Calibration and Development of Safety Performance Functions for Rural Highway Facilities in Idaho

A Thesis

Presented in Partial Fulfillment of the Requirements for the

Degree of Master of Science

with a

Major in Civil Engineering

in the

College of Graduate Studies

University of Idaho

by

Matthew C. Sipple

December 2014

Major Professor: Ahmed Abdel-Rahim, Ph.D.

Authorization to Submit Thesis

This thesis of Matthew C. Sipple, submitted for the degree of Master of Science with a Major in Civil Engineering and titled "Calibration and Development of Safety Performance Functions for Rural Highway Facilities in Idaho," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: ______________________________ Date: ________________
Ahmed Abdel-Rahim, Ph.D.

Committee Members: ______________________________ Date: ________________
Kevin Chang, Ph.D.

______________________________ Date: ________________
Christopher Williams, Ph.D.

Department Administrator: ______________________________ Date: ________________
Richard J. Nielsen, Ph.D.

Discipline's College Dean: ______________________________ Date: ________________
Larry Stauffer, Ph.D.

Final Approval and Acceptance

Dean of the College of Graduate Studies: ______________________________ Date: ________________
Jie Chen, Ph.D.

Abstract

As the reduction of highway fatalities continues to be a point of emphasis for the U.S. Department of Transportation, many state and local agencies have incorporated safety enhancements into planning, design, operations, and maintenance. In 2010, the American Association of State Highway and Transportation Officials (AASHTO) released the Highway Safety Manual (HSM) as a resource to include safety in decision making for transportation professionals. AASHTO developed general safety performance functions (SPF) to predict crash frequencies based on crash data from various jurisdictions, and suggested that agencies calibrate those SPFs to local crash data and/or develop jurisdiction-specific safety performance functions. At the request of the Idaho Transportation Department (ITD), calibration factors were generated to calibrate the HSM SPFs for rural two-lane, two-way highways and rural stop controlled intersections based on Idaho crash history. In addition, new statewide SPFs were developed for the same rural facilities.

Acknowledgements

I would like to thank Dr. Ahmed Abdel-Rahim for his guidance and support during the course of my graduate program and this thesis. I would like to thank the Idaho Transportation Department and the National Institute for Advance Transportation Technology for funding and facilitating this project, respectively. I would also like to acknowledge Dr. Kevin Chang and Dr. Christopher Williams for their contributions along with Nick Schlotthauer for his help during data collection and compilation.

Table of Contents

List of Figures

List of Tables

List of Equations

Preface

The United States Department of Transportation's (USDOT) national strategy "Toward Zero Deaths" has been a successful program for creating a coordinated safety plan focused on bringing the national highway fatality number to zero. Currently this number tops 33,000 deaths per year (1). As a result of this safety plan, many agencies have begun introducing safety analyses into design, planning, operations and maintenance. The American Association of State Highway and Transportation Officials (AASHTO) developed the Highway Safety Manual (HSM) which contains predictive methods to estimate crash frequencies on different highway facility types. The purpose of this project is to develop more reliable crash prediction methods for rural Idaho highway facility types based on the methods described in the HSM.

The scope of the project includes calibrating existing HSM safety performance functions (SPF) for three facility types based on historic Idaho crash data. These facility types are: 1) rural two-lane, two-way highways, 2) rural 3-leg stop controlled intersections, and 3) rural 4-leg stop controlled intersections. In addition to calibration, new safety performance functions for these facility types are to be developed using negative binomial regression and Idaho crash data.

Chapter 1: Introduction

The Highway Safety Manual (HSM) was released by the American Association of State Highway and Transportation Officials (AASHTO) in 2010 as a resource to improve decision making based on the safety performance of highways. The HSM provides tools that allow transportation professionals to quantify the potential effects on roadway safety as a result of decisions made in planning, design, operations, or maintenance. The HSM describes safety in terms of crash frequency, described in crashes per year, and considers this metric during evaluation and estimation. AASHTO has developed safety performance functions (SPF) for several roadway facility types including rural two-lane, two-way roads, rural multilane highways, urban and suburban arterials, 3-leg and 4-leg stop controlled intersections, and 4-leg signalized intersections. These SPFs predict the expected number of crashes in a single year based on the roadway segment's geometric and traffic conditions (2).

AASHTO developed these SPFs based on the most "complete and consistent available data sets" from around the country, however the predicted number of crashes may vary substantially from jurisdiction to jurisdiction. AASHTO recommends that, where data is available, the HSM SPFs be calibrated based on local crash data. The HSM outlines guidelines for calibrating their SPFs using a single calibration factor as well as a method of developing jurisdiction-specific SPFs based on local crash data (2).

The Idaho Transportation Department (ITD) requested that calibration factors be developed for rural two-lane, two-way highway segments, rural 3-leg stop controlled intersections, and rural 4-leg stop controlled intersections. ITD also requested that new jurisdiction-specific SPFs be developed for the above facility types across the entire state of Idaho.

This thesis outlines the methodology and results of developing calibration factors for HSM SPFs and Idaho-specific SPFs based on Idaho crash data. First, background on the HSM and its method for predicting crashes will be presented. Second, a literature review of similar HSM calibration and SPF development projects will be discussed. Third, the methodology for developing the calibration factors and developing the jurisdiction-specific SPFs is presented along with descriptions of the data used. Finally, the results of the calibrations and recommendations based on these results are presented.

Chapter 2: Background

2.1 Highway Safety Manual

The HSM can be a powerful tool for transportation professionals in several areas of design, planning, operations, and maintenance. The HSM provides a quantitative method for considering facility safety. According to the HSM, its tools have several applications such as, identifying sites with the most potential for safety reduction, conducting economic appraisals of improvements and prioritizing projects, and calculating the safety effects of various design alternatives. In addition to the above, Volume 1 of the HSM outlines several more useful applications (3).

2.2 Safety Performance Functions

The HSM defines safety performance functions as "regression equations that estimate the average crash frequency for a specific site type." (2). The HSM SPFs predict average annual crashes based on roadway geometry and the average annual daily traffic (AADT). The SPFs predict crashes using several base geometric conditions, hereinafter referred to as base conditions. The SPF presented in Part C of the HSM for rural two-lane, two-way highways is given as:

$$N_{spf\ rs} = AADT \times L \times 365 \times 10^{-6} \times e^{(-0.312)} \qquad \text{(Eq. 1)}$$

where $N_{spf\ rs}$ is the total predicted crashes per one year and L is the length of the analysis segment. The SPFs for 3-leg and 4-leg stop controlled intersections are given below in Equation 2 and 3, respectively.

$$N_{spf\ 3ST} = exp[-9.86 + 0.79 \times ln(AADT_{maj}) + 0.49 \times ln(AADT_{min})] \qquad \text{(Eq. 2)}$$

$$N_{spf\ 4ST} = exp[-8.56 + 0.60 \times ln(AADT_{maj}) + 0.61 \times ln(AADT_{min})] \qquad \text{(Eq. 3)}$$

where $AADT_{maj}$ is the average annual daily traffic on the major intersection legs and $AADT_{min}$ is the average annual daily traffic on the minor intersection legs (3). It should be noted that rural stop controlled intersections as identified in the HSM are stop controlled on the minor leg approaches only.

As previously stated, the SPFs predict crash frequency based on several base conditions. Those base conditions for rural two-lane, two-way highways are:

- Lane widths are 12 feet
- Shoulders are paved and 6 feet wide
- No horizontal curve
- Level grade
- Driveway density equal to five driveways per mile
- Absence of centerline rumble strips
- Absence of passing lanes
- Absence of two-way left turn lanes (TWLTL)
- A roadside hazard rating of 3
- Absence of any roadway lighting
- Absence of automated speed enforcement

The base conditions for rural intersections are:

- No skew exists on the minor approaches
- Absence of left-turn lanes on all approaches
- Absence of right-turn lanes on all approaches
- Absence of lighting

When facilities fail to meet these base conditions crash modification factors are used to adjust the predicted crash frequency site conditions. These base conditions are described in more detail in Chapter 10 of the HSM (3).

2.3 Crash Modification Factors (CMF)

Crash modification factors (CMF) are used to adjust crash frequencies predicted by the SPFs based on the actual geometric conditions for a segment. The base conditions give CMF values of 1.00, as this indicates that no adjustment to the predicted crashes is needed due to geometric conditions. If a crash modification factor yields a value of 0.95 then this would indicate that the geometric conditions would provide a 5% reduction in crashes. Similarly, if the CMF produced a value above one then the geometric conditions would suggest an increase in crash frequency. The specific details when determining the CMFs' value is not covered in this report, but can be found in Chapter 10 of the HSM (3).

The final number of predicted crashes after applying the CMFs is given by the following equation:

$$N_{predicted} = N_{spf} \times C \times (CMF_1 \times CMF_2 \times \ldots \times CMF_{12}) \qquad \text{(Eq. 4)}$$

where N_{spf} is the number of crashes per year for the base conditions, C is the calibration factor developed for a specific jurisdiction or geographical area, and CMF_i is the crash modification factor for geometric characteristic; i.e. lane width or shoulder width and surface type (3). The calibration factor, C, is discussed further in this section and the Methodology section.

2.4 Calibrating HSM SPFs

Calibrating the HSM SPFs is recommended in order to develop a more reliable prediction model. The HSM SPFs were developed based on different jurisdictions across the country and calibration can address differences in local factors such as driver populations, crash reporting thresholds and crash reporting system procedures. Calibration requires the following steps as described in the HSM (3):

- Step 1: Identify facility types for which the applicable Part C predictive model is to be calibrated.
- Step 2: Select sites for calibration of the predictive model for each facility type.
- Step 3: Obtain data for each facility type applicable to a specific calibration period.
- Step 4: Apply the applicable Part C predictive model to predict total crash frequency for each site during the calibration period as a whole.
- Step 5: Compute calibration factors for use in Part C predictive model.

Detailed descriptions of each of the five steps can be found in Appendix A of the HSM (3). The above steps were followed during calibration of the HSM SPFs for Idaho and are discussed in the Methodology section.

2.5 Deriving Jurisdiction-Specific SPFs

The HSM also suggests that developing jurisdiction-specific SPFs can enhance the reliability of the predictive methods described therein. It states that calibrated HSM SPFs are a sufficient method of predicting jurisdiction crashes, however when local data is available, such as Idaho data is, jurisdiction-specific SPFs can be a more dependable prediction. The HSM outlines guidelines for developing jurisdiction-specific models that can be used under the methods described in Part C of the HSM. These guidelines were taken directly from Appendix A of the HSM and are as follows (3):

- In preparing the crash data to be used for the development of jurisdiction-specific SPFs, crashes are assigned to roadway segments and intersections following the definitions explained in Section A.2.3 and illustrated in Figure A-1 [of the HSM].
- The jurisdiction-specific SPF should be developed with a statistical technique such as negative binomial regression that accounts for the overdispersion typically found in crash data and quantifies an overdispersion parameter so that the model's predictions can be combined with observed crash frequency data using the Empirical-Bayes Method.
- The jurisdiction-specific SPF should use the same base conditions as the corresponding SPF from Part C [of the HSM] or should be capable of being converted to those base conditions.
- The jurisdiction-specific SPF should include the effects of the following traffic flow volumes: average annual daily traffic for roadway segments and major- and minor-road average annual daily traffic for intersections.
- The jurisdiction-specific SPF for any roadway segment facility type should have a functional form in which predicted average crash frequency is directly proportional to the segment length.

There are two acceptable data forms for developing jurisdiction-specific SPFs; data that is completely in base conditions and data for a broader set of conditions (3). In this report, the data was converted into base conditions and is described in detail in the Methodology section.

Chapter 3: Literature Review

The HSM suggests that their safety performance functions be calibrated using local crash data by determining the variable C, described in the previous section. Several state agencies have completed these calibrations. Xie et al. calibrated the HSM SPFs for rural two-lane, two-way highway, rural multilane highway, and urban and suburban arterial facilities in the tate of Oregon. For rural two-lane, two-way highways the calibration factor in Oregon was found to be 0.74. Xie et al. also found calibration factors for rural 3-leg and 4-leg stop controlled intersections to be 0.32 and 0.31, respectively. These calibration factors indicate that the HSM SPFs overestimates crash frequencies in Oregon (4).

Williamson and Zhou developed two calibration factors to be applied to the existing HSM SPFs and an Illinois-specific SPF. The Illinois-specific SPF was developed previous to Williamson and Zhou's calibration. The results of the calibration produced a factor of 1.40 when calibrating the HSM SPF and 1.58 when calibrating the Illinois SPF. Looking solely at the calibration factors, one can conclude that the calibrated HSM SPF fits the local crash data better than the calibrated Illinois SPF. The closer the calibration factor is to 1.00, the better the models prediction. This is reinforced through statistical analysis completed by Williamson and Zhou (5).

Calibration of the HSM SPFs was completed for Missouri by Sun et al. Sun et al. calibrated the SPF for rural two-lane, two-way highways using 196 segments with 100.7 crashes per year over all segments resulting in a calibration factor of 0.82 for Missouri (6).

The HSM states that jurisdiction-specific SPFs can be a more reliable predictive model (2). Young and Park completed a study comparing the performance of uncalibrated and calibrated HSM safety performance functions with jurisdiction-specific SPFs for intersections types in Regina, Saskatchewan. The results of this study proved that jurisdiction-specific SPFs performed the best compared to the uncalibrated and calibrated SPFs (7). The results of Young and Park's study is generally believed to hold true for all facility types including rural two-lane, two-way highways. Understanding the benefits, some local agencies have developed jurisdiction-specific SPFs along with calibrating the existing HSM SPFs. Brimley et al. developed a single factor of 1.16 to calibrate the existing HSM rural two-lane two-way roadway SPF based on crash data for the State of Utah. Brimley et al. also developed four jurisdiction-specific (entire State of Utah) SPFs using the variables and CMFs presented in Chapters 10 through 12 of the HSM with the additional variables including shoulder rumble strips, percent single-unit trucks, and percent multiple-unit trucks (8).

Srinivasan and Carter calibrated and developed safety performance functions for North Carolina. Calibration factors were found for many facility types including rural 3- and 4-leg intersections. Calibration factors were developed for three analysis years (2007, 2008, and 2009) and a three year average was calculated. For 3- and 4-leg stop controlled intersections, Srinivasan and Carter found calibration factors of 0.57 and 0.68, respectively. Calibration factors were also included for other facility types. In addition to the calibration factors, North Carolina-specific SPFs were developed. Srinivasan and Carter estimated SPFs using negative binomial regression for sixteen highway facility types. The Freeman-Tukey R^2 and the Pseudo R^2 goodness-of-fit tests were conducted on the North Carolina SPFs to describe their fit to local data (9).

Mehta and Lou developed calibration factors for two-lane, two-way rural highways and four-lane divided highways for Alabama. The calibration factors were calculated at 1.39 for two-lane, two-way highways and 1.10 for four-lane divided highways using the calibration method presented in the HSM. Along with calibration factors, Mehta and Lou developed four jurisdiction-specific SPFs based on a number of input variables along with several formula types using negative binomial regression. Mehta and Lou used several goodness-of-fit measures to compare their new models. These measures include log-likelihood maximization, Akaike information criterion (AIC), Mean Absolute Deviation (MAD), Mean Prediction Bias (MPB), and Mean Square Prediction Error (MSPE). Mehta and Lou found that a model with five explanatory variables; AADT, segment length, lane width, speed, and analysis year, produced the best prediction of Alabama crash frequencies. The authors describe overfitting of data which can result from including too many input variables in a regression model and they describe how to statistically discourage overfitting (10).

Schrock and Wang developed calibration factors for rural stop controlled intersections along with rural highway segments in Kansas. In addition to the calibration factors, Kansas developed a crash prediction model for rural two-lane highway segments. The results of the calibration of the existing HSM SPFs yielded calibration factors of 0.28 for 3-leg stop controlled intersections and 0.19 for 4-leg stop controlled intersections. The results also produced a calibration factor of 1.48 statewide for rural two-lane highway segments. Several statistical tests were completed to compare the different models' fit to the data. Similar to Mehta and Lou, these tests included Mean Prediction Bias (MPB), and Mean Absolute Deviation (MAD) with the addition of Pearson's *R*. After comparing the models, Schrock and Wang found that the Kansas-specific SPF did not perform as well as the calibrated HSM SPFs. As part of their conclusion, they recommended that the Kansas SPF not be implemented (11).

Haas and Gosse developed several state-specific safety performance functions for the Virginia Department of Transportation (VDOT). Virginia was divided into five different VDOT operations districts and these five districts were combined into three districts that Hass and Gosse determined to be similar in driver and roadway characteristics. SPFs were developed for each of these three regions separately to more accurately model the state's crash frequencies. In addition to the three regions, SPFs were developed for the entire state of Virginia. The models that Hass and Gosse developed were for use in the Federal Highway Administration's SafetyAnalyst software. "SafetyAnalyst incorporates the HSM safety management approaches into computerized analytical tools for guiding the decision-making process." (12). Even though the SPFs were developed for use in SafetyAnalyst, SafetyAnalyst was developed using the HSM safety management approaches. Haas and Gosse compared their models using two statistical goodness-of-fit measures; the coefficient of determination and the Freeman-Tukey R^2. After statistical and graphical comparison, Haas and Gosse recommended VDOT apply Virginia-specific safety performance functions when using SafetyAnalyst. Haas and Gosse also recommended that using the region-specific SPFs within Virginia gave even more reliability to the crash frequency predictions (12). The correlation between SafetyAnalyst and the HSM methods suggests that Hass and Gosse's conclusion should hold true to jurisdiction-specific SPFs developed for use with the HSM methods.

Qin et al. developed jurisdiction-specific SPFs and calibrated the existing HSM SPF for rural two-lane, two-way highway segments in South Dakota. The results of the calibration of the HSM SPF yielded a calibration factor of 1.537 (13). The HSM requires that only two possible types of data should be used during the development of jurisdiction-specific SPFs, described previously (3). Qin et al. developed SPFs by dividing their local data into two types. One was data that met the HSM base conditions and the other was data that met new South Dakota specific base conditions. Four predictive models were developed from the data, one from HSM base condition data, one from South Dakota base conditions, and two from the full data. Negative binomial regression was completed for the separated data sets, and Poisson regression was completed using the full data. In addition, calibration factors were calculated for the South Dakota specific SPFs as a comparison technique. Qin et al. concluded after comparing the calibration factors, the correlation coefficient, and the Mean Absolute Deviation, that a linear Poisson regression model from the full crash data was the best prediction of South Dakota crashes (13).

Russo et al. developed safety performance functions based on the methodologies presented in the HSM for rural two-lane highways in Southern Italy. Three safety performance functions were developed for injury crashes only, deaths only, and injuries plus deaths. These SPFs were validated using residuals analysis. Russo et al. developed these SPFs for only segments that met the base conditions described in the HSM. Using the remaining roadway segments that did not meet the HSM's base conditions, CMFs were calculated in order to convert the segments into base conditions. Russo et al. concluded that the SPF developed to predict injury plus death crashes can replace the other two SPFs. Using an ANOVA test, it was found that the more comprehensive SPF predicted the same percent of death crashes and injury crashes as the more specific SPFs for each crash type (14).

Wang and Abdel-Aty developed SPFs for 3- and 4-leg stop controlled intersections in Florida based on data from 190 intersections. The SPFs were developed using traffic and geometric characteristics only from the major road. Regression was completed using major road AADT, number of intersection legs and number of major through lanes as explanatory variables. Wang and Abdel-Aty found that only major AADT was statistically significant as an explanatory variable. Wang and Abdel-Aty found that developing SPFs based on major AADT only is not an ideal solution (15).

Chen et al. suggested that the HSM calibration process may not be a completely adequate method of transferring SPF from jurisdiction to jurisdiction. Chen et al. tested the transferability of safety performance functions through Bayesian model averaging as a more reliable way to transfer these SPFs. Bayesian model averaging was not considered in this report, and therefore Chen et al. was used exclusively as a reference for goodness-of-fit procedures. Their analysis used Pearson's r, Mean Average Deviation (MAD), Mean Prediction Bias (MPB), and Mean Square Prediction Error (MSPE) to compare the fit of their models after applying Bayesian averaging (16).

Previous to the release of the HSM, many researchers were developing regression models to assist in crash prediction. In 1996, fourteen years prior to the release of the HSM, Poch and Mannering looked at developing predictive models for intersections in Bellevue, Washington. At the time, highway and freeway segments were the only facility types considered for developing predictive models. Poch and Mannering developed negative binomial estimates for all crash, rear-end crash, angle crash, and approach-turn crash types. They found that negative binomial can be useful when identifying traffic and geometric characteristics that effect crash frequencies. Poch and Mannering also suggested that having these predictive models could be very useful when analyzing the effects of intersection improvements (17).

Xie and Zhang suggested that crash prediction models' explanatory variables, such as AADT, may not behave linearly in nature. They suggested that negative binomial generalized additive models (NBGAM) may be a more reliable model development technique than the negative binomial generalized linear model (NBGLM). Xie and Zhang developed a linear NBGLM, a logarithmic NBGLM, and a NBGAM for crash data from 59 3-leg signalized intersections in Toronto, Canada and compared them. They used the Mean Average Deviation (MAD), Mean Square Prediction Error (MSPE), and the Akaike information criterion (AIC) to test the models' fit and predictive quality for the Toronto crash data. Xie and Zhang concluded that the NBGAM fit the data the best with the logarithmic NBGLM performing almost as well, while both outperformed the linear NBGLM (18).

The following tables are a summary of the calibration factors developed by each state. The tables include the calibration factor applied to the existing HSM SPF along with the calibration factors applied to state-specific SPFs, if applicable. Table 3.1 shows the calibration factors for two-lane, two-way highways segments and Tables 3.2 and 3.3 show calibration factors for 3- and 4-leg stop controlled intersections, respectively.

Table 3.1: Calibration Factors for Two-lane Highway Segments

State	**HSM SPF**	**State-Specific SPF**
Oregon	0.74	n/a
Illinois	1.40	1.58
Missouri	0.82	n/a
Utah	1.16	n/a
Alabama	1.39	n/a
Kansas	1.48	n/a
South Dakota	1.54	n/a

Table 3.2: Calibration Factors for 3-Leg Intersections

Calibration Factors for 3-Leg Intersections		
State	**HSM SPF**	**State-Specific SPF**
Oregon	0.31	n/a
North Carolina	0.57	n/a
Kansas	0.28	n/a

Table 3.3: Calibration Factors for 4-Leg Intersections

Calibration Factors for 4-Leg Intersections		
State	**HSM SPF**	**State-Specific SPF**
Oregon	0.32	n/a
North Carolina	0.68	n/a
Kansas	0.19	n/a

Chapter 4: Methodology

4.1 Data Collection

The HSM has extensive data requirements in order for the predictive methods to be the most reliable. These data requirements can be found in Appendix A in Volume 1 of the HSM. Data describing roadway geometry, traffic conditions, and crash data were found using several different sources. ITD provided Microsoft Excel (EXCEL) files that included AADT and many of the geometric conditions required as inputs of the HSM SPFs and their CMFs. ITD also provides an online video log system, hereinafter referred to as Pathways, which allows users to visually inspect most state and federal highways within Idaho boundaries; see Figure 4.1 for user interface (19). Pathways was used to visually gather geometric information that was not provided by the EXCEL files and also used to confirm geometric information provided by ITD's EXCEL Files. Geometric data and their respective sources are described below:

- Microsoft Excel: AADT, Major and Minor AADT for Intersection, Segment Length, Lane Width, Shoulder Type and Width, Passing Lanes, Horizontal Curve Details, Segment Grade
- Pathways: Two-Way Left-Turn Lanes, Driveway Density, Center-lane Rumble Strips, Roadway Lighting, Intersection Locations, Intersection Skews, Intersection Right/Left Turn Lanes

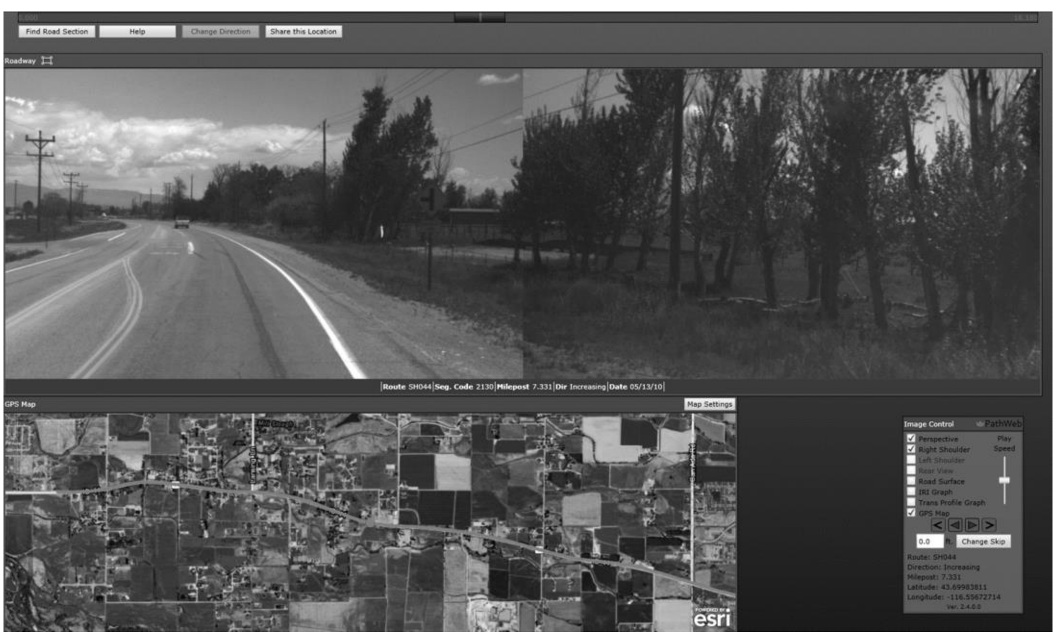

Figure 4.1: User-Interface for Pathways Video Log Tool

Crash data was gathered using ITD's online Web Crash Analysis Reporting System (WebCARS). WebCARS allows the user to search any segment of Idaho state highways and find all crashes within a selected milepost range. Crash data for the selected roadway segments and intersections was averaged between 2003 and 2012. Ten years of crash data were used in order to find a representative number of crashes per year. Crash data is classified in WebCARS as non-junction, in intersection, or intersection related (20). Crash data for two-lane, two-way highway segments were only included if it was classified as "non-junction". Similarly, crash data for the intersections were only included if classified as "in intersection" or "intersection related". It should be noted that crash data and its respective classification is at the discretion of the reporting police officer. It was assumed that crash data were reported correctly in their entirety. All reportable crash types were used when compiling crash data; Fatal, Injury A, Injury B, Injury C, and Property Damage only (20).

Several assumptions were made in order to simplify data collection and limit subjective data inputs. For example, the CMF for roadside hazard rating (RHR) is on a scale of one to seven, with one representing roadside conditions having no or minimal hazard and seven having very dangerous roadside conditions. The model user must select a value based on his/her observation of these conditions. In order to remove this judgment, all segments' RHR were set as the base condition value, 3. The HSM recommends setting RHR equal to 3 when no data is available. Another assumption was that no segment of Idaho state or federal highways were subject to permanent Automated Speed Enforcement, therefore the base condition for the corresponding CMF was used exclusively.

The HSM SPFs predict crashes for a specific time period, referred to as the analysis year (2). In general, all input data for the SPFs should be from the same analysis year, however some assumptions had to be made for this study. Geometric data provided were representative of conditions in 2010. AADT was provided for 2012, which was the most recent and complete year available. In order to address the time period differences, it was assumed that the roadway geometric conditions were unchanged between 2010 and 2013 (the start of this project) other than maintenance related improvements. Regardless of the time period differences, the data are believed to be representative of the most current conditions for rural Idaho roadways and intersections.

4.2 Site Selection

Sites for analysis were selected randomly from the ITD provided EXCEL files as well as utilizing concurrent project work looking at rural two-lane state highways. The ITD geometric data was compiled and separated into homogenous roadway segments. Sites were randomly selected using systematic sampling. Additional sites were added from data collected through a concurrent project at the University of Idaho. These sites were long, continuous segments of State Highway (SH) 7, SH 39, SH 44, SH 50, SH 52, SH 55, SH 75, and SH 81. The long segments were divided into short homogenous segments for use in this project. These sites cannot be considered randomly selected but are still believed to be representative of the different roadway types across the state of Idaho. Some site trimming was necessary after reviewing the segments from the concurrent project. Site trimming included removal of segments with missing or unavailable data and segments of rural two-lane highways that passed through urban areas.

Using Pathways, intersections sites were selected during the data collection process for the highway segments. Once intersections were selected, the required geometric data was obtained using Pathways. Intersection skew, an input for intersection CMFs, was visually approximated using Pathways GPS Map, seen at the bottom of Figure 4.1.

Table 4.1 gives a summary of the highway segments and Table 4.2 give a summary of intersection characteristics used during calibration of the HSM SPFs and developing regression models.

Table 4.1: Summary of Roadway Segment Data

Two-lane, Two-Way Highway Segments			
Average Length (miles)	0.507	No. of Curved Segments	210
Minimum Length	0.004	No. of Straight Segments	237
Maximum Length	7.769	Total Segments	447
Average AADT (veh/day)	3741		
Minimum AADT	250		
Maximum AADT	25000		

Table 4.2: Summary of Intersection Data

3-leg Stop Controlled Intersection		**4-leg Stop Controlled Intersection**	
Average Major AADT (veh/day)	4619	Average Major AADT (veh/day)	6021
Minimum Major AADT	390	Minimum Major AADT	210
Maximum Major AADT	25000	Maximum Major AADT	32500
Average Minor AADT (veh/day)	277	Average Minor AADT (veh/day)	332
Minimum Minor AADT	60	Minimum Minor AADT	40
Maximum Minor AADT	1400	Maximum Minor AADT	1360
Total Intersections	43	Total Intersections	41

4.3 Model Validation

After all the data was collected, the segments were randomly divided 70/30 for fitting the models and testing the predictive capabilities of those models, respectively. The 70 percent was randomly sampled ten times from the full data set to test the variability in each of the calculated parameters, i.e. the calibration factor and the regression coefficients. This was completed for only two-lane highway segments as a test. The results of looking at several random samples showed the averages of each parameter converging toward the parameter values for the full data set. As a result, it was deemed unnecessary to look average parameter values using multiple samples.

Calibration of the existing HSM was completed on a single sample of 70 percent of data as the regression. Once calibration of the HSM SPFs and fitting of the regression equations was complete, statistical analyses were completed to compare the reliability each of the models based on how well they predicted crash frequency as compared to Idaho data. These statistical analyses included the Pearson's *R*, Mean Squared Prediction Error (MSPE), and the Freeman-Tukey R^2 goodness-of-fit test.

4.4 Calibration of HSM SPFs

The HSM suggests that when data is available, SPFs should be calibrated based on jurisdiction or a geographic region's crash data. HSM recommends that the minimum sample size be between 30 and 50 segments. For this study, 313 segments where used for rural two-lane, two-way highways along with 79 and 85 segments for the 3-leg and 4-leg rural intersections, respectively. The calibration factor is described by the following equation:

$$C = \frac{\sum N_{Observed}}{\sum N_{Predicted}} \qquad \text{(Eq. 5)}$$

where $\Sigma N_{Observed}$ is the total number of crashes observed for all selected segments and $\Sigma N_{Predicted}$ is the total number of crashes predicted by the HSM SPF for the same segments. Calibration factors were found for each of the HSM SPFs described in the Background section.

4.5 Developing Jurisdiction-Specific SPFs

Negative binomial regression was completed using R i386 v3.1.1 (R), a statistical analysis software package (21). Before using R, data modifications were required. The observed crashes were converted into base conditions to insure the new SPFs could be applied with the HSM provided CMFs. Equation 6 shows how the observed crashes were modified to be base condition crash frequencies, hereinafter referred to as Method 1:

$$n_{modified} = n_{observed} \div (CMF_{1r} \times CMF_{2r} \times \ldots \times CMF_{12r}) \qquad \text{(Eq. 6)}$$

where $n_{modified}$ is the modified crash frequency for an individual segment, and $n_{observed}$ is the observed crash frequency for an individual segment. Equation 6 was applied to each of the 313 highways segments and the 164 intersections. Equation 6 was used when performing regression using both AADT and segment length (L) as explanatory variables, however the observed crashes were modified further by dividing the individual segment length in order to perform regression using only AADT. Equation 7 gives base condition crash frequency in observed crashes per mile per year and is shown below, hereinafter referred to as Method 2.

$$n_{modified} = n_{observed} \div (L \times \mathrm{CMF}_{1r} \times CMF_{2r} \times \ldots \times CMF_{12r}) \qquad \text{(Eq. 7)}$$

Along with modifying the response variable, crash frequency, the explanatory variables were also modified before regression could be completed. The natural log of AADT was used for all regression analyses and natural log of the segment length was only used when applying Equation 6 in order to be consistent with the HSM format. This produced SPFs in the following formats:

$$N_{Idaho\ rs} = \exp[\beta_0 + \beta_1 \times ln(AADT) + \beta_2 \times ln(L)] \qquad \text{(Eq. 8)}$$

$$N_{Idaho\ rs} = L \times \exp[\beta_0 + \beta_1 \times ln(AADT)] \qquad \text{(Eq. 9)}$$

where β_0, β_1, and β_2 are regressions coefficients and $N_{Idaho\ rs}$ is the Idaho-specific SPF for the rural two-lane, two-way highways. Equation 8 is the regression format that corresponds to the modifications described by Method 1 and Equation 9 corresponds to modification describes by Method 2.

Observed crashes were modified using Method 1 during regression for 3-leg and 4-leg stopped controlled intersections. From these modified observed crashes, the SPF for the two intersection types are formatted as such:

$$N_{Idaho\ 3ST/4ST} = \exp[\beta_0 + \beta_1 \times ln(AADT_{maj}) + \beta_2 \times ln(AADT_{min})] \quad \text{(Eq. 10)}$$

where β_0, β_1, and β_2 are regression coefficients. Notice that the natural logs of major and minor AADT are the explanatory variables, this is consistent with the HSM SPF for the same facility types. β_0, β_1, and β_2 were found for each intersection type and can be found in the Results section.

Chapter 5: Results and Analyses

5.1 HSM SPFs Calibration Results

Based on the methods presented in the HSM, calibration factors for the entire State of Idaho were developed. Roughly 227 miles of state and federal highways were divided into 447 homogenous segments and then analyzed using the HSM Part C method for predicting crash frequency. From these 447 road segments, 313 segments were randomly selected to be used for the calibration and regression. 79 3-leg stop controlled intersections and 85 4-leg stop controlled intersections were used for the analyses.

The HSM SPF for rural two-lane, two-way highway segments estimated 168.95 crashes per year over all selected road segments. The observed total crashes for the same segments totaled 188.40 crashes per year. Using Equation 5, the calibration factor for the state of Idaho's rural two-lane highways was calculated to be 1.115. This indicates that the HSM SPFs under predict total crash frequency by approximately 12%.

The HSM predicted 52.10 crashes per year for 3-leg rural intersections and 69.40 crashes per year for 4-leg intersections. There were only 13.20 observed crashes per year for 3-leg stop controlled intersections, producing a calibration factor of 0.253. For 4-leg stop controlled intersections there were 29.60 crashes per year observed at the selected intersections producing a calibration factor of 0.427. This indicated that the HSM over estimates the crashes per year by 74.7% for 3-leg intersections and 57.3% for 4-leg intersections. The calibration factors for all SPFs can be seen in Table 5.1.

Table 5.1: Calibration Factors for Idaho Rural Highway Facilities

Facility Type	Predicted Crashes	Observed Crashes	Calibration Factor
Two-lane, Two-way highways	168.95	188.40	1.115
3-leg Stop Controlled Intersections	52.10	13.20	0.253
4-leg Stop Controlled Intersections	69.40	29.60	0.427

5.2 Idaho-Specific SPFs Results

Negative binomial regression was completed using R for rural two-lane, two-way highways, rural 3-leg stop controlled intersections, and rural 4-leg stop controlled intersections. The output from R gives regression coefficients for intercept and for any of the explanatory variables (21). These coefficients are represented by β_0, β_1, and β_2 as described in the Methodology section.

Table 5.2 shows the regression coefficients produced by R for the two regression analyses for two-lane highways and the two models for the 3-leg and 4-leg stop controlled intersections. Notice that β_2 was not found for the first regression model for two-lane highways because the AADT was the only variable considered during regression. AADT and segment length were found to be statistically significant explanatory variables during regression for two-lane highway segments, however major and minor road AADT were not significant explanatory variables for 3-leg intersections and minor AADT was not statistically significant for 4-leg intersections. This is believed to be a result of a small sample size for intersection sites and the amount of zero observed crash frequencies.

Table 5.2: Regression Coefficients for Idaho Rural Highway Facilities

Facility Type	β_0	β_1	β_2
Two-lane, Two-way highways #1	-5.7999	0.7371	0.8938
Two-lane, Two-way highways #2	-5.7853	0.7501	n/a
3-leg Stop Controlled Intersections	-6.1502	0.0966	0.6969
4-leg Stop Controlled Intersections	-8.6336	0.8966	0.0458

Equations 11 – 14 are the final Idaho-specific safety performance functions for the two-lane highways using Methods 1 and 2, 3-leg stop controlled intersection and 4-leg stop controlled intersection, respectively. Equations 11 and 12 are simplified forms of Equations 8 and 9, respectively.

$$N_{Idaho\ rs} = L^{0.8938} \times AADT^{0.7371} \times e^{(-5.7999)} \qquad \text{(Eq. 11)}$$

$$N_{Idaho\ rs} = L \times AADT^{0.7501} \times e^{(-5.7853)} \qquad \text{(Eq. 12)}$$

$$N_{Idaho\ 3ST} = \exp[-6.1502 + 0.0966 \times ln(AADT_{maj}) + 0.6969 \times ln(AADT_{min})] \qquad \text{(Eq. 13)}$$

$$N_{Idaho\ 4ST} = \exp[-8.6336 + 0.8966 \times ln(AADT_{maj}) + 0.0458 \times ln(AADT_{min})] \qquad \text{(Eq. 14)}$$

Equations 11 – 14 were compared to the observed crash data along with the uncalibrated and calibrated HSM SPFs to find which model best describes Idaho crash behavior using statistical goodness-of-fit tests. For each of the facility types, statistical analyses were conducted to compare each of the prediction models, uncalibrated and calibrated HSM SPFs and the regression models. These statistical analyses included the Pearson's *R*, Mean Squared Prediction Error (MSPE), and the Freeman-Tukey R^2 goodness-of-fit test.

Pearson's R, or Pearson product moment correlation coefficient, is a measure of the linearity between the observed data and the predicted. A value of 1 would indicated a perfect correlation and a value of zero would suggest that the predictive model has no correlation with the observed data (11, 16). Pearson's R is given by the following equation:

$$r = \frac{\sum_{i=1}^{n}(y_i-\bar{y})\cdot(\widehat{y_i}-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2\cdot\sum_{i=1}^{n}(\widehat{y_i}-\bar{\hat{y}})^2}} \quad \text{(Eq. 15)}$$

where n is the sample size, $\bar{y}$ and $\bar{\hat{y}}$ are the means of the observed crashes and predicted crashes, respectively, and y_i and $\widehat{y_i}$ are the observed and predicted values at site i, respectively.

The MSPE is the sum of the squared difference of the observed crashes and predicted crashes divided by the number of sites (10, 16). MSPE is calculated using the following equation:

$$MSPE = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2 \quad \text{(Eq. 16)}$$

The Freeman-Tukey R^2 goodness-of-fit test can be used as a surrogate to traditional R^2, or coefficient of determination, tests. Since negative binomial regression minimizes log likelihood values when fitting the data, traditional R^2 values are rarely used (9). SPFs developed for North Carolina and Virginia were tested using the Freeman-Tukey R^2 (9, 10). Values of Freeman-Tukey R^2 closer to one will represent a better fit is calculated as follows:

$$R^2 = 1 - \sum_{i=1}^{n}\widehat{e_i}^2 / \sum_{i=1}^{n}(f_i - f_m)^2 \quad \text{(Eq. 17)}$$

where $\widehat{e_i}$ is the residual for site i, f_i is the Freeman-Tukey transformative statistic for site i, and f_m is the mean of the Freeman-Tukey transformative statistics for all sites. The equations for the residuals and the Freeman-Tukey transformative statistic can be found in Srinivasan and Carter (9).

Table 5.3 shows the results of the statistical analysis for rural two-lane, two-way highway segments. Idaho-specific SPF #1 represents the regression model that considered AADT and segment length during the analysis (Method 1) while Idaho-specific SPF #2 represents the regression that only considered AADT (Method 2). From the results of the statistical analysis we can conclude that the Idaho-specific SPFs using Method 1 best represents the crash behavior in the state of Idaho.

Table 5.3: Statistical Comparison for Rural Highway Segments

Prediction Method	**Pearson's R**		**MSPE**		**Freeman-Tukey R^2**	
	Fitted	**Predicted**	**Fitted**	**Predicted**	**Fitted**	**Predicted**
Uncalibrated HSM SPF	0.718	0.531	0.796	0.764	0.473	0.264
Calibrated HSM SPF	0.718	0.531	0.824	0.854	0.480	0.243
Idaho-specific SPF #1	0.739	0.593	0.747	0.666	0.527	0.332
Idaho-specific SPF #2	0.746	0.594	0.817	0.819	0.493	0.272

Based on the criteria for each of the statistical tests, we can see that the Idaho-specific SPF #1 has the second highest Pearson's r value, the lowest MSPE value, and the highest Freeman-Tukey R^2 values. This indicated that it is the best prediction of Idaho crash frequencies.

Figure 5.1 shows the four predictive models plotted with the observed crashes used for testing the predictive quality of each model, approximately 30 percent of the total sites. The crash frequency was plotted with only AADT to be consistent with the HSM. Graphically, the Idaho-specific SPF #1 again best fits the data.

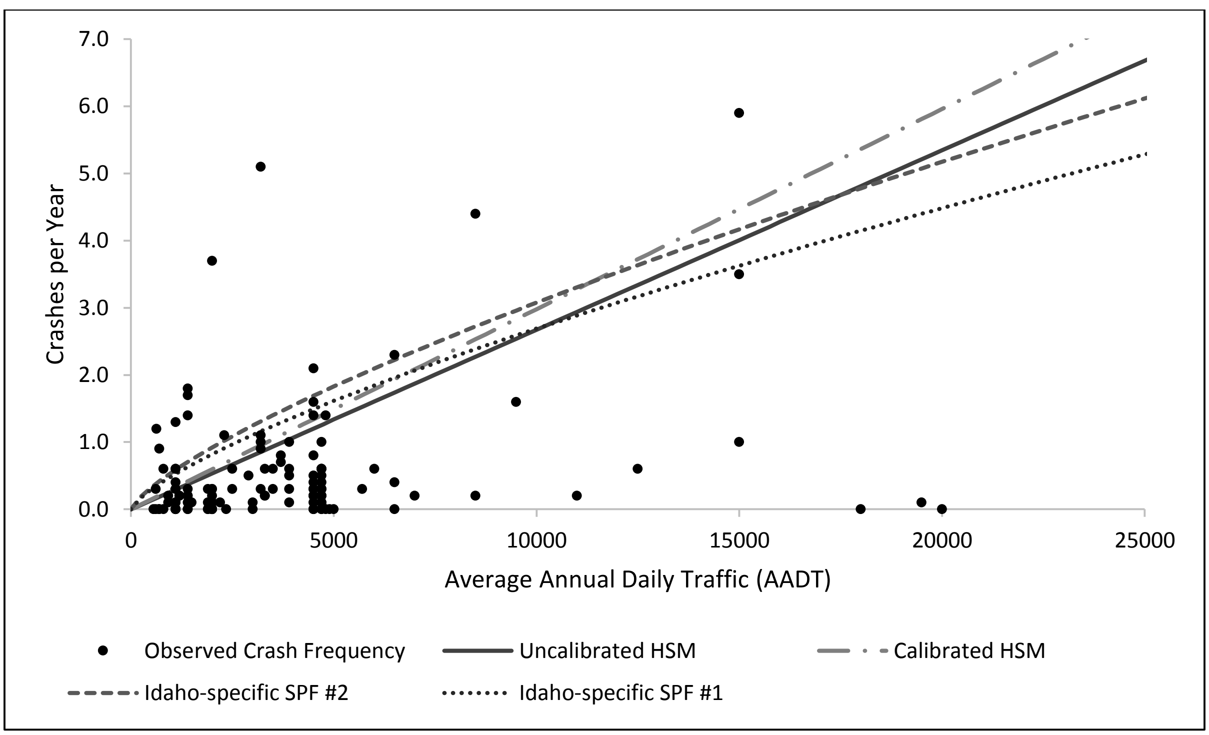

Figure 5.1: Crash Prediction Models for Two-lane, Two-way Highway Segments

Table 5.4 and Table 5.5 show the results of the statistical analysis for rural 3-leg stop controlled intersections and 4-leg stop controlled intersections, respectively. These results indicate that the Idaho-specific SPF best represents the crash behavior for 3-leg intersections and indicates the calibrated HSM SPF is best for 4-leg intersections.

Table 5.4: Statistical Comparison for Rural 3-Leg Stop Controlled Intersections

Prediction Method	**Pearson's R**		**MSPE**		**Freeman-Tukey R^2**	
	Fitted	**Predicted**	**Fitted**	**Predicted**	**Fitted**	**Predicted**
Uncalibrated HSM SPF	0.166	0.350	0.688	0.260	-2.350	-1.252
Calibrated HSM SPF	0.166	0.350	0.109	0.059	-0.090	0.102
Idaho-specific SPF	0.211	0.477	0.102	0.061	-0.018	0.110

Table 5.5: Statistical Comparison for Rural 4-Leg Stop Controlled Intersections

Prediction Method	**Pearson's R**		**MSPE**		**Freeman-Tukey R^2**	
	Fitted	**Predicted**	**Fitted**	**Predicted**	**Fitted**	**Predicted**
Uncalibrated HSM SPF	0.014	-0.096	1.368	1.024	-0.943	-4.398
Calibrated HSM SPF	0.014	-0.096	0.800	0.239	-0.141	-1.053
Idaho-specific SPF	0.063	-0.150	0.861	0.570	-0.350	-2.521

The models are plotted with a scatter plot of the observed crash data used to check the predictive quality of each model. Again, the Idaho-specific SPF best fits the observed crash data for 3-leg intersections, however the calibrated HSM SPF best fits the data the 4-leg intersections. Similar to the two-lane highway segments, the statistical analysis is supported by the graphical conclusions. Figures 5.2 and 5.3 show the predictive models for 3- and 4-leg stop controlled intersection, respectively.

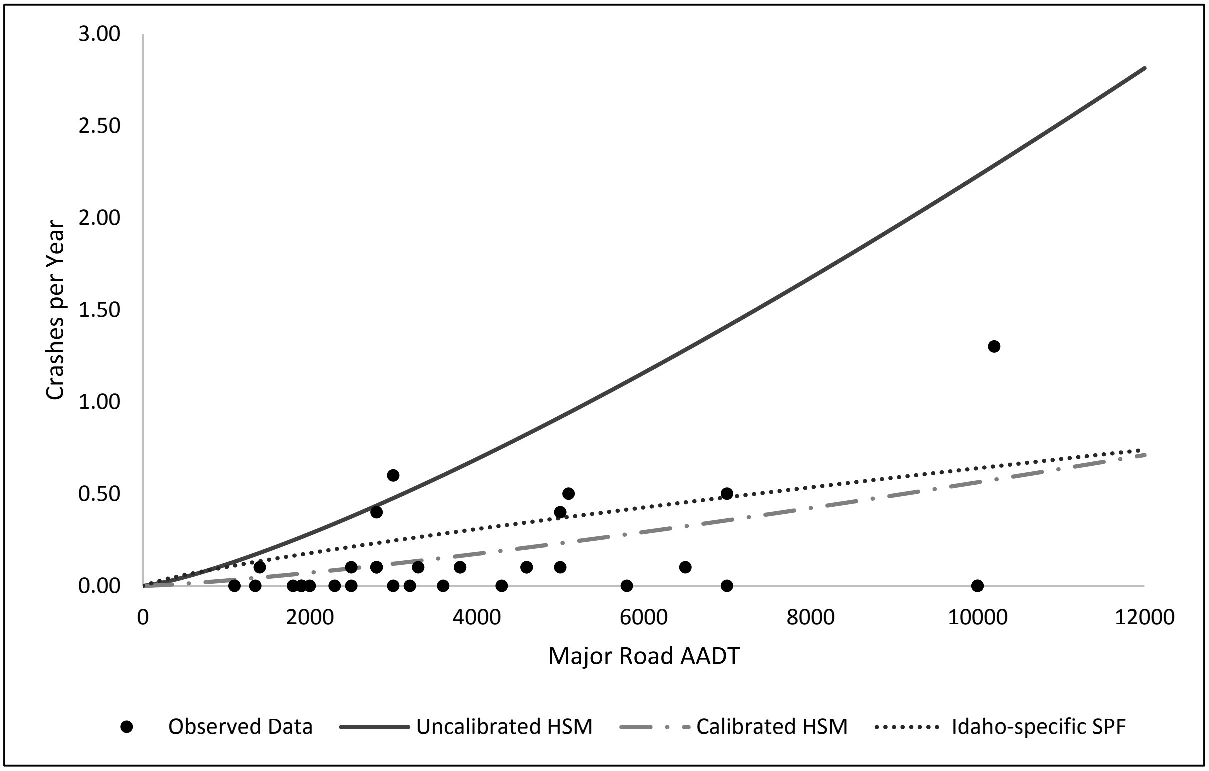

Figure 5.2: Crash Prediction Models for 3-leg Stop Controlled Intersections

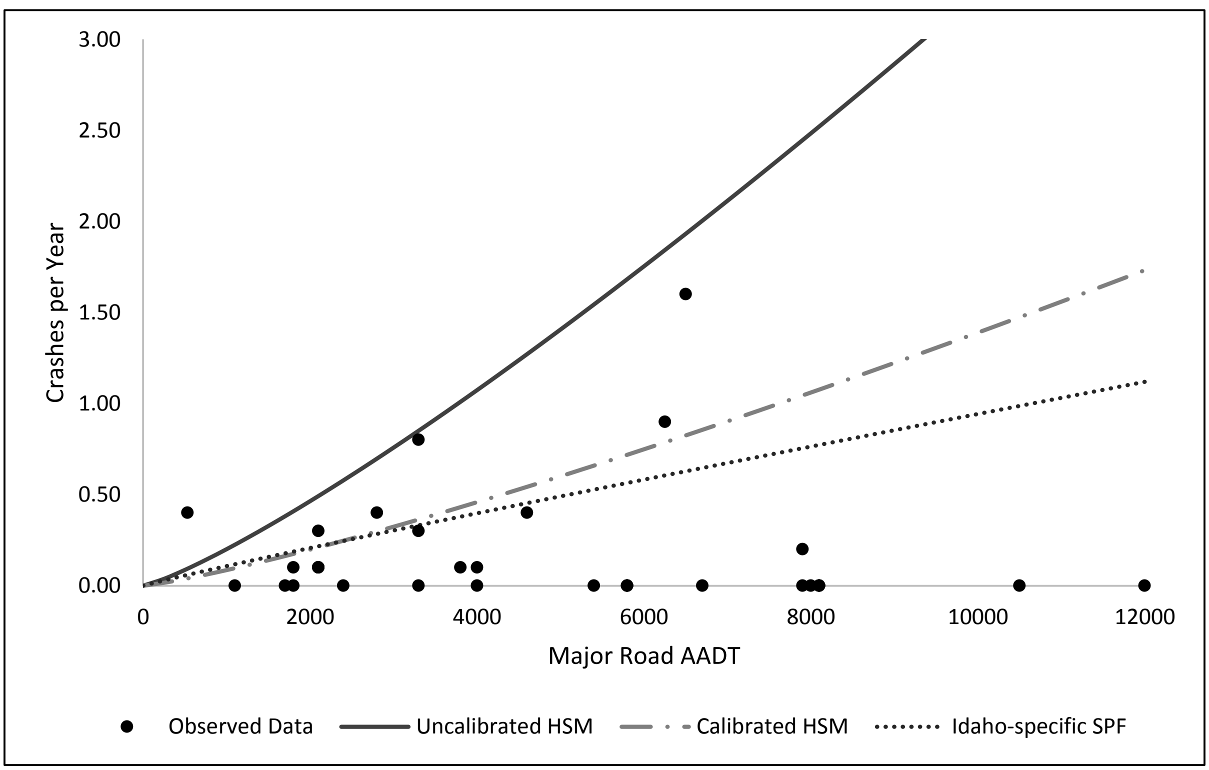

Figure 5.3: Crash Prediction Models for 4-leg Stop Controlled Intersections

Chapter 6: Conclusion

6.1 Discussion and Recommendations

Crash prediction models are becoming more prominent for use during design, planning, operations and maintenance. AASHTO released the Highway Safety Manual as a tool for predicting crash frequency and severity for different highway facility types. Crash frequencies are predicted using safety performance functions which were developed from the most current available crash data across several jurisdictions. The HSM recommends for a more reliable prediction, its models should be calibrated based on local jurisdictions' crash data. It also recommends that, when data is available, jurisdiction-specific SPFs be developed. The Idaho Transportation Department requested calibration of HSM SPFs and development of Idaho-specific SPFs for rural two-lane, two-way highway segment, and rural 3- and 4-leg stop controlled intersections.

Calibration and development of local SPFs was completed using methods presented in the HSM. Negative binomial regression was completed to create the Idaho-specific SPFs. The Idaho-specific SPFs were compared to the uncalibrated and calibrated HSM SPFs using statistical analysis to check reliability of the prediction models. Pearson's *R*, Mean Square Prediction Error, and the Freeman-Tukey R^2 were used for comparison along with graphical inspection. The Idaho-specific SPFs for 3- and 4-leg stop controlled intersections were found to have statistically insignificant explanatory variables, i.e. major and minor road AADT for 3-leg intersections and minor road AADT for 4-leg intersections. The statistical insignificance implies that the regression model could not find a correlation between the explanatory variables and the response variable. In other words, there was no statistical correlation between minor road AADT and crash frequency. This could be a result of small sample size or the number of sites with zero crashes per year. Even though the explanatory variables for the intersections were not all statistically significant, more weight was given to the goodness of fit tests when recommending the models for use by ITD.

As a result of this work, we recommend that the Idaho Transportation Department use the Idaho-specific SPFs for predicting crash frequencies on rural two-lane, two-way state highways and 3-leg intersections. We also recommend that ITD use the calibrated HSM SPF for 4-leg stop controlled intersections. The results show that these are the most reliable prediction methods for rural Idaho facilities.

6.2 Future Work

Additional work could be completed to further improve the reliability of the Idaho prediction models. Due to the statistical insignificance of major and minor AADT for both types of intersections, sample size should be increased to increase the degrees of freedom along with considering different regression types. Increasing the degrees of freedom can increase the likelihood of producing statistically significant explanatory variables. We believe that a larger sample size could yield an Idaho-specific SPF with statistically significant explanatory variables. In addition, these explanatory variables could be statistically significant if a zero-inflated negative binomial regression were completed. The amount of zero crash frequencies observed at these intersections could require the use of zero-inflated negative binomial regression.

Another possible improvement to the prediction models' reliability would be to further divide the jurisdiction. Idaho has two distinct geographical regions within its borders; northern and central Idaho is mountainous and southern Idaho is predominately high desert. Similar to Haas and Gosse for Virginia, SPFs could be developed for these two geographical regions separately.

The HSM suggests that local CMFs be developed if agencies believe there may be a significant effect to crash frequencies. Idaho has many hazards on its highways including deer and elk. Animal populations may have a considerable effect on crash frequencies. Additional work could be done to test this correlation and a CMF based on animal population could be included to further improve reliability.

For Idaho SPFs, AADT and segment length were chosen as explanatory variables for the regression in order to be consistent with the HSM SPFs. Exploring additional explanatory variables was considered out of the scope of this project and thus were not considered for regression. However, more explanatory variables could be considered. For example, researchers could find that horizontal curves have a statistically significant correlation to crash frequencies in Idaho.

References

1. "Toward Zero Deaths." Federal Highway Administration. U.S. Department of Transportation, 15 Oct. 2014. Web. 22 Oct. 2014.
2. Highway Safety Manual, Volume 1 (2010). AASHTO, Washington, D.C.
3. Highway Safety Manual, Volume 2 (2010). AASHTO, Washington, D.C.
4. Xie, F., Gladhill, K., Dixon, K. K., & Monsere, C. M. (2011). Calibration of Highway Safety Manual Predictive Models for Oregon State Highways. *Transportation Research Record: Journal of the Transportation Research Board*, *2241*(1), 19-28.
5. Williamson, M., & Zhou, H. (2012). Develop calibration factors for crash prediction models for rural two-lane roadways in Illinois. *Procedia-Social and Behavioral Sciences*, *43*, 330-338.
6. Sun, C., Brown, H., Edara, P., Claros, B., & Nam, K. A. (2013). Calibration of the Highway Safety Manual for Missouri (No. cmr14-007).
7. Young, J., & Park, P. Y. (2013). Benefits of small municipalities using jurisdiction-specific safety performance functions rather than the Highway Safety Manual's calibrated or uncalibrated safety performance functions. *Canadian Journal of Civil Engineering*, *40*(6), 517-527.
8. Brimley, B. K., Saito, M., & Schultz, G. G. (2012). Calibration of Highway Safety Manual Safety Performance Function. *Transportation Research Record: Journal of the Transportation Research Board, 2279*(1), 82-89.
9. Srinivasan, R., & Carter, D. L. (2011). *Development of Safety Performance Functions for North Carolina* (No. FHWA/NC/2010-09). North Carolina Department of Transportation, Research and Analysis Group.
10. Mehta, G., & Lou, Y. (2013). Calibration and Development of Safety Performance Functions for Alabama. *Transportation Research Record: Journal of the Transportation Research Board*, *2398*(1), 75-82.
11. Schrock, S. D., & Wang, M. H. (2013). Evaluation of Interactive Highway Safety Design Model Crash Prediction Tools for Two-Lane Rural Roads on Kansas Department of Transportation Projects (No. K-TRAN: KU-10-1). Kansas Department of Transportation.
12. Haas, P. R., & Gosse, C. (2010). Development of Safety Performance Functions for Two-Lane Roads Maintained by the Virginia Department of Transportation.
13. Qin, X., Zhi, C., & Vachal, K. (2014). Calibration of Highway Safety Manual Predictive Methods for Rural Local Roads. In *Transportation Research Board 93rd Annual Meeting* (No. 14-1053).

14. Russo, F., Biancardo, S. A., Busiello, M., Dell'Acqua, G., & Coraggio, G. (2014). Crash Severity Prediction Functions on Italian Rural Roads. In *Transportation Research Board 93rd Annual Meeting* (No. 14-4521).
15. Wang, J. & Abdel-Aty, M. (2014). Comparison of Safety Evaluation Approaches for Intersection Signalization in Florida. In *Transportation Research Board 93rd Annual Meeting* (No. 14-0374).
16. Chen, Y., Persaud, B., & Sacchi, E. (2012). Improving transferability of safety performance functions by Bayesian model averaging. *Transportation Research Record: Journal of the Transportation Research Board*, *2280*(1), 162-172.
17. Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, *122*(2), 105-113.
18. Xie, Y., & Zhang, Y. (2008). Crash frequency analysis with generalized additive models. *Transportation Research Record: Journal of the Transportation Research Board*, *2061*(1), 39-45.
19. PathWeb. *PathWays*. http://pathweb.pathwayservices.com/idaho/. Accessed 2013-2014.
20. Idaho Transportation Department. *WebCARS*. http://apps.itd.idaho.gov/apps/webcars/. Accessed 2013-2014.
21. R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Appendix A: Regression Outputs from R i386

The following is the regression output from R for the two-lane, two-way highway segments using Method 1:

```
Call:
glm.nb(formula = Modified.Observed.Crashes ~ ln.AADT. + ln.Length.,
    data = Dataset, init.theta = 17.57167306, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2692  -0.8264  -0.5042   0.7836   3.5144

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.79987    0.73962  -7.842 4.45e-15 ***
ln.AADT.     0.73709    0.08846   8.332  < 2e-16 ***
ln.Length.   0.89375    0.07378  12.114  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(17.5717) family taken to be 1)

    Null deviance: 460.08  on 312  degrees of freedom
Residual deviance: 245.32  on 310  degrees of freedom
AIC: 490.97

Number of Fisher Scoring iterations: 1


              Theta:  17.6
          Std. Err.:  16.8

 2 x log-likelihood:  -482.966
```

The following is the regression output from R for the two-lane, two-way highway segments using Method 2:

```
Call:
glm.nb(formula = Modified.Crashes.Mile.Year ~ ln.AADT., data = Dataset,
    init.theta = 2.479679541, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4830  -1.0268  -0.5002   0.4226   4.0034

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.78526    0.65649  -8.812   <2e-16 ***
ln.AADT.     0.75013    0.07978   9.403   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(2.4797) family taken to be 1)

    Null deviance: 419.84  on 312  degrees of freedom
Residual deviance: 326.36  on 311  degrees of freedom
AIC: 924.36

Number of Fisher Scoring iterations: 1


              Theta:  2.480
          Std. Err.:  0.484

 2 x log-likelihood:  -918.358
```

The following is the regression output from R for the 3-leg stop controlled intersections:

```
Call:
glm.nb(formula = X3.Modifed.Observed ~ ln3Major + ln3Minor, data = Dataset,
    init.theta = 2693.526991, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3294  -0.7171  -0.5695   0.8580   2.5129

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -6.1502     2.5318  -2.429   0.0151 *
ln3Major      0.0966     0.4218   0.229   0.8188
ln3Minor      0.6969     0.5764   1.209   0.2266
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(2693.527) family taken to be 1)

    Null deviance: 63.943  on 78  degrees of freedom
Residual deviance: 59.846  on 76  degrees of freedom
  (6 observations deleted due to missingness)
AIC: 94.491

Number of Fisher Scoring iterations: 1


              Theta:  2694
          Std. Err.:  73669
Warning while fitting theta: iteration limit reached

 2 x log-likelihood:  -86.491
```

The following is the regression output from R for the 4-leg stop controlled intersections:

```
Call:
glm.nb(formula = X4.Modifed.Observed ~ ln4Major + ln4Minor, data = Dataset,
    init.theta = 0.5588683466, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.2777  -0.8939  -0.7227   0.3081   2.4405

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.63357    2.85743  -3.021  0.00252 **
ln4Major     0.89664    0.41505   2.160  0.03075 *
ln4Minor     0.04578    0.42644   0.107  0.91451
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.5589) family taken to be 1)

    Null deviance: 77.905  on 84  degrees of freedom
Residual deviance: 71.053  on 82  degrees of freedom
AIC: 164.03

Number of Fisher Scoring iterations: 1


              Theta:  0.559
          Std. Err.:  0.195

 2 x log-likelihood:  -156.028
```