CALIBRATION OF THE HIGHWAY SAFETY MANUAL AND DEVELOPMENT OF NEW SAFETY PERFORMANCE FUNCTIONS FOR RURAL MULTILANE HIGHWAYS IN KANSAS

by

SYEDA RUBAIYAT AZIZ

BS, Bangladesh University of Engineering and Technology, 2009
MS, Kansas State University, 2013

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Civil Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Abstract

Rural roads account for 90.3% of the 140,476 total centerline miles of roadways in Kansas. In recent years, rural fatal crashes have accounted for about 66% of all fatal crashes. The Highway Safety Manual (HSM) provides models and methodologies for analyzing the safety of various types of highways. Predictive methods in the HSM were developed based on national trends and data from few states throughout the United States. However, these methodologies are of limited use if they are not calibrated for individual jurisdictions or local conditions.

The objective of this study was to analyze the HSM calibration procedures for rural multilane segments and intersections in Kansas. The HSM categorizes rural multilane segments as four-lane divided (4D) and four-lane undivided (4U) segments and rural multilane intersections as three-legged intersections with minor-road stop control (3ST), four-legged intersections with minor-road stop control (4ST), and four-leg signalized intersections (4SG). The number of predicted crashes at each segment was obtained according to the HSM calibration process. Results from calibration of rural segments indicated that the HSM overpredicts fatal and injury crashes by 50% and 65% and underpredicts total crashes by 48% and 64% on rural 4D and 4U segments, respectively. The HSM-given safety performance function (SPF) regression coefficients were then modified to capture variation in crash prediction. The adjusted models for 4D and 4U multilane segments indicated significant improvement in crash prediction for rural Kansas.

Furthermore, Kansas-specific safety performance functions (SPF)s were developed following the HSM recommendations. In order to develop Kansas-specific SPF, Negative Binomial regression was applied to obtain the most suitable model. Several additional variables were considered and tested in the new SPFs, followed by model validation on various sets of locations. The Kansas-specific SPFs are capable of more accurately predicting total and fatal and injury crashes on multilane segments compared to the HSM and the modified HSM models.

In addition to multilane segments, rural intersections on multilane highways were also calibrated according to the HSM methodology. Using crash modification factors for corresponding variables, SPFs were adjusted to obtain final predicted crash frequency at intersections. Obtained calibration factors indicated that the HSM is capable of predicting crashes at intersections at satisfactory level. Findings of this study can be used for improving safety of rural multilane highways.

CALIBRATION OF THE HIGHWAY SAFETY MANUAL AND DEVELOPMENT OF NEW SAFETY PERFORMANCE FUNCTIONS FOR RURAL MULTILANE HIGHWAYS IN KANSAS

by

SYEDA RUBAIYAT AZIZ

BS, Bangladesh University of Engineering and Technology, 2009
MS, Kansas State University, 2013

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Civil Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Sunanda Dissanayake

Abstract

Rural roads account for 90.3% of the 140,476 total centerline miles of roadways in Kansas. In recent years, rural fatal crashes have accounted for about 66% of all fatal crashes. The Highway Safety Manual (HSM) provides models and methodologies for analyzing the safety of various types of highways. Predictive methods in the HSM were developed based on national trends and data from few states throughout the United States. However, these methodologies are of limited use if they are not calibrated for individual jurisdictions or local conditions.

The objective of this study was to analyze the HSM calibration procedures for rural multilane segments and intersections in Kansas. The HSM categorizes rural multilane segments as four-lane divided (4D) and four-lane undivided (4U) segments and rural multilane intersections as three-legged intersections with minor-road stop control (3ST), four-legged intersections with minor-road stop control (4ST), and four-leg signalized intersections (4SG). The number of predicted crashes at each segment was obtained according to the HSM calibration process. Results from calibration of rural segments indicated that the HSM overpredicts fatal and injury crashes by 50% and 65% and underpredicts total crashes by 48% and 64% on rural 4D and 4U segments, respectively. The HSM-given safety performance function (SPF) regression coefficients were then modified to capture variation in crash prediction. The adjusted models for 4D and 4U multilane segments indicated significant improvement in crash prediction for rural Kansas.

Furthermore, Kansas-specific safety performance functions (SPF)s were developed following the HSM recommendations. In order to develop Kansas-specific SPF, Negative Binomial regression was applied to obtain the most suitable model. Several additional variables were considered and tested in the new SPFs, followed by model validation on various sets of locations. The Kansas-specific SPFs are capable of more accurately predicting total and fatal and injury crashes on multilane segments compared to the HSM and the modified HSM models.

In addition to multilane segments, rural intersections on multilane highways were also calibrated according to the HSM methodology. Using crash modification factors for corresponding variables, SPFs were adjusted to obtain final predicted crash frequency at intersections. Obtained calibration factors indicated that the HSM is capable of predicting crashes at intersections at satisfactory level. Findings of this study can be used for improving safety of rural multilane highways.

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I would like to pay my utmost respect to Almighty Allah for granting me this wonderful opportunity to work and to live.

I hereby acknowledge my sincere appreciation and due gratitude to my major professor Dr. Sunanda Dissanayake for academic guidance, mentorship, and continuous support throughout my graduate studies. I am thankful to Dr. Robert Stokes, Dr. Eric Fitzsimmons, Dr. Margaret Rys, and Dr. John Leatherman for serving as my doctoral committee members.

I sincerely appreciate the Kansas Department of Transportation (KDOT) for funding this research. I gratefully thank Ms. Elsit Mandal, Mr. Rex McCommon, and Mr. Leif Holliday for their constant support with providing data to complete this research. I also appreciate the assistance that Austin Juaneman provided for intersection data collection.

Finally, I wish to express my sincere respect and gratitude to my parents, Dr. Syed Azizul Haque and Dr. Nilufar Begum for their encouragement, blessings, and love that helped me to be what I am today. I would like to especially thank my husband Dr. S M Shafiul Alam for his patience, support, and appreciation.

Dedication

This dissertation is dedicated to my late paternal grandfather Prof. Syed Mazharul Haq and late maternal grandfather Prof. Dr. Mufizur Rahman.

Chapter 1 - Introduction

1.1 Background

According to a study published in 2016, motor vehicle crashes were one of the top ten causes of death in the United States in 2013. Relative to 2011, fatal highway crashes increased by 1.7% to 29,989 in 2014, equivalent to an average of 90 daily fatalities (NHTSA, 2015). Despite the decline in fatalities, 32,675 deaths occurred as a result of roadway crashes in the United States in 2014, down from 32,894 in 2013 (NHTSA, 2015).

Rural roads account for 90.3% of the 140,476 total miles of roadway in Kansas (KDOT (a), 2015), and in 2014, rural travel accounted for 48.5% of all vehicle miles (60% for state highways) (KDOT (b), 2015). Figure 1.1 shows the distribution of rural, urban, fatal rural, fatal urban, and total crashes over a 14-year period. In general, Kansas has a low population density and a majority of the roadways are in rural areas. As shown in Figure 1.1, 35% of total crashes occurred on rural roads, while fatal crashes on rural roads accounted for over 66% of the number of total fatal crashes in Kansas during 2014 (KDOT, 2015). Not only in 2014, every year the number of fatal crashes in rural highways always been considerably higher than the fatal crashes on urban highways in Kansas. The time required to respond and transport crash victims potentially determines if the crash is classified as injury or fatal. In rural areas, transportation of severely injured crash victims to hospitals requires 60–120 minutes (NHTSA, 2009). These numbers are a matter of concern for highway safety professionals because they comprise a major proportion of high-level injury crashes in rural areas.

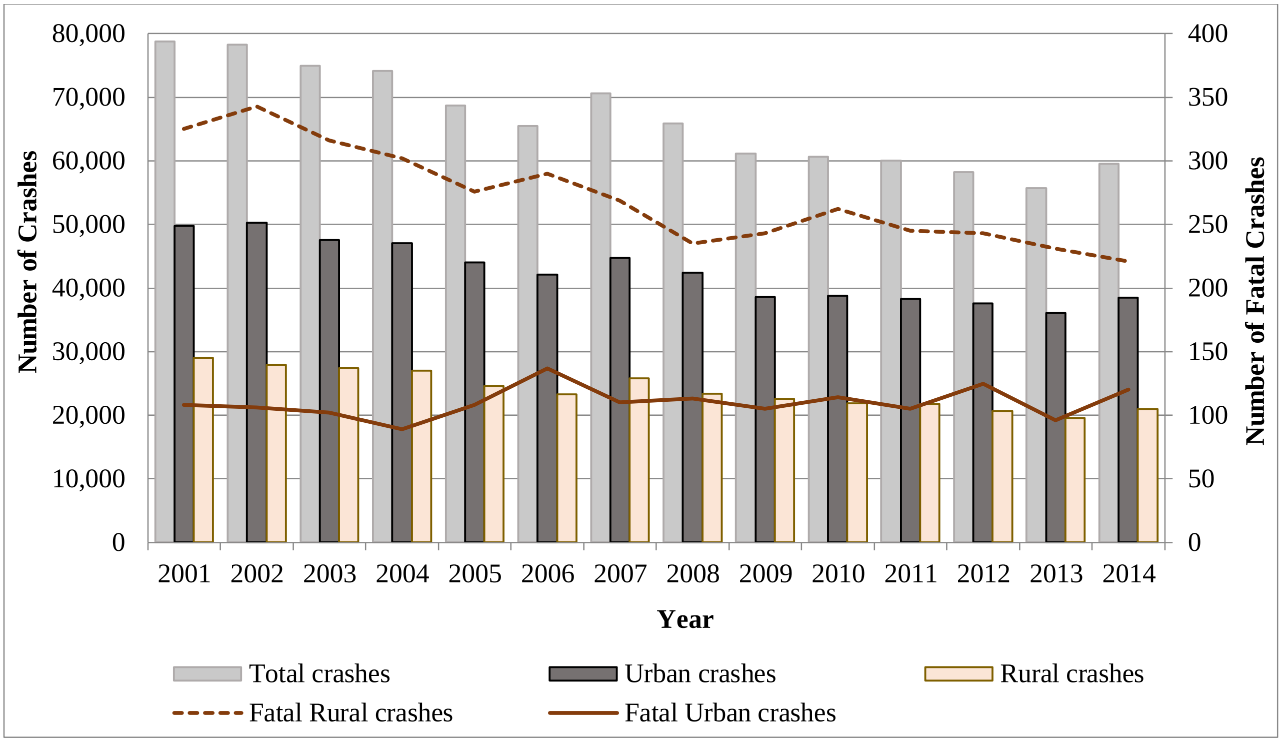


Figure 1.1 Yearly distribution of crashes in Kansas

1.2 Highway Safety Manual

The Highway Safety Manual (HSM) by the American Association of State Highway and Transport Officials (AASHTO) is the culmination of decades of safety research and practices (AASHTO, 2010). The HSM provides models and methodologies for analyzing various types of highways based on safety. The first version, published in 2010, was updated in 2014 with new chapters on predictive methods for freeways and ramps. Procedures to calibrate predictive models are currently available in Part C – Appendix A of the HSM (AASHTO, 2010). Crash predictive methods in the HSM allow planners, designers, and reviewers to comprehensively assess expected safety performance of highway design via methodologies endorsed by the Federal Highway Administration (FHWA). Predictive methods in the HSM were developed based on national trends and statistics from sample states throughout the United States. However, these methodologies are of limited use if they are not calibrated for individual jurisdictions or local conditions. Calibration ensures the most realistic and reliable crash estimates.

1.3 Problem Statement

Safety conditions of highways change over time; therefore, agencies should only use the HSM models that have been calibrated. Uncalibrated models compromise safety estimates, produce unrealistic results, and undermine accountability of highway safety. Even agencies that use their own data to develop SPFs should consider calibrating the models every two to three years in order for results to be comparable to estimates obtained from an agency's records.

An acceptable method to predict crashes for rural multilane highway segments and intersections in Kansas must be developed. Currently, the Kansas Department of Transportation (KDOT) can apply the rural two-lane model given in HSM, because a previous study calibrated such facilities (Lubliner, 2012). KDOT has occasionally requested analysis of a multilane facility, but it cannot be completed without calibration. An effective equation that predicts the number of crashes along a highway and identifies potential high crash locations would enable design engineers to design safer roads while minimizing the cost if, for example, 8-ft. shoulders were determined to be as beneficial as 10-ft. shoulders.

Although calibration procedures are available in the HSM Appendix A, they must be refined or modified to accommodate data availability and roadway, traffic, and crash characteristics in Kansas. The HSM considers only four-lane highways to be categorized as rural multilane. Therefore, this study was limited to calibrations for rural four-lane divided (4D) and four-lane undivided (4U) highways in Kansas. Similar calibration is required on rural multilane intersections, which has not been performed for Kansas till date. So additionally, the rural multilane intersections will be calibrated in this study.

1.4 Objective of the Study

The objective of this dissertation is to analyze the HSM calibration procedures for rural multilane segment and intersection models for Kansas in which rural multilane segments are categorized as 4D and 4U, and intersections are categorized as three-legged intersections with minor-road stop control (3ST) and four-legged intersections with minor-road stop control (4ST). This study utilized the HSM methodology to calibrate the crash predictive method. Since the HSM methodology cannot accurately predict crashes at rural segments, new Kansas-specific models or SPFs were developed and their performances were compared to the HSM-given SPFs.

1.5 Organization of the Dissertation

This dissertation contains six chapters and an appendix. Chapter 1 provides background information regarding the HSM methodology and study objectives. Chapter 2 summarizes past research conducted in similar contexts, and Chapter 3 includes discussion of methodology and data used in this dissertation. Calibration results obtained using the HSM methodology are presented in Chapter 4. Chapter 5 discusses the development of new SPFs, and Chapter 6 summarizes the study with a discussion of future work.

Chapter 2 - Literature Review

This chapter summarizes the review of literature, beginning with initial research reporting the relationship of geometric and surrounding features to crash type, followed by SPFs and the evolution of current crash prediction models (CPM)s. Although the literature review does not include all CPM-related research, it summarizes the most critical sources that have led to development of current prominent methods, including recent research of CPM applications.

2.1 Highway Safety Manual Calibration

A limited number of studies have performed and documented the HSM calibration process. Sun et al. (2006) performed the first study that calibrated the HSM's CPM for two-lane rural highway segments in Louisiana. The CPM used was nearly identical to the current model given in Chapter 10 in the HSM, with the exception that the HSM had additional crash modification factors (CMFs) for rumble strips, lighting, and automated speed enforcement, added after the research by Sun et al. (2006). In addition, the calibration procedure recommended in the draft HSM that was applied to the study differed from the procedure published in the HSM. It is because the procedure required stratification of calibration factors based on traffic volume. Calibration factors were then averaged together for application.

2.1.1 Calibration of Rural Two-lane Two-way Highways

Srinivasan and Carter (2011) developed SPFs for various types of roadway in North Carolina and illustrated how SPFs can improve the decision-making process. The HSM prediction methods were used to compute the calibration factor for total crashes for each facility type. Using data from the crash-reporting database at the North Carolina Department of Transportation (NCDOT), segments within the influence of at-grade intersections and railroad grade crossings (250 ft. on either side of at-grade intersections or railroad grade crossings) were removed. SPFs were estimated for nine crash types identified to be of primary importance to NCDOT. In addition, SPFs for rural two-lane roads were estimated by including site characteristics such as shoulder width/type and terrain. Another SPF was used for network screening. Srinivasan and Carter (2011) also suggested that NCDOT calibrate SPFs developed in this process and/or develop SPFs using Negative Binomial regression.

The study by Sun et al. (2006) utilized the same basic definition for rural two-lane highways in Louisiana, but lack of geometric data required the use of default values for several CMFs, and some data values were not consistent with those experienced in Kansas. Using these data and calibration methodology, a calibration value of 1.63 was determined for the Louisiana highway system. The Louisiana study also validated the CPM using the calibration factor and the Emperical Bayes (EB) procedure. The study demonstrated model accuracy in terms of percent difference between observed and predicted crashes with calibration. Accuracy of the calibrated model without the EB procedure yielded a 5.22% difference. The EB procedure improved model accuracy by 3.06%. Accuracies pertained to the aggregates of all segments modeled in the validation study, but results did not show individual segment accuracy in definable values (Sun et al., 2006).

Xie et al. (2011) calibrated each of the HSM-considered roadway facility type in the Oregon highway system. Using data from 2004–2006 for rural, two-lane, two-way roads, the final calibration factor was determined to be 0.74, which they speculated to be under 1.0 due to less reported property damage only (PDO) crashes since those crashes do not have to be reported to authorities in Oregon. Xie et al. also found that data accumulation was time-consuming, evidenced by a gap in their research because they did not validate newly created calibration factors. Although they followed steps given in the HSM, they did not verify accuracy of the calibrated model for crash prediction.

2.1.2 Calibration of Rural Multilane Highways

As suggested by the HSM, only 4D and 4U facilities are categorized as rural multilane. A review of studies focusing on rural multilane highway calibration using HSM is presented herein.

Sun et al. (2011) calibrated the SPF for rural multilane highway segments, they investigated how calibrated models work in network screening, and they identified potential application issues. Their paper presented results for segments. Among the 600 miles of rural multilane highways in the Louisiana Department of Transportation (LaDOTD) system, some highways were divided into control sections based on highway design features and traffic volumes. All design features and traffic conditions were identical within each control section. Coefficients for basic SPFs were obtained from the HSM, and relevant CMFs were applied to the number of predicted crashes. Obtained calibration parameters indicated that the predicted model from the HSM for rural divided multilane highways underestimated expected crashes. Network screening was performed in conjunction with the Safety Management System introduced in Part B of the HSM. The application indicated that, even without the calibrated safety performance model, commonly used crash frequency methods produce results similar to results of sophisticated models. However, the same thing cannot be said about crash rate methods. Result comparisons of the four screening measures were similar to sample application results presented at the end of Chapter 11 in the HSM (Sun et al., 2011).

Sun et al. (2013) divided segments in Missouri based on Annual Average Daily Traffic (AADT), an important input for HSM given CPMs. Characteristics used to subdivide segments included speed category for urban arterials, median type, effective median width for freeways and rural multilane highways, and horizontal curve radius for rural two-lane highways. After subdivision, some segments were shorter than the desired minimum 0.5 miles for rural segments and 0.25 miles for urban segments. Segments ranged in length between 0.56 and 7.59 miles, with an average length of 2.60 miles. This study considered crash data from 2009 to 2011, and AADT of 2011 was obtained from their database. The total number of vehicle crashes was 715 per year, which significantly exceeded the HSM-recommended 100 crashes per year. A median width of 30 ft. was used for segments with a median barrier, as recommended by the HSM. Segment length was calculated as the average segment length in both directions, excluding interchange limits. Results indicated close agreement between the number of crashes predicted by the HSM and the number of crashes observed in Missouri for those site types (Sun et al., 2013).

Lord et al. (2008) developed a methodology to predict the safety performance of elements in the planning, design, and operation of nonlimited-access rural highways. Models were proposed for the three types of intersections and undivided and divided highway segments by crash type and crash severity. They collected data from databases in California, Minnesota, New York, Texas, and Washington, which they used to develop statistical models and CMFs for intersections and segments as well as a cross-validation study to evaluate the recalibration procedure for jurisdictions other than those for which the models were estimated. They utilized data collected in Texas, California, Minnesota, and Washington to develop models and CMFs, and they used New York data for cross-validation. The collected data included detailed information about geometric design characteristics, traffic flow, and motor vehicle crashes (Lord et al., 2008)

Jalayer et al. (2015) presented a revised method to develop calibration factors for five types of urban and suburban roadways with consideration of recent crash recording threshold (CRT) change, a minimum value to report crashes, in Illinois. Because of a change in 2009 regarding the recording threshold for PDO crashes, the study established a revised method to supplement and adopt a standard approach to develop calibration factors in the HSM, considering impact of the new CRT. The higher the CRT, the fewer recorded PDO crashes. Before and after the threshold change 4D calibration factors were 0.68 and 0.55, respectively. Because the threshold change only affects the total number of crashes and PDO crashes, percentage distributions of fatal and injury crashes before the threshold change were adjusted in order to accurately estimate the total number of fatal and injury crashes. This study provided a revised method to help state and local agencies predict the number of crashes without redeveloping new calibration factors due to change in CRT.

2.2 Development of State-Specific Safety Performance Functions

A unique Oregon study by Xie et al. (2011) developed jurisdiction-specific crash distributions to replace default values in the HSM. Their analysis showed that, on an aggregate level, use of jurisdiction-specific distributions did not significantly affect results compared to HSM default values. However, this analysis did not include quantification of this impact at the project level. Of the statistics provided, Oregon-specific values also did not vary notably from default values in the HSM; therefore, no significant impact was found using Oregon-specific values instead of default values.

Banihashemi (2011) compared CPM calibration to two new SPFs in the state of Washington. Equation 2.1 has the same general form as the rural two-lane SPF in the HSM, and Equation 2.2 has a similar form except that AADT is raised to the power of 1.05. Four new state-specific CMFs were produced and used with the new SPFs in this study: lane width, shoulder width, curve radius, and vertical grade. Results showed that calibration in Washington was identical for any of the new models, but the newer models may be preferable if created specifically for Washington. However, because the original SPF was created using data from Washington and

Minnesota, this model was expected to work just as well as new SPFs. Similar to previous studies, models studied by Banihashemi (2011) assumed default values for a number of CMFs due to data limitations.

$$N_{spf-1-rs} = 0.91705 \times AADT \times L \times 365 \times 10^{-6} \quad (2.1)$$

$$N_{spf-2-rs} = 0.5782 \times AADT^{1.05} \times L \times 365 \times 10^{-6} \quad (2.2)$$

Where,

$AADT$ = average annual daily traffic (vpd), and

L = length of segment (mi).

Qin et al. (2014) applied the HSM methodology to rural two-lane, two-way highway segments in South Dakota. Calibration was based on three years (2009–2011) of crash data from 657 roadway segments, totaling more than 750 miles of roadways. The calibration process established new base conditions, developed SPFs, converted CMFs to base conditions, and substituted default values with state-specific values. Five models were developed and compared based on statistical goodness-of-fit and calibration factors. Results showed that jurisdiction-specific crash type distribution for CMFs drastically differed from crash distribution presented in the HSM. The HSM method without modification was shown to underestimate crashes in South Dakota by 35%. The method based on SPFs developed from a full model demonstrated the best model fit. This study provided important guidance and empirical results regarding calibration of HSM models (Qin et al., 2014).

Mehta and Lou (2013) evaluated applicability of the HSM predictive methods on Alabama data for two-lane, two-way rural roads and 4D highways. They calibrated the HSM-based SPFs using two approaches, and they proposed a new approach that treats the estimation of calibration factors as Negative Binomial regression. Data was taken from the years of 2006 to 2009. In addition, new forms for state-specific SPFs were investigated to identify the best model using Poisson-Gamma regression techniques. Mehta and Lou studied four new model forms and evaluated prediction capabilities of the two calibrated models and four newly developed state-specific SPFs using a validation data set. They considered five performance measures for model evaluation: mean absolute deviance, mean squared prediction error, mean prediction bias, log likelihood value, and Akailke's information criterion (AIC). The study identified a state-specific SPF that accurately fit the Alabama data and outperformed other models, including calibrated SPFs. The best model described mean crash frequency as a function of AADT, segment length, lane width, year, and speed limit. Results showed that the HSM-recommended method for calibration factor estimation performed well, proving to be a straightforward, easily applicable approach even though it was not as good as the best state-specific SPF.

2.3 Crash Prediction Studies in Kansas

Similar to other transportation organizations, KDOT has researched more efficient ways to screen robust system inventories and crash data in order to identify relationships between highway features and safety. In 2009, Najjar and Mandavilli used artificial neural networks (ANNs) to attempt to identify these relationships for Kansas highways. Their research included the six major types of roadway networks in Kansas: rural Kansas Turnpike Authority (KTA), rural two-lane, rural expressway, rural freeway, urban freeway, and urban expressway. The models evaluated total crash rate as well as fatal, injury, and severe injury crash rates. For rural two-lane highways, Najjar and Mandavilli (2009) identified eight variables that affect crashes:

- Section length
- Surface width
- Route class

- Shoulder width (outside)
- Shoulder type (outside)
- AADT
- Average percentage of heavy trucks
- Average speed limit.

ANN models produced by Najjar and Mandavilli (2009) were measured against training, testing, and validation data sets. The overall rural two-lane model produced a R^2 of 0.4655. The total crash rate model was most similar to the HSM model in this research; the R^2-value for the total crash rate ANN model was 0.173.

Lubliner and Shrock (2012) analyzed multiple predictive methods to calibrate rural two-lane segment SPF in Kansas. They initially analyzed all methods published in the HSM to determine method accuracy. Calibrated predictions showed significant improvements compared to uncalibrated predictions, and they were extremely accurate when analyzed at the aggregate level. In order to improve crash prediction accuracy, Lubliner and Shrock analyzed alternative calibration methods, including linear calibration methods that address variables previously shown to positively correlate to highway crashes in Kansas but are not considered in the HSM. Although linear calibration methods did not perform as well on the aggregate level, they were more accurate on the project level. In general, analysis of the HSM rural two-lane segment predictions showed favorable accuracy, leading to recommended inclusion in KDOT's safety evaluation toolbox at the project level. Based on study results, single statewide calibration of total crashes was recommended for aggregate analyses that include multiple sections (Lubliner and Shrock, 2012). However, the study by Lubliner and Shrock (2012) contained a large proportion of animal-related crashes, totaling 58.9% of animal-related crashes in Kansas but only 12.1% animal-related crashes in the HSM crash distribution. Therefore, an additional obtained calibration factor considered only crashes without animals, resulting in a calibration value of 0.557. Final calibrations considered animal crash rates of each segment and county, with the county, or variable, calibration factor working best according to

$$C_{county} = 1.13 \times ACR_{county} + 0.635 \tag{2.3}$$

Where,

C_{county} = calibration factor for a county, and

ACR_{county} = deer crash rate for a county.

Results showed that C_{county} worked best, but they suggested additional research to create a jurisdiction-specific SPF in order to determine if it could more accurately predict crashes on rural Kansas highways compared to the HSM model calibration (Lubliner and Shrock, 2012)

Bornheimer (2011) tested the original HSM CPMs to state-specific calibrated CPMs and new, independent CPMs to determine the best model for rural two-lane highways in Kansas. They collected nearly 300 miles of highway geometric data to create the new models using Negative Binomial regression. The most significant variables in each model were consistently lane width and roadside hazard rating. These models were compared to CPMs calibrated for the HSM using nine validation segments. However, one comparison difficulty was the large amount of animal-related crashes, accounting for 58.9% of crashes on Kansas highways (Bornheimer, 2011).

Analysis results showed that two models work best for Kansas: the variable calibration method in which crashes are predicted using the HSM's CPM and a calibration based on animal crash rates by county that demonstrates high correlation using Pearson's R. The variable calibration method also considers individual county animal crash statistics, thereby accounting for animal crashes. The model was run using the HSM's CPM method and the Interactive Highway

Safety Design Model (IHSDM), requiring in-depth data mining to collect all variables. Equation 2.4 defines the calibration factor, C_{county}, used in the HSM equation, as shown in Equation 2.5.

$$C_{county} = 1.13 \times ACR_{county} + 0.635 \quad (2.4)$$

$$N_{predicted\ rs} = N_{spf\ rs} \times \ \times (CMF_{1r} \times CMF_{2r} \times \ldots \times CMF_{12r}) \quad (2.5)$$

The non-animal model, restated in Equation 2.6, is a new SPF created using only crashes that did not involve an animal. This model had high correlation and low Bayesian Information Criterion (BIC), making it a good candidate. Elimination of animal-related crashes, which were generally out of an engineer's control, improved SPF. The SPF shown in Equation 2.6 also requires roadside hazard rating (RHR), AADT, and length of segment (L), thereby reducing the number of required variables and resulting in less effort to collect data during application (Bornheimer, 2011).

$$N_{pred-no-an} = AADT^{\ 1.01} L^{\ 0.85} e^{\ (-10.07+0.58 \times RHR)} \quad (2.6)$$

Where,

$AADT$ = average annual daily traffic (vpd), and

L = length of segment (mi).

2.4 Sample Size for Calibration Process

Sample size significance and influence also extensively influence the calibration process. Shin et al. (2014) completed the calibration process for SPFs in the HSM for Maryland's DOT in order to determine a statistically reliable sample size for developing local calibration factors (LCFs) and calculating the confidence interval for the range of calibration factors containing 90% of the population (Shin et al., 2014). Study results showed that calibration factor ranges were wider for site types with small populations.

Another study used data from the state of Washington to determine the ideal sample size for calibrating the HSM models and to examine sensitivity in a variety of HSM calibration factor sample sizes in order to evaluate the quality of developed factors (Banihashemi, 2012). Roadway and crash data were obtained for a three-year period (2006–2008). Calibration factors generated from the entire data set for each highway type were considered ideal calibration factors, and factors generated from various data set sizes were compared to the ideal factors. The probability that generated calibration factors fell within 5% and 10% of the ideal calibration factor was calculated. Results of this sensitivity analysis were reviewed and recommendations were derived and presented (Banihashemi, 2012).

2.5 Interactive Highway Safety Design Model

The IHSDM is a suite of software analysis tools used to evaluate the safety and operational effects of geometric design on highways. The IHSDM is a decision-support tool that estimates a highway design's expected safety and operational performance and compares existing or proposed highway designs to relevant design policy values. Results of the IHSDM support decision making in the highway design process. Intended users include highway project managers, designers, and traffic and safety reviewers in state and local highway agencies and engineering consulting firms. The IHSDM, which supports the data-driven safety analysis initiative of the FHWA's Every Day Counts three efforts, includes six evaluation modules: Crash Prediction, Design Consistency, Intersection Review, Policy Review, Traffic Analysis, and Driver/Vehicle.

Qin et al. (2013) developed locally derived IHSDM safety modules for South Dakota and North Dakota by evaluating data availability for rural local roads and tribal rural roads and resolving obstacles to module implementation. After the modules were developed, they used the modules to evaluate design alternatives based on safety performance. This study provided guidance and empirical results regarding calibration of IHSDM models for local agencies, but calibration processes and procedures can be expanded to other highway facilities. The study also recommended that unavailable data, such as curve and driveway density, should be collected to develop more accurate, reliable jurisdiction-specific SPFs. Separate calibration factors may also be considered for regions with distinct features such as mountain versus plain or dry versus wet or as a function of AADT or other characteristics (Qin et al., 2013).

2.6 SafetyAnalyst Prediction Models

SafetyAnalyst, a tool similar to the IHSDM, is associated with Part B of the HSM, which focuses on roadway safety management. SafetyAnalyst utilizes an SPF to predict crashes, but it uses less geometric data and it utilizes several tools to look at an entire network. These tools identify sites that could benefit from safety improvements, diagnose possible reasons for safety problems, suggest improvements and associated costs, prioritize sites that could benefit most according to cost estimates, and perform before and after evaluations. These analyses require the following primary data:

- Segment length
- Area type (rural/urban)
- Number of lanes
- Median type
- Access control
- Traffic volume

The base model for SafetyAnalyst is

$$Crashes = e^{a} \times AADT^{b} \times SL \quad (2.7)$$

Where,

$Crashes$ = predicted crashes per year

$AADT$ = average annual daily traffic (veh/day)

SL = segment length (miles)

a and b = regression parameters

It can also be adjusted with a calibration factor that should be reevaluated annually and a proportion factor if only certain types of crashes are considered. In supportive efforts, a number of states have shared what they have learned and published research regarding development of accurate methods to predict crashes for network analysis. Many states, such as Louisiana, have focused their research on individualized development and calibration of SPFs in SafetyAnalyst (Alluri and Ogle, 2012).

Alluri et al. (2014) studied the two most recent safety analysis tools, the HSM and SafetyAnalyst, which both struggle to meet data requirements for implementation. Many data variables required to derive the HSM calibration factors are currently unavailable in Florida's roadway characteristics inventory (RCI) database. This project attempted to identify and prioritize influential calibration variables for data collection and determine minimum sample sizes in order to estimate reliable calibration factors. For each facility type in the HSM, this project applied the random forest technique to rank required and desired variables based on importance. Variables were categorized as variables of primary importance, variables of secondary importance, and variables of lesser importance. Minimum sample sizes to estimate reliable calibration factors for facility types were also determined, proving that the minimum sample size of 30–50 sites with at least 100 crashes per year, as recommended by the HSM, is insufficient to achieve desired accuracy for nearly all facility types. Compared to the HSM, SafetyAnalyst has fewer and different data requirements. Two major efforts to apply SafetyAnalyst involve conversion of local data into the strict data format required by SafetyAnalyst and development of jurisdiction-specific SPFs. This project developed a software program to convert crash and roadway data for Florida state roads in order to import files used by SafetyAnalyst. This project also developed SPFs for unsignalized intersections in order to supplement those of facilities developed under another project. For example, using Florida data, SafetyAnalyst identified high crash locations. Recommendations for deploying SafetyAnalyst were also provided (Alluri et al., 2014).

Alluri and Ogle (2012) investigated transferability between default SPFs provided by SafetyAnalyst and Georgia-specific SPFs. Georgia-specific SPFs were generated similarly to SafetyAnalyst default SPFs. Sample SPFs were generated for all 17 types of roadway segments; these SPFs predicted the number of crashes as a function of traffic only. Calibrated SafetyAnalyst default SPFs were compared to Georgia-specific SPFs based on the overdispersion parameter. A comparison of overdispersion parameters (k) revealed that Georgia-specific SPFs have higher overdispersion parameters than respective default SPFs. Lower overdispersion parameters increase function reliability by giving more weight to predicted crashes in the EB process (Alluri and Ogle, 2012). When Georgia-specific SPFs demonstrated relatively higher overdispersion values more weight was given to observed crashes than predicted crash frequency. However, while performing EB analysis using default SPFs with relatively low overdispersion values, less weight was given to observed crashes. In general, urban SPFs for Georgia performed slightly better, as evidenced by lower overdispersion parameter values then their default counterparts. Increased understanding of the influence of the overdispersion parameter prompted the researchers to assert that state-specific SPFs with relatively low overdispersion parameters provide better crash prediction results (Alluri and Ogle, 2012).

Chapter 3 - Data and Methodology

This chapter describes the process of calibrating the HSM for rural multilane segments and intersections, including a brief overview of data collection. The methodology of developing new SPFs is also discussed.

3.1 Data

This study utilized highway crash data from the Kansas Crash Analysis and Reporting System (KCARS) database, which consists of all police-reported crashes in Kansas. Geometric characteristics were obtained from the state's highway inventory database, Control Section Analysis System (CANSYS), which also provides traffic data from the year 2013 that was made available in 2014. Therefore, the study duration was 2011–2013.

3.1.1 Kansas Crash Analysis and Reporting System Database

The KCARS database consists of several tables, including ACCIDENTS, DRIVERS, OCCUPANTS, PEDESTRIANS, TRUCKS, VEHICLES, ACCIDENT_CANSYS, SPECIAL_CONDITIONS, TRAFFIC_CONTROLS, IMPAIRMENT_TESTS, SUBSTANCE_ABUSE, CC_DRIVER, CC_ENVIRONMENT, CC_ROADWAY, and CC_VEHICLE. The ACCIDENT table contains details of each crash, such as crash location, light conditions, weather conditions, road surface type, road conditions, road character, road class, road maintenance information, date of crash, time of crash, day of crash, accident class, and manner of collision. The VEHICLE table contains all characteristics pertaining to the vehicle, including vehicle model, vehicle year, registration year, direction of travel, vehicle maneuver, vehicle damage, and number of occupants. The OCCUPANT table consists of age, gender, safety equipment use, injury severity, and ejection information of each occupant in the vehicle. The field

"UAB Code" in ACCIDENT_CANSYS and ACCIDENT tables indicates crashes occurring on rural highways. The ACCIDENTS, DRIVERS, OCCUPANTS, and ACCIDENT_CANSYS tables provide information regarding crashes occurring at rural multilane highways. These tables were combined and queries were used to filter out crashes on rural multilane highways and five levels of crash severities for occupants.

3.1.1.1 Accident Key

KCARS also contains a field that identifies the location and specific identification (ID) number of each crash. Crash ID is a unique value for each crash that can be used to combine crash characteristics from KCARS to other databases, such as CANSYS, in order to add information about highway geometric characteristics.

3.1.1.2 Crash Location

Several fields in KCARS represent crash location, including the county milepost and distance from a named intersection. However, because incident responders do not typically have precise positioning equipment to determine the specific milepost of an incident, this value can contain inaccuracies. Two columns in KCARS provide longitude and latitude of the crash location.

3.1.1.3 Light Condition

The KCARS database also contains information regarding the light condition at the time of the crash. Crash reports categorize light conditions as daylight; dawn; dusk; dark: street light on; dark: no street light; and unknown. This feature was used to obtain crashes occurring during the day or night. For simplification of analysis, crashes occurring at daylight and dawn were considered to be daytime crashes and other crashes were considered to be nighttime crashes.

3.1.1.4 Crash Severity

KCARS contains three main types of crash severity, with injury severity subdivided as follows (KDOT, 2005):

1) Fatal crashes
2) Injury crashes
    - Possible injury
    - Injury, non-incapacitating
    - Disable, incapacitating
3) Property damage only

Each crash is assigned to the most severe level experienced by persons involved.

Fatal injury

A fatal injury is any injury resulting in death to a person within 30 days of the crash. If a person dies after the 30-day period of crash occurrence or dies of a medical condition, the crash is identified as an injury crash and the injury severity is shown as possible injury (KDOT, 2005).

Possible injury

A possible injury is any reported or claimed injury that is not fatal, incapacitating, or non-incapacitating, including momentary unconsciousness, claim of injuries not evident, limping, complaint of pain, nausea, or hysteria (KDOT, 2005).

Injury (non-incapacitating)

A non-incapacitating injury is any injury, other than a fatal injury or incapacitating injury, which is evident to observers at the scene of the crash at which the injury occurred (KDOT, 2005).

Disabled (incapacitating)

An incapacitating injury is any injury, other than fatal, that prevents the injured person from walking, driving, or performing regular activities he/she was capable of before the injury occurred (KDOT, 2005).

Property Damage Only

KDOT considers crashes involving damage to public or private property totaling more than $1,000 threshold with no injuries to be PDO crashes. Multiple-vehicle crashes can have varying severity levels for each vehicle involved in the crash (KDOT, 2012).

3.1.2 Control Section Analysis System

The CANSYS database contains information about the geometrics, condition, and extent of the 10,000-plus miles of roadways in Kansas, as well as a small proportion of local roadways not in the state highway system. CANSYS, which contains data on bridges, access permits, and at-grade rail crossings, supports the work of various bureaus at KDOT, the FHWA, and the Kansas legislature. The KDOT Geometric and Accident Data unit (GAD) maintains CANSYS (KDOT, 2011).

CANSYS data are collected at random intervals from various sources, and the database is typically used for high-level analyses for network screening and trend evaluations. In this study, the data were sorted by route name and county to account for every mile, but no data were counted twice. Based on data requirement, county mile posts of beginning and ending of segments, coordinates of beginning and ending mile posts of segments, lane width, left shoulder width, right shoulder width, median width, side slope (fore slope), and AADT for the year 2013 were obtained from this database. CANSYS also contains the ROUTE_ID, ROUTE_DIR, LANE_CLASS, SHOR_DESC (outer shoulder description), and SHIN_DESC (inner shoulder description).

3.1.2.1 Beginning and Ending Mile Post and Segment Length

Mileposts in Kansas increase from south to north for odd routes and west to east for even routes, as is customary in the United States. KDOT has state mileposts and county mileposts that begin at the state line or county line. In the CANSYS database, beginning and ending mileposts are defined by a crash report or an intersection. Segment length was calculated from the difference in beginning and ending mileposts.

3.1.2.2 Lane Class and City Code

Lane class identifies the type of highway facility, from undivided two-lane segments to divided eight-lane segments. For this study, segments classified as 2 and 3, representing 4U segments and 4D segments, respectively, were filtered out; the remaining segments were not used. The city code ID number dictates whether the segment is urban or rural. Only city code 999 represents a rural segment. This study utilized the FHWA definition of urban, which requires a population to be equal to or larger than 5000 people. Application of “999” under CITY_CITY_NBR, UAB_CITY, and UAB_UACE_HPMS_CODE fields obtained rural locations.

3.1.2.3 Segment Length

The length of segments used was homogeneous in this study. As suggested by the HSM, segment lengths were at least 0.1 miles; only a few of the segments did not meet this requirement and were excluded from the study.

3.1.2.4 AADT

As mentioned, the CANSYS database provided varying AADTs for the year 2013 for calibration of 4D and 4U segments.

3.1.3 Google Maps

Google Maps and Google Earth® were used to obtain information regarding the presence of lighting at segments because this data is not readily available through KDOT. “Street View” in the Google application enabled zooming in order to determine the presence of a light post. Although the resolution was low in both Google applications, light posts were observed. Figure 3.1 shows the Google Maps application to ascertain the presence of lighting at a segment.

Figure 3.1 Using Google Map to obtain presence of lighting

A summary of data sources is shown in Table 3.1. The HSM considers the presence of automated speed enforcement as optional (desired) data. Since Kansas does not have automated speed enforcement, this data was not applicable for Kansas. Once all data were obtained, they were used in accordance with the HSM methodology.

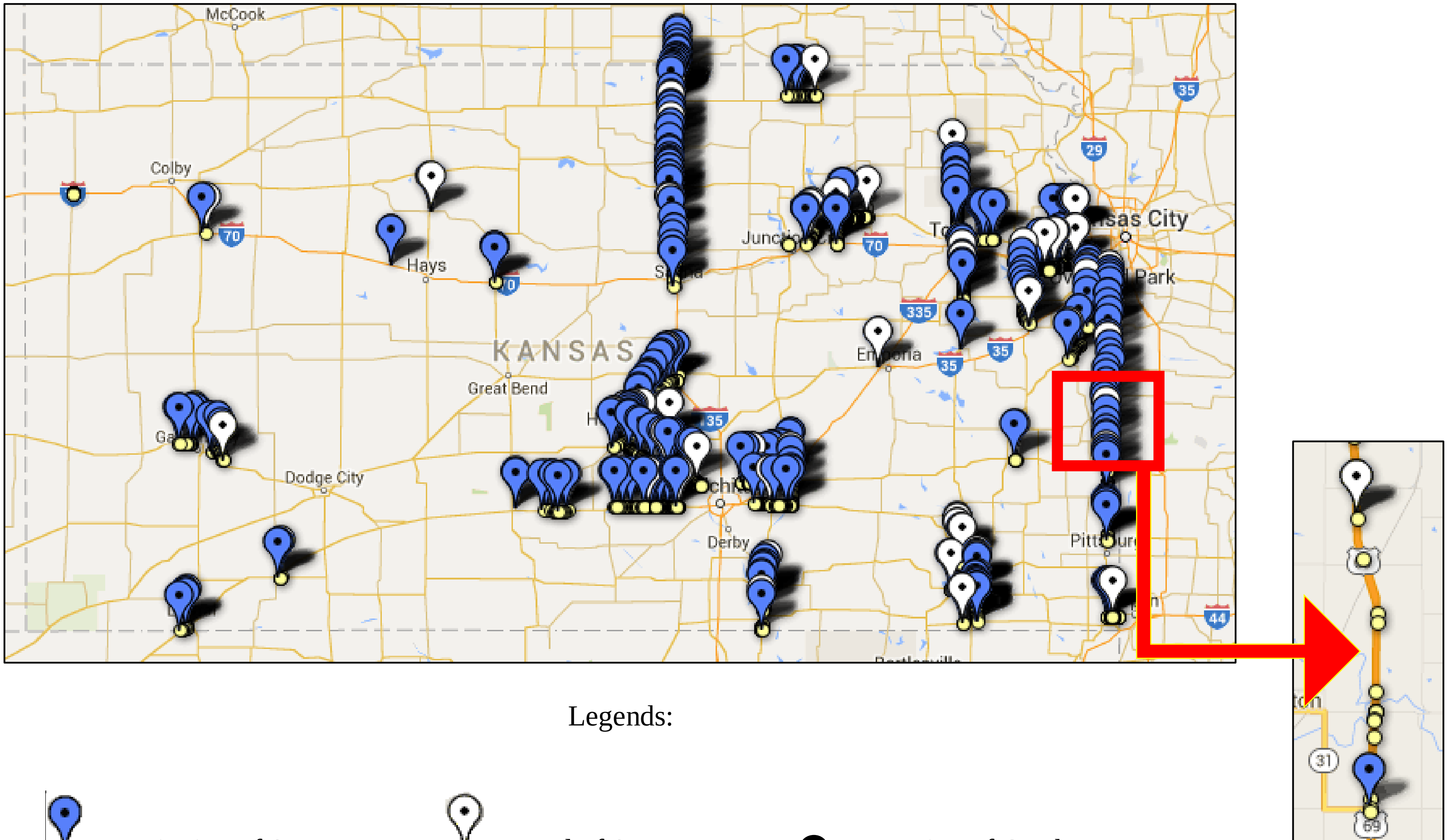
Table 3.1 Data sources for rural four-lane segments

Data Description	Source
AADT	CANSYS
Lane Width	
Median Width	
Shoulder Width	
Side Slope	
Presence of Lighting	Google Maps
Number of Crashes	KCARS
Automated Speed Enforcement	Not Applicable

3.2 Study Segments

The CANSYS database provided a list of rural 4D segments and 4U segments in Kansas. The HSM recommends that segments should be at least 0.1 miles long and contain homogeneous geometry and traffic volume within the length. KDOT uses similar rule of homogeneity for defining their segments within CANSYS database. Using these criterion, a total of 281 4D and 83 4U segments were selected and used for calibration in this study according to the HSM methodology. The number of crashes for all 4D segments was 910 per year, and the number of crashes for 4U segments was 44 per year. Lane width, shoulder width, median width, and side slope were also obtained from the CANSYS database.

Google Maps® was used to show crash locations as well as the beginning and ending of segments, demonstrating that segments were spread throughout Kansas. Figures 3.2 and 3.3 show crash locations at 4D and 4U rural roadway segments in Kansas, respectively. Blue and white markers indicate the beginning and end of segments, respectively, and small dot markers identify crash locations on 4D and 4U highways.



Legends:

- Beginning of Segment
- End of Segment
- Location of Crash

Figure 3.2 Rural 4D segments and crash location map

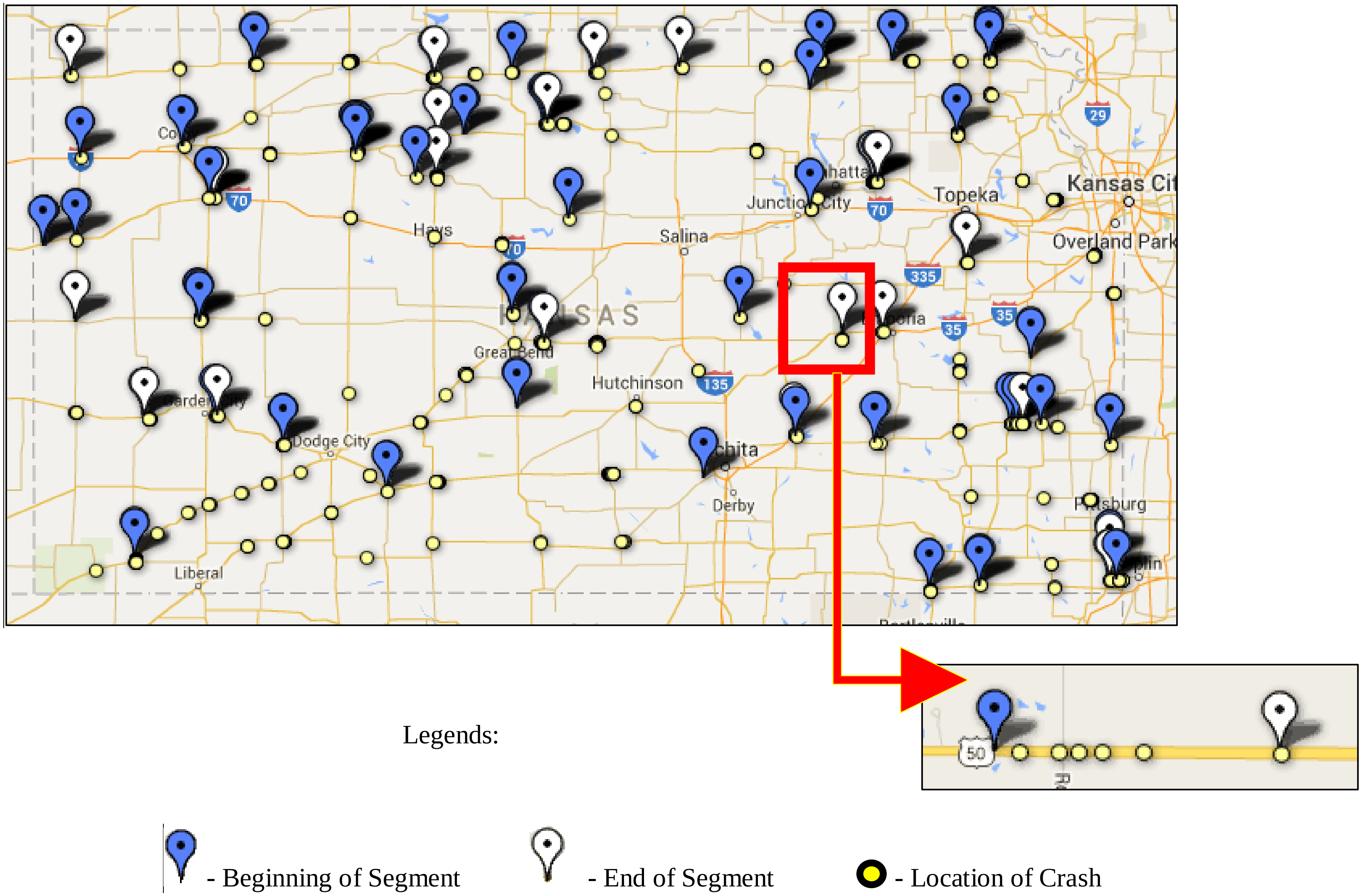


Figure 3.3 Rural 4U segments and crash location map

3.3 Highway Safety Manual Calibration Procedures for Segments

Prediction of the expected number of crashes for an entity given a set of values for input variables follows a three-step process in the HSM. Beginning with an SPF, CMFs and the calibration factor (C) subsequently follow (AASHTO, 2010). The SPF predicts expected crash frequency as a function of AADT and lane width for roadway segments given basic geometrics and traffic conditions. For example, base conditions for a rural four-lane roadway include 12-foot-wide lanes, 8-foot-wide right shoulders (for divided segments), 30-foot-wide median (for divided segments), 1:7 or flatter side slope (for undivided segments), paved 6-foot-wide shoulder (for undivided segments), no lighting, and no automated speed enforcement. Expected crash frequency for sites with characteristics differing from base conditions can be computed by multiplying CMFs that represent each type of change. After all available CMFs are considered, calibration factor C is used as the ultimate adjustment for all other differences, known or unknown, measurable or immeasurable, such as climate, driver and animal populations, CRTs, and crash reporting system procedures. Factor C is the ratio of observed number of crashes to expected number of crashes. This building block structure of the HSM predictive methods enables separate calibration (AASHTO, 2010).

Because the SPF carries the most weight in predicting crashes, SPF calibration may be more critical and effective than other modifications. Ideally, base conditions should be the most representative characteristic of a roadway, guaranteeing a sizable sample in order to develop statistically robust models. However, the most representative roadway type may vary by state or region. If the sample size that satisfies the base conditions is small, SPF calibration may not be rigorous or representative enough for a larger population (AASHTO, 2010).

The standard approach to develop calibration factors in the HSM involves the following steps:

- Identify desired facility types
- Select segments among these types
- Collect required data for those segments
- Apply HSM predictive models
- Compute calibration factors

This research considered rural 4D and 4U segments, and all segments within these categories were selected as analysis locations. Once the site type and locations were selected, methodology given in the HSM was followed for calibration.

3.3.1 Safety Performance Functions

SPFs are regression equations that calculate the dependent variable, or predicted crash frequency, based on independent variables. Because this research focused on utilization of the HSM-specified methods, SPFs in the HSM were used to calculate the number of predicted crashes (AASHTO, 2010).

SPF for a rural four-lane highway segment is estimated as

$$N_{SPF} = e^{[a+b\times ln\ (AADT)\ +\ ln\ (L)]} \tag{3.1}$$

Where,

N_{SPF} = base total expected average crash frequency for the rural segment,

$AADT$ = AADT on the highway segment,

L = Length of highway segment (miles), and

a and b = regression coefficients.

3.3.2 Crash Modification Factors

The SPF was multiplied by CMFs for each independent variable, as described in the HSM (AASHTO, 2010). CMFs only address changes in design and operation characteristics (e.g., lane width and shoulder width) typically under the control of highway engineers and designers. They do not address characteristics such as climate, driver behavior, and CRT (Kweon et al., 2014). Equation 3.2 shows the SPF to obtain predicted number of crashes on 4D and 4U segments in the HSM.

$$N_{Predicted} = N_{spf} \times 1.436 \times (CMF_1 \times CMF_2 \times \ldots\ldots\ldots . CMF_i) \tag{3.2}$$

Where,

$N_{Predicted}$ = Adjusted number of predicted crash frequency,

N_{spf} = Total predicted crash frequency under base condition, and

CMF_i = Crash modification factors.

A CMF greater than 1.0 indicates an expected increase in crashes, demonstrating that the countermeasure decreases safety in that location. A CMF less than 1.0 indicates a reduction in crashes after implementation of the given countermeasure, demonstrating that the countermeasure increases safety in that location.

Chapter 11 in the HSM provides CMFs corresponding to lane width, shoulder width, median width, and side slope. CMF for the presence of lighting was calculated using Equation 3.3. As recommended by the HSM, default proportions of nighttime crashes in the HSM were replaced by Kansas specific crashes.

$$CMF_{Lighting} = 1 - [(1 - 0.72 \times P_{inr} - 0.83 \times P_{pnr}) \times P_{nr}] \tag{3.3}$$

Where,

P_{inr} = Proportion of nighttime crashes for unlighted segments involving fatality or injury,

P_{pnr} = Proportion of nighttime crashes for unlighted segments involving PDO crashes, and

P_{nr} = Proportion of total crashes for unlighted segments that occurring at night.

3.3.3 Calibration Factor

SPFs in the HSM were typically developed using data from jurisdictions and/or time periods rather than where or when such SPFs were desired. For example, default HSM-SPFs for rural multilane highways were developed using data from Texas, California, Minnesota, New York, and Washington from 1991 to 1998. However, the general level of crash frequencies potentially varied substantially from one jurisdiction to another and/or from one year to another due to changes in climate, driver behavior, and CRT and the calibration factor addresses these changes (AASHTO, 2010). Therefore, in order to predict reflecting levels of crash frequencies in jurisdictions and/or years of interest, the predicted number of crash frequencies must be adjusted using calibration factors that are determined for each facility/site type.

Calibration factor (C) was obtained by dividing the number of total observed crashes by the number of total predicted crashes. Observed crash frequencies were obtained from the crash database, and predicted crashes were obtained by the HSM-SPF after applying CMFs. A calibration factor less than 1.0 indicates that the HSM-SPF overpredicted crash frequencies. Therefore, multiplying the factor prediction under base conditions lowers the predictions to match observed frequencies on average. A factor greater than 1.0 indicates underprediction; multiplying the factor increases the predictions to match observed frequencies. Equation 3.4 was used to obtain the calibration factor.

$$C = \frac{\sum \text{all sites observed crashes}}{\sum \text{all sites predicted crashes}} \tag{3.4}$$

3.4 SPF Development

When a calibration factor obtained according to the HSM methodology underpredicts or overpredicts crashes for a particular location, the HSM recommends development of local jurisdiction-specific SPF. This section describes frequently used approaches that could be utilized in developing a new SPF for a roadway facility.

3.4.1 Poisson Regression Model

A Poisson regression model is a generalized linear model, which allows the mean of a population to depend on a linear predictor through a nonlinear link function. This model, which allows the response probability distribution to be any member of an exponential family of distributions, is appropriate for dependent variables that have nonnegative integer values such as 0, 1, 2, etc. Therefore in most cases, Poisson regression can precisely analyze count data. Miaou and Lum (1993) determined the relationship between vehicle crashes and geometric design features of road segments, such as lane width, shoulder width, horizontal curvature, and lane width, therefore, proposed the Poisson regression model, as shown in Equation 3.5.

$$P\,(Y_i = y_i) = p\,(y_i) = \frac{{\mu_i}^{y_i} e^{-\mu_i}}{y_i!}\,,\ (i = 1,2,3,\ldots., n;\ y_i = 0,1,2,3,\ldots) \tag{3.5}$$

Where,

i = A roadway segment (the same roadway segments in other sample periods are considered to be separate roadway segments),

Y_i = The number of crashes for a given time period for roadway segment i,

y_i = The actual number of crashes for a given time period for roadway segment i,

$P(y_i)$ = Probability of crash occurrence for a given time period on roadway segment i, and

μ_i = Mean value of crashes occurring in a given time period as,

$$\mu_i = \mathrm{E}\,(\mathrm{Y_i}) = \theta_i \left[e^{\sum_{j=1}^{k} x_{ij}\beta_j}\right] \tag{3.6}$$

Where,

x_{ij} = The independent j^{th} variable for roadway segment i,

β_j = The coefficient for the j^{th} independent variable, and

θ_i = Traffic exposure for roadway segment i.

For each roadway segment i, x_i independent variables describe geometric characteristics, traffic conditions, and other relevant attributes. Traffic exposure, or the amount of travel during the sample year, can be computed using Equation 3.7.

$$\theta_i = 365 \times AADT_i \times T\% \times l_i \tag{3.7}$$

Where,

$AADT_i$ = Annual average daily traffic (number of vehicles),

$T\%$ = Percentage of all vehicles in traffic stream, and

l_i = Length of road segment.

A Poisson regression model assumes that crash numbers for a given time period for roadway segment (Y_i, i = 1,2,3….,n) are independent of each other and has Poisson distribution with mean μ_i. The expected number of crashes $E(y_i)$ is proportional to motor vehicle travel θ_i. The model ensures that crash frequency is positive, using an exponential function given by Equation 3.8.

$$\lambda_i = \frac{E(y_i)}{\theta_i} = \exp(x_i \beta) \tag{3.8}$$

Where,

λ_i = Crash-involvement frequency

$E(y_i)$ = The expected number of crashes,

x_i = Transpose of covariate vector,

θ_i = Amount of motor vehicle travel, and

β = Vector of unknown regression parameter.

The maximum likelihood method in the SAS GENMOD procedure can be used to estimate parameters of the Poisson regression model for *log (μ)*. One important property of the Poisson regression is that it restricts the mean and variance of the distribution to be equal, written as

$$Var(y_i) = E(y_i) = \mu_i \tag{3.9}$$

Where,

μ_i = Mean of response variable y_i,

$E(y_i)$ = Expected number of response variable, and

$Var(y_i)$ = Variance of response variable y_i.

Using an inappropriate model can affect statistical inference and resulting conclusions. Deviance and a Pearson Chi-square statistic divided by degrees of freedom can be used to detect overdispersion or underdispersion in the data. The degree of freedom can be obtained by reducing the number of parameters estimated in the model from the total number of roadway segments considered for crash prediction modeling.

According to Miaou and Lum (1993), overdispersion could originate from several sources, including uncertainty of vehicle exposure, omitted variables, or a highway environment that is not homogeneous. To account for overdispersion, a scale (dispersion) parameter with respect to the Poisson model can be introduced into the relationship between variance and mean. Although parameter estimates are not affected by the scale parameter, the estimated covariance matrix is affected by this factor, meaning that parameter estimates are not changed, but their standard errors are inflated by the value of scale parameter, wider confidence intervals, higher p-values, and more conservative significance tests than Poisson distribution before the adjustment. Introduction of scale parameters gives a correction term for testing parameter estimates under Poisson distribution but not a different probability distribution. Consideration of a distribution that permits more flexible modeling of the variance is another way to address overdispersion. Hence, use of Negative Binomial regression modeling would be the next step in analysis. The Negative Binomial regression model is more appropriate for overdispersed data because it relaxes the constraint of equal mean and variance. Miaou and Lum (1993) proposed the Negative Binomial regression model specifically for overdispersed data.

3.4.2 Negative Binomial Regression Model

The Negative Binomial regression model is commonly used to develop a crash prediction model. Consider a set of n number segments of a highway. Let Y_i be a random variable that represents the number of vehicles involved in crashes on highway section i during the analysis period. Further, assume the amount of vehicle travel or exposure on this highway segment V_i is also a random variable estimated through a highway sampling system. For each highway segment,

i is a $k \times 1$ vector of explanatory variables, denoted by $x_i = (x_{i1} = 1, x_{i2} \ldots\ldots x_{ik})'$, describing its geometric characteristics, traffic conditions, and other relevant attributors. Given V_i and x_i, crash involvements Y_i, $i = 1,2,3\ldots\ldots, n$ are postulated to be independent and each is Poisson distributed as

$$P\,(Y_i = y_i) = \frac{(\lambda_i \theta_i)^{y_i} e^{-\lambda_i \theta_i}}{y_i!} \tag{3.10}$$

Where,

λ_i = Motor vehicle crash involvement and

θ_i = Exponential of random error.

If the log-linear rate function is used as follows, the model becomes the Negative Binomial regression model that gives the relationship between the expected number of crashes occurring at the i^{th} segment and K number of parameters:

$$\lambda_i = \exp\,(\beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i) \tag{3.11}$$

Where,

λ_i = Number of crashes on highway segment i (with Negative Binomial distribution conditional on ε_i),

β_0 = Constant term,

$\beta_i \ldots \ldots \ldots, \beta_n$= Estimated parameters in vector form,

$x_i \ldots \ldots \ldots, x_n$= Explanatory variables in vector form, and

ε_i = Random error, (exponential is distributed as gamma with mean 1 and variance α^2).

Negative Binomial distribution is a consequence of gamma heterogeneity in Poisson means. The effect of the error term in the Negative Binomial regression model allows for overdispersion of the variance, such that

$$Var\ (y_i) = E(y_i) + \alpha E(y_i)^2 \tag{3.12}$$

Where,

α = The overdispersion parameter,

$E(y_i)$ = Expected mean number of crashes on highway segment i, and

$Var\ (y_i)$ = Variance of the number of crashes y_i.

Variance over the mean is called the overdispersion rate; variance is explained in Equation 3.13.

$$\frac{Var\ (yi)}{E(yi)} = 1 + \alpha E(y_i) \tag{3.13}$$

Where,

$E(y_i)$ = Expected mean number of crashes on highway segment i, and

$Var\ (y_i)$ = Variance of the number of crashes y_i.

If overdispersion α is equal to zero, the Negative Binomial reduces to the Poisson model (Long, 1997). The larger the value of α, the more variability is in the data beyond that associated with mean $E(y_i)$. For the Poisson regression model, coefficients β_1 are estimated by maximizing the log likelihood $\log_e L(\beta)$. The maximum likelihood method in the SAS GENMOD procedure can be used to estimate parameters of the Negative Binomial regression model for $log(\mu)$ and the overdispersion parameter α (Long, 1997).

The HSM has several requirements for making a jurisdiction-specific SPF and for using the Negative Binomial. This model requires the same base conditions as required in the HSM (Section 3.3). Variables such as automated speed enforcement are not prevalent on rural Kansas highways. The model must also include AADT and segment length.

The study conducted by Bornheimer (2012) used two approaches to develop SPF for rural two-lane highway segments. The first approach was identical to the approach used in the HSM.

The second approach, however, utilized known CMFs and actual number of crashes and found the exponent on e, noted as X in Equation 3.14, for each segment. Negative Binomial regression was then run using only that exponent.

$$X = \left(\frac{\mathrm{N}_{known}}{EXPO \times CMF_{combined}}\right) \tag{3.14}$$

Where,

N_{known} = Number of crashes known for the segment and

$CMF_{combined}$ = All CMFs multiplied together.

The other main equation form, shown in Equation 3.15, was considered as an exponential function of the AADT and length, thus allowing predicted crashes to grow exponentially as the AADT increased.

$$A = AADT^{b1} Len^{b2(linear\ terms)} \tag{3.15}$$

With

$$linear\ terms = C_0 + C_1 x_1 + C_2 x_2 + \cdots + C_n x_n \tag{3.16}$$

Where,

A = Annual crash frequency in crashes per segment per year,

$AADT$ = Average annual daily traffic demand,

Len = Street segment length,

x_i = Selected traffic and geometric characteristics, and

b_i, b_2, C_i = Regression coefficients.

This form of equation was created using a reverse method identical to the HSM's CPM model. The level of significance was 0.05, meaning that the model had a confidence level of 95%. Negative Binomial regression was initially run using all available variables, and then it was run again using only variables that had a *p*-value of 0.05 or lower. Thus the final equations to be tested were obtained.

3.4.3 Model Validation Statistics

The following statistical tests were run to determine which models more accurately predicted the number of crashes. They were used in accordance with engineering judgment to discern if the results matched known guidelines.

3.4.3.1 Akaike Information Criterion

The Akaike information criterion (AIC) is a measure of the relative quality of statistical models for a given set of data. For a collection of data models, AIC estimates the quality of each model relative to the other models (Hilbe, 2011). For a set of candidate models for the data, the preferred model has the minimum AIC value, which can be obtained using Equation 3.17.

$$AIC = -2Ln\,(L) + 2k \tag{3.17}$$

Where,

Ln (L) = Model log-likelihood, and

k = Number of predictors.

3.4.3.2 Akaike Information Criterion Corrected

Akaike information criterion corrected (AIC_C) depends on sample size: The smaller the AIC_C value, the better the model. Increasing sample size causes an increasing trend to accept the more complex model when selecting a model based on AICc (Garber and Wu, 2001). The AICc value of the model can be obtained using Equation 3.18.

$$AIC_c = -2Ln\,(L) + 2k + \frac{2k(k+1)}{(n-k-1)} \tag{3.18}$$

Where,

Ln (L) = Model log-likelihood,

k = Number of predictors, and

n = Number of model observations.

3.4.3.3 Bayesian Information Criterion

The Bayesian information criterion (BIC), which is often used in model selection and is based on the likelihood function, accounts for the possibility of overfitting an equation by penalizing equations if too many variables are used. BIC is calculated and given when the Negative Binomial regression is run; therefore, none of the calibration methods contain this value because their CPM equation was already created. Low BIC values indicate better models. The BIC value of the model can be obtained using Equation 3.19.

$$BIC = -2Ln\,(L) + kLn\,(n) \tag{3.19}$$

Where,

$Ln\,(L)$ = Model log-likelihood,

k = Number of predictors, and

n = Number of model observations.

3.4.3.4 Mean Prediction Bias

In this study, the mean prediction bias (MPB) was used to identify overdispersion in each of the models, comparing actual and predicted crashes. The MPB was calculated using Equation 3.20, where a small number indicated less overprediction or underprediction. A positive MPB indicated overprediction, and a negative MPB indicated underprediction (Garber et al., 2011).

$$MPB = \frac{\sum(y_i - x_i)}{n} \tag{3.20}$$

Where,

x_i = Actual number of crashes on a segment,

y_i = Predicted number of crashes on a segment, and

n = Number of segments.

3.4.3.5 Mean Absolute Deviation

The mean absolute deviation (MAD) gave a measure of the average magnitude of variability when each model was compared to the actual number of segments. The MAD's only distinction from the MPB is that negative and positive differences are unable to cancel each other out, either underpredicting or overpredicting the total amount. The MAD was calculated using Equation 3.21.

$$MAD = \frac{\sum |(y_i - x_i)|}{n} \tag{3.21}$$

Where,

x_i = Actual number of crashes on a segment,

y_i = Predicted number of crashes on a segment, and

n = Number of segments.

3.5 Intersection Data

The calibration of rural multilane intersections using HSM methodology pertains to three-leg intersection with minor-road stop control (3ST), four-leg intersection with minor-road stop control (4ST), and four-leg signalized intersection (4SG). To date, the 4SG intersection calibration methodology is not complete in HSM, so only 4ST and 3ST intersections were calibrated in this study. The intersections were preliminarily obtained from the CANSYS database. However, the CANSYS database did not have a complete list of intersections available at the time of this study and most of the required intersection-related information was missing. Therefore, existing intersections were found via Google Maps ®. Figures 3.4 and 3.5 show typical 4ST and 3ST intersections in Google Maps ®, respectively.

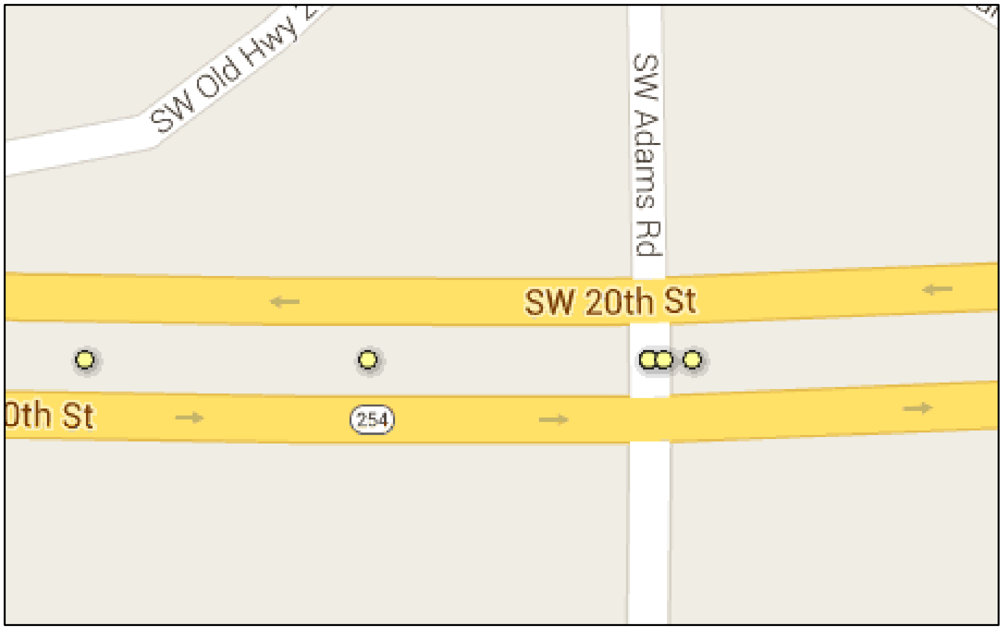


Figure 3.4 4ST intersection with stop control at minor approach

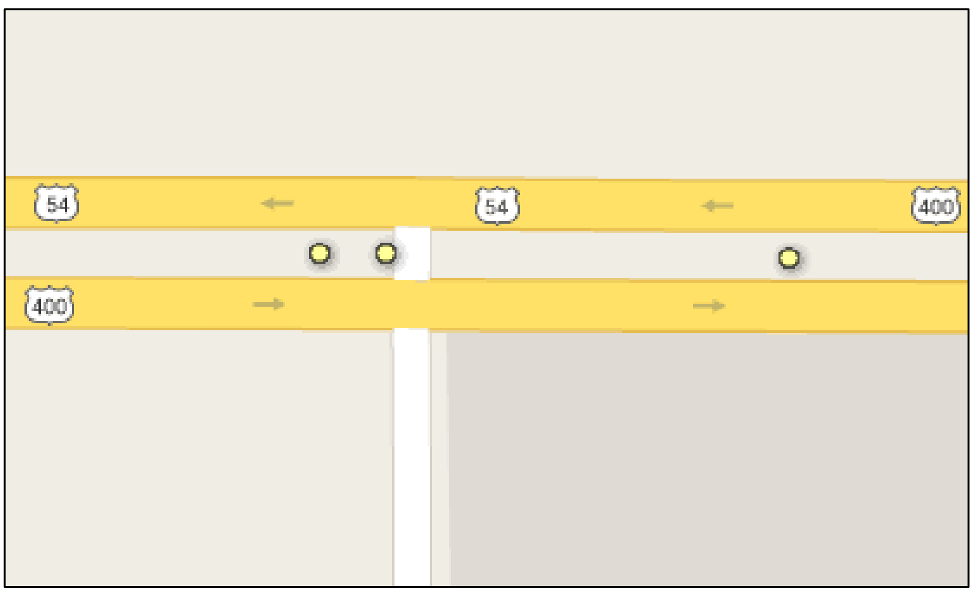


Figure 3.5 3ST intersection with stop control at minor approach

Each intersection was zoomed to Street View in these maps to obtain corresponding intersection skew angle, presence of right turn lane on major road, presence of left turn lane on major road, and presence of lighting posts at intersections. Several intersections were difficult to determine whether they were 3ST or 4ST, so the identified intersections were cross-checked using KDOT-monitored Videologs. Figure 3.6 illustrates use of RoadView Explorer to view intersections through videologs.

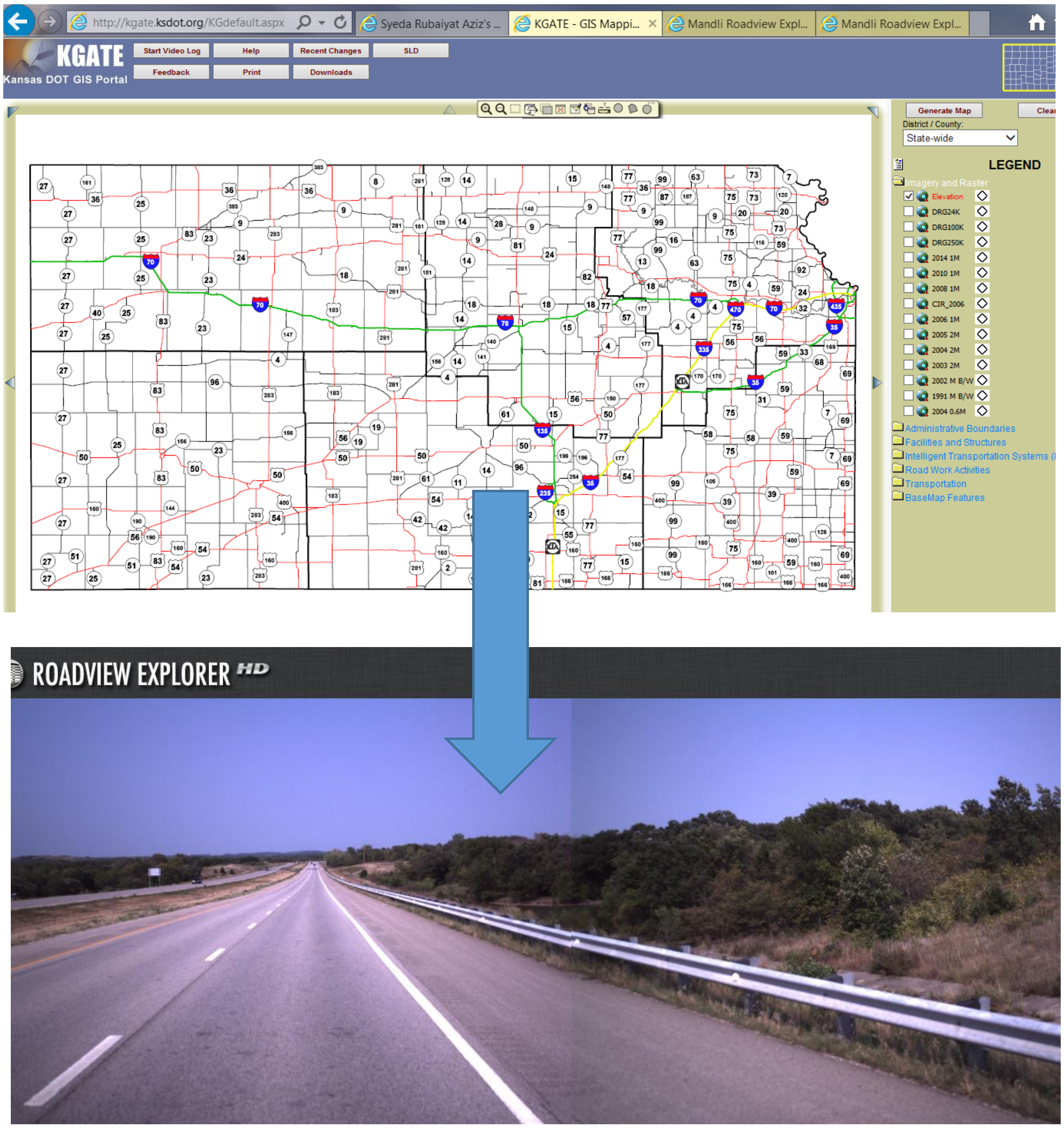


Figure 3.6 Use of KDOT videologs

After completing data collection via Google Maps and KDOT videologs, a total of 199 4ST intersections and 65 3ST intersections at minor approaches were considered in the calibration. Because the HSM provides no precise guidelines regarding the number of observed crashes at intersections, observed crashes at intersections were counted using two methods. The first method considered crashes within an intersection-box of 300 ft. along each approach leading to the intersections regardless of whether or not crashes were intersection-related. Figure 3.7 shows an example of an intersection-box at an intersection. The second method considered the “intersection related” column in the KCARS database, which distinguishes whether or not crashes are intersection related irrespective of crash distance from named intersections.

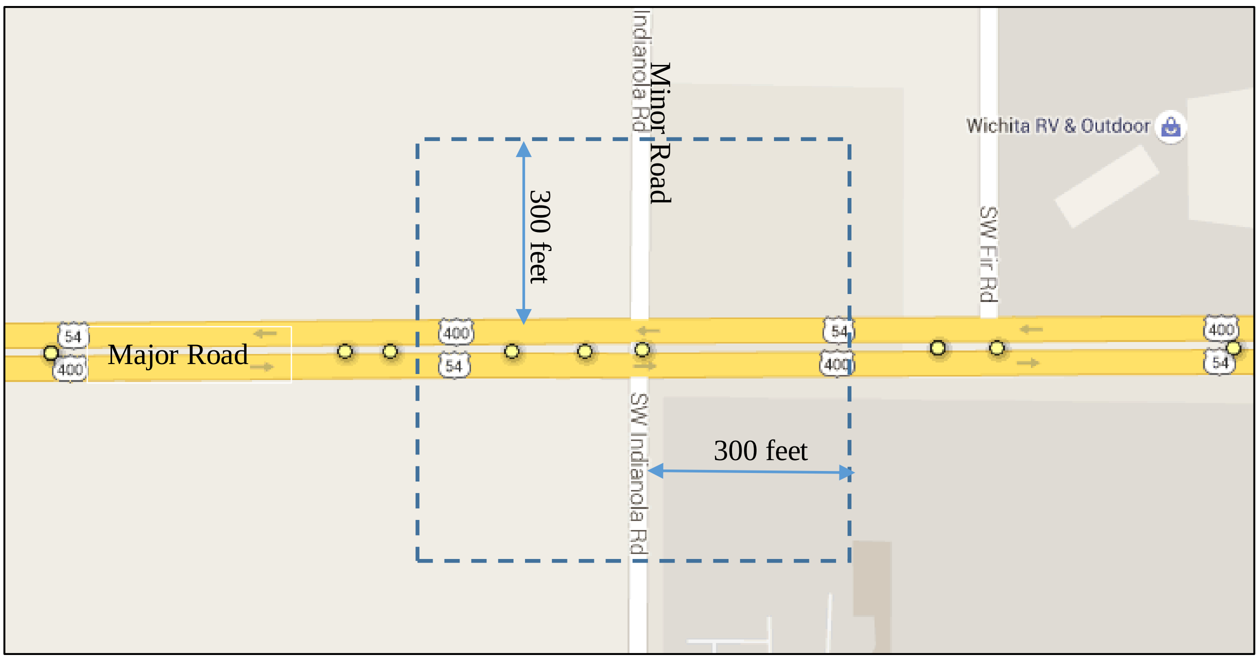


Figure 3.7 **Intersection-box demonstration**

3.6 Highway Safety Manual Calibration Procedures for Intersections

A three-step process for segments was followed to calibrate SPFs in the HSM. The SPF for rural intersections has two alternative functional forms in the HSM: one form considers AADT on major and minor road approaches (Equation 3.22), and the other form considers combined AADT on major and minor road approaches (Equation 3.23).

$$N_{spf\ int} = \exp[a + b \times \ln(AADT_{maj} + c \times \ln(AADT_{min})] \tag{3.22}$$

$$N_{spf\ int} = \exp[a + d \times \ln(AADT_{total})] \tag{3.23}$$

Where,

$N_{spf\ int}$ = SPF estimate of intersection-related expected average crash frequency for base conditions,

$AADT_{maj}$ = AADT (vehicles per day) for major-road approaches,

$AADT_{min}$ = AADT (vehicles per day) for minor-road approaches,

$AADT_{total}$ = AADT (vehicles per day) for major-road and minor-road combined approaches, and

a,b,c,d = regression coefficients.

CMFs for intersection skew angle, presence of right turn lane on major road, presence of left turn lane on major road, and presence of lighting posts were obtained using charts and equations provided in the HSM. SPFs at each intersection were multiplied by corresponding CMFs for all intersection-related attributes.

Chapter 4 - Calibration of HSM Predictive Methods

The HSM recognizes that base formulas and default values originally used to develop CPMs may not be applicable for every jurisdiction or state (AASHTO, 2010). Appendix A of Part C of the HSM describes calibration procedures that can provide meaningful, accurate results for each jurisdiction.

4.1 Distribution and Comparison of Crashes

This section provides crash distributions and compares crash-related attributes. The HSM recommends replacement of selected default values and factors in the calibration methodology, but replacement is not necessary to achieve satisfactory results. Therefore, these results could be used to substitute default values (AASHTO, 2010). Data necessary for this procedure could also be segregated by county or district, thereby providing insight into regions within a state that display unique crash characteristics.

4.1.1 Collision Type

Since collision types in the Kansas Motor Vehicle Accident Report did not match those provided in the HSM, additional sorting was necessary to compare crash numbers. For single vehicle crashes, elements such as collisions with legally parked vehicles, fixed objects, and other objects were assigned the collision type "Ran off Road." Because all of these elements exist outside the normal roadway, a departure from the roadway was assumed to be necessary in order to collide with the objects. "Collisions with Railway Train" was combined with "Other Non-Collision" under the heading "Other Single Vehicle Crash." Table 4.1 shows crashes by collision type for rural four-lane highways in Kansas.

Analysis of collision types is crucial since the types of crashes on Kansas highways could influence how crashes are modeled. More than 30% of segment crashes on Kansas highways were a result of collisions with animals. This percentage is significant because animal collision crashes account for a majority of crashes on Kansas rural four-lane highway segments and because the percentage is significantly higher than the HSM-specified default animal-related crash proportion of 12%.

Table 4.1 Percentage of crashes by collision type for Kansas rural four-lane highways

Collision Type	Year			3-Year Average
	2011	2012	2013	
Animal-related	37.9	39.4	34.1	37.13
Ran-off-Road	29.1	27.8	32.2	29.70
Moving Vehicle	20.7	20.5	20.6	20.60
Rollover	7.45	7.5	8.5	7.82
Other Single Vehicle Crashes	4.6	4.6	4.4	4.53
Pedestrian	0.1	0.0	0.12	0.07
Pedal Cyclist	0.1	0.0	0.0	0.03
Unknown	0.1	0.1	0.1	0.10

4.1.2 Severity Level

Table 4.2 lists crashes on rural 4D highways based on injury severity of vehicle occupants. Injury crashes are further divided into three categories of incapacitating injury, non-incapacitating injury, and possible injury crashes, thus making it five-level injury severity distribution. This distribution was developed by analyzing all crashes in the data set that were not intersection or intersection-related. Each crash was counted only once and was attributed to the highest severity level. Therefore, if a crash had incapacitating injuries and non-incapacitating injuries, it was only counted as incapacitating.

Table 4.2 Crash severity level on four-lane highways

Crash Severity Level	Year						3-Year Average
	2011		2012		2013		
	Count	Percent	Count	Percent	Count	Percent	
Fatal	27	1.5	21	1.4	17	1.5	22
Incapacitating (disabled) Injuries	49	2.7	37	2.4	29	2.5	38
Non-incapacitating Injuries	157	8.7	132	8.5	119	9.9	136
Possible Injuries	96	5.3	80	5.2	65	5.4	80
PDO	1,479	81.7	1,285	82.5	969	80.7	1244

Results from Table 4.2 show that Kansas crashes are typically less severe than those detailed in the default jurisdiction of the HSM (AASHTO, 2010). Approximately 19% of rural four-lane crashes in Kansas resulted in fatality or injury.

Table 4.3 demonstrates distribution by collision type for specific crash severity levels on rural four-lane roadway segments. The same crashes in Tables 4.1 and 4.2 were used for this table, but the crashes were further categorized by type of collision with another vehicle. Once the crashes were categorized as fatal, injury, or PDO, the crashes were assigned using collision types from the Kansas Motor Vehicle Accident Report.

Table 4.3 Crashes by collision type and severity level for four-lane roadways

Collision Type	2011			2012			2013		
	F (%)	I (%)	PDO (%)	F (%)	I (%)	PDO (%)	F (%)	I (%)	PDO (%)
Head-On	20.0	5.4	3.0	20.0	3.9	0.5	23.1	3.0	0.0
Rear End	20.0	45.9	38.1	0.0	46.7	41.6	15.4	50.3	47.3
Angle (side impact)	55.0	38.4	16.8	70.0	35.6	16.3	61.5	28.4	15.9
Sideswipe (opposite direction)	5.0	1.6	1.1	0.0	1.7	0.8	0.0	2.0	0.2
Sideswipe (same direction)	0.0	8.1	33.0	10.0	11.7	32.6	0.0	13.2	29.8
Backed Into	0.0	0.0	1.5	0.0	0.0	1.6	0.0	0.0	0.9
Other	0.0	0.5	6.4	0.0	0.6	5.7	0.0	2.0	5.5
Unknown	0.0	0.0	0.2	0.0	0.0	1.0	0.0	1.0	0.2

4.1.3 Nighttime Crash Proportions

The Kansas Motor Vehicle Accident Report designates five values for light conditions: daylight; dawn; dusk; dark: streetlights on; dark: no streetlights; and unknown. Crashes marked as "unknown" represented a very small portion of the total crashes and may have been a result of undocumented light conditions. In order to determine proportions necessary for Table 11-15 in the HSM, crashes labeled as "unknown" were removed from the count of total crashes. Crashes for daylight and dawn were considered daytime crashes. Crashes in each category are shown in Table 4.4.

Table 4.4 Crash distribution by light condition

Light Condition	Year			3-Year Average
	2011	2012	2013	
Daylight	479	417	523	473
Dawn	65	72	61	66
Dusk	32	27	29	29
Dark (street lights on)	58	75	82	72
Dark (no street lights)	514	475	480	490
Total	1,148	1,066	1,175	1130

Table 4.5 contains nighttime crash proportions for unlighted roadway segments. The HSM provides these proportions in Table 11-15 but recommends obtaining jurisdiction-specific values. As shown in Equation 3.3, the CMF corresponding to the presence of lighting involves proportions of nighttime crashes. These proportions were obtained for rural 4D and 4U highways in Kansas and were compared to HSM default values.

Table 4.5 Proportion of nighttime crashes for rural 4D and 4U highways in Kansas

Roadway Type	Nighttime Crash Proportions	Kansas Highways	HSM Given Default
4D	P_{inr}	0.599	0.426
	P_{pnr}	0.124	0.323
	P_{nr}	0.876	0.677
4U	P_{inr}	0.477	0.255
	P_{pnr}	0.127	0.361
	P_{nr}	0.873	0.639

P_{inr} = proportion of nighttime crashes for unlighted segments involving fatality or injury
P_{pnr} = proportion of nighttime crashes for unlighted segments involving PDO crashes
P_{nr} = proportion of total crashes for unlighted segments occurring at night

4.2 Calibration of Rural Multilane Segments

Study segments were obtained from the CANSYS database in order to calibrate SPFs given in the HSM. The HSM suggests a minimum segment length of 0.1 miles. After applying the length condition, a total of 283 rural 4D segments and 83 4U segments were obtained from the CANSYS database for calibration using the HSM methodology.

From the KCARS database, the number of crash frequencies for all 4D segments was 910 crashes per year and the number of crash frequencies for all 4U segments was 44 crashes per year. Figures 4.1 and 4.2 show crash distributions of 4D and 4U segments, respectively. Total crashes for 4D far exceeded the 100-crashes-per-year requirement, but all 4U segments did not meet this requirement. Therefore, the HSM recommendation to consider all available segments with existing crashes was followed (AASHTO, 2010).

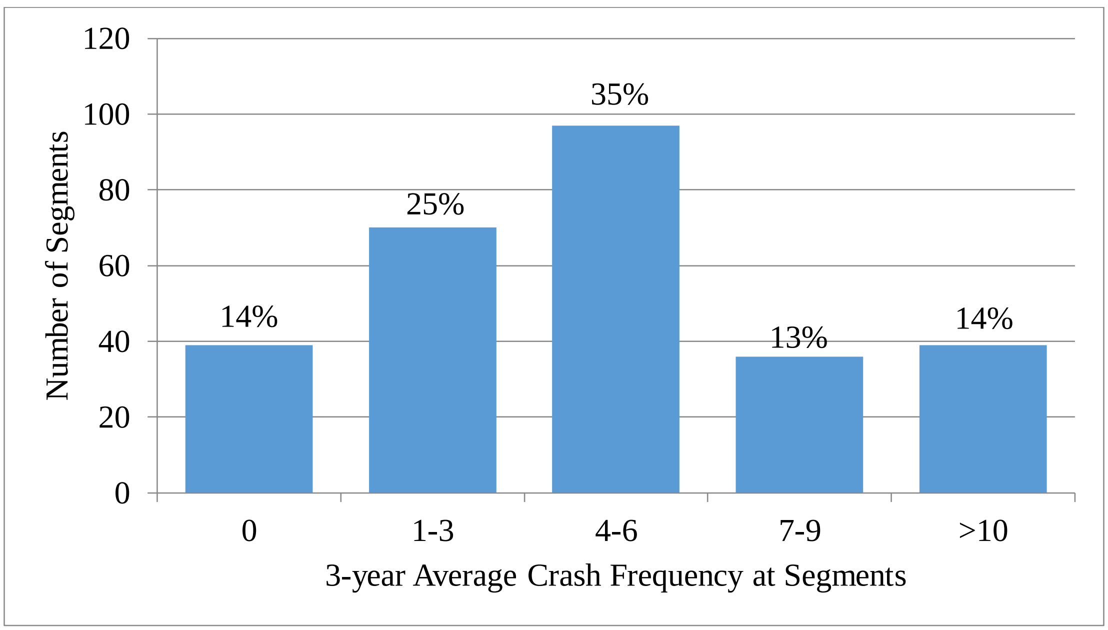


Figure 4.1 Distribution of crash frequency on 4D segments

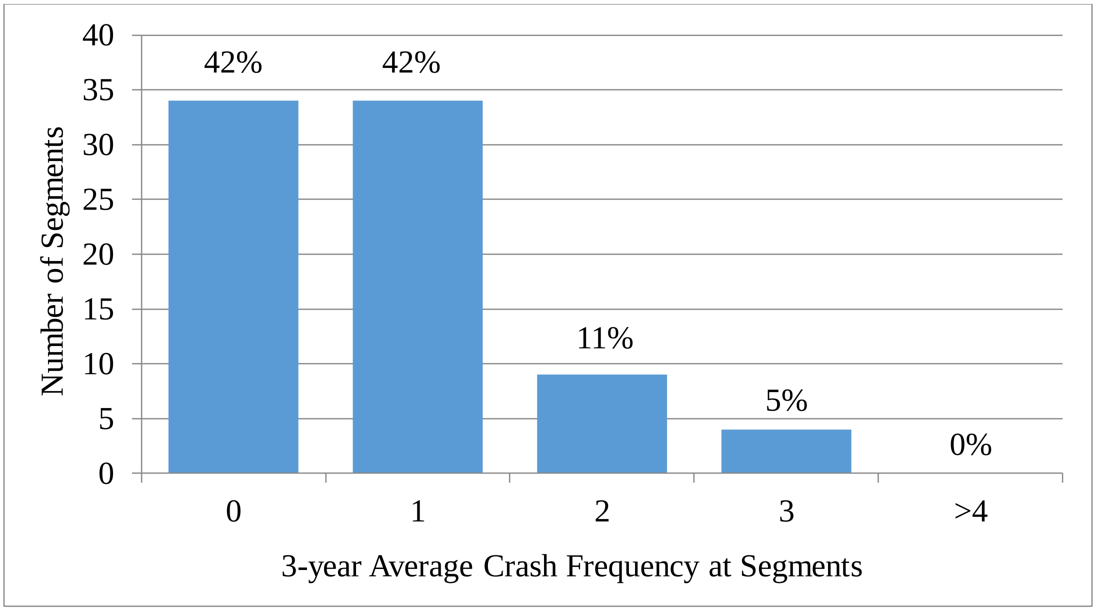


Figure 4.2 Distribution of crash frequency on 4U segments

Descriptive statistics for 4D and 4U segments are shown in Table 4.6. The average length of 4D segments was 1.53 miles, well above the minimum length of 0.1 miles, with segment lengths ranging between 0.1 miles and 8.629 miles. The length-slandered deviation was 1.55 miles. Traffic volumes averaged 8,000 vpd, with a maximum of 31,000 vpd. Segments were relatively uniform with respect to lane and shoulder width, but they showed variation with respect to median width. The average number of crashes was 9.72, with the numbers of crashes ranging from zero to 98. Standard deviation of crashes was 11.90, which was larger than the average. Seventy-eight segments had lighting present, but no automated speed enforcement is currently applicable for highways in Kansas.

The average length of the 4U segments was 0.28 miles, very close to the minimum length of 0.1 miles. Segments ranged in length between 0.1 miles and 0.86 miles. The length standard deviation was 0.16 miles. Traffic volumes averaged 4,114 vpd, with a maximum of 12,600 vpd. Segments were relatively uniform with respect to lane width, but they showed variation with respect to shoulder width. Side slope was required data for rural 4U segments; these segments had a minimum slope of 1:2 and maximum slope of 1:6. The average number of crashes was 1.59, with the numbers of crashes ranging from zero to 11. The standard deviation of crashes was 2.14, which was larger than the average. The total number of crashes was 132 (for three years), or 44 crashes per year, which was less than the HSM's recommendation of 100 crashes per year. Because this study included all possible 4U segments, calibration was performed with these segments. Only 20 segments had lighting present, but no automated speed enforcement is currently applicable for rural undivided highways in Kansas.

After obtaining the observed crash frequency, the next step in the study was to obtain the predicted number of crash frequency. For each segment, the HSM-given SPF was obtained using Equation 3.1. CMFs were obtained for lane width, shoulder width, median width (4D), and side slope (4U) for each segment using charts and equations provided in the HSM (AASHTO, 2010).

Table 4.6 Descriptive statistics for rural four-lane segments

Roadway Type	Description	Average	Minimum	Maximum	Std. Dev.
4D	Length (mile)	1.53	0.1	8.63	1.55
	AADT (vpd)	8,000	490	31,000	4657
	Left lane width (ft.)	12.06	10.99	20.99	0.59
	Right lane width (ft.)	12.06	10.99	20.99	0.59
	Left paved shoulder width (ft.)	5.68	0	9.84	1.43
	Right paved shoulder width (ft.)	9.35	0	9.84	1.84
	Median width (ft.)	30.65	4.92	152.00	15.79
	Number of crashes	9.72	0	98.0	11.90
	Presence of lighting	0.28	0	1	0.44
	Presence of automated speed Enforcement	-	-	-	-
4U	Description	Average	Minimum	Maximum	Std. Dev.
	Length (mile)	0.28	0.1	0.86	0.16
	AADT (vpd)	4,114	460	12,600	2919
	Left lane width (ft.)	12.45	10.00	22.51	1.33
	Right lane width (ft.)	12.45	10.00	22.51	1.33
	Left paved shoulder width (ft.)	5.05	0	10.00	4.68
	Right paved shoulder width (ft.)	4.83	0	10.00	4.66
	Side slope	-	1:2	1:6	-
	Number of crashes	1.59	0	11.0	2.14
	Presence of lighting	0.24	0	1	0.43
	Presence of automated speed Enforcement	-	-	-	-

Table 4.7 shows the 4D segment calculation worksheet from Microsoft Excel. CMFs were obtained from Tables 11-16, 11-17, and 11-18 of Chapter 11 of the HSM for lane widths, shoulder widths, and median widths, respectively (AASHTO, 2010). After applying the CMFs, final N_{spf} for each rural 4D segment was obtained, which was the number of predicted crashes. The summation of predicted crashes for all 283 4D segments was 1,902, and the total number of observed actual crashes was 2,730. A calibration factor of 1.43 was obtained by dividing the total observed crashes by the total predicted crashes; a separate calibration factor was obtained for fatal and injury crashes. Total observed fatal and injury crashes on these segments were 328, and predicted crashes from SPF were 1,008; thus, Equation 3.4 yielded a calibration factor of 0.52. Table 4.8 shows details of calibration factors for 4D segments.

Table 4.7 4D segments sample worksheet

1	ID	PREFIX	ROUTE ID	RTE DIR	BEGIN CO MP	END CO MP	Segement Length (mile)	MED TYPE DESCR	SHOR DESC	AADT SMRY AADT CNT	MED MEDN MED WDTH	MED MEDN MED WDTH (feet)	CMF (Median)	SHLD SHOR SHLDR WDTH	SHLD SHOR SHLDR WDTH (feet)	CMF (Shoulder)	SHLD SHIN SHLDR WDTH	SHLD SHIN SHLDR WDTH (feet)	LANE LN1R LN WDTH	LANE LN1R LN WDTH (feet)	CMF (Lane Width)	LANE LN2R LN WDTH	LANE LN2R LN WDTH (feet)	Presence of Lighting	CMF (Lighting)
2	163	U0005400	U	EB	10.357	11.161	0.804	- Depress	ized, (CA(	2275	2.4	7.87	1.04	3	9.84	1	1.2	3.94	3.66	12.01	1.00	3.66	12.01	n	1.00
3	498	U0006900	U	NB	6.009	6.93	0.921	- Depress	tuminous	2695	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
4	499	U0006900	U	NB	6.93	8.097	1.167	- Depress	tuminous	3420	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
5	500	U0006900	U	NB	8.097	9.067	0.97	- Depress	tuminous	3950	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
6	513	U0006900	U	NB	12.715	13.155	0.44	- Depress	tuminous	2830	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
7	514	U0006900	U	NB	13.155	15.235	2.08	- Depress	ement cor	2830	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
8	515	U0006900	U	NB	15.235	18.273	3.038	- Depress	ement cor	2685	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
9	516	U0006900	U	NB	18.273	22.323	4.05	- Depress	ement cor	2545	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
10	517	U0006900	U	NB	22.323	25.356	3.033	- Depress	ement cor	2420	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
11	641	K0025400	K	EB	0	2.479	2.479	- Depress	tuminous	5650	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
12	644	K0025400	K	EB	2.729	7.957	5.228	- Depress	tuminous	5700	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
13	645	K0025400	K	EB	7.957	10.225	2.268	- Depress	tuminous	5100	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
14	646	K0025400	K	EB	10.225	10.493	0.268	- Depress	tuminous	6000	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
15	648	K0025400	K	EB	10.548	13.157	2.609	- Depress	tuminous	6000	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
16	649	K0025400	K	EB	13.157	13.94	0.783	- Depress	tuminous	6800	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
17	695	U0005400	U	EB	2.985	6	3.015	- Depress	tuminous	8750	5.5	18.04	1.02	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
18	696	U0005400	U	EB	6	8.933	2.933	- Depress	tuminous	7600	5.5	18.04	1.02	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
19	709	U0005400	U	EB	10.716	15.085	4.369	- Depress	tuminous	3155	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
20	710	U0005400	U	EB	15.085	17.191	2.106	- Depress	tuminous	2710	9.1	29.86	1.00	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
21	711	U0005400	U	EB	17.191	17.47	0.279	- Depress	tuminous	1185	6.1	20.01	1.02	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
22	712	U0005400	U	EB	17.47	20.41	2.94	- Depress	tuminous	1185	10.67	35.01	0.99	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
23	713	U0005400	U	EB	20.41	24.405	3.995	- Depress	tuminous	1165	10.67	35.01	0.99	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
24	714	U0005400	U	EB	24.405	25.448	1.043	- Depress	tuminous	1645	10.67	35.01	0.99	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
25	790	U0007700	U	NB	34.985	35.757	0.772	- Depress	tuminous	1295	7	22.97	1.02	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
26	791	U0007700	U	NB	35.757	36.03	0.273	- Depress	tuminous	660	7	22.97	1.02	3	9.84	1	1.8	5.91	3.66	12.01	1.00	3.66	12.01	n	1.00
27	925	K0006600	K	EB	0.656	0.811	0.155	- Depress	tuminous	4020	6.1	20.01	1.02	2.4	7.87	1	1.5	4.92	3.66	12.01	1.00	3.66	12.01	n	1.00
28	926	K0006600	K	EB	0.811	1.247	0.436	- Depress	tuminous	4020	10.4	34.12	1.00	2.4	7.87	1	1.5	4.92	3.66	12.01	1.00	3.66	12.01	n	1.00
29	927	K0006600	K	EB	1.247	1.638	0.391	- Depress	tuminous	4020	6.1	20.01	1.02	2.4	7.87	1	1.5	4.92	3.66	12.01	1.00	3.66	12.01	n	1.00

1	Presence of Lighting	CMF (Lighting)	BEG LON	BEG LAT	END LON	END LAT	Icon Name (Basic)	SECTION No.	No. of Fatal Crash	No. of Injury Crash	No. of Property Damage Crash	Total (F+I+PO)	Total Crashes (MATLAB)	No. of Daytime Crashes	No. of Nighttime Crashes	Night time Fatal Crash	Night time Injury Crash	Night time PDO Crash	No. of Nighttime Crashes	N (spf) [Total Crash]	AADT 2013	N (spf) [Fatal/Injury Crash]	Predicted Total Crashes	Predicted Fatal/Injury Crashes
2	n	1.00	-95.33	37.9221	-95.316	37.9221	arge_blu	1	0	1	5	6	6	3	3	0	0	3	3	0.665	4550	0.373	2.08	1.16
3	n	1.00	-94.705	37.7555	-94.705	37.7681	arge_blu	2	0	1	6	7	7	5	2	0	0	2	2	0.910	5390	0.503	2.73	1.51
4	n	1.00	-94.705	37.7681	-94.705	37.7863	arge_blu	3	0	0	5	5	5	1	4	0	0	4	4	1.481	6840	0.800	4.44	2.40
5	n	1.00	-94.705	37.7863	-94.705	37.8015	arge_blu	4	0	2	4	6	6	1	5	0	2	3	5	1.432	7900	0.764	4.30	2.29
6	n	1.00	-94.702	37.8559	-94.703	37.8622	arge_blu	5	0	0	7	7	7	0	7	0	0	7	7	0.458	5660	0.252	1.37	0.75
7	n	1.00	-94.703	37.8622	-94.705	37.8923	arge_blu	6	0	0	13	13	13	3	10	0	0	10	10	2.164	5660	1.190	6.49	3.57
8	n	1.00	-94.705	37.8923	-94.707	37.9359	arge_blu	7	0	3	20	23	23	8	15	0	1	14	15	2.991	5370	1.652	8.97	4.96
9	n	1.00	-94.707	37.9359	-94.705	37.9938	arge_blu	8	0	2	18	20	20	13	7	0	0	7	7	3.770	5090	2.092	11.31	6.28
10	n	1.00	-94.705	37.9938	-94.707	38.0373	arge_blu	9	0	0	12	12	12	8	4	0	0	4	4	2.678	4840	1.493	8.03	4.48
11	n	1.00	-97.153	37.7961	-97.108	37.7978	arge_blu	10	1	6	19	26	26	15	11	0	1	10	11	5.327	11300	2.750	15.98	8.25
12	n	1.00	-97.103	37.7979	-97.009	37.7982	arge_blu	11	1	11	44	56	56	16	40	1	6	33	40	11.338	11400	5.848	34.01	17.54
13	n	1.00	-97.009	37.7982	-96.971	37.8066	arge_blu	12	0	6	12	18	18	10	8	0	2	6	8	4.377	10200	2.280	13.13	6.84
14	n	1.00	-96.971	37.8066	-96.967	37.8081	arge_blu	13	0	2	5	7	7	6	1	0	0	1	1	0.613	12000	0.315	1.84	0.94
15	n	1.00	-96.966	37.8084	-96.92	37.8161	arge_blu	14	0	3	16	19	19	9	10	0	1	9	10	5.971	12000	3.065	17.91	9.20
16	n	1.00	-96.92	37.8161	-96.905	37.8172	arge_blu	15	0	2	9	11	11	4	7	0	1	6	7	2.043	13600	1.037	6.13	3.11
17	n	1.00	-97.099	37.6795	-97.044	37.6797	arge_blu	16	1	15	35	51	51	19	32	1	6	25	32	10.250	17500	5.085	31.37	15.56
18	n	1.00	-97.044	37.6797	-96.99	37.6795	arge_blu	17	0	13	38	51	51	24	27	0	8	19	27	8.601	15200	4.322	26.32	13.22
19	n	1.00	-96.958	37.6766	-96.881	37.6726	arge_blu	18	0	6	14	20	20	4	16	0	5	11	16	5.095	6310	2.773	15.28	8.32
20	n	1.00	-96.881	37.6726	-96.844	37.6762	arge_blu	19	0	3	6	9	9	1	8	0	2	6	8	2.094	5420	1.155	6.28	3.47

Table 4.7 shows the four-lane divided segment calculation worksheet from Excel. CMFs were obtained from the Tables 11-16, 11-17 and 11-18 of HSM Chapter 11 for lane widths, shoulder widths and median widths respectively (AASHTO, 2010). After applying the CMFs, final N_{spf} for each rural divided segment was obtained, which was the number of predicted crashes. The summation of predicted crashes for all 283 four-lane divided segments was 1,902. The total number of observed actual crashes was 2,730. Finally, a calibration factor of 1.43 was obtained by dividing total observed crashes by total predicted crashes. A separate calibration factor was obtained for fatal and injury crashes. Total observed fatal and injury crashes on these segments were 328 and predicted crashes from SPF were 1,008; thus, Equation 3.4 yielded a calibration factor of 0.52. Table 4.8 shows details of obtaining calibration factor for 4D segments.

Table 4.8 Four-lane divided segments calibration factor calculation

No. of Fatal Crashes	No. of Injury Crashes	Total (Fatal / Injury) Crashes	No. of Property Damage Crashes	Total (Fatal + Injury + Personal Damage Only)	No. of Daytime Crashes	No. of Nighttime Crashes	Nighttime Fatal Crash	Nighttime Injury Crash	Nighttime PDO Crash	No. of Total Nighttime Crashes	Predicted Total Crashes	Predicted Fatal / Injury Crashes
45	483	528	2202	2730	1087	1636	18	185	1433	1636	1901.58	1007.69

$$\text{Total Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{2730}{1901.58} = 1.436$$

$$\text{Fatal and Injury Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{528}{1007.69} = 0.524$$

Table 4.9 Four-lane undivided segments sample worksheet

1	ID	ROUTE ID	PREFIX	RTE DIR	BEGIN CO MP	END CO MP	LENGTH of SEGMENT	LANE CLASS DESCR	SHOR DESC	AADT 2013	SHLD SHOR FORESL OPE	CMF (Fore slope)	SHLD SHOR SHLDR WDTH	CMF (TRA)	Shoulder width (ft)	CMF (WRA)	CMF (Shoulder width)	SHLD SHOL FORESL OPE	SHLD SHOL SHLDR WDTH	LANE LN1R LN WDTH	Lane Width (ft)	CMF (RA)	CMF (Lane Width)	LANE LN2R LN WDTH	LANE LN1L LN WDTH
2	78	U0016900	U	NB	18.647	18.873	0.226	ır lane, u	ıminous	2880	6:01	1.05	3	1	9.84	0.87	0.96	6:01	3	3.66	12.01	1	1	3.66	3.66
3	151	U0005400	U	EB	7.666	8.171	0.505	ır lane, u	ituminous	7910	6:01	1.05	2.8	1	9.19	0.87	0.96	6:01	2.8	4.11	13.48	1	1	4.11	4.11
4	152	U0005400	U	EB	8.171	8.649	0.478	ır lane, u	curb and g	7910	0	1	0	1	0.00	1.50	1.14	0	0	4.11	13.48	1	1	4.11	4.11
5	153	U0005400	U	EB	8.649	9.046	0.397	ır lane, u	ized, (CA	7910	6:01	1.05	3	1	9.84	0.87	0.96	6:01	3	4.11	13.48	1	1	4.11	4.11
6	161	U0005400	U	EB	10.004	10.109	0.105	ır lane, u	ized, (CA	4420	6:01	1.05	3	1	9.84	0.87	0.96	6:01	3	4.11	13.48	1	1	3.66	4.11
7	162	U0005400	U	EB	10.109	10.357	0.248	ır lane, u	ized, (CA	4550	6:01	1.05	3	1	9.84	0.87	0.96	6:01	3	4.11	13.48	1	1	3.66	4.11
8	164	U0005400	U	EB	11.161	11.415	0.254	ır lane, u	ized, (CA	4550	6:01	1.05	3	1	9.84	0.87	0.96	6:01	3	3.66	12.01	1	1	3.66	3.66
9	168	U0005400	U	EB	12.059	12.29	0.231	ır lane, u	curb and g	2740	0	1	0	1	0.00	1.50	1.14	0	0	3.66	12.01	1	1	3.66	3.66
10	169	U0005400	U	EB	12.29	12.56	0.27	ır lane, u	ized, (CA	2740	6:01	1.05	3	1	9.84	0.87	0.96	6:01	3	3.66	12.01	1	1	3.66	3.66
11	180	U0005900	U	NB	12.086	12.406	0.32	ır lane, u	ıminous	1260	4:01	1.12	1.8	1	5.91	1.00	1.00	4:01	1.8	3.66	12.01	1	1	3.66	3.66
12	375	U0005600	U	EB	25.42	25.916	0.496	ır lane, u	curb and g	3860	0	1	0	1	0.00	1.50	1.14	0	0	4.02	13.19	1	1	4.02	4.02
13	401	U0028100	U	NB	17.059	17.344	0.285	ır lane, u	curb and g	5010	0	1	0	1	0.00	1.50	1.14	0	0	3.05	10.01	1.23	1.0621	3.05	3.05
14	402	U0028100	U	NB	17.344	17.59	0.246	ır lane, u	curb and g	4780	0	1	0	1	0.00	1.50	1.14	0	0	3.35	10.99	1.04	1.0108	3.35	3.35
15	501	U0006900	U	NB	9.067	9.296	0.229	ır lane, u	ituminous	7900	6:01	1.05	2.4	1	7.87	1.00	1.00	6:01	2.4	3.66	12.01	1	1	3.66	3.66
16	575	U0007300	U	NB	20.797	20.943	0.146	ır lane, u	inous wit	4000	4:01	1.12	3	1	9.84	0.87	0.96	4:01	3	3.66	12.01	1	1	3.66	3.66
17	582	U0007300	U	NB	22.234	22.517	0.283	ır lane, u	curb and g	2150	0	1	0	1	0.00	1.50	1.14	0	0	3.72	12.20	1	1	3.72	3.72
18	789	U0007700	U	NB	34.13	34.985	0.855	ır lane, u	curb and g	2590	0	1	0	1	0.00	1.50	1.14	0	0	3.66	12.01	1	1	3.66	3.66
19	857	U0005000	U	EB	20.64	20.752	0.112	ır lane, u	ement co	4320	4:01	1.12	3	1	9.84	0.87	0.96	4:01	3	3.66	12.01	1	1	3.66	3.66
20	924	K0006600	K	EB	0.356	0.656	0.3	ır lane, u	curb and g	8040	0	1	0	1	0.00	1.50	1.14	0	0	4.02	13.19	1	1	4.02	4.02
21	930	K0006600	K	EB	3.257	3.682	0.425	ır lane, u	ituminous	8400	6:01	1.05	3	1	9.84	0.87	0.96	6:01	3	3.66	12.01	1	1	3.66	3.66
22	1027	U0040000	U	EB	31.433	31.647	0.214	ır lane, u	ituminous	4970	6:01	1.05	3	1	9.84	0.87	0.96	6:01	3	3.72	12.20	1	1	3.72	3.72
23	1028	U0040000	U	EB	31.647	32.447	0.8	ır lane, u	curb and g	4970	0	1	0	1	0.00	1.50	1.14	0	0	4.02	13.19	1	1	4.02	4.02
24	1057	U0003600	U	EB	14.029	14.245	0.216	ır lane, u	curb and g	2280	0	1	0	1	0.00	1.50	1.14	0	0	4.02	13.19	1	1	4.02	4.02
25	1752	U0008300	U	NB	18.022	18.307	0.285	ır lane, u	curb and g	1480	0	1	0	1	0.00	1.37	1.10	0	0	4.02	13.19	1	1	4.02	4.02
26	2417	U0005000	U	EB	19.133	19.882	0.749	ır lane, u	ituminous	7020	4:01	1.12	3	1	9.84	0.87	0.96	4:01	3	3.66	12.01	1	1	3.66	3.66

1	LANE LN2L LN WDTH	SECT ROUTE	BEG LON	BEG LAT	END LON	END LAT	Presence of Lighting	CMF (lighting)	No. of Fatal Crash	No. of Injury Crash	No. of Property Damage Crash	Total (F+I+PO)	TOTAL (cross-check)	No. of Daytime Crashes	No. of Nighttime Crashes	Night time Fatal Crash	Night time Injury Crash	Night time PDO Crash	No. of Nighttime Crashes	N (spf) [Total Crash]	N (spf) [Fatal/Injury Crash]	Predicted Total Crashes	Predicted Fatal/Injury Crashes
2	3.66	02U00169	-95.24	38.266	-95.238	38.269	y	0.912197	0	0	0	0	0	0	0	0	0	0	0	0.17	0.11	0.47	0.31
3	4.11	01U00054	-95.38	37.922	-95.37	37.922	y	0.912197	0	1	10	11	11	5	6	0	0	6	6	1.25	0.76	3.45	2.11
4	4.11	01U00054	-95.37	37.922	-95.362	37.922	y	0.912197	0	0	1	1	1	0	1	0	0	1	1	1.18	0.72	3.66	2.24
5	4.11	01U00054	-95.362	37.922	-95.354	37.922	n	1	0	0	2	2	2	0	2	0	0	2	2	0.98	0.60	2.98	1.82
6	4.11	01U00054	-95.337	37.922	-95.335	37.922	n	1	0	0	0	0	0	0	0	0	0	0	0	0.13	0.08	0.40	0.25
7	4.11	01U00054	-95.335	37.922	-95.33	37.922	n	1	0	0	0	0	0	0	0	0	0	0	0	0.32	0.20	0.97	0.62
8	3.66	01U00054	-95.316	37.922	-95.311	37.922	n	1	0	0	0	0	0	0	0	0	0	0	0	0.33	0.21	0.99	0.64
9	3.66	01U00054	-95.299	37.922	-95.295	37.922	n	1	0	1	0	1	1	0	1	0	1	0	1	0.16	0.11	0.56	0.37
10	3.66	01U00054	-95.295	37.922	-95.29	37.922	n	1	0	0	1	1	1	1	0	0	0	0	0	0.19	0.13	0.58	0.39
11	3.66	01U00059	-95.17	37.906	-95.17	37.911	n	1	0	0	0	0	0	0	0	0	0	0	0	0.09	0.06	0.31	0.22
12	4.02	05U00056	-98.572	38.353	-98.563	38.353	n	1	0	0	1	1	1	1	0	0	0	0	0	0.53	0.34	1.79	1.16
13	3.05	05U00281	-98.777	38.503	-98.777	38.507	y	0.912197	0	2	2	4	4	2	2	0	0	2	2	0.41	0.26	1.36	0.86
14	3.35	05U00281	-98.777	38.507	-98.777	38.511	y	0.912197	0	0	4	4	4	4	0	0	0	0	0	0.34	0.21	1.05	0.67
15	3.66	06U00069	-94.705	37.801	-94.705	37.805	y	0.912197	0	0	4	4	4	1	3	0	0	3	3	0.56	0.34	1.62	0.99
16	3.66	07U00073	-95.527	39.839	-95.527	39.841	y	0.912197	0	1	2	3	3	2	1	0	0	1	1	0.16	0.10	0.48	0.31
17	3.72	07U00073	-95.527	39.859	-95.527	39.864	n	1	0	0	0	0	0	0	0	0	0	0	0	0.15	0.10	0.51	0.35
18	3.66	08U00077	-96.85	37.85	-96.85	37.862	n	1	0	1	6	7	7	2	5	0	0	5	5	0.57	0.38	1.93	1.29
19	3.66	09U00050	-96.53	38.4	-96.528	38.4	n	1	0	0	0	0	0	0	0	0	0	0	0	0.14	0.09	0.44	0.28
20	4.02	11K00066	-94.71	37.075	-94.705	37.075	n	1	0	1	3	4	4	1	3	0	1	2	3	0.75	0.46	2.57	1.57
21	3.66	11K00066	-94.657	37.075	-94.65	37.075	n	1	0	0	1	1	1	1	0	0	0	0	0	1.12	0.68	3.42	2.08
22	3.72	11U00400	-94.704	37.178	-94.704	37.175	n	1	0	0	0	0	0	0	0	0	0	0	0	0.31	0.19	0.93	0.59
23	4.02	11U00400	-94.704	37.175	-94.704	37.163	n	1	0	0	1	1	1	0	1	0	0	1	1	1.14	0.72	3.89	2.47
24	4.02	12U00036	-101.79	39.767	-101.79	39.765	n	1	0	0	1	1	1	1	0	0	0	0	0	0.12	0.08	0.42	0.28
25	4.02	20U00083	-100.53	39.829	-100.53	39.833	n	1	0	0	1	1	1	0	1	0	0	1	1	0.10	0.07	0.32	0.23

Table 4.9 shows the 4U segment calculation worksheet from Microsoft Excel. CMFs were obtained from Tables 11-11, 11-12, 11-13, and 11-14 of Chapter 11 of the HSM for lane widths, shoulder widths, and side slopes, respectively (AASHTO, 2010). The summation of predicted crashes for all 83 4U segments was 88.23, and the total number of observed actual crashes was 132. A calibration factor of 1.50 was obtained by dividing the total observed crashes by the total predicted crashes; again, a separate calibration factor was obtained for fatal and injury crashes. Total observed fatal and injury crashes on these segments were 20, and predicted crashes from SPF were 56; thus, Equation 3.4 yielded a calibration factor of 0.36. Table 4.10 shows details of calibration factors for 4U segments.

Table 4.10 4U segments sample worksheet

No. of Fatal Crash	No. of Injury Crashes	Total (Fatal / Injury) Crashes	No. of Property Damage Crashes	Total (Fatal + Injury + Personal Damage Only)	No. of Daytime Crashes	No. of Nighttime Crashes	Nighttime Fatal Crash	Nighttime Injury Crash	Nighttime PDO Crash	No. of Total Nighttime Crashes	Predicted Total Crashes	Predicted Fatal / Injury Crashes
0	20	20	112	132	69	63	0	8	55	63	88.28	55.68

$$\text{Total Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{132}{88.28} = 1.495$$

$$\text{Fatal and Injury Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{20}{55.68} = 0.359$$

The calibration factor for total crashes on rural four-lane divided and undivided segments indicates that the HSM underpredicts total crashes by 56% and 50% and overpredicts fatal and injury crashes by 48% and 64% on rural four-lane divided and undivided segments, respectively. In summary, the following Equations 4.1 and 4.2 for 4D segments and Equations 4.3 and 4.4 for 4U can be used for future crash predictions in rural Kansas.

$$N_{Total,\ Predicted} = N_{spf} \times 1.436 \times (CMF_1 \times CMF_2 \times \ldots\ldots\ldots . CMF_i) \tag{4.1}$$

$$N_{F/I,\ \ Predicted} = N_{spf} \times 0.524 \times (CMF_1 \times CMF_2 \times \ldots\ldots\ldots . CMF_i) \tag{4.2}$$

$$N_{Total,\ Predicted} = N_{spf} \times 1.495 \times (CMF_1 \times CMF_2 \times \ldots\ldots\ldots . CMF_i) \tag{4.3}$$

$$N_{F/I,\ Predicted} = N_{spf} \times 0.359 \times (CMF_1 \times CMF_2 \times \ldots\ldots\ldots . CMF_i) \tag{4.4}$$

Where,

$N_{Predicted}$ = Adjusted number of predicted crash frequency,

N_{spf} = Total predicted crash frequency under base condition,

CMF_i = Crash modification factors, and

C_i = Calibration factor.

4.2.1 Modification of HSM-given SPF

Results from the calibration process showed that the HSM methodology underpredicts total crashes on rural multilane highways in Kansas but overpredicts fatal and injury crashes. Therefore, existing SPF given in the HSM was modified to improve crash prediction in rural Kansas. Appendix A of Part C in the HSM describes three components pertaining to SPF modification for a state with available local data. FHWA has funded efforts to develop guidance for this modification (Srinivasan et al., 2013).

In order to increase the accuracy of the HSM procedures, states have been encouraged to customize the procedures with local data (AASHTO, 2010), including developing calibration factors to be applied to default SPFs in the HSM. However, optimum HSM customization for each state requires consideration of factors such as availability of data and resources. Therefore, this research identified a methodology to customize the HSM for Kansas as accurately as its resources allow.

Customization of the HSM is possible through a combination of three components: SPF, CMF, and calibration factor. For example, the HSM typically can be customized with calibration factors calculated from local data, default SPFs, and crash proportions, allowing states that lack available data and resources the opportunity to develop individualized SPFs. However, many other methods can be used to customize the HSM by combining the three components. Although these methods are not explicitly described in the predictive methods of the HSM, they can be inferred from Appendix A and relevant references. Dixon et al. (2012) explored several options related to calibration factors and crash proportions under default SPFs in the HSM. This dissertation developed new regression coefficients for existing HSM-given SPF.

As previously shown in Equation 3.1, the SPF considers segment length and AADT to be independent variables, considering *a* as the intercept of the model and *b* as the parameter estimate for AADT. The original SPF given in the HSM did not show a coefficient for segment length in the model, indicating that 1.0 should be used as a factor in order to obtain the calibration factor. However, while using Kansas-specific data, a new coefficient *p* corresponding to segment length was added to the model, as given in Equation 4.1.

$$N_{SPF} = e^{[a+b\times\ln(AADT)+\mathrm{p}\times\ln(L)]} \quad (4.1)$$

Where,

N_{SPF} = Base total expected average crash frequency for the rural segment,

AADT = AADT on the highway segment,

L = Length of the highway segment (miles), and

a, *b*, and *p* = Regression coefficients.

In order to perform this task, data from the existing set of segments were used to develop a Negative Binomial regression model. Separate models were developed for 4D and 4U segments. Table 4.11 compares regression coefficients given in Chapter 11 of the HSM for 4D and 4U segments with coefficients based on Kansas-specific data.

Table 4.11 Comparison of regression coefficients

Severity Level	**Default HSM Coefficients**			**Kansas-Specific Coefficients** (standard errors)		
	a	b	Coefficient for L	a	b	p
4D						
Total Crashes	-9.025	1.049	1.0	-6.317 (0.631)	0.795 (0.071)	0.898 (0.035)
Fatal and Injury Crashes	-8.837	0.958	1.0	-10.030 (1.133)	1.059 (0.125)	0.399 (0.058)
4U						
Total Crashes	-9.653	1.176	1.0	-6.347 (1.495)	0.822 (0.176)	0.912 (0.227)
Fatal and Injury Crashes	-9.410	1.094	1.0	-8.206 (3.149)	0.817 (0.367)	0.747 (0.439)

Parameter estimates of 4D and 4U differed significantly at all severity levels. The *t*-test was used to determine if slope coefficients obtained for Kansas rural multilane segment data differed from default values at the 0.05 significance level. According to *t*-test results, SPFs in Kansas were statistically different from corresponding default HSM-given SPFs. The newly obtained regression coefficients were used to obtain predicted crashes at each 4D and 4U segment, and then the calibration factor for each facility type was estimated. Calculated calibration factors for 4D facilities were close to 1.0, as shown in Table 4.12; however, a calibration factor of 0.858 was obtained for total and injury crashes on rural 4U segments, which was less than the usual acceptance limit of 1.0 to indicate that the SPF accurately predicts crash frequency for the facility type and matches local conditions. One reason for this low calibration factor could be the small sample size of 4U segments.

Overall, results showed that modification of the SPF with Kansas-specific regression coefficients improved crash frequency prediction on rural 4D roadway segments in Kansas. However, further research must be conducted on 4U segments in order to achieve precise crash prediction, especially for fatal and injury crashes.

Table 4.12 New calibration factors with the modified SPF

Facility Type	Severity	Calibration Factor
4D	Total Crashes	0.956
	Fatal and Injury Crashes	1.002
4U	Total Crashes	1.019
	Fatal and Injury Crashes	0.858

4.3 Calibration of Rural Multilane Intersections

A total of 199 4ST intersections and 65 3ST intersections at minor approach were considered in the calibration for this study. A total of 229 crashes were observed within an intersection-box for all 4ST intersections, and 53 crashes were observed within an intersection-box for all 3ST intersections. Using intersection-related crashes from the KCARS database, 112 and 17 intersection-related crashes were found for 4ST and 3ST intersections, respectively. Both sets of observed crashes were used to obtain two pairs of calibration factors. Figures 4.3 and 4.4 show crash distributions obtained through both methods for 4ST and 3ST intersections, respectively.

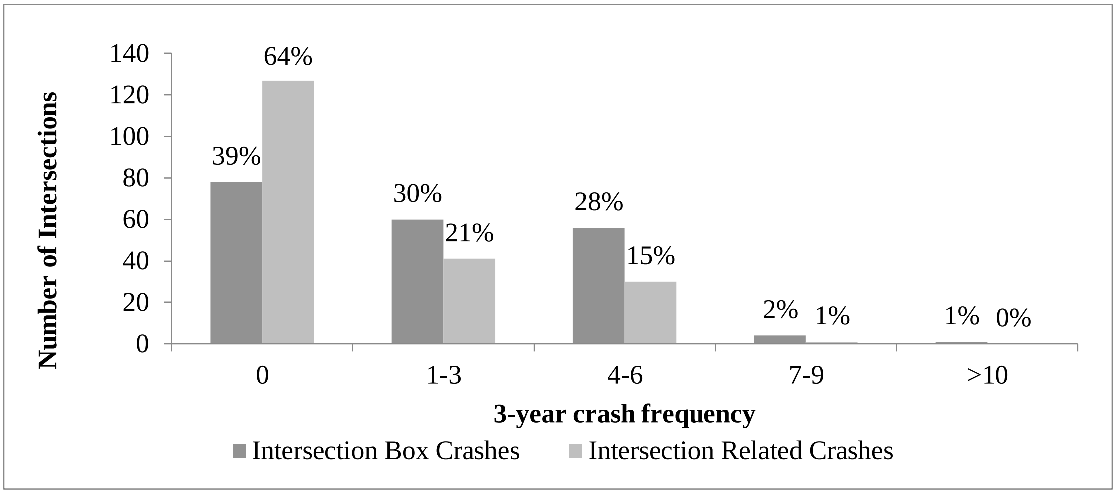


Figure 4.3 Distribution of crash frequency on 4ST intersections

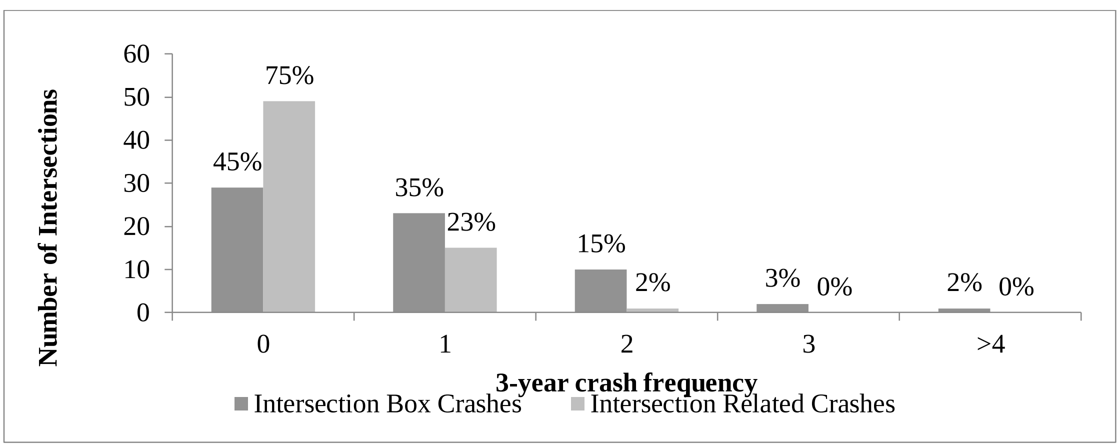


Figure 4.4 Distribution of crash frequency on 3ST intersections

Descriptive statistics for 4ST and 3ST intersections are shown in Table 4.13. For 4ST intersections, the average major road traffic was 7,271 vpd and minor traffic volume was 990 vpd. Some intersections had minor traffic volume as low as 40, but many intersections had high traffic volume of 17,500 vpd. Intersection skew angles averaged 3.92 degrees since most of them were at exact right angles. Only 43 intersections contained right turn lanes, and 30 intersections had lighting posts. The average number of crashes within an intersection-box was 1.15, with the numbers of crashes ranging from zero to 11. Standard deviation of crashes was 1.43, which was larger than the average. Intersection-related crashes from the KCARS database averaged 0.56 crashes, with the numbers of crashes ranging from zero to 5. Standard deviation of crashes was 0.88, which was larger than the average. Automated speed enforcement is not currently applied for 4ST intersections in Kansas, so no corresponding data were obtained.

For 3ST intersections, the average major road traffic was 5,173 vpd and minor traffic volume was 544 vpd. Some intersections had minor traffic volume as low as 20, but many intersections had high traffic volume of 12,600 vpd. Intersection skew angles averaged 1.23 degrees since most of them were exact right angles. Only seven intersections contained right turn lanes, and two intersections had lighting posts. The average number of crashes within an intersection-box was 0.81, with the numbers of crashes ranging between zero and 4. Standard deviation of crashes was 0.92, which was very close to the average. Intersection-related crashes from the KCARS database averaged 0.26 crashes, with the numbers of crashes ranging from zero to 2. Standard deviation of crashes was 0.23, which was less than the average. Automated speed enforcement is not currently applied for 3ST intersections in Kansas, so no corresponding data were obtained.

Table 4.13 Descriptive statistics for rural multilane intersections

Roadway Type	Description	Average	Minimum	Maximum	Std. Dev.
4ST	Major Road AADT (vpd)	7,271	490	17,500	4024
	Minor Road AADT (vpd)	990	40	5,650	1122
	Skew Angle (degrees)	3.92	0	60	12.98
	Presence of Right Turn lane on Major Road	0.21	0	1	0.41
	Presence of Lighting Post	0.15	0	1	0.36
	Number of Crashes within Intersection-box	1.15	0	11	1.43
	Number of Intersection-Related Crashes	0.56	0	5	0.88
3ST	Major Road AADT (vpd)	5,173	490	12,600	3,274
	Minor Road AADT (vpd)	544	20	2,780	543
	Skew Angle (degrees)	1.23	0	30	5.45
	Presence of Right Turn lane on Major Road	0.10	0	1	0.31
	Presence of Lighting Post	0.03	0	1	0.17
	Number of Crashes within Intersection-box	0.81	0	4	0.92
	Number of Intersection-Related Crashes	0.26	0	2	0.23

After obtaining the observed crash frequency, this study obtained the predicted number of crashes. HSM-SPF has two formats for intersection calibration, as previously shown in Equation 3.22 and 3.23. Since major and minor approach AADTs were available, Equation 3.22 was used to obtain predicted crashes at 4ST and 3ST intersections. Charts and equations in the HSM were used to obtain CMFs for intersection skew angle, presence of right turn lane on major road, presence of left turn lane on major road, and presence of lighting posts (AASHTO, 2010).

Table 4.14 shows the 4ST intersection calculation worksheet from Microsoft Excel. CMF factors were obtained from Tables 11-22 and 11-23 and Equations 11-20, 11-21, and 11-22 of Chapter 11 of the HSM for intersection skew angles, left turn lane on major road, right turn lane on major road, and the presence of lighting. After applying the CMFs, final N_{spf} for each rural intersection was obtained, which was the number of predicted crashes. The summation of predicted crashes for all 199 4ST intersections was 252. Using intersection-box (method one), the total number of observed crashes within an intersection-box was 229. A calibration factor of 0.91 was obtained by dividing the total observed crashes by the total predicted crashes. Using method two, a calibration factor of 0.44 was obtained from the total observed 112 intersection-related crashes. A separate calibration factor was obtained for fatal and injury crashes. Total observed fatal and injury crashes on these intersections were 99 from method one and 28 from method two. Calibration factors of 0.74 and 0.21 were obtained from method one and two, respectively, using Equation 3.18. Table 4.15 details calibration factors for 4ST intersections.

Table 4.14 4ST intersection sample worksheet

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Section ID	County Name	Highway No.	Begin County MP	End County MP	Direction (EB/WB/ NB/SB)	Major road AADT (2013)	Minor road AADT (2013)	Number of crashes	Number of int. related crashes	N(SPF)	F/I N(SPF)	No of Entrances/ driveways	No Control (Y/N)	Over-pass Present (Y/N)
2	67	FINNEY	US 50	20.577	23.149	EB	7020	605	4	0	1	1	0	N	N
3	67	FINNEY	US 50	20.577	23.149	EB	7020	400	2	0	1	1	0	N	N
4	67	FINNEY	US 50	20.577	23.149	EB	7020	215	3	0	1	0	0	N	N
5	115	VENWOF	US 24	11.772	11.881	EB	13000	565	0	0	2	1	0	N	N
6	211	PRATT	K 14	26.372	28.287	EB	9040	430	0	0	2	1	0	N	N
7	233	REPUBLIC	US 81	11.135	11.442	NB	5070	455	1	0	1	0	0	N	N
8	270	EDGWICI	US 54	1.48	4.013	EB	11900	925	1	0	3	1	5	N	N
9	1	ALLEN	US 54	10.357	11.161	EB	4550	785	1	0	1	1	7	N	N
10	2	BOURBON	US 69	6.009	6.93	NB	5390	435	0	0	1	0	0	N	N
11	3	BOURBON	US 69	6.93	8.097	NB	6840	665	0	0	1	1	0	N	N

	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	Stop Sign (Y/N)	Signal Present (Y/N)	Yield Sign Present (Y/N)	No. of Legs on Intersection	Lighting Post present (Y/N)	CMF (Lighting)	Right turn lane on Major Road	CMF (Right Turn Lane on Major Road)	CMF (Left Turn Lane on Major Road)	Skew Angle	CMF (Intersection Skew Angle)	No. of Predicted Crashes	No. of Predicted F/I Crashes
2	Y	N	Y	3	N	1.00	N	1.00	0.72	0	1.00	1.04	0.52
3	Y	N	Y	3	N	1.00	N	1.00	0.72	0	1.00	0.87	0.42
4	Y	N	Y	3	N	1.00	N	1.00	0.72	0	1.00	0.66	0.30
5	Y	N	Y	3	N	1.00	Y	0.86	0.72	0	1.00	1.47	0.74
6	Y	N	Y	3	Y	0.90	N	1.00	0.72	0	1.00	1.00	0.49
7	Y	N	N	3	N	1.00	N	1.00	0.72	0	1.00	0.70	0.33
8	Y	N	Y	3	N	1.00	N	1.00	0.72	0	1.00	1.98	1.04
9	Y	N	Y	4	N	1.00	N	1.00	0.72	0	1.00	0.81	0.41
10	Y	N	Y	4	N	1.00	N	1.00	0.72	0	1.00	0.72	0.35
11	Y	N	Y	4	N	1.00	N	1.00	0.72	60	1.10	1.17	0.58

Table 4.15 Calculation of calibration factors for 4ST intersections

Method of Obtaining Observed Crashes at Intersections	No. of Fatal Crashes	No. of Injury Crashes	Total (Fatal / Injury) Crashes	No. of Property Damage Crashes	Total (Fatal + Injury + Personal Damage Only)	No. of Daytime Crashes	No. of Nighttime Crashes	Nighttime Fatal Crash	Nighttime Injury Crash	Nighttime PDO Crash	No. of Total Nighttime Crashes	Predicted Total Crashes	Predicted Fatal / Injury Crashes
1	3	96	99	130	229	62	167	2	17	148	167	252.13	134.67
2	0	28	28	84	112	37	75	0	21	54	75		

Intersection-box (Method 1),

$$\text{Total Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{229}{252.13} = 0.91$$

$$\text{Fatal and Injury Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{99}{134.67} = 0.74$$

Intersection-related crashes (Method 2),

$$\text{Total Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{112}{252.13} = 0.44$$

$$\text{Fatal and Injury Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{28}{134.67} = 0.21$$

Table 4.16 shows the 3ST intersection calculation worksheet from Microsoft Excel. CMFs were obtained from Tables 11-22 and 11-23 and Equations 11-18, 11-19, and 11-22 of Chapter 11 of the HSM for intersection skew angles, left turn lane on major road, right turn lane on major road, and lighting. After applying the CMFs, final N_{spf} for each rural intersection was obtained, which was the number of predicted crashes. The summation of predicted crashes for all 65 3ST intersections was 18.44. Using intersection-box (method one), the total number of observed crashes within an intersection-box was 53. A calibration factor of 2.87 was obtained by dividing the total observed crashes by the total predicted crashes. Using method two, a calibration factor of 0.92 was obtained for the 17 observed intersection-related crashes. A separate calibration factor was obtained for fatal and injury crashes. Total observed fatal and injury crashes on these intersections were 10 from method one and 4 from method two. Calibration factors of 1.16 and 0.47 were obtained from method one and two, respectively, using Equation 3.18. Table 4.17 details calibration factors for 3ST intersections.

Table 4.16 3ST intersection sample worksheet

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	Section ID	County Name	Highway No.	Begin County MP	End County MP	Direction (EB/WB/ NB/SB)	Major road AADT (2013)	Minor road AADT (2013)	Number of all crashes within ints. box of 300	Number of intersection related	N(SPF)	F/I N(SPF)	No of Entrances driveways	No Control (Y/N)	Ove Pr (
2	4	BOURBON	US 69	8.097	9.067	NB	7900	735	1	0	1	0	2	N	
3	27	CHEROKEE	K 66	0.811	1.247	EB	8040	675	1	1	1	0	4	N	
4	29	CHEROKEE	K 66	1.638	2	EB	8040	675	3	1	1	0	4	N	
5	30	BUTLER	K 66	35.757	36.03	NB	8400	1955	2	1	1	1	9	N	
6	30	BUTLER	K 66	35.757	36.03	NB	8400	1955	2	0	1	1	9	N	
7	39	CLOUD	US 81	9.036	12.68	SB	5000	445	1	0	0	0	3	N	
8	40	CLOUD	US 81	12.68	14.168	SB	5000	275	2	0	0	0	0	N	
9	40	CLOUD	US 81	12.68	14.168	SB	5000	20	2	0	0	0	0	N	
10	43	COWLEY	US 77	0	1.977	NB	6340	875	2	0	1	0	2	N	
11	68	FINNEY	US 50	4.931	5.983	EB	4620	345	1	1	0	0	2	N	
12	70	FINNEY	US 83	20.577	23.149	NB	4680	126	1	0	0	0	0	N	

	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	Stop Sign (Y/N)	Signal Present (Y/N)	Yield Sign Present (Y/N)	No. of Legs on Intersection	Lighting Post present (Y/N)	Lighting Post present (Y/N)	CMF (Lighting)	Right turn lane on Major	Right turn lane on Major Road (Y/N)	CMF (Right Turn Lane on	CMF (Left Turn Lane on Major	Skew Angle	CMF (Intersection Skew	No. of Predicted Crashes	No. of Predicted F/I Crashes
2	Y	N	Y	3	N	0	1.00	Y	1	0.86	0.56	0	1.00	0.41	0.19
3	Y	N	Y	3	N	0	1.00	N	0	1.00	0.56	10	1.06	0.51	0.23
4	Y	N	Y	3	N	0	1.00	N	0	1.00	0.56	10	1.06	0.51	0.23
5	Y	N	Y	3	N	0	1.00	N	0	1.00	0.56	0	1.00	0.65	0.31
6	Y	N	Y	3	N	0	1.00	N	0	1.00	0.56	0	1.00	0.65	0.31
7	Y	N	Y	3	N	0	1.00	N	0	1.00	0.56	0	1.00	0.24	0.12
8	Y	N	Y	3	N	0	1.00	N	0	1.00	0.56	30	1.08	0.24	0.11
9	Y	N	Y	3	N	0	1.00	N	0	1.00	0.56	30	1.08	0.13	0.05
10	Y	N	Y	3	N	0	1.00	N	0	1.00	0.56	0	1.00	0.38	0.18
11	Y	N	Y	3	N	0	1.00	N	0	1.00	0.56	0	1.00	0.21	0.10
12	Y	N	N	3	N	0	1.00	N	0	1.00	0.56	0	1.00	0.17	0.08

Table 4.17 Calculation of calibration factors for 3ST intersections

Method of Obtaining Observed Crashes at Intersections	No. of Fatal Crashes	No. of Injury Crashes	Total (Fatal / Injury) Crashes	No. of Property Damage Crashes	Total (Fatal + Injury + Personal Damage Only)	No. of Daytime Crashes	No. of Nighttime Crashes	Nighttime Fatal Crash	Nighttime Injury Crash	Nighttime PDO Crash	No. of Total Nighttime Crashes	Predicted Total Crashes	Predicted Fatal / Injury Crashes
1	0	10	10	43	53	15	38	0	7	31	38	18.44	8.59
2	0	4	4	13	17	8	9	0	1	8	9		

Intersection-box (Method 1),

$$\text{Total Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{53}{18.44} = 2.87$$

$$\text{Fatal and Injury Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{10}{8.59} = 1.16$$

Intersection-related crashes (Method 2),

$$\text{Total Crash, } C_\mathrm{r} = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{17}{18.44} = 0.92$$

$$\text{Fatal and Injury Crash, } C_r = \frac{\text{Total Observed Crashes}}{\text{Total Predicted Crashes}} = \frac{4}{8.59} = 0.47$$

Using observed crashes within an intersection-box (method one), the obtained 0.91 calibration factor for total crashes on rural 4ST intersections indicated precise crash prediction. The HSM underpredicts total crashes on 3ST intersections when considering crashes from method one but showed more precise prediction when considering intersection-related crashes (method two). Fatal and injury crash prediction followed a similar trend for both methods of observed crashes. Results indicated that, using intersection-boxes (method one), the HSM accurately predicts fatal and injury crashes when compared to actual observed crashes on rural 4ST and 3ST intersections.

Chapter 5 - Development of Kansas-Specific New Safety Performance Functions for Rural Four-lane Divided Segments

The objective of this research was to calibrate the HSM for rural multilane highways, including segments and intersections, in Kansas. As discussed in Section 3.3, 4D and 4U segments were calibrated based on the HSM methodologies, and obtained calibration factors indicated that the HSM methodologies underpredict total crashes and overpredict fatal and injury crashes. In addition, the existing SPF given in the HSM was modified, resulting in satisfactory performance for total crash prediction. In order to obtain more reliable crash prediction, this study developed Kansas-specific SPFs and compared them to the HSM calibration and modified SPF method results.

Developing jurisdiction-specific SPFs using data specific to each agency would potentially enhance reliability of the Part C predictive method. The HSM suggests that calibration of the jurisdiction-specific SPF using procedures in Appendix of the HSM may not be necessary within the first two or three years after a jurisdiction-specific SPF is developed, especially if other default values in the HSM Part C models are replaced with locally derived values.

5.1 Model Selection for Kansas-Specific SPFs

The first step in developing a new SPF is to determine which statistical method to use from the multiple statistical methods commonly used to create SPFs according to the literature review. Lord and Mannering (2010) identified promising models to be random-parameter models, finite mixture models, and Poisson and Negative Binomial regression models. The most popular current methods include Poisson regression, ZIP regression, and Negative Binomial regression models. The HSM (AASHTO, 2010) suggests using the Negative Binomial regression procedure because it accounts for overdispersion; however, many studies have successfully used Poisson regression during SPF development. Negative Binomial regression also accurately predicts crashes because it takes into account yearly crash variations and deviation from the normal variance.

Overdispersion occurs when the variance is larger than the sample mean. An overdispersion parameter indicates the statistical reliability of an SPF; a statistically reliable SPF should have an overdispersion parameter close to zero. A Negative Binomial regression model was considered in this study in order to obtain the best-performing model in compliance with the HSM.

5.2 Highway Segments for New SPF Development

Among the 281 4D segments in this study, 200 randomly selected segments were considered for development of a new SPF; the remaining 81 segments were used for model validation. The random selection was performed using a random number generator in Microsoft Excel. All segments varied in length but maintained a minimum length of 0.10 miles.

5.3 New Variables Considered in Kansas-Specific SPFs

SPFs in the HSM incorporate only segment length and AADT of 4D segments. However, the underprediction of total crashes indicated that other variables might be taken into account when predicting crashes for rural multilane segments in Kansas. After evaluating past studies and Kansas-specific attributes, several new variables were identified for consideration in the preliminary stage of SPF development.

Differentiating between correlation and causality is difficult when selecting variables to model crashes. Correlation does not indicate the occurrence by the particular correlated factor. For example, correlation can occur between total crashes on a roadway segment and its length even though the segment length did not cause the crashes. SPFs are simple because they often contain predictive rather than actual causal factors (Lord et al., 2008). Srinivasan and Carter (2011) found that segments within the influence of at-grade intersections and railroad grade crossings (250 ft. on either side of at-grade intersections or railroad grade crossings) significantly affected crash prediction on rural segments. Therefore, all these factors were evaluated in this study during new SPF development. Speed limit, horizontal curve classification, gradient classification, presence of horizontal curve, presence of gradient, roadside hazard rating, medium truck volume, heavy truck volume, presence of rumble strips, and driveway density per mile were potential variables considered in the Kansas-specific SPF development, as listed in Table 5.1. In addition to independent variables given in the HSM, several new variables were considered in the Kansas-specific SPF development.

Table 5.1 Variables in new SPF development

Variable	Data Description	Data Source
L	Segment Length	CANSYS Database
AADT	AADT	
LW	Lane Width	
MW	Median Width	
LSW	Left Shoulder Width	
RSW	Right Shoulder Width	
SS	Side Slope	
SpL	Speed Limit	
PHCur	Presence of Horizontal Curve	
C	Curve Classifications	
G	Presence of Gradient	
PG	Gradient Classifications	
PRs	Presence of Rumble Strips	
HTrc	Heavy Truck Volume	
MTrc	Medium Truck Volume	
TTrc	Total Truck Volume	
RHR	Roadside Hazard Rating	KDOT Videologs
DW	Driveway Density per mile	
PL	Presence of Lighting	Google Map
Crashes	Number of Crashes	KCARS Database
FI Crashes	Number of Fatal and Injury Crashes	

5.3.1 Horizontal Alignment

The CANSYS database provided horizontal curve classifications of roadway segments for this study. KDOT uses the same classification groups as the FHWA (shown in Table 5.2), and roadway segments have uniform alignment within the length. In developing the new SPF, horizontal curve classification initially was a possible variable; for other model variation, however, presence of horizontal curve was considered to be a binomial variable (if present = 1, not present = 0).

Table 5.2 Curve classifications

Curve Classification	Degree of Curvature
A	Under 3.5 degrees (i.e., 0.061 radians)
B	3.5–5.4 degrees (i.e., 0.061–0.094 radians)
C	5.5–8.4 degrees (i.e., 0.096–0.147 radians)
D	8.5–13.9 degrees (i.e., 0.148–0.243 radians)
E	14.0–27.9 degrees (i.e., 0.244–0.487 radians)
F	28 degrees (i.e., 0.489 radians) or more

5.3.2 Vertical Grade

The CANSYS database also provided vertical grades for this study. In developing the new SPF, vertical grade classification initially was a possible variable, but highways in rural Kansas do not contain much grade variation. Therefore, in the later models, presence of vertical grade was considered to be a binomial variable (if present = 1, not present = 0). Table 5.3 lists vertical grade classifications.

Table 5.3 Vertical grade classifications

Grade Classification	Percent Grade
A	0.0–0.4
B	0.5–2.4
C	2.5–4.4
D	4.5–6.4
E	6.5–8.4
F	8.5 or greater

5.3.3 Roadside Hazard Rating

The roadside hazard rating (RHR) is determined based on factors such as side slope, clear zone, and ability of a car to recover if it departed the roadway (Zeeger et al., 1987). Hazard ratings were assigned to each segment by comparing the side slope of the road from the CANSYS database to data from KDOT videologs and Google Street View. Because the topography of Kansas is fairly flat, the RHR for four-lane highways did not vary significantly along segments or among segments; RHR ranged from 1 to 4 (shown in Table 5.4), with 1 being the least hazardous and 4 being extremely hazardous.

Table 5.4 Roadside hazard rating criterion

RHR	Clear Zone Distance	Side slope	Recoverable	Special Features
1	>9 m (30 ft.) from pavement edgeline	flatter than 1:4	Yes	-
2	6 and 7.5 m (20 and 25 ft.) from pavement edge line	approximately 1:4	Marginally Yes	-
3	3 m (10 ft.) from pavement edge line	approximately 1:3 to 1:4	Marginally forgiving	Rough roadside surface
4	1.5 and 3 m (5 and 10 ft.) from pavement edge line	approximately 1:3 or 1:4	Virtually No	May have guardrail, exposed trees, poles, other objects

5.3.4 Speed Limit

Posted speed limit was another variable considered in the development of a new Kansas-specific SPF. Most segments had a posted speed limit of 65 mph, as taken from the CANSYS database. Segments had posted speed limits ranging from 50 to 70 mph.

5.3.5 Driveway Density

Driveway density was determined using aerial photography in Google applications. Driveways onto the highway were counted and considered on a per-mile basis. Field entrances were disregarded because they are not used daily. Few segments had more than five driveways per mile, while many segments did not have any driveways.

Table 5.5 summarizes the mean, maximum, minimum, and standard deviation of data used in development of the new SPF for 4D segments.

Table 5.5 Descriptive statistics of variables

Description	Average	Min	Max	Std. Dev.
Length (mile)	1.53	0.1	8.64	1.54
AADT (vpd)	8,000	490	31,000	4,657
Left lane width (ft)	12.00	10.99	20.99	0.60
Right lane width (ft)	12.00	10.99	20.99	0.60
Left paved shoulder width (ft)	5.67	0	9.84	1.44
Right paved shoulder width (ft)	9.35	0	9.84	1.87
Median width (ft)	30.64	4.92	152.00	15.78
Number of total crashes	9.40	0	56	10.69
Number of fatal and injury crashes	1.79	0	13	2.33
Presence of lighting	0.28	0	1	0.44
Presence of rumble strips	0.70	0	1	0.46
Posted speed limit (mph)	68.44	50	70	5.85
Volume of medium truck (vpd)	480.48	25	930	201.37
Volume of heavy truck (vpd)	124.93	10	360	57.06
Total truck (vpd)	605.18	35	1,150	241.69
Gradient	1.09	0	3	0.98
Presence of gradient	0.63	0	1	0.48
Horizontal curve	1.02	0	5	0.53
Presence of horizontal curve	0.94	0	1	0.24
Roadside hazard rating	1.73	1	4	1.04
No. of driveways per mile	1.04	0	7	2.28

5.4 Correlation Test

Correlation analysis of variables was performed to identify correlation with total observed crashes and total fatal and injury crashes. Variables in the HSM include lane width, shoulder width, median width, total observed crashes, AADT, and length of segments. With the exception of segment length, none of the HSM variables showed strong correlation with total observed crashes at 0.05 level of significance.

Table 5.6 presents results of the correlation study, particularly the correlation of variables to total crashes and total fatal and injury crashes. A positive correlation indicates that as the variable increases, the amount of crashes also increases; a negative correlation indicates that as the variable increases, the number of crashes decreases. A significant correlation indicates a strong relationship between the data. Using a level of significance of 0.05, segment length, AADT, inner shoulder width, posted speed limit, presence of horizontal curve, and gradient class demonstrated statistically significant correlation in both crash categories. The presence of a rumble strip demonstrated significant correlation with total crashes only. Although correlation studies provide insight into the relationship between geometric features and crashes, they do not indicate cause and effect and can potentially be misleading. For example, according to Table 5.6, inner shoulder width has a positive correlation with both types of crashes, indicating that as inner shoulder width increases, the number of crashes increase. However, an increase in shoulder width typically is expected to decrease the number of crashes. Therefore, this relationship could have a confounding factor, thereby negatively affecting the correlation.

Table 5.6 Correlation analysis of variables

Variables	Pearson's Correlation Coefficient (p-value)	
	Total Crashes	Total Fatal and Injury Crashes
Segment Length (mile)	0.71202	0.49684
	(<0.0001)	(<0.0001)
Median Width	0.00447	-0.01558
	(-0.9499)	(-0.8267)
Inner Shoulder Width	0.15335	0.15449
	(0.0302)	(0.0289)
Outer Shoulder Width	0.12061	0.0984
	(0.0889)	(0.1657)
Lane Width	-0.07776	0.00243
	(0.2737)	(0.9728)
Speed Limit	0.38808	0.26553
	(<.0001)	(0.0001)
AADT 2014	0.34422	0.28201
	(<0.0001)	(<0.0001)
Volume of Medium Truck	0.0774	0.04453
	(0.276)	(0.5313)
Volume of Heavy Truck	0.13383	0.06848
	(0.0589)	(0.3353)
Total Truck	0.12882	0.06721
	(0.0691)	(0.3443)
Presence of Horizontal Curve	0.1595	0.14024
	(0.0241)	(0.0476)
Presence of Rumble Strip	0.17244	0.08276
	(0.0146)	(0.244)
Gradient Class	0.1408	0.14038
	(0.0467)	(0.0474)
Presence of Gradient	0.10021	0.09145
	(0.158)	(0.1978)
Presence of Lighting	-0.13094	-0.0653
	(0.0646)	(0.3583)
Driveways per Mile	-0.1287	-0.10798
	(0.0693)	(0.128)
Roadside Hazard Rating	-0.01327	-0.01303
	(0.8521)	(0.8547)

*highlighted variables indicate statistically significant correlation

5.5 New SPFs

Once initial analysis of each variable was complete, the new SPFs were developed. Based on the HSM recommendations, Negative Binomial regression analysis was the model format, and new SPFs were created using SAS.

5.5.1 Total Crashes

The first model for total crashes considered all geometric variables, AADT, segment length, classification of horizontal curve within segments, and classification of vertical gradient. The final model from this iteration was selected using the backward elimination process, including all statistically significant variables, as given in Equation 5.1 and Table 5.7.

$$Predicted\ Crashes/year \quad = e^{[-2.8052\ +\ 0.4849 \times \mathrm{L}\ +\ 0.0001\ \times\ \mathrm{AADT}\ +0.0465\ \times\ \mathrm{SpL}]} \qquad (5.1)$$

Table 5.7 Parameter estimates of model 1 for predicting total crashes

Parameter	**DF**	**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
Intercept	1	-2.8052	0.7943	12.47	0.0004
Segment_Length	1	0.4849	0.0382	161.36	<.0001
Speed_Limit	1	0.0465	0.0121	14.85	0.0001
AADT	1	0.0001	0.0000	73.14	<.0001
Dispersion	1	0.2732	0.0475	-	-

The second model considered the presence of horizontal curves and the presence of vertical gradients within segments instead of their classifications. The final model selected using backward elimination process is given in Equation 5.2 and Table 5.8.

$$Predicted\ Crashes/year = e^{[-3.2541\ +\ 0.4759\ \times \mathrm{L}\ +0.0001 \times \mathrm{AADT}\ +0.4111 \times\ \mathrm{PHCrve}\ +\ 0.0481\ *\ \mathrm{SpL}]} \qquad (5.2)$$

Table 5.8 Parameter estimates of model 2 for predicting total crashes

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-3.254128	0.817401	15.8489	<.0001
Segment_Length	1	0.475918	0.038078	156.2114	<.0001
Speed_Limit	1	0.048098	0.012000	16.0646	<.0001
AADT	1	0.00001	0.00001	67.1577	<.0001
Presence_of_Hor_Curve	1	0.411336	0.198855	4.2788	0.0386
Dispersion	1	0.266896	0.046542	-	-

The third model considered the natural logarithm of segment length, AADT, the presence of horizontal curves, and the presence of vertical gradients within segments. The final model selected using backward elimination process is given in Equation 5.3 and Table 5.9.

$$Predicted\ Crashes/year \quad =$$

$$e^{[-\ 7.6775\ +\ 0.7979 \times \mathrm{Ln(L)} + 0.9259 \times \mathrm{Ln(AADT)} + 0.4479 \times \mathrm{PHCrve} + +\ 0.0169\ \times\ \mathrm{SpL}\ -\ 0.0012\ \times\ \mathrm{MTrc}\ ]} \qquad (5.3)$$

Table 5.9 Parameter estimates of model 3 for predicting total crashes

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-7.6775	0.9155	70.33	<.0001
Ln_length	1	0.7979	0.0452	312.19	<.0001
Speed_Limit	1	0.0169	0.0108	2.45	0.0478
Ln_AADT	1	0.9259	0.0821	127.16	<.0001
Volume_of_Medium_Truck	1	-0.0012	0.0007	3.21	0.0331
Presence_of_Hor_Curve	1	0.3529	0.1837	3.69	0.0447
Dispersion	1	0.1289	0.0305	-	-

The fourth model considered the natural logarithm of segment length, AADT, and total truck volume instead of heavy and medium truck volumes separately. The final model selected using backward elimination process is given in Equation 5.4 and Table 5.10.

$$Predicted\ Crashes/year \quad = e^{[-6.763\ +\ 0.822\times \mathrm{Ln(L)}+0.9259\times \mathrm{Ln(AADT)}+0.4479\times \mathrm{PHCrve}]} \qquad (5.4)$$

Table 5.10 Parameter estimates of model 4 for predicting total crashes

Parameter	**DF**	**Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**
Intercept	1	-6.763654	0.705825	91.8266	<.0001
Ln_length	1	0.822338	0.039485	433.7541	<.0001
Ln_AADT	1	0.925331	0.079098	136.8544	<.0001
Presence_of_Hor_Curve	1	0.447868	0.176145	6.4649	0.0110
Dispersion	1	0.133211	0.031325	-	-

These models were compared using the stepwise selection process, which have similar significant variables with same model coefficients.

5.5.2 Fatal and Injury Crashes

The first model for fatal and injury crashes considered all geometric variables, AADT, segment length, the classification of horizontal curves within segments, and the classification of vertical gradients. The first model from this iteration was selected using backward elimination process, including all statistically significant variables, as given in Equation 5.5 and Table 5.11.

$$Predicted\ Crashes/year = e^{[-5.125\ \ +0.395\times \mathrm{L}\ +0.0001\times \mathrm{AADT}\ +0.190\ \times \mathrm{LW}\ +\ 0.165\ \times \mathrm{RSW}]} \qquad (5.5)$$

Table 5.11 Parameter estimates of model 1 for predicting fatal and injury crashes

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-4.0151	1.2853	9.76	0.0018
Segment_Length	1	0.3979	0.0493	65.23	<.0001
Right_Shoulder_Width	1	0.0687	0.0656	1.10	0.0491
AADT	1	0.0001	0.0000	45.69	<.0001
Lane_Width	1	0.1922	0.0954	4.05	0.0341
Dispersion	1	0.4209	0.1137	-	-

The second model considered the presence of horizontal curves and vertical gradients within segments instead of their classifications. The final model selected using backward elimination process is given in Equation 5.6 and Table 5.12.

$$Predicted\ Crashes/year = e^{[-5.234\ +0.370 \times \mathrm{L}\ +0.0002 \times \mathrm{AADT}\ +0.1176 \times \mathrm{LW}\ +\ 0.028\ \times \mathrm{SpL}]} \tag{5.6}$$

Table 5.12 Parameter estimates of model 2 for predicting fatal and injury crashes

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-5.2338	1.7175	9.29	0.0023
Segment_Length	1	0.3702	0.0532	48.45	<.0001
Speed_Limit	1	0.0275	0.0199	1.92	0.0460
AADT	1	0.0001	0.0000	42.36	<.0001
Lane_Width	1	0.1763	0.0946	3.47	0.0323
Dispersion	1	0.4139	0.1128	-	-

The third model considered the natural logarithm of segment length, AADT, the presence of horizontal curves, and the presence of vertical gradients within segments. The final model selected using backward elimination process is given in Equation 5.7 and Table 5.13.

$$Predicted\ Crashes/year = e^{[-14.3213+0.596\times \mathrm{Ln(L)}+1.320\times \mathrm{Ln(AADT)}+0.259\times\ \mathrm{LW}\ +0.002\times\ \mathrm{MTrc}]} \quad (5.7)$$

Table 5.13 Parameter estimates of model 3 for predicting fatal and injury crashes

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-14.3213	1.8886	57.50	<.0001
Ln_length	1	0.5968	0.0675	78.24	<.0001
Volume_of_Med_Truck	1	0.0017	0.0012	2.08	0.0489
Lane_Width	1	0.2591	0.0868	8.91	0.0028
Ln_AADT	1	1.3205	0.1531	74.41	<.0001
Dispersion	1	0.2617	0.0949	-	-

The fourth model considered the natural logarithm of segment length, AADT, and total truck volume instead of heavy and medium truck volumes separately. The final model selected using backward elimination process is given in Equation 5.8 and Table 5.14.

$$Predicted\ Crashes/year = e^{[-14.264\ +0.585\times Ln(L)+1.297\times Ln(AADT)+0.253\times\ LW\ ]} \quad (5.8)$$

Where,

L= segment length,

LW = lane width,

SpL = speed limit,

$HTrc$ = volume of heavy truck,

MTrc = volume of medium truck,

DW = driveway per mile,

RSW = right shoulder width and

PHCur = presence of horizontal curve.

Table 5.14 Parameter estimates of model 4 for predicting fatal and injury crashes

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-14.2636	1.8990	56.42	<.0001
Ln_length	1	0.5841	0.0673	75.32	<.0001
Lane_Width	1	0.2535	0.0872	8.45	0.0036
Ln_AADT	1	1.2970	0.1530	71.83	<.0001
Dispersion	1	0.2694	0.0969	-	-

5.6 Validation

Once the SPFs were developed, they were validated using a set of roadway segments that differed from segments used to create new SPFs. Statistical tests were run to determine which model was better and could be observed if results match with known guidelines. Bayesian information criterion (BIC), Akaike information criterion (AIC), and corrected Akaike information criterion (AICc) are some of the methods and tests that were performed to obtain best models.

5.6.1 Total Crashes

Table 5.15 compares the goodness of fit for all models developed to predict total crashes, summarizing goodness-of-fit indicators such as log-likelihood, AIC, AICc, and BIC. As shown in the table, model 4 demonstrated an overall better fit than models 1, 2, or 3. The criterion on log-likelihood was not clearly mentioned in most cases, and it alone cannot be used to assess a model. However, a previous study proved that a high log-likelihood is an indication of a better model (Caliendo et al., 2007). AIC, AICc, and BIC indicated smaller values to be the representative of better fit; results indicated that consideration of the natural logarithm of segment length and AADT (Equation 5.4) more accurately explains total crashes on rural multilane highways in Kansas.

Table 5.15 Goodness-of-fit comparison for total crashes model

Criterion	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model Goodness-of-fit criteria**
Deviance/df	1.48	1.29	1.29	1.21	0.8–1.2
Scaled Deviance/df	1.48	1.29	1.29	1.21	0.8–1.2
Pearson Chi-Square/df	2.04	1.05	1.19	1.17	0.8–1.2
Scaled Pearson Chi-Square/df	2.04	1.05	1.19	1.17	0.8–1.2
Log-Likelihood	4984.16	3154.16	3181.53	5233.98	Higher is better
Full Log-Likelihood	-528.78	-545.78	-518.41	5233.98	Higher is better
AIC	1826.56	1125.56	1050.83	1052.13	Smaller is better
AICc	1744.99	1128.92	1051.41	1052.13	Smaller is better
BIC	1592.63	1181.63	1073.92	1068.63	Smaller is better

Each model was run through the validation dataset that consisted of segments not used during development of the new SPF. The number of predicted crashes at each segment, as obtained though the validation process, were plotted against observed crashes. Figures 5.1, 5.2, 5.3, and 5.4 show plots of predicted crashes compared to observed crashes for models 1, 2, 3, and 4, respectively, in order to identify the best model for crash prediction. Ideally, the predicted crashes should be equal or approximately close to the actual observed crashes. The trend line of each plot indicates the plot fit. If predicted crashes and observed crashes are identical, then R^2 is 1.00. Among the four graphs, model 4 demonstrated closest predicted crashes compared to other models; therefore, this model is the best option to predict total crashes on rural four-lane highways in Kansas.

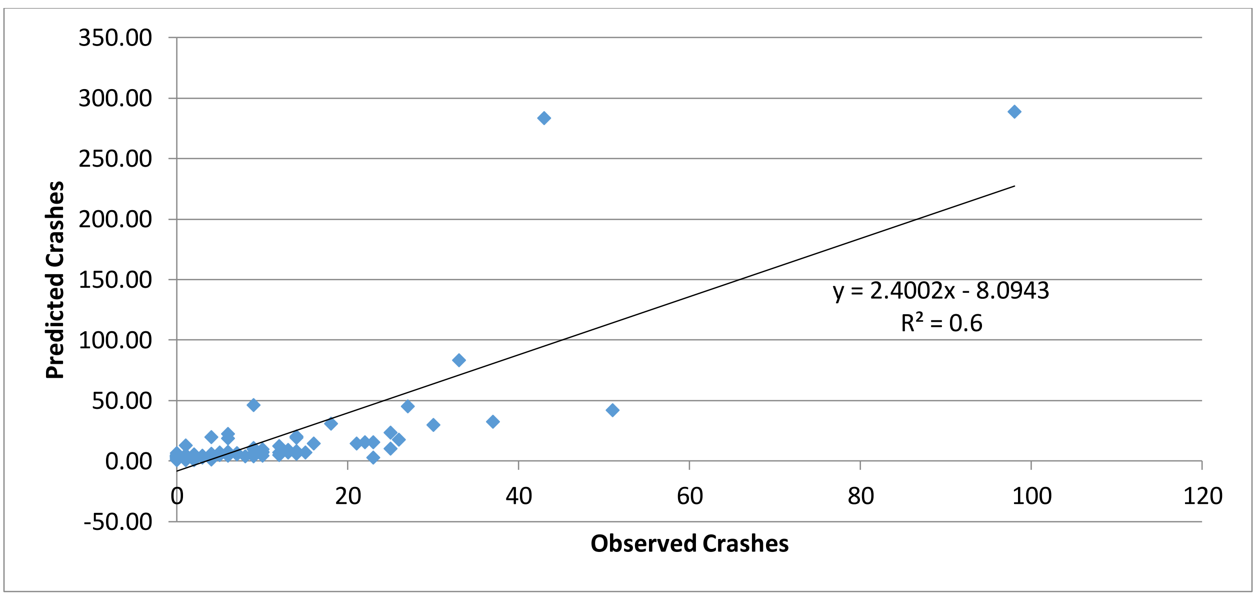


Figure 5.1 Total crashes: model 1 validation plot

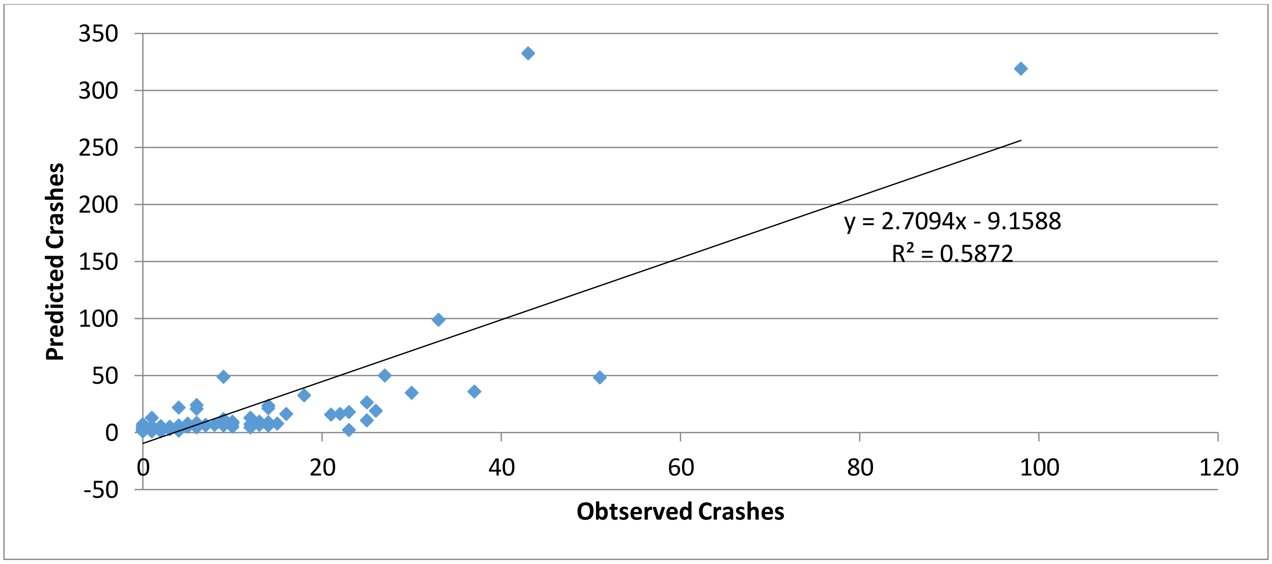


Figure 5.2 Total crashes: model 2 validation plot

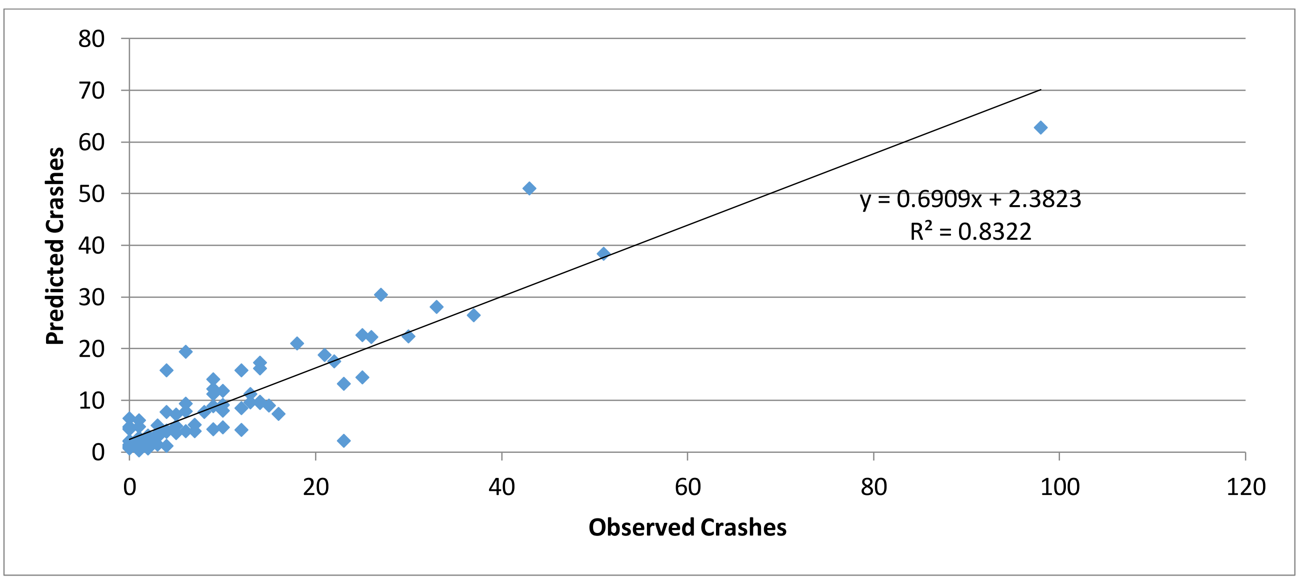


Figure 5.3 Total crashes: model 3 validation plot

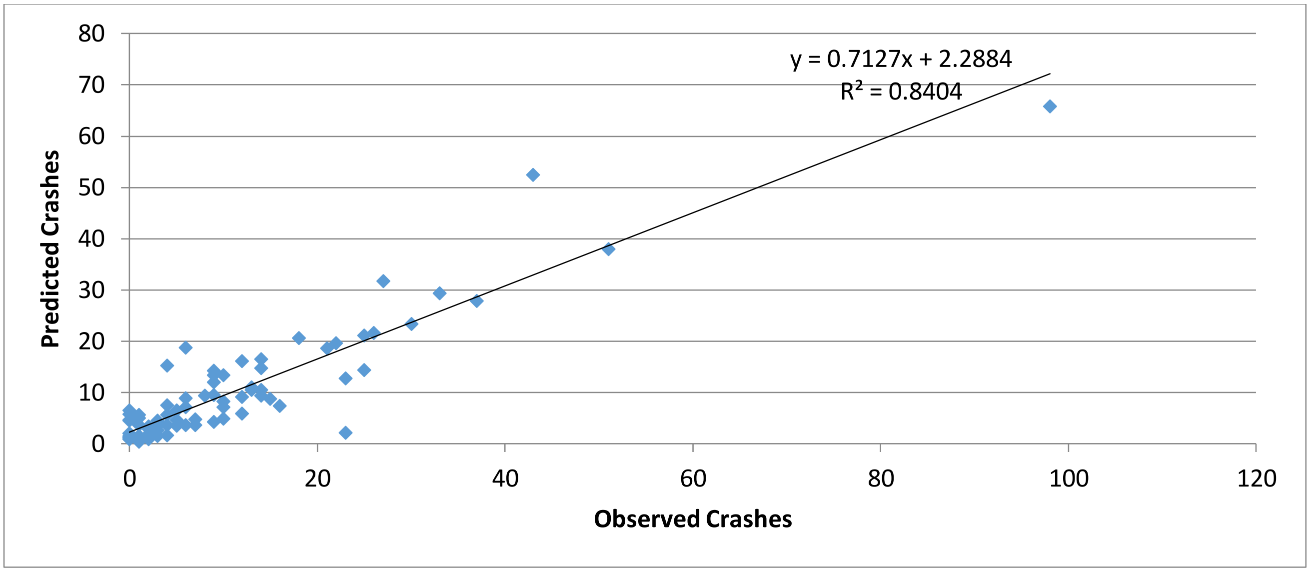


Figure 5.4 Total crashes: model 4 validation plot

5.6.1.1 Outlier Analysis

In order to obtain the best-fitted model, analysis was performed to identify possible outliers or influential data points. Studentized residuals and studentized deleted residuals were used to identify such locations.

Studentized residual

An outlier is a point with a response variable that is far from the implied general regression relationship, thereby requiring a large residual (in absolute value). Studentized residuals (or internally studentized residuals) are defined for each observation, i = 1, 2 ..., n, as an ordinary residual divided by an estimate of its standard deviation.

$$z_i^* = \frac{y_i - \hat{y}_i}{MSE\sqrt{1-h_i}} \tag{5.9}$$

Where,

y_i = Observation I,

$\hat{y}_i$ = Predicted response if observation i is removed from model,

MSE = Mean standard error, and

h_i = Leverage.

Leverage measures the influence of the observation. Any observation with a studentized residual larger than three (in absolute value) is generally deemed an outlier.

Studentized deleted residual

Studentized deleted residual is residual divided by the standard deviation of the residual, or a residual standardized to have a standard deviation of 1. More precisely, the i^{th} standardized residual equals

$$d_i^* = \frac{y_i - \hat{y}_i}{standerd\ error} \quad |d_i^*| > t_{0.001} \tag{5.10}$$

$$df = (n - 1) - (k + 1) \tag{5.11}$$

Where,

n = Sample size, and

k = Number of predictors.

Using outputs of each model, residuals were generated via SAS. Validation segments that showed studentized residual greater than 3 were considered outliers. Segments that showed studentized deleted residual greater than $t_{0.001}$ were also identified as outliers. Figures 5.5, 5.6, 5.7, and 5.8 show plots of predicted total crashes compared to observed crashes after outliers were removed. These figures indicate that removal of outliers improved model fit since R^2 of each plot increased. However, even after outliers were removed, model 4 was still the best model.

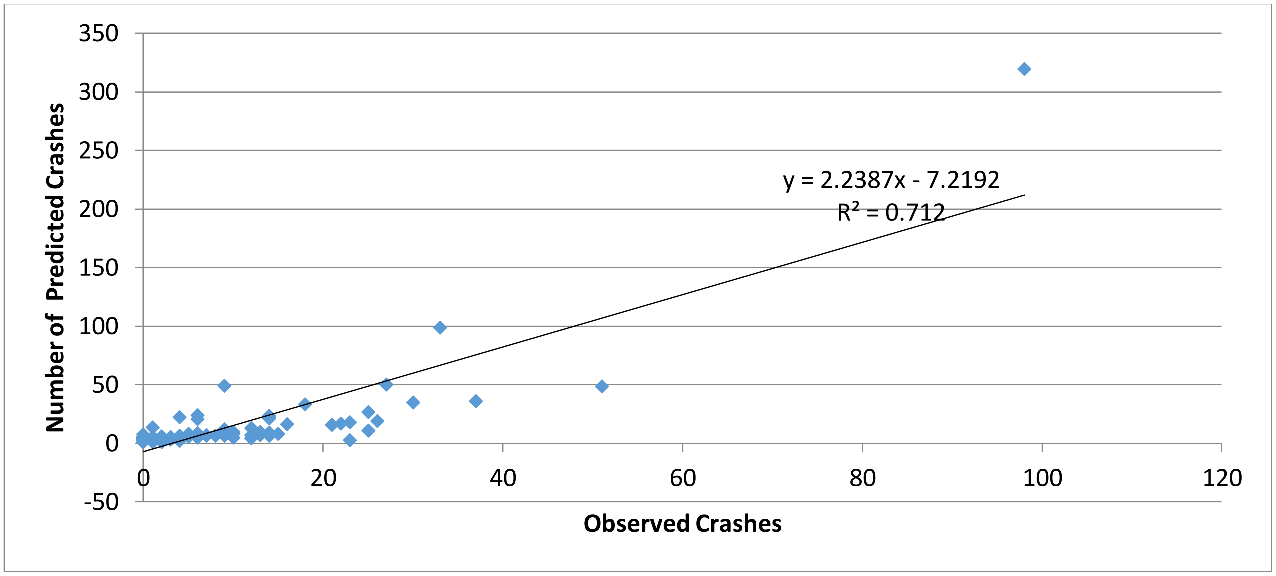


Figure 5.5 Total crashes: model 1 validation plot (without outliers)

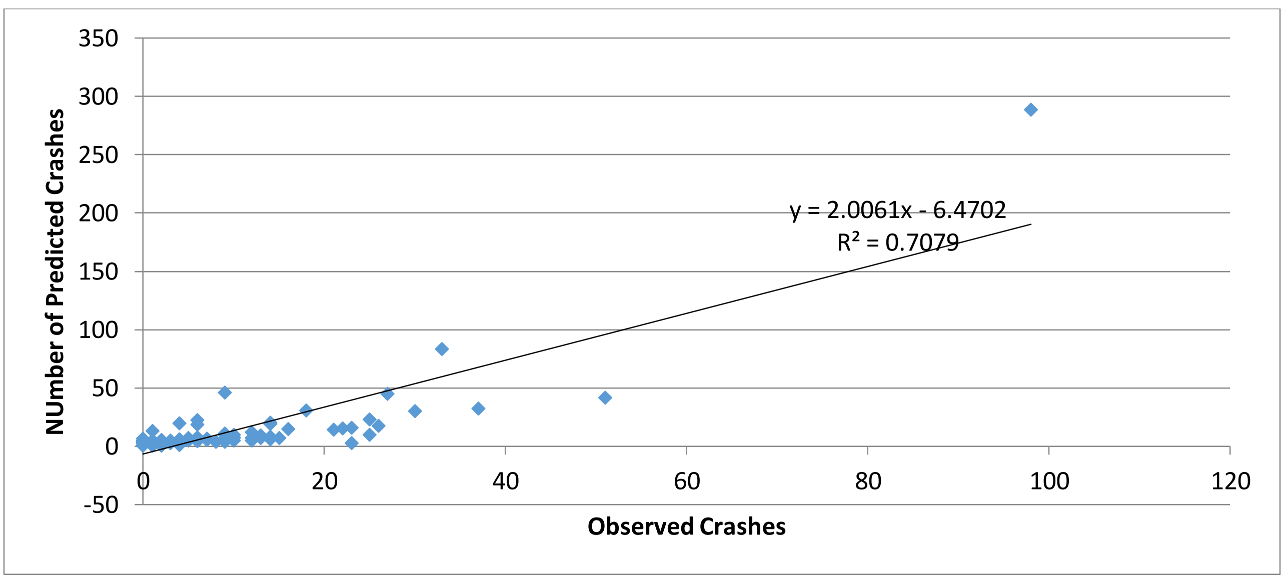


Figure 5.6 Total crashes: model 2 validation plot (without outliers)

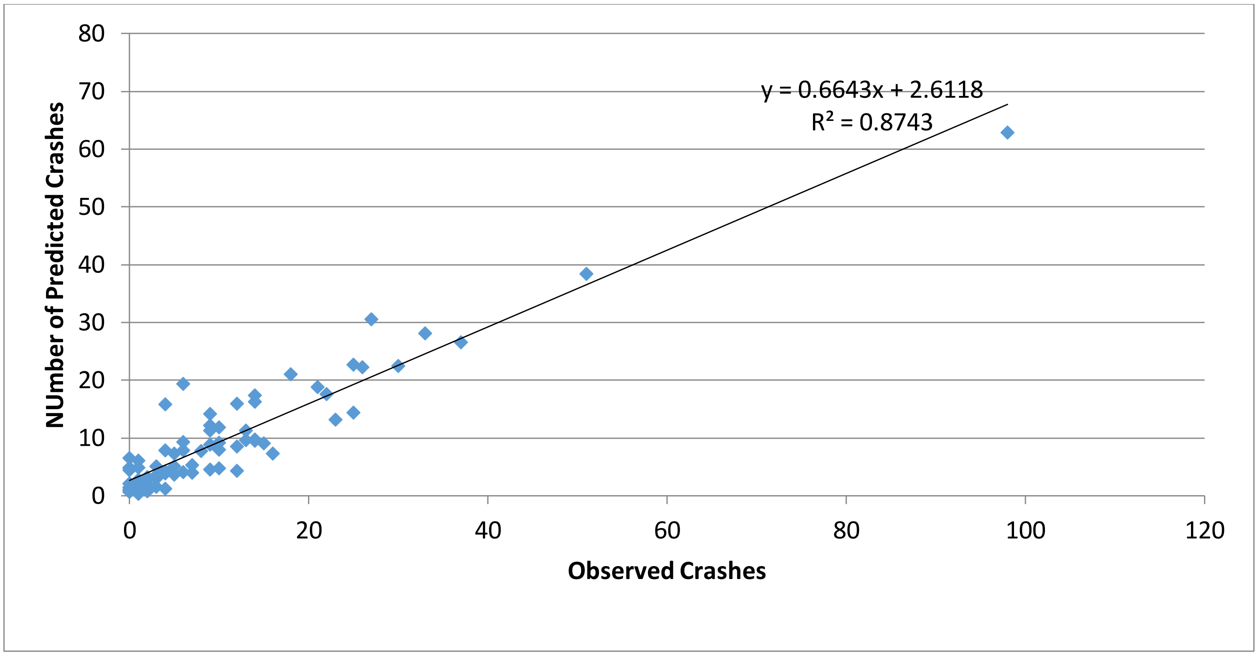


Figure 5.7 Total crashes: model 3 validation plot (without outliers)

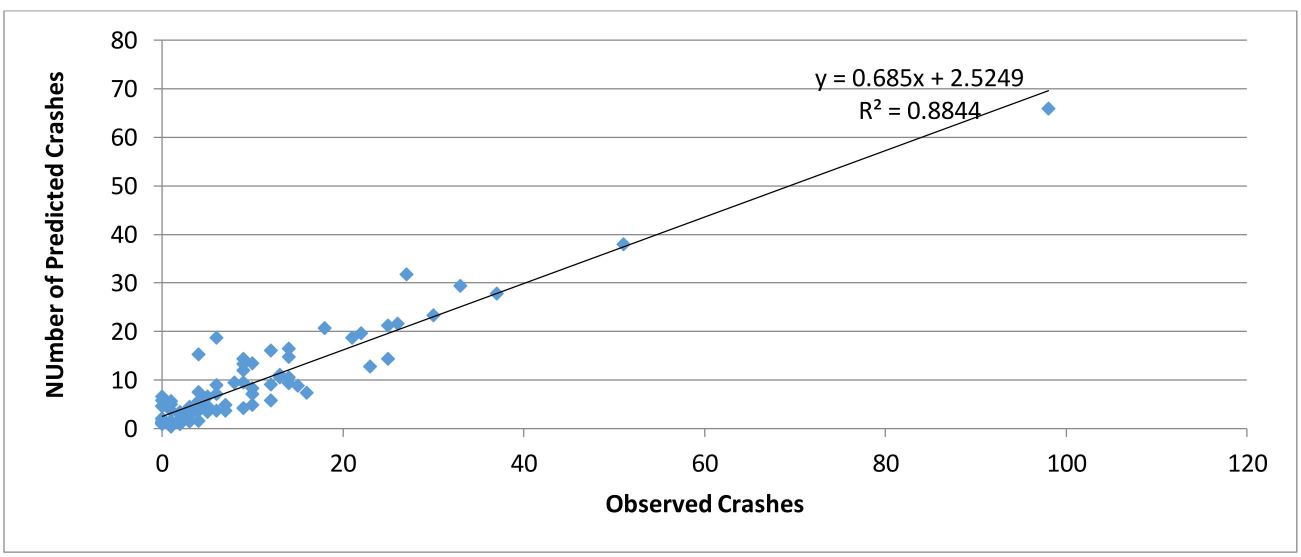


Figure 5.8 **Total crashes: model 4 validation plot (without outliers)**

5.6.2 Fatal and Injury Crashes

Table 5.16 compares goodness of fit for all models developed to predict fatal and injury crashes, summarizing goodness-of-fit indicators such as log-likelihood, AIC, AICc, and BIC. As shown in the table, model 4 had an overall better fit than models 1, 2, and 3 because it had a goodness of fit between 0.8 and 1.2. Although log-likelihood criterion is not clearly mentioned in most cases and it alone cannot be used to assess a model, a previous study (Caliendo et al., 2007) proved that a high log-likelihood indicates a better model. AIC, AICc, and BIC indicate smaller values to be the representative of better fit. Results indicated that consideration of the natural logarithm of segment length and AADT more accurately explains fatal and injury crashes on rural multilane highways in Kansas.

Table 5.16 Goodness-of-fit comparison of fatal and injury crash models

Criterion	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model Goodness-of-fit criteria**
Deviance/df	1.06	1.07	1.11	1.12	0.8–1.2
Scaled Deviance/df	1.06	1.07	1.11	1.12	0.8–1.2
Pearson Chi-Square/df	1.05	1.08	1.09	1.15	0.8–1.2
Scaled Pearson Chi-Square/df	1.05	1.08	1.09	1.15	0.8–1.2
Log-Likelihood	-21.35	-23.59	-13.02	-15.56	Higher is better
Full Log-Likelihood	-314.22	-316.46	-305.88	-308.43	Higher is better
AIC	640.45	644.92	625.77	626.85	Smaller is better
AICc	640.88	645.35	626.36	627.16	Smaller is better
BIC	660.24	664.71	648.86	643.35	Smaller is better

Each model was run through the validation dataset that consisted of segments not used during the development of new SPF. The number of predicted crashes at each segment, as obtained though the validation process, were plotted against observed crashes. Figures 5.9, 5.10, 5.11, and 5.12 show plots of predicted fatal and injury crashes compared to observed crashes corresponding to models 1, 2, 3, and 4, respectively. Results from model 4 demonstrated closest predicted crashes, so this model is the best option to predict fatal and injury crashes on rural four-lane highways in Kansas.

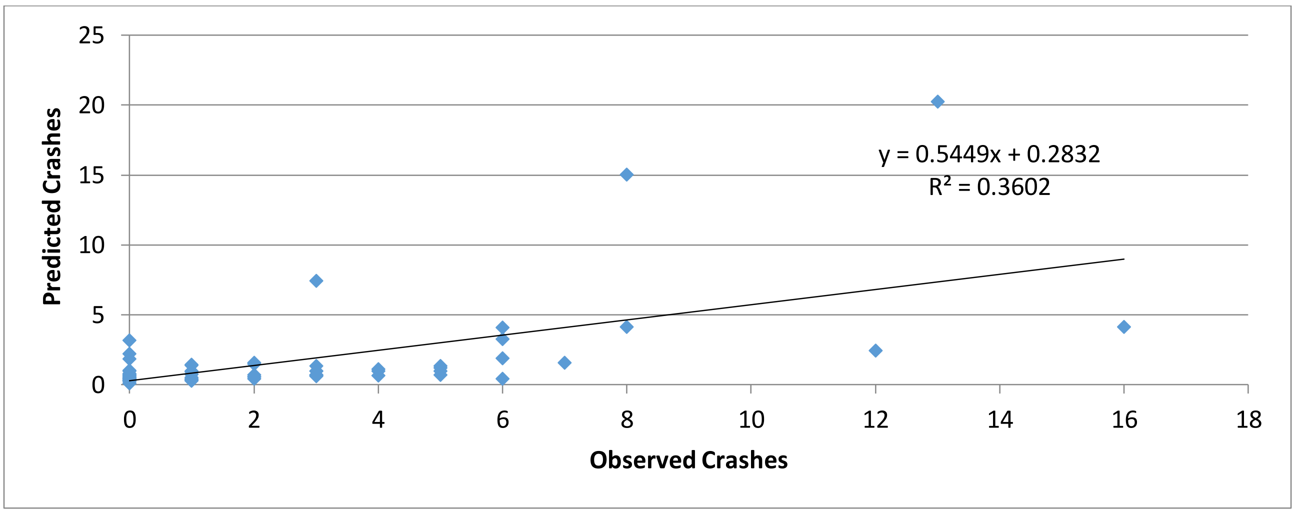


Figure 5.9 Fatal and injury crashes: model 1 validation plot

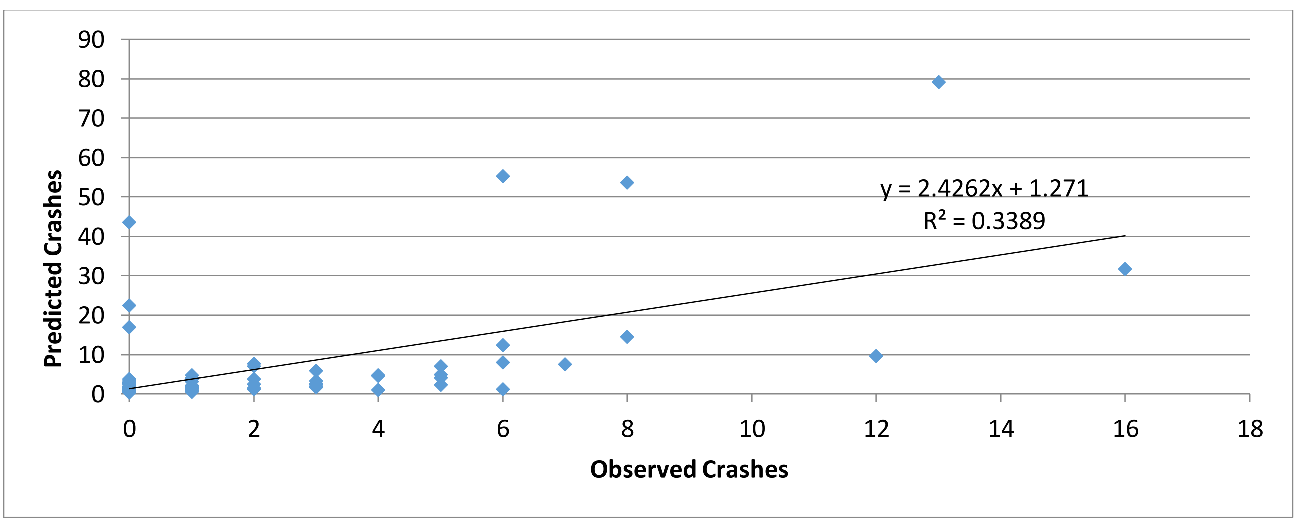


Figure 5.10 Fatal and injury crashes: model 2 validation plot

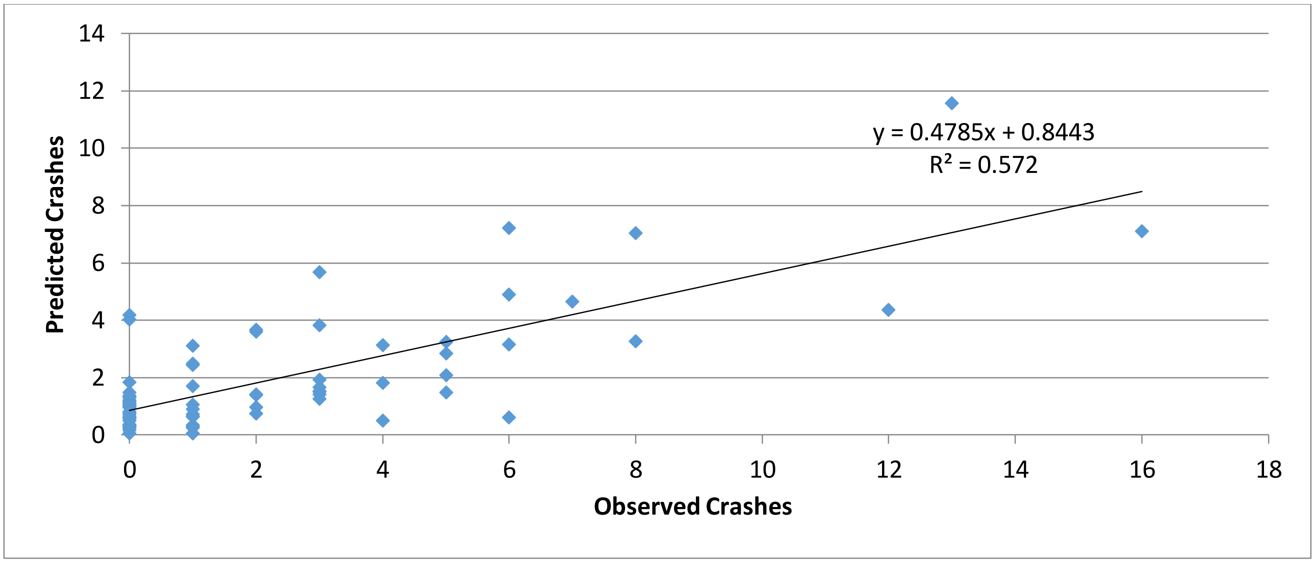


Figure 5.11 Fatal and injury crashes: model 3 validation plot

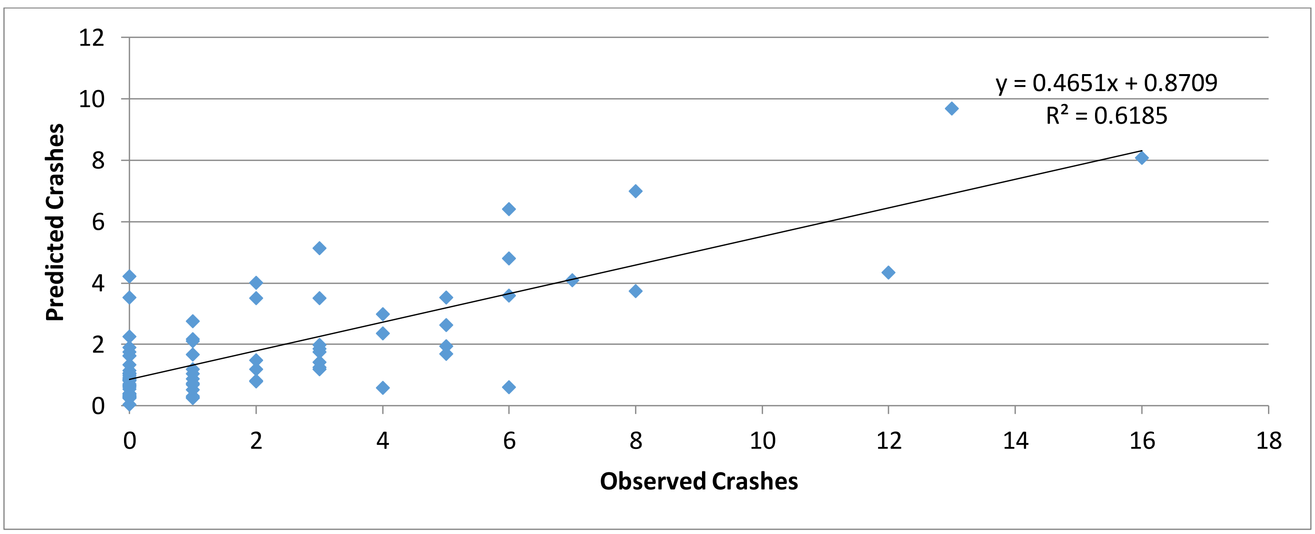


Figure 5.12 Fatal and injury crashes: model 4 validation plot

5.6.2.1 Outlier Analysis

Residuals were generated via SAS using model outputs. Validation segments showing studentized residual greater than 3 were considered outliers. Segments showing studentized deleted residual greater than $t_{0.00}$ were also identified as outliers. Figures 5.13, 5.14, 5.15, and 5.16

show plots of predicted total crashes compared to observed crashes after outliers were removed. As shown in the figures, removing the outliers improved model fit since R^2 of each plot increased. However, even after outliers were removed, model 4 was still the best model.

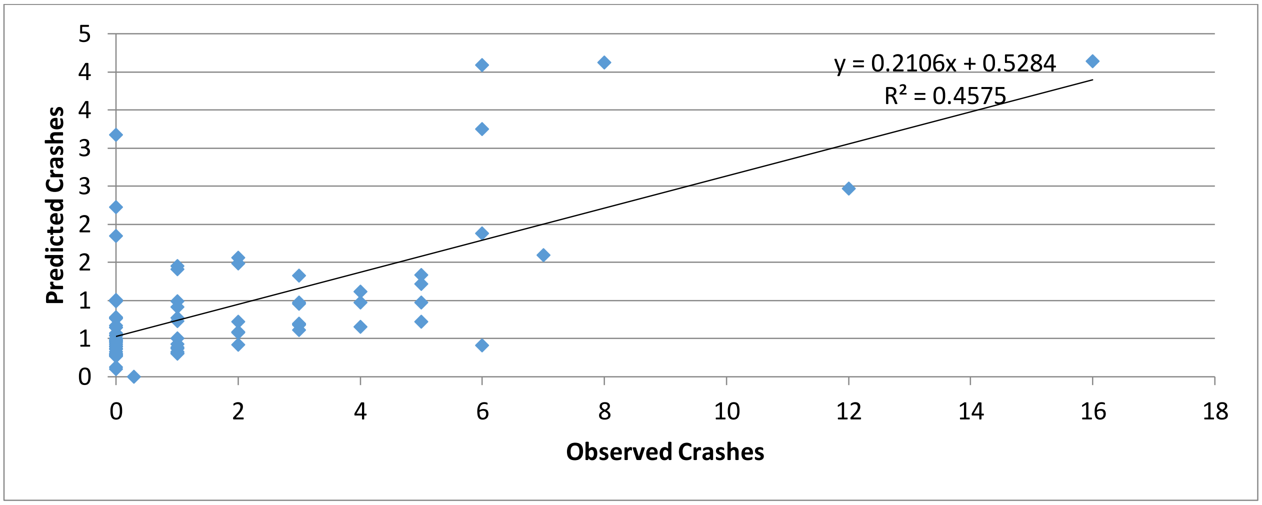


Figure 5.13 Fatal and injury crashes: model 1 validation plot (without outliers)

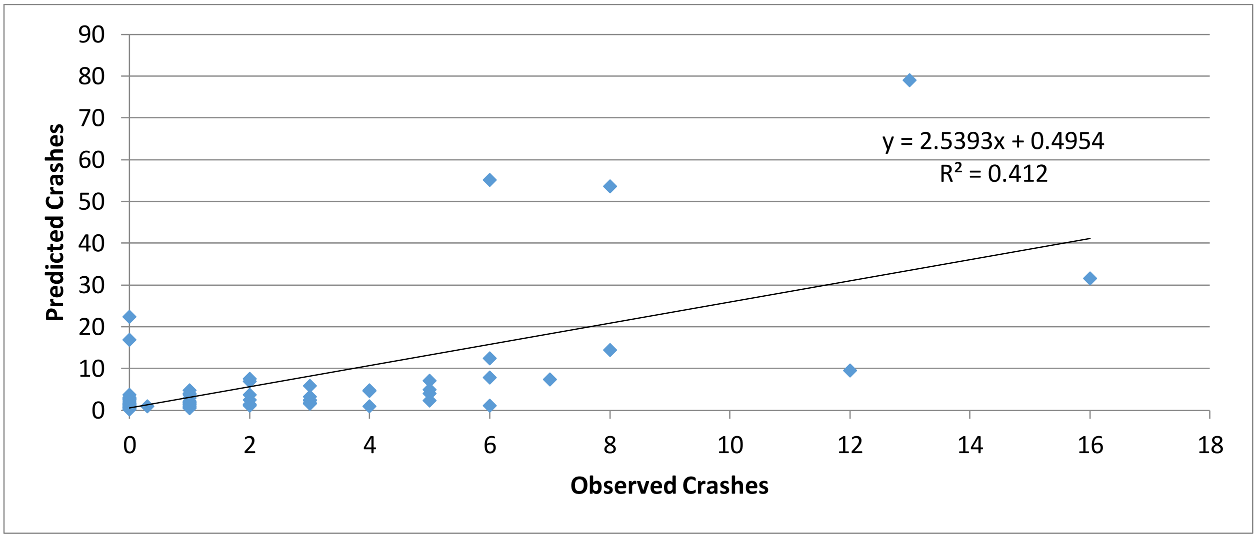


Figure 5.14 Fatal and injury crashes: model 2 validation plot (without outliers)

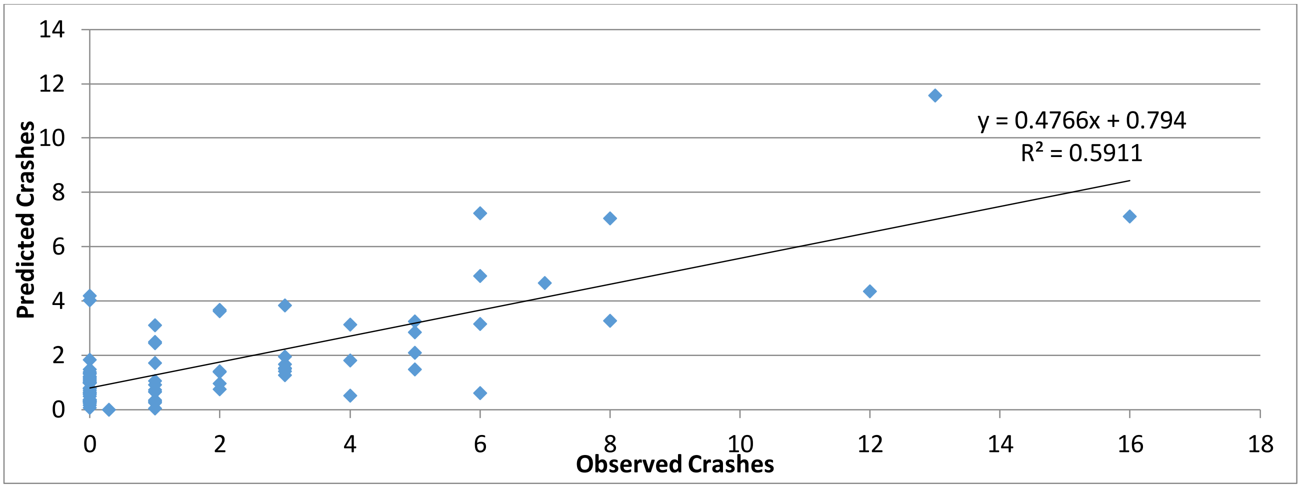


Figure 5.15 Fatal and injury crashes: model 3 validation plot (without outliers)

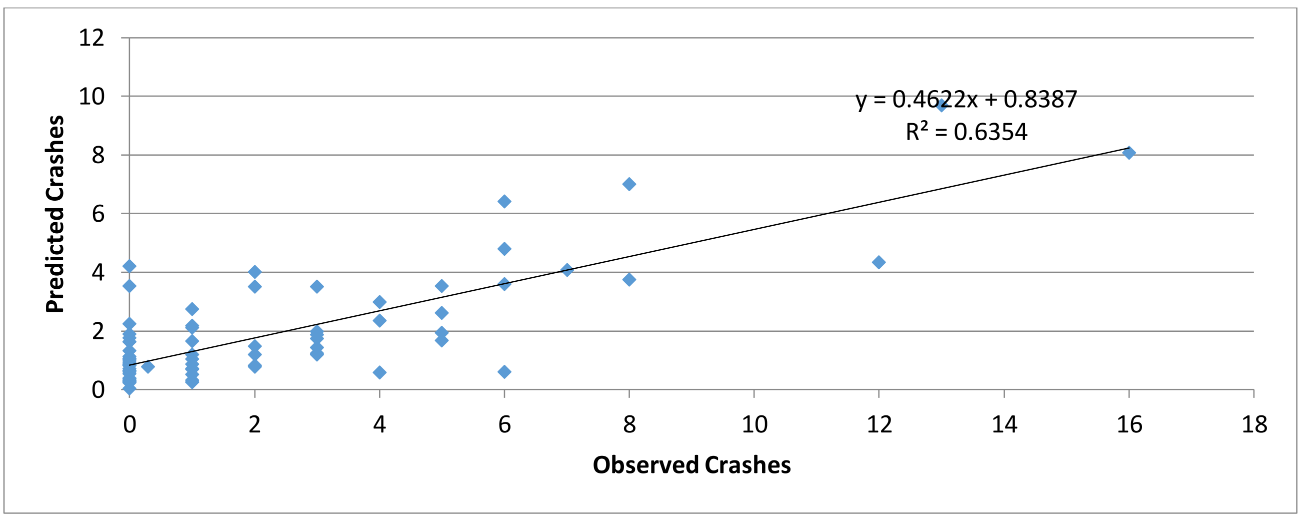


Figure 5.16 Fatal and injury crashes: model 4 validation plot (without outliers)

5.7 Comparison of New SPF to HSM-given SPF

The new Kansas-specific SPFs were compared to the HSM calibration and modified HSM-given SPF for predicted crashes. Errors in prediction compared to observed data were obtained to calculate Mean Prediction Bias (MPB), Mean Absolute Deviation (MAD) and Mean Squared Predicted Error (MSPE).

Table 5.17 compares model statistics in which statistical parameters close to 0 indicate a desirable model and good fit of the data (Garber et al., 2011). Since positive MPB indicates overprediction and negative MPB indicates underprediction, the new SPF showed least underprediction for total crashes, and the modified HSM model showed least underprediction for fatal and injury crashes. The new SPF showed smallest MAD, indicating the best fit for predicting total crashes and fatal and injury crashes. Similarly, smaller MSPE indicates a better fit, and the Kansas-specific SPF showed optimal results. Therefore, the new Kansas-specific SPF for four-lane divided highway segments more accurately predicts total and fatal and injury crashes for rural Kansas.

Table 5.17 Comparison of model statistics

Model	**MPB**	**MAD**	**MSPE**
HSM Total Crashes	-3.43	4.50	53.69
Modified HSM Total Crashes	-0.80	3.94	38.80
New SPF Total Crashes	-0.73	3.89	37.67
HSM Fatal and Injury Crashes	1.67	2.09	13.61
Modified HSM Fatal and Injury Crashes	-0.12	1.65	5.60
New SPF Fatal and Injury Crashes	-0.27	1.40	4.12

Chapter 6 - Summary, Conclusions, and Recommendations

6.1 Summary and Conclusions

The HSM is commonly used to predict crash frequency for highway facilities using SPFs that were developed based on available crash and other data throughout several states. The HSM recommends that models be calibrated based on crash data from the local jurisdiction in order to obtain a more reliable crash prediction. An acceptable method to predict crashes for rural multilane highway segments and intersections in Kansas must be developed, if calibration does not lead to accurate predictions. Prior to this study, the Kansas Department of Transportation (KDOT) could apply the rural two-lane model given in the HSM, but rural multilane highways in Kansas were lacking a reliable crash prediction methodology. KDOT has occasionally requested analysis of a multilane facility, but it could not be be completed without calibration. The objective of this research was to calibrate the HSM for rural multilane highways in Kansas that include 4D and 4U segments and 4ST and 3ST intersections. As discussed in Section 3.3, 4D and 4U segments were calibrated based on the HSM methodologies. Obtained calibration factors indicated that the HSM methodologies underpredict total crashes and overpredict fatal and injury crashes. The corresponding calibration factors can be used for future crash prediction.

Several default regression factors and crash proportions are utilized in the HSM calibration methodology. A comparison of Kansas crash proportions based on severity, daytime/nighttime condition, and collision type revealed significant differences between these proportions and default crash proportions in the HSM. The HSM-given SPF regression coefficients were therefore modified to capture variations in crash predictions, to better suit Kansas conditions. The SPFs with new coefficients were multiplied by CMFs to obtain the predicted crash frequency. The adjusted models for 4D and 4U facilities indicated significant improvement in crash prediction compared to HSM crash prediction for rural Kansas.

Kansas-specific SPFs were developed in this study according to the HSM recommendations. Development of jurisdiction-specific SPFs using individual agency data typically enhanced reliability of Part C predictive methods in the HSM. The HSM suggests, however, that calibration of jurisdiction-specific SPF using procedures in the Appendix A of the HSM may not be necessary within the first two or three years after development, particularly if other default values in the HSM Part C models are replaced with locally derived values.

Analysis results showed two models that would work best for the state of Kansas. One model predicts total crashes better, and the other model predicts fatal and injury crashes better. The model for predicting total crashes includes segment length, AADT, and the presence of horizontal curves. The model for predicting fatal and injury crashes in Kansas, included segment length, AADT, and lane width as significant variables. This model showed smallest BIC, AIC, and AICc and high log-likelihood in the goodness-of-fit tests.

The newly developed SPFs were also compared to the HSM-given SPF and adjusted SPF using statistical parameters Mean Prediction Bias, Mean Absolute Deviation and Mean Squared Prediction Error, leading to the conclusion that the new Kansas-specific SPF for 4D highway segments reliably predicts total and fatal and injury crashes in rural Kansas. This model fits Kansas data better than the HSM-given SPF and modified SPF, thereby enabling prediction closest to actual conditions. However, if geometric data are not readily available, then the modified SPF would be a better alternative because it has fewer data requirements than the other models.

In addition to segments, this study calibrated multilane intersections. The HSM methodology was followed to obtain the number of predicted crashes at 4ST and 3ST intersections.

Observed crashes at intersections were considered using two methods: intersection-boxes and intersection-related crashes. This study found that intersection-box crashes (method one) is predicting the fatal and injury crashes comparatively close to actual observed crashes on rural 4ST and 3ST intersections.

The number of predicted crashes at segments and intersections can be used for several situations such as: comparing facilities under past or future traffic volumes, checking the alternative designs for an existing facility, designing a new facility under future traffic volumes, estimating effectiveness of countermeasures after a period of implementation and estimating effectiveness of a proposed countermeasure on an existing facility prior to implementation.

This research will help private, county, and state agencies identify possible factors that may influence rural crash occurrence and help determine if a countermeasure could reduce rural fatalities. Calibration of the HSM predictive model for multilane facilities will help transportation practitioners reduce the number of fatalities on rural roadways in Kansas. Development of reliable crash prediction methodology will ultimately save lives in Kansas and reduce the number of crashes and fatalities on rural multilane roadways and intersections.

6.2 Recommendations and Future Work

Additional work could further improve the reliability of Kansas-specific crash prediction models, including considering additional CMFs and determining their effect on crash prediction. The HSM suggests that local CMFs be developed if agencies believe that factor has a significant effect to crash frequencies. Because Kansas highways are not geographically similar in all districts or even counties and terrain differences exist throughout the state, development of county-specific and zone-specific (north, south, east, and west) calibration factors for rural multilane segments and intersections may be checked to verify whether separate calibration increases the reliability of crash prediction. The literature review included studies that have considered separate analysis for multiple zones within a single state.

Sample size was the biggest challenge while analyzing rural multilane intersections. In future work sample size should be increased to increase the degrees of freedom and allow consideration of various regression types in order to increase the likelihood of statistically significant explanatory variables. The database of highway intersections should also continue to be expanded until it includes all geometric features of the Kansas highway system. In addition, methodologies described in this dissertation would provide closer crash prediction with a larger sample size.

References

AASHTO. 2010. Highway Safety Manual. Volume 1. American Association of State Transportation Officials. Washington, D.C.

Alluri, P., and Ogle, J. 2012. Effects of state-specific SPFs, AADT Estimations, and Overdispersion Parameters on Crash Predictions Using SafetyAnalyst. In Proceedings from the 2012 TRB 91st Annual Meeting. Washington, D.C.

Alluri, P., Saha, D., Liu, K., and Gan, A. 2014. Improved Processes for Meeting the Data Requirements for Implementing the Highway Safety Manual (HSM) and SafetyAnalyst in Florida.

Banihashemi, M. 2012. Sensitivity Analysis of Data Set Sizes for Highway Safety Manual Calibration Factors. In Transportation Research Record: Journal of the Transportation Research Board, No. 2279, Transportation Research Board of the National Academies, Washington, D.C., pp. 75–81.

Banihashemi, M. 2011. Highway Safety Manual, New Model Parameters vs. Calibration of Crash Prediction Models. 90th Annual Meeting of the Transportation Research Board Compendium (CD-ROM), Washington D.C.

Bornheimer, C. 2011. Finding A New Safety Performance Function for Two-Way, Two-Lane Highways in Rural Areas, Master's Thesis, University of Kansas, Lawrence, KS.

Bornheimer, C., Schrock, S. Wang, M., and Lubliner, H. 2012. Developing a Regional Safety Performance Function for Rural Two-Lane Highways. In Proceedings from the 2012 TRB 91st Annual Meeting. Washington, D.C.

Caliendo, C., Guida, M., and Parisi, A. 2007. A Crash-prediction Model for Multilane Roads. Accident Analysis & Prevention, 39(4), pp. 657–670.

Garber, N. J., and Wu, L. 2001. Stochastic Models Relating Crash Probabilities with Geometric and Corresponding Traffic Characteristics Data (No. UVACTS-5-15-74,). Center for Transportation Studies, University of Virginia, Charlottesville, VA.

Garber, N. J., Rivera, G., and Lim, I. K. 2011. Safety Performance Functions for Intersections in Virginia. Annual Meeting of the Transportation Research Board Compendium (CD-ROM), Washington D.C. (No. 11-0418).

Heron, M. 2016. Deaths: Leading causes for 2010. National Vital Statistics Reports, 65(2), pp. 1–95. http://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_02.pdf. Accessed on March 4, 2016.

Hilbe, J. M. 2011. Negative Binomial Regression, Cambridge University Press.

Jalayer, M., Zhou, H., Williamson, M., and LaMondia, J. J. 2015. Developing Calibration Factors for Crash Prediction Models with Consideration of Crash Recording Threshold Change. In Transportation Research Board 94th Annual Meeting (No. 15-2298).

KDOT. 2005. Motor Vehicle Accident Report Coping Manual. Kansas Department of Transportation. Topeka, KS.

KDOT. 2011. Kansas Data Support- The Data Team. Kansas Department of Transportation. Topeka, KS. http://www.ksdot.org/burTrafficSaf/reports/ksshspdatasupport.pdf Accessed on April 12, 2015.

KDOT. 2012. Kansas Facts Book - Definitions. Kansas Department of Transportation. Topeka, KS. http://www.ksdot.org/Assets/wwwksdotorg/bureaus/burTransPlan/prodinfo/2012factsbook/Definitions.pdf Accessed on April 12, 2015.

KDOT. 2012. KDOT Quick Facts – January 2012. Kansas Department of Transportation. Topeka, KS. http://www.ksdot.org/Assets/wwwksdotorg/pdf_files/2011-quick-facts.pdf Accessed on April 4, 2015.

KDOT. 2013. 2012 Kansas Traffic Accident Facts. Kansas Department of Transportation. Topeka, KS. http://www.ksdot.org/Assets/wwwksdotorg/bureaus/burTransPlan/prodinfo/2012factsbook/RuralUrban.pdf Accessed on April 4, 2015.

KDOT. 2015. Kansas Transportation Quick Facts. Kansas Department of Transportation. Topeka, KS. http://www.ksdot.org/Assets/wwwksdotorg/Headquarters/PDF_Files/QuickFacts.pdf Accessed on March 14, 2016.

KDOT. 2015. Kansas Mileage and Travel by Functional Classification – 2014. Kansas Department of Transportation. Topeka, KS. https://www.ksdot.org/Assets/wwwksdotorg/bureaus/burTransPlan/prodinfo/Mileage_Travel/FunClass2014.pdf Accessed on March 14, 2016.

Kweon, Y. J., Lim, I. K., Turpin, T. L., and Read, S. W. 2014. Development and Application of Guidance to Determine the Best Way to Customize the Highway Safety Manual for Virginia. In Transportation Research Board 93rd Annual Meeting (No. 14-2630).

Long, J. S. 1997. Regression Models for Categorical and Limited-Dependent Variables. Thousand Oaks, California: SAGE Publications, Inc.

Lord, D., and Mannering, F. 2010. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. Transportation Research Part A: Policy and Practice, 44(5), pp. 291–305.

Lord, D., Geedipally, S. R., Persaud, B. N., Washington, S. P., van Schalkwyk, I., Ivan, J. N., Lyon, C., and Jonsson. T. 2008. National Cooperative Highway Research Program

Document 126: Methodology to Predict the Safety Performance of Rural Multilane Highways. (Web Only). NCHRP, Transportation Research Board, Washington, D.C.

Lubliner, H. 2011. Evaluation of the Highway Safety Manual Crash Prediction Model for Rural Two-Lane Highway Segments in Kansas. Ph.D. dissertation, University of Kansas, Lawrence, KS.

Lubliner, H., and Schrock. S. D. 2012. Calibration of the Highway Safety Manual Prediction Method for Rural Kansas Highways. In Transportation Research Board 91st Annual Meeting (No. 12-2835).

Mayora, J. M. P., and Rubio, R. L. 2003. Relevant Variables for Crash Rate Prediction in Spain's Two Lane Rural Roads. In Transportation Research Board 82nd Annual Meeting, pp. 10–18.

Mehta, G., and Lou, Y. 2013. Safety Performance Function Calibration and Development for the State of Alabama: Two-Lane Two-Way Rural Roads and Four-Lane Divided Highways. In Transportation Research Record: Journal of the Transportation Research Board, No. 2398 (1), Transportation Research Board of the National Academies, Washington, D.C., pp. 75–82.

Meyer, E. 2006. Assessing the Effectiveness of Deer Warning Signs. Kansas Department of Transportation, Topeka, KS.

Miaou, S., and Lum. H. 1993. Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. (No. CONF-930120-2). Oak Ridge National Lab., TN.

Najjar, Y., and Mandavilli, S. 2009. Data Mining the Kansas Traffic-Crash Data Base. Report No. K-TRAN: KSU-05-6. Kansas Department of Transportation, Topeka, KS.

NHTSA. 2016. Traffic Safety Fact 2014. National Highway Traffic Safety Administration, Washington, D.C. http://www-nrd.nhtsa.dot.gov/Pubs/812234.pdf Accessed on March 14, 2016.

NHTSA. 2009. Traffic Safety Fact 2009. National Highway Traffic Safety Administration, Washington, D.C. http://www-nrd.nhtsa.dot.gov/Pubs/811402.pdf Accessed on April 4, 2015.

Qin, X., Zhi, C., and Vachal, K. 2014. Calibration of Highway Safety Manual Predictive Methods for Rural Local Roads. In Proceedings of 93rd Annual Meeting of Transportation Research Board, Washington, D.C. (No. 14-1053).

Shin, H., Lee, Y., and Dadvar. S. 2014. The Development of Local Calibration Factors for Implementing the Highway Safety Manual in Maryland. Report No. MD-14-SP209B4J, Maryland State Highway Administration.

Srinivasan, R., and Carter. D. 2011. Development of Safety Performance Functions for North Carolina. Publication FHWA-NC-2010-09. FHWA, U. S. Department of Transportation.

Sun, C., Brown, H., Edara, P., Claros, B., and Nam, K. 2013. Calibration of the Highway Safety Manual for Missouri. Missouri Department of Transportation; Missouri Department of Transportation; Mid-America Transportation Center; Research and Innovative Technology.

Sun, X., Li, Y. Magri, D., and Shirazi, H. 2006. Application of Highway Safety Manual Draft Chapter. In Transportation Research Record: Journal of the Transportation Research Board, No. 1950, Transportation Research Board of the National Academies, Washington D.C., pp. 55–64.

Sun, X., Magri, D., Shirazi, H. H., Gillella, S., and Li. L. 2011. Application of Highway Safety Manual: Louisiana Experience with Rural Multilane Highways. In TRB 90th Annual Meeting Compendium of Papers DVD. Transportation Research Board of the National Academies, Washington, D.C., pp. 1–13.

Xie, F., Gladhill, K. Dixon, K., and Monsere, C. 2011. Calibrating the Highway Safety Manual Predictive Models for Oregon State Highways. In Transportation Research Record: Journal of the Transportation Research Board, No. 2241, Transportation Research Board of the National Academies, Washington, D.C., pp. 19–28.

Zegeer, C. V., Hummer, J., Reinfurt, D., Herf, L., and Hunter, W. 1987. Safety Effects of Cross Section Design for Two Lane Roads-Volume 1: final report, FHWA, McLean, VA.

Appendix A - Calibration Data

Table A.1 List of locations for 4D segment calibration

ID	Route Id	Prefix	Route Dir	Begin County Milepost	End County Milepost	Segment Length (mile)	AADT 2013
1	001U0005400-EB	U	EB	10.357	11.161	0.804	4550
2	006U0006900-NB	U	NB	6.009	6.93	0.921	5390
3	006U0006900-NB	U	NB	6.93	8.097	1.167	6840
4	006U0006900-NB	U	NB	8.097	9.067	0.97	7900
5	006U0006900-NB	U	NB	12.715	13.155	0.44	5660
6	006U0006900-NB	U	NB	13.155	15.235	2.08	5660
7	006U0006900-NB	U	NB	15.235	18.273	3.038	5370
8	006U0006900-NB	U	NB	18.273	22.323	4.05	5090
9	006U0006900-NB	U	NB	22.323	25.356	3.033	4840
10	008K0025400-EB	K	EB	0	2.479	2.479	11300
11	008K0025400-EB	K	EB	2.729	7.957	5.228	11400
12	008K0025400-EB	K	EB	7.957	10.225	2.268	10200
13	008K0025400-EB	K	EB	10.225	10.493	0.268	12000
14	008K0025400-EB	K	EB	10.548	13.157	2.609	12000
15	008K0025400-EB	K	EB	13.157	13.94	0.783	13600
16	008U0005400-EB	U	EB	2.985	6	3.015	17500
17	008U0005400-EB	U	EB	6	8.933	2.933	15200
18	008U0005400-EB	U	EB	10.716	15.085	4.369	6310
19	008U0005400-EB	U	EB	15.085	17.191	2.106	5420
20	008U0005400-EB	U	EB	17.191	17.47	0.279	2370
21	008U0005400-EB	U	EB	17.47	20.41	2.94	2370
22	008U0005400-EB	U	EB	20.41	24.405	3.995	2330
23	008U0005400-EB	U	EB	24.405	25.448	1.043	3290
24	008U0007700-NB	U	NB	34.985	35.757	0.772	2590
25	008U0007700-NB	U	NB	35.757	36.03	0.273	1320
26	011K0006600-EB	K	EB	0.656	0.811	0.155	8040
27	011K0006600-EB	K	EB	0.811	1.247	0.436	8040
28	011K0006600-EB	K	EB	1.247	1.638	0.391	8040
29	011K0006600-EB	K	EB	1.638	2	0.362	8040
30	011K0006600-EB	K	EB	2	3.257	1.257	8400
31	015U0008100-NB	U	NB	21.037	21.164	0.127	5160
32	015U0008100-NB	U	NB	21.164	24.053	2.889	5160
33	015U0008100-SB	U	SB	0	0.489	0.489	5590
34	015U0008100-SB	U	SB	0.489	1	0.511	5590
35	015U0008100-SB	U	SB	1	1.944	0.944	5350
36	015U0008100-SB	U	SB	1.944	4.011	2.067	5350
37	015U0008100-SB	U	SB	4.011	5.085	1.074	4800
38	015U0008100-SB	U	SB	5.085	9.036	3.951	4800
39	015U0008100-SB	U	SB	9.036	12.68	3.644	5000
40	015U0008100-SB	U	SB	12.68	14.168	1.488	5000
41	015U0008100-SB	U	SB	14.168	16.624	2.456	5720
42	015U0008100-SB	U	SB	19.074	21.037	1.963	6460
43	018U0007700-NB	U	NB	0	1.977	1.977	6340
44	018U0007700-NB	U	NB	8.532	8.985	0.453	12000
45	018U0007700-NB	U	NB	8.985	11.587	2.602	10600
46	018U0007700-NB	U	NB	11.587	12	0.413	10600
47	018U0007700-NB	U	NB	12.015	13.053	1.038	11900
48	018U0007700-NB	U	NB	13.053	14.6	1.547	12700
49	018U0007700-NB	U	NB	14.6	14.88	0.28	12700
50	018U0007700-NB	U	NB	14.88	16.535	1.655	12700
51	019U0006900-NB	U	NB	10.698	11.726	1.028	10100
52	019U0006900-NB	U	NB	11.726	12.422	0.696	9870
53	019U0006900-NB	U	NB	12.422	12.618	0.196	9870
54	019U0006900-NB	U	NB	12.618	12.728	0.11	9870
55	019U0006900-NB	U	NB	12.728	12.845	0.117	6480
56	019U0006900-NB	U	NB	12.845	13.047	0.202	6480
57	023K0001000-EB	K	EB	16.153	17.613	1.46	23200
58	023K0001000-EB	K	EB	20.968	21.113	0.145	24000
59	023K0001000-EB	K	EB	21.113	21.476	0.363	24000
60	023U0005900-NB	U	NB	0	3.043	3.043	5310
61	023U0005900-NB	U	NB	3.043	6.543	3.5	7140
62	023U0005900-NB	U	NB	6.543	10.2	3.657	8930
63	028U0005000-EB	U	EB	4.931	5.983	1.052	4940
64	028U0005000-EB	U	EB	5.983	9.864	3.881	7790
65	028U0005000-EB	U	EB	9.864	9.98	0.116	7790
66	028U0005000-EB	U	EB	20.149	20.577	0.428	7020
67	028U0005000-EB	U	EB	20.577	23.149	2.572	7020

ID	Route Id	Prefix	Route Dir	Begin County Milepost	End County Milepost	Segment Length (mile)	AADT 2013
68	028U0005000-EB	U	EB	23.149	25.535	2.386	4620
69	028U0005000-EB	U	EB	25.535	26.823	1.288	4080
70	028U0008300-NB	U	NB	21.419	21.939	0.52	4680
71	030U0005900-NB	U	NB	18.761	21.1	2.339	3280
72	030U0005900-NB	U	NB	21.1	23.3	2.2	2650
73	030U0005900-NB	U	NB	23.3	24.503	1.203	5200
74	030U0005900-NB	U	NB	24.503	26.516	2.013	4820
75	031K0001800-EB	K	EB	15.552	15.659	0.107	12300
76	031K0001800-EB	K	EB	15.659	15.839	0.18	12300
77	031K0001800-EB	K	EB	15.839	18.177	2.338	12300
78	031K0017700-NB	K	NB	13.768	14.016	0.248	7370
79	031K0017700-NB	K	NB	14.016	14.48	0.464	7370
80	040U0005000-EB	U	EB	1.255	1.477	0.222	4280
81	040U0005000-EB	U	EB	1.729	1.943	0.214	4800
82	043U0007500-NB	U	NB	0	2.002	2.002	14700
83	043U0007500-NB	U	NB	2.002	2.991	0.989	12800
84	043U0007500-NB	U	NB	2.991	7	4.009	12200
85	043U0007500-NB	U	NB	7	7.999	0.999	10200
86	043U0007500-NB	U	NB	7.999	16.628	8.629	9750
87	044U0002400-EB	U	EB	2.198	2.4	0.202	6840
88	044U0002400-EB	U	EB	2.4	3.054	0.654	6840
89	044U0002400-EB	U	EB	3.054	4.05	0.996	6670
90	044U0002400-EB	U	EB	4.05	6.516	2.466	5230
91	044U0002400-EB	U	EB	6.516	7.276	0.76	5060
92	046K0001000-EB	K	EB	0	1.006	1.006	24000
93	046K0001000-EB	K	EB	1.006	2.477	1.471	27800
94	046K0001000-EB	K	EB	2.477	3.447	0.97	27900
95	046K0001000-EB	K	EB	7.472	7.862	0.39	31000
96	046U0006900-NB	U	NB	0	1.521	1.521	17400
97	046U0016900-NB	U	NB	0.501	1.005	0.504	16700
98	046U0016900-NB	U	NB	2.19	2.327	0.137	21800
99	046U0016900-NB	U	NB	3.933	4.195	0.262	21800
100	048U0005400-EB	U	EB	0	2.065	2.065	5610
101	048U0005400-EB	U	EB	2.065	5.568	3.503	5540
102	048U0005400-EB	U	EB	5.568	6.203	0.635	5540
103	048U0005400-EB	U	EB	23.259	23.694	0.435	6440
104	048U0005400-EB	U	EB	23.694	26.635	2.941	6320
105	048U0005400-EB	U	EB	26.635	29.671	3.036	6060
106	048U0005400-EB	U	EB	29.671	34.735	5.064	5880
107	048U0005400-EB	U	EB	34.735	36.747	2.012	6130
108	052U0002400-EB	U	EB	11.272	11.663	0.391	11000
109	052U0002400-EB	U	EB	11.663	11.772	0.109	11000
110	052U0002400-EB	U	EB	11.772	11.881	0.109	11000
111	052U0002400-EB	U	EB	11.881	12.39	0.509	11000
112	052U0002400-EB	U	EB	12.39	13.1	0.71	11000
113	052U0002400-EB	U	EB	13.1	14.34	1.24	11000
114	052U0002400-EB	U	EB	14.34	14.626	0.286	11000
115	052U0002400-EB	U	EB	14.727	14.844	0.117	13000
116	052U0002400-EB	U	EB	14.844	15.1	0.256	13000
117	052U0002400-EB	U	EB	15.1	16.39	1.29	13000
118	052U0002400-EB	U	EB	16.39	17.1	0.71	13000
119	052U0002400-EB	U	EB	17.1	17.604	0.504	13000
120	052U0002400-EB	U	EB	17.604	17.713	0.109	13000
121	052U0002400-EB	U	EB	17.713	17.824	0.111	14100
122	052U0002400-EB	U	EB	17.824	17.931	0.107	14100
123	052U0002400-EB	U	EB	17.931	18.234	0.303	14100
124	052U0002400-EB	U	EB	18.234	18.357	0.123	14100
125	052U0002400-EB	U	EB	18.357	19.537	1.18	14100
126	052U0002400-EB	U	EB	19.616	19.718	0.102	14100
127	054U0006900-SB	U	SB	0	2.012	2.012	4840
128	054U0006900-SB	U	SB	2.012	3.703	1.691	4770
129	054U0006900-SB	U	SB	3.703	7.63	3.927	4770
130	054U0006900-SB	U	SB	7.63	10.335	2.705	4940
131	054U0006900-SB	U	SB	10.335	12.826	2.491	5430
132	054U0006900-SB	U	SB	12.826	16.411	3.585	5410
133	054U0006900-SB	U	SB	16.411	19.052	2.641	5300
134	054U0006900-SB	U	SB	19.052	22.295	3.243	5300
135	054U0006900-SB	U	SB	22.295	25.353	3.058	7200
136	055U0004000-EB	U	EB	37.332	38.649	1.317	3710
137	056U0005000-EB	U	EB	5.569	5.923	0.354	5620
138	058U0003600-EB	U	EB	0	1	1	2430
139	058U0003600-EB	U	EB	1	6.724	5.724	3830
140	059K0006100-NB	K	NB	0	0.85	0.85	6870

ID	Route Id	Prefix	Route Dir	Begin County Milepost	End County Milepost	Segment Length (mile)	AADT 2013
141	059K0006100-NB	K	NB	0.85	1.143	0.293	6870
142	059K0006100-NB	K	NB	1.143	1.66	0.517	5590
143	059K0006100-NB	K	NB	1.66	5.38	3.72	5590
144	059K0006100-NB	K	NB	5.38	9.25	3.87	5680
145	059K0006100-NB	K	NB	9.25	9.596	0.346	5950
146	059K0006100-NB	K	NB	9.596	10.145	0.549	5950
147	059K0006100-NB	K	NB	10.145	12.708	2.563	5950
148	059K0006100-NB	K	NB	12.708	14.367	1.659	4130
149	059K0006100-NB	K	NB	14.367	14.483	0.116	4130
150	059K0015300-NB	K	NB	0	0.664	0.664	2010
151	059U00081B1-NB	U	NB	0	2.562	2.562	4410
152	060U0005400-EB	U	EB	18.016	18.586	0.57	4610
153	060U0005400-EB	U	EB	18.586	19.029	0.443	3350
154	061U0006900-NB	U	NB	0	0.98	0.98	7200
155	061U0006900-NB	U	NB	0.98	6.022	5.042	7530
156	061U0006900-NB	U	NB	6.022	9.037	3.015	8740
157	061U0006900-NB	U	NB	9.037	12.128	3.091	9100
158	061U0006900-NB	U	NB	12.128	16.128	4	9410
159	061U0006900-NB	U	NB	16.128	20.25	4.122	13500
160	061U0006900-NB	U	NB	20.25	22.062	1.812	15500
161	061U0006900-NB	U	NB	22.062	23.4	1.338	15500
162	061U0006900-NB	U	NB	23.4	24.402	1.002	17400
163	061U0016900-NB	U	NB	6.451	7.244	0.793	4380
164	061U0016900-NB	U	NB	21.121	23.877	2.756	12400
165	061U0016900-NB	U	NB	23.877	27.441	3.564	12400
166	063U0007500-NB	U	NB	20.664	20.915	0.251	5230
167	063U0007500-NB	U	NB	33.493	35.557	2.064	6140
168	063U0016000-WB	U	WB	26.887	26.992	0.105	6080
169	063U0016000-WB	U	WB	26.992	27.89	0.898	6080
170	063U0016600-EB	U	EB	18.505	19.159	0.654	5920
171	063U0016600-EB	U	EB	19.159	19.352	0.193	5920
172	063U0016600-EB	U	EB	24.597	24.8	0.203	9970
173	063U0016900-NB	U	NB	4.607	5.892	1.285	6800
174	063U0016900-NB	U	NB	5.892	6.437	0.545	5750
175	063U0016900-NB	U	NB	6.437	6.584	0.147	5750
176	063U0016900-NB	U	NB	6.584	6.684	0.1	5750
177	063U0016900-NB	U	NB	7.093	8.849	1.756	5750
178	063U0016900-NB	U	NB	8.849	8.98	0.131	5750
179	063U0016900-NB	U	NB	9	9.139	0.139	4870
180	063U0016900-NB	U	NB	17.068	18.053	0.985	6550
181	063U0016900-SB	U	SB	6.684	6.834	0.15	5750
182	063U0016900-SB	U	SB	6.834	7.093	0.259	5750
183	063U0016900-SB	U	SB	9.139	9.309	0.17	4870
184	063U0040000-EB	U	EB	2.064	2.689	0.625	3020
185	070K0003100-NB	K	NB	32.077	32.328	0.251	490
186	070U0007500-NB	U	NB	24.57	25.082	0.512	6790
187	070U0007500-NB	U	NB	25.082	27.354	2.272	10300
188	070U0007500-NB	U	NB	27.444	27.591	0.147	10300
189	070U0007500-NB	U	NB	27.591	31.11	3.519	10500
190	072U0008100-NB	U	NB	0	4.037	4.037	7960
191	072U0008100-NB	U	NB	4.037	10.234	6.197	6990
192	072U0008100-NB	U	NB	10.234	11.434	1.2	6990
193	072U0008100-NB	U	NB	22.485	24.28	1.795	5590
194	072U0008100-NB	U	NB	24.28	24.494	0.214	5590
195	072U0008100-SB	U	SB	11.434	12.127	0.693	6310
196	072U0008100-SB	U	SB	12.127	12.458	0.331	6310
197	072U0008100-SB	U	SB	12.458	17.904	5.446	5900
198	072U0008100-SB	U	SB	17.904	18.449	0.545	5900
199	072U0008100-SB	U	SB	18.449	19.664	1.215	5220
200	072U0008100-SB	U	SB	19.664	19.967	0.303	5220
201	072U0008100-SB	U	SB	19.967	22.485	2.518	5220
202	075U0002400-WB	U	WB	3.327	3.565	0.238	12600
203	075U0002400-WB	U	WB	3.565	4.253	0.688	12600
204	075U0002400-WB	U	WB	4.253	12.77	8.517	12600
205	076K0006100-NB	K	NB	1.065	1.192	0.127	4530
206	076U0005400-EB	U	EB	26.372	28.287	1.915	5470
207	076U0005400-EB	U	EB	28.287	30.309	2.022	5610
208	078K0001400-WB	K	WB	16.656	17.143	0.487	8280
209	078K0001400-WB	K	WB	17.143	18.381	1.238	8280
210	078K0001400-WB	K	WB	18.381	19.13	0.749	9040
211	078K0001400-WB	K	WB	19.13	19.239	0.109	9040
212	078K0006100-NB	K	NB	26.075	26.767	0.692	2830
213	078K0006100-NB	K	NB	41.974	42.6	0.626	7360

ID	Route Id	Prefix	Route Dir	Begin County Milepost	End County Milepost	Segment Length (mile)	AADT 2013
214	078K0006100-NB	K	NB	42.6	45.5	2.9	7110
215	078K0006100-NB	K	NB	45.5	47.922	2.422	6870
216	078K0009600-EB	K	EB	23.784	27.704	3.92	5760
217	078K0009600-EB	K	EB	27.704	28.62	0.916	5370
218	078K0009600-EB	K	EB	28.62	33.462	4.842	9450
219	078K0009600-EB	K	EB	33.462	37.642	4.18	9340
220	078K0009600-EB	K	EB	37.642	38.684	1.042	10600
221	078U0005000-WB	U	WB	24.288	24.499	0.211	5980
222	078U0005000-WB	U	WB	24.499	28.499	4	5980
223	078U0005000-WB	U	WB	31.533	32.504	0.971	8300
224	078U0005000-WB	U	WB	33.541	35.084	1.543	4120
225	078U0005000-WB	U	WB	35.084	35.561	0.477	4140
226	079U0008100-NB	U	NB	0	0.911	0.911	5160
227	079U0008100-NB	U	NB	0.911	2.984	2.073	5160
228	079U0008100-NB	U	NB	2.984	9.088	6.104	4890
229	079U0008100-NB	U	NB	9.088	10.162	1.074	4890
230	079U0008100-NB	U	NB	10.162	10.736	0.574	4900
231	079U0008100-NB	U	NB	10.74	10.956	0.216	4900
232	079U0008100-NB	U	NB	10.956	11.135	0.179	4900
233	079U0008100-NB	U	NB	11.135	11.442	0.307	5070
234	079U0008100-NB	U	NB	11.442	11.564	0.122	5070
235	079U0008100-NB	U	NB	12.018	12.141	0.123	5390
236	079U0008100-NB	U	NB	12.143	12.355	0.212	5390
237	079U0008100-NB	U	NB	12.355	13.033	0.678	3800
238	079U0008100-NB	U	NB	13.033	13.293	0.26	3800
239	079U0008100-NB	U	NB	13.293	13.605	0.312	3800
240	079U0008100-SB	U	SB	13.605	13.733	0.128	3620
241	079U0008100-SB	U	SB	13.733	14.37	0.637	3620
242	079U0008100-SB	U	SB	14.37	14.711	0.341	3620
243	079U0008100-SB	U	SB	14.711	16.932	2.221	3620
244	079U0008100-SB	U	SB	16.932	17.458	0.526	3620
245	079U0008100-SB	U	SB	17.458	19.564	2.106	3620
246	079U0008100-SB	U	SB	19.564	21.152	1.588	3620
247	079U0008100-SB	U	SB	21.152	24.141	2.989	3620
248	079U0008100-SB	U	SB	24.141	24.654	0.513	3620
249	081K0001800-EB	K	EB	0	0.671	0.671	12300
250	081K0017700-NB	K	NB	0	4.969	4.969	7370
251	082K0001800-EB	K	EB	21.403	21.714	0.311	870
252	084U0028100-NB	U	NB	11.382	11.622	0.24	2260
253	085K0014000-EB	K	EB	16.594	16.769	0.175	3030
254	085U0008100-NB	U	NB	18.797	22.548	3.751	8500
255	085U0008100-NB	U	NB	22.548	24.62	2.072	7960
256	087K0009600-EB	K	EB	0	1.139	1.139	10600
257	087K0009600-EB	K	EB	1.139	2.045	0.906	10600
258	087K0009600-EB	K	EB	2.296	2.541	0.245	10100
259	087K0009600-EB	K	EB	2.547	10.813	8.266	10100
260	087K0009600-EB	K	EB	10.813	11.841	1.028	10900
261	087K0009600-EB	K	EB	11.841	14.588	2.747	12200
262	087K0025400-EB	K	EB	8.295	10.319	2.024	12200
263	087U0005400-EB	U	EB	0	0.98	0.98	6130
264	087U0005400-EB	U	EB	0.98	1.48	0.5	6470
265	087U0005400-EB	U	EB	1.48	4.013	2.533	7770
266	087U0005400-EB	U	EB	4.013	7.031	3.018	7820
267	087U0005400-EB	U	EB	7.031	9.1	2.069	9650
268	087U0005400-EB	U	EB	9.1	9.3	0.2	11100
269	087U0005400-EB	U	EB	9.3	10.1	0.8	11100
270	087U0005400-EB	U	EB	10.1	11.07	0.97	11900
271	088U0005400-EB	U	EB	0	2.741	2.741	6190
272	088U0005400-EB	U	EB	2.741	3.04	0.299	6190
273	088U0005400-EB	U	EB	3.04	3.34	0.3	6190
274	088U0005400-EB	U	EB	3.34	3.635	0.295	6310
275	089U0007500-NB	U	NB	0	2.256	2.256	10500
276	089U0007500-SB	U	SB	2.256	2.46	0.204	10500
277	089U0007500-SB	U	SB	23.846	27.85	4.004	14700
278	101U0003600-EB	U	EB	26.445	27.534	1.089	2330
279	101U0003600-EB	U	EB	27.534	30.525	2.991	2430
280	103U0007500-NB	U	NB	0	1.967	1.967	5830
281	103U0040000-EB	U	EB	22.389	22.748	0.359	3740

Table A.2 List of locations for 4U segment calibration

ID	Route Id	Prefix	Route Dir	Begin County Milepost	End County Milepost	Segment Length (mile)	AADT 2013
1	058U0003600-EB	U	EB	7.287	7.422	0.135	4120
2	071U0002400-EB	U	EB	31.383	31.524	0.141	2890
3	100U0004000-EB	U	EB	4.494	4.64	0.146	705
4	095U0005600-EB	U	EB	13.148	13.248	0.1	2300
5	095U0005600-EB	U	EB	13.248	13.348	0.1	2300
6	071U0002400-EB	U	EB	30.7	30.811	0.111	1390
7	071U0002400-EB	U	EB	30.811	30.911	0.1	1390
8	071U0002400-EB	U	EB	30.468	30.592	0.124	1390
9	071U0002400-EB	U	EB	30.592	30.7	0.108	1390
10	001U0005900-NB	U	NB	12.156	12.406	0.25	1460
11	100U0004000-EB	U	EB	4.163	4.394	0.231	520
12	100U0004000-EB	U	EB	4.394	4.494	0.1	520
13	093U0028100-NB	U	NB	12.073	12.226	0.153	2510
14	093U0028100-NB	U	NB	12.226	12.426	0.2	2510
15	071U0002400-EB	U	EB	30.107	30.468	0.361	1390
16	005U0005600-EB	U	EB	25.439	25.711	0.272	3510
17	005U0005600-EB	U	EB	25.711	25.911	0.2	3510
18	029U0005400-EB	U	EB	18.114	18.214	0.1	3300
19	063U0016600-EB	U	EB	24.495	24.597	0.102	10600
20	001U0005400-EB	U	EB	10.004	10.109	0.105	4410
21	009U0005000-EB	U	EB	20.64	20.752	0.112	4380
22	082K0001800-EB	K	EB	21.287	21.403	0.116	875
23	006U0006900-NB	U	NB	9.067	9.2	0.133	9410
24	031K0001800-EB	K	EB	15.417	15.552	0.135	12400
25	001U0005400-EB	U	EB	12.059	12.194	0.135	3040
26	097U0002400-EB	U	EB	21.657	21.809	0.152	2860
27	079U0003600-EB	U	EB	16.127	16.323	0.196	3500
28	032U0004000-EB	U	EB	0	0.216	0.216	3450
29	011U0040000-EB	U	EB	32.201	32.327	0.126	5440
30	011U0040000-EB	U	EB	32.327	32.447	0.12	5440
31	001U0005400-EB	U	EB	10.109	10.357	0.248	4330
32	001U0005400-EB	U	EB	11.161	11.295	0.134	4330
33	001U0005400-EB	U	EB	11.295	11.415	0.12	4330
34	075U0002400-EB	U	EB	12.8	12.912	0.112	12700
35	075U0002400-EB	U	EB	12.912	13.054	0.142	12700
36	063U0016600-EB	U	EB	24.231	24.495	0.264	10600
37	058U0003600-EB	U	EB	6.998	7.287	0.289	4120
38	005U0028100-NB	U	NB	6.07	6.245	0.175	6730
39	005U0028100-NB	U	NB	6.245	6.365	0.12	6730
40	075U0002400-EB	U	EB	13.054	13.268	0.214	12100
41	075U0002400-EB	U	EB	13.268	13.398	0.13	12100
42	001U0005400-EB	U	EB	8.649	9.046	0.397	7270
43	097U0002400-EB	U	EB	21.241	21.657	0.416	2860
44	001U0005400-EB	U	EB	8.171	8.349	0.178	7270
45	001U0005400-EB	U	EB	8.349	8.649	0.3	7270
46	001U0005400-EB	U	EB	7.666	7.798	0.132	7270
47	001U0005400-EB	U	EB	7.798	7.945	0.147	7270
48	001U0005400-EB	U	EB	7.945	8.171	0.226	7270
49	087K0004200-EB	K	EB	14.936	15.04	0.104	7070
50	008U0007700-NB	U	NB	34.196	34.384	0.188	2620
51	008U0007700-NB	U	NB	34.384	34.584	0.2	2620
52	008U0007700-NB	U	NB	34.584	34.721	0.137	2620
53	008U0007700-NB	U	NB	34.721	34.842	0.121	2620
54	008U0007700-NB	U	NB	34.842	34.985	0.143	2620
55	058U0003600-EB	U	EB	6.724	6.825	0.101	3870
56	055U0004000-EB	U	EB	37.205	37.332	0.127	3750
57	055U0004000-EB	U	EB	38.649	38.8	0.151	3750
58	011K0006600-EB	K	EB	3.257	3.424	0.167	8200
59	093U0028100-NB	U	NB	11.886	12.073	0.187	2510
60	075U0002400-EB	U	EB	14.27	14.47	0.2	8120
61	037U0005400-EB	U	EB	12.305	12.509	0.204	2930
62	011U0040000-EB	U	EB	31.433	31.647	0.214	5440
63	002U0016900-NB	U	NB	18.652	18.873	0.221	2700
64	057K0001500-NB	K	NB	27.471	27.696	0.225	1220
65	066U0003600-EB	U	EB	3	3.249	0.249	3470
66	066U0003600-EB	U	EB	2.748	3	0.252	2870
67	028U0005000-EB	U	EB	19.882	20.149	0.267	7810
68	001U0005400-EB	U	EB	12.29	12.56	0.27	3040
69	047U0005000-EB	U	EB	14	14.159	0.159	2430
70	047U0005000-EB	U	EB	14.159	14.286	0.127	2430
71	055U0004000-EB	U	EB	36.894	37.048	0.154	3750

ID	Route Id	Prefix	Route Dir	Begin County Milepost	End County Milepost	Segment Length (mile)	AADT 2013
72	055U0004000-EB	U	EB	37.048	37.182	0.134	3750
73	047U0005000-EB	U	EB	13.665	14	0.335	1940
74	075U0002400-EB	U	EB	14.47	14.649	0.179	4900
75	075U0002400-EB	U	EB	14.649	14.784	0.135	4900
76	075U0002400-EB	U	EB	14.784	14.949	0.165	4900
77	056U0005000-EB	U	EB	4.892	5.109	0.217	5680
78	056U0005000-EB	U	EB	5.109	5.369	0.26	5680
79	056U0005000-EB	U	EB	5.369	5.569	0.2	5680
80	028U0005000-EB	U	EB	19.133	19.482	0.349	7810
81	028U0005000-EB	U	EB	19.482	19.679	0.197	7810
82	028U0005000-EB	U	EB	19.679	19.782	0.103	7810
83	028U0005000-EB	U	EB	19.782	19.882	0.1	7810

Table A.3 List of locations for 4ST intersections calibration

Intersection ID	Section ID	County Name	Highway No.	Begin County MP	End County MP	Direction (EB/WB/NB/SB)	Number of all crashes within ints. box of 300 ft	Number of intersection related crashes (only)
1	67	FINNEY	US 50	20.577	23.149	EB	4	0
2	67	FINNEY	US 50	20.577	23.149	EB	2	0
3	67	FINNEY	US 50	20.577	23.149	EB	3	0
4	115	LEAVENWORTH	US 24	11.772	11.881	EB	0	0
5	211	PRATT	K 14	26.372	28.287	EB	0	0
6	233	REPUBLIC	US 81	11.135	11.442	NB	1	0
7	270	SEDGWICK	US 54	1.48	4.013	EB	1	0
8	1	ALLEN	US 54	10.357	11.161	EB	1	0
9	2	BOURBON	US 69	6.009	6.93	NB	0	0
10	3	BOURBON	US 69	6.93	8.097	NB	0	0
11	4	BOURBON	US 69	8.097	9.067	NB	1	0
12	10	BUTLER	K 254	0	2.479	EB	1	1
13	10	BUTLER	K 254	0	2.479	EB	5	3
14	11	BUTLER	K 254	2.729	7.957	EB	3	1
15	11	BUTLER	K 254	2.729	7.957	EB	2	1
16	11	BUTLER	K 254	2.729	7.957	EB	1	0
17	11	BUTLER	K 254	2.729	7.957	EB	4	3
18	11	BUTLER	K 254	2.729	7.957	EB	4	2
19	12	BUTLER	K 254	7.957	10.225	EB	2	2
20	12	BUTLER	K 254	7.957	10.225	EB	2	1
21	13	BUTLER	K 254	10.225	10.493	EB	3	2
22	14	BUTLER	K 254	10.548	13.157	EB	2	0
23	16	BUTLER	US 400	2.985	6	EB	0	0
24	16	BUTLER	US 400	2.985	6	EB	2	0
25	16	BUTLER	US 400	2.985	6	EB	2	1
26	17	BUTLER	US 400	6	8.933	EB	11	5
27	17	BUTLER	US 400	6	8.933	EB	2	0
28	20	BUTLER	US 400	17.191	17.47	EB	0	0
29	21	BUTLER	US 400	17.47	20.41	EB	2	2
30	21	BUTLER	US 400	17.47	20.41	EB	1	1
31	21	BUTLER	US 400	17.47	20.41	EB	0	0
32	22	BUTLER	US 400	20.41	24.405	EB	0	0
33	22	BUTLER	US 400	20.41	24.405	EB	1	1
34	22	BUTLER	US 400	20.41	24.405	EB	0	0
35	22	BUTLER	US 400	20.41	24.405	EB	1	1
36	23	BUTLER	US 400	24.405	25.448	EB	1	0
37	24	BUTLER	US 77	34.985	35.757	NB	2	1
38	25	BUTLER	US 77	35.757	36.03	NB	0	0
39	27	CHEROKEE	K 66	0.811	1.247	EB	0	0
40	29	CHEROKEE	K 66	1.638	2	EB	1	0
41	30	CHEROKEE	K 66	2	3.257	EB	3	2
42	31	CLOUD	US 81	21.037	21.164	NB	1	0
43	32	CLOUD	US 81	21.164	24.053	NB	1	0
44	32	CLOUD	US 81	21.164	24.053	NB	2	1
45	32	CLOUD	US 81	21.164	24.053	NB	0	0
46	33	CLOUD	US 81	0	0.489	SB	1	0
47	35	CLOUD	US 81	1	1.944	SB	1	0
48	36	CLOUD	US 81	1.944	4.011	SB	0	0
49	37	CLOUD	US 81	4.011	5.085	SB	0	0
50	38	CLOUD	US 81	5.085	9.036	SB	0	0

Intersection ID	Section ID	County Name	Highway No.	Begin County MP	End County MP	Direction (EB/WB/NB/SB)	Number of all crashes within ints. box of 300 ft	Number of intersection related crashes (only)
51	38	CLOUD	US 81	5.085	9.036	SB	0	0
52	38	CLOUD	US 81	5.085	9.036	SB	1	0
53	39	CLOUD	US 81	9.036	12.68	SB	0	0
54	39	CLOUD	US 81	9.036	12.68	SB	1	0
55	41	CLOUD	US 81	14.168	16.624	SB	0	0
56	41	CLOUD	US 81	14.168	16.624	SB	1	0
57	42	CLOUD	US 81	19.074	21.037	SB	1	0
58	43	COWLEY	US 77	0	1.977	NB	2	2
59	43	COWLEY	US 77	0	1.977	NB	0	0
60	45	COWLEY	US 77	8.985	11.587	NB	1	0
61	45	COWLEY	US 77	8.985	11.587	NB	1	1
62	45	COWLEY	US 77	8.985	11.587	NB	1	0
63	47	COWLEY	US 77	12.015	13.053	NB	0	0
64	48	COWLEY	US 77	13.053	14.6	NB	2	2
65	49	COWLEY	US 77	14.6	14.88	NB	2	0
66	50	COWLEY	US 77	14.88	16.535	NB	0	0
67	51	COWLEY	US 77	10.698	11.726	NB	1	1
68	52	CRAWFORD	US 69	11.726	12.422	NB	0	0
69	91	JACKSON	US 75	7	7.999	NB	2	1
70	91	JACKSON	US 75	7.999	16.628	NB	2	2
71	91	JACKSON	US 75	7.999	16.628	NB	1	0
72	91	JACKSON	US 75	3.054	4.05	EB	0	0
73	91	JACKSON	US 75	3.054	4.05	EB	1	0
74	91	JACKSON	US 75	4.05	6.516	EB	2	2
75	97	JACKSON	US 24	6.516	7.276	EB	2	2
76	150	MARSHALL	US 36	1	6.724	EB	2	0
77	150	MARSHALL	US 36	1	6.724	EB	1	0
78	150	MARSHALL	US 36	1	6.724	EB	0	0
79	203	POTTAWATOMIE	US 24	3.565	4.253	WB	1	1
80	204	POTTAWATOMIE	US 24	4.253	12.77	WB	7	1
81	204	POTTAWATOMIE	US 24	4.253	12.77	WB	7	1
82	204	POTTAWATOMIE	US 24	4.253	12.77	WB	1	1
83	204	POTTAWATOMIE	US 24	4.253	12.77	WB	0	0
84	204	POTTAWATOMIE	US 24	4.253	12.77	WB	0	0
85	204	POTTAWATOMIE	US 24	4.253	12.77	WB	0	0
86	204	POTTAWATOMIE	US 24	4.253	12.77	WB	1	0
87	204	POTTAWATOMIE	US 24	4.253	12.77	WB	1	0
88	204	POTTAWATOMIE	US 24	4.253	12.77	WB	0	0
89	204	POTTAWATOMIE	US 24	4.253	12.77	WB	2	0
90	204	POTTAWATOMIE	US 24	4.253	12.77	WB	0	0
91	204	POTTAWATOMIE	US 24	4.253	12.77	WB	1	0
92	204	POTTAWATOMIE	US 24	4.253	12.77	WB	1	0
93	205	PRATT	K 61	1.065	1.192	NB	0	0
94	208	RENO	K 14	16.656	17.143	WB	1	0
95	209	RENO	K 14	17.143	18.381	WB	2	1
96	210	RENO	K 14	18.381	19.13	WB	0	0
97	216	RENO	K 96	23.784	27.704	EB	1	0
98	216	RENO	K 96	23.784	27.704	EB	1	0
99	216	RENO	K 96	23.784	27.704	EB	0	0
100	218	RENO	K 96	28.62	33.462	EB	2	1
101	218	RENO	K 96	28.62	33.462	EB	0	0
102	218	RENO	K 96	28.62	33.462	EB	0	0
103	219	RENO	K 96	33.462	37.642	EB	2	2
104	219	RENO	K 96	33.462	37.642	EB	2	2
105	219	RENO	K 96	33.462	37.642	EB	0	0
106	220	RENO	K 96	37.642	38.684	EB	0	0
107	222	RENO	US 50	24.499	28.499	WB	2	0
108	222	RENO	US 50	24.499	28.499	WB	1	0
109	222	RENO	US 50	24.499	28.499	WB	2	2
110	226	REPUBLIC	US 81	0	0.911	NB	0	0
111	227	REPUBLIC	US 81	0.911	2.984	NB	1	1
112	228	REPUBLIC	US 81	2.984	9.088	NB	2	2
113	228	REPUBLIC	US 81	2.984	9.088	NB	1	1
114	228	REPUBLIC	US 81	2.984	9.088	NB	1	1
115	228	REPUBLIC	US 81	2.984	9.088	NB	0	0
116	228	REPUBLIC	US 81	2.984	9.088	NB	0	0
117	228	REPUBLIC	US 81	2.984	9.088	NB	0	0
118	229	REPUBLIC	US 81	9.088	10.162	NB	0	0
119	230	REPUBLIC	US 81	10.162	10.736	NB	1	1
120	233	REPUBLIC	US 81	11.135	11.442	NB	3	3
121	234	REPUBLIC	US 81	11.442	11.564	NB	0	0

Intersection ID	Section ID	County Name	Highway No.	Begin County MP	End County MP	Direction (EB/WB/NB/SB)	Number of all crashes within ints. box of 300 ft	Number of intersection related crashes (only)
122	235	REPUBLIC	US 81	12.018	12.141	NB	1	1
123	237	REPUBLIC	US 81	12.355	13.033	NB	0	0
124	237	REPUBLIC	US 81	12.355	13.033	NB	0	0
125	238	REPUBLIC	US 81	13.033	13.293	NB	0	0
126	242	REPUBLIC	US 81	14.37	14.711	SB	1	1
127	243	REPUBLIC	US 81	14.711	16.932	SB	1	0
128	243	REPUBLIC	US 81	14.711	16.932	SB	0	0
129	245	REPUBLIC	US 81	17.458	19.564	SB	0	0
130	245	REPUBLIC	US 81	17.458	19.564	SB	0	0
131	246	REPUBLIC	US 81	19.564	21.152	SB	0	0
132	246	REPUBLIC	US 81	19.564	21.152	SB	0	0
133	247	REPUBLIC	US 81	21.152	24.141	SB	0	0
134	247	REPUBLIC	US 81	21.152	24.141	SB	0	0
135	247	REPUBLIC	US 81	21.152	24.141	SB	0	0
136	248	REPUBLIC	US 81	24.141	24.654	SB	0	0
137	251	ROOKS	K 18	21.403	21.714	EB	0	0
138	252	RUSSELL	US 281	11.382	11.622	NB	3	3
139	256	SEDGWICK	K 96	0	1.139	EB	3	2
140	257	SEDGWICK	K 96	1.139	2.045	EB	3	3
141	259	SEDGWICK	K 96	2.547	10.813	EB	1	1
142	259	SEDGWICK	K 96	2.547	10.813	EB	2	2
143	259	SEDGWICK	K 96	2.547	10.813	EB	0	0
144	259	SEDGWICK	K 96	2.547	10.813	EB	0	0
145	259	SEDGWICK	K 96	2.547	10.813	EB	1	0
146	259	SEDGWICK	K 96	2.547	10.813	EB	0	0
147	260	SEDGWICK	K 96	10.813	11.841	EB	1	0
148	261	SEDGWICK	K 96	11.841	14.588	EB	1	0
149	262	SEDGWICK	K 254	8.295	10.319	EB	1	0
150	262	SEDGWICK	K 254	8.295	10.319	EB	2	1
151	268	SEDGWICK	US 54	9.1	9.3	EB	5	3
152	272	SEWARD	US 54	2.741	3.04	EB	1	1
153	277	SHAWNEE	US 75	23.846	27.85	SB	1	0
154	277	SHAWNEE	US 75	23.846	27.85	SB	3	1
155	277	SHAWNEE	US 75	23.846	27.85	SB	2	2
156	277	SHAWNEE	US 75	23.846	27.85	SB	1	0
157	277	SHAWNEE	US 75	23.846	27.85	SB	2	1
158	277	SHAWNEE	US 75	23.846	27.85	SB	0	0
159	278	WASHINGTON	US 36	26.445	27.534	EB	2	0
160	279	WASHINGTON	US 36	27.534	30.525	EB	2	2
161	279	WASHINGTON	US 36	27.534	30.525	EB	1	1
162	279	WASHINGTON	US 36	27.534	30.525	EB	0	0
163	279	WASHINGTON	US 36	27.534	30.525	EB	1	0
164	280	WILSON	US 75	0	1.967	NB	2	1
165	280	WILSON	US 75	0	1.967	NB	0	0
166	280	WILSON	US 75	0	1.967	NB	2	1
167	1	ALLEN	US 54	7.666	8.171	EB	0	0
168	2	ALLEN	US 54	8.171	8.649	EB	0	0
169	4	ALLEN	US 54	10.004	10.109	EB	0	0
170	6	ALLEN	US 54	11.161	11.415	EB	0	0
171	10	ANDERSON	US 169	18.652	18.873	NB	0	0
172	11	BARTON	US 56	25.439	25.911	EB	1	1
173	14	BARTON	US 281	17.344	17.588	NB	3	2
174	16	BROWN	US 73	20.797	20.943	NB	3	2
175	17	BROWN	US 73	22.234	22.517	NB	0	0
176	23	CHEROKEE	US 400	31.433	31.647	EB	1	1
177	24	CHEROKEE	US 400	32.201	32.447	EB	0	0
178	25	CHEROKEE	US 400	32.201	32.447	EB	0	0
179	26	CHEYENNE	US 36	14.029	14.245	EB	1	0
180	29	FINNEY	US 50	19.882	20.149	EB	0	0
181	31	GEARY	K 18	15.417	15.552	EB	3	1
182	34	GRAHAM	US 24	17.525	18.178	EB	1	1
183	38	HASKELL	US 56	4.982	5.162	EB	0	0
184	39	JACKSON	US 75	16.628	16.832	NB	2	2
185	46	LYON	US 50	4.892	5.569	EB	4	1
186	56	OSAGE	US 56	22.825	23.015	EB	1	1
187	56	OSAGE	US 56	22.825	23.015	EB	2	2
188	57	OSBORNE	US 24	30.107	30.468	EB	1	1
189	61	OSBORNE	US 24	31.187	31.374	EB	3	2
190	62	OSBORNE	US 24	31.383	31.524	EB	1	1
191	64	POTTAWATOMIE	US 24	12.8	13.054	EB	2	0
192	66	POTTAWATOMIE	US 24	14.47	14.949	EB	3	2

Intersection ID	Section ID	County Name	Highway No.	Begin County MP	End County MP	Direction (EB/WB/NB/SB)	Number of all crashes within ints. box of 300 ft	Number of intersection related crashes (only)
193	68	REPUBLIC	US 36	16.127	16.323	EB	0	0
194	73	ROOKS	US 24	28.009	28.153	EB	0	0
195	77	STAFFORD	US 281	12.073	12.426	NB	0	0
196	78	STEVENS	US 56	13.148	13.348	EB	0	0
197	80	THOMAS	US 24	21.657	21.809	EB	0	0
198	81	WALLACE	US 40	4.163	4.494	EB	0	0
199	82	WALLACE	US 40	4.494	4.64	EB	0	0

Table A.4 List of locations for 3ST intersections calibration

Intersection ID	Section ID	County Name	Highway No.	Begin County MP	End County MP	Direction (EB/WB/NB/SB)	Number of all crashes within ints. box of 300 ft	Number of intersection related crashes (only)
1	2	ALLEN	US 54	8.171	8.649	EB	0	0
2	2	ALLEN	US 54	8.171	8.649	EB	0	0
3	3	ALLEN	US 54	8.649	9.046	EB	0	0
4	3	ALLEN	US 54	8.649	9.046	EB	0	0
5	5	ALLEN	US 54	10.109	10.357	EB	0	0
6	5	ALLEN	US 54	10.109	10.357	EB	2	0
7	7	ALLEN	US 54	12.059	12.194	EB	1	0
8	13	BARTON	US 281	17.059	17.344	NB	0	0
9	13	BARTON	US 281	17.059	17.344	NB	2	1
10	14	BARTON	US 281	17.344	17.588	NB	1	0
11	15	BOURBON	US 69	9.067	9.2	NB	3	0
12	18	BUTLER	US 77	34.196	34.584	NB	2	1
13	24	CHEROKEE	US 400	32.201	32.447	EB	0	0
14	27	DECATUR	US 83	18.045	18.307	NB	1	1
15	32	GOVE	US 40	0	0.216	EB	0	0
16	33	GRAHAM	US 24	16.458	16.77	EB	0	0
17	33	GRAHAM	US 24	16.458	16.77	EB	0	0
18	33	GRAHAM	US 24	16.458	16.77	EB	0	0
19	33	GRAHAM	US 24	16.458	16.77	EB	0	0
20	34	GRAHAM	US 24	17.525	18.178	EB	0	0
21	40	JEWELL	US 36	14.93	15.402	EB	0	0
22	47	MARION	K 15	27.471	27.696	NB	1	1
23	49	MARSHALL	US 36	6.998	7.287	EB	1	0
24	53	MONTGOMERY	US 75	1.201	1.325	NB	1	0
25	53	MONTGOMERY	US 75	1.201	1.325	NB	0	0
26	54	NEMAHA	US 36	2.748	3	EB	0	0
27	55	NEMAHA	US 36	3	3.249	EB	0	0
28	55	NEMAHA	US 36	3	3.249	EB	0	0
29	58	OSBORNE	US 24	30.468	30.7	EB	0	0
30	61	OSBORNE	US 24	31.187	31.374	EB	0	0
31	65	POTTAWATOMIE	US 24	13.054	13.398	EB	1	0
32	65	POTTAWATOMIE	US 24	13.054	13.398	EB	4	0
33	67	POTTAWATOMIE	US 24	14.47	14.949	EB	2	2
34	73	ROOKS	US 24	28.009	28.153	EB	0	0
35	73	ROOKS	US 24	28.009	28.153	EB	0	0
36	77	STAFFORD	US 281	12.073	12.426	NB	0	0
37	81	WALLACE	US 40	4.163	4.494	EB	0	0
38	81	WALLACE	US 40	4.163	4.494	EB	0	0
39	82	WALLACE	US 40	4.494	4.64	EB	0	0
40	82	WALLACE	US 40	4.494	4.64	EB	0	0
41	4	BOURBON	US 69	8.097	9.067	NB	1	0
42	27	CHEROKEE	K 66	0.811	1.247	EB	1	1
43	29	CHEROKEE	K 66	1.638	2	EB	3	1
44	30	BUTLER	K 66	35.757	36.03	NB	2	1

Intersection ID	Section ID	County Name	Highway No.	Begin County MP	End County MP	Direction (EB/WB/NB/SB)	Number of all crashes within ints. box of 300 ft	Number of intersection related crashes (only)
45	30	BUTLER	K 66	35.757	36.03	NB	2	0
46	39	CLOUD	US 81	9.036	12.68	SB	1	0
47	40	CLOUD	US 81	12.68	14.168	SB	2	0
48	40	CLOUD	US 81	12.68	14.168	SB	2	0
49	43	COWLEY	US 77	0	1.977	NB	2	0
50	68	FINNEY	US 50	4.931	5.983	EB	1	1
51	70	FINNEY	US 83	20.577	23.149	NB	1	0
52	80	GEARY	US 50	15.552	15.659	EB	1	0
53	91	JACKSON	US 75	7.999	16.628	NB	2	1
54	91	JACKSON	US 75	7.999	16.628	NB	1	0
55	91	JACKSON	US 75	7.999	16.628	NB	1	0
56	91	JACKSON	US 75	7.999	16.628	NB	0	0
57	113	KINGMAN	US 24	26.635	29.671	EB	1	1
58	115	KINGMAN	US 24	34.735	36.747	EB	0	0
59	180	MONTGOMERY	US 169	6.437	6.584	NB	1	1
60	197	OTTOWA	US 81	12.458	17.904	SB	1	0
61	198	OTTOWA	US 81	22.485	24.28	NB	1	1
62	204	OTTOWA	US 24	18.449	19.664	SB	1	0
63	241	RENO	US 50	24.499	28.499	WB	1	1
64	241	RENO	US 50	24.499	28.499	WB	1	1
65	294	SEDGWICK	US 54	10.1	11.07	EB	1	1