Connecting South Dakota and the Nation

SD2013-04-F

South Dakota
Department of Transportation
Office of Research

Calibration of HSM Predictive Methods on State and Local Rural Highways

Study SD2013-04

Final Report

SDSU

Prepared by
South Dakota State University
Department of Civil and Environmental Engineering
Brookings, SD 57007

February 2022

DISCLAIMER

ACKNOWLEDGEMENTS

This work was performed under the direction of the SD2013-04 Technical Panel:

Lee Axdahl................Department of Public Safety
Mike Behm .... Project Development
Harvey Fitzgerald .... Research
Nicole Frankl .... Project Development
David Huft .... Research
Sharon Johnson. Federal Highway Administration
Jess Powell.............................. Operations Support
John Rehorst .... Operations
Cliff Reuer ....SDLTAP
Kelly Serr......................National Weather Service
Megan Steever ....Road Design
Andy Vandel........................ Project Development

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. SD2013-04-F	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Calibration of HSM Predictive Methods on State and Local Rural Highways		5. Report Date 02/22/2022 6. Performing Organization Code	
7. Author(s) Xiao Qin, Zhi Chen, Mohammad Razuar Shaon		8. Performing Organization Report No.	
9. Performing Organization Name and Address South Dakota State University Department of Civil and Environmental Engineering Brookings, SD 57007		10. Work Unit No. HRY304 11. Contract or Grant No. 311179	
12. Sponsoring Agency Name and Address South Dakota Department of Transportation Office of Research 700 East Broadway Avenue Pierre, SD 57501-2586		13. Type of Report and Period Covered Final Report 10/18/2013 – 06/30/2015 14. Sponsoring Agency Code	
15. Supplementary Notes An executive summary is published separately as SD2013-04-X.			

16. Abstract

Performance-based safety goals and objectives are more attainable with the use of the Highway Safety Manual (HSM). However, the safety performance functions (SPFs) in the HSM are not accurate because they are not calibrated to local conditions. Additionally, each SPF and crash modification factor (CMF) assumes a set of base site conditions which might not be realistic or representative of local roadways. The calibration procedures provided in HSM Part C Appendix A should therefore be modified to accommodate local data availability as well as roadway, traffic, and crash characteristics. Furthermore, a set of base conditions applicable to local highways should be determined.

This study presents the application of the HSM to rural state highways in South Dakota and provides important guidance and empirical results regarding how to calibrate HSM models. Calibration guidelines have been developed to facilitate future calibration activities. The calibration was based on crash data from three roadway segment types and five highway intersection types during a five-year period (2008-2012). The calibration process includes establishing new base conditions, developing SPFs, converting CMFs to base conditions, and deriving calibration factors. Considering the sample size, base conditions, and prediction accuracy, four state-specific SPF models have been developed for two segment types (two-way two-lane and multilane divided) and two intersection types (two-way two-lane three-leg STOP and two-way two-lane four-leg STOP), respectively. The HSM SPF's are retained for other highway facility types. The values of calibration factors for crashes are derived by injury severity scale (i.e., killed, type A, B, or C injury and property damage only crashes). Results show that the number of crashes observed in South Dakota can be drastically different from those predicted by the HSM.

The HSM models in this study underestimate the total crash count on roadway segments for all rural state highway segment types in South Dakota. On the other hand, the models overestimate the total crash count at intersections for all the intersection types in South Dakota except multilane three-leg STOP sites. The HSM models overestimate fatal and injury crashes in all highway facility types. Property damage-only crashes are sometimes overestimated by 200 percent and sometimes underestimated by 72 percent. The state-specific SPF models developed from South Dakota's safety data perform better for facility types that are not adequately predicted by the HSM method. The recommendation is to use a hybrid method with both HSM and state-specific models for certain segment and intersection types to achieve high crash prediction accuracy.

17. Keywords highway safety; safety performance function; crash modification factor; calibration		18. Distribution Statement No restrictions. This document is available to the public from the sponsoring agency.	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 97	22. Price

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

TABLE OF ACRONYMS

Acronym	Definition
AADT	Annual Average Daily Traffic
AASHTO	American Association of State Highway and Transportation Officials
C/ Cr / Ci	Calibration Factor (*r* for intersection and *i* for segment)
CMF	Crash Modification Factor
DD	Driveway Density
FI	Fatal + Injury Crashes (including K, A, B and C)
FI*	Fatal + Injury* Crashes (including K, A, and B)
GIS	Geographic Information System
GLM	Generalized Linear Model
GOF	Goodness-of-fit
HSM	Highway Safety Manual
IHSDM	Interactive Highway Safety Design Model
LOESS	Locally Weighted Scatterplot Smoothing
LRS	Linear Reference System or Linear Referencing System
LW	Lane Width
MRM	Mileage Reference Marker
NB	Negative Binomial (distribution)
NSTRI	Non-State Trunk Road Inventory
PC	Point of Curve
PHF	Peak Hour Factor
PT	Point of Tangent
PDO	Property Damage Only
PVI	Point of Vertical Intersection
RIS	Roadway Inventory System
RM3ST	Rural Multilane Three-Leg Intersection with STOP Control on the Minor Road Approach
RM4ST	Rural Multilane Four-Leg Intersection with STOP Control on the Minor Road Approach
RM4SG	Rural Multilane Four-Leg Signalized Intersection
RM4D	Rural Multilane Divided Segment
RM4U	Rural Multilane Undivided Segment
RT	Rural Two-lane Two-way Segment
RT3ST	Rural Two-lane Three-Leg Intersection with STOP Control on the Minor Road Approach
RT4ST	Rural Two-lane Four-Leg Intersection with STOP Control on the Minor Road Approach
SAE	Sum of Absolute Errors
SD	South Dakota
SDARS	South Dakota Accident Records System
SDDOT	South Dakota Department of Transportation
SDDPS	South Dakota Department of Public Safety
SPF	Safety Performance Function
SW	Shoulder Width
WVC	Wildlife-vehicle Collision

1. EXECUTIVE SUMMARY

1.1. INTRODUCTION

To meet the economic needs of local communities while also keeping motorists safe, transportation agencies need to set safety goals and develop methods that will ultimately reduce fatalities and serious injuries. The Highway Safety Manual (HSM) of the American Association of State Highway and Transportation Officials (AASHTO) provides statistically sound safety analysis methods that have been developed over decades of highway safety research.

The predictive models provided by the HSM offer reliable estimates of expected crash frequencies for specific roadway segments and intersections. However, because safety conditions change over time, the models should be calibrated to avoid compromising safety estimates, producing unrealistic results, or undermining agency accountability. Calibration is also necessary if the results are to be compared to an agency's estimates based on historical crash data, as the HSM models were developed from safety data collected in multiple states.

The South Dakota Department of Transportation (SDDOT) has implemented the HSM guidelines in its project development and planning processes. For example, SDDOT uses HSM models to compare safety design alternatives, evaluate site-specific safety issues, and plan future safety projects. Specifically, HSM methods are used to screen South Dakota (SD) roadways to identify problem areas for further safety review. This report presents a South Dakota version of HSM models and provides guidance for future calibration activities. The principal findings are based on calibrating the HSM predictive models for South Dakota rural state highway segments and intersections. The calibrated South Dakota HSM models are critical complementary to the HSM models. For the highway sites whose crash observations substantially deviate from the HSM prediction, the state-specific models provide more accurate estimates, and should therefore be used to make more informed decisions regarding highway safety improvements.

1.2. RESEARCH OBJECTIVES

The main objectives of this project are to:

1) calibrate the HSM predictive models to South Dakota data for all rural facility types where Safety Performance Functions (SPFs) are available and provide guidance on the selection of Crash Modification Factors (CMF)
2) develop guidelines for future calibration of HSM predictive methods

The calibration method for the HSM predictive method followed HSM Volume 2 Part C – Appendix A. Necessary modifications to the HSM SPFs and CMFs were discussed and justified by rigorous engineering and statistical analyses. The scope of the calibration was limited to the rural state highway segment and intersection types listed in Table 1-1. For rural two-lane two-way highways, undivided highway segments (RT), three-leg stop-controlled (RT3ST), and four-leg stop-controlled (RT4ST) intersections were calibrated. For rural multilane highways, both undivided (RM4U) and divided highway segments (RM4D), three intersection types including three-leg stop (RM3ST), four-leg stop (RM4ST), as well as four-leg signalized (RM4SG) intersections were calibrated.

Table 1-1 Facility Types Used for HSM Predictive Model Calibration

Facility Type	Segment	Intersection
Rural Two-lane Two-way Highways	Undivided (RT)	Three-Leg Stop (RT3ST) Four-Leg Stop (RT4ST)
Rural Multilane Highways	Undivided (RM4U) Divided (RM4D)	Three-Leg Stop (RM3ST) Four-Leg Stop (RM4ST) Four-Leg Signalized (RM4SG)

1.3. TASK DESCRIPTIONS

Thirteen tasks were performed to accomplish this project:

1.3.1. Meet with Technical Panel

Meet with the project's technical panel in Pierre, SD to review the project's scope and work plan.

On February 7, 2014, the research team met with the project's technical panel at the SDDOT office in Pierre, SD to discuss the project's scope, work plan, and data needs. The research team presented the required and optional data items for the calibration and SDDOT agreed to provide such data.

1.3.2. Review and Summarize Literature

Review and summarize literature pertinent to the calibration of Highway Safety Manual predictive methods, including information about other state DOT efforts to develop or calibrate HSM Safety Performance Functions.

The literature related to the calibration of HSM predictive methods was reviewed and synthesized. The review was focused on the criteria and methods for creating homogeneous highway segments, the development of agency-specific safety performance functions, the application of crash modification factors, and the derivation of calibration factors.

1.3.3. Apply HSM Calibration

Apply the HSM Part C calibration technique to sample data provided by SDDOT.

The calibration procedure illustrated in Figure 1-1 was implemented. HSM models and equations were used without alteration, and all the steps recommended in the HSM were strictly followed. Local calibration factors were calculated for segments and intersections using South Dakota safety data.

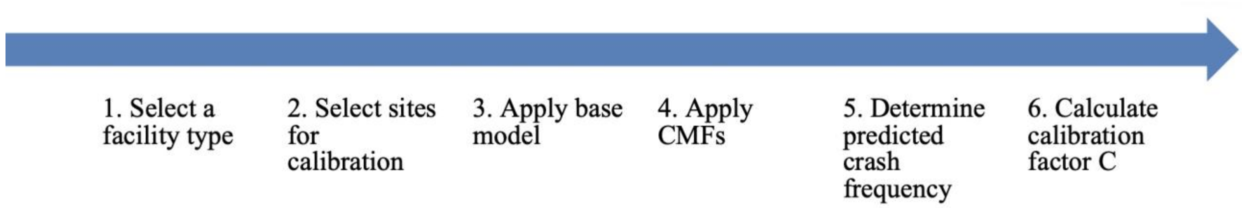


Figure 1-1 HSM Calibration Procedure

1.3.4. Review Base Conditions

Review base conditions for SPFs and define a set of base conditions appropriate for SPF and CMF models to be used on South Dakota highways.

The base conditions for SPFs of a highway facility type represent the most prevailing attributes for that facility. Since the HSM SPFs were developed with data from other states – not including South Dakota – the HSM base conditions may inaccurately represent South Dakota roadways and compromise calibration accuracy. After reviewing the same set of controlling variables used to define the HSM base conditions, new base conditions were established if differences were found.

1.3.5. Identify Facility Types for Calibration

Identify which facility type SPFs will be calibrated based on the available data and needs of Project Development staff.

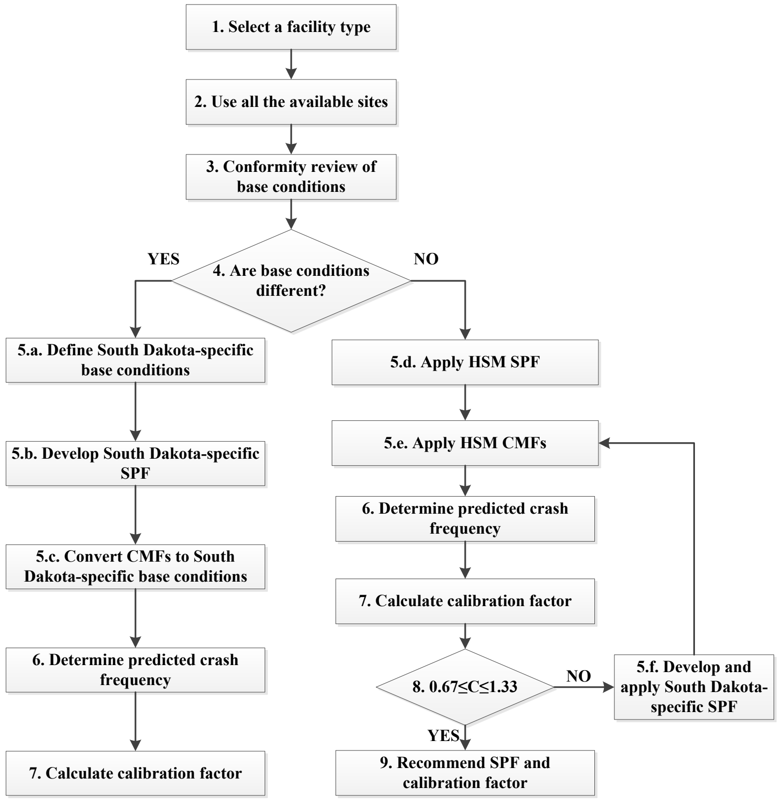
Not all facility types in the HSM are available in South Dakota; thus, calibration was not performed for facility types in which no data were available. Due to inadequate information on local roads, only rural state highways were calibrated for this project.

1.3.6. Determine Needed HSM Modifications

Determine any necessary modifications to HSM procedures and calculate an appropriate sample size for calibration.

Figure 1-2 presents the modifications made to the HSM predictive method.

Figure 1-2 Modified HSM Predictive Method



For the facility types with South Dakota-specific base conditions, new SPFs were developed with all the base condition sites. Accordingly, the HSM CMFs were converted because they represent the changes in crash frequency from base conditions.

According to Section A.1.1.2 in the Appendix, HSM Part C, the desirable minimum sample size for the calibration dataset is 30 to 50 sites, and these sites should have a minimum of 100 observed crashes per year. This minimum crash count requirement was difficult to be met for South Dakota safety data if 30 or 50 sites were randomly selected because of extremely low number of crashes at most sites in South Dakota. In this project, all available data were used to determine calibration factors.

1.3.7. Evaluate Crash Modification Factors

Determine which CMFs are appropriate for use in calibration and provide guidance on future application of CMFs.

Since the values of CMFs in the HSM are determined for a specified set of base conditions, when the base conditions change, the CMFs will change accordingly. In this project, the CMF for shoulder width was adjusted due to the change of shoulder width in the base conditions. The adjustment procedure was subsequently developed and the other CMFs remained unchanged. No new CMFs for the implementation of a specific safety countermeasure have been evaluated and implemented in this project because such information was either not available or not provided by SDDOT.

1.3.8. Submit Technical Memorandum

Submit a technical memorandum and make a presentation to the technical panel summarizing the literature search and propose a methodology for approval.

A technical memorandum was delivered to the technical panel on October 9th, 2014 and a presentation was made at SDDOT in Pierre on October 14th, 2014. The technical memorandum includes a literature review, findings from the South Dakota-specific data, calibration methodologies, and proposed modifications. The comparison with and without modifications was also included. The need for regional calibrations was not pursued due to the mixed results of the calibration factors across the state. The attendees acknowledged the challenge of defining boundaries between regions due to the absence of explicit spatial patterns. In exchange, the calibration of crashes by severity level was added to the project scope. The addendum proposal to calibrate crashes by injury severity scale was developed and delivered to the panel via email on October 28th, 2014.

1.3.9. Determine Need for Regional Calibration

Upon approval, apply a statistical technique to determine whether SPFs need to be developed for different regions, such as Black Hills versus non-Black Hills roads, local and state roads, or other categories deemed important by the technical panel or through the literature review.

To investigate the need for region-specific calibration factors, the ratio of the observed and predicted crash frequency, aka a calibration factor, was computed and mapped for all state highway segments and intersections. After reviewing the calibration factor map, the research team and the technical panel were unable to identify strong spatial patterns and therefore, could not define explicit boundary lines between regions without additional information and more in-depth analysis. The development of region-specific calibration was decided to be replaced by the injury severity calibration, as explained in Task 8.

1.3.10. Calibrate Safety Performance Functions

Calibrate the Highway Safety Manual SPFs to South Dakota's data using the proposed methodology.

No additional comments were received from the panel for the technical memorandum. The research team continued to develop and refine agency-specific SPFs for South Dakota rural state highways. The results were summarized and the findings were included as part of the final report.

1.3.11. Develop Guidelines for Future Calibration

Develop guidelines that include procedures for the future calibration of SPFs.

Following the format presented in the HSM's Volume 2 Part C, Appendix A.1 ― Calibration of the Part C Predictive Models, the guidelines for future HSM calibration at SDDOT have been developed. An example calibration factor calculation has been included. Additional information for SDDOT data sources, steps for data preparation and reduction, and recommendations have been described in detail.

1.3.12. Prepare Final Report

Prepare a final report summarizing research methodology, findings, conclusions, and recommendations.

A draft final report documents the project's results, findings, data requirements, methodologies, conclusions, and recommendations. This report was submitted to the SDDOT technical panel for review on

February 8, 2015. On March 23, 2015, the research team received the comments from the panel. The revised final report was delivered to SDDOT on May 1, 2015.

1.3.13. Make Executive Presentation

Make an executive presentation to the SDDOT Research Review Board at the conclusion of the project.

The project PI will present the findings of the study, including a recommended methodology, comparative tables, and analysis to the SDDOT Research Review Board at the conclusion of the project. This presentation will summarize project activities and any conclusions or recommendations that emerge from the research.

1.4. FINDINGS

1.4.1. Data Collection and Processing

Calibration requires extensive data collection for each specific type of highway. Following the required and desirable data items for calibration in the HSM, the available data elements in South Dakota have been reviewed, processed, and prepared. The roadway data were provided by SDDOT, and the crash data were provided by the South Dakota Department of Public Safety (SDDPS).

The data required for calibrating the HSM predictive method were collected from intersection and segment shapefiles, the Roadway Inventory System (RIS), and the South Dakota Accident Records System (SDARS). The intersection shapefile was readily available for processing, but the segment shapefile required additional work such as the creation of homogeneous segments based on traffic and geometric features. In this process, the RIS event tables of features such as Annual Average Daily Traffic (AADT), surface type, median, horizontal curve, vertical curve and speed limit were merged into one single table. The roadway was then segmented wherever a feature changed. A roadway shapefile was created for homogeneous segments using the SDDOT linear reference system. After creating homogenous segments, crashes between 2008 and 2012 were assigned to intersections and segments.

Several key data elements were redefined after the project's technical panel deliberated. In this study, a rural state highway refers to a US or South Dakota (non-Interstate) highway that is located outside city boundaries. A rural state intersection refers to an intersection between two rural state highways or a rural state highway and a federal-aid non-state highway. Segment-related crashes include wildlife-vehicle collisions (WVCs). An intersection-related crash is defined as: 1) a crash that happens within a 100-foot radius of the center of the intersection; 2) any crash occurring within a 200-foot radius from the intersection that is coded in the accident database as "intersection" or "intersection-related". All critical issues encountered and steps taken to prepare the final dataset for calibration are documented in detail in Chapter 6: Findings.

1.4.2. Methodology

The HSM recommends using three indicators to predict the number of crashes for a given roadway segment or intersection: Safety Performance Function (SPF), Crash Modification Factor or Function (CMFs), and calibration factor (C). The SPF predicts crash frequency as a function of annual average daily traffic (AADT) for roadway segments with basic geometric and traffic conditions. The crash modification factor (CMF) is a measure of the safety effectiveness of a particular treatment or design element which is not different from the basic conditions described in the SPF. After all available CMFs have been considered, the calibration factor serves as the ultimate adjustment for all the other known or unknown, measurable or immeasurable differences, such as climate, driver populations, animal populations, crash reporting thresholds, and crash reporting system procedures. The calibration factor is the ratio of the observed number of crashes to the expected or predicted number of crashes (e.g., a value larger than one suggests underestimation of the HSM predicted value).

Each of the three indicators yields the opportunity for calibration if more accurate results are desired. The base conditions defined in the HSM may not be representative of South Dakota. If the base conditions are not representative of South Dakota, the CMFs and calibration factors can scale the safety performance up or down to correlate with the base conditions. The modification strategies include:

a) Defining South Dakota-specific base conditions: all South Dakota rural state highways were screened using the controlling variables for base conditions specified in the HSM. The highway segments or intersections with attributes that are different from the HSM base conditions were defined as state-specific base conditions;

b) Developing South Dakota-specific SPFs: South Dakota-specific SPFs were developed using the same statistical techniques applied in the HSM (i.e., negative binomial generalized linear regression model). It is believed that agency-developed SPFs were more accurate than SPFs in the HSM.

c) Converting CMFs to South Dakota-specific base conditions: The development of two methods allowed for the adjustment of CMF's to state-specific base conditions. One method treated a CMF as a scale factor and the other treated a CMF as a function of AADT.

1.4.3. Results

The HSM predictive method was used to calculate calibration factors for total crashes and for crashes of different injury severities, as shown in Table 1-2.

Table 1-2 Calibration Factors for Total Crashes and Crashes by Injury Severity

Facility Type[1]		Calibration Factor			
		Total	FI	FI[3]	PDO
Segment	RT	1.18	0.56	0.72	1.47
	RM4U	1.14	0.38	0.46	2.52
	RM4D	1.57	0.30	0.26	3.50
Intersection	RT3ST	0.55	0.55	0.73	0.55
	RT4ST	0.33	0.35	0.45	0.31
	RM3ST	1.36	0.53	0.31	2.16
	RM4ST	0.56	0.56	0.59	0.57
	RM4SG[2]	0.04	0.00	0.00	0.07

1. Refer to the table of acronyms and abbreviations.
2. Results are not reliable because only one site is available for calibration. RM4SG is not included in any discussion hereinafter.

3. The KABCO (killed, type A, B, or C injury, and O for property damage only) scale is used, but includes only KAB crashes; crashes with severity level C (possible injury) and O are not included.

South Dakota-specific predictive methods were developed for RT, RM4D, RT3ST, and RT4ST. The state-specific base conditions for these four facility types are given in Table 1-3, which lists only the base conditions that are different from those in the HSM. Complete base conditions of rural two-lane two-way and rural multilane highway facilities are shown in Appendix A (Table 11-1 and Table 11-4).

Table 1-3 South Dakota-Specific Base Conditions by Facility Type

Facility Type	RT	RM4D	RT3ST	RT4ST
South Dakota Base Conditions	Shoulder Width = 4 ft.	Lane Width= 13 ft. Shoulder Width = 4 ft.	No Deviations	No Deviations

The state-specific SPFs were developed only with the new base condition sites. The negative binomial regression model, the same methodology used to develop all HSM SPFs, was used to develop the state-specific SPFs. Table 1-4 presents the state-specific SPFs for each of the four facility types.

Table 1-4 Summary of South Dakota-Specific SPFs

Segment		Intersection	
RT	$AADT \times L \times 365 \times 10^{-6} \times e^{(-0.1101)}$	**RT3ST**	$e^{-9.93+0.66\times ln(AADT_{maj})+0.52\times ln(AADT_{min})}$
RM4D	$e^{(-19.7106+2.4597\times \ln(AADT)+\ln(L))}$	**RT4ST**	$e^{-10.55+0.79\times ln(AADT_{maj})+0.51\times ln(AADT_{min})}$

After establishing the state-specific SPFs, CMFs must be converted accordingly as they represent quantitative changes in predicted crash frequencies that result from site characteristic variations from base conditions. The base conditions for intersections were the same as these in the HSM, while the South Dakota base conditions of rural roadway segments were different. The adjusted CMF for RT shoulder width is given in Table 1-5. The adjusted CMF for RM4U shoulder width is given in Table 1-6.

Table 1-5 Adjusted CMF for Shoulder Width for Rural Two-lane Two-way Segments

Shoulder Width	AADT		
	<400	400-2000	>2000
0 feet	1.078	$1.078 + 1.4 \times 10^{-4}(AADT - 400)$	1.304
2 feet	1.049	$1.049 + 5 \times 10^{-5}(AADT - 400)$	1.130
4 feet	1.000	1.000	1.000
6 feet	0.980	$0.980 - 6.9 \times 10^{-5}(AADT - 400)$	0.869
8 feet or more	0.961	$0.961 - 1.28 \times 10^{-4}(AADT - 400)$	0.756

Table 1-6 Adjusted CMF for Shoulder Width for Rural Multilane Divided Segments

SW	CMF
0 feet	1.082
2 feet	1.037
4 feet	1.000
6 feet	0.954
8 feet or more	0.917

Calibration factors were calculated after the state-specific SPFs and CMFs were applied. Table 1-7 summarizes the calibration results.

Table 1-7 Calibration Factors Following South Dakota-Specific Predictive Methods

Facility Type	Calibration Factor
RT3ST	1.29
RT4ST	1.11
RT	1.05
RM4D	1.23

1.5. CONCLUSIONS

This report summarizes the effort to calibrate the HSM predictive method for different segment and intersection types on South Dakota's rural state highways. Beginning with a strict implementation of the predictive methods documented in Part C of the HSM, the calibration factors based on both total number of crashes and crashes by severity level were calculated for each highway facility type studied in this project. The KABCO scale (referred to in Table 1-2), obtained from the "Person" Table in SDARS by identifying the highest severity level of all persons involved, presents crash injury severity. Predicted crash injuries were estimated using the HSM recommended method. For a rural two-lane highway, predicted crash frequency by injury severity was obtained by applying the HSM crash severity proportion to the total number of predicted crashes. For a rural multilane facility, the predicted crash frequency by injury severity was calculated using the SPFs for different severity levels. The results show large variations in the calibration factors across injury severity and facility type. In general, the HSM models underestimated the total number of crashes on roadway segments for all South Dakota rural state highway types in this study; however, they overestimated the total number of crashes at intersections for all the South Dakota intersection types except RM3ST sites. The HSM models overestimated fatal and injury crashes (FI) for all highway facility types. For property damage only (PDO) crashes, the estimates ranged from overestimation (C_i=0.31 for RT4ST) to underestimation (C_i=3.50 for RM4D).

Large differences remained between prediction and observation for individual sites after calibration factors were applied. The presence of large deviations prompted the establishment of local base conditions and the development of agency-specific models. Among all facility types, RT and RM4D had base conditions that were different from those in the HSM. Due to the sample size constraint, state-specific SPFs were developed only for RT, RM4D, RT3ST, and RT4ST. HSM SPFs were recommended for all other highway facility types (i.e., RM4U, RM3ST, RM4ST, and RM4SG). All new calibration factors for RT, RM4D, RT3ST, and RT4ST were found to be closer to one when compared with the values following the HSM method. A combination of state-specific and HSM models resulted in specific recommendations for each highway facility type. Specific recommendations are discussed in the next section.

In conclusion, the HSM predictive method was assessed for its ability to provide reliable crash predictions when used with South Dakota data. When the HSM method was inadequate in providing close estimates, modifications were introduced (i.e. establishing local base conditions, developing state-specific SPFs, and adjusting CMFs accordingly). After this calibration, predicting crashes on South Dakota's rural state highways is now more accurate.

1.6. IMPLEMENTATION RECOMMENDATIONS

The following implementation recommendations can be made based on the findings in this project.

1.6.1. Rural Two-lane Two-Way Segments (RT)

The state-specific SPF and the adjusted CMF for shoulder width should be used for rural two-lane two-way highway segments. A calibration factor of 1.05 should be used.

1.6.2. Rural Multilane Undivided Segments (RM4U)

The HSM SPF and CMFs should be followed for rural multilane undivided highway segments. A calibration factor of 1.14 should be used.

1.6.3. Rural Multilane Divided Segments (RM4D)

The state-specific SPF and adjusted CMF for shoulder width should be used for rural multilane divided highway segments. A calibration factor of 1.23 should be used.

1.6.4. Rural Two-Lane Three-Leg Intersections with STOP Control (RT3ST)

Rural two-lane three-leg stop controlled intersections can use the HSM SPF with a calibration factor of 0.55 or the state-specific SPF with a calibration factor of 1.29, given the marginal calibration improvement.

1.6.5. Rural Two-Lane Four-Leg Intersections with STOP Control (RT4ST)

Rural two-lane four-leg stop-controlled intersections can use HSM SPF with a calibration factor of 0.33 or the state-specific SPF with a calibration factor of 1.11, given the marginal calibration improvement.

1.6.6. Rural Multilane Three-Leg Intersections with STOP Control (RM3ST)

Due to the small sample size, calibration is not recommended for rural multilane three-leg stop-controlled intersections. More data should be collected, but in the meantime, it is recommended to use the HSM predictive method for rural multilane three-leg intersections with stop control.

1.6.7. Rural Multilane Four-Leg Intersections with STOP Control (RM4ST)

Rural multilane four-leg stop-controlled intersections should use the HSM SPF and CMFs with a calibration factor of 0.56. Predicted crash frequency should be used with caution, however, as this type of facility has 71 sites and just 125 crashes between 2008 and 2012; this is substantially lower than the HSM recommendation of 30-50 sites with at least 100 crashes per year.

1.6.8. Rural Multilane Four-Leg Signalized Intersections

Due to the small sample size, calibration is not recommended for rural multilane four-leg signalized intersections. More data should be collected, but in the meantime, it is recommended to use the HSM predictive method for rural multilane four-leg signalized intersections.

1.6.9. Calibration by Injury Severity

The HSM SPFs and CMFs should be used along with calibration factors by injury severity for RT, RM4U, RM4D, RT3ST and RT4ST. For RM4ST sites, the calibration factor should be used with caution because of the relatively small sample size (71 sites) and low crash frequencies observed. It is not recommended to use injury severity calibration for RT3ST and RM4SG sites because of the extremely small sample sizes (19 RT3ST sites and 1 RM4SG site) and low crash frequencies observed. More data from the two facility types should be collected in the future.

1.7. Data Collection

Although all required data items for intersections are available in the SDDOT "StatetoState_and_StatetoNonStateFedAid" file, data items such as *driveway density, horizontal curve superelevation, roadside design, lighting* and *number of passing lanes and two-way left-turn lanes,* are not available for roadway segments. The missing data items should be collected in future research for a more accurate prediction. The complete list of data available in South Dakota can be found in Table 6-4 List of CMFs for Rural Two-lane Two-way Highway Facilities and

Table 6-5 List of CMFs for Rural Multilane Highway Facilities.

1.8. Data Integration

The data source tables pertaining to intersections have been integrated into an intersections toolbox that is currently in use to apply the HSM predictive methods at SDDOT. The parameters in SPFs can be easily modified in the toolbox to implement state-specific models; however, there is no such convenience for road segments, as those data items are stored in separate tables. The process of integrating the segment data tables developed in this project was complicated because dynamic segmentation cannot be performed without access to the SDDOT's linear referencing system (LRS); therefore, alternative computer programs (e.g., R, MS Excel, ArcGIS) were employed to merge files. These disjointed steps can be integrated and streamlined in a GIS environment with the aid of SDDOT LRS. The data reduction criteria and specific file merging procedures are detailed in Section 6.2 Data Processing.

1.9. Calibration Updates

The HSM recommends that *"the new values of the calibration factors be derived at least every two to three years, and some users may prefer to develop calibration factors on an annual basis*" (1). There are no other standards regarding how often the calibration should be updated. Due to the complexity of calibration, it is recommended that the new values of the calibration factors be derived every three years or when crash frequency or injury severity distribution has been significantly changed for a specific facility (e.g., ±10%).

2. PROBLEM STATEMENT

Defining safety performance expectations is a challenge for transportation agencies. The Highway Safety Manual (HSM) provides guidance for safety analyses using scientific and statistically sound methods (*1*). Given the expense of engineering studies and limited funding, safety reviews based on expected safety performance are a useful way to identify hot spots in a highway network as well as site-specific safety problems. Predictive crash models, as formulated in Equation 1, can pinpoint sites that have a good chance of crash reduction based on decades of safety research and statistical analysis.

$$N_{predicted} = N_{spf} \times C \times (CMF_1 \times \cdots \times CMF_n) \quad (2\text{-}1)$$

where $N_{predicted}$ is the predicted average crash frequency for a site, N_{spf} is the predicted average crash frequency for base conditions for a site, also called safety performance function (SPF), and C is the calibration factor (C_r is for a roadway segment and C_i is for an intersection). A series of crash modification factors (CMFs) account for changes in the number of crashes due to specific site characteristics or safety treatments. Locations where the actual crash count is higher than the predicted crash count need to be further investigated for safety improvements.

Because safety conditions change over time and conditions in South Dakota (or other places) may differ from the conditions assumed in the HSM models, agencies should use calibrated HSM models. Uncalibrated models compromise safety estimates, produce unrealistic results, and undermine accountability. Even agencies that use their own data to develop SPFs should consider calibrating the models every two to three years. HSM models must be calibrated for the results to be comparable to the estimates obtained from an agency's records.

The South Dakota Department of Transportation (SDDOT) has implemented HSM guidelines in its project development and planning processes. Specifically, HSM SPFs and CMFs are used to screen South Dakota (SD) roadways to find problem areas for further safety review. SDDOT also uses HSM models to compare safety design alternatives, evaluate site-specific safety issues, and program and plan future safety projects. Although calibration procedures are available in the HSM in Appendix A, they need to be refined or modified to accommodate South Dakota's data availability, as well as roadway, traffic, and crash characteristics. It is imperative to develop a South Dakota version of the HSM models by using proper calibration methods and to provide guidance for future calibration activities.

3. LITERATURE REVIEW

Decades of research has demonstrated that highway safety analysis can be improved substantially when using a scientific, systematic, consistent, and proactive approach. The HSM provides ways to predict crash frequencies for various facility types (*1*); however, because calibration leads to a more accurate prediction, the HSM models need to be calibrated if they are to be comparable to an agency's estimates. Although calibration procedures are available in HSM Part C, Appendix A, they should be refined or modified to accommodate for local data availability and roadway, traffic, and crash characteristics.

The calibration process accounts for different safety effects due to driver population, environmental variables, and other unobserved or unmeasured factors. The HSM recommends the use of the Safety Performance Function (SPF), Crash Modification Factor or Function (CMF), and calibration factor (C) to predict the number of crashes for a given roadway segment or intersection.

In the HSM predictive method, the SPF predicts crash frequency as a function of AADT for roadway segments with basic geometric and traffic conditions. However, the base conditions defined in the HSM may not be the most prevailing for local conditions. For example, the rural two-lane, two-way road SPF assumes the following base conditions in the HSM:

- 12-foot lane width
- 6-foot shoulder width
- paved shoulder
- a 3-point roadside hazard rating
- 5 driveways per mile
- level grade with no horizontal curvature
- no vertical curvature
- no centerline rumble strips
- no passing lanes
- no two-way left-turn lanes
- no lighting
- no automated speed enforcement

Note that not many sites on rural local roads have paved shoulders, let alone a 6-foot paved shoulder. Additionally, not many rural highways are without horizontal curves. Each of the aforementioned components should be calibrated if more accurate results are desired. A review of existing HSM studies identified the four areas that are most relevant to this project: highway segmentation, safety performance functions, crash modification factors or functions, and calibration factors or calibration functions.

3.1. Segmentation

The HSM requires that a site should be either an intersection or a homogeneous roadway segment. A roadway segment is defined as a part of roadway which is not interrupted by intersection and consists of homogeneous geometric and traffic control features (*1*). For roadway data, an important step is to create homogeneous segments. Homogeneity is the key for a successful development or implementation of SPFs for roadway segments.

Segmentation of a roadway network based on multiple variables can result in very short roadway segments (*2*). Shorter roadway segments are undesirable because roadway features associated with crash risk may not be prominent given a very short distance or given that crash location information may not be accurate enough to appropriately assign each crash. Moreover, when linking crash data with roadway data, the presence of short segments can lead to a large number of segments with zero crashes, which can become problematic for proper statistical inference.

According to the HSM, the rules of dividing a highway include the beginning or ending of a horizontal curve, point of vertical intersection (PVI) of a vertical curve, passing lane or two-way left turn lane and changes with respect to AADT, lane width, shoulder width, driveway density and roadside hazard rating (*1*). There is no prescribed minimum segment length for application of the predictive models, but there is a suggestion to have a segment length of at least 0.10 miles. Recognizing the potential issue of creating short segments, Part C provides the following guidance: "*When dividing roadway facilities into small homogenous roadway segments, limiting the segment length to a minimum of 0.10 miles will decrease data collection and management efforts*." and "*When dividing roadway facilities into small homogenous roadway segments, limiting the segment length to a minimum of 0.10 miles will minimize calculation efforts and not affect results*." (*1*)

Aside from the HSM, researchers have adopted a variety of approaches for segmentation. Miaou and Lum suggested that short sections less than or equal to 80 meters could create a bias in the estimation of linear models except when using Poisson models (*3*). Ogle et al. demonstrated that short segment lengths, less than 160 m, cause uncertain results in a crash analysis (*4*). Qin et al. studied the relationship between segmentation and safety screening analysis using different lengths of sliding windows (*5*), concluding that short segments or extremely long segments created a bias in the identification of sites with safety problems. In a motorway study using sample data from Italy, Cafiso et al. (*6*) used the general estimating equations (GEE) methodology to evaluate three models with the five following approaches to segmentation: 1) homogeneous segments with respect to AADT and curvature using AADT and curvature as explanatory variables; 2) two curves and two tangents within each segment, avoiding short segments when using a single curve; 3) segments having constant AADT; 4) segments having a constant length; and 5) all the variables used in the stepwise procedure are constant within each segment with their original value. Approaches two and four returned the best results.

As can be seen, there are no prescribed rules, but there is some general guidance on how to divide a highway and how to set the minimum segment length. As a set of variables described in the HSM are required to be the same within each segment, creating homogeneous segments based on one or two specific variables will contradict with the definition of segments in the HSM. A balance between segment homogeneity and length must be maintained. Although the minimum segment length can be decided after applying all the required roadway geometric characteristics such as horizontal curve, vertical curve etc., the minimum length of 0.01 mile was considered for this study because hundredth of a mile is usually used as the distance unit for locating a crash.

3.2. Safety Performance Function

The safety performance function (SPF) is a regression equation for estimating the predicted average crash frequency of individual roadway segments or intersections (*1*). SPFs have been studied in different forms, either in the specification of the equation or in the number of variables. Banihashemi (*7*) used roadway segments that satisfy the base conditions to develop two respective base models: $L \times AADT$ and $L \times AADT^p$, where L is the segment length, $AADT$ is the annual average daily traffic, and p is power coefficient. Kononov et al. (*8*) evaluated an alternative approach for choosing a functional form of the SPF. Authors used related traffic flow parameters such as speed and density and chose the SPF developed in sigmoid and exponential functional forms using neural network (NN). The results show that the NN-generated SPF has less bias and a better fit when compared with power-function SPFs that were developed in the generalized linear model (GLM) framework with a negative binomial (NB) error structure.

Brimley et al. (*9*) hypothesized new variables to have a measurable correlation for total crash frequencies which were not used in the HSM SPF for rural two-lane two-way roads. Results showed that among those new variables tested, speed limit and the percentage of multiple unit trucks have a significant correlation with crash frequency. Martinelli et al. (*10*) employed an alternative method in an attempt to develop predictive crash models. The study used the model proposed in the HSM prototype chapter (*11; 12*) with variables including:

- segment length
- AADT
- lane width
- shoulder width
- roadside hazard rating
- driveway density
- horizontal curvature
- grade rate for crest vertical curves
- percent grade for straight grade

To obtain the SPF, the authors substituted values of variables corresponding to base conditions, except for AADT and segment length. This method seems to address the low sample issue by including all sites, but it is essentially the development of a full-scale crash prediction model, which is different from the approach used in the HSM predictive method. Developing a full-scale model is certainly more complicated than developing a model with just AADT and length (*13*). Model specification, variable correlation, and interaction need to be carefully considered when more variables are involved.

In the SPFs developed for the HSM, the predicted average crash frequency is for a roadway segment or intersection under base conditions, and the independent variables are the AADTs of the roadway segment, (and, for roadway segments, the segment length) (*1*). This definition considerably limits the flexibility of the model form and variables to be included. Here, the base conditions should be the most representative segments amongst all types, guaranteeing a sizable sample for developing statistically robust models. Because the most representative roadway type may vary from state to state, it is necessary to check if the most prevailing conditions conform to the HSM base conditions before applying the HSM SPFs.

In a recent study, Abdel-Aty et al. (*14*) developed statewide SPFs for various subtypes of multilane roadway and freeway segments in Florida using a simple SPF form which contained only AADT as the sole explanatory variable. The calibration factors were applied to the default SPFs, and the calibrated SPFs could match statewide SPFs at conditions with average AADT levels; however, at low or high AADT levels, crash frequency was either overestimated or underestimated. To account for this discrepancy, Florida was arbitrarily divided either by region or by district, and region-specific or district-specific SPFs were developed. A population group-level calibration factor was put in place of the state-level calibration factor.

The previous studies have shown that adopting new function forms, including new variables, and using complicated statistical methodologies for SPFs may improve the prediction power. But the tradeoff is the increased complexity of the model. A complicated model can be data demanding as well. Considering these limitations, in this study, only the SPF coefficients were calibrated with local data so that the state-specific models could be compared to the HSM predictive models.

3.3. Crash Modification Factor

Each CMF represents one type of change being made from the base conditions. CMFs could be in the format of a scale factor or a function for a specific site, based on its characteristics. Crash modification functions are useful when a treatment has a varying safety effect at sites with different characteristics. For instance, a crash modification function for a horizontal curve is a function of curve length and degree of curve. Sun et al. (*11*) created a database of the available variables such as segment length, ADT, lane width, and shoulder type and width, while setting other variables to be the same as the base conditions. When the empirical Bayes method is used and the calibration parameter is a function of ADT, the differences between the observed and predicted crash frequencies are well within the 5% range, meaning the results are satisfactory. A sensitivity analysis was performed by collecting additional data to test the effect of driveway density and horizontal curves. It was concluded that omitting one or more insignificant variables in the model calibration will not compromise the model's accuracy, but will instead help to alleviate the burden of collecting additional data.

Kweon and Lim (*15*) developed CMFs for various treatments in Virginia. Some of the CMFs were already included in the HSM, while some were not. Before-and-after studies (naïve before-after, before-after with comparison group, and before-after with empirical Bayes) were used to develop CMFs. For some treatments, a cross-sectional method was applied to estimate CMFs, and a generalized linear model (GLM) with a negative binomial distribution (NB) was chosen as the estimation method. Among various CMFs developed for a given treatment with multiple methods, the CMF with the lowest standard error was chosen for application in Virginia. For CMFs included in the HSM, most Virginia-specific CMFs show similar values while a few carry different values and even opposite trends.

The two main resources for CMFs are the HSM (volume 3, Part D) (*1*) and the FHWA CMF Clearinghouse (www.cmfclearinghouse.com) (*16*). The CMF Clearinghouse is a comprehensive database of all the CMFs available for a given safety treatment, including all the treatments and CMFs in the HSM. Despite the star-rating provided by the CMF Clearinghouse website, the quality of the CMF cannot be guaranteed. On the contrary, each CMF in the HSM passed a rigorous evaluation and was considered to be reliable.

3.4. Calibration Factor

Both the SPF and CMFs account for the safety effects of measured variables. The unmeasured factors can be estimated via an overall calibration factor or function. A calibration factor (C) is the ratio of the expected crash frequency (calculated by the SPF multiplied by all the available CMFs) to the observed crash frequency. C can be directly estimated through the Interactive Highway Safety Design Model (IHSDM) once the data are imported. The HSM recommends using C to adjust for regional differences (*10; 17*). The literature shows that this factor can be affected by sample size as well as other site characteristics. Banihashemi (*7*) evaluated the quality of calibration factors generated from datasets of different percentages of a complete data set. The methodology for conducting this sensitivity analysis relied on the assumption that calibration factors calculated based on different subsets of the same dataset are normally distributed. The author also evaluated different data size percentages for various highway types that fell within a 5% to 10% limit of the ideal calibration factor.

Mehta and Lou (*18*) treated calibration factor as a special case of the NB regression model, but the results showed that the HSM-recommended method outperforms the proposed method for estimating the calibration factor. When unmeasured errors represented by the calibration factor are correlated with observable variables such as AADT, the calibration function can be more effective than a single ratio in describing the trend or pattern. This argument is similar to that of the crash modification factor vs. crash modification function. In another study, Sun et al. (*11*) treated the calibration parameter as a function of AADT and set different calibration values for different ranges of AADT. The trend of the ratio of the estimation to observation may change considerably, making it difficult to be represented by a single value. However, compared to the large number of papers adopting the single value as the calibration factor (*9; 10;*

17), little research has studied how a well-defined calibration function can greatly improve the power of prediction models.

To avoid the sampling bias mentioned in the previous studies, all South Dakota safety data provided by SDDOT were used to derive the calibration factors. Statistical analysis can answer whether or not a calibration should be considered as a factor or as a function of some site characteristics.

4. OBJECTIVES

4.1. Calibrate HSM Models

Calibrate HSM predictive models to South Dakota's data for all rural facility types where SPFs are available and provide guidance on the selection of CMFs.

All available safety data for rural state highway facilities were provided by SDDOT, including three segment types and five intersection types. Following the HSM calibration procedures and models, values were derived for calibration factors for these facility types using South Dakota's data. After a comprehensive review, highway facilities with base conditions different from HSM base conditions were identified, and state-specific safety performance functions were developed. A method for converting the HSM CMFs, based on the new base conditions, was proposed and implemented. After implementing these enhancements to the HSM predictive method, new values were derived for calibration factors, and significant improvements were observed for all highway facility types.

4.2. Develop Calibration Guidelines

Develop guidelines for future calibration of HSM predictive methods.

Step-by-step procedures have been developed to guide future calibration activities at SDDOT. The guidelines, presented in Appendix B, include data requirements, calibration procedures, sample size recommendations, and an example for rural multilane undivided highway segments with three-year crash data. Diagrams, tables, and equations are supplied when appropriate in order to make the guidelines more reader-friendly.

Even with these guidelines, the current data processing method for integrating multiple tables is complex. A more streamlined process should be implemented through dynamic segmentation using the linear referencing system. We anticipate the same process can be replicated and simplified using the SDDOT linear referencing system with the aid of the GIS unit.

5. TASK DESCRIPTIONS

5.1. Meet with Technical Panel

Meet with the project's technical panel in Pierre, SD to review the project scope and work plan.

On February 7, 2014, the research team met with the project's technical panel at the SDDOT office in Pierre, SD to discuss the project's scope, work plan, and data needs. The research team presented the required and optional data items for the calibration, and SDDOT agreed to provide all available data.

5.2. Review and Summarize Literature

Review and summarize literature pertinent to the calibration of Highway Safety Manual predictive methods, including information about other state DOT efforts to develop or calibrate HSM Safety Performance Functions.

The literature related to the calibration of HSM predictive methods was reviewed and synthesized. The review was focused on the criteria and methods for creating homogeneous highway segments, developing agency-specific safety performance functions, applying crash modification factors, and deriving calibration factors.

5.3. Apply HSM Calibration

Apply the HSM Part C calibration technique to sample data provided by SDDOT.

According to the HSM, the calibration of the Part C SPFs will yield satisfactory results that compare with predictive models developed with agency safety data. To prove this concept, all HSM equations were used without modification and all steps recommended in the HSM were strictly followed. The list of data requirements for each of the 10 intersection types and five segment types was submitted to SDDOT. After receiving the raw data tables from SDDOT, the tables were processed and prepared in the format needed for calibration. The local calibration factors for rural state highway segments and intersections were calculated. A greater-than-one calibration factor means more crashes were observed than predicted, while a smaller-than-one calibration factor means less crashes were observed than predicted. These calibration factors were compared with those calculated by the South Dakota-specific models as described in Task 6: Determine Needed HSM Modifications.

5.4. Review Base Conditions

Review base conditions for SPFs and define a set of base conditions appropriate for SPF and CMF models used on South Dakota highways.

The base conditions for SPFs of a highway facility type represent the most prevailing highway attributes within the facility. Since the HSM SPFs were developed using data from states that do not include South Dakota, it is likely that these base conditions do not represent South Dakota roadways, thus compromising calibration accuracy. All roadway segments and intersections within each facility type defined in the HSM (e.g. rural two-lane two way roads, rural multilane highways) were reviewed in this task. Using the same controlling variables (e.g. lane width, shoulder width and type, etc.), we determined new base conditions if they were different from those in the HSM. For example, if the prevailing rural two-lane two-way highway shoulder width was 4-feet wide, we used a 4-foot shoulder width rather than a 6-foot shoulder width (as recommended in the HSM) for one of the base conditions. The HSM default base conditions were applied when data were either not available or not applicable for certain variables like automated speed enforcement, red-light cameras, lighting, passing lanes, side slopes, on-street parking, traffic signal timing, etc.. A set of suitable base conditions for South Dakota highways was prepared in order to develop state-specific SPFs.

5.5. Identify Facility Types for Calibration

Identify which facility type SPFs will be calibrated, based on the available data and needs of Project Development staff.

The first edition of the HSM contains 10 intersection types and five segment types, some of which are not available in the predominantly rural state of South Dakota. Some facility types in South Dakota constitute a very small portion of the highway system, such as rural multilane signalized intersections and rural multilane four-leg intersections with stop control.

Another concern is that data availability is highly disproportional between the state highway system and local roads. After reviewing with the technical panel all available data, it was decided that the project be limited to rural state highway calibration and that all facility types be provided by SDDOT. The facility types that were identified for calibration are as follows:

- For rural two-lane two-way highways, undivided highway segments (RT), three-leg stop-controlled (RT3ST) and four-leg stop-controlled (RT4ST) intersections.
- For rural multilane highways, both undivided (RM4U) and divided highway segments (RM4D), three intersection types including three-leg stop (RM3ST), four-leg stop (RM4ST), as well as four-leg signalized (RM4SG) intersections.

5.6. Determine Needed HSM Modifications

Determine any necessary modifications to HSM procedures and calculate an appropriate sample size for calibration.

Modifications to HSM predictive methods were proposed in **Error! Reference source not found.**. C onformity review is the process of examining whether the prevailing roadway characteristics in South Dakota are the same as those in the HSM. If the characteristics are different, new base conditions should be established. For the facility types with the South Dakota-specific base conditions, new SPFs were developed with all the base condition sites. Subsequently, the HSM CMFs were converted because they are related to the base conditions. South Dakota-specific SPFs have also been developed for the highway facility types under the circumstance when the calibration factor is considerably smaller than or greater than one.

According to Section A.1.1.2 of the Appendix to HSM Part C, the desirable minimum sample size for the calibration data set is 30 to 50 sites. These sites should have a minimum total of 100 observed crashes per year. It was difficult for South Dakota to meet this requirement when the sites were randomly selected. In this project, all available data was used to determine the calibration factor.

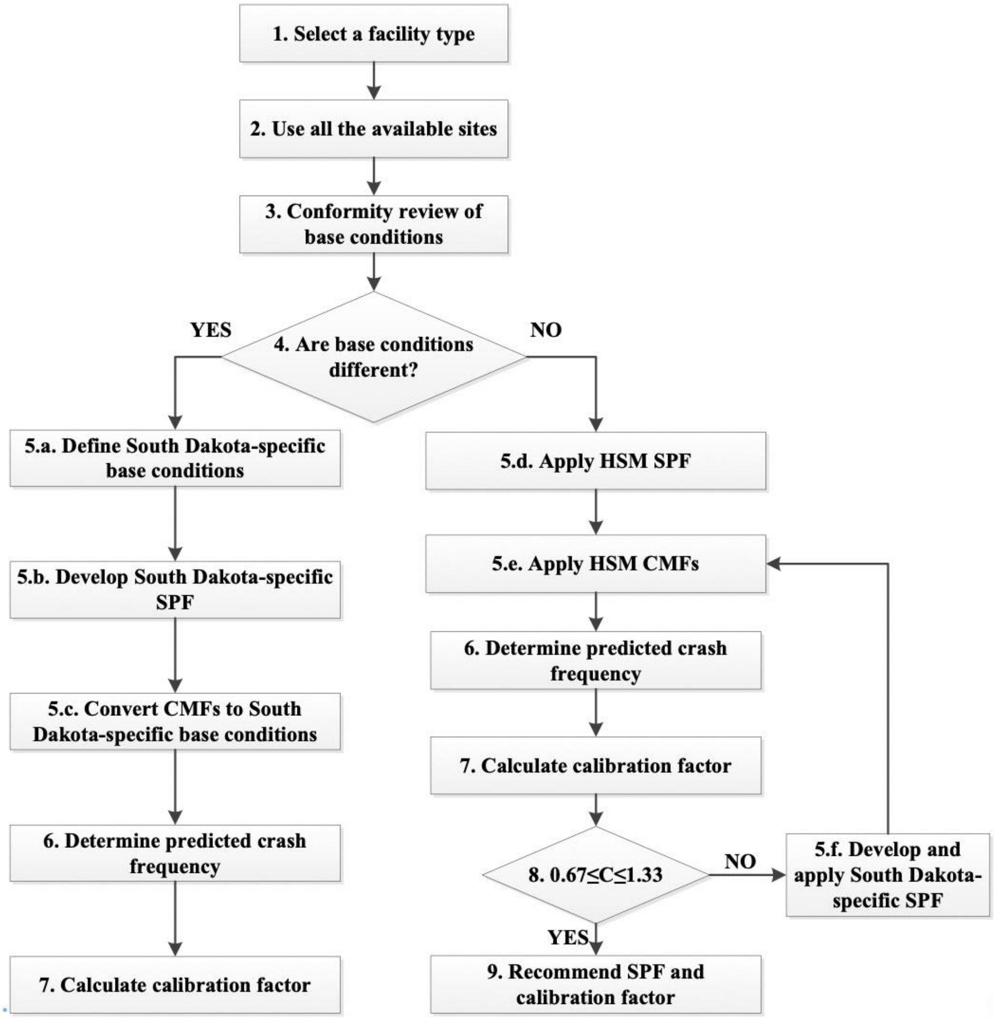


Figure 5-1 Modified HSM Predictive Method

5.7. Evaluate Crash Modification Factors

Determine which CMFs are appropriate for calibration and provide guidance on the future application of CMFs.

Since the values of CMFs in the HSM are determined for a specified set of base conditions, they will change when base conditions change. In this project, the CMF for shoulder width was adjusted due to the change in shoulder width for the base conditions. The adjustment procedure was subsequently developed, and the other CMFs remained unchanged. The project did not evaluate or implement new CMFs for specific safety countermeasures because the SDDOT did not provide that information.

5.8. Submit Technical Memorandum

Submit a technical memorandum and present to the technical panel a summary of the literature review. Propose a methodology for approval.

A technical memorandum was delivered to the technical panel on October 9th, 2014, and a presentation was made at SDDOT in Pierre on October 14th, 2014. The technical memorandum included a literature review, South Dakota data findings, calibration methodologies, and proposed modifications. Discussion regarding the need for regional calibrations was not pursued, as results are not uniform throughout the state. The meeting attendees acknowledged the challenge to define boundaries between regions because of the absence of explicit spatial patterns; thus, it was decided that the calibration of crashes by severity level be added to the project scope. Per the panel's request, the addendum proposal to calibrate crashes by injury severity scale was delivered to the panel via email on October 28th, 2014.

5.9. Determine Need for Regional Calibration

Upon approval, apply a statistical technique to determine whether SPFs should be developed for different regions (e.g. Black Hills roads versus non-Black Hills roads, local roads versus state roads) or for other categories deemed important by the technical panel or by the literature review.

One proposed modification in Task 5.6 is to stratify the calibration factor by region or by another category if a large variation is found within a similar state facility. To investigate the need for region-specific calibration factors, the ratio of the observed to predicted crash frequency (i.e. calibration factor), was calculated and mapped for all state highway segments and intersections. After reviewing the calibration factor map, the research team and the technical panel were unable to identify clear spatial patterns; therefore, there was no way to define explicit boundary lines between regions without a more in-depth analysis. It was decided that region-specific calibration factors be developed to replace the injury severity calibration, as explained in Task 8.

5.10. Calibrate Safety Performance Functions

Calibrate the Highway Safety Manual SPFs to South Dakota's data using the proposed methodology.

No additional comments have been received from the technical panel for the technical memorandum. The research team continued to develop and refine agency-specific SPFs for South Dakota rural state highways. The results were summarized and the findings were presented as part of the final report.

5.11. Develop Guidelines for Future Calibration

Develop guidelines that include procedures for future calibration of SPFs.

Following the format in HSM Volume 2 Part C, Appendix A.1 — Calibration of the Part C Predictive Models, the guidelines for future HSM calibration at SDDOT have been developed. The guidelines include an example calibration factor calculation. Additionally, information for SDDOT data sources as well as procedures for data preparation and reduction are presented in detail.

5.12. Prepare Final Report

Prepare a final report summarizing research methodology, findings, conclusions, and recommendations.

A draft final report has been prepared documenting project results, findings, data requirements, methodologies, conclusions, and recommendations. This report was submitted to the SDDOT technical panel on February 8th, 2015. On March 23th, 2015, the research team received the panel's comments. A revised report was delivered to SDDOT on May 1, 2015.

5.13. Make Executive Presentation

Make an executive presentation to the SDDOT Research Review Board at the conclusion of the project.

The project PI will present the findings of the study, including a recommended methodology, comparative tables, and analysis to the SDDOT Research Review Board at the conclusion of the project. This presentation will summarize project activities and any conclusions or recommendations that emerged from this research project.

6. FINDINGS

6.1. Data Sources

Highway segment and intersection data along with historical crash information are required to calibrate the HSM predictive models. The roadway inventory data used in this study were supplied by SDDOT and the South Dakota Department of Public Safety.

6.1.1. Intersection Data

SDDOT provided the intersection data in two shapefiles (a popular geospatial vector data format for geographic information system (GIS) software): 1) "all intersection" and 2) "StatetoState_and_StatetoNon StateFedAid". The shapefile of all intersections in South Dakota has very limited information, and therefore is not adequate for HSM calibration. However, the "StatetoState_and_StatetoNonStateFedAid" shapefile includes all intersections between one state highway and another state highway and all intersections between one state highway and a federal-aid non-state highway. The dataset has information on the intersection's lane type (two-lane or multilane), traffic control type, AADT of major and minor roads, first and second skew angles, left and right turn lane counts, and existence of lighting.

6.1.2. Roadway Segment Data

Roadway segment data were obtained from the Roadway Inventory System (RIS), which is maintained by SDDOT. RIS has six (6) subsystems including Mileage Reference Marker (MRM), Roadway Features, Intersection Inventory, Traffic Inventory, GIS Data Extract, Maintenance Cost Inventory and RIS validation. Prior to requesting data items, a data requirement study was completed by reviewing both the previous literature and the HSM. Based on the data requirement list provided by the research team, SDDOT provided event tables for each highway, as shown in Table 6-1.

Table 6-1 Provided RIS Subsystem Event Tables and Key Variables

Event Table	Key Variables
MRM	Mileage, MRM
A - Admin System Data	City Code, Functional Class, Rural-Urban Code
B - Surface Property Data	Surface Type, Surface Width, Shoulder Width
C - Median Data	Rumble Strip, Median Type, Median Width
L - Horizontal Curve Data	Degree Curve, Deflection Angle, Spiral Length
N - Vertical Curve Data	Elevation, Grade Value, Length In/Out
P - Speed Limit Data	Speed Limit
Traffic Data	Current Annual Average Daily Traffic, Peak Hour Factor

6.1.3. Crash Data

Five-year crash data (2008 to 2012) were collected from SDARS, the database of motor vehicle crash information that provides the statistics necessary to identify problems, assist in countermeasures, and evaluate applied countermeasures to promote safe roadways. Multiple event tables depict crash information from all aspects, including Accident, Vehicle, and Person. The Accident table details information such as area type (rural/urban), crash type, crash location, etc. The information on the vehicle(s) and occupant(s) involved in the crash are contained in the Vehicle and Person table. These tables can be related by a common primary index called "AccidentSeqID".

6.2. Data Processing

To generate homogeneous highway segments or intersections that met the HSM requirements without creating extremely short roadway sections, the data event tables containing various highway and traffic information need to be processed, integrated, and reduced to a format that is appropriate for calibration. The crash data were joined to either the segment table or the intersection table to produce the crash count for each site. The following section introduces the data processing of the shapefile.

6.2.1. Intersection

The intersection shapefile provided by SDDOT can be used without processing, but intersection-related crashes need to be spatially joined to individual intersections to create an intersection crash count.

6.2.1.1. Intersection-Related Crash Assignment

A crash event can be joined to an intersection based on the spatial distance between the two items. The spatial join function provided in ArcMap requires both crash and intersection datasets to be in a shapefile format. Crash data in a CSV format can be converted to a point shapefile by importing its geographic coordinate information (i.e., latitude and longitude fields) with the "Add XY Data" function. After successfully importing the crash data, intersection and crash locations can be joined based on their proximity.

Figure 6-1 illustrates the intersection-related crashes in relation to state highway intersections. In the figure, intersections on state highways are marked by red and blue dots, and only intersections marked by red dots are included in the scope of this study. Red dots represent intersections with approaches as either a state highway or a highway included in the Non-State Trunk Road Inventory (NSTRI) with federal aid. Blue dots represent intersections with the other approach as a highway included in Non-State Trunk Road Inventory (NSTRI) without federal aid, such as highways administrated by towns or counties. This figure also illustrates how crashes were assigned to each intersection in the study. Each circle has a different radius, 100 feet and 200 feet, respectively. Red crosses represent intersection-related crashes and blue crosses represent non-intersection-related crashes. Crashes within 100 feet of the center of intersection are treated as intersection-related. Crashes within 200 feet of the center of the intersection with an ILTIntersectFlag as "TRUE" are also treated as intersection-related. All crashes outside of the 200 feet radius are not considered intersection-related.

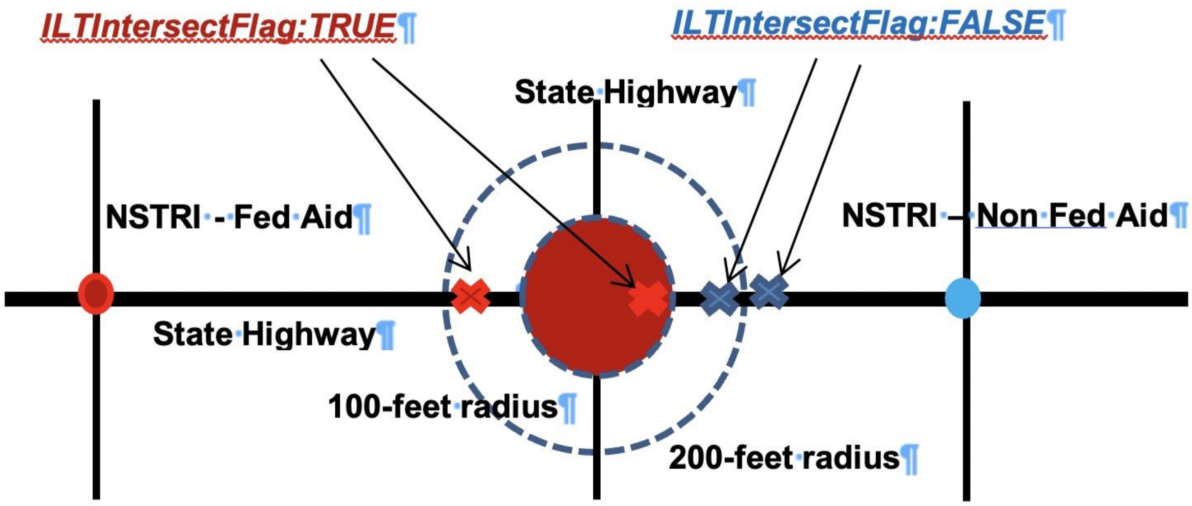


Figure 6-1 Illustration of Intersection-Related Crashes

Although the spatial join function connects the two locations based on the "closest" distance, it is important to know that one intersection can be closest to more than one crash. Therefore, when joining crash locations to intersections, it is possible for a crash to be joined to multiple intersections, resulting in over-counting crashes. To ensure a crash is uniquely joined to an intersection, the crash and intersection shapefile should be defined as "Target Features" and "Join Features," respectively. In other words, the intersection attributes are joined to a crash. The "CLOSEST" option is used as the joining rule. "Search Radius" is defined as 200 feet. When the spatial join operation is completed, a new field called "Distance Field Name" is created to measure the distance between the target feature and the closest joined feature. It is convenient to post-process the data if necessary, as all the key crash attributes (e.g., rural/urban, animal collision, intersection flag, injury severity, manner of collision, etc.) are maintained. Table 6-2 presents the relationship between the distance and the intersection flag for all crashes that occurred within 200 feet of an intersection in a rural area.

Table 6-2 Crash Counts by Distance and Intersection-Related Flag

Closest Distance	ILTIntersect Flag TRUE	ILTIntersect Flag FALSE	Total
100 feet ≤ Distance ≤ 200 feet	59	17*	76
Distance ≤ 100 ft.	155*	572*	727*
Total	214	589	803
*SDDOT definition of intersection-related crashes.			

According to SDDOT's definition of intersection-related crashes, 744 crashes were included in the HSM calibration.

6.2.2. Segment

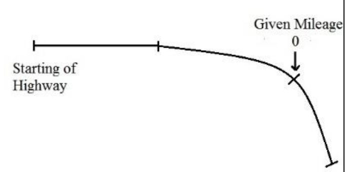
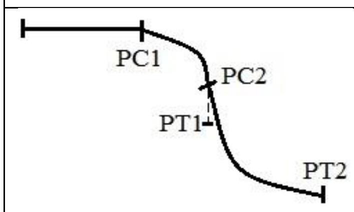
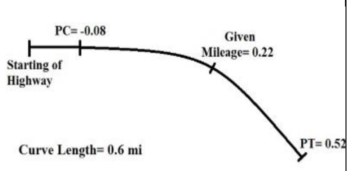
Processing highway segment data are more complicated than processing intersection data because of the definition of homogeneity. Various highway geometric attributes required by the HSM are stored in several event tables in RIS. To combine all the roadway information required for the HSM calibration into one dataset, these event tables need to be merged. The following sections discuss the problems encountered, the actions taken, and the major steps taken to convert the event tables that were eligible for merging.

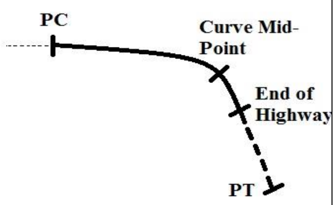
Four clarifications should be made before proceeding with the data processing:

- Rural roads are defined by “City Code” in the Admin dataset.
- The “mileage” of the last segment can be calculated by the MRM dataset.
- The “mileage” field in the horizontal curve file is the midpoint of each horizontal curve.
- The “mileage” field in the vertical curve file is the point of vertical intersection (PVI) of a vertical curve.

6.2.2.1. Horizontal Curve

Table 6-3 List of Problems and Actions

Problem	**Description**	**Action**
	Highway starting with a curve with given zero (0) mileage. There are 12 highways have this kind of data issue.	Considered half of the length of curve when the curve is starting with zero mileage.
	Overlap of horizontal curves. There are 342 pairs of curves (684) which are overlapping.	Equally divide the overlapped length into both of the curves.
	The PC of first curve is negative. This situation can occur when a highway starts with a tangent and the following curve has a large curve length. [PC= Given mileage- (curve length/2)]. There are 20 cases.	The starting tangent was neglected and the PC of the curve was considered at zero (0) mileage. Now, the maximum curve length changes from 11.04 miles to 4.155 miles and the average curve length is 0.214 mile as oppose to 0.226 mile before.

Problem	Description	Action
	The calculated PT of the last curve exceeds the maximum mileage given in MRM file. There are 17 highways with calculated PT exceeding the maximum mileage.	PT of last curve was considered as maximum mileage given in MRM file. PT will not be calculated using curve length but be considered as the highest given mileage for the highway.

In previous studies, researchers investigated the safety effect of horizontal curve and grade combination on rural two-lane roads (*19*) and the impact of spatial relationship on horizontal curve safety (*20*). Based on the recommendations from the HSM and previous studies (*19; 20*), threshold values below were used to exclude extremely short segments and low curvature horizontal curves. The threshold values are as follows:

1. The threshold value used as minimum degree of curve is 1 (degree). As we processed the data table to create segments with continuous mileage, curves with a less than 1 degree of curve were considered as tangents instead of curves.
2. Bauer et al. noted in her study, curves with less than 0.01 miles were extremely short segments that can represent horizontal feature of the segment (14). Another reason for choosing 0.01 mile as the minimum length is that hundredth of a mile is usually used as the distance unit for locating a crash. Any curve less than 0.01 mile long was considered a tangent and was then merged with an adjacent tangent in the processed table.

All curves and tangents were calculated using mileage value with six digits after the decimal point. After calculating the mileage information and length of each curve and tangent, the mileage and length were rounded to three digits after the decimal point to remain consistent with the other data tables. Considering all the above recommendations, the number of horizontal curves was 3,667 and the number of tangents was 3,608.

6.2.2.2. Vertical Curve

Similar to the horizontal curve table, the vertical curve is a point file that contains both MRM and mileage values for each vertical curve. The given mileage is the Point of Vertical Intersection (PVI) of each vertical curve. The table has 29,720 observations, 2,111 of which have zero length.

The Vertical Curve file was converted to a table with continuous mileage, and had the same types of issues as the horizontal curve event table. For example, twelve curves had negative mileage information for the beginning of the curve. 3,203 pairs of curves overlapped. Treatments for these issues were the same as the recommendations for the horizontal curve.

The HSM recommends a new segment start at the PVI of each vertical curve so that each segment should have a uniform grade value. Considering this recommendation, each vertical curve was split at the PVI. The sign of the grade (e.g., + is upgrade and – is downgrade) was not critical because when the grade in one direction is positive, it must be negative in the other direction for two-way traffic, therefore offsetting any overall positive or negative safety effects.

The HSM also recommends a 3% minimum grade for any vertical curves. In this study, any grade less than 3% was considered to be level. Similar to horizontal curves, 0.01 mile was considered as the minimum threshold value for vertical curve length, resulting in 7,077 grade segments and 3,421 level segments.

6.2.3. Traffic Data

The traffic data table contains 1,867 observations, and each has an MRM value. All event tables include mileage, and thus mileage information was used to merge tables needed to create a comprehensive calibration database. The original traffic data table did not include mileage information, so a mileage check was conducted to retrieve the mileage value for any given MRM in the table. Among the 1,867 given MRM values, all mileage values could be retrieved from other tables. Figure 6-2 illustrates the data table clipping procedure in a finer scale, also called dynamic segmentation.

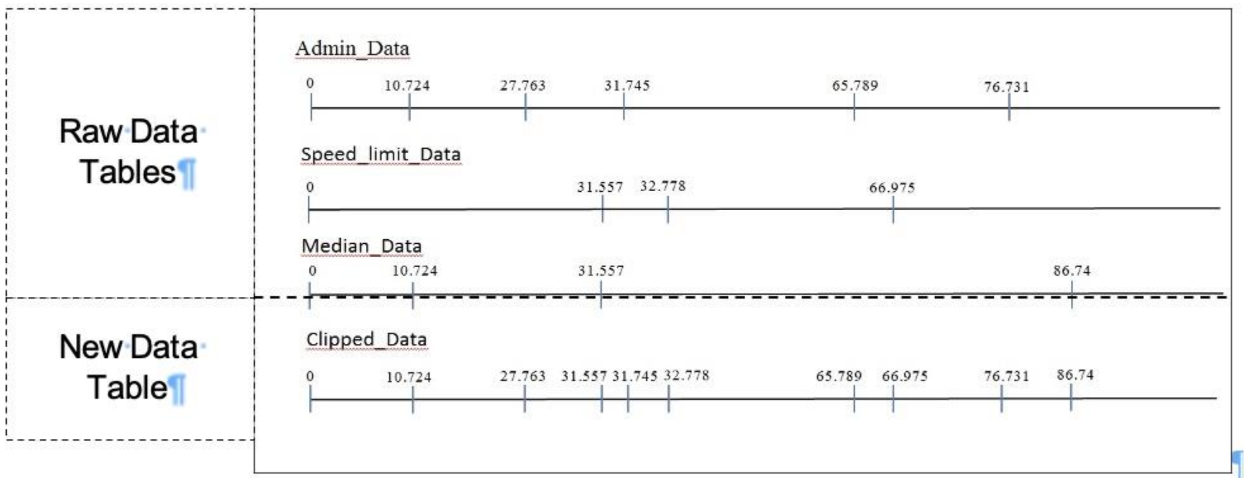
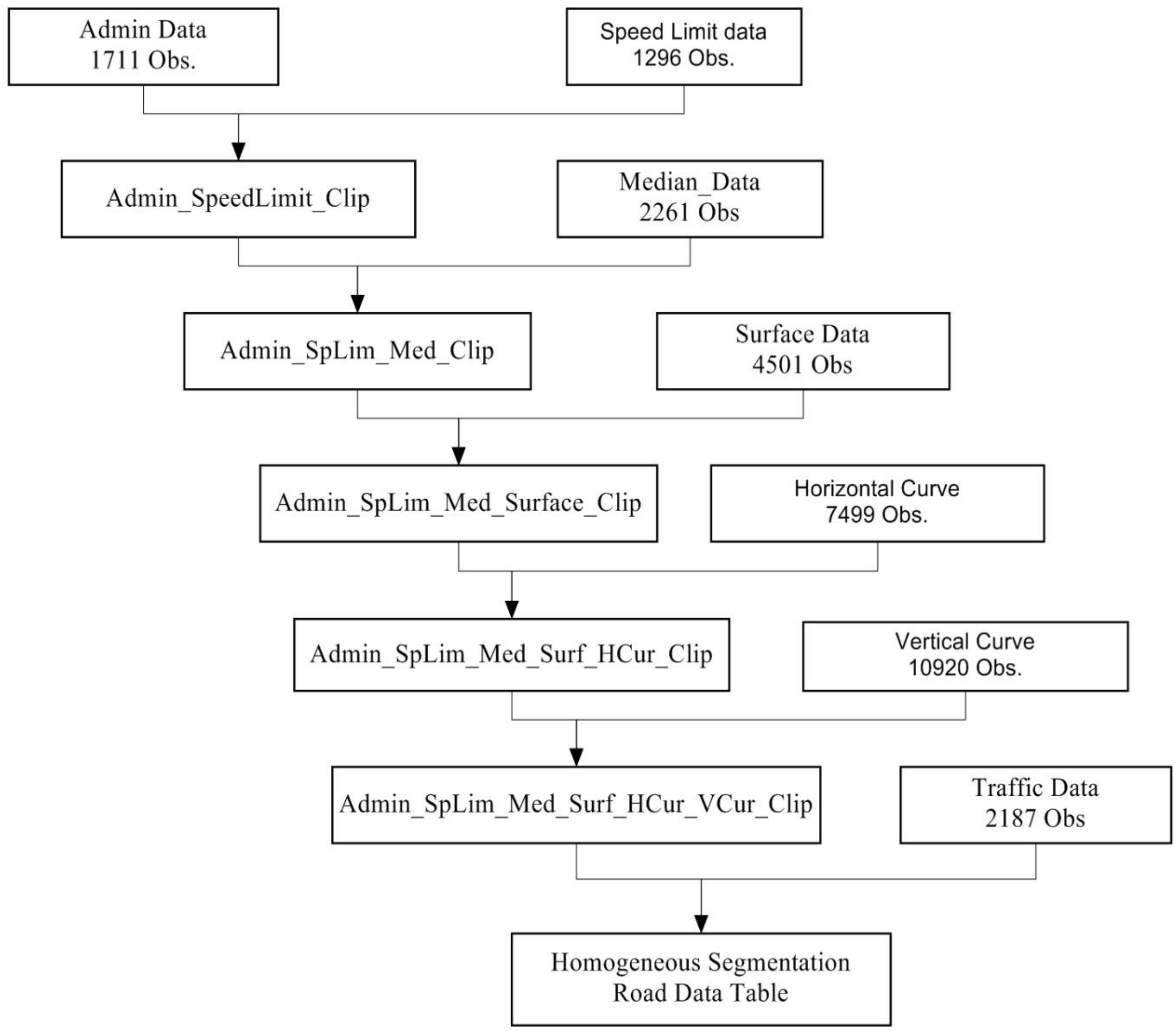


Figure 6-2 Event table merging technique with mileage information

6.2.4. Event Tables Merging Process

By clipping two or more data tables based on mileage value, the clipped file has created a new segment on each unique mileage available in the data tables used for clipping. The sequence of the merge was important, as different sequences may produce a different number of segments. The merging principle proceeded from coarse to fine (see Figure 6-3) to avoid creating short segments at the starting of merging sequence and improve computational efficiency.



[Note: Horizontal and Vertical curve file in this chart represents the processed event table for clipping]

Figure 6-3 RIS Event Table Merging Sequence

Figure 6-3 illustrates the data table merging sequence to obtain homogeneous roadway segment table. As discussed earlier, the main rationale considered for table merging went from coarse to fine to avoid creating short segments. The table merging scripts were written in such a way that it can only merge two separate data tables at once. In the RIS inventory data tables, “A-Admin System Data” and “P-Speed Limit Data” had the lowest number of observations to represent state highway roadway network. This means that both were coarse data tables in the RIS data inventory. So, the merging sequence starts with “A-Admin System Data” and “P-Speed Limit Data” data tables. After merging the two files, it followed the descending order of number of observations in one data table for merging other data tables. The traffic data table was merged at the end of the sequence because it does not have the mileage value. For this reason, the traffic data table was merged with other data table based on “MRM” value. The final roadway table is generated by clipping event tables from the RIS subsystems based on the mileage value. Each observation in this dataset has uniform attribution. Each highway is segmented according to the change in the roadway features in different data tables. The segment length is calculated using the mileage value. A roadway shapefile was created based on this table by using the linear referencing system (LRS) at SDDOT.

6.2.5. Lane Width Calculation

One of the important data items to apply HSM predictive models is the lane width. The lane width information was not directly available in any of the RIS inventory data tables. The lane width can be simply calculated by dividing surface width by number of lanes. Surface width information was available in "B-Surface Property Data" as "SurfaceWidthNbr" (Item 66) and number of lane information was available in "C-Median Data" as "NumberLanesNbr" (Item 87). After joining the data tables from RIS inventory, both surface width and number of lane information was available in homogeneous segmentation road data table. The formula used to calculate lane width can be expressed as follows:

$$Lane\ Width = \frac{SurfaceWidthNbr}{NumberLanesNbr} \tag{6-1}$$

6.2.6. Segment Crash Assignment

All crashes that occurred on state roads outside the urban boundary were filtered from the original crash dataset to be linked with the roadway shapefile. ArcMap was used to assign crashes to each segment based on their spatial distance. Similarly, to avoid the situation of joining a crash to multiple roadway links, the crash dataset was set as target features and the roadway shapefile was set as joined features in "Spatial Join". Each crash was linked with its closest segment. All key crash characteristics (i.e., area type, animal collision, injury severity, manner of collision, etc.) were maintained which made it more convenient to run attribute-specific queries. Only 0.35% (or 46 out of 13,337) crashes have the distance from the centerline of the roadway longer than 51 feet. Therefore, a 50-foot buffer distance was assumed for each crash when finalizing the linkage between the crash dataset and the roadway dataset because crashes that are more than 50 feet from the centerline are probably wrong. All crashes, including animal collisions, within a 50-foot buffer distance of each roadway segment were linked with that segment. Figure 6-4 illustrates the join procedure.

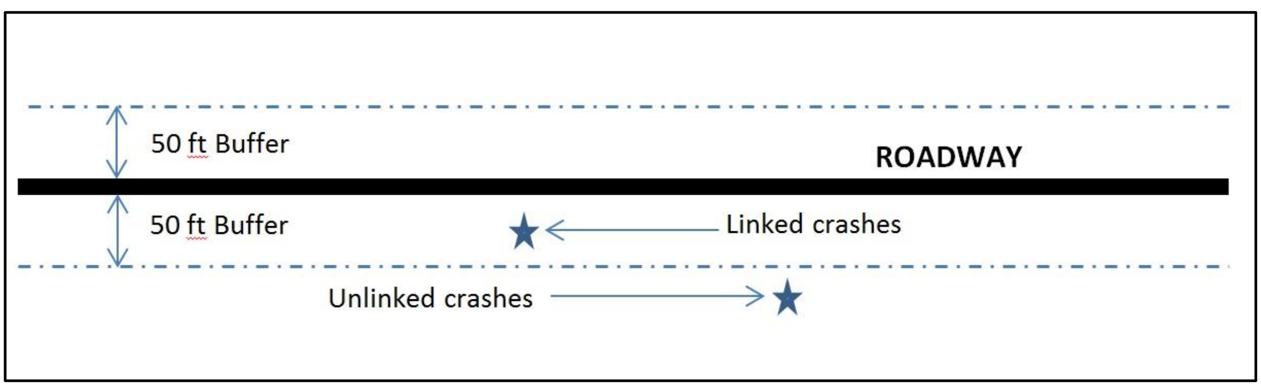


Figure 6-4 Illustration of Buffer Distance Used to Link Crashes with Segments

Following the following steps, the database was condensed even further:

- The total roadway file which included both rural and urban roads was used in the joining stage, meaning a few of the crashes linked with those segments have a city flag. After removing those crashes, the total number of joined segment crashes is 21,767.

- Some of the roadway segments were specified as being state highways divided by one lane. Those segments included 22 crashes and were 8.861 miles long. The HSM does not have such a category, so those segments were removed for consistency.

6.3. HSM Calibration

In this section, the calibration factors were calculated using the HSM procedures that have not been modified. The HSM recommends following certain steps to calibrate the HSM predictive models. For both rural two-lane two-way and multilane facilities, there are 6 steps to follow when calibrating HSM predictive models. A summary of these steps is illustrated in Figure 6-5.

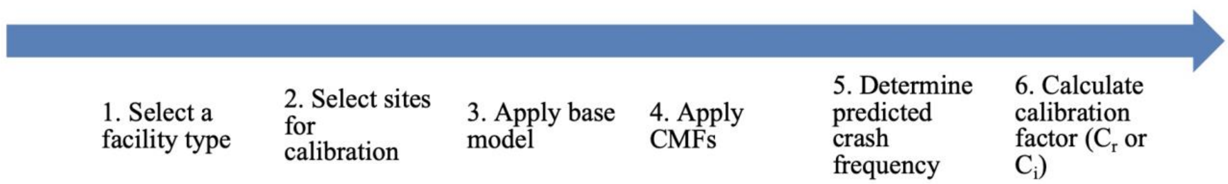


Figure 6-5 HSM Calibration Procedure

6.3.1. HSM Predictive Methods

For any facility type, the HSM provides the following predictive model equation:

$$N_{predicted} = N_{spf} \times C \times (CMF_1 \times \cdots \times CMF_n) \quad (6\text{-}2)$$

where $N_{predicted}$ is the predicted average crash frequency for a site, N_{spf} is the predicted average crash frequency for the base conditions for a site, which is also called safety performance function (SPF). C is the calibration factor (Cr is for a roadway segment and Ci is for an intersection). A series of crash modification factors (CMFs) account for changes in the number of crashes due to site characteristics or safety treatments that are different from base conditions.

The base conditions need to be identified to calculate the SPF. Tables in the appendix present the base conditions, SPF functions, and CMF list for each facility type.

Table 6-4 and

Table 6-5 list the availability of CMF data in South Dakota.

Table 6-4 List of CMFs for Rural Two-lane Two-way Highway Facilities

Type	CMF	Description*	Data Availability
Roadway Segments	CMF_{1r}	Lane Width	X
Roadway Segments	CMF_{2r}	Shoulder Width and Type	X
Roadway Segments	CMF_{3r}	Horizontal Curve	X
Roadway Segments	CMF_{4r}	*Horizontal Curve Superelevation*	

Roadway Segments	CMF_{5r}	Grades	X
Roadway Segments	CMF_{6r}	*Driveway Density*	
Roadway Segments	CMF_{7r}	*Centerline Rumble Strip*	X
Roadway Segments	CMF_{8r}	*Passing Lanes*	
Roadway Segments	CMF_{9r}	Two-way left-turn Lanes	
Roadway Segments	CMF_{10r}	*Roadside Design*	
Roadway Segments	CMF_{11r}	*Lighting*	
Roadway Segments	CMF_{12r}	*Automated Speed Enforcement*	Not applied
Intersections	CMF_{1i}	*Intersection Skew Angle*	X
Intersections	CMF_{2i}	Intersection Left-turn Lanes	X
Intersections	CMF_{3i}	Intersection Right-turn Lanes	X
Intersections	CMF_{4i}	Lighting	X

Table 6-5 List of CMFs for Rural Multilane Highway Facilities

Road Type	CMF	Description	Data Availability
Rural Multilane Undivided Segments (RM4U)	CMF_{1ru}	Lane Width	X
Rural Multilane Undivided Segments (RM4U)	CMF_{2ru}	Shoulder Width and Type	X
Rural Multilane Undivided Segments (RM4U)	CMF_{3ru}	*Side Slopes*	X
Rural Multilane Undivided Segments (RM4U)	CMF_{4ru}	Lighting	
Rural Multilane Undivided Segments (RM4U)	CMF_{5ru}	*Automated Speed Enforcement*	Not applied
Rural Multilane Divided Segments (RM4D)	CMF_{1rd}	Lane Width	X
Rural Multilane Divided Segments (RM4D)	CMF_{2rd}	Right Shoulder Width	X
Rural Multilane Divided Segments (RM4D)	CMF_{3rd}	*Median Width*	X
Rural Multilane Divided Segments (RM4D)	CMF_{4rd}	Lighting	
Rural Multilane Divided Segments (RM4D)	CMF_{5rd}	*Automated Speed Enforcement*	Not applied
Rural Multilane Intersections	CMF_{1i}	*Intersection Skew Angle*	X
Rural Multilane Intersections	CMF_{2i}	Intersection Left-turn Lanes on Major Road	X
Rural Multilane Intersections	CMF_{3i}	Intersection Right-turn Lanes on Major Road	X
Rural Multilane Intersections	CMF_{4i}	Lighting	X

6.3.2. HSM Calibration Procedure

The calibration procedure recommended in the HSM was followed when calibrating predictive models for both segments and intersections. The procedure involved five steps:

Step 1—Identify facility types for which the applicable Part C predictive model is to be calibrated.

Three facility types were identified: rural two-lane two-way segments (RT), rural multilane undivided segments (RM4U), and rural multilane divided segments (RM4D). In this study there are five intersection types: rural two-lane three-leg intersections with STOP control (RT3ST), rural two-lane four-leg intersections with STOP control (RT4ST), rural multilane three-leg intersections with STOP control (RM3ST), rural multilane four-leg intersections with STOP control (RM4ST), and rural multilane four-leg signalized intersections (RM4SG).

Step 2—Select sites for calibration of the predictive model for each facility type.

The HSM suggests that the desirable minimum sample for each facility type be 30 to 50 sites and have more than 100 crashes per year. Therefore, a desirable total would consist of more than 500 crashes in five years. The HSM also suggests using all available sites for calibration if there are fewer than 30 sites for a specific type.

The numbers of rural multilane undivided highway segments were 1,211 with 940 crashes. All the facility types fulfilled the HSM recommended number of sites and number of total crashes. Within each facility type, all segments were considered to calculate the calibration factor.

Among all intersection types, only RT4ST had more than 500 crashes in five years. It was more accurate to use the entire dataset for calibration than using a sample that could harbor selection bias. Hence, for all facility types, the entire dataset was used for calibration. Although RM4SG had only one site, the calibration was still developed for reference.

Step 3—Obtain data applicable to a specific calibration period for each facility type.

The observed crash frequencies from 2008-2012 were obtained for both the segments and the intersections under consideration. Most of the data items required by the HSM for calibration were available for each segment, referred to in Table 6-4 and

Table 6-5. A unit value of one was used for unavailable CMFs. The intersection data contained all necessary information for applying the predictive models.

Step 4—Apply the applicable Part C predictive model to predict total crash frequency for each site during the whole calibration period.

The SPF equations provided in the HSM were used to calculate the total crash number prediction for the base conditions. The CMFs were applied to the predicted total to account for the conditions that deviate from the base conditions.

Step 5—Compute calibration factors for use in Part C predictive model.

The following equation was used to compute the calibration factors for each facility type:

$$C_r \ or \ C_i = \frac{\sum_{all\ sites} observed\ crashes}{\sum_{all\ sites} predicted\ crashes} \qquad (6\text{-}3)$$

6.3.3. HSM Calibration Factor

The calibration procedure was conducted for each facility type. The calibration results for both segments and intersections were summarized in Table 6-6.

Table 6-6 Calibration Results for Segments and Intersections

Segment/Intersection	Facility Type	Number of Facilities	Crashes Observed (2008-2012)	Crashes Predicted (5 years)	Calibration Factor
Segments (miles)	RT	6,362	10,418	8,861	1.18
Segments (miles)	RM4U	152	940	822	1.14
Segments (miles)	RM4D	634	1,791	1,139	1.57
	Grand Total	7,149	13,149	10,822	1.22
Intersections (number)	RT3ST	337	170	309	0.55
Intersections (number)	RT4ST	582	415	1,276	0.33
Intersections (number)	RM3ST	19	26	19	1.36
Intersections (number)	RM4ST	71	125	222	0.56
Intersections (number)	RM4SG	1	2	52	0.04
	Grand Total	1,010	738	1,878	0.40

In summary, about 81 percent of the segments are rural two-lane two-way. Crashes occur on South Dakota rural state highway segments at an average 22 percent higher rate than what is predicted by the HSM. Although the calibration factors vary between different types of highway facilities, the deviation is relatively small, ranging from 1.14 to 1.57. It was found that crashes on rural state highway intersections occur 60 percent less than what is predicted by the HSM. The calibration factors, however, change drastically among different types of intersections. RM3ST has a calibration factor of 1.36, whereas the calibration factor for RT4ST is only 0.33. Note that the calibration reliability can be affected by a small sample size. There are only 19 RM3ST facility types and one signalized intersection in this study; therefore, the calibration factors for these types of facilities may not be reliable.

6.4. Analysis and Discussion

Although the calibration factor has been estimated for each state highway facility type in South Dakota, individual site performance is unknown. It is possible that the calibration factor may vary considerably among the sites within the same facility type; if so, new information that may explain the within-facility variability should be introduced to further classify the current calibration factor.

A logical starting point is to review the calibration factor for each site over the entire state highway system and identify any spatial patterns that may support a regional stratification. The regional factor is often considered the surrogate measure for conditions other than those used in the HSM predictive methods (e.g. weather, animal population, terrain, crash reporting threshold and criteria). Next, several roadway attributes that are included in RIS are not included in the HSM predictive methods (e.g. posted speed limit, the degree of curve, grade); these factors contribute to crash occurrence and they might also explain variability within the same facility. Lastly, the effect of the variables already included in the HSM predictive methods (e.g. AADT, lane width, shoulder width) should be evaluated because the HSM SPFs and CMFs were not developed with local data. Furthermore, the quantitative relationship presented as the coefficients of SPF does not necessarily hold in South Dakota. The following analysis and discussion provide factual information and statistical evidence to support any modification to the current HSM calibration.

6.4.1. Geographic Distribution of Calibration Factors

The calibration factor of each intersection was calculated and mapped in

Figure 6-6. The size of the circle denotes the calibration factor value. Most calibration factors are lower than one, suggesting that the HSM predictive method overestimates crash frequency for most intersection sites in South Dakota. Moreover, no visible spatial clusters or patterns can be discerned (e.g. East of the Missouri River vs. West of the Missouri River, or in the vicinity of large cities such as Sioux Falls and Rapid City). It is concluded that a region-specific model may not be effective in distinguishing different safety performances among intersections.

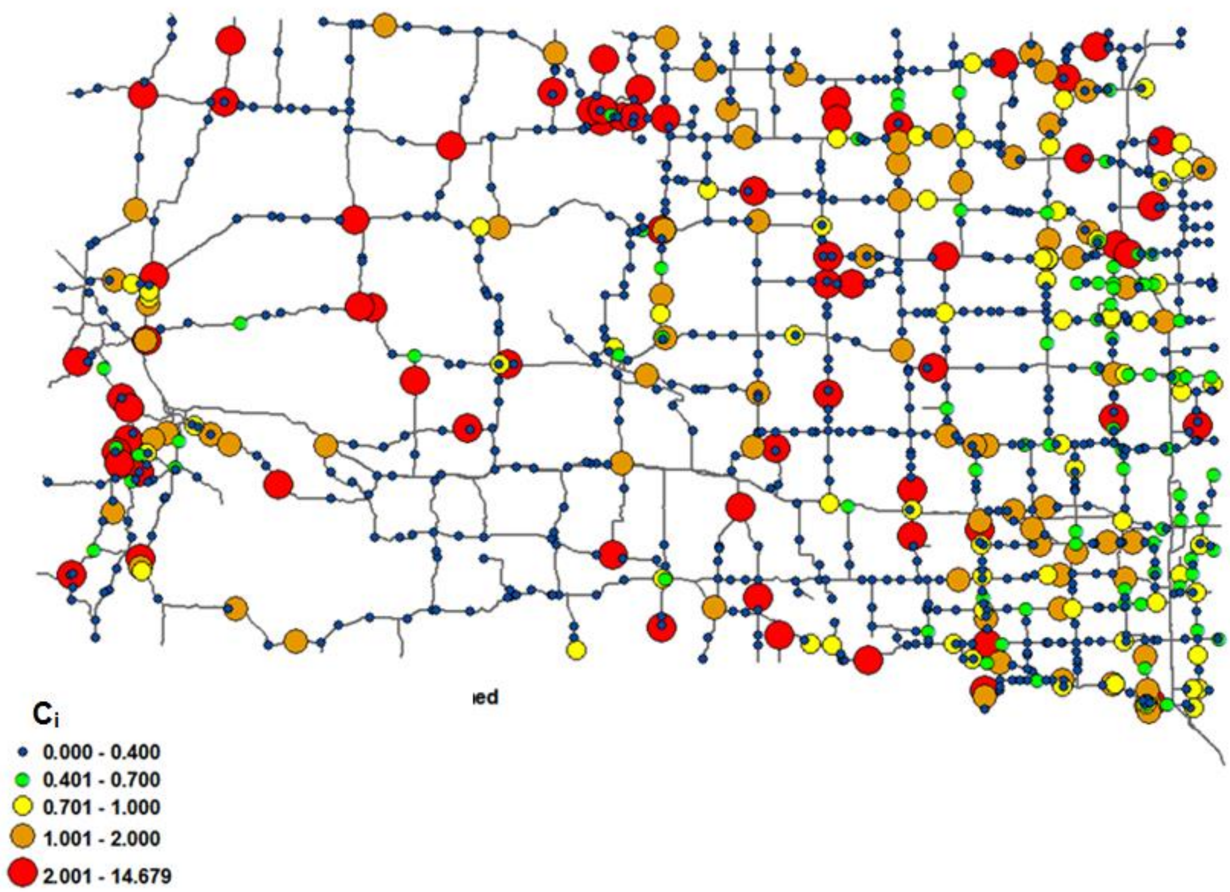


Figure 6-6 Distribution of Intersection Calibration Factor

Figure 6-7 presents the map of calibration factors for each highway segment. Unlike intersections, the spatial distribution suggests there may be two distinctive zones for segments in South Dakota. In western South Dakota, the predicted number of segment crashes is overestimated, whereas it is underestimated in eastern South Dakota. The denser highway network in southeastern South Dakota may contribute to the average calibration factor of 1.22. Despite noticeable patterns, the mixed results indicate that it can be challenging to draw boundaries between regions.

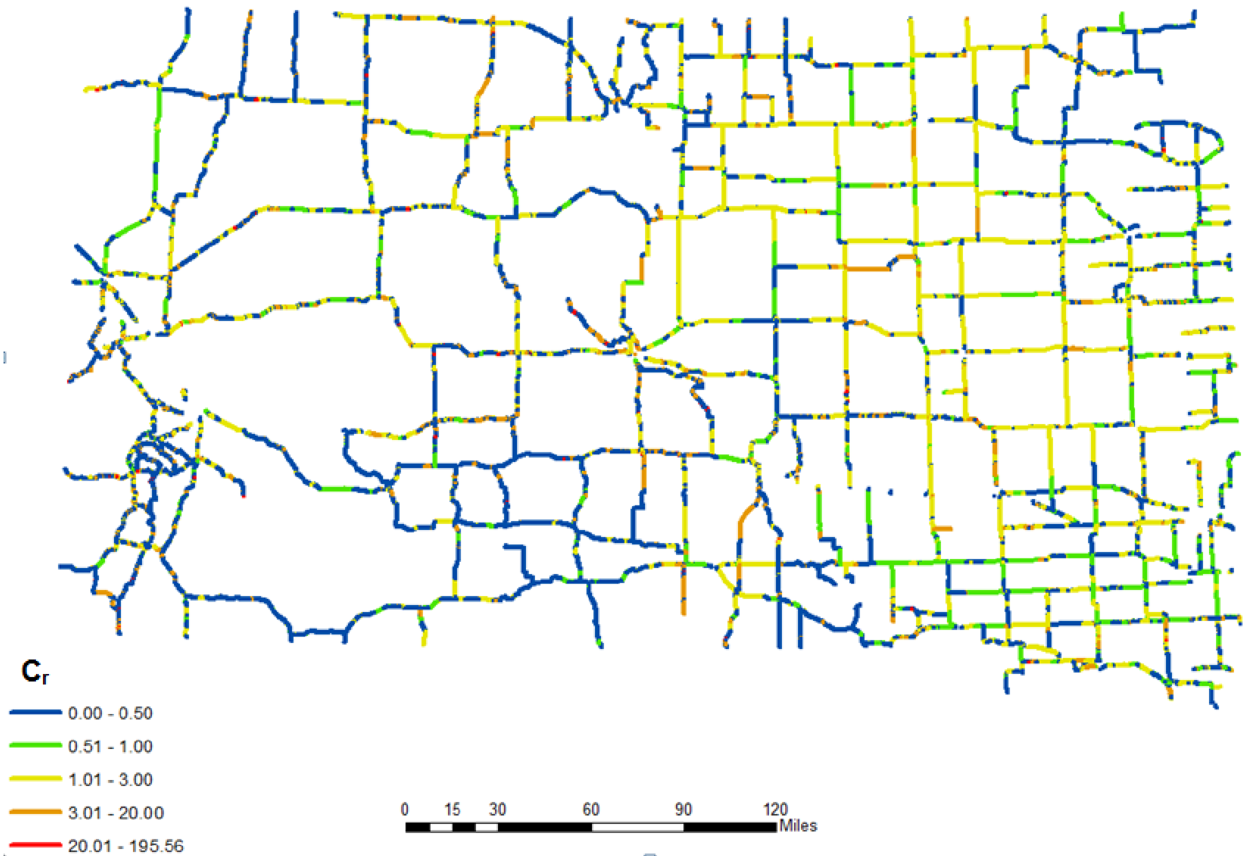


Figure 6-7 Distribution of Segment Calibration Factor

The calibration factor maps reveal a large disparity between intersections and segments. Compared with intersection crashes, segment crashes disclose some features that may be attributable to regional effects (i.e. weather, terrain, animal population). However, it is unclear on how much better a region-specific SPF will be in comparison to a state SPF and the challenge of defining explicit boundary lines between areas is daunting due to the lack of information and in-depth analysis. Dividing the state by geographic areas and developing region-specific SPFs were not further pursued in this project.

6.4.2. Factors Contributing to Calibration

Spatial analysis indicates whether or not the dataset should be divided to better represent the geograhic disparities. The difference between observation and prediction reveal whether or not the relationship between the predicted crash frequency and the explanatory variables should be changed for higher accuracy. Namely, adding new variables to the current SPFs or calibrating SPFs with local data. In this exploration, “residual” was used to examine which factors contribute to the difference between observation and prediction, which was formulated as the observation minus the prediction multiplied by Cr (or Ci). It is a measure for within-facility variation and indicates how dispersed the data can be.

Beginning with intersections, four types of intersections were considered: RT3ST, RT4ST, RM3ST, and RM4ST. RM4SG was not included because there is only one site. The normal Q-Q plots for residuals in Figure 6-8 present two remarkably different patterns for two-lane two-way highways and multilane highways, respectively. Note that the Q-Q plot shows the residual quantiles against the theoretical normal quantiles. When the plots are generally aligned on a straight line, the residuals approximately follow the normal distribution. Denoted by green and orange dots in Figure 6-8, residuals of intersections on two-lane facilities are not normally distributed, but are skewed to the right. This pattern suggests that the HSM predictive model for rural two-lane intersections does not competently account for data skewness. In this event, a prediction model has to be calibrated with local data. Denoted by purple or blue dots in Figure 6-8, residuals of intersections on multilane facilities approximately formed a straight line, suggesting strong normality. The HSM prediction models for multilane intersections can effectively correct the skewness in crash data and do not require a new model.

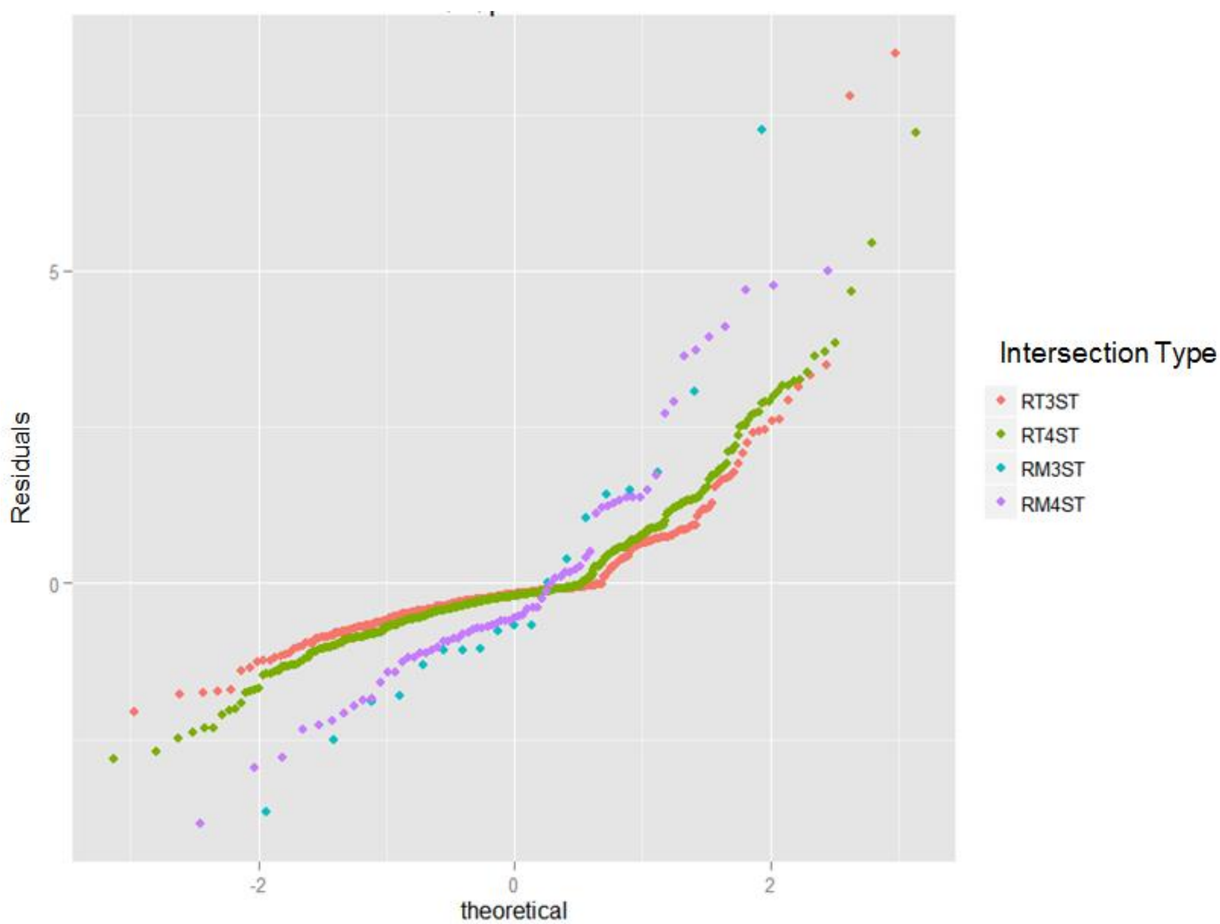


Figure 6-8 Normal Q-Q Plot for Residuals by Intersection Type

Figure 6-9 shows the correlation between the residual and intersection characteristics. According to the correlation matrix, none of the correlation coefficients between the residual and highway characteristics are large enough to warrant a significant linear dependence. Among them, the number of left turn lanes has the highest correlation coefficient. Different lighting conditions have different residual distributions. Both prompt further analysis.

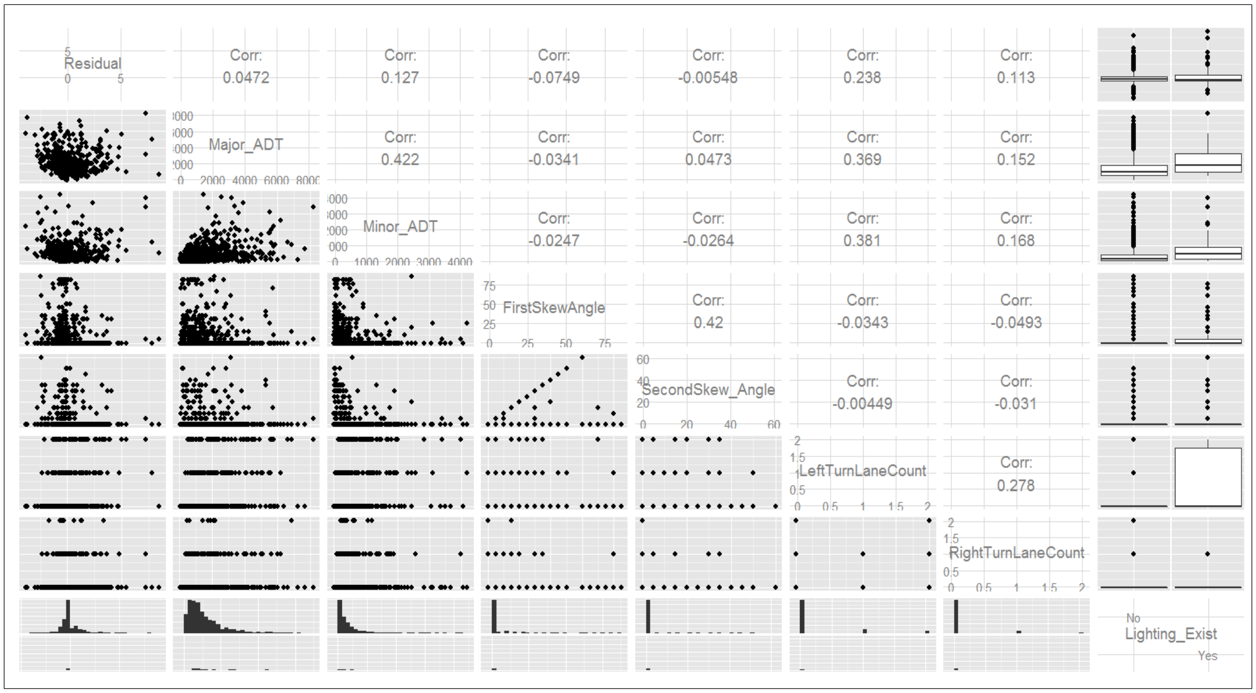


Figure 6-9 Correlation between Residual and Intersection Characteristics

Since the left turn lane count is a categorical variable, the boxplots were generated by intersection type in Figure 6-10. At first glance, the HSM prediction generally overestimates the crash frequency for base condition intersections (i.e., without left turn lanes). A closer review indicates that the residuals increase from negative to positive as the number of left turn lanes increases. This is similar to the change that takes place when the HSM prediction goes from underestimation to overestimation. Note that there is already a left-turn lane CMF in the HSM that declines from one as the number of left-turn lanes increases.

The boxplots of residuals for lighting conditions in Figure 6-11 were produced to depict the relationship between the residuals and lighting conditions. Again, the HSM prediction generally overestimates the crash frequency for intersections with base conditions (i.e., no light). A significant difference is not present between the "no light" condition and "light" condition on rural two-lane highway intersections, but is present at multilane highway intersections, which may be attributed to the randomness caused by the small sample size for multilane intersections.

Since the number of left-turn lanes has been taken care of by a CMF in the HSM and few sites were available in the multilane intersections, the adjustment for the number of left-turn lanes and light condition at intersections was not pursued further in this project.

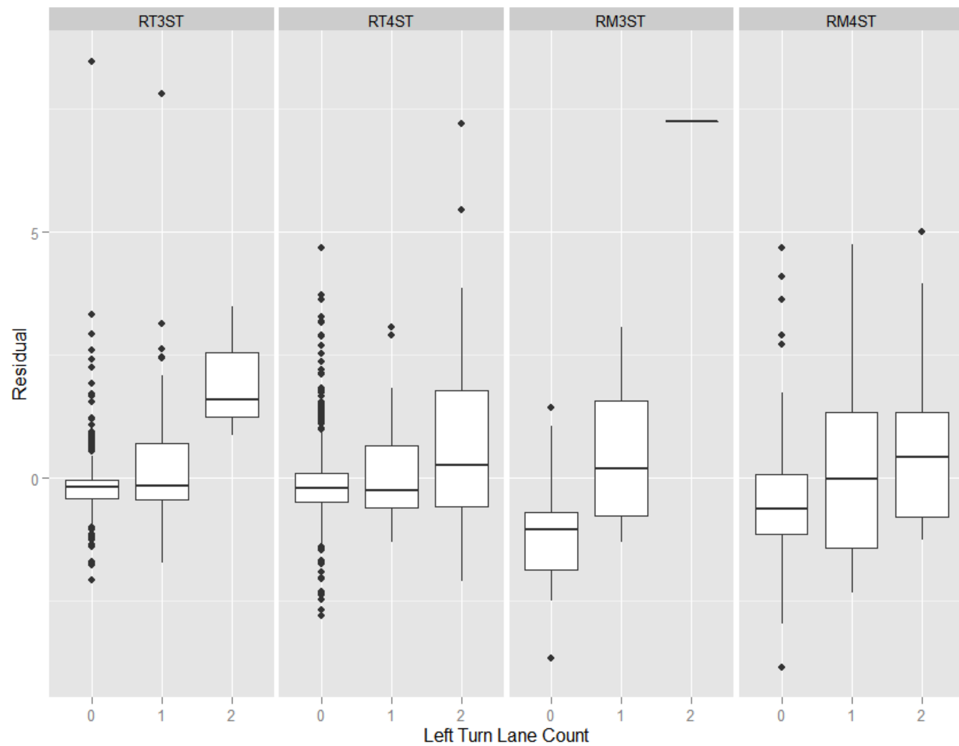


Figure 6-10 Boxplot of Residual by Left Turn Lane Count

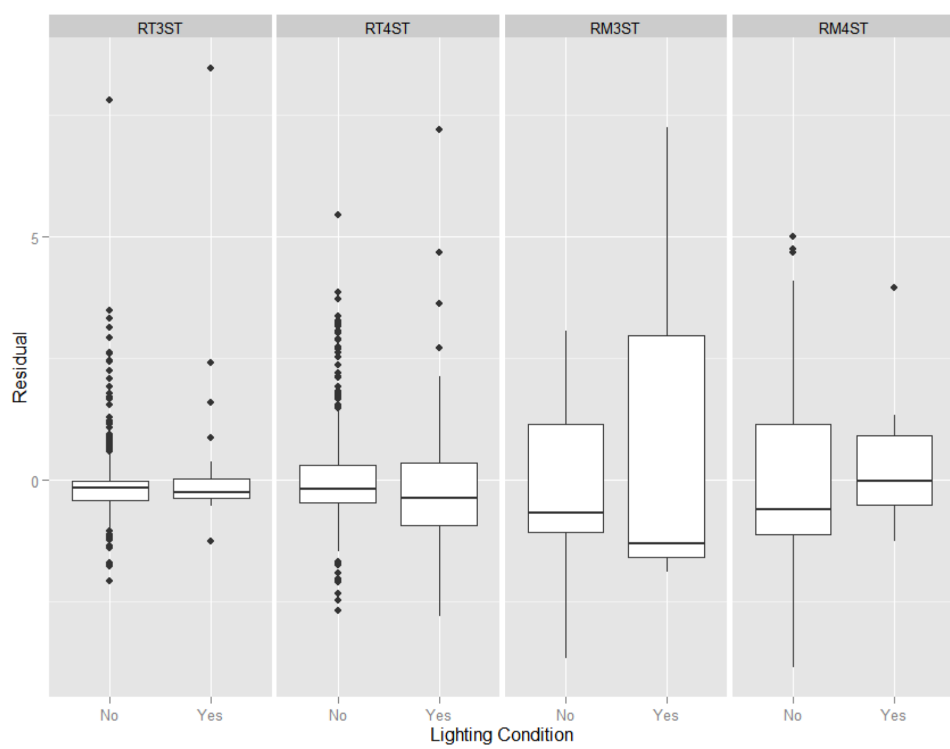


Figure 6-11 Boxplot of Residual by Lighting Condition for Different Intersection Types

Similar analyses were conducted on rural highway segments for RT, RM4U, and RM4D. First, the correlation matrix plots between residual and segment characteristics were produced in Figure 6-12, Figure 6-13, and Figure 6-14. The four largest correlation coefficient values between residual and segment characteristics are marked by a red circle. None of the correlation coefficients is larger than 0.3, indicating a weak linear correlation between residual and segment characteristics.

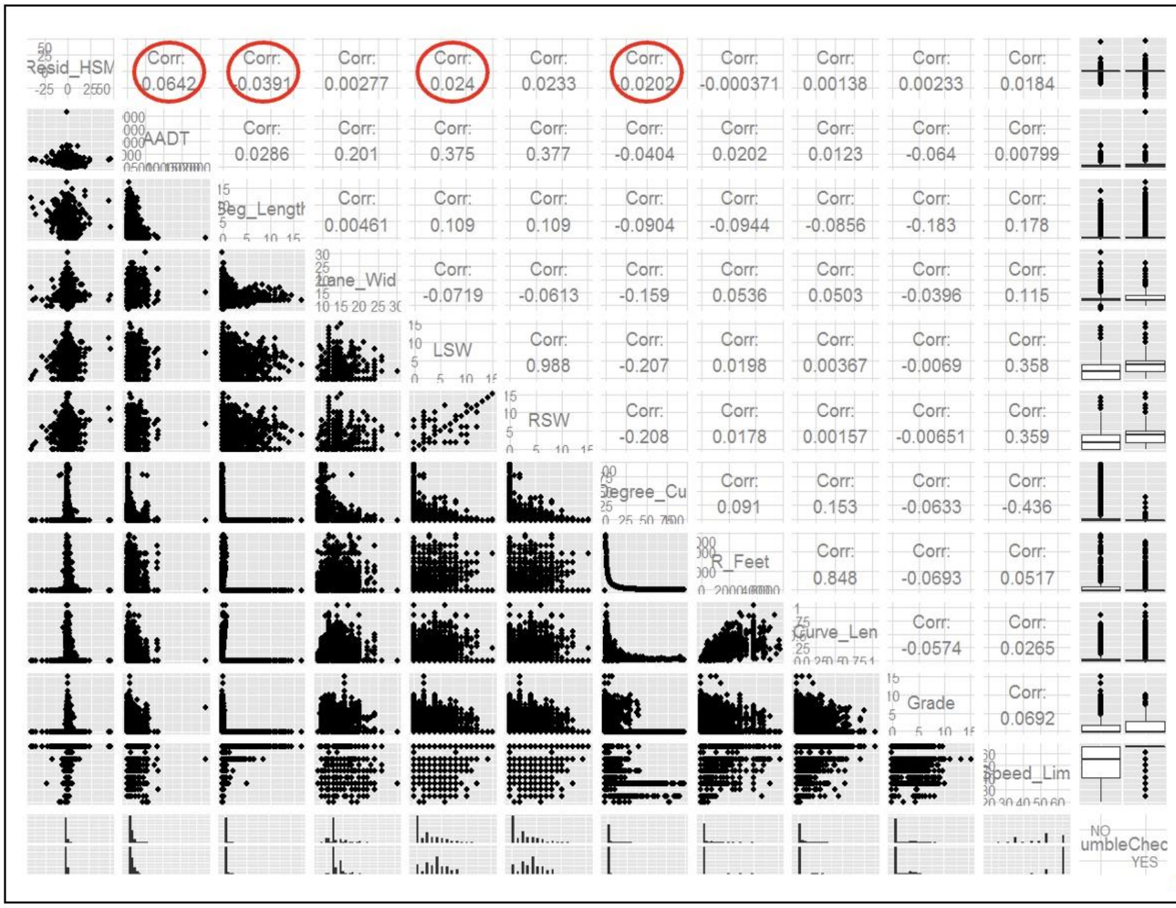


Figure 6-12 Correlation between Residuals and Segment Characteristics (RT)

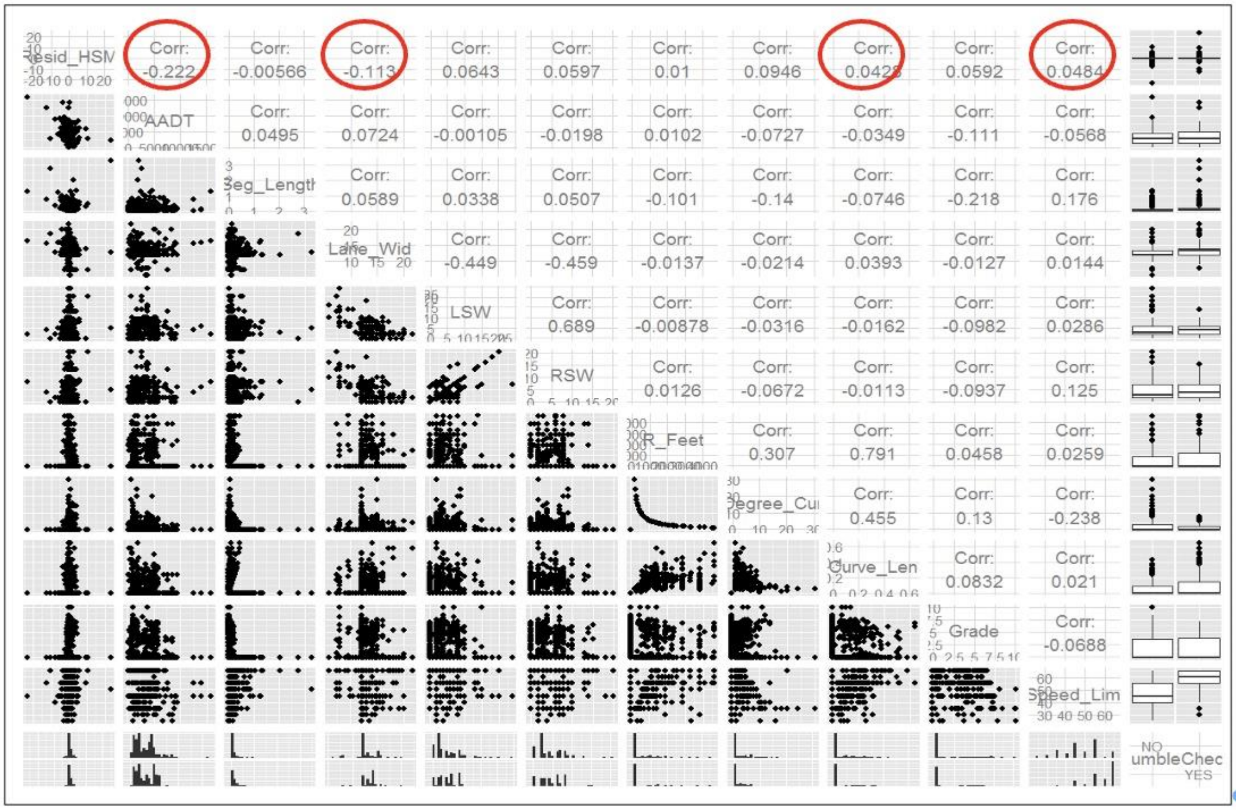


Figure 6-13 Correlation between Residuals and Segment Characteristics (RM4U)

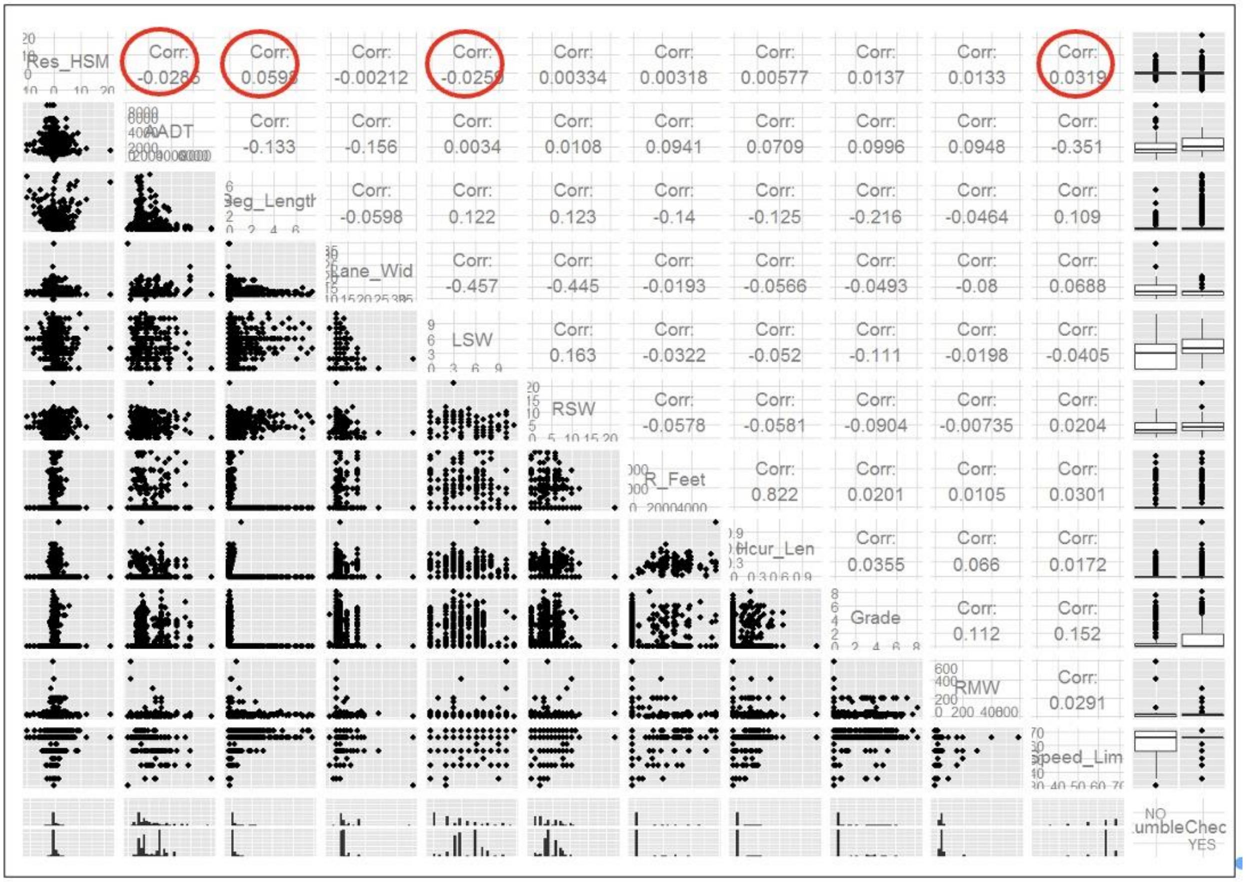


Figure 6-14 Correlation between Residuals and Segment Characteristics (RM4D)

Linear correlation may not be adequate to depict the possible non-linear dependence between residual and segment characteristics. The nonlinearity can be revealed by the locally weighted scatterplot smoothing (LOESS) curve. The residuals were plotted with LOESS from Figure 6-15 to Figure 6-17 for the four variables (i.e., AADT, segment length, left shoulder width, and the degree of curve) with the largest correlation coefficients in each of the three segment types to investigate any non-linear effect.

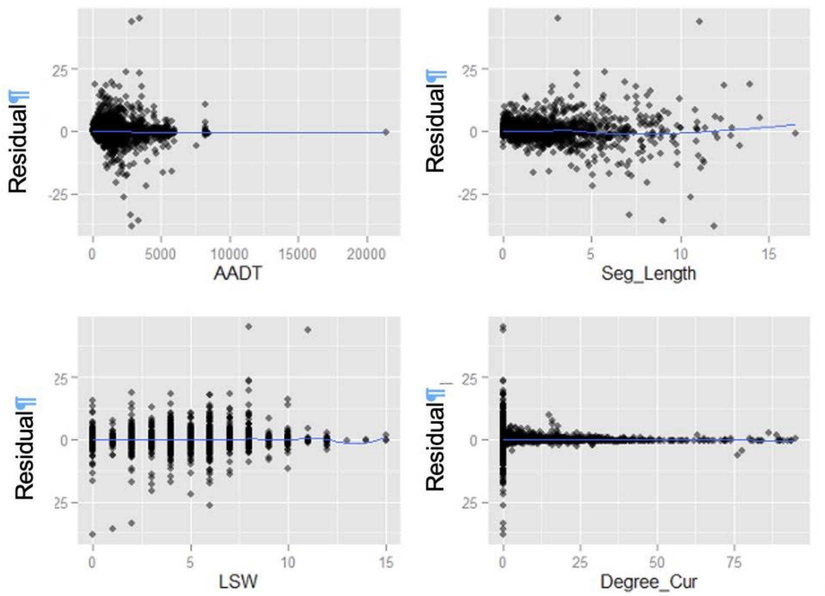


Figure 6-15 Residual Plot with Correlated Variables (RT)

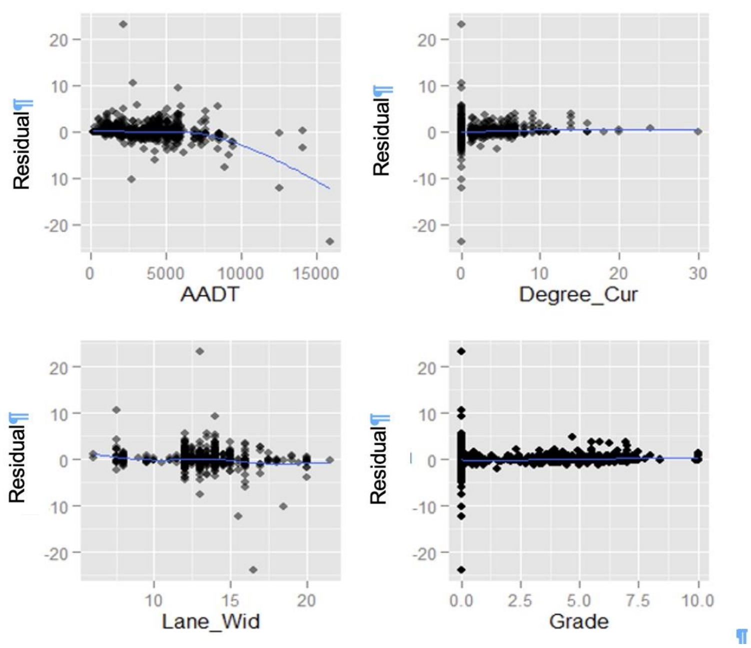


Figure 6-16 Residual Plot with Correlated Variables (RM4U)

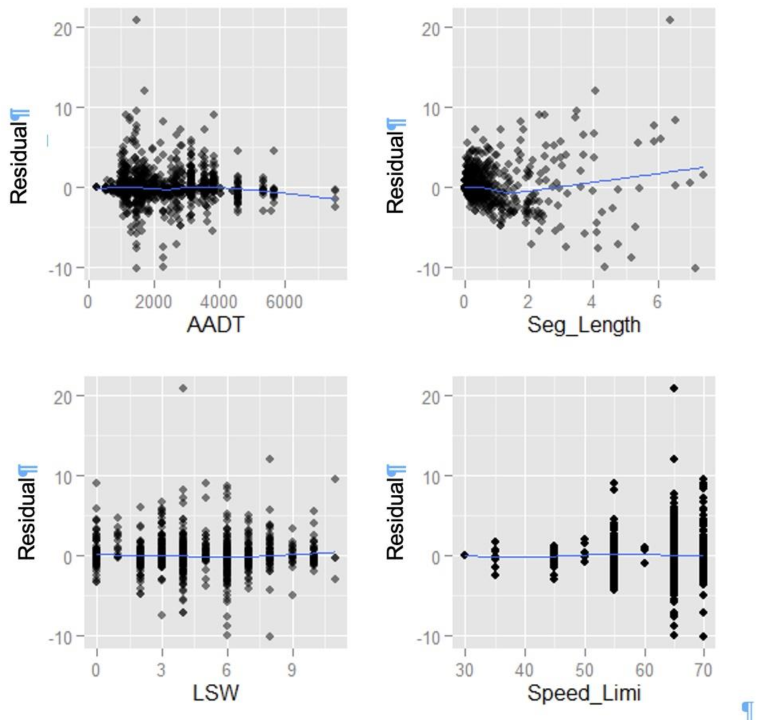


Figure 6-17 Residual Plot with Correlated Variables (RM4D)

In these plots, LOESS shows a downward trend between residuals and AADT for rural multilane segments. A closer review of this relationship suggests the LOESS fit curves were bent down because of outliers in high AADT region. Removing these outliers led to no apparent dependence between residuals and AADT. No distinct trends were found between residuals and other variables.

After extensive statistical and spatial analysis, some patterns and trends have been discovered from the perspective of either geography or contributing factor. But no pattern or trend is substantially enough to warrant the development of region-specific SPFs or add any change to the current HSM predictive method except for calibrating intersection SPFs for rural two-lane two-way intersections with South Dakota data. As mentioned at the beginning of the section, the series of analysis is based on the residual which is the observation minus the prediction multiplied by Cr (or Ci) and the prediction is calculated by strictly following the HSM predictive method. In the next section, base conditions in South Dakota are established and discussed; and appropriate modifications are recommended and carried out based on the outcome presented in this section and base conditions in the next section.

6.5. Modifications in the HSM Predictive Method

The HSM calibration process in Part C of the Predictive Methods Appendix A was reviewed. Figure 6-18 illustrates the procedure of modified HSM predictive method. The first step is to select a facility type for modification. Step 2 is to use all the available sites. In the third step, the most prevailing conditions of this facility type in South Dakota are identified as state-specific base conditions. The state-specific base conditions are compared with those defined in the HSM in Step 4. If they are different, Step 5.a, 5.b and 5.c are executed. Otherwise, Step 5.d and 5.e are followed. The calibration factor is derived following Step 6 and 7. If the value is between 0.67 and 1.33, suggesting the observed crash frequency is above or below the predicted crash frequency by one-third, the HSM SPF and the corresponding calibration factor are recommended in step 9. Otherwise, state-specific SPF should be developed in Step 5.f. Following Step 5.e, Step 6, and Step 7, the new calibration factor is derived.

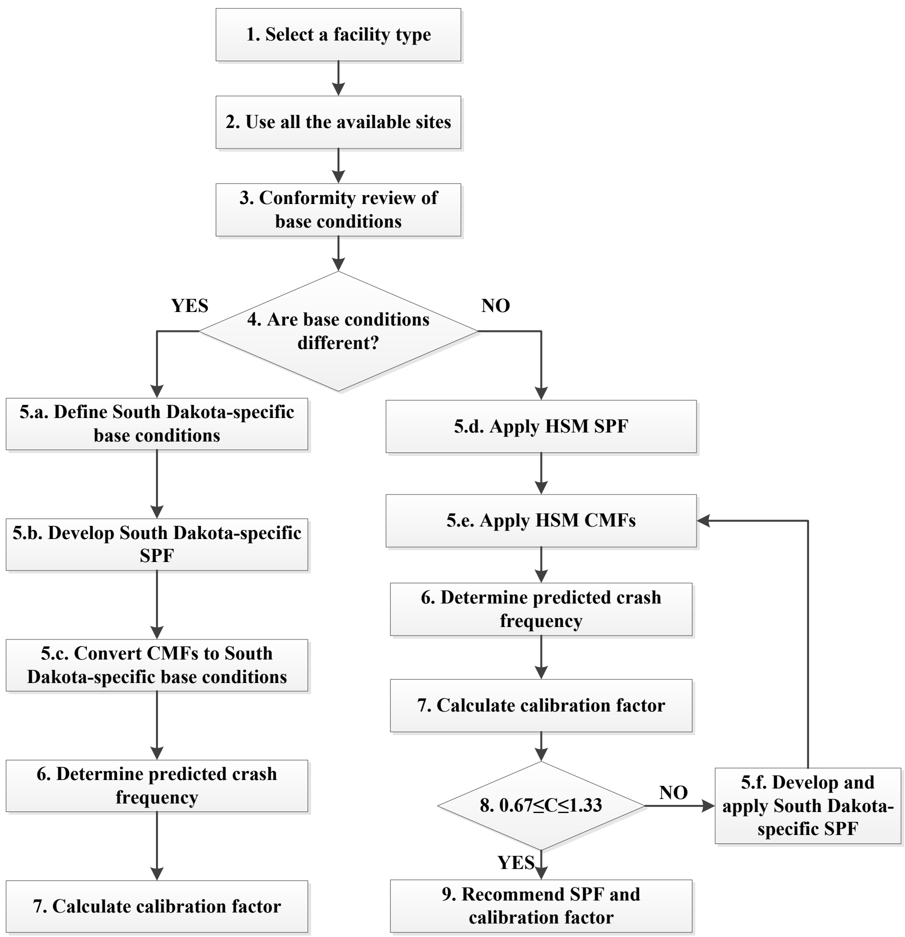


Figure 6-18 Modified HSM Predictive Methods

6.5.1. Safety Performance Function (SPF)

For each intersection and segment facility type, base conditions should first be defined and SPFs should be calibrated with the sites satisfying the base conditions. The HSM suggests that the agency-specific base conditions be designed to represent the most common characteristics of facilities. SPFs should be developed using statistically valid methods by using the local data that represent base conditions.

6.5.1.1. Base Conditions

Base conditions for different facility types in the HSM can be found in Table 11-1 and Table 11-4 in Appendix A. For rural RT3ST intersections, the most representative base conditions in South Dakota, or 162 out of 337 sites, are the same as the HSM base. For South Dakota RT4ST intersections, 409 intersections out of 582 have the same base conditions as the HSM base.

Table 6-7 summarizes the site counts and observed crash counts for the entire dataset and for the sites with base conditions only.

Table 6-7 Summary of the Whole Dataset and Dataset with Base Conditions

Intersection Type	Total Site Counts	Total Observed Crashes	Base Conditions Site Counts	Base Conditions Observed Crashes
RT3ST	337	172	162 (48%)	53 (31%)
RT4ST	582	416	409 (70%)	235 (56%)

Lane width and shoulder width are the two roadway features that can have a wide range of values. Note that the lane width is not directly available in RIS but was calculated by dividing the surface width by the number of lanes. Details are explained in Section 6.2.4. Shoulder width was directly retrieved from RIS. The length distribution for various combinations of lane and shoulder width was used to identify the conditions with the largest share of mileage. The segment length distribution for different lane and shoulder width are provided for the three segment types in

Table 6-8, Table 6-9, and

Table 6-10**.** Cells in green indicate the largest total length. Note that the average number for left and right shoulder width was considered as the shoulder width for both rural two-lane two-way and multilane undivided facilities.

Representing the maximum length, the state-specific base conditions were determined from the cell values. In

Table 6-8, the maximum length for lane width and shoulder width combination found is 12-foot lane and 4-foot shoulder. In Table 6-11, the maximum length for the combination is 12-foot lane width and 2-foot shoulder width, and in Table 6-12, the maximum length for the combination is 13-foot lane width and 4-foot shoulder width. For RM4D facility, the highest segment length was found with 30 feet median width.

Table 6-10 shows the facility length distribution with lane width and shoulder width after filtering the sites with 30 feet median width. Table 6-11 displays the new base conditions for each segment type. Base conditions for other geometric characteristics such as lane width, curve information, grade, etc. were the same as the HSM recommended base conditions.

Table 6-8 Length Distribution for Rural Two-lane Two-way Segments

LW	SW													
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**>12**
9	0.24			1.66										
10	19.35	0.35	23.98		13.48	5.83	35.18		0.19					
10.5		0.03												
11	67.84	25.79	31.22		5.51	15.20		8.89						
11.5	4.21		12.00			0.22								
12	145.82	122.15	563.20	402.40	750.37	252.78	230.93	128.30	96.29	30.76	32.53	41.71	13.87	0.31
12.5	4.56													
13	67.72	20.54	102.50	196.54	27.34	21.80	12.62	27.54	21.30		7.46			
13.5	3.43					13.88								
14	71.45	6.88	50.18	65.14	358.38	176.36	461.14	86.86	129.27	25.78	8.31			1.19
14.5								0.03						
15	89.93	0.96	0.80	10.05	9.28	41.85	17.88	54.79	14.73		16.22			0.40
15.5	0.72													
16	101.98	0.20	2.19	0.41	1.94	0.65								
16.5			1.44	0.10	0.27									
17	57.24	1.74	0.20		0.42									
17.5	0.84		0.11	0.82	0.13									
18	66.96	0.39	1.27	0.92	3.28	0.59	0.42		0.22	0.08	0.11			
18.5					0.05									
19	5.44		0.40		0.09	0.67	0.21							
19.5			0.04											
20	24.84	0.17	3.91		0.90		0.32		0.36					
21	0.33	0.41	0.08	1.59		0.07					0.21			
22	0.61	0.25	1.97	0.06	2.11	0.39								
23	0.64													
24	0.02		1.30	0.32	0.04	0.11								
26		0.15				0.25	0.05							
30				0.02										
Total	734.18	180.011	796.772	680.024	1173.59	530.646	758.753	306.408	262.343	56.612	64.855	41.714	13.867	1.905

Table 6-9 Length Distribution for Rural Multilane Undivided Segments

LW	SW												
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**>11**
6											0.10		0.13
7.5													2.48
8							0.22	0.83	0.08		0.22	0.95	0.56
9									0.19				
9.5											0.43		
10							0.14	0.22					
10.5													0.17
11				0.56									
11.5					0.07			0.32					
12	1.94		11.87	0.79	2.44	6.85	2.64	1.58	4.72	1.56	0.37		
12.5		0.08	0.03		0.20	0.21	0.18		1.22				
13	0.27		2.14	0.07	1.15	5.55	5.61	0.45	1.84				0.18
13.5	0.13	1.69	0.71	0.20	0.40	0.19	2.07	0.39	2.52				
14	2.85	0.05	0.35	1.95	1.68	3.13	1.89		1.14	0.25			
14.5		0.07		1.07		0.20							
15	0.58	0.15	1.70		0.84	0.06	1.11	0.02					
15.5			0.10		1.06								
16	1.76	0.74	2.13	2.10		0.22	0.19				0.13		
16.5						1.27							
17			0.61					1.07					
17.5	0.24		0.39	0.02									
18			0.36										
18.5	2.78												
19			0.29										
19.5	0.26												
20			0.29	0.23	0.29								
Total	10.81	2.79	20.97	6.98	8.13	17.67	14.05	4.87	11.71	1.81	1.25	0.95	3.52

Table 6-10 Length Distribution for Rural Multilane Divided Segments

Sum of Length	Right Shoulder Width											
Lane Width	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**>10**
10	0.20		0.25						0.19			
11			0.05	0.05								
12	2.50		16.21	55.35	44.52	5.69	27.96	20.75	60.37	29.43	23.36	4.15
12.5							0.12			1.08		
13	0.13		11.33	0.50	136.72	0.49	53.83	11.77	4.62	0.86	9.05	
13.5	0.45								0.45			
14	0.05	1.25	0.54	5.13	0.59		10.16		0.60			
14.5			0.77				0.77					
15	16.52	1.75	0.42		0.32	6.56	16.36	9.07			2.29	
16	0.46		4.79				0.08					
16.5												0.16
17			0.03									
18	27.68	2.28	1.42	0.31			0.24					
19	0.61											
20	0.18		0.20									
24			0.05									
34			0.22									
Total	48.79	5.27	36.27	61.33	182.15	12.75	109.51	41.59	66.24	31.37	34.69	4.59

Table 6-11 South Dakota-Specific Base Conditions by Segment Type

RT	RM4U	RM4D
Shoulder Width = 4 feet	Shoulder Width = 2 feet	Lane Width= 13 feet Shoulder Width = 4 feet

The statistics summaries of state-specific base conditions are shown in Table 6-12. The proportion of observed crashes is comparable to the proportion of length for both the RT and RM4D base conditions. In the base condition dataset for RM4U, the share of observed crashes (3.3%) is less than half of the length proportion, 7.81%. Additionally, the number of observed crashes is only 31 and the sample size is 75.

Table 6-12 Statistics Summary of the Whole Datasets and Datasets for Base Conditions

	Whole Dataset			Dataset for Base Conditions		
Type	**Length (mi)**	**Observed Crashes**	**Sample Size**	**Length (mi)**	**Observed Crashes**	**Sample Size**
RT	6,361.531	10,418	16,828	750.372 (11.8%)	949 (9.11%)	1,002 (5.95%)
RM4U	152.158	940	1,210	11.88 (7.81%)	31 (3.3%)	75 (6.2%)
RM4D	634.246	1,791	1,619	97.427 (15.36%)	110 (6.14%)	209 (12.9%)
Grand Total	7,147.935	13,149	19,657	859.679 (12.01%)	1,090 (8.29%)	1,286 (6.5%)

6.5.1.2. Developing SPFs

Similar to the SPFs in the HSM, the negative binomial regression analysis was used to develop these SPFs. All equation forms of SPFs for each facility type can be found in Table 11-2 and Table 11-5 in Appendix A. For comparison purposes, the equation forms of South Dakota-specific SPFs were the same as those in the HSM but coefficients were calibrated with local data. For intersections, only the AADTs of the major approach and minor approach were included in the SPFs. AADT and segment length were the only two variables in the SPFs for segment facilities. Model results show that all explanatory variables are statistically significant. SPF was not developed for RM3ST intersections as there are very few such sites, and not developed for RM4ST intersections as the HSM has successfully corrected the data skewness (see Figure 6-8). No SPF was developed for RM4U because it had a very small sample size and very few crashes. The SPFs for two intersection types and two segment types are provided in Table 6-13 along with the HSM SPFs.

Table 6-13 South Dakota-Specific SPFs and HSM SPFs

Facility Types	HSM SPFs	South Dakota-Specific SPFs
Intersection		
RT3ST	$e^{-9.86+0.79\times ln(AADT_{maj})+0.49\times ln(AADT_{min})}$	$e^{-9.93+0.66\times ln(AADT_{maj})+0.52\times ln(AADT_{min})}$
RT4ST	$e^{-8.56+0.60\times ln(AADT_{maj})+0.61\times ln(AADT_{min})}$	$e^{-10.55+0.79\times ln(AADT_{maj})+0.51\times ln(AADT_{min})}$
Segment		
RT	$AADT \times L \times 365 \times 10^{-6} \times e^{(-0.312)}$	$AADT \times L \times 365 \times 10^{-6} \times e^{(-0.1101)}$
RM4D	$e^{(-9.025+1.049\times \ln(AADT)+\ln(L))}$	$e^{(-19.7106+2.4597\times \ln(AADT)+\ln(L))}$

Figure 6-19 through Figure 6-22 illustrate the difference between the HSM SPFs and state-specific SPFs. SPFs for intersections involve major AADT (AADT on the major road) and minor AADT (AADT on the minor road). To visualize both, two figures were created with one being fixed and the other being varied and the average AADT is the fixed value.

For RT3ST, the average major AADT and minor AADT are 1,172 and 307, respectively. **Error! Reference source not found.** shows the trend of predicted crash frequency against major AADT when the minor AADT is 307; Figure 0-1 shows the trend of predicted crash frequency against minor AADT when the major AADT is 1,172. Both figures show that the predicted crash frequency using the HSM SPF is always greater than that of state-specific SPF, and the difference increases as AADT increases.

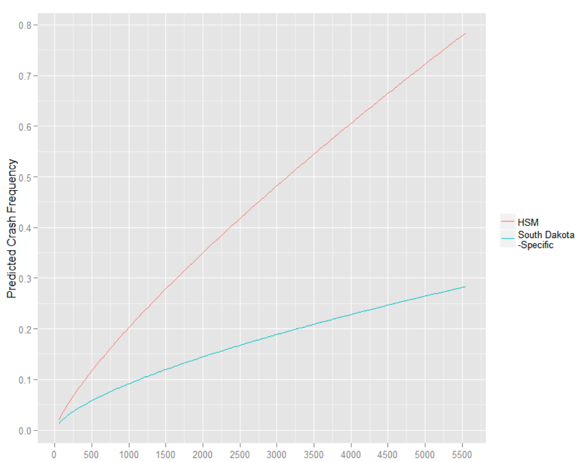


Figure 6-19 HSM SPF and State-specific SPF vs. Major AADT (RT3ST)

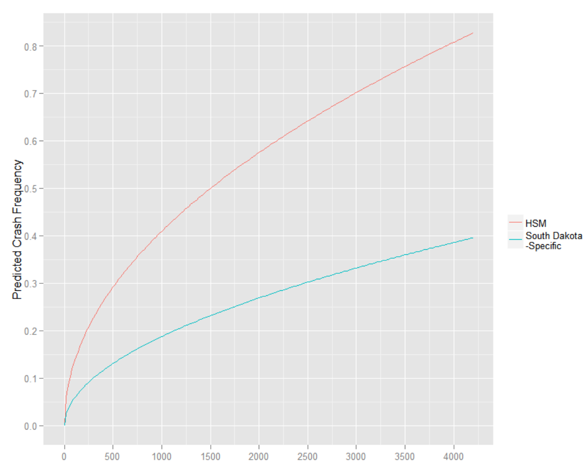


Figure 6-19 HSM SPF and State-specific SPF vs. Minor AADT (RT3ST)

For RT4ST, the average major AADT and minor AADT are 1,299 and 373, respectively. Error! Reference source not found. **shows the trend of predicted crash frequency against major AADT when the minor AADT is 373;**

shows the trend of predicted crash frequency against minor AADT when the major AADT is 1,299. Both figures show that the predicted crash frequency using the HSM SPF is always greater than state-specific SPF, and the difference increases as AADT increases.

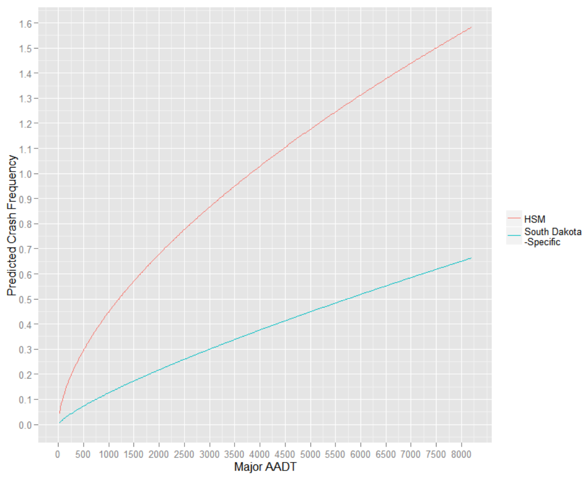


Figure 6-20 HSM SPF and State-specific SPF vs. Major AADT (RT4ST)

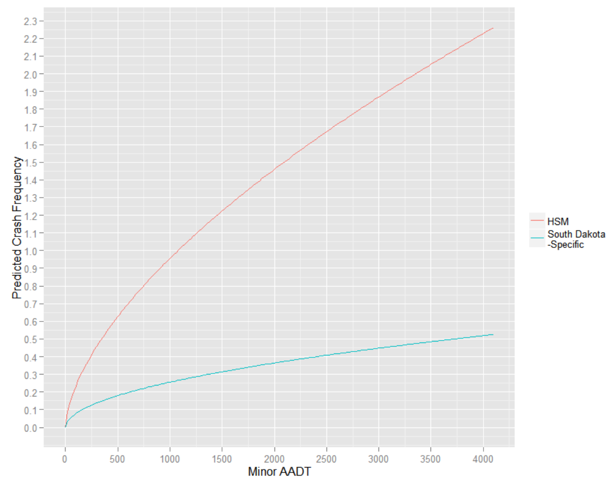


Figure 6-21 HSM SPF and State-specific SPF vs. Minor AADT (RT4ST)

SPFs were developed using the database that included all the attributes identified as “base conditions". The comparison between the HSM and state-specific SPFs was plotted in Figure 6- and Figure 6-22 for RT and RM4D, respectively. **Error! Reference source not found.**3 shows the state-specific SPF consistently predicts higher number of crashes compared to the HSM SPF. **Error! Reference source not**

found.4 shows that the state-specific SPF predicts fewer crashes in the lower bound of AADT but more crashes in the high bound of AADT.

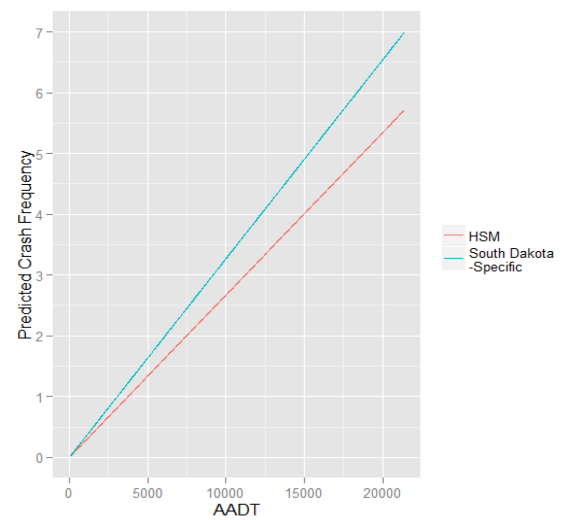


Figure 6-23 HSM SPF and State-specific SPF vs. AADT (RT)

Segment length = 1 mile

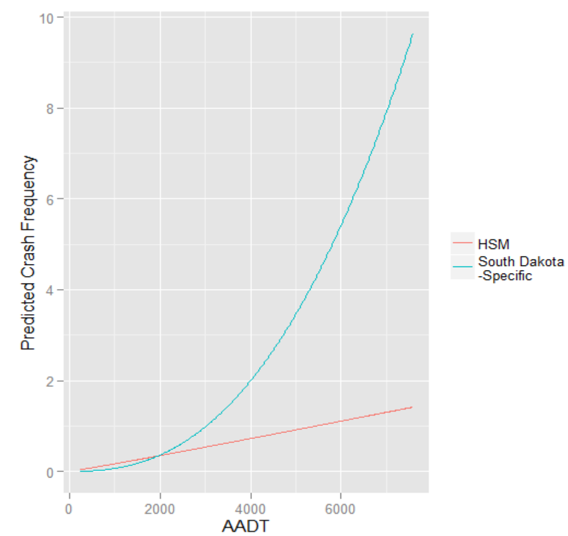


Figure 6-22 HSM SPF and State-specific SPF vs. AADT (RM4D)

Segment length = 1 mile

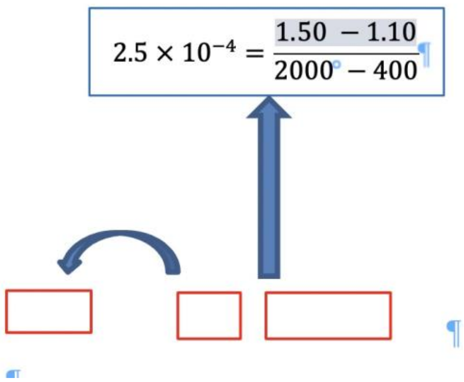
6.5.2. Crash Modification Factor (CMF)

After defining and establishing base conditions, CMFs must be converted because they represent quantitative changes in predicted crash frequencies resulting from site characteristic variations from base conditions. The base conditions for South Dakota intersections do not differ from the HSM bases; hence no changes were made for CMFs. However, the South Dakota base conditions of rural roadway segments are different from those defined in the HSM; hence CMFs were modified accordingly.

The new base lane width is 13 feet for rural multilane divided highway facilities. In the HSM, the CMF for lane width is the same for a width of 12 feet or larger. Since 12 feet is the base lane width in the HSM, the CMF for lane width is intact when the base lane width becomes 13 feet.

The shoulder width for both rural two-lane two-way and multilane divided state highways in the new base conditions is 4 feet. The CMF for shoulder width is divided into three domains based on AADT as shown in Table 11-3 in Appendix A. For AADT less than 400 vehicles/day and over 2,000 vehicles/day, the CMF value is a scale factor from the study conducted by Zegeer et al. (*21*). The CMF for AADT between 400 and 2000 vehicles/day is formulated as a linear equation where the variable is (AADT - 400), the intercept is the CMF value for AADT less than 400, and the slope is $\frac{CMF_{>2000} - CMF_{<400}}{2000-400}$. The conversion of CMFs according to the new base conditions was performed using the CMF values in the HSM. When the CMF value is a scale factor, the new base condition is reset to one, and the others can be adjusted by a corresponding multiplier. The CMF for the transition zone is then adjusted using the two scale values. The adjusted CMF is provided in Table 6-14 (refer to Table 11-3 for the original CMF).

Table 6-14 Adjusted CMF for Shoulder Width for Rural Two-lane Two-way Segments



	AADT		
Shoulder Width	**<400**	**400-2000**	**>2000**
0 feet	1.10	$1.10 + 2.5 \times 10^{-4}(AADT - 400)$	1.50
2 feet	1.07	$1.07 + 1.43 \times 10^{-4}(AADT - 400)$	1.30
4 feet	1.02	$1.02 + 8.125 \times 10^{-5}(AADT - 400)$	1.15
6 feet	1.00	1.00	1.00
8 feet or more	0.98	$0.98 - 6.875 \times 10^{-5}(AADT - 400)$	0.87

The CMF for shoulder width for rural multilane divided highway facilities was calculated the same way as the two-lane two-way highway. For shoulder width, all CMF values were adjusted with corresponding multipliers by resetting the CMF value for the new base conditions to one. The conversion of CMF for shoulder width in the multilane divided highway is provided in

Table 6-15 in which the original CMF value for 4 feet was reset to 1 and the other CMF values were adjusted by multiplying $\frac{1}{1.09}$.

Table 6-15 Adjusted CMF for Shoulder Width for Rural Multilane Divided Segments

Shoulder Width	CMF (HSM)	CMF (South Dakota-Specific)
0 feet	1.18	$\frac{1.18}{1.09} = 1.082$
2 feet	1.13	$\frac{1.13}{1.09} = 1.037$
4 feet	1.09	1.00
6 feet	1.04	$\frac{1.04}{1.09} = 0.954$
8 feet or more	1.00	$\frac{1}{1.09} = 0.917$

6.5.3. Calibration Factor

After the state-specific SPFs and CMFs were calculated, the calibration factors were derived again. Table 6-16 summarizes the calibration results. Only facility types with available state-specific SPFs and CMFs are included. The table shows that all the calibration factors using the state-specific SPFs and CMFs are closer to one when compared to those using the HSM method. The larger the calibration factor, the more the prediction deviates from the observation. From this perspective, the state-specific predictive models outperform the HSM ones.

Table 6-16 Calibration Factors

Facility Type		Observed Crashes (2008-2012)	Predicted Crashes		Calibration Factor	
			HSM Original Model	State-Specific Model	HSM Original Model	State-Specific Model
Intersection	RT3ST	170	309	131.61	0.55	1.29
	RT4ST	415	1,276	374.25	0.33	1.11
Segment	RT	10,418	8,860.31	9,898.45	1.18	1.05
	RM4D	1,791	1,138.77	1,455.17	1.57	1.23

6.5.4. Measures of Prediction Accuracy

After applying the calibration factor, the overall prediction is equal to the observation. However, for individual sites, some predictions may overestimate crash frequency while others underestimate. The residual is the difference between observation and prediction. The residual analysis helps to assess the overall prediction accuracy of the model for all the sites and the sum of absolute errors (SAE) is the accuracy measurement. The equation of SAE is as follows:

$$SAE = \sum_{i=1}^{N} |Y_i - \widehat{Y_i}| \quad (4)$$

where Y_i is the observation and $\widehat{Y_i}$ is the prediction.

Table 6-17 presents the SAE values that were calculated for the residuals of both state-specific and HSM methods. For two-lane two-way highways, the difference is so small that it can be neglected for both segments and intersections; however, the difference is rather significant for multilane divided highways.

Table 6-17 SAE for Residuals

	Facility Type	Crashes (2008-2012)	SAE	
			HSM Original Model	South Dakota-Specific Model
Intersection	RT3ST	170	193.80	193.89
	RT4ST	415	401.82	400.86
Segment	RT	10,418	8,715.65	8,686.92
	RM4D	1,791	1,419.63	1,673.58

In general, the state-specific models did not show superiority over the HSM models in terms of the prediction accuracy, and in the case of RM4D, it was worse. Although the specific models were developed from South Dakota sites, the proportion of the base condition sites and the CMFs can contribute to the value of SAE. Moreover, SAE is a scale-dependent measure and is sensitive to errors calculated at different scales (in this case, the number of crashes). For example, if the errors of majority sites are very small (i.e., much smaller than one), the error of 10 crashes or more at one site may significantly affect the value of SAE.

An alternative way is to look at the cumulative density function and examine where the large error or residual may occur. RM4D sites were plotted because of their larger SAE. The comparison of residuals between the HSM and the state-specific models shown in Figure 6-23 suggests both models have high prediction accuracy as a large portion of residuals are at or around zero. 99% of the residuals are within the range of -5 to +5 and the distributions are almost identical.

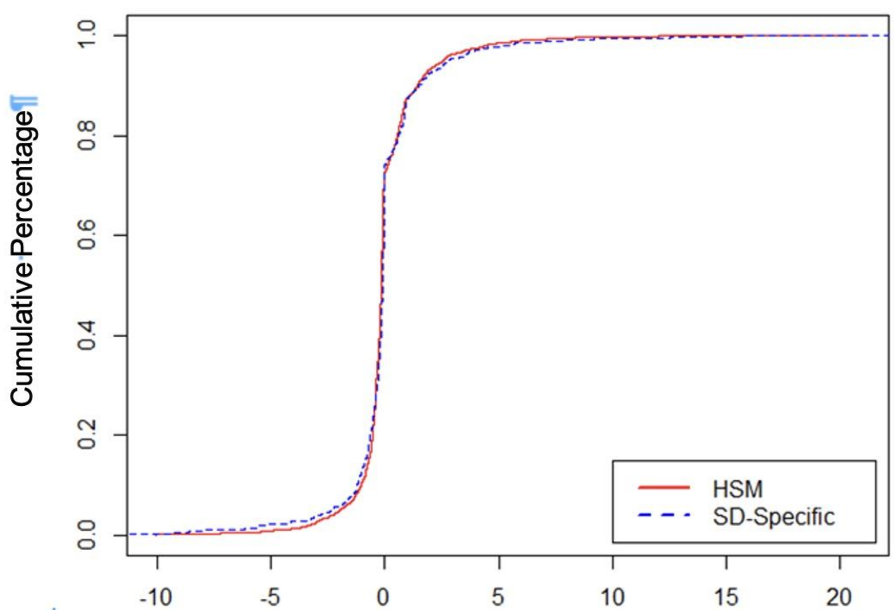


Figure 6-23 Comparison of Residuals using Cumulative Density Function

A closer look reveals that the state-specific model generated more negative residuals, which explains the larger SAE value. Later, only the RM4D sites with base conditions were compared and the SAE value for the HSM model is 163.25 whereas the SAE value for the state-specific prediction model is 152.89, suggesting more accurate prediction by the state-specific model. But the performance is inconsistent for the rest of the segments, resulting in an overall larger SAE by the state-specific model.

6.6. Calibration By Injury Severity

It is necessary to develop calibration factors in accordance with the differing injury severity levels across South Dakota. This chapter discusses where the crash injury severity data were retrieved, what procedures were followed, and how they were followed to calculate the calibration factors for each crash injury severity level by highway facility.

6.6.1. Crash Severity Data

In the crash data obtained from the "Accident" table provided by SDDOT, the severity of each crash is categorized by three levels: Fatal, Injury, and PDO (Property Damage Only). Table 20 summarizes the distribution of severity levels by facility type. Eighty-five percent of the total crashes that occurred between 2008 and 2012 in South Dakota are considered PDO. The percentages of all Fatal and Injury crashes are 0.97% and 14.27%, respectively. The percentage of fatal crashes ranges from 0.34% to 1.07% within different segment types. When comparing segment types, RM4U has the highest percentage of combined Fatal and Injury crashes (21.06%), while RM4D has the lowest percentage of those crashes. A small percentage (2.30%) of intersection-related crashes were Fatal. PDO crashes dominate among segment-related crashes; the percentage of PDO crashes (55.69%) is slightly greater than that of Injury crashes (42.01%). Among RT3ST, RT4ST, and RM4SG, all of which have relatively large crash counts, the percentages of Injury crashes are close, ranging from 40.59% to 45.60%. This is also true for PDO crashes among the three intersection types.

Table 6-18 Distribution of Crash Injury Severity

Facility Type		Total Crashes	Fatal		Injury		PDO	
Segments	RT	10,418	111	(1.07%)	1,492	(14.32%)	8,815	(84.61%)
	RM4U	940	10	(1.06%)	188	(20.00%)	742	(78.94%)
	RM4D	1,791	6	(0.34%)	197	(11.00%)	1,588	(88.67%)
	Total	13,149	127	(0.97%)	1,877	(14.27%)	11,145	(84.76%)
Intersections	RT3ST	170	1	(0.59%)	69	(40.59%)	100	(58.82%)
	RT4ST	415	13	(3.13%)	179	(43.13%)	223	(53.73%)
	RM3ST	26	0	(0.00%)	5	(19.23%)	21	(80.77%)
	RM4ST	125	3	(2.40%)	57	(45.60%)	65	(52.00%)
	RM4SG	2	0	(0.00%)	0	(0.00%)	2	(100.00%)
	Total	738	17	(2.30%)	310	(42.01%)	411	(55.69%)

The HSM recommends calibrating the crash prediction method based on the KABCO scale to get calibration factors based on injury severity. The KABCO scale consists of five crash severity categories designated as Fatal (K), Incapacitating Injury (A), Non Incapacitating Injury (B), Possible Injury (C) and Property Damage Only (O). As discussed in Section 6, SDARS manages the crash dataset, and multiple tables depict detailed crash information in different aspects. The injury severities of all persons involved in each crash are available in the KABCO scale within the "Person" table. In this table, each person has a unique "PersonSeqID" and all persons involved in the same crash share one "AccidentSeqID". The injury severity of each crash can be obtained in the KABCO scale by identifying the highest severity level of all persons involved in that crash. The injury severity information of each person provided in the "Person" table was coded by consecutive numbers from the least severe injury to the most severe injury to find the most severe injury of each crash (see Table 6-19). Aside from the five levels included in the KABCO scale, there are three additional levels that describe the injury status of each person: “wild animal hit”, “not applicable”, and “unknown”. The three levels were assumed to be “no injury”.

Table 6-19 Injury Severity Coding

Injury Status	Injury Severity	Coded Number
Fatal injury	K	5
Incapacitating Injury	A	4
Non-incapacitating Injury	B	3
Possible Injury	C	2
No injury	O	1
Wild animal hit	O	1
Not applicable	O	1
Unknown	O	1

After coding each injury severity in the “Person” table, the maximum coded number for each unique “AccidentSeqID” was calculated. For example, a crash event involves three persons and their injury severities are incapacitating injury (4), possible (2) and wild animal hit (1). The maximum coded number for this crash will be 4, and the injury severity of this crash will be designated as incapacitating injury (4). The severity of each crash that occurred in a site was obtained by linking the crash data of all facilities with the “Person” table based on shared “AccidentSeqID”. The three-level injury severity data were compared with the KABCO-scale severity obtained from the “Person” table for segment-related and intersection-related crashes, as presented in Table 6-20 and Table 6-21, to examine the consistence of severity designation in the two tables.

Table 6-20 Comparison of Injury Severities of Segment-Related Crashes

	K	A	B	C	O	Total
Fatal	127					127
Injury		562	723	592		1,877
PDO					11,145	11,145
Total	127	562	723	592	11,145	13,149

Table 6-21 Comparison of Injury Severities of Intersection-Related Crashes

	K	A	B	C	O	Total
Fatal	17					17
Injury		88	118	104		310
PDO					417	417
Total	17	88	118	104	417	744

Tables 6-22 and 6-23 show that all K and O injury severities are linked to fatal and PDO crashes, respectively, and all A, B, and C injury severities are linked to injury crashes. This indicates that the injury severity in the crash data and "Person" tables are consistent, and that the previous assumption that "Wild animal hit", "Not applicable" and "Unknown" all qualify as "No injury" crashes is correct.

6.6.2. Calibration by Injury Severity

After crash observations by severity were obtained, it was necessary to identify the crash prediction by severity to calculate the calibration factor. In the HSM for rural multilane facilities, SPFs are available for total crashes, Fatal + Injury (KABC) crashes, and Fatal + Injury* (KAB) crashes are used to calculate predicted crash counts by severity level. However, for rural two-lane two-way facilities, SPFs are available only for total crashes. The problem was solved by using the distributions for crash severity level to generate the crash counts of differing severity levels.

The CMFs do not change for different severity levels at the same site, so the same CMF values were used to calculate the predicted crash frequencies of different severity levels at the same site. Therefore, the distribution for crash severity level can be directly applied to the predicted total crash counts to get the predicted crash frequency of different injury severities for rural two-lane two-way facilities. For rural multilane facilities, predicted PDO crash frequency was calculated as the difference between predicted total crash frequency and Fatal + Injury crash frequency.

The distributions for crash severity level of observed crash counts and predicted crash counts following HSM procedures were compared and are presented in Table 6-22. It should be noted that the predicted crash counts for rural two-lane two-way facilities are derived from the default severity distribution given in the HSM, while rural multilane facilities have SPFs by severity level to calculate the predicted crash counts of different severity levels. The comparison results show that the two distributions are similar for intersections, except RM3ST and RM4SG which have a small sample of crashes. However, a significant deviation is found between the distributions for segments. Generally, much smaller proportions of FI (Fatal + Injury) and FI* (Fatal + Injury*) crashes are found in South Dakota. Thus, with the same amount of total crashes, much less FI and FI* crashes have occurred in South Dakota.

Table 6-22 Comparison Between Observed and HSM Severity Distributions

Facility Type	Observed			HSM		
	FI	FI*	PDO	FI	FI*	PDO
Segments						
RT	15.4%	10.8%	84.61%	32.1%	17.6%	67.9%
RM4U	21.0%	16.1%	78.9%	64.1%	40.2%	35.9%
RM4D	11.3%	7.4%	88.7%	60.1%	44.1%	39.9%
Intersections						
RT3ST	41.2%	29.4%	58.8%	41.5%	22.3%	58.5%
RT4ST	46.3%	30.8%	53.7%	43.1%	22.3%	56.9%
RM3ST	19.2%	7.7%	80.8%	47.4%	31.6%	52.6%
RM4ST	48.0%	34.4%	52.0%	48.4%	32.6%	51.6%
RM4SG	0.0%	0.0%	100.0%	40.4%	13.5%	59.6%
*Using the KABCO scale, these include only KAB crashes. Crashes with severity level C (possible injury) are not included.						

Calibration factors were calculated as the ratio of the observed to predicted crash counts. The calibration factors by severity level for all segments and intersections are presented in Table 6-25. The calibration factors for FI and FI*crashes are all smaller than 1, while the calibration factors for PDO crashes and total crashes are all larger than 1. This finding indicates that although a higher number of total crashes occurred in South Dakota than predicted, the number of fatal and injury-causing crashes was lower than predicted for South Dakota. As calibration factors for FI and FI* crashes become smaller than 1, those for PDO crashes get larger beyond 1. Among three segment types, RT calibration factors for FI and FI* crashes are the smallest (0.56 and 0.72) and the calibration factor for PDO crashes is the largest (1.47). The distinction between calibration factors for FI and FI* crashes and for PDO crashes is related to the different severity distribution for segment-related crashes shown in Table 6-23. For intersections, except RM3ST and RM4SG, the calibration factors of different severity levels are close to each other and are also close to those for total crashes. The distributions for severity level in South Dakota are similar and therefore consistent with distributions for HSM predicted crashes.

Table 6-23 Calibration Factor for Different Severity Levels

	Observed Crashes			Predicted Crashes			Calibration Factor			
Facility Type	**FI**	**FI***	**PDO**	**FI**	**FI***	**PDO**	**FI**	**FI***	**PDO**	**Total[a]**
Segment										
RT	1,603	1,127	8,815	2,841	1,558	6,011	0.56	0.72	1.47	**1.18**
RM4U	198	152	742	526	330	295	0.38	0.46	2.52	**1.14**
RM4D	203	133	1,588	684	502	454	0.30	0.26	3.50	**1.57**
Intersection										
RT3ST	70	50	100	128	68	181	0.55	0.73	0.55	**0.55**
RT4ST	192	128	223	550	284	726	0.35	0.45	0.31	**0.33**
RM3ST	5	2	21	9	6	10	0.53	0.31	2.16	**1.36**
RM4ST	60	43	65	107	72	114	0.56	0.59	0.57	**0.56**
RM4SG	0	0	2	21	7	31	0.00	0.00	0.07	**0.04**

a. Combined calibration factor for the total crash counts.
b. Using the KABCO scale, these include only KAB crashes. Crashes with severity level C (possible injury) are not included.

7. CONCLUSIONS

This report summarizes the effort to calibrate the HSM predictive methods for available segment and intersection types on rural South Dakota highways. The calibration began with a review of existing SDDOT safety data sources that follow the HSM application data requirements, and continued with a faithful implementation of the HSM SPFs and CMFs. The calibration concluded by making appropriate modifications to highway facility types that otherwise could not make adequate predictions with the HSM predictive method. The recommendation is to use a hybrid method including both HSM models and state-specific models for specific segment and intersection types in order to achieve high crash prediction accuracy. Guidelines for future calibration have been included in this report.

The data required for calibrating the HSM predictive methods were collected from intersection and segment shapefiles, RIS, and the South Dakota Accident Records System (SDARS). Preparing the segment shapefile required the creation of homogeneous segments based on traffic and geometric features in which the RIS event tables (which include features such as AADT, surface type, median, horizontal curve, vertical curve, and speed limit) were merged into a single table. The roadway was then segmented wherever a feature changed. After creating homogenous sites, crashes that occurred between 2008 and 2012 were assigned to intersections and segments. The HSM predictive methods were employed to predict crashes for all facility types. The calibration factor for each facility type was computed as the ratio of observed crashes to predicted crashes. Table 7-1 summarizes the calibration factors. It was clear that the calibration factor varied remarkably across the facility types. The HSM underestimated the total number of crashes on roadway segments for all South Dakota rural state highway types included in this study; on the other hand, it overestimated the total number of crashes at intersections for all South Dakota intersection types except RM3ST.

Table 7-1 Summary of HSM Calibration Factors

Facility Type				Facility Count*	Observed Crashes	Calibration Factor (C)
Segments	Rural Two-Lane	RT	2-lane undivided	6,362	10,418	1.18
Segments	Rural Multilane	RM4U	4-lane undivided	152	940	1.14
Segments	Rural Multilane	RM4D	4-lane divided	634	1,791	1.57
Intersections	Rural Two-Lane	RT3ST	3-leg, minor STOP	337	170	0.55
Intersections	Rural Two-Lane	RT4ST	4-leg, minor STOP	582	415	0.33
Intersections	Rural Multilane	RM3ST	3-leg, minor STOP	19	26	1.36
Intersections	Rural Multilane	RM4ST	4-leg, minor STOP	71	125	0.56
Intersections	Rural Multilane	RM4SG	4-leg, signalized	1	2	0.04

** For segments, the number is the length (miles); for intersections, the number is the frequency*

After applying the calibration factors, considerable differences still existed between prediction and observation for individual sites within the same highway facility type across the state. The geographic distribution of calibration factors for individual sites was plotted to identify any spatial pattern that may support region-specific models rather than a single state model. While no apparent pattern was found for intersections, clusters were observed for segments. Despite mixed results, the prediction pattern of segment crashes appears to be overestimated in Western South Dakota but underestimated in Eastern South Dakota. It is challenging to divide South Dakota into regions due to the calibration factor and unknown benefits gained through this effort; therefore, it was decided that a single state model should remain.

The roadway characteristics were evaluated in the crash prediction models to find novel ways of improving prediction accuracy. Prediction accuracy, also called the residual, was measured by the difference between the observation and the prediction. Because of the combined effects of both a limited sample size and an overall weak correlation between the residual and each studied attribute, only rural two-lane intersections with stop control (RT3ST and RT4ST) were recommended for the development of state-specific performance functions. The nature of base conditions in South Dakota is another justification for developing South Dakota models. Base conditions represent the most prevailing roadway or traffic characteristics within each facility type. A conformity review revealed that the base conditions for RT, RM4D, and RM4U were different from the ones defined in the HSM. Subsequently, SPFs were developed for RT, RM4D, RT3ST, and RT4ST. RM4U was excluded due to its small sample size.

The base conditions for RT are the same as those in the HSM except for the 4-foot shoulder width. For RM4D, the new base lane width and shoulder width are 13 feet and 4 feet, respectively. According to the HSM, a 12-foot lane width has the same CMF value as a 13-foot width; therefore, no change is needed for the CMF lane width. The CMF shoulder width, however, should be adjusted for both RT and RM4D. The calibration results, which show that the calibration factor is now almost one, show that the state-specific predictive method provides a more accurate prediction of crash frequency than the HSM method.

In addition to calibrating the total number of crashes for each facility type, calibration factors by severity level were calculated as the ratio of the observed and predicted crash frequency by injury severity. The injury severity by KABCO scale was obtained from the "Person" table in SDARS by identifying the highest severity level of all occupants involved in a crash. For rural multilane facilities, the HSM crash frequencies by injury severity were calculated by SPFs of different injury severities. Rural two-lane highway crash frequencies were obtained by applying the HSM severity distribution to the total number of predicted crashes. The calibration factors are summarized in Table 7-2. The calibration factor shows a large diversity across facility type and injury severity level. For all fatal and injury crashes (FI), the HSM overestimated for all highway facility types. For property damage-only (PDO) crashes, the HSM estimates ranged from overestimation (0.31) for RT4ST to underestimation (3.50) for RM4D.

Facility Type				Facility Count[a]	Calibration Factor		
					FI	FI[b]	PDO
Segments	Rural Two-Lane	RT	2-lane undivided	6,362	0.56	0.72	1.47
	Rural Multilane	RM4U	4-lane undivided	152	0.38	0.46	2.52
		RM4D	4-lane divided	634	0.30	0.26	3.50
Intersections	Rural Two-Lane	RT3ST	3-leg, minor STOP	337	0.55	0.73	0.55
		RT4ST	4-leg, minor STOP	582	0.35	0.45	0.31
	Rural	RM3ST	3-leg, minor STOP	19	0.53	0.31	2.16

	Multilane	RM4ST	4-leg, minor STOP	71	0.56	0.59	0.57
		RM4SG	4-leg, signalized	1	0.00	0.00	0.07
a. For segments, the number is the length (miles); for intersections, the number is the frequency. b. These include only KAB crashes. Crashes with severity level C (possible injury) are not included.							

Table 7-2 Summary of Calibration Factors by Injury Severity

8. IMPLEMENTATION RECOMMENDATIONS

This section presents the recommendations for implementation of the HSM calibration at SDDOT.

8.1. Rural Two-lane Two-Way Segments

The state-specific SPF and adjusted CMF for shoulder width should be used for rural two-lane two-way highway segments with a calibration factor of 1.05 for improved crash prediction accuracy.

The shoulder width in the South Dakota base conditions for RT sites is 4 feet, which is different from the 6-foot shoulder width defined in the HSM. Since the base conditions have changed, the state-specific SPF is now: $N_{spf\ RT} = AADT \times L \times 365 \times 10^{-6} \times e^{(-0.1101)}$ (Table 6-15). CMF for shoulder width was adjusted accordingly, and is presented in Table 6-14. The calibration factor is 1.05 as shown in Table 6-16.

8.2. Rural Multilane Undivided Segments

For rural multilane undivided highway segments, the HSM SPF and CMFs should be followed, and a calibration factor of 1.14 is recommended.

Although the South Dakota base conditions of RM4U are different from the HSM base conditions, there were only 31 crashes occurred at base conditions sites between 2008 and 2012. The crash frequency is too low to develop any statistically reliable SPF. The calibration factor following the HSM predictive method is 1.14 as shown in Table 7-1.

8.3. Rural Multilane Divided Segments

For improved crash prediction accuracy, the state-specific SPF and adjusted CMF for shoulder width should be used for rural multilane divided highway segments with a calibration factor of 1.23.

The shoulder width in the South Dakota base conditions for RT sites is 4 feet, which is different from the 6-foot shoulder width defined in the HSM. Since the base conditions have changed, the state-specific SPF is now: $N_{spf\ RM4D} = e^{(-19.7106+2.4597\times\ln(AADT)+\ln(L))}$ (Table 6-15). CMF for shoulder width was adjusted accordingly as presented in

Table 6-15. The calibration factor is 1.23 as shown in Table 6-16.

8.4. Rural Two-Lane Three-Leg Intersections with STOP Control

Either the HSM SPF with a calibration factor of 0.55 or the state-specific SPF with a calibration factor of 1.29 is acceptable for rural two-lane three-leg stop controlled intersections, as the calibration improvement is marginal.

The South Dakota base conditions for RT3ST sites are the same as the HSM base conditions. Also, the overall prediction accuracy is similar for that of the HSM model and the state-specific model; therefore, either can be applied for RT3ST. The calibration factors for using the HSM SPF and the state-specific SPF are 0.55 and 1.29, as shown in Table 6-16.

8.5. Rural Two-Lane Four-Leg Intersections with STOP Control

The calibration improvement for rural two-lane four-leg stop-controlled intersections is marginal, meaning either the HSM SPF with a calibration factor of 0.33 or the state-specific SPF with a calibration factor of 1.11 is acceptable.

The South Dakota base conditions for RT4ST sites are the same as the HSM base conditions, and the overall prediction accuracy of the two models is similar; therefore, either can be applied for RT4ST. The calibration factors for using the HSM SPF and state-specific SPF are 0.33 and 1.11, as shown in Table 6-16.

8.6. Rural Multilane Three-Leg Intersections with STOP Control

For rural multilane three-leg intersections with stop control, the HSM SPF and CMFs should be followed along with a calibration factor of 1.36, as shown in Table 7-1. The predicted crash frequency should be used with extreme caution, however, because this category failed to meet the HSM recommended calibration sample size requirement. The current dataset has 19 sites and 26 crashes from 2008 to 2012 while the HSM recommendation is to have more than 30 sites and at least 100 crashes per year. More data should be collected in the future to warrant any calibration. In the meantime, the HSM predictive method for rural multilane three-leg intersections with stop control should be used with a calibration factor of one.

8.7. Rural Multilane Four-Leg Intersections with STOP Control

The HSM SPF and CMFs should be followed along with a calibration factor of 0.56 for rural multilane four-leg stop-controlled intersections, as shown in Table 7-1. The predicted crash frequency should be used with caution, however, because this type of facility had 71 sites, representing 125 crashes from 2008 to 2012. This amount is lower than the HSM-recommended 30 to 50 sites with at least 100 crashes per year.

8.8. Rural Multilane Four-Leg Signalized Intersections

Due to the major limitation of a small sample size, calibration is not recommended for rural multilane four-leg signalized intersections. More data should be collected in the future. In the meantime, the HSM predictive method for rural multilane four-leg signalized intersections is recommended.

In the current SDDOT dataset, there is only one RM4SG site with two crashes in five years. This is substantially lower than the HSM-recommended 30 to 50 sites with at least 100 crashes per year.

8.9. Calibration by Injury Severity

The process for predicting crashes by injury severity is reiterated here because of the huge difference in methodology between rural multilane and two-lane facilities. The HSM provides the SPFs for both FI and FI* injury severities for rural multilane facilities. The predicted PDO crashes can be calculated as the difference between the predicted total and predicted FI crashes. For rural two-lane facilities, the HSM provides the proportion of crash frequency by severity level. It is recommended that the HSM SPF and CMFs be used along with calibration factors by injury severity for RT, RM4U, RM4D, RT3ST and RT4ST, as shown in Table 7-2; however, for RM4ST, the calibration factor should be used with caution. It is not recommended to use calibration by injury severity for RM3ST and RM4SG.

8.10. Data Collection

As SPFs are developed for the sites that conform to base conditions, CMFs account for any variation of site characteristics from the base conditions. Although all required data items for intersections are available in the SDDOT "StatetoState_and_StatetoNonStateFedAid" file, data items such as *driveway density, horizontal curve superelevation, number of passing lanes, two-way left-turn lanes, roadside design,* and *lighting* are not available for roadway segments. These missing data items should be collected for a more accurate prediction. The complete list of available South Dakota data can be found in Table 6-4 List of CMFs for Rural Two-lane Two-way Highway Facilities and

Table 6-5 List of CMFs for Rural Multilane Highway Facilities.

8.11. Data Integration

The data source tables pertaining to intersections have been integrated in an intersections toolbox that is currently in use to apply the HSM predictive methods at SDDOT. The parameters in SPFs can easily be modified in the toolbox to implement state-specific models; however, there is no such convenience for road segments, as those data items are stored in separate tables. The process of integrating segment data tables was complicated because dynamic segmentation could not be performed without access to SDDOT's linear referencing system (LRS); instead, alternative computer programs (e.g., R, MS Excel, and ArcGIS) were employed to merge files. These disjointed steps can be streamlined, simplified, and implemented in a GIS environment with the aid of SDDOT LRS by the GIS unit. The data reduction criteria and specific file merging steps and procedures are detailed in Section 6.2 Data Processing.

8.12. Calibration Updates

As road conditions change over time, the amount and types of crashes will also change. To maintain the integrity and accuracy of calibration results, it is recommended to evaluate existing base conditions that are significantly different. The HSM states, "*it is recommended that the new values of the calibration factors be derived at least every two to three years, and some users may prefer to develop calibration factors on an annual basis*" (1). There is no other guidance regarding how often the calibration should be updated. Due to the complexity of calibration, it is recommended that new values for calibration factors be derived every three years, or when crash frequency or injury severity distribution is significantly changed for a specific facility (e.g., ±10%).

9. RESEARCH BENEFITS

The calibrated HSM crash prediction models for the implemented facility types reflected local safety conditions in South Dakota. The calibration of HSM predictive models also sought to quantify the empirical benefits of safety conditions in South Dakota. SDDOT can benefit from this research in the following respects:

1. Safety Data Preparation: SDDOT maintains roadway geometric characteristics data and traffic information in separate data tables stored in RIS. The crash database was maintained by SDARS, which is completely separate from RIS. To develop a crash prediction model, all the roadway geometric characteristics data and traffic data need to be merged together along with the crash assignment for each site. This process involves separate data tables and includes a procedure to integrate all necessary tables for HSM calibration. This data integration process can be repeated for a specific area or for the entire state when developing a single database appropriate for other safety analyses.
2. Agency-developed SPFs: This research includes the development of South Dakota-specific SPFs for predicting crashes. The SPFs developed with South Dakota data were more accurate and reliable for predicting crash frequencies under South Dakota-specific base conditions.
3. CMF Adjustment: For South Dakota-specific SPFs, the CMFs were adjusted to the new base conditions in order to obtain expected crash frequency for a site. The adjusted CMFs can reflect the trend for the effect of countermeasures on new base conditions.
4. Accurate Crash Prediction: One of the most substantial benefits of calibrating HSM predictive methods across all facility types is the ability to provide more accurate safety benefits. This can be done by comparing the predicted and expected numbers of predicted crash frequencies over a multiple-year time horizon. This can also help in comparing, selecting, and recommending safety design alternatives, evaluating site-specific safety issues, and budgeting and planning future safety projects.
5. Development of Calibration Guidelines: The guidelines for future calibration outline the data requirements, identify key components for calibration, and recommend calibration time intervals. SDDOT can obtain reliable, correct, and consistent results for the South Dakota version of HSM safety predictions when following these guidelines.
6. Identification of Future Safety Data Needs: The application and calibration of the HSM predictive method requires extensive data items in order to quantify safety conditions. This research investigated data requirements and prepared a list of data items needed for calibration. A future recommendation to collect new data items was proposed after developing state-specific SPFs, as not all required data items for calibration were available in RIS data tables. By collecting the previously unavailable data items, the prediction accuracy of original or modified HSM predictive methods would be greatly improved. Higher prediction accuracy helps to obtain more accurate calibration factors, which will in turn contribute to a more reliable predication.

10. REFERENCES

[1] AASHTO. *Highway Safety Manual*. AASHTHO, 2010.

[2] Resende, P. T. V., and R. F. Benekohal. Effects of roadway section length on accident modeling.In *Traffic congestion and traffic safety in the 21st century: Challenges, innovations, and opportunities*, 1997.

[3] Miaou, S.-P., and H. Lum. Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention,* Vol. 25, No. 6, 1993, pp. 689-709.

[4] Ogle, J. H., P. Alluri, and W. Sarasua. MMUCC and MIRE: the role of segmentation in safety analysis.In *Proceedings of the Paper Presented at the 90th Annual Meeting of the Transportation Research Board*, 2011.

[5] Qin, X., and A. Wellner. Segment length impact on highway safety screening analysis.In *Transportation Research Board 91st Annual Meeting*, 2012.

[6] Cafiso, S., C. D'agostino, and B. Persaud. Investigating the influence of segmentation in estimating safety performance functions for roadway sections.In *92th Annual Meeting of the Transportation Research Board, Washington, DC*, 2013.

[7] Banihashemi, M. Highway Safety Manual, new model parameters vs. calibration of crash prediction models.In *Compendium of the 90th Annual Meeting of the Transportation Research Board*, 2011.

[8] Kononov, J., C. Lyon, and B. K. Allery. Relation of flow, speed, and density of urban freeways to functional form of a safety performance function. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 2236, No. 1, 2011, pp. 11-19.

[9] Brimley, B. K., M. Saito, and G. G. Schultz. Calibration of Highway Safety Manual Safety Performance Function. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 2279, No. 1, 2012, pp. 82-89.

[10] Martinelli, F., F. La Torre, and P. Vadi. Calibration of the Highway Safety Manual's Accident Prediction Model for Italian Secondary Road Network. *Transportation Research Record*, No. 2103, 2009, pp. 1-9.

[11] Sun, X., Y. Li, D. Magri, and H. H. Shirazi. Application of highway safety manual draft chapter: Louisiana experience. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 1950, No. 1, 2006, pp. 55-64.

[12] Harwood, D. W., F. Council, E. Hauer, W. Hughes, and A. Vogt. Prediction of the expected safety performance of rural two-lane highways.In, 2000.

[13] Lord, D., and J. A. Bonneson. Development of accident modification factors for rural frontage road segments in Texas. *Transportation Research Record*, No. 2023, 2007, pp. 20-27.

[14] Abdel-Aty, M. A., P. C. Lee, P. Eng, J. Park, J.-H. Wang, M. Abuzwidah, and S. Al-Arifi. Validation and Application of Highway Safety Manual (Part D) in Florida.In, Florida Department of Transportation, 2014.

[15] Kweon, Y.-J. L., In-Kyu. Development of Safety Performance Functions for Multilane Highway and Freeway Segments Maintained by the Virginia Department of Transportation

In, Virginia Department of Transportation, 2014.

[16] USDOT. *Crash Modification Factors Clearinghouse*. http://www.cmfclearinghouse.org/. Accessed Dec 4, 2014.

[17] Xie, F., K. Gladhill, K. K. Dixon, and C. M. Monsere. Calibration of Highway Safety Manual Predictive Models for Oregon State Highways. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 2241, No. 1, 2011, pp. 19-28.

[18] Mehta, G., and Y. Lou. Calibration and Development of Safety Performance Functions for Alabama. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 2398, No. 1, 2013, pp. 75-82.

[19] Bauer, K. M., and D. W. Harwood. Safety effects of horizontal curve and grade combinations on rural two-lane roads.In *Proceedings of the 92nd Transportation Research Board Annual Meeting*, 2013. pp. 1-22.

[20] Findley, D. J., J. E. Hummer, W. Rasdorf, C. V. Zegeer, and T. J. Fowler. Modeling the impact of spatial relationships on horizontal curve safety. *Accident Analysis & Prevention,* Vol. 45, 2012, pp. 296-304.

[21] Zegeer, C. V., R. C. Deen, and J. G. Mayes. The Effect of Lane and Shoulder Widths on Accident Reductions on Rural, Two-Lane Roads. 1980.

11. APPENDIX A Selected HSM Tables

Table 11-1 Base Conditions for Rural Two-lane Two-way Highway Facilities

Roadway Segments	
Lane Width (LW)	12 feet
Shoulder Width (SW)	6 feet
Shoulder Type	Paved
Roadside Hazard Rating (RHR)	3
Driveway Density (DD)	5 driveways per mile
Horizontal Curvature	None
Vertical Curvature	None
Centerline Rumble Strips	None
Passing Lanes	None
Two-way left-turn Lanes	None
Lighting	None
Automated Speed Enforcement	None
Grade Level	0%
Intersections	
Intersection Skew Angle	0˚
Intersection Left-turn Lanes	none on approaches without stop control
Intersection Right-turn Lanes	none on approaches without stop control
Lighting	None

Table 11-2 SPF for Rural Two-lane Two-way Highways

Facility Type	N_{spf} [1]
RT	$AADT \times L \times 365 \times 10^{-6} \times e^{(-0.312)}$
RT3ST	$\exp[-9.86 + 0.79 \times \ln(AADT_{maj}) + 0.49 \times \ln(AADT_{\min})]$
RT4ST	$\exp[-8.56 + 0.60 \times \ln(AADT_{maj}) + 0.61 \times \ln(AADT_{\min})]$
RT4SG	$\exp[-5.13 + 0.60 \times \ln(AADT_{maj}) + 0.20 \times \ln(AADT_{\min})]$
	N_{spf} = Predicted total crash frequency per year for roadway segment/ intersection for base condition AADT = Average Annual Daily Traffic volume (vehicle per day) on segment L = Length of roadway segment (miles) $AADT_{maj}$ = AADT (vehicle per day) on major road $AADT_{min}$ = AADT (vehicle per day) on minor road

Table 11-3 CMF for Shoulder Width for Rural Two-lane Two-way Segments

	AADT		
Shoulder Width	<400	400-2000	>2000
0 feet	1.10	$1.10 + 2.5 \times 10^{-4}(AADT - 400)$	1.50
2 feet	1.07	$1.07 + 1.43 \times 10^{-4}(AADT - 400)$	1.30
4 feet	1.02	$1.02 + 8.125 \times 10^{-5}(AADT - 400)$	1.15
6 feet	1.00	1.00	1.00
8 feet or more	0.98	$0.98 - 6.875 \times 10^{-5}(AADT - 400)$	0.87

Table 11-4 Base Conditions for Rural Multilane Highways

Undivided Roadway Segments (RM4U)	
Lane Width (LW)	12 feet
Shoulder Width (SW)	6 feet
Shoulder Type	Paved
Side slopes	1v:7h or flatter
Lighting	None
Automated Speed Enforcement	None
Divided Roadway Segments (RM4D)	
Lane Width (LW)	12 feet
Right Shoulder Width	8 feet
Median Width	30 feet
Lighting	None
Automated Speed Enforcement	None
Intersections (RM3ST and RM4ST)	
Intersection Skew Angle	0˚
Intersection Left-turn Lanes	0˚, except on stop-controlled approaches
Intersection Right-turn Lanes	0˚, except on stop-controlled approaches
Lighting	None

Table 11-5 SPF for Rural Multilane Highway Facilities

Facility Type	$\boldsymbol{N_{spf}}$
RM4U /RM4D	$AADT^b \times L \times e^a$
RM3ST/ RM4ST/ RM4SG	$\exp[a + b \times \ln(AADT_{\text{maj}}) + c \times \ln(AADT_{\text{min}})]$ or $\exp[a + d \times \ln(AADT_{\text{total}})]$
	N_{spf} = Predicted total crash frequency per year for roadway segment/ intersection for base condition AADT = Average Annual Daily Traffic volume (vehicle per day) on segment L = Length of roadway segment (miles) $AADT_{maj}$ = AADT (vehicle per day) on major road $AADT_{min}$ = AADT (vehicle per day) on minor road

12. APPENDIX B: HSM Predictive Models Calibration Guide

The calibration procedure includes five steps as illustrated in Figure 12-1:

Step 1: Identify facility types for which the appropriate predictive models are to be calibrated.

Step 2: Select sites from selected facility types for calibration.

Step 3: Obtain data for each facility type during a specified calibration period (e.g., 3-year or 5-year).

Step 4: Estimate crash frequency for each site using the HSM predictive method.

Step 5: Compute calibration factors.

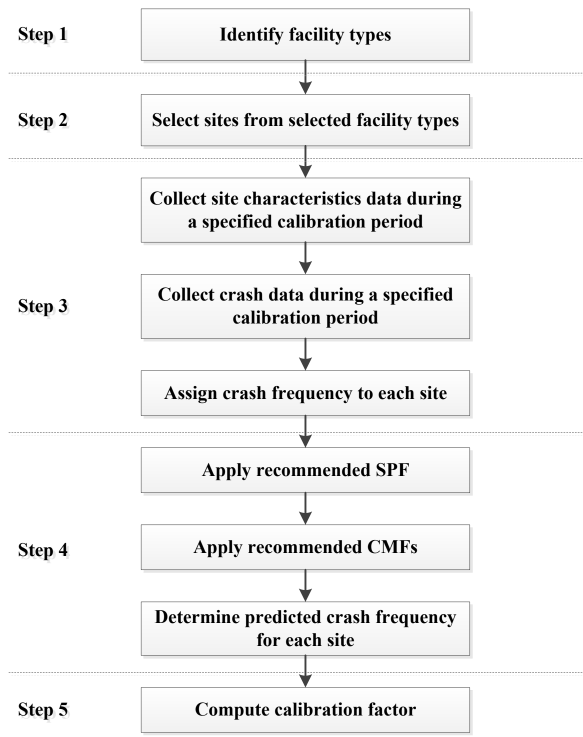


Figure 12-1 Procedure of HSM Predictive Models Calibration

12.1. Step 1: Identify facility types for which the appropriate models are to be calibrated.

The identified facility types for calibration include rural two-lane two-way segments, rural multilane undivided segments, rural multilane divided segments, rural two-lane three-leg intersections with STOP control, rural two-lane four-leg intersections with STOP control and rural multilane four-leg intersections with STOP control.

12.2. Step 2: Select sites from selected facility types for calibration.

The minimum sample size for the calibration data recommended by the HSM is 30 to 50 sites. The HSM also recommends that the entire group of the calibration sites should represent at least 100 crashes per year. In this guide, all sites can be used for the calibration because 1) all sites have the same data items,

2) selection bias resulted from the sampling procedure can be avoided; and 3) calibration bias due to low crash counts can be mitigated.

12.3. Step 3: Obtain data for each facility type during a specified calibration period.

12.3.1. Roadway Segments

All required roadway geometric and traffic characteristics along with crash count for each segment need to be stored in a single data table.

12.3.1.1. Join Site Characteristics Data Tables

All HSM calibration required site characteristics data are stored in eight RIS event tables. They are:

1. A-Admin System Data
2. B-Surface Property Data
3. C-Median Data
4. L-Horizontal Curve Data
5. N-Vertical Curve Data
6. P-Speed Limit Data
7. Traffic Data
8. MRM data

Join all data tables to get one data table appropriate for HSM predictive model application. The table merging sequence is provided in Figure 6-3 RIS Event Table Merging Sequence. Table merging script written in statistical analysis software R is provided to join data tables in such a way that it can only join two data tables at a time. All steps conducted to merge all data tables are summarized below:

Step a. Export all data tables from RIS to a comma separated value file and then import them to Excel and save in the R working directory.

Step b. Using R script to merge "A-Admin System Data" and "P-Speed Limit data". Name the output file from R as "Admin_SpeedLimi_Clip". Note the output file format is comma separated.

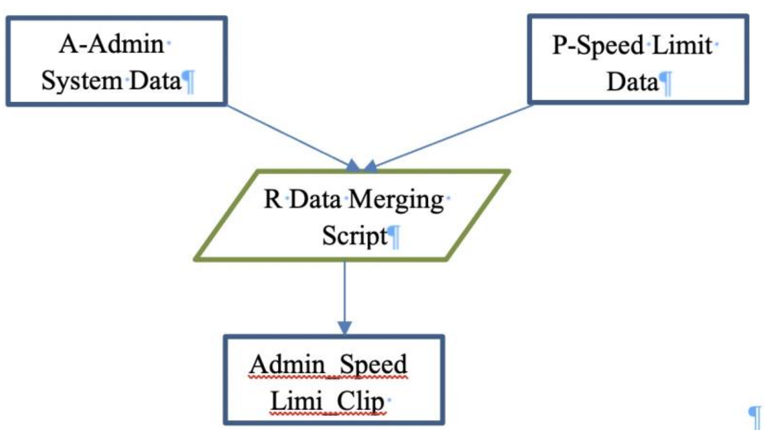


Figure 12-2 Data Table Merging Mechanism.

Step c. Open “Admin_SpeedLimi_Clip” in Excel and save it as an Excel file in the R working directory of the computer.

Step d. Merge next set of data tables using the same R script but user need to rename input data tables and change merging column numbers based on two input data tables to get all the attributes. Change input data table names and column numbers in merging script to get all data attribute after data table merging.

Step e. Use “Admin_SpeedLimi_Clip” and “C-Median Data” as input file in R merging script and name the output data table from R as “Admin_SpeedLimi_Median_Clip.

Step f. Perform Step 3 with “Admin_SpeedLimi_Median_Clip” data table. Perform Step 4 with next set of tables in the merging sequence. Use “Admin_SpeedLimi_Median_Clip” and “B-Surface Property Data” as input data tables and name the output data table from R as “Admin_SpeedLimi_Median_Surface_Clip”.

Step g. Repeat Step 3 and Step 4 following the table merging sequence provided in Figure 6-3 until Traffic data.

Step h. After joining Admin data, speed limit data, median data, surface property data, horizontal curve and vertical curve data, the merged file needs to be joined with Traffic data. Using the data table merging script sample provided in Appendix, use these two data tables as input data tables. The output data table is “Homogeneous Segmentation Road Data Table”.

Step i. Using simple MS-Excel formula, the maximum mileage of each highway is collected from MRM data table. The maximum mileage is then used to calculate segment length for the last segment of each highway.

12.3.1.2. Merge Crash Data

Create roadway network shapefile, a workable file format in ArcMap using linear referencing system developed by SDDOT. Accident table contains crash occurrence location geographic coordinates e.g. Longitude and Latitude. Using location coordinates, convert crash data to point shapefile in ArcMap. This crash data shapefile can be used for both segments and intersections.

To obtain crash frequency for each site, follow the steps below:

Step a. Import both roadway segment shapefile and crash data shapefile to ArcMap.

Step b. Create a unique identification number for each highway segment in segment shapefile.

Step c. Use the “Spatial Join” toolbox as illustrated in Figure 12-3, a build-in analysis tool in ArcMap, to join each segment to a crash with “CLOSEST” as the match option and calculate join distance by creating a new attribute called “Join_Dist” in the “Spatial Join” tool as shown in Figure 12-4.

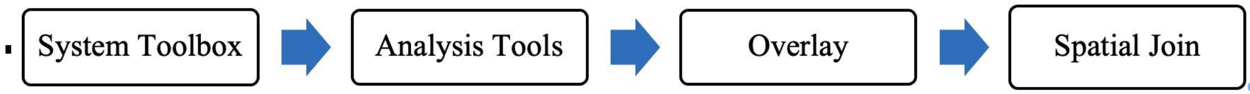


Figure 12-3 Spatial Join Toolbox in ArcMap

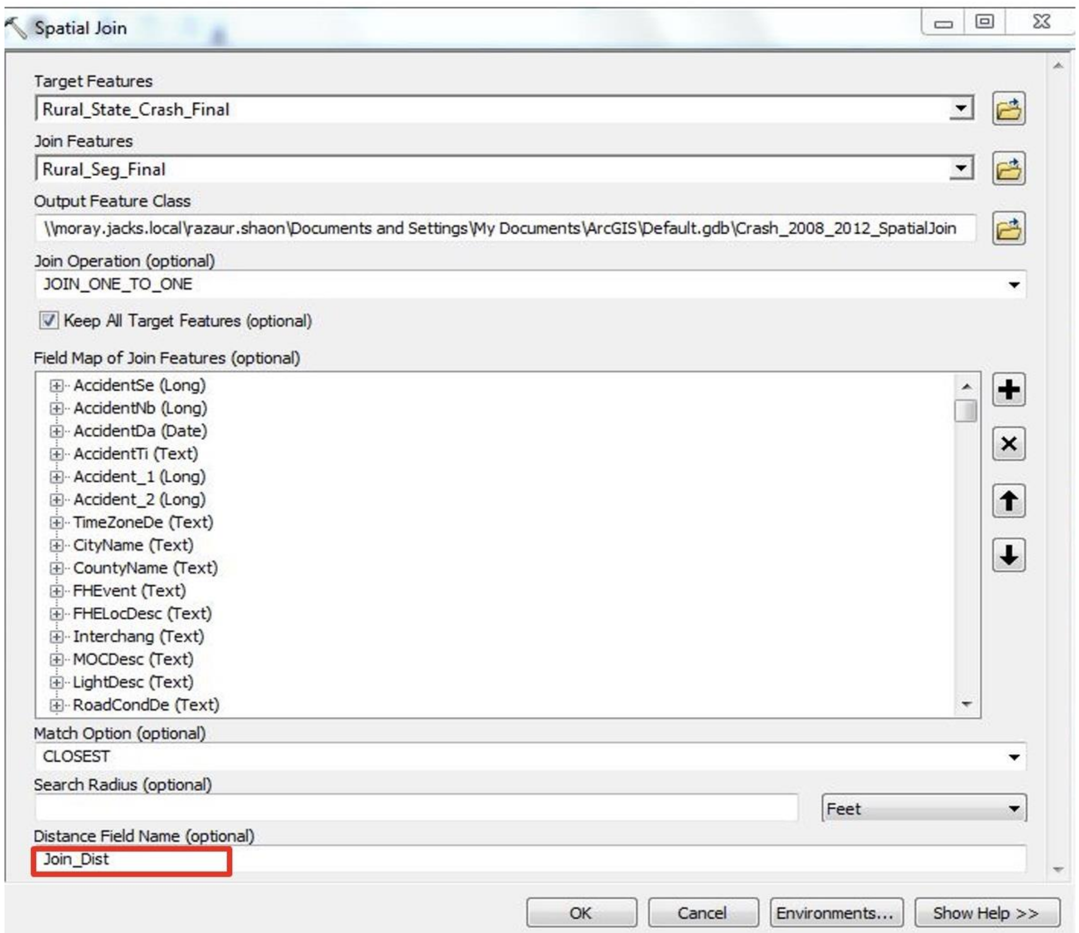
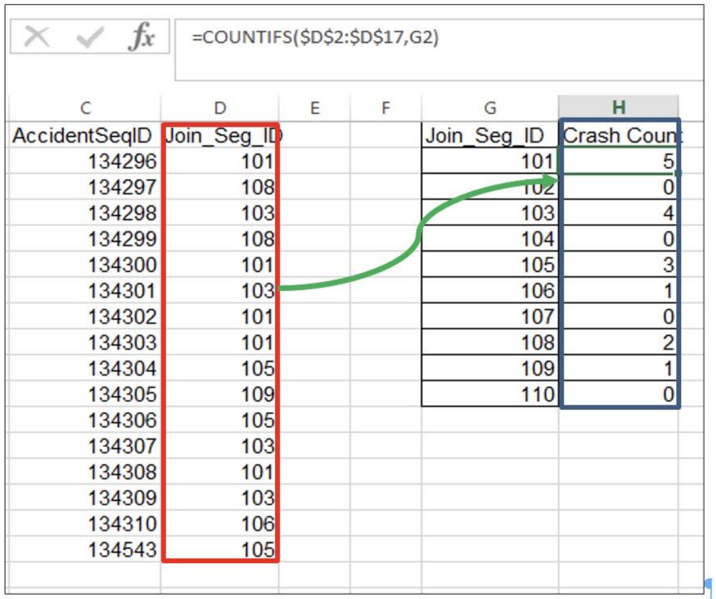


Figure 12-4 Spatial Join Tool Features in ArcMap

Step d. Open spatial join output feature class “dbf” file in MS-Excel and save it in a MS-Excel file.

Step e. Establish join distance threshold value by generating join distance count pivot table in Excel. Filter crashes within join distance threshold (e.g. 50 feet join distance threshold) value.

Step f. After joining roadway data with crash data, each crash has a segment ID. In roadway dataset, use “COUNTIFS” formula in MS-Excel to count the number of crashes occurred on each unique highway segment ID. Figure 12-5 illustrates an example of “COUNTIFS” formula in MS-Excel.



=COUNTIFS($D$2:$D$17,G2)

AccidentSeqID	Join_Seg_ID			Join_Seg_ID	Crash Count
134296	101			101	5
134297	108			102	0
134298	103			103	4
134299	108			104	0
134300	101			105	3
134301	103			106	1
134302	101			107	0
134303	101			108	2
134304	105			109	1
134305	109			110	0
134306	105				
134307	103				
134308	101				
134309	103				
134310	106				
134543	105				

Figure 12-5 Example of "COUTIFS" Formula in MS-Excel

12.3.2. Intersections

"StatetoState_and_StatetoNonStateFedAid" shapefile has all required roadway and traffic data for calibration. To obtain the crash count for each intersection, follow the steps below:

Step a. Import both "StatetoState_and_StatetoNonStateFedAid" shapefile and crash data shapefile to ArcMap.

Step b. Use "Spatial Join" toolbox as shown in Figure 12-3, a build-in analysis tool in ArcMap, to join each intersection to a crash with "CLOSEST" as the match option and calculate join distance by creating a new attribute called "Join_Dist" in the "Spatial Join" tool similar to the example shown in Figure 12-4.

Step c. Open the output file with the extension of "dbf" in MS-Excel software and save it as a MS-Excel file.

Step d. Filter crashes to obtain appropriate intersection-related crashes following two criteria: 1) a crash that happens within a 100-ft. radius of the center of the intersection; 2) any crash occurring within a 200-ft. radius from the intersection that is coded in the accident database as intersection or intersection related (i.e., ILTIntersectFlag=TRUE and non-WVC).

Step e. After joining intersection data to crash data, each crash now has a "Grouping_Number" field. In the intersection worksheet exported from the shapefile, use "COUNTIFS" formula in MS-Excel to count the number of crashes related to one unique "Grouping_Number".

12.4. Step 4: Estimate crash frequency for each site using the HSM predictive method.

Follow the HSM predictive method and apply appropriate SPF models and CMFs to each site with the site characteristics obtained in Step 3. In this step, only use SPF and CMFs. The expected average crash frequency obtained by applying the predictive method is the annual crash frequency, and the calibration duration needs to be considered in the calculation.

12.4.1. Rural Two-lane Two-way Segments

For rural two-lane two-way roadway segments, state-specific base conditions shown in Table 12-1 are different from those defined in the HSM. Use state-specific SPF and CMFs. If there is an adjustment for CMF, use adjusted CMFs.

Table 12-1 South Dakota-Specific Base Conditions for Rural Two-lane Two-way Segments

Road Characteristic	Base Condition
Lane Width (LW)	12 feet
Shoulder Width (SW)	4 feet*
Shoulder Type	Paved
Roadside Hazard Rating (RHR)	3
Driveway Density (DD)	5 driveway per mile
Horizontal Curvature	None
Vertical Curvature	None
Centerline Rumble Strips	None
Passing Lanes	None
Two-way left-turn Lanes	None
Lighting	none
Automated Speed Enforcement	none
Grade Level	0%
* This value is different from that defined in the HSM.	

Safety Performance Function

State-specific SPF is presented in Equation 12-1:

$$N_{spf} = AADT \times L \times 365 \times 10^{-6} \times e^{(-0.1101)} \quad (12\text{-}1)$$

Where:

N_{spf}= predicted total crash frequency for roadway segment base conditions;

AADT = average annual daily traffic volume (vehicles per day); and

L = length of roadway segment (miles).

Crash Modification Factors

The CMF for shoulder width is adjusted and presented in Table 12-2. All the other CMFs remain unchanged, and users can refer to Chapter 10 in the HSM to find them.

Table 12-2 South Dakota-Specific CMF for Shoulder Width for Rural Two-lane Two-way Segments

	AADT		
Shoulder Width	<400	400-2,000	>2,000
0 feet	1.10	$1.10 + 2.5 \times 10^{-4}(AADT - 400)$	1.50
2 feet	1.07	$1.07 + 1.43 \times 10^{-4}(AADT - 400)$	1.30
4 feet	1.02	$1.02 + 8.125 \times 10^{-5}(AADT - 400)$	1.15
6 feet	1.00	1.00	1.00
8 feet or more	0.98	$0.98 - 6.875 \times 10^{-5}(AADT - 400)$	0.87

12.4.2. Rural Multilane Undivided Segments

For rural multilane undivided roadway segments, state-specific base conditions are the same as those defined in the HSM. Apply original HSM SPF and CMFs in HSM Chapter 11.

12.4.3. Rural Multilane Divided Segments

Both lane width and shoulder width in base conditions of South Dakota rural multilane divided segments are different from the HSM. State-specific base conditions are shown in Table 12-3. Under different base conditions, use state-specific SPF and CMFs. If there is an adjustment for CMF, use adjusted CMFs.

Table 12-3 South Dakota-Specific Base Conditions for Rural Multilane Divided Segments

Road Characteristic	Base Condition
Lane Width (LW)	13 feet*
Shoulder Width (SW)	4 feet*
Shoulder Type	paved
Side slopes	1v:7h or flatter
Lighting	none
Automated Speed Enforcement	none
* This value is different from that defined in the HSM.	

Safety Performance Function

State-specific SPF is presented in Equation 12-2:

$$N_{spf} = e^{(-19.7106 + 2.4597 \times \ln(AADT) + \ln(L))} \quad (12\text{-}2)$$

Where:

N_{spf} = predicted total crash frequency for roadway segment base conditions;

$AADT$ = average annual daily traffic volume (vehicles per day); and

L = length of roadway segment (miles).

Crash Modification Factors

Adjust the CMF for shoulder width accordingly. The adjusted CMF for shoulder width is presented in Table 12-4. The CMF for lane width remains the same as the original HSM CMF because it is larger than 12 ft. All the other CMFs in the HSM should be applied.

Table 12-4 South Dakota-Specific CMF for Shoulder Width for Rural Multilane Divided Segments

Shoulder Width	CMF
0 feet	0.082
2 feet	1.037
4 feet	1.000
6 feet	0.954
8 feet or more	0.917

12.4.4. Rural Two-Lane Three-Leg Intersections with STOP Control

The base conditions of rural two-lane three-leg intersections with STOP control in South Dakota are the same as those defined in the HSM. Either HSM SPF or state-specific SPF in Equation 12-3 can be applied. Use original HSM CMFs in Chapter 10.

$$N_{spf} = e^{-9.93+0.66\times ln(AADT_{maj})+0.52\times ln(AADT_{min})} \quad (12\text{-}3)$$

Where:

N_{spf} = predicted total crash frequency for roadway segment base conditions;

$AADT_{maj}$ = average annual daily traffic volume (vehicles per day) for major-road approaches; and

$AADT_{min}$ = average annual daily traffic volume (vehicles per day) for minor-road approaches.

12.4.5. Rural Two-Lane Four-Leg Intersections with STOP Control

The base conditions of rural two-lane three-leg intersections with STOP control in South Dakota are the same as those defined in the HSM. Either HSM SPF or state-specific SPF in Equation 10-4 can be applied. Use original HSM CMFs in Chapter 10.

$$N_{spf} = e^{-10.55+0.79\times ln(AADT_{maj})+0.51\times ln(AADT_{min})} \quad (12\text{-}4)$$

Where:

N_{spf} = predicted total crash frequency for roadway segment base conditions;

$AADT_{maj}$ = average annual daily traffic volume (vehicles per day) for major-road approaches; and

$AADT_{min}$ = average annual daily traffic volume (vehicles per day) for minor-road approaches.

12.4.6. Rural Multilane Four-Leg Intersections with STOP Control

The base conditions haven't been reviewed due to the small data size. It's assumed that the base conditions in South Dakota are the same as those defined in the HSM. SPF and CMFs in Chapter 11 of HSM should be applied.

12.5. Step 5: Compute calibration factors

The final step is to compute the calibration factor as:

$$C_r \text{ or } C_i = \frac{\sum_{all\ sites} observed\ crashes}{\sum_{all\ sites} predicted\ crashes} \quad (12\text{-}5)$$

12.6. Example of Calibration Factor Calculation

Eight rural multilane divided roadway segments have AADT and segment length shown in Column 1 and 2 in Table 12-5. This example is intended solely to illustrate the computation as the number of sites here is below the recommended sample size, which is 30 to 50. All the other site characteristics of these eight sites are the same as state-specific base conditions in Table 12-3, except the lane width and shoulder width, as shown in Column 4 and 6 in Table 12-5.

Table 12-5 Example of Calibration Factor Computation

1	2	3	4	5	6	7	8	9	10
AADT	Length	SPF Prediction	Lane Width	CMF_{1r}	Shoulder Width	CMF_{2r}	Years of Data	Predicted Average Crash Frequency	Observed Crash frequency
1200	1.1	0.114	11	1.03	2	1.037	3	0.364	2
3100	2.3	2.450	12	1	6	0.954	3	7.013	4
1600	1.4	0.293	10	1.23	4	1	3	1.082	1
2300	0.8	0.409	12	1	6	0.954	3	1.171	2
800	0.5	0.019	10	1.09	4	1	3	0.062	1
1700	1.5	0.365	11	1.0425	4	1	3	1.140	3
1500	2.0	0.357	11	1.0375	6	0.954	3	1.061	2
2500	1.8	1.130	13	1	2	1.037	3	3.515	4
						Sum		**15.407**	**19**
						Calibration Factor (C_r)		**1.233**	

The SPF for rural multilane divided roadway segments from Equation 12-2 is:

$$N_{spf} = e^{(-19.7106+2.4597\times\ln(AADT)+\ln(L))}$$

Where:

N_{spf} = predicted total crash frequency for roadway segment base conditions;

$AADT$ = average annual daily traffic volume (vehicles per day); and

L = length of roadway segment (miles).

For the first segment in the example, the predicted crash frequency for base conditions is:

$$N_{base} = e^{(-19.7106+2.4597\times\ln(1200)+\ln(1.1))} = 0.113 \text{ crashes/year}$$

This segment has 11-foot lane width and 2-foot shoulder width. Based on Table 10-8 in the HSM, $CMF_{1r} = 1.01 + 2.5 \times 10^{-5} \times (1200 - 800) = 1.03$. And CMF_{2r} is 1.037 based on Table 12-4.

The predicted crash frequency without calibration is:

$N = N_{base} \times CMF_{1r} \times CMF_{2r} \times (\text{number of years of data}) = 0.113 \times 1.03 \times 1.037 \times 3 = 0.364$ crashes in three years, shown in Column 9.

Similar calculations were conducted for each segment in the table.

The calibration factor is computed as:

$$C_r = \frac{\sum_{all\ sites} observed\ crashes}{\sum_{all\ sites} predicted\ crashes} = \frac{19}{15.407} = 1.233.$$

13. APPENDIX C: R Data-merging Script

```
##Set working directory
setwd("C:\\Users\\razaur.shaon\\Desktop\\HSM")
##Required library to load excel file
require(XLConnect)

##Load data tables
SurfaceXL<- readWorksheet(loadWorkbook("Admin_Calc_Final.xlsx"),sheet=1,header =
TRUE,forceConversion = TRUE)
SurfaceXL.names <- names(SurfaceXL)
names(SurfaceXL)
fix(SurfaceXL)

AdminXL<- readWorksheet(loadWorkbook("Speed Limit_Final_Calc.xlsx"),sheet=1,header =
TRUE,forceConversion = TRUE)
AdminXL.names <- names(AdminXL)
names(AdminXL)
fix(AdminXL)

##Filter & Sort Highways
surface.highway <- unique(SurfaceXL$Highway)
length(surface.highway)  #171
data.frame(surface.highway)
admin.highway <- unique(AdminXL$Highway)
length(admin.highway)           #171
data.frame(admin.highway)
highway <-
merge(data.frame(surface.highway),data.frame(admin.highway),by.x="surface.highway",by.y="admin.hi
ghway")
names(highway)
class(highway)
fix(highway)
highway.nbr <- nrow(highway) #171 identical highways
highway.nbr
highway[3,1]
nrow(highway)
highway[2]

highway[n,1]
n=1
n=2
sort(c(2,1,3))
```

```
i=1
j=1
class(temp)
fix(temp)

for (n in 1:highway.nbr)
{
		SurfaceXL.temp <- SurfaceXL[SurfaceXL$Highway == highway[n,1],]
		AdminXL.temp <- AdminXL[AdminXL$Highway == highway[n,1],]
		SurfaceXL.sort <- SurfaceXL.temp[order(SurfaceXL.temp$Mileage),]
		AdminXL.sort <- AdminXL.temp[order(AdminXL.temp$Mileage),]
		mileage <- unique(c(SurfaceXL.sort$Mileage,AdminXL.sort$Mileage))
		mileage <- sort(mileage)
		mileage.nbr <- length(mileage)
		SurfaceXL.sort.rows <- nrow(SurfaceXL.sort)
		AdminXL.sort.rows <- nrow(AdminXL.sort)
		temp <- matrix(NA, nrow=mileage.nbr,ncol=22)
		temp <- data.frame(temp)
    colnames(temp)[1:10] <- c( "Sadmin_ID","Slimit_ID","Highway","DataClassCode", "Mileage",
    "Admin_MRMNbr", "Slimit_MRMNbr", "Admin_DisplacementNbr",
"Slimit_DisplacementNbr","Speed_Limit")
		colnames(temp)[11:22] <- colnames(SurfaceXL.sort)[7:18]

		temp[,"Mileage"] <- mileage
		temp[,"DataClassCode"] <- 1
		temp[,"Highway"] <- as.character(highway[n,1])

				for (i in 1:SurfaceXL.sort.rows)
				{
						if (i == SurfaceXL.sort.rows)
						{
						SurfaceXL.index <- which (temp[,"Mileage"] >=
SurfaceXL.sort$Mileage[i])
						}
						else
						{
						SurfaceXL.index <- which (temp[,"Mileage"] >=
SurfaceXL.sort$Mileage[i]
											& temp[,"Mileage"] <
SurfaceXL.sort$Mileage[i+1])
						}

		temp[SurfaceXL.index,c("Sadmin_ID","Admin_MRMNbr","Admin_DisplacementNbr")] <-
SurfaceXL.sort[i,c( "Sadmin_ID", "MRMNbr", "DisplacementNbr" )]
```

```
                    temp[SurfaceXL.index,11:22] <- SurfaceXL.sort[i,7:18]
          }

          for (j in 1:AdminXL.sort.rows )
          {
                    if (j == AdminXL.sort.rows)
                    {
                    AdminXL.index <-  which (temp[,"Mileage"] >=
AdminXL.sort$Mileage[j])
                    }
                    else
                    {
                    AdminXL.index <- which (temp[,"Mileage"] >=
AdminXL.sort$Mileage[j]
                                                            & temp[,"Mileage"] <
AdminXL.sort$Mileage[j+1])
                    }
                    temp[AdminXL.index,c("Slimit_ID","Slimit_MRMNbr",
"Slimit_DisplacementNbr","Speed_Limit")]  <- AdminXL.sort[j,c( "Slimit_ID", "MRMNbr",
"DisplacementNbr","SpeedLimitNbr" )]

          }

          if (n == 1)
          {
          join.table <- temp
          }                    else
          {
          join.table <- rbind(join.table,temp)
          }
}
fix(join.table)

write.csv(data.frame(join.table),"export4.csv")
```

Sample 2: Data Table Merging Script for Traffic Data

```
setwd("C:\\Users\\razaur.shaon\\Desktop\\HSM")
require(XLConnect)
merge1<- readWorksheet(loadWorkbook("Semi_Clip_part2.xlsx"),sheet=1,header =
TRUE,forceConversion = TRUE)
merge1.names <- names(SurfaceXL)
```

```
names(merge1)
fix(merge1)
freememory()

###Input Admin Data
curve<- readWorksheet(loadWorkbook("Traffic_2.xlsx"),sheet=1,header = TRUE,forceConversion =
TRUE)
curve.names <- names(curve)
names(curve)
fix(curve)
ncol(curve)
##Filter & Sort Highways
merge1.highway <- unique(merge1$Highway)
length(merge1.highway)  #171
data.frame(merge1.highway)
curve.highway <- unique(curve$Highway)
length(curve.highway)#171
data.frame(curve.highway)
highway <-
merge(data.frame(merge1.highway),data.frame(curve.highway),by.x="merge1.highway",by.y="curve.hig
hway")
names(highway)
class(highway)
fix(highway)
highway.nbr <- nrow(highway) #171 identical highways
highway.nbr
highway[3,1]
nrow(highway)
highway[2]

highway[n,1]
n=1
n=2
sort(c(2,1,3))
i=1
j=1
class(temp)
fix(temp)

## Merge Tables of Surface & Admin Data
for (n in 1:highway.nbr)
{
        merge1.temp <- merge1[merge1$Highway == highway[n,1],]
        curve.temp <- curve[curve$Highway == highway[n,1],]
```

```
            merge1.sort <- merge1.temp[order(merge1.temp$Mileage),]
            curve.sort <- curve.temp[order(curve.temp$MRMNbr),]
            Merge_Mileage <- merge1.sort$Mileage
            Merge_Mileage <- sort(Merge_Mileage)
            Merge_MRM <- merge1.sort$Join5_MRM
            Merge_MRM <- sort(Merge_MRM)
            MRM.nbr <- length(Merge_Mileage)
            merge1.sort.rows <- nrow(merge1.sort)
            curve.sort.rows <- nrow(curve.sort)
            temp <- matrix(NA, nrow=MRM.nbr,ncol=61)
            temp <- data.frame(temp)
    colnames(temp)[1:7] <- c( "Mer2_ID","Highway","DataClassCode","Traffic_ID", "Merge_MRM",
"Merge_Mileage","Mer_Displacement")
            colnames(temp)[8:52] <- colnames(merge1.sort)[7:51]
            colnames(temp)[53:61] <- colnames(curve.sort)[8:16]
            temp[,"Merge_MRM"] <- Merge_MRM
            temp[,"Merge_Mileage"] <- Merge_Mileage
            temp[,"DataClassCode"] <- 1
            temp[,"Highway"] <- as.character(highway[n,1])

for (i in 1:merge1.sort.rows)
                        {
                                    if (i == merge1.sort.rows)
                                    {
                                    merge1.index <- which (temp[,"Merge_Mileage"] >=
merge1.sort$Mileage[i])
                                    }
                                    else
                                    {
                                    merge1.index <- which (temp[,"Merge_Mileage"] >=
merge1.sort$Mileage[i]
                                                                                        &
temp[,"Merge_Mileage"] < merge1.sort$Mileage[i+1])
                                    }
                                    temp[merge1.index,c("Mer2_ID","Mer_Displacement")] <-
merge1.sort[i,c( "Mer2_ID", "Join5_Displacement")]
                                    temp[merge1.index,8:52] <- merge1.sort[i,7:51]
                        }




for (j in 1:curve.sort.rows )
                        {
```

```
                    if (j == curve.sort.rows)
                    {
                    curve.index <-  which (temp[,"Merge_MRM"] >=
curve.sort$MRMNbr[j])
                    }
                    else
                    {
                    curve.index <- which (temp[,"Merge_MRM"] >=
curve.sort$MRMNbr[j]
                                                    & temp[,"Merge_MRM"]
< curve.sort$MRMNbr[j+1])
                    }

                    temp[curve.index,"Traffic_ID"] <- curve.sort[j, "Traffic_ID"]
                    temp[curve.index,53:61] <- curve.sort[j,8:16]
             }



             if (n == 1)
             {
             join.table <- temp
             }              else
             {
             join.table <- rbind(join.table,temp)
             }
}
fix(join.table)
write.csv(data.frame(join.table),"export120.csv")
```