this horizontal curve was certainly affected by the intersection and interchange. However, there was no indication of the intersection from the data itself. The intersection was only identified visually in Google Earth. This curve was eventually removed from the final dataset.

In addition, the process of data validation in Google Earth was also helpful to identify winter closure. With the "Roads" layer activated, Google Earth provided sections of roadway that are closed during winter season. Figure 6 presented an example of winter closure information in Google Earth. During this data screening process, if a curve was found to be within a section of roadway with the "closed winters" label, it was tagged with an indicator variable. After the process of validating the database in Google Earth,

the final dataset shrank to 1,755 horizontal curve segments.

To study design consistency with successive curves, mending and extending the final database was necessary. Thus, the process of improving database quality will be shown in the following steps. First, based on an overview of all 4416 curves in Google Earth, the potential recovery database includes improper locations at PC or PT or both, poor GPS coordinates, passing zones in a two-lane rural highway, and intersections without/with signs. The countermeasures involved fixing the improper locations by using Google Earth measurements and recovering the curves with passing zones and intersections without signs. Google Earth measurements consist of four procedures.

- I. Identified the GPS coordinates of PC and PT,
- II. Converted the latitude and longitude to UTM, and calculated curve length,
- III. Measured deflection angle, and calculated radius,
- IV. Fixed milepost of PC and PT.

Figure 7 presented an example of manual measurements for determination of horizontal curves in Google Earth. In the example shown below, a new PT location was identified by engineering judgment, while the coordinates of new locations were consequentially gained. The point of intersection (PI) was found by drawing two extending lines along the upstream and downstream tangent length. Deflection angle was measured by using the ruler tool in Google Earth. Then, the curve radius was calculated based on deflection angle equations which have already been shown in the previous section.

According to the methods applied for recovering the database, 1318 curves (out of 4416) have been proved and fixed for this study so far with the limited working time. The

remaining database is also able to possibly be fixed by using a similar method in future research.

4.1.3 Final Horizontal Curve Segment Entity Database

Due to the limited number of validated horizontal curve segments, the research team decided to analyze each tangent-curve-tangent as an entity instead of long segments with multiple successive horizontal curves. To consider the effect of successive curve segments on design consistency, the research team created an algorithm to find at least three successive curves. The middle of three successive curves was analyzed as the studied curve. The upstream and downstream tangent length was calculated based on UDOT Milepost. If tangent length is too short or too long to be satisfied with roadway design requirements, these entities would be excluded from the database. UDOT RMOI (2011) recommended the maximum permissible rate of superelevation as 6%. The average of roadway cross slope is between 1.5% and 3% based on requirements of snow plows and ice clearance operations in Utah. Thus, the minimum tangent length was found around 300 feet (0.057 mile) under a design speed of 65 mph, and cross slope changed from 1.5% to 6%. Table 3 shows the minimum tangent length results at different design speeds. The maximum tangent length was determined to be around 20,000 feet (3.79 mile) by considering the distribution of tangent lengths. Ng and Sayed (2004) also provided the maximum tangent length as around 20,000 feet. The final observations included 582 remaining entities.

4.1.4 AADT and Post Speed

The average annual daily traffic (AADT) data are stored in a GIS shape file and captured through UDOT's data portal. The AADT data were allowed to import into ArcGIS software and were converted into a Comma Separated Value (CSV) format. The data in CSV format were brought into Excel and merged into each horizontal curve based on route number and UDOT milepost. Compared with the length of horizontal curves, the length of road segments within AADT data are longer along rural, two-lane highways. Thus, most of the horizontal curves often completely fell within one of these long segments. In some instances, horizontal curves belong to two different roadway segments with different AADTs. This situation always occurred at those horizontal curves with an intersection. In this case, the weighted average of two different AADTs were calculated. However, horizontal curves with intersections were eventually dropped from the final dataset and were not included in the analysis due to the influence of intersections. The AADT data were merged into final entity data for 7 years (2008 to 2014). The natural logarithm of AADT data was calculated according to the average value of AADT in 7 years.

The posted speed limit data were stored in a GIS shape file as polylines in the UDOT data portal. The data with 'N' direction was excluded and the data with 'P' direction represents the speed in both directions on all nondivided routes. The posted speed limit data were officially published in 2015. The posted speed limit data were extracted to an Excel file from the shapefile in ArcGIS. The four major locations of the posted speed limit data were captured, including the middle point of upstream tangent length, the point of curvature, the point of tangent, and the middle point of downstream

tangent length. The posted speed limit data were also utilized for approximately estimating superelevation inside of the horizontal curve as design speed.

4.2 Roadway Inventory Data File

Cross slope and vertical grade data were provided by a Mobile LiDAR data collection machine. These data were recorded every 0.1 mile along either the increasing milepost direction or decreasing milepost direction. In terms of cross slope validation, four alternative methods were created to calculate the cross slope and merge the result as superelevation into the final database based on the milepost. Alternative 1 is the average of all cross slopes inside of the horizontal curve either in the increasing or decreasing milepost direction. Alternative 2 is the average of the cross slopes at the middle of the horizontal curve in both directions. Alternative 3 is the average of the cross slopes at the middle of the horizontal curve in both directions, with the signs of the cross slope in the decreasing milepost direction being reversed. Alternative 4 is the average of all cross slopes inside of the horizontal curve, with the signs of the cross slope in the decreasing milepost direction being reversed.

The UDOT Roadway Design Manual of Instruction (RMOI) indicates the maximum permissible rate of superelevation is 6% because of Utah's weather conditions. AASHTO (2011) records the minimum radii for design superelevation rates, design speeds, and the maximum superelevation as 6%. To roughly validate the cross slope from Mobile LiDAR, AASHTO's superelevation was calculated based on the posted speed limit and the radius of each horizontal curve from the final database. The results of the four alternatives were used to compare with AASHTO's superelevations. By computing

the difference between the superelevation of the 4 alternatives and the AASHTO values for each horizontal curve, the research team decided to use difference values of 1, 1.5, 2, and 2.5 as a comparable reference. Figure 8 shows the percentage of confidence level for superelevation using each of the 4 reference difference values among the 4 alternatives.

Alternative 2 has the highest confidence level compared with the other 3 calculation methods. However, alternative 3 was eventually selected for calculating the validated cross slope. 74% of the samples fall below the reference difference value of 2. This alternative is more realistic than alternative 2, because the signs of the cross slope in the different direction interfered with the final calculation. In addition, the team also checked the horizontal curve direction based on the signs of the cross slope from 284 samples. Under alternative 2, 99% of the samples have the correct direction when compared with the real horizontal curve directions in Google Earth.

In terms of vertical grade validation, the main purpose is to identify signs of grade and the different types of vertical alignment and straight grades inside of the horizontal curves. The Team possesses the raw original grade database from Mobile LiDAR, as well as the verified grade with absolute integer values which were provided by the UDOT data portal. Thus, the first step was to test the signs of vertical grades by combining verified grade data and raw LiDAR grade data to estimate the signs of grades. This involved evaluating the signs of the raw grade data for both directions of travel for a given road segment to identify likely positive or negative grades, and applying those signs to the verified grade data. After figuring out the signs of the verified value, five critical grades at locations inside of the horizontal curves were captured from the new corrected verified grade database. These five locations included the point of curvature, one quarter of the

curve length, the middle of the curve, three quarters of the curve length, and the point of the tangent. Vertical alignment profiles were approximately plotted based on horizontal distances and vertical elevations of these five critical locations.

Then, by checking Web Navigator video, which visually provided all data within each 0.01-mile interval from the LiDAR database, the types of vertical alignment profiles were observed by engineering judgment. The four classical types of vertical alignments have been summarized and presented by AASHTO (2010). Figure 9 illustrates these four types of vertical alignments which include the type 1 crest, type 1 sag, type 2 crest, and type 2 sag. According to the characteristics of these vertical curves, the team was able to approximately validate the grade values with signs and identify them. Figures 10 and 11 present an example plot of the vertical alignment and the video image captured from the Web Navigator website, respectively. These two figures clearly indicate that the vertical curve at this example location is a type 1 sag curve, in which the initial grade is negative and the final grade is positive. After the process of validation, almost all principal curves were applicable, except 5 principal curve segments had wrong coordinates which could not be found on video. A total of 516 principal curves involved at least one type of vertical curve. In addition, 62 horizontal curves were built on level ground.

4.3 Crash Data File

Total crash and roadway departure crash data between the years 2008 and 2014 were obtained from the UDOT traffic safety division. The safety division provided crash files which include attributes describing the manner of collision, roadway junction feature, contributing circumstance, and crash locations for collisions occurring in rural areas of

Utah. The crash location information consists of the route number, milepost, and GPS coordinates. In the crash files, each row represents the crash ID, vehicle ID, number of vehicles involved in each crash, year in which the crash occurred, route number, GPS locations, sequence of events, roadway junction features, and driver contributing circumstance. These variables were identified by the Utah DI-9 instruction manual, which recorded Utah police report codes and provided UDOT connecting communities' code table listings.

The total crash data were directly merged into the horizontal curve database by each year. Before merging the horizontal curve database with the roadway departure crash data, these crashes must first be defined and identified. UDOT crash rollups defined roadway departure crashes based on the following attribute description:

- Roadway junction features do not include four legs, T, Y, five legs or more, roundabout, ramp intersection with crossroad, bike/pedestrian path intersection, and
- Driver contributing circumstance only contains ran off road, or the sequence of events includes ran off road right, left, crossed median, and collision with fixed object.

According to the FHWA roadway departure crash definition, the research team created two alternative roadway departure crash definitions based on the crash attributes. Among the 3 alternatives, Alternative UDOTRWD (described by the crash rollup definition above), Alternative UURWD1, and Alternative UURWD2, the last two alternatives included situations in which the following conditions were present:

- Driver contributing circumstance condition included “failed to keep proper lane”

- Sequence of events includes “travel in the opposite direction” where the crash occurs with “other motor vehicle in transport,” and the manner of collision includes either “head on” or “sideswipe opposite direction”

The UURWD2 definition also considered overturn events listed in the “sequence of events 1” variable.

All three roadway departure crash alternative definitions are shown in Table 4. The roadway departure crashes of all three alternatives were identified by making queries based on the definitions. The crash data merging methods utilized both the research team’s algorithm and the safety data merge program developed by Chongkai. Both methods were found to be the equivalent to each other. Figure 12 shows the roadway departure crash frequency comparisons among 7 years (2008-2014) with different roadway departure crash definitions for crashes occurring in Utah. In this thesis, the roadway departure crash data based on the Alternative UURWD2 definition were selected as the dependent variable because it represented the most comprehensive roadway departure crash definition.

4.4 Variable Definitions and Descriptive Statistics

The final entity database was manipulated and utilized for studying the sensitive relationship between crashes and alignment index measures. Tables 5, 6, and 7 present data descriptions for all variables which include general attributes, horizontal attributes, vertical attributes, horizontal alignment index attributes, vertical alignment index attributes, and crash attributes. General attributes included 12 variables, which are presented in Table 5. Horizontal attributes contained 15 variables, which are shown in

Table 6. Six significant horizontal alignment index variables and the relative indicators are shown in the horizontal alignment index attributes in Table 6. In Table 7, vertical alignment index attributes include one vertical alignment index and 6 types of vertical curve or straight grade inside of horizontal curves. Vertical attributes presented grades at critical locations and the relative indicator variables. All variables were explored in the modeling study.

Tables 8, 9, and 10 provide descriptive statistics summaries of 578 horizontal curve segments for general and crash attributes, horizontal attributes and horizontal alignment index attributes, and vertical attributes and vertical alignment index attributes. The average total crash frequency was 0.773 per horizontal curve in 7 years. The average roadway departure crash frequency was 0.391. The number of left-turning horizontal curves studied was similar to the number of right turning curves, which was found by identifying the direction based on increasing mileposts. Due to crash counts recorded in both travel directions, curve directions could not significantly impact safety in this study. However, curve directions were able to be used to verify the reliability of the cross slope validation. In the horizontal disaggregate data table, the length of selected curve segments (with the tangent length) ranged from 300 feet to 20,000 feet. Curve segments with spiral curves, composite curves, and curves with lengths less than 300 feet and greater than 20,000 feet were excluded in this study. In the vertical disaggregate data table, more than 50% of the horizontal curves have straight grade segments. The grade ranged from -10 to 9 for most of the critical locations. The analysis and results will be presented in the following chapter.

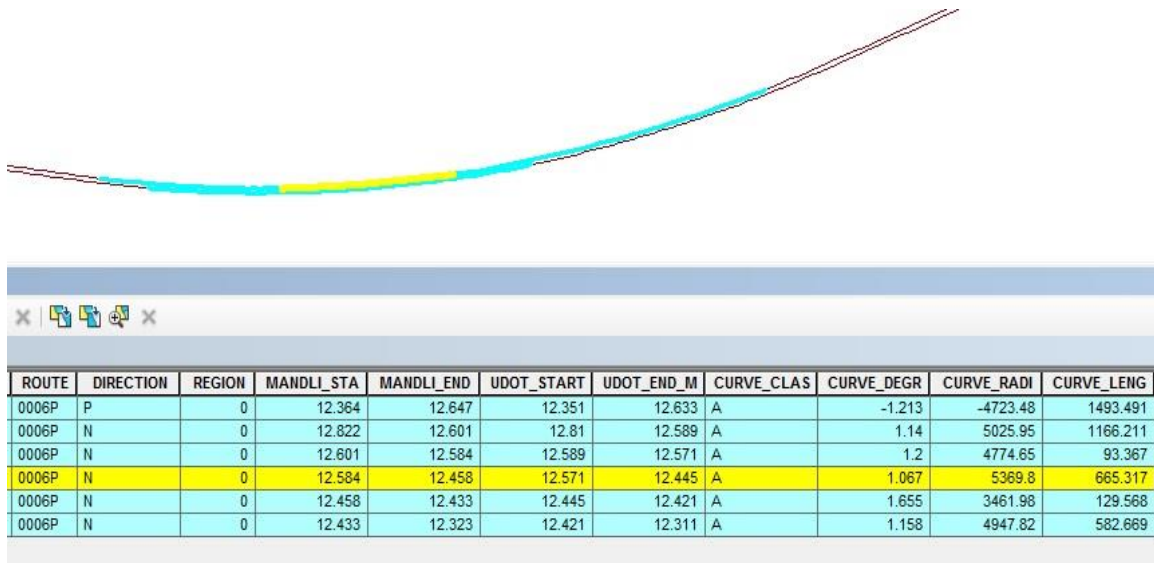


Figure 1. Example of “broken” segments of a horizontal curve.

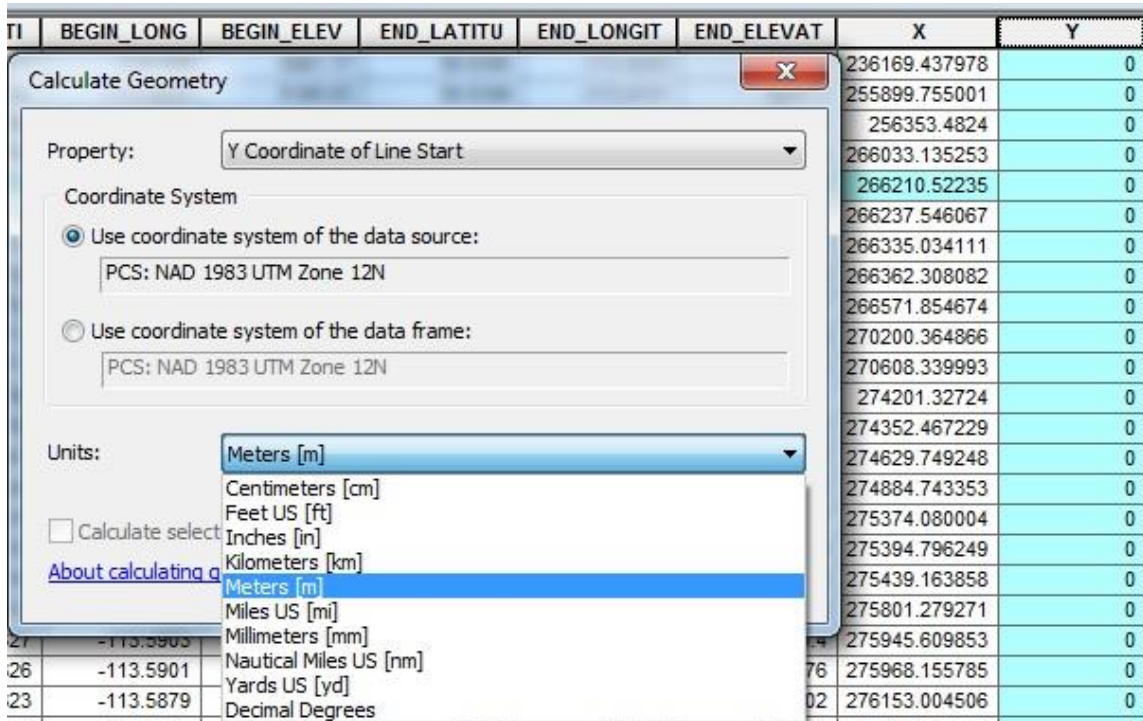


Figure 2. A screen capture of “Calculate Geometry” tool in ArcGIS for converting coordinates from degrees into meters.



Figure 3. An example of curve with accurate information.



Figure 4. An example of discrepancy between curve length and GPS coordinates.



Figure 5. An example of curve at or near intersection.

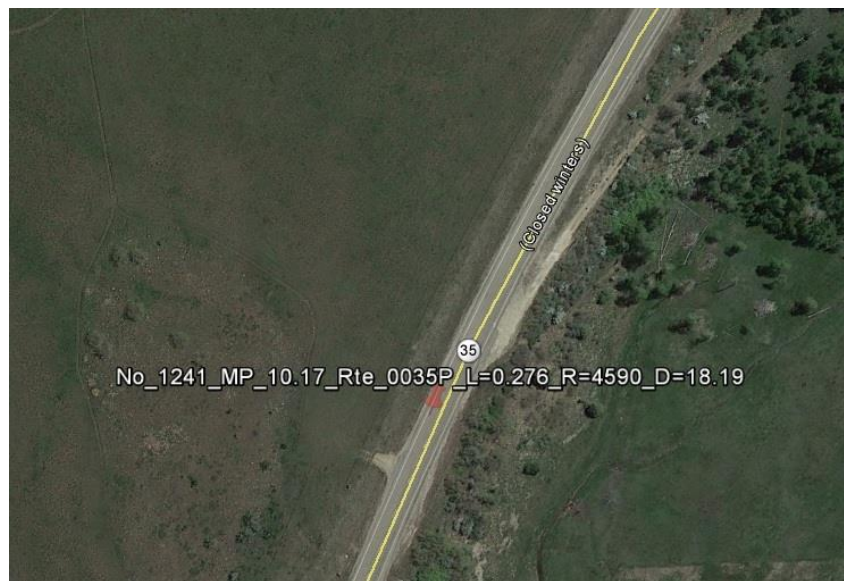


Figure 6. Example of winter closure information in Google Earth.

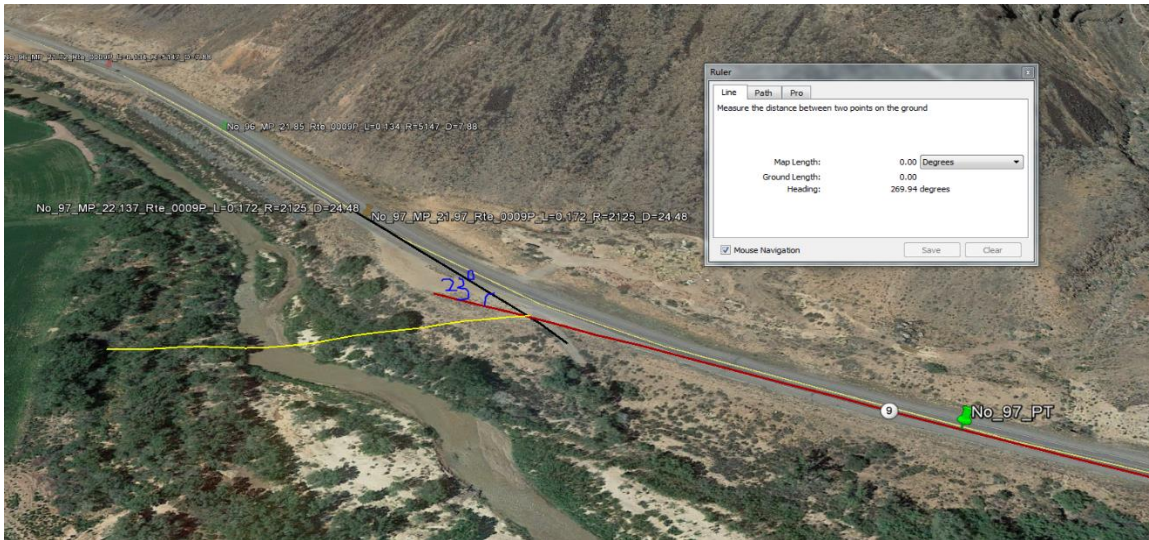


Figure 7. An example of manual measurement in Google Earth.

Table 3. Minimum Design Tangent Length

e(%)	Minimum Tangent Length (ft)						
	Design Speed (mph)						
	35	40	45	50	55	60	65
1.5	197.2	210.8	226.1	244.8	260.1	272	283.9

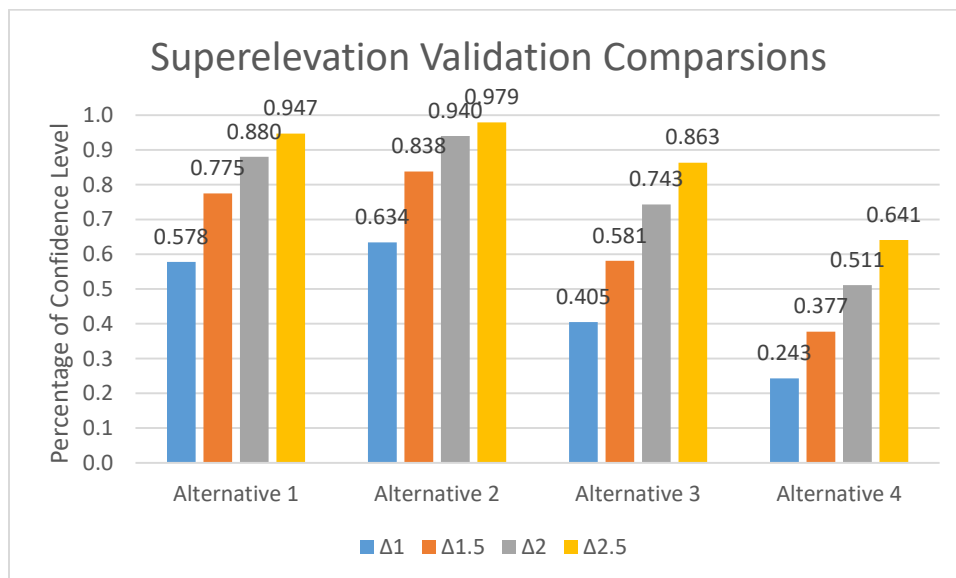


Figure 8. Superelevation validation comparisons

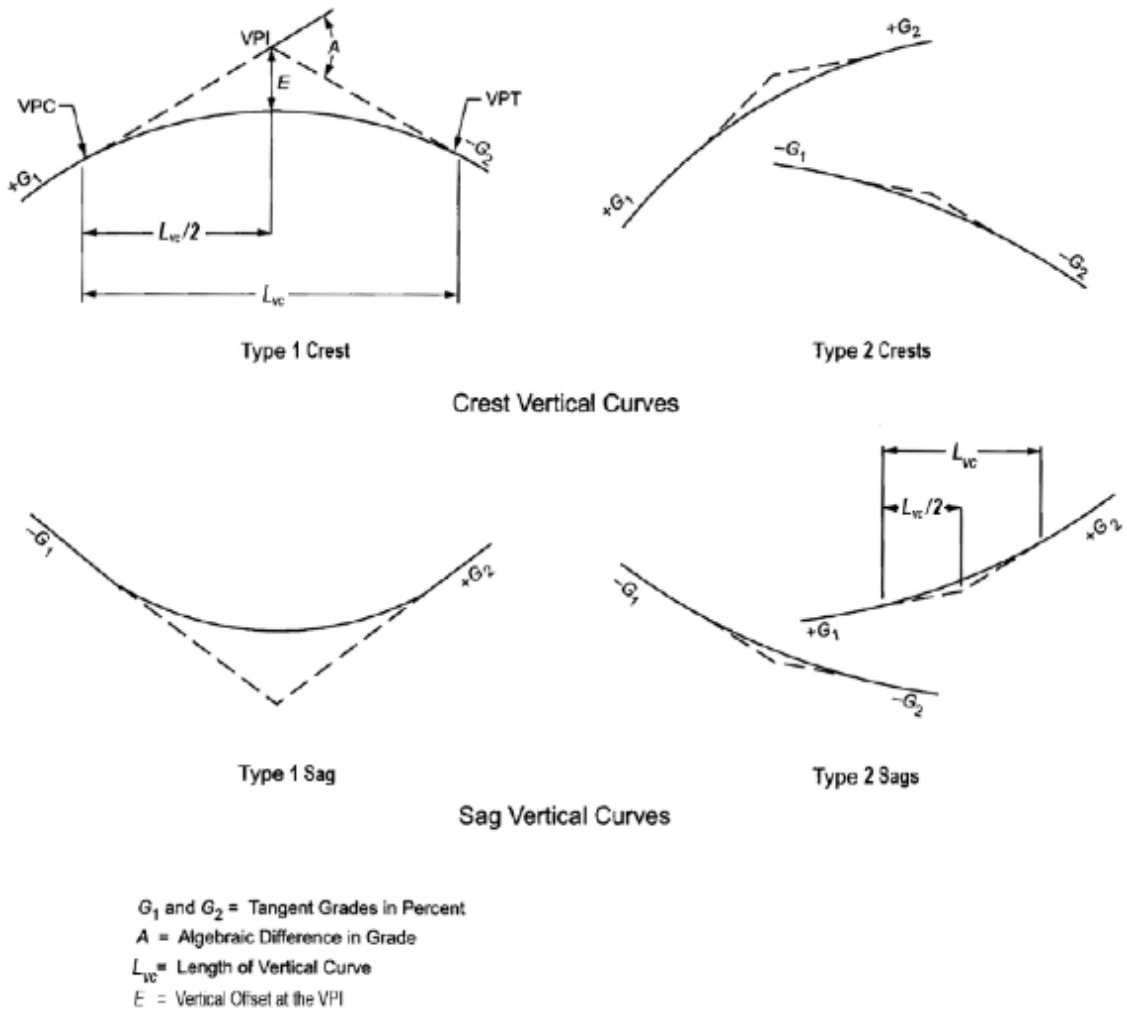


Figure 9. Types of vertical alignments

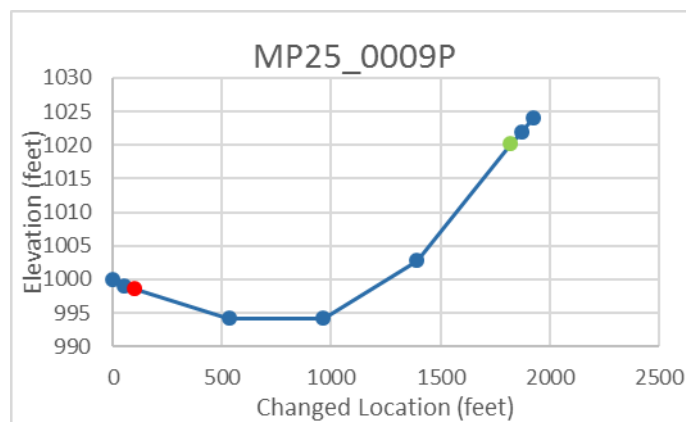


Figure 10. Example of vertical alignment inside of horizontal curve.



Figure 11. Example of video image from Web Navigator.

Table 4. Roadway Departure (RWD) Crash Descriptions

<i>DI-9 Box</i>	<i>Definition</i>	UDOTRWD	UURWD1	UURWD2	Relationship	
28	Roadway/Junction Feature	NOT 4-Leg, T, Y, 5-Leg or More, Roundabout, Ramp Intersection with Crossroad, Bike/Ped Path Intersection			OR	AND
17 (1 or 2)	Driver Contributing Circumstance	Ran Off Road				
		N/A	Failed to Keep Proper Lane			
	Sequence of Events 1	ROR Right, Left, Crossed Median/Centerline, Collision with Fixed Object		OR	AND	
		N/A	Opposite Direction (Head on, Sideswipe Opposite Direction)			
			Overturn			

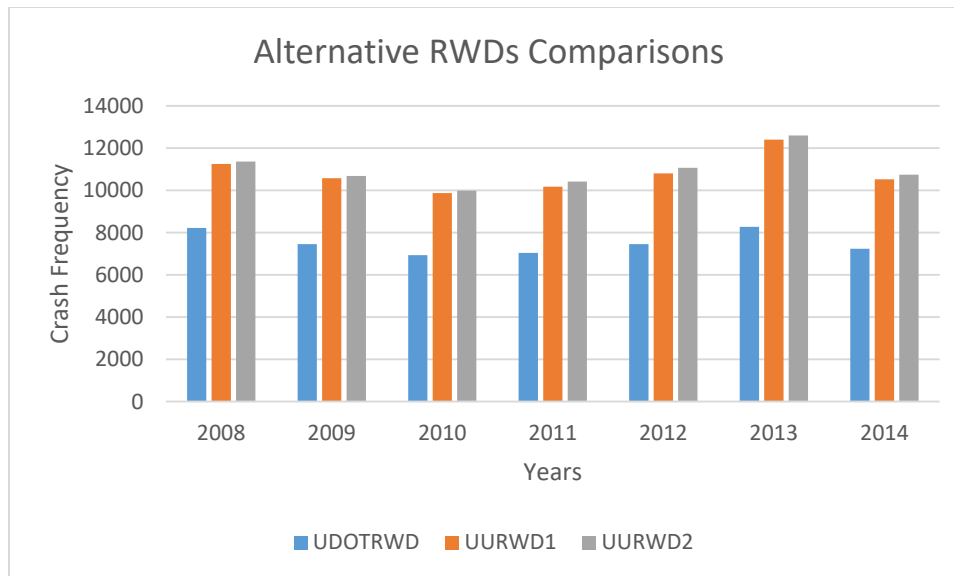


Figure 12. Alternative roadway departure (RWD) crash frequency comparisons among 7 years (2008-2014) in Utah

Table 5. General and Crash Variables Descriptions

General Attributes	
Variables	Definition
C_Dir	Curve Direction
Ln_CL	Natural log of length of horizontal curve in feet
Ln_AADT	Natural logarithm of total AADT (From 2008 to 2014)
Fri_Fatr	The side-friction factors are employed in horizontal curve with post speed, superelevation and radius
PS	Post Speed Limit(mi/h)
PS_30	1= Post Speed Limit at 30 mi/h, 0 = Others
PS_35	1= Post Speed Limit at 35 mi/h, 0 = Others
PS_40	1= Post Speed Limit at 40 mi/h, 0 = Others
PS_45	1= Post Speed Limit at 45 mi/h, 0 = Others
PS_50	1= Post Speed Limit at 50 mi/h, 0 = Others
PS_55	1= Post Speed Limit at 55 mi/h, 0 = Others
PS_60	1= Post Speed Limit at 60 mi/h, 0 = Others
PS_65	1= Post Speed Limit at 65 mi/h, 0 = Others
Crash Attributes	
T_Crash	Total Crash between 2008 and 2014
T_RWD	UU Roadway Departure Crash Type II between 2008 and 2014

Table 6. Horizontal Variables Descriptions

Horizontal Attributes	
Variables	Definition
CL	Distance between PC and PT (Curve segment length, mi)
USM_CL	Upstream Curve Length in the previous curve (mi)
DSM_CL	Downstream Curve Length in the next curve (mi)
USM_TL	Upstream Tangent Length (Distance between PT of previous curve and PC of the tested curve, mi)
DSM_TL	Downstream Tangent Length (Distant between PT of the tested curve and PC of next curve, mi)
Radius	Curve Radius (feet)
USM_R	Upstream Curve Radius (feet)
DSM_R	Downstream Curve Radius (feet)
Curve_D	Degree of curve
USM_D	Degree of Upstream Curve
DSM_D	Degree of Downstream Curve
D_Ang	Deflection Angle
USM_D_Ang	Deflection Angle of Upstream Curve
DSM_D_Ang	Deflection Angle of Downstream Curve
Super_e	The cross slope at the middle of horizontal curve
MCDC	Maximum changed in degree of curve
ACDC	Average changed in degree of curve
MCDA	Maximum changed in deflection angle
ACDA	Average changed in deflection angle

Table 6. (Continued)

Horizontal Alignment Indices Attributes	
Variables	Definition
CCR	The curvature change rate
DC	The degree of curvature
Avg_R	The average radius of curvature
CRR	The changed radius rate
RRR	The ratio of radius and total radii
RTR	The ratio of tangent length over radius
RTR_USM	The ratio of upstream tangent length over radius
RTR_DSM	The ratio of downstream tangent length over radius
RRR_G	1= good design consistency criteria on RRR, 0= other design consistency criteria on RRR
RRR_F	1= fair design consistency criteria on RRR, 0= other design consistency criteria on RRR
RRR_P	1= poor design consistency criteria on RRR, 0= other design consistency criteria on RRR
RTR_G	1= good design consistency criteria on RTR, 0= other design consistency criteria on RTR
RTR_F	1= fair design consistency criteria on RTR, 0= other design consistency criteria on RTR
RTR_P	1= poor design consistency criteria on RTR, 0= other design consistency criteria on RTR

Table 7. Vertical Variables Descriptions

Vertical Attributes	
Variables	Definition
Ai	Absolute gradient difference (%)
G_PC	Grade at point curvature of the horizontal curve (%)
G_1_4CL	Grade at point curvature of the horizontal curve (%)
G_1_2CL	Grade at point curvature of the horizontal curve (%)
G_3_4CL	Grade at point curvature of the horizontal curve (%)
G_PT	Grade at point curvature of the horizontal curve (%)
G_USM_50ft	Grade at 50 feet before PC of the horizontal curve (%)
G_USM_100ft	Grade at 100 feet before PC of the horizontal curve (%)
G_DSM_50ft	Grade at 50 feet after PT of the horizontal curve (%)
G_DSM_100ft	Grade at 100 feet after PT of the horizontal curve (%)
Avg_G	Average Grade inside of the horizontal curve (%)
G_-9_-4	1= grade between -9 and -4 % inside of horizontal curve,0= otherwise
G_-4_0	1= grade between -4 and 0 % inside of horizontal curve,0= otherwise
G_0_4	1= grade between 0 and 4 % inside of horizontal curve,0= otherwise
G_4_9	1= grade between 4 and 9 % inside of horizontal curve,0= otherwise
HC_VC	1= vertical curves on horizontal curve, 0=otherwise
Vertical Alignment Indices Attributes	
VCCR	Vertical curvature change rate
CCR_Combo	The sum of the horizontal curvature change rate and the vertical curvature change rate
Pos_G	Positive straight grade on horizontal curve (%)
Neg_G	Negative straight grade on horizontal curve (%)
TI_Crest	1= type 1 crest curve on horizontal curve, 0= otherwise
TII_Crest	1= type 2 crest curve on horizontal curve, 0= otherwise
TI_Sag	1= type 1 sag curve on horizontal curve, 0= otherwise
TII_Sag	1= type 2 sag curve on horizontal curve, 0= otherwise

Table 8. Summary Descriptive Statistics for General and Crash Disaggregate Data

Variable	Obs.	Mean	Standard deviation	Min	Max
General Attributes					
Curve_Dir	578	0.510	0.500	0	1
CL_mi	578	0.206	0.121	0.047	1.147
Ln_AADT	578	6.523	0.889	4.508	8.594
Post Speed	578	55.908	7.003	30	65
Speed_30	578	0.002	0.042	0	1
Speed_35	578	0.019	0.137	0	1
Speed_40	578	0.024	0.154	0	1
Speed_45	578	0.062	0.242	0	1
Speed_50	578	0.107	0.310	0	1
Speed_55	578	0.464	0.499	0	1
Speed_60	578	0.073	0.260	0	1
Speed_65	578	0.249	0.433	0	1
Crash Attributes					
Tot_Crash	578	0.773	1.538	0	14
Tot_RWD	578	0.391	0.972	0	12

Table 9. Summary Descriptive Statistics for Horizontal Disaggregate Data

Horizontal Attributes					
Variable	Obs.	Mean	Standard deviation	Min	Max
CL_mi	578	0.206	0.121	0.047	1.147
USM_CL_mi	578	0.199	0.113	0.055	1.055
DSM_CL_mi	578	0.202	0.122	0.047	1.248
USM_TL_mi	578	0.474	0.520	0.057	3.214
DSM_TL_ft	578	2551.098	2726.809	300.540	18231.840
Radius_ft	578	2420.676	1331.414	318.281	8872.976
USM_R_ft	578	2385.378	1303.047	335.340	8872.976
DSM_R_ft	578	2353.732	1313.492	318.281	10568.230
Curve_Deg	578	3.225	2.121	0.646	18.003
USM_Deg	578	3.263	2.169	0.646	17.087
DSM_Deg	578	3.294	2.053	0.542	18.003
D_Ang	578	31.959	22.326	3.428	177.492
USM_D_Ang	578	32.383	22.425	4.969	177.492
DSM_D_Ang	578	33.177	23.041	3.428	148.884
Super_e	578	0.039	0.016	0	0.089
MCDC	578	2.031	1.97	0.021	14.543
ACDC	578	1.489	1.505	0.021	13.673
MCDA	578	25.163	21.361	0.871	135.103
ACDA	578	17.854	15.352	0.709	98.127
Horizontal Alignment Indices Attributes					
CCR	578	82.992	63.277	5.126	450.527
DC	578	9.281	8.996	0.675	69.914
Avg_R	578	0.452	0.194	0.112	1.267
CRR	578	1.011	0.317	0.277	2.092
RRR	578	1.104	0.408	0.478	3.614
RTR	578	1.164	0.955	0.108	6.325
RTR_USM	578	1.140	1.184	0.078	8.010
RTR_DSM	578	1.188	1.276	0.100	9.404

Table 10. Summary Descriptive Statistics for Vertical Disaggregate Data

Vertical Alignment Indices Attributes					
Variable	Obs.	Mean	Standard deviation	Min	Max
VCCR	578	0.001	0.002	0	0.012
CCR_Combo	578	0.017	0.012	0.001	0.085
Pos_G	578	0.246	0.431	0	1
Neg_G	578	0.362	0.481	0	1
Type I_Crest	578	0.047	0.211	0	1
Type II_Crest	578	0.130	0.336	0	1
Type I_Sag	578	0.026	0.159	0	1
Type II_Sag	578	0.080	0.271	0	1
HC_VC	578	0.270	0.460	0	2
Vertical Attributes					
Ai	578	1.036	2.157	0	13
G_PC	578	-0.090	2.841	-10	9
G_1_4CL	578	-0.170	2.836	-10	9
G_1_2CL	578	-0.159	2.795	-10	8
G_3_4CL	578	-0.234	2.827	-10	8
G_PT	578	-0.258	2.788	-10	8
G_USM_50ft	578	-0.071	2.852	-10	9
G_USM_100ft	578	-0.087	2.855	-10	9
G_DSM_50ft	578	-0.270	2.799	-10	8
G_DSM_100ft	578	-0.282	2.816	-10	8
Avg_G	578	1.904	1.763	0	10
HC_G_-9_-4	578	0.067	0.251	0	1
HC_G_-4_0	578	0.521	0.500	0	1
HC_G_0_4	578	0.367	0.482	0	1
HC_G_4_9	578	0.043	0.204	0	1

CHAPTER 5

DATA ANALYSIS RESULTS

This chapter includes model estimation results and interpretations. First, all eight major horizontal and vertical alignment indices are evaluated individually with respect to their relationship to the expected number of roadway departure crashes. The purpose of this step is to identify the approximate sensitivity between safety and each alignment index “alone,” and test four different types of specifications regarding the usage of the natural logarithm of both average annual daily traffic and curve length. Second, all variables and alignment indices are evaluated as part of a more comprehensive model specification by using a negative binomial model. Third, zero-inflated Poisson and negative binomial models are estimated to determine how well they fit the data compared to a standard, negative binomial model. The final statistical results in terms of a recommended model will be presented at the end of this section.

5.1 Relationship between Roadway Departure Crashes and

Individual Design Consistency Measures

Before testing the models, the AADT and horizontal curve length, as two major variables in this study, will be transformed in three alternative specifications. Wu et al. (2013) has implemented specifications similar to this study. First, the natural logarithm of

average AADT for 7 years and the natural logarithm of curve length (in miles) are applied as two predictor variables. Second, the natural logarithm of AADT is deployed as a predictor variable, while the natural logarithm of the curve length is regarded as an offset variable (exposure variable). The coefficient of the offset or exposure variable is restricted to one. This specification was used in the crash prediction models which are suggested in the HSM (AASHTO 2010). Moreover, a number of research studies related to the design consistency topic have applied homogeneous segments to overcome heteroskedasticity in regression model analysis (Anderson et al., 1999; Appelt et al., 2000; Ng et al., 2003; Butsick et al., 2015). The third specification regarded the natural logarithm of AADT as the exposure variable and the natural logarithm of segment length as a predictor variable. This common specification in crash modeling has also been implemented by Anastasopoulos et al. (2008) and Kopits and Cropper (2005). In the last specification, the natural logarithm of curve length and AADT are combined as exposure variables (Miaou et al., 2003; Miaou and Song, 2005; Wu et al., 2013).

Eight models for relating each individual alignment index to roadway departure crashes under four alternative specifications are presented in Table 11. Considering the coefficient of each variable for all eight models, Alternative IV has the lowest Pseudo R^2 for each “alone” model, which represents the lowest predictive power. Meanwhile, the P-value and coefficient of each alignment index is obviously insignificant. Wu et al. (2013) also indicated the combined exposure variables between curve length and AADT produced lower predictive power compared with the last three formulations. Then, comparing Alternatives I, II, and III, Alternatives I and II have more predictive power than Alternative III, which means the Pseudo R^2 values of Alternatives I and II are

larger than Alternative III. Meanwhile, each alignment index in Alternative III is less significant than the other two alternatives. This means Alternative III, with Ln AADT as the exposure variable, will be thrown out in the final combination models. Even though the variables in Alternatives I and II had similar P-values, this pattern did not prove to be a predictable assumption when applied in the final combination models. This test cannot be strongly distinguished from the more predictive specification until iterating a reasonable model with combinations of predictive variables. Thus, final models, with or without curve length as an exposure variable, will be finalized after comparing the combination models by employing a negative binomial modeling method.

Analyzing the eight individual alignment index variables, Avg_R and CRR have a negative correlation with roadway departure crashes, and the other six variables have positive correlation in the models. The results of the models with Avg_R and CRR agreed with our expectations regarding the coefficient signs. Increases in the average of the radius, or the radius of the tested curve, reduced roadway departure crashes on the horizontal curve segments. Obviously, the results of the model with RRR have a completely opposite explanation from CRR, since the value of RRR is the inverse of the CRR. For alignment index RTR, either increasing the tangent length or decreasing the radius of the tested curve was estimated to cause more roadway departure crashes. It was noteworthy that CRR, RRR, and RTR have more predictive power compared with the remaining alignment index variables. The results of models with CCR, VCCR, and CCR_Combo indicated that the CCR has higher predictive power than VCCR, which means the horizontal alignment may have more weight than the vertical alignment in CCR_Combo. Although all eight variables have significant P-values in the individual

models, not all variables were significant in the final combination models. The final models will be presented in the following section.

5.2 Relationship between Roadway Departure Crashes and All Design Consistency Measures

The sensitivity relationship between the expected number of roadway departure crashes and all types of alignment indices, estimated in Alternatives I and II, were investigated in this section. Among the aforementioned alignment index variables, the values of CRR and RRR were inverse to each other and had the most predictive power. Thus, the CRR and RRR variables will be used exclusively to distinguish between two different base models, which may include the other six alignment indices and geometric variables. Ultimately, a standard negative binomial regression model with more than 100 model combinations was generated in Model A (including CRR) and Model B (including RRR).

The final models were presented under two different alternative specifications in Table 12 and Table 13. Table 12 and Table 13 presented all statistically significant design consistency measures in Model A and Model B, respectively. To summarize the final results in Model A and Model B, the final alignment indices in both Alternative II-A and Alternative II-B have more predictive power and more statistically significant impact than Alternative I-A and Alternative I-B. In addition, the Pseudo R^2 values in Alternative II-A or B were slightly larger than in Alternative I-A or B. This demonstrated that the Alternative II specification, with the natural logarithm of the horizontal curve length as the offset variable, had a better fit and was used in the final

model to explain design consistency measures. Thus, the final model specification used the natural logarithm of the horizontal curve length as an offset variable.

In model Alternative II-A, it was found that higher values of the tested curve radius or changed radius rate (CRR) significantly reduced the expected number of roadway departure crashes. The estimated coefficient for VCCR indicated that a higher vertical grade change per mile resulted in an increase in the expected frequency of roadway departure crashes. Other significant geometric variables include the maximum change in degree of curvature (MCDC), HC_VC, and Average grade. Higher values of MCDC, representing the sudden change in the degree of the curve, increased the expected number of roadway departure crashes. Horizontal curves with a vertical curve indicator showed a negative impact on expected roadway departure crashes, which might be explained by drivers being less distracted when driving along a horizontal curve with a vertical curve. Thus, a higher VCCR and a lower CRR caused an inconsistent design.

In model Alternative II-B, a similar result was found to Alternative II-A, except with the opposite explanation for RRR. VCCR was also significant in this model. ACDC, which is an extra geometric variable, was shown in this model compared with Alternative II-A. The results indicated that a higher average change in the degree of curvature increased the expected frequency of roadway departure crashes. Overall, the model log-likelihood for Alternative II-A and Alternative II-B were almost the same, as were the pseudo R^2 values. The results indicate that the expected number of roadway departure crashes is most affected by three major alignment indices: CRR, RRR, and VCCR. However, considering the preponderance of zero crashes in the model (mentioned in the methods section), a zero-inflated negative binomial model will be tested in the following

section which will indicate whether the alignment indices are influenced by a zero-state process.

5.3 Exploring “Excessive” Zero Roadway Departure Crashes and All Design Consistency Measures

The relationship between the expected number of roadway departure crashes and all design consistency measures was explored by using a zero-inflated negative binomial model in order to explore the influence of “excessive” zero crashes. Figure 13 illustrates the frequency distribution of roadway departure crashes on all of the principle horizontal curve segments. Almost 80% of the segments (443 out of 578) experienced zero crashes during the 7-year period. A zero-inflated distribution will be satisfied with this assumption appropriately. All the variables in Model A and Model B will be re-tested by using zero-inflated negative binomial models. The two processes in the ZINB include the zero-state and non-zero state for roadway departure crashes. Some of the variables that influence the non-zero crash state process will potentially impact the safety performance effects of design consistency measures in the standard negative binomial models.

Table 14 showed that all of the variables in Model A were tested in a ZINB model, and some of them were significant in the zero-state crash process. All of the variables in Model A were significant in the non-zero state of this model. The predictive power of these variables had no apparent differences compared with Model B, even though Avg_G was insignificant because of the small number of observations for the roadway departure crash process. In the logit part of this model (zero-state), the CRR, VCCR, Avg_G, and natural logarithm of AADT were included in the final ZINB model. Ln_AADT, CRR,

and Avg_G were significant at nearly the 90% confidence level, but VCCR had an insignificant influence on the zero-state crash process. Nonetheless, VCCR was still kept in the final model to improve its predictive performance. Ln_AADT had a negative impact on zero-state crashes, indicating that roadways with higher AADT would be more likely to have roadway departure crashes, which is expected. CRR has a negative coefficient sign, which means higher CRR is more likely to result in observing roadway departure crashes (non-zero crash state). Higher Avg_G is also more likely to result in observing roadway departure crashes. The Vuong test directly demonstrated that ZINB-A is better than ZIP-A, as expected. The model comparison between ZINB-A and NB-A will be presented in the next subsection.

Table 15 showed that all variables in Model B were tested using a ZINB model, and some of them were significant in the zero-state crash process. The explanation for the predictive power in the non-zero state of this model was similar to that given for the ZINB-A. The indicator RRR_P and RRR_F replaced CRR in the logit part of the ZINB. Values classifying RRR as poor design consistency and fair design consistency are more likely to result in a zero-crash state. The results of these two variables intuitively conflict with the finding in Califso et al. (2009). The reason for keeping RRR_P and RRR_F is because it has a more significant impact on the final model compared to RRR. The additional explanatory benefit for keeping these two indicator variables is that it allows testing different thresholds for design consistency classification (poor, fair, and good) based on the safety design criteria method mentioned in Califso et al. (2009). Other variables in the logit part of ZINB-B have a similar explanation to those given for ZINB-A. The Vuong test directly demonstrated that ZINB-B is better than ZIP-B, as expected.

The model comparison between ZINB-B and NB-B will be presented in the next subsection.

5.4 Model Selections

This subsection will determine the best final model among the zero-inflated negative binomial models (ZINB-A and ZINB-B) and the negative binomial models (NB-A and NB-B). To summarize the difference between the models, Figure 14 shows the predicted probability of different roadway departure crash frequencies with different models. Due to only having one site with 12 roadway departure crashes (only 0.1% probability) in the database, it was excluded in this figure. The NB-A and NB-B models produced no roadway departure crashes for 58.9% and 58.3% of observed sites, respectively. This was higher than the 57.8% and 56.6% of sites with no roadway departure crashes in the ZINB-A and ZINB-B models, respectively, although all probability of those models are lower than 76.6% of actual observations. The models generally underestimated the frequency of the zero-state condition, but overestimated the expected frequency of roadway departure crashes.

To quantitatively select the best model, the Vuong test, AIC, and BIC were utilized for comparing models. The Vuong test indicated that the value of V is more than 99% significant in the ZINB model compared with the ZIP model, meaning that the ZINB model is more favored to be applied than the ZIP model, which was intuitively expected.

Model selection between the ZINB and NB models was determined by utilizing the AIC and BIC tests. Table 16 shows the results of the fitted models between NB and

ZINB for the roadway departure crash model with design consistency measures. ΔA and ΔB represent the difference calculated by the value of AIC and BIC in NB and the value of both tests in ZINB. The AIC test for Model A and Model B did not present a significant difference between ZINB and NB, since ΔA and ΔB are both less than 2.5. However, the superiority of the NB model was proven by the BIC test. The absolute values of ΔA and ΔB were more than 10, which revealed a very strong difference between ZINB and NB based on the Raftery' rule. The lower BIC value in the NB model indicated that it is highly favorable over the ZINB model for Model A and Model B. The NB model was adopted as the final model selection, which indicated that homogeneous segments are associated with the overdispersion in observed roadway departure crashes, and the excessive zero-crash state is not properly fitted with the ZINB model.

Even though the values of AIC and BIC in NB-A are lower than in NB-B, it cannot simply prove that the most favorable model is NB-A because the parameters in these two models are not completely the same and have their own special interpretations for the relationships between roadway departure crashes and design consistency measures. Eventually, there are two final models with different parameters that have a significant impact on roadway departure crashes. The final parameters with Model NB-A include the natural logarithm of average annual daily traffic, the changed radius rate, vertical curvature change rate, maximum change in degree of curvature, the indicator variable for vertical curves in horizontal curves, and average grade. The final parameters with Model NB-B include the natural logarithm of average annual daily traffic, the ratio of average radius over radii, vertical curvature change rate, maximum change in degree of curvature,

average change in degree of curvature, the indicator variable for vertical curves in horizontal curves, and average grade.

Table 11. Negative Binomial Models with Individual Design Consistency Measures

Alternative	I	II	III	IV
AI Variable	CCR			
Pseudo R ²	0.086	0.082	0.011	0.006
Coefficient	0.004	0.004	0.003	0.003
Standard Error	0.001	0.001	0.001	0.001
P-Value	0.003	0.001	0.017	0.021
AI Variable	DC			
Pseudo R ²	0.081	0.082	0.001	0.002
Coefficient	0.022	0.031	0.02	0.014
Standard Error	0.01	0.01	0.011	0.01
P-Value	0.028	0.001	0.051	0.159
AI Variable	Avg_R			
Pseudo R ²	0.081	0.08	0.006	0.001
Coefficient	-0.991	-1.339	-0.539	-0.365
Standard Error	0.466	0.463	0.475	0.459
P-Value	0.033	0.004	0.256	0.427
AI Variable	CRR			
Pseudo R ²	0.106	0.106	0.028	0.021
Coefficient	-1.458	-1.6	-1.336	-1.24
Standard Error	0.284	0.283	0.294	0.29
P-Value	0	0	0	0
AI Variable	RRR			
Pseudo R ²	0.108	0.107	0.031	0.024
Coefficient	0.972	1.055	0.933	0.883
Standard Error	0.183	0.183	0.196	0.195
P-Value	0	0	0	0
AI Variable	RTR			
Pseudo R ²	0.082	0.077	0.007	0.002
Coefficient	0.217	0.223	0.136	0.137
Standard Error	0.094	0.098	0.094	0.094
P-Value	0.021	0.023	0.146	0.146
AI Variable	VCCR			
Pseudo R ²	0.079	0.074	0.004	0
Coefficient	0.016	0.016	0.001	0.001
Standard Error	0.01	0.01	0.01	0.001
P-Value	0.084	0.092	0.916	0.886
AI Variable	CCR_Comb			
Pseudo R ²	0.087	0.083	0.01	0.006
Coefficient	0.004	0.005	0.003	0.003
Standard Error	0.001	0.001	0.001	0.001
P-Value	0.001	0.001	0.019	0.022

Table 12. Model A with Design Consistency Measures

NB Models Parameter	Alternative I- A			Alternative II- A		
	Coeff.	Std.Err	P- Value	Coeff.	Std.Err	P- Value
Ln_AADT	1.034	0.123	0	1.058	0.124	0
CRR	- 1.325	0.272	0	- 1.416	0.270	0
VCCR	0.031	0.016	0.044	0.035	0.015	0.024
MCDC	0.087	0.360	0.018	0.092	0.037	0.012
HC_VC	- 0.537	0.351	0.126	- 0.650	0.347	0.061
Avg_G	0.095	0.046	0.04	0.098	0.047	0.038
Ln_CL_Mi	0.675	0.186	0	1	(offset)	
Constant	- 6.108	0.946	0	- 5.675	0.920	0
/lnalpha	- 0.081	0.266		- 0.078	0.268	
Alpha	0.922	0.245		0.925	0.248	
No.of Observation		578			578	
LR chi2		116.37(7)			119.82 (6)	
Prob > chi2		0			0	
Pseudo R2		0.127			0.13	
chibar2(01)		42.14			41.32	
Log likelihood		-397.87			-399.386	

Table 13. Model B with Design Consistency Measures

NB Models Parameter	Alternative I- B			Alternative II- B		
	Coeff.	Std.Err	P- Value	Coeff.	Std.Err	P- Value
Ln_AADT	1.020	0.123	0	1.047	0.124	0
RRR	1.033	0.225	0	1.089	0.225	0
VCCR	0.032	0.015	0.038	0.036	0.015	0.016
MCDC	0.261	0.149	0.079	0.261	0.151	0.084
ACDC	- 0.330	0.208	0.113	- 0.325	0.211	0.123
HC_VC	- 0.584	0.350	0.096	- 0.723	0.346	0.037
Avg_G	0.117	0.047	0.012	0.121	0.048	0.011
Ln_CL_Mi	0.624	0.185	0.001	1	(offset)	
Constant	- 8.464	0.955	0	- 8.120	0.946	0
/lnalpha	- 0.055	0.264		- 0.049	0.266	
Alpha	0.946	0.250		0.952	0.253	
No.of Observation		578			578	
LR chi2		114.59(8)			116.99 (7)	
Prob > chi2		0			0	
Pseudo R2		0.126			0.127	
chibar2(01)		42.07			41.3	
Log likelihood		-398.756			-400.802	

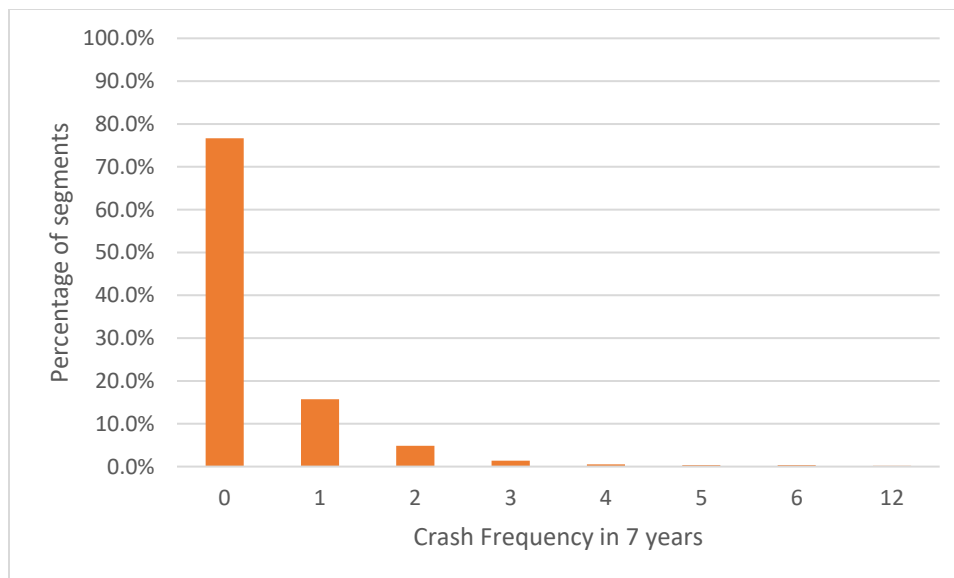


Figure 13. Roadway departure crashes distribution

Table 14. Zero-Inflated Negative Binomial Model A with Design Consistency Measures

Models	ZINB-A		
Parameter	Coefficient	Standard Error	P-Value
Ln_AADT	0.966	0.125	0
CRR	-1.660	0.281	0
VCCR	0.030	0.015	0.047
MCDC	0.111	0.037	0.003
HC_VC	-0.696	0.337	0.039
Avg_G	0.046	0.049	0.352
Constant	-4.612	0.964	0
ln_CL_mi	1	(offset)	
Inflation model	Logit		
Ln_AADT	-3.304	1.879	0.079
CRR	-14.019	6.879	0.042
VCCR	-1.212	9.243	0.896
Avg_G	-5.131	3.158	0.104
Constant	33.520	17.872	0.061
/lnalpha	-0.246	0.292	0.4
Alpha	0.782	0.229	
Number Of Observation	578		
Zero Observation	443		
LR chi2	54.12 (6)		
Prob > chi2	0		
Log likelihood	-393.4361		
ZINB vs ZIP	Pr>chibar2 = 0.0001		
ZINB vs NB	Pr>z =0.024		

Table 15. Zero-Inflated Negative Binomial Model B with Design Consistency Measures

Models		ZINB-B	
Parameter	Coefficient	Standard Error	P-Value
Ln_AADT	0.967	0.132	0
RRR	1.305	0.240	0
VCCR	0.032	0.015	0.03
MCDC	0.307	0.153	0.045
ACDC	-0.386	0.216	0.073
HC_VC	-0.791	0.337	0.019
Avg_G	0.075	0.051	0.139
Constant	-7.597	1.024	0
ln_CL_mi	1	(offset)	
Inflation model		Logit	
Ln_AADT	-1.579	0.995	0.112
RRR_P	3.525	2.583	0.172
RRR_F	4.771	2.449	0.051
VCCR	-0.440	0.537	0.413
Avg_G	-2.474	1.540	0.108
Constant	8.832	6.493	0.174
/lnalpha	-0.257	0.304	0.398
Alpha	0.774	0.235	
Number of Observation		578	
Zero Observation		443	
LR chi2		55.87(7)	
Prob > chi2		0	
Log likelihood		-394.813	
ZINB vs ZIP		Pr>chibar2 = 0.0000	
ZINB vs NB		Pr>z =0.0187	

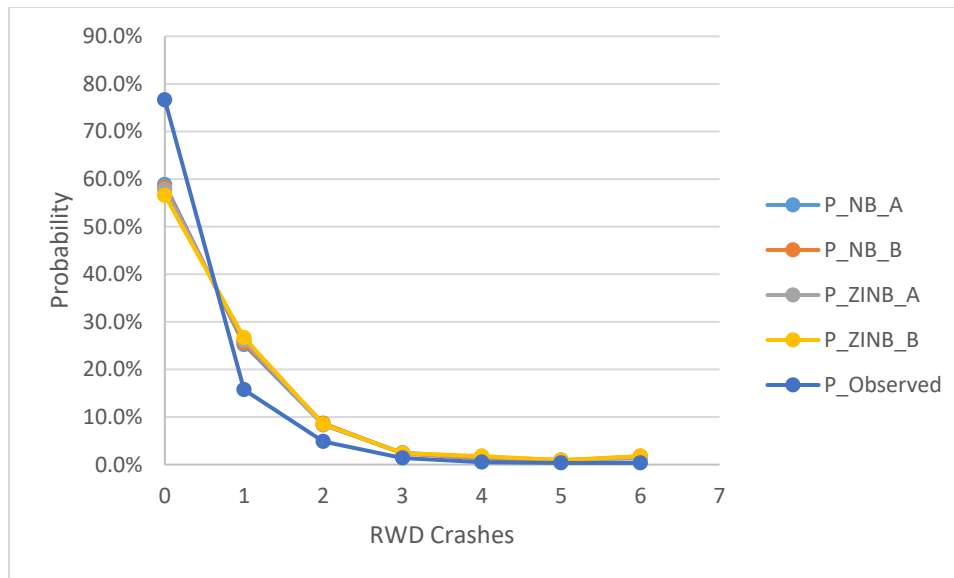


Figure 14. Probability of roadway departure (RWD) crashes among different models

Table 16. Comparisons between Zero-Inflated Negative Binomial and Negative Binomial in Model A and Model B (578 Observations)

Models	NB- A	ZINB- A	ΔA	NB- B	ZINB- B	ΔB
Degree of Freedom	8	13	-5	9	15	-6
AIC	814.773	812.872	1.901	819.605	819.624	-0.019
BIC	849.65	869.547	-19.897	858.841	885.02	-26.179

CHAPTER 6

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

6.1 Summary

Roadway departure crashes are one of the most frequent causes of traffic fatalities in the U.S., leading to over half of all traffic fatalities every year. The AASHTO Highway Safety Manual (HSM) still needs to be improved and updated with new analytical methodologies and techniques for predicting crash frequency on road segments or at intersections. As part of this effort, researchers have evaluated the safety performance effects of design consistency measures, analyzing roadway design attributes with respect to driver expectancy. Geometric alignment indices, as a type of design consistency measure, have been shown to affect safety performance and are intuitively linked to roadway departure crashes. The literature review conducted in this study summarized previous efforts to analyze roadway crashes by different geometric alignment indices and quantitative modeling methodologies.

The objective of this study was to explore the relationship between geometric design consistency measures and the expected number of horizontal curve roadway departure crashes on rural, two-lane highways by using count models. The data used for analysis in this thesis were provided by the Utah DOT and collected using a Mobile LiDAR system. This technology is able to accurately capture geospatial and roadway

inventory data. However, the data were processed in such a way as to support asset management, and the accuracy of certain data elements was at a level consistent with that need and inconsistent with safety analysis. Additional data processing was required to improve the accuracy and reliability of the database for modeling purposes. Vertical grade and cross slope variables were processed and tested before implementing the final models. The final database consisted of 578 principle curve entities, which have a combined length of 900 miles total from 37 highway routes in Utah. Each principle curve entity consists of three successive horizontal curves. Roadway departure crashes were identified by combining Federal Highway Administration and Utah Department of Transportation roadway departure crash definitions. A total of 217 roadway departure crashes were identified on study curve locations between 2008 and 2014.

Negative binomial modeling was employed for modeling the frequency of roadway departure crashes to discover the sensitive geometric variables. To explore the “excessive” zeroes in the database, zero-inflated negative binomial (ZINB) models were also employed to compare with the standard negative binomial model. In the modeling approach, all eight alignment indices (CCR, DC, AVG_R, RRR, CRR, RTR, VCCR, and CCR_Combo) and other general geometric variables were tested in the model, and two exposure variables (the natural logarithm of AADT and horizontal curve length) were tested in different model specifications during the process.

6.2 Findings and Conclusions

This analysis offers a statistical approach to identify the geometric design consistency variables that impact the frequency of roadway departure crashes. The

process of analytical study consisted of four steps, which include the evaluation of individual alignment indices, the determination of final model specifications with different exposures, the identification of geometric design consistency measures, and the selection of the best models. All findings and conclusions will be shown as follows:

6.2.1 Finding and Conclusion 1

Eight alignment indices explored from past safety research were all individually significant in this roadway departure crash study. The estimated effect of these alignment indices are summarized below. Based on the horizontal alignment indices, the higher value of the change radius rate (CRR) and the average radius of curvature (Avg_R) reduced the frequency of roadway departure crashes. Vice versa, the higher value of the ratio of average radius over radii (RRR) increased the frequency of roadway departure crashes. For the ratio of tangent length over radius (RTR), either increasing tangent length or decreasing the radius of the tested curve causes a potential increase in the frequency of roadway departure crashes. In addition, the higher value of the curvature change rate (CCR) and the degree of curvature (DC) increased the frequency of roadway departure crashes. The results indicated that all individual horizontal alignment indices affected the frequency of roadway departure crashes. Among them, CRR and RRR have more influence than the other four alignment indices. Based on vertical alignment indices, even though the higher value of vertical curvature change rate (VCCR) and the composite alignment index (CCR_Comb) may increase the frequency of roadway departure crashes, it provided lower predictive power than the other six horizontal alignment indices. This result indicated that the vertical alignment indices had a smaller effect on the frequency

of roadway departure crashes than the horizontal alignment indices. This finding agrees with the results of previous studies.

6.2.2 Finding and Conclusion 2

The four model specifications were compared based on the goodness-of-fit and predictive power for each variable. The model specification with the natural logarithm of curve length as an offset variable was selected as the best model. It also was used in the crash prediction models which are suggested in the HSM (AASHTO 2010). This model specification had the most predictive power and statistically significant impact on the tested variables compared to the other three model specifications. In addition, this model specification accounts for expected roadway departure crash frequencies increasing with longer curve lengths by assuming that the crash rate is proportional to the curve length (the effect of the offset variable).

6.2.3 Finding and Conclusion 3

The two best models were found by utilizing negative binomial regression analysis. One model's (Model A) final parameters include the natural logarithm of average annual daily traffic (AADT), the changed radius rate (CRR), vertical curvature change rate (VCCR), maximum change in degree of curvature (MCDC), the indicator variable for a vertical curve in a horizontal curve, and average grade. It was found that the higher changed radius rate reduced the expected frequency of roadway departure crashes significantly, which might indicate that higher CRR improves the driver performance consistently. Coefficients for VCCR indicated that the higher vertical grade

change per mile resulted in a higher expected frequency of roadway departure crashes. The higher value of MCDC increased the expected roadway departure crash frequency. The horizontal segments with vertical curvature showed decreases in expected roadway departure crash frequencies. This might be explained by the possibility that it is less distracting/complicated to drive along a horizontal curve with vertical curvature due to its more complex geometry. Thus, higher VCCR and lower CRR indicate an inconsistent design.

The other model's (Model B) final parameters include the ratio of average radius over radii (instead of the changed radius rate), average change in degree of curvature (ACDC), and the rest of variables were the same as the first model. This model provided similar results to Model A, except the opposite effect for RRR. VCCR was also significant in this model. ACDC is the additional geometric variable in this model compared with Model A. It indicated the higher average change in degree of curvature increased the expected frequency of roadway departure crashes. This finding would imply that some geometric elements from alignment indices, including the horizontal curve radius, degree of curvature, and vertical grade, significantly affect both design consistency and crash frequency.

6.2.4 Finding and Conclusion 4

The standard negative binomial model was more favorable to fit the data in this study than the zero-inflated negative binomial model, as indicated by the Vuong test, AIC, and BIC. The zero-inflated negative binomial model may have been adversely impacted by the low sample-mean and small sample size bias, since the dataset of this study is very

small.

6.3 Recommendations

Though this study found valuable information on the sensitivity relationship between roadway departure crashes and geometric design consistency measures, there are a variety of suggestions for improvement in future research. The future recommendations are as follows:

- To overcome heterogeneity caused by temporal and spatial changes, a mixed-effect negative binomial model might be utilized.
- To avoid a large amount of “excessive zeros” in the database, better data from other States might also be applied.
- To increase the number of observations of roadway departure crashes at more locations, the horizontal curve determination algorithm needs to be improved. In addition, a series of successive curve segments need to be implemented instead of a principle curve segment.
- To discuss the roadway departure crashes influenced by horizontal curves with vertical alignment, vertical curve length needs to be determined and the different terrain should be identified.
- To find safety design criteria for RRR in a roadway departure crash study, Cafiso et al. (2009) suggested the threshold of RRR was able to be identified by using linear correlation with speed profile thresholds, which is safety criterion I and safety criterion II (Lamm et al., 1987). Therefore, design speed and 85th percentile speed should be collected in future work.

- To explore the relationship between geometric design alignment and roadway departure crash severity levels, roadway departure crash severity need to be categorized and utilized by a multinomial logit model, which is widely applicable for discrete choice modeling to explore the severity distribution function.
- Speed and performance considerations should also be explored using other design consistency measures (e.g., speed profile, driver workload). These design consistency measures may causes indirectly effect of safety performance.

REFERENCES

- AASHTO. (2010). *Highway safety manual*. American Association of State Highway and Transportation Officials, Washington, D.C.
- Abbess, C., Jarrett, D., and Wright, C. C. (1981). "Accidents at blackspots: Estimating the effectiveness of remedial treatment, with special reference to the" regression-to-mean" effect." *Traffic Engineering & Control*, 22, 535-542.
- Aguero-Valverde, J. (2013). "Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates." *Accident Analysis & Prevention*, 50, 289-297.
- Alexander, G. J., and Lunenfeld, H. (1986). *Driver expectancy in highway design and traffic operations*. US Department of Transportation, Federal Highway Administration, Office of Traffic Operations, Report No. FHWA-TO-86-1, Washington, D.C.
- Anastasopoulos, P. C., Tarko, A. P., and Mannering, F. L. (2008). "Tobit analysis of vehicle accident rates on interstate highways." *Accident Analysis & Prevention*, 40(2), 768-775.
- Anderson, I., Bauer, K., Harwood, D., and Fitzpatrick, K. (1999). "Relationship to safety of geometric design consistency measures for rural two-lane highways." *Transportation Research Record: Journal of the Transportation Research Board*, 1658, 43-51.
- Appelt, V. (2000). "New approaches to the assessment of the spatial alignment of rural roads-apparent radii and visual distortion." *2nd International Symposium on Highway Geometric Design*.
- Blincoe, L., Miller, T. R., Zaloshnja, E., and Lawrence, B. A. (2015). *The economic and societal impact of motor vehicle crashes, 2010 (Revised)*. US Department of Transportation, National Highway Traffic Safety Administration, Report No. DOT HS 812 013, Washington, D.C.
- Butsick, A. J., Jovanis, P. P., and Wood, J. (2015). "Modeling safety effects of geometric design consistency on two-lane rural roads using mixed effects negative binomial regression." *Transportation Research Board 94th Annual Meeting*, No. 15-0797.

- Cafiso, S., and Cava, G. (2009). "Driving performance, alignment consistency, and road safety: Real-world experiment." *Transportation Research Record: Journal of the Transportation Research Board*, 2102, 1-8.
- Castro, M., Pardillo-Mayora, J. M., and Sanchez, J. F. (2005). "Alignment indices as a tool to evaluate safety and design consistency in two lane rural roads." *3rd International Symposium on Highway Geometric Design*.
- Easa, S., and You, Q. (2009). "Collision prediction models for three-dimensional two-lane highways: Horizontal curves." *Transportation Research Record: Journal of the Transportation Research Board*, 2092, 48-56.
- Ellis, N. C. (1972). *Driver expectancy: Definition for design*. Texas Transportation Institute, Texas A&M University, College Station, Texas.
- Faghri, A., and Harbeson, M. (1999). "A knowledge-based GIS approach to the evaluation of design consistency of horizontal alignments." *Transportation Research Record: Journal of the Transportation Research Board*, 1658, 1-8.
- Fitzpatrick, K., and Collins, J. (2000). "Speed-profile model for two-lane rural highways." *Transportation Research Record: Journal of the Transportation Research Board*, 1737, 42-49.
- Fitzpatrick, K., Wooldridge, M. D., Tsimhoni, O., Collins, J. M., Green, P., Bauer, K. M., Parma, K. D., Koppa, R., Harwood, D. W., Anderson, I., Krammes, R. A., and Poggioli, B. (2000). *Alternative design consistency rating methods for two-lane rural highways*. US Department of Transportation, Federal Highway Administration, Report No. FHWA-RD-99-172, Washington, D.C.
- Gibreel, G. M., Easa, S. M., Hassan, Y., and El-Dimeery, I. A. (1999). "State of the art of highway geometric design consistency." *Journal of Transportation Engineering*, 125(4), 305-313.
- Glennon, J. C., and Harwood, D. W. (1978). "Highway design consistency and systematic design related to highway safety." *Transportation Research Record: Journal of the Transportation Research Board*, 681, 77-88.
- Greene, W. H. (2000). *Econometric analysis*. 4th ed. Prentice Hall, New Jersey.
- Hauer, E. (1997). *Observational before-after studies in road safety--estimating the effect of highway and traffic engineering measures on road safety*. Pergamon, Tarrytown, New York.
- Hauer, E., Council, F., and Mohammedshah, Y. (2004). "Safety models for urban four-lane undivided road segments." *Transportation Research Record: Journal of the Transportation Research Board*, 1897, 96-105.

- Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press, New York, New York.
- Hosseinpour, M., Yahaya, A. S., and Sadullah, A. F. (2014). "Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian Federal Roads." *Accident Analysis & Prevention*, 62, 209-222.
- Jalayer, F., De Risi, R., De Paola, F., Giugni, M., Manfredi, G., Gasparini, P., Topa, M. E., Yonas, N., Yeshitela, K., and Nebebe, A. (2014). "Probabilistic GIS-based method for delineation of urban flooding risk hotspots." *Natural Hazards*, 73(2), 975-1001.
- Jalayer, M., and Zhou, H. (2016). "Overview of safety countermeasures for roadway departure crashes." *Institute of Transportation Engineers. ITE Journal*, 86(2), 39-46.
- Jones, B., Janssen, L., and Mannering, F. (1991). "Analysis of the frequency and duration of freeway accidents in Seattle." *Accident Analysis & Prevention*, 23(4), 239-255.
- Jovanis, P. P., and Chang, H.-L. (1986). "Modeling the relationship of accidents to miles traveled." *Transportation Research Record: Journal of the Transportation Research Board*, 1068, 42-51.
- Kopits, E., and Cropper, M. (2005). "Traffic fatalities and economic growth." *Accident Analysis & Prevention*, 37(1), 169-178.
- Lambert, D. (1992). "Zero-inflated Poisson regression, with an application to defects in manufacturing." *Technometrics*, 34(1), 1-14.
- Lamm, R., and Choueiri, E. M. (1987). "Recommendations for evaluating horizontal design consistency based on investigations in the state of New York." *Transportation Research Record: Journal of the Transportation Research Board*, 1122, 68-78.
- Lamm, R., Psarianos, B., and Mailaender, T. (1999). *Highway design and traffic safety engineering handbook*. McGraw-Hill, New York, New York.
- Lee, J., and Mannering, F. (2002). "Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis." *Accident Analysis & Prevention*, 34(2), 149-161.
- Lord, D., Washington, S. P., and Ivan, J. N. (2005). "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory." *Accident Analysis & Prevention*, 37(1), 35-46.

- Lord, D., Washington, S., and Ivan, J. N. (2007). "Further notes on the application of zero-inflated models in highway safety." *Accident Analysis & Prevention*, 39(1), 53-57.
- Messer, C. J. (1980). "Methodology for evaluating geometric design consistency." *Transportation Research Record: Journal of the Transportation Research Board*, 757, 7-14.
- Messer, C. J., Brackett, Q., and Mounce, J. M. (1981). *Highway geometric design consistency related to driver expectancy*. US Department of Transportation, Federal Highway Administration, Report No. FHWA-RD-81-035, Washington, D.C.
- Miaou, S.-P. (1994). "The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions." *Accident Analysis & Prevention*, 26(4), 471-482.
- Miaou, S.-P., and Lum, H. (1993). "Modeling vehicle accidents and highway geometric design relationships." *Accident Analysis & Prevention*, 25(6), 689-709.
- Miaou, S.-P., and Song, J. J. (2005). "Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion and spatial dependence." *Accident Analysis and Prevention*, 37(4), 699-720.
- Miaou, S.-P., Song, J. J., and Mallick, B. K. (2003). "Roadway traffic crash mapping: A space-time modeling approach." *Journal of Transportation and Statistics*, 6(1), 33-58.
- Miranda-Moreno, L., Fu, L., Saccomanno, F., and Labbe, A. (2005). "Alternative risk models for ranking locations for safety improvement." *Transportation Research Record: Journal of the Transportation Research Board*, 1908, 1-8.
- Morrall, J. F., and Talarico, R. J. (1994). "Side friction demanded and margins of safety on horizontal curves." *Transportation Research Record: Journal of the Transportation Research Board*, 1435, 145-152.
- Ng, J. C. W., and Sayed, T. (2004). "Effect of geometric design consistency on road safety." *Canadian Journal of Civil Engineering*, 31(2), 218-227.
- Poch, M., and Mannering, F. (1996). "Negative binomial analysis of intersection-accident frequencies." *Journal of Transportation Engineering*, 122(2), 105-113.
- Post, T. J., Alexander, G. J., and Lunenfeld, H. (1981). *A user's guide to positive guidance*. US Department of Transportation, Federal Highway Administration, Report No. FHWA-TO-81-1, Washington, D.C.

- Raftery, A. E. (1995). "Bayesian model selection in social research." *Sociological Methodology*, 25, 111-163.
- Reurings, M., Janssen, T., Eenink, R., Elvik, R., Cardoso, J., and Stefan, C. (2006). *Accident prediction models and road safety impact assessment: A state-of-the-art*. SWOV Institute for Road Safety Research, Leidschendam, Netherlands.
- Saito, M., Knecht, C. S., Schultz, G.G., and Cook, A.A. (2015). *Crash prediction modeling for curved segments of rural two-lane two-way highways in Utah*. Utah Department of Transportation, Report No. UT-15.12, Salt Lake City, Utah.
- Shankar, V., Mannering, F., and Barfield, W. (1995). "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies." *Accident Analysis & Prevention*, 27(3), 371-389.
- Tang, J., and Zakhor, A. (2011). "3D object detection from roadside data using laser scanners." *Proc., SPIE 7864, Three-Dimensional Imaging, Interaction, and Measurement*, International Society for Optics and Photonics.
- Tarris, J., Poe, C., Mason Jr, J., and Goulias, K. (1996). "Predicting operating speeds on low-speed urban streets: Regression and panel analysis approaches." *Transportation Research Record: Journal of the Transportation Research Board*, 1523, 46-54.
- U.S. Department of Transportation Federal Highway Administration. (2015). "Roadway Departure Safety." <http://safety.fhwa.dot.gov/roadway_dept/> (Feb.16, 2017).
- Utah Department of Public Safety Highway Safety Office. (2014). "2014 Utah Crash Facts." <<http://site.utah.gov/dps-highwaysafe/wp-content/uploads/sites/22/2015/02/overviewFactSheet2014.pdf>>
- Vuong, Q. H. (1989). "Likelihood ratio tests for model selection and non-nested hypotheses." *Econometrica: Journal of the Econometric Society*, 57(2), 307-333.
- Washington, S. P., Karlaftis, M. G., and Mannering, F. (2010). *Statistical and econometric methods for transportation data analysis*. CRC Press, Boca Raton, Florida.
- Wooldridge, M. D., Fitzpatrick, K., Harwood, D. W., Potts, I. B., Elefteriadou, L., and Torbic, D. J. (2003). *Geometric design consistency on high-speed rural two-lane roadways*. Transportation Research Board, National Cooperative Highway Research Program, Report No. 502, Washington, D.C.
- Wu, K. F., Donnell, E. T., Himes, S. C., and Sasidharan, L. (2013). "Exploring the association between traffic safety and geometric design consistency based on vehicle speed metrics." *Journal of Transportation Engineering*, 139(7), 738-748.