Identifying High Risk Locations of Animal-Vehicle Collisions on Washington State Highways

WA-RD 752.1
(TNW 2010-04)

Yinhai Wang
Yunteng Lao
Yao-Jan Wu
Jonathan Corey

October 2010

Washington State
Department of Transportation
Office of Research & Library Services

WSDOT Research Report

TRANS NOW

Final Research Report
Agreement T4118 Task 34
High Risk Loc for AVC
TNW 2010-04
TransNow Budget 61-8379

Identifying High Risk Locations of Animal-Vehicle Collisions on Washington State Highways

by

Yinhai Wang	Yunteng Lao	Yao-Jan Wu	Jonathan Corey
Associate Professor	Graduate Research Assistant	Graduate Research Assistant	Graduate Research Assistant

Department of Civil and Environmental Engineering
University of Washington
Seattle, Washington 98195-2700

Washington State Transportation Center (TRAC)
University of Washington, Box 354802
1107 NE 45th Street, Suite 535
Seattle, Washington 98105-4631

Washington State Department of Transportation Technical Monitor
Kelly McAllister
Transportation Planning Specialist, Environmental Services Office

Sponsored by

Washington State Department of Transportation
Paula J. Hammond, Secretary
Olympia, Washington 98504-7372

Transportation Northwest (TransNow)
University of Washington
135 More Hall, Box 352700
Seattle, Washington 98195-2700

and in cooperation with
U.S. Department of Transportation
Federal Highway Administration

October 2010

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
WA-RD 752.1 / TNW 2010-04		

4. TITLE AND SUBTITLE	5. REPORT DATE
Identifying High Risk Locations of Animal-Vehicle Collisions on Washington State Highways	October 2010
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Yinhai Wang, Yunteng Lao, Yao-Jan Wu, and Jonathan Corey	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
Transportation Northwest Regional Center X (TransNow) Box 352700, 129 More Hall University of Washington Seattle, Washington 98195-2700 Washington State Transportation Center (TRAC) University of Washington, 354802 1107 NE 45th Street, Suite 535 Seattle, Washington 98105-4631	11. CONTRACT OR GRANT NO. Agreement T4118, Task 34 TransNow Contract 61-8379
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Washington State Dept of Transp Transportation Building, MS 47372 Olympia, Washington 98504-7372 Rhonda Brooks, 360-705-7945 U.S. Department of Transportation Office of the Secretary of Transp. 400 7th St. SW Washington, DC 20590	Final Research Report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This study was conducted in cooperation with University of Washington, Washington State Department of Transportation and the U.S. Department of Transportation

16. ABSTRACT

Animal-vehicle collisions (AVCs) have been increasing with increases in both animal populations and motor vehicle miles of travel and have become a major safety concern nationwide. Most previous AVC risk studies have not considered factors related to human behavior or the spatial distribution of animal populations in depth because of missing datasets or the poor quality of data. The two common sources of data—the Collision Report (CRpt) and Carcass Removal (CR) datasets—are often found significantly different. To address these data issues, two approaches were followed in this research. In the first approach, a fuzzy logic-based data mapping algorithm was developed to obtain a more complete AVC dataset from the CRpt and CR data. In comparison to the original CR dataset, the combined dataset increased the number of AVC records by 13~22 percent. This combined dataset was used to develop and calibrate a microscopic probability (MP) model that can explicitly consider drivers' behaviors and the spatial distributions of animal populations. In the second approach, a Diagonal Inflated Bivariate Poisson (DIBP) regression model was developed to fit the two datasets simultaneously. The DIBP model can effectively identify the overlapping parts of the two datasets and quantify the impacts of road and environmental factors on AVCs.

Both proposed models used the CRpt and CR data collected from ten selected study routes in Washington state. The MP model results showed that variables including number of lanes and animal habitat areas are significantly associated with the probability of animals crossing the highway. Two factors, speed limit and truck percentage, have impacts on the probability of a driver's ineffective response. A wider median may decrease the probability of an animal failing to avoid a collision. The DIBP results showed that speed limit, restrictive access control, and roadway segment length have an increasing relationship with AVCs. Furthermore, hotspots (high risk roadway segments) were identified for all the study routes on the basis of the modeling and data analysis results. These quantitative results will help WSDOT develop countermeasures to AVCs.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Animal vehicle collision, accident data, fuzzy logic, accident risk modeling, human factors.	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22616

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
None	None		

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and accuracy of the data presented herein. This document is disseminated through the Transportation Northwest (TransNow) Regional Center under the sponsorship of the U.S. Department of Transportation UTC Grant Program and through the Washington State Department of Transportation. The U.S. Government assumes no liability for the contents or use thereof. Sponsorship for the local match portion of this research project was provided by the Washington State Department of Transportation. The contents do not necessarily reflect the views or policies of the U.S. Department of Transportation or Washington State Department of Transportation. This report does not constitute a standard, specification, or regulation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Animal-vehicle collisions (AVCs) have been increasing nationally, with increases in both animal populations and motor vehicle miles of travel, and AVCs have become a major safety concern nationwide and in Washington state. State Farm Insurance reported that deer and vehicle collisions in Washington state increased by 15 percent over the five years from 2002 to 2007. Therefore, a good understanding of the relationship between AVCs and their associated factors, such as roadway geometry and traffic characteristics, is important in order for transportation agencies to select effective countermeasures against AVCs. High quality AVC data and reasonable explanatory models are essential for scientific investigations of this issue.

Collision Report (CRpt) data and Carcass Removal (CR) data are often used in AVC studies. However, previous studies have found that these two commonly used datasets are very different from each other. This implies that both datasets are incomplete records of ground-truth AVCs, and analyses based solely on the CRpt data or the CR data may result in biased results. Therefore, this data issue must be properly addressed in AVC modeling and statistical analysis. Two approaches to deal with the data problem were studied in this research. The first approach was to combine the two datasets to obtain a more complete dataset that could be used by conventional econometric models for AVC analysis. The second approach was aimed at developing a modeling structure that could use both datasets simultaneously.

In the first approach, a fuzzy logic-based mapping algorithm was developed for merging the two datasets. This proposed mapping algorithm can identify the intersections of the two datasets so that duplications can be avoided. Additionally, because records at the intersection of the two datasets contain more information about the same accidents, this more detailed AVC dataset can enable more thorough analyses. For the selected period of the study routes, about 27 percent to 37 percent of the CRpt data could be matched to the CR data. The union of the two datasets contained more samples than either of the two original datasets. In comparison to the original CR dataset, the combined dataset increased the number of records by 13 percent to 22 percent. The proposed fuzzy logic-based mapping algorithm also matched records in a manner consistent with that of experts, as found in an evaluation survey. The survey was conducted at the Washington State Department of Transportation (WSDOT), where experts were asked to provide their judgment about whether the selected data pairs matched. The verification results showed that the accuracy of the proposed algorithm was approximately 90 percent for the limited pairs of data included in the survey. The fuzzy logic-based mapping algorithm was proved appropriate for enhancing the quality of AVC data. The improved dataset will definitely benefit AVC risk modeling and statistical analysis.

Using the recovered data from the fuzzy logic-based mapping algorithm, descriptive statistical analysis and hypothesis testing were conducted to achieve a better understanding of the characteristics of AVCs and the explanatory variables significantly associated with AVCs. Impacts due to location and season, roadway geometric factors, traffic features, and animal distribution characteristics were analyzed in greater detail. Some factors, such as speed limit, annual average daily traffic (AADT), percentage of trucks, and animal distribution, were found to have significant impacts on AVCs. The statistical analysis results identified five segments as high-risk locations worthy of immediate attention:

- US 2 between mileposts 297.42 and 297.97
- US 12 between mileposts 356.61 and 356.94
- I 90 between milepost 84.15 and 84.35
- US 97 between mileposts 284.2 and 295.81
- US 395 between mileposts 214.35 and 217.62.

Econometric models are frequently used to evaluate the impacts of explanatory variables on AVC risks. However, most existing AVC risk studies have not considered human factors or the spatial distribution of animal population (habitat), although they play a crucial role in the probability of an AVC occurrence. Therefore, a microscopic probability (MP) model, consisting of the probability of a driver's ineffective response and the probability of encountering an animal, was applied in this study. Nineteen explanatory variables, including speed limit and shoulder width, were considered in this MP model. Variables such as number of lanes and whether a highway section routed through animal habitats were found to be significantly associated with the probability of encountering an animal. Increasing median width was found to be significantly associated with decreases in the probability of an animal's failing to avoid the collision. Two of the variables, speed limit and percentage of trucks, were found to be significantly correlated with an increase in the probability of a driver's ineffective response.

In the second approach, a diagonal inflated bivariate Poisson (DIBP) regression model simultaneously takes into account both the CR and AVC datasets in our modeling work. Our models' estimation results indicated that the DIBP model outperformed the double Poisson (DP) model, bivariate Poisson (BP) model, and zero-inflated DP (ZIDP) model. Functionally, the DIBP model was able to handle not only under- or over-dispersed count data but also to model paired datasets with correlation.

The principal findings from applying the MP and DIBP regression models can be used to develop countermeasures against AVCs. The findings are summarized as follows:

- Animal priority habitat areas, particularly for white-tailed deer, are significantly associated with the probability of animal crossings.
- Roadway segments with a speed limit higher than 50 mph correspond to a higher rate of AVCs, which may be due to drivers' failure to react quickly enough to animal crossings.
- There is a correlation between increased highway access control of and reduced probability of AVCs.
- The probability of drivers' reaction failure becomes lower when the percentage of trucks is higher than 5 percent.
- The probability of collision with a crossing animal decreases when the number of lanes increases.
- Increased median width decreases the occurrence of AVCs.

On the basis of the modeling results and the geospatial data, this study further identified AVC hot spots for all study routes. The investigation used the online State Route (SR) Web tool developed by WSDOT and found that rural area roadway sections and deer habitat areas had a higher AVC frequency. Roadway design, such as shoulder width, could be a controllable factor related to AVCs that could be used to mitigate such collisions.

Overall, the research team would like to make the following recommendations to WSDOT:

- For the purpose of improving AVC data, it would be helpful to include specific animal types in the CRpt. With this information, the CRpt data and the CR data can be better matched. Beginning in 2010, WSDOT is using descriptions for deer, elk, and moose in its collision database.
- Plans for new highways should avoid bisecting high quality animal habitats. Highways that through animal habitats have negative effects on animal activities and ecology. In addition, animal movements between bisected habitats increase highway-crossing activities and hence increase the probability of AVCs.
- For existing highways that pass through high-density animal habitats, engineering solutions can be applied to reduce AVC risk. These solutions should focus on reducing either animal-vehicle interactions or the probability of drivers' failure in responding to animal presence.
- Solutions aimed at reducing the probability of encountering animals are desirable. These solutions include preventing animal crossing movements with fence or grade separation at hot crossing spots.
- Solutions are also needed to reduce the probability of drivers' failure to react. These solutions include installing warning systems that can alert drivers to pay more attention when highway-crossing animals are detected and setting the speed limit lower than 50 mph to increase drivers' available perception and reaction time when animals are present.
- This study also found that when the percentage of trucks is above 5 percent, AVC risk is lower. This is likely related to the fact that when more trucks are present, traffic speed is typically slower and/or drivers are more careful in driving. Other factors such as increased noise or visual effects may also play a role by deterring wildlife crossing attempts. However, more data and analysis are desirable to confirm this finding.

CHAPTER 1. RESEARCH BACKGROUND

1.1 RESEARCH BACKGROUND

The national number of animal-vehicle collisions (AVCs) has been rising year after year with continuing increases in both deer populations and motor vehicle traffic (Curtis and Hedlund, 2005). Romin and Bissonettee (1996) reported that at least 500,000 deer-vehicle collisions occurred nationwide in 1991. In Washington state, approximately 3,000 collisions with deer and elk occur annually on state highways (Wagner and Carey, 2006). The number of insurance claims resulting from AVCs has grown correspondingly, indicating that the losses from these accidents have grown. Identifying potential countermeasures against AVCs has become a complex, worldwide, interdisciplinary issue with important implications for traffic safety and the environment.

Over the past two decades, biologists, engineers, and others have gained a better understanding of the impacts of transportation facilities on wildlife. Studies have determined that one of the best solutions to preventing AVCs is to avoid separating animal habitats when new roads are constructed. Although it may not always be possible to create a transportation network without habitat segmentation, care should be taken with regard to the highest density habitats. To reduce the effects of transportation on habitat areas, environmental impact analysis on habitat is now frequently incorporated into the planning stages of transportation facilities. However, the planning and construction of most existing highways did not include the preservation of habitat connectivity. The consequent lack of connectivity may cause the animals to interact with vehicles when animals move between habitat fragments, and these movement conflicts may become AVCs (WSDOT, 2007).

The increasing number of AVCs suggests that countermeasures against AVCs are immediately needed. Over the past decades, many countermeasures have been attempted to reduce AVCs in the U.S. These countermeasures include expensive engineering solutions to connect fragmented habitats and cost-effective ways to warn drivers or animals when potential conflicts are detected. For example, the U.S. Highway 93 improvement project (between Evaro and Polson, Montana) employed the context-sensitive design concept and serves as a great example of how engineering solutions can connect fragmented habitats. According to Frazier (2001), 42 animal crossings and 14 miles of fence were planned over the 56-mile corridor to channel animal movements and reduce their conflicts with highway traffic. Of course, the construction and maintenance costs for such proposed installations are high.

In addition to roadway and environmental factors, human factors are also relevant to AVCs. Most AVCs could have been avoided if drivers had been vigilant. A watchful driver can often see an animal at the roadside or on the road soon enough during the daytime. However, human detection of emerging animals becomes more difficult when the light is dim, especially at night. Because of this and the fact that many animals are most active during the evening and early morning hours, approximately 90 percent of deer-vehicle collisions happen between dusk and dawn (TranSafety, 1997). To help reduce such collisions, automatic detection-based warning systems have been developed and implemented for AVC mitigation. There are two types of warning systems (Huijser and McGowen, 2003): one warns drivers when animals are detected within the conflict range and the other repels wildlife by means of sounds, lights, or scent when vehicles are present. Depending on operating principles and effective distances, these systems may also require the installation of fencing along substantial portions of the affected road segments.

No matter which approach is adopted, it is essential to understand where the high-risk AVC locations are and which roadway factors are associated with AVCs in order to effectively allocate the limited safety improvement resources. Data analysis and statistical modeling based on real-life data are common techniques for dealing with these issues. However, data quality has been a great concern for analysts because of various data collection procedures. Typically, two types of data are available for AVC studies: Collision Report (CRpt) data and Carcass Removal (CR) data. Ideally, these two sets of data should correspond exactly. However, previous studies (for example Knapp et al., 2007 and Huijser et al., 2007) found that these two datasets are significantly different. These differences imply that not all animal carcasses were removed and reported by transportation agencies. Additionally, not all collisions resulted in submitted reports (reports are only required if a specific threshold of property damage has been reached). For these reasons, either dataset would underestimate the actual number of AVCs to some extent. However, if AVC data can be fully recovered by combining these two datasets, or if a collision model is able to consider both datasets together, then AVC analysis can be more accurately conducted and the analysis results can be more reliable. Another issue researchers have to face is choosing suitable statistical models for analyzing AVCs and identifying their causal factors. Generalized linear models (GLMs) have been used to model collisions in many studies (see for example, Miaou, 1994 and Kim et al., 2007). However, most accident modeling studies have not considered human factors, which are known to play a critical role in the crash event (Wang et al., 2003). Moreover, few studies have considered the effects of the spatial distribution of animal populations along the highway.

In summary, there are two major challenges in identifying high-risk AVC locations: the first is to recover the AVC data from the two datasets, and the second is to develop probability models for quantifying the relationships between explanatory variables and AVC risk. This research project used the Collision Report (CRpt) data and Carcass Removal (CR) from the Washington State Department of Transportation (WSDOT) to describe the process of identifying high-risk AVC locations.

1.2 RESEARCH OBJECTIVE

In this project, we developed a data mapping methodology for recovering AVC data and then used the recovered data for AVC risk models. A probability model that considers the driver's response and presence of an animal was developed and calibrated by using different regression forms, including Poisson and negative binomial. High AVC risk locations were identified on the basis of the results from both the descriptive statistical analyses and AVC risk modeling. Roadway factors associated with AVCs were also identified from the probability models. More specifically, the objectives of this study were as follows:

- Design a data mapping approach to maximally recover AVC data from the CRpt and animal CR datasets.
- Build a relational database that stores the recovered AVC data, highway geometric data, traffic data, and animal habitat data.
- Develop new modeling approaches that can consider the occurrence mechanisms of AVCs to quantify the relationships between AVCs and roadway and environmental variables.
- Identify high AVC risk locations on selected Washington state highways.

1.3 REPORT ORGANIZATION

There are nine chapters in this report. The remainder of this report is organized as follows:

Chapter 2 reviews related research on accident data quality and accident modeling. Chapter 3 introduces the AVC study routes in Washington State and our data collection plan. This is followed by Chapter 4, which presents methodologies for database design, data recovery, statistical analysis, and risk modeling. In chapters 5 through 7, details of our data recovery, statistical analysis, and risk modeling efforts are described, respectively; these are based on data from the study routes in Washington state. The high AVC risk locations are identified in Chapter 8. Finally, Chapter 9 provides research conclusions and summarizes recommendations.

CHAPTER 2. STATE OF THE ART

AVC data recovery and risk modeling were the two major challenges in this research project. This chapter reviews the existing studies relevant to these two challenges.

2.1 DATA RECOVERY

In most AVC studies, two types of AVC data are usually used: the CRpt data and the CR data, as described in a National Cooperative Highway Research Program (NCHRP) report by the Western Transportation Institute (Huijser et al., 2007). Since the records compiled in the two datasets are collected by different agencies (in Washington state, the Washington State Patrol and WSDOT) using different methods, data integration and interpretation are issues. Therefore, most previous AVC studies have used either the CRpt data or the CR data, treating the two datasets separately. For example, Hubbard et al., (2000), Malo et al., (2004) and Seiler (2005) conducted their AVC analyses on the basis of the CRpt data, whereas Reilley and Green (1974), Allen and McCullough (1976), and Knapp and Yi (2004) employed the CR data in their research.

In a survey conducted by the research team in 2009, carcass removal professionals at the WSDOT estimated that over 90 percent of the carcasses removed from the road were likely involved in traffic accidents. Therefore, these two sets of data should have significant overlap. However, previous studies (Romin and Bissonette, 1996; Knapp et al., 2007; Huijser et al., 2007) have found that they are significantly different. This suggests that the two sets of data complement each other and should be combined to improve the quality of AVC data. Analyses based solely on the CRpt data or the CR data may result in biased results.

The fuzzy logic-based data-mapping algorithm has proved to be an effective way to deal with problems related to linguistic vagueness and human factors (Zhao, 1997). Fuzzy logic mapping algorithms have been widely used in various fields of transportation engineering, such as ramp metering (Taylor and Meldrum, 1998), speed control systems (Rao and Saraf, 1995), and map matching issues (Syed and Cannon, 2004; Mohammed et al., 2006). Generally, the fuzzy logic-mapping algorithm involves three major steps (Chen and Pham, 2001): (1) fuzzification: converting the quantitative inputs into natural language variables, (2) rule evaluation: implementing the mapping logic, and (3) defuzzification: converting the qualitative rule outcomes into a numerical output. Here, our fuzzy logic-based mapping algorithm will be explained in Section 4.2 AVC Data Recovery.

2.2 ACCIDENT MODELING

Most traffic collision models have been developed on the basis of statistical regression techniques. Different regression models are developed depending on the diverse characteristics of collision data in different situations. As one of the most traditional and basic methods, the Poisson regression model has been widely used for collision count data analysis (e.g., Jovanis and Chang, 1986; Miaou et al., 1992; Miaou and Lum, 1993; Miaou, 1994). However, despite having a significant advantage in accurate modeling capability (Maher and Summersgill, 1996), Poisson models are inadequate for handling over-dispersed data, which have a variance greater than the mean (Maycock and Hall, 1984). A well-recognized issue with the Poisson regression model is that the Poisson regression model should be reserved for use in situations where the sample variance is approximately equal to the sample mean.

Other models, including Poisson-lognormal models (Miaou et al., 2005; Lord and Miranda-Moreno, 2008; Aguero-Valverde and Jovanis, 2008) and negative binomial (NB) regression (or Poisson-gamma) (Miaou, 1994; Shankar et al., 1995; Poch and Mannering, 1996; Maher and Summersgill, 1996; Milton and Mannering, 1998; Chin and Quddus, 2003; Wang et al., 2003; Lord, 2006; El-Basyouny and Sayed, 2006; Donnell and Mason, 2006; Malyshkina and Mannering, 2010; Daniels et al., 2010) have been proposed for over-dispersed collision data.

Besides the overdispersion phenomenon, a collision dataset may occasionally be underdispersed, i.e., the variance is smaller than the mean. Under-dispersion may exist in a dataset with a very low sample variance (Oh et al., 2006). Gamma regression models

(Winkelmann and Zimmermann, 1995; Oh et al., 2006) have been proposed to fit the under-dispersed data. Additionally, the Conway–Maxwell–Poisson (COM-Poisson) distribution was introduced for modeling either over- or under-dispersed count data (Shmueli et al., 2005; Kadane et al., 2006; Lord et al., 2007).

Another issue in collision data is the phenomenon of an apparent excess of zeros. In this situation, zero-inflated models (Poisson and NB) have been suggested for modeling collision data with excessive zeros in the observations (Shankar et al., 1997; Garber and Wu, 2001; Lee and Mannering, 2002; Kumara and Chin, 2003; Miaou and Lord, 2003; Rodriguez et al., 2003; Shankar et al., 2003; Noland and Quddus, 2004; Qin et al., 2004; and Lord et al., 2005). Recently, several innovative accident models, including random parameter models, finite-mixture/Markov switching models, neural networks, Bayesian neural networks, and support vector machines, have been used in accident analysis research. A comprehensive review of the accident models mentioned above can be found in the paper by Lord and Mannering (2010).

Most of the regression models described above share one common characteristic: they are univariate Poisson- (or Gamma-) based models. Recently, multivariate Poisson (MVP) regression models (Miaou and Song, 2005; Ma and Kockelman, 2006; Park and Lord, 2007), multivariate zero-inflated Poisson (MVZIP) regression models (Li et al., 1999), or multivariate Poisson-lognormal regression models (Lord, 2007; Karim and Tarek, 2009) have been proposed for modeling different count data that are correlated. As a special case of MVP regression models, bivariate Poisson (BP) regression models can be used for paired count datasets. However, BP and other MVP regression models cannot handle over- or underdispersed count data. In order to concurrently utilize both AVC and carcass removal data while also providing a better model to describe AVCs, a diagonal inflated bivariate Poisson (DIBP) regression model (Karlis and Ntzoufras, 2005) was applied to AVC modeling in this research.

These regression models have been used to model collisions in many previous studies. However, another issue is that most previous accident modeling studies have failed to consider factors related to human behavior, despite its critical role in crash events (Wang et al., 2003). Furthermore, when these models have been applied to AVC analysis, few studies have considered the effects of animal population distribution along the highway. Therefore, to get a better understanding of AVCs and their associated factors, a microscopic probability (MP) model that can explicitly consider human factors and the spatial distribution of animal population is desirable. When the DIBP regression model and the MP model are used in conjunction, they can provide a more detailed explanation of AVCs from different perspectives.

CHAPTER 3. STUDY ROUTES AND DATA

3.1 Route Selection

Ten highway routes (US 2, SR 8, US 12, SR 20, I-90, US 97, US 101, US 395, SR 525, and SR 970) with varying AVC rates in the past five years were chosen as the study routes following recommendations from WSDOT.

3.2 Data Types

Data from different sources used in this research and are listed as follows.

(1) Collision Report data

Reported collisions between vehicles and non-domestic animals were extracted from the traffic accident records maintained by WSDOT. This dataset was also extracted from the Washington state accident files provided by the Highway Safety Information System (HSIS) (HSIS, 2009). However, since a significant portion of accidents is not reported, this dataset is only a subset of actual animal-vehicle accidents. Collision reports are only required for incidents that cause damage values greater than a particular threshold. The threshold value is high enough that only large animal collisions are likely to be reported.

(2) Carcass Removal Data

WSDOT maintenance employees record the location— by milepost, date, weather, animal type, sex, and age— of every deer and elk carcass removed from state highways (Myers et al., 2007). Given that carcasses may also be removed by un-authorized parties and that some animals leave the right-of-way after a collision, this dataset is also a subset of all animal-vehicle accidents and may complement the CRpt dataset to some extent.

(3) Highway Geographic Information System (GIS) map

This dataset contains locations and curvatures of state highways in the GIS format.

(4) Deer Distribution Data

Data were supplied by the Washington Department of Fish and Wildlife through WSDOT. This data contain GIS-based species distribution data for Mule Deer (Mule Deer Foundation, unpublished Data), Elk (Rocky Mountain Elk Foundation, unpublished Data), and White-tailed Deer (Washington Gap Analysis Project, 1997).

(5) SURVEY DATA

The research team conducted two surveys to collect input from WSDOT maintenance employees. The first survey was used to determine threshold values for the CRpt and CR data. The other survey was used to verify the quality of the data recovery algorithm.

(6) PRIORITY HABITAT AND SPECIES DATABASE

This database contains location data for deer and elk habitats in Washington state. These data were provided by the WDFW.

(7) WEYWILD: A COMPILATION OF WILDLIFE HABITAT INFORMATION FOR THE PACIFIC NORTHWEST

This dataset is a compiled database derived from 20 sources of species habitat information for southwestern Washington.

(8) WILDLIFE HABITAT MATRICES

This tool, derived from the Johnson and O'Neil (2001) assessment of wildlife habitat relationships for Washington and Oregon, provides tabular data on the vegetation types, vegetation structures, important habitat elements, population structures, and historical trends of all terrestrial vertebrates in the state.

Data sources (1) through (5) were mainly used for the analysis, whereas data sources (6) through (8) were used for reference. Table 3-1 shows the years of data covered by each of the five major data types used in this research.

Table 3-1: Data collection information

Data	Data Time Covered	Date Received	Providing Agency
Collision Report Data	2000-2006	Apr. 2008 (Jan. 2009 update)	HSIS
Roadlog Data	2002-2006	Apr. 2008 (Jan. 2009 update)	HSIS
Carcass Removal Data	1999-2007	Jul. 2008	WSDOT
Survey Data		Feb. 2008-Mar. 2009	WSDOT
Deer Distribution Data		Jul. 2009	WSDOT & WDFW

CHAPTER 4. METHODOLOGY

4.1 DATABASE DESIGNS

For this research, most of the original data were received in the Microsoft Excel format. To manage the data efficiently, Microsoft SQL Server 2008 was employed. Before data were imported into the data server, the database had to be designed to ensure storage and operational efficiencies. The Entity/Relationship (E/R) diagram method (Garcia-Molina et al., 2002) was used to design the two AVC study databases, one for the CRpt data and the other for the CR data. The designed E/R diagrams of these two databases were then converted to relational schemas. Structured Query Language (SQL) was used to manipulate the data.

4.1.1 Reported AVC Data

The E/R diagram of the CRpt data is shown in Figure 4-1. Relational schemas of the CRpt database were converted from the E/R diagram design. This database consisted of three tables: *vehicle*, *accident*, and *road*. These relational tables included the actual variables, and each variable was referred to as an attribute. If an attribute or a set of attributes had a unique value for each row in a table, then this attribute or set of attributes was qualified to serve as a key for the table. The primary key for each table is underlined in the E/R diagram (Figure 4-1). A dictionary of all the attributes is provided in the Appendix.

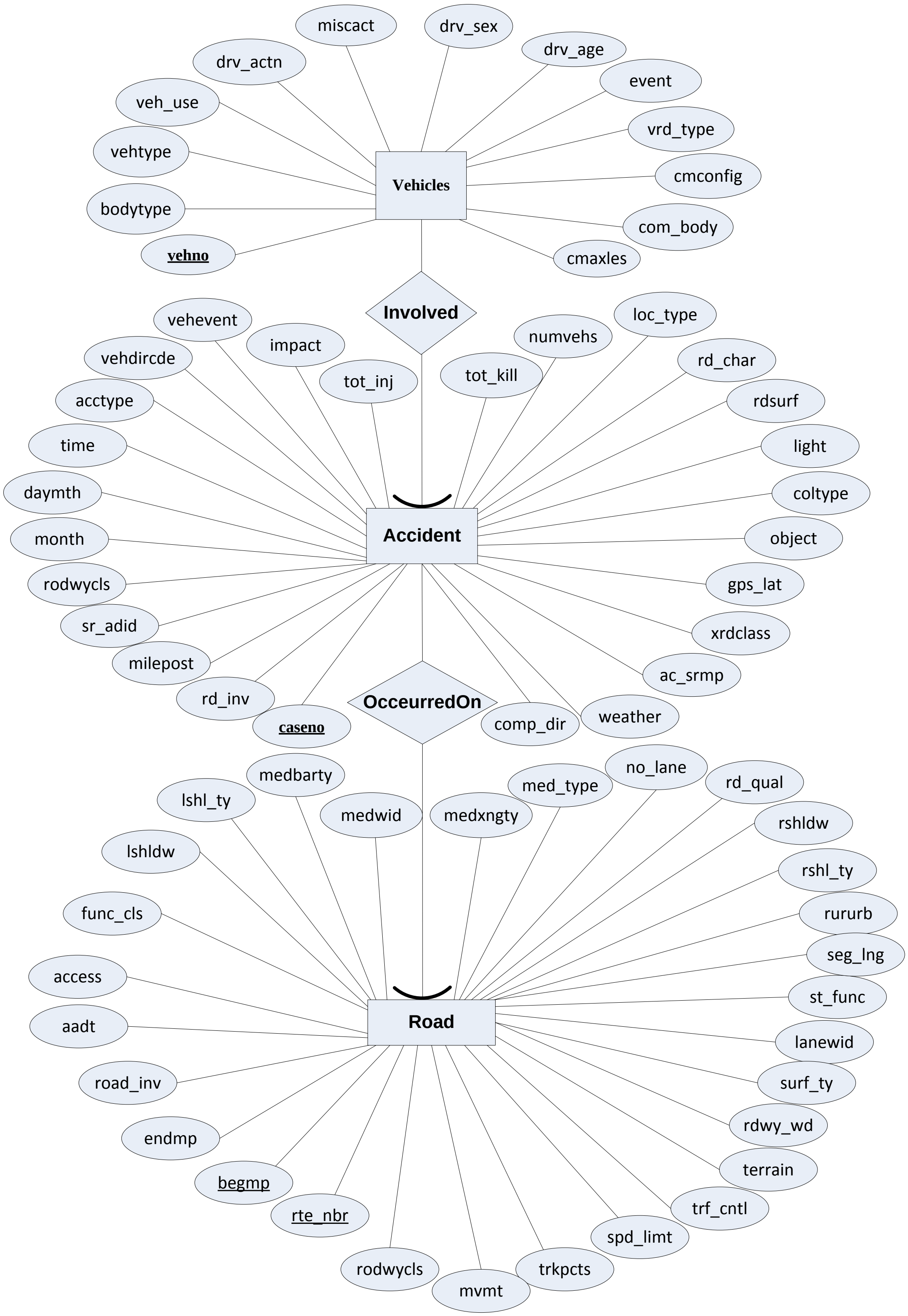


Figure 4-1: The E/R diagram for CRpt data

Each row in the CRpt table, e.g., the *accident* table, corresponded to a collision that occurred on a segment of one of the study routes. Each accident was uniquely identified by its case number (*caseno)*. Thus, *caseno* served as the primary key for the *accident* table. Similarly, the combination of route number (*rte_nbr)* and beginning milepost (*begmp)* was chosen as the primary key for the road table, and vehicle number (*vehno)* was chosen for the vehicle table. The relationship *HappenedOn* linked the *accident* and *road* tables, while *involved* linked the *accident* and *vehicle* tables. The following relational schemas were converted from the E/R diagram, with the primary key attributes underlined:

Vehicle (vehno, bodytype, vehtype, veh_use, drv_actn, miscact, drv_sex, drv_age, event, vrd_type, cmconfig, com_body, cmaxles, caseno)

Accident (caseno, rd_inv, milepost, sr_adid, rodwycls, month, daymth, time, acctype, vehdircde, vehevent, impact, tot_inj, tot_kill, numvehs, loc_type, rd_char, rdsurf, light, coltype, object, gps_latx, gps_laty, gps_latz, xrdclass, ac_srmp, weather, com_dir, rte_nbr, begmp)

Road (rte_nbr, begmp, endmp, road_inv, aadt, access, func_cls, lshldw, lshl_typ, medbarty, medwid, medxngty, med_type, no_lane, rd_qual, rshldw, rshl_typ, rururb, seg_lng, st_func, lanewid, surf_typ, rdwy_wd, terrain, trf_cntl, spd_limt, trkpcts, mvmt, rodwycls)

4.1.2 Carcass Removal Data

The E/R diagram of the CR data is shown in Figure 4-2. The CR database consisted of two tables: *animal* and *road*. The primary key of each table is underlined in the E/R diagram (Figure 4-2). The attributes explanation is provided in the Appendix.

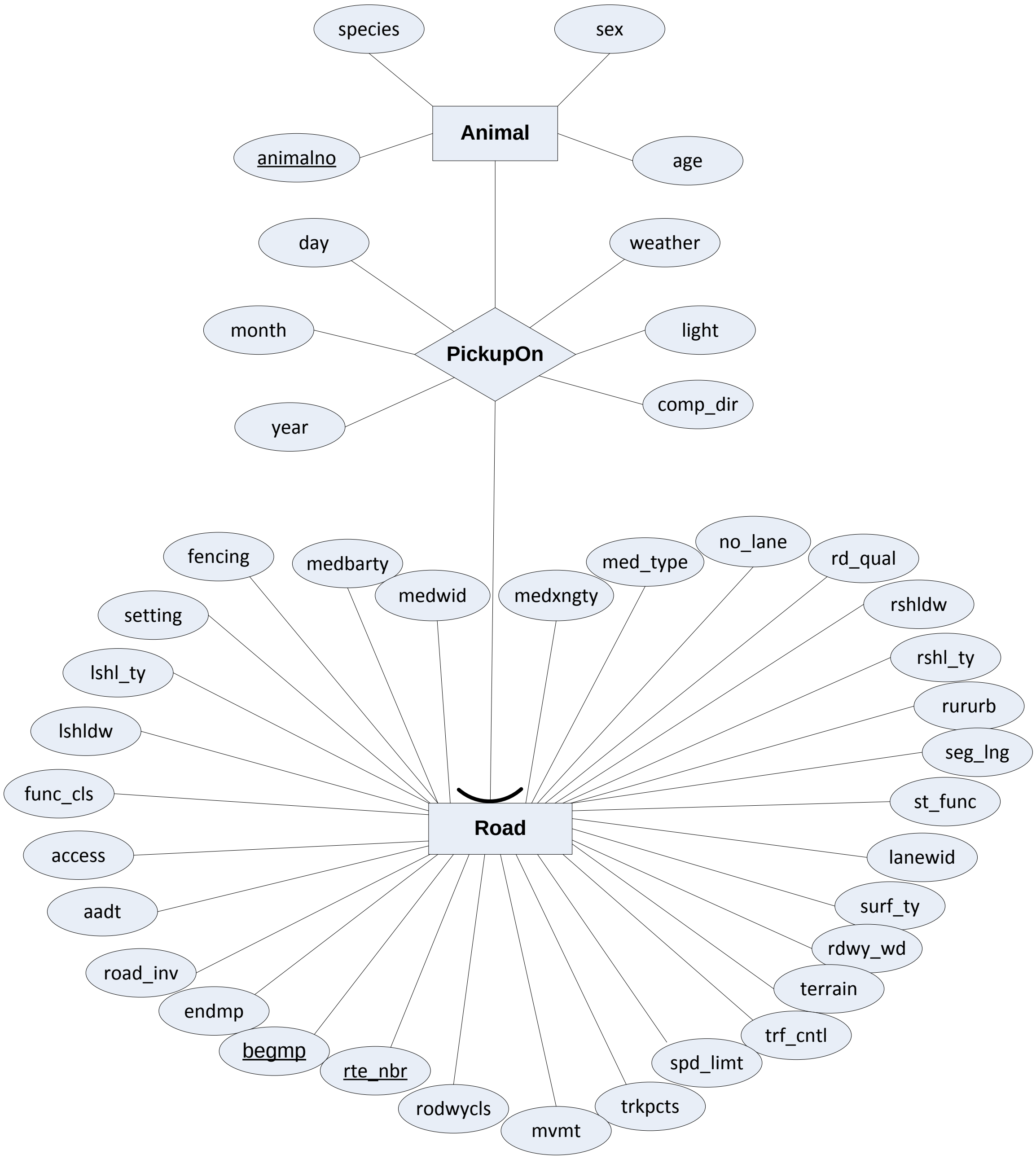


Figure 4-2: The E/R diagram for CR data

Each row in the CR database table corresponded to one animal that was picked up from a segment of one of the studied routes. Similar to the CRpt database, the combination of *rte_nbr* and *begmp* was chosen as the primary key for the *road* table; *AnimalNo* served as the primary key for the *animal* table. The relationship *pickupOn* linked the *animal* and *road* tables. The following relational schemas are produced from the E/R diagram, with the primary key attributes underlined:

Animal (animalno, species, sex, age, day, month, year, weather, light, com_dir, rte_nbr, begmp)

Road (rte_nbr, begmp, endmp, road_inv, aadt, access, func_cls, lshldw, lshl_typ, setting, fencing, medbarty, medwid, medxngty, med_type, no_lane, rd_qual, rshldw, rshl_typ, rururb, seg_lng, st_func, lanewid, surf_typ, rdwy_wd, terrain, trf_cntl, spd_limt, trkpcts, mvmt, rodwycls)

4.2 Fuzzy Logic-based AVC Data Recovery

High quality AVC data were crucial for this project. However, as mentioned earlier, previous studies (Romin and Bissonette, 1996; Knapp et al., 2007; Huijser et al., 2007) found that CRpt data and CR data are significantly different. This suggests that the two sets of data complement each other and could be combined to improve the quality of AVC data.

Whether two datasets match is not a question that can be answered with precise quantitative techniques. Rather, it requires both qualitative and quantitative inferences. Fuzzy logic-based data matching has proved to be an effective way to deal with such problems related to linguistic vagueness and human factors (Zhao, 1997). Generally, the fuzzy logic-based approaches involve three major steps (Chen and Pham, 2001):

- Fuzzification: converting the quantitative inputs into natural language variables
- Rule evaluation: implementing the mapping logic
- Defuzzification: converting the qualitative outcomes into a numerical output.

4.2.1 Fuzzification

Three attributes were used in the data mapping process: animal type, collision time, and collision location. The animal categories for reported AVC data and CR data were a little different. The "non-domestic" animal type reported in the AVC data was matched with the three deer types and elk in the CR data. After the animal types had been matched, this algorithm considered only "date difference" and "location difference" as the inputs.

Date difference referred to the difference between the date when the carcass was collected and the date when the collision was recorded in the CRpt dataset. Note that the date recorded in the CR dataset should have been the same date or later as that in the CRpt database because a carcass cannot be collected until after the collision has happened. Therefore, the date difference was mathematically defined as:

$$\text{Date difference} = \text{Date in the CR dataset} - \text{Date in the CRpt dataset} \qquad (4\text{-}1)$$

Location difference was the milepost difference between the CRpt location and the location where the carcass was collected. The route numbers in a data pair had to be identical before mileposts could be compared. Therefore, the location difference was defined as the absolute difference between the milepost in the CRpt dataset and the milepost in the CR dataset:

$$\text{Location difference} = |\,\text{Milepost in the CRpt dataset} - \text{Milepost in the CR dataset}\,| \qquad (4\text{-}2)$$

These inputs were then translated into four fuzzy classes based on the level of difference: small, medium, big, and very big (S, M, B, and VB). The categories small to big were used to avoid confusion with the output class that uses low to high and would result in overlapping abbreviations such as L and VL and the introduction of ambiguity. VB represented the situation in which the input exceeded the critical range. For example, if the location difference was only considered within 3 miles, a 5-mile difference would be marked as VB.

A membership function for each class needed to be determined during the fuzzification step. A membership function calculates describes the membership degree, defined as the association level of a datum with a predefined category. . The value of membership degree ranges from zero to one. Most research (Taylor et al., 1998; Nikunja, 2006; Naso et al., 2006) has assumed the membership function to be a triangle for simplification and has designed it on the basis of subjective experiences. However, the triangular membership function may be too simple to accurately reflect reality. Therefore, this study adopted a survey-based method (Li and Yen, 1995) to determine the membership functions for the fuzzy classes. Details about the determination process of membership function will be described later.

4.2.2 Rule Design

Fuzzy logic rules are needed for mapping inputs to outcomes. Eleven rules, shown in Table 4-1, were designed for this algorithm. The default rule weights reflected the relative importance of the rules. As mentioned earlier, the two inputs were milepost difference and date difference. The matching degree (*MD*) between the CRpt and the CR datasets was the outcome that was represented by six fuzzy classes: very very low (VVL), very low (VL), low (L), medium (M), high (H), and very high (VH). For example, VVL represented the situation in which *MD* as very close to zero, and the candidate data pair was too different to be a possible matching pair.

The *MD* decreased with an increase in milepost difference or date difference. Rules 1 through 9 covered normal matching conditions. For example, Rule 9 could be interpreted as follows: if the milepost difference is large, and the date difference is large, then their mapping degree is very low. Rules 10 and 11 set *MD* to VVL if either of the inputs was outside the range limits.

4.2.3 Defuzzification

The defuzzification process converts the qualitative rule outcome into a numerical output. The centroid defuzzification method (a.k.a. Center-of-Area, or gravity method) (Runkler, 1996; Taylor and Meldrum, 1998) was used to determine the *MD* in this research:

$$MD = \frac{\sum_{i=1}^{n} w_i c_i I_i}{\sum_{i=1}^{n} w_i I_i} \tag{4-3}$$

where w_i was the rule weight representing the importance of rule i; c_i was the centroid of the output class i; and I_i was the implicated area of the output class i. The centroid of each output class is defined in Table 4-2. Note that if the output classes included VVL, the output MD was set to zero. In this study, a data pair was regarded as a match if $MD \geq 0.5$.

Table 4-1: Rule base for fuzzy mapping algorithm

Rule	Default Rule Weight	Rule Premise: Milepost Difference	Rule Premise: Date Difference	*MD* Classes
1	1	S	S	VH
2	1	S	M	H
3	1	S	B	M
4	1	M	S	H
5	1	M	M	M
6	1	M	B	L
7	1	B	S	M
8	1	B	M	L
9	1	B	B	VL
10	1	VB	–	VVL
11	1	–	VB	VVL

Table 4-2: Centroid value for output classes

	VH	H	M	L	VL	VVL
c_i	1	0.8	0.6	0.4	0.2	0

4.3 Statistical Analysis

Descriptive statistics for AVCs on the study routes during a five-year period from 2002 to 2006 are summarized in Chapter 6. Non-parametric analyses and *t*-tests were conducted to identify factors that have significant impacts on AVC risks. The AVC distributions were analyzed in four groups—spatial and temporal characteristics, roadway geometric characteristics, traffic characteristics, and animal distribution characteristics—and are described in Section 6.2.

Statistical hypothesis testing is used to decide whether a difference in a population parameter, e.g., mean, variance, or proportion, between two or more groups is significant (Washington et al., 2003). In this study, the t-test was used to compare the means of two groups and ANOVA (or F-test) was used to compare the means of more than two groups.

4.4 RISK MODELING

4.4.1 Microscopic Probability Model

In this study, a microscopic probability (MP) modeling approach developed by Wang (1998) was applied. An important advantage of this approach is its capability to consider the mechanism of accident occurrence in risk modeling. This approach has been successfully applied in many studies of accident risks (see for example Siddique, 2000; Wang et al., 2003; Kim et al., 2007) and has achieved favorable results.

A representative AVC process is difficult to accurately model and interpret because of the multiple subjective and objective factors involved and limited data supports. However, a simplified process can be formulated by considering two significant AVC contributors: failed responses from drivers, such as insufficient deceleration, or swerving, and failed responses from animals, such as freezing or running in the wrong direction. These two factors interact with each other so that an AVC may be caused by either one or both. An AVC is avoidable if a driver applies early and strong deceleration or an animal has an instant and powerful reaction. Additionally, when vehicles travel along roadways, animals' crossing or following presence could be another important component for modeling an AVC process. Therefore, the occurrence of an AVC is conditioned on the presence of an animal in the roadway, the ineffective response of the arriving driver, and the animal's failure to escape. The MP AVC risk model included three components, following the approach of Wang (1998): the probability of a hazardous crossing of an animal (P_o), the probability of the animal failing to escape (P_{af}), and the probability of an ineffective response by the driver (P_{vf}). The probability for a randomly selected vehicle to have an AVC on a certain roadway section would then be the product of P_o, P_{af}, and P_{vf}:

$$P_{AVC} = P_o \cdot P_{af} \cdot P_{vf} \tag{4-4}$$

Since P_o, P_{af} and P_{vf} are not directly observable, we needed to estimate P_o, P_{af}, and P_{vf}, respectively.

4.4.1.1 P_{af} and P_{vf} Formulations

It is assumed that a driver cannot avoid a collision if his/her necessary perception reaction time (NPRT) is longer than the available perception reaction time (APRT). The APRT refers to the time available to a driver to complete his/her perception and response under a given condition. The NPRT is the ability-oriented minimum required perception/reaction time. The NPRT typically varies from person to person. Both the APRT and the NPRT are random variables and are assumed to follow normal distributions. Since a normal distribution does not have a closed form for cumulative probability calculation, the Weibull distribution was used instead. The NPRT was assumed to follow the Weibull (α, λ) distribution, and the APRT was assumed to follow the Weibull (α, γ) distribution. Therefore, λ and γ were the scale parameters for the NPRT and APRT distributions, respectively. The Weibull distribution shape parameter α was chosen to be 3.25 in this study because it has been empirically verified that when α=3.25, the Weibull distribution is a very good approximation to normal distribution (Kao, 1960; Plait, 1962). By using the assumed distributions for the APRT and the NPRT, P_{vf} could be calculated as:

$$P_{vf} = \int_0^\infty \int_{t_{av}}^\infty f(\lambda, t) f(\gamma, t_{av}) dt dt_{av} = \int_0^\infty e^{-\lambda t_{av}^\alpha} \alpha \gamma t_{av}^{\alpha-1} e^{-\gamma t_{av}^\alpha} dt_{av} = \frac{1}{1+\lambda/\gamma} \tag{4-5}$$

where t_{av} was the variable used to represent the APRT. Equation (4-5) shows that P_{vf} is only dependent on λ/γ and has no relationship to α. Since the parameters λ and γ are positive variables, λ/γ can be related to various factors by using an exponential link function, as shown in Equation (4-6). Correspondingly, P_{vf} can be written in the form of Equation (4-7).

$$\lambda/\gamma = e^{-\boldsymbol{\beta}_{\mathbf{vh}}\mathbf{x}_{\mathbf{vh}}} \tag{4-6}$$

$$P_{vf} = \frac{1}{1+e^{-\beta_{vh}x_{vh}}} \tag{4-7}$$

In Equations (4-6) and (4-7), β_{vh} and x_{vh} are vectors of unknown parameters and explanatory variables, respectively, related to P_{vf}. Variables affecting drivers' task load and action complexity need to be included in x_{vh}.

Similarly, P_{af} can be written in the form of Equation (4-8).

$$P_{af} = \frac{1}{1+e^{-\beta_{ah}x_{ah}}} \tag{4-8}$$

In Equation (4-8), β_{ah} and x_{ah} are vectors of unknown parameters and explanatory variables, respectively, related to P_{af}. Variables affecting animals' actions need to be included in x_{ah}.

4.4.1.2 P_o Formulation

An animal becomes an obstacle for vehicles if its highway-crossing movement interrupts the smooth movement of vehicles. Only when a highway-crossing movement occurs within a certain period may the animal become an obstacle to the arriving vehicle. This period is called "effective time." As the arrival of an obstacle is discrete, nonnegative, and random, it is assumed to have a Poisson arrival process. In such a process, intervals between arrivals are independent and follow the same exponential distribution (Pitman, 1993). Let us consider a disturbance, j, whose arrival rate is η_{dj} and effective time is t. Then, its density function is:

$$f(t) = \eta_{dj}e^{-\eta_{dj}t} \qquad \text{for t>0} \tag{4-9}$$

According to the memoryless property of the exponential distribution (Pitman, 1993), the probability of having disturbance j within t_{dj} is independent of the time waited. Therefore, the probability for an arriving vehicle encountering disturbance j within t_{dj} can be calculated by Equation (4-10).

$$P_{dj} = \int_0^{t_{dj}} \eta_{dj} e^{-\eta_{dj} t} dt = 1 - e^{-\eta_{dj} t_{dj}} \tag{4-10}$$

The risk of an AVC is the sum of the risk of individual vehicles encountering obstacle animals. Since any of the disturbances occurring in t_{dj} may result in an AVC, the per vehicle probability of encountering an obstacle animal, P_o, is equivalent to the probability that at least one disturbance occurs within the effective period. Therefore, P_o can be formulated as:

$$P_o = 1 - \prod_{j=1}^{J} (1 - P_{dj}) \tag{4-11}$$

Replacing P_{dj} with Equation (4-10), a P_o becomes:

$$P_o = 1 - e^{-\sum_j \eta_{dj} t_{dj}} \tag{4-12}$$

In Equation (4-12), $\sum_j \eta_{dj} t_{dj}$ should always be positive and dependent on a set of variables. Thus, an exponential link function can be employed to reflect the effects of the explanatory factors, as shown in Equation (4-13).

$$\sum_j \eta_{dj} t_{dj} = e^{\boldsymbol{\beta}_o \mathbf{x}_o} \tag{4-13}$$

P_o then becomes:

$$P_o = 1 - e^{-e^{\boldsymbol{\beta}_o \mathbf{x}_o}} \tag{4-14}$$

In Equations (4-13) and (4-14), $\boldsymbol{\beta_o}$ and $\boldsymbol{x_o}$ are vectors of unknown parameters and explanatory variables of disturbance frequency, respectively. $\boldsymbol{\beta_o}$ does not change with location, while $\boldsymbol{x_o}$ does. Animal habitat integrity, size, and animal population are very likely contributing variables to $\boldsymbol{x_o}$.

4.4.1.3 Integrated AVC Risk Model

Substituting Equations (4-7), (4-8), and (4-14) into Equation (4-4), and adding subscripts denoting roadway section (*i*) to the variables, yields an integrated AVC risk model formulation, as shown in Equation (4-15). The model contains not only road environment related factors, but also factors related to human and animal behaviors. The inclusion of human and animal factors is one of the major distinctions of the proposed approach in comparison with most existing accident risk models. Obviously the human and animal factors are very macroscopic, as details of individual collisions are not generally available.

$$P_{AVCi} = P_{oi} P_{afi} P_{vfi} = \frac{1 - e^{-e^{\beta_o x_{oi}}}}{(1 + e^{-\beta_{af} x_{afi}})(1 + e^{-\beta_f x_{fi}})} \tag{4-15}$$

4.4.1.4 AVC Risk Model Estimation

To simplify the problem, it is assumed that individual vehicles within a traffic flow have a consistent AVC risk, P_{AVCi}. Thus, the number of AVCs occurring within this flow follows binomial distribution:

$$P(n_i) = \binom{f_i}{n_i} P_{AVCi}^{n_i} (1 - P_{AVCi})^{f_i - n_i} \tag{4-16}$$

where f_i is annual traffic volume that can be calculated from the annual average daily traffic (AADT) of roadway section *I* and n_i is the number of AVCs occurring within f_i.

Since it is very rare for an AVC to occur, P_{AVCi} should be very small while traffic volume f_i is very large for the given span of time. Thus, the Poisson distribution is a good approximation to the binomial distribution (Pitman, 1993):

$$P(n_i) = \frac{m_i^{\,n_i} \cdot e^{-m_i}}{n_i!} \tag{4-17}$$

with the Poisson distribution parameter:

$$m_i = E(n_i) = f_i \cdot P_{AVCi} \tag{4-18}$$

The Poisson distribution model has commonly been used to predict the number of accidents. It is usually the first choice when modeling traffic accidents because of the nonnegative, discrete, and random features of accidents. Since Poisson distributions have only one distribution parameter, the mean and variance need to be the same. However, in most of the cases, accident data are over-dispersed, and the applicability of the Poisson model is therefore limited. An easy way to overcome this difficulty is to add an independently distributed error term, ε_i, to the log transformation of Equation (4-18). That is:

$$\ln m_i = \ln(f_i P_{AVCi}) + \varepsilon_i \tag{4-19}$$

We assume that $\exp(\varepsilon_i)$ is a Gamma distributed variable with mean 1 and variance δ. Substituting Equation (4-19) into Equation (4-17) yields

$$P(n_i \mid \varepsilon_i) = \frac{e^{(-f_i P_{AVCi} \exp(\varepsilon_i))} \cdot (f_i P_{AVCi} \exp(\varepsilon_i))^{n_i}}{n_i!} \tag{4-20}$$

Integrating ε_i out of Equation (4-20), we can directly derive a negative binomial distribution model as the following:

$$P(n_i) = \frac{\Gamma(n_i + \theta)}{\Gamma(n_i + 1)\Gamma(\theta)} \left(\frac{\theta}{f_i \cdot P_{AVCi} + \theta}\right)^{\theta} \left(\frac{f_i P_{AVCi}}{f_i \cdot P_{AVCi} + \theta}\right)^{n_i} \tag{4-21}$$

where $\theta = 1/\delta$. The expectation of this negative binomial distribution equals the expectation of the Poisson distribution shown in Equation (4-18). The variance is now:

$$V(n_{ik}) = E(n_{ik})[1 + \delta E(n_{ik})] \tag{4-22}$$

Since δ can be larger than 0, the constraint that the mean be equal to the variance in the Poisson model is removed. Thus, the negative binomial distribution can deal with the over-dispersed data.

4.4.2 DIBP Regression Model

4.4.2.1 Bivariate Poisson Regression Model

Figure 4-3 shows the relationships between two types of data, the reported AVC data (left circle) and the CR data (right circle) typically collected by transportation agencies. There are three regions of interest: Z_1, Z_2, and Z_3. Z_1 represents the CRpt with no corresponding CR data. Z_2 represents the CR data with no counterparts in the CRpt data. The records contained in both the CRpt data and the CR data are represented by Z_3. This area is the overlapping portion of the two datasets.

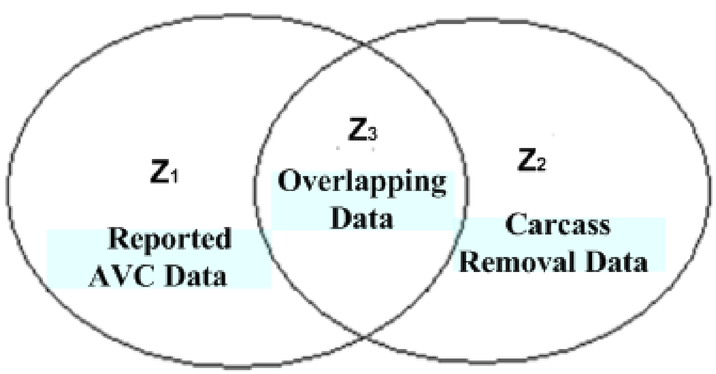


Figure 4-3: Relationship between the reported AVC and CR datasets

Let us assume that the count datasets Z_1, Z_2, and Z_3 follow independent Poisson distributions with parameters (means) λ_1, λ_2, and λ_3, respectively. Then the reported CRpt dataset $X = Z_1+Z_3$ and the carcass removal dataset $Y = Z_2+Z_3$ follow a bivariate Poisson distribution, BP (λ_1, λ_2, λ_3), with a joint probability mass function defined as follows (Karlis and Ntzoufras, 2005):

$$f_{BP}(x, y \mid \lambda_1, \lambda_2, \lambda_3) = e^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^x}{x!} \frac{\lambda_2^y}{y!} \sum_{i=0}^{\min(x,y)} \binom{x}{i} \binom{y}{i} i! \left(\frac{\lambda_3}{\lambda_1 \lambda_2}\right) \tag{4-23}$$

The BP distribution is appropriate for modeling two random variables with positive dependence, which is the case for the CRpt and CR datasets. Its marginal distributions of X and Y follow Poisson distributions with $E(X) = \lambda_1+\lambda_3$ and $E(Y) = \lambda_2+\lambda_3$, respectively. Moreover, $COV(X, Y) = \lambda_3$, and hence λ_3, is a measure of dependence between the CRpt dataset and the CR dataset.

In the BP model, λ_1, λ_2, and λ_3 can each be related to various explanatory variables by using exponential link functions. Therefore, the BP regression model can take the following form:

$$\begin{aligned} &(X_i, Y_i) \sim BP(\lambda_{1i}, \lambda_{2i}, \lambda_{3i}), \\ &\ln(\lambda_{ij}) = \boldsymbol{\omega}_{ji}^T \boldsymbol{\beta}_j \end{aligned} \tag{4-24}$$

where $i = 1, \ldots, n$, is the roadway segment number, $\boldsymbol{\omega}_{ji}$ is the vector of explanatory variables for roadway segment i, $\boldsymbol{\beta}_j$ is the corresponding coefficient vector for Z_j, and j = 1, 2, and 3. In this study, the roadway segments were separated by consistent geometric factors.

4.4.2.2 Diagonal Inflated BP (DIBP) Regression Model

A major disadvantage of the BP model is that its marginal distributions cannot handle over-dispersed or under-dispersed data because its marginal distributions are Poisson distributions that require the mean and the variance to be equal (Karlis and Ntzoufras, 2005). The DIBP model proposed by Karlis and Ntzoufras (2005) can be used to fix this problem. This model uses a more general form developed on the basis of zero-inflated models, and the probabilities of the diagonal elements are inflated in the probability table. Note that the BP model and the ZIDP model are special cases of the diagonal inflated model (Karlis and Ntzoufras, 2005). The DIBP model can be defined on the basis of the BP regression model as follows:

$$f_{IBP}(x,y)=\begin{cases}(1-p)\,f_{BP}(x,y\,|\,\lambda_1,\lambda_2,\lambda_3), & x\neq y\\ (1-p)\,f_{BP}(x,y\,|\,\lambda_1,\lambda_2,\lambda_3)+pf_D(x\,|\,\theta,J), & x=y\end{cases} \tag{4-25}$$

where $f_D(x\,|\,\theta,J)$ is the probability mass function of a discrete distribution $D(x;\,\theta)$. $D(x;\,\theta)$ can be a Poisson, geometric, or a simple discrete distribution. That is, the data process has a probability of 1-p to follow a BP distribution. Note that when p=0, the DIBP model is simply the BP model. $f_D(x\,|\,\theta,J)$ can be defined as

$$f_D(x\,|\,\theta,J)=\begin{cases}\theta_x & for \quad x=0,1,\ldots,J\\ 0 & for \quad x\neq 0,1,\ldots,J\end{cases} \tag{4-26}$$

where $\sum_{x=0}^{J}\theta_x=1$. If J = 0, then the model in Equation (4-25) become a zero-inflated model. The marginal distributions of a DIBP regression model are mixtures of distributions with one Poisson component. For example, the marginal distribution of X is

$$f_{IBP}(x)=(1-p)\,f_{P_0}(x\,|\,\lambda_1+\lambda_3)+pf_D(x\,|\,\theta) \tag{4-27}$$

where $f_{Po}(x\,|\,\lambda)$ is the Poisson probability mass function with parameter $\lambda_1+\lambda_3$. The marginal distributions of the DIBP regression model can model either under-dispersed or over-dispersed count data, depending on the definition of $D(x;\,\theta)$. For example, if J=1, $\lambda_1+\lambda_3$=1 and p=0.5, then the resulting distribution is under-dispersed. If J=0 (the simplest case of zero-inflated models), then the resulting distribution is over-dispersed. This implies that the DIBP regression model is more flexible than the BP regression model and hence a clearly better choice for modeling the AVC data in this study.

4.4.2.3 Model Estimation Using the EM Algorithm

The parameters in most MVP or related models are difficult to estimate because of the computational issues involved in their applications (Karlis and Ntzoufras, 2005; Ma and Kockelman, 2006). However, recent developments in statistical software models and computer hardware have provided several ways to estimate BP models. In this study, the

Expectation-Maximization (EM) approach (Dempster and Rubin 1977) was investigated for estimating the parameters in the DIBP regression model.

EM is an algorithm for estimating the maximum likelihood (ML) values of model parameters when data that contain missing values are used. The basic idea of the EM method is to alternately perform an expectation (E) step and a maximization (M) step until the best set of parameters is obtained. An E step calculates the expectation of the likelihood by including the latent variables, while an M step calculates the maximum likelihood estimates of the parameters by maximizing the expected likelihood derived from the E step (Borman, 2009). Details of the EM algorithm can be found in (Karlis, 2003; Karlis and Ntzoufras, 2005).

4.4.3 Goodness of Fit Measures

To evaluate the explanatory and predictive power of a model, three measures of goodness of fit were adopted here for model comparison: ρ^2 (Ben-Akiva and Lerman, 1985), Akaike's information criterion (AIC) (Akaike, 1974) and Bayesian information criterion (BIC) (Schwarz, 1978; Liddle, 2007). These three measures are described as follows.

ρ^2 (rho-squared) is the log-likelihood ratio index and is used to evaluate a model’s goodness of fit for random, discrete, and sporadic count data (Ben-Akiva and Lerman, 1985; Chin and Quddus, 2003). The index is formulated as

$$\rho^2 = 1 - \frac{\ln L(\hat{\boldsymbol{\beta}})}{\ln L(0)} \tag{4-27}$$

where $L(\hat{\boldsymbol{\beta}})$ is the maximum likelihood estimation of the compared model and $L(0)$ is the initial maximum likelihood estimation of the same model with only the constant term.

AIC is another measure of goodness of fit for a statistical model (Akaike, 1974). AIC is often used for model selection. The model with the lowest AIC is considered the best model. In general, AIC is formulated as follows:

$$AIC = 2k - 2\ln(L) \tag{4-28}$$

where k is the number of parameters in the model and L is the model's maximum likelihood estimation.

BIC (Schwarz, 1978) is also a criterion used for model selection among a group of models with different numbers of parameters. In comparison to AIC, BIC has a stronger penalty for additional parameters. Similarly, the model with the lowest BIC is considered the best model. BIC is calculated as follows:

$$BIC = -2\ln(L) + k\ln N \tag{4-29}$$

where N is the number of observations in the data and L represents the model's maximum likelihood estimation.

These three measures are consistent in most cases. They can be used individually or jointly in model selections. In this study, AIC was used to quantify goodness of fit for the MP model. and all the three measures were used for the DIBP model.

CHAPTER 5. AVC Data Recovery

As mentioned in Research Background, high-quality data are essential for meaningful data analysis and risk modeling. This chapter focuses on the implementation of a fuzzy logic-based mapping algorithm for recovering the AVC data from the CRpt and CR datasets.

Figure 5-1 shows the total numbers of records in each dataset over the five-year period for each study route. Apparently, the CRpt and CR datasets are different. The number of CR records is typically more than the numbers of CRpt data on each route except for US 101. Hence, using the CRpt data or the CR data alone is likely to underestimate the frequency of AVCs.

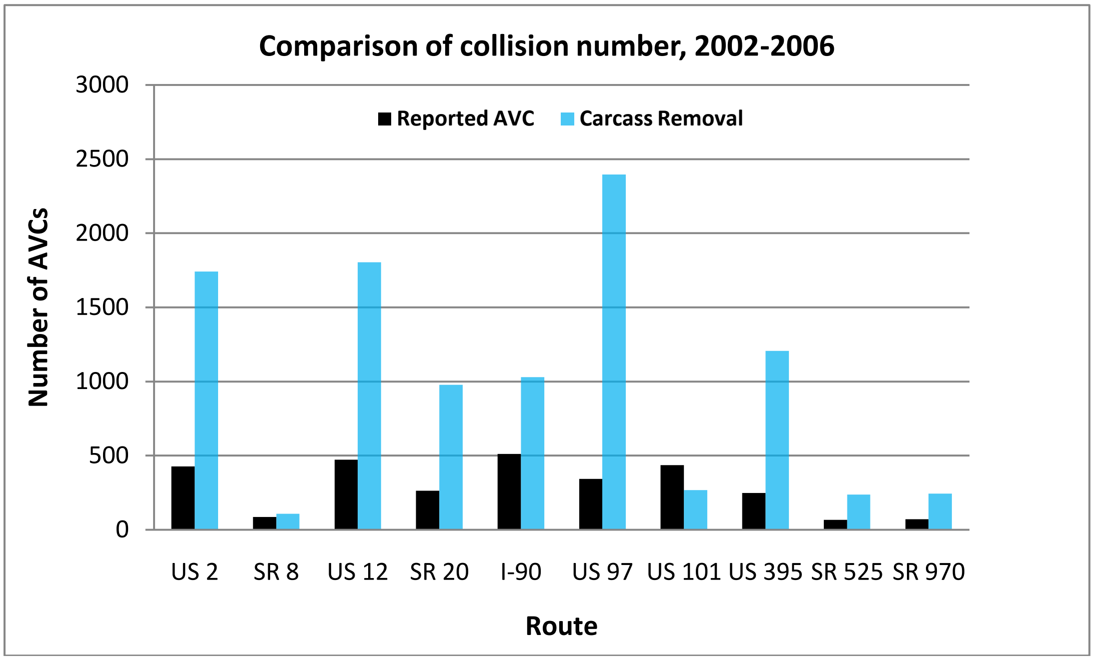


Figure 5-1: Comparison of the total number of AVCs between the CRpt and the CR datasets for each study route from 2002 to 2006

To improve overall data quality, combining both datasets is required. The fuzzy logic-based mapping algorithm introduced in Section 4.2 was used for data recovery. “Recovery” refers to the procedure to combine the two datasets to form a more complete dataset for AVC research. In the following sections, the determination of membership functions in fuzzy logic is first introduced. This is followed by the mapping results. Finally, verification of the algorithm by an evaluation survey is described.

5.1 Determination of Membership Function

The CR and CRpt datasets differ significantly and have different sources, so it is difficult to find people familiar with both datasets. Because the CRpt data are more precise in location and date, as well as more physically and directly tied to incident location, the CRpt data were chosen as a baseline for comparison to the application of fuzzy logic to the CR data. Therefore, the decision was made to survey only people with expertise in the CR data. Before the fuzzy logic-based mapping algorithm was applied, the membership functions had to be determined. To determine some threshold values for the membership functions, a survey was conducted of WSDOT employees who had collected CR data for more than three years to obtain information about how experts made their judgments about the corresponding variables. The survey was conducted from February 5 to March 3, 2009. The survey contained eight questions, including four questions directly related to the determination of the fuzzy membership function. Questions included, "What is the average location difference between the carcass and the actual collision?" and "What is the largest location difference between the carcass and the actual collision?" Similar questions about the date difference were also included.

Forty-eight of the 54 responses received were valid. Incomplete questionnaires were considered invalid. From each expert's input, we were able to understand how he or she judged the date and location differences and the corresponding threshold values he or she used. Figure 5-2 illustrates the fuzzification process of an expert. For example, if a location difference was smaller than the expert's expected location difference, then the current data pair's location difference was small, in this expert's opinion. The location difference of this same data pair may have been considered as big in another expert's view. The measured differences in experts' judgment offered a solid foundation for building the membership functions.

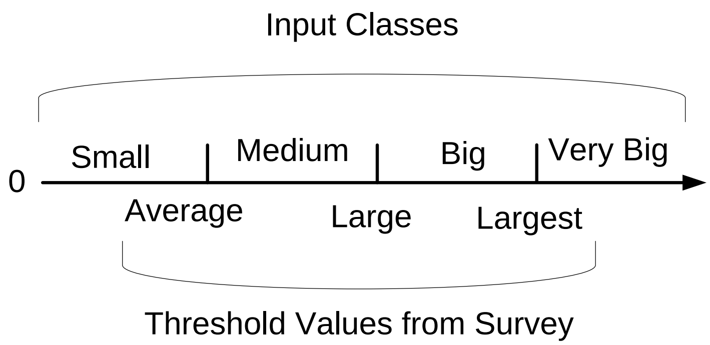


Figure 5-2: Determination of fuzzy classes

The degree of membership of input value u in fuzzy class A_i (i=1, 2, 3, 4, representing the classes of S, M, B, VB, respectively) can be calculated by using the membership function for class A_i. In this study, u was the value of date difference or location difference. The membership function was constructed as shown in Equation (5-1) by using the survey inputs from the WSDOT experts.

$$f_i(u) = \frac{n_{i,u}}{K} \tag{5-1}$$

where $n_{i,u}$ is the number of observations of $u \in A_i$ for class i and K is the total number of observations (valid survey responses) for all classes (K = 48 in our study).

The results for the constructed membership functions are shown in figures 5-3 to 5-5. Figure 5-3 shows the membership function for location difference between the AVC and CR datasets. For example, approximately 56 percent of the employees regarded 1 mile as a Big difference, while 38 percent thought that it was a Medium difference, and about 6 percent regarded it as a Small difference.

Figures 5-4 and 5-5 show the membership functions for date differences on weekdays and weekends, respectively. When an AVC happens during a weekend, the carcass is often collected on the following Monday or Tuesday. The date differences were slightly larger on weekends.

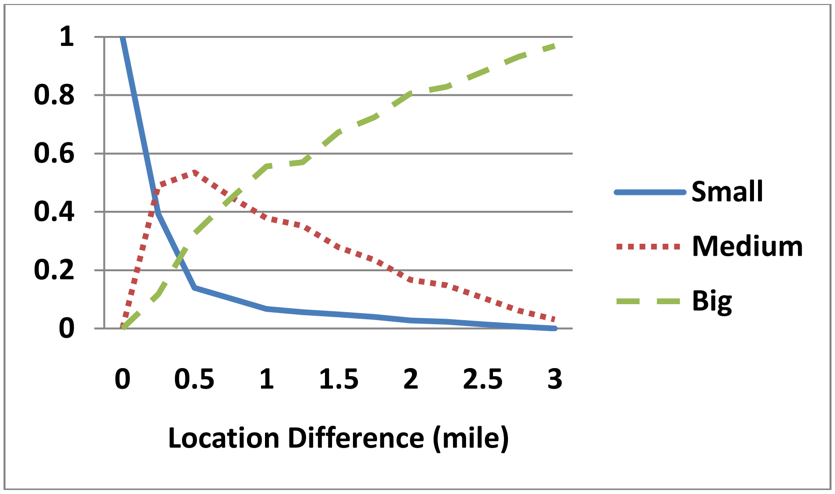


Figure 5-3: Membership function for location difference

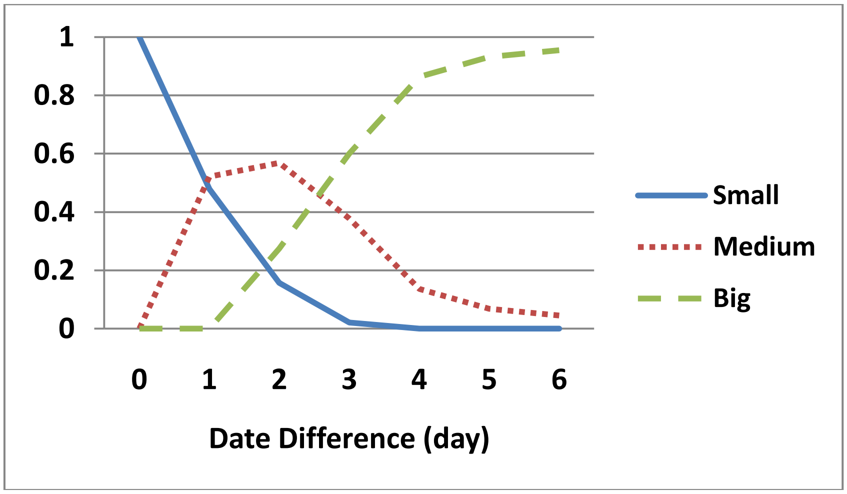


Figure 5-4: Membership function for time difference on weekdays

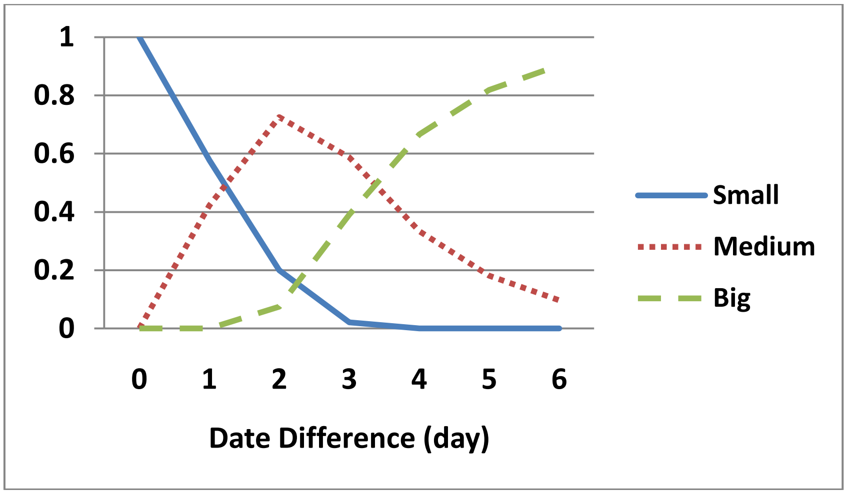


Figure 5-5: Membership function for time difference on weekends

5.2 MAPPING RESULTS

The fuzzy logic-based mapping algorithm described in Section 4.2 was immediately applicable to combining the two datasets with the calibrated membership functions. Five years of reported AVC and CR data for the ten study routes were combined by using this approach. However, to merge the two datasets, we needed to identify their intersections so that the same accidents would not be recorded twice in the combined dataset.

As shown in Table 5-1, the fuzzy logic-based mapping algorithm identified matches at between 27 percent and 37 percent for each year. Thus the union of the two datasets expanded the data breadth. In comparison to the original CRpt dataset, the union dataset contained about 310 percent to 420 percent more records, as shown in the Improved Percentage column.

Table 5-1: Data mapping results for the study routes in five years (2002~2006)

Year	Total Number of Records		Matched	Matching	Union	Improved
	CRpt data	CR data	Data Pairs	Percentage	Datasets	Percentage
2002	529	1989	158	29.9%	2360	346%
2003	508	1935	163	32.1%	2280	349%
2004	529	1800	145	27.4%	2184	313%
2005	544	2484	198	36.4%	2830	420%
2006	533	2112	162	30.4%	2483	366%

5.3 ALGORITHM VERIFICATION

After the proposed algorithm had been implemented, a major step was to verify whether the algorithm was able to reasonably imitate the experts' decision process and produce a quality combined dataset. However, because no ground-truth AVC data were available, it was nearly impossible to validate the performance of the algorithm by using the existing datasets. Therefore, an evaluation survey was conducted from March 5 to March 23, 2009. Again, the survey subjects were WSDOT employees who had collected CR data for more than three years. Each survey subject was asked to judge whether the data pairs listed on the questionnaire matched. The disparity between the experts' results and the algorithm results was a measure of the credibility of the proposed algorithm.

Thirteen data pairs were extracted from the CRpt and CR datasets and included in the survey questionnaire. These pairs were considered representative of both the day and location differences between the two datasets. As shown in Table 5-2, information about Route, Milepost, Weekday, Month, and Day from the data pairs was also provided on the survey questionnaire. Experienced WSDOT highway maintenance staff were invited to fill in the questionnaire. The matching degree for each of the 13 data pairs was computed on the basis of the expert's input and then compared to the fuzzy logic-based mapping algorithm's output. The last three columns of Table 5-2 show the matching degrees for both the survey results and the fuzzy logic-based mapping algorithm, as well as the percentage of error between matching results. In the Matching Degree column, the gray cells indicate that the data pair should refer to the same collision (the matching degree of a data pair should be 50 percent or higher to be marked as a match).

Table 5-2: Survey and algorithm matching percentages for different data pairs

No	**Route**	**CRpt Data**				**Carcass Removal Data**				**Matching Degree (%)**		e_i*
		Milepost	Weekday	Month	Day	Milepost	Weekday	Month	Day	Survey	Algorithm	
1	2	302.1	Thu	Oct.	20	302	Thu	Oct.	20	100	96	0.04
2	2	327.2	Wed	May	25	325	Mon	Jun.	20	8	25	0.17
3	12	118.14	Mon	Feb.	14	118	Tue	Feb.	15	88	86	0.02
4	20	24.77	Wed	Oct.	26	24.1	Wed	Oct.	26	58	74	0.16
5	20	8.1	Thu	Nov.	10	5.5	Fri	Nov.	18	0	24	0.24
6	90	257.27	Sun	Sep.	25	257	Thu	Sep.	29	69	51	0.18
7	90	55.2	Sun	Jul.	31	56	Mon	Aug.	1	88	64	0.24
8	90	32.88	Thu	Mar.	31	34	Sat	Apr.	2	50	52	0.02
9	97	25.5	Wed	Jul.	20	24	Mon	Jul.	25	46	31	0.15
10	97	299.02	Sun	Sep.	10	299.7	Mon	Oct.	3	35	35	0
11	195	84.53	Mon	Nov.	14	83	Thu	Nov.	17	54	40	0.14
12	395	231.44	Fri	Apr.	29	233.8	Thu	May	12	12	24	0.12
13	970	2.21	Tue	Nov.	22	2	Wed	Nov.	23	96	82	0.14

* e_i is the absolute percentage of error between the matching results

The table shows that the survey and algorithm results agreed in all cases except data pair No. 11, which the experts concluded was a match but the algorithm rejected. If the survey results are assumed accurate, then the accuracy rate (AR) for the proposed algorithm is as follows:

$$AR = \frac{N_{accurate}}{N_{total}} = 12/13 = 92.3\% \quad (5\text{-}2)$$

where $N_{accurate}$ is the number of data pairs correctly matched by the algorithm, and N_{total} is the total number of the data pairs evaluated. The matching rate of 92.3 percent is considered to be a very encouraging result, given the complexity of this issue.

The Mean Absolute Error (MAE), a quantity used to measure how close forecasts or predictions are to the eventual outcomes (Morris, 1986), was used as the error indicator. The MAE of the proposed algorithm can be calculated by using Equation (5-3):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|(f_i - y_i)\right| = \frac{1}{n}\sum_{i=1}^{n}\left|e_i\right| = 12\% \quad (5\text{-}3)$$

where f_i is the result estimated by the fuzzy logic-based data mapping algorithm; y_i is the ground truth matching degree values calculated from the survey result; and e_i is the MAE between the algorithm result and the survey result. The calculated error for each surveyed data pair is listed in the last column of Table 5-2.

CHAPTER 6. DATA ANALYSIS

Descriptive analysis and hypothesis testing were applied to identify factors that may have significant impacts on the AVC risk. Descriptive analysis is separated into four categories: spatial and temporal factor analysis, roadway geometry analysis, traffic characteristic analysis, and deer distribution analysis. The combined data from Chapter 5 are used in the analysis of the first three categories whereas CR data and animal habitat data are applied in the deer distribution analysis since the CRpt data does not provide detailed animal type information. In the statistical analysis testing, a t-test is used to compare the means of two groups.

6.1 Descriptive Analysis

6.2.1 Spatial and Temporal Factor Analysis

The spatial and temporal distributions of AVCs were analyzed by counting the number of AVCs that occurred on each of the ten study routes during different periods (year, month, and weekday).

Figure 6-1 shows the number of AVCs that occurred from 2002 to 2006 by route. Figure 6-2 presents the number of AVCs per mile by route for the same period. From these two figures, one can see that US 97, followed by US 12 and US 2, had the highest number of AVCs, while SR 970 had the highest AVC rate (number of AVCs per mile) during the same five years.

Figures 6-3 through 6-5 show the total number of AVC records in relation to time (year, month, and weekday). Figure 6-3 shows that the number of collected AVC records for the different study routes did not follow a predictable pattern. Interestingly, the numbers of records for US 2, US 97, and I-90 followed a climbing trend and reached their maximums in 2005 but dropped in 2006. The reasons for this need further study.

In Figure 6-4, October and November have more AVC records than other months. This could be because of treacherous driving conditions and higher deer activities in autumn. Even though roadway condition during winter can be worse than in autumn, drivers tend to be more vigilant, and animal activity level is usually lower. Given these factors, autumn could be the most AVC-prone season.

Figure 6-5 shows the number of AVC records by weekday. The results show that Mondays are associated with more AVC records. This may be because many deer killed on Fridays, Saturdays, and Sundays are not picked up and recorded until the following Monday.

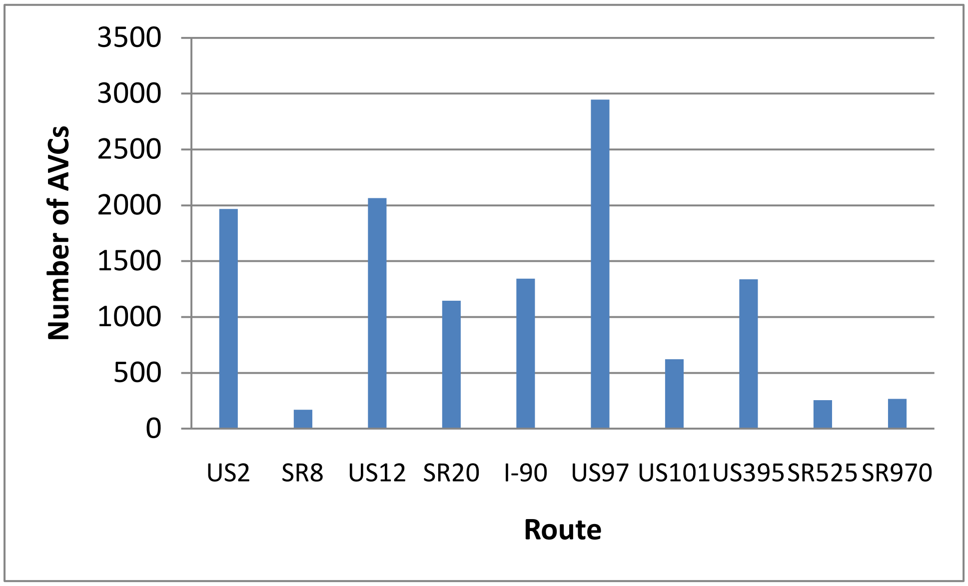


Figure 6-1: Total number of AVCs, by route (2002-2006)

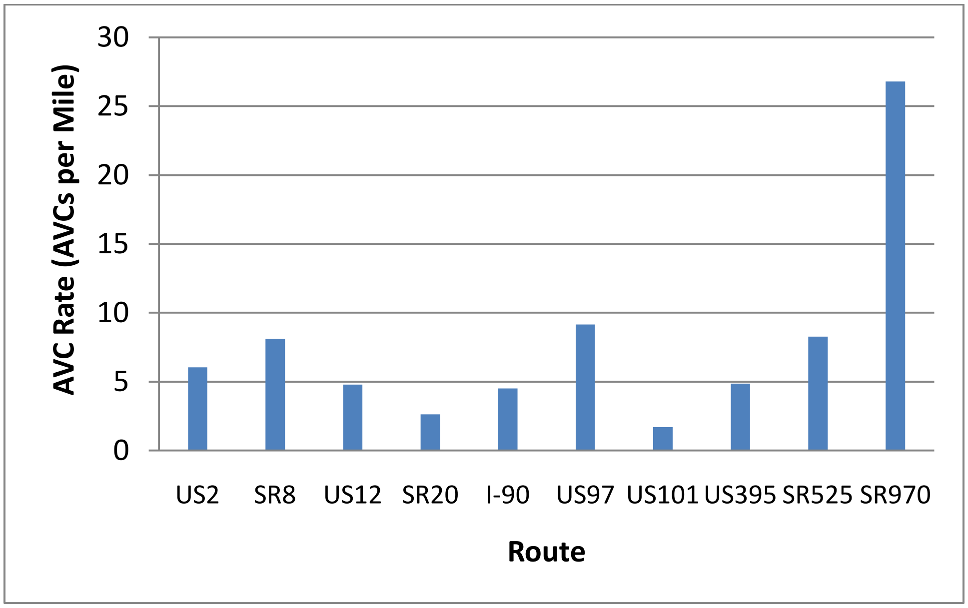
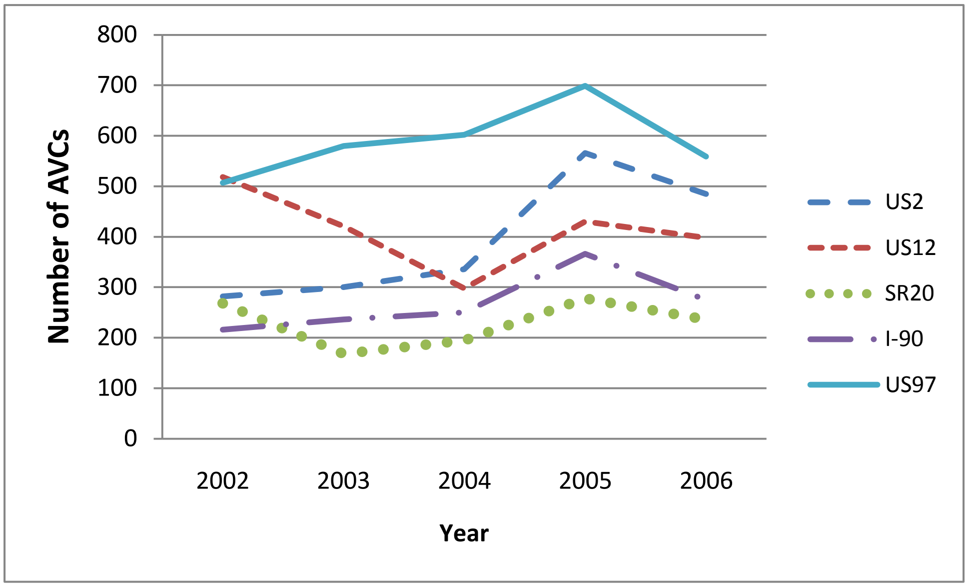
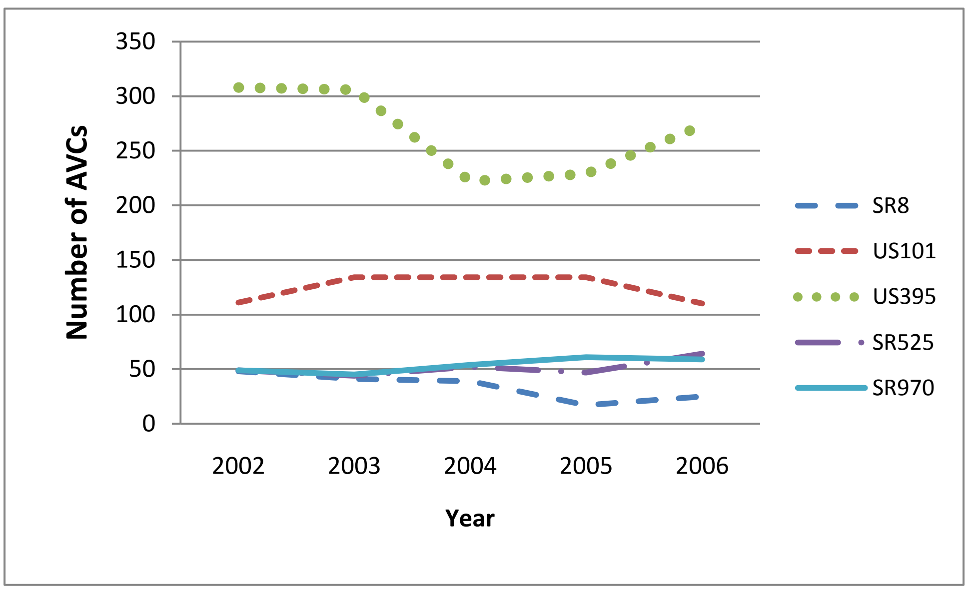


Figure 6-2: Average AVC rate (number of AVCs per mile), by route (2002-2006)



(a)



(b)

Figure 6-3: Total number of AVCs, by year (2002-2006)

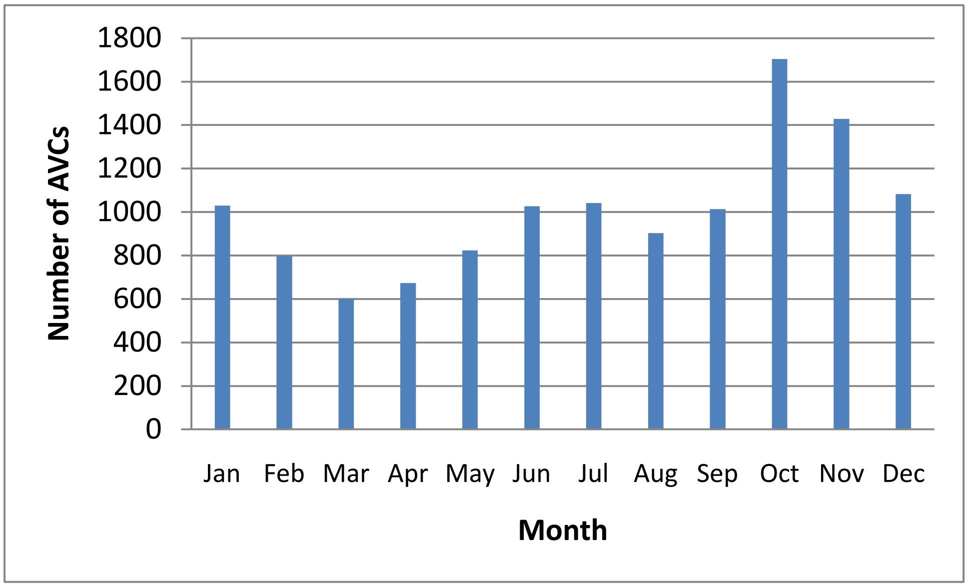


Figure 6-4: Total number of AVCs, by month (2002-2006)

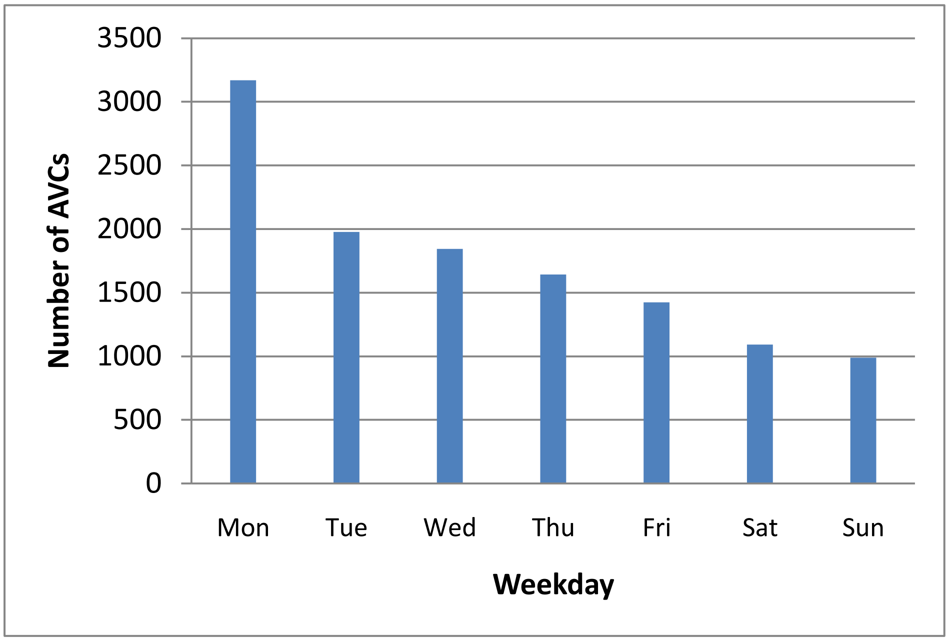


Figure 6-5: Total number of AVCs, by weekday (2002-2006)

6.2.2 Roadway Geometry Analysis

For this analysis, roadways were separated into different segment units. Each segment unit was consistent in terms of certain characteristics, such as lane width, shoulder type, and median width, and a new section was defined each time any of the characteristics changed (HSIS, 2009). Then the associations of AVCs with different roadway geometric characteristics, such as grade, number of lanes, and terrain type, were analyzed. Two figures were produced for each characteristic type: one showing the total number of AVCs (figures 6-6 through 6-8) and the other illustrating the total number of AVCs divided by the total segment length (figures 6-9 through 6-11). The latter procedure, called "normalization," allowed comparisons of different groups while controlling for segment length. For comparison purposes, the normalized figures provide more intuitive inference than the non-normalized ones, but this report contains the non-normalized figures (figures 6-6 through 6-8) for readers' reference.

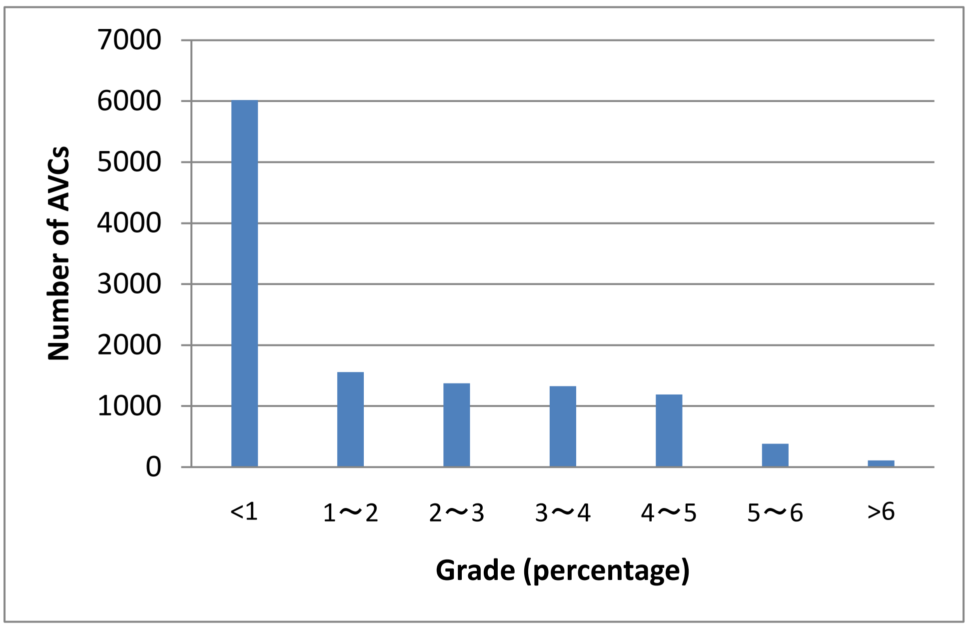


Figure 6-6: Total number of AVCs, by grade (2002-2006)

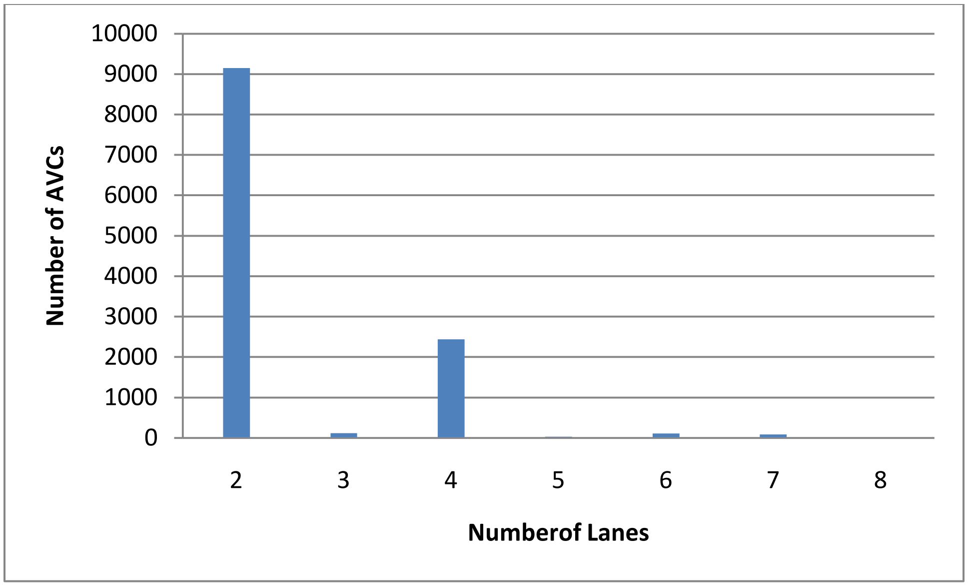


Figure 6-7: Total number of AVCs, by number of lanes (2002-2006)

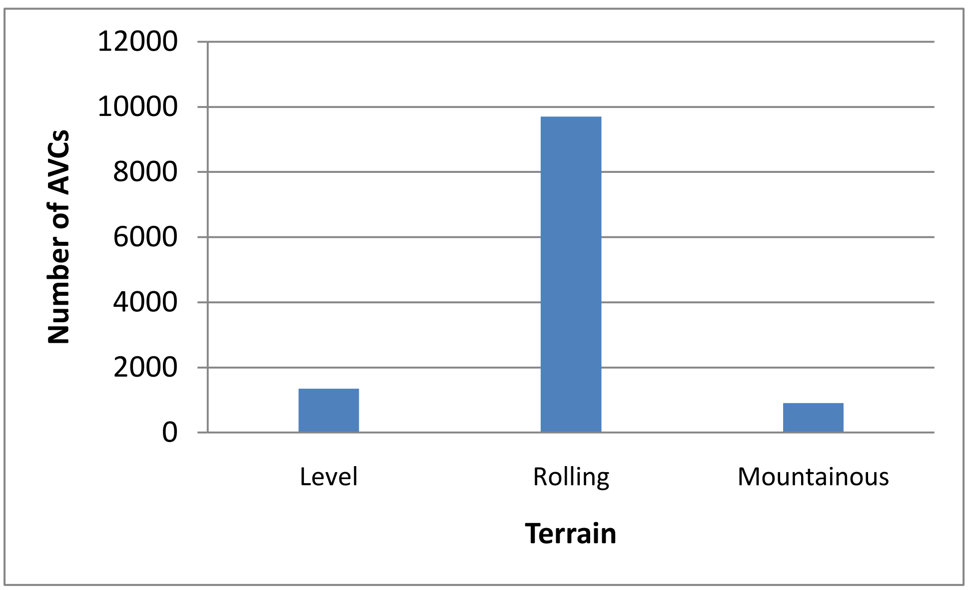


Figure 6-8: Total number of AVCs, by terrain type (2002-2006)

Figure 6-9 shows no clear trend in the statistical results for grade. However, when the grade is greater than 5 percent, the number of AVCs starts to decline. This could be a result of the vertical grade's impacts on vehicle speed and stopping sight distance. It could also be caused by animals' preference to move on gentler slopes. The tendency to build roads on the shallowest available slopes may put animals and vehicles in the same location. Figure 6-10 shows that the impact of the number of lanes on AVC rate does not create a constant trend. Interestingly, six-lane roadways are associated with the lowest number of AVCs per mile. Figure 6-11 shows that rolling terrain tends to have higher numbers of AVCs. This may be a result of drivers' sight restrictions caused by rolling terrain. In addition, the higher AVC rate may be with a product of more deer activity in rolling terrain areas, as indicated by the deer distribution data.

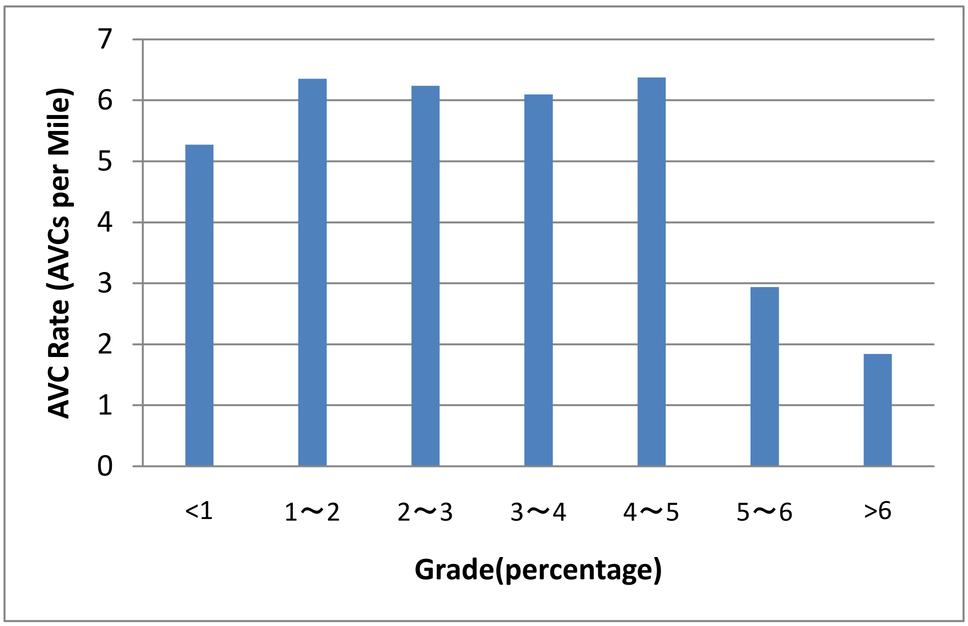


Figure 6-9: AVC rate (number of AVCs per mile), by grade (2002-2006)

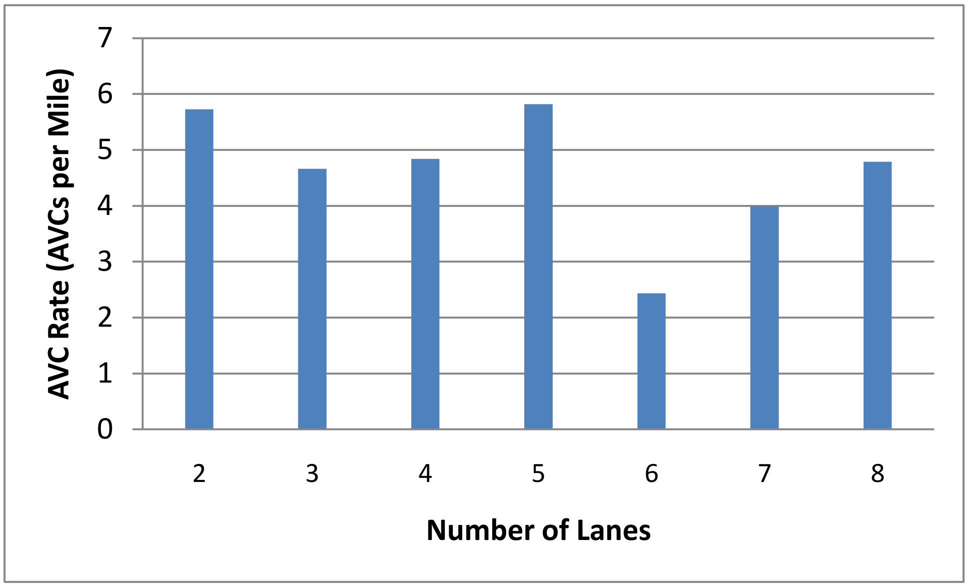


Figure 6-10: AVC rate (number of AVCs per mile), by number of lanes (2002-2006)

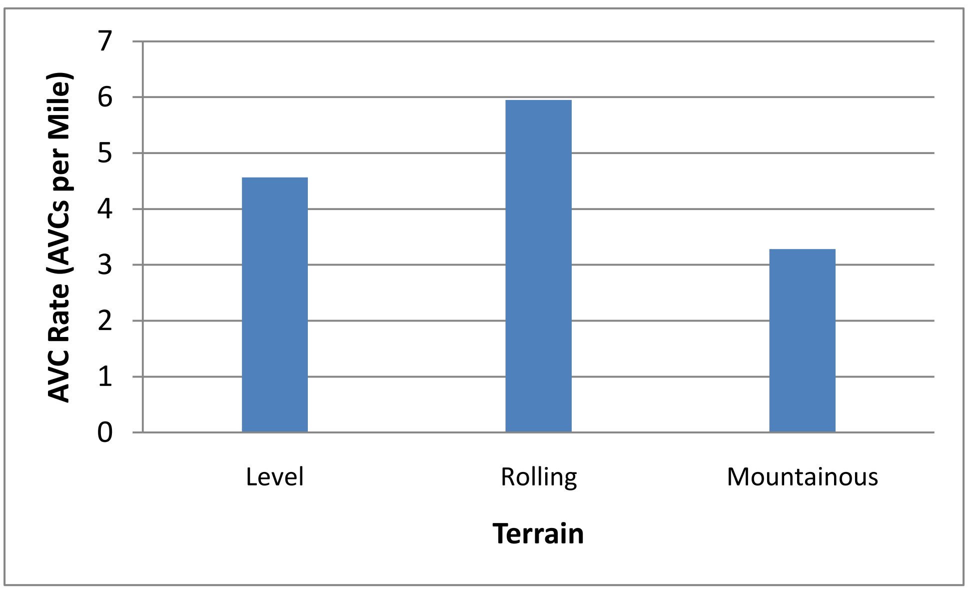


Figure 6-11: AVC rate (number of AVCs per mile), by terrain type (2002-2006)

Note that the sample sizes for level terrain, rolling terrain, and mountainous terrain were 1915, 7542, and 1009 records, respectively

6.2.3 Traffic Characteristic Analysis

The relationships of AVCs to various traffic characteristics, including the speed limit, annual average daily traffic (AADT), and percentage of trucks, were also analyzed. Similarly to Section 6.1.2, figures 6-12 through 6-14 show the total number of AVCs for each type of traffic characteristic, while figures 6-15 through 6-17 plot the results normalized by the total segment length.

Some key points were concluded from the results shown in figures 6-15 through 6-17:

- As a general trend, the AVC rate increases with speed limits until 60 mph (as shown in Figure 6-15). The only exception is a 50 mph speed limit. This could be caused by random variation or biases where 50 mph speed limits are used. The AVC rates for speed limits of 65 mph and 70 mph are not as high as those for the 60 mph speed limit. The reason is likely related to the enhanced access control for high speed roads with 65 mph or higher speed limits.

- As shown in Figure 6-16, the AVC rate is relatively low when AADT is lower than 4,000 and shows a declining trend after AADT=25,000, except for the AADT range between 50,000 and 55,000 because of smaller sample size. Further research is needed to collect more data to investigate this issue.

- Figure 6-17 shows that the AVC rate decreases when the percentage of trucks becomes higher than 5 percent. This may be because more trucks slow traffic speed or make more noise that drives animals away from highways. Moreover, drivers tend to drive more carefully when more trucks are present.

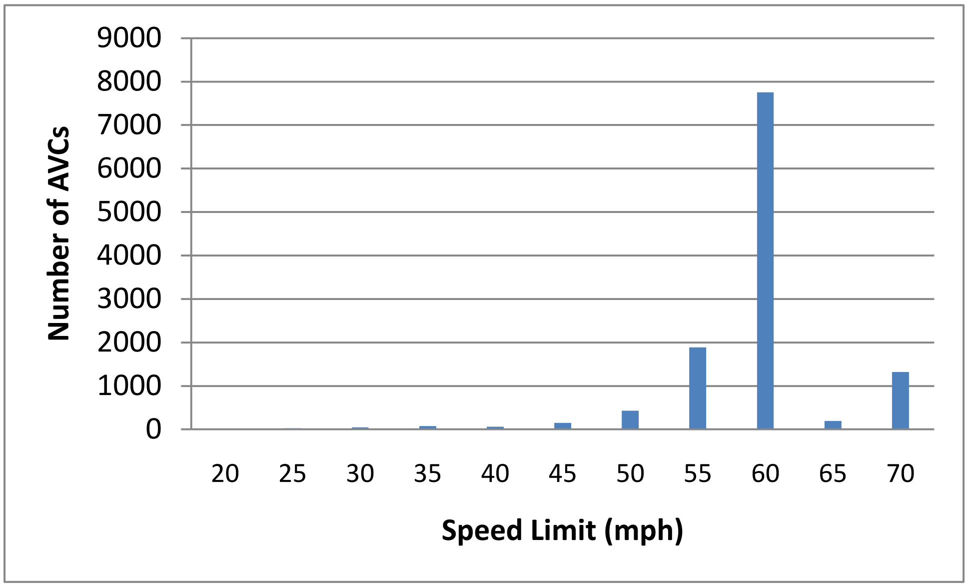


Figure 6-12: Total number of AVCs, by speed limit (2002-2006)

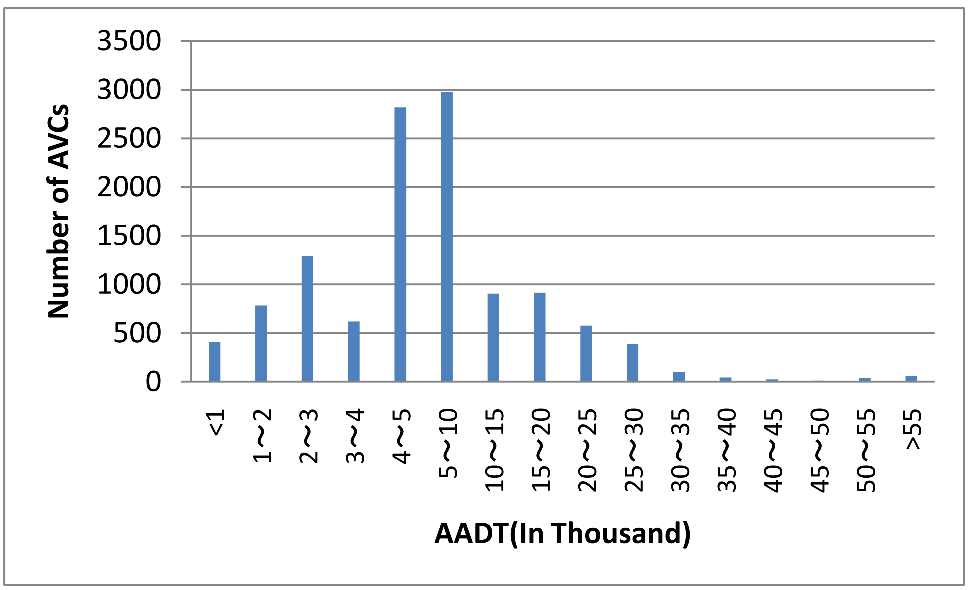


Figure 6-13: Total number of AVCs, by AADT (2002-2006)

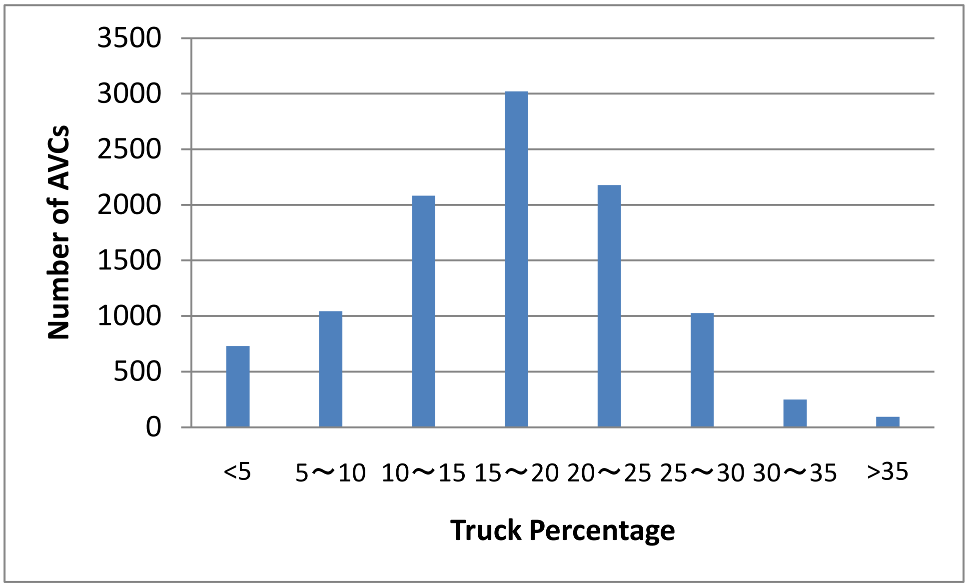


Figure 6-14: Total number of AVCs, by percentage of trucks (2002-2006)

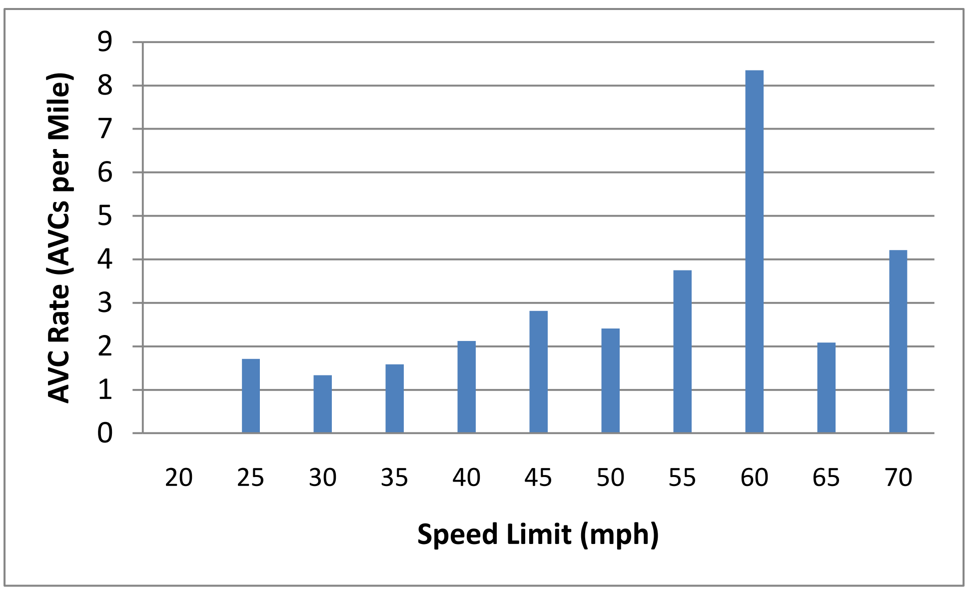


Figure 6-15: AVC rate (number of AVCs per mile), by speed limit (2002-2006)

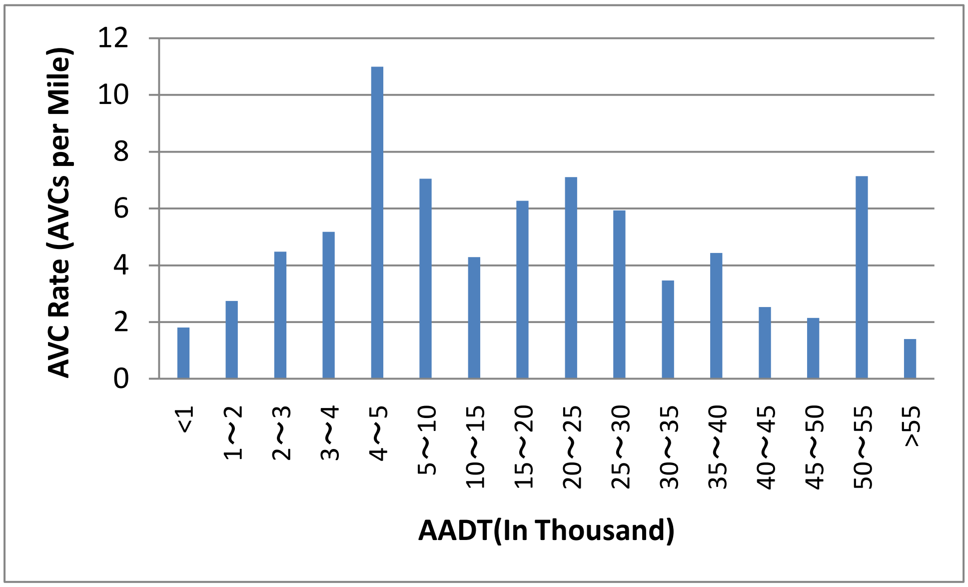


Figure 6-16: AVC rate (number of AVCs per mile), by AADT (2002-2006)

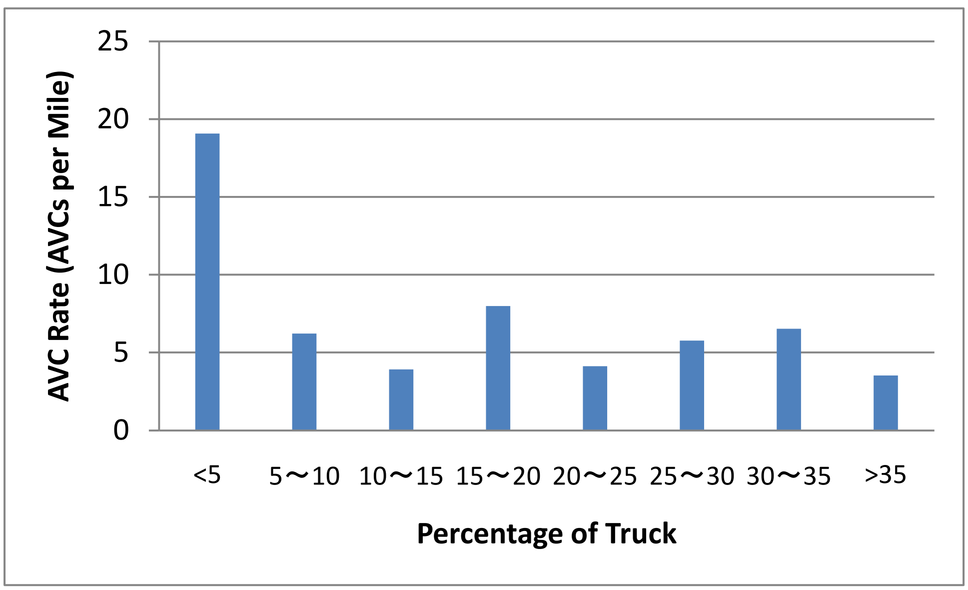


Figure 6-17: AVC rate (number of AVCs per mile), by percentage of trucks (2002-2006)

6.2.4 Deer Distribution Analysis

The geographic distributions of carcass pickups and their relationship to deer habitats were analyzed. Because of a lack of detailed information about deer types in the CRpt data, only CR data were geographically compared with deer population distribution (habitat) data retrieved from WSDOT and WDFW.

Table 6-1 shows the carcass pickup records for different deer types from 2002 to 2006. Figure 6-18 shows the distribution of mule deer carcass pickup locations and mule deer habitat areas in Washington state. Note that black-tailed deer are a subspecies of mule deer and are therefore included in the mule deer map.

Table 6-1: Carcass records for different deer types (2002-2006)

Animal Type	Carcass Records					
	2002	2003	2004	2005	2006	Total
Mule Deer	1307	1339	1282	1528	1201	6657
Black-tailed Deer	850	753	874	903	723	4103
White-tailed Deer	867	915	795	1368	1335	5280
Elk	90	93	97	135	101	516

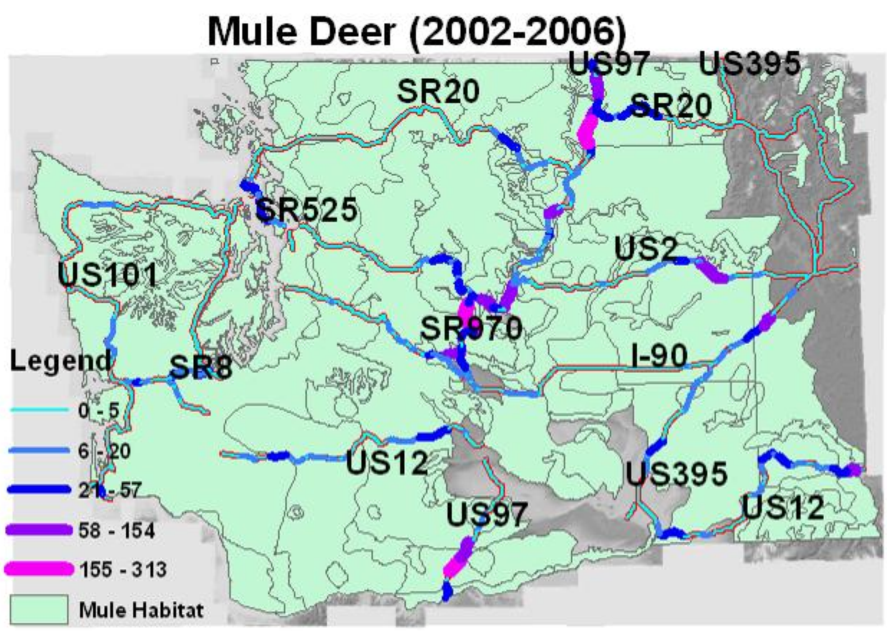


Figure 6-18: Comparison between the mule deer carcass pickup locations and their habitat

Figure 6-19 shows the white-tailed deer habitat areas in Washington state. Figure 6-20 illustrates the overlap between the distribution of white-tailed deer carcass pickup locations and white-tailed deer habitat areas[1] in Washington state. Similarly, Figure 6-21 shows the pickup locations of elk carcasses and elk habitat areas in Washington state. From figures 6-18, 6-19, and 6-21, one can see that the habitat areas of the three types of animals cover all of the locations with high carcass pickups. Most carcasses were found and picked up in the deer habitat areas. This indicates that the AVC records are strongly related to animal population distribution.

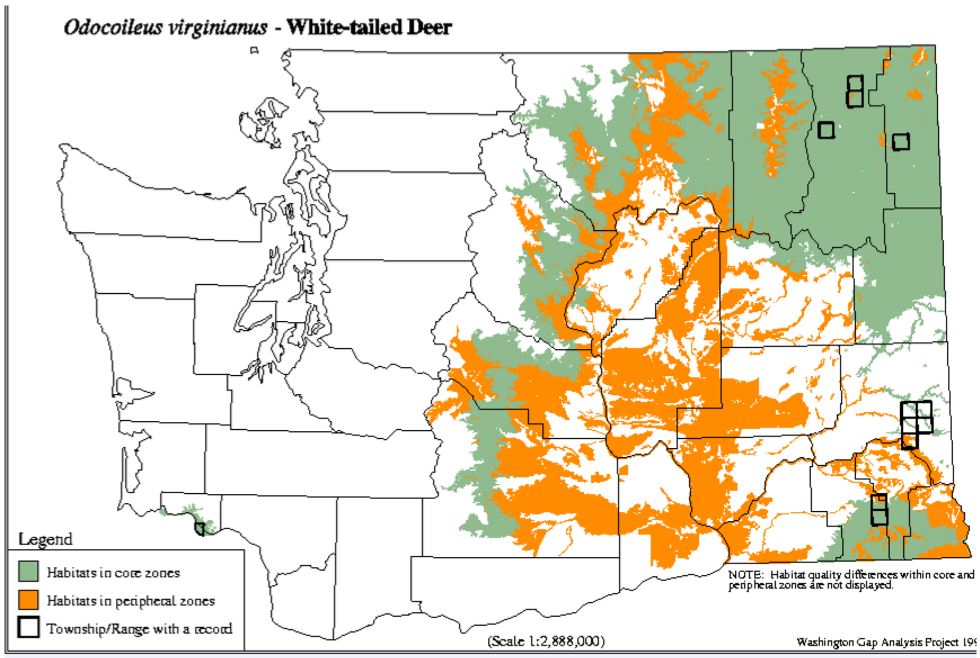


Figure 6-19: White-tailed deer habitat areas

[1] The GIS map in Figure 6-20 was converted from the raster image in Figure 6-19, since GIS data for the white-tailed deer habitats were not available from WSDOT. The other maps were created directly from GIS data.

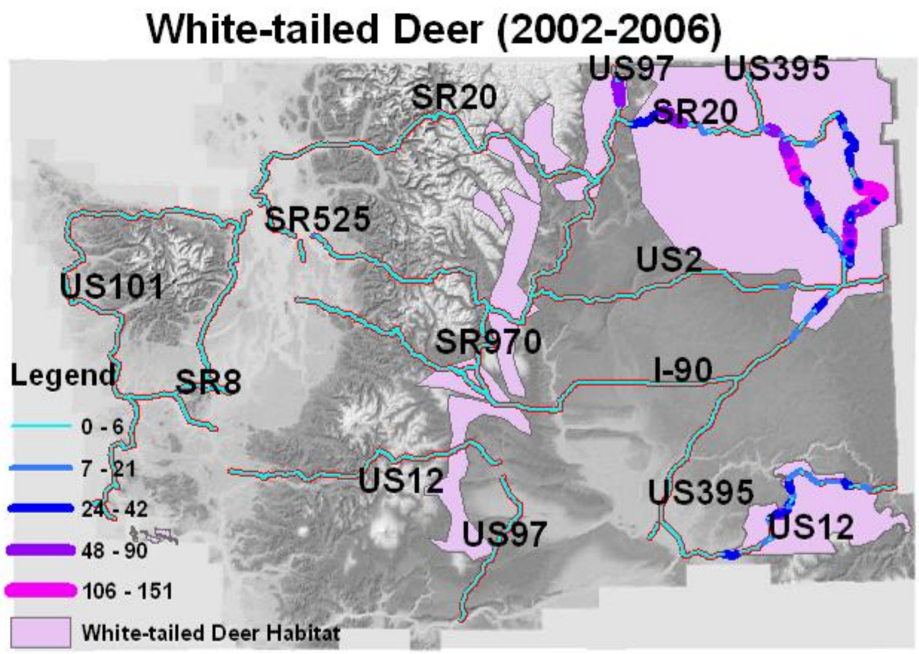


Figure 6-20: Comparison between white-tailed deer carcass pickup locations and their habitat areas

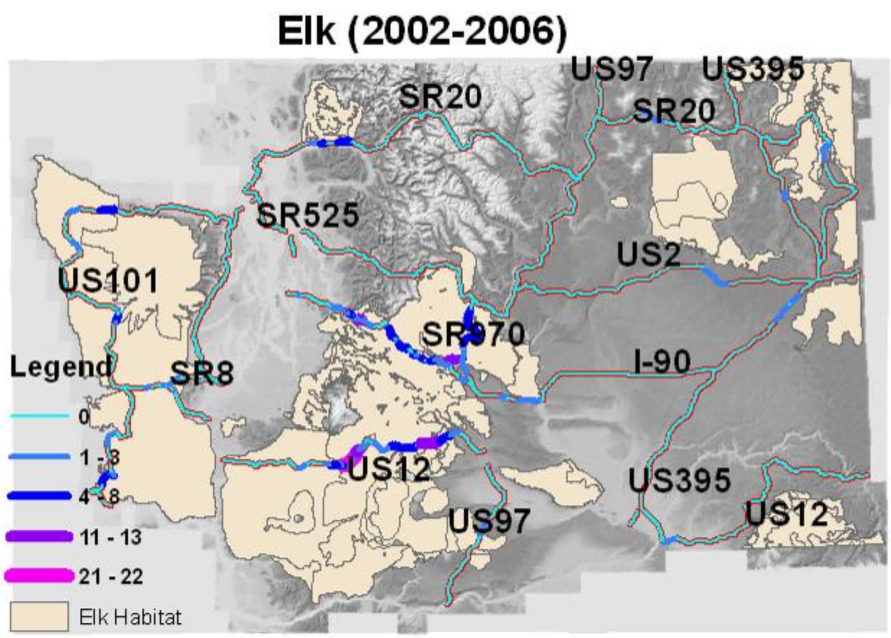


Figure 6-21: Comparison between the elk carcass pickup locations and their habitat areas

6.2 STATISTICAL ANALYSIS

For this analysis, a t-test was used to compare the means of two groups. The combined data from Chapter 5 were used in this analysis.

6.3.1 Tested Variables

Table 6-2 shows the analysis variables used for the t-test. All the variables in this table are dummy variables. The dummy value is described in the last column.

Table 6-2: Tested variables

Independent Variable	Variable Description	Dummy value
rururb	Rural area	1 yes; 0 no
rolling	Terrain type: rolling	1 yes; 0 no
mountainous	Terrain type: mountainous	1 yes; 0 no
spd_limtlevel	Posted speed limit>50mph	1 yes; 0 no
trkpctslevel	Percentage of trucks>5%	1 yes; 0 no
ver_curve	Grade percentage>3%:	1 yes; 0 no
hor_curve	Horizontal curve degree>3%	1 yes; 0 no
White-tailedArea	white-tailed deer habitat	1 yes; 0 no
MuleArea	mule deer habitat	1 yes; 0 no
ElkArea	elk habitat	1 yes; 0 no

6.3.2 t-Test

For each segment, the number of AVCs was divided by the segment length to compute the AVC rate. Given the dummy variables listed in Table 6-2, all the segments were separated into two groups based on the criteria being tested. The mean value of each group was examined with a t-test to determine whether the two group means had a significant difference. Table 6-3 shows the t-test results. Only those significant at the p=0.05 significance level are listed in Table 6-3.

Table 6-3: t-test results

Variable	Groups	N	Average AVC rate (AVCs per mile)	p-value
Rururb	Rural	7933	2.50	0.00
	Urban	2534	1.07	
terrain_rolling	Rolling	7542	2.50	0.00
	Others	2925	1.27	
terrain_mou	Mountainous	1009	1.65	0.02
	Others	9458	2.21	
spd_limtlevel	>50mph	7208	2.86	0.00
	Others	3259	0.59	
trkpctslevel	Percentage of trucks>5%	8014	1.21	0.00
	Others	2453	5.24	
White-tailedArea	Yes	3259	3.61	0.00
	No	7208	1.50	

The results shown in Table 6-3 can be interpreted as follows:

- The average AVC rate for rural areas is significantly higher than that in urban areas. This is likely because the animals' habitats are mostly in rural areas.
- The average AVC rate for rolling terrain is significantly higher than that of other terrain types. However, the AVC rate of mountainous terrain is significantly lower than that of other types.
- Higher speed limit segments tend to have higher average AVC rates. This may be because drivers' perception and reaction times, as well as animals' escape time, are shortened when vehicle speed is higher.
- When the percentage of trucks is low, the collision rate tends to be high. This result is similar to the conclusions drawn from Figure 6-17: trucks tend to reduce traffic speed and make louder noise that may drive animals away.
- The average AVC rate of the white-tailed deer habitat areas is significantly higher than that of other habitat areas. This indicates that the white-tailed deer habitat

locations are more strongly related to AVC rates. This may because white-tailed deer have higher densities in these areas, and possibly that they prefer valley bottoms, where roads are often located.

CHAPTER 7. ACCIDENT FREQUENCY MODELING

Two accident modeling approaches were used. The first approach was the fuzzy logic-based method, introduced in Section 4.2, which combined the CRpt and the CR datasets and then applied the combined data to the MP models introduced in Section 4.4.1 to quantify the relationships between explanatory variables and AVC frequency. The second approach employed the DIBP regression model, introduced in Section 4.4.2, to fit both data sources simultaneously without combining the two datasets. A comparison of the MP model and DIBP regression model results can be found in Section 7.4.

An open source statistical analysis package, R (http://www.r-project.org/, 2009), was used to estimate the models. A popular tool, R provides effective programming capabilities for developing customized functions to estimate various econometric models. Additionally, the R community at large can create and distribute modules for use in specific applications.

7.1 DATA DESCRIPTION

Table 7-1 lists all explanatory variables used in the modeling process. Most of the quantitative and dummy variables were selected from or created on the basis of the observed data, such as AADT. Some variables, such as legal speed limit (z3), were divided into two types, discrete and continuous. Variable z3 was a dummy variable that took the value 1 when the posted speed limit was greater than 50 mph and was 0 otherwise. Either the dummy value of the legal speed limit or its quantitative value was used in the modeling process. Variable z5 is a dummy variable indicating whether the width of the median is greater than six feet. This variable is important because median width is relevant to the refuge space needed for deers or elks to escape from AVCs. Variables such as z14, z15, and z16 were created to represent habitats of different animal types.

The minimum, maximum, mean, and standard deviation (S.D.) of each variable are also included in Table 7-1. One can see that both the CRpt data and the CR data were over-dispersed, as indicated by the variance being higher than the mean.

Table 7-1: Description of explanatory variables in the models

	Variable	Min	Max	Mean	S.D.
X*	Number of reported AVCs per segment #	0	22	0.24	0.81
Y**	Number of carcasses per segment #	0	95	0.94	3.88
z1	Annual average daily traffic (in thousands)	0.31	148.8	13.85	19.76
z2	Restrictive access control (Yes: 1; No: 0)	0	1	0.24	0.43
z3	Legal speed limit (mph)	20	70	52.76	10.79
z3’	Speed level (>50mph: 1; otherwise: 0)	0	1	0.69	0.46
z4	Truck percentage (%)	0	52.28	14.05	8.29
z4’	TruckPerLevel (>5%:1; otherwise: 0)	0	1	0.77	0.42
z5	Median width (> 6 feet: 1; others: 0)	0	1	0.34	0.48
z6	Total number of lanes	1	9	2.79	1.24
z7	Roadway length (feet)	0.01	6.99	0.22	0.4
z8	Terrain type (Rolling: 1; Otherwise:0)	0	1	0.720	0.45
z9	Terrain type (Mountainous:1; Otherwise:0)	0	1	0.096	0.30
z10	Lane width (feet)	10	20	12.5	1.88
z11	Left shoulder width (feet)	0	18	2.44	2.04
z12	Right shoulder width (feet)	0	20	4.03	3.52
z13	Rural area (Rural:1; Urban:0)	0	1	0.758	0.43
z14	White-tailed deer habitat (Yes: 1; No: 0)	0	1	0.31	0.46
z15	Mule deer habitat (Yes: 1; No: 0)	0	1	0.51	0. 50
z16	Elk habitat (Yes: 1; No: 0)	0	1	0.31	0. 46
z17	Horizontal curve (Curve degree>3: 1; otherwise: 0)	0	1	0.16	0.36
z18	Vertical curve (Grade percentage>3%: 1; otherwise: 0)	0	1	0.22	0.42

* Specific to CRpt data only; ** Specific to CR data only; # Dependent variable; Min: Minimum; Max: Maximum.

Table 7-2 is a cross-tabulation for the AVC data. The columns are the numbers of reported AVCs on a roadway segment, and the rows are the number of carcasses picked up from a roadway segment. A cell at row *m* and column *n* represents the number of roadway segments that have *m* picked-up carcasses and *n* reported AVCs in the five-year study period. From this table, one can see that most roadway segments had zero records in both datasets. That is the (0, 0) cell has the largest number. This indicates that for most segments, no AVCs were observed during the study period. Among segments with at least one record in both the CRpt and CR datasets, the (1, 1) and (2, 1) cells contain the largest numbers of segments. Similarly, among the segments with at least two records in each dataset, the (2, 2) cell contains the most records. Therefore, the diagonal cells, cells (0, 0), (1, 1), and (2, 2) should be expected to play important roles in the datasets.

Table 7-2 Cross-tabulation for AVC and CR data

		Number of Reported AVCs								Cumulated
		0	1	2	3	4	5	6	>6	Record
Number of Carcasses	0	**6698**	361	77	21	10	3	2	2	7174
	1	301	**67**	22	10	6	3	0	1	410
	2	228	69	28	5	6	2	2	0	340
	3	81	35	9	7	1	0	0	1	134
	4	63	26	10	5	1	2	1	0	108
	5	35	17	8	1	2	1	0	2	66
	6	26	17	7	7	0	0	2	0	59
	7	15	14	7	4	1	1	2	0	44
	8	17	8	7	4	2	0	0	0	38
	>8	81	64	43	31	22	13	10	16	280
Cumulated Record		7545	678	218	95	51	25	19	22	8653

7.2 RESULTS OF MP MODEL

7.2.1 Estimation Results

Table 7-3 shows detailed information about the three fitted models: the first two models were MP models estimated with Poisson regression, the last one was the best-fitted MP models estimated with negative binomial regression. Variables significantly associated with the probability of encountering an animal, P_o, the probability of an animal's failure to escape from being hit by the vehicle, P_{af}, and the probability of a driver's ineffective response, P_{vf}, were chosen as the explanatory variables in the models. Ps_i and NB_i represented the fitted Poisson and negative binomial regression models, respectively. i was the indicator for the base model (i=0) or non-base model (i=1). A base model can easily be obtained from the non-base model by setting all parameters to zero except the constant term. The AIC for each model was calculated and is shown in the last column of Table 7-3.

Table 7-3: Details of fitted models

		Model details				Evaluation Criterion
		P_0	P_{af}	P_{vf}	Par.	AIC
1	Ps_0	-1	-1	1	2	46892
2	Ps_1	-1, *z2,z6,z13, z14*	-1, *z5*	1, *z3',z4'*	8	29894
3	NB_1*	-1, *z2,z6,z13, z14*	-1, *z5*	1, *z3',z4'*	9	20772

*Best-fit model. Notes: "—" indicates no value; Par. is the number of parameters; "-1" indicates no constant term; "1" refers to the existence of a constant term.

Given the AIC values, the negative binomial regression models outperformed the corresponding Poisson regression models. Table 7-4 shows the estimated coefficients and test results for the explanatory variables of P_o, P_{af}, and P_{vf} in both models Ps_1 and NB_1. All of the listed variables in the Ps_1 and NB_1 models were significant at the p=0.05 significance level. The function *dispersiontest()* in the R package AER can be used to test equidispersion in Poisson against the alternative of overdispersion (Cameron and Trivedi, 1998; Kleiber and Zeileis, 2008). The hypothesis test using *dispersiontest()* assumed the δ value in equation (4-21) to be 0. The estimate results, δ value is 0.99 and the p value is 0.00, showed that δ was significantly greater than 0.

Because δ was significantly greater than 0, the combined AVC data were over-dispersed. In this case, Ps_1 could not be used because, as a Poisson model, it assumed that the mean and variance of the AVC data were the same. NB_1 was the suitable choice for this study. According to the AIC values in Table 7-3, NB_1 fit the AVC data better than Ps_1. Hence, the following discussion and analysis are mainly based on results of the NB_1 model.

The estimated coefficients and p-values in Table 7-4 show the degree of association between the explanatory variables and AVCs. The sign of each coefficient indicates the increasing (positive sign) or decreasing (negative sign) impacts on the AVC risk.

Seven variables were found to significantly affect AVCs. Among these variables, two of them, i.e., z3 and z4, were associated with the probability of drivers' ineffective response. The only variable that affected P_{af} was z5. The remaining four variables had significant impacts on the probability of encountering a disturbance animal: z2, z6, z13, and z14.

Table 7-4: Estimated coefficients for explanatory variables in NB_1 model

		Variable description	Coef.	Std. error	t value	p value
P_0	*z2*	Restrictive access control (Yes: 1; No: 0)	-0.521	0.168	-3.107	0.002
	z6	Total number of lanes	-0.714	0.051	-14.074	0.000
	z13	Rural area (Rural:1; Urban:0)	1.082	0.142	7.618	0.000
	z14	White-tailed deer habitat (Yes: 1; No: 0)	1.036	0.153	6.791	0.000
P_{af}	*z5*	Median width (> 6 feet: 1; others: 0)	-0.654	0.259	-2.523	0.012
P_{vf}	Con.	Constant	-13.535	0.199	-68.118	0.000
	z3	Speed level (>50mph: 1; otherwise: 0)	1.260	0.105	12.025	0.000
	z4	TruckPerLevel (>5%:1; otherwise: 0)	-0.453	0.132	-3.432	0.001

Notes: Coef.: coefficients

7.2.2 Interpretation of P_o Results

As shown in Table 7-4, the four variables found to affect P_o significantly were *z2* (restrictive access control), *z6* (total number of lanes), *z13* (Rural or Urban), and *z14* (white-tailed deer habitat). *z2* and *z6* were dependent on roadway design, while *z13* and *z14* were related to animal activities.

Among the roadway design factors, total number of lanes was found to have a very significant decreasing impact on P_o (Coef.= -0.714, t=-14.074). This may be because roadway sections with more travel lanes are typically wider, and an animal's crossing movement is more likely to be impeded by vehicles' motion, lights, and noise. Hence, animals are more hesitant to cross over a wider roadway.

Restrictive access control was found to have a decreasing effect on P_o (Coef. = - 0.521, t=-3.107). This implies that AVC frequency is lower for roadways with access control. This may be because fencing and barriers along access-controlled roadways may effectively prevent animals from crossing the roadway at random locations.

Among the factors related to animal activity areas, the estimated coefficient of *z13* indicated that if a highway section is in a rural area, then P_o tends to increase (Coef.=1.082, t=7.618). This may be due to higher animal populations and activity levels in rural areas. In addition, if a highway section bisects white-tailed deer range, a driver on this section will have a higher probability of encountering a deer. White-tailed deer habitat was the most significant variable among the three animal habitat variables (Coef.=1.036, t=6.791). In comparison with white-tailed deer habitats, the range of mule deer habitats is much broader, covering most of Washington state, which may be the reason that the variable for mule deer distribution was not significant. The variable elk-area (elk habitat) was also not significant in the model. This may be because the total number of collisions with elk was a small portion of the AVC records for the study period.

7.2.3 Interpretation of P_{af} Results

Among the factors affecting the probability of an animal's failure to avoid the collision, P_{af}, a median width of greater than 6 feet was found to have a significant decreasing effect on P_f (Coef.=-0.654, t=-2.523). This indicates that it may be easier for the animals to evade oncoming vehicles if the median is broader. Wider medians may also affect an animal's perception of the space as safe and as a destination to run toward to avoid a collision.

7.2.4 Interpretation of P_{vf} Results

Among the factors affecting the probability of drivers' ineffective response, P_{vf}, two explanatory variables that were found to affect P_{vf} significantly were *z3* (speed limit level) and *z4* (truck percentage level). The speed limit level had a positive sign for the estimated coefficient (Coef.=1.260, t=12.025). This implies that when a highway segment has a speed limit of greater than 50 mph, the probability of a driver's ineffective response will increase. An explanation may be that drivers tend to drive faster under a higher speed limit. A vehicle running at a higher speed requires a longer stopping distance. Hence, when an animal is perceived, reaction time for a faster vehicle is shorter.

The truck percentage level was found to have an increasing impact on the probability of a driver's effective response (decreased failure to avoid collision Coef.=-0.453, t=-3.432). This can be explained in several aspects. Presumably, drivers pay more attention and drive at a lower speed when truck volume is more than 5 percent of total traffic. Routes with high truck percentages are likely to be busy routes with other protective features such as fences. Additionally, trucks tend to have taller profiles, which provide drivers longer sight lines and animals have more time to spot them in return. Another possible reason may be that trucks usually make more noise, which may drive animals away from the roadway; however, more research will be necessary to confirm this finding and the mechanism by which it functions.

7.2.5 Summary

Among all the significant variables, speed limit, rural area, and white-tailed deer habitat have increasing effects on AVC frequency. The remaining three variables, *restrictive access control*, *truck percentage*, *and total number of lanes* reduce the probability of AVC risk when the variable values increase. These results from this model may help elucidate the causes of AVCs and enable development of countermeasures against such collisions.

7.3 RESULTS OF THE DIBP MODEL

7.3.1 Estimation Results

To help compare different models, the model details and evaluation criterion for DP, BP, DIBP, and ZIDP are listed in Table 7-5. In order to compare the effects of different J values on the DIBP model, three models—DIBP0, DIBP1, and DIBP2—with different J values were also estimated.

The BP model had a better fit than the traditional DP model. The DIBP and ZIDP models generally had better fits because they considered the zero inflated portion. Overall, the DIBP1 model was considered the best-fitted model because it had the highest ρ^2, lowest AIC, and lowest BIC. In comparison with DIBP1, DIBP2 did not show any improvement in its log-likelihood, even when the *J* value was larger. That is because the number of records in the (2, 2) cell of Table 7-2 was relatively small. Therefore, the control of the parameter *J* depends on the diagonal cell values in the AVC-CR cross-tabulation, as well as goodness-of-fit measures. The mixing proportions (*p*) in the last column indicate that the data in the diagonal of the AVC cross-tabulation should be over 66 percent. This result accorded with the statistical result in Table 7-2, where the sum of the diagonal value is about 79 percent of the total data.

Table 7-6 shows estimated values of θ and λ in the DIBP models; θ values indicate the proportion of the corresponding diagonal cells in the mixing proportion data; λ values indicate the proportion of the three regions in Figure 1. All models show $\theta_0 >$ 0.99, indicating that more than 99 percent of the mixing proportion data had zero AVC records and less than 1 percent of the mixing proportion data had at least one AVC record for both datasets. This result is a direct consequence of the statistics in Table 7-2, where both datasets show large numbers (more than 6698) of zero-accident roadway segments. Note that the value of λ_3 represents the average number of overlapped records per road segment. For the DIBP1 results, the overlapping percentage in the reported AVC data is about 13 percent (0.0664/(0.0664+0.4605)). Table 7-7 shows the coefficient, standard deviation, *t*-value, and p-value for each explanatory variable for λ_1, λ_2, and λ_3. All the listed variables are statistically significant at a 5 percent significance level.

Table 7-5: Details for the six fitted models

	Model details					Evaluation Criterion				
	λ_1	λ_2	λ_3	J	Para.	LL	ρ^2	AIC	BIC	p
DP	-z5-z10-z14-z15	-z1- z5-z10-z15-z16	—	—	25	-21802	0.313	43654	43852	—
BP	-z5-z10-z14-z15	-z1- z5-z10-z15-z16	-z5- z6-z10-z12-z15	—	37	-21173	0.333	42421	42715	—
DIBP0	-z5-z10-z14-z15	-z1- z5-z10-z15-z16	-z5- z6-z10-z12-z15	0	38	-17283	0.456	34642	34944	0.6612
DIBP1*	-z5-z10-z14-z15	-z1- z5-z10-z15-z16	-z5- z6-z10-z12-z15	1	39	-17275	0.456	34628	34938	0.6637
DIBP2	-z5-z10-z14-z15	-z1- z5-z10-z15-z16	-z5- z6-z10-z12-z15	2	40	-17275	0.456	34630	34948	0.6637
ZIDP	-z5-z10-z14-z15	-z1- z5-z10-z15-z16	—	—	27	-17415	0.451	34884	35099	0.6659

*Best-fitted model; -z10 represents variable z10 is not used in the model; (—): the parameter is set zero; *LL*: Log-likelihood; Par.: number of parameters; *p*: mixing Proportion.

Table 7-6: Estimated values of θ and λ in DIBP models

Models	θ estimation			Mean of parameter λ		
	θ_0	θ_1	θ_2	λ_1	λ_2	λ_3
DIBP0	1	—	—	0.4608	1.9205	0.0659
DIBP1*	0.9976	0.0023	—	0.4605	1.9359	0.0664
DIBP2	0.9976	0.0023	0.0000	0.4606	1.9359	0.0664

Table 7-7: The DIBP1 model for AVC

Explanatory variables	λ_1				λ_2				λ_3			
	Coef.	st.err	t-value	p-value	Coef.	st.err	t-value	p-value	Coef.	st.err	t-value	p-value
Constant	-2.904	0.093	-31.067	0.000	-3.164	0.101	-31.369	0.000	-26.763	2.769	-9.665	0.000
Annual average daily traffic (in thousands)	0.013	0.001	13.556	0.000	—	—	—	—	0.069	0.005	14.827	0.000
Restrictive access control (Yes: 1; No: 0)	-1.141	0.032	-35.753	0.000	-0.986	0.062	-15.973	0.000	-2.036	0.136	-14.988	0.000
Legal speed limit (mph)	0.043	0.001	30.302	0.000	0.060	0.002	33.129	0.000	0.068	0.007	10.298	0.000
Truck percentage (%)	-0.049	0.001	-39.589	0.000	-0.011	0.003	-4.055	0.000	-0.069	0.004	-16.417	0.000
Total number of lanes	-0.198	0.020	-9.761	0.000	-0.395	0.017	-22.882	0.000	—	—	—	—
Roadway segment length (feet)	0.499	0.009	58.069	0.000	0.471	0.028	17.042	0.000	0.912	0.030	30.785	0.000
Terrain type (Rolling: 1; Otherwise: 0)	-0.302	0.029	-10.543	0.000	0.105	0.044	2.417	0.016	-1.925	0.096	-20.152	0.000
Terrain type (Mountainous: 1; Otherwise: 0)	-0.958	0.037	-25.646	0.000	-0.182	0.066	-2.755	0.006	-2.027	0.182	-11.159	0.000
Left shoulder width (feet)	0.036	0.004	9.718	0.000	0.038	0.004	8.836	0.000	0.092	0.012	7.416	0.000
Right shoulder width (feet)	0.034	0.003	12.466	0.000	0.032	0.003	11.340	0.000				
Rural or Urban (Urban:0; Rural:1)	0.560	0.046	12.114	0.000	0.780	0.049	15.790	0.000	19.984	0.232	86.172	0.000
White-tailed deer habitat (Yes: 1; No: 0)	—	—	—	—	0.973	0.088	11.005	0.000	1.607	2.743	0.586	0.558
Elk habitat (Yes: 1; No: 0)	0.203	0.018	11.162	0.000	—	—	—	—	1.417	0.078	18.102	0.000

7.3.2 Model Interpretation

Table 7-7 shows the DIBP1 model results, indicating the potential factors contributing to AVCs. In contrast to regular Poisson accident models, the DIBP model contains three dependent variables, λ_1, λ_2 and λ_3. λ_1 and λ_2 consider the effects on the CRpt and the CR, respectively, whereas λ_3 accounts for the combined effects on the overlapping CRpt and the CR datasets. The significance and interpretation of the explanatory variables for each dependent variable are presented below.

Among the traffic elements, three variables were found to contribute significantly to the occurrence of AVCs. The estimated coefficients showed that speed limit was the most significant variable affecting the occurrence of AVCs (λ_1: coef.= 0.043, t=30.302; λ_2: coef.=0.06, t=33.129; λ_3: coef.=-2.036, t=10.298). Higher speed limits tend to increase the likelihood of AVCs. This may be because drivers travel at higher speeds under a higher speed limit, and high-speed vehicles require longer stopping distances. Therefore, drivers may not be able to stop quickly enough to avoid colliding with an animal on the road. This finding is consistent with most AVC related research that has concluded that higher speed limits increase AVC rates (Rolley and Lehman, 1992; Allen and McCullough, 1976).

AADT was found to have an increasing relationship with λ_1 (Coef.= 0.013, t=13.556) and λ_3 (Coef.=0.069, t=14.827) but to not significantly affect λ_2. In the analysis described in Chapter 6, we found that collisions increase with traffic volume, up to a point. After that, higher traffic volumes may actually repel animal attempts to cross and may not significantly affect AVCs anymore. Overall, a higher AADT may increase the likelihood of AVCs because higher volumes elevate the level of accident exposure. This result is consistent with the accident research conducted by Chin and Quddus (2003). This explains why the AADT variable was found to significantly contribute to AVC occurrences in the overlapping portion of the reported AVC and CR data but to not significantly affect λ_2.

A higher truck percentage was found to decrease AVC for all λ*'s*. One reason may be that drivers are more cautious when more trucks are on the road. Another reason may be that trucks are generally louder, which may scare animals away. Trucks also tend to have better driver visibility forward, which could provide more time for drivers to react. This result is similar to the motor vehicle accident research by Milton and Mannering (1998), which identified a decreasing relationship between truck percentage and accident probability.

Among the geometric design elements, five variables were significantly associated with the occurrence of AVCs. Roadway segments with restrictive access control tend to have lower accident risks, with a t-ratio of -35.753 for λ_1, -15.973 for λ_2, and -15.973 for λ_3. This is not surprising because animals may find it more difficult to access highways protected by physical obstructions.

The variable total number of lanes was found to be statistically significant at a 5 percent significance level for λ_1 (t=-9.761) and λ_2 (t=-22.882) but not for λ_3. A roadway becomes wider with more lanes. On wider roads, drivers have a better ability to see roadway-crossing animals earlier and more room to avoid AVCs. In addition, wider roadway sections are typically associated with higher traffic volumes and more noise, which are likely to keep animals away. A lower animal crossing frequency, together with enhanced driver vision and enlarged reaction space, are likely reasons for the reduction in AVC risk.

Increasing roadway segment length was found to proportionally increase the occurrence of AVCs (λ_1: coef.= 0.499, t=58.069; λ_2: coef.=0.471, t=17.042; λ_3: coef.=0.912, t=30.785). This is not surprising because the longer the roadway segment is, the more likely it is to divide a habitat into small areas between which animals need to interact. Similarly, when more vehicle-miles are traveled on longer segments for the same traffic, number of lanes, etc., the risk of an AVC occurring on the segment will increase.

Both left and right shoulders were found to have an increasing impact on AVCs for λ_1 (t=9.718 for the left, t=12.466 for the right) and λ_2 (t=8.836 for the left, t=11.340

for the right). Generally, drivers have broader views on roadways with shoulders. However, drivers also tend to drive faster when shoulders are present. This estimation result implies that the net impact of shoulder width is to increase AVCs.

In terms of area types, three variables were found to have significant impacts on the occurrence of AVCs: rolling, mountainous, and rural. In comparison with level terrain, rolling areas are associated with low numbers of reported AVCs (λ_1: coef=-0.302, *t*=-10.543) and AVCs in the overlapping portion of the reported AVC and CR datasets (λ_3: coef.=-1.925, *t*=-20.152). However, rolling areas were found to be associated with a higher number of CRs (λ_2: coef.=0.105, *t*=2.417). The contradiction in coefficient results may be caused by drivers being less likely to report AVCs when they hit animals in rolling areas, which may hide collision sites from view.

In comparison with level terrain, mountainous areas tend to have a low likelihood of AVCs and CR (λ_1: coef.= -0.958, *t*=-25.646; λ_2: coef.=-0.182, *t*=-2.755;λ_3: coef.=-2.027 *t*=-11.159). This may be because in mountainous areas, people drive more carefully, and vehicles are also slower. Another possible reason is that carcasses may not be easily found or require removal when they come to rest in areas off of roadways. Moreover, the valleys and tunnels in mountainous areas may impede animals' movements because these geometric characteristics may physically separate different habitats.

In comparison to highways in urban areas, those in rural areas were found to have more reported AVC and CR records in both datasets (λ_1: coef=0.560, *t*=12.114; λ_2: coef=0.780, *t*=15.790; λ_3: coef=19.984, t=86.172). This was unsurprising because animals are expected to be more active and populous in rural areas. However, looking at the overlapping portion of the two datasets makes this "rural effect" more obvious (λ_3: coef=19.984, t=86.172). This result highlights rural AVCs as a potential focus for future AVC research.

In terms of high-density animal distribution areas, white-tailed deer habitat areas were found to be associated with a higher λ_1 (*t*=11.005). Elk habitat was also found to have an increasing impact on λ_1 (*t*=11.162) and λ_3 (*t*=18.102). It makes sense that the areas with higher density animal distributions tend to have a higher AVC rate. However, mule deer habitat was not found to significantly affect the likelihood of AVCs. One main reason may be that the mule deer populations are distributed relatively uniformly and widely in Washington state, covering a large portion of the study routes.

In summary, speed limit, restrictive access control, and roadway segment length were the most significant explanatory variables affecting all λ*'s* (the absolute values of their *t* ratios were over 10). Therefore, in areas where the highway crosses the habitat of non-domestic animals, especially deer, transportation agencies should further examine speed limit and access control options to develop suitable countermeasures. Constructing fences and crossing infrastructure (e.g., tunnels and bridges) along and within the hot spots would be helpful for connecting segmented animal habitats and preventing animals from interacting with vehicles in areas with frequent AVCs (Donaldson, 2007).

7.3.3 Summary

This section describes the application of a DIBP regression model to fit two datasets concurrently. With the lowest AIC and BIC, the DIBP model outperformed other models, such as the DP model, BP model and ZIDP model, in this study. The overlapping portion (13 percent) of the two datasets in this study can facilitate a deeper investigation into the occurrence of AVCs.

The DIBP model identified the factors contributing to AVCs. Three dependent variables (λ_1, λ_2 and λ_3) were each linked with a group of explanatory variables, including traffic elements, geometric design factors, and geographic characteristics associated with AVCs. Speed limit and AADT were found to have an increasing effect on the likelihood of AVCs. Among geographic characteristics, wider shoulder widths and rural area segments tended to be associated with higher numbers of reported AVCs. Areas with dense animal distributions, such as white-tailed deer and elk habitats, were also found to increase the probability of AVC occurrence.

7.4 MODEL DISCUSSION

The results of the MP model and DIBP regression model were similar. Most variables significant in the MP model were also significant in the DIBP model. Moreover, these significant variables had similar impacts on the AVCs, whether increasing or decreasing. However, the two models had some differences, as stated below.

First, the input dataset for the MP model was a combined dataset, whereas the input dataset for the DIBP regression model used two separated datasets, the reported AVC data and CR data. In other words, the DIBP regression model was able to model these two datasets concurrently and did not require the data recovery process.

Second, both models had different significant variables. For example, the terrain type and shoulder width variables were significant in the DIBP model but not in the MP model. This may be because of the difference in dataset integrity.

Last, the interpretation principles of these two models were different. The MP model was based on the idea that AVCs are caused by three components: the hazardous crossing of an animal, the ineffective response of an animal to avoid the collision, and the ineffective response of the driver. The MP model therefore sought variables significantly related to these three components. On the other hand, the DIBP model considered the two datasets as three components: $\lambda 1$ (the CRpt data portion), $\lambda 2$ (the CR data portion), and $\lambda 3$ (the overlapping portion). The DIBP regression model was thus able to effectively explore the variables significantly related to these three components.

Overall, the MP model and the DIBP regression model can provide a more detailed explanation from different perspectives. The MP model is a better choice if only one dataset is available, whereas the DIBP regression model is a better choice if both the CRpt and CR datasets are available (but not combined).

CHAPTER 8. IDENTIFICATION OF HIGH AVC RISK LOCATIONS

Identifying AVC-prone areas can help better allocate limited resources to reduce AVC accidents. This chapter focuses on identifying hot spot locations on the basis of two datasets: the CR data and the combined data. In this research, a hot spot location was defined as a segment having a number of collisions that exceeded a specified threshold. This value was set as the "number of records" at the 99th percentile of all the non-zero records.

8.1 HOT SPOT IDENTIFICATION BY DEER TYPE

The CR data contained information for identifying high deer-mortality locations by deer type. The threshold values used to determine hot spots for mule deer, white-tailed deer, and elk were 79, 90, and 13, respectively. Tables 8-1, 8-2, and 8-3 list all the locations where the number of carcass pickups for each deer type was greater than the corresponding threshold value. This information will assist WSDOT transportation engineers in identifying the hot spots for specific types of AVCs.

Table 8-1: High mule deer carcass locations

Route	Segment Location*		Records	Route	Segment Location		Records
	BARM	EARM			BARM	EARM	
2	122.23	132.91	98	97	284.20	289.82	275
97	14.45	21.94	243	97	289.82	294.86	313
97	21.94	30.76	154	97	295.81	299.86	152
97	149.01	166.01	193	97	306.59	315.62	111

* BARM: Begin Accumulated Route Mileage; EARM: End Accumulated Route Mileage

Table 8-2: High white-tailed deer carcass locations

Route	Segment Location*		Records	Route	Segment Location		Records
	BARM	EARM			BARM	EARM	
20	425.02	436.00	128	395	220.33	228.09	151
395	214.35	217.62	147				

* BARM: Begin Accumulated Route Mileage; EARM: End Accumulated Route Mileage

Table 8-3: High elk carcass locations

Route	Segment Location* BARM	EARM	Records	Route	Segment Location BARM	EARM	Records
12	123.24	127.58	21	12	167.67	176.27	13
12	128.67	131.06	22				

* BARM: Begin Accumulated Route Mileage; EARM: End Accumulated Route Mileage

8.2 HOT SPOT IDENTIFICATION BY OVERALL AVCS

The combined dataset introduced in Chapter 5 was used for identifying hot spots for all animal types. The geographical distribution of AVCs in the combined datasets for the ten study routes are plotted in Figure 8-1. One can easily observe that portions of SR 97 had more AVCs along the route, indicating the severity of the AVC problem along this corridor.

As in Section 8.1, the value of 99th percentile of all the non-zero records was used to establish the threshold value for hot spot identification for all animal types. This threshold was 105. Table 8-4 lists the segments with record numbers higher than this threshold.

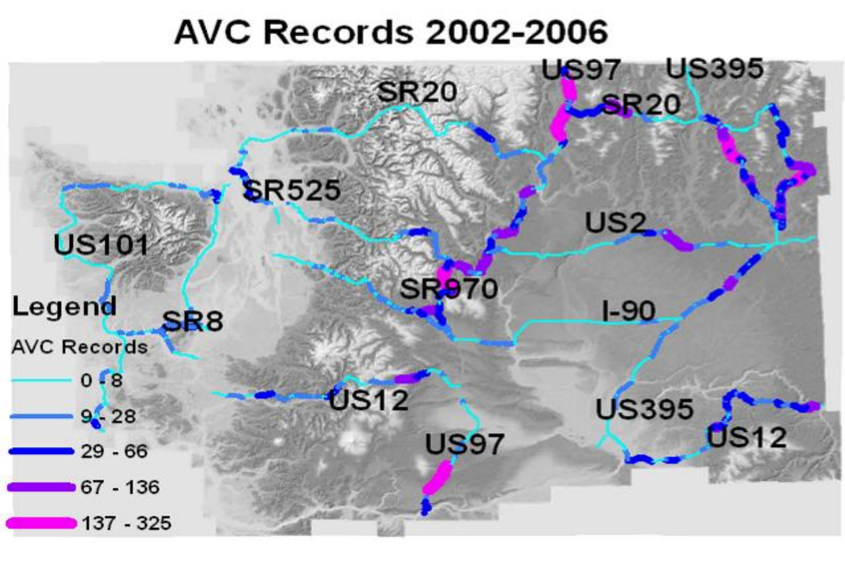


Figure 8-1: AVC location distribution for the ten study routes

Table 8-4: High AVC record locations

Route	Segment Location* BARM	Segment Location* EARM	Records	Route	Segment Location BARM	Segment Location EARM	Records
2	122.23	132.91	116	97	149.01	166.01	215
2	297.97	302.08	112	97	284.20	289.82	277
2	316.28	320.45	156	97	289.82	294.86	325
20	425.02	436.00	136	97	295.81	299.86	155
90	246.75	251.24	105	97	306.59	315.62	194
97	14.45	21.94	259	395	214.35	217.62	157
97	21.94	30.76	167	395	220.33	228.09	162

* BARM: Begin Accumulated Route Mileage; EARM: End Accumulated Route Mileage

To compare segments of different lengths, "number of AVC records per mile" was more suitable than "total number of records" Table 8-5 lists all of the segments with higher numbers of AVCs and AVCs per mile. The criteria here for selecting a hotspot were (1) Minimum records=10 and (2) Minimum records per mile =35. In this way, 1 percent of segments (25 out of a total 2548 segments) was selected. The segments in Table 8-5 are ranked by the AVC records per mile.

The segments in Table 8-4 and Table 8-5 are identified as the hot spots for the study routes. These segments should be priorities for further research into the specific factors that contribute to AVCs and potential preventive measures.

In a preliminary investigation of the causes of the hot spots, the first eight hot spots in Table 8-5 were chosen for a more detailed analysis. Figure 8-2 shows the locations of the first seven hot spots. Note that the locations ranked second, fourth, and sixth were adjacent to each other along US 97 and were therefore regarded as one location; the locations ranked seventh and eighth were also adjacent along US 395 and also were regarded as one location. Hence, in total, five locations were investigated. Figures 8-3 through 8-7 are photos downloaded from the WSDOT's SR Web (WSDOT, 2010) of these five locations.

Table 8-5: Segments with high AVC records and AVC records per mile

Rank	Route	BARM	EARM	L(mile)	Records	Records/mi	Mule/mi	W-deer/mi	B-deer/mi
1	2	297.42	297.97	0.55	36	65.45	0	61.82	0
2	97	289.82	294.86	5.04	326	64.68	62.30	0.60	0
3	12	356.61	356.94	0.33	21	63.64	3.03	51.52	0
4	97	294.86	295.81	0.95	57	60.00	60.00	1.05	0
5	90	84.15	84.35	0.2	12	60.00	45.00	5.00	0
6	97	284.2	289.82	5.62	278	49.47	48.93	0.36	0
7	395	214.35	217.62	3.27	158	48.32	0.92	44.95	0
8	395	217.62	217.89	0.27	13	48.15	0	48.15	0
9	97	134.72	135.04	0.32	15	46.88	46.88	0	0
10	90	103.91	104.35	0.44	18	40.91	38.64	0	2.27
11	12	356.98	358.2	1.22	49	40.16	4.10	34.43	0
12	2	294.81	296.41	1.6	64	40.00	0	35.00	0
13	97	204.36	205.66	1.3	52	40.00	39.23	0	0
14	970	3.73	6.33	2.6	102	39.23	28.08	0	0.38
15	97	295.81	299.86	4.05	154	38.02	37.53	0.49	0
16	97	13.16	13.85	0.69	26	37.68	1.45	0	37.68
17	2	316.28	320.45	4.17	156	37.41	0	33.33	0
18	90	81.84	82.49	0.65	24	36.92	29.23	0	0
19	90	81.84	82.49	0.65	24	36.92	29.23	0	0
20	2	294.56	296.41	1.85	67	36.22	0	31.35	0
21	90	103.85	104.35	0.5	18	36.00	34.00	0	2
22	2	314.52	316.28	1.76	62	35.23	0	33.52	0
23	90	81.04	81.41	0.37	13	35.14	21.62	0	2.70

* BARM: Begin Accumulated Route Mileage; EARM: End Accumulated Route Mileage; L: section length; Records/mi: total AVC records per mile; Mule/L: Mule deer carcass per mile; W-deer/L: white-tailed deer carcass per mile; B-deer/L: black-tailed deer carcass per mile

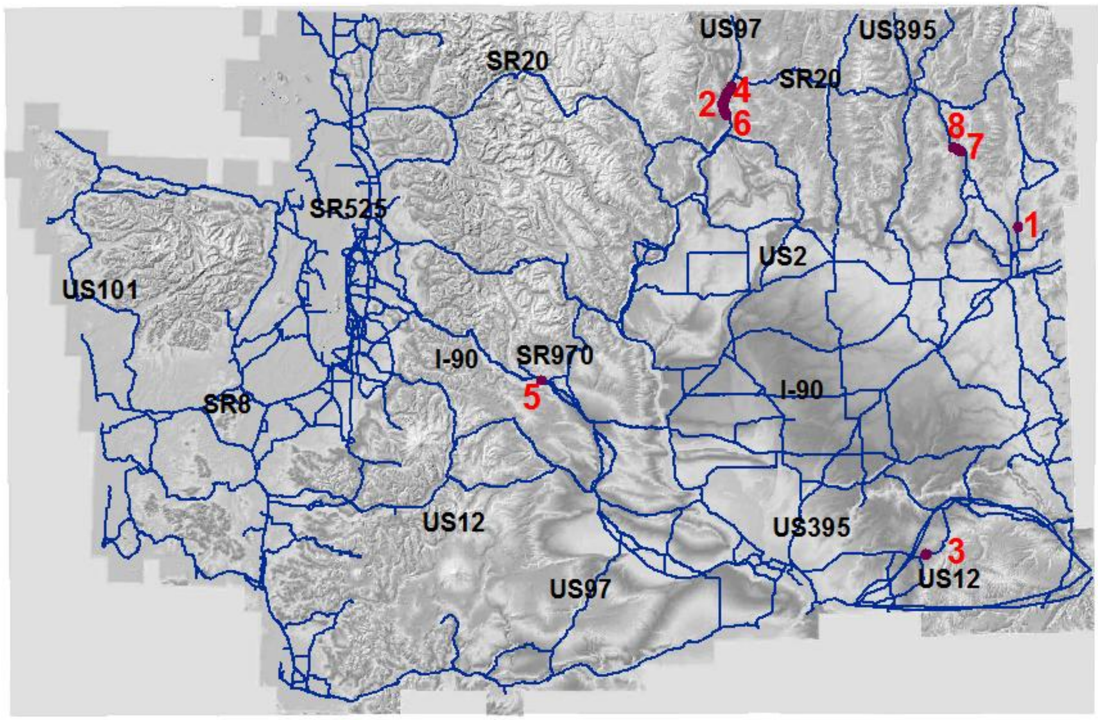


Figure 8-2: The locations of the first seven hot spots

As shown in Figure 8-2, the segments ranked No. 1 (on US 2), No. 3 (on US 12), and No. 7/8 (on US 395) were within the white-tailed deer habitat areas (see Figure 6-19). Not surprisingly, as shown in Table 8-5, the white-tailed deer carcass per mile rate for these three segments was high: 61.82 for the No. 1 segment, 51.52 for the No.3 segment, and 44.95 for the No.7/8 segment. Similarly, the segments ranked No. 2/4/6 on US 97, and No. 5 on I-90 were within the mule deer habitat areas (see Figure 6-18). The mule deer carcass per mile rate for these four segments was also high: 62.3 for No. 2 segment, 60.0 for No. 4 segment, 45.0 for the No. 5 segment, and 48.93 for the No. 6 segment.

In figures 8-3 through 8-7, one can find some roadway factors that may increase the probability of AVCs within these hot spots. Most factors are consistent with the modeling results:

- These hot spot segments are mostly located in rural areas, which is consistent with the findings of the proposed probability models, which show rural areas to have an increasing relationship with the AVCs.
- Most of these hot spot segments have two lanes, except US 2 and I-90 with four lanes. This is similar to the discovery of the models discussed in Charter 7: animal crossing frequency is lower at wider roadway sections.
- These hot spot segments seem to have sufficient shoulder widths. Unfortunately, wide shoulder width does not necessarily make AVC risk lower, as may have been thought. As discussed in Chapter 7, wider shoulder width enables earlier detection of crossing animals but is also associated with higher speeds. The net impact is to increase AVC risk. This finding is similar to the finding from the DIBP model that more AVCs may occur at segments with wider shoulders.

Figure 8-3: Segment photo at ARM= 297.88, US 2

Figure 8-4: Segment photo at ARM= 289.57, US 97

Figure 8-5: Segment photo at ARM= 356.77, US 12

Figure 8-6: Segment photo at ARM= 84.15, I-90

Figure 8-7: Segment photo at ARM= 216.6, US 395

CHAPTER 9 CONCLUSIONS AND RECOMMENDATIONS

9.1 CONCLUSIONS

The CRpt and CR data are often used for AVC studies. However, these two datasets are very different from each other, indicating that neither of the datasets is complete. Combining these two datasets together can significantly improve the quality of AVC data. In this study, we implemented two approaches to handle the multiple dataset issue and interpret the relationships between explanatory variables and AVC risk from different perspectives. The first approach was to use a fuzzy logic-based data mapping algorithm to combine the two datasets into a more complete set of data. On the basis of these data, a microscopic probability-based approach was proposed. The other approach was to use a DIBP regression model to fit the two datasets simultaneously. The results were encouraging and provided important insights into the causes of AVCs.

For the first approach, the proposed fuzzy logic-based data-mapping algorithm effectively combined the two datasets and avoided duplicating records in the combined dataset. For the ten study routes, about 27 percent to 37 percent of the CRpt data were matched to the CR data. The union of the two datasets significantly increased the number of samples and hence enhanced the quality of AVC data. In comparison to the original CR dataset, the new union dataset increased the number of records by 13 percent to 22 percent.

The threshold values used by the proposed algorithm were based on expert inputs collected through a survey. The effectiveness of the fuzzy logic-based data mapping algorithm was also verified by using the expert judgment data obtained from an evaluation survey. The verification results showed that the accuracy of the proposed algorithm was approximately 90 percent for the limited pairs of data included in the survey. The improved dataset will definitely benefit AVC risk modeling and statistical analysis. Because the design of the membership functions is adaptive in nature, the fuzzy logic-based data mapping algorithm can be easily transferred for data mapping applications in other areas.

On the basis of the improved dataset, an MP model was developed to assess the impacts of explanatory variables on AVC risk. The advantage of this proposed modeling approach over most existing ones is that it explicitly takes human factors and the spatial distribution of animal habitat into account. Explanatory variables that may affect the probability of a driver's ineffective response, P_{vf}, the probability of an animal's failure to avoid a collision, P_{af}, and the probability of an animal's presence, P_o, were estimated by the maximum likelihood estimation method. Seven variables were found to be significant.

For the second approach, a DIBP regression model was developed to quantitatively study AVCs by using the CRpt and CR datasets simultaneously. As an inflated version of the BP regression model, the DIBP model outperformed other models, as indicated by the AIC and BIC scores in this study in comparison to scores from the DP model, BP model, and ZIDP model. Functionally, the DIBP model not only can handle under- or over- dispersed count data but also can model paired datasets with correlation.

In this study, the DIBP model was demonstrated to be effective at modeling AVCs. The methodology developed in this study is general and may be applied to model other types of accidents with two datasets of similar characteristics. Although the DIBP model can be used to predict and assess the causal factors of AVCs on the basis of two different datasets concurrently, accident causation will need to be further investigated. Transferability testing will also be needed when this model is applied to other locations.

The principal findings from the MP and DIBP regression models can be used to develop countermeasures against AVCs. The findings are summarized as follows:

- In terms of traffic element factors, speed limit and AADT were found to have an increasing effect on the likelihood of AVCs, especially, when the speed limit is higher than 50 mph. The percentage of trucks was found to have a decreasing effect on the likelihood of AVCs.
- In terms of geometric characteristics, wider shoulder width and rural areas tend to be associated with higher frequencies of AVCs. Restrictive access control, higher numbers of lanes, and wider median width were found to be associated with lower AVC risks.

- In terms of geographic characteristics, areas within animal habitats, such as white-tailed deer areas, were found to increase the probability of AVCs occurring. Rural areas were also found to have an increasing relationship with the occurrence of AVCs.

The study identified AVC hotspots for all study routes. The most severe hotspots were interpreted on the basis of the modeling results and geospatial data. The locations of rural areas and deer habitat areas were found to have a high correlation with AVCs.

9.2 RECOMMENDATIONS

On the basis of the findings of this study, the research team would like to make the following recommendations:

- For the purpose of improving AVC data, it would be helpful to record specific animal types in the AVC report. With this information, the CRpt data and the CR data could be better mapped.
- Plans for new highways should avoid bisecting high quality animal habitats. Highways going through high quality animal habitats have negative effects on animal activities and ecology. In addition, animal movements between bisected habitats increase highway-crossing activities and hence increase the probability of AVCs.
- For existing highways that pass through high-density animal habitats, engineering solutions can be applied to reduce AVC risk. These solutions should focus on reducing either animal-vehicle interactions or the probability of driver failure in responding to animal presence.
- Solutions aimed at reducing the probability of encountering animals are desirable. These solutions include preventing animal crossing movements with grade separation at crossing hot spots and installing devices that can repel animals when vehicles are present.

- Solutions are also needed to reduce the probability of drivers’ failure to react. These solutions include installing warning systems that can alert drivers to pay more attention when highway-crossing animals are detected and setting speed limits lower than 50 mph to increase drivers’ available perception and reaction time when animals are present.
- This study also found that when the percentage of trucks was above 5 percent, AVC risk is lower. More data and analysis are needed to confirm this finding.

ACKNOWLEDGMENTS

The authors are grateful for the financial support to this project from the Washington State Department of Transportation (WSDOT). The authors wish to express sincere appreciation to WSDOT's Environmental Services Office and Research Office personnel, specifically Kelly McAllister and Rhonda Brooks, for their valuable suggestions and help with the data collection and surveys. The authors also want to acknowledge Highway Safety Information System (HSIS) staff member Yusuf Mohamedshah for his help with data collection. Special thanks also go to Barrett Welford Taylor at the University of Washington for his editing work, and Heather Turner from University of Warwick for her suggestions on model calculation.

REFERENCES

1. Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models, Transp Res Rec 2061, 55–63.
2. Akaike, H. 1974. "A new look at the statistical model identification". IEEE Transactions on Automatic Control 19 (6): 716–723.
3. Allen, R.E. and D.R. McCullough. 1976. Deer-Car Accidents in Southern Michigan. Journal of Wildlife Management, Volume 40, pp. 317–325.
4. Ben-Akiva, M. E., & Lerman, S. R. 1985. Discrete choice analysis: Theory and application to travel demand. MIT Press series in transportation studies, 9. Cambridge, Mass: MIT Press. pp:167-168.
5. Borman, S. 2009. The Expectation Maximization Algorithm - A short tutorial, http://www. seanborman.com/publications/EM_algorithm.pdf.
6. Cameron, A.C. and Trivedi, P.K. 1998. Regression Analysis of Count Data. Cambridge: Cambridge University Press. p.77-79
7. Chen, G. and Pham, T. T.,. 2001. Introduction to Fuzzy Sets, Fuzzy Logic, and Fuzzy Control Systems, CRC Press, pp.147-150.
8. Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Anal. Prev. 35, 253–259.
9. Curtis, P.D. and J.H. Hedlund. 2005. Reducing Deer-Vehicle Crashes. Wildlife Damage Management Fact Sheet Series. Cornell Cooperative Extension, Ithaca, N.Y.
10. Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining variation in safety performance of roundabouts. Accid. Anal. Prev. 42(2), 292-402.
11. Dempster, A, Laird, N, Rubin, D., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. J R Stat Soc Series B Stat Methodol 39, 1–38.
12. Donaldson, B.M., 2007. Use of Highway Underpasses by Large Mammals and Other Wildlife in Virginia: Factors Influencing Their Effectiveness. Transportion Research Record 2011, 157–164.
13. Donnell, E.T., Mason, J.M., 2006. Predicting the frequency of median barrier crashes on Pennsylvania interstate highways. Accid. Anal. Prev. 38 (3), 590–599.
14. El-Basyouny,, K., Sayed, T., 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. Transportation Research Record 1950, 9-16.
15. Frazier. L., 2001. Context Sensitive Design for U.S. 93 from Evaro to Polson. Montana Department of Transportation. Presentation file available online at http://environment.transportation.org/pdf/CONTEXT_SENSITIVE_US_93.pdf.
16. Garber, N.J., and Wu, L. 2001. Stochastic Models Relating Crash Probabilities with Geometric and Corresponding Traffic Characteristics Data (UVACTS-5-15-74). Charlottesville, VA: Center for Transportation Studies, University of Virginia.
17. Garcia-Molina, H., J.D. Ullman, and J. Widom. 2002. Database Systems: The Complete Book. Prentice-Hall, Inc. Upper Saddle River, New Jersey.
18. HSIS. 2009. http://www.hsisinfo.org/. Accessed June 15, 2009
19. http://www.r-project.org/. Accessed September 5, 2009

20. Huijser, M.P. and P.T. McGowen. 2003. Overview of Animal Detection and Animal Warning Systems. Proceedings of the 2003 International Conference on Ecology & Transportation. New York.
21. Huijser, M.P., J. Fuller, M.E. Wagner, A. Hardy, and A.P. Clevenger. 2007. Animal-vehicle Collision Data Collection: A Synthesis of Highway Practice. National Cooperative Highway Research Board Program: Synthesis 370. Transportation Research Board, Washington, DC.
22. Hubbard, M.W. B.J. Danielson and R.A. Schmitz. 2000. Factors influencing the location of deer-vehicle accidents in Iowa. Journal of Wildlife Management, Volume 64, pp. 707–712
23. Johnson, David H. and Thomas A. O'Neil. 2001. Wildlife-habitat relationships in Oregon and Washington. Oregon State University Press.
24. Jovanis, P. P., and Chang, H. 1986. Modeling the Relationship of Accidents to Miles Traveled, Transportation Research Record 1068, 42-51
25. Kadane, J.B., Shmueli, G., Minka, T.P., Borle, S., Boatwright, P., 2006. Conjugate analysis of the Conway–Maxwell–Poisson distribution. Bayesian Analysis 1, 363–374.
26. Kao, J. H. K. 1960, A summary of some new techniques in failure analysis, Proceedings of Sixth National Symposium on Reliability and Quality Control, Washington DC, 190-201.
27. Karlis, D., 2003. An EM algorithm for multivariate Poisson distribution and related models. J Appl Stat 30(1), 63-77.
28. Karlis, D., Ntzoufras, I., 2005. Bivariate Poisson and diagonal inflated Poisson regression models in R. J Stat Softw 14(10).
29. Karim, E., Tarek, S., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accid. Anal. Prev. 41 (4), 820-828.
30. Kim, J., Wang, Y., and Ulfarsson, G. 2007. Modeling the Probability of Freeway Rear-End Crash Occurrence. ASCE J. Transp. Eng., 133 (1), 11-19.
31. Kleiber, Christian and Zeileis , Achim, 2008. Applied Econometrics with R, Springer-Verlag, New York. ISBN 978-0-387-77316-2. P134
32. Knapp, K.K., C. Lyon, A. Witte, and C. Kienert. 2007. Crash or Carcass Data: A Critical Definition and Evaluation Choice. In Transportation Research Record: Journal of the Transportation Research Board, No. 2019, Transportation Research Board of the National Academies, Washington, D.C., pp 189-196.
33. Knapp K.K. and Yi X. 2004. Deer-Vehicle Crash Patterns and. Proposed Warning Sign Installation Guidelines. Transportation Research Board, Washington, DC.
34. Kumara, S.S.P., Chin, H.C., 2003. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. Traffic Inj. Prev. 3 (4), 53–57.
35. Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accid. Anal. Prev. 34 (2), 149–161.
36. Li, C.C., Lu, J.C., Park, J., Kim, K., Brinkley, P.A. & Peterson, J.P. 1999. Multivariate zeroinflated Poisson models and their applications. Technometrics, 41(1), 29-38.

37. Li HX, VC Yen, 1995. fuzzy sets and fuzzy decision-making, Boca Raton: CRC Press, pp. 93-99.
38. Liddle, A. R. 2007. Information criteria for astrophysical model selection. Monthly Notices of the Royal Astronomical Society: Letters 377:1, L74-L78.
39. Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37(1), 35-46.
40. Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accid. Anal. Prev. 38 (4), 751–766.
41. Lord, D., Guikema, S.D., Geedipally, S., 2007. Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes. Accident Analysis & Prevention.
42. Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. In: Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC.
43. Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Saf Sci 46 (5), 751–770.
44. Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Accid. Anal. Prev. 44 (5), 291-305.
45. Ma, J., Kockelman, K., 2006. Bayesian multivariate Poisson regression for models of injury count, by severity, Transp Res Rec 1950, 24–34.
46. Maher, M.J., Summersgill I.A. 1996. Comprehensive methodology for the fitting of predictive accident models. Accident Analysis & Prevention, 28(3) 281-296.
47. Malo, J.E., F. Suarez and A. Diez, 2004. Can we mitigate animal–vehicle accidents using predictive models?. Journal of Applied Ecology, Volume 41. pp. 701–710.
48. Malyshkina, N., Mannering, F., 2010. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents, Accid. Anal. Prev. 42 (1), 131–139.
49. Maycock, G.; Hall, R. D. 1984. Accidents at 4-arm roundabouts. Laboratory Report LR1120. Crowthorne, Berks, U.K.: Transport Research Laboratory.
50. Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accid. Anal. Prev. 26 (4), 471–482.
51. Miaou, S.-P., Bligh, R.P., Lord, D., 2005 Developing median barrier installation guidelines: a benefit/cost analysis using Texas data. Transp Res Rec 1904, 3–19.
52. Miaou, S., Hu, P.S., Wright, T., Rathi, A.K., and Davis, S.C. 1992. Relationship between truck accidents and highway geometric design: a Poisson regression approach, Transportation Research Record 1376, 10-18.

53. Miaou, S.P., Lord, D. 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and bayes versus empirical bayes. Transportation Research Record 1840, 31-40
54. Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accid. Anal. Prev. 25 (6), 689–709.
55. Miaou, S.P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accid. Anal. Prev. 37 (4), 699–720.
56. Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation 25 (4), 395–413.
57. Mohammed A., Noland, Robert B. and Ochieng, Washington Y. 2006. A High Accuracy Fuzzy Logic Based Map Matching Algorithm for Road Transport. Journal of Intelligent Transportation Systems, 10:3,103-115.
58. Morris H. DeGroot. Probability and statistics. Addison-Wesley Pub. Co., 1986. pp. 209.
59. Myers, W.L. Chang,, W.Y. Germaine, S.S. Vander Haegen, W.M. Owens,T.E. 2007. Analysis of Deer and Elk-Vehicle Collision Sites along State Highways in Washington State. Research Report for Washington Department of Fish and Wildlife. Olympia, Washington.
60. Naso, D.; Scalera, A.; Aurisicchio, G.; Turchiano, B. 2006. Removing spike noise from railway geometry measures with a fuzzy filter. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions. Volume 36, Issue 4, pp. 485–494.
61. Nikunja K. Swain, 2006. “A Survey of Application of Fuzzy Logic in Intelligent Transportation Systems (ITS) and Rural ITS”- Southeast Con, Proceedings of IEEE, pp 85-89.
62. Noland, R.B., and Quddus, M.A. 2004. A spatially disaggregate analysis of road casualties in England. Accident Analysis and Prevention, 36(6), 973-984.
63. Oh, J.,Washington, S.P., Nam, D., 2006. Accident prediction model for railway highway interfaces. Accid. Anal. Prev. 38 (2), 346–356.Park, E.S.,
64. Pitman J. 1993. Probability, Springer-Verlag New York, Inc.
65. Plait, A. 1962. The Weibull Distribution – with the tables, Industrial Quality Control, No. 19, 19-26.
66. Poch, M., Mannering F. 1996. Negative binomial analysis of intersection accident frequencies, Journal of Transportation Engineering, 122(2), 105-113.
67. Qin, X., Ivan, J.N., Ravishankar, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accident Analysis and Prevention 36 (2), 183–191.
68. Rao, D. H., S. S. Saraf. 1995. Study of Defuzzification Methods of Fuzzy Logic Controller for Speed Control of a DC Motor, IEEE Transactions, pp. 782–787.
69. Reilley, R.E. and H.E. Green. 1974. Deer Mortality on a Michigan Interstate Highway. Journal of Wildlife Management, Volume 38, pp. 16–19.

70. Rodriguez, D.A., Rocha, M., Khattak, A.J., and Belzer, M.H. 2003. Effects of truck driver wages and working conditions on highway safety: Case study. Transportation Research Record 1833, 95-102.
71. Rolley, R. E., Lehman, L. E., 1992. Relationships Among Raccoon Road- Kill Surveys, Harvests, and Traffic. Wildlife Society Bulletin, Vol. 20, 313–318.
72. Romin, L. A., and J. A. Bissonette. 1996. Deer–Vehicle Collisions: Status of State Monitoring Activities and Mitigation Efforts. Wildlife Society Bulletin, Vol. 24, No. 2, pp. 276–283.
73. Runkler, T. A.. 1996. Extended Defuzzification Methods and Their Properties. IEEE Transactions. pp. 694–700.
74. Schwarz, Gideon E. 1978. "Estimating the dimension of a model". Annals of Statistics 6 (2): 461–464.
75. Seiler A. 2005. Predicting locations of moose–vehicle collisions in Sweden. J Appl Ecol 42: 371–382
76. Shankar,V., Mannering, F., Barfield,W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal.Prev. 27 (3), 371–389.
77. Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accid. Anal. Prev. 29 (6), 829–837.
78. Shankar V.N., Ulfarsson, G.F., Pendyala, R.M., and Nebergall, M.B. 2003. Modeling crashes involving pedestrians and motorized traffic. Safety Science, 41(7), 557-640.
79. Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., Boatwright, P., 2005. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. Journal of the Royal Statistical Society, Part C 54, 127–142.
80. Siddique, Z. Q. 2000. Accident Risk Modeling of Vehicle-to-Bicycle and Vehicle-to-Pedestrian Accidents at Four-Legged Signalized Intersections, Master Thesis, Asian Institute of Technology.
81. Syed, S., and Cannon, M. E. 2004. Fuzzy logic-based map matching algorithm for vehicle navigation system in urban canyons. Proceedings of the Institute of Navigation (ION) National Technical Meeting, San Diego, CA. pp. 26–28.
82. Taylor, C. and D. Meldrum. 1998. Fuzzy Ramp Metering - Design Overview and Simulation Results. In Transportation Research Record: Journal of the Transportation Research Board, No. 1634, Transportation Research Board of the National Academies, Washington, D.C., pp 10-18.
83. TranSafety, Inc. 1997. Deer-Vehicle Collisions are Numerous and Costly. Do Countermeasures Work? Road Management & Engineering Journal. Available online at http://www.usroads.com/journals/rmj/9705/rm970503.htm.
84. Wagner, P. and M. Carey. 2006. Connecting Habitats and Improving Safety. WSDOT 2007-09 Transportation Research Project Problem Statements.
85. Wang, Y., 1998. Modeling Vehicle-to-Vehicle Accident Risks Considering the Occurrence Mechanism at Four-Legged Signalized Intersections .Ph.D. Dissertation, University of Tokyo

86. Wang, Y, Ieda, H., and Mannering, F.L. 2003. Estimating Rear-End Accident Probabilities at Signalized Intersections: An Occurrence-Mechanism Approach. J. Transp. Eng. 129(4), 1-8.
87. Washington, S.P., Karlaftis, M. G., & Mannering, F. L. 2003. Statistical and econometric methods for transportation data analysis. Boca Raton: Chapman & Hall/CRC, Boca Raton, Florida, 242-243.
88. Winkelmann and Zimmermann, 1995. R. Winkelmann and K. Zimmermann, Recent developments in count data modeling: theory and applications, J. Econ. Surveys 9, pp. 1–24
89. WSDOT (Washington State Department of Transportation). 2006. Transportation Plan Update. Accessed online at http://www.wsdot.wa.gov/planning/wtp/documents/HealthEnvironment.htm. Sept. 2007.
90. WSDOT, 2010. http://www.srview.wsdot.wa.gov/. access 04/24/2010
91. Zhao, Y. 1997. Vehicle location and navigation system. MA: Artech House, Norwood, MA.

APPENDIX. ATTRIBUTE EXPLANATION

The attributes (variables) of vehicles, accidents, roads, and animals in the database designs in Chapter 4 are described as follows.

A. Vehicles

List of variables for vehicle subfile:

Variable Name	Variable Description	Variable Type
bodytype	vehicle model	Char (3)
cmaxles	comm carrier num of axles	Char (2)
cmconfig	comm carrier config	Char (1)
com_body	comm carrier cargo body	Char (1)
drv_actn	drviver action	Char (2)
drv_age	drviver age	Char (2)
drv_sex	drviver sex	Char (1)
event	sequence of events	Char (2)
miscact	drviver misc action	Char (2)
veh_use	vehicle usage	Char (2)
vehno	vehicle sequence number	Char (8)
vehtype	vehicle type	Char (2)
vrd_type	roadway type	Char (1)

B. Accidents

List of variables for accident subfile:

Variable Name	Variable Description	Variable Type
ac_srmp	state route milepost	Char (1)
acctype	accident year	Char (2)
caseno	accident report number	Char (10)
coltype	accident type	Char (2)
comp_dir	compass direction	Char (2)
daymth	accident day of month	Char (2)
gps_latx	GPS x	Numeric (8)
gps_laty	GPS y	Numeric (8)
gps_latz	GPS z	Numeric (8)
impact	impact location	Char (2)
light	light condition	Char (1)
loc_type	location type	Char (1)
milepost	ccum route milepost (arm)	Numeric (8)
month	accident month	Char (2)
numvehs	number of vehicle	Numeric (8)
object	object struck	Char (2)
rd_char	roadway characteristics	Char (1)
rd_inv	roadway inventory	Char (11)
rdsurf	roadway surface	Char (1)
rodwycls	roadway class	Char (2)
rte_nbr	state route number	Char (3)
sr_adid	state route additional	Char (3)
time	accident time	Char (4)
tot_inj	number of persons injured	Numeric (8)
tot_kill	number of persons killed	Numeric (8)
vehdircde	vehicle direction	Char (1)
vehevent	vehicle movement	Char (1)
weather	weather condition	Char (1)
xrdclass	cross road class type	Char (1)

C. *Roads*

List of variables for Road subfile:

Variable Name	Variable Description	Variable Type
aadt	aver annual daily traffic	Numeric (8)
access	access control type	Char (1)
begmp	begin milepost	Numeric (8)
endmp	calculated ending milepost	Numeric (8)
func_cls	federal func class	Char (2)
lanewid	lane width	Numeric (8)
lshl_typ	left shoulder type	Char (1)
lshldw	left shoulder width	Numeric (8)
med_type	median type	Char (1)
medbarty	median barrier type	Char (2)
medwid	median width	Numeric (8)
medxngty	median crossing type	Char (1)
mvmt	million vehicle miles travelled	Numeric (8)
no_lane	number of lanes	Numeric (8)
rd_qual	related road quality	Char (8)
road_inv	route type id	Char (11)
rodwycls	roadway classification	Char (2)
rshl_typ	right shoulder type	Char (1)
rshldw	right shoulder width	Numeric (8)
rte_nbr	route number	Char (3)
rururb	rural urban	Char (1)
seg_lng	segment length	Numeric (8)
spd_limt	legal speed limit	Numeric (8)
st_func	state function class	Char (2)
surf_typ	surface type	Char (1)

terrain	terrain type	Char (1)
trf_cntl	intersection control type	Char (2)
trkpcts	truck percentage	Numeric (8)

D. *Animals*

List of variables for animal subfile:

Variable Name	Variable Description	Variable Type
age	animal age	Char (1)
animalno	animal number	Numeric (8)
comp_dir	compass direction	Char (2)
day	collection date	Numeric (8)
light	light condition	Char (2)
month	collection month	Numeric (8)
sex	animal sex	Char (1)
species	animal type	Char (2)
weather	weather condition	Char (2)
year	collection year	Numeric (8)