Development of Safety Performance Functions for Undivided Rural Two-Lane Roadways

by

Paige Martz

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

(Civil and Environmental Engineering)

At the

University of Wisconsin – Madison

2017

Abstract

The objective of the research was to develop a methodology for creating safety performance functions specific to the state of Wisconsin, using undivided rural two-lane roadways. The methodology was designed to complement the practices of Wisconsin Department of Transportation (WisDOT) and work with the database WisDOT uses. The considered variables included curve presence, lane width, right shoulder total width, right shoulder paved width, truck percent, international roughness index, and posted speed. Negative Binomial regression models were developed using a dataset of 7,911 segments. Cumulative residual plots identified significant points in the AADT range where the SPFs could be split into two. Results show using two models, split by AADT at 7,500 vehicles per day (vpd) for all crash types and 7,000 vpd for fatal and injury crashes, were more accurate models than using a single grouping of AADT. Weather-related crashes were found to reduce the dispersion parameter and Akaike information criterion when comparing models including weather-related crashes and models excluding weather-related crashes. SPFs were developed for weather-related crashes: one with AADT $\leq$ 6,000 vpd and one with AADT > 6,000 vpd. Six SPFs were developed for undivided rural two-lane roadways, including all crash types and fatal and injury crashes.

Acknowledgements

I would first like to thank my advisor, Dr. David Noyce, for his guidance throughout the thesis project and Master's program. I would like to thank Ms. Andi Bill, Dr. Madhav Chitturi, and Dr. Ghazan Khan for their collaboration and assistance on the project. I would also like to thank Dr. Soyoung Ahn and Dr. Bin Ran for serving on my thesis committee. Next, I would like to thank Wisconsin Department of Transportation for their support of this research project. Finally, I must express my gratitude to my family and friends for their endless support and encouragement throughout my years of study.

Table of Contents

List of Figures

List of Tables

1. Introduction

The Highway Safety Manual (HSM) provides safety performance functions (SPFs) that can be used to predict crashes on facilities, such as undivided rural two-lane roadways and four-way stop-controlled intersections. However, these equations were developed from limited data. For example, the rural two-lane two-way roadway SPF was developed using data from Minnesota, Washington, Michigan, and California which does not support a strong variety of locations or even jurisdictions *(1)*. Different locations and jurisdictions can contain different characteristics, like terrain, driver behaviors, crash reporting, and roadway maintenance. Development of SPFs specific to a jurisdiction can improve the accuracy of crash predictions and safety analysis.

The developed SPFs can be used a few different ways *(2)*. On the network screening level, SPFs can be used to identify locations with promise by estimating the predicted number of crashes for a particular facility type with a particular set of characteristics. SPFs can be used to rank sites in order of need for safety improvements and can be used for high level planning and design. The expected safety impacts of design changes can be determined from SPFs. State specific SPFs can be added to Safety Analyst, IHSDM, and spreadsheet tools. SPFs can also be utilized to compare predicted crashes of alternatives for a site or to compare design features or specific crash types with more detailed SPFs, for example curve crashes or nighttime crashes. Finally, crash modification factors (CMFs) can be estimated directly from SPFs.

Weather-related crashes can add a layer of uncertainty and increase the error in SPFs, in part, because weather occurs randomly. Adverse roadway conditions, like snow, ice, or rain, can be problematic to drivers. In the winter, temperatures can drop into the negative degrees (Fahrenheit) creating ice on the roadway and drivers can lose control of their vehicles. Therefore, there is a need to determine what factors affect weather-related crashes and a necessity to develop a method of identifying locations of interest for roadway improvements in order to reduce the number of weather-related crashes that occur. The Wisconsin Department of Transportation (WisDOT) has already implemented programs to improve the safety of roadways under the Highway Safety Improvement Plan (HSIP) *(3)*. HSIP "funds highway safety projects at sites that have experienced a high crash history" *(3)*. One location that benefitted was the west-north ramp of the Marquette Interchange in downtown Milwaukee *(4)*. The ramp received a high friction surfacing treatment because crash records showed a high number of crashes, especially on rainy days. After the installation of the treatment, the number of crashes per year significantly dropped. Having additional weather-related crash information could help in the efforts with HSIP as well as providing more information for planning and design.

1.1 Objectives

The objective of the research was to develop a methodology for creating SPFs in the state of Wisconsin using undivided rural two-lane roadways as the first facility. In the process of developing SPFs for undivided rural two-lane roadways, it was found that weather-related crashes affected the goodness-of-fit of the SPFs. Because of this, weather-related crashes were analyzed more closely in relation to geometric features and weather variables. Through the analysis of weather-related crashes, SPFs were developed for undivided rural two-lane roadways specific to weather-related crashes.

1.2 Tasks

To complete the objectives, data were first collected from Meta-Manager, a data management system developed by WisDOT and then processed to only include undivided rural two-lane segments. As a part of this filtering process, segments with any null values input for any of the influencing variables were removed along with segments that contained outlying values. The influencing variables include annual average daily traffic (AADT), segment length, lane width, shoulder width, curve presence, truck percent, international roughness index (IRI), and posted speed. The HSM guidelines for jurisdiction-specific SPF development were followed along with the use of the Negative Binomial regression model to statistically analyze the filtered dataset. Preliminary and final SPFs were developed for all crash types and fatal and injury crashes. A weather-related crash analysis was also performed after finding that weather-related crashes affected the goodness-of-fit of the preliminary SPFs for all crash types. A basic statistical analysis was performed on weather-related variables, including road conditions and weather conditions at the occurrence of the weather-related crashes. SPFs were developed for the weather-related crashes using the same influencing factors as the two previously developed SPFs.

1.3 Scope

The scope of work consists of developing SPFs for undivided rural two-lane roads, specifically for all crash types, fatal and injury crashes, and weather-related crashes using the Meta-Manager database. One limitation of the scope is the use of the Meta-Manager database. Only variables included in the database could be used in the development of the SPFs. For example, the database does not include any information on curves aside from curve presence. A further limitation involves the fatal and injury crash SPF. Not all of the intersection-related crashes could be removed from the fatal and injury crash total which may affect the finalized SPFs.

1.4 Thesis Organization

The following document is organized into 8 primary chapters. Chapter 1 introduces the topic and outlines the objectives. Chapter 2 includes the literature review that provides background to the thesis topic. Chapter 3 covers the data collection and processing. Chapter 4 discusses the selected model and outlines the methodology. Chapter 5 outlines and explains the goodness-of-fit parameters. Chapter 6 discusses the model results and applications for the developed models. Chapter 7 includes the analysis and model results for the weather-related crashes. The conclusion is covered in Chapter 8. At the end of the document, a list of references and appendices are included.

2. Literature Review

2.1 HSM recommended methods and models

The HSM provides a straight forward, step-by-step procedure on how to use the predictive model. One of the benefits of the HSM predictive model is that it allows for crash estimation when no observed crash data is available or no other predictive model is available *(1)*. Additionally, it addresses regression-to-the-mean bias and the reliance on crash data for any one site is reduced by using crash data from multiple similar sites. All of the SPFs in the HSM were produced by using the Negative Binomial distribution which is better at modeling crash data that tends to have high variability. The most basic form of an SPF is given in Equation 1:

$$N_{predicted} = N_{SPF\ x} \times (CMF_{1x} \times CMF_{2x} \times \ldots CMF_{yx}) \times C_x \qquad (1)$$

Where $N_{predicted}$ is the predicted model estimate of crash frequency for a certain year on a specific site type, $N_{SPF\ x}$ is the predicted average crash frequency for base conditions with the SPF and site type in crashes per year, CMF_{yx} is the CMF specified to the site type, and C_x is the calibration factor to adjust for site type local conditions.

The HSM gives a concise and easy to follow step-by-step process of how to use the predictive method. The Empirical Bayes (EB) method is a part of the predictive method and is used to combine the estimation from the statistical model with the observed crash frequency at a certain site. It is the chosen method of the HSM, chosen over the Hierarchical Bayes method and the Full Bayes method. It is advantageous to use because once a calibrated model is developed for a specific site type, the method can be easily applied.

2.2 Calibration versus Development

As discussed previously, the HSM provides a procedure on how to use crash estimation when no observed crash data are available or no other predictive model is available *(1)*. The HSM predictive model addresses regression-to-the-mean bias and all of the SPFs in the HSM were produced by using the Negative Binomial distribution which is efficient at modeling crash data with high variability.

A limitation of the HSM predictive method is that it accounts for state wide or community wide differences in behavioral factors like driving age and alcohol usage but does not account for it on a site-by-site basis. The effects of weather are not specifically addressed, although they are accounted for indirectly through the calibration process. AADT volumes are accounted for; however, the proportions of trucks and/or motorcycles nor traffic volume variations throughout the day are not. Additionally, the HSM assumes the effects of individual geometric design and traffic control features are independent and ignores potential interactions between them.

The development of jurisdiction-specific SPFs can improve the accuracy of crash predictions, but development can be limited by the quality and amount of available data. Garber and Rivera suggest the development of state specific SPFs, when possible, to better represent the environment and driving behaviors in each state *(5)*. Developing jurisdiction-specific SPFs can also “provide the opportunity to examine alternative functional forms (depending on the data) rather than using the default forms in the HSM and Safety Analyst” *(6)*. From the many models that were developed in a large number of studies, it was proven that the number of expected crashes depends on specific variables that may differ from state to state. Furthermore, jurisdictions can deviate from the HSM identified SPF facility types: rural two-lane, two-way roads, rural multilane highways, urban and suburban arterials, freeways and interchanges. For example, a jurisdiction could create an SPF for low-speed rural two-lane roads and high-speed rural two-lane roads.

2.3 Statistical Models to Develop SPFs

Some of the first models in SPF development related crash rates to site characteristics, other than AADT *(7)*. The models, however, assumed that crash rate and exposure were linearly related, so the relationship would need to be addressed by including exposure as an independent variable. Earlier SPFs used conventional linear regression to show the relationship between crash rate and site characteristics. Linear regression assumes that crash rate is normally distributed, while studies since then have shown this to be incorrect. Linear regression can also output negative values, which does not pair well with crash counts which are non-negative integers.

Poisson and Negative Binomial regression are two of the most popular models used in SPF development. The main limitation of the Poisson regression model is that the mean and variance are considered equal. The variance is often found to exceed the mean with crash data, which results in overdispersion *(8)*. Because of the overdispersion, “standard errors of the parameter estimates tend to be understated” which could cause the parameters to be considered statistically significant when they actually are not *(9)*. Negative Binomial regression accounts for overdispersion making it the more widely used model for SPF development, along with being the recommended model by the HSM *(1)*. The Negative Binomial model relies on an exponential function that makes the relationship between the dependent (predicted crashes) and independent variables (AADT, site characteristics,…) log linear *(7)*. By this relationship, the model can be estimated using Generalized Linear models (GLM) which is consistent with the predictive models mentioned in the HSM. One version of a Negative Binomial regression model with a Gamma-Poisson mixture is shown in Equation 3.

$$f(y;\mu,\theta) = \frac{\Gamma(y+\theta)}{y!\Gamma(\theta)} \frac{\mu^y \theta^y}{(\mu+\theta)^{y+\theta}} \tag{3}$$

Where θ is the shape parameter, μ is the mean, and $\Gamma()$ is the gamma function *(10)*. The variance is shown in Equation 4.

$$V(\mu) = \mu + \frac{\mu^2}{\theta} \tag{4}$$

When the dispersion parameter nears infinity, the Poisson distribution is approached. Equation 5 shows the basic form of the Negative Binomial regression model used in the project.

$$N_{SPF} = L \times e^{\alpha_0} \times AADT^{\alpha_1} \tag{5}$$

Where the variables α_0 and α_1 are coefficients, L represents segment length, and N_{SPF} represents the predicted number of annual crashes. Some of the different functional forms of SPFs used by different states based on the Negative Binomial Regression model are demonstrated in the following section, Section 2.4.

More recently, some researchers do not encourage the use of the GLM Negative Binomial relationship, claiming the model does not always represent the relationship between the predicted crashes and site characteristics. Hauer suggests a method to choose the most convenient function for each independent variable separately *(11)*. This could mean taking the exponent of any of the dependent variables and/or simply adding the dependent variables together.

Another similar type of model is the truncated Poisson regression model which takes into account the maximum number of uses of a design parameter *(12)*. Washington et al. describe how the model can be used, for example, when including the number of times per week when an in-vehicle navigation system is used. The data would be truncated at seven which is the maximum number of uses during any given week. Zero-inflated Poisson and Negative Binomial regression models are also other options *(12)*. These models account for excess zero count data for two types of cases. The first case results from failing to observe an event during the observation period, while the second case results from the incapability to experience an event.

Other new methods have been developed using neural networks (GAM) *(13)(14)*. Neural network models are a subset of nonlinear models and are based on an observed, single variable response found to be dependent on a corresponding one-dimensional input *(14)*. The models are not limited to a predetermined functional form or distribution assumptions. The newer models are considered to be time consuming and have some limitations based on generalizability *(15)*.

2.4 Case Studies

The section provides an overview of SPF development and HSM calibration efforts of other agencies. Basic information is given for each study, including the regression method used, variables considered, facilities used, and results. In the case of a study using calibration, it can be given that the HSM method was used unless otherwise specified.

2.4.1 Alabama

The study used calibration and development to create SPFs for Alabama, specifically two-lane two-way rural roads and four-lane divided highways *(16)*. Using random sampling, 6000 sites were selected for the study and four years of data were collected. AADT, segment length, lane width, speed limit, dummy year, and shoulder width were considered in the model development process. Using Negative Binomial regression techniques, four different SPFs were developed. The models are shown in Equations 6 through 9, respectively, as models 1 through 4. The first model uses the HSM base SPF form, seen in Equation 6.

$$\widehat{\mu_i} = \beta_0 AADT_i^{\beta_1} SL_i \tag{6}$$

Where μ_i is the estimated expected number of crashes per year for site i, $AADT_i$ and SL_i are the annual average daily traffic and the segment length for site i, respectively, and β_0 and β_1 are parameters to be estimated. The second model uses the formula of SPF developed in the United Kingdom for two-lane two-way rural roads, in Equation 7.

$$\widehat{\mu_i} = \beta_0 AADT_i^{\beta_1} SL_i^{\beta_2} e^{(\frac{bn_i}{SL_i})} \tag{7}$$

Where n_i is the number of minor junctions within site i and β_0, β_1, β_2, and b are parameters to be estimated. Model 3 is the form used by the Connecticut Transportation Institute, as seen in Equation 8.

$$\widehat{\mu_i} = e^{(\beta_0 + \beta_1 DY_i + \beta_2 lnAADT_i + \beta_3 lnSL_i + \beta_4 LW_i + \beta_5 S_i)} \tag{8}$$

Where DY_i is the dummy variable to account for effects of year on intercept, LW_i is the lane width for site i, and S_i is the speed limit (mph). Council and Stewart developed the fourth model, Equation 9, based on California, Minnesota, North Carolina, and Washington data:

$$\widehat{\mu_i} = e^{(\beta_0 + \beta_1 SW_i + \beta_2 LW_i)} AADT_i^{\beta_3} SL_i \tag{9}$$

Where SW_i represents the shoulder width.

For model validation, five measures were found: Akiake Information Criteria (AIC), likelihood value (LL), Mean Absolute Deviation (MAD), Mean Square Prediction Error (MSPE), and Mean Prediction Bias (MPB). Based on the AIC and the LL, Model 3 was determined to be the best model to use when developing new SPFs for two-lane two-way rural roads and four-lane divided highways. The crash rates were found to be affected statistically by lane width, speed limit, and year.

2.4.2 Illinois

SPF models for intersections and roadway segments were developed for Illinois, in two different studies. The initial step for developing the SPFs included collecting crash data between 2001 and 2005 *(17)*. The information taken from each of the crashes included the case ID number, county, crash type, mile, route name, and the TS route. Additionally, it was necessary to know the AADT, beginning station, county, ending station, functional class, key route, median type, number of lanes, segment length, township, and urban code for each of the sites. To organize the collected data, the crash dataset and the roadway dataset were merged to provide a single location for all of the information. Specifically, the segments were separated by type within the following groups: rural two-lane highway, rural multilane undivided highway, rural multilane divided highway, rural freeway with four lanes, rural freeway with six or more lanes, urban two-lane highway, urban one-way arterial, urban multilane undivided highway, urban multilane divided highway, urban freeway with four lanes, urban freeway with six lanes, and urban freeway with eight or more lanes.

To develop SPFs for segments, Tegge et al. used segment length, traffic volume, and regression parameters. Developed from past research, the fundamental form of a road segment SPF is shown in Equation 5. The study used the EB method to compare sites with and without the treatments for each site type. Microsoft Excel Visual Basic software was used for SPF estimation and implementation of the intersection and segment SPF base equations. The following variables were found to be statistically significant in model development: shoulder width, IRI, number of lanes, lane width, median type, one or two way direction, rut depth, speed limit, surface type, area type, number of intersections, and AADT. Furthermore, the next set of variables had a positive relationship with crash rate: AADT, access control, IRI, number of lanes, lane width, median type, and one or two way. The following variables were shown to have a negative relationship with crash frequency: functional class, shoulder type, shoulder width, rut depth, speed limit, and surface type. The main limitations in developing jurisdiction SPFs were incomplete datasets since some variables that were not included in the dataset may have had an impact on the overall outcome of the models. Additionally, the data that were available may have produced some error in the models.

2.4.3 Kansas

An SPF for rural two-way two-lane highways was developed and calibrated for Kansas to predict the number of future crashes on the facility *(18)*. Crash data were collected between 2005 and 2007. The following site characteristic data were collected and considered in model development: shoulder width, lane width, AADT from 2007, roadside hazard rating, horizontal curvature, vertical curvature, and segment length. The Negative Binomial regression model was used, along with SPSS software and a 95 percent confidence interval. In developing the SPF, animal crashes were excluded due to the difficulty of finding their exact locations. Equation 10 is the modeled SPF.

$$N_{pred-no-an} = AADT^{1.01} L^{0.85} e^{(-10.07+0.58\times RHR)} \tag{10}$$

Where $N_{pred\text{-}no\text{-}an}$ is the number of predicted crashes without animals and RHR is the roadside hazard rating. Roadside hazard rating was found to be significant during the model development. After comparing the different models, using test statistics, it was shown that the HSM calibrated model had better results than the developed SPF. Therefore, it was recommended to use the calibrated model over the developed model.

2.4.4 North Carolina

The study used two SPF approaches: calibrating the HSM SPFs and developing North Carolina state specific SPFs *(8)*. The facility types include the following:

- Rural two-lane roads
- Rural freeways – four-lanes – outside the influence of interchanges
- Rural freeways – +six-lanes – outside the influence of interchanges
- Rural freeways – four-lanes – within the influence of interchanges
- Rural freeways – +six-lanes – within the influence of interchanges
- Rural multilane divided roads
- Rural multilane undivided roads
- Urban two-lane roads
- Urban freeway – four-lanes – outside the influence of interchanges
- Urban freeway – six-lanes – outside the influence of interchanges
- Urban freeway – +eight-lanes – outside the influence of interchanges
- Urban freeway – four-lanes – within the influence of interchanges
- Urban freeway – six-lanes – within the influence of interchanges
- Urban freeway – +eight-lanes – within the influence of interchanges
- Urban multilane divided roads
- Urban multilane undivided roads

In developing state specific SPFs, data were collected including geometric characteristics of segments, traffic volumes, and crashes. The following includes specified segment conditions selected in the study: the segments must be more than 250 feet away from at-grade intersections or rail road crossings, AADT must be greater than 500 vehicles per day (vpd) for each segment, and there must be a minimum of 100 crashes per year. The data used in development were from 2004 to 2008. A Negative Binomial regression method and SAS software were utilized. The study used Equation 5, the basic Negative Binomial model. Another equation, Equation 11, was modeled for only rural two-lane roads by including additional variables, other than AADT:

$$Y = L \times exp\{\alpha + \beta_1 \times f_1(AADT) + \beta_2 f_2(X_2) + \beta_3 f_3(X_3) + \beta_4 f_4(X_4) + \cdots\} \qquad (11)$$

Where f_1, f_2, f_3 and f_4 are functions of independent variables, X_1, X_2, X_3 and X_4 are independent variables, and α, β_1, β_2, β_3 and β_4 are estimated parameters. The following variables were included in the rural two-lane road model: AADT, terrain, shoulder width, and shoulder type. Each facility had an SPF developed, each, for total crashes, fatal and injury crashes (KABC), fatal and injury crashes (KAB), property-damage-only crashes, lane departure crashes, single vehicle crashes, multi-vehicle crashes, wet crashes, and night crashes.

2.4.5 Pennsylvania

State specific SPFs were developed based on Negative Binomial regression models *(19)*. Models were developed for rural two-lane intersection facilities and segment facilities. The crash data used were from 2005 to 2012 which included 170,720 crashes along 21,340 segments. The variables considered in model development included AADT, segment length, posted speed, left paved shoulder width, right paved shoulder width, access density, horizontal curve density, degree of curve per mile, length of curve per mile, roadside hazard rating, presence of passing zone, presence of centerline rumble strips, presence of shoulder rumble strips presence of curve warning pavement marking, presence of intersection warning pavement marking, and presence of “aggressive driving dots” *(19)*. The following function, Equation 12, was developed to predict the number of crashes at any segment i.

$$N_{cr,pr} = Length \times AADT^{0.754} \times e^{-5.934} \times e^{0.101RHR6,7} \times e^{0.091RHR4,5} \times e^{-0.239PZ} \times e^{-0.188SRS} \times e^{0.008AD} \times e^{0.03HCD} \times e^{0.002DCPM} \quad (12)$$

Where $N_{cr,pr}$ is the predicted total crash frequency on the segment (crashes/year), length is the length of segment (miles), AADT is the annual average daily traffic on the segment (vpd), RHR6, 7 is the roadside hazard rating on the segment of 6 or 7 (1 if RHR is 6 or 7; 0 otherwise), RHR4, 5 is the roadside hazard rating on the segment of 4 or 5 (1 if RHR is 4 or 5; 0 otherwise), PZ is the presence of a passing zone in the segment (1 if present; 0 otherwise), SRS is the presence of shoulder rumble strips in the segment (1 If present; 0 otherwise), AD is the access density in the segment, total driveways and intersections per mile of segment length (Access Points/Mile), HCD is the horizontal curve density in the segment, number of curves in the segment per mile (Hor. Curves/Mile), and DCPM is the total degree of curvature per mile in the segment, the sum of degree of curvature for all curves in the segment divided by segment length in miles (Degrees/100 ft/Mile).

Two SPFs were developed for the facility type: total crashes and fatal and injury crashes. Some of the variables included in the HSM functions were excluded due to lack of data, limited confidence in data quality, and/or little variation in data across segments. One of the most important findings in the study was that the presence of a right turn was found associated with an increase in the predicted crash frequency, contrary to what is mentioned in the HSM. Additionally, roadside hazard rating of 6 or 7, roadside hazard rating of 4 or 5, access density, horizontal curve density, and degree of curve per mile had a positive relationship with crash rate. The presence of a passing zone and the presence of shoulder rumble strips, on the other hand, had a negative relationship with crash frequency.

2.4.6 South Caraguas, Puerto Rico

An SPF was developed for shoulder rumble strips on freeways, in South Caraguas, using the Negative Binomial model and the EB method *(20)*. The study originally started out with the objective to develop a CMF for the rumble strip treatment, but concluded with the development of an SPF. A Negative Binomial model was chosen because it is known to offer better fitted models than a Poisson regression model. Additionally, the EB method includes the Negative Binomial parameter that can be obtained from regression-to-the-mean. The first part of the process involved identifying a group of 43 untreated segments that had similar characteristics to the treated segments. The major data needs included number of lanes (4 to 6 lanes), lane width (12 feet), average segments AADT's, average crashes for segments, and total untreated segments. A period of 2 to 3 years was used. The data were then cleaned by eliminating the inaccurate or incomplete data. The model form is shown in Equation 13.

$$E(\mu) = \beta_0 \times X_1^{\beta 1} \times X_2^{\beta 2} \tag{13}$$

Where E(μ) is the number of crashes per year, X_1 represents the segment length (kilometers), X_2 is the average AADT (vpd), and $\beta_0, \beta_1, \beta_2$ is the regression parameters. AADT and segment length were the two variables used in model development

2.4.7 Utah

The study developed a state specific SPF and calibrated the HSM SPF for rural two-lane two-way roadway segments using data from 2005 to 2007 *(21)*. A total of 426 crashes were reported on the 157 road segments. The EB method and the Negative Binomial model were used because they account for the overdispersion parameter. The data needs were determined by referencing the HSM data requirements. Instead of simply modifying the HSM model, a few new variables were included like speed limit, presence or absence of shoulder rumble strips, and the percentage of single unit and multiple unit trucks. A calibration factor of 1.16 was added into the HSM model to obtain Equation 14.

$$N_{local} = AADT \times L \times 3.09 \times 10^{-4} \quad (14)$$

Where N_{local} is the predicted number of crashes per year. The Negative Binomial model was first developed using SAS software, with a 75% confidence level, to formulate the general relationships between the independent variables and crash frequencies. Then, the model with a 95% confidence level was developed, shown in Equation 15.

$$N_{SPF} = AADT^{0.840} \times \exp[-12.1 + (0.450)(L) - (0.0271)(CT) + (0.0824)(Speed)] \quad (15)$$

Where CT represents combo-unit trucks. Model results showed that longitudinal grade, percentage of single unit trucks, shoulder width, and lane width had no significant impact on crash rate. Driveway density and speed limit were shown to have a positive relationship with crash frequency. Additionally, the absence of shoulder rumble strips consistently had a negative correlation with crash rate.

2.4.8 Virginia

Virginia Department of Transportation developed SPFs specific to the state of Virginia *(22)*. The study focused on multilane highway and freeway segments with data collected from 2004 to 2008, consisting of 20,235 multilane highways and 2,905 directional freeways. The highways and freeways were divided into subtypes as follows:

- Rural multilane undivided highway segments
- Rural multilane divided highway segments
- Urban multilane undivided arterial segments
- Urban multilane divided arterial segments

- Rural freeway segments—4 lanes
- Rural freeway segments—6+ lanes
- Rural freeway segments within an interchange area—4 lanes
- Rural freeway segments within an interchange area—6+ lanes
- Urban freeway segments—4 lanes
- Urban freeway segments—6 lanes
- Urban freeway segments—8+ lanes
- Urban freeway segments within an interchange area—4 lanes
- Urban freeway segments within an interchange area—6 lanes
- Urban freeway segments within an interchange area—8+ lanes

Each model was developed for fatal and injury crashes and all crash types. The Negative Binomial regression models that were used to develop the SPFs follow the form of Equation 5. The variables used in the model were AADT and segment length.

Variations were detected within the state of Virginia implying the developed state SPF could not represent all the regions so district group SPFs were recommended to reflect those variations. Eight subtypes of the multilane highway segments were defined differentiating between urban/rural, divided/undivided, and eastern/western. Accordingly, district SPFs were developed.

2.5 Influencing Factors

Section 2.5 discusses the influencing factors chosen for SPF development in this research effort. The decisions were based on available data, WisDOT input, previously summarized case studies, and the following literature. Based on past studies, aside from AADT and segment length, curve presence, IRI, lane width, posted speed, truck percent, and shoulder width have proven to be statistically significant in developing SPFs. The following briefly explains why each variable is significant to the model.

Many studies have proven that curve presence is a valid factor in predicting crash frequency. Some of the curve attributes that affect crash frequency include curve radius and speed change on the curve. Crash frequency has shown to decrease as curve radius increases *(23)*. Additionally, crashes involving trucks were seen to increase as horizontal curvature increased *(24)*. In relation to speed change, curves that required speed reduction had higher crash rates than curves that did not *(25)*. Zegeer et al. stated that "speed [was] a definite factor, perhaps in both the occurrence and also the severity of crashes on curves" *(26)*. The individual attributes of curves affect crashes, making the knowledge of curve presence along roadway segments important in model development.

IRI is a quantitative variable that measures the roughness of a travelway. The roughness is measured on a scale starting at zero, the smoothest possible value, and increasing as roughness increases. Tegge et al. determined that IRI had a positive relationship with crash frequency after performing a multivariate analysis *(17)*. This means that as IRI increases, crash frequency increases. Another study also found that IRI had a positive relationship with crash frequency *(27)*. Studies show that IRI is a valid variable to consider in crash model development.

Studies have confirmed that lane width affects crash frequency. One study found that "average lane width was estimated to be…negatively related to crash frequency for rural non-freeway segments" *(9)*. Additionally, widening lanes on curves could affect crash frequency *(26)*. Hauer stated that there was a connection between lane width and safety due to two reasons: wider lanes have larger separation "between vehicles moving in adjacent lanes" which "may provide a wider buffer to adsorb the small random deviations of vehicles moving from their intended path" and "a wider lane may [provide] more room for correction in near-accident circumstances" *(28)*. For single lanes, the width where "drivers do not feel the need to shift to the right when meeting an oncoming truck was deemed appropriate." Studies provide evidence that lane width affects crash frequency.

As far as posted speed, the relationship with crash frequency varies. It can be confirmed that the longer it takes a vehicle to stop, the larger the probability of crash involvement. Furthermore, a higher speed leads to a larger "proportion of accidents that are reportable and get reported" *(29)*. Tegge et al. found that as speed limit increases, crash frequency decreases, showing that there is a relationship with speed *(17)*. On the other hand, Saito et al. found that speed limit had a positive relationship with crash rate *(21)*. Another more involved study found that crash frequency increases as speed limit increases, as well *(30)*. The Federal Highway Administration describes how many studies have found the relationship between the two factors inconclusive *(31)*. The bottom line is that evidence shows speed is a factor affecting crash frequency.

Truck percent has a relationship with crash frequency. Ran and Lee found that truck percent had a nonlinear relationship with crash frequency. Drivers appeared to be more cautious around low truck percents and, again, at truck percents around 16 to 18 percent *(32)*. In the same way, Dissanayake and Amarasingha noted an increase in truck percent corresponding to an increase in crash frequency *(33)*. The percent of trucks on the travelway affects crash frequency.

In many studies, shoulder width was found to have a relationship with crash frequency. Strathman et al. established there was a positive relationship between the two for rural roadway sections and Stamatiadis et al. found "a relationship between shoulder width and crashes," as well *(9,34)*. It was also determined that widening paved shoulders and even adding unpaved shoulders to curves could reduce crashes *(26)*. Shoulder width has a relationship to crash frequency.

In addition to determining the factors related to crash frequency, segment length was analyzed. Hauer and Bamfo stated that "road sections shorter than 0.1 miles should either be reassembled into longer road sections or removed from the data base" *(35)*. This was, in part, due to the practice of police locating crashes within the nearest tenth of a mile. Segments less than 0.1 miles often contain intersections and may more frequently contain zero crashes. Lord et al. stated, "crash data characterized by a preponderance of zeros" could be caused by "analyses conducted with small time or spatial scales" *(36)*. Additionally, Koorey found that a "feature of the road in one segment triggered a crash officially located in another segment" *(37)*. It seemed to be a particular issue in remote rural locations, "where precise location by the attending traffic officer may be less likely." Given the findings from other studies, any segments less than 0.1 miles were removed from the database.

2.6 Weather-Related Crashes

This section provides a summary of the factors used in other weather-related research efforts, along with factors that were found to be significant. Many studies have investigated how different weather states and geometric conditions affect crash occurrence. Any combination of the following variables was used: road geometrics, temperature, rain, snow, wind speed, daylight, seasons, and fog *(38-56)*.

Specifically with road geometrics, studies used speed limit, grade, and curvature. Khattak et al. found curved roads to increase the occurrence of single-vehicle crashes *(45)*. Grades, hillcrests, and bottoms on straight roads in normal weather were found to increase two-vehicle crashes when compared to single-vehicle crashes. AADT and the interaction between speed limit change and left shoulder width were found to increase the number of multi-vehicle crashes *(48)*. Another study found median width, longitudinal grade, and horizontal curvature to be significant in both dry and snowy weather *(56)*.

The weather variables could be as specific as using the average temperature, daily average temperature in the month of January, or speed of the maximum wind gust. Many of the studies found that crashes were significantly influenced by weather conditions. Jung et al. found the following variables to be statistically significant during model development for rainy weather crashes: rainfall intensity, wind speed, roadway terrain, driver's gender, and safety belt *(43)*. A study in California found that wet weather largely effected crashes in California, most commonly single-vehicle and head-on crashes *(44)*. Qiu and Nixon found that crash rate had the tendency to increase during precipitation and that snow had a larger effect on crash rates than rain *(49)*. Other studies found that rainfall and temperature were positively correlated with the number of injury crashes *(39,55,57)*. Contrarily, Eisenberg found a negative relationship between monthly precipitation and monthly fatal crashes, and additionally, a positive relationship between daily precipitation and daily fatal crashes *(58)*. Fridstrom et al. found that snow had a negative relationship with crashes, frost had a negative impact on crashes, and daylight had a variable relationship with crash counts *(57)*.

Few studies show the direct relationship between weather-related crashes and road geometry. More often, studies analyze the relationship between weather and crashes or weather and crash severity.

2.7 Summary

To summarize the literature review, SPF development was chosen over HSM calibration because jurisdiction specific developed SPFs can improve the accuracy of crash predictions and use statistical regression models that better fit the data. Furthermore, the significance of roadway/intersection/interchange characteristics has been proven to change from state to state and even jurisdiction to jurisdiction. Finally, jurisdictions can deviate from the HSM identified SPF facility types if jurisdiction specific SPFs are developed.

The Negative Binomial regression model and Poisson regression models are the most commonly considered models for SPF development. The Negative Binomial regression model was chosen for SPF development in the research project because it accounts for overdispersion which often matches the relationship between crash rate and variables, like AADT. The case studies summarize research efforts in developing SPFs for road segments and look at the commonly used statistical models, crash data, variables considered, and results.

The variables selected for consideration in model development include AADT, segment length, curve presence, IRI, lane width, posted speed, truck percent, and shoulder width. The variables were found to have relationships with crash rate. Similarly, weather-related crashes were analyzed by many studies with respect to road geometrics, temperature, rain, snow, wind speed, daylight, seasons, and fog. Geometric variables including grade, curvature, and shoulder width were found to have an effect in relation to crashes and weather-related variables. Very little research was found to analyze geometric variables specifically with weather-related crashes.

3. Data Processing Methodology

The data processing chapter, Chapter 3, is the proposed methodology for developing Wisconsin state specific SPFs.

Roadway and safety data are to be assembled from Meta-Manager, a data management system developed by WisDOT to aid in the development of projects and programs that improve the State Trunk Highway Network. The database contains components such as pavement and bridge conditions, Six Year Highway Improvement Program information, highway geometric and attribute information, crash data, and highway capacity. Specifically for this project, crash data and roadway characteristic data were utilized. Data from the May 2016 Meta-Manager update were used, as well. Crash data were collected from the previous 5 year period for all crash types and fatal and injury crashes. The data processing methodology includes three main considerations: data assembly and cleaning, choosing influencing factors and their thresholds, and types of crashes used to develop the SPF. Each consideration is included in a subsection. The fourth subsection discusses how to apply CMFs to the developed SPFs.

3.1 Data Assembly and Cleaning

The following explains the process of filtering the data for any specific roadway type. First, the Excel files for the five regions need to be combined: NC, NE, NW, SE, and SW. This step involves copying and pasting the five separate Excel files into a single Excel file. The database can then be cleaned by removing segments with no AADT values and ramp segments. Ramp segments can be identified by using ArcGlobe to visually classify each segment that is defined in the database as a one-way highway or an undivided one-lane roadway. Any segment that gives some uncertainty will be removed. For example, one segment in Meta-Manager is a ferry route across a river, so it would be removed. Any segments that are mislabeled should be deleted, as well. Figure 1 shows possible types of segments that should be removed.

FIGURE 1 Possible segment types to remove

The top left image shows a segment that is both on a ramp and not on a ramp, the top right image shows a ramp segment, the bottom left image shows a segment on a roundabout, and the bottom right image shows a mislabeled segment. All four types of segments were removed from the database.

Next, identify the peer groups (“SFTY_TRVL_CLS_CD”) and functional classifications (“FCLASS”) of the target roadway segment. For example, rural segments are identified under the following peer groups:

- 410: rural 2-lane highways with AADT ≤ 2,000 vpd

- 420: rural 2-lane highways with 2,000 vpd < AADT ≤ 7,000 vpd
- 430: rural 2-lane highways with AADT > 7,000 vpd
- 440: low speed 2-lane highways

Segments with functional classifications less than 49 are considered rural and any number equal to or greater than 49 is considered urban. Remove the peer groups and functional classifications that do not apply to the specified segment type. Finally, each of the influencing factors should be evaluated for null values. The segments with null values for any of the influencing factors should be removed or the analyst could also assign likely values to the null cells.

Curve presence is calculated from the two columns labelled “HCURLE40” and “HCURLGT40.” “HCURLE40” is curves per mile posted 40 mph or less and “HCURLGT40” is curves per mile posted more than 40 mph. The columns contain ratings of good (G), fair (F), and poor (P) which represent the calculated amount of curvature a segment’s geometry contains. A curve rated good has no curvature and a curve rated fair or poor has curvature present. Each segment is given a value of no curve present (0) or curve present (1) depending on the rating. For example, if both columns contain a G, then the segment is given a 0 in the “CurvePresence” column. If the segment has a G in one column and an F or P in the other column, a 1 is listed in the “CurvePresence” column. If the segment has an F or P in both columns, a 1 is listed in the “CurvePresence” column. Next, lane width is calculated by dividing the travelway width (“TRWAYWD”) by the number of lanes.

Meta-Manager has crash data labeled under different categories, or columns. Total crashes are calculated by summing the following columns.

- fatal crashes (“MMGR_FATAL_CRSH_TOT”)

- incapacitating injury crashes (“MMGR_INCAP_INJ_CRSH_TOT”)
- non-incapacitating injury crashes (“MMGR_NONINCAP_INJ_CRSH_TOT”)
- possible injury crashes (“MMGR_PSBL_INJ_CRSH_TOT”)
- property-damage-only crashes (“MMGR_PD_ONLY_CRSH_TOT”)

Intersection-related crashes are then removed from the total. To do this, multiply the “INTPROP” column by the number of total crashes and subtract the calculated value from the total crashes. Intersection-related crashes cannot be removed from the fatal and injury crash total because the “INTPROP” column can only be applied to the total crashes. Deer crashes are not included in the total crashes used in model development. Fatal and injury crashes are calculated by summing, across the rows, fatal crashes, incapacitating injury crashes, non-incapacitating crashes, and possible injury crashes.

3.2 Choosing Influencing Factors and Their Thresholds

Multiple factors need to be considered in choosing which factors to include in the analysis:

- Engineering judgement and previous research
- Data availability: What roadway characteristics are available?
- Data reliability: Data may be available, but is it reliable? Paved shoulder width is considered to be more reliable than total shoulder width.
- Confounding factors: Total shoulder width and paved shoulder width are confounding variables. Therefore, the analyst has to decide, based on other factors, which variable to include in SPF development. One may want to use paved shoulder width instead of total shoulder width.

- Intended use: The analyst should consider what the SPFs will be used for. Who will be utilizing the SPFs (DOT, consultant, researcher, etc.)? For example, if the SPF will be used for design work, IRI may not make sense. If the SPF is used for analyzing roadway designs, there would not be an existing IRI value to use in comparing the new design to the existing roadway.

Below is a list of the influencing factors considered in developing undivided rural two-lane roadway segment SPFs, along with their database column labels:

- AADT: "HSTL_AADT_5_YR"
- Segment length: "PDP_MILE"
- Curve presence: "CurvePresence"
- IRI: "IIRIAV_BY"
- Lane width: "LaneWidth"
- Posted speed: "PTDSPEED"
- Truck percent: "TRKYR_1"
- Right total shoulder width: "RSHTOTWD"
- Right shoulder paved width: "RSHPAVWD"

3.2.1 Outlier Analysis

Once the database has been filtered and all segments with null significant variables removed, plot each of the influencing factors with the five year crash total. This step will provide a way in identifying obvious outliers. In this step, it can also be determined if a threshold is needed for any of the influencing variables. For example, a threshold could be set for shoulder width or lane width, based on design standards for the state. The plots used in the outlier analysis can be found in Appendix A.

3.3 Types of Crashes Used to Develop the SPF

The analyst needs to decide what types of crashes to include in developing the SPF, like weather-related crashes, deer crashes, and even severity types. Weather-related crashes are included in a single column in Meta-Manager. If included in the SPF, they will alter the precision in the prediction capabilities of the SPF. Deer crashes are not included in Meta-Manager. To include deer crashes in the SPF, the analyst would need to peruse the crash reports of each defined roadway segment. Additionally, SPFs are developed for severity types including total crashes and fatal and injury crashes. Does the analyst think it important to develop an SPF specifically for incapacitating injuries or non-incapacitating injuries, for example? The types of crashes included in the analysis should be determined based on intended use of the SPFs and availability of data.

3.4 Using CMFs with the Developed SPFs

CMFs can be used with the developed SPFs, just like in the HSM. This means that CMFs can be multiplied to the SPF to further alter the output of predicted crashes per year. However, the analyst using the SPFs should only utilize CMFs that have not already been considered in the data analysis. For example, suppose the following SPF was developed using the list of variables described in Section 3.2. The hypothetical SPF is shown below:

$$N_{SPF} = N \times L \times AADT^{\alpha_0} \times e^{\alpha_1 + (\alpha_2 LaneWidth + \alpha_3 ShoulderPavedWidth + \alpha_4 TruckPercent)}$$

The α_0 through α_4 are the regression coefficients. The SPF shows that lane width, shoulder width, and truck percent were found to be statistically significant in the dataset. This also means that curve presence, IRI, posted speed, and right shoulder total width were not found to be statistically significant in the dataset. The factors that were not found to be significant should not be used as CMFs. A CMF that addresses lighting or roadside hazard rating, for example. could be used as a CMF. It should also be mentioned that the use of CMFs should be based on the analyst's judgement.

4. Model Development and Methodology

4.1 Choosing the Method

Two SPF methods were evaluated in order to identify the method that best fit the needs of WisDOT: HSM calibration method and HSM guidelines for jurisdiction development. The HSM proposed guidelines for jurisdiction-specific SPF development was chosen, as opposed to the HSM calibration method. The HSM calibration method, outlined in the literature review, is beneficial because it can be directly compared with other states and with HSM values and only has a single base model. However, the method can be limited in evaluating the influence of predictor variables with regards to crash prediction and safety. It can also be difficult to identify appropriate combinations of variables when determining which variables to include and not include in the model. For example, the HSM has a set list of CMFs that are applied to the base model, regardless of how much each variable effects the dataset or regardless of how much a variable not listed as a CMF affects the dataset.

The proposed methodology for developing Wisconsin state specific SPFs, which follows the HSM guidelines for jurisdiction development, is similar to the HSM calibration method, but differs in some important areas. First, the HSM contains a set of facility types for roadway segments and intersections: rural two-lane two-way roads, rural multilane highways, urban and suburban arterials, and freeways *(1)*. The proposed methodology determines facilities based on the needs of WisDOT analysts, so the facility types may or may not overlap with those in the HSM. Next, the HSM identifies base conditions for the SPFs, while the proposed method does not require base conditions. The developed SPFs will also be able to provide basic comparison to developed SPFs from other states. Lastly, it was assumed that the values available for each of the influencing factors found in Meta-Manager were correct unless otherwise found incorrect through the filtering process and outlier analysis. Commonly, influencing factor values are validated on a location-by-location basis; however, this common practice was not followed in order to provide a more expedient methodology.

4.2 Choosing the Regression Model

The literature review provides overviews of each of the types of models that have been used in developing SPFs. The Negative Binomial regression model was chosen for the methodology because it accounts for overdispersion which has made it the more widely used model for SPF development and is recommended by the HSM *(8)*. Overdispersion is the existence of variability in a dataset. It is important to check if there is variability in the dataset used in SPF development because it may affect which model will be used. If overdispersion is present, the Negative Binomial regression method is ideal. If overdispersion is not present, meaning the mean of the data is equal to the variance, then the Poisson regression method would be more relevant. Equation 15 shows the basic form of the Negative Binomial regression model used in the project.

$$N_{SPF} = N \times L \times AADT^{\alpha_0} \times e^{\alpha_1 + (\alpha_2 X_1 + \alpha_3 X_2 + \cdots + \alpha_m X_n)} \qquad (15)$$

Where the variable α_0 is the coefficient of AADT and α_1 through $\alpha_{(m)}$ are coefficients for each of the influencing factors (X_1 through X_n), L represents segment length (miles), N_{SPF} represents the predicted number of crashes per year, and N is the number of total years of crash data.

5. Goodness-of-Fit Measures

The Negative Binomial regression model was chosen because of its ability to account for overdispersion. In addition to choosing a model, goodness-of-fit methods were selected in order to develop an optimal crash prediction model. Three parameters were used to analyze goodness-of-fit of the developed models: AIC, dispersion parameter, and cumulative residual (CURE) plots. The R package was used to develop the models. R^2 was not used as a goodness-of-fit measure because it is more commonly used with and is more accurate when used with linear regression models *(12)*.

5.1 Akaike Information Criterion

AIC is used to evaluate the goodness-of-fit of a model. Equation 16 shows the general case of AIC, where k is the number of parameters in the statistical model and L is the maximized log-likelihood for the model *(10)*.

$$\mathrm{AIC} = 2k - 2\ln(L) \tag{16}$$

The lower the value of AIC, the better the model. On the other hand, a higher log-likelihood value would indicate a better model. When comparing AIC values, the larger the difference, the stronger the evidence for one model over another model.

5.2 Dispersion Parameter

The dispersion parameter indicates the amount of variation found in data. Larger values are preferred. It will be calculated in the R package and summarized with the results of the model *(59)*. The dispersion parameter is used in conjunction with the generalized linear model.

5.3 Cumulative Residual Plots

CURE plots are figures that show how well a model fits the data. Residuals are defined as the “differences between the observed and fitted values of the response” and, when plotted cumulatively, demonstrate the suitability of a regression model *(60)*. The cumulative residuals are plotted on the y-axis. The observed crashes are found in the database and the predicted crashes are calculated using the developed models. The x-axis consists of one of the variables used to develop the model, filtered in ascending order. AADT is most commonly used on the x-axis. The 95th percentile confidence limits are plotted on the figure, as well *(59)*. Any residuals that fall outside the confidence limits show an area where the model does not fit the data well *(60)*. It is commonly accepted for up to 5% of the residuals to fall outside of the limits. This term is otherwise known as percent CURE deviation. Percent CURE deviation will be 0% when the model fits the data closely. When the model is not biased, cumulative residuals will oscillate tightly around the x-axis, or zero. Long up or down regions indicate areas of bias *(61)*. Any large jumps between residuals indicate areas where there may be outliers in the data. The CURE plots were created using The Calibrator Excel tool *(62)*.

6. Developing SPFs for Undivided Rural Two-Lane Roads

Chapter 6 discusses the process of developing SPFs for the undivided rural two-lane road facility, which includes data processing and model results.

6.1 Data Processing

Roadway and safety data were assembled from Meta-Manager. Meta-Manager data from May 2016 were utilized. The crash data were collected from the previous 5 year period, 2010-2014, for all crash types. Based on past studies, aside from AADT and segment length, curve presence, IRI, lane width, posted speed, truck percent, and shoulder width have proven to be statistically significant in developing SPFs. The following explains the process of filtering the data for undivided rural two-lane two-way roadway facilities.

The database contained a total of 19,623 segments, after combining the Excel files for the five regions. The database was first cleaned by removing segments with no AADT values and ramp segments. Ramp segments were identified by using ArcGlobe to visually classify each segment that was defined in the database as a one-way highway or an undivided one-lane roadway.

Undivided rural two-lane segments were identified by using peer groups ("SFTY_TRVL_CLS_CD") and functional classifications ("FCLASS"). Rural peer groups were 410: rural 2-lane highways with AADT ≤ 2,000 vpd, 420 (rural 2-lane highways with 2,000 vpd < AADT ≤ 7,000 vpd), 430 (rural 2-lane highways with AADT > 7,000 vpd), and 440 (low speed 2-lane highways). Segments with functional classifications less than 49 are considered rural and segments with functional classifications greater than or equal to 49 are considered urban. To filter the database for undivided rural two-lane segments, peer groups 410, 420, 430, and 440 were each selected. For peer group 440, the functional classes equal to or greater than 49 were removed. Then, any segments remaining that were not undivided or two-lane were removed. Finally, each of the influencing factors was evaluated for null values. Some of the null values overlapped with other null values. The total number of segments remaining for model development was 7,911. Figure 2 shows a breakdown of the segments.

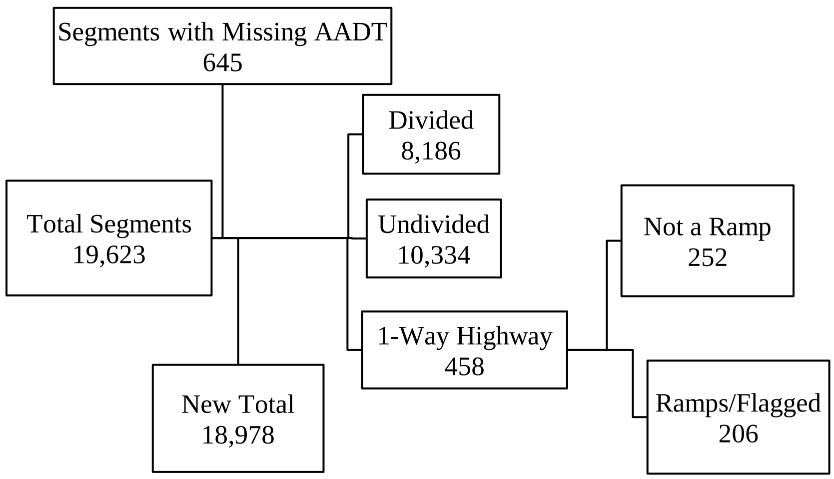


FIGURE 2 Basic break down of segments within Meta-Manager

As seen in Figure 2 Basic break down of segments within Meta-Manager, there were a total of 645 segments with missing AADT removed from the database. Remaining, there were 8,186 divided segments, 10,334 undivided segments, and 458 1-way highway segments. The database was then filtered to undivided rural two-lane roadway segments. Table 1 shows the order of filtering that was used.

TABLE 1 Order of filtering used in database

Filter	Number of segments after applying filter
Total segments in database	19,623
Removal of ramps and null AADT	18,978
Removal of all peer groups except 410, 420, 430, 440	10,670
Removal of all functional classes from peer group 440 that were greater than 49	9,731
Removal of segments with length less than 0.10 miles	9,344
Removal of segments classified as divided or as one-lane highways	8,805
Removal of segments with null truck percent	8,794
Removal of segments with null shoulder widths	8,790
Removal of segments with null curve presence values	8,694
Removal of segments with null IRI values	8,691
Removal of two segments with obvious posted speed outliers. One segment had a speed of 0 mph and one had a speed of 70 mph	8,689
Removal of segments with a lane width outside of the range 10ft to 12ft	8,322
Removal of segments with a right shoulder paved width outside of the range 0ft to 6ft	7,978
Removal of segments outside truck percent threshold of 0% to 30%	7,911

Curve presence, lane width, total crashes, and fatal and injury crashes were each calculated. Table 2 shows the mean, minimum, maximum, standard deviation, and median values of all the influencing factors included in the final SPF development. Lane width, right shoulder paved width, and truck percent have set thresholds that are based on design parameters in the state of Wisconsin, engineering judgement from WisDOT engineers, and outlier analysis. The plots used in the outlier analysis can be found in Appendix A. Paved shoulder width was used, instead of total shoulder width, because it was determined to be more reliable by WisDOT engineers.

TABLE 2 Descriptive statistics of variables

Variable	Mean	Minimum	Maximum	Standard Deviation	Median
AADT	3556	90	21800	2557	3020
*Curve Presence	0.25	0	1	0.43	0
Posted Speed (mph)	53.52	25	65	5.48	55
Right Shoulder Paved Width (feet)	2.79	0	6	1.46	3
Segment Length (mile)	0.99	0.10	2.60	0.36	1
Lane Width (feet)	11.86	10	12	0.36	12
Truck Percent (%)	11.63	0	30	3.30	11

*Note: Curve Presence is either 0 (no curve present) or 1 (yes there is a curve present).

IRI was removed from consideration in model development because the variable did not coincide well with WisDOT engineers' expected uses of the SPFs. As mentioned previously, when comparing an existing road to a proposed design or a proposed design to a proposed design, IRI cannot be used because there is no IRI value for a proposed road.

6.2 Model Results

As noted previously, SPFs were developed without IRI as a variable and with weather-related crashes included in the total crashes. For evaluation purposes models were also developed for all crash types for

- Models containing IRI and weather-related crashes
- Models containing no IRI, no weather-related crashes, and intersection-related crashes
- Models containing no IRI, weather-related crashes, and intersection-related crashes

The three model results can be found in Appendix B. Additionally, a model was created for fatal and injury crashes with IRI included in the analysis. The results can also be found in Appendix

B. Because there is no easy way of determining what ratio of weather-related crashes make up the fatal and injury crash total, no models were developed without weather-related crashes removed from the fatal and injury crash total.

6.2.1 All Crash Types

The preliminary steps in model creation involved developing a model with all AADT values. A CURE plot was then created to ascertain what AADT value(s) would create two, or more, better fitting models. Figure 3 shows the CURE plot for the model containing all AADT and Table 3 shows the model results. The x-axis is AADT in vehicles per day. There were 7,911 segments used in the analysis. In the table, the estimate is used to infer the value of an unknown statistic and the standard error measures the statistical accuracy of the estimate. The z-value represents how many standard deviations the variable is from the mean. The p-value indicates the level of significance. The ideal significance of the variables is 95%, so any p-value less than 0.05 is considered significant. The standardized coefficient is a calculated value that can be used to determine which variables are more important in the development of the regression model. Larger absolute values show a higher importance.

TABLE 3 Model results with all AADT for all crash types

Variable	Estimate	Standard Error	z-Value	$p\ (>\lvert z\rvert)$	Standardized Coefficient	AIC	Dispersion Parameter
All AADT							
(Intercept)	-2.974	0.357	-8.320	$< 2 \times 10^{-16}$	-	32417	4.031
AADT	0.839	0.015	55.926	$< 2 \times 10^{-16}$	0.204		
Segment Length	0.873	0.023	38.361	$< 2 \times 10^{-16}$	-0.023		
Lane Width	-0.158	0.030	-5.291	1.80×10^{-7}	-0.006		
Paved Shoulder Width	-0.062	0.007	-8.647	$< 2 \times 10^{-16}$	-0.004		
Posted Speed	-0.009	0.002	-5.101	1.22×10^{-7}	-0.017		
Truck Percent	-0.010	0.003	-3.655	0.000257	0.342		
Curve Presence	0.354	0.024	14.779	$< 2 \times 10^{-16}$	0.107		

AADT, segment length, and curve presence have positive estimates. Respectively, this means that vehicle volume increases with increasing crash rate, segment length increases with increasing crash rate, and there is an increased crash rate with the presence of a curve on a road segment. Lane width, paved shoulder width, posted speed, and truck percent have negative estimates. Lane width decreases as crash rate increases, paved shoulder width decreases as crash frequency increases, there is an increased crash rate with lower speeds, and truck percent decreases as crash rate increases. Curve presence, lane width, and paved shoulder width all match the expectation set by the literature review. Since truck percent and posted speed did not have a set expectation in the literature review, it is interesting to note the trend. All of the variables are significant to the 95th percentile confidence interval. The AIC value is 32,417 and the dispersion parameter is 4.031. Truck percent, AADT, and curve presence have the largest standard coefficient values.

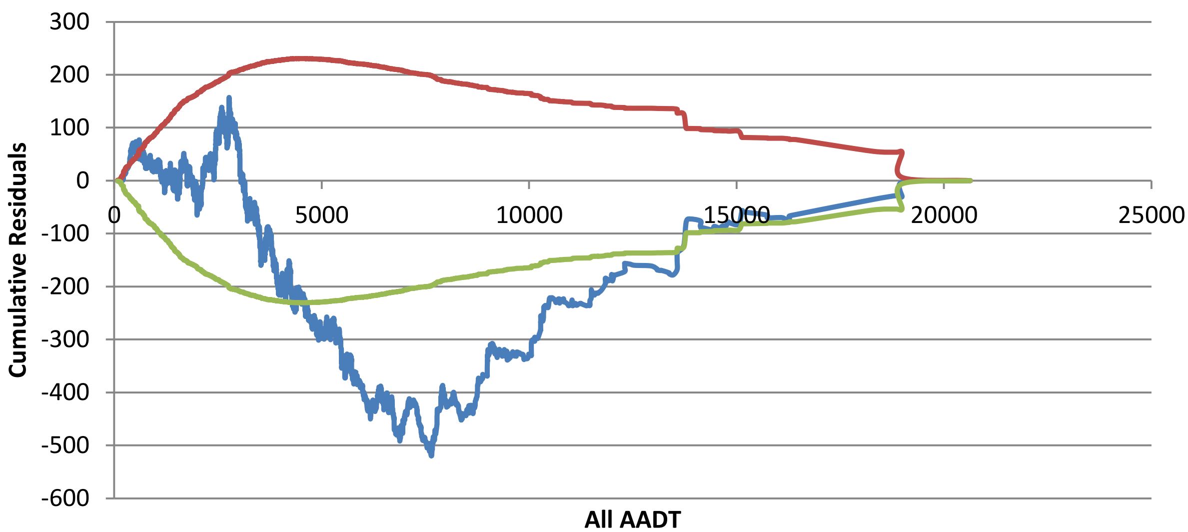


FIGURE 3 CURE plot for the model with all AADT for all crash types

The CURE plot, in Figure 3, has a percent CURE deviation of 29%. The CURE plot shows that the model under predicts crashes from 3,000 vpd to about 7,500 vpd. The model then starts to over predict crashes starting at about 7,500 vpd. Because of this clear dip, the model was split into two models at an AADT value of 7,500 vpd. The two models were developed and evaluated. Table 5 shows the results for the two models. There were 7,371 segments used in the analysis of the model with AADT less than or equal to 7,500 vpd and 540 segments used in the analysis of the model with AADT greater than 7,500 vpd.

Results from Table 3 and Figure 3 were compared to those of the models including intersection-related crashes in the crash total, located in Appendix B. Table 4 shows the difference in values between the two models.

TABLE 4 Comparison of SPFs, with all AADT, with and without weather-related crashes

	AIC	**Dispersion Parameter**	**Percent CURE Deviation**
(A) weather-related crashes and no intersection-related crashes	32,417	4.031	29%
(B) weather-related crashes and intersection-related crashes	37,520	3.0442	53%
(C) no weather-related crashes and intersection-related crashes	31,740	2.5355	57%

As seen in Table 4, the AIC value is smaller, the dispersion parameter is larger, and the percent CURE deviation is smaller when comparing (A) to (B), or no intersection-related crashes to intersection-related crashes. It is important to make sure that intersection-related crashes are removed from the data because, aside from being completely different facility types, they greatly affect the goodness-of-fit parameters. The AIC value is larger, the dispersion parameter is larger, and the percent CURE deviation is relatively smaller when comparing (B) to (C), or weather-related crashes to no weather-related crashes. The AIC value suggests the SPF without weather-related crashes fits the data better than the SPF with weather-related crashes; however, the dispersion parameter and percent CURE deviation conclude otherwise.

TABLE 5 Model results with AADT split for all crash types

Variable	Estimate	Standard Error	z-Value	p (> |z|)	Standardized Coefficient	AIC	Dispersion Parameter
AADT ≤ 7,500 vpd							
(Intercept)	-2.878	0.369	-7.810	5.71×10^{-15}	-	29107	4.083
AADT	0.797	0.017	45.633	$< 2 \times 10^{-16}$	0.225		
Segment Length	0.872	0.025	34.861	$< 2 \times 10^{-16}$	-0.023		
Lane Width	-0.141	0.030	-4.641	3.46×10^{-6}	-0.008		
Paved Shoulder Width	-0.061	0.008	-7.994	1.31×10^{-15}	-0.005		
Posted Speed	-0.009	0.002	-4.556	5.22×10^{-6}	-0.017		
Truck Percent	-0.009	0.003	-3.034	0.00242	0.408		
Curve Presence	0.351	0.025	14.275	$< 2 \times 10^{-16}$	0.124		
AADT > 7,500 vpd							
(intercept)	-8.994	1.166	-7.711	1.25×10^{-14}	-	3294	4.144
AADT	1.292	0.125	10.350	$< 2 \times 10^{-16}$	0.030		
Segment Length	0.901	0.056	16.197	$< 2 \times 10^{-16}$	-0.005		
Paved Shoulder Width	-0.052	0.021	-2.431	0.0151	-0.002		
Posted Speed	-0.010	0.004	-2.439	0.0147	1.659		

The variable estimates follow the same negative and positive trends as in Table 3 and are all significant to the 95th percentile confidence limit. It is not necessarily surprising that most of the variables were found significant to the 95th percentile confidence limit because each of the variables was found to affect data significantly in past studies, as mentioned in the literature review. Lane width, truck percent, and curve presence were not found significant to the 95th percentile in the model with AADT greater than 7,500 vpd and is most likely the reason why there was a significant break point at an AADT of 7,500 vpd. The AIC value and dispersion parameter is 29,107 and 4.083, respectively, for the model with AADT less than or equal to

7,500 vpd. The AIC value and dispersion parameter is 3,294 and 4.144, respectively, for the model with AADT greater than 7,500 vpd. AADT, truck percent, and curve presence are the most important variable predictors for the model with AADT less than or equal to 7,500 vpd. For the model with AADT greater than 7,500 vpd, posted speed has the largest standardized coefficient. Figure 4 and Figure 5 show the CURE plots for the two final models with AADT split at 7,500 vpd.

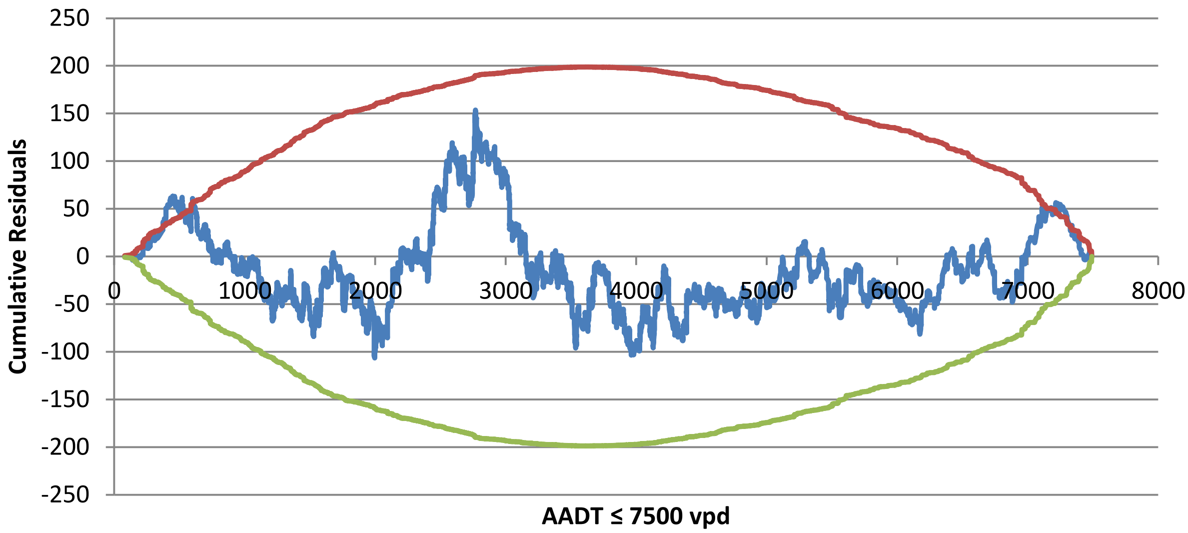


FIGURE 4 CURE plot for the model with AADT ≤ 7,500 vpd for all crash types

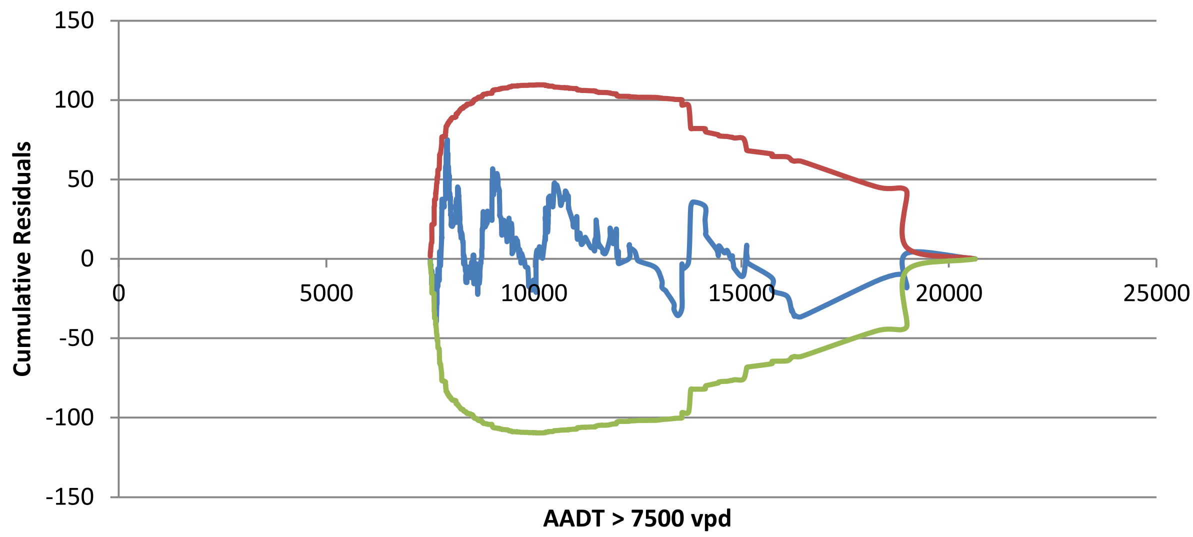


FIGURE 5 CURE plot for the model with AADT > 7,500 vpd for all crash types

Figure 4 and Figure 5 have percent CURE deviations of 3% and 0%, respectively. Both plots show the data oscillating around the x-axis. The percent CURE deviation values are below the recommended 5% which shows that both models fit the data well. The final two models are found in Equation 17 and Equation 18.

$$N_{AADT\leq 7500} = N \times L \times AADT^{\alpha_0} \times e^{\alpha_1+(\alpha_2(LW)+\alpha_3(PSW)+\alpha_4(PS)+\alpha_5(TP)+\alpha_6(CP))} \qquad (17)$$

$$N_{AADT>7500} = N \times L \times AADT^{\alpha_0} \times e^{\alpha_1+(\alpha_2(PSW)+\alpha_3(PS))} \qquad (18)$$

Where LW is lane width, PSW is paved shoulder width, PS is posted speed, TP is truck percent, and CP is curve presence. The α values correspond to the estimates in the table. For example, α_0 in Equation 17 is the value 0.797 from Table 5.

Figure 6 shows a regression tree analysis, developed by Khan, which validates splitting the SPFs at a significant break point in AADT *(63)*. Regression tree analysis shows variable significance through the tree hierarchy which can make it more intuitive in deciding which variables to include in a model. The data splits into two nodes where there is a statistically significant difference. Generally, “at each intermediate node, an observation goes to the left branch only if the condition is satisfied” *(10)*. At each node, this splitting process continues to occur until no statistically significant splits can be made. The regression tree analysis, Figure 6, was developed using the LogWorth statistic in the JMP pro software which is predictive analytics software that can be used in conjunction with SAS. The LogWorth statistic is calculated in Equation 19 *(64)*.

$$-\log_{10}(p - value) \tag{19}$$

The response variable, total crashes, is normalized by segment length, in miles. In the figure, the bold number indicates the total number of segments and the number in parenthesis and italics represents the mean.

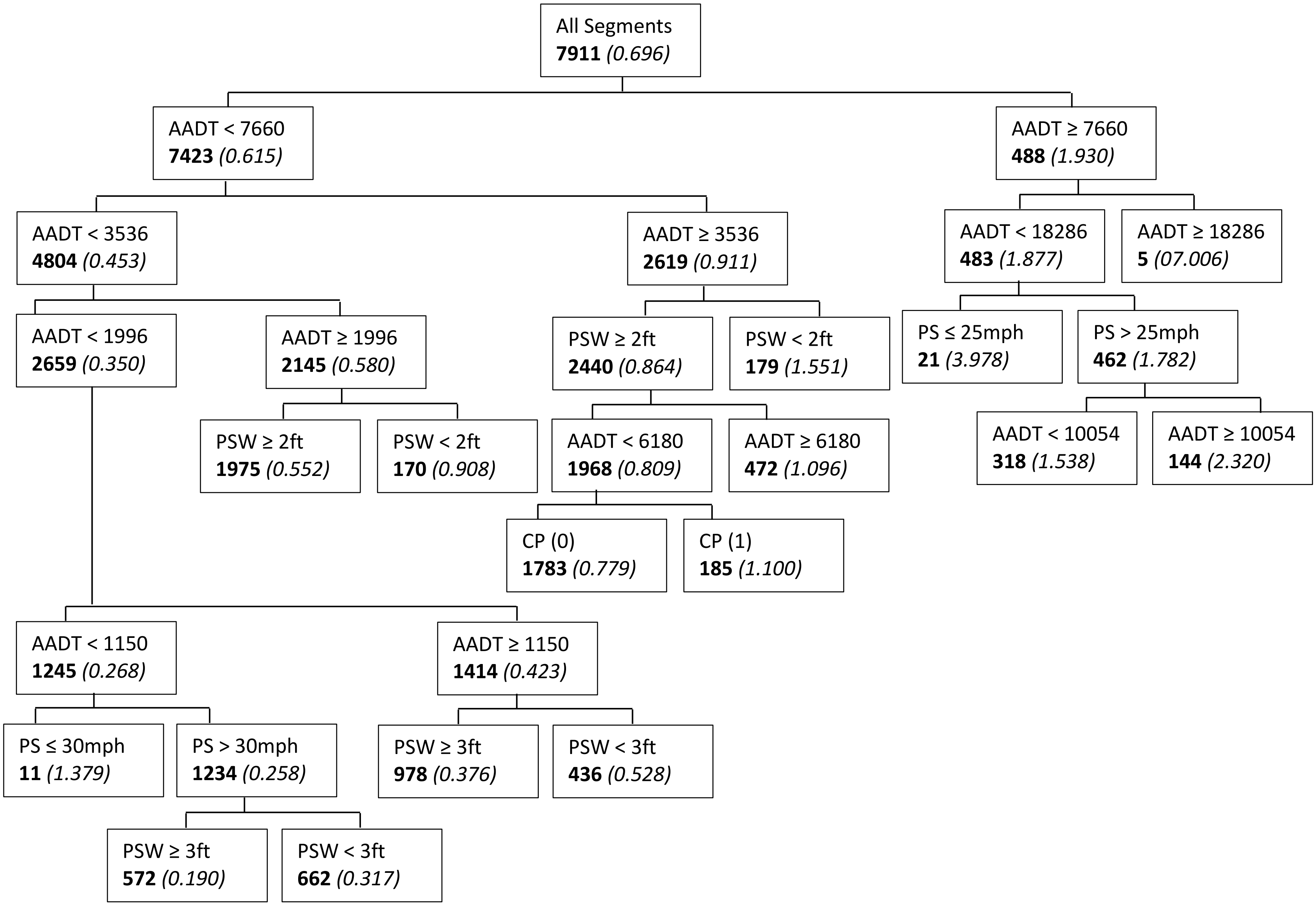


FIGURE 6 Regression tree analysis for total crashes

In Figure 6, the first split at AADT of 7,660 vpd validates the split used with the two SPFs developed for all crash types. The regression tree analysis shows that the first significant split in the data is AADT. The AADT is split at 7,660 vpd which is similar to the split point of the SPFs at 7,500 vpd using the CURE plot.

6.2.2 Fatal and Injury Crashes

The process used to develop the models for all crash types was used to develop the models for fatal and injury crashes. A model was developed for all AADT. The model results are located in Table 6 and the CURE plot can be seen in Figure 7.

TABLE 6 Model results with all AADT for fatal and injury crashes

Variable	Estimate	Standard Error	z-Value	p (> $\|z\|$)	Standardized Coefficient	AIC	Dispersion Parameter
All AADT							
(Intercept)	-4.739	0.453	-10.452	$< 2 \times 10^{-16}$	-	25769	2.781
AADT	0.941	0.019	48.311	$< 2 \times 10^{-16}$	0.224		
Segment Length	0.620	0.027	23.353	$< 2 \times 10^{-16}$	-0.036		
Lane Width	-0.164	0.040	-4.141	3.46×10^{-5}	-0.009		
Paved Shoulder Width	-0.058	0.009	-6.413	1.42×10^{-10}	-0.013		
Truck Percent	-0.019	0.004	-5.133	2.85×10^{-7}	0.431		
Curve Presence	0.288	0.031	9.187	$< 2 \times 10^{-16}$	0.185		

The variable estimates follow the same negative and positive trends as in Table 3. The variables are all significant to the 95th percentile confidence limit. The AIC value and dispersion parameter is 25,769 and 2.781, respectively. Truck percent, AADT, and curve presence have the largest standard coefficient values.

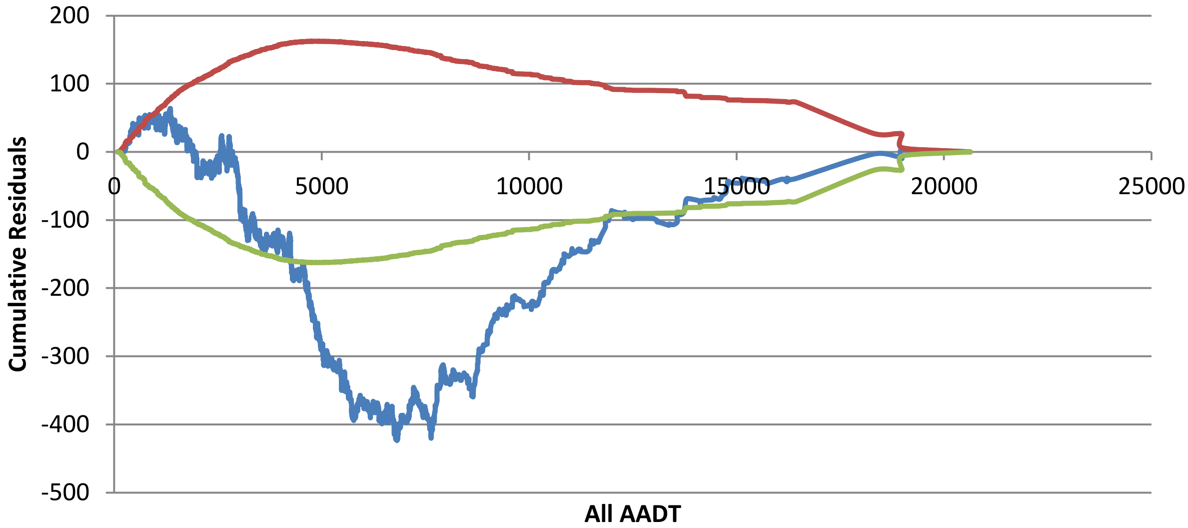


FIGURE 7 CURE plot for the model with all AADT for fatal and injury crashes

The CURE plot, in Figure 7, has a percent CURE deviation of 34%. The CURE plot shows a large spike in over predicted crashes from 3,100 vpd to about 7,000 vpd. The model then under predicts crashes starting at about 7,000 vpd. Because of this clear dip, the model was split into two models at an AADT value of 7,000 vpd. The two models were developed and evaluated. Table 7 shows the results for the two models. There were 7,243 segments used in the analysis of the model with AADT less than or equal to 7,000 vpd and 668 segments used in the analysis of the model with AADT greater than 7,000vpd.

TABLE 7 Model results with AADT split for fatal and injury crashes

Variable	Estimate	Standard Error	z-Value	p (> |z|)	Standardized Coefficient	AIC	Dispersion Parameter
AADT ≤ 7,000 vpd							
(Intercept)	-4.567	0.468	-9.763	$< 2 \times 10^{-16}$	-	22123	2.883
AADT	0.865	0.024	36.391	$< 2 \times 10^{-16}$	0.271		
Segment Length	0.655	0.031	21.403	$< 2 \times 10^{-16}$	-0.035		
Lane Width	-0.130	0.041	-3.191	0.00142	-0.013		
Paved Shoulder Width	-0.063	0.010	-6.316	2.68×10^{-10}	-0.014		
Truck Percent	-0.0170	0.004	-4.154	3.27×10^{-5}	0.500		
Curve Presence	0.267	0.033	8.218	$< 2 \times 10^{-16}$	0.219		
AADT > 7,000 vpd							
(intercept)	-9.548	1.257	-7.597	3.04×10^{-14}	-	3619	2.689
AADT	1.238	0.135	9.141	$< 2 \times 10^{-16}$	0.029		
Segment Length	0.532	0.053	10.041	$< 2 \times 10^{-16}$	-0.002		
Truck Percent	-0.015	0.009	-1.670	0.0949	0.214		
Curve Presence	0.254	0.126	2.011	0.0443	0.070		

The variable estimates follow the same negative and positive trends as in Table 3 and are all significant to the 95th percentile confidence limit except for truck percent in the model with AADT greater than 7,000 vpd which is significant to the 90th percentile confidence limit. The AIC value and dispersion parameter is 22,123 and 2.883, respectively, for the model with AADT less than or equal to 7,000 vpd. The AIC value and dispersion parameter is 3,619 and 2.689, respectively, for the model with AADT greater than 7,000 vpd. AADT, truck percent, and curve presence are the most important variable predictors for the model with AADT less than or equal to 7,000 vpd. For the model with AADT greater than 7,000 vpd, truck percent has the largest standardized coefficient. Figure 8 and Figure 9 show the CURE plots for the two final models with AADT split at 7,000 vpd.

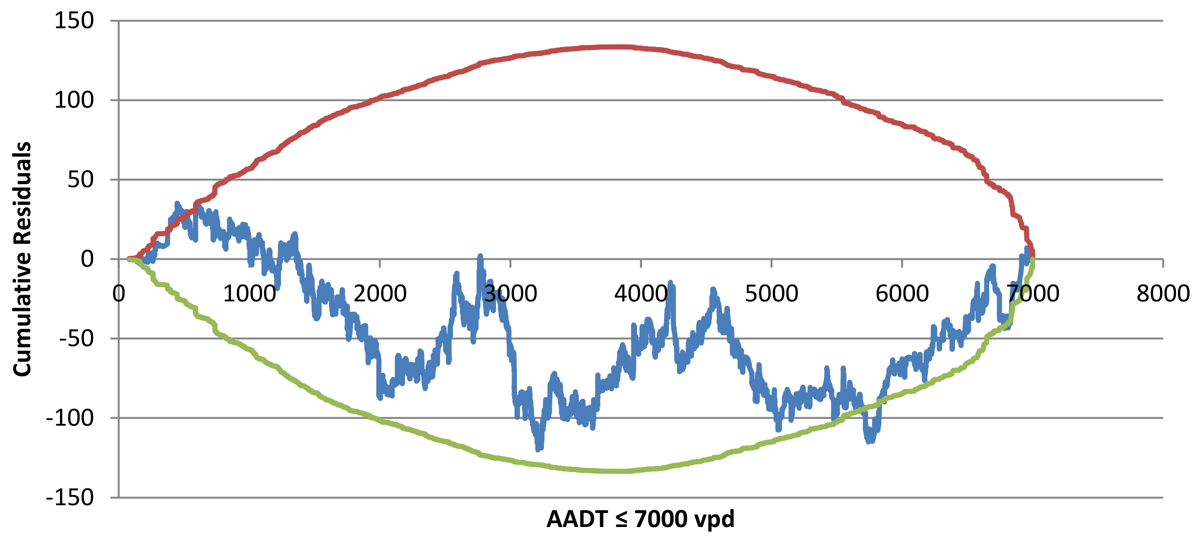


FIGURE 8 CURE plot for the model with AADT ≤ 7,000 vpd for fatal and injury crashes

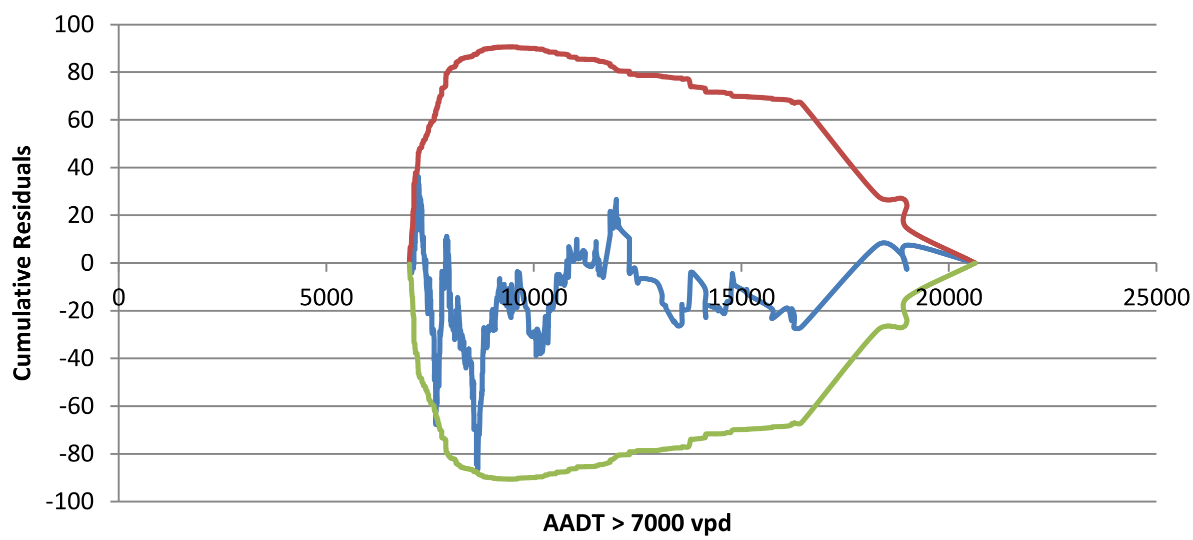


FIGURE 9 CURE plot for the model with AADT > 7,000 vpd for fatal and injury crashes

Figure 8 and Figure 9 have percent CURE deviations of 3% and 1%, respectively. The percent CURE deviation values are below the recommended 5% which shows that both models fit the data well. The final two models are found in Equation 19 and Equation 20.

$$N_{AADT \leq 7000} = N \times L \times AADT^{\alpha_0} \times e^{\alpha_1 + (\alpha_2(LW) + \alpha_3(PSW) + \alpha_4(TP) + \alpha_5(CP))} \quad (19)$$

$$N_{AADT>7000} = N \times L \times AADT^{\alpha_0} \times e^{\alpha_1 + (\alpha_2(TP) + \alpha_3(CP))} \tag{20}$$

7. Developing SPFs for Weather-Related Crashes

Chapter 7 discusses the development of an SPF specific to weather-related crashes. While developing the SPFs for undivided rural two-lane roads for all crash types and fatal and injury crashes, it was noted that weather-related crashes altered the goodness-of-fit parameters. Table 4 compares the goodness-of-fit parameters of the SPFs, with all AADT, created with and without weather-related crashes included in the total number of crashes. See Appendix B for the complete results of the three models compared in Table 4: model with weather-related crashes and no intersection-related crashes, model with weather-related crashes and intersection-related crashes, and model with no weather-related crashes and intersection-related crashes.

Because of the difference in the goodness-of-fit values, weather-related crashes were analyzed separately. Weather-related crashes were evaluated for relationships with roadway characteristics and compared to the SPFs developed for undivided rural two-lane roads with all crash types. Data processing is discussed in the first section, basic statistics in the second section, model results in the third section, and model applications in the fourth section.

7.1 Data Processing

The same filtered and processed Meta-Manager data were used to analyze the weather-related crashes as with all crash types and fatal and injury crashes. Additionally, the same 7,911 road segments were analyzed in relation to AADT, segment length, right shoulder paved width, curve presence, IRI, truck percent, and posted speed. Meta-Manager has a column specifically for weather-related crashes, titled “MMGR_WTHR_CRSH_TOT.” According to Meta-Manager, a weather-related crash is any crash with weather recorded as rain, snow, fog/smoke, or sleet/hail, as well as road conditions of wet, snow/slush, ice, or sand/mud/dirt/oil.

7.2 Basic Statistics for Weather-Related Variables

This section discusses general statistics of the weather-related crashes. Specific crash data were obtained from the WisTransportal website, a site that contains data from all reported crashes in Wisconsin. Each reported crash is given a unique accident number that is linked to a MV400 form. The crashes in the undivided rural two-lane road dataset were queried by using these accident numbers. The MV400 form allows crashes to be categorized by weather conditions and road conditions. There were a total of 37,464 crashes for undivided two-lane rural roadways. The percentage designated as weather-related crashes was 38.8%, or 14,522 crashes. A portion of the weather-related crash reports were not available, so 38.6% (14,448 crashes) of the overall crash total was used in the basic statistical analysis. 29.0% of weather-related crashes were fatal and injury crashes compared to 54.0% of the total crashes, which includes weather-related crashes, were fatal and injury crashes. Figure 10 and Figure 11 show the percentage of each type of weather condition and each type of road condition as a part of the total weather-related crashes.

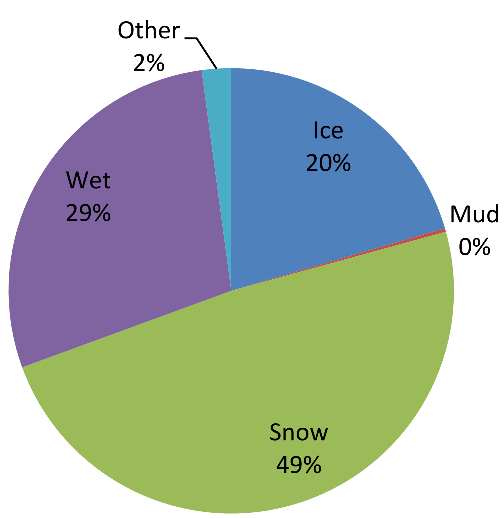


FIGURE 10 Percent of the types of road conditions

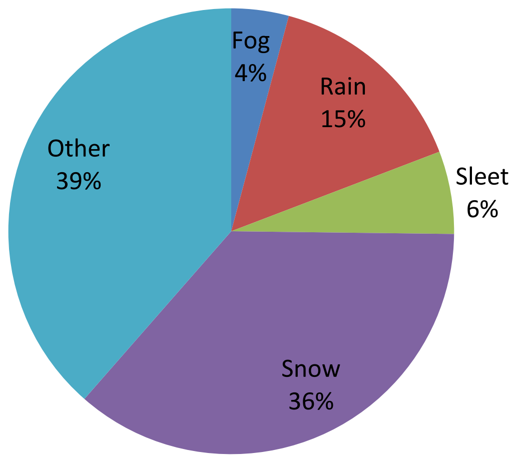
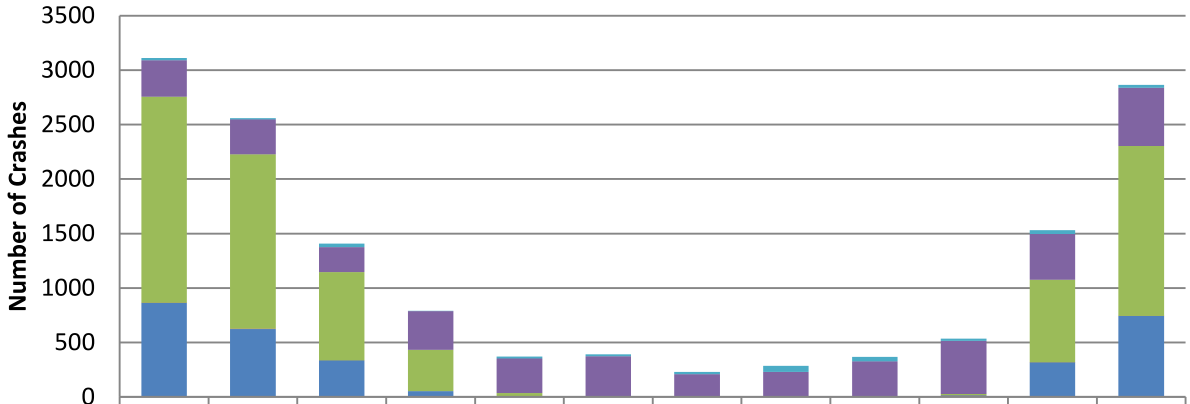


FIGURE 11 Percent of the types of weather conditions

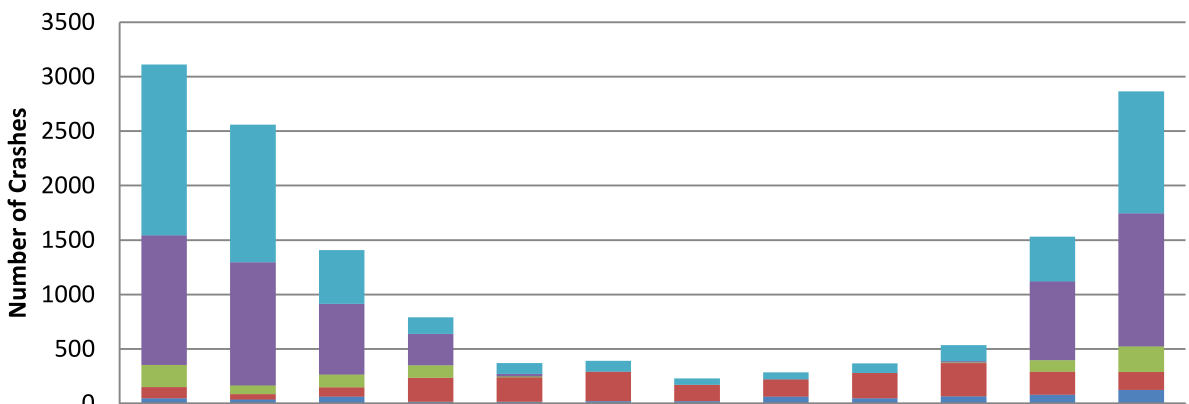
The category labelled "Other" consists of categories not considered to be either weather-related road conditions or weather-related weather conditions. For Figure 10, the "Other" category accounts for the crashes that had a weather-related weather condition but not a weather-related road condition. In Figure 11, the "Other" category consists of the crashes that had weather-related road conditions but not weather-related weather conditions. For the purpose of this statistical analysis, the "Other" category will not be the focus of the investigation. Figure 10 shows that crashes occurring on snowy roads are the most common at 49%, followed by wet roads with 29%, then icy roads with 20%. Figure 11 shows that crashes occur most commonly in snowy weather at 36%, followed by rain with 15% of the crashes, sleet with 6%, and fog with 4%. In both Figure 10 and Figure 11 crashes occur more commonly in snowy conditions. Figure 12 shows the weather-related crashes, categorized by road conditions, by month. Figure 13 categorizes the weather-related crashes by weather conditions, grouped by month. Figure 14 shows total crashes assembled by month.



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Other	19	12	31	4	20	18	22	57	43	21	35	25
Wet	334	321	229	356	316	367	201	223	322	488	420	535
Snow	1892	1599	810	378	28					10	758	1560
Mud	1	3	2		4	6	8	6	4	2	1	1
Ice	864	624	335	54	4					15	317	743

Month

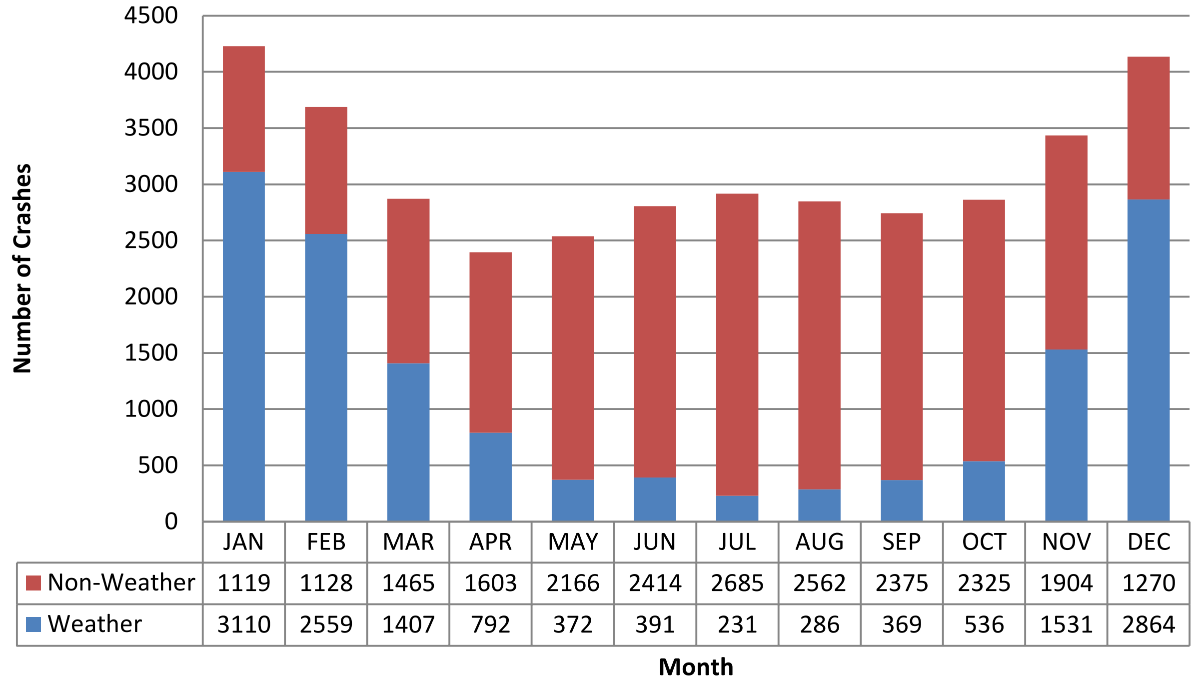
FIGURE 12 Weather-related crashes grouped by road conditions, per month



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Other	1568	1262	493	154	102	100	61	64	90	146	410	1120
Snow	1189	1132	650	289	23					12	724	1219
Sleet	202	80	116	112	9					6	105	237
Rain	103	50	86	220	221	269	150	160	232	307	211	163
Fog	48	35	62	17	17	22	20	62	47	65	81	125

Month

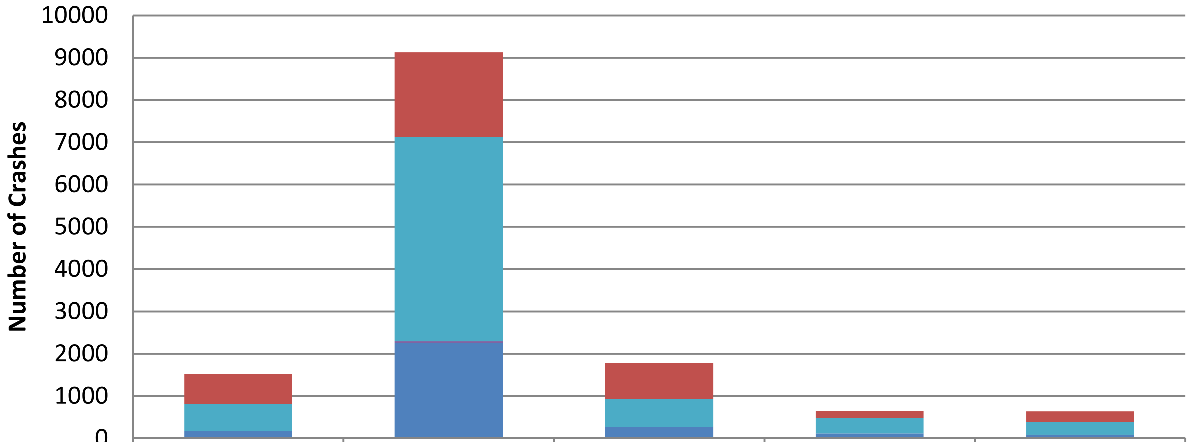
FIGURE 13 Weather-related crashes grouped by weather conditions, per month



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Non-Weather	1119	1128	1465	1603	2166	2414	2685	2562	2375	2325	1904	1270
Weather	3110	2559	1407	792	372	391	231	286	369	536	1531	2864

FIGURE 14 Total crashes grouped by non-weather or weather crashes, per month

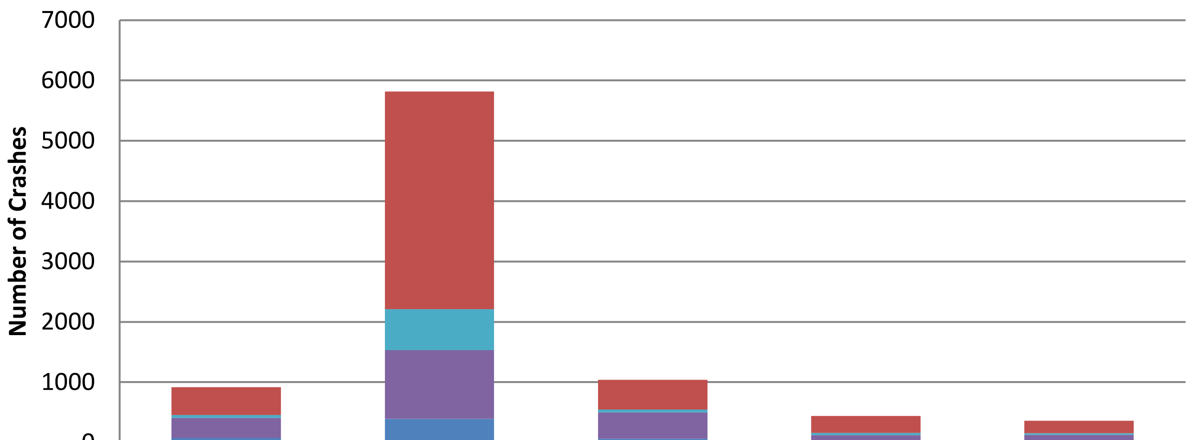
Figure 12 and Figure 13 show that a higher number of weather-related crashes occur during months that are known for experiencing snow in Wisconsin. The months, April through October, have relatively low crash totals, predominately consisting of crashes that occur on rainy days. Crashes occurring in snowy road conditions and/or snowy weather conditions are more common in January, February, March, November, and December. Comparatively, Figure 14 shows that non-weather-related crashes make up a higher percentage of April through October months. Figure 15 shows crashes, categorized by road conditions, grouped by crash type. Figure 16 shows crashes, categorized by weather conditions, by crash type. Figure 17 shows total crashes split by crash type. The five crash types included were the most common out of all types of crashes in the dataset. SSOP and SSS stand for sideswipe/opposite direction and sideswipe/same direction, respectively.



	Angle	No Collision	Rear End	SSOP	SSS
Wet	703	2001	853	165	251
Snow	645	4831	658	369	296
Mud	3	32	2		1
Ice	164	2259	267	111	85

Crash Type

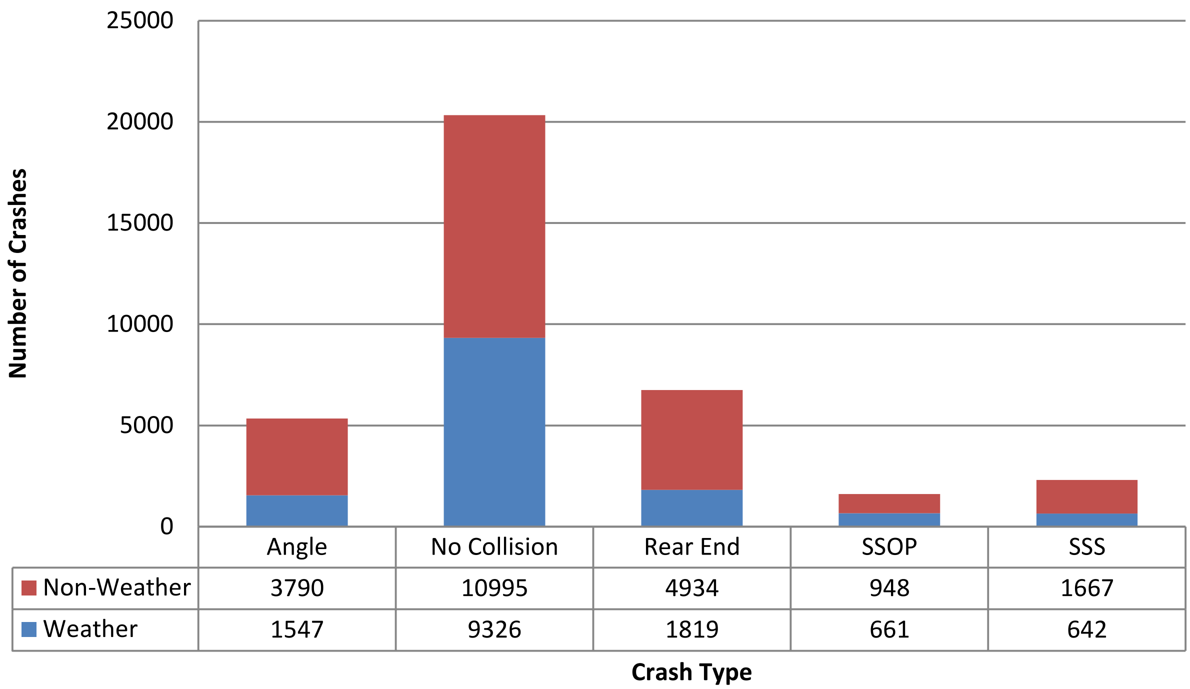
FIGURE 15 Weather-related crashes grouped by road conditions, per crash type



	Angle	No Collision	Rear End	SSOP	SSS
Snow	462	3609	493	281	207
Sleet	50	680	50	37	23
Rain	330	1134	431	94	111
Fog	76	395	66	28	19

Crash Type

FIGURE 16 Weather-related crashes grouped by weather conditions, per crash type



	Angle	No Collision	Rear End	SSOP	SSS
Non-Weather	3790	10995	4934	948	1667
Weather	1547	9326	1819	661	642

FIGURE 17 Total crashes grouped by non-weather or weather crashes, per month

Figure 15 and Figure 16 show that no collision crashes, meaning no collision with another vehicle, are the most common in weather-related crashes. Snow makes up the highest proportion of no collision crashes for both road condition and weather condition. After snow, rain/wet is the next most common condition across all crash types. Figure 17 shows that there is no weather-related crash type that does not follow the trend of the crash types for total crashes.

The basic statistical analysis shows that weather-related crashes most commonly occur with snowy weather conditions or snowy road conditions. Non-weather-related crashes occur more often during April through October, while weather-related crashes occur more frequently during November, December, January, and February. However, this may be because a larger number of weather events may occur during November, December, January, and February, meaning fewer clear weather days during those months. The trends in crash type occur similarly between non-weather-related crashes and weather-related crashes, with no collision crashes with the highest proportion of crashes. Snowy weather conditions and snowy road conditions make up the largest proportion of all crash types, with rain next, followed by icy road conditions.

Solely focusing on weather-related variables will not explain why there are a number of crashes occurring during any weather-related event. It is challenging to pinpoint if a crash occurred only because of weather or if a crash occurred because of another factor. For example, if a vehicle rounded a curve in a road during snowy conditions and crashed, was the crash caused because of poor visibility, low surface friction due to the snow, or something else? The next section, Section 7.3, describes weather-related SPFs that were developed specifically with geometric roadway variables. The variables were used because they could be compared to the previously developed SPFs and also because they could be changed, at any specific site, in order to address noticeable trends in weather-related crashes. As mentioned previously, one example of weather-related crash trends involves the west-north ramp of Marquette Interchange in downtown Milwaukee. The ramp received a high friction surface treatment because there was a noticeable trend of crashes occurring on rainy weather days.

7.3 Model Results

The preliminary steps in model creation involved developing a model with all AADT values. A CURE plot was then created to ascertain what AADT value(s) would create two, or more, better fitting models. Table 8 shows the model results and Figure 18 shows the CURE plot for the model containing all AADT. The x-axis, in Figure 18, is AADT in vehicles per day.

TABLE 8 Model results with all AADT for weather-related crashes

Variable	Estimate	Standard Error	z-Value	$p\ (> \lvert z \rvert)$	Standardized Coefficient	AIC	Dispersion Parameter
All AADT							
(Intercept)	-4.578	0.463	-9.896	$< 2 \times 10^{-16}$	-	26215	2.861
AADT	0.957	0.019	49.276	$< 2 \times 10^{-16}$	0.241		
Segment Length	0.684	0.028	24.675	$< 2 \times 10^{-16}$	-0.034		
Lane Width	-0.157	0.039	-4.053	5.05×10^{-5}	-0.008		
Paved Shoulder Width	-0.052	0.009	-5.743	9.30×10^{-9}	-0.004		
Posted Speed	-0.006	0.002	-2.849	0.00438	-0.049		
Truck Percent	-0.020	0.004	-5.463	4.69×10^{-8}	0.498		
Curve Presence	0.341	0.031	11.126	$< 2 \times 10^{-16}$	0.184		

The variable estimates follow the same negative and positive trends as in Table 3 and are all significant to the 95th percentile confidence limit. It is not possible to compare the two models more directly because the SPF for all crash types, Table 3, does not contain intersection-related crashes while the SPF for only weather-related does contain intersection-related crashes. As stated at the end of Section 3.1, intersection-related crashes can only be removed from the overall crash total. With this in mind, it can still be evaluated that both models find the same variables to be significant in crash prediction. The AIC value and dispersion parameter is 26,220 and 2.853, respectively, for the preliminary weather-related model. Truck percent, AADT, and curve presence have the largest standard coefficient values.

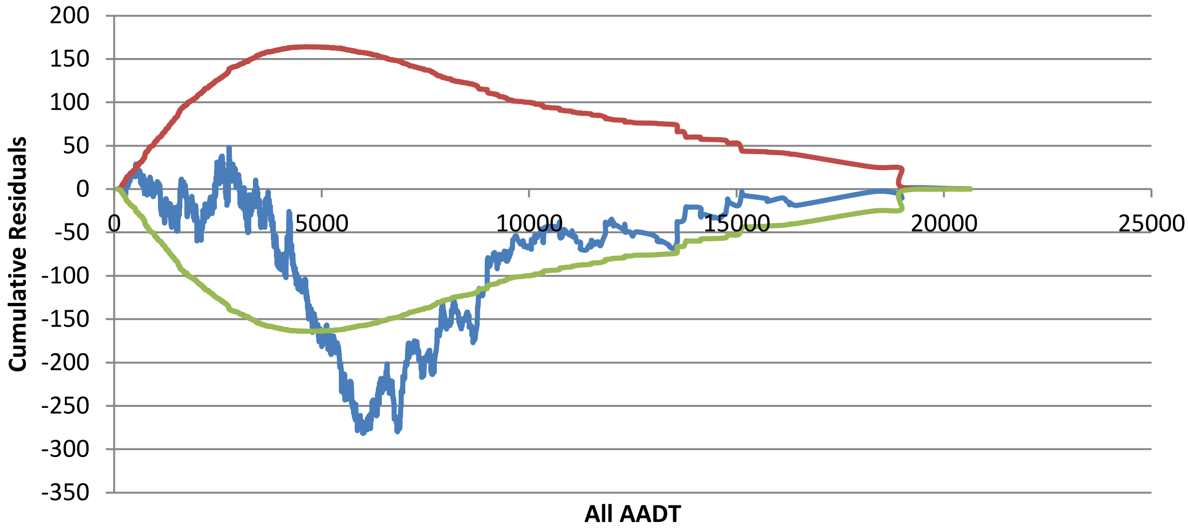


FIGURE 18 CURE plot for the model with all AADT for weather-related crashes

The percent CURE deviation is 19%. The CURE plot shows that the model under predicts crashes from about 4,000 vpd to about 6,000 vpd. The model then starts to over predict crashes starting at 6,000 vpd. Because of this clear dip, the model was split into two models at an AADT value of 6,000 vpd. The two models were developed and evaluated. Table 9 shows the results for two models with AADT split at 6,000 vpd. The first model contains AADT values less than or equal to 6,000 vpd and the second model contains AADT values greater than 6,000 vpd. There were 6,838 segments used in the analysis of the model with AADT less than or equal to 6,000 vpd and 1,073 segments used in the analysis of the model with AADT greater than 6,000 vpd.

TABLE 9 Model results with AADT split for weather-related crashes

Variable	Estimate	Standard Error	z-Value	p (> |z|)	Standardized Coefficient	AIC	Dispersion Parameter
AADT ≤ 6,000 vpd							
(Intercept)	-4.748	0.498	-9.531	$< 2 \times 10^{-16}$	-	20802	2.824
AADT	0.924	0.026	35.435	$< 2 \times 10^{-16}$	0.265		
Segment Length	0.667	0.033	19.982	$< 2 \times 10^{-16}$	-0.032		
Lane Width	-0.120	0.041	-2.969	0.00299	-0.012		
Paved Shoulder Width	-0.056	0.010	-5.382	7.37×10^{-8}	-0.005		
Posted Speed	-0.006	0.003	-2.234	0.02547	-0.062		
Truck Percent	-0.022	0.004	-5.145	2.68×10^{-7}	0.653		
Curve Presence	0.349	0.033	10.422	$< 2 \times 10^{-16}$	0.235		
AADT > 6,000 vpd							
(intercept)	-2.536	2.060	-1.231	0.21835	-	5414	3.027
AADT	1.047	0.097	10.831	$< 2 \times 10^{-16}$	0.047		
Segment Length	0.705	0.050	14.229	$< 2 \times 10^{-16}$	-0.071		
Lane Width	-0.413	0.158	-2.611	0.00904	-0.004		
Posted Speed	-0.009	0.004	-2.546	0.01088	0.394		
Curve Presence	0.184	0.084	2.192	0.02835	0.081		

The variable estimates follow the same negative and positive trends as in Table 3 and are all significant to the 95th percentile confidence limit. The AIC value and dispersion parameter is 20,802 and 2.824, respectively, for the model with AADT less than or equal to 6,000 vpd. The AIC value and dispersion parameter is 5,414 and 3.027, respectively, for the model with AADT greater than 6,000 vpd. AADT, truck percent, and curve presence are the most important variable predictors for the model with AADT less than or equal to 6,000 vpd. For the model with AADT greater than 6,000 vpd, posted speed has the largest standardized coefficient. Figure 19 shows the CURE plot for the model with AADT less than or equal to 6,000 vpd. Figure 20 shows the CURE plot for the model with AADT greater than 6,000 vpd.

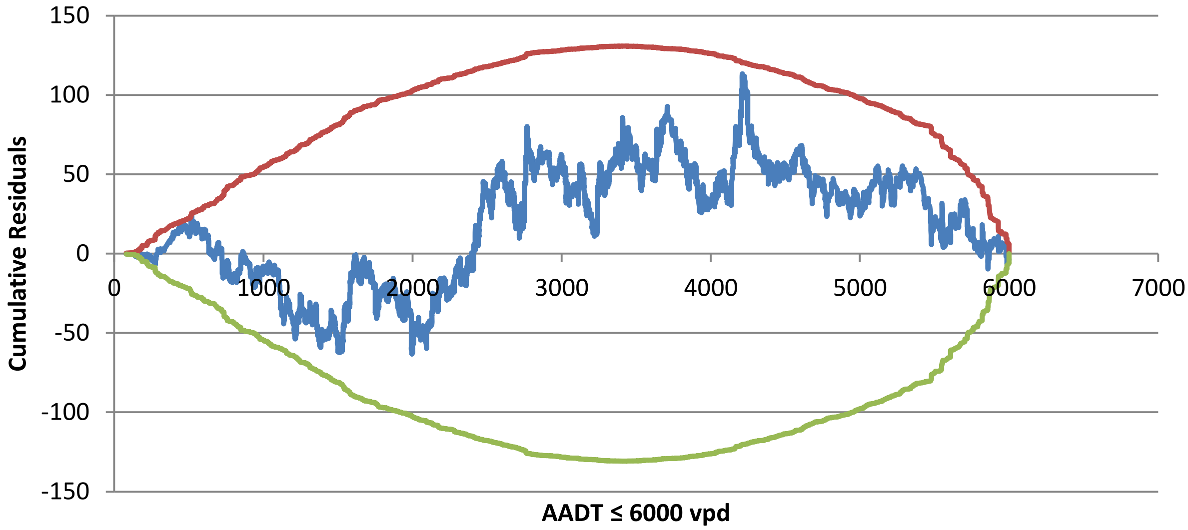


FIGURE 19 CURE plot for the model with AADT ≤ 6000 vpd for weather-related crashes

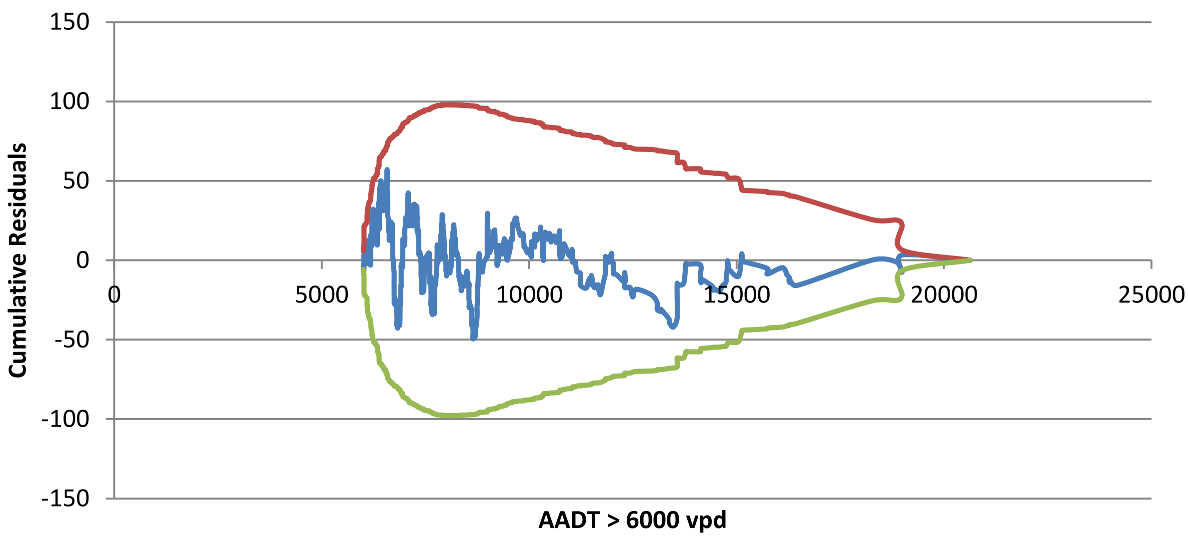


FIGURE 20 CURE plot for the model with AADT > 6000 vpd for weather-related crashes

The percent CURE deviation is 0% for both Figure 19 and Figure 20. The percent CURE deviation values are below the recommended 5% which shows that both models fit the data well. The final two models are found in Equation 22 and Equation 23.

$$N_{AADT\leq 6000} = N \times L \times AADT^{\alpha_0} \times e^{\alpha_1+(\alpha_2(LW)+\alpha_3(PSW)+\alpha_4(PS)+\alpha_5(TP)+\alpha_6(CP))} \qquad (22)$$

$$N_{AADT>6000} = N \times L \times AADT^{\alpha_0} \times e^{\alpha_1 + (\alpha_2(LW) + \alpha_3(PS) + \alpha_4(CP))} \quad (23)$$

7.4 Model Applications

The proposed weather-related SPFs have a few potential applications. The weather-related SPFs can be used alongside the total crash and fatal and injury SPFs, developed in Chapter 6. The SPFs can be used in planning and design, to determine expected safety impacts of design changes, and to compare proposed alternatives for a site. The weather-related SPFs can also be used to identify locations with promise. In this way, locations in need of safety improvements specifically related to weather can be identified. To do so, the observed number of crashes at a site would be compared to the predicted number of crashes, calculated by the SPFs. The percent change would be calculated, Equation 24.

$$Percent\ Change = \frac{Observed - Predicted}{Predicted} \times 100 \quad (24)$$

Any site with the observed number of crashes larger than the predicted number of crashes, or percent change greater than zero, would be flagged. Each of the flagged sites would be ranked in order of the largest percent change to the smallest percent change. The flagged sites at the top of the list would have top priority for potential improvements. This method would provide another way of determining how to allocate available funding.

8. Conclusions

The objective of the project was to develop a methodology for creating SPFs in the state of Wisconsin. The methodology was developed using the HSM proposed jurisdiction development guidelines illustrated by developing SPFs for undivided rural two-lane roadway segments.

Crash and roadway segment data were filtered using a series of outlined steps. AADT, segment length, curve presence, IRI, lane width, posted speed, truck percent, and shoulder width were proven to be statistically significant in developing SPFs. Outliers were identified by plotting each of the variables versus total crashes and the types of crashes (weather-related, deer, etc) to include in the SPFs will be left to the judgement of the analyst. The Negative Binomial model and backwards stepwise regression was used to create the models within the R package. Three common goodness-of-fit parameters were chosen to evaluate the regression models: AIC, dispersion parameter, and CURE plots.

AADT values were used to split the data at significant break points which ended up being 7,500 vpd for all crash types and 7,000 vpd for fatal and injury crashes. CURE plots were used to visually identify the AADT value at which to split the data. A regression tree analysis was used to corroborate the use of splitting the preliminary SPFs by AADT. As a result of the methodology, four models were developed for undivided rural two-lane roadway segments: one with AADT ≤ 7,500 vpd for all crash types, one with AADT > 7,500 vpd for all crash types, one with AADT ≤ 7,000 vpd for fatal and injury crashes, and one with AADT > 7,000 vpd for fatal and injury crashes. The two SPFs for all crash types and the SPF with AADT ≤ 7,000 vpd for fatal and injury crashes were significant to the 95th percentile confidence limit and the SPF with

AADT > 7,000 vpd for fatal and injury crashes was significant to the 90th percentile confidence limit.

Weather-related crashes were then analyzed and compared to the SPFs developed. Weather-related crashes were found to increase the dispersion parameter and the AIC value when added to the total number of crashes for all crash types. Two SPFs were developed for weather-related crashes: one with AADT ≤ 6,000 vpd and one with AADT > 6,000 vpd. The models were significant to the 95th percentile confidence limit. When comparing the all AADT model with only weather-related crashes and the all AADT model including weather related crashes, the influencing variables were found to be the same.

8.1 Future Research

The methodology of the data processing will be used to develop further Wisconsin state specific SPFs, including, but not limited to, interchange ramps, freeway segments, and interchange segments for six-lane freeways, four-lane rural freeways, and four-lane urban freeways. Additionally, the reliability of the fatal and injury crash SPFs could be investigated by finding and removing each intersection-related crash that could not be removed from the fatal and injury crash total. In another light, regression tree analysis could be used to develop SPFs instead of CURE plots. Other influencing factors could, also, be considered in developing SPFs like using a regional variable or using more specific curve information like curve length or curve radius.

With regards to weather-related crashes, much more analysis can be performed with the regression models. This completed research only analyzed roadway geometry; however, analyzing the effects of weather-related variables (weather conditions, roadway conditions, etc) would provide a bigger picture of the factors effecting weather-related crashes.

9. References

1. AASHTO Highway Safety Manual. American Association of State Highway and Transportation Officials. First Edition. Washington, D.C., 2010.
2. Srinvasan, R. and Bauer, K. Safety Performance Function Development Guide: Developing Jurisdiction Specific SPFs. Federal Highway Administration. FHWA-SA-14-005.
3. Wisconsin Strategic Highway Safety Plan 2014-2016. Wisconsin Department of Transportation. 2014.
4. *Crossroads*. Wisconsin Transportation Information Center. Spring 2014. http://epdfiles.engr.wisc.edu/pdf_web_files/tic/crossroads/xrds_2014_1.pdf .
5. Garber, N. and Rivera, G. Safety Performance Functions for Intersections on Highways Maintained by the Virginia Department of Transportation. Report No. FHWA/VTRC 11-CR 1, Federal Highway Administration, 2010.
6. Srinivasan, R., Carter, D. and Bauer, K. Safety Performance Function Decision Guide: SPF Calibration vs SPF Development. Federal Highway Administration, 2013, Report No. FHWA-SA-14-004.
7. Srinivasan, S., Haas, P., Dhakar, N., Hormel, R., Torbic, D. and Harwood, D. Development and Calibration of Highway Safety Manual Equations for Florida Conditions. Florida Department of Transportation, 2011.
8. Srinivasan, R. and Carter, D. Development of Safety Performance Functions for North Carolina. Federal Highway Administration, FHWA/NC/2010-09, 2011.
9. Strathman, J. G., Duecker, K. J., Zhang, J. and Williams, T. Analysis of Design Attributes and Crashes on the Oregon Highway System. Federal Highway Administration, 2001, Report No. FHWA-OR-RD-02-01.
10. Khan, G., Bill, A. R., Chitturi, M. V. and Noyce, D. A. Safety Evaluation of Horizontal Curves on Rural Undivided Roads. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2386, 2013, pp. 147-157.
11. Hauer, E. Statistical Road Safety Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 2004, No.1897, pp.81-87.
12. Washington, S.P., Karlaftis, M.G. and Mannering, F.L. *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman & Hall/CRC, Boca Raton, 2003.
13. Kononov, J., Bailey, B., and Allery, B.K. Relationships between Safety and Both Congestion and Number of Lanes on Urban Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, 2008, No.2083. pp.26-39.
14. Xie, Y., Lord, D., and Zhang, Y. Predicting Motor Vehicle Crashes using Bayesian Neural Network Models. *Accident Analysis and Prevention*, 2007, Vol.39, pp.922-933.
15. Williamson, M. and Zhou, H. Develop Calibration Factors for Crash Prediction Models for Rural Two-Lane Roadways in Illinois. 8th International Conference on Traffic and Transportation Studies, Changsha, China, 2012.
16. Mehta, G. and Lou, Y. Safety Performance Function Calibration and Development for the State of Alabama: Two-Lane Two-Way Rural Roads and Four-Lane Divided Highways. 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.
17. Tegge, R., Jo, J. and Ouyang, Y. Development and Application of Safety Performance Functions for Illinois. Federal Highway Administration, 2010, Report No. FHWA-ICT-10-066.

18. Lubliner, H., Bornheimer, C., Schrock, S., Wang, M. and Fitzsimmons, E. Evaluation of Interactive Highway Safety Design Model Crash Prediction Tools for Two-Lane Rural Roads on Kansas Department of Transportation Projects. Kansas Department of Transportation, 2014, KU-10-1R.
19. Donnell, E., Gayah, V. and Jovanis, P. Safety Performance Functions. Federal Highway Administration, 2014, FHWA-PA-2014-007-PSU WO 1.
20. Rios, B. and Javier, D. Development of Crash Modification Factors for Rumble Strips Treatment for Freeway Applications: Phase I Development of Safety Performance Functions. Presented at the Latin American and Caribbean Conference for Engineering and Technology, Guayaquil, Ecuador, 2014.
21. Saito, M., Brimley, B. and Schultz, G. Transportation Safety Data and Analysis Volume 2: Calibration of the Highway Safety Manual and Development of New Safety Performance Functions. The Utah Department of Transportation, 2011.
22. Kweon, Y. and Lim, I. Development of Safety Performance Functions for Multilane Highway and Freeway Segments Maintained by the Virginia Department of Transportation. Federal Highway Administration, 2014, FHWA/VCTIR 14-R14.
23. Khan, G., A. R. Bill, M. Chitturi, and D. A. Noyce. Horizontal Curves, Signs, and Safety. In Transportation Research Record: Journal of the Transportation Research Board, No. 2279, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 124–131.
24. Miaou, S.-P., and H. Lum. Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. In Transportation Research Record 1407, TRB, National Research Council, Washington, D.C., 1993, pp. 11–23.
25. Anderson, I. B. and R. A. Krammes. Speed Reduction as a Surrogate for Accident Experience at Horizontal Curves on Rural Two-Lane Highways. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1701, 2000.
26. Zegeer, C., R. Stewart, D. Reinfurt, F. Council, T. Neuman, E. Hamilton, T. Miller, and W. Hunter, Cost Effective Geometric Improvements for Safety Upgrading of Horizontal Curves, Report No. FHWA-RD-90-021, Federal Highway Administration, 1990.
27. Chan, C. Y., Huang, B., Yan, X. and Richards, S. Investigating Effects of Asphalt Pavement Conditions on Traffic Accidents in Tennessee Utilizing Pavement Management System (PMS). *Journal of Advanced Transportation*, Vol. 44, 2010, pp. 150-161.
28. Hauer, E. Shoulder Width and Safety. March 2000. https://www.researchgate.net/publication/261363624_Lane_width_and_safety_Literature
29. Hauer, E. Speed and Safety. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2103, 2009, pp. 10-17.
30. Elvik, R., Christensen, P. and Amundsen, A. Speed and Road Accidents: An Evaluation of the Power Model. The Institute of Transport Economics (TOI), 2004, TOI Report 740/2004.
31. Methods and Practices for Setting Speed Limits: An Informational Report. Federal Highway Administration, 2012, FHWA-SA-12-004.
32. Ran, R. and Lee, C. Effects of Geometric and Traffic Factors on Frequencies of Truck-Involved Crashes on Ontario Highways. Presented at 2015 CTRF Annual Conference. Montreal, Quebec. 2015.

33. Dissanayake, S. and Amarasingha, N. Effects of Geometric Design Features on Truck Crashes on Limited-Access Highways. Mid-America Transportation Center. 2012.
34. Stamatiadis, N., Pigman, J., Sacksteder, J., Ruff, W. and Lord, D. Impact of Shoulder Width and Median Width on Safety. NCHRP Report 633. Transportation Research Board, 2009.
35. Hauer, E. and Bamfo, J. Two Tools for Finding What Function Links the Dependent Variable to the Explanatory Variables, Proceedings International Cooperation on Theories and Concepts in Traffic Safety. Lund, Sweden, 1997.
36. Lord, D., Washington, S. P. and Ivan, J. Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and Prevention*, 2005, Vol. 37, pp.35-46.
37. Koorey, G. Road Data Aggregation and Section Considerations for Crash Analysis, *Transportation Research Record: Journal of the Transportation Research Board*, No. 2103, Washington, D.C., 2009, pp. 61-68.
38. El-Basyouny, K., Barua, S., Islam, M.T. and Li, R. Assessing the Effect of Weather States on Crash Severity and Type by Use of Full Bayesian Multivariate Safety Models. *Transportation Research Record: Journal of Transportation Research Board*, 2014, No.2432, pp.65-73.
39. Van Schalkwyk, I., Mitra, S. and Washington, S. Incorporating Weather Into Region-Wide Safety Planning Prediction Models. Transportation Research Board 85th Annual Meeting 2006 Compendium of Papers, Washington, D.C., 2006, pp.1-17.
40. Yu, R. and Abdel-Aty, M. Developing Safety Performance Functions for Mountainous Freeway. 16th Road Safety on Four Continents Conference. Beijing, China.
41. Yu, R. and Abdel-Aty, M. Developing Hierarchical Bayesian Safety Performance Functions Using Real-Time Weather and Traffic Data. 16th Road Safety on Four Continents Conference. Beijing, China, 2013.
42. Shankar, V., Mannering, F. and Barfield, W. Effect of Roadway Geometrics and Environmental Factors on rural Freeway Accident Frequencies. *Accident Analysis and Prevention*, 1995, Vol.27, pp.371-389.
43. Jung, S., Qin, X. and Noyce, D.A. Rainfall Effect on Single-Vehicle Crash Severities Using Polychotomous Response Models. *Accident Analysis and Prevention*, 2010, Vol.42, pp.213-224.
44. Satterthwaite, S.P. An Assessment of Seasonal and Weather Effects on the Frequency of Road Accidents in California. *Accident Analysis and Prevention*, 1976, Vol.8, pp.87-96.
45. Khattak, A.J., Kantor, P. and Council, F.M. Role of Adverse Weather in Key Crash Types on Limited-Access Roadways. *Transportation Research Record: Journal of the Transportation Research Board*, 1998, No.1621, pp.10-19.
46. Andrey, J. Long-Term Trends in Weather-related Crash Risks. *Journal of Transport Geography*, 2010, Vol.18, pp.247-258.
47. Jung, S., Qin, X. and Noyce, D.A. Injury Severity of Multivehicle Crash in Rainy Weather. *Journal of Transportation Engineering*, 2012, Vol.138, pp.50-59.
48. Jung, S., Qin, X. and Noyce, D.A. Modeling Highway Safety and Simulation in Rainy Weather. Transportation Research Record: Journal of the Transportation Research Board, 2012, No.2237, pp.134-143.

49. Qiu, L. and Nixon, W. Effects of Adverse Weather on Traffic Crashes. Transportation Research Record: Journal of the Transportation Research Board, 2008, No.2055, pp.139-146.
50. Khattak, A.J. and Knapp, K.K. Snow Event Effects on Interstate Highway Crashes. Journal of Cold Regions Engineering, 2001, Vol.15, No.4, pp.219-229.
51. Hermans, E., Brijs, T., Stiers, T. and Offermans, C. The Impact of Weather Conditions on Road Safety Investigated on an Hourly Basis. Transportation Research Board 85th Annual Meeting, Washington, D.C., 2006.
52. El Basyouny, K. and Kwon, D. Assessing Time and Weather Effects on Collision Frequency by Severity in Edmonton Using Multivariate Safety Performance Functions. *Transportation Research Board 91st Annual Meeting*, Washington, D.C., 2012.
53. Brijs, T., Karlis, D. and Wets, G. Studying the Effect of Weather Conditions on Daily Crash Counts Using a Discrete Time-Series Model. *Accident Analysis and Prevention*, 2008, Vol. 40, No. 3, pp.1180-1190.
54. Usman, T., Fu, L. and Miranda-Moreno, L.F. A Disaggregate model for Quantifying the Safety Effects of Winter Road Maintenance Activities at an Operational Level. *Accident Analysis and Prevention*, 2012, Vol. 48, pp.368-378.
55. Bergel-Hayat, R., Debbarh, M., Antoniou, C. and Yannis, G. Explaining the Road Accident Risk: Weather Effects. *Accident Analysis and Prevention*, 2013, Vol.60, pp.456-465.
56. Ahmed, M.M., Abdel-Aty, M. and Yu, R. Assessment of Interaction of Crash Occurrence, Mountainous Freeway Geometry, Real-Time Weather, and Traffic Data. *Transportation Research Record: Journal of the Transportation Research Board*, 2012, No. 2280, pp.51-59.
57. Fridstrom, L., Ifver, J., Ingebrigtsen, S., Kulmala, R. and Thomsen, L.K. Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in Road Accident Counts. *Accident Analysis and Prevention*, 1995, pp.1-20.
58. Eisenberg, D. The Mixed Effects of Precipitation on Traffic Crashes. *Accident Analysis and Prevention*, 2004, Vol.36, pp.637-647.
59. Lin, D.Y., Wei, L.J. and Ying, Z. Model-Checking Techniques Based on Cumulative Residuals. *Biometrics*, 2004, Vol.58, pp.1-12.
60. Hauer, E. Safety Performance Functions: A Workshop. 2013.
61. Thompson, S., Roche, J., Rivera, Y., Gross, F. and Green, E. The Calibrator: How to Calibrate and Evaluate Safety Performance Functions. FHWA. 21 July 2016.
62. The Calibrator. Federal Highway Administration. 2016.
63. Khan, G. Personal interview. 12 January 2017.
64. Partition Method. JMP. Accessed 14 April 2017. http://www.jmp.com/support/help/Statistical_Details_8.shtml#1277120

10. Appendix

10.1 Appendix A

Scatter plots used in outlier analysis for SPFs

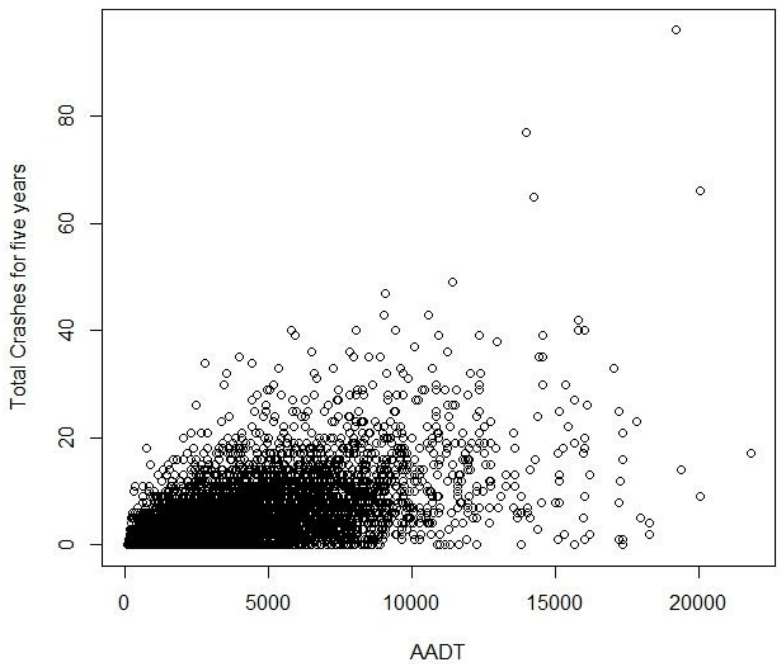


FIGURE 1A AADT plotted against total crashes for five years

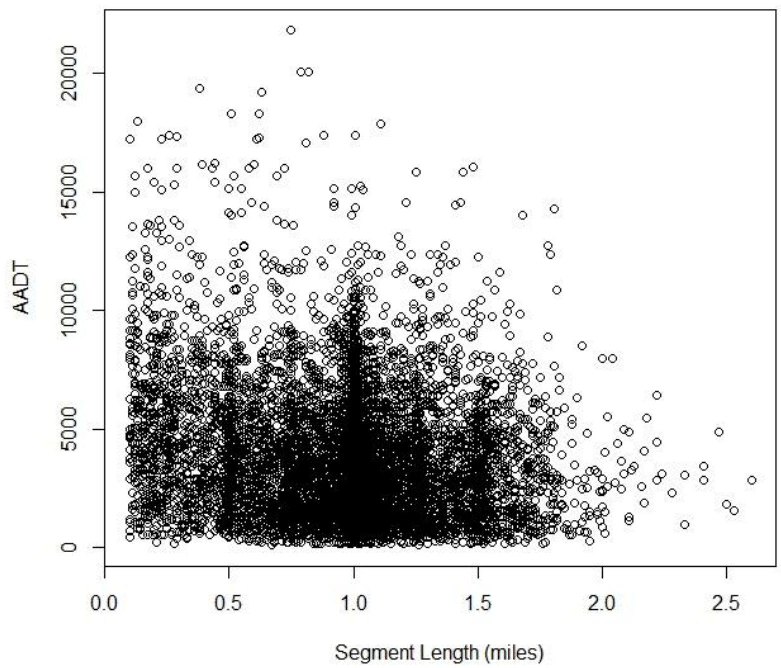


FIGURE 2A Segment length plotted with AADT

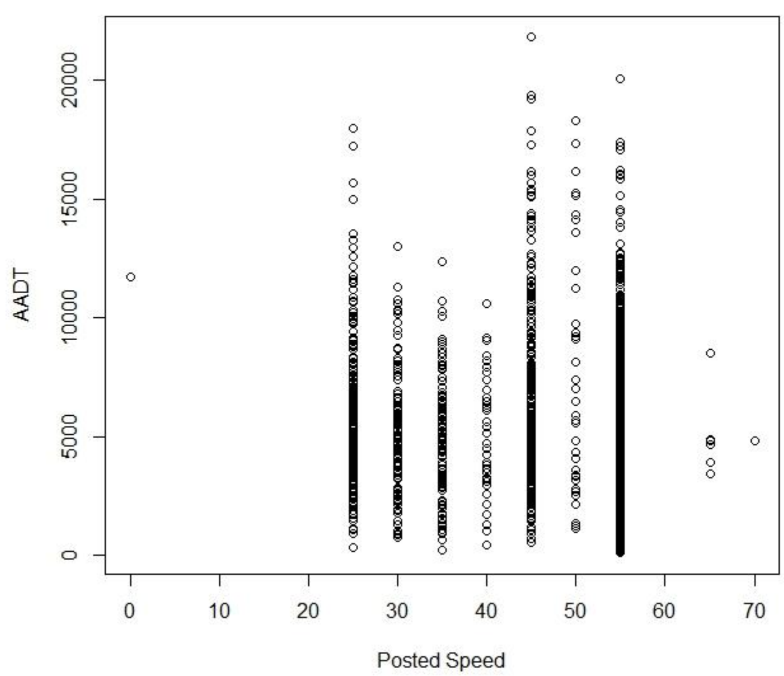


FIGURE 3A Posted speed versus AADT

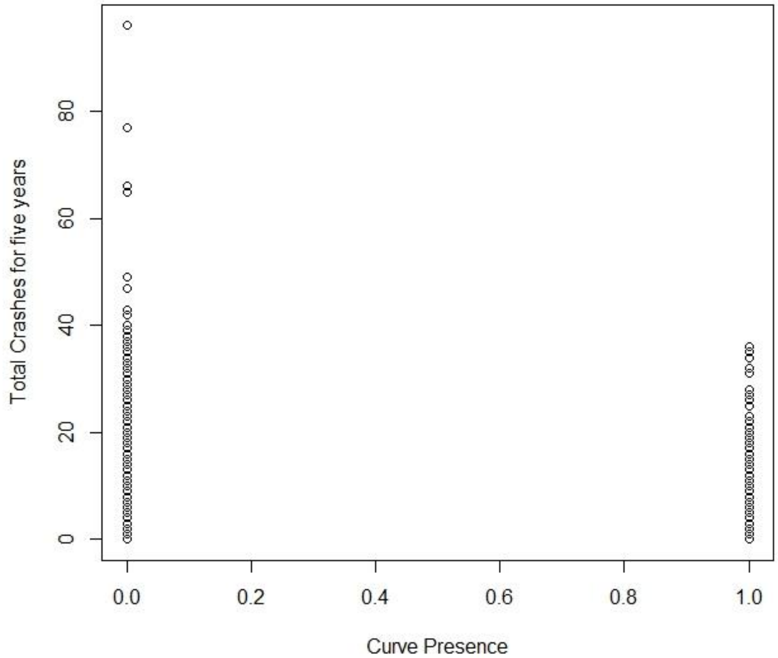


FIGURE 4A Curve presence (yes “1” or no “0”) versus total crashes for five years

TABLE 1A Minimum, maximum, and average values for curve presence

Minimum	Maximum	Average
0	1	0.240

TABLE 2A Number of segments for each curve presence value

	0	1
Number of Segments	6607	2084

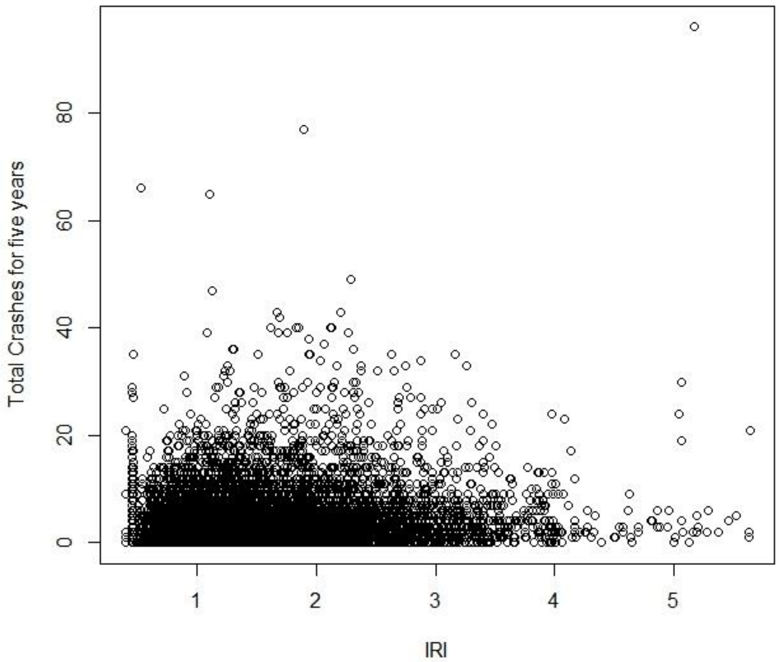


FIGURE 5A IRI versus total crashes for five years

TABLE 3A Minimum, maximum, and average values for IRI

Minimum	Maximum	Average
0.402	5.632	1.591

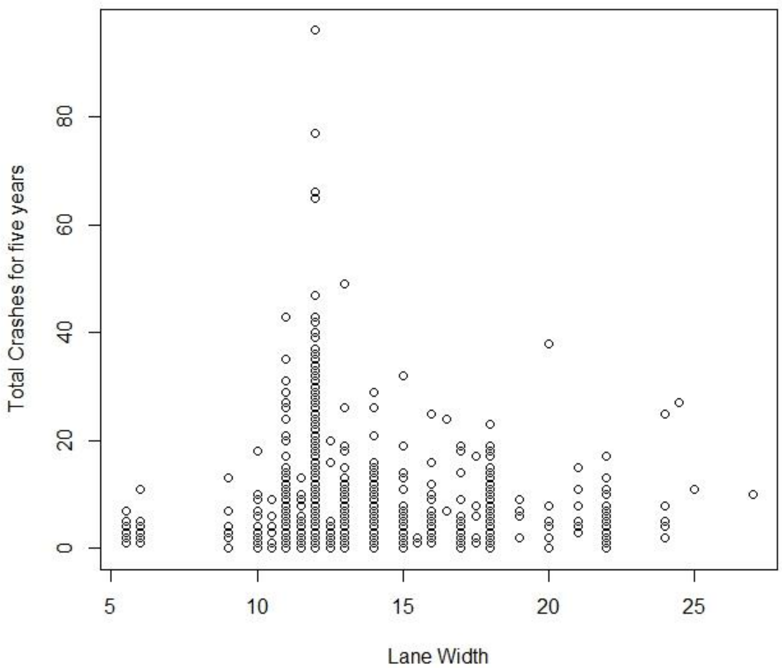


FIGURE 6A Lane width (ft) versus total crashes for five years

TABLE 4A Minimum, maximum, and average values for lane width

Minimum	Maximum	Average
5.5ft	27ft	12ft

TABLE 5A Number of segments for each lane width value

	5.5ft	6ft	9ft	10ft	10.5ft	11ft	11.5ft	12ft	12.5ft
Number of Segments (Number of Crashes)	8 (24)	16 (54)	10 (42)	28 (94)	9 (32)	1053 (3458)	39 (135)	7195 (36121)	29 (83)
	13ft	**14ft**	**15ft**	**15.5ft**	**16ft**	**16.5ft**	**17ft**	**17.5ft**	**18ft**
Segments (Crashes)	42 (361)	66 (425)	29 (162)	2 (3)	28 (180)	2 (31)	25 (140)	6 (35)	61 (406)
	19ft	**20ft**	**21ft**	**22ft**	**24ft**	**24.5ft**	**25ft**	**27ft**	
Segments (Crashes	6 (28)	6 (57)	6 (46)	17 (101)	5 (44)	1 (27)	1 (11)	1 (10)	

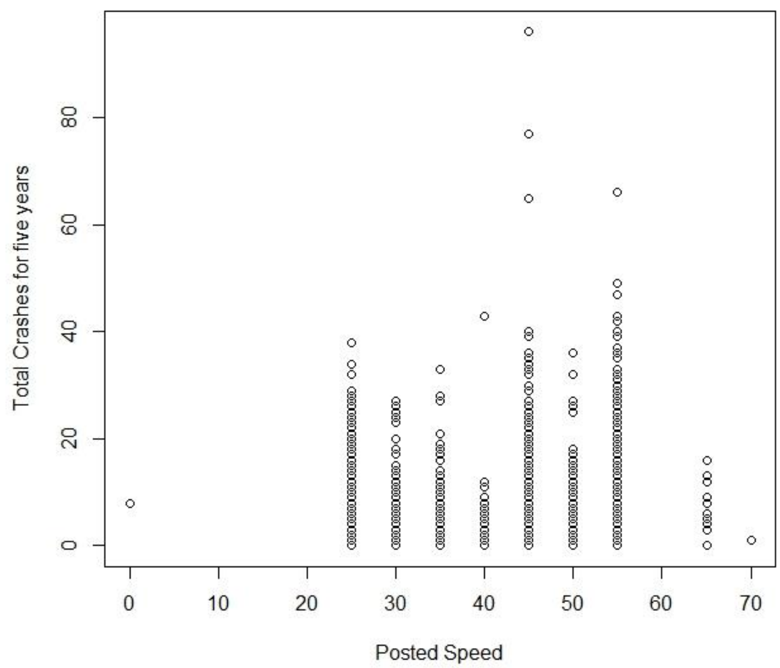


FIGURE 7A Posted speed (mph) versus total crashes for five years

TABLE 6A Minimum, maximum, and average values for posted speed

Minimum	Maximum	Average
0mph	70mph	53mph

TABLE 7A Number of segments for each posted speed value

	0mph	25mph	30mph	35mph	40mph
Number of Segments (Number of Crashes)	1 (8)	251 (1897)	123 (767)	151 (803)	42 (214)
	45mph	**50mph**	**55mph**	**65mph**	**70mph**
Segments (Crashes)	339 (3038)	64 (613)	7704 (34664)	15 (105)	1 (1)

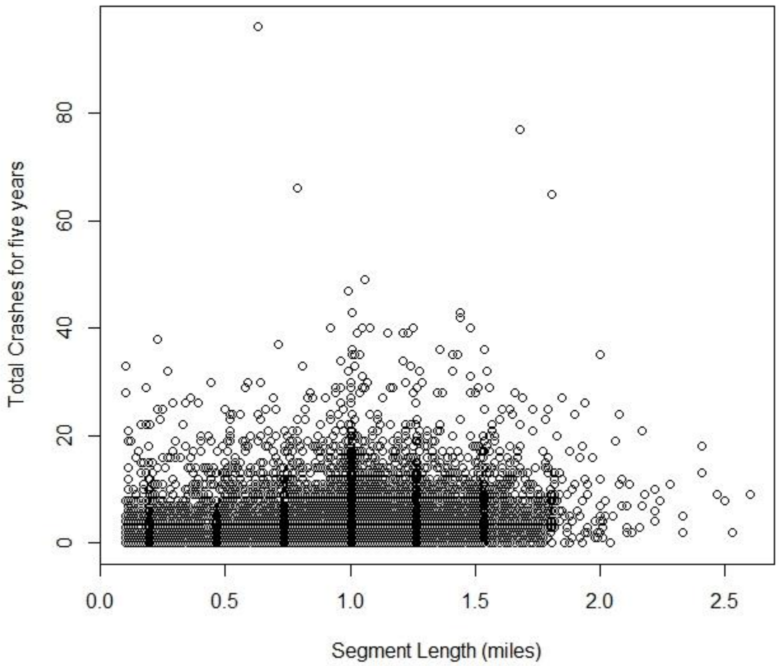


FIGURE 8A Segment length (miles) versus total crashes for five years

TABLE 8A Minimum, maximum, and average values for segment length

Minimum	Maximum	Average
0.10mi	2.60mi	0.96mi

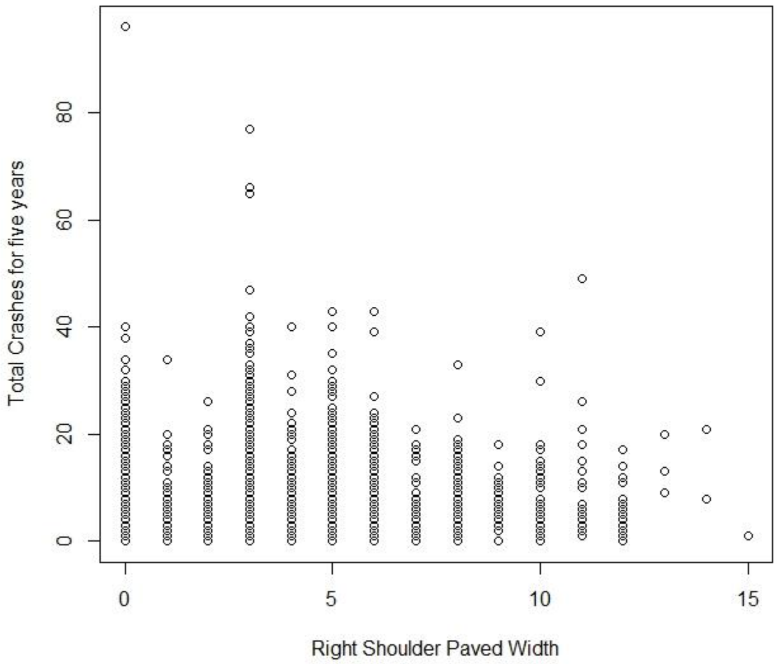


FIGURE 9A right shoulder paved width (ft) versus total crashes for five years

TABLE 9A Minimum, maximum, and average values for right should paved width

Minimum	Maximum	Average
0ft	15ft	3ft

TABLE 10A Number of segments for each right shoulder paved width value

	0ft	1ft	2ft	3ft	4ft	5ft	6ft	7ft
Number of Segments (Number of Crashes)	1454 (6438)	169 (631)	207 (880)	5099 (24202)	419 (1978)	736 (4217)	243 (1415)	48 (275)
	8ft	**9ft**	**10ft**	**11ft**	**12ft**	**13ft**	**14ft**	**15ft**
Segments (Crashes)	183 (1142)	28 (180)	58 (346)	19 (205)	22 (129)	3 (42)	2 (29)	1 (1)

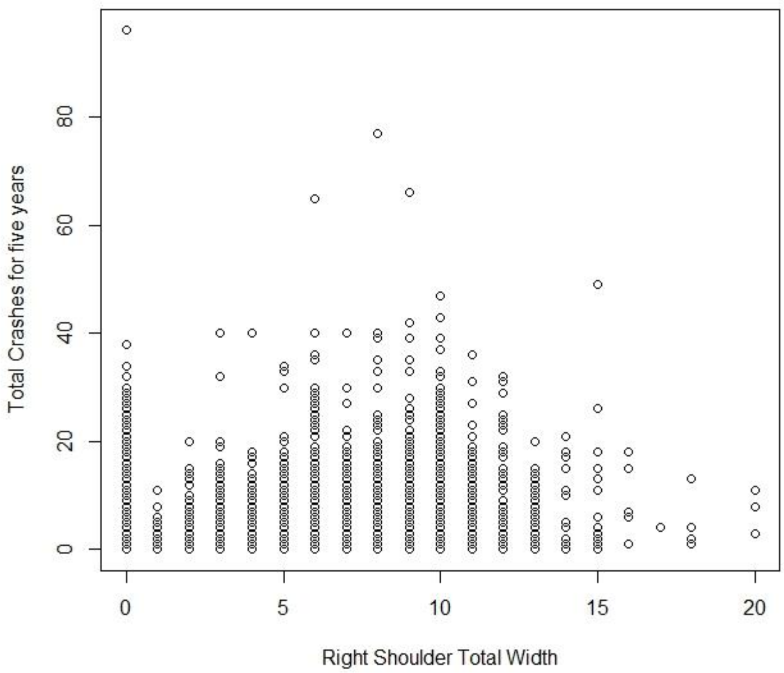


FIGURE 10A Right shoulder total width (ft) versus total crashes for five years

TABLE 11A Minimum, maximum, and average values for right shoulder total width

Minimum	Maximum	Average
0ft	20ft	7ft

TABLE 12A Number of segments for each right shoulder total width value

	0ft	**1ft**	**2ft**	**3ft**	**4ft**	**5ft**	**6ft**
Number of Segments (Number of Crashes)	454 (3480)	69 (187)	321 (898)	424 (1462)	339 (1216)	475 (1885)	1750 (7685)
	7ft	**8ft**	**9ft**	**10ft**	**11ft**	**12ft**	**13ft**
Segments (Crashes)	965 (3681)	1499 (7217)	767 (3805)	1137 (7349)	311 (1902)	81 (661)	53 (304)
	14ft	**15ft**	**16ft**	**17ft**	**18ft**	**20ft**	
Segments (Crashes)	15 (124)	16 (158)	5 (47)	1 (4)	6 (23)	3 (22)	

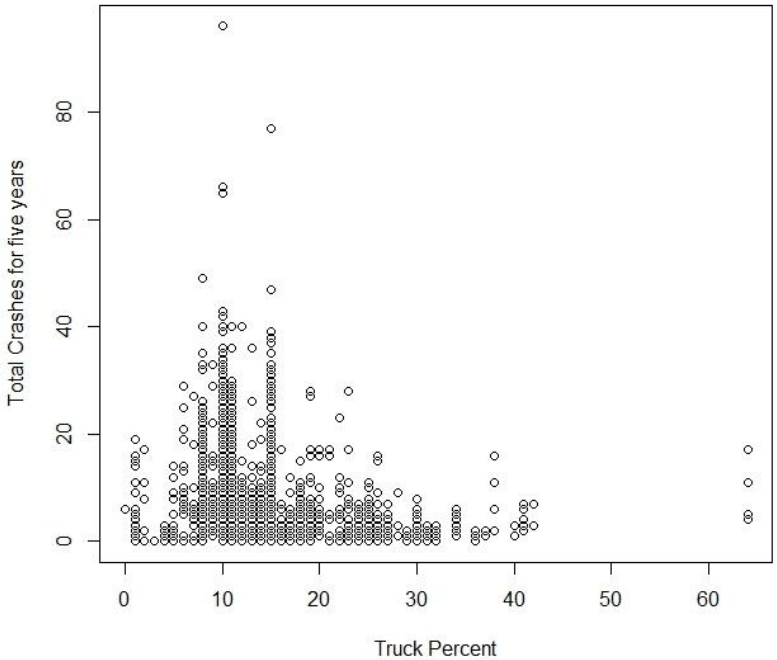


FIGURE 11A Truck percent versus total crashes for five years

TABLE 13A Minimum, maximum, and average values for truck percent

Minimum	Maximum	Average
0%	64%	12%

TABLE 14A Number of segments for each truck percent value

	0%	**1%**	**2%**	**3%**	**4%**	**5%**	**6%**
Number of Segments (Number of Crashes)	1 (6)	32 (179)	6 (40)	3 (0)	5 (7)	19 (88)	19 (190)
	7%	**8%**	**9%**	**10%**	**11%**	**12%**	**13%**
Segments (Crashes)	17 (107)	1577 (6000)	32 (302)	526 (5602)	3990 (16216)	69 (331)	42 (245)
	14%	**15%**	**16%**	**17%**	**18%**	**19%**	**20%**
Segments (Crashes)	63 (353)	1853 (10720)	29 (97)	32 (96)	44 (198)	85 (380)	17 (99)
	21%	**22%**	**23%**	**24%**	**25%**	**26%**	**27%**
Segments (Crashes)	11 (57)	25 (104)	14 (100)	12 (44)	30 (106)	26 (116)	19 (64)
	28%	**29%**	**30%**	**31%**	**32%**	**34%**	**36%**
Segments (Crashes)	3 (13)	5 (5)	18 (43)	11 (15)	10 (15)	13 (35)	7 (4)
	37%	**38%**	**40%**	**41%**	**42%**	**64%**	
Segments (Crashes)	2 (3)	6 (39)	3 (7)	7 (29)	3 (13)	5 (42)	

10.2 Appendix B

Developed Models

Eight SPFs were developed for undivided rural two-lane roadway segments in order to evaluate how IRI and weather-related crashes affected the fit of the SPFs. The following shows a breakdown of the four models.

- All crash types
    - Model including IRI as a variable and including weather-related crashes in the crash total
    - Model not including IRI as a variable and including weather-related crashes in the crash total (results located in Chapter 6, Section 6.2.1)
    - Model not including IRI as a variable, including weather-related crashes in the total, and including intersection-related crashes in the total
    - Model not including IRI as a variable, not including weather-related crashes in the total, and including intersection-related crashes in the total
- Fatal and injury crashes
    - SPF including IRI as a variable
    - SPF not including IRI as a variable (results located in Chapter 6, Section 6.2.2)
    - Note: weather-related crashes were included in the crash total because, in Meta-Manager, there is no way to discern the difference between crash type, specifically property-damage-only crashes and fatal and injury crashes. Intersection-related crashes are also included in the crash total for a similar reason.

10.2.1 All Crash Types

The following shows the table of coefficients and significance for each SPF and the CURE plots.

TABLE 1B Model results for all AADT with IRI and weather-related crashes

Variable	Estimate	Standard Error	z-Value	p (> |z|)	AIC	Dispersion Parameter
All AADT						
(Intercept)	-3.162	0.363	-8.721	$< 2 \times 10^{-16}$	32410	4.046
AADT	0.837	0.015	55.862	$< 2 \times 10^{-16}$		
Segment Length	0.876	0.023	38.486	$< 2 \times 10^{-16}$		
Lane Width	-0.150	0.030	-5.003	5.66×10^{-7}		
Paved Shoulder Width	-0.058	0.007	-8.076	6.71×10^{-16}		
IRI	0.037	0.012	3.057	0.002239		
Posted Speed	-0.009	0.002	-4.754	2.00×10^{-6}		
Truck Percent	-0.010	0.003	-3.405	0.000661		
Curve Presence	0.350	0.024	14.633	$< 2 \times 10^{-16}$		

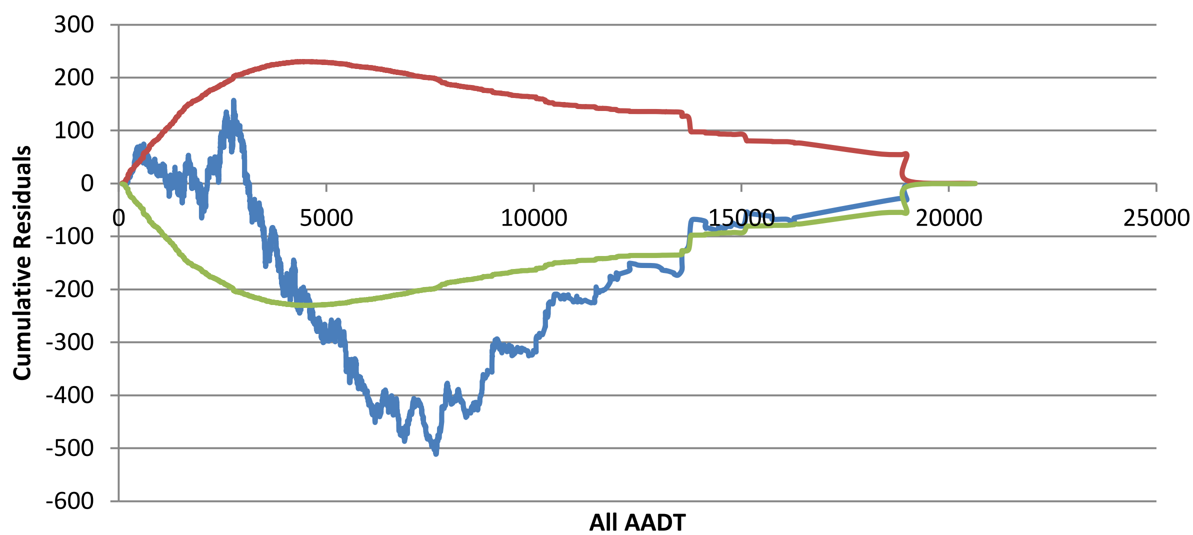


FIGURE 1B CURE plot for all AADT with IRI and weather-related crashes

The percent CURE deviation of Figure 1B is 29%. Table 2B shows the results with the AADT split of 7,500 vpd and Figure 2B and Figure 3B show the CURE plots for the two models.

TABLE 2B Model results for AADT splits with IRI and weather-related crashes

Variable	Estimate	Standard Error	z-Value	p (> $\lvert z \rvert$)	AIC	Dispersion Parameter
AADT ≤ 7,500 vpd						
(Intercept)	-3.021	0.374	-8076	6.67×10^{-16}	29104	4.091
AADT	0.797	0.0175	45.646	$< 2 \times 10^{-16}$		
Segment Length	0.873	0.025	34.927	$< 2 \times 10^{-16}$		
Lane Width	-0.135	0.031	-4.438	9.10×10^{-6}		
Paved Shoulder Width	-0.058	0.008	-7.534	4.94×10^{-14}		
IRI	0.028	0.018	2.224	0.02616		
Posted Speed	-0.009	0.002	-4.344	1.40×10^{-5}		
Truck Percent	-0.009	0.003	-2.913	0.00358		
Curve Presence	0.349	0.025	14.182	$< 2 \times 10^{-16}$		
AADT > 7,500 vpd						
(intercept)	-8.583	1.176	-7.301	2.86×10^{-13}	3291	4.204
AADT	1.217	0.129	9.462	$< 2 \times 10^{-16}$		
Segment Length	0.915	0.056	16.416	$< 2 \times 10^{-16}$		
Paved Shoulder Width	-0.048	0.021	-2.257	0.0240		
IRI	0.089	0.039	2.278	0.0227		
Posted Speed	-0.008	0.004	-1.838	0.0661		

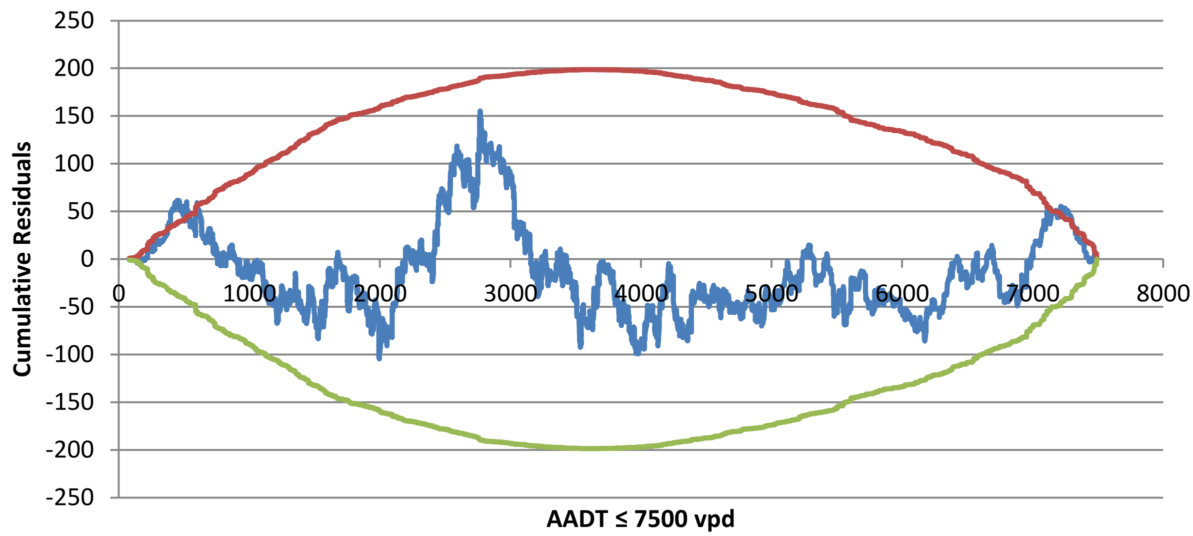


FIGURE 2B CURE plot for AADT ≤ 7,500 vpd with IRI and weather-related crashes

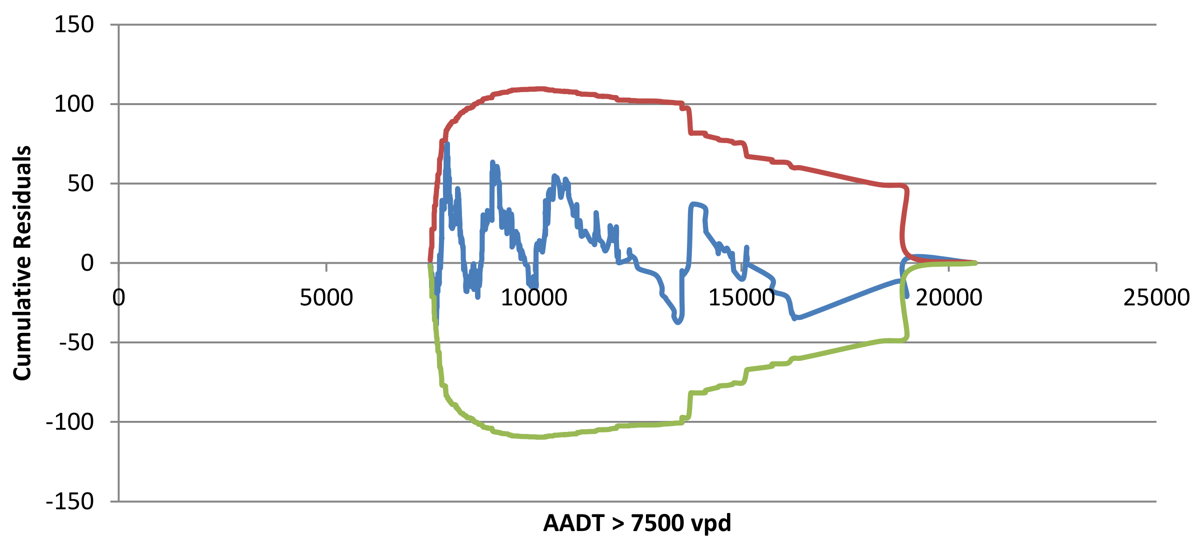


FIGURE 3B CURE plot for AADT > 7,500 vpd with IRI and weather-related crashes

The percent CURE deviation for Figure 2B and Figure 3B are 3% and 0%, respectively. It should be noted that the SPFs with weather-related crashes and not including IRI are located in Section 6.2.

The next set of tables and figures provide the results for models including total crashes without intersection-related crashes removed. This analysis was completed in order to show the effect of removing intersection-related crashes from the total number of crashes. The analysis was also used to allow for comparison of crash totals with and without weather-related crashes. Table 3B shows the results for the model with all AADT, no IRI, weather-related crashes included, and intersection-related crashes included. Figure 4B shows the respective CURE plot.

TABLE 3B Model results for all AADT, no IRI, weather-related crashes, and intersection-related crashes

Variable	**Estimate**	**Standard Error**	**z-Value**	**p (> $\|z\|$)**	**AIC**	**Dispersion Parameter**
			All AADT			
(Intercept)	-3.072	0.352	-8.739	$< 2 \times 10^{-16}$	37520	3.0442
AADT	0.949	0.015	64.015	$< 2 \times 10^{-16}$		
Segment Length	0.622	0.021	30.149	$< 2 \times 10^{-16}$		
Lane Width	-0.154	0.029	-5.237	1.63×10^{-7}		
Paved Shoulder Width	-0.064	0.007	-9.222	$< 2 \times 10^{-16}$		
Posted Speed	-0.015	0.002	-9.164	$< 2 \times 10^{-16}$		
Truck Percent	-0.020	0.003	-7.181	6.93×10^{-13}		
Curve Presence	0.289	0.024	12.119	$< 2 \times 10^{-16}$		

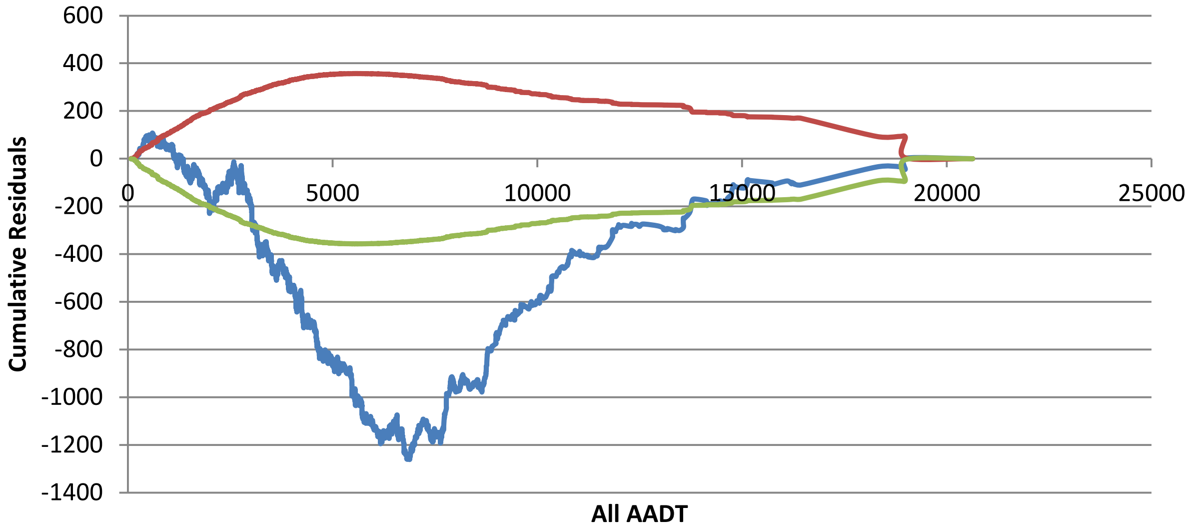


FIGURE 4B CURE plot for all AADT with no IRI, weather-related crashes, and intersection-related crashes

The percent CURE deviation is 53%. Table 4B shows the model results for all AADT, no IRI, no weather-related crashes included, and intersection-related crashes included. Figure 5B shows the respective CURE plot for the results in Table 4B.

TABLE 4B Model results for all AADT, no IRI, no weather-related crashes, and intersection-related crashes

Variable	Estimate	Standard Error	z-Value	p (> |z|)	AIC	Dispersion Parameter
All AADT						
(Intercept)	-3.299	0.414	-7.959	1.73×10^{-15}	31740	2.5355
AADT	0.947	0.018	54.076	$< 2 \times 10^{-16}$		
Segment Length	0.595	0.024	24.624	$< 2 \times 10^{-16}$		
Lane Width	-0.151	0.035	-4.324	1.53×10^{-5}		
Paved Shoulder Width	-0.070	0.008	-8.543	$< 2 \times 10^{-16}$		
Posted Speed	-0.020	0.002	-10.611	$< 2 \times 10^{-16}$		
Truck Percent	-0.021	0.003	-6.298	3.01×10^{-10}		
Curve Presence	0.251	0.028	8.917	$< 2 \times 10^{-16}$		

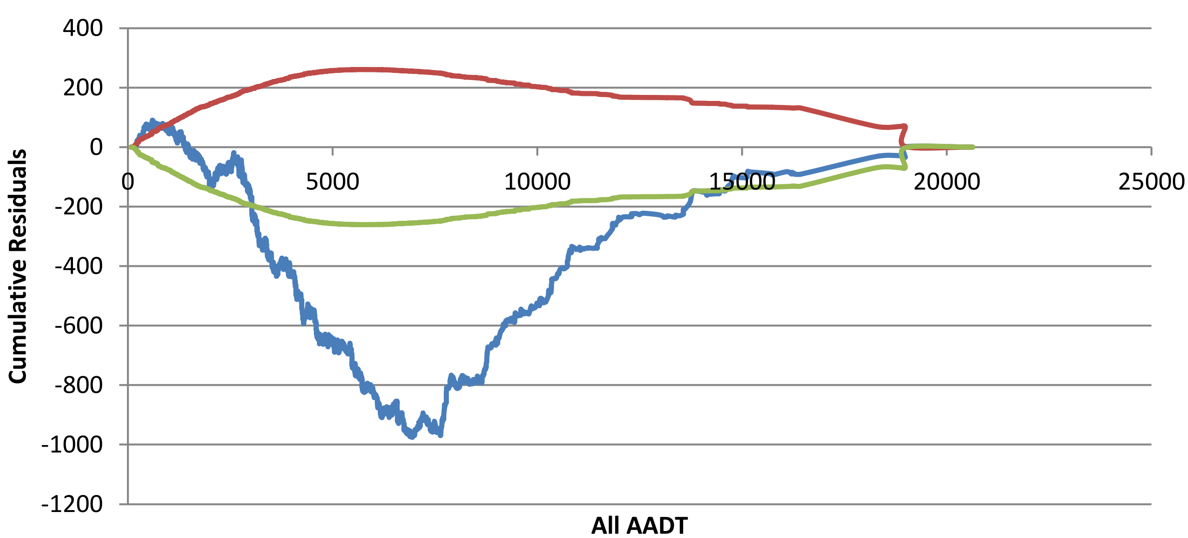


FIGURE 5B CURE plot for all AADT with no IRI, no weather-related crashes, and intersection-related crashes

The percent CURE deviation for Figure 5B is 57%. When the two intersection-related models are compared, the model without weather included in the total has a smaller dispersion parameter, a lower AIC value, and a larger percent CURE deviation.

10.2.2 Fatal and Injury Crashes

Table 7B shows the model results for fatal and injury crashes with the inclusion of IRI. Weather-related crashes and intersection-related crashes are included in the model because, with the available data, there is no way of identifying which crashes are fatal and injury or property damage only. Figure 10B shows the CURE plot for the respective model.

TABLE 7B Model results for all AADT with IRI for fatal and injury crashes

Variable	Estimate	Standard Error	z-Value	p (> $\|z\|$)	AIC	Dispersion Parameter
All AADT						
(Intercept)	-4.955	0.458	-10.828	$< 2 \times 10^{-16}$	25760	2.800
AADT	0.936	0.019	48.124	$< 2 \times 10^{-16}$		
Segment Length	0.628	0.027	23.597	$< 2 \times 10^{-16}$		
Lane Width	-0.152	0.040	-3.833	0.000126		
Paved Shoulder Width	-0.053	0.009	-5.748	9.03×10^{-9}		
IRI	0.053	0.016	3.437	0.000588		
Truck Percent	-0.018	0.004	-4.784	1.72×10^{-6}		
Curve Presence	0.283	0.031	9.034	$< 2 \times 10^{-16}$		

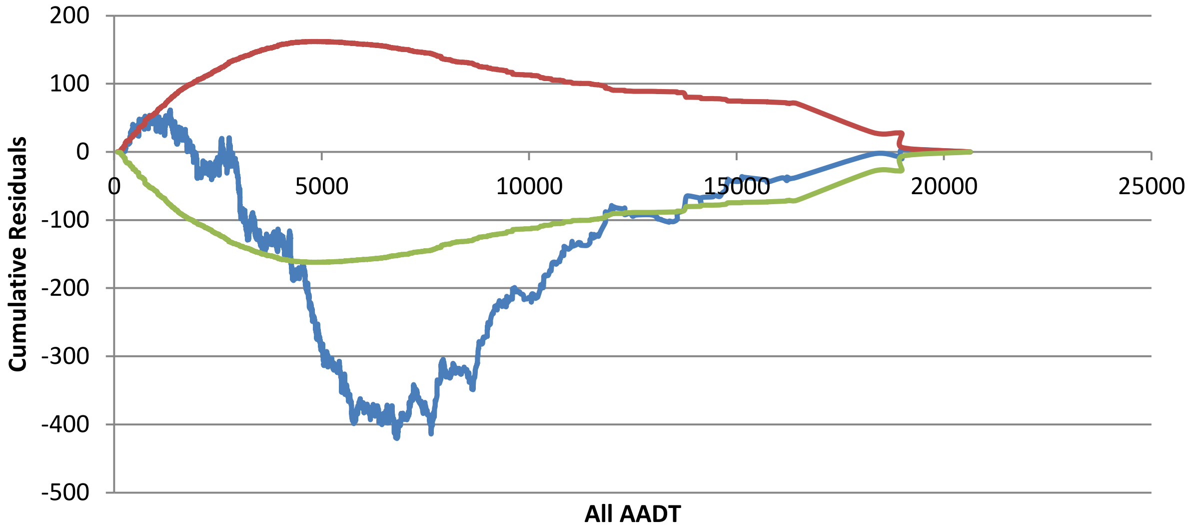


FIGURE 10B CURE plot for all AADT with IRI for fatal and injury crashes

The percent CURE deviation is 34% with an AADT split close to 7,000 vpd. Table 8B shows the results for the model with AADT less than or equal to 7,000 vpd and for the model with AADT greater than 7,000 vpd. Figure 11B and Figure 12B show the CURE plots for the respective models.

TABLE 8B Model results for split AADT with IRI and no weather-related crashes

Variable	Estimate	Standard Error	z-Value	p (> |z|)	AIC	Dispersion Parameter
AADT ≤ 7,000 vpd						
(Intercept)	-4.567	0.468	-9.763	$< 2 \times 10^{-16}$	22123	2.883
AADT	0.865	0.024	36.391	$< 2 \times 10^{-16}$		
Segment Length	0.655	0.031	21.403	$< 2 \times 10^{-16}$		
Lane Width	-0.130	0.041	-3.191	0.00142		
Paved Shoulder Width	-0.063	0.010	-6.316	2.68×10^{-10}		
Truck Percent	-0.017	0.004	-4.154	3.27×10^{-5}		
Curve Presence	0.267	0.033	8.218	$< 2 \times 10^{-16}$		
AADT > 7,000 vpd						
(intercept)	-9.070	1.199	-7.567	3.82×10^{-14}	3599	2.845
AADT	1.131	0.133	8.478	$< 2 \times 10^{-16}$		
Segment Length	0.583	0.054	10.843	$< 2 \times 10^{-16}$		
IRI	0.192	0.040	4.745	2.08×10^{-6}		
Curve Presence	0.236	0.124	1.904	0.057		

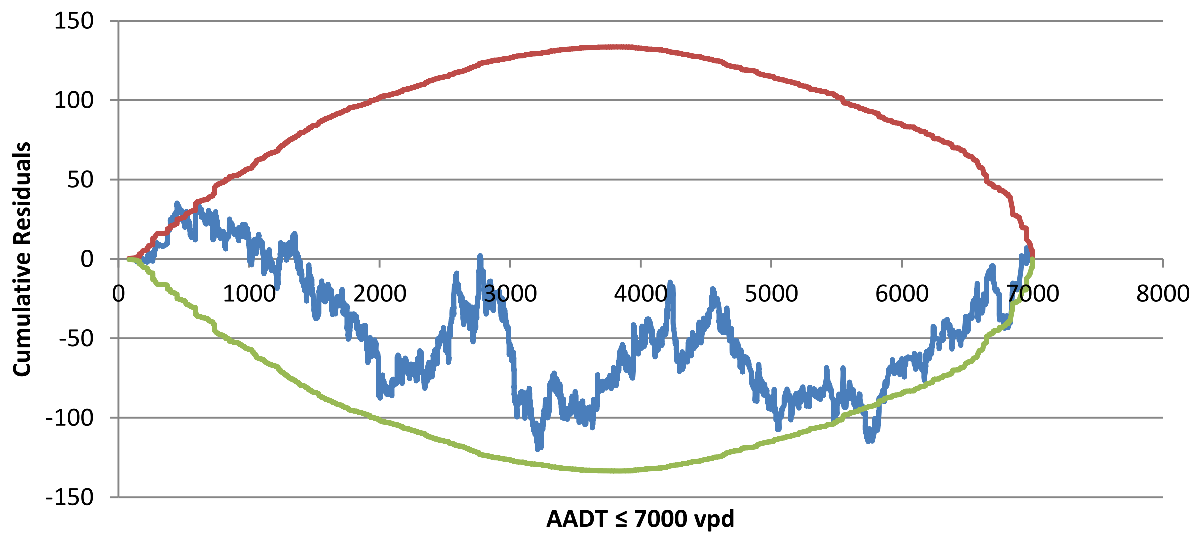


FIGURE 11B CURE plot for AADT ≤ 7,000 vpd with IRI for fatal and injury crashes

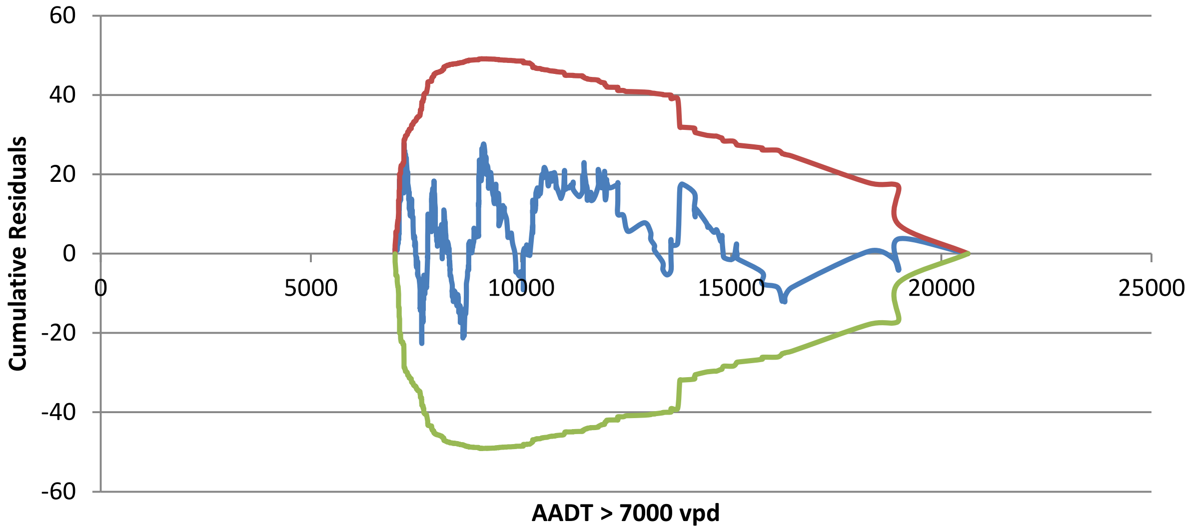


FIGURE 12B CURE plot for AADT > 7,000 vpd with IRI for fatal and injury crashes

The percent CURE deviation is 3% and 0%, respectively, for Figure 11B and Figure 12B.