FHWA-NJ-2019-007

Calibration/Development of Safety Performance Functions for New Jersey

FINAL REPORT

October 2019

Submitted by

Kaan Ozbay, Ph.D.
Professor
Department of Civil and Urban Engineering
Tandon School of Engineering
New York University

Hani Nassif, Ph.D.
Professor
Department of Civil and Environmental Engineering
Rutgers University

Bekir Bartin, Ph.D.
Department of Civil and Environmental Engineering
Rutgers University

Chuan Xu, Ph.D.
Visiting Research Scholar
C2SMART Center,
New York University

Abhinav Bhattacharyya, M.Sc.
Graduate Research Assistant
Department of Civil and Urban Engineering
Tandon School of Engineering
New York University

NJDOT Research Project Manager
Priscilla Ukpah

In cooperation with

New Jersey
Department of Transportation
Bureau of Research
And
U.S. Department of Transportation
Federal Highway Administration

DISCLAIMER STATEMENT

TECHNICAL REPORT
STANDARD TITLE PAGE

Report No. FHWA-NJ-2019-007	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle FINAL REPORT Calibration/Development of Safety Performance Functions for New Jersey		5. Report Date October 2019 6. Performing Organization Code Rutgers / NYU
7. Author(s) Kaan Ozbay, Ph.D., Hani Nassif, Ph.D., Bekir Bartin, Ph.D., Chuan Xu, Ph.D. and Abhinav Bhattacharyya, M.Sc.		8. Performing Organization Report No.
9. Performing Organization Name and Address Rutgers University 96 Frelinghuysen Road Piscataway, NJ 08854	New York University 15 Metrotech Center, 6th Floor Brooklyn, NY 11201	10. Work Unit No. 11. Contract or Grant No.
12. Sponsoring Agency Name and Address New Jersey Department of Transportation PO 600 Trenton, NJ 08625	Federal Highway Administration U.S. Department of Transportation Washington, D.C.	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

Safety Performance Functions (SPFs) in the Highway Safety Manual (HSM) were developed using historic crash data collected in different states. Because local state or geographic conditions vary, to make the SPFs better accommodate the local data, two strategies are usually undertaken: the first strategy is to calibrate SPFs provided in HSM so that the contents of HSM can be fully leveraged and the second strategy is to develop location-specific SPFs regardless of the predictive modeling framework in the HSM.

The main objective of this research project is to (1) calibrate the SPFs provided in the HSM using New Jersey (NJ) data and (2) develop new NJ-specific SPFs as appropriate. The facility types considered for this research project include segments and intersections of rural two-lane two-way, rural multilane, and urban and suburban roads. The following tasks were completed to achieve the main project objective:

- Conducted an in-depth review of the relevant studies in the literature
- Identified the key sources of data required for calibration and development of SPFs. These include roadway characteristics data, traffic volume data, and crash data.
- Developed a computer code to read and process the compiled database to (a) filter out inconsistent data entries, (b) identify facility types, (c) execute roadway segmentation process, (d) assign crash statistics for each facility, and (e) generate a complete database for each facility type to be used in calibration and/or development of SPFs.
- Provided recommendations to improve data collection and recording practices that would facilitate easier data extraction required for the SPF calibration/development process.

17. Key Words Safety Performance Functions, Highway Safety Manual, Calibration Factor, Segments, Intersections			
19. Security Classif (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 171	22. Price

Form DOT F 1700.7 (8-69)

ACKNOWLEDGEMENTS

The authors wish to thank various staff members without whom this research would not have been possible. Special thanks to the project customer Daniel LiSanti of the Bureau of Safety, Bicycle and Pedestrian Programs, the project manager Priscilla Ukpah of the Bureau of Research and Amanda Gendek, Manager, the Bureau of Research, Chris Zajac of the Bureau of Transportation Data and Support, Tan Virgilio of the Bureau of Safety, Bicycle and Pedestrian Programs and Dr. Camille Kamga, Director of University Transportation Research Center (UTRC) Region II at City University of New York. The authors would like to thank DRIVE Engineering for their invaluable help in the data extraction process. The authors also thank Keith Skilton from the FHWA for his support.

TABLE OF CONTENTS

	Page
EXECUTIVE SUMMARY	1
INTRODUCTION	10
Objectives	11
LITERATURE REVIEW	15
Background	15
Calibration of SPFs	16
Development of SPFs	17
Data Requirements	18
The Choice Whether to Calibrate or Develop SPFs	20
Previous Work	20
Interviews	33
Findings	34
AVAILABLE DATA SOURCES	37
Traffic Volume Data	37
Sensor Database	37
TMC Database	38
Roadway Features Data	39
Straight Line Diagrams Database	39
NJ Roads Centerlines Geographic Information Systems Geodatabase	41
Google Street View	42
Voyager Safety Crash Data	42
DATA PREPARATION AND CLEANING	44
Traffic Volume Data	44
Inconsistencies in the Sensor Database	44
Missing AADT Data	45
Roadway Features Data	45
Preparing Intersection Database Using the SLD Database	45
Inconsistencies and Missing Information in the SLD Database	52
Voyager Safety Crash Database	54
PROCESSING DATA	57
Generating The Intersection Database	57
Generating the Homogeneous Segments Database	58
RURAL TWO-LANE TWO-WAY SEGMENTS	61
Data Requirements	61
Gathering and Processing the Roadway Feature Dataset	62
Extracting Horizontal Curvature Data	62
Automatic Identification of Segments	65
Calibration Results	67
Development Results	68
RURAL TWO-LANE TWO-WAY INTERSECTIONS	71
Data Requirements	71
Gathering and Processing the Roadway Feature Dataset	71
Automatic Identification of Intersection Types	72

Manual Data Extraction and Validation 73
Intersection Crash Frequency and Traffic Counts 76
R2 Intersection Dataset 76
Detailed Description of Sampling Results 78
Calibration Results........ 81
Development Results........ 84
RURAL MULTILANE SEGMENTS........ 87
Data Requirements........ 87
Gathering and Processing the Roadway Feature Dataset........ 87
Automatic Identification of Segments 88
Calibration Results........ 90
URBAN AND SUBURBAN SEGMENTS 92
Data Requirements........ 92
Gathering and Processing the Roadway Feature Dataset........ 93
Automatic Identification of Segment Types 93
Manual Data Extraction and Validation 95
Calibration Results........ 99
Development Results........ 100
URBAN AND SUBURBAN INTERSECTIONS 104
Data Requirements........ 104
Gathering and Processing the Roadway Feature Dataset........ 105
Automatic Identification of Intersection Types 105
Manual Data Extraction and Validation 106
Intersection Crash Frequency and Traffic Counts 112
Urban and Suburban Intersection Dataset........ 112
Detailed Description of Sampling Results 114
Calibration Results........ 119
Development Results........ 122
CONCLUSIONS........ 127
REFERENCES........ 133
APPENDIX A – SPFS DEVELOPED BY OTHER STUDIES........ 139

LIST OF FIGURES

Page

Figure 1. Calibration process recommended by the HSM [2] 17
Figure 2: NJDOT sensor database and its key data elements 38
Figure 3: A sample of TMC count data provided in *.csv format.......... 39
Figure 4: NJ GIS Links database and its key data elements 41
Figure 5: Two types of 3-way intersections 47
Figure 6: Identifying intersection types using NJ GIS Links database 47
Figure 7: Decision flowchart for identifying location and type of intersections 48
Figure 8: Example signalized one-directional 3-way intersection on Route 22.......... 50
Figure 9: Example of a missing link in the NJ GIS Links database.......... 51
Figure 10: Center two-way left-turn lane on Route 41 in Deptford Township 53
Figure 11: Decision flowchart for determining missing SRI and milepost information .. 56
Figure 12: Generating the intersection database for each facility type 58
Figure 13: Generating homogeneous segments database for each facility type 59
Figure 14: Hypothetical roadway 64
Figure 15: Flowchart for automatic detection of R2U.......... 66
Figure 16: CURE plot of R2U segments with respect to AADT and segment length.... 68
Figure 17: Observed vs. predicted crashes – NJ-specific SPF for R2 segments 69
Figure 18: Observed vs. predicted crashes – calibrated SPF for R2 segments 69
Figure 19: Flowchart for identifying the type of data-ready R2 intersections 72
Figure 20: Importing and visualizing the R2 intersections 74
Figure 21: Selecting an intersection from the list.......... 74
Figure 22: Top view of the selected intersection 75
Figure 23: Street view of selected intersection 76
Figure 24: Spatial distribution of R23ST intersections.......... 78
Figure 25: Spatial distribution of R23SG intersections 79
Figure 26: Spatial distribution of R24ST intersections.......... 80
Figure 27: Spatial distribution of R24SG intersections 81
Figure 28: CURE plot of R23ST with respect to $AADT_{maj}$ and $AADT_{min}$.......... 83
Figure 29: CURE plot of R24ST with respect to $AADT_{maj}$ and $AADT_{min}$.......... 83
Figure 30: CURE plot of R24SG with respect to $AADT_{maj}$ and $AADT_{min}$ 84
Figure 31: Automatic identification of RM segments.......... 89
Figure 32: CURE plots for R4U with respect to AADT and segment length 90
Figure 33: CURE plots for R4D with respect to AADT and segment length 91
Figure 34: Urban and suburban segment types 92
Figure 35: Flowchart for automatically identifying urban and suburban segments 94
Figure 36: Presence of a driveway (minor residential) in the segment 97
Figure 37: Presence of a connecting road in the segment 97
Figure 38: Example of a minor commercial driveway 98
Figure 39: Example of a parking space 98
Figure 40: Example of a parking spot marked X.......... 99
Figure 41: CURE plots for (a) U2U, (b) U4U, and (c) U4U 100
Figure 42: Flowchart of process for automatically identifying urban and suburban intersections 106

Figure 43: A rural-like intersection 108
Figure 44: Presence of lighting in an intersection 109
Figure 45: A signalized intersection 109
Figure 46: An example of a three-leg intersection 111
Figure 47: A sign indicating that left-turns are allowed on green 111
Figure 48: No-turn-on-red sign 112
Figure 49: The spatial distribution of U3ST intersections 115
Figure 50: The spatial distribution of U3SG intersections 116
Figure 51:The spatial distribution of U4ST intersections 117
Figure 52: The spatial distribution of U4SG intersections 118
Figure 53: CURE plots for U3ST 121
Figure 54: CURE plots for U3SG 121
Figure 55: CURE plots for U4ST 122
Figure 56: CURE plots for U4SG 122

LIST OF TABLES

Page

Table 1: Facility types used in the HSM 13
Table 2: Data requirements for HSM safety predictions [6] 19
Table 3: Pros and cons of the two different SPF strategies 20
Table 4: General available data sources for New Jersey 37
Table 5: Description of key SLD tables 40
Table 6: Key data elements of the Voyager Safety Crash database 42
Table 7: Crash data statistics 43
Table 8: Examples of sensor database errors 44
Table 9: Comparison of pt_intersection and pt_int_approach tables 46
Table 10: Summary of conventional intersections found in pt_int_approach table 49
Table 11: Type of 3-way intersections found in the pt_int_approach table 50
Table 12: Summary of additional conventional intersections 52
Table 13: Summary of additional 3-way intersections by type and signalization 52
Table 14: Statistics on Voyager crash data missing SRI and milepost information 55
Table 15: Data requirements for R2U segments 61
Table 16: R2U segments statistics 67
Table 17: Development results of SPF for R2 segments 68
Table 18: Data requirements for R2 intersections 71
Table 19: Data elements used to identify intersection attributes 73
Table 20: Sample size of preliminary selection and final selection 76
Table 21: Sample size by county of preliminary selection and final selection 77
Table 22: Summary of intersection AADT data 77
Table 23: SPFs of three types of R2 intersections in the HSM 82
Table 24: Calibration factors of R2 intersections 83
Table 25: Development results for R23ST 84
Table 26: Development results for R23SG 85
Table 27: Development results for R24ST 85
Table 28: Development results for R24SG 85
Table 29: Summary of SPFs for R2 intersections 86
Table 30: Data requirements for RM segments 87
Table 31: RM segments statistics 89
Table 32: Data requirements for urban and suburban segments 92
Table 33: Urban and suburban segment statistics 94
Table 34: Data extracted for urban and suburban segments 94
Table 35: Calibration factors of urban and suburban segments 100
Table 36: Estimated coefficients of developed SPFs 101
Table 37: SPF development results for U2U 101
Table 38: SPF development results for U4U 102
Table 39: SPF development results for U4D 102
Table 40: Vehicle-pedestrian and vehicle-bicycle collision factors at urban and suburban segments 103
Table 41: Data requirements for R2 intersections 104
Table 42: Number of left-turn approaches at a signalized intersection 110
Table 43: Number of left-turn approaches at a minor stop-controlled intersection 110

Table 44: Sample size of preliminary selection and final selection 112
Table 45: Sampled size by county of preliminary selection and final selection 113
Table 46: Summary of intersection AADT data 113
Table 47: Multi-vehicle collision SPFs for urban intersections in the HSM.................. 119
Table 48: Single-vehicle collision SPFs for urban intersections in the HSM 119
Table 49: Vehicle-pedestrian collision SPFs for urban intersections in the HSM........ 120
Table 50: Vehicle-bicycle collision SPFs for urban intersections in the HSM.............. 120
Table 51: Calibration factors of urban and suburban Intersections 121
Table 52: SPF functions developed for total, multi and single vehicle collisions......... 123
Table 53: SPF development results for total, multi and single vehicle collisions at U3ST 123
Table 54: SPF development results for total, multi and single vehicle collisions at U3SG 124
Table 55: SPF development results for total, multi and single vehicle collisions at U4ST 124
Table 56: SPF development results for total, multi and single vehicle collisions at U4SG 125
Table 57: SPF development for vehicle-bicycle collisions at urban and suburban intersections 125
Table 58: SPF development for urban intersections – PV 126
Table 59: SPF development results for vehicle-pedestrian collisions at U3SG and U4SG 126
Table 60: Sample size of each facility type used for calibration and development 129
Table 61: Calibration factors for each facility type 130
Table 62: Developed SPFs for segments 131
Table 63: Developed SPFs for intersections.. 131

EXECUTIVE SUMMARY

The predictive models provided by the Highway Safety Manual (HSM) are based on safety performance functions (SPFs), which are statistical regression models based on observed crash data from similar facility types that estimate the predicted average crash frequency for the base conditions. To account for differences between the base conditions and the specific conditions of a facility site, crash modification factors (CMFs) are utilized to adjust the prediction to account for the geometric design and traffic control features of the specific site.

The SPFs in the HSM were developed using historic crash data collected over a number of years at sites of the same facility types in various states. To make the SPFs better accommodate local data, the first strategy is to calibrate SPFs provided in HSM and the second strategy is to develop location-specific SPFs regardless of the predictive modeling framework in the HSM.

The main objective of this research project is to (1) calibrate the SPFs provided in the HSM using New Jersey (NJ) data and (2) develop new NJ-specific SPFs as necessary.

The facilities considered for this research project include the following types:

Facility	**Segments**	**Intersections**
Rural Two-Lane Two-Way Roads	Undivided rural 2-lane 2-way segments (R2U)	Unsignalized 3-way intersections (R23ST)
		Unsignalized 4-way intersections (R24ST)
		Signalized 3-way intersections (R23SG)
		Signalized 4-way intersections (R24SG)
Rural Multilane Highways	Undivided rural multilane highways (R4U)	Unsignalized 3-way intersections (RM3ST)
	Divided rural multilane highways (R4D)	Unsignalized 4-way intersections (RM4ST)
		Signalized 3-way intersections (RM3SG)
		Signalized 4-way intersections (RM4SG)
Urban and Suburban Arterials	Undivided 2-lane arterials (U2U)	Unsignalized 3-way intersections (U3ST)
	Undivided 3-lane arterials with a center two-way left-turn lane (U3T)	Unsignalized 4-way intersections (U4ST)
	Undivided 4-lane arterials (U4U)	Signalized 3-way intersections (U3SG)
	Divided 4-lane arterials (U4D)	Signalized 4-way intersections (U4SG)
	Undivided 4-lane arterials with a center two-way left-turn lane (U5T)	

Several tasks were completed to achieve the main project goals. An outline of this report follows, including succinct descriptions of these tasks.

The background on the SPF calibration and development process is presented in the Background subsection of the Literature Review. The required and desired data for calibration according to the HSM are presented in Table 2 of the Data Requirements subsection. The research team conducted an in-depth review of the studies in the literature that focused on the calibration of the SPFs used in the HSM and the development of new SPFs. A detailed review of these studies and the related calibration and estimation issues are presented in the Previous Work subsection. The key findings from this comprehensive literature review are presented in the Findings subsection. The team also interviewed researchers who conducted similar research projects for the Departments of Transportation (DOTs) of Pennsylvania, South Carolina, Missouri, Kansas, Kentucky, and New York, as discussed in the Interviews subsection.

It should be noted that the process of SPF calibration and/or development is highly data driven. There are 60 variables used in the HSM predictive models, 41 of which are required, and the remainder of which are desirable. The majority of these 60 variables are not commonly collected, such as the number of driveways by type, horizontal and vertical curvature, number of left- and right-turn lanes at intersections, parking information, etc. Because collecting or extracting these data and also integrating data from various sources require a considerable amount of time and man-hours, it is crucial to acquire as much data as possible automatically from existing databases. This requires identifying all possible relevant databases maintained by the New Jersey Department of Transportation (NJDOT) and automatically extracting the required/desired data without having to spend excessive amounts of time on manual extraction.

The research team, being experienced with many of the databases maintained by NJDOT, was able to identify the available data sources. As mentioned in the Available Data Sources section, existing data are grouped into three categories by type: (1) traffic volume, (2) roadway features, and (3) roadway crashes. Information regarding each data source is summarized in Table 4.

Traffic volume data include the sensor database maintained by the New Jersey Traffic Monitoring Program at NJDOT, and hourly turning movement counts collected at various intersections.

Roadway features data were extracted from three data sources: the Straight Line Diagrams (SLD) database, Geographic Information Systems (GIS) maps, and Google Street View. NJDOT's SLD database is the richest source of information for roadway features. This database was provided by NJDOT in MS Access® format. It includes various tables on different geometric and operational aspects of NJ roadways, as shown in Table 5.

The crash data were provided by NJDOT for 2011 to 2015. Table 6 lists the key data elements used as part of this study. Note that the database only includes the reportable crashes. A reportable crash is one that results in the injury or death of any person, or damage to property of any one person in excess of $500. As discussed in the Findings subsection of the Literature Review, some studies discussed the impact of the crash reporting threshold on the scale of the calibration factors. Crash data from Washington and California were used in developing the HSM predictive models, and any deviation from the baseline crash reporting threshold would affect the estimation results for property damage crashes.

Also note that the location of a crash, namely the Standard Route Identifier (SRI) and milepost, is usually the most incomplete part of the crash database, making it difficult, if not impossible, to provide accurate location data. Similar issues were reported by other researchers, as mentioned in the Literature Review section.

The Data Preparation and Cleaning section describes the cleaning process for the three available datasets described above. This section also discusses various observed inconsistencies and inaccurate information in the available data sources. The most time-consuming tasks in this process were preparing the final intersection database and crash databases.

As to the intersection database, correct identification of intersection types and their locations are crucial because intersections are used both in determining homogeneous roadway segments and generating the list of intersections per facility type. To be specific, when the task is to determine homogeneous roadway segments, any point along the roadway that would make it possible for vehicles to make a turn is labeled as an intersection. In addition to the conventional intersections, median left-turns/U-turns, circles, jug handles, and interchanges are also deemed as intersections. In contrast, when the task is to generate the list of intersections per facility type, the focus is to identify only the conventional 3-way and 4-way intersections, and to dismiss any other types of intersections.

As to preparing the crash database, as mentioned above, the crash location information is usually the most incomplete part of the database. Crash entries are often missing SRI or milepost information, or latitude and longitude readings. Latitude and longitude readings of crashes are of utmost importance for assigning crashes to intersections. To determine whether a crash is intersection-related, one must determine if its location is within 250 ft. of the intersection, as suggested the HSM [(2)]. The milepost information for crashes is deemed too coarse for such calculations. To assign crashes to homogeneous segments, however, SRI and milepost information is required, because determining which segment a crash should be assigned to is not straightforward when using the latitude and longitude readings. Inspection of the available crash database revealed that even though latitude and longitude information is available for nearly 95 percent of crashes, there are many instances in which the SRI or milepost information is missing. Table 14 shows the statistics for missing information in the crash database for 2011 to 2015. The research team developed a code in C programming language to determine the missing SRI or milepost information using the available latitude and longitude data, as shown in Figure 11.

Once the available datasets were gathered and cleaned, the research team developed a computer code in R programming language to read and process the compiled database to (a) filter out inconsistent data entries, (b) identify facility types, (c) execute the roadway segmentation process, (d) assign crash statistics for each facility, and (e) generate a complete database for each facility type to be used in calibration and/or development of SPFs. The generic procedure for generating an intersection database per facility type is shown in Figure 12. Generating a homogeneous roadway database for each facility was a more challenging process than that for an intersection database. A homogeneous segment starts at an intersection or at any point where various geometric and operational features of the roadway change at either direction of the facility. The generic procedure for generating homogeneous database per facility type is shown in Figure 13.

In addition to the automatically generated data, the research team conducted an extensive data extraction process to meet the data requirements of the HSM. The following table lists the data elements manually extracted by the research team because they were required by the HSM but not available in the existing databases.

Facility Type	**Extracted Data**
R2 intersections	Number of left- and right-turn lanes Presence of lighting
Urban and suburban intersections	Number of left- and right-turn lanes Presence of right-turn-on-red signs Presence and type of left-turn signal phasing Presence of lighting
Urban and suburban segments	Number of driveways by type On-street parking by type

As mentioned in the Findings subsection of the Literature Review section, among the required dataset, the horizontal curvature data (i.e., curve radius and length) for two-lane two-way rural roadways stand out as the most problematic. With few exceptions, state DOTs do not have a curvature dataset in a usable format that can be easily integrated into the process of finding homogeneous segments. The lack of this dataset led researchers to assume default values for horizontal curvature CMFs, omit sections with horizontal curvature, or manually extract this information using Google Earth™ (e.g., references or built-in plans).

The research team developed a clustering method for automatically estimating horizontal curvature data and CMFs using GIS roadway shapefiles. The clustering method identifies distinct sections on a roadway, either curved or tangent, based on the proximity of the approximated curvature values of data points from GIS roadway centerline shapefiles, and calculates horizontal curvature data and the corresponding CMFs. In fact, a comparison of this clustering method with the manual extraction method and the use of mobile asset vehicles was presented in a journal paper [(78)]. The results of this paper showed that the CMFs estimated by the clustering method were within 12.2 percent and 15.5 percent of those produced by the mobile asset vehicle and the manual data extraction method, respectively. In addition, the sensitivity of the manually extracted horizontal curvature data was examined by conducting 15 additional independent trials. The average percent difference in the calculated CMFs between trials was 15.5 percent. The study therefore concluded that the clustering method can produce CMF estimates as accurate as those yielded by the two other methods and can do so much more efficiently in terms of time and cost.

The table below presents the sample size for each facility type used for calibration and/or development of SPFs for segments and intersections separately.

It should be mentioned that the dataset used for calibration and development includes not only state but also local roads in New Jersey.

SEGMENTS		
Facility Type	**Sample Size**	**SPF**
R2U	756	Calibration and Development
R4U	32	Calibration
R4D	34	Calibration
U2U	459	Calibration and Development
U3T	n/a	n/a
U4U	514	Calibration and Development
U4D	387	n/a
U5T	n/a	Calibration and Development
INTERSECTIONS		
Facility Type	**Sample Size**	**SPF**
R23ST	314	Calibration and Development
R24ST	149	Calibration and Development
R23SG	15	Development
R24SG	45	Calibration and Development
RM3ST	3	n/a
RM4ST	1	n/a
RM3SG	0	n/a
RM4SG	6	n/a
U3ST	227	Calibration and Development
U4ST	121	Calibration and Development
U3SG	164	Calibration and Development
U4SG	209	Calibration and Development

Model calibration in the HSM is performed by applying a multiplicative factor to the given SPF so that the aggregate number of predicted crashes is equal to the aggregate number of observed crashes within a jurisdiction. A calibration factor allows the SPF to keep its original model form. As discussed in the appendix to Part C of the HSM, selected samples are used to find the calibration factor that will make the aggregate predicted crash frequency equal to the observed total in the jurisdiction. The HSM recommends using a minimum sample size of 30 to 50 sites [(2)].

As seen in the table above, among the original facility types, two segment types, namely U3T and U5T, and rural multilane intersections could not be included in the calibration and development process due to lack of data. As explained in the Urban and Suburban Segments section, U3T and U5T segment types that include center two-way left-turn lanes are not indicated in the SLD database. A future version of the SLD database should have a separate table indicating whether roadway segments include two-way left-turn lanes. Regarding rural multilane intersections, the team could not identify an adequate number of such intersections for calibration. Therefore, rural multilane intersections are not included in the report. Similar data issues were observed for rural multilane segments. As noted in the Multilane Segments section, the scarcity of multilane rural roadway data is to be expected because NJ is a densely populated state, and, based on the SLD database, 86.8 percent of its roadway segments are in urban areas.

To calculate a calibration factor, the observed crash frequency and the predicted crash frequency for each intersection and segment are required. The observed crash frequency was calculated for each segment as explained in the Processing Data section. The predicted crash frequency can be calculated using the SPF and the corresponding CMF values given in the HSM.

The Calibrator tool developed by the Federal Highway Administration (FHWA) was used to calculate the calibration factors and measure their goodness of fit [(77)].

The table below presents the calibration factors calculated for each facility type.

SEGMENTS			
Facility Type	**Calibration Factor**	**Standard Error**	**Coefficient of Variation**
R2U	1.55	±0.12	0.08
R4U	1.12	±0.42	0.38
R4D	1.70	±0.80	0.47
U2U	1.264	±0.14	0.11
U3T	n/a	n/a	n/a
U4U	1.097	±0.15	0.13
U4D	1.596	±0.21	0.13
U5T	n/a	n/a	n/a
INTERSECTIONS			
Facility Type	**Calibration Factor**	**Standard Error**	**Coefficient of Variation**
R23ST	0.88	±0.08	0.09
R24ST	0.88	±0.11	0.13
R23SG*	-	-	-
R24SG	0.85	±0.16	0.18
RM3ST	n/a	n/a	n/a

RM4ST	n/a	n/a	n/a
RM3SG	n/a	n/a	n/a
RM4SG	n/a	n/a	n/a
U3ST	2.61	$\pm$0.29	0.11
U3SG	3.60	$\pm$0.36	0.10
U4ST	1.66	$\pm$0.25	0.15
U4SG	4.25	$\pm$0.40	0.09

Note: *R23SG intersection type is not included in the HSM.

In addition, to assess the validity of the calculated calibration factors, the team used cumulative residual (CURE) plots, which are simply graphs of the cumulative residuals (observed minus predicted crashes) against variables of interest. The residuals between the estimated and observed values are assumed to be independent random variables. It is presumed that the CURE plots should be within the expected limits of an unbiased random walk, i.e., plus/minus two standard deviations. CURE plots for each facility type are presented in their respective sections throughout the report.

Using the same data used for calibration, the research team developed NJ-specific SPFs for facilities with a sufficient number of data points using the crash data from 2011 to 2015. SPFs were estimated based on the negative binomial model suggested by the HSM. The model estimation was performed in R statistical package.

The tables below present the SPFs developed for segments and intersections per facility type and crash type, where applicable, respectively.

Segment	Crash Type	Developed SPFs
R2U	Total	$N_{TOT\ R2U} = exp[-6.41 + 0.83. ln(AADT) + 0.86. ln(L)]$
U2U	Total	$N_{TOT\ U2U} = exp[-9.798 + 1.188. ln(AADT) + ln(L)]$
	Multi-Vehicle	$N_{MV\ U2U} = exp[-14.411 + 1.641. ln(AADT) + ln(L)]$
	Single-Vehicle	$N_{SV\ U2U} = exp[-3.977 + 0.435. ln(AADT) + ln(L)]$
U4U	Total	$N_{TOT\ U4U} = exp[-12.01 + 1.432. ln(AADT) + ln(L)]$
	Multi-Vehicle	$N_{MV\ U4U} = exp[-13.794 + 1.59. ln(AADT) + ln(L)]$
	Single-Vehicle	$N_{SV\ U4U} = exp[-6.961 + 0.751. ln(AADT) + ln(L)]$
U4D	Total	$N_{TOT\ U4D} = exp[-3.00 + 0.543. ln(AADT) + ln(L)]$
	Multi-Vehicle	$N_{MV\ U4D} = exp[-3.363 + 0.558. ln(AADT) + ln(L)]$
	Single-Vehicle	$N_{SV\ U4D} = exp[-4.687 + 0.543. ln(AADT) + ln(L)]$

Details regarding the coefficients and the statistical significance of the estimates can be found in the relevant sections throughout the report.

It should be noted that the calculated calibration factors and the developed SPFs are being embedded in the safety analysis spreadsheet used by the NJDOT staff. The spreadsheet is being modified so that users can either select the SPFs provided in the HSM and apply the calculated calibration factors or simply use the NJ-specific SPFs developed by the research team.

Intersection	Crash Type	Developed SPFs
R23ST	Total	$N_{spf\ 3ST} = exp[-6.139 + 0.498. ln(AADT_{maj}) + 0.296. ln(AADT_{min})]$
R23SG	Total	$N_{spf\ 3SG} = exp[-12.140 + 1.184. ln(AADT_{maj}) + 0.281. ln(AADT_{min})]$
R24ST	Total	$N_{spf\ 4ST} = exp[-3.716 + 0.159. ln(AADT_{maj}) + 0.426. ln(AADT_{min})]$
R24SG	Total	$N_{spf\ 4SG} = exp[-5.811 + 0.345. ln(AADT_{maj}) + 0.526. ln(AADT_{min})]$
U3ST	Total	$N_{TOT\ U3ST} = exp[-5.855 + 0.434. ln(AADT_{maj}) + 0.384. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U3ST} = exp[-6.892 + 0.483. ln(AADT_{maj}) + 0.429. ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U3ST} = exp[-4.895 + 0.283. ln(AADT_{maj}) + 0.219. ln(AADT_{min})]$
U3SG	Total	$N_{TOT\ U3SG} = exp[-7.553 + 0.693. ln(AADT_{maj}) + 0.321. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U3SG} = exp[-8.019 + 0.713. ln(AADT_{maj}) + 0.336. ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U3SG} = exp[-8.093 + 0.676. ln(AADT_{maj}) + 0.141. ln(AADT_{min})]$
	Vehicle-Pedestrian	$N_{ped\ U3SG} = exp[-18.636 + 1.145 \times ln(AADT_{tot}) + 0.0898 \times PedVol]$
U4ST	Total	$N_{TOT\ U4ST} = exp[-8.269 + 0.743. ln(AADT_{maj}) + 0.343. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U4ST} = exp[-8.959 + 0.752. ln(AADT_{maj}) + 0.392ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U4ST} = exp[-8.359 + 0.724. ln(AADT_{maj}) + 0.142. ln(AADT_{min})]$
U4SG	Total	$N_{TOT\ U4SG} = exp[-9.593 + 0.968. ln(AADT_{maj}) + 0.308. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U4SG} = exp[-10.307 + 1.022. ln(AADT_{maj}) + 0.317. ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U4SG} = exp[-5.804 + 0.424. ln(AADT_{maj}) + 0.173. ln(AADT_{min})]$
	Vehicle-Pedestrian	$N_{ped\ U4SG} = exp[-18.935 + 1.245. ln(AADT_{tot}) + 0.874. \ln(PedVo)l - 0.135. MaxLanes]$

In closing, it should be emphasized that calibration and development of SPFs are highly data driven tasks and that the quality of data is of utmost importance to ensure the validity of results. The observed issues related to the existing databases are presented in the Data Preparation and Cleaning section. The following is a brief of list of the most important issues that should be addressed by NJDOT:

- Although the SLD database is the richest source of information on roadway features, as with any database, it is not without errors. The research team detected errors or inconsistent information in the pavement width, highway type, number of lanes, and shoulder type tables.
- As mentioned above, there is no indication in the SLD database whether roadway segments include two-way left-turn lanes in the center. The lack of such information prevents users from identifying U3T and U5T segment types.

- Information for intersections is provided in two separate tables, namely *pt_intersection* and *pt_int_approach*. Some overpasses are coded as intersections in the database. Furthermore, the team found some incorrect data on control type and number of legs. The research team developed a sophisticated programming code to correct some of these errors with the use of NJ GIS roadway maps.
- The research team was not able to find any available database on driveways. This information is required for urban and suburban segments. The team conducted a manual data extraction to gather this information.
- The location information in the crash database poses a significant problem. As mentioned in the Data Preparation and Cleaning section, a significantly high percentage of crashes are missing SRI and milepost information. In addition, very few crashes include latitude and longitude information entered by police. NJDOT post-processes the raw crash data and geocodes the latitude and longitude of crashes using any available location-related information such as cross street names, addresses, and SRI and milepost, if reported. Following this geocoding process nearly 95 percent of crashes are assigned latitude and longitude values. It should be noted that this process yields the most probable coordinate readings, and that deviation in location by even small distances makes a difference in labeling a crash as intersection- or segment-related.
- The linear referencing system (LRS) table of NJ roadways includes information such as roadway SRI, start milepost, end milepost, route type, etc. The research team found that the mileposts for some links are off by 0.01 to 0.09 miles. As per our e-mail correspondence with the Bureau of Information Management and Technology on May 14, 2019, it was determined that the LRS is out of date and the department is in the process of updating it.

INTRODUCTION

Highway transportation has been the backbone of mobility and economic development in the U.S. and throughout the world, yet motor vehicle crashes have always been a leading safety issue on our roadways. For example, according to National Highway Traffic Safety Administration (NHTSA) motor vehicle crashes were the leading cause of death for age 11 and for every age 16 through 24 in 2014 [(1)]. NHTSA's mission is to reduce deaths and injuries from motor vehicle crashes. Although the number of fatal crashes in the U.S. has declined by 24 percent within the last decade, the total number of police-report crashes increased from 2011 to 2014, mostly driven by the increase in property damage crashes. In addition, the number of injury crashes increased from 2013 to 2014 by 3.6 percent [(1)]. According to the latest crash statistics for New Jersey (NJ)[1], fatal crashes and total crashes increase by 3.3 percent and 4.0 percent, respectively, from 2013 to 2014, while injury crashes decreased by 3.5 percent.

These statistics prove that there is still more work to be done to reduce the number of motor vehicle crashes. By relying on the abundant data sources related to traffic and motor vehicle crashes, state department of transportation (DOT) agencies have been utilizing science-based and state-of-the-practice methods and approaches to determine the causes of crashes and make reliable decisions on how to reduce their numbers. This new data-driven and scientific approach was initiated by the American Association of State Highway and Transportation Officials (AASHTO) in 1998 with the approval of its Strategic Highway Safety (SHS) plan, developed by the AASHTO Standing Committee on Highway Traffic Safety with the assistance of the Federal Highway Administration (FHWA), the National Highway Traffic Safety Administration (NTHSA), and the Transportation Research Board Committee on Transportation Safety Management, which has been the driving force in directing these safety-related efforts in the U.S. AASHTO's SHS plan includes strategies in 22 key emphasis areas that affect highway safety. Each of the emphasis areas includes specific strategies and an outline of what is needed to implement each strategy. As a result of these efforts, in order to assist transportation agencies to integrate safety into their decision making processes, the Highway Safety Manual (HSM) was published in 2010, providing a comprehensive approach and a set of analytical tools and methods for the integration of safety considerations into highway planning, design, operations, and maintenance. The HSM outlines and provides examples of the following applications [(2)]:

- Identifying sites with the most potential for reduction in crash frequency or severity.
- Identifying factors contributing to crashes and associated potential countermeasures to address these issues.
- Conducting economic appraisals of potential improvements and prioritizing projects.
- Evaluating the crash reduction benefits of implemented treatments.

[1] http://www.state.nj.us/transportation/refdata/accident/pdf/njcrash.pdf

- Estimating the potential effects on crash frequency and severity of planning, design, operations, and policy decisions.

The HSM is composed of four sections. The third section of the HSM, Part C, provides a predictive method for estimating the expected number of crashes for base conditions, alternative conditions, or proposed new roadway designs. The predictive method is applied for a given time period, traffic volume, and geometric design characteristics of a facility. This predictive method is most applicable when developing and evaluating various solutions to reduce the number of motor vehicle crashes at a specific location. Currently, the HSM's predictive model can be used for the segments and intersections of rural two-lane roads, rural multilane highways, and urban and suburban arterials.

The predictive model provided by the HSM is based on the Safety Performance Function (SPF), which is a statistical regression model based on observed crash data from similar facility types and estimates the predicted average crash frequency for the base conditions. The estimate can also be categorized by crash severity or collision type distribution. To account for differences between the base conditions and the specific conditions of the facility site, crash modification factors (CMFs) are utilized to adjust the prediction to account for the geometric design and traffic control features of the specific site.

SPFs in the HSM were developed using historic crash data collected over a number of years at sites of the same facility type in various states. <u>Because the SPFs provided in the HSM were developed using data from different states, it is more than likely that they cannot be transferred directly for use in other locations and times</u>. Thus, HSM's predictive model often needs to be calibrated to capture local state or geographic conditions. Moreover, accident frequencies for similar facility types can also vary from one jurisdiction to another, since their locations differ in climate, driver population and characteristics, accident reporting threshold, accident reporting practices, and other contributing factors.

To make the SPFs better accommodate the local data, two strategies are usually employed:

- The first strategy is to calibrate SPFs provided in the HSM so that the contents of the HSM can be fully leveraged.
- The second strategy is to develop location-specific SPFs regardless of the predictive modeling framework in the HSM.

Objectives

The main objective of this research project is to (1) calibrate the SPFs provided in the HSM using NJ data or (2) develop new NJ-specific SPFs as appropriate. The facility types considered for this research project include, but are not limited to, the following:

- Rural two-lane roads
- Rural four-lane divided roads
- Rural four-lane undivided roads

- Two-lane urban and suburban arterials
- Three-lane (with center two-way left-turn lanes [TWLTL]) urban and suburban arterials
- Four-lane divided urban and suburban arterials
- Four-lane undivided urban and suburban arterials
- Five-lane (with center TWLTL) urban and suburban arterials
- Three-leg minor road stop-controlled intersections on rural two-lane roads
- Four-leg minor road stop-controlled intersections on rural two-lane roads
- Three-leg signalized intersections on rural two-lane roads **(not addressed in HSM)**
- Four-leg signalized intersections on rural two-lane roads
- Three-leg minor road stop-controlled intersections on rural four-lane roads
- Four-leg minor road stop-controlled intersections on rural four-lane roads
- Three-leg signalized intersections on rural four-lane roads **(not addressed in HSM)**
- Four-leg signalized intersections on rural four-lane roads
- Three-leg minor road stop-controlled intersections on urban and suburban arterials
- Four-leg minor road stop-controlled intersections on urban and suburban arterials
- Three-leg signalized intersections on urban and suburban arterials
- Four-leg signalized intersections on urban and suburban arterials

Table 1 presents the list of facility types for which the HSM current has specific SPFs. The table also shows the list of facilities that this study focuses on.

Calibrating the SPFs used in the predictive models of the HSM requires data from a limited number of sites (for each facility type) from NJ, using the methods suggested in the Part C of the HSM. Developing NJ-specific SPFs provides more accurate results, but requires data from a larger sample of sites, and also involves the application of generalized linear models. Several tasks were completed to achieve the main project objective which is “to either calibrate the SPFs provided in the HSM using New Jersey (NJ) data or develop new NJ-specific SPFs.” These are:

- Conducted an in-depth review of the studies in the literature that focused on the calibration of the SPFs used in the HSM and the development of new SPFs, and identified the related calibration and development issues.
- Identified the key sources of data required for calibration and development of SPFs. These include roadway characteristics data, traffic volume data, and crash data.
- Developed a computer code to read and process the compiled database to: (a) filter out inconsistent data entries, (b) identify facility types, (c) execute the roadway segmentation process, (d) assign crash statistics for each facility, and (e) generate a complete database for each facility type to be used in the calibration and/or development of SPFs.
- Provided recommendations and activities that include: (a) improvements to data collection and recording practices that would lead to easier data extraction

required for the SPF calibration/development process, and (b) a workshop that demonstrates the step-by-step approach for using the SPFs for the New Jersey Department of Transportation (NJDOT) staff and other interested parties.

Table 1: Facility types used in the HSM

Rural Two-Lane Two-Way Roads				
Type	**HSM**	**NJ**	**Acronyms**	**Definition**
Roadway Segments	√	√	R2U	Rural two-lane roads
Intersections	√	√	R23ST	Three-leg minor road stop-controlled intersections on rural two-lane roads
		√	R23SG	Three-leg signalized intersections on rural two-lane roads
	√	√	R24ST	Four-leg minor road stop-controlled intersections on rural two-lane roads
	√	√	R24SG	Four-leg signalized intersections on rural two-lane roads
Rural Multilane Highways				
Type	**HSM**	**NJ**	**Acronyms**	**Definition**
Roadway Segments	√	√	R4U(RMU)	Rural four-lane undivided roads
	√	√	R4D(RMD)	Rural four-lane divided roads
			R4F	Rural four-lane freeways
			R6+F	Rural six+ lane freeways
Intersections	√	√	RM3ST	Three-leg minor road stop-controlled intersections on rural four-lane roads
		√	RM3SG	Three-leg signalized intersections on rural four-lane roads
	√	√	RM4ST	Four-leg minor road stop-controlled intersections on rural four-lane roads
	√	√	RM4SG(R4 4SG)	Four-leg signalized intersections on rural four-lane roads
Urban and Suburban Arterials				
Type	**HSM**	**NJ**	**Acronyms**	**Definition**
Roadway Segments	√	√	U2U	Two-lane urban and suburban arterials
	√	√	U3T	Three-lane (with center TWLTL) urban and suburban arterials
	√	√	U4U(UMU)	Four-lane undivided urban and suburban arterials
	√	√	U4D(UMD)	Four-lane divided urban and suburban arterials
	√	√	U5T	Five lane (with center TWLTL) urban and suburban arterials
Intersections	√	√	U3ST	Three-leg minor road stop-controlled intersections on urban and suburban arterials
	√	√	U4ST	Four-leg minor road stop-controlled intersections on urban and suburban arterials
	√	√	U3SG	Three-leg signalized intersections on urban and suburban arterials
	√	√	U4SG	Four-leg signalized intersections on urban and suburban arterials

The following section presents background on the HSM predictive methodology and the SPFs, along with an extensive review of the existing DOT studies.

LITERATURE REVIEW

This section first presents background on the SPFs and the HSM's predictive methodology.

Background

The HSM published by AASHTO provides quantitative tools in a useful form to facilitate improved decision making based on safety performance. SPFs, which are commonly developed to correlate traffic, geometric, and environmental characteristics with crash frequencies, are the fundamental components of the HSM. In the HSM, SPFs are functions of annual average daily traffic (AADT) and/or road length only, and provide estimates of the average crash frequency for a certain facility under base conditions.

The HSM utilizes SPFs in two key applications for which they can be used by jurisdictions to make better safety decisions. One application is to use SPFs as part of "network screening" to identify sections that may have the best potential for improvements. Part B of the HSM is dedicated to this type of application. The second application is to use SPFs to determine the safety impacts of design changes at the "project level." Part C of the HSM, which is composed of Chapters 10, 11, and 12, is dedicated to this type of application.

SPFs in the HSM were developed using historic crash data collected over a number of years at sites of the same facility type in various states. For example, Chapter 10 of Part C includes SPFs for rural two-lane two-way roads that were developed using state data from Minnesota, Washington, Michigan, and California. Chapter 11 includes SPFs for rural multilane highways using data from Texas, California, Minnesota, New York, and Washington. Chapter 12 includes SPFs for roadway segments in urban and suburban arterials developed using data from Minnesota, Michigan, and Washington. The SPFs for intersections in urban and suburban arterials were estimated using data from Minnesota, North Carolina, Florida, and Ontario, Toronto [(3)].

For example, the following SPF is used in the HSM (page 11–17, Chapter 11) for predicting the total number of crashes on an undivided roadway segment on a rural multilane highway:

$$N_{spf} = e^{a+b\ln(AADT)+\ln(L)} \tag{1}$$

Where, N_{spf} is the base predicted total number of crashes on an undivided segment of rural multilane highway per year, *AADT* is annual average daily traffic on roadway segment in vehicles per day, *L* is the segment length in miles, and *a* and *b* are regression coefficients.

This specific SPF and those used for other facility types are used to predict the crash frequency for a "base" condition. CMFs are used to adjust the predictions for situations that differ from that base condition. Specifically, a CMF is a constant or equation that represents the change in estimated average crash frequency due to a change in one specific condition (when all other conditions and site characteristics remain constant) [4]. CMFs vary by treatments as well as facility types. The CMFs of treatments such as changing lane width, adding lanes by narrowing existing lanes and shoulders, widening paved shoulder, modifying shoulder type, and providing raised median are proved in HSM based on multiple before–after studies. CMF can be presented as follows:

$$CMF = \frac{N_A}{N_B} \tag{2}$$

Where, N_A is the expected number of crashes at a facility after one or more safety measures are implemented, and N_B is the expected number of crashes at the same facility without any safety treatments. CMF less than 1.0 indicates fewer crashes due to the safety measures being taken. In HSM, SPFs and CMFs are used jointly to predict the average crash frequencies for selected sites using the following equation:

$$N_{pred} = C \times N_{spf} \times \prod CMF_k \tag{3}$$

Where, N_{pred} is the predicted average crash frequency, N_{spf} is the crash expectation under the base condition estimated by specific SPFs, CMF_k is the CMF for the treatment *k*, and C is the local calibration factor.

<u>Calibration of SPFs</u>

As mentioned earlier, the base SPFs presented in the HSM safety prediction methodologies were developed with data from specific highway agencies. SPFs in the HSM employ a local calibration factor, *C*, to "…account for the expected variations jurisdiction and time period for which the predictive models were developed and the jurisdiction and time period to which they are applied by HSM users" (2). The local calibration factor, *C*, can be expressed mathematically as follows:

$$C = \frac{\sum_{i=1}^{S} N_{obv,i}}{\sum_{i=1}^{S} N_{pre-unadj,i}} \tag{4}$$

Where, *C* is the local calibration factor for a certain facility type, $N_{obv,i}$ is the observed crash frequency for site *i* of a certain facility type during the study period, and $N_{pred-unadj,i}$ is the predicted crash frequency for site *i* of a certain facility type. This predicted value is unadjusted, since the C_i is not utilized in the SPF. *S* is the total number of sites for that facility type.

A calibration factor, *C*, of 1.0 means the predicted number of crashes is the same as the observed crash frequency. In other words, no calibration for local conditions is required. A *C* of less than 1.0 indicates that the observed crashes for a certain facility type in a region are fewer than the base model crash frequencies. A *C* of over 1.0 suggests that crash frequencies for a facility in a study location are greater than those for the base model. The computed *C* values are meant to be used to adjust HSM base model results to local conditions (5).

The calibration method suggested by the HSM involves the 5-step process shown in Figure 1.

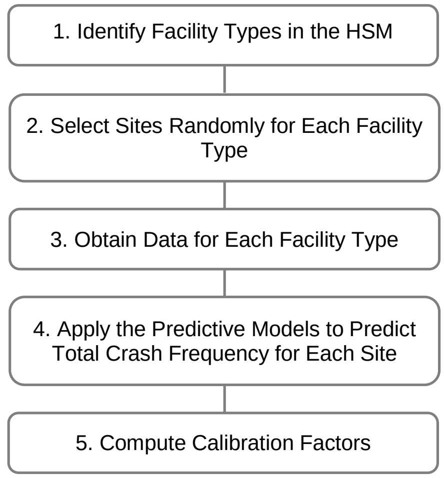


Figure 1. Calibration process recommended by the HSM [2]

The calibration procedure commences with Step 1, the decision by the agency as to which facility types need to be calibrated. Step 2 involves selecting a random sample of sites representing the given facility type until reaching a sample size adequate to achieve a desired accuracy level. In Step 3, data are collected and compiled for each specific site selected to compute unadjusted predicted crash frequency (i.e., C = 1.0). In Step 4, the unadjusted predicted crash frequency for each group of sites relevant to a given SPF is calculated applying the necessary and jurisdiction-specific CMFs to the HSM SPF base model. In Step 5, the sum of all unadjusted predicted crash frequencies for each group of sites is compared with the respective sum of crash frequencies observed at these sites for the same period, and the ratio between the observed and unadjusted predicted crash frequencies is the estimated calibration factor. It should be noted that the calibration process is repeated for each year for which crash and traffic volume data are available.

<u>Development of SPFs</u>

The process for developing SPFs is similar to that for calibrating them, with two important exceptions. The first is that developing state-specific SPFs requires more data to achieve statistically significant results. The second is that when developing SPFs, the analyst must be particularly carefully in selecting the functional structure and the explanatory variables included in the model.

Poisson models [14, 15] and negative binomial (NB) models [16-18] have been widely used to develop SPFs. It is widely recognized that Poisson models outperform the standard regression models in handling the nonnegative, random, and discrete features of crash counts [19, 20]. Despite the improved performance, however, the constraint of the mean being equal to the variance in Poisson models is often violated by over-dispersed crash data. Alternatively, NB models are used to accommodate this over-dispersion issue by incorporating an independently distributed error term. The specification of the NB models is expressed as follows:

$$N_{obv,i} \sim Negbin(\theta_i, r) \tag{5}$$

$$\ln(\theta_i) = \beta_0 + \sum_{p=1}^{P} \beta_p X_{pi} \tag{6}$$

Where, $N_{obv,i}$ denotes the crash frequency at the site *i* in a given time period, θ_i is the expectation of $N_{obv,i}$, X_{pi} are site-specific explanatory variables (e.g., AADT, road length, and lane width), β_0 and β_p are coefficients to be estimated, and *r* is the dispersion parameter of the NB models.

However, with the assumption of independent observations, neither the Poisson models nor the NB models address any inherent correlation of crash data. To complement the Poisson models and NB models, random effect models have been proposed in previous studies to account for the potential heterogeneity across homogeneous groups [(21-23)]. In addition, random parameter models that can be viewed as extensions of random effect models are developed to incorporate the variability of both the intercept and the variable coefficients across observations and thus provide more flexibility for handling the heterogeneity [(24-28)]. More recently, hierarchical models have become the preferred method to accommodate a multilevel data structure [(29-35)]. Hierarchical models can accommodate the heterogeneity among different groups and have the ability to incorporate variables at the specific levels where impacts of specific variables occur [(36)]. In addition, spatial autocorrelation can be implemented in crash observations. To model spatially correlated observations, generalized estimating equations (GEEs) are often the most widely preferred method [(37-40)]. However, GEEs have the limitation of setting the same correlation matrix for all intersection groups, and thus cannot reflect the discrepancies in correlations among different groups of intersections. Conditional autoregressive (CAR) models can provide more flexibility in specifying the magnitude of correlation, and have been recommended in many recent studies [(41-44)]. The CAR models capture the spatial dependence using the spatial error specification [(45)].

Data Requirements

In each of the HSM's safety prediction methodologies – whether calibration, development, or simply the application of SPFs, – the first step is to divide a roadway or project into homogenous roadway segments or intersections. Homogeneity means the geometric characteristics and the traffic flow along a segment do not vary over the study period. For example, there is a roadway segment that starts from mile point 0.1 mile and ends at mile point 0.2 mile. Within that 0.1-mile segment, all geometric characteristics, such as number of lanes, shoulder width, median type, and other variables should remain the same over the study period, or the researchers should redefine the segments to make them as homogeneous as possible [(5)].

The most important variables regarding site characteristic and traffic flow used in HSM safety predictions are shown in Table 2. Table 2 also identifies, for both roadway segments and intersections, the specific data elements needed to use each of the HSM's safety prediction methodologies.

Table 2: Data requirements for HSM safety predictions [6]

Data Elements	Data Requirements per Facility Type																	
	Rural Two-Lane Two-way Roads				Rural Multilane Highways					Urban and Suburban Arterials								
	R2U	R23ST	R24ST	R24SG	R4U	R4D	RM3ST	RM4ST	RM4SG	U2U	U3T	U4D	U4U	U5T	U3ST	U4ST	U3SG	U4SG
AADT of Major Road	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★
AADT of Minor Road		★	★	★			★	★	★						★	★	★	★
Segment Length	★				★	★				★	★	★	★	★				
Lane Width	★				★	★												
Shoulder Width	★				★	★												
Shoulder Type	★				★													
Horizontal Curve Data	★																	
Vertical Grades	☆																	
Driveway Density	☆									★	★	★	★	★				
Centerline Rumble Strips	☆																	
Passing Lanes	☆																	
Four-Lane Section	☆																	
TWLTLs	★																	
Roadside Hazard Rating	☆																	
Side Slope					★													
Roadside Fixed Object Density										☆	☆	☆	☆	☆				
Median Type and Width						★				★	★	★	★	★				
Lighting	☆	★	★	★	☆	☆	★	★		☆	☆	☆	☆	☆	★	★	★	★
Posted Speed Limit										★	★	★	★	★				
Automated Speed Enforcement	☆				☆	☆				☆	☆	☆	☆	☆			★	★
Intersection Skew Angle		☆	☆				☆	☆										
Left-Turn Signal Phasing																	★	★
Intersection Left-Turn Lane		★	★	★			★	★							★	★	★	★
Intersection Right-Turn Lane		★	★	★			★	★							★	★	★	★
Right-Turn-on-Red Prohibited																	★	★
On-Street Parking Type										★	★	★	★	★				
Maximum Lanes for Pedestrian Crossing																	☆	☆
Pedestrian Volumes																	☆	☆
Bus Stops within 1000 ft																	☆	☆
Schools within 1000 ft																	☆	☆
Alcohol Sales Establishments within 1,000 ft																	☆	☆

Note: ★ required, ☆ desired

However, as stated in the Maryland study [5] , states' datasets are not currently built for the HSM. Many of the variables used in HSM, such as AADT on minor roads, driveway density, liquor store density, and others, are not commonly collected; thus, collecting required/desired data was the most difficult task in the Maryland study [5].
The required crash data are:

- Date (year)

- Location (milepost, latitude and longitude readings)
- Severity level (fatal/injury/property damage only)
- Relationship to intersection (at-intersection/intersection-related/not-intersection-related)
- Type of crash

The Choice Whether to Calibrate or Develop SPFs

As stated earlier, because the SPFs provided in the HSM are developed using data from various states, it is more than likely that they cannot be transferred directly for use in a specific state. There are two possible strategies for making the SPFs better accommodate local data:

- The first strategy is to calibrate SPFs provided in HSM so that the contents of HSM can be fully leveraged. States that calibrated SPFs provided in the HSM include Florida (13), Maryland (5), Oregon (6), Missouri (8), and Louisiana (9).
- The second strategy is to develop location-specific SPFs regardless of the predictive modeling framework in the HSM. States that developed their own SPFs include Illinois (46) and Virginia (47).

The pros and cons of these two strategies are summarized in Table 3.

Table 3: Pros and cons of the two different SPF strategies

	Calibrate SPFs Provided in HSM	Develop Location-Specific SPFs
Pros	• Makes the best use of the predictive modeling framework in HSM • Requires less sample data	• Provides more flexibility to accommodate local data • Provides crash estimates for facilities not included in HSM
Cons	• Provides less flexibility to accommodate the data in the new locations • Cannot provide crash estimates for facilities not included in HSM	• Requires more sample data to achieve statistically robust results • Requires additional work for model development, evaluation, and comparison

There are also states that adopt both strategies, such as Idaho (7), Alabama (10), Kansas (48), North Carolina (12) and Utah (49).

Previous Work

This section provides a detailed review of the existing studies conducted for various state DOTs. The jurisdiction-specific SPFs developed by these studies are shown in Appendix A.

Trieu et al. (52) evaluated the HSM calibration criteria using two-lane two-way undivided urban arterial roadway (2U) segments and proposed a method to improve the accuracy and predictions of the calibration factors. A sensitivity analysis using different sized calibration sets was performed by using Monte-Carlo simulation to re-sample sites. The researchers used roadway geometry, traffic volume, and crash data from six counties in northern New Jersey for a 3-year period, from 2009 to 2011. NJDOT's Straight Line Diagram (SLD) database was used to assemble the required data for the 2U segments. The 2U segments were divided into 0.5-mile homogeneous sites. Google Earth and Google Street View tools were used to collect the remaining information required for calibration. A total of 372 sites were created. The expected number of crashes was calculated using the corresponding SPFs of the HSM. The number of sites drawn for each calibration set was based on varying percentages of the 372 sites. The authors used six different percentages from 10 percent to 60 percent in increments of 10 percent. Monte-Carlo simulation was used to select random samples from the full dataset. Sites were selected with no regard to the criterion of a minimum of 100 crashes per year to be included in the analysis; later results indicated that most sites met this criterion. For each randomly selected sample the authors compared the percent error of the calibration factor with respect to the target calibration factor. In this case the target calibration factor was that calculated using all 372 sites.

The target calibration factor was computed as 1.99 when all 372 sites were used. It was shown that at 50 percent of the full dataset all calibration factors were within 10 percent of the target calibration factor. In contrast, at 10 percent (i.e., a sample size of 37, which falls within the HSM site-count criterion of 30 to 50 sites), the probability of the calibration factor being within 10 percent of the target fell to 67 percent. The results indicate that the HSM's criterion of 30 to 50 sites is insufficient for calculating a calibration factor, at least for two-lane undivided urban segments. The authors also examined how the average AADT distribution differed between calibration factors with small and large error terms.

Green et al. (53) developed a methodology to create and maintain a full inventory of all intersection types in Kentucky using ArcMap. Using the developed database, the authors developed SPFs for various intersection types. The developed database contained precise location information as well as several safety and operational attributes for each point of an intersection. Once the database was created for all intersections, it was divided into 20 categories based on urbanization degree (rural or urban), the number of approaches, the type of traffic control, and whether the intersection was divided or undivided. Approximately one-third of the full dataset had sufficient attribute data. There was a total of 182,384 intersections. The authors performed negative binomial regression analysis using 5-year crash data from 2009 to 2014. They estimated SPFs for all crash types and for fatal and serious injury crashes (KAB) separately.

Storm and Richfield (54) developed calibration factors for the segments and intersections of two-lane two-way rural highways and rural expressways in Minnesota. Although the HSM provides SPFs for both divided and undivided segments of rural multilane highways, in Minnesota, rural expressways differ from those included in the HSM, where they are defined as typically high-speed multilane facilities with a depressed grass median, a limited number of at-grade intersections, and occasional interchanges at high-volume cross streets. MnDOT's intersection and section toolkit data were utilized to identify these segments and intersections of the two selected roadway facilities. Crash data for 3 years, from 2009 to 2011, were used in the analyses. From the database of two-lane two-way rural highway intersection sites, the authors used 100 sites to calculate calibration factors, except 4-way signalized intersections, of which there were only 33 in total. Similarly, for the segments, they used a total of 115 sites to calculate calibration factors. From the database of rural multilane expressways, they used a number of sites varying from 31 to 93 for 4-way stop-controlled intersections and 50 sites for segments. Data collection was performed manually using Google Earth aerial images. MnDOT Geographic Information Systems (GIS) shapefiles with traffic volumes were used to identify volumes for segments and intersections. For intersections, if an intersection was missing the AADT value for an approach, the authors assumed a fixed value for that approach based on the area population. Using the extracted data, separate calibration factors were calculated for two regions in Minnesota for each facility type. The authors also performed a cumulative residuals (CURE) analysis with respect to AADT and segment length to determine if the model bias for each facility type was within the acceptable range. A CURE plot shows the cumulative residuals of the data within the 95-percent confidence interval, or the upper and lower bound limits equal to plus or minus two standard deviations. If the plotted residuals are clustered around zero, it indicates that there is greater strength in the model in predicting the number of crashes for the given facility [(55)]. The calculated calibration factors varied from 0.41 to 0.58 for two-lane two-way segments, and from 0.45 to 1.22 for two-lane two-way intersections, for two different areas within Minnesota. Similarly, the calibration factors for rural expressway segments varied from 0.53 to 0.69, and those for the intersections varied from 0.39 to 2.32.

Shankar and Madanat [(56)] developed SPFs for roadway segments, intersections, and ramps in California. They developed two separate SPFs, namely Type 1 and Type 2. Type 1 SPFs use only AADT as the explanatory variable, whereas Type 2 SPFs include both ADT and geometric variables.

Type 1
$$\lambda_i = L_i . e^{\alpha} . ADT_i^{\beta} \tag{7}$$

Type 2
$$\lambda_i = L_i . e^{\alpha} . ADT_i^{\beta} e^{\sum_{j=1}^{n} \gamma_i Z_{ij}} \tag{8}$$

Where, λ_i is the expected number of crashes for facility type *i*; *L* is segment length; *α*, *β,* and γ_j are model parameters; and Z_{ij} is the area vector of geometric effects. The equations assume that length linearly affects the expected crash rate for a roadway segment. Also, the length variable is not present in the estimation equation for intersections, which are defined as fixed-length ranges of 0.05 miles from the centerline of the intersecting roadway.

A total of 11 SPF classes were developed based on rural–urban distinctions and the number of lanes: (1) two-lane rural, (2) four-lane rural, (3) four-plus-rural, (4) multilane undivided rural, (5) multilane divided, (6) two-lane urban, (7) four-lane urban, (8) five-to-seven lane urban, (9) eight or more lane urban, (10) multilane undivided urban, and (11) multilane divided urban. Functional classifications of roadways were based on ADT counts. For example, an upper ADT bound of 35,000 was used to define rural interstate freeways. Comparatively, a lower ADT bound of 13,000 was used for urban state freeways and expressways. Also, a lower ADT bound of 3,000 was used for urban non-freeways/non-expressways, including arterials. It should be noted that the roadway functional classification used in this study does not coincide with those used in the HSM. The homogenous segments used in this study had varying lengths; 60 percent had lengths less than or equal to 0.1 miles. In addition to segments, a total of 17,200 intersection data were assembled. Intersections were extracted from the database using the mainline sections within 0.05 mile with respect to the centerline milepost. Mainline attributes such as shoulder widths, number of lanes, and roadside treatments (e.g., median barrier, guardrail) were integrated to form a comprehensive intersection geometric attribute dataset. Ten SPFs of Type 1 and Type 2 were developed for five severity types (i.e., fatality, severe injury, visible injury, complaint of pain, and property damage only) and total crash counts (rural multilane divided highways did not have enough sites for development). As a result, 120 SPFs were developed for roadway segments. It should be noted that there are more severity types included in this study than considered in the HSM.

The integrated dataset used in this study consisted of 15,162 centerline miles and 50,893.55 lane miles. A total of 40,541 roadway segments (excluding intersection ranges), with average lane length of 1.032 miles and segment length of 0.277 miles, constituted this network.

Crash data for the period 2005 to 2010 were used to develop SPFs, and 2011 to 2012 crash data were used for prediction testing.

The main drawback of this study is the intersection SPFs, for which only the mainline estimated crash data is provided. In other words, the study did not use cross street crash data, since they were not available. Also, model transferability tests conducted on roadway segments indicated that unobserved effects remain in the models. Despite these unobserved effects, the predictive effectiveness of Type 2 SPFs compared to Type 1 SPFs was significant.

Srinivasan et al. [(13)] estimated the calibration factors for the segments and intersections of all facilities included in the HSM for the years 2005 through 2008. The segment-level SPFs for total crashes in rural facilities were not calibrated, because property damage crashes are not fully recorded by the long-form crash reports in Florida’s crash analysis reporting system. Also, the equations for fatal and injury crashes were not calibrated for multilane undivided rural segments, due to lack of adequate data. As a result of this data-reporting issue, the calibration of segment-level SPFs for urban and suburban roadways was similarly only performed for fatal and injury crashes. The calibration factor was not calculated for each of five collision types individually, but for the sum of these five components. Similarly, at the intersection level, calibration factors were only computed for all collision types, rather than for each individual SPF.

The roadway characteristic data were collected through the Florida Roadway Characteristics Inventory (RCI) database maintained by FDOT. The majority of the required roadway characteristics data were obtained from the RCI database. For those that were not available through the RCI, such as grade, centerline rumble strips, roadside hazard rating, driveway density, and side slope, the recommended HSM default values were used. Sensitivity analyses for various ranges of these assumed default values were also conducted to assess the variation in the calculated calibration factors, and the results did not indicate any substantially different predicted crash rates. In this study, in identifying homogeneous segments, a minimum segment length was implemented for rural and urban segments, 0.10 and 0.04 miles respectively. This study also estimated the calibration factors with respect to geographic segmentation, namely seven Florida districts and four different population density levels. However, in the analyses, any segments that included intersections or curves were removed from the sample. This approach inevitably limits the applicability of the estimated calibration factors.

In addition, this study also developed Florida-specific SPFs for the segments of rural two-lane two-way roadways and urban and suburban four-lane divided arterials. Using the 80 percent of the data already extracted, the authors developed SPFs for these facilities using negative binomial regression, and the remaining 20 percent of the data was used for testing. However, it was reported that there was no system-wide improvement in the accuracy of crash prediction.

Kweon and Lim [(47)] developed SPFs for multilane highway and freeway segments using 5 years (2004 to 2008) of data collected from 20,235 multilane highway segments and 2,905 directional freeway segments in Virginia. Statewide SPFs were developed for 4 subtypes of multilane highway segments and 10 subtypes of freeway segments. VDOT district-group SPFs were developed for 4 subtypes of multilane highway segments. The study used VDOT's roadway management system called the Roadway Network System (RNS), which includes the roadway inventory, accident database, and traffic monitoring system. The data used for this study were extracted from this database for the years 2004 to 2008. The functional form of the developed SPFs for the segments followed that used by the HSM. A negative binomial regression was conducted to estimate the model parameters. The predicted crash rate functions for multilane highway segments were for two directions, and those for freeway segments were for one direction. Separate SPFs were estimated for total crashes and for fatal and injury crashes.

Garber and Rivera [(57)] developed SPFs for urban and rural intersections in Virginia. A total of 8,010 urban and 10,346 rural intersection data were extracted from the Highway Traffic Records Information System (HTRIS) and were later replaced by the RNS database. Crash data for the years 2003 to 2007 were used. The functional form of the intersection-level SPF is as follows:

$$\lambda_i = e^{\alpha}.AADT_{major,i}^{\beta_1} AADT_{minor,i}^{\beta_2} \tag{9}$$

Where, λ_i is the expected number of crashes for intersection i, $AADT_{major}$ and $AADT_{minor}$ are the annual average daily traffic at the major and minor approaches of the intersection, respectively, and α, β_1, and β_2 are model parameters. Several test statistics were used to evaluate the developed SPFs, including the mean prediction bias (MPB), the mean absolute deviation (MAD), the mean square error (MSE), the mean squared prediction error (MSPE), the Pearson's product moment correlation coefficients between observed and predicted crash frequencies, the dispersion parameters, and the Freeman Tukey R_{FT}^2 coefficients. Seventy percent of the data was used to fit the SPFs, and the remaining data were used for testing the transferability.

Savolainen et al. (58) developed SPFs for four urban intersections in Michigan, namely three-leg (3ST) and four-leg (4ST) minor stop-controlled and signalized intersections (3SG, 4SG). In addition to the available traffic volume, roadway geometry, and crash database for the years 2008 to 2012, they performed extensive data collection to obtain missing information such as major and minor AADT, intersection skew angle, presence and type of left-turn phasing, presence of on-street parking and bus stops, etc. A threshold value of 0.04 miles from the intersection was used to identify crashes as intersection-related. The database created included 353 samples for 3ST intersections, 350 for 4ST, 210 for 3SG, and 349 for 4SG. Because pedestrian and cyclist volumes were not available, the study developed models for pedestrian and bicycle crashes based on AADT for different severity types. The authors developed two types of SPFs: Type 1 was in the functional form suggested by the HSM, and Type 2 included detailed operational and geometric variables. Type 1 SPFs were developed for fatality/injury and property damage only crash severity types separately, with regional variables included in the model. The authors used state-specific percentages to calculate crash rates for different collision types. Type 2 SPFs for multi-vehicle and single-vehicle crashes and fatal/injury and property damage severity crashes were based on the following base conditions: no left-turn or right-turn lanes on the major road, zero skew angle, and no intersection lighting present. Similar to the Type 1 SPFs, these included regional indicators in the model. These SPFs were developed separately for intersections of only two-way streets and for intersections with at least one one-way intersecting street. CMFs were estimated for AADT, region, presence of median, presence of intersection lighting, number of lanes, posted speed limit, right-turn-on-red prohibition, and left- and right-turn lane presence.

Robicheaux and Wolshon (9) calculated the calibration factors for rural two-lane two-way, rural multilane divided and undivided, and urban and suburban two-lane, four-lane divided and undivided roadway segments in Louisiana. The roadway design and historical crash data for the years 2009 to 2011 were used in the analyses. The calibration factors for these segments varied between 0.62 and 2.54. This study did not include all possible geometric design and traffic control variables required by the HSM, due to the lack of these variables in the available state-maintained roadway database. However, the researchers extracted these data for randomly selected segments using Google Earth, and reported changes in the calibration factors of 20 to 60 percent.

Garber et al. [58] developed SPFs for rural and urban two-lane two-way roadway segments in Virginia. These SPFs were estimated for total crashes and combined fatal and injury crashes using a negative binomial distribution separately for three geographic regions in Virginia. The same functional form shown in equation (1) was followed. Roadway geometry, volume, and crash data for the years 2003 to 2007 were used in the model estimation, using sites where no geometric or operational changes were implemented within the time period. Similar to the study by Garber and Rivera [57], 70 percent of the data was used to fit the SPFs, and the remaining data were used for testing the transferability. The authors examined the fit of each SPF, R^2 and R^2_{FT} values were calculated, and the transferability of the models was evaluated using the MSPE.

Sipple [60] calculated the calibration factors for rural two-lane two-way roadway segments and 3ST and 4ST intersections in Idaho. In this study, state-specific SPFs were also developed for the same facilities. Most geometric and volume data were obtained from a state-maintained database. The necessary data that were not available in this database were collected using video logs. Crash data for the selected roadway segments and intersections were averaged from the 2003 to 2012 crash databases. The geometric data were from 2010, and the AADT values were from 2012. It was also assumed that geometric conditions were unchanged between these years. Identification of segments and intersections was performed using the intersection flag in the data, assuming that crash data were reported accurately. Roadside hazard rating (RHR) values were assumed as the base condition given in the HSM (i.e., RHR = 3). The sample size for the segments analysis was 447 sites, for which the average segment length was 0.507 miles. The sample sizes for 3ST and 4ST intersections were 79 and 85, respectively. After all the required data were compiled, the segments were randomly divided for fitting the models and testing the predictive capabilities of those models, respectively. The 70 percent was randomly sampled 10 times from the full dataset to test the variability in each of the calculated parameters, i.e., the calibration factor and the regression coefficients. This was completed for only two-lane highway segments as a test. The results of examining several random samples showed the averages of each parameter converging toward the parameter values for the full dataset. Once calibration of the HSM SPFs and fitting of the developed state-specific SPF equations was complete, statistical analyses were performed to compare the reliability of the models based on how well they predicted crash frequency as compared to Idaho data. These statistical analyses included the Pearson's R, MSPE, and the R^2_{FT} values. Based on the results, it was found that the developed SPFs for 3ST and 4ST were not statistically significant.

Williamson and Zhou [61] calculated the calibration factor for rural two-lane two-way roadway segments in Illinois using the crash data for the years 2007 to 2009. They also used a state-specific SPF developed by an earlier study and concluded that the calibrated HSM SPFs better fit the observed crash data. Collection or extraction of horizontal curvature data was not mentioned in the study.

Sun et al. [62] evaluated the application of the HSM for rural two-lane two-way roadway segments using Louisiana data prior to the publication of the first edition of the HSM.

However, the recommended HSM calibration procedure was only partly followed, due to the lack of required geometric data. The study was able to create a database with ADT, lane width, and shoulder type and width, but RHR, driveway density, and horizontal and vertical curvature information were assumed as default values. The database used included 4,123 homogeneous sections with an average site length of 3.25 miles. The crash data from 1999 to 2001 were used, and an average calibration factor of 1.63 was calculated. The study then collected two additional variables, namely horizontal and vertical curvature and average driveway density. The researchers then examined the differences between the observed and estimated crash counts using six different analysis scenarios designed to calculate the impact of including these additional variables. It was shown that with the inclusion of these variables the difference between the observed and predicted crash counts was reduced from 5.8 percent to 3.3 percent.

Banihashemi [63] evaluated the quality of the calibration factors generated from datasets of varying sizes by observing the changes in the calculated calibration factor using different percentages of the complete dataset. The study focused on rural two-lane two-way, multilane, and urban and suburban arterial roadway segments. The roadway and crash data from 2006 to 2008 were used in the analyses. All the required data for calibration was readily available from various sources, except on-street parking, driveways, roadside objects, and side slopes. The study used default values for these missing variables. The methodology for conducting the sensitivity analysis was based on the assumption that the calibration factors calculated using different subsets of the overall dataset followed a normal distribution. Subsets' percentages were chosen as 5, 10, 15, 20, 30, and 50 percent of the overall dataset. The probabilities that the estimated calibration factor fell within 5 percent and 10 percent of the ideal calibration factors were calculated. These probabilities varied greatly for roadway segment types.

Troyer et al. [55] presented the efforts undertaken by Ohio DOT to implement the HSM SPFs through data collection and calibration. It was stated that Ohio DOT had not collected all the data required for the HSM calibration procedure. The department had only recently started collecting additional data elements such as parking, driveways, and roadway curvature. The department elected to first calibrate the HSM SPFs and then decide whether to develop state-specific SPFs based on the evaluation of the calculated calibration factors. In order to evaluate the applicability of the calibration factors, the study generated CURE plots, which illustrate how well the model estimates the observed crashes and whether the explanatory variable fits the data for the entire dataset. The study included all 18 facilities included in the first edition of the HSM. Since most required data were not available, an extensive data extraction was performed to collect all required data elements using online maps, video logs, and GIS maps. The calibration factors, calculated as averages of multiple random samples, varied from 0.25 to 3.71. The study concluded that the three-quarters of the HSM SPFs calibrated fulfilled Ohio DOT's needs, and no state-specific SPF development was required for those facilities.

Tegge et al. [64] developed state-specific SPFs for 17 segment types and 10 intersection types suggested by SafetyAnalyst, a software tool developed by the FHWA. The authors used the negative binomial regression model to estimate the SPFs. They used 5 years of crash data, from 2001 to 2005, but only 1 year of roadway data, due to limited availability. Therefore, the developed SPFs estimate the number of crashes for a 5-year period. Separate SPFs were developed for fatal crashes, two types of injury crashes, and fatal and injury crashes.

Persaud et al. [(65)] evaluated the transferability of HSM SPFs to urban signalized intersections in Toronto, Canada. The analyses dataset included AADT data for 137 three-leg and 1,691 four-leg intersections, as well as crash data by collision type and severity, for a 6-year period, 1999-2004. The evaluation included only two collision types: all multi-vehicle collisions, and rear-end collisions for all severity types combined. Calibration factors for adjusting the HSM models to local conditions were first estimated and then applied to the HSM models to predict collisions for local sites in Toronto. These model predictions were then compared to those developed using local site data by using various goodness of fit and other performance measures, including the value of the recalibrated overdispersion parameter, MAD, and CURE plots. The analyses' results indicated that the performance of the HSM SPFs was mixed. Applying the algorithm with base models estimated from local data produced marginally better predictions than applying the algorithm with HSM base models calibrated to local data.

Persaud et al. [(66)] developed SPFs for 10 intersection types for urban roadways in Colorado. SPFs were developed separately for total crashes and fatal/injury crashes. The study used a sample of 543 intersections in total. Five years of crash data for the period 2000 to 2004 were collected for each of the 10 intersection types. CURE plots were used to evaluate the estimated models. The study suggests recalibrating the developed SPFs as new data become available. It also suggests recalibrating the overdispersion parameter, since it not only indicates how well the recalibrated SPF is fitting the data, but can also be used in the empirical Bayes methodology.

Alluri et al. [(67)] identified and prioritized the influence of various variables on the calibration factor by ranking the variables using the random forest technique. The analyses were conducted using data for seven roadway segment types and three intersection types in Florida. The results showed that aside from AADT and length, driveway density and shoulder width were the most influential variables for rural two-lane two-way roads. Similarly, shoulder width and median width were ranked most influential for rural multilane roadways. This study did not include curvature or vertical grade data, hence these variables were not ranked. For urban and suburban 4SG intersections, the number of approaches with right-turn lanes and the presence of bus stops were the most influential variables after AADT. Similarly, for 3ST intersections in rural and urban/suburban roadways, the third and fourth most influential variables were the number of approaches with left- and right-turn lanes, respectively. In a similar study, Findley et al. [(68)] reported that for rural two-lane two-way roadways, AADT, curve radius, and curve length were the most important variables in predicting curve crashes using the HSM model.

Mehta and Lou [10] calibrated the HSM SPFs for rural two-lane two-way and divided four-lane roadways in Alabama, and also developed state-specific SPFs in four different functional forms. The prediction capabilities of the two calibrated models and the four newly developed SPFs were evaluated using a validation dataset. Data included multiple traffic- and crash-related variables for the period 2006 to 2009. Homogeneous segments were created based on facility type, number of lanes, lane width, shoulder width, median width, speed limit, and AADT. Horizontal and vertical curvatures were not considered in generating homogeneous segments. The state-specific SPFs were developed to estimate the total number of crashes, and separate models for different severity and collision types were not considered. Five performance measures were used for model evaluation: MAD, MSPE, MPB, log-likelihood value, and the Akaike's information criterion (AIC) (50). Two approaches were used to estimate the calibration factor. The first was the HSM-recommended method, while the second treated the calibration factor estimation as a special case of SPF development. The analyses' results indicated that the calibrated HSM SPFs performed satisfactorily, but the state-specific ones yielded superior predictions.

Kim et al. [69] conducted the calibration of HSM SPFs and the development of state-specific SPFs for urban and suburban arterial segments in Alabama. The study used the available roadway and crash dataset for the period 2007 to 2009. The available dataset included roadway segments of 0.01 miles containing variables including the number of lanes, AADT, functional class, total crash frequencies and crash type, presence of medians, median type, and presence of TWLTL. Homogeneous segments were generated by combining these 0.01-mile segments based on AADT, number of lanes, speed limit, median type and width, and presence of TWLTL. The average length of the generated homogenous segments was 0.14 mile, with minimum and maximum lengths of 0.01 miles and 2.14 miles, respectively. The dataset included 2,613 samples of 2U, 479 samples of 3T, 1,054 samples of 4U, 3,153 samples of 4D, and 1,598 samples of 5T. The calibration factors were given 90 percent confidence levels, assuming that the calibration factor obtained from a limited number of samples follows a hypergeometric distribution, following the method proposed by Shin et al. [5]. The study concluded that the HSM-recommended 30 to 50 samples were not enough to identify the calibration factor. For each facility, separate SPFs were developed for multi-vehicle and single-vehicle collisions. The performance of the models was evaluated using MAD, MPD, AIC, and Bayesian information criterion (BIC) measures. The analyses indicated mixed success for the developed SPFs.

Shin et al. [5] calculated the calibration factors for all 18 facilities included in the HSM using data from 2008 to 2010 in Maryland. The study stresses the challenges in extracting the data required for the HSM calibration procedure. The data limitations included (a) the lack of precision measurement, reporting, and data collection tools, (b) inadequate coverage of traffic data, (c) incomplete data, (d) lack of roadway inventory data, and (e) data integration and interoperability. The most common data hurdles identified were the AADT on minor roadways, curvature information, and the signal phasing data. The study found that the calibration factors for 15 out of 18 facility types were less than 1.0. Of particular note, the calibration factors for intersections were extremely low. The study also pointed out the effect of crash reporting thresholds on the calibration factors, especially for property damage type crashes. For example, in the case of Maryland, law enforcement officers are required to fill out a crash report only if there is an injury involved or if one of the involved vehicles needs to be towed. Therefore, the authors suggested that for Maryland, only the calibration factors for fatal and injury crashes should be utilized.

A different study by Shin et al. [(70)] proposed an approach to determine a statistically reliable sample size for calculating local calibration factors using Maryland data from 2008 to 2010. The proposed approach was based on the finite population correction (FPC) factor, which determined the minimum sample sizes by considering trade-offs among the desired error levels of the estimated calibration factors, confidence levels, and sample standard deviations. In many statistical analyses the underlying assumption is that the samples are taken from an infinite population and that they are selected with replacement. However, this is not always the case. These assumptions do not present much of a problem when the sample size, *n*, is small relative to the population size, *N* – for example when *n* is less than 5 percent to 10 percent of *N*. However, when *n* is larger, it is best to apply a correction to the formulas used to compute standard error. This correction, the FPC, is calculated as follows:

$$FPC = \sqrt{\frac{N-n}{N-1}} \tag{10}$$

The sample sizes obtained by using the proposed approach were compared with those suggested by the HSM, and the study concluded that the HSM-based sample sizes yielded inconsistent reliabilities, depending on facility types.

Xie and Chen [(71)] first calculated the calibration factors for the SPFs of signalized intersections of urban and suburban arterials in the HSM using Massachusetts data from the period 2009 to 2012. The results indicated that the calibration factors for 3SG and 4SG intersections are substantially greater than 1.0, suggesting that the observed crashes at these two types of intersections are significantly higher than those predicted using the HSM SPFs. The study also developed new SPFs for urban and suburban intersections in Massachusetts for multiple- and single-vehicle, vehicle-bicycle, and vehicle-pedestrian crashes. In addition, because the HSM SPFs for vehicle-pedestrian collisions at signalized intersections require daily pedestrian volumes, in this study regression models were developed to estimate daily pedestrian volumes.

Donnell et al. [(72)] developed state-specific SPFs for roadway segments and intersections of rural two-lane two-way roadways in Pennsylvania. The study used 8 years of crash data, from 2005 to 2012. Because the available roadway data was from 2008 to 2012, linear interpolation was performed to account for the AADT for the missing years, i.e., 2005 to 2008. Various roadway data such as curvature, intersection skew angle, presence of passing zones, RHR, rumble strips, driveway density, and left- or right-turn lanes at intersections were extracted using Google Earth and PennDOT's video photologs. A total of 21,340 unique roadway segments and 683 intersections were used to develop two separate SPFs, for total crashes and fatal/injury crashes. However, the developed SPFs were not in the same form as those in the HSM, in which the study allows many of the variables included in the CMFs inside the model.

Donnell et al. [(73)] extended their previous study and developed state-specific, regionalized SPFs for rural two-lane two-way, multilane, and urban and suburban arterial roadway segments and intersections in Pennsylvania. Separate SPFs were developed for total and fatal/injury type crashes for each facility type and for each of the 12 districts in the state. The analyses' results showed that, when an adequate sample of roadway segments or intersections was available for modeling, district-level SPFs, with county adjustment factors, outperformed other regional or statewide models based on their predictive power. The study used crash data from 2010 to 2014 and roadway geometric and operational data from 2013. AADT data were available for the years 2009 and 2013, and thus the study used linear interpolation to estimate the AADT values for the interim years, and used linear extrapolation for the year 2014. Similar to the earlier study, the missing data were extracted using online mapping tools and PennDOT's video logs. The performance of the developed models was tested based on the RMSE measure.

Dixon et al. [(6)] calibrated the HSM SPFs for all roadway facility types in Oregon. The study extracted random samples for each facility type, with sample sizes for facilities ranging from 19 to 491. Crash data from 2004 to 2006 were used in the analyses. ODOT had horizontal curvature and vertical grade data available, so the researchers did not have to manually extract this information. However, it was determined that most minor-leg AADT values were missing, thus a multivariate linear regression model was used to estimate the minor AADT using independent variables such as land use, distance to the nearest freeway, presence of a centerline, etc. Other required data were extracted using ODOT video logs and online maps. For any data that could not be obtained or extracted, default values suggested by the HSM were used. Calibration factors varied from 0.15 to 1.43. The study also stresses the importance of the crash reporting differences between states, since Oregon is a self-reporting state. That is, when a person is involved in a crash that does not involve any injuries or require towing, the drivers involved are not legally required to report the crash. This results in underrepresentation of property damage crashes in the database. A similar argument was presented in Shin et al. [(70)].

Saito et al. [(49)] calibrated the HSM SPFs for rural two-lane two-way roadway segments in Utah, and also developed four state-specific SPFs using the negative binomial technique and one state-specific SPF using the full Bayesian modeling technique. The authors used crash data from 2005 to 2007, using 157 sample segments. Regarding geometric details, the study did not use curved roadways and eliminated segments within 250 feet of a horizontal curve. Therefore, only tangent road segments were used in the analyses. Rather than using crash data for individual years, the authors calculated the calibration factor using the total number of crashes during the years 2005 to 2007. The overall calibration factor for the studied roadway facility was found to be 1.16. Evaluation of the developed SPFs was performed based on the BIC measure. The researchers reported that the calibrated HSM SPF performed well compared to the developed models, yet argued that it did not provide a substantial benefit, because of the additional data required to implement it.

Srinivasan and Carter [3] developed state-specific SPFs for 16 roadway segments and 9 crash types using data from North Carolina. This study also calibrated the HSM SPFs for divided rural multilane road segments, all urban road segments, and 8 of 10 intersections included in the HSM. Roadway operational and geometric details and historical crash data from 2007 to 2009 were used in the calibration analyses. The analyses' results indicated that the calibration factors did not vary significantly from year to year; however, the calibration factors for urban two-lane roads with TWLTL and four lane divided and undivided roadway segments had significantly high calibration factors, in the range of 3.45 to 4.45. Roadway operational and geometric details and historical crash data from 2004 to 2008 were used in the model development. The roadway types selected for the model developed were chosen with the anticipation that they would be used within SafetyAnalyst. Therefore, some of the selected roadway types are not currently present in the HSM. In addition, the functional forms of the developed models were selected so that they were consistent with those used in SafetyAnalyst. It should be noted that the functional forms of SPFs in HSM and SafetyAnalyst differ from each other. The developed SPFs were evaluated using Freeman Tukey R_{FT}^2 and pseudo-R^2 coefficients.

Sun et al. [8] calibrated the HSM SPFs of five segment types and eight intersection types for Missouri. Three years of geometric, operational, and crash data from 2009 to 2011 were used in the analyses. The random sampling technique applied ensured geographic representativeness across the state. The extraction of the required data included examining video logs for roadside features, estimating horizontal curve parameters using CAD, reviewing Street View photographs to verify inventories and configuration, and measuring median widths using aerial photographs. The study also discussed certain challenges encountered during calibration, including data availability, finding a sufficient sample size for certain site types, maintaining a balance between segment homogeneity and minimum segment length, and excluding inconsistent crash data. The number of sites used in the analyses varied between 35 and 196, and the calibration factors calculated for 3 years varied between 0.28 and 1.98.

Lubliner et al [48] calibrated the HSM SPFs of rural two-lane two-way roads in Kansas, and also developed state-specific SPFs. The data used for the calibration and development of SPFs were obtained from a KDOT-maintained database that included shoulder width, lane width, urbanization degree, and AADT for 2007. The required information that was not available in this database – such as horizontal curvature, vertical grades, and the information to determine the RHR – was extracted from as-built plans, and also from online maps. In addition, KDOT maps were utilized to gather traffic counts to estimate AADT for years other than 2007. The study used crash data for 2005-2007. It was stated in the study that some of the most difficult data to collect were the AADT values for minor roads for intersections. Therefore, the study estimated these values using and ad hoc method called "travel shed." A statewide calibration factor of 1.48 was calculated for road segments. The calibration factors for three-leg and four-leg stop-controlled intersections were calculated as 0.28 and 0.19, respectively. The level of crash prediction accuracy was not satisfactory for the KDOT; therefore, it was decided that the next step would be to develop state-specific SPFs.

Dissanayake and Aziz (74) continued the work performed by Lubliner et al. (48), and calibrated the HSM SPFs of rural multilane divided and undivided road segments and stop-controlled intersections for Kansas. Geometric and operational data were obtained from the KDOT database for the period 2011-2013. A total of 281 divided and 83 undivided rural multilane road segments and 199 four-leg and 65 three-leg intersections were used for calibration. The calibration factor for divided segments was calculated as 1.436 and 0.524 for the total number of crashes and fatal/injury crashes, respectively. For undivided segments, these were calculated as 1.495 and 0.359. The calibration factors for four-leg stop-controlled intersections were 0.44 and 0.21 for total number of crashes and fatal/injury crashes, respectively. For three-leg stop-controlled intersections these were calculated as 0.92 and 0.47. Using the same dataset, the study developed state-specific SPFs for segments only in a form similar to that used in the HSM, for total crashes and fatal/injury crashes separately. The study also developed four distinct SPFs for total and fatal/injury crashes for segments using additional variables such as lane width, speed limit, shoulder width, etc. The evaluation of these models was performed using BIC and AIC measures.

Interviews

The team conducted in-person interviews with researchers who conducted similar research projects for the DOTs of Pennsylvania, South Carolina, Missouri, Kansas, Kentucky, and New York. The objective of these interviews was to learn the current state of practice and identify possible hardships in conducting similar analyses.

The following is a list of important notes from these interviews:

- Among the software tools used for SPF calibration are the Interactive Highway Safety Design Model (IHSDM), Smart Spreadsheets, and the Calibrator tool. It was unanimously agreed that the available databases do not include all of the data elements required by the HSM, and that manual data extraction was required. The research team was told that many graduate and undergraduate students participated to manually extract the required data elements. The most common tools used for the manual data extraction process were state-developed video referencing tools and Google Maps.
- The team was told that crash locations as reported by police officers were often found to be erroneous when compared with detailed crash reports, especially at intersections.
- While most studies relied on the urbanization code – i.e., rural or urban – included in the available dataset, the Missouri study determined the urbanization degree based on the population density database.
- None of the researchers interviewed investigated the validity of the AADT data for segments or for the major and minor legs at intersections. Those interviewed

mentioned that they used the AADT shown in the database and did not further research the proximity of AADT stations to the intersections.

Findings

Since the publication of the first edition of the HSM, many states have attempted to estimate local calibration factors or to develop state-specific SPFs. As evidenced from the review of the previous work, although the calibration factor, as shown in equation (4), is a straightforward ratio between the estimated and observed crash frequencies, the major hurdle in calculating the calibration factor is the use of the SPFs provided by the HSM. In order to apply the corresponding CMFs, various geometric and traffic datasets are required.

There are 60 variables used in the HSM predictive models, 41 of which are required and 19 of which are desirable. Thus, data collection and extraction pose the most difficult challenge, whether the objective is calibration or development. As stated in Shin et al [(5)], states' readily available datasets are not currently built for a seamless application of the HSM guidelines. The majority of the 60 variables are not commonly collected, such as driveway density, side slope, roadside hazard rating, liquor store density, pedestrian volumes at intersections, horizontal and vertical curvature, number of left- and right-turn lanes at intersections, etc. Collecting or extracting these data and also integrating data from various sources requires considerable time and man-hours.

Based on the literature review, among the required datasets, the horizontal curvature data (i.e., curve radius and length for two-lane two-way rural roadways) stand out as the most problematic. With a few exceptions – i.e., Oregon DOT [(6)] and Ohio DOT [(55)] – state DOTs do not have a curvature dataset in a usable format that can be easily integrated into the process of finding homogeneous segments. The lack of this dataset has led researchers to assume default values for horizontal curvature CMFs or to omit sections with horizontal curvature (e.g., references [9,10,13,49, 67]) or to manually extract this information using Google Earth (e.g., references [8, 60, 62, 72]) or built-in plans [(48)].

In addition, although state DOTs conduct comprehensive roadway traffic count programs, the available AADT datasets are often insufficient to meet the requirements of the HSM guidelines. As mentioned in the Introduction section, the AADT variable is utilized in all SPFs. The most common issue raised in previous studies is the lack of minor leg AADT at intersection facilities, especially in rural areas. In order to overcome this hurdle, some researchers have attempted to use statistical models to estimate AADT values at intersections where there are no available data [(6, 48)]. Also, because AADT values are not typically updated every year at every intersection and roadway segment, many researchers use interpolated values for missing years, as suggested by the HSM.

It has been noticed by the research team that none of the previous studies has alluded to the reliability of the available counts at intersections. As mentioned earlier, the base

SPFs for intersections include two AADT values, one for the major and one for the minor approach. Reliability of the AADT values can be assured when one has the turning movement counts collected at the target intersection. However, given the well-known limitations in obtaining these counts, it would be very unusual to have AADT values obtained from sensors located immediately at each approach before an intersection. In reality, the AADT value for each approach is often estimated using the counts collected by a detector located at the upstream of the approach leg. One would expect that the reliability of the AADT value would diminish in proportion to the distance of a detector from the target intersection. This issue shall be investigated in the analyses carried out in this research project.

Another issue that was raised in the previous studies is the fact that the urbanization classification of a roadway segment as defined by the HSM might not always match the urbanization code in a state's database (9, 48). According to the HSM, "rural" refers to any segment outside an urban area, with a population of less than 5,000 people. However, the HSM leaves the final decision to the users, as there can be exceptions to the rule. For example, Lubliner et al. (48) found that some segments that can be classified as rural under the HSM were observed to have characteristics of an urban area, such as on-street parking, sidewalks, and downtown-style developments. Therefore, the research team followed the urbanization classification of segments as coded in the SLD database. The SLD database is being used by various departments in NJDOT, and we have deemed it best to follow the same values in order to avoid confusion when applying either the calibrated or state-specific SPFs in safety analyses.

The most frequently discussed issue in the research is the minimum segment length criterion set by the HSM. The HSM suggest that the homogenous roadway segments be no less than 0.10 miles. However, as evidenced in the previous studies, this criterion is hardly ever met, especially for urban roadway segments. For example, Shrinivasan et al. (13) used a minimum of 0.04 miles for urban roadway segments in Florida, and Dixon et al. (6) used a minimum of 0.07 miles for urban segments in Oregon. Similarly, Shin et al. (5) reported that over 60 percent of rural roadway segments and over 80 percent of urban roadway segments in their study were shorter than 0.1 miles.

Some studies also discussed the impact of the crash reporting threshold on the scale of the calibration factors. Crash data from Washington and California were used in developing the HSM predictive models, and any deviation from the baseline crash reporting threshold would affect the estimation results for property damage crashes. For example, in Oregon crashes with over $1,500 of property damage are reported, while these values are $700 and $750 in Washington and California, respectively (6). Currently this threshold is $500 in NJ. According to Titze and Faron (75), approximately 65 percent of the states have a higher threshold than NJ, which is one of only 15 states with a threshold below $1,000, including Washington and California.

As seen in the literature review, some studies solely focused on calibrating the SPFs presented in the HSM, and some attempted to develop state-specific SPFs for some or all of the facilities in the HSM. Those who only worked on calibration did not elaborate on their findings or discuss the possible direction that the state agency would take based on the calibration results in terms of either applying these calibration factors or developing state-specific SPFs. The reported calibration factors in the previous studies varied considerably, from 0.15 in Dixon et al (6) to 4.45 in Srinivasan and Carter (3). In addition, most studies did not present a thorough analysis of the calibration factors in terms of their statistical significance, such as coefficient of variance, CURE plots, etc.

Although the choice of whether to calibrate the HSM SPFs or develop state-specific SPFs is not discussed in detail in many of the previous studies, the main objective of these studies was presented as achieving the most accurate crash frequency predictions. As stated in Srinivasan and Carter (12), the future application of SPFs should be considered when making this decision. Although the HSM suggests developing state-specific SPFs if more accurate crash prediction results are sought, this strategy has not been supported by the findings of some studies. For example, after performing both the calibration process and SPF development for Florida, Srinivasan et al. (13) found that the crash prediction did not improve overall when state-specific SPFs were used. Saito et al (49) found that both approaches produced similar results in Utah. And finally, Mehta and Lou (10) found that while both approaches yielded reliable results for Alabama rural roadways, the calibration process was easier to apply.

AVAILABLE DATA SOURCES

The available data sources are grouped into three categories by type: (1) Traffic Volume, (2) Roadway Features, and (3) Roadway Crashes. Information for each data source is summarized in Table 4.

Table 4: General available data sources for New Jersey

Data Type	Data Sources	Dates
Traffic Volume	Sensor Database	2009–2017
	Daily Turning Movement Counts (TMC) Database	2015–2018
	Weekly TMC Database	2017–2018
Roadway Features	Straight Line Diagrams (SLDs)	2017
	NJ Roads Centerlines GIS Geodatabase	2018
	Google Street View	–
Roadway Crashes	Voyager Safety Crash Database	2011–2015

Traffic Volume Data

The New Jersey Traffic Monitoring Program of NJDOT maintains traffic detector stations that are used to collect continuous and short-term traffic count data on state roadways. This data collection process is conducted by the Bureau of Transportation Data and Safety (BTDS) in accordance with the FHWA Traffic Monitoring Guide (TMG) and the AASHTO Guidelines for Traffic Data Programs. The traffic counting program is designed to utilize, at a minimum, 48-hour short-term counts to produce estimates of AADT. The permanent or continuous site elements consist of approximately 95 Traffic Volume/Speed Stations (TVS), 90 permanent Weigh-in-Motion (WIM) System sites, and 50 Major Stations of Volume and Classification, with data collected monthly on 7-days' duration. The Short-Term Count Program includes about 5,500 Volume (48-hour) sites counted in a 3-year cycle, and about 4,000 ramps counted in a 6-year cycle. In addition, approximately 400 special counts are performed each year as needed. These special counts are normally requested by NJDOT engineers to support projects and/or investigations. These counts include volume, classification, or turning movement counts at select intersections.

Traffic volume data were obtained from two databases provided by NJDOT: the NJDOT sensor database and the turning movement counts (TMC) database. The subsections that follow describe these two traffic volume databases.

<u>Sensor Database</u>

The sensor database provided by NJDOT consists of two separate tables: *Stations* and *Counts*. Figure 2 shows the key data elements available in each table. The sensor volume dataset is generated by combining each table using the unique ID numbers of sensors. This process was performed by developing a C programming code that reads each table, relates each entry using the sensor ID number, and outputs the relevant information as listed in Figure 1.

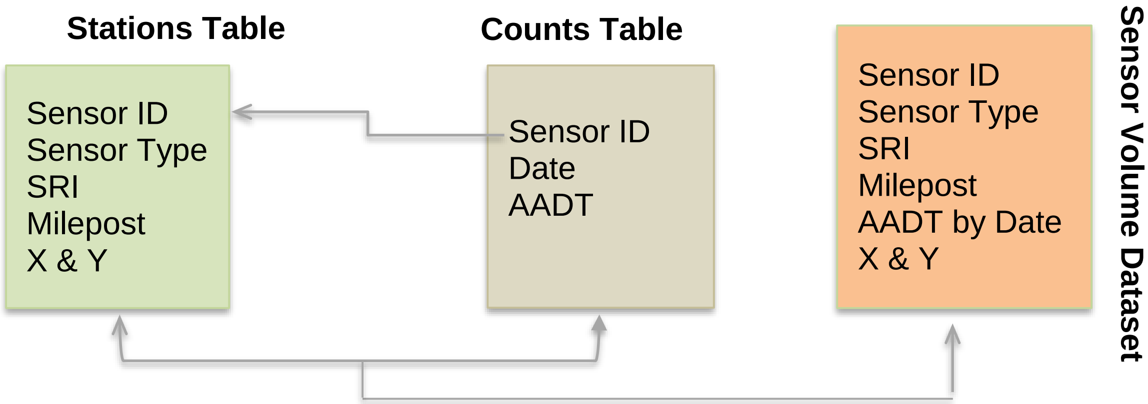


Figure 2: NJDOT sensor database and its key data elements

The sensor volume dataset includes AADT values collected from 10,959 sensors from 2009 to 2017. As noted earlier, the majority of these sensors are used for short-term traffic counting, the duration of which ranges from 2 to 3 days. These counts are then converted to AADT values. The short-term traffic count readings are repeated at each designated location every 3 years. Therefore, aside from continuous counting stations, the majority of the sensors have AADT values for specific years between 2009 and 2017.

TMC Database

The TMC database provided by NJDOT includes hourly movement counts collected at various intersections. The TMC database is formed by combining data from two sources. The first set of TMC data was provided by Mobility and Systems Engineering department in MS Excel format. This dataset covers a total of five intersections: four intersections on NJ 18 and one intersection on US 130. Data were available for 1 week in April and 1 week in July 2018 at NJ 18 intersections, and for 1 week in October 2017 at the US 130 intersection. The second set of TMC data was provided by the Bureau of Transportation Data and Safety. This data was provided in *.csv format, covers the years 2015 to 2018, and includes counts at 138 intersections collected over 12-hour periods. The following key data elements are included: date, start and end hours, Standard Route Identifier (SRI), milepost and functional class of primary and secondary approaches of the intersection, and hourly counts for each legal turning movement over a 12-hour period.

The difficulty in using these datasets was twofold. The first issue was that each turning count was provided in a separate *csv file, and the second was that the counts were available in a detailed yet unusable format. A sample of this data is shown in Figure 3.

```
COUNT_ID|15t-102
TABLE|Count_Event|29|Count_ID|Count_Date|Count_Time_From|Count_Time_To|Rt_NS_SRI|Rt_NS_MP|Rt_NS_FuncClass|Rt_EW_SRI|Rt_EW_MP|Rt
DATA|Count_Event||15t-102|3/17/2015|6:00:00 AM|6:00:00 PM|14271031__|0|19|00000046__|26.45|14|driveways|Old Wolfe Rd|US 46|US 4
TABLE|Count_Data|6|Count_ID|Dir|Approach|Period|Classification|Count|
DATA|Count_Data||15t-102|WB|TH|7:00:00 AM|ALL|87|
DATA|Count_Data||15t-102|WB|RT|7:00:00 AM|ALL|0|
DATA|Count_Data||15t-102|WB|U|7:00:00 AM|ALL|0|
DATA|Count_Data||15t-102|NB|LT|7:00:00 AM|ALL|0|
DATA|Count_Data||15t-102|NB|TH|7:00:00 AM|ALL|0|
DATA|Count_Data||15t-102|NB|RT|7:00:00 AM|ALL|51|
DATA|Count_Data||15t-102|NB|U|7:00:00 AM|ALL|0|
DATA|Count_Data||15t-102|EB|LT|7:00:00 AM|ALL|0|
DATA|Count_Data||15t-102|EB|TH|7:00:00 AM|ALL|265|
DATA|Count_Data||15t-102|EB|RT|7:00:00 AM|ALL|2|
DATA|Count_Data||15t-102|EB|U|7:00:00 AM|ALL|0|
DATA|Count_Data||15t-102|SB|LT|7:15:00 AM|ALL|0|
DATA|Count_Data||15t-102|SB|TH|7:15:00 AM|ALL|0|
DATA|Count_Data||15t-102|SB|RT|7:15:00 AM|ALL|0|
DATA|Count_Data||15t-102|SB|U|7:15:00 AM|ALL|0|
DATA|Count_Data||15t-102|WB|LT|7:15:00 AM|ALL|9|
DATA|Count_Data||15t-102|WB|TH|7:15:00 AM|ALL|145|
DATA|Count_Data||15t-102|WB|RT|7:15:00 AM|ALL|0|
DATA|Count_Data||15t-102|WB|U|7:15:00 AM|ALL|0|
DATA|Count_Data||15t-102|NB|LT|7:15:00 AM|ALL|0|
DATA|Count_Data||15t-102|NB|TH|7:15:00 AM|ALL|0|
DATA|Count_Data||15t-102|NB|RT|7:15:00 AM|ALL|44|
DATA|Count_Data||15t-102|NB|U|7:15:00 AM|ALL|0|
DATA|Count_Data||15t-102|EB|LT|7:15:00 AM|ALL|0|
DATA|Count_Data||15t-102|EB|TH|7:15:00 AM|ALL|261|
DATA|Count_Data||15t-102|EB|RT|7:15:00 AM|ALL|3|
DATA|Count_Data||15t-102|EB|U|7:15:00 AM|ALL|0|
```

Figure 3: A sample of TMC count data provided in *.csv format

In order to combine the TMC data files and make them more useful, the research team developed a C programming code that can read these data files and produce a summary of AADT for major and minor approaches. In addition, the code developed also locates each intersection's latitude and longitude information using the *pt_int_approach* table of the SLD database, as explained in the following subsection.

Roadway Features Data

Roadway features data were extracted from three data sources: the SLD database, GIS maps, and Google Street View.

<u>Straight Line Diagrams Database</u>

NJDOT's SLD database is the richest source of information on roadway features. This database was provided by NJDOT in MS Access format. It includes various tables on different geometric and operational aspects of NJ roadways. The key tables of interest for this study and the information within each table are presented in Table 5.

Detailed information on how this dataset is used to locate intersection and homogeneous segments and information regarding data processing is presented later in this section.

Table 5: Description of key SLD tables

Table Name	Fields										Notes
ln_f_system	ID	SRI	MP_S	MP_E	f_system_code						f_system_code 1: interstate, 2: principal arterial/freeway/ expressway, 3: principal arterial, 4: minor arterial, 5: major collector, 6: minor collector, 7: local
ln_urban_code	ID	SRI	MP_S	MP_E	is_urban (Y/N)						
ln_highway_type	ID	SRI	MP_S	MP_E	type						type 0: undivided, 1: divided, 2: dual-dual
ln_median_type	ID	SRI	MP_S	MP_E	type						type 0: none, 1: grass, 2: curbed, 3: positive, 4: painted
ln_median_width	ID	SRI	MP_S	MP_E	width (ft)						
ln_shoulder_type	ID	SRI	MP_S	MP_E	location	X1	X2	Y1	Y2	Park	location 1: right, 2: left, 3: middle, 4: right middle, 5: left middle park metered, prohibited, unlimited
ln_shou_width	ID	SRI	MP_S	MP_E	width (ft)						
ln_pave_width	ID	SRI	MP_S	MP_E	width (ft)						
ln_lane_count	ID	SRI	MP_S	MP_E	no_of_lanes						
ln_speed	ID	SRI	MP_S	MP_E	speed limit (mph)						
pt_intersection	ID	SRI	MP	SRI_X	MP_X	type	Name of crossing approach 1	Name of crossing approach 2			type 0: unsignalized, 1: signalized, 2: interchange, 3, 4: circle, 5: median, 6,7,8: jug handles
pt_int_approach	ID	SRI	MP	skew	type	X	Y	Signal (Y/N)			type 3-way, 4-way, other
pt_highway_lighting	ID	SRI	MP	location	offset (ft)	X	Y				location 1: right, 2: left, 3: middle,

NJ Roads Centerlines Geographic Information Systems Geodatabase

The geodatabase of NJ roads centerlines GIS dataset was downloaded from the NJ Geographic Information Network website [76]. The road centerlines geodatabase was built off of the existing NJDOT roadway network data by expanding coverage to include alleys and private roads, physical segmentation of geographic features at intersections, and jurisdictional boundaries and zip code boundaries, and by adding attribute values for alternate road names and linear referencing attributes [76].

The centerlines geodatabase includes various tables, but the two most important are:

1) *Tran_road_NJ_Tran_road_centerline_NJ shapefile*: This is the shapefile of the entire NJ roadway network. There are nearly 490,000 links in this shapefile, along with various link attributes, the most important of which are latitude and longitude readings of links' start and end nodes, and the roadway SRI and roadway name.

2) *Tran_road_LRS_NJ table*: This is a linear referencing system (LRS) table for all roadways in NJ that includes information such as roadway SRI, start milepost, end milepost, route type, etc. This table, however, does not include the latitude and longitude readings of links' start and end nodes.

These two tables can be merged using a common field name SEG_GUID in MS Access, as shown in
Figure **4**. After merging the two tables, the result was the NJ GIS Links database, which contains the SRI, name, start and end milepost, and latitude and longitude readings of each link's start and end nodes. The dataset contains nearly 490,000 roadway links.

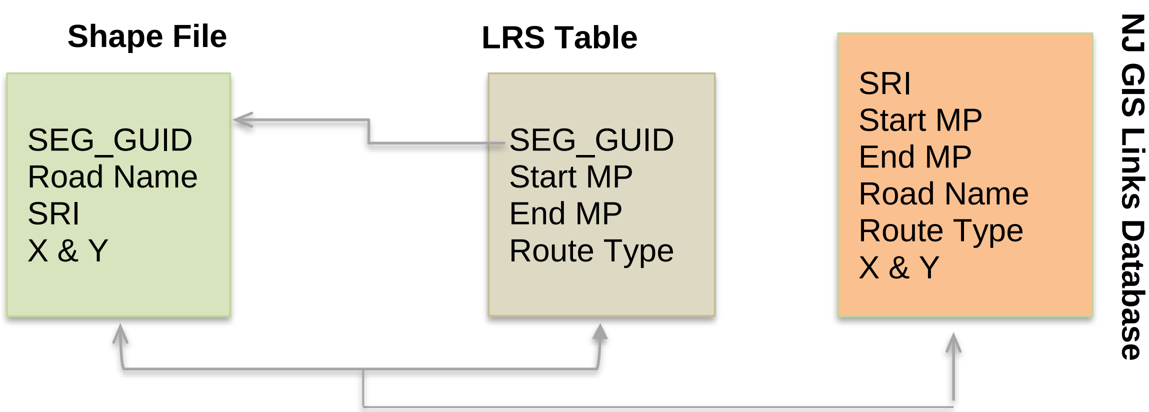


Figure 4: NJ GIS Links database and its key data elements

Google Street View
The Street View service of Google Maps was frequently used both to collect the roadway feature data missing in the available datasets and to verify the available data, especially for intersections, as discussed in the relevant subsections that follow.

Voyager Safety Crash Data

The crash data were provided by NJDOT for 2011 to 2015. There is much valuable information provided in this dataset. Table 6 lists the key data elements used as part of this study.

Table 6: Key data elements of the Voyager Safety Crash database

Data Element	Notes
Crash ID	
SRI	
Milepost	
Date	
Time	
Severity	Fatality, Injury, Property Damage Only
Severity code	Complaint of Pain, Incapacitated, Killed, Moderate Injury, Property Damage Only
Collision type	
Crash type	Animal, Backing, Encroachment, Fixed Object, Left Turn/U-Turn, Non-fixed Object, Opposite Direction (Head-On), Opposite Direction (Sideswipe), Overturned, Pedalcyclist, Pedestrian, Rail Car-Vehicle, Right Angle, Same Direction (Head-On), Same Direction (Sideswipe), Struck Parked Vehicle, Other, Unknown
Number of vehicles	
Number of fatalities	
Number of injuries	
Number of pedestrian fatalities	
Number of pedestrian injuries	
Latitude and Longitude	

It should be noted that the crash database only includes reportable crashes. A reportable crash is one that results in the injury or death of any person or damage to property of any one person in excess of $500. It should also be noted that the location of a crash, namely the SRI and milepost, are usually the most incomplete part of the crash database, making it difficult, if not impossible, to provide accurate location data. Table 7 shows the number of crashes per year, along with the percentage of crashes with complete SRI and milepost information. As the table shows, 40 to 50 percent of the crashes do not include this information.

In NJ, the police are equipped with GPS devices that can record the latitude and longitude of crashes. This information is supposed to be added to the crash report; however, the percentage of crashes for which this information is actually included in the raw crash data is very low. NJDOT post-processes the raw crash data and geocodes crashes with missing coordinates using SRI and milepost and cross street names. After this post-processing, data on nearly 95 percent of crashes from 2011 to 2015 include latitude and longitude readings, as shown in Table 7.

Table 7: Crash data statistics

Year	**Number of Crashes**	**SRI & Milepost**	**Latitude & Longitude**
2011	295,093	53.77 %	93.67 %
2012	284,059	53.48 %	93.69 %
2013	289,423	52.92 %	94.23 %
2014	290,209	53.78 %	94.29 %
2015	270,939	61.68 %	95.48 %

The information gathered from the three data sources described in this section can be used to generate the data required for the calibration/development of SPFs for intersections and homogeneous segments per facility types included in the HSM. However, before generating these required datasets, the compiled data should be cleaned and corrected as described in the section that follows.

DATA PREPARATION AND CLEANING

The objective of this section is to describe the cleaning process for data obtained from the three available datasets described in the previous section, and also to discuss several observed inconsistencies and instances of inaccurate information in these data sources.

Traffic Volume Data

AADT is a key variable in the HSM's crash prediction model. The manual requires AADT at homogeneous segments and at major and minor road approaches of intersections for the calibration/development process. As mentioned above, the AADT data used in this study originates from two sources: the sensor database and the TMC database. Data cleaning was only required for the sensor database.

<u>Inconsistencies in the Sensor Database</u>

Four typical error types were identified in the sensor database, as depicted in Table 8. The first two types, namely the sensor type and direction, do not have any effect on the AADT values, while the last two types could lead to erroneous AADT values. Regarding Type 3 errors, there are 685 records that show multiple AADT values for a detector in the same year. Understanding well that these are not errors per se, the research team used the average AADT value for each year in the analyses. Type 4 errors were identified for 63 records; these records were removed from the analyses.

Table 8: Examples of sensor database errors

<table>
<tr><th>Error Types</th><th>Description</th><th>Example</th></tr>
<tr><td>Type 1</td><td>Same station with different station type</td><td>
<table>
<tr><th>SI_STATIO</th><th>COUNT_Y</th><th>STATION_</th><th>DIRECTIO</th></tr>
<tr><td>4-3-016</td><td>01-Jan-13</td><td>Volume 48hr</td><td>East/West</td></tr>
<tr><td>4-3-016</td><td>01-Jan-10</td><td>major</td><td>East/West</td></tr>
<tr><td>4-3-016</td><td>01-Jan-10</td><td>major</td><td>East/West</td></tr>
<tr><td>4-3-016</td><td>01-Jan-11</td><td>major</td><td>East/West</td></tr>
<tr><td>4-3-016</td><td>01-Jan-11</td><td>major</td><td>East/West</td></tr>
<tr><td>4-3-016</td><td>01-Jan-12</td><td>major</td><td>East/West</td></tr>
<tr><td>4-3-016</td><td>01-Jan-12</td><td>Classificatior</td><td>East/West</td></tr>
</table>
</td></tr>
<tr><td>Type 2</td><td>Same station with different direction</td><td>
<table>
<tr><th>SI_STAT</th><th>EV_AADT_</th><th>EV_AA</th><th>EV_A</th><th>COUNT_YEAR</th><th>STATION_DESC</th><th>DIRECTION</th></tr>
<tr><td>3-3-032</td><td>0</td><td></td><td>66506</td><td>01-Jan-09</td><td>Weigh in Motion</td><td>South</td></tr>
<tr><td>3-3-032</td><td>0</td><td>0</td><td>68493</td><td>01-Jan-10</td><td>Weigh in Motion</td><td>West</td></tr>
<tr><td>3-3-032</td><td>0</td><td>0</td><td>61557</td><td>01-Jan-11</td><td>Weigh in Motion</td><td>South</td></tr>
<tr><td>3-3-032</td><td>0</td><td>0</td><td>63802</td><td>01-Jan-14</td><td>Weigh in Motion</td><td>North/South</td></tr>
<tr><td>3-3-032</td><td>0</td><td>0</td><td>60982</td><td>01-Jan-15</td><td>Weigh in Motion</td><td>South</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td>Type 3</td><td>Same year, multiple observations</td><td>
<table>
<tr><th>SI_STATION_NUM</th><th>EV_AADT_</th><th>EV_AADT_DI</th><th>EV_AADT_DIR2</th><th>COUNT_YEAR</th><th>STATION_DESC</th><th>DIRECTION</th></tr>
<tr><td>5-3-013</td><td>18763</td><td>9241</td><td>9522</td><td>01-Jan-10</td><td>Volume 48hrs</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>20771</td><td>9685</td><td>11086</td><td>01-Jan-10</td><td>Volume 48hrs</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>22530</td><td>12560</td><td>9970</td><td>01-Jan-10</td><td>Volume 48hrs</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>18188</td><td>8575</td><td>9613</td><td>01-Jan-10</td><td>Volume 48hrs</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>20530</td><td>10194</td><td>10336</td><td>01-Jan-13</td><td>Volume 48hrs</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>20151</td><td>9476</td><td>10675</td><td>01-Jan-10</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>20945</td><td>10105</td><td>10840</td><td>01-Jan-10</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>19277</td><td>9446</td><td>9831</td><td>01-Jan-10</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>21820</td><td>10361</td><td>11459</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>19828</td><td>9539</td><td>10289</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>20495</td><td>9751</td><td>10744</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>19390</td><td>9502</td><td>9888</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>20480</td><td>9971</td><td>10509</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>20428</td><td>9725</td><td>10703</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>18455</td><td>9694</td><td>8761</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>18548</td><td>8655</td><td>9893</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>19995</td><td>9105</td><td>10890</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>18620</td><td>8603</td><td>10017</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
<tr><td>5-3-013</td><td>19485</td><td>8966</td><td>10519</td><td>01-Jan-11</td><td>major</td><td>North/South</td></tr>
</table>
</td></tr>
<tr><td>Type 4</td><td>One direction AADT is missing and 0 value for total AADT</td><td>
<table>
<tr><th>SI_STATION_</th><th>EV_AADT_TC</th><th>EV_AADT_DI</th><th>EV_AADT_DI</th><th>COUNT_YEAI</th><th>STATION_DE</th><th>DIRECTION</th></tr>
<tr><td>1-1-026</td><td>43505</td><td>20641</td><td>22864</td><td>01-Jan-10</td><td>Weigh in Mo</td><td>North/South</td></tr>
<tr><td>1-1-026</td><td>40667</td><td>21666</td><td>19001</td><td>01-Jan-11</td><td>Weigh in Mo</td><td>North/South</td></tr>
<tr><td>1-1-026</td><td>0</td><td>20363</td><td>0</td><td>01-Jan-13</td><td>Weigh in Mo</td><td>North/South</td></tr>
<tr><td>1-1-026</td><td>43372</td><td>20733</td><td>22639</td><td>01-Jan-15</td><td>Weigh in Mo</td><td>North/South</td></tr>
</table>
</td></tr>
</table>

<u>Missing AADT Data</u>

As mentioned earlier, AADT data from sensors is not available for all years from 2009 to 2017. As suggested by the HSM, an estimated AADT for each year of the evaluation period was interpolated or extrapolated as appropriate. The default rules suggested by the HSM for estimating AADT for the missing data are [(2)]:

- If two or more years of AADT data are available, the AADTs for intervening years are computed by interpolation.
- The AADTs for years before the first year for which data are available are assumed to be equal to the AADT for that first year.
- The AADTs for years after the last year for which data are available are assumed to be equal to that for the last year.

Following the HSM's guidelines, the research team conducted the AADT interpolation or extrapolation for missing years and prepared a complete AADT database.

Roadway Features Data

As mentioned earlier, the SLD database is a rich source of data on the operational and geometric characteristics of NJ's roadways. However, the research team found several inconsistencies and errors. The following subsection provides a detailed description of the data cleaning process performed for the SLD diagrams.

<u>Preparing Intersection Database Using the SLD Database</u>

Correct identification of intersection types and locations is crucial because intersections are used both in determining homogeneous roadway segments and in generating the list of intersections per facility type. Here, the term "intersection" takes on a different meaning depending on the task at hand. To be specific, when the task is to determine homogeneous roadway segments, any point along the roadway that would make it possible for vehicles to make a turn is labeled as an intersection. Roadways should be split at these points to determine homogeneous segments. Within this task, in addition to the conventional definition of intersection, median left-turns/U-turns, circles, jug handles, and interchanges are also deemed as intersections. In contrast, when the task is to generate the list of intersections per facility type, the focus is to identify only the conventional 3-way and 4-way intersections, while dismissing any other types of intersections.

There are two tables in the SLD database that include information regarding intersections, namely *pt_intersection* and *pt_int_approach*. A comparison of these two tables is provided in Table 9.

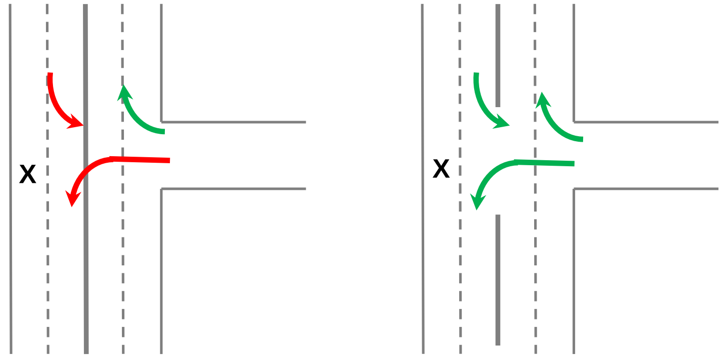
Table 9: Comparison of pt_intersection and pt_int_approach tables

pt_intersection	**pt_int_approach**
Advantages ▪ Includes cross approach's SRI, milepost, and name ▪ Includes not only intersections but also interchanges, median turns, jug handles ▪ Includes the cross approach entry with its name even if no SRI or milepost exists **Disadvantages** ▪ Does not include latitude and longitude ▪ Does not include skew angle ▪ Does not specify number of approaches	**Advantages** ▪ Includes latitude and longitude ▪ Includes skew angle ▪ Specifies the type of intersection and number of approaches **Disadvantages** ▪ Classifies interchanges and circles as intersections ▪ Does not include cross approach if it does not have an SRI, thus type of intersection is not always correct ▪ Does not include cross approach's name

It can be seen in Table 9 that neither table includes complete information on intersections. The *pt_int_approach* table has one entry for each approach in the database. The major advantage of this table is the fact that it includes the latitude and longitude and skew angle. It also includes the type of intersection, namely whether it is a 3-way, 4-way, or 5-or-more-way intersection. However, because intersection type is determined simply by the number of approaches and the table does not include any approach that does not have an SRI, intersection type is sometimes reported inaccurately.

The major advantage of the *pt_intersection* table is that it includes each cross approach's SRI, milepost, and name, and even when the approach does not have an SRI, the table includes the name of the approach. However, the table does not include latitude and longitude or intersection type.

The common drawback of both tables is the fact that the type of 3-way intersection cannot be determined when the roadway is divided. That is, as shown in Figure 5, when a roadway is divided at a 3-way intersection, in cases in which turns to and from the major approach are restricted by the median, the intersection is labeled as one-directional 3-way, while in cases in which turns are not restricted, the intersection is labeled as two-directional 3-way.



One-directional 3-way intersection

Two-directional 3-way intersection

Figure 5: Two types of 3-way intersections

Identifying 3-way intersections based on their type is important only when the task is to generate a list of intersections per facility type that are then used for calibration/development of SPFs for intersections. The reason is that, as mentioned in the next subsection, in the available Voyager Safety crash database, direction of roadway on which a crash occurs is rarely reported. Consequently, when crashes are assigned to 3-way intersections regardless of type, there is an inherent error in the calibration or development procedure. That is, suppose a crash occurs at the intersection along the southbound direction as shown by the “X” in Figure 5. Let us also suppose that the direction of the crash is not reported in the crash database. Then it would be correct to assign this crash to the intersection provided that the intersection type is a two-directional 3-way. However, if the type of intersection is one-directional 3-way, then this assignment incorrectly assigns the crash to the intersection, even though this crash is not due to the intersection, which will bias the results of calibration/development.

In order to correctly identify the type and location of intersections, the research team developed a C programming code that processed both tables and also makes use of the NJ GIS Links database. Type of intersection can be easily identified using this database, as shown in Figure 6. In the map shown in the left of the figure, there are two one-directional 3-way intersections. In the NJ GIS Links database, the type of such intersections can be easily identified, as shown in the right of the figure.

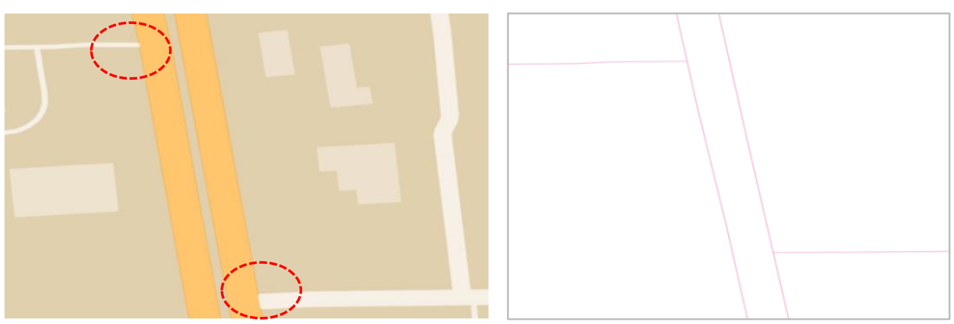


Figure 6: Identifying intersection types using NJ GIS Links database

The decision flowchart for identifying the location and type of intersection using the NJ GIS Links database along with the relevant SLD tables is shown in Figure 7.

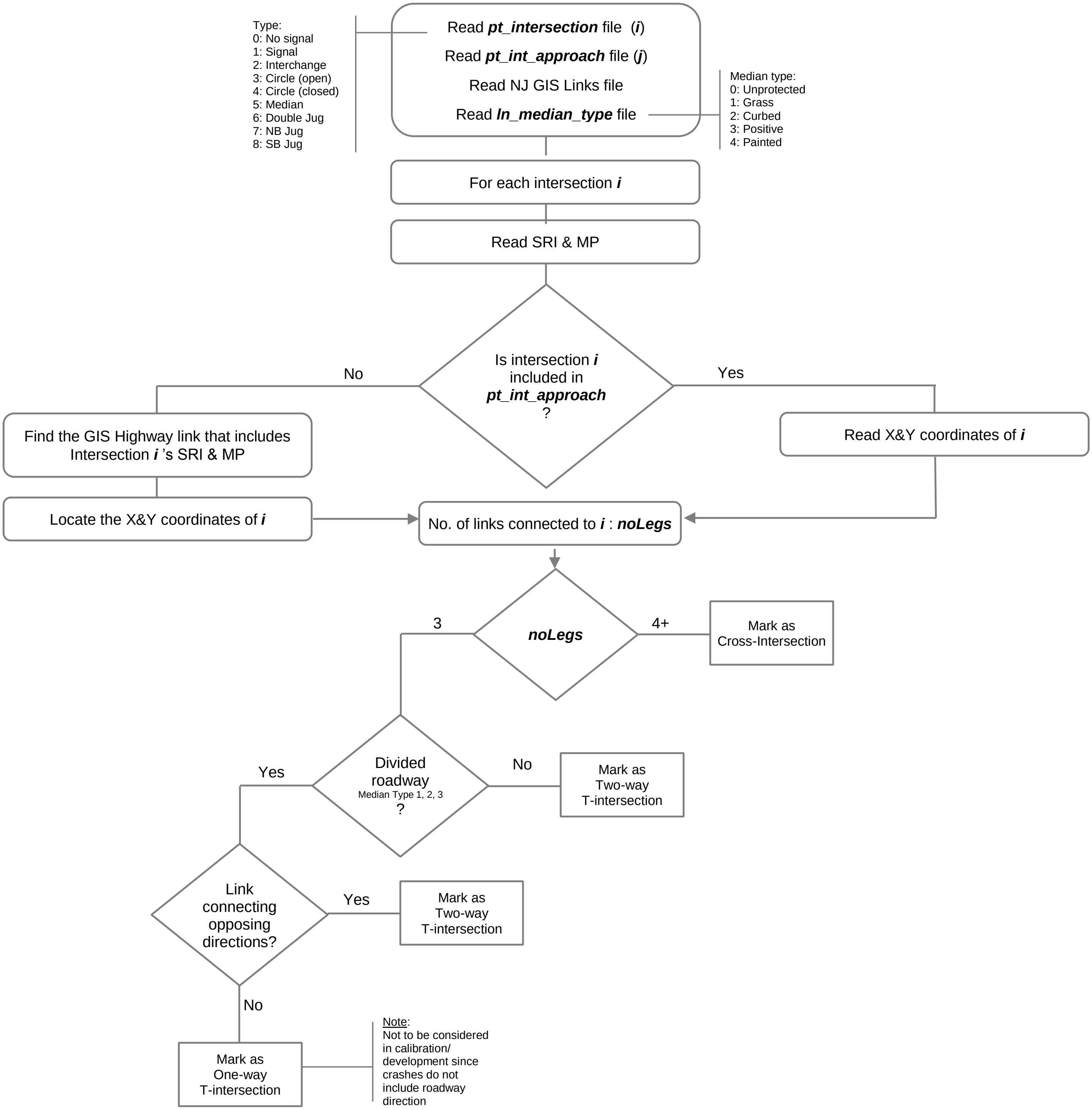


Figure 7: Decision flowchart for identifying location and type of intersections

This process can be briefly explained as follows. For each intersection in the *pt_intersection* table, the code first checks whether the same intersection is included in the *pt_int_approach* table by matching its SRI and milepost. When a match is found, the code then reads the latitude and longitude of the intersection from the *pt_int_approach* table, finds the corresponding node in the NJ GIS Links database, and determines the number of approaches connected to it. Otherwise, it uses the SRI and milepost of the intersection, and searches within the NJ GIS Links database to match the corresponding node. Using the cross approach's name, the code verifies whether this is the correct node or not, reads its latitude and longitude, and determines the number of approaches connected to the node. If this number, i.e., *noLegs*, is equal to 3, the intersection is labeled as a 3-way intersection. In order to determine if it is one-directional or two-directional, the code first checks whether or not the intersection is located at a divided roadway using the *ln_median_type* table. If the roadway is undivided, it is labeled as a two-directional intersection. Otherwise, the code checks whether there exists a link at the intersection that connects the opposing directional links of that roadway, making it possible for vehicles to make a left turn to and from the major approach. If such a link exists, the intersection is labeled as a two-directional 3-way; otherwise it is labeled as a one-directional 3-way.

It should be mentioned that due to the high number of iterations and matching within multiple large databases, the code takes nearly 10 hours to execute on a workstation with an i7 3.30 GHz processor. In the end, the code outputs a total of 296,384 intersections, consisting of conventional 3-way, 4-way, and 5-or-more-way intersections, medians, jug handles, circles, and interchanges.

280,685 of these intersections were matched in the *pt_int_approach* table. Filtering out interchanges and circles, 279,530 of these intersections include conventional intersections, jug handles, and median left/U-turns. Among these intersections, the number of 3-way, 4-way, and 5-or-more-way intersections that also appear in the *pt_int_approach* table is 275,366. A summary of these intersection types is shown in Table 10.

Table 10: Summary of conventional intersections found in pt_int_approach table

Intersection Type	Signalized	Unsignalized	Total
3-way	2,944	198,641	201,585
4-way	8,785*	64,306*	73,091*
5-or-more-way	570	823	1,393

Table 11 presents a summary of 3-way intersections by type and signalization. As shown, there is a total of 1,168 signals that are one-directional 3-way intersections. As mentioned earlier, these intersections were removed from the list of intersections used for SPF calibration/development.

Table 11: Type of 3-way intersections found in the pt_int_approach table

3-way Type	Signalized	Unsignalized	Total
One-directional	105	1,513	1,618
Two-directional	2,839*	197,128*	199,967*

Attention should be given to the number of signalized one-directional 3-way intersections, as highlighted in Table 11. Since this type of intersection is not conventional, the research team investigated the output in detail to detect any possible errors in the decision process illustrated in Figure 7. It was revealed that there are, in fact, examples of such intersections. For example, as shown in Figure 8, the signalized intersection on Route 22 at milepost 48.26 consists of two separate 3-way intersections with a concrete barrier prohibiting left turns. Since each of these two 3-way intersections serve one direction of the main roadway, they are removed from the intersection database.

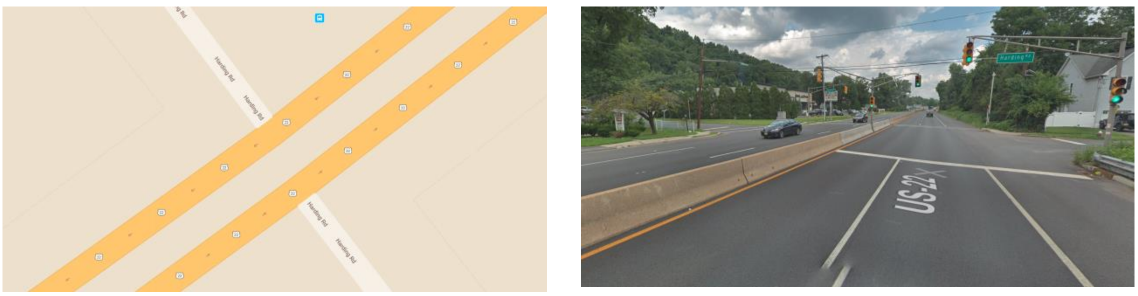


Figure 8: Example signalized one-directional 3-way intersection on Route 22

Other examples similar to the intersections shown in Figure 8 are Route 20 at milepost 1.32, Route 9 at milepost 107.4, and Route 30 at milepost 50.18.

In addition, there are also cases where an intersection was labeled as signalized but in reality is unsignalized (e.g., Route 30 at milepost 57.94).

Several of these intersections were labeled as signalized one-direction 3-way for two reasons. The first reason is that the NJ GIS Links database sometimes has missing links. For example, Figure 9 shows the intersection on Route 1 at milepost 40.16, where the minor leg does not have a corresponding link in the GIS database. The second reason is that there is occasional inaccurate information in the *ln_highway_type* and *ln_median_type* tables, where the roadway is reported as divided but in reality it is not. For example, Route 1 at milepost 59.64 is reported as divided, and because the algorithm searches for a link connecting the opposing directions of the primary approach, it cannot it labels it at one-directional; yet in reality Route 1 that specific milepost is undivided.

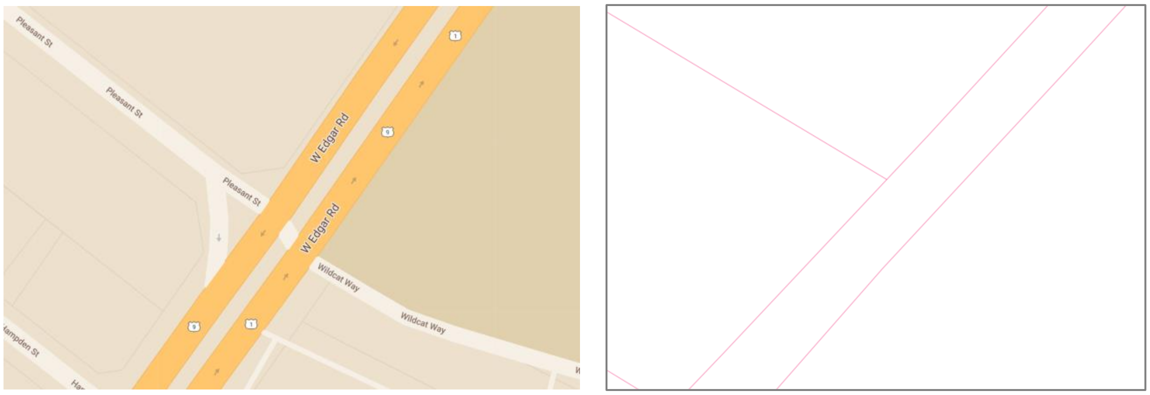


Figure 9: Example of a missing link in the NJ GIS Links database

Because it would be very time-consuming to sift through these intersections and determine which ones are in fact one-directional 3-ways and which are not, the research team decided to remove them from the intersections database to be used in the calibration/development process.

In addition to the intersections listed in the *pt_int_approach* table, there are a total of 15,699 additional intersections in the *pt_intersection* database. The developed code finds the exact location, i.e., latitude and longitude, of these intersections using the SRI and milepost. Due to the inaccuracies in the milepost values reported in the NJ GIS Links file, as discussed later in this section, the code verifies the intersection by matching the name of the crossing approach. If the name of the crossing approach given in the *pt_intersection* table is matched with the name in the NJ GIS Links database, the code marks the intersection as "*verified,*" otherwise it is marked as "*unknown.*".

Table 12 shows the summary statistics for the additional 15,699 intersections found. The table also shows the number of intersections verified by type. Among the 15,699 intersections found, 7,642 were verified. It should be mentioned that "unverified" does not connote that the intersection location and type are not accurate, but simply that its exact location could not be verified by the cross street name. However, since perusing the unverified database is not time efficient, the research team decided not to include these intersections in the calibration / verification process. The research team realized that the mileposts in the LRS might vary from 0.01 to 0.09 miles. Through our e-mail correspondence with the Bureau of Information Management and Technology on May 14, 2019, it was found that the LRS is out of date and the department is in the process of updating it. Once the modified LRS database becomes available, these intersections can be located very accurately.

Table 12: Summary of additional conventional intersections

Intersection Type	Signalized	Unsignalized	Total
3-way	594	8,906	9,500
3-way verified	179	5,063	5,242
4-way	394	2,781	3,175
4-way verified	209*	1,684*	1,893*
5-or-more-way	32	144	176
5-or-more-way verified	20	104	124

Table 13 presents a summary of the additional 3-way intersections by type and signalization.

Table 13: Summary of additional 3-way intersections by type and signalization

3-way Type	Signalized	Unsignalized	Total
One-directional	121	685	806
One-directional Verified	22	177	199
Two-directional	473	8,221	8,694
Two-directional Verified	157*	4,886*	5,043*

The reasons for the presence of signalized one-directional 3-way intersections were presented earlier; similar reasons apply to those in Table 13.

In summary, using the developed code, the research team was able to identify the locations and types of intersections by combining the information gathered from the pt_int_approach and pt_intersection tables and the NJ GIS Links database. It should be noted that the full list of 296,384 intersections, consisting of conventional 3-way, 4-way, and 5-or-more-way intersections; medians; jug handles; circles; and interchanges are used in determining the homogeneous roadway segments. As to the list of intersections to be used in the calibration / development process, the HSM covers only 3-way and 4-way intersections, therefore, those used in the final list are indicated by asterisks in Table 10, Table 11, Table 12, and Table 13. The final list to be used in the calibration / development process includes a total of 279,994 intersections, among which 74,984 are 4-way intersections (8,994 signalized and 65,990 unsignalized) and 205,010 are 3-way intersections (2,994 signalized and 202,014 unsignalized). Note that, as discussed in the relevant sections that follow, the number of intersections used in the calibration / development process is much lower than these values, based on the facility types included in the HSM and the availability of traffic volume data.

Inconsistencies and Missing Information in the SLD Database

The subsections that follow describe examples of inconsistent and missing information in some of the SLD database tables. There could be other instances of such cases discovered if a thorough quality check was performed. It should be noted that such cases are expected in large and comprehensive databases such as the SLD, and that the objective of this subsection is to point out these cases for future reference for

NJDOT. It should also be noted that the research team used the SLD database for 2017 in our analyses and that these cases might not appear in the later versions of the database.

Parking

As discussed in the urban and suburban facilities section, parking information is required for the calibration of SPFs for roadway segments. The SLD table named "*ln_shoulder_type"* does in fact include a field that indicates the type of parking as (i) metered, (ii) prohibited, or (iii) unlimited, as shown in Table 5. However, further investigation of this database revealed that the information provided is far from complete, and the majority of the roadways are not included. For example, there is no information given for Route 27 between mileposts 0.0 and 1.444, which corresponds to Princeton downtown, where there is metered parking. Also, in some cases the table reports unlimited parking, yet no parking is shown in Google Street View (e.g., Route 1 between mileposts 37.608 and 37.714).

Based on our calculations, about 12.2 percent of the state roadways are included in the *ln_shoulder_type* table.

Center Two-Way Left-Turn Lanes

Another task related to urban and suburban segments is to determine the roadway segments that include a center two-way left-turn lane. Investigation revealed that it is not possible to use the SLD database to determine whether the roadway segment includes a center two-way left-turn lane. For example, the segment on Route 41 between mileposts 2.37 and 2.41 includes a center two-way left-turn lane, as shown in Figure 10. The SLD database, however, marks this section as two-lane with a median width of 12 ft. Another example is the segment on Central Avenue in Clark Township between mileposts 4.08 and 4.17. This segment has two lanes in each direction, with a center two-way left-turn lane; however, the SLD database shows four lanes total with no median.

Figure 10: Center two-way left-turn lane on Route 41 in Deptford Township

Pavement Width

Pavement width information is available in the *ln_pave_width* table of the SLD database, as shown in Table 5. This value is the total width across all traveling lanes excluding the shoulder width. However, there are cases in which the pavement width values do not match the actual measurements in Google Earth, especially when the roadway is divided. For example, on Route 27 at milepost 20.0, where the roadway is divided with painted median, the SLD database reports the pavement width as 48 ft in the northbound direction and 36 ft in the southbound direction, with shoulder widths of 0 and a median width of 12 ft. There are two through lanes in each direction on this section. The total pavement width should therefore be 84 ft, but when measured on Google Earth, this value is around 55 ft. Another example is Route 41 between milepost 2.37 and 2.41. This section has two through lanes and one center two-way left-turn lane. It is indicated as being undivided in the SLD's *ln_highway_type* table. The pavement width for this segment is given as 24 ft, number of lanes as two, and median width as 0, yet Google Earth measurements show the pavement width as being 40 ft.

Lane Count

Lane count information is available in the *ln_lane_count* table of the SLD database. The research team identified several cases of inconsistencies between the actual number of lanes and the number reported in the SLD database. For example, for Jacksonville Road in Springfield Township in Burlington County (SRI 03000670), the database shows four lanes between mileposts 7.15 and 9.0; however, Google Street View shows that there are actually only two lanes.

Highway Type

As shown in Table 5, the *ln_highway_type* table shows whether a roadway is undivided, divided, or dual-dual. Most entries are given for the primary direction, but there are some entries for the secondary direction. For example, for Route 26, i.e., SRI 00000026__, the table reports that between mileposts 0.0 and 0.22 the roadway is divided, and that it is undivided between mileposts 0.22 and 2.54. There is also information on Route 26 Southbound, i.e., SRI 00000026_S, for which the table reports that between mileposts 2.315 and 2.54 the roadway is undivided. Assuming that the milepost information for the southbound direction corresponds to parent milepost range 0.0 to 0.225, this information is conflicting. Similar cases can be found for SRI 00000027__ and 00000027_S.

In addition, as mentioned in the previous subsection, there are some cases in which the roadway is reported as being divided when it is in fact undivided, or vice versa.

Voyager Safety Crash Database

The key data elements of the Voyager Safety crash database are presented in Table 6. As mentioned earlier, crash location is usually the most incomplete part of the database. There are crash entries that are missing SRI, milepost, or latitude and longitude readings. NJDOT post-processes the raw crash data and geocodes the latitude and longitude of crashes using any location-related information reported, such as cross street names, addresses, SRI, or milepost. As shown in Table 7, following this geocoding process, nearly 95 percent of crashes include latitude and longitude readings.

Knowing the latitude and longitude of crashes is of utmost importance for assigning crashes to intersections per facility type. To determine whether a crash is intersection-related, one must determine if its location is within 250 ft. of the intersection, as per the HSM's recommendations (2). The milepost information for crashes is deemed too coarse for such calculations. One milepost is 528 feet, and any inaccuracies in milepost information would lead to incorrect assignment of crashes to intersections. As mentioned in the preceding subsection, the latitude and longitude values of intersections are available in the final intersection database. Thus, one could easily calculate the distance between a crash and an intersection using its latitude and longitude readings. This calculation is presented in the next section.

To assign crashes to homogeneous segments, however, SRI and milepost information is required, because determining which segment a crash should be assigned to is not straightforward when using the latitude and longitude readings, especially when segments of different roadways are located close together. Inspection of the available crash database revealed that even though latitude and longitude information is available for nearly 95 percent of crashes, there are many instances in which the SRI or milepost information is missing. Table 14 shows the statistics for missing information for the crash database between 2011 and 2015.

The research team developed a code in C programming language to determine the missing SRI or milepost information using the available latitude and longitude data. This code uses the NJ GIS Links database, which includes the start and end latitude and longitude and start and end milepost of each roadway link in the state. The code follows the decision flowchart shown in Figure 11 to determine the missing SRI and milepost information.

Table 14: Statistics on Voyager crash data missing SRI and milepost information

Year	Total Number of Crashes	Number Missing SRI and/or Milepost Information
2011	295,093	127,786
2012	284,059	123,806
2013	289,423	128,398
2014	290,209	126,476
2015	270,939	99,131

It should be noted that the total size of crash data for 5 years is approximately 266 MB, with 605,000 crashes with missing information. The GIS Links database has 490,000 links. Therefore, it takes nearly 18 hours for the code to identify all missing SRI and milepost information. However, this is a one-time process. At this point, crash data for 2011 to 2015 has been processed, and nearly 99 percent of the missing information has been restored.

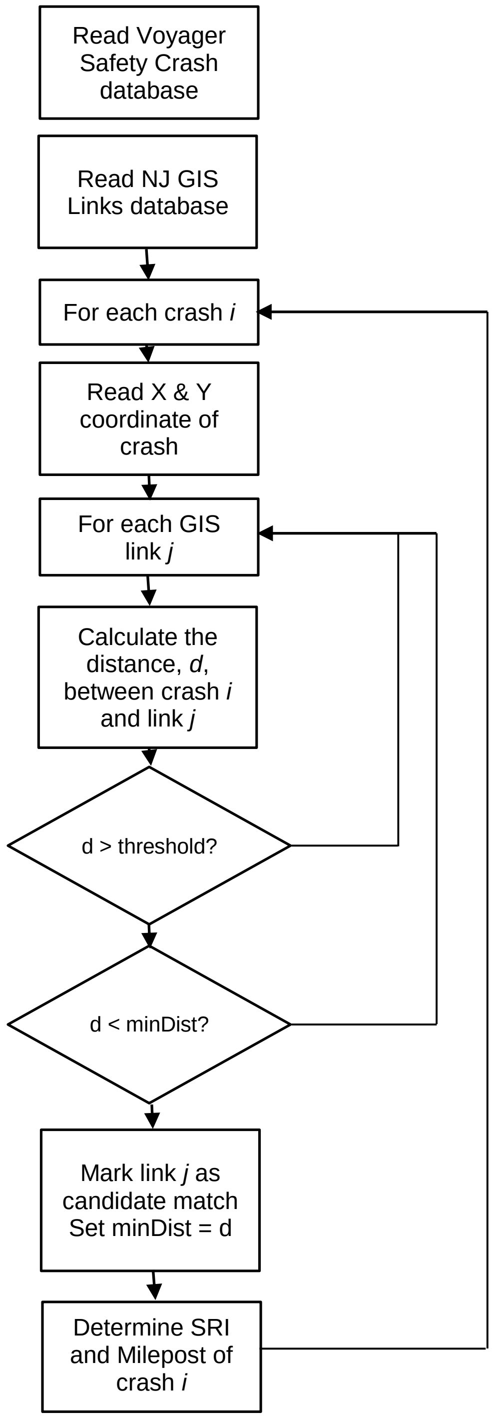


Figure 11: Decision flowchart for determining missing SRI and milepost information

PROCESSING DATA

Generating The Intersection Database

Figure 12 shows the procedure for generating the intersection database per facility type. As shown in Table 5, SLD tables *pt_intersection* and *pt_int_approach* include the location of each intersection, type of intersection (i.e., 3-way or 4-way), whether it is signalized or not, the major and minor roadway SRI and milepost, cross street names, skew angle, and latitude and longitude readings. Using the SRI and milepost of the major approach of the intersection, one can determine the urbanization degree of the intersection using the *ln_urban_code* table, determine its facility type (e.g., arterial or not) using the *ln_f_system* table, and determine the number of lanes using the *ln_lane_count* table.

There are various intersection-related data required by the HSM per facility type, such as the number of left-turn and right-turn lanes, no turn on red, signal phase type (e.g., left-turn protected or not), and presence of lighting. Any information that was not provided in the SLD tables or was missing for a particular intersection was extracted using Google Street View images. This procedure is explained in detail within the relevant sections of this report dealing with intersections.

Once the geometric and operational information for intersections per facility type is gathered, the traffic volume and crash data can be assigned to each intersection using the latitude and longitude readings.

Crash data are assigned to intersections based on the distance between the intersection and the crash location using the Haversine distance formula, shown in equation (11). Based on the HSM's definition, a crash is labelled as intersection-related if it occurs on an intersection approach within 250 ft. of the intersection.

$$d = 2.r.arcsin\left(\sqrt{sin^2\left(\frac{X_1 - X_2}{2}\right) + \cos X_1 . \cos X_2 . sin^2\left(\frac{Y_1 - Y_2}{2}\right)}\right) \tag{11}$$

Where, *d* is distance in kilometers, *r* is the radius of the sphere (6,371 km), and X_1, Y_1 and X_2, Y_2 are the latitude and longitude readings of points 1 and 2.

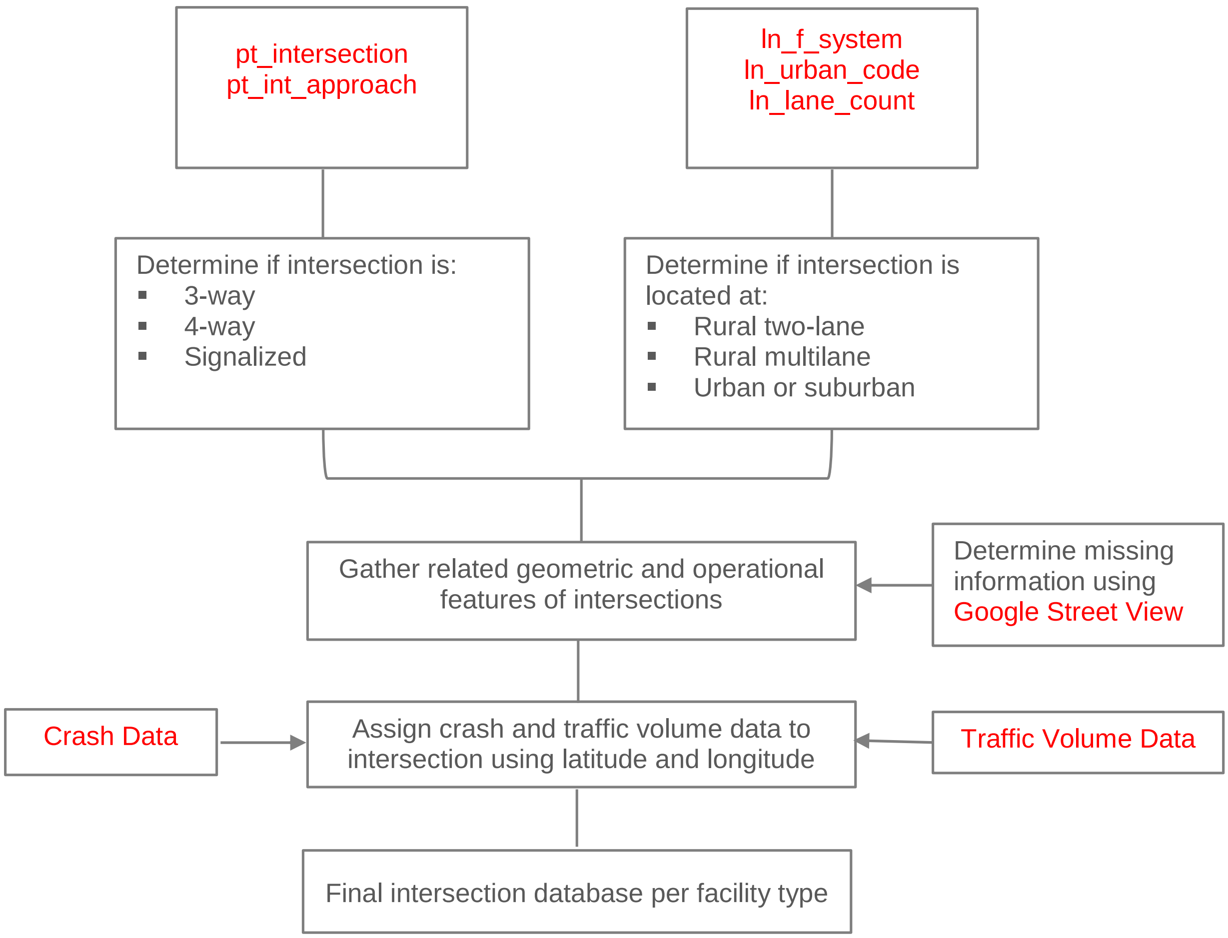


Figure 12: Generating the intersection database for each facility type

It is worth mentioning that traffic volume data are rarely available for each intersection, except when there is TMC data. Therefore, for each intersection, the closest available traffic volume data are determined for major and minor approaches by matching the SRI and milepost values of sensor stations and the intersection.

Generating the Homogeneous Segments Database

Determining roadway segments that have similar geometric and operational features was a more challenging process than that for intersections. A homogeneous segment starts at an intersection or at any point where one of various geometric and operational features of the roadway changes at either direction of the facility. Figure 13 shows the procedure for generating the homogeneous segments database per facility type.

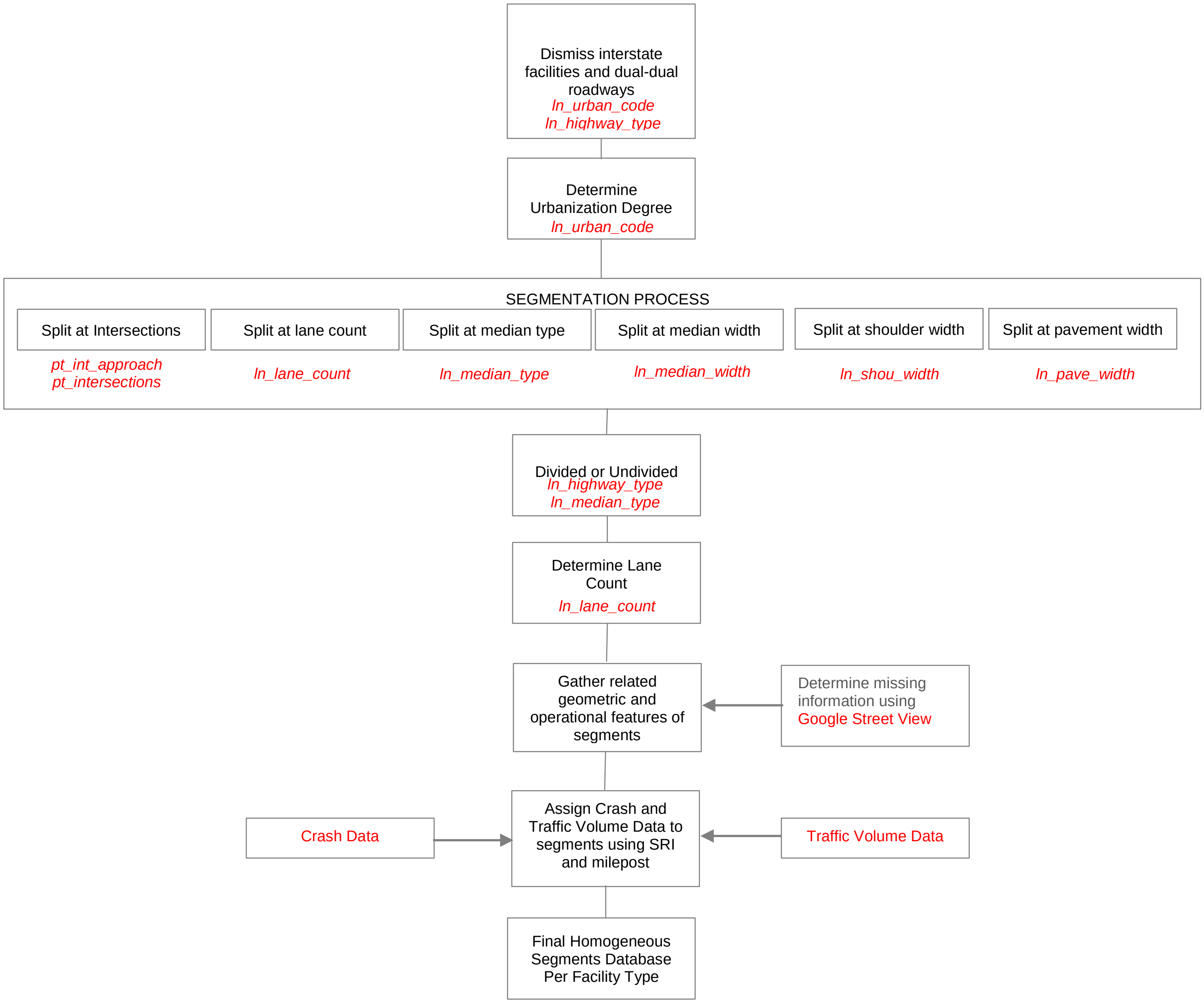


Figure 13: Generating homogeneous segments database for each facility type

As shown in Table 6, the SLD tables include the essential roadway feature data required to generate homogeneous segments per facility type. It should be mentioned that the procedure shown in Figure 13 is generic and varies per facility type, as per the HSM's guidelines. A more focused procedure for each facility type is presented in the relevant sections of this report.

The procedure first eliminates any roadway of interstate functional type and dual-dual roadways using the *ln_f_system* and *ln_highway_type* tables. Whether the roadway is rural or urban is determined using the *ln_urban_code* table. Then the roadways are split at intersections. Intersection locations are obtained from the *pt_intersection* and *pt_int_approach* tables. Next, the roadway segments are further divided where there is a change in lane count, median type, median width, shoulder width, or pavement width, based on the data obtained from the *ln_lane_count*, *ln_median_type*, *ln_median_width*, *ln_shou_width,* and *ln_pave_width* tables, respectively. Later, these segments are labeled as divided or undivided based on the data available in the *ln_median_type* table. It should be mentioned that whether a roadway is divided or not is also specified in the *ln_highway_type* data, as shown in Table 6. However, the SLD labels a roadway as divided if the median type is painted, i.e., flush median. In contrast, the HSM considers painted medians as undivided facilities. Therefore, based on the data in *ln_median_type*, if the median is curbed, positive, or grass, the segment is labeled as divided, and if the median is unprotected or painted, it is labeled as undivided.

There are various segment-related data required by the HSM per facility type, some of which are not available in the SLD tables. These data were extracted using Google Street View images. This procedure is explained in detail within the relevant sections of this report dealing with segments.

Once the homogeneous segments are generated and the relevant geometric and operational information is gathered, the traffic volume and crash data can be assigned to each segment using the SRI and milepost information. In this case, the latitude and longitude information is not used, since it is not straightforward to automatically determine which roadway segment a crash should be assigned to.

RURAL TWO-LANE TWO-WAY SEGMENTS

The HSM's crash frequency predictive model specifies SPFs for the following segment for a rural two-lane two-way road site: undivided rural two-lane two-way roadway segment (R2U). It defines an undivided roadway segment (2U) as a roadway consisting of two lanes with a continuous cross-section providing two directions of travel in which the lanes are not physically separated by either distance or a barrier. In addition, the definition includes a section with three lanes where the center lane is a TWLTL or a section with added lanes in one or both directions of travel to provide increased passing opportunities (for example, passing lanes, climbing lanes, and short four-lane sections).

This section presents a detailed description of the data requirements, data processing, and extraction of additional required data, and of the results of SPF calibration and development.

Data Requirements

The data required for calibration of R2U segment predictive models as specified by the HSM are presented in Table 15.

Table 15: Data requirements for R2U segments

DATA ELEMENT	REQUIRED	DESIRABLE	SOURCE
Segment Length			Segmentation of SLD
AADT			Sensor Database
Horizontal Curve Length			GIS Centerline Data
Horizontal Curve Radius			
Presence of Spiral Transition			
Superelevation Variance for Horizontal			
Percent Grade			
Lane Width			SLD->*ln_pave_width*
Shoulder Type			SLD->*ln_shoulder_type*
Shoulder Width			SLD->*ln_shou_width*
Lighting			Google Street View
Driveway Density			SLD->*pt_intersection*
Presence of Passing Lane			SLD->*ln_passing_zone*
Presence of Short Four-Lane Section			SLD->*ln_lane_count*
Presence of Center Two-Way Left Turn			
Presence of Centerline Rumble Strip			SLD->*rumble_strip*
Roadside Hazard Rating			
Use of Automated Speed Enforcement			

As shown in Table 15, three data elements that are required by the HSM are not readily available in the roadway features dataset. These are horizontal curve length, horizontal curve radius, and presence of center two-way left-turn lane. The last is not deemed problematic, since the presence of center two-way left-turn lanes is rare in NJ. In addition, after the homogeneous segments were determined, the research team perused the dataset and did not detect any center two-lane left-turn lanes. Horizontal curve length and horizontal curve radius, however, are crucial not only in the segmentation process but also in calculating the corresponding CMFs. As mentioned in the Findings subsection of the Literature Review section, among the datasets required by the HSM, the horizontal curvature data stand out as the most problematic. With few exceptions, state DOTs do not have a curvature dataset in a usable format that can be easily integrated into the process of finding homogeneous segments. The lack of this dataset led the researchers to assume default values for horizontal curvature CMFs, omit sections with horizontal curvature, or manually extract these data.

The research team decided to extract the horizontal curvature data for R2 segments using the GIS centerline map of NJ roadways. This approach minimized the amount of manual effort needed and increased the accuracy of data extraction. It was found that horizontal data extraction using Google Earth, as was done in previous studies, is prone to errors, especially in detecting and measuring radii of compound curves. In order to extract this important dataset, the team used a novel clustering-based approach, as explained in detail below.

Gathering and Processing the Roadway Feature Dataset

The data required for R2U segments shown in Table 15 are available through the roadway features, traffic volume, and NJ GIS Links databases described in the Available Data Sources section, with the exception of horizontal curvature data and the presence of center two-way left-turn lane. As mentioned above, the detailed investigation of the final dataset revealed no center two-way left-turn lanes. The subsection that follows describes the identification of horizontal curves using approximated curvature values of data points from the GIS roadway centerline maps.

Extracting Horizontal Curvature Data

The proposed approach is centered on the idea that distinct road sections, either curved or straight (i.e., tangent), have the same curvature values, and that one could detect these distinct sections based on their curvature.

Let us consider the graphical representation of a road by discrete points on a GIS map, as shown in part (a) of Figure 14. Suppose that the study route of length L is discretized by $i = 1\ to\ n$ points, dividing the route into $n - 1$ number of intervals of varying length Δ_i, where $\sum_{i=1}^{n-1} \Delta_i = L$. The forward finite-difference approximation of the first order derivative at any point i can be written as:

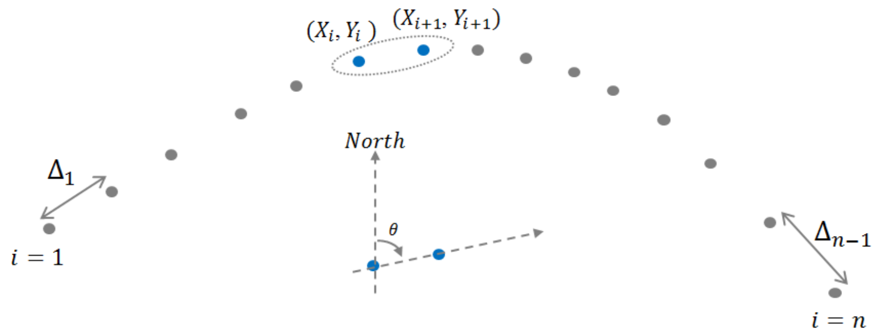
$$f'(i) \approx \frac{Y_{i+1} - Y_i}{X_{i+1} - X_i} \tag{12}$$

Where, X and Y are the latitude and longitude readings of points. Note that equation (12) is analogous to the tangent of the bearing angle θ between points i and $i + 1$, shown in the inset of part (a) of Figure 14. Similarly, the finite-difference approximation of the second order derivative at point i can be written as:

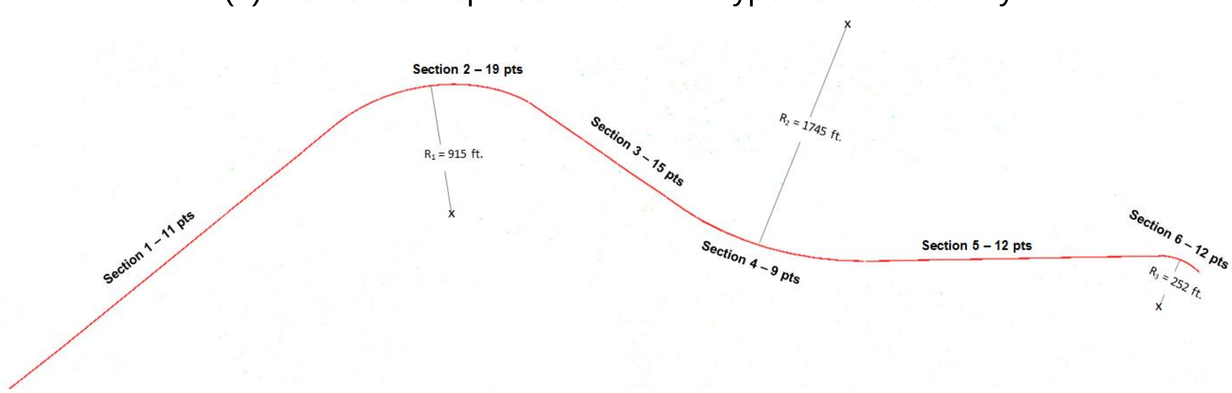
$$f''(i) \approx \frac{f'(i) - f'(i-1)}{X_{i+1} - X_i} \tag{13}$$

Using the approximated first and second order derivatives, curvature of point *i* can be calculated as:

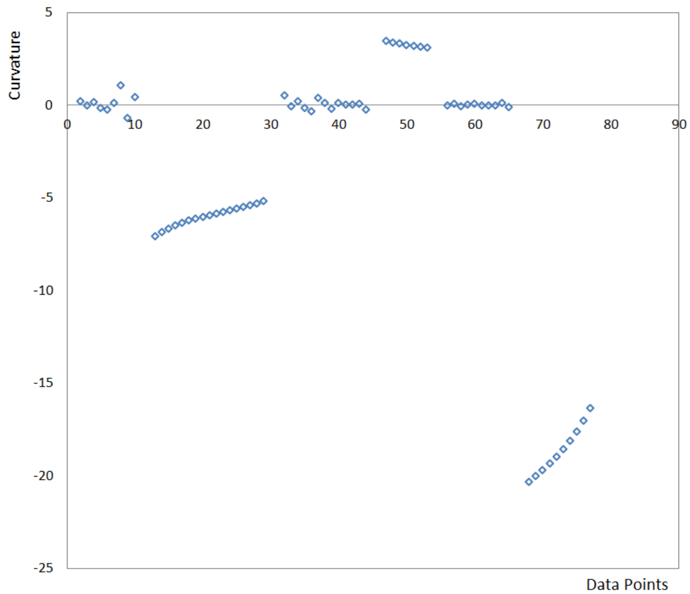
$$\kappa_i \approx \frac{f''(i)}{[1 + f'(i)^2]^{\frac{3}{2}}} \tag{14}$$



(a) Discretized representation of a hypothetical roadway



(b) Curvature information



(c) Calculated curvature value of points along the hypothetical roadway shown in (b)

Figure 14: Hypothetical roadway

Although κ_i are approximations, one can assert that points within a distinct roadway section, curved or tangent, will have similar values. In order to depict this proposition, consider the hypothetical roadway drawn in Google Earth as shown in part (b) of Figure 14. The roadway includes three tangent and three curved sections. Each section is discretized by points automatically in Google Earth. The number of points on each section is also shown in part (b) of Figure 14. The curved sections 2, 4, and 6 are parts of circles with radii 915 ft, 1,745 ft, and 252 ft, respectively.

Curvature κ of each point is calculated using equation (14) by using the latitude and longitude readings extracted from Google Earth, and plotted in part (c) of Figure 14. As seen in the figure, the curvatures of points within distinct sections are clustered closely. This plot clearly demonstrates that the approximated curvature values of discrete data points can be used to detect horizontal curves. For example, as expected, the points within tangent sections 1, 3, and 5 have curvature values clustered around zero; whereas, the points within the curved sections have curvature values distinctly different from zero.

Note that the calculated curvature values of individual points along a perfectly drawn circle vary, as shown in part (c) of Figure 14. This is due to the representation of a continuous line with discrete points in space. However, as mentioned earlier, one would expect that κ of points on the same curve would be close to each other.

The objective of accurately determining the sections that have similar values is similar to that of the common clustering algorithm, which is based on minimum within-group distance criterion [78].

Once the distinct sections are identified by the clustering, the radius R of each horizontal curve can then be calculated using the Chord Method [79], as follows:

$$R = \frac{M^2 + (LC^2/4)}{2M} \tag{15}$$

Where, LC is the length of the chord between PC and PT, and M is the middle ordinate, i.e., the distance between the middle point of the chord and the centerline of the curve. The chord method eliminates the need for determining the deflection angle. Note that the start point of each cluster (i.e., distinct road section) is the PC, and the end point is the PT. Knowing these points, the variables required to estimate R can be easily calculated.

In order to apply this approach, the proposed method first processes the latitude and longitude readings of the vertices of roadway centerlines from the NJ GIS Links database, and calculates the finite-difference approximation of curvature at each data point, as shown in Equation (14). These curvature data are then analyzed using the modified global *K*-means clustering algorithm, developed in C programming language. The proposition of this method is that curvature of points within distinct sections, i.e., curved or tangent sections, are clustered closely, as shown in part (c) of Figure 14. Therefore, the approximated curvature data of discrete GIS data points can be used to identify horizontal curves. The details of this approach can be found in Bartin et al. [78], in which the validity of the clustering method is examined with respect to two other methods, namely the mobile access vehicle method and manual horizontal curvature extraction using satellite images. It was shown that the rate of correct identification of curves by the clustering method is 95 percent on average. The undetected curves were found to be slight and short horizontal curves, which are usually not well described by discrete data points, especially in low-resolution GIS maps.

Automatic Identification of Segments

Using the method for extraction of horizontal curvature data described above, the research team was able to assign horizontal curvature radius and length values to all roadway segments in the SLD database.

Homogeneous segments were generated by following the process described in the Processing Data section. The flowchart for this process is shown in Figure 15. A generic flowchart for automatic identification of homogeneous segments is in Figure 13. R2 segments are determined by filtering based on urbanization degree, median type, and lane count, and by splitting when there are intersections or horizontal curvature within segments.

Once homogeneous segments are identified, using the traffic volume dataset, distance between each segment and the closest traffic sensor is calculated. Table 16 presents the summary statistics on R2 homogeneous segments. As shown, there were a total of 13,886 homogenous R2 segments found. However, the HSM suggests using segments of 0.1 mile or longer for calibration and development purposes. The team found 5,847 R2 segments with lengths longer than 0.1 mile, 756 of which include an AADT station within the segment.

The research team used the 756 segments for analysis, due to the quality of AADT information expected from the stations located within segments.

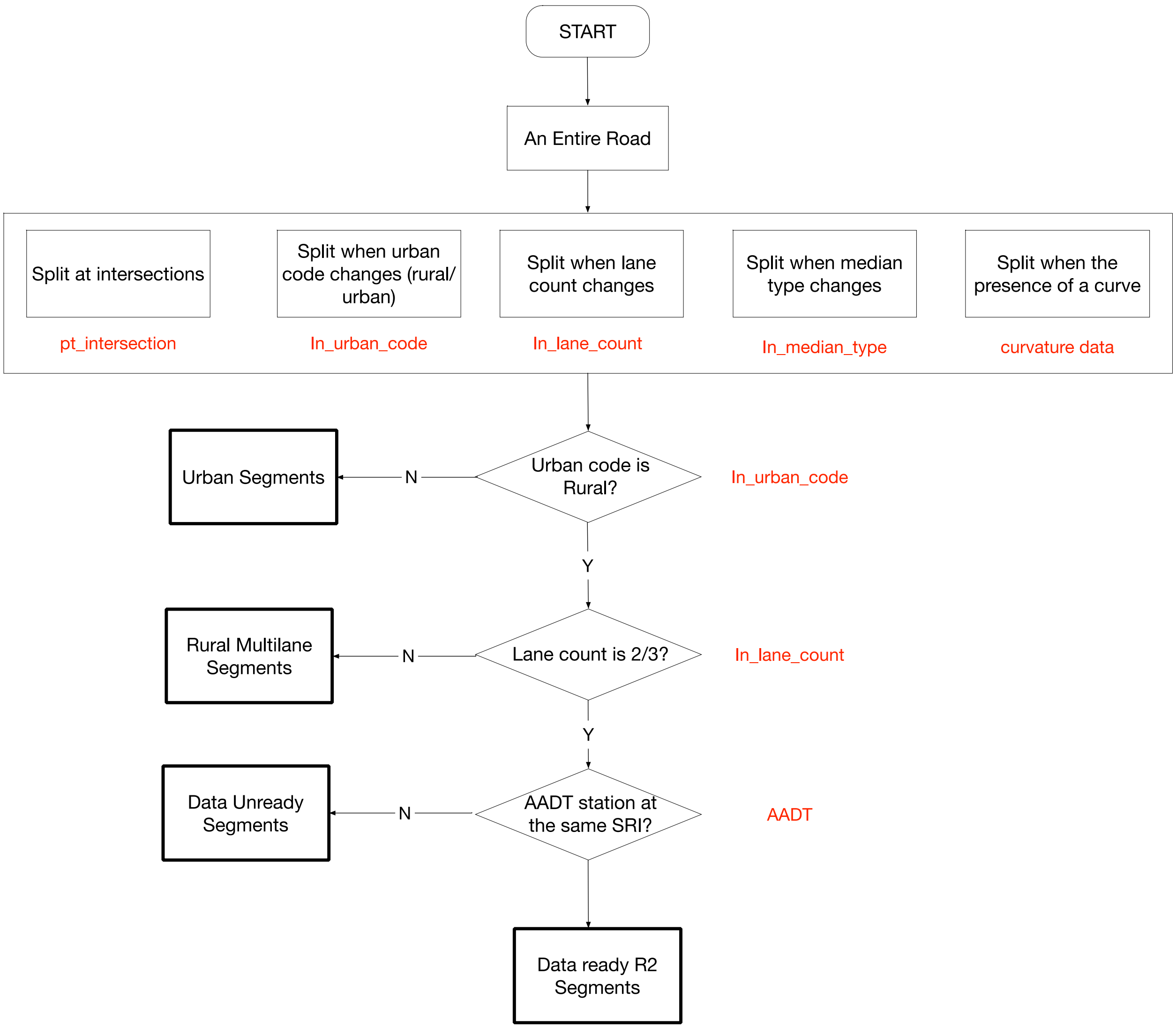


Figure 15: Flowchart for automatic detection of R2U

Table 16: R2U segments statistics

R2U	Number of R2Us	Average Distance between AADT Station and the R2U	Average Number of Intersections between AADT Station and the R2U
Total	13,886	1.03	0.78
> 0.1 mile	5,847	0.97	1.42
AADT station within	756	0.13	0

Calibration Results

Chapter 10 of the HSM provides a crash prediction model for R2 segments. The base SPF includes the AADT and segment length (*L*) as covariates:

$$N_{spf\ R2U} = \text{AADT} \times L \times 365 \times 10^{-6} \times e^{(-0.312)} \quad (16)$$

The reliable AADT range is given as [0 – 17,800 veh/day] by the HSM.

The SPF shown in equation (16) calculates the predicted crash frequency for the base conditions. The base conditions for R2 segments are 12-ft lanes, 6-ft paved shoulder width with no horizontal curvature, and center two-way left-turn lane. If the attributes of an intersection are different from the base conditions, CMFs are applied. Detailed information on CMFs for R2 segments is presented in Chapter 10 of the HSM.

Model calibration in the HSM is performed by applying a multiplicative factor to the given SPF so that the aggregate number of predicted crashes is equal to the aggregate number of observed crashes in a jurisdiction. A calibration factor allows the SPF to keep its original model form. As discussed in the appendix to Part C of the HSM, selected samples are used to find the calibration factor that will make the aggregate predicted crash frequency equal to the observed total in the jurisdiction. The HSM recommends using a minimum of 30 to 50 sites that are selected without regard to their crash frequencies [(2)].

In order to calculate a calibration factor, the observed crash frequency and the predicted crash frequency for each intersection are required. The observed crash frequency was calculated for each segment as explained in the Processing Data section. The predicted crash frequency can be calculated using the SPF and the corresponding CMF values given in the HSM.

CMFs for shoulder width, shoulder type, lane width, and curve radius and length were calculated and multiplied by the predicted crash frequency for the base conditions to calculate the predicted crash frequency. Then calibration factors were calculated using equation (2).

The Calibrator tool developed by the FHWA was used to calculate the calibration factor and measure its goodness of fit [77]. The calibration factor for R2U segments was calculated as 1.55 with a standard error of 0.12 and a coefficient of variation of 0.08. According to the FHWA report, a reasonable upper threshold for the coefficient of variation of a calibration factor is 0.10 to 0.15. Thus, the calibration factor can be considered acceptable.

Figure 16 presents the CURE plots with respect to AADT and segment length. The results show that the cumulative residuals with respect to AADT stray from the region defined by the lower and upper bounds, which suggests that using the SPF from the HSM with the calibration factor is not a desired fit to the data at hand. Therefore, development of an SPF is warranted for R2 segments.

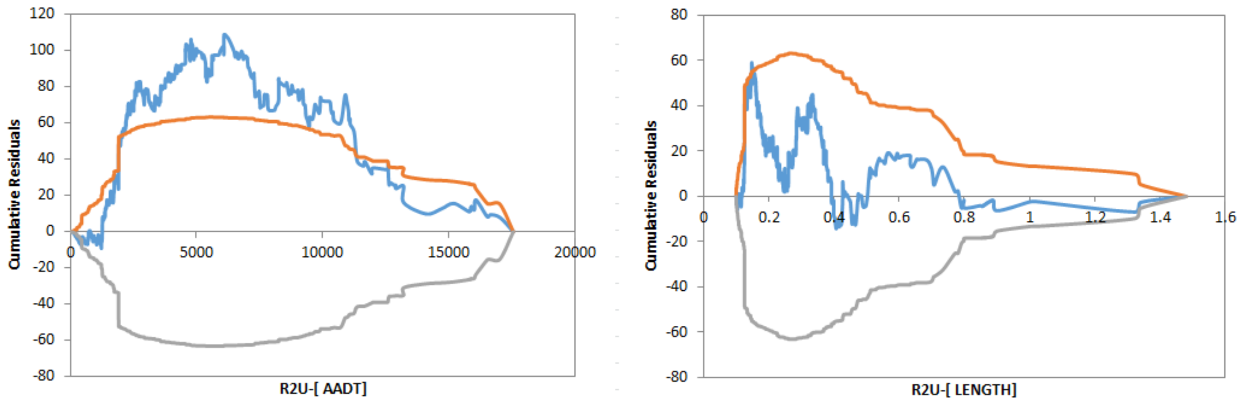


Figure 16: CURE plot of R2U segments with respect to AADT and segment length

Development Results

The SPF for R2 segments was developed using the available datasets, with 756 homogeneous segments, based on the negative binomial model suggested by the HSM. The model estimation was performed in R statistical package. The developed function is shown in equation (17), and the statistical results are shown in Table 17.

$$N_{spf\ R2U} = e^{-6.405}\mathrm{AADT}^{0.827}L^{0.8644} \tag{17}$$

Table 17: Development results of SPF for R2 segments

Variable	Estimate	Std	z-	Pr(> |z|)
Intercept	-6.40549	0.30042	-21.32	< 2e-16***
Log(AADT)	0.82670	0.03550	23.59	< 2e-16***
Log(Length)	0.8644	0.05924	14.59	< 2e-16***
Overdispersion parameter, k	0.664			

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.
3,394 dof, AIC: 4740.9

The SPF for roadway segments on rural two-lane highways are applicable to the AADT range from zero to 27,000 vehicles per day.

In order to compare the developed SPF with the calibrated SPF from the HSM, the research team analyzed the comparison of predicted versus observed crash frequencies. Figure 17 shows the accuracy performance of the developed SPF, where the red bars indicate the observed frequency of crashes and the black outlined bars indicate the crash frequency predicted by the NJ-specific SPF. The results suggest that the developed SPF yields crash frequency predictions very close to the observed crash frequencies.

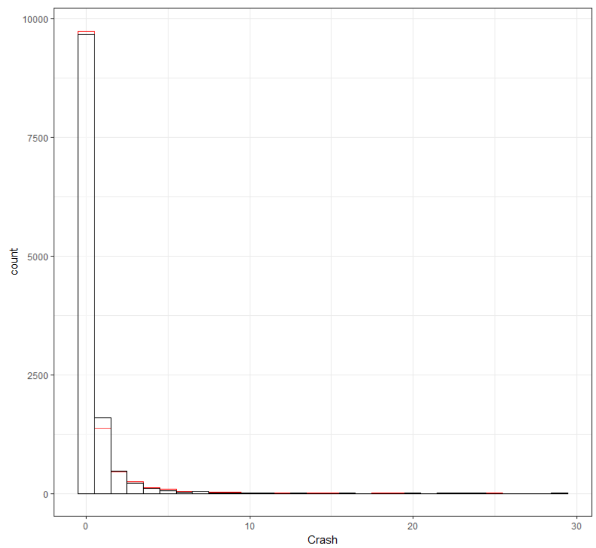


Figure 17: Observed vs. predicted crashes – NJ-specific SPF for R2 segments

Figure 18 shows the accuracy performance of the calibrated HSM SPF. It can be observed that the calibrated SPF prediction is significantly different from the observed crash frequencies, especially at zero crashes.

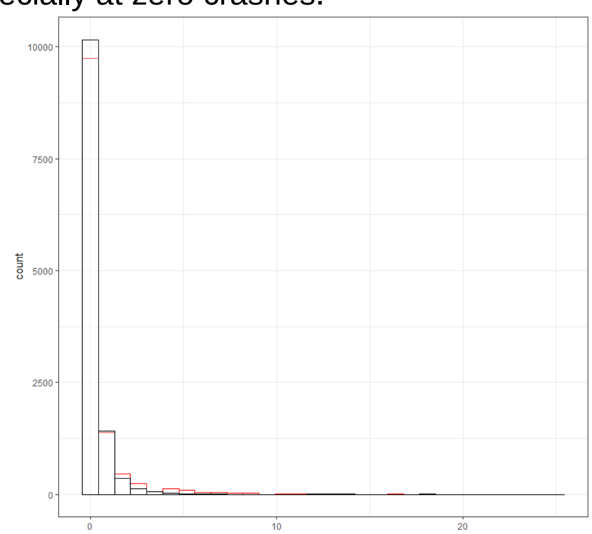


Figure 18: Observed vs. predicted crashes – calibrated SPF for R2 segments

Based on the results of the comparison, the use of the developed SPF for R2 segments is recommended for NJDOT.

RURAL TWO-LANE TWO-WAY INTERSECTIONS

The HSM's crash frequency predictive model specifies SPFs for three types of intersections for rural two-lane two-way rural roads (R2)[(2)] :

(1) Three-leg stop-controlled intersections (R23ST)
(2) Four-leg stop-controlled intersections (R24ST)
(3) Four-leg signalized intersections (R24SG)

This section presents a detailed description of the data requirements, data processing, and extraction of additional required data, and the results of SPF calibration and development.

Data Requirements

The required data for calibration of R2 intersection SPFs as specified by the HSM are presented in Table 18.

Table 18: Data requirements for R2 intersections

DATA ELEMENT	REQUIRED	DESIRABLE	SOURCE
Number of Intersection Legs			Intersection Database
Type of Traffic Control			Intersection Database
AADT for Major Road			Sensor Database
AADT for Minor Road			Sensor Database
Intersection Skew Angle			Intersection Database
Number of Approaches with Left Turn			Google Street View
Number of Approaches with Right Turn			Google Street View
Lighting			Google Street View

As shown in Table 18, the required dataset includes AADT for both major and minor roads, lighting, and intersection left- and right-turn lanes; the desired dataset includes intersection skew angle only. The base conditions that apply to the SPFs in the HSM are zero skew angle, no intersection left-turn or right-turn lanes, and no lighting present.

Gathering and Processing the Roadway Feature Dataset

The roadway feature dataset was used to identify the type of intersection and extract the information necessary for calculating CMFs. However, as mentioned earlier, as with any large database, there were observed errors in the SLD database. The most common types of such errors regarding R2 intersections were: (1) some overpasses were identified as intersections and (2) data such as lighting information was sometimes missing. In addition, the SLD database does not include the number of left- and right-turn lanes at an intersection, which are required variables for the calibration of R2 intersection SPFs. Therefore, the roadway feature data was used primarily to identify the type of intersection. Consequently, the research team conducted a manual data extraction process to verify the information in the SLD database and extract the missing variables using Google Earth.

Automatic Identification of Intersection Types

A flowchart of the process for automatic identification of R2 intersection types is provided in Figure 19.

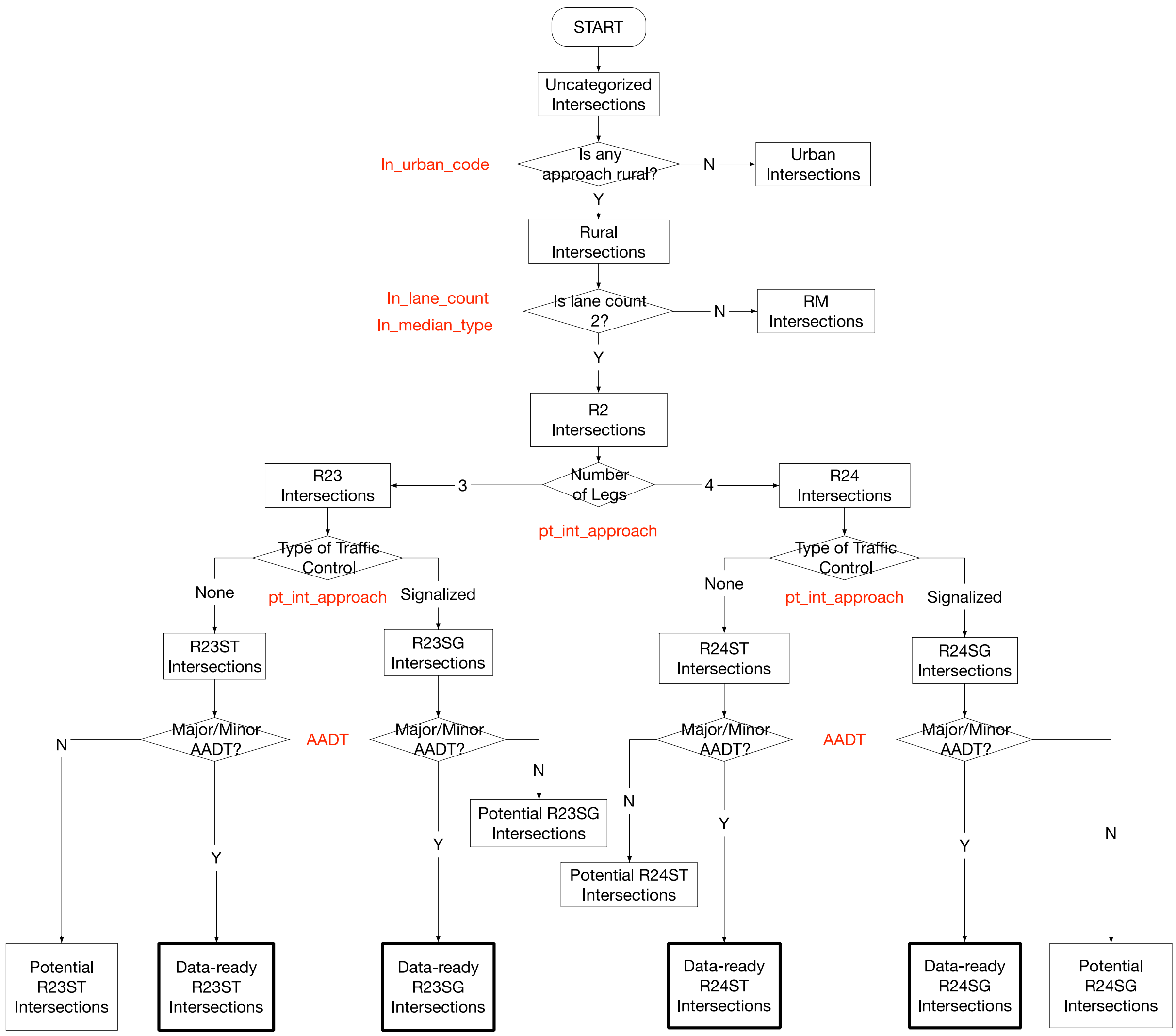


Figure 19: Flowchart for identifying the type of data-ready R2 intersections

Four roadway features are needed to automatically identify the type of intersection (i.e., R23ST, R23SG, R24ST, and R24SG). These are: (1) number of intersection legs, (2) urban/rural attribute, (3) lane count, and (4) type of traffic control. These features are available in the available roadway features and intersection database, as shown in Table 19. It should be mentioned that this step is only preliminary and that a manual process is required to finalize the identification of intersection types, as described in the next subsection.

Urbanization for an SRI and milepost can be determined using the *In_urban_code* table. As shown in Table 5, this table provides the SRI and start and end milepost, and the urban/rural feature, between the starting and end points of each roadway link. The field “is_urban” = Y indicates an urban area, while “is_urban” = N indicates a rural area.

Table 19: Data elements used to identify intersection attributes

Data Element	**Level**	**Current Data Source**
Urban Code	Approach	*ln_urban_code*
Lane Count	Approach	*ln_lane_count* *In_median_type*
Number of Intersection Legs	Intersection	*pt_int_approach* *pt_intersection* NJ GIS Links database
Type of Traffic Control	Intersection	*pt_int_approach* *pt_sign*

Lane count information is found in the table *In_lane_count*. Similar to the urban/rural attribute, this table also consists of SRI, starting and end mileposts, and the lane count of the road segment between the starting and ending points. The field “descr” indicates how many lanes the intersection has. However, the lane count recorded in *In_lane_count* depends on whether the road is divided. If the road is divided, the “descr” field only records the lane count of one direction, while if it is undivided, the “descr” field records the total lane count for both directions. Thus, to identify the total lane number of an approach, the table *In_median_type* is also needed. The structure of the median type dataset is also similar to the urban/rural attribute, and the start and end mileposts are used to mark the road segments with the same median type.

Number of intersection legs and type of traffic control is available in the intersection database that was generated by the research team using the *pt_intersection* and *pt_int_approach* tables, as explained in the Data Preparation and Cleaning section.

For each approach of the intersection, the nearest AADT station was found, as explained in the Processing Data section. Only the intersections for which AADT stations exist on both pairs of approaches were selected as final candidates. $AADT_{maj}$ was defined as the pair of approaches with larger AADT, and $AADT_{min}$ was defined as the pair of approaches with smaller AADT. When the nearest AADT station was found, the distance to the intersection and the number of intersections between the AADT station and the intersection were also calculated.

<u>Manual Data Extraction and Validation</u>

Google Earth was utilized to extract additional data and check the validity of the data obtained from RF2. The steps of manual data collection and validation are as follows:

(1) Import the results of automatic identification of intersection types into Google Earth, and label the intersection using “Int_Id” (see Figure 20).

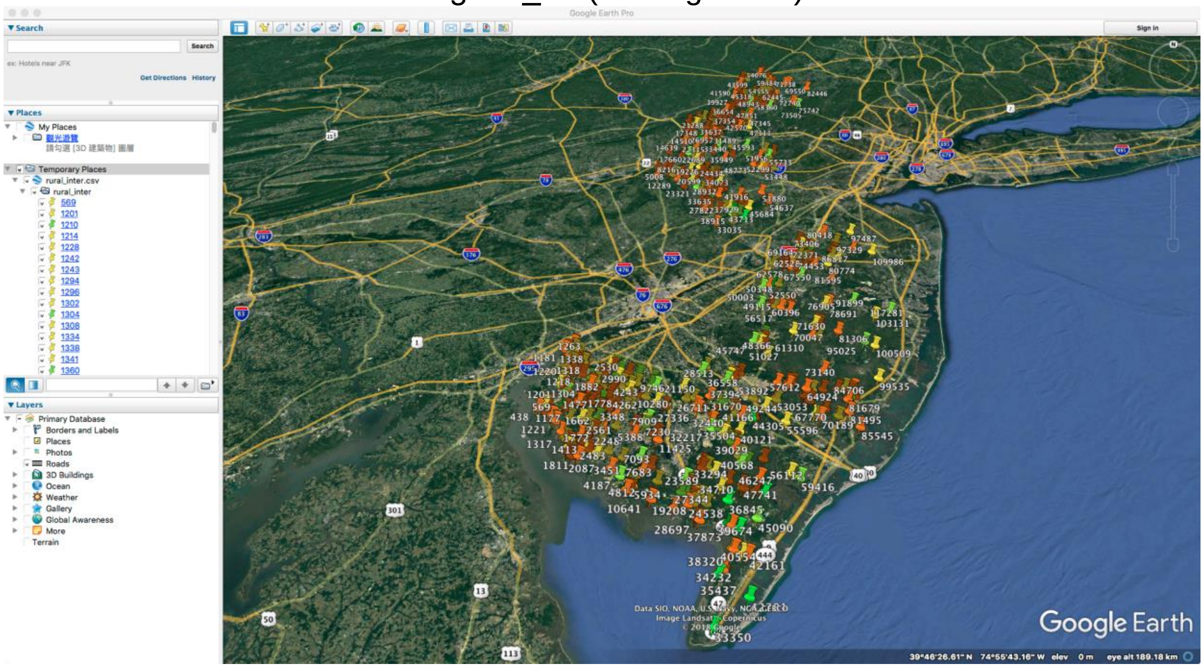


Figure 20: Importing and visualizing the R2 intersections

(2) Select one of the intersections from the intersection list (see Figure 21).

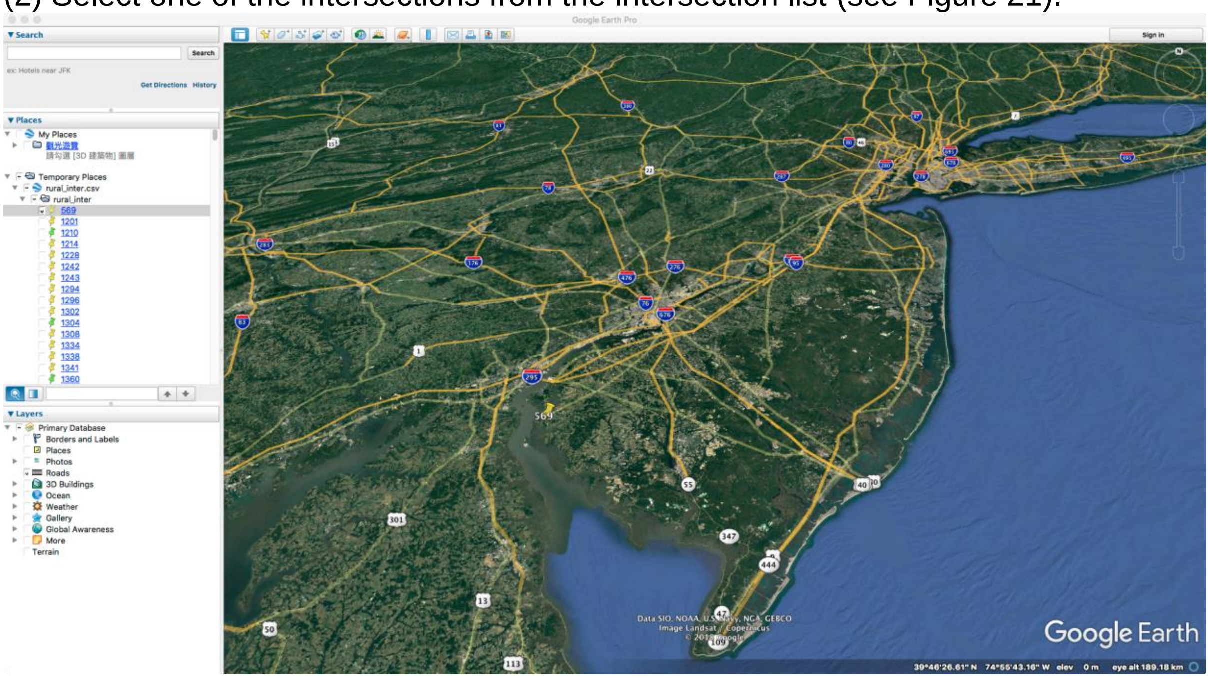


Figure 21: Selecting an intersection from the list

(3) Go to the top view of the selected intersection (see Figure 22):

a. Check whether it is an overpass
b. Validate the number of lanes

c. Validate the number of legs
d. Extract the number of left-turn lanes and number of right-turn lanes
e. Confirm the major direction and minor direction

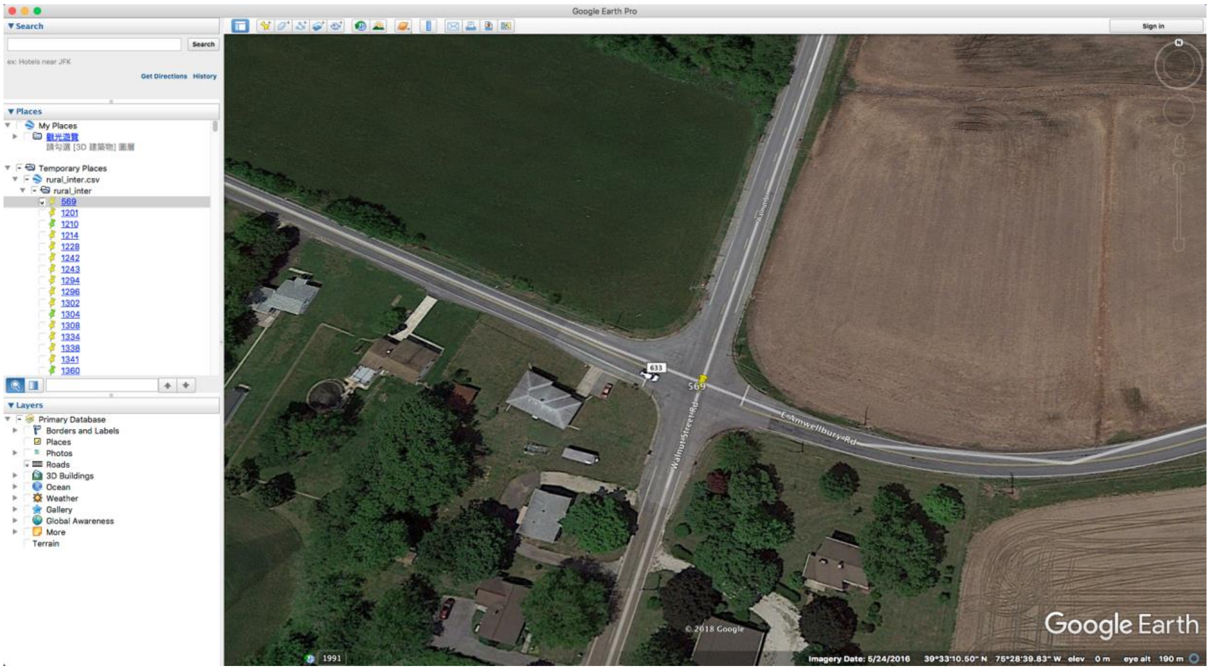


Figure 22: Top view of the selected intersection

(4) Go to the Street View of the selected intersection (see Figure 23):
a. Check whether there is a Street View
b. Validate the signalized attribute
c. Extract whether there is a light
d. Extract whether it is minor stop-controlled

Figure 23: Street view of selected intersection

Based on the validated data, decide the type of intersection and label the intersection with the new intersection type, if appropriate.

Intersection Crash Frequency and Traffic Counts

As mentioned in the Processing Data section, a threshold distance was selected to determine whether to assign a crash as intersection-related. As recommended in the HSM, 250 feet was set as the distance threshold: all crashes that occurred within 250 feet of the identified intersections were counted. Because the coordinates of both crashes and intersections are in longitude and latitude, we used the Great Circle (WGS84 ellipsoid) distance, as shown in equation (11). In addition, for each intersection approach, the closest sensors were found using the latitude and longitude values for sensor stations and intersections.

R2 Intersection Dataset

Following the data processing and extraction process, a total of 523 R2 intersections were identified, as shown in Table 20. Note that in addition to the three intersection types in the HSM, Table 20 also includes rural three-leg signalized intersections, R23SG. These intersections are summarized by county in Table 21.

Table 20: Sample size of preliminary selection and final selection

Type	R23ST	R23SG	R24ST	R24SG
Preliminary Sample Size	422	21	220	94
Final Sample Size	314	15	149	45

Table 21: Sample size by county of preliminary selection and final selection

County	Preliminary Sample Size	Final Sample Size
Atlantic	66	40
Bergen	0	0
Burlington	73	61
Camden	7	3
Cape May	23	19
Cumberland	114	83
Essex	0	0
Gloucester	25	17
Hudson	0	0
Hunterdon	71	45
Mercer	12	9
Middlesex	1	1
Monmouth	25	14
Morris	5	4
Ocean	9	8
Passaic	0	0
Salem	112	77
Somerset	10	6
Sussex	101	74
Union	0	0
Warren	103	62

The quality of AADT data for major and minor approach at an intersection depends on the distance between the AADT station and the target intersection, and on the number of other intersections between the AADT station and the target intersection. The average intersection number between major/minor AADT station and the target intersection, along with the average distance between major/minor AADT station and the target intersection, are summarized in Table 22.

Table 22: Summary of intersection AADT data

Type	Average Major AADT	Average Minor AADT	Average Number of Intersections between Major Station and Target Intersection	Average Number of Intersections between Minor Station and Target Intersection	Average Distance between Major Station and Target Intersection (miles)	Average Distance between Minor Station and Target Intersection (miles)
R23ST	4,703	1,109	1.17	1.30	0.75	1.22
R23SG	13,720	5,414	0.86	1.27	0.48	0.63
R24ST	4,453	958	1.00	1.13	0.68	1.09
R24SG	10,969	3,594	1.68	1.27	0.60	0.61

Detailed Description of Sampling Results

Three-Leg Stop-Controlled Intersections (R23ST): The final sample set contains 314 R23ST intersections (see Figure 24). For those intersections, the average AADT on major roads is 4,703; the average AADT on minor roads is 1,109; the average intersection count between AADT station and target intersection on major roads is 1.17; the average intersection count between AADT station and target intersection on minor roads is 1.30; the average distance between AADT station and target intersection on major roads is 0.75 miles; and the average distance between AADT station and target intersection on minor roads is 1.22 miles.

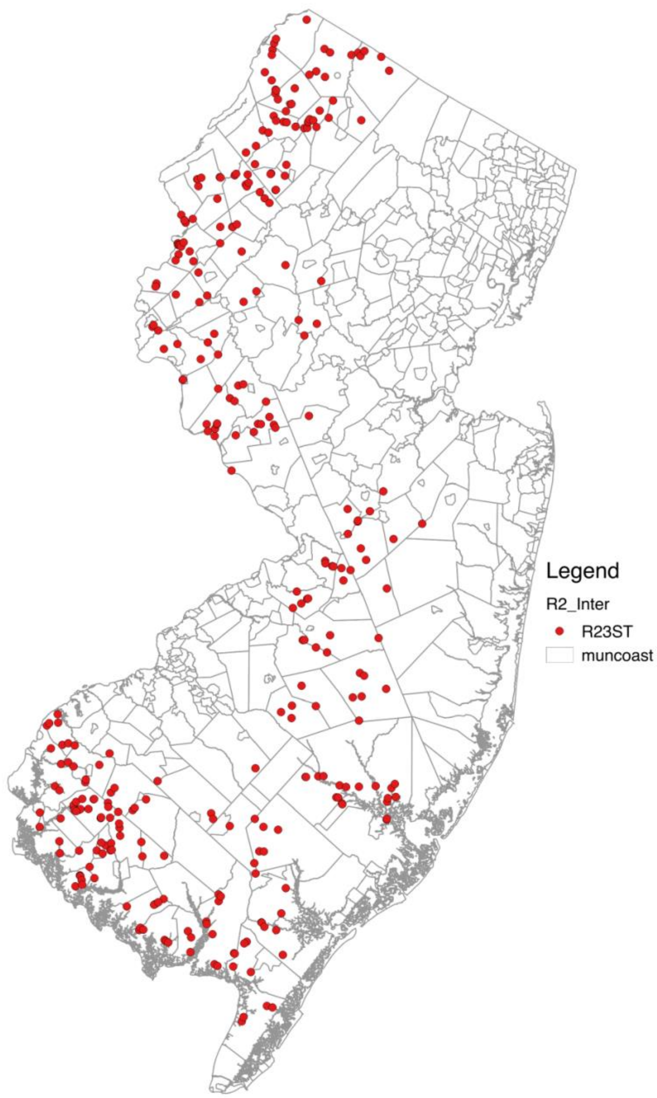


Figure 24: Spatial distribution of R23ST intersections

Three-Leg Signalized Intersections (R23SG): The final sample set contains 15 R23SG intersections (see Figure 25). For those intersections, the average AADT on major road is 13,720; the average AADT on minor road is 5,414; the average intersection count between AADT station and target intersection on major road is 0.86; the average intersection count between AADT station and target intersection on minor road is 1.27; the average distance between AADT station and target intersection on major road is 0.48 miles; and the average distance between AADT station and target intersection on minor road is 0.63 miles.

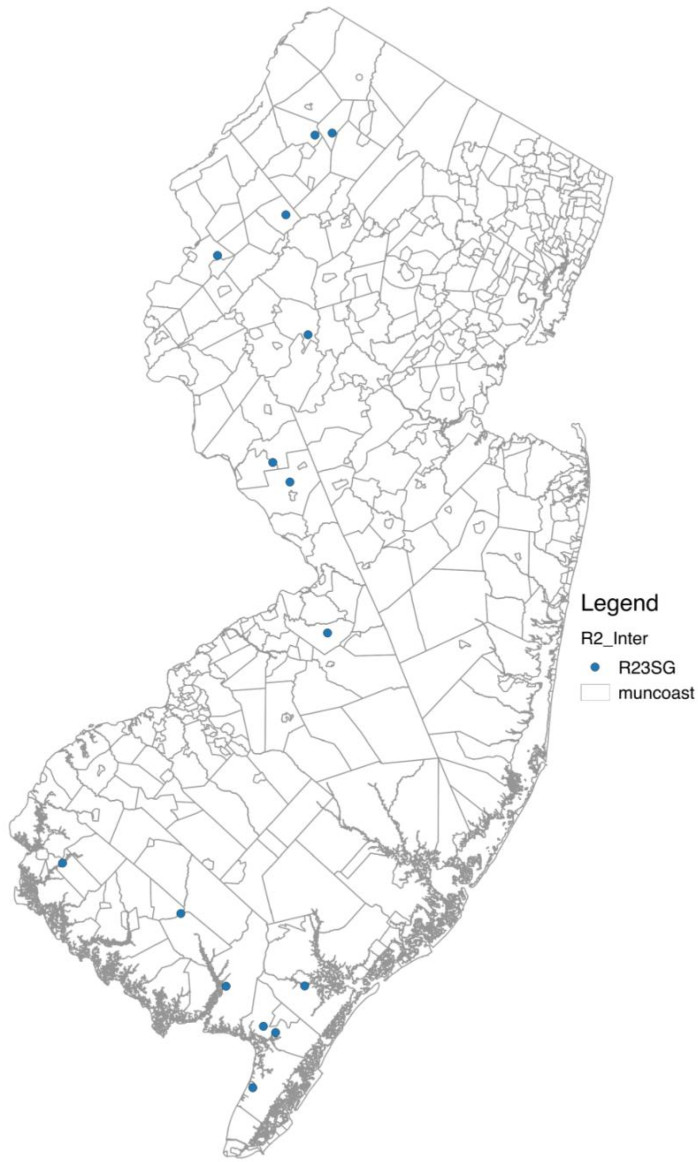


Figure 25: Spatial distribution of R23SG intersections

<u>Four-leg Stop-Controlled Intersections (R24ST)</u>: The final sample set contains 149 R24ST intersections (see Figure 27). For those intersections, the average AADT on major road is 4,453; the average AADT on minor road is 958; the average intersection count between AADT station and target intersection on major road is 1.00; the average intersection count between AADT station and target intersection on minor road is 1.13;

the average distance between AADT station and target intersection on major road is 0.68 miles; and the average distance between AADT station and target intersection on minor road is 1.09 miles.

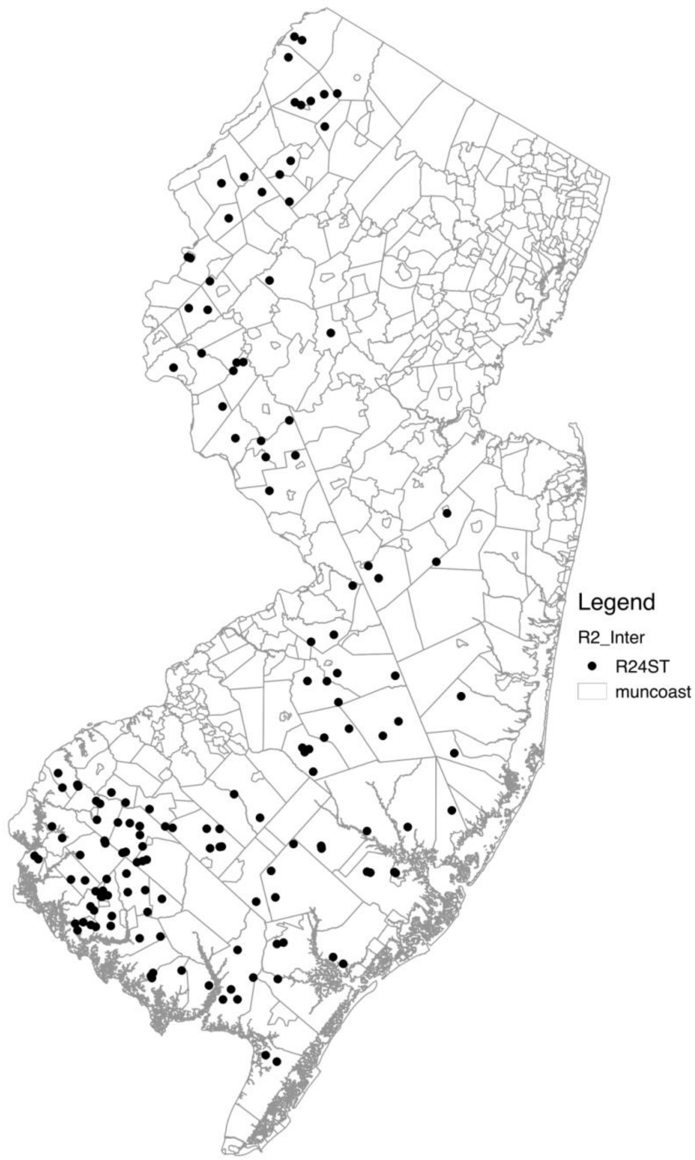


Figure 26: Spatial distribution of R24ST intersections

Four-leg Signalized Intersections (R24SG): The final sample set contains 45 R24SG intersections (see Figure 27). For those intersections, the average AADT on major road is 10,969; the average AADT on minor road is 3,594; the average intersection count between AADT station and target intersection on major road is 1.68; the average intersection count between AADT station and target intersection on minor road is 1.27; the average distance between AADT station and target intersection on major road is 0.60 miles; and the average distance between AADT station and target intersection on minor road is 0.61 miles.

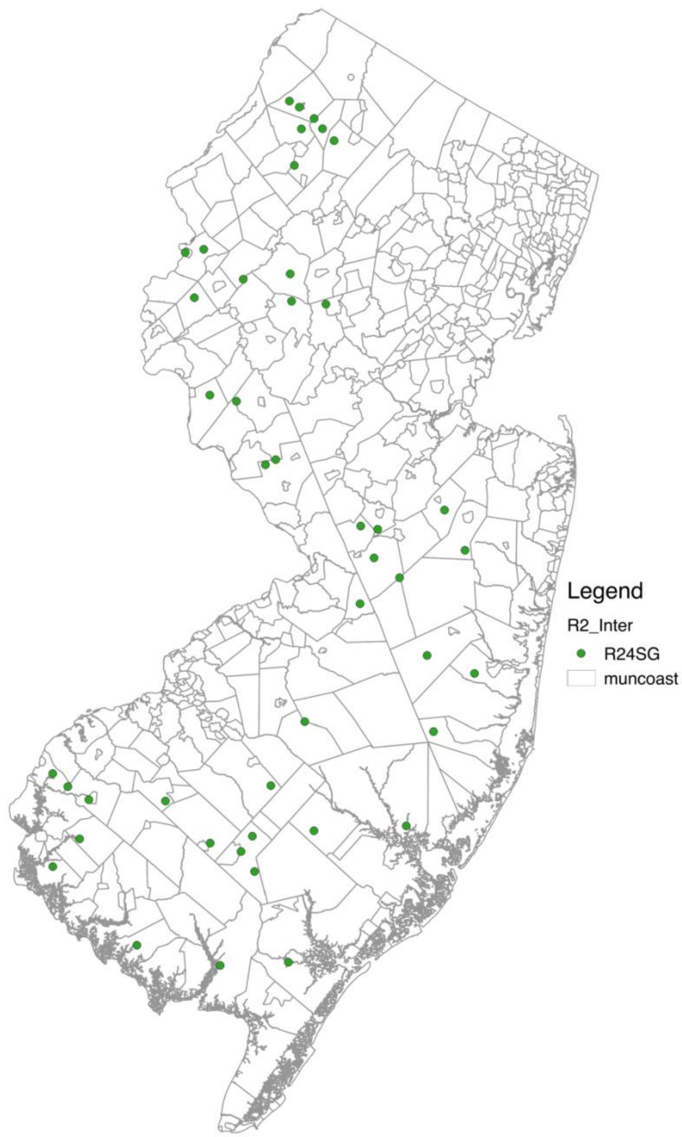


Figure 27: Spatial distribution of R24SG intersections

Calibration Results

Chapter 10 of the HSM provides a crash prediction model for R2 intersections. The base SPFs for R23ST, R24ST, and R24SG include the $AADT_{maj}$ and $AADT_{min}$ variables. The HSM also specifies the reliable AADT range for an intersection and cautions that the application of SPFs to sites with AADTs substantially outside this range may not provide reliable results (see Table 23).

Table 23: SPFs of three types of R2 intersections in the HSM

Type	SPF	Reliable AADT Range
R23ST	$N_{spf\ 3ST} = exp[-9.86 + 0.79. ln(AADT_{maj}) + 0.49. ln(AADT_{min})]$	$AADT_{maj} \in [0,19500]$ $AADT_{min} \in [0,4300]$
R24ST	$N_{spf\ 4ST} = exp[-8.56 + 0.60. ln(AADT_{maj}) + 0.61. ln(AADT_{min})]$	$AADT_{maj} \in [0,14700]$ $AADT_{min} \in [0,3500]$
R24SG	$N_{spf\ 4SG} = exp[-5.13 + 0.60. ln(AADT_{maj}) + 0.20. ln(AADT_{min})]$	$AADT_{maj} \in [0,25200]$ $AADT_{min} \in [0,12500]$

The SPFs shown in Table 23 calculate the predicted crash frequency for the base conditions. The base conditions for R2 intersections are zero skew angle, no right- or left-turn lanes, and no lighting. If the attributes of an intersection are different from the base conditions, CMFs are applied. For R2 intersections, four CMFs are used – intersection skew angle, intersection left-turn lanes, intersection right-turn lanes, and lighting. Detailed information on CMFs for R2 intersections is presented in Chapter 10 of the HSM.

Model calibration in the HSM is performed by applying a multiplicative factor to the given SPF so that the aggregate number of predicted crashes is equal to the aggregate number of observed crashes in a jurisdiction. A calibration factor allows the SPF to keep its original model form. As discussed in the appendix to Part C of the HSM, selected samples are used to find the calibration factor that will make the aggregate predicted crash frequency equal to the observed total crashes in the jurisdiction. The HSM recommends using a minimum of 30 to 50 sites that are selected without regard to their crash frequencies [(2)].

In order to calculate a calibration factor, the observed crash frequency and the predicted crash frequency for each intersection are required. The observed crash frequency was calculated for each intersection as explained in the Processing Data section. The predicted crash frequency can be calculated using the SPF and the corresponding CMF values given in the HSM.

CMFs for intersection skew angle, number of approaches with left-turn lanes, number of approaches with right-turn lanes, and presence of lighting were calculated and multiplied by the predicted crash frequency for the base conditions to calculate the predicted crash frequency. Then calibration factors were calculated using equation (2).

The results of calibration factors are shown in Table 24**.** The Calibrator tool developed by the FHWA was used to calculate the calibration factors and measure their goodness of fit [77]. According to the FHWA report, a reasonable upper threshold for the coefficient of variation of a calibration factor is 0.10 to 0.15. In this respect, the results shown in Table 24 are found to be acceptable.

Table 24: Calibration factors of R2 intersections

Distance	Calibration Factor	Standard Error	Coefficient of Variation
R23ST	0.88	±0.08	0.09
R24ST	0.88	±0.11	0.13
R24SG	0.85	±0.16	0.18

Another approach to assessing the validity of the calculated calibration factor is the CURE plot, which is simply the graph of the cumulative residuals (observed minus predicted crashes) against a variable of interest. The residuals between the estimated and observed values are assumed as independent random variables. CURE plots are based on a loose version of the central limit theorem. It is expected that the CURE plots should be within the expected limits of an unbiased random walk, i.e., plus/minus two standard deviations. The CURE plots of R23ST, R25ST, and R24SG intersection types with respect to major and minor AADTs are shown in Figure 28, Figure 29, and Figure 30, respectively.

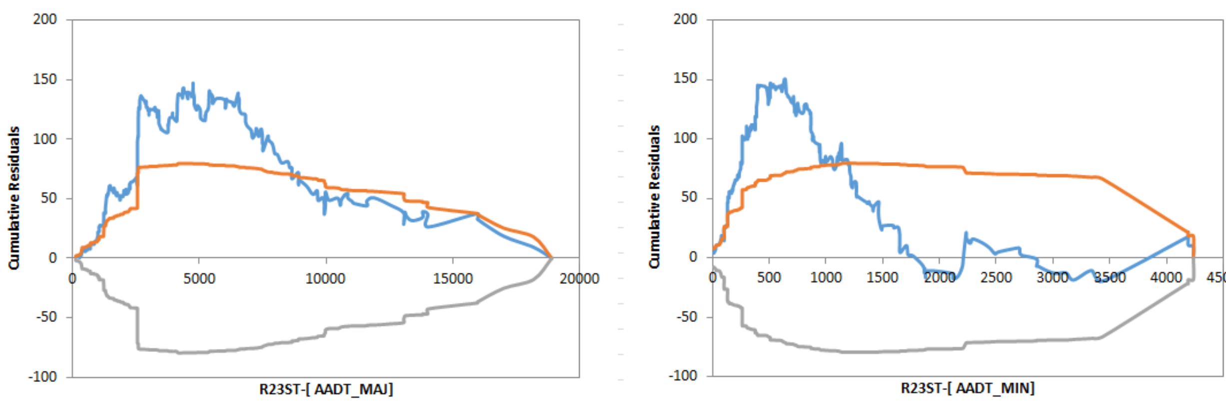


Figure 28: CURE plot of R23ST with respect to $AADT_{maj}$ and $AADT_{min}$

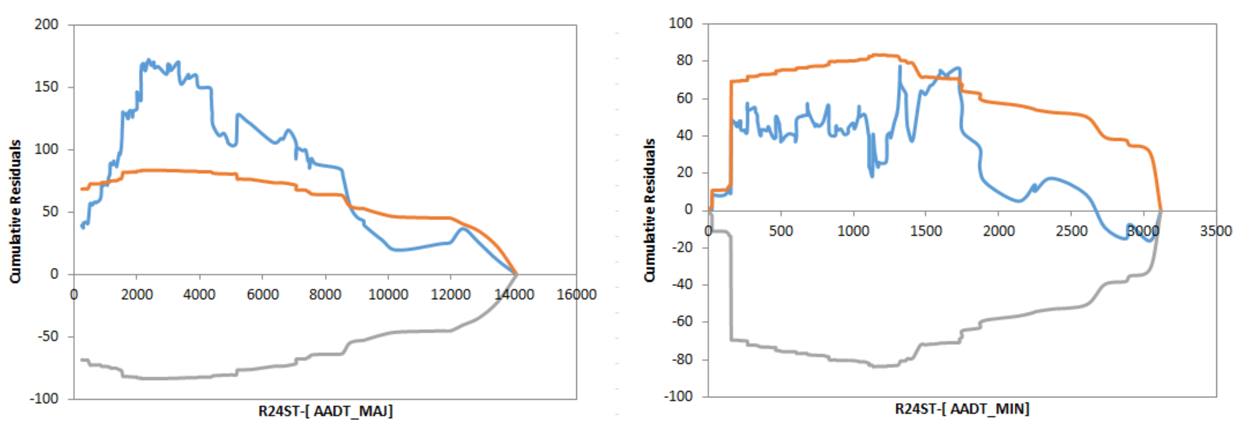


Figure 29: CURE plot of R24ST with respect to $AADT_{maj}$ and $AADT_{min}$

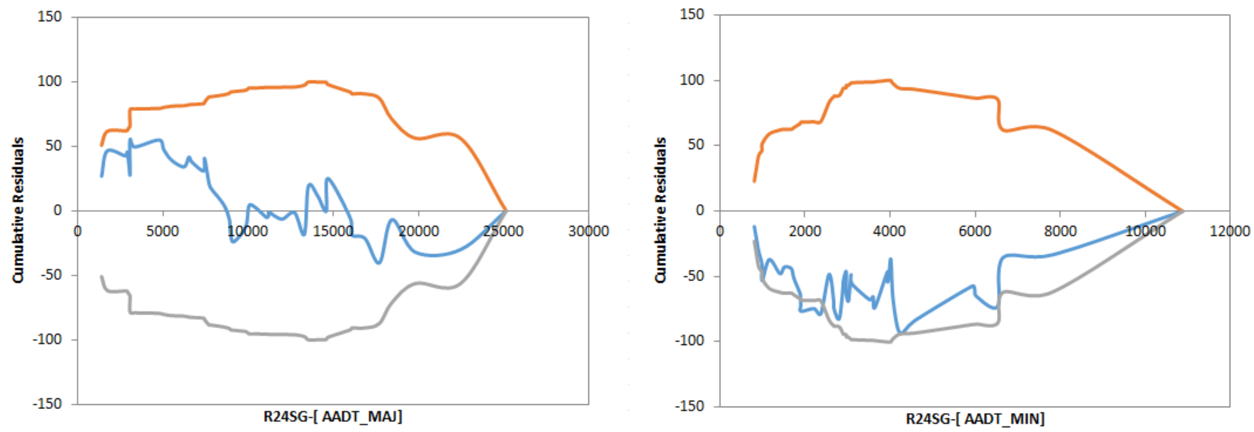


Figure 30: CURE plot of R24SG with respect to $AADT_{maj}$ and $AADT_{min}$

Development Results

The SPFs for R2 intersections were developed using the available datasets, based on the negative binomial model suggested by the HSM. The model estimation was performed in R statistical package. The development results for R23ST, R23SG, R24ST, and R24SG are shown in Table 25, Table 26, Table 27, and Table 28, respectively.

Table 25: Development results for R23ST

Variable	Estimate	Std Error	z-value	Pr(> |z|)
Intercept	-6.1386	0.34741	-17.613	< 2e-16 ***
Log($AADT_{MAJ}$)	0.49775	0.04194	11.869	< 2e-16 ***
Log($AADT_{MIN}$)	0.29571	0.03284	9.004	< 2e-16 ***
Overdispersion Parameter, k	0.619			

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.
1,569 dof, AIC: 4006.5

$$N_{spf\ 3ST} = exp[-6.139 + 0.498.\,ln(AADT_{maj}) + 0.296.\,ln(AADT_{min})] \qquad (18)$$

This SPF is applicable to an $AADT_{MAJ}$ range from zero to 33,750 vehicles per day and $AADT_{MIN}$ range from zero to 13,550 vehicles per day.

Table 26: Development results for R23SG

Variable	**Estimate**	**Std Error**	**z-value**	**Pr(> |z|)**
Intercept	-12.1399	1.7339	-7.001	2.53e-12***
Log($AADT_{MAJ}$)	1.1840	0.1811	6.539	6.20e-11***
Log($AADT_{MIN}$)	0.2809	0.1339	2.098	0.0359 .
Overdispersion Parameter, k	0.128			

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.
74 dof, AIC: 361.69

$$N_{spf\ 3SG} = exp\left[-12.140 + 1.184.\,ln\left(AADT_{maj}\right) + 0.281.\,ln(AADT_{min})\right] \quad (19)$$

This SPF is applicable to an $AADT_{MAJ}$ range from zero to 27,000 vehicles per day and $AADT_{MIN}$ range from zero to 9,750 vehicles per day. Note that the HSM does not have an SPF for R3SG intersection type.

Table 27: Development results for R24ST

Variable	**Estimate**	**Std Error**	**z-value**	**Pr(> |z|)**
Intercept	-3.71636	0.43042	-8.634	< 2e-16***
Log($AADT_{MAJ}$)	0.15934	0.04728	3.370	7.5e-04***
Log($AADT_{MIN}$)	0.42567	0.04962	8.579	< 2e-16***
Overdispersion Parameter, k	0.71			

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.
744 dof, AIC: 2415

$$N_{spf\ 4ST} = exp[-3.716 + 0.159.\,ln\left(AADT_{maj}\right) + 0.426.\,ln(AADT_{min})] \quad (20)$$

This SPF is applicable to an $AADT_{MAJ}$ range from zero to 24,500 vehicles per day and $AADT_{MIN}$ range from zero to 5,300 vehicles per day.

Table 28: Development results for R24SG

Variable	**Estimate**	**Std Error**	**z-value**	**Pr(> |z|)**
Intercept	-5.81141	0.77483	-7.500	6.37e-14***
Log($AADT_{MAJ}$)	0.34479	0.06612	5.215	1.84e-7***
Log($AADT_{MIN}$)	0.52624	0.07042	7.473	7.85e-14***
Overdispersion Parameter, k	0.218			

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:., 224 dof, AIC: 1106

$$N_{spf\ 4SG} = exp[-5.811 + 0.345.\, ln\left(AADT_{maj}\right) + 0.526.\, ln(AADT_{min})] \qquad (21)$$

This SPF is applicable to an $AADT_{MAJ}$ range from zero to 31,250 vehicles per day and $AADT_{MIN}$ range from zero to 13,250 vehicles per day.

Table 29 presents the R2 intersection SPFs developed for NJ and also the SPFs available in the HSM.

Table 29: Summary of SPFs for R2 intersections

Type	Source	SPF
R23ST	HSM	$N_{spf\ 3ST} = exp[-9.86 + 0.79 \times ln(AADT_{maj}) + 0.49 \times ln(AADT_{min})] \times CMF_{skew} \times CMF_{LT} \times CMF_{RT} \times CMF_{light}$
R23ST	Development	$N_{spf\ 3ST} = exp[-6.139 + 0.498.\, ln\left(AADT_{maj}\right) + 0.296.\, ln(AADT_{min})]$
R24ST	HSM	$N_{spf\ 4ST} = exp[-8.56 + 0.60 \times ln(AADT_{maj}) + 0.61 \times ln(AADT_{min})] \times CMF_{skew} \times CMF_{LT} \times CMF_{RT} \times CMF_{light}$
R24ST	Development	$N_{spf\ 4ST} = exp[-3.716 + 0.159.\, ln\left(AADT_{maj}\right) + 0.426.\, ln(AADT_{min})]$
R24SG	HSM	$N_{spf\ 4SG} = exp[-5.13 + 0.60 \times ln(AADT_{maj}) + 0.20 \times ln(AADT_{min})] \times CMF_{LT} \times CMF_{RT} \times CMF_{light}$
R24SG	Development	$N_{spf\ 4SG} = exp[-5.811 + 0.345.\, ln\left(AADT_{maj}\right) + 0.526.\, ln(AADT_{min})]$
R23SG	HSM	n/a
R23SG	Development	$N_{spf\ 3SG} = exp\left[-12.140 + 1.184.\, ln\left(AADT_{maj}\right) + 0.281.\, ln(AADT_{min})\right]$

RURAL MULTILANE SEGMENTS

The HSM's crash frequency predictive model specifies SPFs for the following two segment types for rural multilane roadways:

(1) Rural four-lane undivided segment (R4U)
(2) Rural four-lane divided segment (R4D)

The HSM defines the term "multilane" as a roadway facility with four or more through lanes. The rural multilane (RM) facilities may have occasional grade-separated interchanges, but these are not to be the primary forms of access and egress [(2)].

In order to calculate the calibration factors for R4U and R4D segments for NJ-specific conditions, the required dataset was first gathered and processed. This section presents a detailed description of the data requirements, data processing, and the results of calibration and development.

Data Requirements

The required data for calibration of RM segment predictive models as specified by the HSM are presented in Table 30.

Table 30: Data requirements for RM segments

DATA ELEMENT	REQUIRED	DESIRABLE	SOURCE
Segment Length	■		Segmentation of SLD
AADT	■		Sensor Database
Lane Width	■		SLD->*ln_pave_width*
Shoulder Width	■		SLD->*ln_shou_width*
Lighting		■	
Use of Automated Speed Enforcement		■	
For Undivided Segments Only			
Sideslope	■		
For Divided Segments Only			
Median Width	■		SLD->*ln_median_type* SLD->*ln_median_width*

As shown in Table 30, all required data except sideslope information was available in the compiled datasets. The research team used the default value of 1V:7H or flatter for sideslope, as assumed in the baseline conditions for undivided facilities in the HSM.

Gathering and Processing the Roadway Feature Dataset

Because the required data elements for RM segments were available in the compiled roadway feature dataset, there was no need for any manual extraction. However, the research team manually verified the output of the automatic identification of RM segments using Google Maps Street View.

Automatic Identification of Segments

Homogeneous segments were generated by following the process described in the Processing Data section. The flowchart for this process is presented in Figure 31. A generic flowchart for automatic identification of homogeneous segments is shown in Figure 13. RM segments were determined by filtering based on urbanization degree, median type, median width, shoulder width, and lane count, and by splitting where an intersection exists.

It should be mentioned that whether a roadway is divided or not is indicated in the *ln_highway_type* table, as shown in Table 5. The SLD classifies roadways with painted medians as divided facilities. However, the HSM identifies roadways with painted medians, i.e., flush medians, as undivided facilities. Therefore, using the median type information in the *ln_median_type* table, segments with painted medians are classified as undivided facilities in the automatic segment identification process.

Once homogeneous segments were identified, using the traffic volume dataset, distance between each segment and the closest traffic sensor was calculated. Table 31 presents the summary statistics on RM homogeneous segments. It can be seen that there is a total of only 45 homogenous R4D and 27 homogeneous R4U segments. However, the HSM suggests using segments of 0.1 mile or longer for calibration and development purposes. The team found 35 R4D segments of 0.1 mile or longer, 12 of which include a sensor station within the segment, and 19 R4U segments of 0.1 mile or longer, none of which include a sensor within the segment.

Due to the limited number of RM segments, the research team used all the available homogeneous segments.

Figure 31: Automatic identification of RM segments

Table 31: RM segments statistics

R4D	**Number of R2Ds**	**Average Distance between AADT Station and R4D**
Total	45	0.35
< 0.1 mile	34	0.36
AADT station within	12	0.17
R4U	**Number of R2Us**	**Average Distance between AADT Station and R4U**
Total	47	0.38
< 0.1 mile	32	0.36
AADT station within	0	n/a

Calibration Results

Chapter 11 of the HSM provides a crash prediction model for RM segments. The base SPF includes the AADT and segment length (L) as covariates:

$$N_{spf\ R4U} = \mathrm{AADT}^{1.176} \times L \times 365 \times 10^{-6} \times e^{(-9.653)} \quad (22)$$

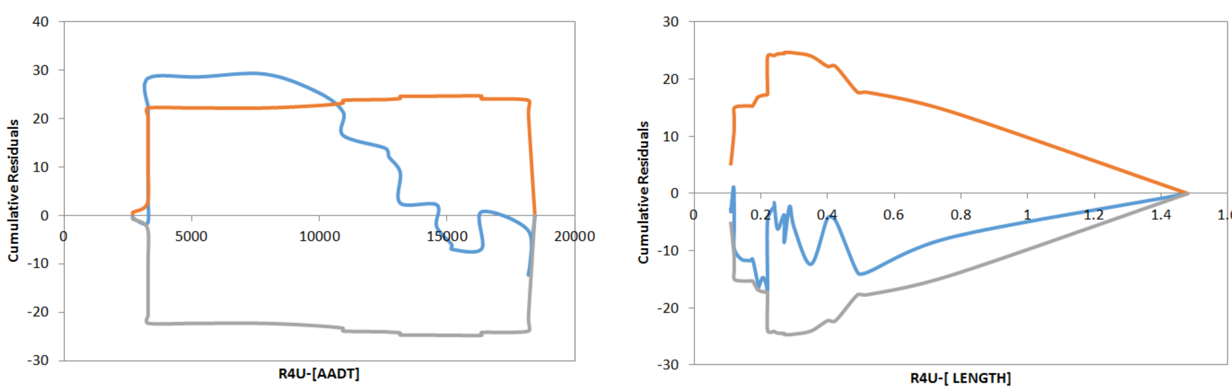
$$N_{spf\ R4D} = \mathrm{AADT}^{1.049} \times L \times 365 \times 10^{-6} \times e^{(-9.025)} \quad (23)$$

The reliable AADT range is specified by the HSM as [0 – 33,200 veh/day] for R4U and [0 – 89,300 veh/day] for R4D segments.

SPFs shown in equations (22) and (23) calculate the predicted crash frequency for the base conditions. The base conditions for R4U segments are 12-ft lanes, 6-ft paved shoulder width, no lighting, and 1V:7H or flatter sideslope; those for R4D segments are 12-ft lanes, 8-ft shoulder width, 30-ft median, and no lighting.

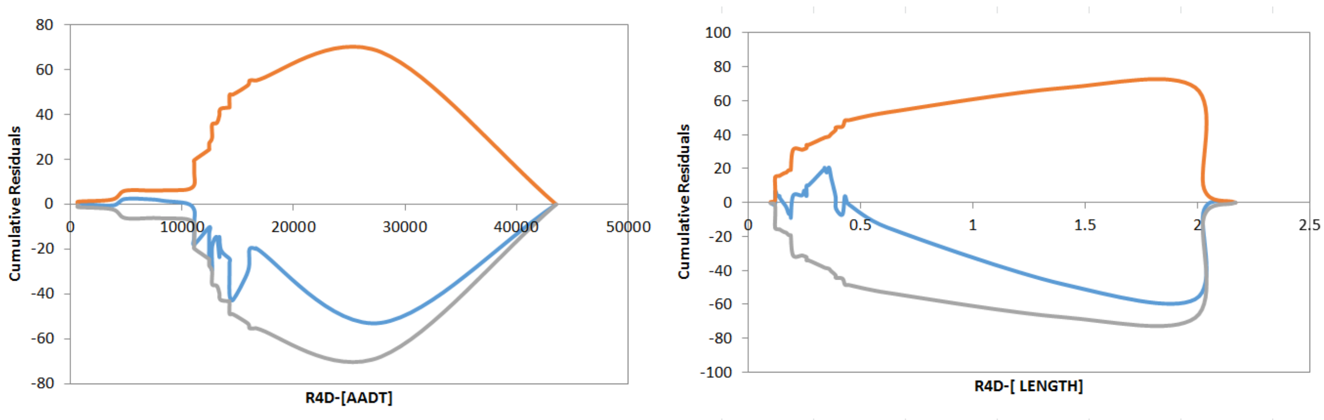
It is clear from the results shown in Table 31 that the sample size for RM segments is nearly sufficient enough for calibration, however, it is not enough for development of state-specific SPFs. This result is to be expected, since NJ is a densely populated state, and based on the SLD database, 86.8 percent of its roadway segments are in urban areas.

The Calibrator tool developed by the FHWA was used to calculate the calibration factors and measure their goodness of fit. The calibration factor for R4U was calculated as 1.10 with a standard error of 0.42 and coefficient of variation of 0.38. The calibration factor for R4D was calculated as 1.70 with a standard error of 0.80 and coefficient of variation of 0.47. As can be concluded from these values, the calibration factors do not yield the required statistical significance. The CURE plots for R4U and R4D with respect to AADT and segment length are shown in Figure 32 and Figure 33, respectively.



(a) AADT (b) Segment Length

Figure 32: CURE plots for R4U with respect to AADT and segment length



(a) AADT (b) Segment Length

Figure 33: CURE plots for R4D with respect to AADT and segment length

It can be seen from the CURE plots that the cumulative residuals are within the allowable upper and lower bounds, which signifies that the calibrated SPFs for R4U and R4D segments are statistically acceptable based on the available dataset. However, the coefficients of variation of the calibration factors for R4U and R4D were found to be 0.73 and 0.36. These values are well above the acceptable range of 0.10 to 0.15 as suggested by the FHWA (77). Because the available dataset is not large enough for development, the only alternative is the use of the calibration factors presented above, despite the subpar coefficient of variation values.

URBAN AND SUBURBAN SEGMENTS

The HSM's crash frequency predictive model specifies SPFs for the following types of urban and suburban segments: two-lane undivided (U2U), three-lane with center two-way left-turn lane (U3T), four-lane undivided (U4U), four-lane divided (U4D), and five-lane with center two-way left-turn lane (U5T), as shown in Figure 34.

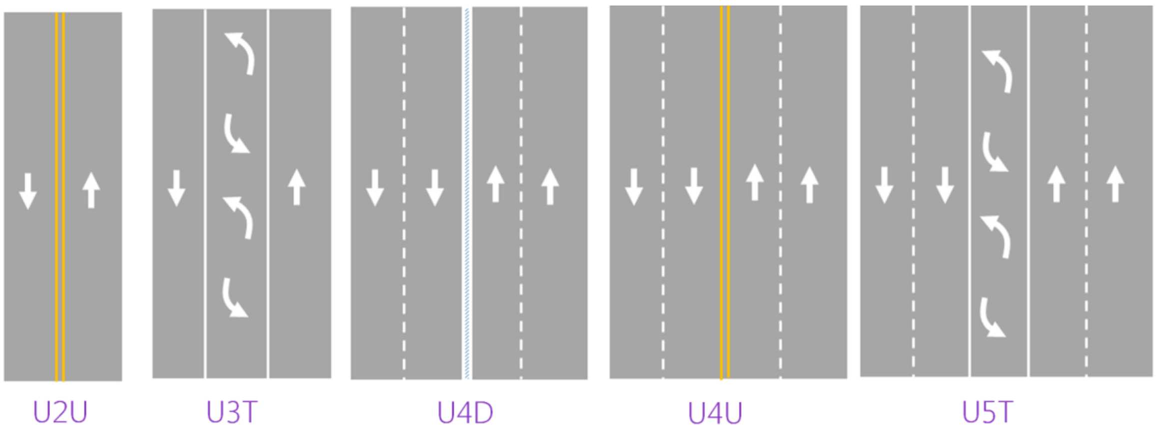


Figure 34: Urban and suburban segment types

This section presents a detailed description of the data requirements, data processing, and extraction of additional required data, and the results of SPF calibration and development.

Data Requirements

The required data for calibration of urban and suburban arterial segment SPFs as specified by the HSM are presented in Table 32.

Table 32: Data requirements for urban and suburban segments

DATA ELEMENT	REQUIRED	DESIRABLE	SOURCE
Segment Length			Segmentation of SLD
Number of Through Lanes			SLD-> *lane_count*
Presence of Median			SLD-> *ln_median_type*
Presence of Center Two-Way Left Lane			
AADT			Sensor Counts
Number of Driveways by Land Use Type			Manual Extraction
Posted Speed Limit			SLD-> *ln_speed*
Presence of On-Street Parking			Manual Extraction
Type of On-Street Parking			Manual Extraction
Roadside Fixed Object Density			
Lighting			Google Street View
Presence of Automated Speed Enforcement			

As seen in Table 32, the number of driveways per land use type, presence of on-street parking, and type of on-street parking, which are required by the HSM, are not readily available in the compiled roadway feature dataset. Thus, a comprehensive manual extraction was conducted to obtain these data elements.

Gathering and Processing the Roadway Feature Dataset

The next two subsections describe the automatic identification of urban and suburban homogeneous segments using the compiled database and the manual data extraction process to obtain the missing required data elements using Google Maps Street View.

Automatic Identification of Segment Types

Homogeneous segments were generated by following the process described in the Processing Data section. The flowchart of this process for automatic identification of urban and suburban intersections is presented in Figure 35. However, identification of U3T and U5T is not possible using the SLD database, since the center two-way left-turn lanes are not indicated in the database. It was recommended that the research team use the lane usage information available on the NJDOT website instead, but this dataset is not sufficient either, and is also limited to state routes only. Therefore, the automatic identification process only identifies the U2U, U4U, and U4D segment types.

It should be mentioned again that segments with painted medians are classified as undivided facilities, following the HSM guidelines.

Once homogeneous segments were identified, using the traffic volume dataset, distance between each segment and the closest traffic sensor was calculated. Table 33 presents the summary statistics on the number of homogeneous urban and suburban arterial segments. It can be seen that there are 275 U4U and 242 U4D segments where the AADT is of length more than 0.01 mile.

The extracted data for each segment type is presented in Table 34. It should be mentioned that for U4U and U4D the team conducted data extraction for the segments shown in Table 33. These segments are longer than 0.1 mile, with a sensor present within the segment. In order to increase the sample size, the team also extracted data for segments that are longer than 0.1 mile where the sensor is not within the segment but is less than 0.15 miles in proximity.

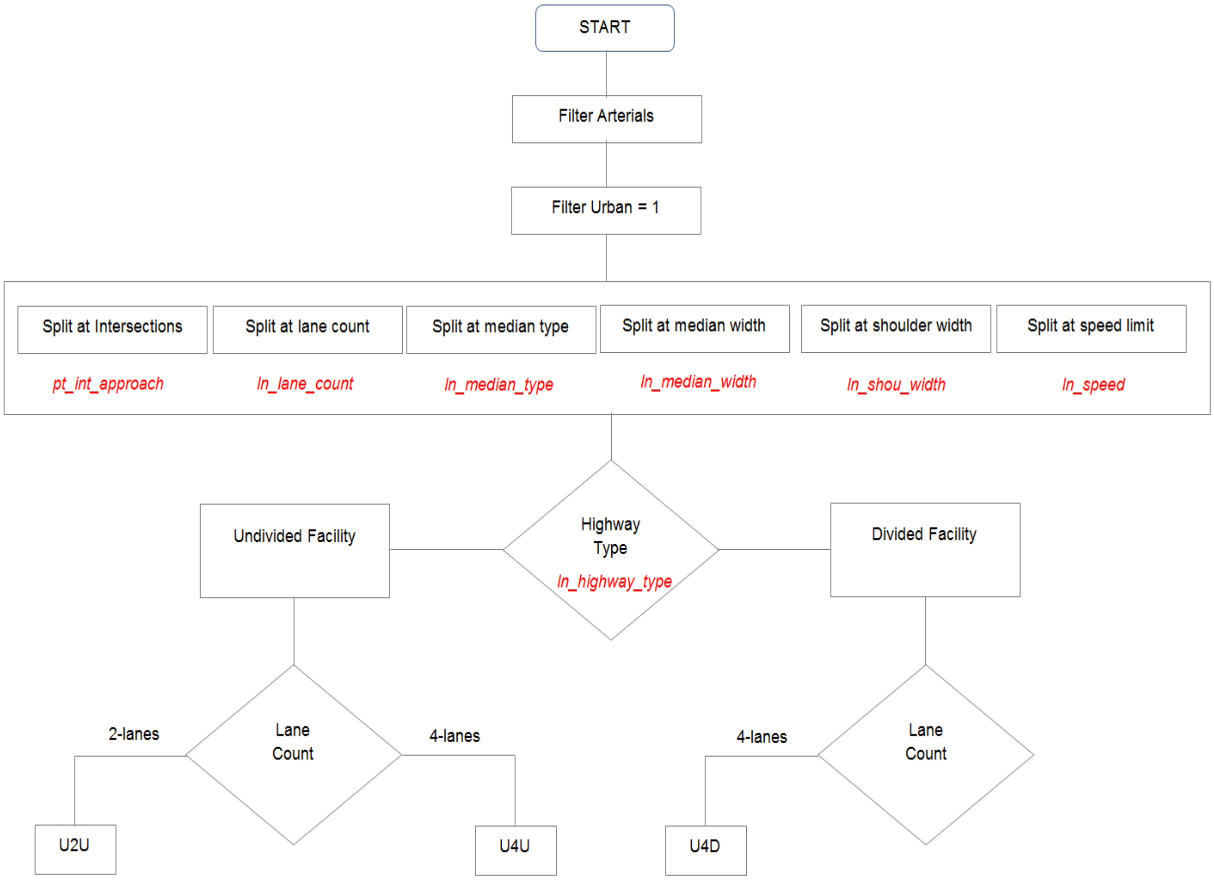


Figure 35: Flowchart for automatically identifying urban and suburban segments

Table 33: Urban and suburban segment statistics

Site Type	No. of Segments	No. of Segments of Length > 0.1 Mile	No. of Segments with AADT of Length > 0.1 Mile
2-lane undivided (U2U)	36,008	11,610	1,708
3-lane with TWLTL (U3T)	N/A	N/A	N/A
4-lane undivided (U4U)	6,062	1,675	275
4-lane divided (U4D)	2,833	1,315	242
5-lane with TWLTL (U5T)	N/A	N/A	N/A

Table 34: Data extracted for urban and suburban segments

Site Type	**Data Size**
2-lane undivided (U2U)	459
4-lane undivided (U4U)	514
4-lane divided (U4D)	387

Manual Data Extraction and Validation

As mentioned in the previous section, some of the required data elements were not available in the dataset, so manual extraction was conducted by the research team to obtain them. Google Maps was utilized to extract additional data and check the validity of the available data. The steps of manual data collection and validation are described in this section.

The objective of this data processing effort is twofold:

First is to *verify* the following attributes of urban and suburban segments obtained from SLDs:

1. Number of lanes in the segment
2. Type of segment (divided or undivided)
3. Whether the segment is, in fact, urban

Second is to *extract* the following attributes:

1. Presence or absence of roadway lighting
2. Total number of driveways in the roadway segment
3. Number of driveways by type (see Section C, point 8)
4. Total number of on-street parking spaces in the segment

The HSM classifies driveways as major commercial, minor commercial, major industrial/institutional, minor industrial/institutional, major residential, minor residential, or other.

The HSM classifies parking as parallel or angle parking, and requires the proportion of curb length within which parking is allowed to the total curb length of the segment.

The detailed steps of this process are as follows:

1. Open Google Maps and copy-paste the latitude and longitude of the start and end point of each segment. Use the Google Driving Route option from the start to the end point of segment.
2. Using the Satellite View option of Google Maps for the current segment, check whether it is in fact an actual segment and record in the worksheet as follows:
If it is not a segment, move to the next segment in the database. Otherwise leave this field blank and continue.
3. Check the number of lanes and whether divided or undivided. If the actual number of lanes in the segment does not match the segment type[2] of the current tab or an abnormal situation is observed, record "Y" and the actual number of lanes in the "*Lane Error/Abnormality*" field, **stop** and move to the next intersection. Otherwise leave this field blank and continue.

[2]Note: U2U should have 2 lanes, and the segment should be undivided. U3T should have 3 lanes.

4. Using the Street View option of Google Maps for the current segment, if no Street View of the current segment is available, record "N" in "*Street View*" field, **stop** and move to the next intersection. Otherwise leave this field blank and continue.
5. If the segment seems to be located in a rural area, record "Y" in field "*Possibly Rural,*" then continue. Note that this is a subjective decision. Otherwise leave this field blank and continue.
6. Extract driveway information (See Figure 36)
7. Count and record the number of driveways along the segment from the start to the end point coordinates.
8. While counting driveways, be mindful of other roads that might connect to the segment; they should be counted as driveways (See Figure 37)
9. Note the type of driveway

During the process of counting driveways, clearly mark what category each falls into. The HSM classifies driveways in the following manner:

- Major commercial
- Minor commercial
- Major industrial/institutional
- Minor industrial/institutional
- Major residential
- Minor residential
- Other

Any driveway that serves sites with 50 or more parking spaces is a major driveway. Otherwise note the driveway as “minor.” Driveways can be readily classified as minor or major from a quick review of aerial photographs that show parking areas or through user judgment based on the character of the establishment served by the driveway. Commerical driveways provide access to establishments that serve retail customers (See Figure 38). Industrial/Institutional driveways serve factories, warehouses, schools, hospitals, churches, offices, public facilities, and other places of employment. Commercial sites with no restriction on access along an entie property frontage are generally counted as two driveways. Residential driveways serve single and multi-family dwellings.

10. Extract on-street parking information. Note and count the number of on-street parking spaces along the segment (See Figure 39). Also, note that the parking spots marked with an X should not be counted (See Figure 40)

Figure 36: Presence of a driveway (minor residential) in the segment

Figure 37: Presence of a connecting road in the segment

Figure 38: Example of a minor commercial driveway

Figure 39: Example of a parking space

Figure 40: Example of a parking spot marked X

Calibration Results

Chapter 12 of the HSM provides a crash prediction model for urban and suburban segments. Crash prediction for segments is conducted for 5 different crash types, namely multi-vehicle non-driveway collisions (N_{brmv}), single-vehicle crashes (N_{brsv}), multi-vehicle driveway-related collisions (N_{brdwy}), vehicle-pedestrian collisions (N_{pedr}), and vehicle-bicycle collisions (N_{biker}). CMFs are applied only to the first three collision types.

$$N_{spf\ rs} = N_{brmv} + N_{brsv} + N_{brdwy} \tag{24}$$

$$N_{br} = N_{spf\ rs}.CMF_{1r}.CMF_{2r}.CMF_{3r}.CMF_{4r}.CMF_{5r} \tag{25}$$

Where, N_{br} is the predicted average crash frequency of a segment (excluding vehicle-pedestrian and vehicle-bicycle crashes), CMF_{1r} is on-street parking, CMF_{2r} is roadside fixed objects, CMF_{3r} is median width, CMF_{4r} is lighting, and CMF_{5r} is auto speed enforcement.

Note that N_{pedr} and N_{biker} are calculated as a proportion of N_{br}, as follows:

$$N_{pedr} = N_{br}.f_{pedr} \tag{26}$$

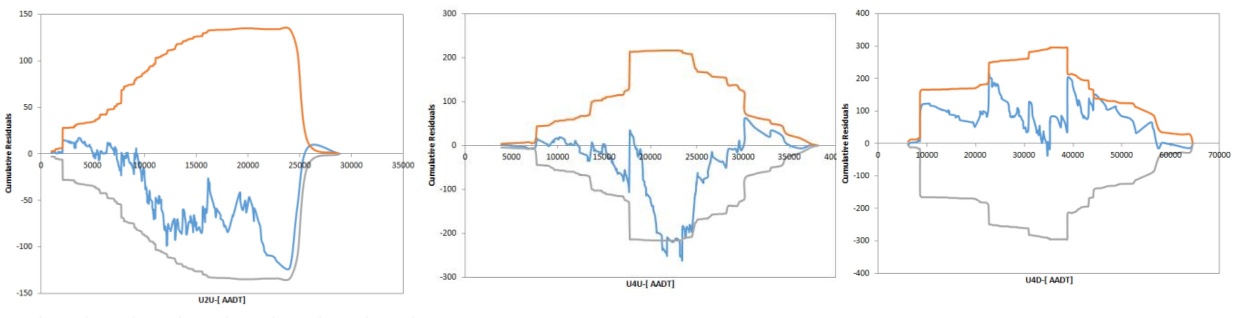
$$N_{biker} = N_{br}.f_{biker} \tag{27}$$

Using the compiled dataset, the calibration factors for each facility were calculated, as shown in Table 35.

Table 35: Calibration factors of urban and suburban segments

Segment Type	Calibration Factor	Standard Error	Coefficient of Variation
U2U	1.264	±0.14	0.11
U4U	1.097	±0.15	0.13
U4D	1.596	±0.21	0.13

The coefficients of variation for U2U, U4U, and U4D are 0.11, 0.13, and 0.13, respectively. It is suggested in the literature[3] that a reasonable upper threshold for the coefficient of variation is 0.10 to 0.15. The team also generated the CURE plots for each segment type, as shown in Figure 41.



(a) (b) (c)

Figure 41: CURE plots for (a) U2U, (b) U4U, and (c) U4U

Development Results

The research team performed the SPF development process for urban and suburban segments U2U, U4U, and U4D using the data extracted for each segment type. Data size for each segment type is shown in Table 34**.**

The research team followed the SPF form used in the HSM that links the expected number of crashes to AADT and segment length, shown here:

$$N_{spf} = e^{a+b.\ln(AADT)+\ln(L)} \quad (28)$$

[3] Lyon, C., Persaud, B., and Gross, F. (2016). “The Calibrator: An SPF Calibration and Assessment Tool User Guide.” FHWA-SA-17-016.

Note that because the crash database does not indicate whether or not a crash was driveway-related, the developed SPFs could not be grouped into driveway-related and non-driveway-related multi-vehicle crashes as is done in the HSM. The SPFs are for total crashes, multi-vehicle crashes, and single-vehicle crashes. Also note that the SPFs for vehicle-pedestrian and vehicle-bicycle SPFs are not developed, but f_{pedr} and f_{biker} values are determined specifically for NJ.

Table 36 shows the estimated coefficients of the developed SPFs for each segment type and crash type. The AADT range for U2U is [0 – 29,000 veh/day], for U4U is [0 – 48,600 veh/day], and for U4D is [0 – 84,100 veh/day]. The detailed statistical results are shown in Table 37, Table 38, and Table 39.

Table 36: Estimated coefficients of developed SPFs

Segment	Crash Type	SPF
U2U	Total	$N_{TOT\ U2U} = exp[-9.798 + 1.188. ln(AADT) + ln(L)]$
	Multi-Vehicle	$N_{MV\ U2U} = exp[-14.411 + 1.641. ln(AADT) + ln(L)]$
	Single-Vehicle	$N_{SV\ U2U} = exp[-3.977 + 0.435. ln(AADT) + ln(L)]$
U4U	Total	$N_{TOT\ U4U} = exp[-12.01 + 1.432. ln(AADT) + ln(L)]$
	Multi-Vehicle	$N_{MV\ U4U} = exp[-13.794 + 1.59. ln(AADT) + ln(L)]$
	Single-Vehicle	$N_{SV\ U4U} = exp[-6.961 + 0.751. ln(AADT) + ln(L)]$
U4D	Total	$N_{TOT\ U4D} = exp[-3.00 + 0.543. ln(AADT) + ln(L)]$
	Multi-Vehicle	$N_{MV\ U4D} = exp[-3.363 + 0.558. ln(AADT) + ln(L)]$
	Single-Vehicle	$N_{SV\ U4D} = exp[-4.687 + 0.543. ln(AADT) + ln(L)]$

Table 37: SPF development results for U2U

	Variable	Estimate	Std Error	z-value	Pr(> |z|)
Total	Intercept	-9.79816	0.67509	-14.51	<2e-16 ***
	Log(AADT)	1.1888	0.07258	16.38	<2e-16 ***
	k	1.01	df: 1879 AIC: 4362.6		
Multi-Vehicle	Intercept	-14.4113	0.90904	-15.85	<2e-16 ***
	Log(AADT)	1.64054	0.09711	16.89	<2e-16 ***
	k	1.47	df: 1879 AIC: 3513.9		
Single-Vehicle	Intercept	-3.97695	0.80255	-4.955	7.22e-07***
	Log(AADT)	0.43518	0.08718	4.992	5.99e-07***
	k	0.701	df: 1879 AIC: 2230.3		

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.

Table 38: SPF development results for U4U

	Variable	Estimate	Std Error	z-value	Pr(> |z|)
Total	Intercept	-12.0081	1.2848	-9.346	<2e-16 ***
	Log(AADT)	1.4318	0.1304	10.984	<2e-16 ***
	k	1.81	df: 1094 AIC: 3925.5		
Multi-Vehicle	Intercept	13.7937	1.4572	-9.466	<2e-16 ***
	Log(AADT)	1.5912	0.1478	10.769	<2e-16 ***
	k	2.27	df: 1094 AIC: 3524.1		
Single-Vehicle	Intercept	-6.9611	1.5328	-4.541	5.59e-06 ***
	Log(AADT)	0.7514	0.1553	4.838	1.31e-06 ***
	k	0.90	df: 1094 AIC: 1678.1		

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.

Table 39: SPF development results for U4D

	Variable	Estimate	Std Error	z-value	Pr(> |z|)
Total	Intercept	-3.0009	0.94179	-3.186	0.00144 **
	Log(AADT)	0.54256	0.09147	5.931	3.01e-09 ***
	k	1.48	df: 899 AIC: 4285		
Multi-Vehicle	Intercept	-3.3629	1.0992	-3.060	0.00222 **
	Log(AADT)	0.5581	0.1068	5.227	1.73e-07 ***
	k	2.07	df: 899 AIC: 3876.1		
Single-Vehicle	Intercept	-4.6871	1.1111	-4.218	2.46e-05 **
	Log(AADT)	0.5429	0.1072	5.066	4.07e-07 ***
	k	0.715	df: 899 AIC: 2005.4		

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.

The AADT ranges applicable to the SPFs developed for roadway segments on urban and suburban arterials are as follows: Zero to 33,000 vehicles per day for U2U, zero to 48,750 vehicles per day for U4U and zero to 84,500 vehicles per day for U4D.

Table 40 presents the vehicle-pedestrian collision adjustment factor f_{pedr} and the vehicle-bicycle collision adjustment factor f_{bike}, calculated using the 2011 to 2015 NJ crash database.

Table 40: Vehicle-pedestrian and vehicle-bicycle collision factors at urban and suburban segments

Collision Type	Segment Type	Posted Speed ≤ 30 mph	Posted Speed > 30 mph
Vehicle-Pedestrian (f_{bike})	U2U	0.0208	0.0085
	U4U	0.0252	0.0109
	U4D	0.0128	0.0071
Vehicle-Bicycle (f_{pedr})	U2U	0.0068	0.0045
	U4U	0.0048	0.0029
	U4D	0.0029	0.0019

URBAN AND SUBURBAN INTERSECTIONS

The HSM's crash frequency predictive model specifies SPFs for four types of urban and suburban intersections:

(1) Three-leg intersections with stop control on the minor-road approach (U3ST)
(2) Three-leg signalized intersections (U3SG)
(3) Four-leg intersections with stop control on the minor-road approaches (U4ST)
(4) Four-leg signalized intersections (U4SG)

This section presents a detailed description of the data requirements, data processing, and extraction of additional required data, and the results of SPF calibration and development.

Data Requirements

The required data for calibration of urban and suburban intersection SPFs as specified by the HSM are presented in Table 41.

Table 41: Data requirements for R2 intersections

DATA ELEMENT	REQUIRED	DESIRABLE	SOURCE
Number of Intersection Legs			Intersection Database
Type of Traffic Control			Intersection Database
AADT for Major Road			Sensor Database
AADT for Minor Road			Sensor Database
Number of Approaches with Left-Turn Lanes			Google StreetView
Number of Approaches with Right-Turn			Google StreetView
Lighting			Google StreetView
For Signalized Intersections Only			
Type of Left Turn Phasing			Google StreetView
Use of RTOR			Google StreetView
Use of Red Light Cameras			N/A for New Jersey
Pedestrian Volume			Default Values from HSM
Max Number of Lanes Crossed by Pedestrians			Google StreetView
Presence of Bus Stops within 1,000 feet			NJ Transit GIS Map
Presence of Schools within 1,000 feet			OpenStreet Map
Presence of Alcohol Sales Est. within 1,000 feet			

As seen in Table 41, the *required* dataset for all urban and suburban intersections includes AADT for both major and minor roads, lighting, and intersection left- and right-turn lanes; for signalized intersections, left-turn signal phasing, and right-turn-on-red-prohibited data are also required. The *desired* dataset includes maximum lanes for pedestrian crossing, pedestrian volumes, bus stops within 1,000 ft, schools within 1,000 ft, and alcohol sales establishments within 1,000 ft. Note that among the data listed in Table 41, only the presence of alcohol sales establishments was not extracted.

The base conditions that apply to the SPFs in the HSM are no intersection left-turn or right-turn lanes and no lighting present.

Gathering and Processing the Roadway Feature Dataset

The roadway feature dataset was used to identify the type of intersections and extract the information necessary for calculating CMFs. Some of the required data – namely the number of left- and right-turn lanes at each intersection, left-turn signal phasing, and right-turn-on-red-prohibited flag – are not included in any of the available databases. Therefore, the roadway feature data was used primarily to automatically identify the type of intersection. Consequently, the research team conducted a manual data extraction process to verify the information in the SLD database and extract the missing variables using Google Earth.

Automatic Identification of Intersection Types

The flowchart of the process for automatic identification of urban and suburban intersections is presented in

Figure 42.

Following the same generic process described in the Processing Data section, urban intersections were automatically identified using the information provided in the available dataset, namely urbanization, lane count, number of intersection legs, and type of traffic control. These features are included in the available roadway features and intersection database, as shown in Table 19. Then, the type of intersection were preliminarily identified as follows:

- Intersections with urban (Urban Code), less than 6 lanes (Lane Count), three-leg (Number of Intersection Legs), and NA (Type of Traffic Control) were labeled as U3ST candidates.
- Intersections with urban (Urban Code), less than 6 lanes (Lane Count), three-leg (Number of Intersection Legs), and Signalized (Type of Traffic Control) were labeled as U3SG candidates.
- Intersections with urban (Urban Code), less than 6 lanes (Lane Count), four-leg (Number of Intersection Legs), and NA (Type of Traffic Control) were labeled as U4ST candidates.
- Intersections with urban (Urban Code), less than 6 lanes (Lane Count), four-leg (Number of Intersection Legs), and Signalized (Type of Traffic Control) were labeled as U4SG candidates.

Note that this is only a preliminary step and that a manual process is required to finalize the identification of intersection types, as described in the next subsection.

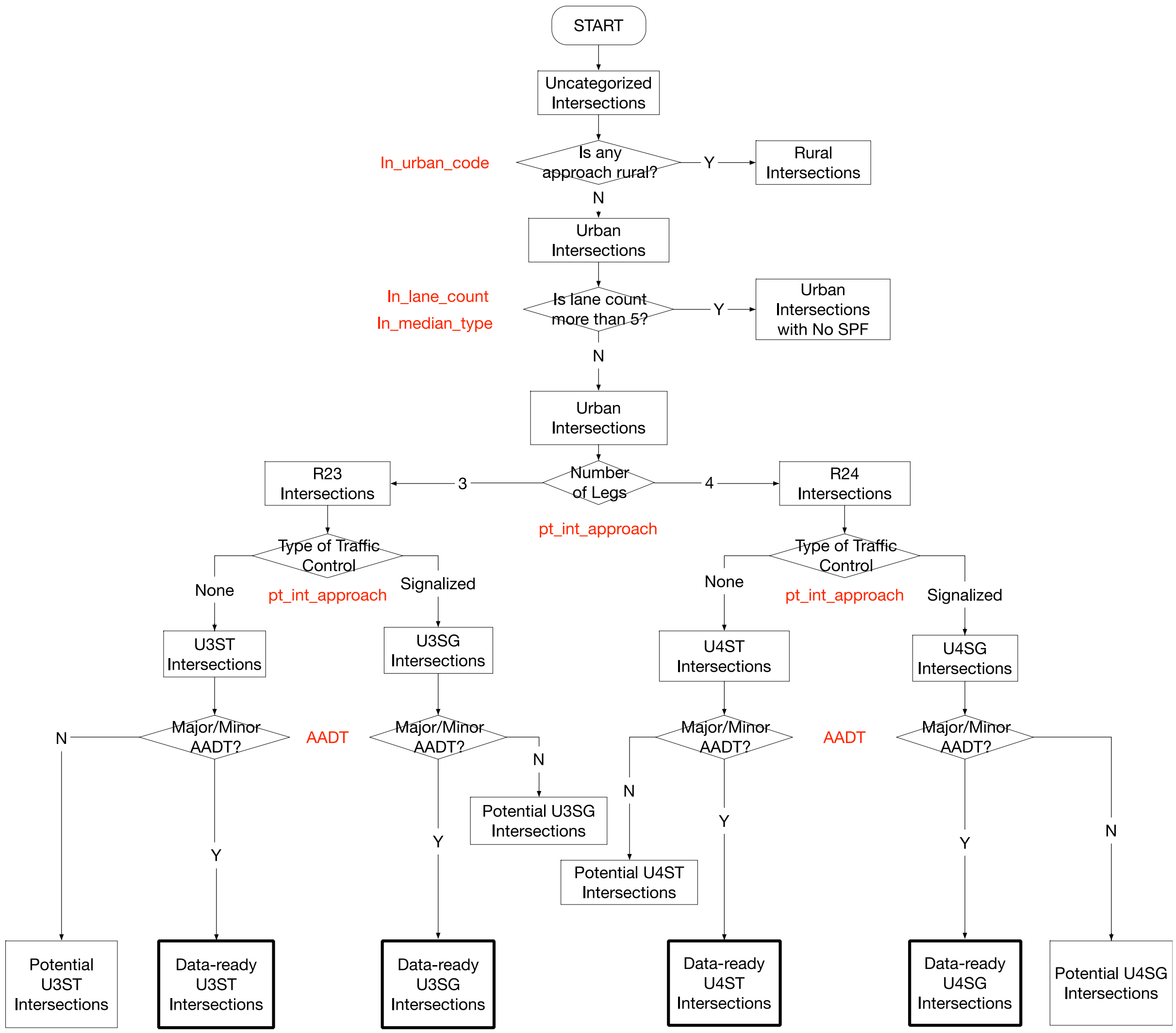


Figure 42: Flowchart of process for automatically identifying urban and suburban intersections

Manual Data Extraction and Validation

Google Maps was utilized to extract additional data and check the validity of the data obtained from RF2. The steps of manual data collection and validation are as follows:

First is to verify the following attributes of urban intersections obtained from the roadway features and intersection database:

1. Number of intersection legs (three-leg or four-leg)
2. Type of traffic control (signalized or unsignalized)
3. Misclassification of a bridge overpass as an intersection

4. Stop signs only on minor approaches (for unsignalized intersections only)

Second is to extract the following attributes:

1. Presence of lighting
2. Total number of approaches with left-turn lanes
3. Total number of approaches with right-turn lanes
4. Presence of left-turn signal phase (for signalized intersections only)
5. Type of left-turn signal phase (for signalized intersections only)
6. Presence of right-turn-on-red (for signalized intersections only)
7. Presence of red-light cameras (for signalized intersections only)

The detailed steps are as follows:

1. Copy and paste the latitude and longitude of each intersection in the Google Map search box.
2. Using the Satellite View option of Google Maps for the current intersection, check whether it is in fact an actual intersection and record in the worksheet as follows: If it is an overpass, record “Y” in the "*Overpass*" field, stop and move to the next intersection in the database. Otherwise, leave this field blank and continue.
3. Check the number of intersection legs. If the actual number of intersection legs does not match the intersection type[4] of the current tab or an abnormal situation is observed, record "Y" and the actual number of legs in “*Leg Error/Abnormality*” field, stop and move to the next intersection. Otherwise leave this field blank and continue.
4. Using the Street View option of Google Maps for the current intersection, if no Street View of the current intersection is available, record "N" in "*Street View*" field, stop and move to the next intersection. Otherwise leave this field blank and continue.
5. If the intersection seems to be located in a rural area, record "Y" in field "*Possibly Rural,*" then continue. Note that this is a subjective decision. Otherwise leave this field blank and continue. For example: a rural-like intersection (39.55941638, -75.48753948) marked as urban in the straight line diagrams database (See Figure 43).
6. If the number of lanes of any approach is six or more, record "Y" in "*Lane Number Error*" field, **stop** and move to the next intersection. Otherwise leave this field blank and continue.
7. Extract presence of lighting information (See Figure 44): If there are light poles at the intersection, record "Y" in "*Light*" field, otherwise, record "N."
8. Verify whether the intersection is signalized or not. If the current intersection type is signalized (i.e., U3SG or U4SG) and the intersection is in fact unsignalized, record “N” in in the "*Signalized*" field. Similarly, if the current intersection type is unsignalized (i.e., U3ST or U4ST) and the intersection is in fact signalized, record “Y” in the “*Signalized*” field (See Figure 45). Note that some traffic lights are flashing yellow/red. Intersections with these kinds of traffic lights should also be recorded as "N," for example see intersection (39.37669654, -74.82407489).

[4]Note: U3SG/U3ST should have 3 legs; U4ST/U4SG should have 4 legs.

9. Extract the total number of left-turn approaches. For signalized intersections, count all approaches with left-turn lane(s), and record this number in the "*LT Approaches*" field (see Table 42). For a minor stop-controlled intersection, count the number of approaches with left-turn lane(s) on major roads without stop-control, and record in the "*LT Approaches*" field (see Table 43).

10. Extract the total number of right-turn approaches. This is similar to the previous step. For signalized intersections, count all approaches with right-turn lanes, and record in the "*RT Approaches*" field. For stop-controlled intersections, count approaches with right-turn lanes for major roads without stop-control, and record in the "*RT Approaches*" field.

11. Record the maximum number of lanes crossed by pedestrians by checking each crosswalk on each approach.

Figure 43: A rural-like intersection

Figure 44: Presence of lighting in an intersection

Figure 45: A signalized intersection

Table 42: Number of left-turn approaches at a signalized intersection

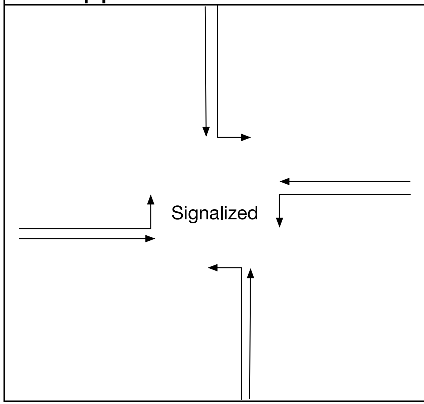
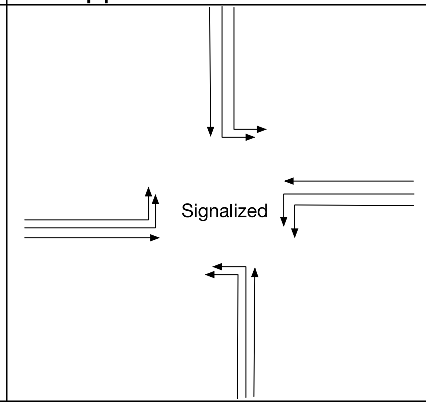
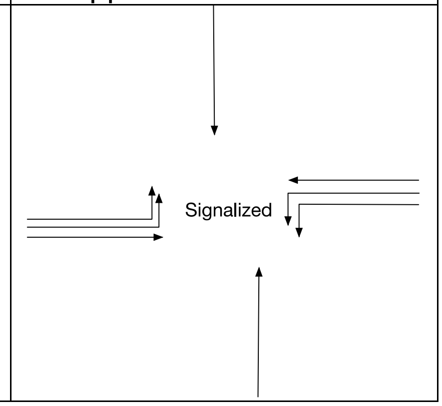
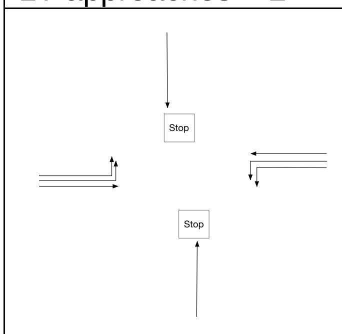
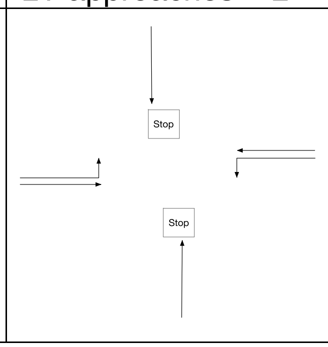
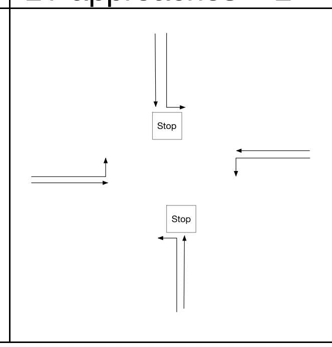
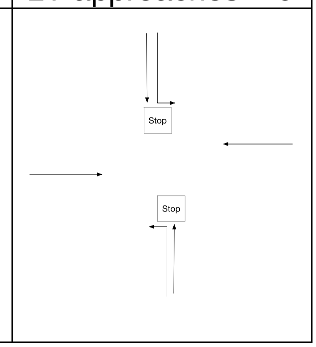
LT approaches = 4	LT approaches = 4	LT approaches = 2
		

Table 43: Number of left-turn approaches at a minor stop-controlled intersection

LT approaches = 2	LT approaches = 2	LT approaches = 2	LT approaches = 0
			

For U3ST and U4ST only:

11. Using the names of major and minor approaches, check which approach is stop-controlled:

If only minor road is stop-controlled, record "Y" in the "*Minor Stop Control*" field;
If only major road is stop-controlled, record "R"[5] in the "*Minor Stop Control*" field;
If an intersection is all-way stop-controlled, record "A"[6] in the "*Minor Stop Control*" field and move to the next intersection.

For U3SG and U4SG only:

12. Extract intersection left-turn signal phasing type. For each approach, determine whether the left-turn signal type is *Permissive*, *Protected/Permissive,* or *Protected,* and record this information in the "*LT Signal Phase*" field. In order to identify the type of left-turn signal phase using the Street View images of Google Maps, follow these guidelines:

[5] **R** means reverse.

[6] **A** means all-way stop-controlled.

Permissive:

- U4SG: All traffic signals have three lights, and no additional sign is used.
- U3SG:
    - For an approach with through lanes (like approaches 1 and 2 in Figure 46), if all traffic signals have three lights and no additional sign is used, the approach should be identified as permissive.
    - An approach without through lanes (like approach 3 in Figure 46) should not be recognized as a permissive approach.

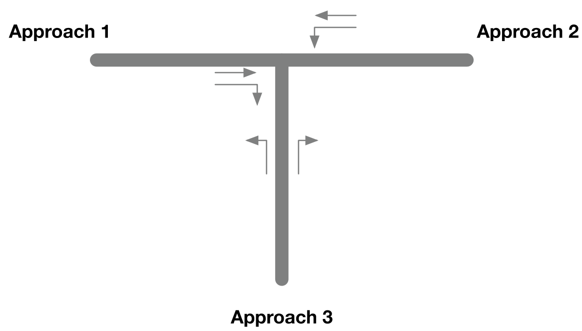


Figure 46: An example of a three-leg intersection

Protected/Permissive: One of the traffic signals has four lights and there is a sign indicating that left-turns are allowed on green, as shown in Figure 47.

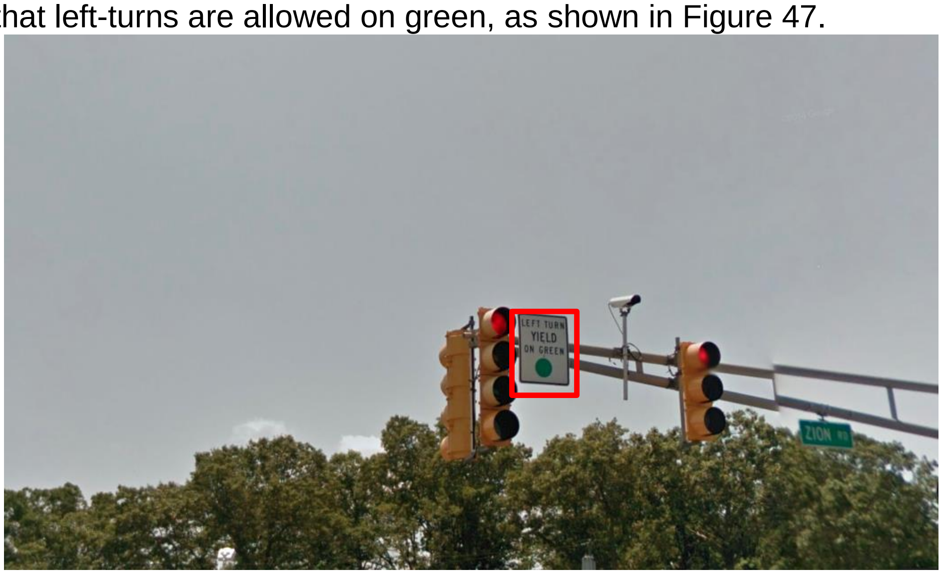


Figure 47: A sign indicating that left-turns are allowed on green

Protected: One of the traffic signals has four lights, and there are no additional signs.

- U4SG: Any traffic signal has four lights, and there are no additional signs.
- U3SG:
    - For an approach with through lanes (like approaches 1 and 2 in Figure 46), if any traffic signal has four lights and there are no additional signs, the approach should be identified as protected.
    - An approach without through lanes (like approach 3 in Figure 46) should be identified as a protected approach.

13. Extract right-turn-on-red information. Record the number of signalized intersection approaches for which right-turn-on-red is prohibited (see Figure 48) in the "*NO RTOR*" field.

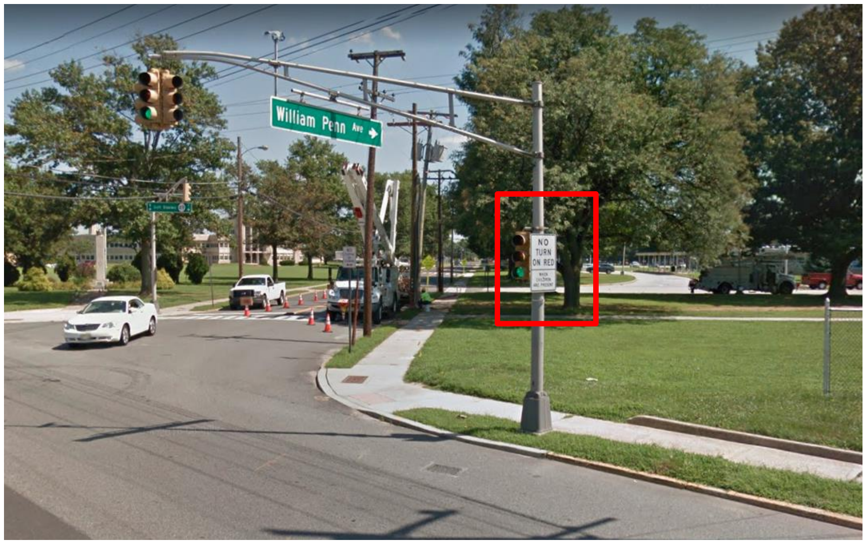


Figure 48: No-turn-on-red sign

Intersection Crash Frequency and Traffic Counts

The process of assigning traffic counts and crash data to urban and suburban intersections is similar to that for R2 intersections. Details of this process are presented in the Processing Data section.

Urban and Suburban Intersection Dataset

Following the data processing and extraction procedure, a total of 721 urban and suburban intersections were identified, as shown in Table 44. These intersections are also summarized by county in Table 45.

Table 44: Sample size of preliminary selection and final selection

Type	U3ST	U3SG	U4ST	U4SG
Preliminary Sample Size	1,592	755	1,471	1,961
Manually Reviewed	420	377	610	310
Final Sample Size	227	164	121	209

The quality of AADT data for major and minor approach at an intersection depends on the distance between the station and the target intersection, and the number of intersections between the AADT station and the target intersection. Therefore, the average intersection number between major/minor AADT station and the target intersection, and the average distance between major/minor AADT station and the target intersection, are summarized in Table 46.

Table 45: Sampled size by county of preliminary selection and final selection

County	Preliminary Sample Size	Final Sample Size
Atlantic	398	55
Bergen	492	42
Burlington	388	51
Camden	591	91
Cape May	183	39
Cumberland	281	27
Essex	298	40
Gloucester	390	51
Hudson	141	13
Hunterdon	84	8
Mercer	398	56
Middlesex	337	38
Monmouth	325	25
Morris	234	29
Ocean	217	31
Passaic	248	30
Salem	115	12
Somerset	166	22
Sussex	109	13
Union	236	21
Warren	148	27

Table 46: Summary of intersection AADT data

Type	Average Major AADT	Average Minor AADT	Average Intersection Number between Major station and the target intersection	Average Intersection Number between Minor Station and the Target Intersection	Average Distance between Major Station and the Target Intersection (mile)	Average Distance between Minor Station and the Target Intersection (mile)
U3ST	8,609	2,310	0.0000	0.0044	0.483	0.472
U3SG	15,591	7,832	0.348	0.1402	0.401	0.606
U4ST	7,140	1,427	0.165	0.0279	0.351	0.415
U4SG	16,212	8,402	0.005	0.0000	0.426	0.419

Detailed Description of Sampling Results

Three-Leg Stop-Controlled Intersections (U3ST): The final sample set contains 227 U3ST intersections (see Figure 49). For those intersections, the average AADT on major road is 8,609; the average AADT on minor road is 2,310; the average intersection count between AADT station and target intersection on major road is 0.00; the average intersection count between AADT station and target intersection on minor road is 0.004; the average distance between AADT station and target intersection on major road is 0.48 miles; and the average distance between AADT station and target intersection on minor road is 0.47 miles.

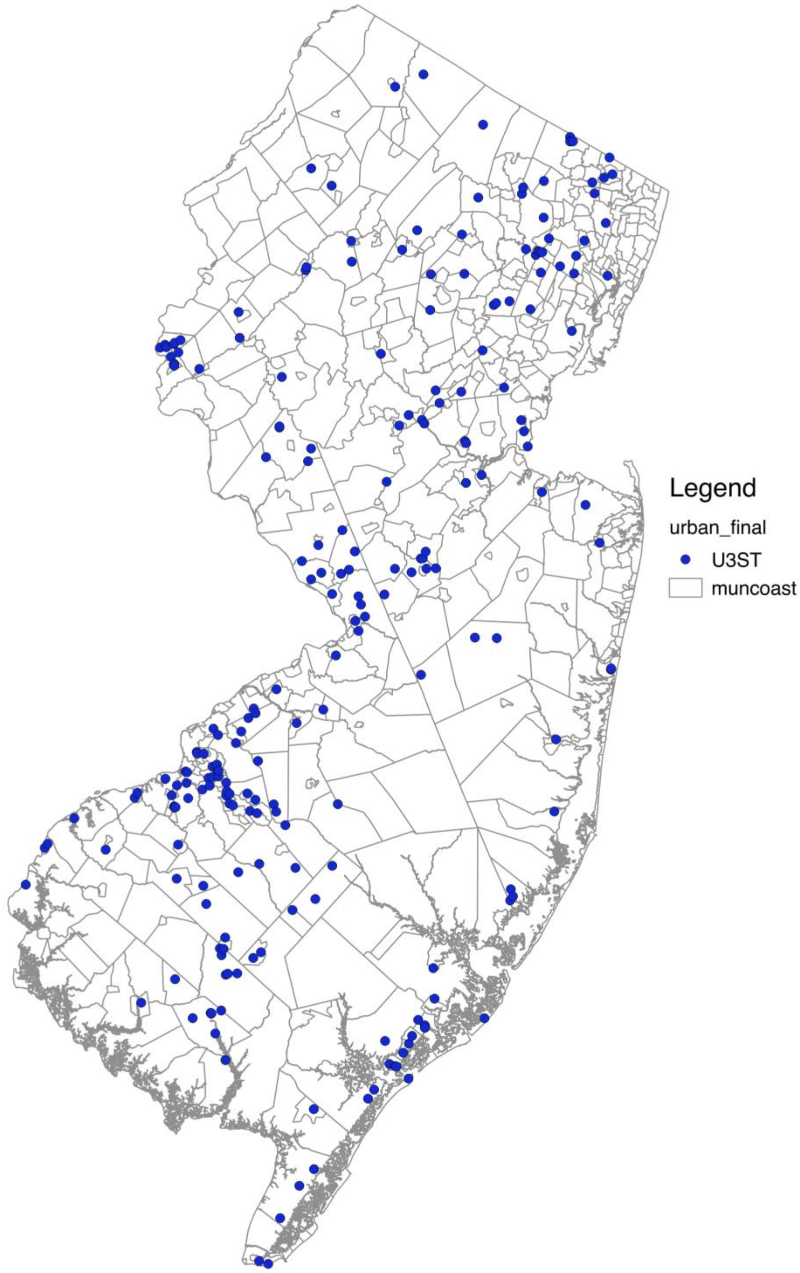


Figure 49: The spatial distribution of U3ST intersections

Three-Leg Signalized Intersections (U3SG): The final sample set contains 164 U3SG intersections (see Figure 50). For those intersections, the average AADT on major road is 15,519; the average AADT on minor road is 7,832; the average intersection count between AADT station and target intersection on major road is 0.35; the average intersection count between AADT station and target intersection on minor road is 0.14; the average distance between AADT station and target intersection on major road is

0.40 miles; and the average distance between AADT station and target intersection on minor road is 0.61 miles.

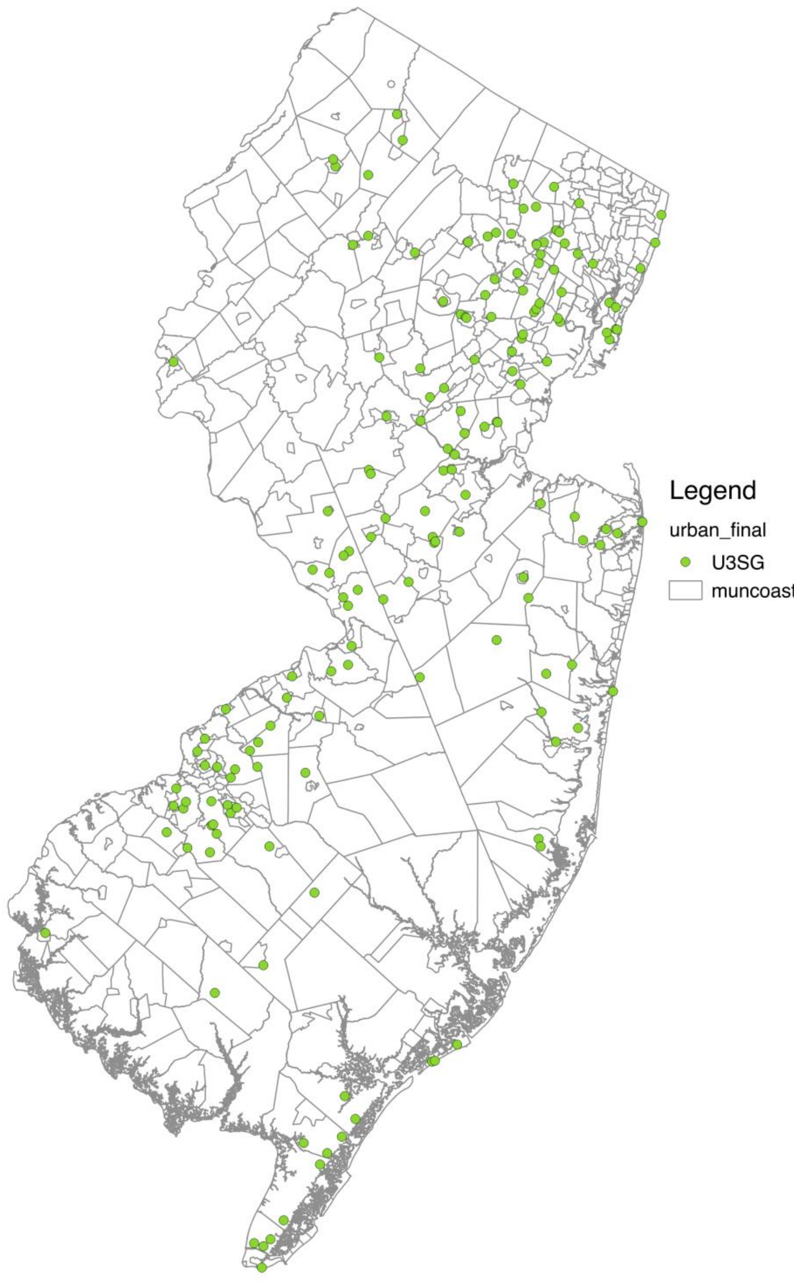


Figure 50: The spatial distribution of U3SG intersections

Four-leg Stop-Controlled Intersections (U4ST): The final sample set contains 121 U4ST intersections (see Figure 51). For those intersections, the average AADT on major road is 7,140; the average AADT on minor road is 1,427; the average intersection count between AADT station and target intersection on major road is 0.17; the average intersection count between AADT station and target intersection on minor road is 0.03; the average distance between AADT station and target intersection on major road is 0.35 miles; and the average distance between AADT station and target intersection on minor road is 0.42 miles.

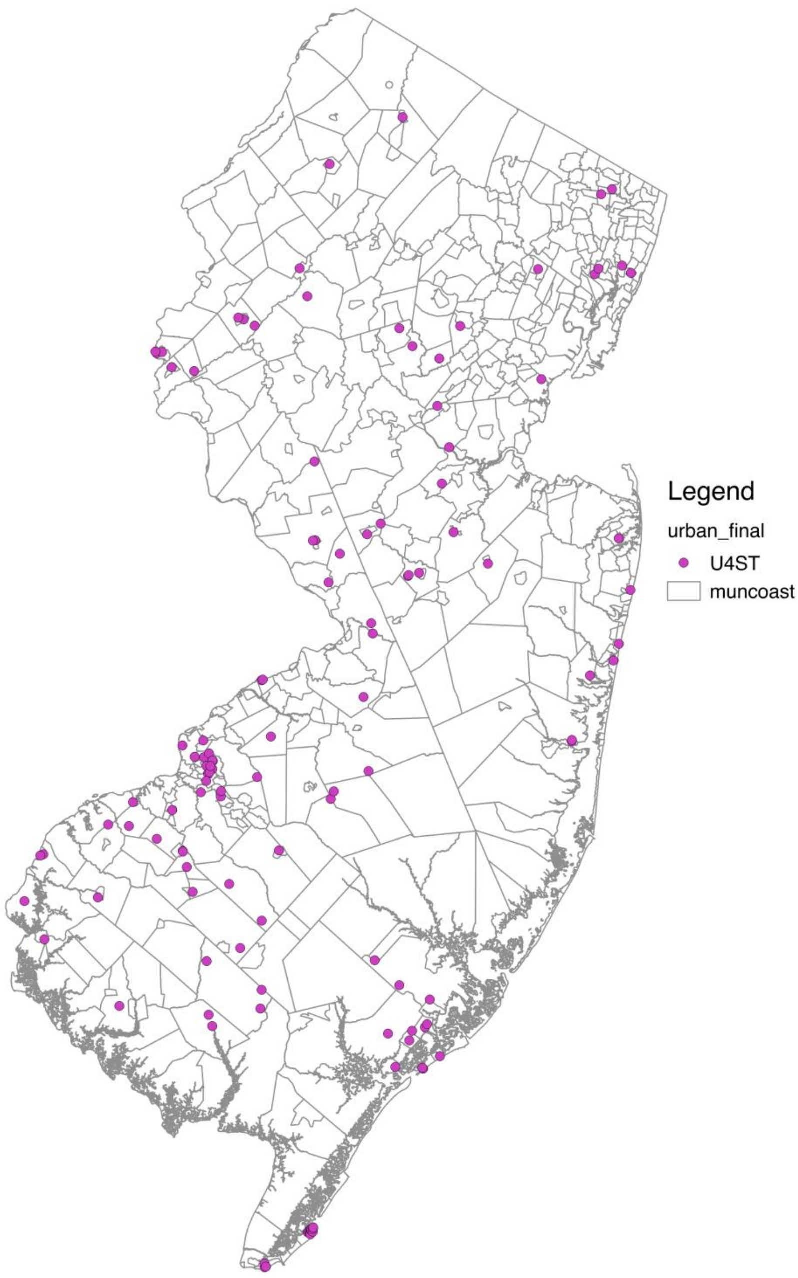


Figure 51:The spatial distribution of U4ST intersections

Four-leg Signalized Intersections (U4SG): The final sample set contains 209 U4SG intersections (see Figure 52). For those intersections, the average AADT on major road is 16,212; the average AADT on minor road is 8,402; the average intersection count between AADT station and target intersection on major road is 0.005; the average intersection count between AADT station and target intersection on minor road is 0.00; the average distance between AADT station and target intersection on major road is 0.43 miles; and the average distance between AADT station and target intersection on minor road is 0.42 miles.

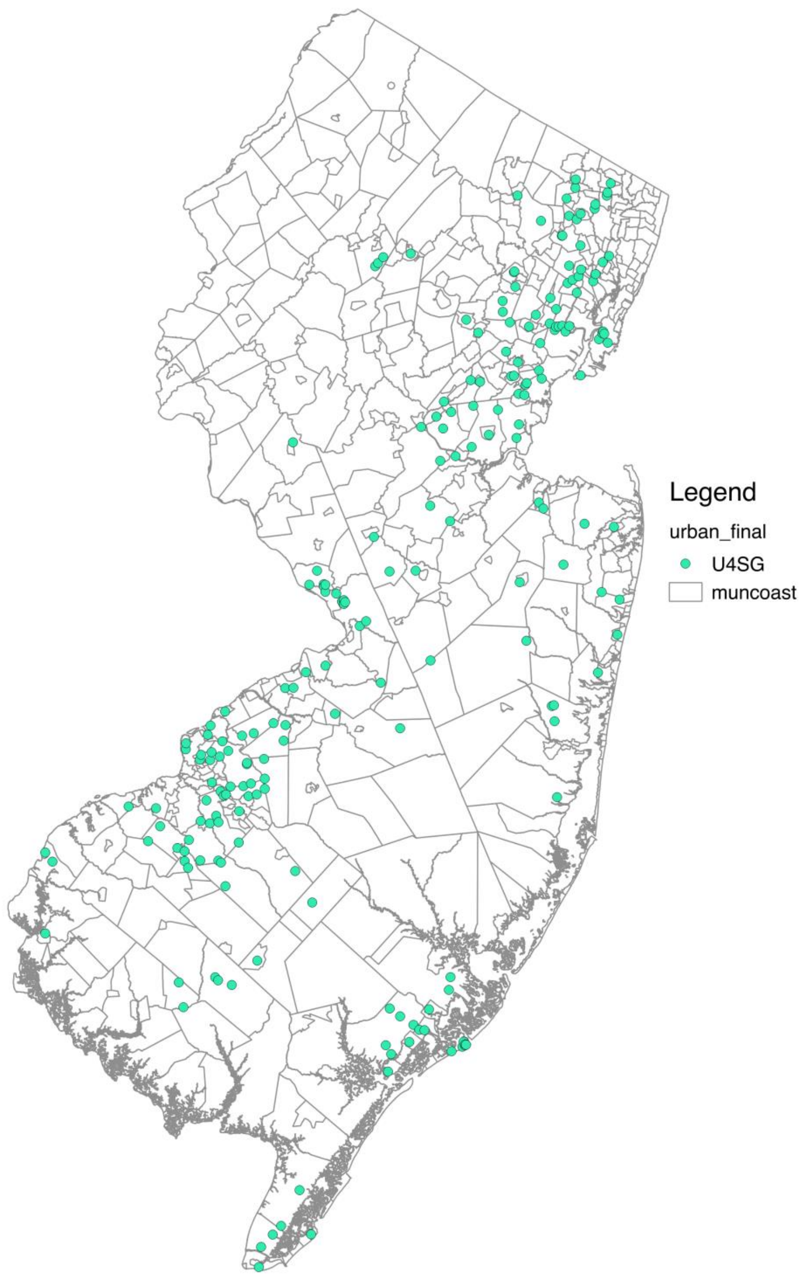


Figure 52: The spatial distribution of U4SG intersections

Calibration Results

Chapter 12 of the HSM provides a crash prediction model for R2 intersections. The crashes are grouped based on collision type as multi-vehicle, single-vehicle, vehicle-pedestrian, and vehicle-bicycle collisions. The base SPFs for U3ST, U4ST, U4ST, and U4SG multi- and single-vehicle collisions include the $AADT_{maj}$ and $AADT_{min}$ variables, as shown in Table 47 and Table 48.

Table 47: Multi-vehicle collision SPFs for urban intersections in the HSM

Type	SPF	Reliable AADT Range
U3ST	$N_{bimv\ 3ST} = exp[-13.36 + 1.11 \times ln(AADT_{maj}) + 0.41 \times ln(AADT_{min})]$	$AADT_{maj} \in [0{,}45700]$ $AADT_{min} \in [0{,}9300]$
U3SG	$N_{bimv\ 3SG} = exp[-12.13 + 1.11 \times ln(AADT_{maj}) + 0.26 \times ln(AADT_{min})]$	$AADT_{maj} \in [0{,}46800]$ $AADT_{min} \in [0{,}5900]$
U4ST	$N_{bimv\ 4ST} = exp[-8.90 + 0.82 \times ln(AADT_{maj}) + 0.25 \times ln(AADT_{min})]$	$AADT_{maj} \in [0{,}58100]$ $AADT_{min} \in [0{,}16400]$
U4SG	$N_{bimv\ 4SG} = exp[-10.99 + 1.07 \times ln(AADT_{maj}) + 0.23 \times ln(AADT_{min})]$	$AADT_{maj} \in [0{,}80200]$ $AADT_{min} \in [0{,}49100]$

Table 48: Single-vehicle collision SPFs for urban intersections in the HSM

Type	SPF	Reliable AADT Range
U3ST	$N_{bisv\ 3ST} = exp[-6.81 + 0.16 \times ln(AADT_{maj}) + 0.51 \times ln(AADT_{min})]$	$AADT_{maj} \in [0{,}45700]$ $AADT_{min} \in [0{,}9300]$
U3SG	$N_{bisv\ 3SG} = exp[-9.02 + 0.42 \times ln(AADT_{maj}) + 0.40 \times ln(AADT_{min})]$	$AADT_{maj} \in [0{,}46800]$ $AADT_{min} \in [0{,}5900]$
U4ST	$N_{bisv\ 4ST} = exp[-5.33 + 0.33 \times ln(AADT_{maj}) + 0.12 \times ln(AADT_{min})]$	$AADT_{maj} \in [0{,}58100]$ $AADT_{min} \in [0{,}16400]$
U4SG	$N_{bisv\ 4SG} = exp[-10.21 + 0.68 \times ln(AADT_{maj}) + 0.27 \times ln(AADT_{min})]$	$AADT_{maj} \in [0{,}80200]$ $AADT_{min} \in [0{,}49100]$

$$N_{spf\ int} = N_{bimv} + N_{bisv} \tag{29}$$

Where, N_{bimv} and N_{bisv} are the predicted average number of multiple-vehicle and single-vehicle collisions for base conditions, respectively.

$$N_{bi} = N_{spf\ int}.CMF_1 \ldots CMF_6 \tag{30}$$

Where, N_{bi} is the predicted average crash frequency of an intersection excluding vehicle-pedestrian and vehicle-bicycle collisions, and CMF_1 through CMF_6 are the CMFs for intersection left-turn lanes, left-turn signal phasing, right-turn lanes, right-turn-on-red, lighting, and red-light cameras, respectively.

Vehicle-pedestrian collisions SPFs for each intersection type are presented in Table 49.

Table 49: Vehicle-pedestrian collision SPFs for urban intersections in the HSM

Type	SPF
3SG	$N_{pedbase} = exp[-6.6 + 0.05. ln(AADT_{TOT}) + 0.24. ln(AADT_{min}/AADT_{TOT})] + 0.41. \ln(PedVol) + 0.09. n_{lanesx}$
4SG	$N_{pedbase} = exp[-9.53 + 0.4. ln(AADT_{TOT}) + 0.26. ln(AADT_{min}/AADT_{TOT})] + 0.45. \ln(PedVol) + 0.04. n_{lanesx}$
3ST	$N_{pedi} = N_{bi}. 0.021$
4ST	$N_{pedi} = N_{bi}. 0.022$
	Note: $AADT_{TOT} = AADT_{maj} + AADT_{min}$, $PedVol$ is the daily pedestrian volume crossing all intersection legs, and n_{lanesx} is the maximum number of traffic lanes crossed by pedestrians.

Note that CMF for vehicle-pedestrian collisions only apply at signalized intersections.

$$N_{pedi} = N_{pedbase}. CMF_1. CMF_2. CMF_3 \quad (31)$$

Where, CMF_1 is for the number of bus stops near an intersection, CMF_2 is for presence of a school near an intersection, and CMF_3 is for the presence of alcohol sales establishments near an intersection.

Table 50: Vehicle-bicycle collision SPFs for urban intersections in the HSM

Type	SPF = $N_{bi}. f_{bike}$
3ST	$N_{bike} = N_{bi}. 0.016$
3SG	$N_{bike} = N_{bi}. 0.011$
4SG	$N_{bike} = N_{bi}. 0.018$
4ST	$N_{bike} = N_{bi}. 0.015$

The total number of predicted crashes for an intersection, $N_{predicted\ int}$, can then be calculated as:

$$N_{predicted\ int} = N_{bi} + N_{pedi} + N_{bike} \quad (32)$$

The results of the calibration factors are shown in Table 51**.** The Calibrator tool developed by the FHWA was used to calculate the calibration factors and measure their goodness of fit [77]. According to the FHWA report, a reasonable upper threshold for the coefficient of variation of a calibration factor is 0.10 to 0.15. In that respect, the results shown in Table 51 are found to be acceptable.

Table 51: Calibration factors of urban and suburban Intersections

Type	Calibration Factor	Standard Error	Coefficient of Variation
U3ST	2.61	±0.29	0.11
U3SG	3.60	±0.36	0.10
U4ST	1.66	±0.25	0.15
U4SG	4.25	±0.40	0.09

The CURE plots for U3ST, U3SG, U4ST, and U4SG type intersections with respect to major and minor AADTs are shown in Figure 53 through Figure 56.

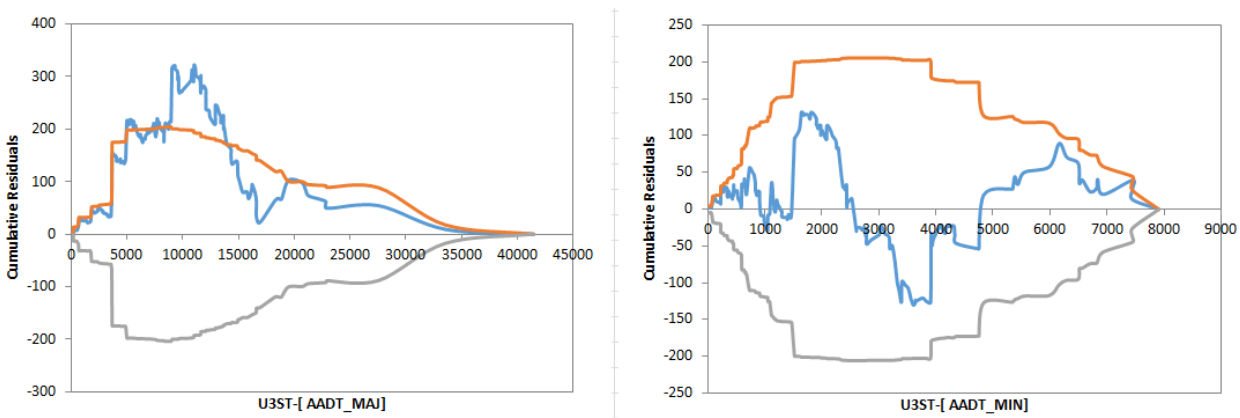


Figure 53: CURE plots for U3ST

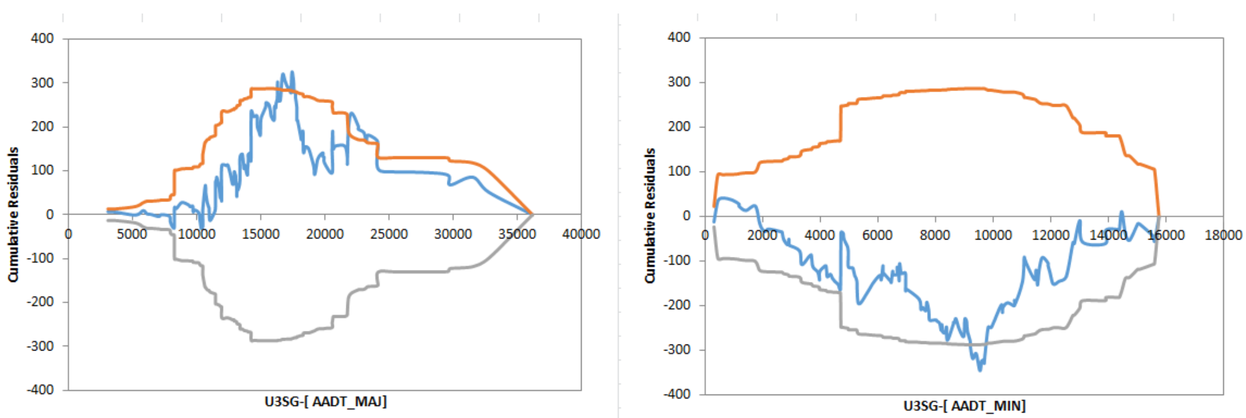


Figure 54: CURE plots for U3SG

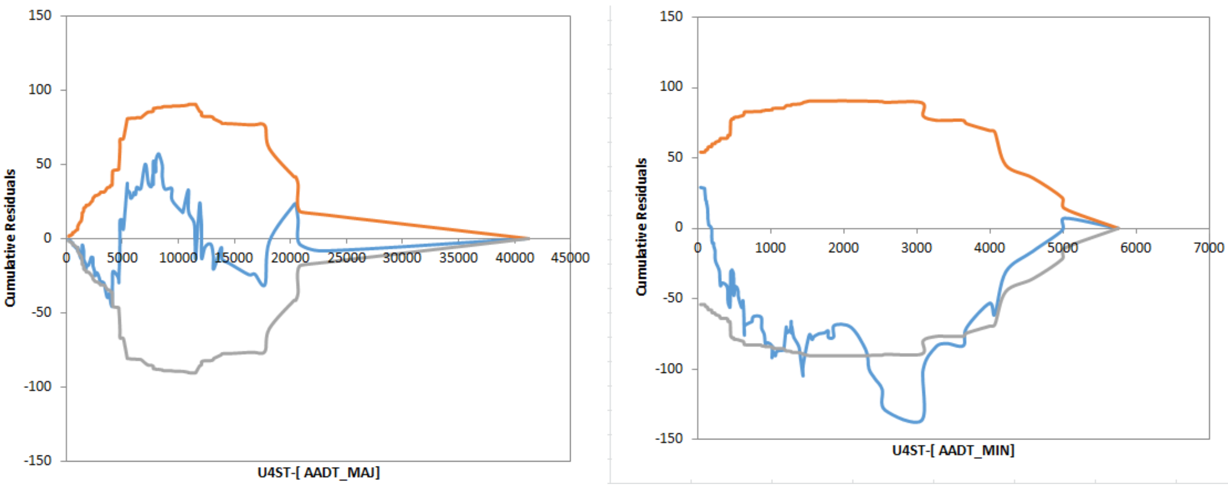


Figure 55: CURE plots for U4ST

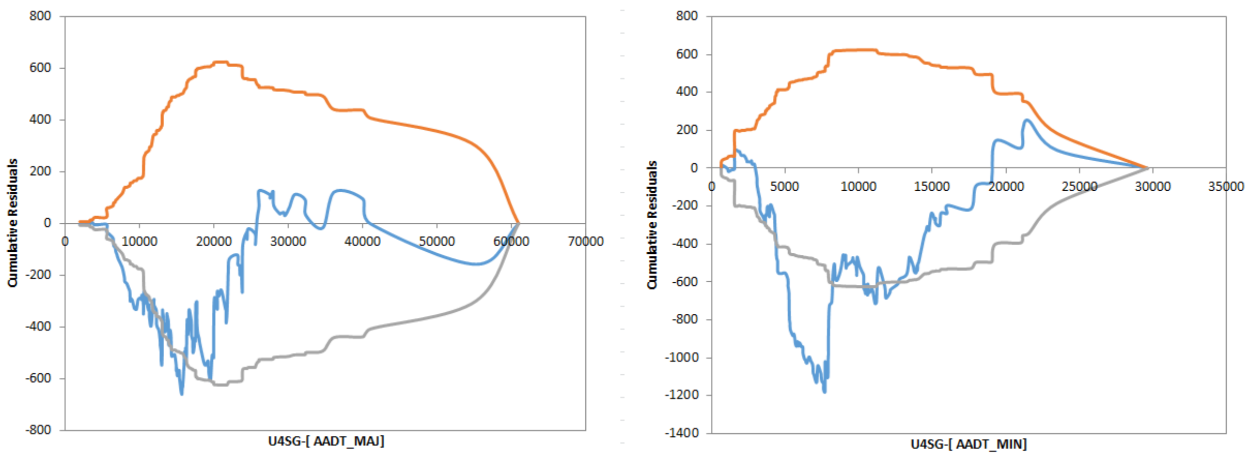


Figure 56: CURE plots for U4SG

It can be observed that with the exception of U4ST, the cumulative residuals tend to stray from the acceptable lower and upper bounds. In addition, the calibration factors for U3SG and U4SG are significantly high. The results suggest that the development of SPFs is warranted. The following subsection presents the SPF development results.

Development Results

The SPFs for urban and suburban intersection multi-vehicle and single-vehicle collisions were developed using the available datasets, based on the negative binomial model suggested by the HSM. The model estimation was performed in R statistical package. The SPF development equations for total, multi-vehicle and single-vehicle collisions for various intersection types are shown in Table 52. The statistical results of each developed function for multi-vehicle and single-vehicle collisions at U3ST, U4ST, U3SG, and U4SG are presented in Table 52 through Table 56.

Table 52: SPF functions developed for total, multi and single vehicle collisions

Segment	Crash Type	SPF
U3ST	Total	$N_{TOT\ U3ST} = exp[-5.855 + 0.434. ln(AADT_{maj}) + 0.384. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U3ST} = exp[-6.892 + 0.483. ln(AADT_{maj}) + 0.429. ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U3ST} = exp[-4.895 + 0.283. ln(AADT_{maj}) + 0.219. ln(AADT_{min})]$
U3SG	Total	$N_{TOT\ U3SG} = exp[-7.553 + 0.693. ln(AADT_{maj}) + 0.321. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U3SG} = exp[-8.019 + 0.713. ln(AADT_{maj}) + 0.336. ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U3SG} = exp[-8.093 + 0.676. ln(AADT_{maj}) + 0.141. ln(AADT_{min})]$
U4ST	Total	$N_{TOT\ U4ST} = exp[-8.269 + 0.743. ln(AADT_{maj}) + 0.343. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U4ST} = exp[-8.959 + 0.752. ln(AADT_{maj}) + 0.392 ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U4ST} = exp[-8.359 + 0.724. ln(AADT_{maj}) + 0.142. ln(AADT_{min})]$
U4SG	Total	$N_{TOT\ U4SG} = exp[-9.593 + 0.968. ln(AADT_{maj}) + 0.308. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U4SG} = exp[-10.307 + 1.022. ln(AADT_{maj}) + 0.317. ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U4SG} = exp[-5.804 + 0.424. ln(AADT_{maj}) + 0.173. ln(AADT_{min})]$

Table 53: SPF development results for total, multi and single vehicle collisions at U3ST

	Variable	Estimate	Std Error	z-value	Pr(> |z|)
Total	Intercept	-5.85500	0.43892	-13.340	<2e-16 ***
	Log($AADT_{MAJ}$)	0.43426	0.05039	8.618	<2e-16 ***
	Log($AADT_{MIN}$)	0.38403	0.03590	10.696	<2e-16 ***
	k	0.883	df: 1079 AIC: 4220.5		
Multi-Vehicle	Intercept	-6.89172	0.51519	-13.377	<2e-16 ***
	Log($AADT_{MAJ}$)	0.48314	0.05820	8.301	<2e-16 ***
	Log($AADT_{MIN}$)	0.42899	0.04076	10.525	<2e-16 ***
	k	1.085	df: 1079 AIC: 3772.3		
Single-Vehicle	Intercept	-4.89471	0.66523	-7.358	1.87e-13 ***
	Log($AADT_{MAJ}$)	0.28321	0.07801	3.630	0.000283 ***
	Log($AADT_{MIN}$)	0.21852	0.05616	3.891	9.99e-05 ***
	k	1.238	df: 1879 AIC: 1926		

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.

Table 54: SPF development results for total, multi and single vehicle collisions at U3SG

	Variable	Estimate	Std Error	z-value	Pr(> |z|)
Total	Intercept	-7.55286	0.72757	-10.381	<2e-16 ***
	Log($AADT_{MAJ}$)	0.69293	0.07706	8.992	<2e-16 ***
	Log($AADT_{MIN}$)	0.32068	0.04378	7.325	2.39e-13 ***
	k	0.417	df: 784 AIC: 4550.3		
Multi-Vehicle	Intercept	-8.01883	0.77259	-10.379	<2e-16***
	Log($AADT_{MAJ}$)	0.71255	0.08158	8.734	<2e-16***
	Log($AADT_{MIN}$)	0.33593	0.04646	7.231	4.79e-13***
	k	0.458	df: 784 AIC: 4377.3		
Single-Vehicle	Intercept	-8.09349	1.36612	-5.924	3.13e-09 ***
	Log($AADT_{MAJ}$)	0.67635	0.14129	4.787	1.69e-06 ***
	Log($AADT_{MIN}$)	0.14110	0.07929	1.779	0.0752 .
	k	0.453	df: 784 AIC: 1799.6		

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.

Table 55: SPF development results for total, multi and single vehicle collisions at U4ST

	Variable	Estimate	Std Error	z-value	Pr(> |z|)
Total	Intercept	-8.26895	0.54191	-15.26	<2e-16***
	Log($AADT_{MAJ}$)	0.74256	0.05840	12.71	<2e-16***
	Log($AADT_{MIN}$)	0.34296	0.04303	7.97	1.59E-15 ***
	k	0.507	df: 569 AIC: 1892.7		
Multi-Vehicle	Intercept	-8.9590	0.59658	-15.017	<2e-16***
	Log($AADT_{MAJ}$)	0.75245	0.06390	11.775	<2e-16***
	Log($AADT_{MIN}$)	0.39214	0.04633	8.464	<2e-16***
	k	0.502	df: 569 AIC: 1687.6		
Single-Vehicle	Intercept	-8.35851	1.09410	-7.640	2.18e-14 ***
	Log($AADT_{MAJ}$)	0.72436	0.11811	6.133	8.64e-10 ***
	Log($AADT_{MIN}$)	0.14217	0.08641	1.645	0.0999 .
	k	1.832	df: 569 AIC: 833.23		

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.

Table 56: SPF development results for total, multi and single vehicle collisions at U4SG

	Variable	Estimate	Std Error	z-value	Pr(> |z|)
Total	Intercept	-9.59343	0.44441	-21.587	<2e-16***
	$Log(AADT_{MAJ})$	0.96824	0.05427	17.842	<2e-16***
	$Log(AADT_{MIN})$	0.30818	0.04068	7.576	3.57e-14 ***
	k	0.338	df: 969 AIC: 6388.1		
Multi-Vehicle	Intercept	-10.30717	0.45984	-22.415	<2e-16 ***
	$Log(AADT_{MAJ})$	1.02248	0.05580	18.325	<2e-16 ***
	$Log(AADT_{MIN})$	0.31697	0.04175	7.593	3.14e-14***
	k	0.347	df: 969 AIC: 6199.9		
Single-Vehicle	Intercept	-5.80447	0.85712	-6.772	1.27e-11 ***
	$Log(AADT_{MAJ})$	0.42360	0.10523	4.026	5.68e-05 ***
	$Log(AADT_{MIN})$	0.17335	0.08069	2.148	0.0317 *
	k	0.470	df: 969 AIC: 2407.7		

Signif. Codes: 0.001:***; 0.01:**; 0.05:*; 0.1:.

The $AADT_{MAJ}$ and $AADT_{MIN}$ ranges for U3ST are [0 – 44,750 veh/day] and [0 – 15,250 veh/day], respectively; for U3SG are [0 – 36,250 veh/day] and [0 – 20,000 veh/day], respectively; for U4ST are [0 – 41,250 veh/day] and [0 – 11,500 veh/day], respectively, and for U4SG are [0 – 62,500 veh/day] and [0 – 31,750 veh/day], respectively.

Table 57 presents f_{bike}, the vehicle-bicycle collision adjustment factor calculated using the 2011 to 2015 NJ crash database.

Table 57: SPF development for vehicle-bicycle collisions at urban and suburban intersections

Type	Bicycle Crash Adjustment Factors f_{bike}
U3ST	0.011
U3SG	0.008
U4ST	0.009
U4SG	0.007

Table 58 presents the developed SPFs for vehicle-pedestrian collisions at urban and suburban intersections. Note that the values for the variable *PedVol*, i.e., daily pedestrian volume, used in the estimation process was adopted from the default values given in the HSM, since pedestrian volumes are not available for most intersections.

The statistical results of the developed SPFs for the vehicle pedestrian collisions at U3SG and U4SG are presented in Table 59.

Table 58: SPF development for urban intersections – PV

Type	SPF/Pedestrian Crash Adjustment Factor
U3ST	0.0134
U3SG	$N_{ped\ U3SG} = exp[-18.636 + 1.145 \times ln(AADT_{tot}) + 0.0898 \times PedVol]$
U4ST	0.0142
U4SG	$N_{ped\ U4SG} = exp[-18.935 + 1.245.\,ln(AADT_{tot}) + 0.874.\ln(PedVo)l - 0.135.\,MaxLanes]$

Table 59: SPF development results for vehicle-pedestrian collisions at U3SG and U4SG

	Variable	Estimate	Std Error	z-value	Pr(> |z|)
U3SG	Intercept	-18.6357	3.5278	-5.283	1.27e-07 ***
	Log($AADT_{TOT}$)	1.1446	0.3498	3.272	0.00107 **
	Log(PedVol)	0.8982	0.1438	6.248	4.15e-10 ***
	k	1.16	df: 969 AIC: 585.52		
U4SG	Intercept	-18.93484	2.03255	-9.316	< 2e-16 ***
	Log($AADT_{TOT}$)	1.24547	0.20014	6.223	4.88e-10 ***
	Log(PedVol)	0.87426	0.10465	8.354	< 2e-16 ***
	MaxLanes	-0.13539	0.06785	-1.995	0.046 *
	k	1.35	df: 969 AIC: 1243.8		

CONCLUSIONS

SPFs in the HSM were developed using historic crash data collected over a number of years at sites of the same facility type in different states. Because the SPFs provided in the HSM are developed using data from various states, it is more than likely that they cannot be transferred directly to other locations and times. Thus, the HSM's predictive model often needs to be calibrated to capture local state or geographic conditions. Moreover, accident frequencies for similar facility types can also vary from one jurisdiction to another, since their locations differ in climate, driver population and characteristics, accident reporting threshold, accident reporting practices, and other contributing factors.

To make the SPFs better accommodate the local data, two strategies are usually undertaken:

- The first strategy is to calibrate SPFs provided in the HSM so that the contents of HSM can be fully leveraged.
- The second strategy is to develop location-specific SPFs, regardless of the predictive modeling framework in the HSM.

The main objective of this research project is to either (1) calibrate the SPFs provided in the HSM using New Jersey (NJ) data or (2) develop new, NJ-specific SPFs. The facility types considered for this research project include segments and intersections of rural two-lane two-way, rural multilane, and urban and suburban roads.

Calibrating the SPFs used in the predictive models of the HSM requires data from a limited number of sites (for each facility type) from NJ, using the methods suggested in Part C of the HSM. Developing NJ-specific SPFs would provide more accurate results, but requires data from a larger sample of sites, and also involves the application of generalized linear models.

The research team completed several tasks to achieve the main project objective:

- Conducted an in-depth review of the studies in the literature that focused on the calibration of the SPFs used in the HSM and the development of new SPFs, and identified the related calibration and development issues.
- Identified the key sources of data required for calibration and development of SPFs. These include roadway characteristics data, traffic volume data, and crash data.
- Developed a computer code to read and process the compiled database to (a) filter out inconsistent data entries, (b) identify facility types, (c) execute the roadway segmentation process, (d) assign crash statistics for each facility, and (e) generate a complete database for each facility type to be used in calibration and/or development of SPFs.
- Provided recommendations and activities that include:
  (a) Improvements to data collection and recording practices that would facilitate easier data extraction required for the SPF calibration/development process.

(b) A workshop that demonstrates the step-by-step approach for using the SPFs for the NJDOT staff and other interested parties.

As mentioned in the Available Data Sources section, existing data are grouped into three categories by type: (1) Traffic Volume, (2) Roadway Features, and (3) Roadway Crashes. Information regarding each data source is summarized in Table 4.

Traffic volume data include the sensor database, maintained by the New Jersey Traffic Monitoring Program at NJDOT, and hourly TMC collected at various intersections.

Roadway features data were extracted from three data sources: the Straight Line Diagrams (SLD) database, Geographic Information Systems (GIS) maps, and Google StreetView. NJDOT's SLD database is the richest source of information for roadway features. This database was provided by NJDOT in MS Access™ format. It includes various tables on different geometric and operational aspects of NJ roadways, as shown in Table 5.

The crash data were provided by NJDOT for the years 2011 to 2015. Table 6 lists the key data elements used as part of this study.

Once the available datasets were gathered and cleaned, the research team developed a computer code in R programming language to read and process the compiled database to (a) filter out inconsistent data entries, (b) identify facility types, (c) execute the roadway segmentation process, (d) assign crash statistics for each facility, and (e) generate a complete database for each facility type to be used in calibration and/or development of SPFs.

The generic procedure for generating the intersection database per facility type is shown in Figure 12. Generating the homogeneous roadway database for each facility was a more challenging process than that for the intersection database.

A homogeneous segment starts at an intersection or at any point where various geometric and operational features of the roadway change at either direction of the facility. The generic procedure for generating homogeneous database per facility type is shown in Figure 13.

In addition to the automatically generated data, the research team conducted an extensive data extraction process to meet the data requirements of the HSM, as described within the respective section for each facility type.
Furthermore, the research team developed a novel clustering method for automatically estimating horizontal curvature data and CMFs using GIS roadway shapefiles, as presented in the Rural Two-Lane Two-Way Segments section.

Table 60 presents the sample size for each facility type used for calibration and/or development of SPFs for segments and intersections separately.

Table 60: Sample size of each facility type used for calibration and development

SEGMENTS		
Facility Type	**Sample Size**	**SPF**
R2U	756	Calibration and Development
R4U	32	Calibration
R4D	34	Calibration
U2U	459	Calibration and Development
U3T	n/a	n/a
U4U	514	Calibration and Development
U4D	387	Calibration and Development
U5T	n/a	n/a
INTERSECTIONS		
Facility Type	**Sample Size**	**SPF**
R23ST	314	Calibration and Development
R24ST	149	Calibration and Development
R23SG	15	Calibration
R24SG	45	Calibration and Development
RM3ST	3	n/a
RM4ST	1	n/a
RM3SG	0	n/a
RM4SG	6	n/a
U3ST	227	Calibration and Development
U4ST	121	Calibration and Development
U3SG	164	Calibration and Development
U4SG	209	Calibration and Development

It can be seen in Table 60 that among the original facility types, two segment types, namely U3T and U5T, and rural multilane intersections could not be included in the calibration and development process due to lack of data.

In order to calculate a calibration factor, the observed crash frequency and the predicted crash frequency for each intersection and segment are required. The observed crash frequency was calculated for each segment as explained in the Processing Data section. The predicted crash frequency can be calculated using the SPF and the corresponding CMF values given in the HSM.

The Calibrator tool developed by the FHWA was used to calculate the calibration factors and measure their goodness of fit [77].

Table 61 presents the calibration factors calculated for each facility type.

Table 61: Calibration factors for each facility type

SEGMENTS			
Facility Type	**Calibration Factor**	**Standard Error**	**Coefficient of Variation**
R2U	1.55	±0.12	0.08
R4U	1.12	±0.42	0.38
R4D	1.70	±0.80	0.47
U2U	1.264	±0.14	0.11
U3T	n/a	n/a	n/a
U4U	1.097	±0.15	0.13
U4D	1.596	±0.21	0.13
U5T	n/a	n/a	n/a
INTERSECTIONS			
Facility Type	**Sample Size**		
R23ST	0.88	±0.08	0.09
R24ST	0.88	±0.11	0.13
R23SG*	-	-	-
R24SG	0.85	±0.16	0.18
RM3ST	n/a	n/a	n/a
RM4ST	n/a	n/a	n/a
RM3SG	n/a	n/a	n/a
RM4SG	n/a	n/a	n/a
U3ST	2.61	±0.29	0.11
U3SG	3.60	±0.36	0.10
U4ST	1.66	±0.25	0.15
U4SG	4.25	±0.40	0.09

Note: *R23SG intersection type is not included in the HSM.

In addition, to assess the validity of the calculated calibration factors, the team used the CURE plots, which are simply graphs of the cumulative residuals (observed minus predicted crashes) against variables of interest. The residuals between the estimated and observed values are assumed to be independent random variables. It is expected that the CURE plots should be within the expected limits of an unbiased random walk, i.e., plus/minus two standard deviations. CURE plots for each facility type are presented in their respective sections throughout the report.

Using the same data used for calibration, the research team developed NJ-specific SPFs for facilities with a sufficient number of data points using the crash data from 2011 to 2015. SPFs were estimated based on the negative binomial model suggested by the HSM. The model estimation was performed in R statistical package.

Table 62 and Table 63 present the developed SPFs for segments and intersections per facility type and collision type, where applicable, respectively.

Table 62: Developed SPFs for segments

Segment	Crash Type	Developed SPFs
R2U	Total	$N_{TOT\ R2U} = exp[-6.41 + 0.83. ln(AADT) + 0.86. ln(L)]$
U2U	Total	$N_{TOT\ U2U} = exp[-9.798 + 1.188. ln(AADT) + ln(L)]$
	Multi-Vehicle	$N_{MV\ U2U} = exp[-14.411 + 1.641. ln(AADT) + ln(L)]$
	Single-Vehicle	$N_{SV\ U2U} = exp[-3.977 + 0.435. ln(AADT) + ln(L)]$
U4U	Total	$N_{TOT\ U4U} = exp[-12.01 + 1.432. ln(AADT) + ln(L)]$
	Multi-Vehicle	$N_{MV\ U4U} = exp[-13.794 + 1.59. ln(AADT) + ln(L)]$
	Single-Vehicle	$N_{SV\ U4U} = exp[-6.961 + 0.751. ln(AADT) + ln(L)]$
U4D	Total	$N_{TOT\ U4D} = exp[-3.00 + 0.543. ln(AADT) + ln(L)]$
	Multi-Vehicle	$N_{MV\ U4D} = exp[-3.363 + 0.558. ln(AADT) + ln(L)]$
	Single-Vehicle	$N_{SV\ U4D} = exp[-4.687 + 0.543. ln(AADT) + ln(L)]$

Table 63: Developed SPFs for intersections

Intersection	Crash Type	Developed SPFs
R23ST	Total	$N_{spf\ 3ST} = exp[-6.139 + 0.498. ln(AADT_{maj}) + 0.296. ln(AADT_{min})]$
R23SG	Total	$N_{spf\ 3SG} = exp[-12.140 + 1.184. ln(AADT_{maj}) + 0.281. ln(AADT_{min})]$
R24ST	Total	$N_{spf\ 4ST} = exp[-3.716 + 0.159. ln(AADT_{maj}) + 0.426. ln(AADT_{min})]$
R24SG	Total	$N_{spf\ 4SG} = exp[-5.811 + 0.345. ln(AADT_{maj}) + 0.526. ln(AADT_{min})]$
U3ST	Total	$N_{TOT\ U3ST} = exp[-5.855 + 0.434. ln(AADT_{maj}) + 0.384. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U3ST} = exp[-6.892 + 0.483. ln(AADT_{maj}) + 0.429. ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U3ST} = exp[-4.895 + 0.283. ln(AADT_{maj}) + 0.219. ln(AADT_{min})]$
U3SG	Total	$N_{TOT\ U3SG} = exp[-7.553 + 0.693. ln(AADT_{maj}) + 0.321. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U3SG} = exp[-8.019 + 0.713. ln(AADT_{maj}) + 0.336. ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U3SG} = exp[-8.093 + 0.676. ln(AADT_{maj}) + 0.141. ln(AADT_{min})]$
	Vehicle-Pedestrian	$N_{ped\ U3SG} = exp[-18.636 + 1.145 \times ln(AADT_{tot}) + 0.0898 \times PedVol]$
U4ST	Total	$N_{TOT\ U4ST} = exp[-8.269 + 0.743. ln(AADT_{maj}) + 0.343. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U4ST} = exp[-8.959 + 0.752. ln(AADT_{maj}) + 0.392 ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U4ST} = exp[-8.359 + 0.724. ln(AADT_{maj}) + 0.142. ln(AADT_{min})]$
U4SG	Total	$N_{TOT\ U4SG} = exp[-9.593 + 0.968. ln(AADT_{maj}) + 0.308. ln(AADT_{min})]$
	Multi-Vehicle	$N_{MV\ U4SG} = exp[-10.307 + 1.022. ln(AADT_{maj}) + 0.317. ln(AADT_{min})]$
	Single-Vehicle	$N_{SV\ U4SG} = exp[-5.804 + 0.424. ln(AADT_{maj}) + 0.173. ln(AADT_{min})]$
	Vehicle-Pedestrian	$N_{ped\ U4SG} = exp[-18.935 + 1.245. ln(AADT_{tot}) + 0.874. \ln(PedVo)l - 0.135. MaxLanes]$

The calculated calibration factors and the developed SPFs are being embedded in the safety analysis spreadsheets used by the NJDOT staff. These spreadsheets were modified so that users can either select the SPFs provided in the HSM and apply the calculated calibration factors, or simply use the NJ-specific SPFs developed by the research team.

REFERENCES

1. National Highway Transportation Safety Administration, *Traffic Safety Facts 2013*. 2015.
2. AASHTO, *Highway Safety Manual.* Washington, DC, 2010. 529.
3. Srinivasan, R., Carter, D., and Bauer, K., *Safety Performance Function Decision Guide: SPF Calibration vs SPF Development*. 2013, Report FHWASA-14-004. Federal Highway Administration, Washington, DC.
4. Lord, D., and Bonneson, J.A., *Role and application of accident modification factors within highway design process.* Transportation Research Record: Journal of the Transportation Research Board, 2006. 1961(1): p. 65-73.
5. Shin, H., Lee, Y.J., and Dadvar, S., *The Development of Local Calibration Factors for Implementing the Highway Safety Manual in Maryland*. Publication SP209B4J, Office of Traffic and Safety (OOTS), Maryland State Highway Administration, 2014.
6. Dixon, K., et al., *Calibrating the Highway Safety Manual Predictive Methods for Oregon Rural State Highways*. 2011.
7. Abdel-Rahim, A., et al., *Calibration and Development of Safety Performance Functions for Rural Highway Facilities in Idaho*. 2014.
8. Sun, C., et al., *Calibration of the Highway Safety Manual for Missouri*. 2014.
9. Robicheaux, B., and Wolshon, B., *Calibration of the Louisiana Highway Safety Manual*. 2015.
10. Mehta, G., and Lou, Y., *Safety Performance Function Calibration and Development for the State of Alabama: Two-Lane Two-Way Rural Roads and Four-Lane Divided Highways*, in *92nd Transportation Research Board Annual Meeting, Washington, DC*. 2013.
11. St-Aubin, P., Saunier, N., and Miranda-Moreno, L.F., *Road User Collision Prediction Using Motion Patterns Applied to Surrogate Safety Analysis.* Transportation Research Board 93rd Annual Meeting, 2014(14-5363).
12. Srinivasan, R., and Carter, D.L., *Development of Safety Performance Functions for North Carolina*. 2011, North Carolina Department of Transportation, Research and Analysis Group Chapel Hill, NC.
13. Srinivasan, S., et al., *Development and Calibration of Highway Safety Manual Equations for Florida Conditions*. 2011.
14. Jones, B., Janssen, L., and Mannering, F., *Analysis of the frequency and duration of freeway accidents in Seattle.* Accident Analysis and Prevention, 1991. 23(4): p. 239-255.
15. Miaou, S.P., and Lum, H., *Modeling vehicle accidents and highway geometric design relationships.* Accident Analysis and Prevention, 1993. 25(6): p. 689-709.
16. Abdel-Aty, M.A., and Radwan, A.E., *Modeling traffic accident occurrence and involvement.* Accident Analysis and Prevention, 2000. 32(5): p. 633-642.
17. Miaou, S.P., *The relationship between truck accidents and geometric design of road sections – Poisson versus negative binomial regressions.* Accident Analysis and Prevention, 1994. 26(4): p. 471-482.

18. Poch, M., and Mannering, F., *Negative binomial analysis of intersection-accident frequencies.* Journal of Transportation Engineering-Asce, 1996. 122(2): p. 105-113.
19. Joshua, S.C., and Garber, N.J., *Estimating truck accident rate and involvements using linear and Poisson regression models.* Transportation Planning and Technology, 1990. 15(1): p. 41-58.
20. Maher, M.J., and Summersgill, I., *A comprehensive methodology for the fitting of predictive accident models.* Accident Analysis and Prevention, 1996. 28(3): p. 281-296.
21. Chin, H.C., and Quddus, M.A., *Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections.* Accident Analysis and Prevention, 2003. 35(2): p. 253-259.
22. Shankar, V., et al., *Evaluating median crossover likelihoods with clustered accident counts: An empirical inquiry using the random effects negative binomial model.* Transportation Research Record, 1998(1635): p. 44-48.
23. Wang, X., et al., *A systematic approach for hazardous intersection identification and countermeasure development.* Journal of Transportation Engineering, 2014. 140(6): p. 0401-4022.
24. Anastasopoulos, P.C., and Mannering, F.L., *A note on modeling vehicle accident frequencies with random-parameters count models.* Accident Analysis and Prevention, 2009. 41(1): p. 153-159.
25. El-Basyouny, K., and Sayed, T., *Accident prediction models with random corridor parameters.* Accident Analysis and Prevention, 2009. 41(5): p. 1118-1123.
26. Venkataraman, N.S., et al., *Model of relationship between interstate crash occurrence and geometrics.* Transportation Research Record: Journal of the Transportation Research Board, 2011. 2236(1): p. 41-48.
27. Venkataraman, N., Ulfarsson, G.F., and Shankar, V.N., *Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type.* Accident Analysis and Prevention, 2013. 59: p. 309-318.
28. Venkataraman, N., et al., *A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies.* Analytic Methods in Accident Research, 2014. 2(0): p. 12-20.
29. Huang, H., and Abdel-Aty, M., *Multilevel data and Bayesian analysis in traffic safety.* Accident Analysis and Prevention, 2010. 42(6): p. 1556-1565.
30. Ahmed, M., et al., *Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway.* Accident Analysis and Prevention, 2011. 43(4): p. 1581-1589.
31. Kim, D.G., et al., *Modeling crash outcome probabilities at rural intersections: Application of hierarchical binomial logistic models.* Accident Analysis and Prevention, 2007. 39(1): p. 125-134.
32. Lenguerrand, E., Martin, J.L., and Laumon, B., *Modelling the hierarchical structure of road crash data – Application to severity analysis.* Accident Analysis and Prevention, 2006. 38(1): p. 43-53.

33. Jones, A.P., and Jorgensen, S.H., *The use of multilevel models for the prediction of road accident outcomes.* Accident Analysis and Prevention, 2003. 35(1): p. 59-69.
34. Xie, K., et al., *Corridor-level signalized intersection safety analysis in Shanghai, China using Bayesian hierarchical models.* Accident Analysis and Prevention, 2013. 50: p. 25-33.
35. Chen, Y., and Persaud, B., *Methodology to develop crash modification functions for road safety treatments with fully specified and hierarchical models.* Accident Analysis & Prevention, 2014. 70(0): p. 131-139.
36. Gelman, A., and Hill, J., *Data analysis using regression and multilevel/hierarchical models*. Analytical Methods for Social Research. 2007, Cambridge, New York: Cambridge University Press. xxii, 625 p.
37. Abdel-Aty, M., and Wang, X., *Crash estimation at signalized intersections along corridors: Analyzing spatial effect and identifying significant factors.* Transportation Research Record, 2006(1953): p. 98-111.
38. Lord, D., and Persaud, B.N., *Accident prediction models with and without trend – Application of the generalized estimating equations procedure.* Highway and Traffic Safety: Crash Data, Analysis Tools, and Statistical Methods, 2000(1717): p. 102-108.
39. Wang, X., and Abdel-Aty, M., *Temporal and spatial analyses of rear-end crashes at signalized intersections.* Accident Analysis and Prevention, 2006. 38(6): p. 1137-1150.
40. Hutchings, C.B., Knight, S., and Reading, J.C., *The use of generalized estimating equations in the analysis of motor vehicle crash data.* Accident Analysis and Prevention, 2003. 35(1): p. 3-8.
41. Guo, F., Wang, X.S., and Abdel-Aty, M.A., *Modeling signalized intersection safety with corridor-level spatial correlations.* Accident Analysis and Prevention, 2010. 42(1): p. 84-92.
42. Aguero-Valverde, J., and Jovanils, P.P., *Analysis of road crash frequency with spatial models.* Transportation Research Record, 2008(2061): p. 55-63.
43. El-Basyouny, K., and Sayed, T., *Urban arterial accident prediction models with spatial effects.* Transportation Research Record, 2009(2102): p. 27-33.
44. Song, J.J., et al., *Bayesian multivariate spatial models for roadway traffic crash mapping.* Journal of Multivariate Analysis, 2006. 97(1): p. 246-273.
45. Narayanamoorthy, S., Paleti, R., and Bhat, C.R., *On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level.* Transportation Research Part B: Methodological, 2013. 55: p. 245-264.
46. Tegge, R.A., Jo, J.H., and Ouyang, Y., *Development and application of safety performance functions for Illinois*, in *ICT-10-066 UILU-ENG-2010-2006*. 2010.
47. Kweon, Y.J., and Lim, I.K., *Development of Safety Performance Functions for Multilane Highway and Freeway Segments Maintained by the Virginia Department of Transportation*. 2014.
48. Lubliner, H., et al., *Evaluation of Interactive Highway Safety Design Model Crash Prediction Tools for Two-Lane Rural Roads on Kansas Department of Transportation Projects*. 2014, Kansas Department of Transportation.

49. Saito, M., Brimley, B.K., and Schultz, G.G., *Transportation Safety Data and Analysis. Volume 2: Calibration of the Highway Safety Manual and Development of New Safety Performance Functions*. 2011.
50. Akaike, H., *A new look at the statistical model identification.* Automatic Control, IEEE Transactions on, 1974. 19(6): p. 716-723.
51. Srinivasan, R., and Bauer, K., *Safety Performance Function Development Guide: Developing Jurisdiction-Specific SPFs*. 2013, FHWA-SA-14-005. Federal Highway Administration Office of Safety, Washington, DC.
52. Trieu, V., Park, S., and McFadden, J., *Use of Monte Carlo simulation for a sensitivity analysis of Highway Safety Manual calibration factors.* Transport. Res. Rec., 2435 (2014), pp. 1-10
53. Green, E. R., Blackden, C.L., and Fields, M.A., *Spatial Database for Intersections.* Kentucky Transportation Center Research Report, Report No. KTC-15-11/SPR14-480-1F. 2015.
54. Storm , R., and Richfield, V., *Phase I-II of the Minnesota Highway Safety Manual Calibration*. Technical Memorandum, 2014. Available online at http://www.dot.state.mn.us/trafficeng/safety/docs/mnhsmcalibration.pdf. Accessed on June 10, 2017.
55. Troyer, D., Bradbury, K., and Juliano, C., *Strength of the variable calculating and evaluating safety performance function calibration factors for the State of Ohio. Transport. Res. Rec*., 2515, pp. 86-93.
56. Shankar, V., and Madanat, S., *Methods for Identifying High Collision Concentrations for Identifying Potential Safety Improvements: Development of Safety Performance Functions for California*. California Department of Transportation Research Report, Report No. CA15-2317, 2015.
57. Garber, N.J., and Rivera, G., *Safety Performance Functions for Intersections on Highways Maintained by the Virginia Department of Transportation*. Virginia Department of Transportation Research Report, Report No. FHWA/VTRC 11-CR1, 2010.
58. Garber, N.J., Haas, P.R., and Gosse, C., *Development of Safety Performance Functions for Two-Lane Roads Maintained by the Virginia Department of Transportation*. Virginia Department of Transportation Research Report, Report No. FHWAY/VTRC 10-R25, 2010.
59. Savolainen, P.T., Gates, T., Lord, D., Geedipally, S., Rista, E., Barrette, T., Russo, B.J., and Hamzeie, R., *Michigan Urban Trunkline Intersections Safety Performance Functions (SPFs) Development and Support*. Michigan Department of Transportation Research Report. Report No. RC-1628, 2015.
60. Sipple, M.C., *Calibration and Development of Safety Performance Functions for Rural Highway Facilities in Idaho*. Master Thesis, University of Idaho, 2014.
61. Williamson, M., and Zhou, H., *Develop calibration factors for crash prediction models for rural two-lane roadways in Illinois.* Procedia Soc. Behave. Sci., 43 (2012), pp. 330-338.
62. Sun, X, Li, Y., Magri, D., and Shirazi H.H., *Application of the highway safety manual draft chapter: Louisiana experience. Transport. Res. Rec*., 1950 (2006), pp. 55-64.

63. Banihashemi, M. *Sensitivity analysis of data set sizes for Highway Safety Manual calibration factors. Transport. Res. Rec*., 2279 (2012), pp. 75-81.
64. Tegge, R.A., Jo, J.H., and Ouyang, Y., *Development and Application of Safety Performance Functions for Illinois***.** Report No. ICT-10-066 Illinois Department of Transportation, Springfield, Illinois, 2010.
65. Persaud, B., Saleem, T., Faisal, S., Lyon, C., Chen, Y., and Sabbaghi, A., "Adoption of Highway Safety Manual Predictive Methodologies for Canadian Highways." *Proceedings from the 2012 Conference of the Transportation Association of Canada*, Fredericton, New Brunswick, October 2012.
66. Persaud and Lyon, Inc., and Felsburg Holt & Ullevig, *Safety Performance Functions for Intersections*. Colorado Department of Transportation Research Report. Report No. CDOT-2009-10, 2009.
67. Alluri, P., Saha, D., and Gan, A., M*inimum sample sizes for estimating reliable Highway Safety Manual (HSM) calibration factors. J. Transport. Saf. Sec.*, 8 (1) (2016), pp. 56-74, 10.1080/19439962.2014.978963.
68. Findley, D., Zegeer, C., Sundstrom, C., Hummer, J., and Rasdorf, W., *Applying the Highway Safety Manual to two-lane road curves. Transport. Res. Forum*, 51 (3) (2012), pp. 25-38
69. Kim, J., Anderson, M., and Gholston, S., *Modeling Safety Performance Functions for Alabama Urban and Suburban Arterials.* International Journal of Traffic and Transportation Engineering, Vol. 4, No. 3, (2015), pp. 84-93.
70. Shin, H., Dadvar, S., and Lee, Y.J., *Results and Lessons from Local Calibration Process of the Highway Safety Manual for the State of Maryland*. Transport. Res. Rec., 2515 (2015), pp. 104-114.
71. Xie, Y., and Chen, C., *Calibration of Safety Performance Functions for Massachusetts Urban and Suburban Intersections*. Massachusetts Department of Transportation Research Report. 2016.
72. Donnell, E., Gayah, V., and Jovanis, P., *Safety Performance Functions.* Pennsylvania Department of Transportation Research Report. Report No. FHWA-PA-2014-007-PSU WO 1, 2014.
73. Donnell, E., Gayah, V., and Li, L., Regionalized *Safety Performance Functions.* Pennsylvania Department of Transportation Research Report. Report No. FHWA-PA-2016-001-PSU WO 017, 2016.
74. Dissanayake, S., and Aziz, S.R., *Calibration of the Highway Safety Manual and Development of New Safety Performance Functions for Rural Multilane Highways in Kansas*. Kansas Department of Transportation Research Report, Report No. K-TRAN: KSU-14-3, 2016.
75. Titze, C., and Faron, W., *Crash Records Collection Policy Examination and Best Practices Review.* New Jersey Department of Transportation Research Report. NJ-2017-001.
76. New Jersey Geographic Information Network Website. https://njgin.state.nj.us/NJ_NJGINExplorer/jviewer.jsp?pg=ROADS. Accessed on February 5, 2019.
77. Lyon, C., Persaud, B., and Gross, F., The Calibrator: An SPF Calibration and Assessment Tool User Guide. (2016). FHWA-SA-17-016.

78. Bartin, B., Ozbay, K., and Chuan, X., *Extracting Horizontal Curvature Data from GIS Maps: Clustering Method.* Transportation Research Record: Journal of the Transportation Research Board. (2019). DOI: 10.1177/0361198119850789.
79. Findley, D.J., and Foyle, R.S., *Procedure for Identification and Investigation of Horizontal Curves with Insufficient Superelevation Rates*. FHWA-NC-2009-08. North Carolina Department of Transportation, Raleigh, NC.

APPENDIX A – SPFs DEVELOPED BY OTHER STUDIES

RURAL TWO-LANE TWO-WAY ROADWAYS

Saito et al. (2011) Utah

$$N = exp[-12.06 + 0.84 * \ln(AADT) + 0.45.L - 0.0271.CT + 0.0824.S]$$

Where, N is the predicted number of total annual crashes, $AADT$ is annual average daily traffic, L is segment length in miles, CT is percentage of combo-unit trucks in percentages, and S is speed limit in miles per hour.

Srinivasan and Carter (2010) North Carolina

$$N = L.exp[-4.0852 + 0.583 * \ln(AADT)]$$

Where, N is the predicted number of total annual crashes, $AADT$ is annual average daily traffic, and L is segment length in miles.

Lubliner et al. (2014) Kansas

$$N = \exp[-10.07 + 1.01.\ln(AADT) + 0.85.\ln(L) + 0.58.RHR]$$

Where, N is the predicted number of total annual crashes excluding animal crashes[7], $AADT$ is annual average daily traffic, L is segment length in miles, and RHR is roadside hazard rating, as defined in the HSM.

Banihashemi (2011) Washington

$$N = L.exp[-8.46345 + 1.05 * \ln(AADT)]$$

Shankar and Madanat (2015) California

$$N = L.exp[-5.13 + 0.68 * \ln(AADT)]$$

Garber et al. (2010) Virginia

$$N = L.exp[-5.71 + 0.744 * \ln(AADT)]$$

Sipple (2014) Idaho

$$N = L.exp[-5.7853 + 0.7501 * \ln(AADT)]$$

Mehta and Lou (2013) Alabama

$$N = L.exp[-7.135 + 0.916 * \ln(AADT)]$$

[7] Lubliner et al (2014) stated that the number of animal crashes was significantly high in Kansas and it was difficult to determine where these crashes would occur.

RURAL MULTILANE DIVIDED ROADWAYS

Srinivasan and Carter (2010) North Carolina

Total Crashes, $N = exp[-5.8986 + 0.7673 * \ln(AADT) + lnL]$

Fatal/Injury Crashes, $N = exp[-5.89869773 + 0.5584 * \ln(AADT) + lnL]$

Where, N is the predicted number of total annual crashes, $AADT$ is annual average daily traffic, and L is segment length in miles.

Dissanayake and Aziz (2016) Kansas

Total Crashes, $N = exp[-6.317 + 0.795 * \ln(AADT) + 0.898. lnL]$

Fatal/Injury Crashes, $N = exp[-10.030 + 1.0599 * \ln(AADT) + 0.399. lnL]$

Shankar and Madanat (2015) California

Total Crashes, $N = L. exp[-4.36 + 0.60 * \ln(AADT)]$

Fatal/Injury Crashes, n/a

Kweon and Lim (2014) Virginia

Total Crashes, $N = L. exp[-7.47 + 0.88 * \ln(AADT)]$

Fatal/Injury Crashes, $N = L. exp[-8.05 + 0.84 * \ln(AADT)]$

Tegge et al. (2010) Illionis

Total Crashes, n/a

Fatal/Injury Crashes, $N = L. exp[-7.767 + 0.923 * \ln(AADT)]$

Mehta and Lou (2013) Alabama $N = L. exp[-7.706 + 0.974 * \ln(AADT)]$

RURAL MULTILANE UNDIVIDED ROADWAYS

Srinivasan and Carter (2010) North Carolina

Total Crashes, $N = L.exp[-5.0970 + 0.7309 * \ln(AADT)]$

Fatal/Injury Crashes, $N = L.exp[-5.7277 + 0.6868 * \ln(AADT)]$

Where, N is the predicted number of total annual crashes, $AADT$ is annual average daily traffic, and L is segment length in miles.

Dissanayake and Aziz (2016) Kansas

Total Crashes, $N = exp[-6.347 + 0.822 * \ln(AADT) + 0.912. \ln(L)]$

Fatal/Injury Crashes, $N = exp[-8.206 + 0.817 * \ln(AADT) + 0.747. \ln(L)]$

Shankar and Madanat (2015) California

Total Crashes, $N = L.exp[-4.49 + 0.60 * \ln(AADT)]$

Fatal/Injury Crashes, n/a

Kweon and Lim (2014) Virginia

Total Crashes, $N = L.exp[-6.91 + 0.82 * \ln(AADT)]$

Fatal/Injury Crashes, $N = L.exp[-8.03 + 0.84 * \ln(AADT)]$

Tegge et al. (2010) Illionis

Total Crashes, n/a

Fatal/Injury Crashes, $N = L.exp[-3.005 + 0.259 * \ln(AADT)]$

URBAN TWO-LANE UNDIVIDED ARTERIALS (2U)

Srinivasan and Carter (2010) North Carolina[8]

Total Crashes, $N = L.exp[-6.5287 + 0.8777 * \ln(AADT)]$

Fatal/Injury Crashes, $N = L.exp[-7.6112 + 0.8801 * \ln(AADT)]$

Property Damage Only, $N = L.exp[-7.3781 + 0.9212 * \ln(AADT)]$

Shankar and Madanat (2015) California[2]

Total Crashes, $N = L.exp[-7.09 + 0.98 * \ln(AADT)]$

Fatal/Injury Crashes, n/a

Property Damage Only, $N = L.exp[-8.81 + 1.11 * \ln(AADT)]$

Tegge et al. (2010) Illionis[2]

Total Crashes, n/a

Fatal/Injury Crashes, $N = L.exp[-3.557 + 0.462 * \ln(AADT)]$

Property Damage Only, n/a

Kim et al. (2015) Alabama

Total Crashes (Multiple Vehicle), $N = L.exp[-17.062 + 1.957 * \ln(AADT)]$

Total Crashes (Single Vehicle), $N = L.exp[-5.926 + 0.487 * \ln(AADT)]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a

---

[8] Not specified whether divided or not.

URBAN THREE-LANE WITH TWLTL (3T)

Kim et al. (2015) Alabama

Total Crashes (Multiple Vehicle), $N = L.exp[-8.258 + 1.073 * \ln(AADT)]$

Total Crashes (Single Vehicle), $N = L.exp[-11.442 + 1.065 * \ln(AADT)]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a

URBAN FOUR-LANE UNDIVIDED ARTERIALS (4U)

Kim et al. (2015) Alabama

Total Crashes (Multiple Vehicle), $N = L.exp[-21.383 + 2.479 * \ln(AADT)]$

Total Crashes (Single Vehicle), $N = L.exp[-5.713 + 0.498 * \ln(AADT)]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a

URBAN FOUR-LANE DIVIDED ARTERIALS (4D)

Srinivasan et al. (2011) Florida

Total Crashes, n/a

Fatal/Injury Crashes, $N = L.exp[-11.010 + 1.185 * \ln(AADT)]$

Property Damage Only, n/a

Kim et al. (2015) Alabama

Total Crashes (Multiple Vehicle), $N = L.exp[-11.037 + 1.294 * \ln(AADT)]$

Total Crashes (Single Vehicle), $N = L.exp[-5.570 + 0.446 * \ln(AADT)]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a

URBAN FIVE-LANE WITH TWLTL ARTERIALS (5T)

Kim et al. (2015) Alabama

Total Crashes (Multiple Vehicle), $N = L.exp[-3.762 + 0.539 * \ln(AADT)]$

Total Crashes (Single Vehicle), $N = L.exp[-1.923 + 0.010 * \ln(AADT)]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a

URBAN MULTILANE DIVIDED ROADWAYS[9]

Srinivasan and Carter (2010) North Carolina

Total Crashes, $N = L.exp[-4.1053 + 0.6750 * \ln(AADT)]$

Fatal/Injury Crashes, $N = L.exp[-4.5964 + 0.6113 * \ln(AADT)]$

Property Damage Only, $N = L.exp[-5.0670 + 0.7297 * \ln(AADT)]$

Shankar and Madanat (2015) California

Total Crashes, $N = L.exp[-7.11 + 1.01 * \ln(AADT)]$

Fatal/Injury Crashes, n/a

Property Damage Only, $N = L.exp[-7.23 + 0.97 * \ln(AADT)]$

Kweon and Lim (2014) Virginia

Total Crashes, $N = L.exp[-9.14 + 1.07 * \ln(AADT)]$

Fatal/Injury Crashes, $N = L.exp[-10.19 + 1.06 * \ln(AADT)]$

Property Damage Only, n/a

[9] This definition does not exist in the HSM – there are 5 segment categories for urban/suburban arterials in the manual

Tegge et al. (2010) Illionis

Total Crashes, n/a

Fatal/Injury Crashes, $N = L.exp[-6.206 + 0.761 * \ln(AADT)]$

Property Damage Only, n/a

URBAN MULTILANE UNDIVIDED ROADWAYS[10]

Srinivasan and Carter (2010) North Carolina

Total Crashes, $N = L.exp[-5.2849 + 0.8114 * \ln(AADT)]$

Fatal/Injury Crashes, $N = L.exp[-5.8160 + 0.756 * \ln(AADT)]$

Property Damage Only, $N = L.exp[-6.4065 + 0.8813 * \ln(AADT)]$

Shankar and Madanat (2015) California

Total Crashes, $N = L.exp[-5.86 + 0.91 * \ln(AADT)]$

Fatal/Injury Crashes, n/a

Property Damage Only, $N = L.exp[-6.13 + 0.89 * \ln(AADT)]$

Kweon and Lim (2014) Virginia

Total Crashes, $N = L.exp[-7.88 + 0.94 * \ln(AADT)]$

---

[10] Lubliner et al (2014) stated that the number of animal crashes was significantly high in Kansas and it was difficult to determine where these crashes would occur.

10 Not specified whether divided or not.

[10] This definition does not exist in the HSM – there are 5 segment categories for urban/suburban arterials in the manual

Fatal/Injury Crashes, $N = L.exp[-10.36 + 1.09 * \ln(AADT)]$

Property Damage Only, n/a

Tegge et al. (2010) Illionis

Total Crashes, n/a

Fatal/Injury Crashes, $N = L.exp[-4.876 + 0.647 * \ln(AADT)]$

Property Damage Only, n/a

Donnell et al. (2014) Pennsylvania[11]

Total Crashes, $N = exp\left[-6.337 + 0.479 * \ln\left(AADT_{major}\right) + 0.362. \ln(AADT_{minor}) - 0.33. LT_{major} + \right.$

$\left. -0.507. RT_{major}\right]$

Where, LT_{major} is the presence of an exclusive left turn on major leg, and RT_{major} is the presence of an exclusive right turn on major leg.

Sipple (2014) Idaho

$$N = exp\left[-6.1502 + 0.0966 * \ln\left(AADT_{major}\right) + 0.6969. \ln(AADT_{minor})\right]$$

Sabbaghi (2010) Ontario, Canada

Total Crashes, $N = exp\left[-18.01 + 1.82 * \ln\left(AADT_{major}\right) + 0.175. \ln(AADT_{minor})\right]$

[11] The definition of 3ST in Donnell et al. (2014) differs from the one included in the HSM. 3ST in the HSM has stop control on minor legs only. 3ST in this study has stop control on both approaches. However, in the following study by Donnell et al. (2016), this definition was reverted to the HSM standard with the same coefficients shown here.

RURAL TWO-LANE TWO-WAY INTERSECTIONS – 3SG[12]

Donnell et al. (2014) Pennsylvania

Total Crashes, $N = exp\left[-6.813 + 0.451 * \ln\left(AADT_{major}\right) + 0.349. \ln(AADT_{minor}) + 0.02. S_{major} - 0.433. W_{major} - 0.345. W_{minor}\right]$

Where, S_{major} is the posted speed limit on major leg, and W_{major} and W_{minor} are the presence of a crosswalk on major and minor legs, respectively.

Sipple (2014) Idaho

$$N = exp\left[-8.6336 + 0.8966 * \ln\left(AADT_{major}\right) + 0.0458. \ln(AADT_{minor})\right]$$

[12] Currently not included in the HSM

RURAL TWO-LANE TWO-WAY INTERSECTIONS – 4ST

Donnell et al. (2014) Pennsylvania

Total Crashes, $N = exp\left[-6.359 + 0.528 * \ln\left(AADT_{major}\right) + 0.275. \ln(AADT_{minor}) + 0.007. \theta\right]$

Where, $S_{major}$is the posted speed limit on major leg and θ is intersection skew angle.

Sabbaghi (2010) Ontario, Canada

Total Crashes, $N = exp\left[-10.79 + 0.56 * \ln\left(AADT_{major}\right) + 0.82. \ln(AADT_{minor})\right]$

RURAL TWO-LANE TWO-WAY INTERSECTIONS – 4SG

Donnell et al. (2014) Pennsylvania

Total Crashes, $N = exp[-5.353 + 0.313 * \ln(AADT_{major}) + 0.25. \ln(AADT_{minor}) + 0.025. S_{major} + 0.014. S_{minor} + 0.216. RT]$

Where, $S_{major}$and S_{minor} are posted speed limits on major and minor legs, respectively, and RT is the presence of exclusive right-turn lane on either major approach.

RURAL MULTILANE ROADWAY INTERSECTIONS – 3ST

Donnell et al. (2016) Pennsylvania

Total Crashes, $N = exp\left[-8.072 + 0.509 * \ln\left(AADT_{major}\right) + 0.509.\ln(AADT_{minor})\right]$

Fatal/Injury Crashes, $N = exp\left[-7.830 + 0.459 * \ln\left(AADT_{major}\right) + 0.459.\ln(AADT_{minor})\right]$

Property Damage, n/a

RURAL MULTILANE ROADWAY INTERSECTIONS – 4ST

Donnell et al. (2016) Pennsylvania

Total Crashes, $N = exp\left[-4.342 + 0.334 * \ln\left(AADT_{major}\right) + 0.264.\ln(AADT_{minor})\right]$

Fatal/Injury Crashes, $N = exp\left[-3.248 + 0.217 * \ln\left(AADT_{major}\right) + 0.152.\ln(AADT_{minor})\right]$

Property Damage, n/a

RURAL MULTILANE ROADWAY INTERSECTIONS – 4SG

Donnell et al. (2016) Pennsylvania

Total Crashes, $N = exp\left[-3.563 + 0.389 * \ln\left(AADT_{major}\right) + 0.134. \ln(AADT_{minor})\right]$

Fatal/Injury Crashes, $N = exp\left[-3.301 + 0.291 * \ln\left(AADT_{major}\right) + 0.133. \ln(AADT_{minor})\right]$

Property Damage, n/a

URBAN INTERSECTIONS – 3ST

Garber and Rivera (2010) Virginia

Total Crashes, $N = exp\left[-5.4696 + 0.4874 * \ln\left(AADT_{major}\right) + 0.1985. \ln(AADT_{minor})\right]$

Fatal/Injury Crashes, $N = exp\left[-7.4642 + 0.5791 * \ln\left(AADT_{major}\right) + 0.2091. \ln(AADT_{minor})\right]$

Property Damage Only, n/a

Savalainen et al. (2015) Michigan

Total Crashes, n/a

Fatal/Injury Crashes, $N = exp\left[-11.89 + 0.75 * \ln\left(AADT_{major}\right) + 0.42. \ln(AADT_{minor})\right]$

Property Damage Only, $N = exp\left[-13.06 + 0.99 * \ln\left(AADT_{major}\right) + 0.46. \ln(AADT_{minor})\right]$

Persaud et al. (2009) Colorado[13]

Total Crashes, $N = exp\left[-10.6568 + 0.8999 * \ln\left(AADT_{major}\right) + 0.3019. \ln(AADT_{minor})\right]$

Fatal/Injury Crashes, $N = exp\left[-11.6429 + 0.7689 * \ln\left(AADT_{major}\right) + 0.8642. \ln(AADT_{minor})\right]$

Property Damage Only, n/a

Xie and Chen (2016) Massaschusettes

Total Crashes (Multiple Vehicles), $N = exp\left[-20.02 + 1.66 * \ln\left(AADT_{major}\right) + 0.54. \ln(AADT_{minor})\right]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a

[13] This model was specific to urban four-lane divided road intersections

URBAN INTERSECTIONS – 4ST

Garber and Rivera (2010) Virginia

Total Crashes, $N = exp\left[-6.0723 + 0.4558 * \ln\left(AADT_{major}\right) + 0.347. \ln(AADT_{minor})\right]$

Fatal/Injury Crashes, $N = exp\left[-7.6917 + 0.5001 * \ln\left(AADT_{major}\right) + 0.3695. \ln(AADT_{minor})\right]$

Property Damage Only, n/a

Savalainen et al. (2015) Michigan

Total Crashes, n/a

Fatal/Injury Crashes, $N = exp\left[-9.21 + 0.56 * \ln\left(AADT_{major}\right) + 0.38. \ln(AADT_{minor})\right]$

Property Damage Only, $N = exp\left[-7.94 + 0.50 * \ln\left(AADT_{major}\right) + 0.43. \ln(AADT_{minor})\right]$

Persaud et al. (2009) Colorado[14]

Total Crashes, $N = exp\left[-13.481 + 0.981 * \ln\left(AADT_{major}\right) + 0.6658. \ln(AADT_{minor})\right]$

Fatal/Injury Crashes, $N = exp\left[-14.0091 + 0.7689 * \ln\left(AADT_{major}\right) + 0.8512. \ln(AADT_{minor})\right]$

Property Damage Only, n/a

Xie and Chen (2016) Massachusetts

Total Crashes (Multiple Vehicles), $N = exp\left[-8.70 + 0.31 * \ln\left(AADT_{major}\right) + 0.86. \ln(AADT_{minor})\right]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a

---

[14] This model was specific to urban two-lane undivided road intersections

URBAN INTERSECTIONS – 3SG

Garber and Rivera (2010) Virginia

Total Crashes, $N = exp\left[-6.543 + 0.6591 * \ln\left(AADT_{major}\right) + 0.2119. \ln(AADT_{minor})\right]$

Fatal/Injury Crashes, $N = exp\left[-8.4268 + 0.7147 * \ln\left(AADT_{major}\right) + 0.2481. \ln(AADT_{minor})\right]$

Property Damage Only, n/a

Savalainen et al. (2015) Michigan

Total Crashes, n/a

Fatal/Injury Crashes, $N = exp\left[-5.35 + 0.43 * \ln\left(AADT_{major}\right) + 0.13. \ln(AADT_{minor})\right]$

Property Damage Only, $N = exp\left[-5.96 + 0.57 * \ln\left(AADT_{major}\right) + 0.19. \ln(AADT_{minor})\right]$

Persaud et al. (2009) Colorado[15]

Total Crashes, $N = exp\left[-10.552 + 0.7596 * \ln\left(AADT_{major}\right) + 0.5425. \ln(AADT_{minor})\right]$

Fatal/Injury Crashes, $N = exp\left[-11.0639 + 0.7215 * \ln\left(AADT_{major}\right) + 0.5027. \ln(AADT_{minor})\right]$

Property Damage Only, n/a

Persaud et al. (2012) Toronto, Canada

Total Crashes (Multi-vehicle), $N = exp\left[-6.60248 + 0.6177 * \ln\left(AADT_{major}\right) + 0.3874. \ln(AADT_{minor})\right]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a

Xie and Chen (2016) Massachusetts

[15] This model was specific to urban four-lane divided road intersections

Total Crashes (Multiple Vehicles), $N = exp\left[-8.30 + 0.81 * \ln\left(AADT_{major}\right) + 0.21.\ln(AADT_{minor})\right]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a

URBAN INTERSECTIONS – 4SG

Garber and Rivera (2010) Virginia

Total Crashes, $N = exp\left[-7.6234 + 0.6742 * \ln\left(AADT_{major}\right) + 0.3453.\ln(AADT_{minor})\right]$

Fatal/Injury Crashes, $N = exp\left[-8.5256 + 0.6477 * \ln\left(AADT_{major}\right) + 0.3579.\ln(AADT_{minor})\right]$

Property Damage Only, n/a

Savalainen et al. (2015) Michigan

Total Crashes, n/a

Fatal/Injury Crashes, $N = exp\left[-7.90 + 0.71 * \ln\left(AADT_{major}\right) + 0.17.\ln(AADT_{minor})\right]$

Property Damage Only, $N = exp\left[-8.44 + 0.80 * \ln\left(AADT_{major}\right) + 0.26.\ln(AADT_{minor})\right]$

Persaud et al. (2009) Colorado[16]

Total Crashes, $N = exp\left[-17.4479 + 1.5811 * \ln\left(AADT_{major}\right) - 0.2585.AADT_{major}/10000\right]$

Fatal/Injury Crashes, $N = exp\left[-20.6848 + 1.8508 * \ln\left(AADT_{major}\right) + 0.4547.\ln(AADT_{minor}) - 0.3743.AADT_{major}/10000\right]$

Property Damage Only, n/a

Persaud et al. (2012) Toronto, Canada

Total Crashes (Multi-vehicle), $N = exp\left[-7.48131 + 0.5661 * \ln\left(AADT_{major}\right) + 0.5581.\ln(AADT_{minor})\right]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a

[16] This model was specific to urban four-lane divided road intersections

Xie and Chen (2016) Massachusetts

Total Crashes (Multiple Vehicles), $N = exp\left[-11.79 + 0.92 * \ln\left(AADT_{major}\right) + 0.52.\ln(AADT_{minor})\right]$

Fatal/Injury Crashes, n/a

Property Damage Only, n/a