STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV 10/98)

Lock Data on Form

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA19-3027		

4. TITLE AND SUBTITLE

Implementation of SPF methods to Identify High Collision Concentration Locations

5. REPORT DATE

November 30, 2019

6. PERFORMING ORGANIZATION CODE

7. AUTHOR

Praveen Vayalamkuzhi, Aditya Medury, Offer Grembek, Venky Shankar, SangHyouk Oum, Camille Salem, Lin Yang, Jiajian Lu

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS

UC Berkeley Safe Transportation Research & Education Center
2614 Dwight Way, #7374
Berkeley, CA 94720-7374

10. WORK UNIT NUMBER

11. CONTRACT OR GRANT NUMBER

65A0633

12. SPONSORING AGENCY AND ADDRESS

California Department of Transportation
Division of Research and Innovation, MS-83
1227 O Street
Sacramento CA 95814

13. TYPE OF REPORT AND PERIOD COVERED

Final Report

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT

The California Department of Transportation (Caltrans) ranks transportation safety as a top priority to save lives by improving safety throughout the State Highway System (SHS). Caltrans is establishing a unique role as a leader in state-specific highway safety improvements and capacity building at the national level. To effectively manage transportation safety along the SHS, it is essential to monitor traffic collisions and traffic volumes along three infrastructure types: (i) segments; (ii) intersections; and (iii) ramps. Caltrans uses the Traffic Accident Surveillance and Analysis System (TASAS) Transportation Network System (TSN) to manage infrastructure assets, traffic volumes, and police-reported traffic collisions.

17. KEY WORDS	17. DISTRIBUTION STATEMENT
	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161

19. SECURITY CLASSIFICATION *(of this report)*	20. NUMBER OF PAGES	21. COST OF REPORT CHARGED
Unclassified	129	N/A

DISCLAIMER STATEMENT

Implementation of SPF methods to Identify High Collision Concentration Locations

Final Technical Report

Prepared by the University of California Berkeley
Safe Transportation Research and Education Center

for the

California Department of Transportation

November 30, 2019

Acknowledgments

The authors would like to thank the California Department of Transportation for their support of this project. We especially acknowledge the support, guidance, and collaboration of Thomas Schriber, Dean Samuelson, John Ensch, Aaron Truong, Eric Wong, and Vladimir Poroshin at Caltrans. We also deeply appreciate the work of Jerry Kwong of the Division of Research and Innovation for facilitating the project from its inception and through the final report.

TABLE OF CONTENTS

Acknowledgments 4
TABLE OF CONTENTS 5
Chapter 1. Introduction 8
Chapter 2. Overview of Safety Performance Functions 10
2.1. Highway Safety Manual 10
2.1.1. Safety Performance Functions 10
2.1.2. Network Screening 11
2.1.3. Network Screening Methods 12
2.2. Potential for Safety Improvement 13
2.3. Past California-Specific Highway Safety Research 14
2.4. SPF Data Requirement 15
2.4.1. Highway 15
2.4.2. Intersection 15
2.4.3. Ramp 15
2.4.4. Collision Data 15
2.5. Summary 16
Chapter 3. Understanding the Current Practice of Safety Analysis – Table C 17
3.1. Overview of Table C 17
3.2. Mapping Existing Process 17
3.2.1. Questionnaire Survey 18
3.2.2. Stakeholder Interviews 19
3.3. Process Mapping 20
3.3.1. Relationship Map 20
3.3.2. Cross-Functional Process Map 21
3.3.3. Flow Chart Map 22
3.4. Discussion on Table C Maps 23
3.5. Summary 24
Chapter 4. California-Specific Safety Performance Functions (SPF) 25
4.1. Need for California-specific SPF 25
4.2. Applications of SPF 25
4.3. Development of SPF 25

4.4. Facility-based SPF 25
4.5. Injury Severity-based SPF 26
4.6. Steps in Developing SPFs 26
4.6.1. Facility Type 27
4.6.2. Data Requirements and Compilation 27
4.6.3. Data Challenges 29
4.6.4. Functional Form 29
4.6.5. Mathematical Modeling 31
4.6.6. Model Checks 31
4.7. Data Analysis 32
4.7.1. Estimation Process 34
4.7.2. Clustering Approach 34
4.8. California-specific SPFs 35
4.8.1. Segment SPF 35
4.8.2. Intersection SPFs 61
4.8.3. Ramp SPF 70
4.9. Potential for Safety Improvement 73
4.9.1. Empirical Bayes Method 74
4.10. Network Screening Methods 75
4.10.1. Sliding Window 75
4.10.2. Peak Searching 76
4.11. SPF Implementation Challenges 77
4.12. Chapter Summary 77
Chapter 5. SPF Implementation - Design and Development of the SPF Tool 78
5.1. Desired Functionality of the Excel Macro Spreadsheet Tool 78
5.2. Development of the SPF Tool 79
5.2.1. Data Input 79
5.2.2. Data Analysis 81
5.2.3. Report Generation 83
5.3. Summary 86
Chapter 6. Summary and Conclusions 87
6.1. Summary 87
6.2. Conclusions 88

6.3. Project Takeaway 88
References 89
Appendix A – Data Structure used for SPF Development 90
I. Infrastructure Data 90
II. Collision Data 95
Appendix B –Total Collision SPFs Developed 96
This appendix shows the statistically significant variables in total collision SPFs, and the corresponding route and county clustering results. 96
1. Segment SPF 96
Class – 1 Rural 2 Lane Non-Freeway Undivided 96
Class – 2 Rural 2 Lane Non-Freeway Divided 97
Class – 3 Rural 3+ Lane Non-Freeway Undivided 99
Class – 4 Rural 3+ Lane Non-Freeway Divided 100
Class – 5 Rural 2-4 Lane Freeway Undivided 101
Class – 6 Rural 2-4 Lane Freeway Divided 103
Class – 7 Rural 5+ Lane Freeway 105
Class – 8 Urban 2 Lane Non-Freeway Undivided 106
Class – 9 Urban 2 Lane Non-Freeway Divided 108
Class – 10 Urban 3+ Lane Non-Freeway Undivided 109
Class – 11 Urban 3+ Lane Non-Freeway Divided 111
Class – 12 Urban 2-7 Lane Freeway Undivided 113
Class – 13 Urban 2-7 Lane Freeway Divided 114
Class – 14 Urban 8+ Lane Freeway Divided 116
Class – 15 Rural L/R Alignment Freeway 118
Class – 16 Urban L/R Alignment Freeway 119
Class – 17 L/R Alignment Non-Freeway 120
2. Intersection SPF 121
Rural Intersection SPF 121
Urban Signalized SPF 123
Urban Un-signalized SPF 125
3. Ramp SPF 127
Appendix C – SPF Tool Version -1 Process Flowchart 129

Chapter 1. Introduction

The California Department of Transportation (Caltrans) ranks transportation safety as a top priority to save lives by improving safety throughout the State Highway System (SHS). Caltrans is establishing a unique role as a leader in state-specific highway safety improvements and capacity building at the national level. To effectively manage transportation safety along the SHS, it is essential to monitor traffic collisions and traffic volumes along three infrastructure types: (i) segments; (ii) intersections; and (iii) ramps. Caltrans uses the Traffic Accident Surveillance and Analysis System (TASAS) Transportation Network System (TSN) to manage infrastructure assets, traffic volumes, and police-reported traffic collisions. Previous Caltrans studies—"Data Requirements for Safety Studies" and "SPF Tool Enhancement"—have established the data needs and opportunities for safety analysis and evaluation. These studies build upon the work described in the report, "Methods for Identifying High Collision Concentrations for Identifying Potential Safety Improvements," by identifying the existing data structure and the limitations of the study.

The purpose of this study was to support Caltrans in transitioning from the development and update of Safety Performance Function (SPF) to SPF implementation. The idea behind this is to develop new SPFs based on available infrastructure and collision data, and later develop an MS Excel macro spreadsheet tool that is flexible enough to make use of any progress related to network screening capabilities. This was achieved through four overarching objectives: (i) re-estimate/develop SPFs with most recent data; (ii) design and develop an MS Excel macro spreadsheet tool that can be used to conduct SPF-based network screening; (iii) incorporate all Caltrans-reviewed SPFs equations into the spreadsheet tool, so that it can be used by selected Caltrans expert users; and (iv) provide guidelines for developing additional SPFs, re-calibrating existing SPFs, and creating a roadmap for incorporating such SPFs into the spreadsheet tool. As part of the study, we developed the MS Excel macro spreadsheet tool with re-estimated SPFs based on total collisions and injury-based—Fatal + Severe + Visible (FSV) injury collisions—for identifying high collision concentration locations (HCCLs). Development of this spreadsheet tool will allow Caltrans to better make use of existing data collection efforts for improving safety through state-of-the-art network screening practices.

The main documents/reports reviewed as part of this project encompass the following:

- Highway Safety Manual (HSM) 2010
- Methods for identifying High Collision Concentrations for Identifying Potential Safety Improvements: Development of Safety Performance Functions for California
- Data Requirements for Safety Studies

After reviewing these reports and other relevant information, the following research questions were compiled:

- What is the current practice used by Caltrans for identifying HCCL?
- Can we use Safety Performance Functions based on HSM for California's SHS? If not, how can we develop California-specific SPFs?
- What is the data availability for developing California-specific SPFs, and how often it is updated?
- How can we utilize the developed SPFs for optimal network screening?

This report provides answers to the above questions, and includes an overview and mapping of the current practice used by Caltrans to identify high crash concentration locations. Additionally, the SPF spreadsheet tool developed as part of this project can be modified or updated using advanced safety performance functions when available.

The following chapters describe the tasks conducted as part of this study. Chapter 2 describes the overview of safety performance functions and network screening approaches. Chapter 3 describes the existing practice of identifying HCCLs. Chapter 4 explains the step-by-step process in SPF development and the challenges in data analysis. Chapter 5 describes the development of the MS Excel macro spreadsheet tool and elaborates on the stages in the tool process. Chapter 6 presents the conclusion. Appendix A includes a summary of data structure with technical specifications used for SPF development for safety screening and advanced safety studies. Appendix B includes model outputs for total collision SPFs. And Appendix C presents the SPF Tool Version 1.X flow chart.

Chapter 2. Overview of Safety Performance Functions

Prior to developing California-specific safety performance functions for network screening, it is important to review the best practices currently being used nationwide and research work in practice. This chapter summarizes safety performance functions based on the Highway Safety Manual (HSM), definitions used in this study, and other relevant information. In addition, this chapter provides the data requirements for developing SPFs for three infrastructure facilities—segment, intersection, and ramp—for the California State Highway System.

This overview focuses on the application and approach for developing site-specific SPFs and other relevant information. Additionally, we describe the types of SPFs, data structure describing the infrastructure, and collision data requirements for developing these types of SPFs and their significance.

2.1. Highway Safety Manual

The Highway Safety Manual (HSM) provides new and advanced analytical tools and techniques for quantifying the potential effects on crashes as a result of decisions made in planning, design, operations and maintenance. The HSM is a resource document that is used nationwide to help transportation professionals conduct safety analyses in a technically sound and consistent manner, thereby improving decisions made based on safety performance. The HSM describes techniques for safety analysis. One of these techniques is quantitative predictive analysis, which calculates an expected number and severity of crashes at sites with similar geometric and operational characteristics based on existing and future conditions, or roadway design alternatives, to improve highway safety. Applications of HSM considered in this study are: (i) estimate potential crash frequency and severity on highway networks; and (ii) estimate potential effects on crash frequency and severity of planning, design, operations and policy decisions.

Two of the HSM's four parts were used in this study to develop California-specific SPFs for identifying HCCLs though network screening—Part B: Road Safety Management Process – Network Screening, and Part C: Predictive Method – Safety Performance Functions (SPFs).

2.1.1. Safety Performance Functions

Safety Performance Functions are mathematical relationships between roadway attributes and crashes. There are two types of SPFs based on data availability as follows:

(i) Type I SPFs

Type 1 SPFs include functional forms in which the independent variables include an intercept and average daily traffic (ADT). The functional form for type 1 for the segments is shown in Eqn. (2.1):

$$\lambda_i = \text{length}_i * e^{\alpha} * \text{ADT}_i^{\beta} \quad (2.1)$$

where, 'λ' is the expected number of collisions; 'α' is the intercept; and 'β' is the coefficient of ADT (captures the effect of variable on the number of collisions).

Generally, the length of the segment is assumed to linearly affect the expected crash rate for a roadway segment and is considered an offset variable.

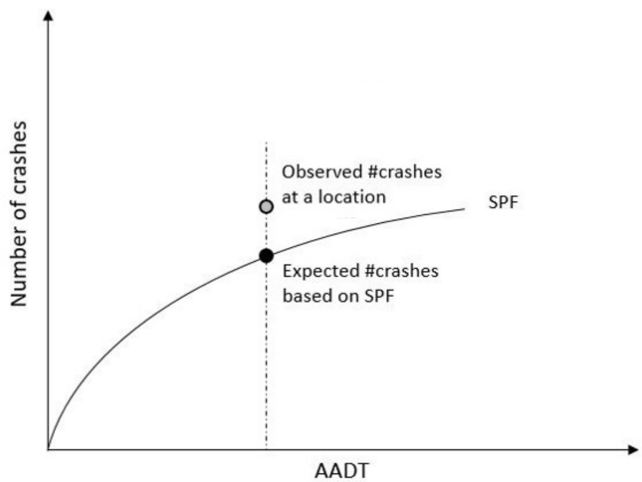

Figure 2.1 Graphical representation of a typical Type I Safety Performance Function

(ii) Type II SPFs

In type 2 SPFs, the estimating equation includes roadway geometry variables and intersection design elements in addition to the length and ADT effects. Therefore, given a vector of geometric effects Z_{ij} and associated coefficients γ_{j}, the functional form for segments is shown in Eqn. (2.2):

$$\lambda_i = length_i * e^{\alpha} * \mathrm{ADT}_i^{\beta} * e^{\sum_{j=1}^{l} \gamma_j Z_{ij}} \tag{2.2}$$

where, 'λ' is the expected number of collisions; 'α' is the intercept; and 'β' is the coefficient of ADT; 'Z' is the geometric and other site characteristics, while 'γj' is the coefficient of 'Z'.

2.1.2. Network Screening

Network screening is the process of reviewing the State Highway System to identify and rank sites based on the potential for reducing average crash frequency. This is the first activity undertaken in the road safety management process. HSM 2010 identifies five major steps in network screening as follows:

i. Establish Focus—Identify the purpose or intended outcome of the network screening analysis.
ii. Identify Network and Establish Reference Populations—Specify the type of sites or facilities being screened and identify groupings of similar sites or facilities.
iii. Select Performance Measures—The performance measure is selected as a function of the screening focus and the data and analytical tools available.
iv. Select Screening Method—There are three principle screening methods: ranking, sliding window, and peak searching.
v. Screen and Evaluate Results—The final step in the process is to conduct the screening analysis and evaluate the results.

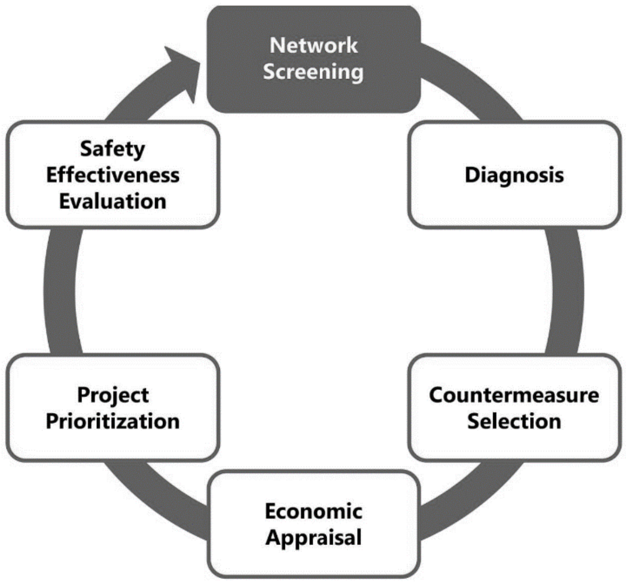

Figure 2.2 Schematic diagram of the road safety management process

2.1.3. Network Screening Methods

This study is particularly focused on adopting a suitable approach for identifying HCCLS based on network screening methods suggested by HSM 2010. Based on the suitability for California SHS and other relevant researches, two network screening methods were identified: the sliding window and peak searching methods. The section below describes the application of both methods in network screening.

2.1.3.1. Sliding Window Approach

In this method a window of a specified length is conceptually moved along the road segment from beginning to end in increments of a specified size (typically 0.10 mi or equal to the length of the roadway segment for small segments). Screening calculations are performed for each 'window' and segments are ranked by most critical window. After all segments are ranked according to the respective highest subsegment value, those segments with the greatest potential for reduction in crash frequency or severity are studied in detail to identify potential countermeasures. Windows may overlap adjacent road segments that are not identical in terms of traffic volumes and geometry

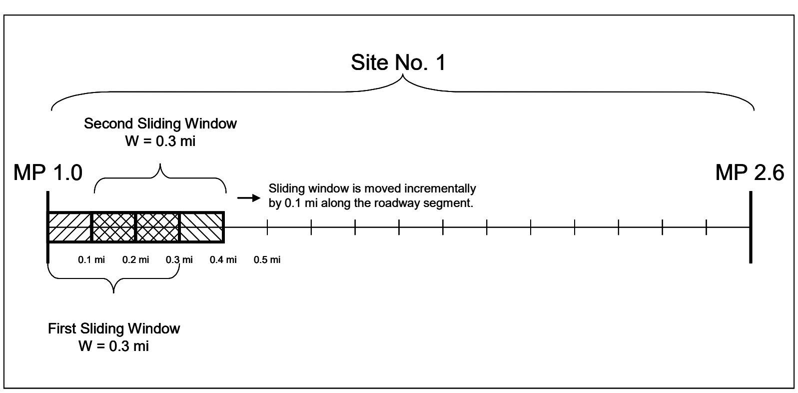

Figure 2.3 Schematic diagram of the sliding window method of network screening

2.1.3.2. Peak Searching Approach

In this method, each individual roadway segment is subdivided into windows of similar length, potentially growing incrementally in length until the length of the window is equal to the length of the entire roadway segment. The windows do not span multiple roadway segments. The first step in the peak searching method is to divide a given roadway segment (or ramp) into 0.1-mi windows. The windows do not overlap, with the possible exception that the last window may overlap the previous one. If the segment is less than 0.1 mi in length, then the segment length is equal to the window length.

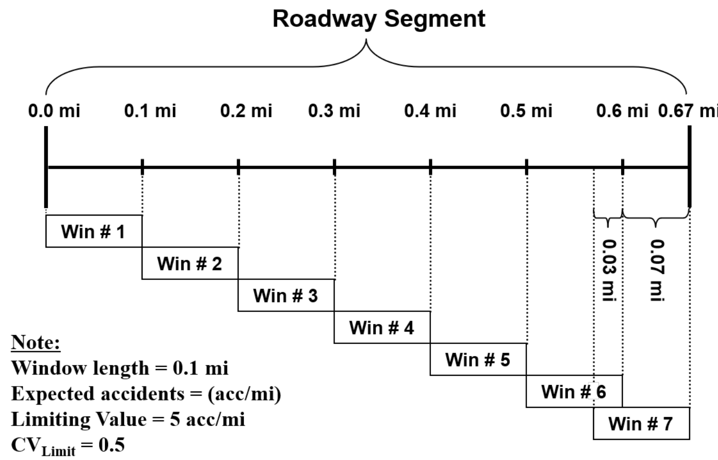

Figure 2.4 Schematic diagram of the peak searching method

The precision of the performance measure is assessed by calculating the coefficient of variation (CV) of the performance measure, where CV is a ratio of root of variance of performance measure to the performance measure.

A large CV indicates a low level of precision in the estimate, while a small CV indicates a high level of precision in the estimate. The calculated CV is compared with a specified limiting CV. If the calculated CV is less than or equal to the CV limiting value, the performance measure meets the desired precision level, and the performance measure for a given window can potentially be considered for use in ranking the segment. If the calculated CV is greater than the CV limiting value, the window is automatically removed from further consideration.

2.2. Potential for Safety Improvement

Potential for Safety Improvement (PSI) estimates how much the long-term crash frequency could be reduced at a site. Based on the network screening method, PSI is estimated as the difference between estimated crashes based on the Empirical Bayes (EB) approach and predicted using SPF (N_p) as shown in Figure 2.4. The EB estimate is a weighted average of the site's observed crash count and crashes expected at similar sites using a safety performance function.

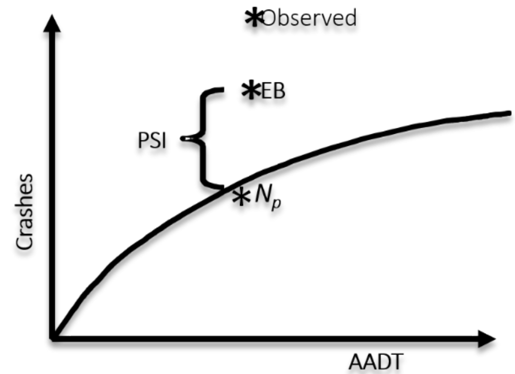

Figure 2.4 Potential for safety improvement of a site

2.3. Past California-Specific Highway Safety Research

Earlier attempts to develop Type 1 and Type 2 SPFs for segments, intersections and ramps for California SHS were made by Venky and Sameer (2015). Over 13,000 centerline miles of road segments, over 17,000 intersections, and the entire ramp system with ramp metered subsets were evaluated. The SPFs were estimated using 2005-2010 historic data. Severity data was developed using SWITRS definitions, including property damage only, complaint of pain, visible injury, severe, and fatal injury. A total of 60 Type 1 SPFs were developed for the five major severity outcomes, and another 60 Type 2 SPFs were also developed. Twelve Type 1 and Type 2 SPFs were developed for intersections, while twelve Type 1 and Type 2 SPFs were also developed for ramps. Model transferability tests were conducted to evaluate parameter stability across years. In addition, model predictive measures of effectiveness were evaluated for 2011-2012 out of model estimation samples. The study concluded that Type 2 SPFs were superior to Type 1 SPFs.

In 2017, a study by UC Berkeley SafeTREC attempted to understand the data needs for developing SPF. This project seeks to develop a roadmap for integrating data sources within as well as outside of Caltrans and to improve the overall quality of available data for SPF development, which will eventually increase the effectiveness of network screening employed for identifying high collision concentration locations. The visualization of the project is shown in Figure 2.5. The study identified the need of additional geometric data for developing advanced SPF and provided data structure with technical specifications.

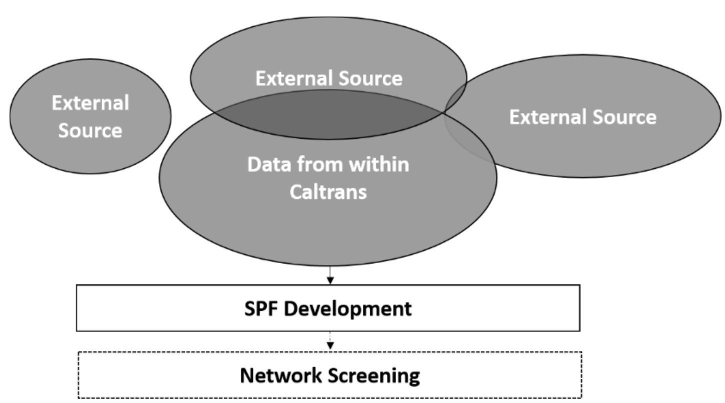

Figure 2.5 Visualization of the SPF data needs project

2.4. SPF Data Requirement

TASAS-TSN is the departmental database used to maintain and link traffic census, collision, and highway inventory data for the State Highway System (SHS). The TASAS branch also maintains accident data in the TSN database for all collisions which occur on, or are associated with, a State highway facility. Combining the highway inventory and accident data allows Caltrans to identify highway locations for safety investigation. Two data structures for the three facility types are currently available in the TASAS-TSN—highway, intersection, and ramp infrastructure data, in addition to collision data.

Based on the HSM, past research and data availability in the TASAS-TSN, infrastructure data required for safety analyses includes location information, geometric or design characteristics, traffic volume and additional characteristics for all three facility types: highway segments, intersection, and ramp. This information varies according to facility type. Collision data include location information and crash severity. Both infrastructure and collision data are explained in the following sections.

2.4.1. Highway

Highway segment infrastructure data requirements includes location information such as district, county, route (including suffix if any) and post mile (including prefix and suffix, if any). The main geometric characteristics of the highway include number of lanes, shoulder type and width, median type and width, travel way width, length of segment, and functional class. Average daily traffic (ADT) is considered as the traffic volume. In addition to these details, information on highway group, population group, lighting condition, break description, and operation characteristics is also provided.

2.4.2. Intersection

Intersection infrastructure data is similar to that of the highway with the addition of information on cross street. In this case infrastructure data includes district, county, route (including suffix if any) and post mile (including prefix and suffix, if any) of the mainline, in addition to geometric characteristics, traffic volume and additional features of the mainline as well as the cross street.

2.4.3. Ramp

Ramp infrastructure location information is similar to highway and intersection data structure, and design characteristics include on/off ramp type, and design type. Additional information includes highway group and population code.

2.4.4. Collision Data

Collision data structure includes location information including district; county; route (including suffix if any), and post mile (including prefix and suffix, if any). To represent the facility type in the collision data, there is a field named ‘file type.’ In addition, collision description including date and time, lighting condition, and severity level are also required based on the data availability.

2.5. Summary

This chapter provides an overview of Safety Performance Functions (SPFs), network screening methods and data requirements. Additional insights are provided on ranking sites and prioritizing sites or locations based on Potential for Safety Improvement (PSI). The following chapter will describe the current approach being used by Caltrans for identifying high collision concentration locations.

Chapter 3. Understanding the Current Practice of Safety Analysis – Table C

Prior to developing California-specific Safety Performance Functions for network screening, it is important to identify the current practice in place to identify high collision concentration locations (HCCLs). This chapter summarizes the existing practice—Table C—used in the state of California for identifying high crash concentration locations on the SHS. In addition, this chapter explains the production of Table C through interaction between entities within Caltrans and its application in identifying HCCLs. Limited documentation, in conjunction with Caltrans' desire to evaluate the value of transitioning to other network screening methods, resulted in an effort to identify the entities that contribute to, or are a part of, this process. The research team introduced a new technique of process mapping that is used in this project to better understand the processes involved in the production Table C. This chapter concludes by mapping the Table C report as well as identifying key personnel involved in the production of this report.

3.1. Overview of Table C

Table C is the existing practice used in the state of California to identify HCCLs on the SHS. To identify accident rates along different highway facilities—segments, intersections, and ramps—which are significantly higher than the statewide average, periods of 36, 24, 12, 6, and 3 months are used. Generally, the Table C report is generated quarterly, but it can also be generated by special request. The process begins at the start of the route within a district. The first 0.2 miles segment is analyzed, and a significance test at 99.5% of significance factor is performed for highway segments, intersections and ramps. Accident investigators are required to examine those locations in the final output with locations that experience 4 or more accidents and are significant in either the 3-, 6-, or 12-month periods, subsequently these locations are labeled "REQ" in the output table.

3.2. Mapping Existing Process

To better understand the current practice, an attempt was made by the research team to map the Table C process used by Caltrans to identify HCCLs across California. Although necessary and specific knowledge about each of the individuals who are routinely involved in the generation of Table C was available, there was no documentation to provide a consistent and comprehensive understanding of the entire production process. With respect to the underlying assumptions that govern the Table C analysis, no information was readily available within the agency. The "Summary Report of Task Force's Findings and Recommendations" conducted by the Caltrans task force in 2002 helped to provide a theoretical background behind the algorithms used to process the collision data, but was limited in terms of identifying the key activities involved in the process. The "State Highway Safety Improvement Program Guideline" (HSIP 2017) describes the steps following HCCL identification by traffic safety engineers to assess these locations for potential improvement. The research team started at the point of the crash, which triggers the data used for HCCL identification. The next step was reviewing the police collision reports that are completed according to the California Highway Patrol Collision Investigation Manual. This is the most critical data point for this process, and it is collected by police departments across the state, and shared with Caltrans by the California Highway Patrol, an affiliated agency. Obtaining documentation about the entities that contribute to, or are a part of, the process can facilitate an understanding of the interactions and information flow that govern the production of Table C, as well as the impact of the entities interfacing with Table C. Therefore, the research team applied the three process maps—relationship maps, cross-functional maps, and flowchart maps—to visualize the process involved in identification of HCCLs using Caltrans' Table C report.

To map the processes, it was necessary to collect data directly from relevant personnel across different entities within the organization, due to the lack of comprehensive documentation available at Caltrans covering the entire process related to the production of the Table C report. This was achieved by a two-stage methodology (i) Stage 1: Questionnaire survey, and (ii) Stage 2: Stakeholder interview—adopted within Caltrans.

3.2.1. Questionnaire Survey

The first stage in the data collection involved a questionnaire survey, which was conducted primarily to obtain preliminary information on the process and identifying key personnel involved using Qualtrics. Two main components were considered in preparing the questionnaires. First, the questions included in the survey were designed to obtain information relevant to the relationship map. The relationship map requires inputs from key offices within the relevant divisions involved in the process, along with their respective inputs and outputs. The questions included each respondent's corresponding office, the services/deliverables he or she provides, and the services/deliverables received regarding Table C. Second, the relevant subjects of the questionnaire were identified. In total, 80 questionnaires were circulated among Caltrans entities—Division of Research, Innovation and System Information (DRISI), Division of Traffic Operations, and Districts within Caltrans—to acquire as much information as possible to understand the role and responsibility of each entity, and key personnel involved. Figures 3.1, 3.2 and 3.3 shows the questionnaire surveys that were developed for DRISI, Traffic Ops, and Districts, respectively.

Table C/Table C Wet Process Questionnaire (DRISI)	DRISI

Contact Information:	Your name	Title	Office & Division	Email & Phone number

In your response below, please include in your deliverables pertaining to: Infrastructure, Crash, Traffic Data, and Table C Report Generation						
Item	Question	Office in Charge of Providing	-->	Services/Deliverables -->	Receiving Office	
		Office	Contact Information		Office	Contact Information
1	What services/deliverables do offices in your division provide to entities outside DRISI with respect to Table C/Table C Wet process? **(From DRISI)**					
2	What services/deliverables do entities outside DRISI provide to your offices with respect to Table C/Table C Wet process? **(To DRISI)**					
3	What services/deliverables does each office provide to other offices in DRISI division with respect to Table C/Table C Wet process? **(Within DRISI)**					
4	Which office inputs data into TASAS?			Inputs Collected Data	TASAS	
Please add any additional comments						

Figure 3.1 Questionnaire survey – DRISI

Table C/Table C Wet Process Questionnaire (Traffic Operations)	Traffic Ops

Contact Information:	Your name	Title	Office & Division	Email & Phone number

In your response below, please include in your responses Reviewing, Approving, Distributing, and Following up on Table C Report

Item	Question	Office in Charge of Providing	-->	Services/Deliverables -->	Receiving Office	
		Office	Contact Information		Office	Contact Information
1	What services/deliverables do offices in your division provide to entities outside Traffic Operations with respect to Table C/Table C Wet process? **(From Traffic Operations)**					
2	What services/deliverables do entities outside Traffic Operations provide to your offices with respect to Table C/Table C Wet process? **(To Traffic Operations)**					
3	What services/deliverables does each office provide to other offices in Traffic Operations division with respect to Table C/Table C Wet process? **(Within Traffic Operations)**					
4	What criteria are used to prioritize HCCLs and which office is responsible for the prioritization					
5	Are the rate group tables updated and by which office?					
	Please add any additional comments					

Figure 3.2 Questionnaire survey – Traffic Operations

Table C/Table C Wet Process Questionnaire (District)	District

Contact Information:	Your name	Title	Office & Division	Email & Phone number

In your response below, please include in your responses Receiving, Distributing, Reviewing, and Reporting Back Table C Report

Item	Question	Office in Charge of Providing	-->	Services/Deliverables -->	Receiving Office	
		Office	Contact Information		Office	Contact Information
1	What services/deliverables do offices in your division provide to entities outside District with respect to Table C/Table C Wet process? **(From District)**					
2	What services/deliverables do entities outside District provide to your offices with respect to Table C/Table C Wet process? **(To District)**					
3	What services/deliverables does each office provide to other offices in District division with respect to Table C/Table C Wet process? **(Within District)**					
4	Which office is responsible for carrying out site visits to assess whether the site can be improved					
5	What are the criteria used in assessing whether the HCCLs requie improvement or not?					
6	Which office is responsible for updating the infrastructure data once an improvement has been made?					
	Please add any additional comments					

Figure 3.3 Questionnaire survey – District

Data collected from DRISI, Traffic Ops and Districts using the questionnaire survey were then processed. The output from this stage was followed by one-on-one interviews with key personnel involved in the production of Table C in Stage 2.

3.2.2. Stakeholder Interviews

The second stage in the data collection involved stakeholder interviews. This was implemented based on the responses to the questionnaire survey, relevant information regarding the specific offices within the divisions, and collection of their respective input/output. A preliminary relationship map was developed using the available data. Interestingly, the preliminary relationship map revealed inconsistent and often contradictory data. The responses were also used to identify three key individuals who are instrumental to the production of the Table C report. Based on information gathered from the questionnaire, interviews were scheduled with the three key personnel from the Collision Coding Unit and IT-Client Support Server of DRISI, and District, to obtain more specific information.

The interview process was iterative, in which each interviewee provided a piece of information, which was used to develop three different process maps. While developing these maps, gaps in information emerged. The process of interviews/correspondence with the three key personnel was repeated until all gaps were eliminated. In summary, the three interviews along with follow-up communications for clarification provided enough data for mapping of the three maps.

3.3. Process Mapping

The process mapping technique used in this study relies on three types of maps—relationship maps, cross-functional process maps, and flowchart process maps. These maps are used to generate necessary information across a range of industries for a variety of purposes. The sections below describe each of these maps, and how they were developed in the current context.

3.3.1. Relationship Map

Relationship maps are used to show responsibilities and expectations between organizations or among different entities within the same organization. This type of map can also help identify the input required and output produced by an organization, divide the organization into individual components working together, show what each entity of an organization produces, and can be used to familiarize members of an organization with the entities and products involved.

The first step in creating the relationship map was to identify the relevant parties involved in the process, using collected data. The main divisions involved in generating Table C are as follows: California Highway Patrol (CHP), Caltrans Division of Research, Innovation and System Information (DRISI), Caltrans Division of Traffic Operations, and the Caltrans Districts, in addition to some key Information Technology (IT) personnel, SWITRS, and TASAS-TSN. The second step was to determine how the entities interact with each other. This was accomplished by establishing one of the entities as the supplier and the other as the customer. Some of these entities were responsible for several different deliverables within the Table C process. To further refine the responsibility for each deliverable, DRISI was then divided into the Collision Coding Unit (CCU) and the IT-Client Support Server.

Once the main entities were identified, the input/output of each and the interactions among these entities were mapped. The mapping was conducted by gradually placing entities from left to right—the left side comprised of entities involved early on in the process, and the right side consisting of entities involved in the end of the process. The arrows connecting these entities represent deliverables generated by one entity and received by another as indicated by the direction of the arrows. The relationship map (Fig. 3.4) shows the key stakeholders (DRISI, Traffic Ops, CCU) and resources (State-wide Integrated Traffic Records System (SWITRS), Traffic Accident Surveillance and Analysis System-Transportation System Network (TASAS-TSN) involved in producing Table C, in addition to the interactions among them.

Table C Relationship Map

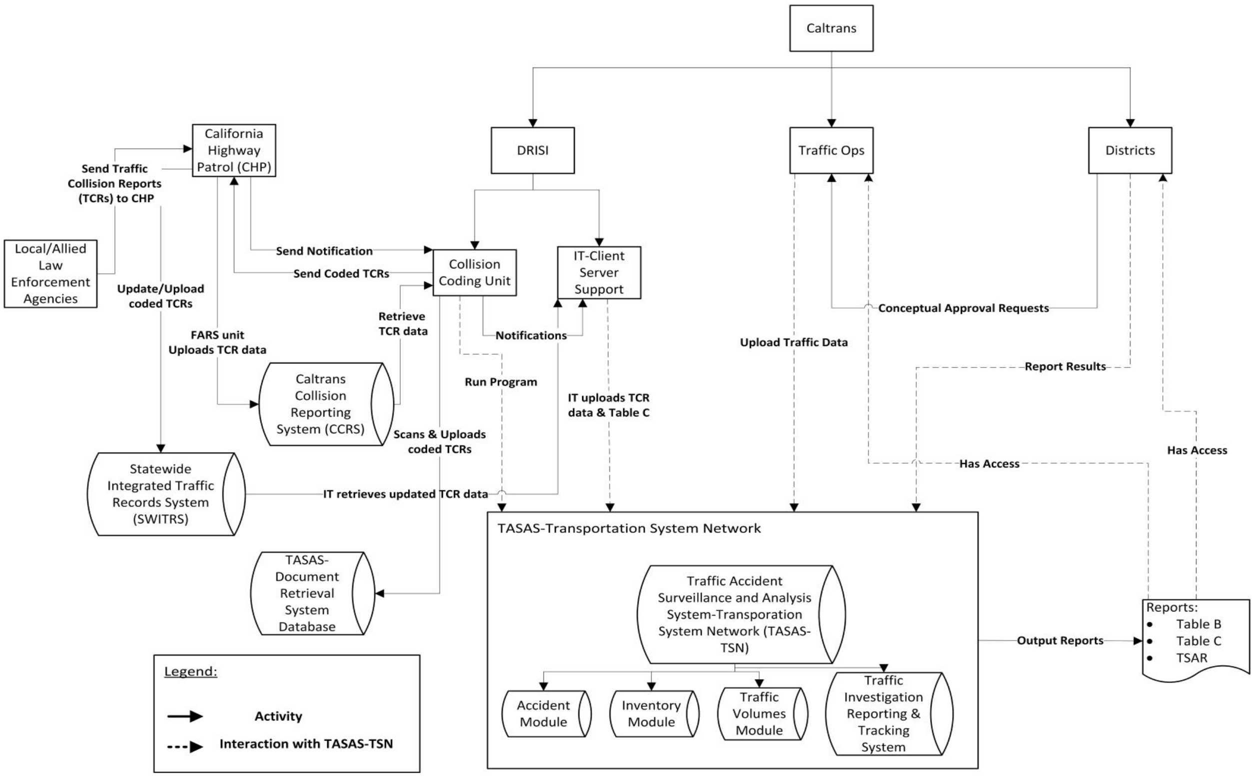

Figure 3.4 Relationship map developed for Table C production

3.3.2. Cross-Functional Process Map

Cross-functional process maps show workflow, which is composed of activities that are performed by entities within an organization. This type of map illustrates how work activities flow within a certain entity of an organization and the handoffs between the organizations. It is also used to show the beginning and end of a process, while highlighting the specific activities for which each part of an organization is responsible and identifying which parts of an organization interact with each other.

Cross-functional process maps are also known as swim lane diagrams because each entity in the map is represented by a horizontal band stacked on top of other bands, similar to a competition swimming pool viewed from above. Creating a cross-functional process map involves the entities involved in the process, and the processes for which each entity is responsible. The entities involved in the process along with their inputs/outputs can be obtained from the relationship map which was completed first. In the case of the HCCL identification process using Caltrans' Table C report, the processes within each entity were obtained through interviews. The focus was on one entity at a time to help obtain the full set of activities for which each individual entity is responsible. After collecting the relevant information, the cross-functional process map was developed as shown in Fig. 3.5.

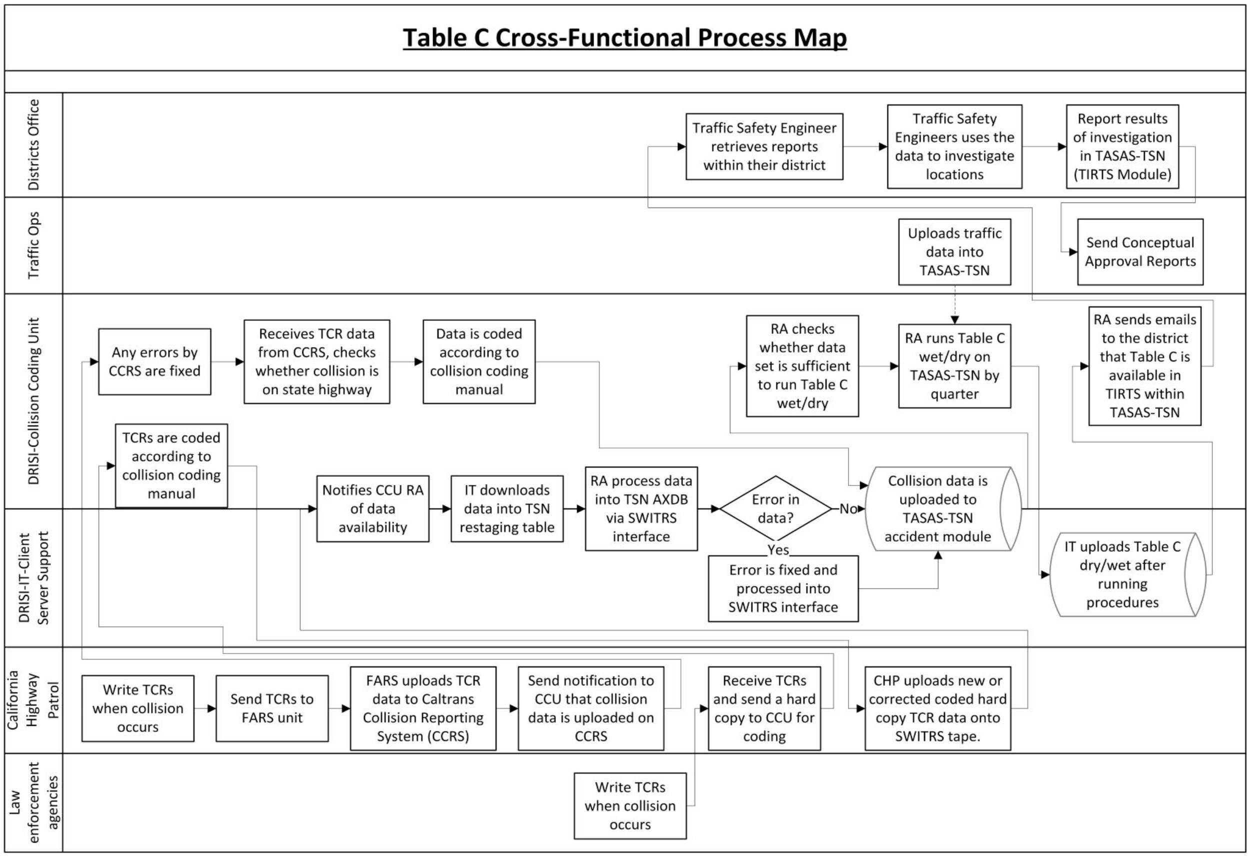

Figure 3.5 Cross-functional process map developed for Table C Production

The top horizontal band in the map shows the entity responsible for the final output, in this case, the Districts. Additional bands were allocated for the entities responsible for the various processes consecutively until the beginning process was reached. Within each band the activities performed by that entity are placed in order from left to right connected by arrows. The final activity performed by that entity is connected with an arrow to the starting activity to be performed by another entity in another swim lane.

3.3.3. Flow Chart Map

Flowcharts are used to represent graphically the sequence of activities involved in producing an output or in providing a service. This type of map captures value adding activities, as well as non-value adding activities—delays due to inspection, approval processes.

Flowchart maps are intended to show the most granular level of workflow. To reach the desired level of granularity, the work comprising different activities must be mapped out. Fig. 3.6 shows the flowchart map—the activities performed by each of the different entities were already mapped in the cross-functional diagram. In the flowchart map, it is necessary to divide up these activities into more detailed work tasks. Additional data was collected via correspondence to identify the steps involved in each activity in the Table C development process. After collecting the required data, the work comprising all the activities was placed in order from left to right (beginning to end).

Table C Flowchart Map

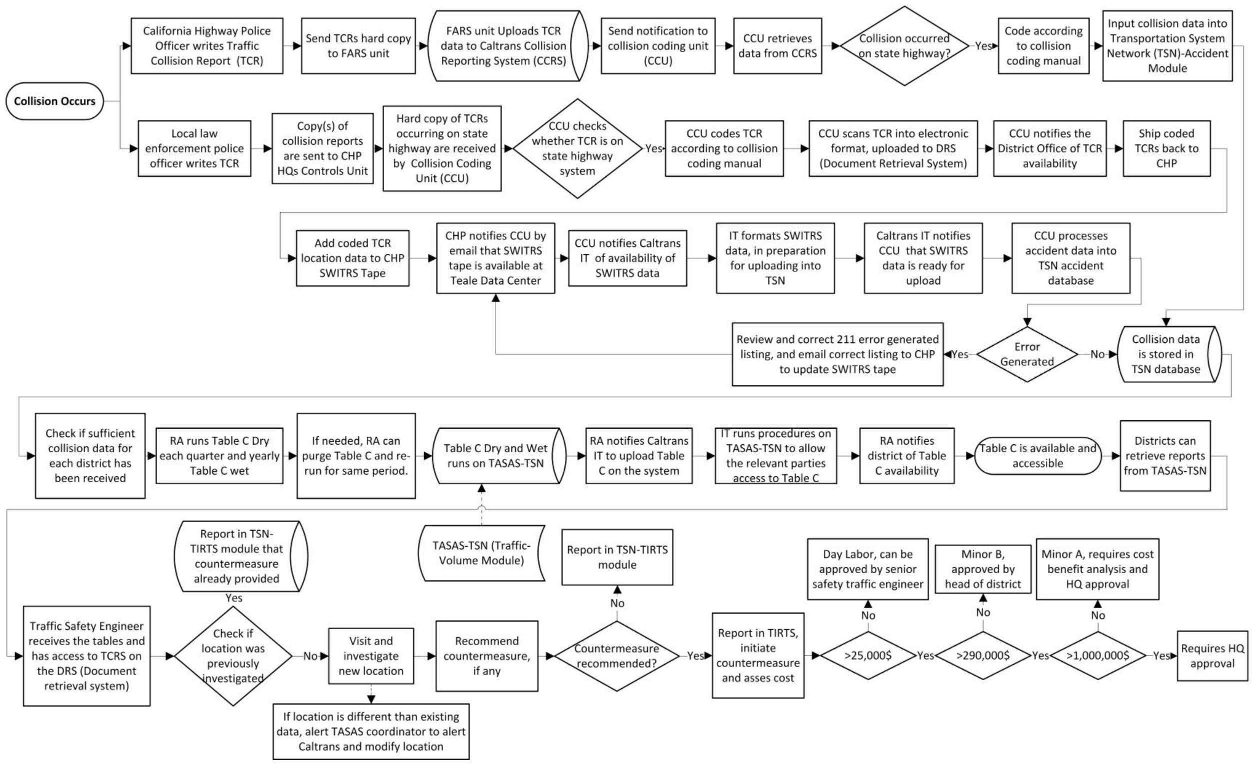

Figure 3.6 Flowchart map developed for Table C production

3.4. Discussion on Table C Maps

The level of detail presented in the three maps increases from the relationship map to the cross-functional process map, to the flowchart map. The relationship map (Fig. 3.4) shows the key stakeholders (e.g., DRISI, Traffic Ops, CCU) and resources (SWITRS, TASAS-TSN) involved in producing Table C, in addition to the interactions among them. The cross-functional map (Fig. 3.5) is most useful for addressing the boundaries of the Table C process (start/end)—i.e. the core decision-making process and the points of handoff between different offices and divisions. This map also helps identify the locations of key activities. Finally, the flowchart process map (Fig. 3.6) provides in-depth information on each activity shown in the cross-functional process map and describes the roles and responsibilities of each unit/entity and the next steps.

Each of the maps developed for this documentation process contributed to other insights and together provided the agency with a robust description of the legacy process. From an organizational perspective, these maps are useful in teaching new employees about various organization processes while also acting as a basis to improve the process and reduce the required lead time. Each type of map is designed to show different aspects of the modeled process, ranging from high-level interaction among entities (relationship map) to detailed, task-level representation (flowchart map). Following are specific insights on the functional capabilities of each map:

Relationship Map:

The relationship map helps to identify individual stakeholders who directly interact with the TASAS-TSN database. This functionality provides the appropriate level of detail to identify the entities who need to be involved when revising the decision-making process, to ensure backward and forward compatibility with the input/output.

Cross-functional Process Map:

The cross-functional process map provides key insights in assigning responsibilities to key players, and isolating handoff activities. This structure provides the necessary granularity to identify specific points of interventions where new stakeholders/resources can be integrated into the decision-making process. For example, the current workflow for Table C does not incorporate information about pavement quality that is periodically collected by the office of pavement management. This information can be conceptually integrated by introducing a new swim-lane corresponding to the pavement management office and identifying the appropriate handoff points.

Flowchart Map:

The flowchart map provides a task-by-task representation of the overall process. This level of detail allows the introduction of automation and quality control features to ensure the reliability and robustness of system performance. For example, if additional variables from crash narratives (e.g., manual vs automated mode of vehicle) are suggested in the future, the flowchart map can pinpoint the specific points of intervention where these changes need to be executed.

3.5. Summary

This chapter summarizes the current practice of identifying HCCLs on the state highway system by Caltrans. This provides a good starting point to understand the responsibilities of each of the entities involved in the safety investigation from the collision occurrence to the proper countermeasure. The next chapter will provide valuable information about the development of Safety Performance Functions (SPFs), and the data challenges in developing SPFs.

Chapter 4. California-Specific Safety Performance Functions (SPF)

This chapter describes the need for developing California-specific SPFs. The chapter also outlines the step-by-step process of developing SPFs, the data required as well as results of developed SPFs. The chapter concludes with the challenges in SPF implementation for identifying high collision concentration locations.

4.1. Need for California-specific SPF

The SPFs developed in the first edition of HSM 2010 are based on data from certain jurisdictions and won't be applicable for other jurisdictions. Hence HSM suggests two options to use SPFs—calibrating the existing SPFs or developing site-specific SPFs—based on certain conditions. To best apply the SPFs for identifying high crash concentration locations along the California SHS, Caltrans' maintains infrastructure and collision data is well suited for the second option of developing California-specific SPFs.

4.2. Applications of SPF

This section will provide a description of application of SPFs specifically for California. SPFs are useful in identifying HCCLs, but the challenge is understanding the need and the specifications required—facility-based SPF and/or injury-based SPF—which depends mainly upon data availability.

4.3. Development of SPF

While the main goal of this project is implement existing SPFs for identifying HCCLs, the research team also made an effort to re-estimate and develop new California-specific SPFs that should be included in the tool, which will be developed as part of this project. The two types of SPFs are based on HSM 2010 and described in more detail in Section 2.1.1. They are as follows:

- Type 1 SPFs include functional forms in which the independent variables include an intercept and average daily traffic (ADT).
- In Type 2 SPFs, the estimating equation includes roadway geometry variables and intersection design elements in addition to the length and ADT effects.

Based on the discussions/interviews with the Caltrans' safety experts in the Division of Traffic Operations, these two types of SPFs were developed based on facility type and injury severity.

4.4. Facility-based SPF

Based on the three facilities along the California state highway system—segments, intersections, and ramps—three SPFs were developed as described them in the following sections.

4.5. Injury Severity-based SPF

The project also made an extensive effort to evaluate and optimize the methodology adopted for identifying HCCLs. The research team explored the possibility of using different combinations of injury collisions. Thus, three combinations of injury collisions were considered in developing facility-based SPFs as follows and described in the subsequent sections:

i. Total collision (TOT): includes fatal, severe, visible, compliant of pain and property damage only
ii. Fatal plus Severe (FS): a combination of fatal and severe injury collisions
iii. Fatal plus Severe plus Severe (FSV): a combination of fatal severe and visible injury collisions

Injury collisions and SPF category considered initially are shown in Table 4.1.

Table 4.1 Preliminary Injury-based SPF combinations

No.	**SPF Category**	**Injury Type**
1	All Collision	Total
2	Individual Severity Levels	Fatal
		Severe
		Visible
		Complaint of Pain
		PDO
3	All Injury	Total without PDO
4	Fatal & Severe	Fatal plus severe
5	Non-Severe Injury	Visible plus Complaint of Pain plus PDO

4.6. Steps in Developing SPFs

Developing California-specific SPFs is a five-step process as shown in Fig. 4.1. The process starts with identifying the facility type for SPF development. The following sections will describe each step.

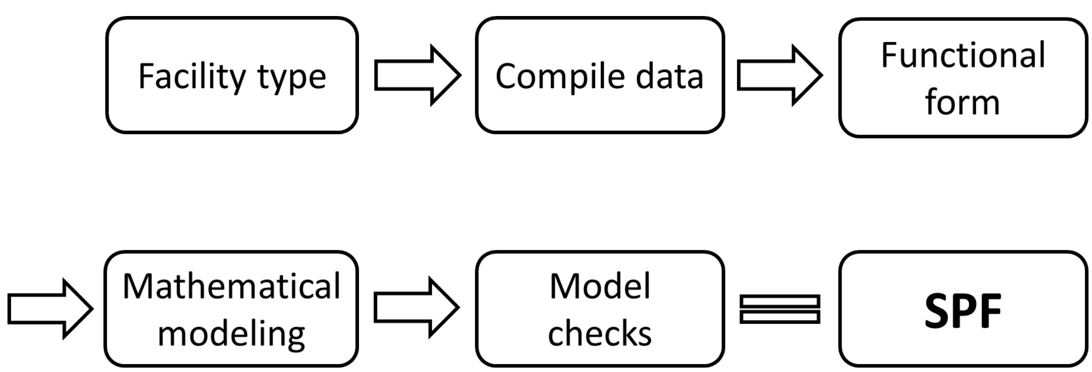

Figure 4.1 Flowchart of workflow in SPF development

4.6.1. Facility Type

The first and foremost step in the SPF development is the identification of facility type. In this project, SPFs were developed for the three infrastructure facility types—segment, intersection and ramp—along the state highway system within California as shown in Table 4.2.

Table 4.2 Summary statistics of Facility Type

State Highway System (~50,000 Miles)		
Segment (~13000 centerline miles)	Intersection (~16500)	Ramp (~14,000)

4.6.2. Data Requirements and Compilation

Data play a vital role in the development of SPFs. For re-estimating/developing California-specific safety performance functions two data sets are required: infrastructure and collision data. Infrastructure consists of three facility types—segment, intersection and ramp—while collision data provides relevant information pertaining to the collision occurrence.

As part of this project, infrastructure and collision data extracted from TASAS-TSN for the seven-year period between 2010 and 2017 were considered for initial SPF development efforts. Finally, the analysis converges on the most recently available five-year period between 2013 and 2017, using available infrastructure and collision data. The data dictionary used in the project is provided in *Appendix A*. Data from 2013 through 2015 were used for training the model, while data from 2016 and 2017 were used to test the developed model.

4.6.2.1. Segment infrastructure data

The infrastructure data considered incorporates location and geometric features of the highway system with different levels of aggregation. The infrastructure data structure for the highway includes data fields such as functional class, which helps to group segments, and begin and end date of the segment to identify active segments. The key variables of this data are *county, route, route suffix, post mile prefix, begin post mile, end post mile* and the *post mile suffix*. Segment infrastructure data dictionary used in this project is provided in *Appendix A*.

4.6.2.2. Intersection infrastructure data

In the case of intersection data, new location information was identified, including begin and end information for county, route, route suffix, post mile prefix, post mile, and end post mile based on the override length, in addition to the main location information. The key variables of this data are *begin county, begin route, begin route suffix, begin post mile prefix, begin post mile, begin post mile suffix, main county, main route, main route suffix, main post mile prefix, main post mile, main post mile suffix,* *and end county, end route, end route suffix, end post mile prefix, end post mile, end post mile suffix.* The intersection infrastructure data dictionary used in this project is provided in *Appendix A*.

4.6.2.3. Ramp infrastructure data

The infrastructure data considered incorporates ramp location and design features. The ramp infrastructure data is similar to existing structure except that the ramp description is added. In this case ramp collisions are marked as point locations, since the length of the ramp is unknown. The key variables of this data are *county, route, route suffix, post mile prefix, post mile,* and *post mile suffix.* The ramp infrastructure data dictionary used in this project is provided in *Appendix A*.

4.6.2.4. Collision data

Collision data considered includes location information, facility type, date and time of collision, highway group, population code and severity level. Each component of collision information should be assigned with one of five levels of collision severity: fatal, severe injury, visible injury, complaint of pain, and property damage only. The data field *file type* helps to identify the facility type where the collision occurred, and therefore only a single collision data file is required. The key variables of this data are *county, route name, route suffix, post mile prefix, post mile, post mile suffix, accident date* and the *accident time*. The collision data dictionary used for developing SPFs is provided in *Appendix A*.

District crash frequency is shown in Table 4.3. As shown in the table, the combination of fatal and severe collisions contributes only tiny portion of total crashes, while property damage only (PDO) collisions account for over half of crashes in all districts. Table 4.4 shows the summary of collision based on facility type.

Table 4.3 District Crash Frequency 2013-2017

Districts	Fatal	Severe	Visible	Complaint of Pain	PDO	Total
1	2%	6%	15%	17%	61%	11129
2	2%	5%	15%	18%	60%	10059
3	1%	3%	10%	23%	63%	53443
4	1%	2%	9%	24%	65%	168908
5	1%	3%	10%	20%	66%	36559
6	1%	3%	11%	20%	65%	44855
7	0%	1%	8%	22%	68%	251514
8	1%	2%	10%	22%	64%	101985
9	2%	5%	16%	11%	66%	3404
10	1%	3%	12%	22%	62%	38179
11	1%	3%	12%	28%	57%	63228
12	0%	2%	9%	23%	66%	74052

Table 4.4 Summary of Collision based on Facility Type

Facility type	Year				
	2013	**2014**	**2015**	**2016**	**2017**
Segment	123,526	124,401	136,085	161,737	163,627
Intersection	7,091	7,372	7,772	8,400	7,497
Ramp	20,721	20,981	21,861	24,185	23,769

4.6.3. Data Challenges

Frequency of data updates is a major challenge, in addition to the following:

1. Inconsistent data updates
   - It is necessary to identify the most appropriate ADT (mainline and cross-street) for every year
2. Missing attributes—geometric characteristics including number of lanes
3. Placeholders and outliers ADTs

4.6.4. Functional Form

The functional form is specified as a logarithmic function representation of the event rate, in the case of SPFs, it is the number of crashes occurring each year. The functional forms considered for Type 1 and 2 SPFs for all the three-facility types—segment, intersection and ramp—are described in the subsequent sub-sections.

4.6.4.1. Segment

In the case of segments, the length of the segment is used as an offset in the case of Type 1 SPFs, which implies that the coefficient for segment length is unity. For Type 2 SPFs, the estimating equation includes roadway geometry variables and intersection design elements, in addition to length and ADT effects. Therefore, given a vector of geometric effects Zij and associated coefficients γjj, functional forms considered for Type 1 and 2 segment SPFs are as shown in equations 4.1 and 4.2 respectively.

Type 1 segment SPF:

$$\lambda_i = length_i * e^{\alpha} * ADT_i^{\beta} \tag{4.1}$$

where, 'λ' is the expected number of collisions; 'α' is the intercept; and 'β' is the coefficient of ADT (captures the effect of variable on the number of collisions). Length is assumed as an offset variable.

Type 2 segment SPF:

$$\lambda_i = length_i * e^{\alpha} * \mathrm{ADT}_i^{\beta} * e^{\sum_{j=1}^{l} \gamma_j Z_{ij}} \tag{4.2}$$

where, 'λ' is the expected number of collisions; 'α' is the intercept; and 'β' is the coefficient of ADT; 'Z' is the geometric and other segment characteristics, while 'γj' is the coefficient of 'Z'.

4.6.4.2. Intersection

In the case of intersections, mainline and cross-street traffic volumes are considered separately for Type 1 and 2 SPFs. For Type 2 intersection SPFs, the estimating equation includes roadway geometry variables and intersection design elements in addition to the ADT effects. Therefore, given a vector of geometric effects Zij and associated coefficients γjj, functional forms considered for Type 1 and 2 intersection SPFs are as shown in equations 4.3 and 4.4 respectively.

Type 1 intersection SPF:

$$\lambda_i = e^{\alpha} * \mathrm{ADT}_{(main)i}^{\beta 1} * \mathrm{ADT}_{(x\text{-}street)i}^{\beta 2} \quad (4.3)$$

where, 'λ' is the expected number of collisions; 'α' is the intercept; '$\beta 1$' is the coefficient of mainline ADT; and '$\beta 2$' is the coefficient of cross-street ADT (captures the effect of variable on the number of collisions).

Type 2 intersection SPF:

$$\lambda_i = e^{\alpha} * \mathrm{ADT}_{(main)i}^{\beta 1} * \mathrm{ADT}_{(x\text{-}street)i}^{\beta 2} * e^{\sum_{j=1}^{l} \gamma_j Z_{ij}} \quad (4.4)$$

where, 'λ' is the expected number of collisions; 'α' is the intercept; and 'β' is the coefficient of ADT; 'Z' is the geometric and other site characteristics, while 'γj' is the coefficient of 'Z'.

4.6.4.3. Ramp

In the case of ramps, only traffic volume is considered for Type 1 SPFs since ramp length infrastructure data is not available at this point. Type 2 ramp SPF considers ramp geometry variables and design elements in addition to ADT effects. Therefore, given a vector of geometric effects Zij and associated coefficients γjj, functional forms considered for Type 1 and 2 intersection SPFs are as shown in equations 4.5 and 4.6 respectively.

Type 1 ramp SPF:

$$\lambda_i = e^{\alpha} * ADT_i^{\beta} \quad (4.5)$$

where, 'λ' is the expected number of collisions; 'α' is the intercept; and 'β' is the coefficient of ADT (captures the effect of variable on the number of collisions).

Type 2 ramp SPF:

$$\lambda_i = e^{\alpha} * \mathrm{ADT}_i^{\beta} * e^{\sum_{j=1}^{l} \gamma_j Z_{ij}} \quad (4.6)$$

where, 'λ' is the expected number of collisions; 'α' is the intercept; and 'β' is the coefficient of ADT; 'Z' is the geometric and other ramp characteristics, while 'γj' is the coefficient of 'Z'.

4.6.5. Mathematical Modeling

Crash occurrences are rare and random events, therefore count data modeling—Poisson regression and negative binomial regression—of Generalized Linear Model were used in this project.

Negative binomial distribution is more general than Poisson distribution because it has a variance that is greater than its mean, making it suitable for count data that do not meet the assumptions of the Poisson distribution (mean is equal to variance).

This project considered different statistical tools for analysis. Statistical tool STATA as well as R package were used for mathematical modeling.

4.6.6. Model Checks

Measures to evaluate the performance of models used are mean absolute deviation, mean squared prediction error, and root mean square error which are explained below.

4.6.6.1. Mean absolute deviation

The mean absolute deviation (MAD) of a set of data is the average distance between each data point and the mean value of the data set. MAD gives the difference in prediction of the models in an absolute format, as given by Eqn. (4.7). A value closer to zero shows that the model will perform well, when compared with the observed data.

$$\text{MAD} = 1/n\sum_{i=1}^{n}\left|x_i - \bar{x}\right| \tag{4.7}$$

4.6.6.2. Mean squared prediction error

The mean square percentage error (MSPE) is a measure of accuracy of the model in statistics, as given by Eqn. (4.8). It usually expresses error as a percentage.

$$\text{MSPE} = 1/n\sum_{i=1}^{n}\left|\frac{x_i - \bar{x}}{x_i}\right|^2 \tag{4.8}$$

Where, ‘ x_i ’ denotes the i^{th} individual value and ‘$\bar{x}$’ is the mean value.

4.6.6.3. Root mean square error

The root-mean-square error (RMSE) is a frequently used measure of the differences between values (sample or population values) predicted by a model or an estimator and the values observed.

4.6.6.4. Variable significance

Variable significance considers whether a variable considered for the analysis is statistically significant. In this project only variables that were statistically significant at the 5% significance level were included in the model.

4.7. Data Analysis

Based on the above steps, detailed data analysis plan considered in this project for developing Type 1 and 2 SPFs is shown in tables 4.5 and 4.6.

Table 4.5 Data analysis plan for Type 1 SPFs

Functional Component	Variables Considered Previously	Additional Considerations for SPF Implementation	Plan	Evaluation Criteria
Roadway segment	Length & ADT	- Length of segment as a explanatory as well as offset variable - Collision data to be segregated as with and without PDO	Estimate and test the model with most recent data: SPF estimation - 2013-15 SPF testing - 2016-17	1. Assess statistical significance of new and existing variables 2. Evaluate the suitability of new SPF categories (sample sizes, predictive ability)
Intersection	ADT (mainline and cross-street)	- Intersection type and control conditions - Functional classes of highway - With and without PDO - Remove cross-street ADT and identify potential proxies		
Ramps and ramp metering	ADT	- Ramp configuration - Functional classes of highway - With and without PDO		

Table 4.6 Data analysis plan for Type 2 SPFs

Functional Component	Variables Considered Previously	Additional Considerations for SPF Implementation	Plan	Evaluation Criteria
Roadway segment	• Geometric characteristics • Year dummies (time) • Route dummies • County dummies (spatial)	- Collision data to be segregated as with and without PDO - Include additional geometric characteristics	Estimate and test the model with most recent data: SPF estimation - 2013-15 SPF testing - 2016-17	1. Assess statistical significance of new and existing variables 2. Evaluate the suitability of new SPF categories (sample sizes, predictive ability)
Intersection	• Geometrics of mainline • Intersecting roadway characteristics • Attributes of intersection – intersection geometry • Traffic signal control type • Turn lane treatment	- Intersection type and control conditions - Functional classes of highway - With and without PDO		
Ramps and ramp metering	• Ramp control type • Presence of HOV lane • On-ramp/off-ramp	- Ramp configuration - Functional classes of highway - With and without PDO		

Data preparation stage which includes data cleaning is the most important element of the data analysis. The step-by-step procedure is as follows:

I. Required data files

- Clean infrastructure data of each facility type separately
- All facility type should have observations corresponding to the analysis period (usually 5 years). Here the training data is 2010-2014 (5 years) and the test data is 2015-2017 (3 years). Later in the final stage, SPFs were developed using 2013-2015 data.

II. Data Cleaning

- From the infrastructure data files, remove the variables which are not required (see data dictionary in *Appendix A*), and check for completeness of each variable and for the analysis period.

III. Merging data sets

a) Merge required variables from the infrastructure data file to the corresponding fields (Postmile) and create dummy variables for the route, year and county—Version 1.0
b) Merge the collision data with the Version 1.0 infrastructure data file for the entire analysis period and create a full data set for each facility type to be ready for analysis—Version 2.0

Identify the best fit models based on their predicted performance (Type 1 or Type 2 SPF based on injury type) for each facility type. Performance tests were performed as described in Section 4.6.6 to identify the best fit model.

Table 4.7 Summary of Data Analysis based on SPF Class

SPF Class	Segment		Intersection						Ramp	
			Rural		Urban Signalized		Urban Unsignalized			
	Total	FSV	Total	FSV	Total	FSV	Total	FSV	Total	FSV
Total				I1		I2		I3		R1
Rural 2 Lane Non-Freeway Undivided		H1								
Rural 2 Lane Non-Freeway Divided		H2								
Rural 3+ Lane Non-Freeway Undivided		H3								
Rural 3+ Lane Non-Freeway Divided		H4								
Rural 2-4 Lane Freeway Undivided		H5								
Rural 2-4 Lane Freeway Divided		H6								
Rural 5+ Lane Freeway		H7								
Urban 2 Lane Non-Freeway Undivided		H8								
Urban 2 Lane Non-Freeway Divided		H9								
Urban 3+ Lane Non-Freeway Undivided		H10								
Urban 3+ Lane Non-Freeway Divided		H11								
Urban 2-7 Lane Freeway Undivided		H12								
Urban 2-7 Lane Freeway Divided		H13								
Urban 8+ Lane Freeway Divided		H14								
Rural L/R Alignment Freeway		H15								
Urban L/R Alignment Freeway		H16								
L/R Alignment Non-Freeway		H17								

4.7.1. Estimation Process

The following stages were considered in the estimation process of the three facility-based SPFs:

- Stepwise elimination process to identify significant variables at 5% significance level
- Stepwise regression results many route and county dummies
- Clustering approach to combine routes and counties with similar coefficients

4.7.2. Clustering Approach

This project adopts K-means clustering. The K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. The 'means' in the K-means refers to averaging of the data; that is, finding the centroid. K-means clustering was used in this project to group large number of statistically significant 'Routes' and 'Counties' based on their mean values for all three facility-based SPFs.

4.8. California-specific SPFs

California-specific SPF were achieved following the above five steps. The sections below summarize the results of all facility-based and injury-based SPFs developed.

4.8.1. Segment SPF

Data for roadway segments were assembled for the entire state network consisting of over 50,000 lane miles of roadway. Infrastructure and collision data for the most recent five-year period (2013-2017) were collected from TASAS. Segment SPFs excluded intersection ranges.

4.8.1.1. Segment SPF Class

Segments are classified into 17 classes based on population group, number of lanes, traffic volume, functional class, and highway group. Table 4.8 provides the results of segment SPF classes.

Table 4.8 Segment SPF Class

SPF Class	SPF Class
1	Rural 2 Lane Non-Freeway Undivided
2	Rural 2 Lane Non-Freeway Divided
3	Rural 3+ Lane Non-Freeway Undivided
4	Rural 3+ Lane Non-Freeway Divided
5	Rural 2-4 Lane Freeway Undivided
6	Rural 2-4 Lane Freeway Divided
7	Rural 5+ Lane Freeway
8	Urban 2 Lane Non-Freeway Undivided
9	Urban 2 Lane Non-Freeway Divided
10	Urban 3+ Lane Non-Freeway Undivided
11	Urban 3+ Lane Non-Freeway Divided
12	Urban 2-7 Lane Freeway Undivided
13	Urban 2-7 Lane Freeway Divided
14	Urban 8+ Lane Freeway Divided
15	Rural L/R Alignment Freeway
16	Urban L/R Alignment Freeway
17	L/R Alignment Non-Freeway

4.8.1.2. Segmentation

Segmentation is a key process in the development of segment SPF, which generates homogenous highway segments. Homogeneity is typically defined based on location and geometric characteristics, as shown in Table 4.9. The segmentation process includes two main stages as follows:

- Removing overlap from discontinuous intersection buffers
- Merging two contiguous segments only if merge criteria is identical for all years

Table 4.9 Location and geometric characteristics considered for Segmentation

Location	Geometric
County	Number of lanes
Route	Lane width
Route Suffix	Inside shoulder width
PM Prefix	Outside shoulder width
PM Suffix	Median width
Population group	Design speed
Begin and End PM	Intersection influence area (N distance)

Intersection buffer removal stage:

1. Create a table of temporary intersection buffers that are homogenous
   - If no discontinuity, use intersection information as-is
   - If discontinuity is present, create copies of the intersection corresponding to each merge scenario
2. Check for overlapping intersections, and aggregate them
3. Remove the aggregated intersections from segments

Segment length distributions were examined by SPF class and are shown in Table 4.10. As shown in the table, the 25th percentile of the segment lengths are less than 0.1 miles and median is approximately less than or equal to 0.2 miles. This has implications for network screening. After segmentation it was observed that 30% of segments that are longer than 3 miles are included in the TASAS infrastructure data itself.

Table 4.10 Distribution of Segments Length based on Segmentation

SPF Class #	SPF Class	#obs	Segment Length				
			Min	25th Percentile	Median	99th Percentile	Max
1	Rural 2 Lane Non-Freeway Undivided	12196	0.001	0.06	0.22	4.13	16.40
2	Rural 2 Lane Non-Freeway Divided	886	0.001	0.02	0.06	0.92	6.33
3	Rural 3+ Lane Non-Freeway Undivided	729	0.001	0.05	0.13	1.37	2.74
4	Rural 3+ Lane Non-Freeway Divided	1097	0.001	0.03	0.08	2.21	5.57
5	Rural 2-4 Lane Freeway Undivided	1380	0.001	0.08	0.24	3.42	6.03
6	Rural 2-4 Lane Freeway Divided	3016	0.001	0.03	0.14	6.63	25.60
7	Rural 5+ Lane Freeway	947	0.001	0.03	0.11	4.03	9.17
8	Urban 2 Lane Non-Freeway Undivided	2737	0.001	0.02	0.07	1.35	2.33
9	Urban 2 Lane Non-Freeway Divided	724	0.001	0.02	0.04	0.65	1.67
10	Urban 3+ Lane Non-Freeway Undivided	846	0.001	0.01	0.02	0.48	0.74
11	Urban 3+ Lane Non-Freeway Divided	4197	0.001	0.01	0.03	0.63	2.42
12	Urban 2-7 Lane Freeway Undivided	240	0.001	0.05	0.17	2.27	2.75
13	Urban 2-7 Lane Freeway Divided	5294	0.001	0.03	0.10	3.07	12.00
14	Urban 8+ Lane Freeway Divided	5632	0.001	0.04	0.11	1.88	9.32
15	Rural L/R Alignment Freeway	379	0.001	0.07	0.26	4.25	13.20
16	Urban L/R Alignment Freeway	419	0.002	0.05	0.12	1.59	2.76
17	L/R Alignment Non-Freeway	296	0.001	0.00	0.02	1.55	2.18

4.8.1.3. Summary of Segment SPFs

After processing the segment infrastructure and collision data through each of the SPF development steps, Type 1 and 2 SPFs were developed for Total and FSV collisions. FS-based SPFs fail to provide good results due to limited observations. Furthermore, based on the performance measures, Type 2 FSV SPF performs better than the Total collision SPF.

Three main points observed from the model outputs are as follows:

- Sites based on total collisions are likely to be influenced by PDOs and complaint of pain
- Fatal + severe collisions may yield few collisions for investigators to recommend countermeasures
- Performance of sites identified using FSV collisions can provide higher resolution for investigation purposes, while limiting influence of PDOs

Segment SPF Class 1 – Rural 2 Lane Non-Freeway Undivided

In the case of Rural 2 Lane Non-Freeway Undivided SPF, FSV-based SPF is identified and is shown in Eqn. 4.9. Furthermore, the model estimates, including goodness-of-fit measures, clustering results of route and county for the Rural 2 Lane Non-Freeway Undivided SPF are shown in tables 4.11, 4.12, and 4.13.

$$Predicted\ Crashes\ (Np) = exp\,(-6.800 + 0.737 * X1 + 0.939 * X2 - 0.498 * X3 - 0.500 * X4 - 0.154 * X5 - 0.162 * X6 + 0.82 * X7 - 0.224 * X8 - 0.225 * X9 - 0.151 * X10 - 0.030 * X11 - 0.274 * X12 + 0.577 * X13 + 1.103 * X14 + 0.270 * X15 + 0.299 * X16 - 0.250 * X17 - 0.578 * X18 + 0.094 * X19\,) \quad (4.9)$$

Table 4.11 Significant Variables - FSV Rural 2 Lane Non-Freeway Undivided SPF

Code	Variable	Estimate
	Intercept	-6.8
X1	ADT	0.737
X2	Segment length	0.939
X3	Terrain: F	-0.498
X4	Terrain: R	-0.5
X5	Right surface type: H	-0.154
X6	Right surface type: B	-0.162
X7	Median type: A	0.82
X8	Year_1	-0.224
X9	Year_2	-0.225
X10	Year_3	-0.151
X11	Year_4	-0.03
X12	Route_cluster_1	-0.274
X13	Route_cluster_2	0.577
X14	Route_cluster_3	1.103
X15	Route_cluster_4	0.27
X16	County_cluster_1	0.299
X17	County_cluster_2	-0.25
X18	County_cluster_3	-0.578
X19	County_cluster_4	0.094
	Theta (overdispersion)	2.8673
	AIC	198288
	Log-likelihood	-99126.87

Considering the statistically significant 20 routes and 20 counties, it is difficult to incorporate all of these into the model. Therefore, the clustering approach was adopted as described in Section 4.7.2. Routes and counties were each clustered into four groups, as shown in Tables 4.12 and 4.13, respectively.

Table 4.12 Route Cluster in FSV Rural 2 Lane Non-Freeway Undivided SPF

1	2	3	4
12	199	263	96
152	26	2	67
138	79		121
62	39		180
	74		18
	227		38
	178		
	127		

Table 4.13 County Cluster in FSV Rural 2 Lane Non-Freeway Undivided SPF

1	2	3	4
SIE	STA	IMP	HUM
SCR	LAS	SHA	LAK
	NEV	SIS	CAL
	SM	KER	ORA
	MON	INY	SCL
	SLO		
	FRE		
	MNO		

Segment SPF Class 2 - Rural 2 Lane Non-Freeway Divided

In the case of Rural 2 Lane Non-Freeway Divided SPF, the FSV-based SPF is identified and shown in Eqn. 4.10. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Rural 2 Lane Non-Freeway Divided SPF as shown in tables 4.14, 4.15, and 4.16.

$$Predicted\ Crashes\ (Np) = exp\ (-1.597 + 1.163 * X1 + 0.352 * X2 - 1.621 * X3 - 0.351 * X4 - 1.379 * X5\ ) \tag{4.10}$$

Table 4.14 Significant Variables - FSV Rural 2 Lane Non-Freeway Divided

Code	Variable	Estimate
	Intercept	-1.597
X1	Segment length	1.163
X2	Route_cluster_1	0.352
X3	Route_cluster_2	-1.621
X4	County_cluster_1	-0.351
X5	County_cluster_2	-1.379
	Theta (overdispersion)	0.2622
	AIC	262745
	Log-likelihood	-131365.61

Table 4.15 Route Cluster in FSV Rural 2 Lane Non-Freeway Divided

1	2
101	36
199	96
76	165
50	3
37	227
	58
	62

Table 4.16 County Cluster in FSV Rural 2 Lane Non-Freeway Divided

1	2
LAK	AMA
SJ	MPA
NAP	TUL
SON	MNO
MON	
SB	
VEN	
FRE	
MAD	
RIV	

Segment SPF Class 3 - Rural 3+ Lane Non-Freeway Undivided

In the case of Rural 3+ Lane Non-Freeway Undivided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.11. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Rural 3+ Lane Non-Freeway Undivided SPF as shown in tables 4.17, 4.18, and 4.19.

$$Predicted\ Crashes\ (Np) = exp\ (-7.708\ +\ 0.613 * X1\ +\ 1.184 * X2\ -\ 0.273 * X3\ -\ 0.483 * X4\ -\ 0.155 * X5\ -\ 0.166 * X6\ +\ 0.319 * X7\ -\ 0.091 * X8\ +\ 0.060 * X9\ ) \tag{4.11}$$

Table 4.17 Significant Variables - FSV Rural 3+ Lane Non-Freeway Undivided

Code	Variable	Estimate
	Intercept	-7.708
X1	ADT	0.613
X2	Segment length	1.184
X3	Terrain: F	-0.273
X4	Median barrier: Z	-0.483
X5	Year 1	-0.155
X6	Year 2	-0.166
X7	Route_Cluster_1	0.319
X8	Route_Cluster_2	-0.091
X9	County_Cluster_1	0.060
	Theta (overdispersion)	0.6345
	AIC	230015
	Log-likelihood	-113951.25

Table 4.18 Route Cluster in FSV Rural 3+ Lane Non-Freeway Undivided

1	2
2	33
18	46
41	101
49	108
88	
168	
175	

Table 4.19 County Cluster in FSV Rural 3+ Lane Non-Freeway Undivided

1
HUM
MPA
SD
SIS
NAP
VEN
KER
RIV

Segment SPF Class 4 - Rural 3+ Lane Non-Freeway Divided

In the case of Rural 3+ Lane Non-Freeway Divided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.12. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Rural 3+ Lane Non-Freeway Divided SPF as shown in tables 4.20, 4.21, and 4.22.

$$Predicted\ Crashes\ (Np) = exp\,(-6.535 + 0.485 * X1 + 1.143 * X2 + 0.288 * X3 - 0.623 * X4 - 0.847 * X5 - 0.033 * X6 + 0.415 * X7 - 0.140 * X8 - 0.448 * X9 + 0.148 * X10\,) \quad (4.12)$$

Table 4.20 Significant Variables - FSV Rural 3+ Lane Non-Freeway Divided

Code	Variable	Estimate
	Intercept	-6.535
X1	ADT	0.485
X2	Segment length	1.143
X3	Left surface type: C	0.288
X4	Median barrier: Z	-0.623
X5	Median barrier: G	-0.847
X6	Year 3	-0.033
X7	Route_cluster_1	0.415
X8	Route_cluster_2	-0.140
X9	County_cluster_1	-0.448
X10	County_cluster_2	0.148
	Theta (overdispersion)	0.6622
	AIC	228596
	Log-likelihood	-114285.76

Table 4.21 Route Cluster in FSV Rural 3+ Lane Non-Freeway Divided

1	2
49	46
67	99
76	126
199	138

Table 4.22 County Cluster in FSV Rural 3+ Lane Non-Freeway Divided

1	2
COL	CAL
KER	HUM
LAS	LA
MNO	RIV
	SBD
	TRI
	VEN

Segment SPF Class 5 - Rural 2-4 Lane Freeway Undivided

In the case of Rural 2-4 Lane Freeway Undivided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.13. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Rural 2-4 Lane Freeway Undivided SPF as shown in tables 4.23, 4.24, and 4.25.

$$Predicted\ Crashes\ (Np) = exp\ (-8.503 + 0.668 * X1 + 1.169 * X2 - 0.254 * X3 - 0.847 * X4 - 0.132 * X5 - 0.762 * X6 + 0.070 * X7 - 0.234 * X8 + 0.188 * X9) \quad (4.13)$$

Table 4.23 Significant Variables - FSV Rural 2-4 Lane Freeway Undivided

Code	Variable	Estimate
	Intercept	-8.503
X1	ADT	0.668
X2	Segment length	1.169
X3	Right surface type: H	-0.254
X4	Right surface type: B	-0.847
X5	Year 2	-0.132
X6	Route_cluster_1	-0.762
X7	Route_cluster_2	0.070
X8	County_cluster_1	-0.234
X9	County_cluster_2	0.188
	Theta (overdispersion)	0.6479
	AIC	229491
	Log-likelihood	-114734.59

Table 4.24 Route Cluster in FSV Rural 2-4 Lane Freeway Undivided

1	2
46	99
83	101
104	154
124	199
166	

Table 4.25 County Cluster in FSV Rural 2-4 Lane Freeway Undivided

1	2
BUT	ALP
	AMA
	LAK
	MPA
	PLU
	SAC
	SCR
	SD
	TUO

Segment SPF Class 6 - Rural 2-4 Lane Freeway Divided

In the case of Rural 2-4 Lane Freeway Divided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.14. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Rural 2-4 Lane Freeway Divided SPF as shown in tables 4.26, 4.27, and 4.28.

$$Predicted\ Crashes\ (Np) = exp\,(-8.882 + 0.656 * X1 + 1.083 * X2 + 0.009 * X3 - 0.701 * X4 - 0.590 * X5 + 0.277 * X6 + 0.373 * X7 + 0.194 * X8 + 0.689 * X9 + 0.326 * X10 + 0.064 * X11 - 1.015 * X12 - 0.469 * X13 - 0.156 * X14 - 0.168 * X15 + 0.262 * X16 - 0.700 * X17 - 0.123 * X18 + 0.221 * X19 - 0.610 * X20 - 0.250 * X21) \quad (4.14)$$

Table 4.26 Significant Variables - FSV Rural 2-4 Lane Freeway Divided

Code	Variable	Estimate
	Intercept	-8.882
X1	ADT	0.656
X2	Segment length	1.083
X3	Design speed	0.009
X4	Terrain: F	-0.701
X5	Terrain: R	-0.590
X6	Right surface type: H	0.277
X7	Right surface type: C	0.373
X8	Left surface type: H	0.194
X9	Left surface type: M	0.689
X10	Left surface type: C	0.326
X11	Median barrier: N	0.064
X12	Median type: G	-1.015
X13	Median type: Q	-0.469
X14	Year 1	-0.156
X15	Year 2	-0.168
X16	Route_cluster_1	0.262
X17	Route_cluster_2	-0.700
X18	Route_cluster_3	-0.123
X19	County_cluster_1	0.221
X20	County_cluster_2	-0.610
X21	County_cluster_3	-0.250
	Theta (overdispersion)	0.719
	AIC	225752
	Log-likelihood	-112853.16

Table 4.27 Route Cluster in FSV Rural 2-4 Lane Freeway Divided

1	2	3
4	40	1
8	62	14
10	86	15
36	135	50
37	246	101
80		395
178		
215		

Table 4.28 County Cluster in FSV Rural 2-4 Lane Freeway Divided

1	2	3
DN	IMP	BUT
HUM	INY	FRE
SAC	KER	GLE
	KIN	MAD
	LAS	SOL
	NEV	STA
	SHA	TUL
	SIS	
	SUT	
	TEH	

Segment SPF Class 7 - Rural 5+ Lane Freeway

In the case of Rural 5+ Lane Freeway SPF, the FSV-based SPF is identified and is shown in Eqn. 4.15. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Rural 5+ Lane Freeway SPF as shown in tables 4.29, 4.30, and 4.31.

$$Predicted\ Crashes\ (Np) = exp\,(-8.926 + 0.662 * X1 + 1.160 * X2 + 0.279 * X3 + 0.587 * X4 + 0.554 * X5 + 0.119 * X6 - 0.538 * X7 - 0.153 * X8 - 0.166 * X9 - 0.061 * X10 - 0.269 * X11 + 0.187 * X12) \tag{4.15}$$

Table 4.29 Significant Variables - FSV Rural 5+ Lane Freeway

Code	Variable	Estimate
	Intercept	-8.926
X1	ADT	0.662
X2	Segment length	1.160
X3	Left surface type: H	0.279
X4	Left surface type: M	0.587
X5	Left surface type: C	0.554
X6	Median barrier: N	0.119
X7	Median type: Q	-0.538
X8	Year 1	-0.153
X9	Year 2	-0.166
X10	Route_Cluster_1	-0.061
X11	County_Cluster_1	-0.269
X12	County_Cluster_2	0.187
	Theta (overdispersion)	0.6564
	AIC	229222
	Log-likelihood	-114596.98

Table 4.30 Route Cluster in FSV Rural 5+ Lane Freeway

1
5
71
80
152
395

Table 4.31 County Cluster in FSV Rural 5+ Lane Freeway

1	2
BUT	HUM
MON	MEN
SIS	RIV
TUL	SCR
	SD

Segment SPF Class 8 - Urban 2 Lane Non-Freeway Undivided

In the case of Urban 2 Lane Non-Freeway Undivided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.16. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Urban 2 Lane Non-Freeway Undivided SPF as shown in tables 4.32, 4.33, and 4.34.

$$Predicted\ Crashes\ (Np) = exp\,(-10.369 + 0.776 * X1 + 1.177 * X2 + 0.332 * X3 + 0.886 * X4 + 0.568 * X5 + 0.612 * X6 + 0.573 * X7 - 0.132 * X8 - 0.584 * X9 - 0.228 * X10 + 0.325 * X11 + 1.177 * X12 - 0.117 * X13) \quad (4.16)$$

Table 4.32 Significant Variables - FSV Urban 2 Lane Non-Freeway Undivided

Code	Variable	Estimate
	Intercept	-10.369
X1	ADT	0.776
X2	Segment length	1.177
X3	Right surface type: H	0.332
X4	Right surface type: O	0.886
X5	Left surface type: M	0.568
X6	Left surface type: C	0.612
X7	Median type: B	0.573
X8	Year_2	-0.132
X9	Route_cluster_1	-0.584
X10	Route_cluster_2	-0.228
X11	Route_cluster_3	0.325
X12	Route_cluster_4	1.177
X13	County_cluster_1	-0.117
	Theta (overdispersion)	0.6629
	AIC	228526
	Log-likelihood	-114247.90

Table 4.33 Route Cluster in FSV Urban 2 Lane Non-Freeway Undivided

1	2	3	4
12	1	2	173
32	16	4	200
46	33	9	
111	43	18	
116	68	35	
142	108	49	
145	132	70	
184		74	
223		79	
395		94	
		154	
		178	

Table 4.34 County Cluster in FSV Urban 2 Lane Non-Freeway Undivided

1
FRE
KER
KIN
MAD
RIV
STA

Segment SPF Class 9 - Urban 2 Lane Non-Freeway Divided

In the case of Urban 2 Lane Non-Freeway Divided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.17. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Urban 2 Lane Non-Freeway Divided SPF as shown in tables 4.35, 4.36, and 4.37.

$$Predicted\ Crashes\ (Np) = exp\,(-9.135 + 0.701 * X1 + 1.224 * X2 + 0.341 * X3 - 0.916 * X4 - 0.728 * X5 + 0.347 * X6 + 0.131 * X7 - 0.216 * X8) \quad (4.17)$$

Table 4.35 Significant Variables - FSV Urban 2 Lane Non-Freeway Divided

Code	Variable	Estimate
	Intercept	-9.135
X1	ADT	0.701
X2	Segment length	1.224
X3	Left surface type: M	0.341
X4	Median barrier: Y	-0.916
X5	Route_cluster_1	-0.728
X6	Route_cluster_2	0.347
X7	County_cluster_1	0.131
X8	County_cluster_2	-0.216
	Theta (overdispersion)	0.6179
	AIC	231069
	Log-likelihood	-115524.51

Table 4.36 Route Cluster in FSV Urban 2 Lane Non-Freeway Divided

1	2
63	4
166	38
184	49
218	74
247	88

Table 4.37 County Cluster in FSV Urban 2 Lane Non-Freeway Divided

1	2
ORA	BUT
SAC	GLE
SD	MON
VEN	

Segment SPF Class 10 - Urban 3+ Lane Non-Freeway Undivided

In the case of Urban 3+ Lane Non-Freeway Undivided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.18. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Urban 3+ Lane Non-Freeway Undivided SPF as shown in tables 4.38, 4.39, and 4.40.

$$Predicted\ Crashes\ (Np) = exp\ (-9.186 + 0.726 * X1 + 1.217 * X2 - 0.292 * X3 + 1.172 * X4 - 0.662 * X5 + 0.083 * X6) \tag{4.18}$$

Table 4.38 Significant Variables - FSV Urban 3+ Lane Non-Freeway Undivided

Code	Variable	Estimate
	Intercept	-9.186
X1	ADT	0.726
X2	Segment length	1.217
X3	Terrain: F	-0.292
X4	Median type: A	1.172
X5	Route_cluster_1	-0.662
X6	County_cluster_1	0.083
	Theta (overdispersion)	0.6202
	AIC	230912
	Log-likelihood	-115448.23

Table 4.39 Route Cluster in FSV Urban 3+ Lane Non-Freeway Undivided

1
111
185

Table 4.40 County Cluster in FSV Urban 3+ Lane Non-Freeway Undivided

1
ORA

<u>*Segment SPF Class 11- Urban 3+ Lane Non-Freeway Divided*</u>

In the case of Urban 3+ Lane Non-Freeway Divided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.19. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Urban 3+ Lane Non-Freeway Divided SPF as shown in tables 4.41, 4.42, and 4.43.

$$Predicted\ Crashes\ (Np) = exp\ (-8.848 + 0.718 * X1 + 1.131 * X2 - 0.250 * X3 - 0.104 * X4 - 1.030 * X5 + 0.185 * X6 - 0.882 * X7 - 0.166 * X8 - 1.307 * X9 - 0.391 * X10 - 0.257 * X11 - 1.077 * X12 - 1.387 * X13 - 0.657 * X14 + 0.120 * X15 + 0.089 * X16 + 0.394 * X17) \tag{4.19}$$

Table 4.41 Significant Variables - FSV Urban 3+ Lane Non-Freeway Divided

Code	Variable	Estimate
	Intercept	-8.848
X1	ADT	0.718
X2	Segment length	1.131
X3	Terrain: F	-0.250
X4	Right surface type: H	-0.104
X5	Median barrier: Y	-1.030
X6	Median barrier: F	0.185
X7	Median type: F	-0.882
X8	Median type: H	-0.166
X9	Median type: G	-1.307
X10	Median type: K	-0.391
X11	Median type: J	-0.257
X12	Median type: Q	-1.077
X13	Median type: M	-1.387
X14	Route_cluster_1	-0.657
X15	Route_cluster_2	0.120
X16	County_cluster_1	0.089
X17	County_cluster_2	0.394
	Theta (overdispersion)	0.6739
	AIC	227850
	Log-likelihood	-113906.01

Table 4.42 Route Cluster in FSV Urban 3+ Lane Non-Freeway Divided

1	2
47	1
72	2
82	17
90	35
138	39
255	74
395	88
	99

Table 4.43 County Cluster in FSV Urban 3+ Lane Non-Freeway Divided

1	2
LA	DN
MON	HUM
RIV	SF
SD	
SLO	
SM	
VEN	

Segment SPF Class 12 - Urban 2-7 Lane Freeway Undivided

In the case of Urban 2-7 Lane Freeway Undivided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.20. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Urban 2-7 Lane Freeway Undivided SPF as shown in tables 4.44, 4.45, and 4.46.

$$Predicted\ Crashes\ (Np) = exp\ (-9.017 + 0.691 * X1 + 1.226 * X2 + 0.110 * X3 - 0.230 * X4 + 0.251 * X5 + 0.222 * X6) \tag{4.20}$$

Table 4.44 Significant Variables - FSV Urban 2-7 Lane Freeway Undivided

Code	Variable	Estimate
	Intercept	-9.017
X1	ADT	0.691
X2	Segment length	1.226
X3	Year 4	0.110
X4	Route_cluster_1	-0.230
X5	Route_cluster_2	0.251
X6	County_cluster_1	0.222
	Theta (overdispersion)	0.6160
	AIC	231261.00
	Log-likelihood	-115622.64

Table 4.45 Route Cluster in FSV Urban 2-7 Lane Freeway Undivided

1	2
1	18
65	70
111	154

Table 4.46 County Cluster in FSV Urban 2-7 Lane Freeway Undivided

1
DN
HUM
SD

Segment SPF Class 13 - Urban 2-7 Lane Freeway Divided

In the case of Urban 2-7 Lane Freeway Divided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.21. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Urban 2-7 Lane Freeway Divided SPF as shown in tables 4.47, 4.48, and 4.49.

$$Predicted\ Crashes\ (Np) = exp\,(-6.994 + 0.505*X1 + 1.089*X2 - 0.002*X3 + 0.305*X4 + 0.700*X5 + 0.428*X6 + 0.193*X7 + 0.283*X8 - 0.525*X9 - 0.329*X10 + 0.139*X11 - 1.059*X12 - 0.571*X13 - 0.204*X14 - 0.216*X15 - 0.131*X16 - 0.448*X17 + 0.197*X18 + 0.372*X19 + 0.635*X20 - 0.448*X21 - 0.158*X22 + 0.189*X23) \quad (4.21)$$

Table 4.47 Significant Variables - FSV Urban 2-7 Lane Freeway Divided

Code	Variable	Estimate
	Intercept	-6.994
X1	ADT	0.505
X2	Segment length	1.089
X3	Design speed	-0.002
X4	Right surface type: H	0.305
X5	Right surface type: M	0.700
X6	Right surface type: C	0.428
X7	Left surface type: H	0.193
X8	Left surface type: C	0.283
X9	Median barrier: Z	-0.525
X10	Median barrier: I	-0.329
X11	Median barrier: F	0.139
X12	Median type: G	-1.059
X13	Median type: Q	-0.571
X14	Year 1	-0.204
X15	Year 2	-0.216
X16	Year 3	-0.131
X17	Route_cluster_1	-0.448
X18	Route_cluster_2	0.197
X19	Route_cluster_3	0.372
X20	Route_cluster_4	0.635
X21	County_cluster_1	-0.448
X22	County_cluster_2	-0.158
X23	County_cluster_3	0.189
	Theta (overdispersion)	0.7077
	AIC	226524
	Log-likelihood	-113237.07

Table 4.48 Route Cluster in FSV Urban 2-7 Lane Freeway Divided

1	2	3	4
190	8	4	2
	17	10	51
	41	18	76
	49	44	105
	52	242	215
	60		259
	67		780
	70		
	78		
	80		
	91		
	99		
	110		
	118		
	180		
	198		
	210		
	580		
	710		

Table 4.49 County Cluster in FSV Urban 2-7 Lane Freeway Divided

1	2	3
GLE	BUT	DN
KER	FRE	ORA
SHA	MAD	SCR
	PLA	SF
	SBD	TUO
	TEH	
	TUL	

Segment SPF Class 14 - Urban 8+ Lane Freeway Divided

In the case of Urban 8+ Lane Freeway Divided SPF, the FSV-based SPF is identified and is shown in Eqn. 4.22. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Urban 8+ Lane Freeway Divided SPF as shown in tables 4.50, 4.51, and 4.52.

$$Predicted\ Crashes\ (Np) = exp\ (-9.170 + 0.654 * X1 + 1.121 * X2 + 0.008 * X3 - 0.003 * X4 - 0.346 * X5 + 0.245 * X6 + 0.380 * X7 + 0.167 * X8 + 0.684 * X9 + 0.334 * X10 + 0.119 * X11 - 1.003 * X12 + 0.192 * X13 + 0.230 * X14 + 0.090 * X15 - 0.138 * X16 + 0.132 * X17 - 0.398 * X18 - 1.531 * X19 - 0.195 * X20 - 0.208 * X21 - 0.127 * X22 + 0.333 * X23 - 0.417 * X24 - 0.086 * X25 + 0.137 * X26) \quad (4.22)$$

Table 4.50 Significant Variables - FSV Urban 8+ Lane Freeway Divided SPF

Code	Variable	Estimate
	Intercept	-9.170
X1	ADT	0.654
X2	Segment length	1.121
X3	Design speed	0.008
X4	Median width	-0.003
X5	Terrain: F	-0.346
X6	Right surface type: H	0.245
X7	Right surface type: C	0.380
X8	Left surface type: H	0.167
X9	Left surface type: M	0.684
X10	Left surface type: C	0.334
X11	Median barrier: C	0.119
X12	Median barrier: Y	-1.003
X13	Median barrier: J	0.192
X14	Median barrier: F	0.230
X15	Median barrier: Q	0.090
X16	Median type: Q	-0.318
X17	Median type: R	0.132
X18	Median type: T	-0.398
X19	Median type: S	-1.531
X20	Year 1	-0.195
X21	Year 2	-0.208
X22	Year 3	-0.127
X23	Route_cluster_1	0.333
X24	Route_cluster_2	-0.417
X25	County_cluster_1	-0.086
X26	County_cluster_2	0.137
	Theta (overdispersion)	0.6895
	AIC	227664
	Log-likelihood	-113804.03

Table 4.51 Route Cluster in FSV Urban 8+ Lane Freeway Divided SPF

1	2
4	24
8	85
10	163
51	
78	
91	
105	
215	
605	

Table 4.52 County Cluster in FSV Urban 8+ Lane Freeway Divided SPF

1	2
SBD	LA
SM	SCL
	SF
	YOL

<u>Segment SPF Class 15 - Rural L/R Alignment Freeway</u>

In the case of Rural L/R Alignment Freeway SPF, the FSV-based SPF is identified and is shown in Eqn. 4.23. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Rural L/R Alignment Freeway SPF as shown in tables 4.53, 4.54, and 4.55.

$$Predicted\ Crashes\ (Np) = exp\ (-8.970 + 0.699 * X1 + 1.130 * X2 + 0.007 * X3 - 0.797 * X4 - 0.635 * X5 + 0.125 * X6 + 0.418 * X7 - 0.186 * X8 + 0.196 * X9 - 0.133 * X10 - 0.684 * X11 - 0.130 * X12 - 0.090 * X13 - 0.149 * X14) \quad (4.23)$$

Table 4.53 Significant Variables - FSV Rural L/R Alignment Freeway SPF

Code	Variable	Estimate
	Intercept	-8.970
X1	ADT	0.699
X2	Segment length	1.130
X3	Design speed	0.007
X4	Terrain: F	-0.797
X5	Terrain: R	-0.635
X6	Right surface type: H	0.125
X7	Right surface type: C	0.418
X8	Left surface type: BLANK	-0.186
X9	Median barrier: K	0.196
X10	Median type: J	-0.133
X11	Median type: Q	-0.684
X12	Year 2	-0.130
X13	Route_cluster_1	0.090
X14	County_cluster_1	-0.149
	Theta (overdispersion)	0.6756
	AIC	228058
	Log-likelihood	-114013.17

Table 4.54 Route Cluster in FSV Rural L/R Alignment Freeway SPF

1
80
99

Table 4.55 County Cluster in FSV Rural L/R Alignment Freeway SPF

1
BUT

Segment SPF Class 16 - Urban L/R Alignment Freeway

In the case of Urban L/R Alignment Freeway SPF, the FSV-based SPF is identified and is shown in Eqn. 4.24. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Urban L/R Alignment Freeway SPF as shown in tables 4.56, 4.57, and 4.58.

$$Predicted\ Crashes\ (Np) = exp\,(-2.837 + 1.079 * X1 + 0.343 * X2 + 0.065 * X3 + 0.260 * X4 + 0.186 * X5 + 1.366 * X6 + 0.561 * X7 - 0.445 * X8 - 0.041 * X9 + 1.010 * X10 + 0.348 * X11 - 0.514 * X12 + 0.696 * X13) \quad (4.24)$$

Table 4.56 Significant Variables - FSV Urban L/R Alignment Freeway SPF

Code	Variable	Estimate
	Intercept	-2.837
X1	Segment length	1.079
X2	Terrain: F	0.343
X3	Terrain: R	0.065
X4	Right surface type: H	0.260
X5	Right surface type: B	0.186
X6	Right surface type: C	1.366
X7	Median barrier: Q	0.561
X8	Highway group: R	-0.445
X9	Year 3	-0.041
X10	Route_cluster_1	1.010
X11	Route_cluster_2	0.348
X12	County_cluster_1	-0.514
X13	County_cluster_2	0.696
	Theta (overdispersion)	0.4244
	AIC	244589
	Log-likelihood	-122279.70

Table 4.57 Route Cluster in FSV Urban L/R Alignment Freeway SPF

1	2
37	1
73	5
80	15
91	57
101	110
280	118
580	134
680	

Table 4.58 County Cluster in FSV Urban L/R Alignment Freeway SPF

1	2
KER	LA
MON	RIV
SIS	SAC
SOL	SD

Segment SPF Class 17 - L/R Alignment Non-Freeway

In the case of L/R Alignment Non-Freeway SPF, the FSV-based SPF is identified and is shown in Eqn. 4.25. Furthermore, the model estimates including goodness-of-fit measures for the L/R Alignment Non-Freeway SPF is shown in table 4.59. Route and county dummies are not statistically significant.

$$Predicted\ Crashes\ (Np) = exp\ (-8.342 + 0.620 * X1 + 1.195 * X2 - 0.270 * X3 + 0.188 * X4 - 0.106 * X5) \tag{4.25}$$

Table 4.59 Significant Variables - FSV L/R Alignment Non-Freeway SPF

Code	Variable	Estimate
	Intercept	-8.342
X1	ADT	0.620
X2	Segment length	1.195
X3	Right surface type: C	0.270
X4	Left surface type: C	0.188
X5	Median type: J	-0.106
	Theta (overdispersion)	0.6303
	AIC	230563
	Log-likelihood	-115274.31

Appendix B provides the model outputs of Segment SPFs based on total collisions developed as part of this project.

4.8.2. Intersection SPFs

Intersections are generally defined as fixed length ranges of 250 feet from the centerline of the intersecting roadway. In this project, actual values of override length were accounted for to define intersection and to merge collision data with the infrastructure data. Explanatory variables considered for intersection SPFs based on the availability of data are provided in Table 4.60, and the summary statistics of the SPF category are shown in Table 4.61.

Table 4.60 Explanatory Variables considered for Intersection SPF

Explanatory Variables	
ADT – Mainline & Cross street	Highway group
Number of lanes – Main lane & Cross street	Flow description - Main lane & Cross street
Design type	County
Control type	Route
Population group	Year
Light condition	

Table 4.61 Summary Statistics of the Injury-based SPF

SPF Category	Mean	Std. Dev.	Min.	Max.
Total	15.39	14.08	0	133
FSV	2.21	2.44	0	17
FS	0.42	0.76	0	5

4.8.2.1. Intersection SPF Class

To better reflect the actual site condition, intersection SPFs are classified into three classes based on the population group and control type as follows:

- Rural Intersection SPF
- Urban Signalized Intersection SPF
- Urban Un-Signalized Intersection SPF

The rural intersection SPF class is created based on the observations with the population group as 'Rural' —field name 'SPFI_POPULATION_GROUP with value 'R'—in the intersection infrastructure data. Due to limited observations, further classification based on control conditions was not possible.

If the population group is 'Urban' and the control condition is 'Signalized'—field name 'SPFI_POPULATION_GROUP with value 'U' and field name 'SPFI_CONTROL_CODE' with values 'J' through 'P'—in the intersection infrastructure data, then it is Urban Signalized Intersection SPF Class, and if the field name 'SPFI_CONTROL_CODE' is with values other than 'J' through 'P'—it is Urban Un-Signalized Intersection SPF Class.

4.8.2.2. Summary of Intersection SPFs

After processing the intersection infrastructure and collision data through each of the SPF development steps, Type 1 and 2 SPFs were developed with Total and FSV collisions. FS-based SPF fails to provide to good results due to limited observations. Furthermore, based on the performance measures, Type 2 FSV SPF performs better than the Total collision SPF.

Three main points observed from the model outputs are as follows:

- Sites based on total collisions are likely to be influenced by PDOs and complaint of pain
- Fatal + severe collisions may yield few collisions for investigators to recommend countermeasures
- Performance of sites identified using FSV collisions can provide higher resolution for investigation purposes while limiting influence of PDOs

The following sub-sections provide results of the FSV SPFs for each intersection SPF class—Rural, Urban Signalized, and Urban Un-Signalized.

Rural Intersection SPF

In the case of Rural Intersection SPF, the FSV-based SPF is identified and is shown in Eqn. 4.26. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Rural Intersection SPF as shown in tables 4.62, 4.63, and 4.64.

$$Predicted\ Crashes\ (Np) = exp\,(-8.505 + 0.511 * X1 + 0.231 * X2 - 0.291 * X3 - 0.423 * X4 - 0.254 * X5 + 0.183 * X6 + 0.117 * X7 + 0.094 * X8 + 0.193 * X9 + 0.189 * X10 - 0.716 * X11 + 1.424 * X12 + 0.309 * X13 + 0.973 * X14 + 0.494 * X15 + 0.757 * X16 + 0.776 * X17 - 0.380 * X18 - 0.860 * X19\,) \tag{4.26}$$

Table 4.62 Significant Variables - FSV Rural Intersection SPF

Code	FSV	
	Variables	Estimate
B0	Intercept	-8.505
X1	Mainline adt	0.511
X2	Cross street adt	0.231
X3	Design: IS	-0.291
X4	Design: IT	-0.423
X5	Design: IY	-0.254
X6	Cross street lanes amt	0.183
X7	Mainline lanes_amt	0.117
X8	Year_3: 2015	0.094
X9	Year_4: 2016	0.193
X10	Year_5: 2017	0.189
X11	Route_cluster_1	-0.716
X12	Route_cluster_2	1.424
X13	Route_cluster_3	0.309
X14	Route_cluster_4	0.973
X15	Route_cluster_5	0.494
X16	Route_cluster_6	0.757
X17	County_cluster_1	0.776
X18	County_cluster_2	-0.380
X19	County_cluster_3	-0.860
	Theta (overdispersion)	1.158
	AIC	24737.000
	Log-likelihood	-12347.559
	RMSE	0.334
	MAD	0.157
	MAPE	195.782

Considering the statistically significant 29 routes and 13 counties, it is difficult to incorporate all of these into the model. Hence, the clustering approach was adopted as described in Section 4.7.2, and routes were clustered into six groups and counties into three, as shown in tables 4.63 and 4.64, respectively.

Table 4.63 Route Cluster in FSV Rural Intersection SPF

1	2	3	4	5	6
126	371	88	79	138	78
203	60	4	74	121	113
395		18	184	2	25
211		152	76	99	37
				20	38
				26	41
				201	243
				43	

Table 4.64 County Cluster in FSV Rural Intersection SPF

1	2	3
SCR	MAD	SIS
VEN	SUT	TEH
SM	INY	
STA	RIV	
	SHA	
	IMP	
	MOD	

Urban Signalized Intersection SPF

In the case of Urban Signalized Intersection SPF, the FSV-based SPF is identified and is shown in Eqn. 4.27. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Urban Signalized Intersection SPF are shown in tables 4.65, 4.66, and 4.67.

$$Predicted\ Crashes\ (Np) = exp(-7.003 + 0.504 * X1 - 0.568 * X2 - 0.189 * X3 - 0.474 * X4 - 0.114 * X5 + 0.083 * X6 + 0.091 * X7 + 0.346 * X8 + 0.363 * X9 + 0.514 * X10 + 0.091 * X11 + 0.081 * X12 - 1.054 * X13 - 0.320 * X14 - 1.449 * X15 + 1.529 * X16 - 0.804 * X17 - 1.626 * X18 + 0.485 * X19 + 0.551 * X20 + 0.436 * X21 - 0.893 * X22 - 1.325 * X23 - 0.317 * X24) \quad (4.27)$$

Table 4.65 Significant Variables - FSV Urban Signalized Intersection SPF

Code	FSV	
	Variables	Estimate
	Intercept	-7.003
X1	Main_adt	0.504
X2	Control: K	-0.568
X3	Control: L	-0.189
X4	Design: IS	-0.474
X5	Design: IT	-0.114
X6	Cross street_adt	0.083
X7	Cross street lanes_amt	0.091
X8	Main_flow: P	0.346
X9	Main_flow: W	0.363
X10	Cross street_flow: R	0.514
X11	Year_4	0.091
X12	Year_5	0.081
X13	Route_cluster_1	-1.054
X14	Route_cluster_2	-0.320
X15	Route_cluster_3	-1.449
X16	Route_cluster_4	1.529
X17	Route_cluster_5	-0.804
X18	Route_cluster_6	-1.626
X19	Route_cluster_7	-0.485
X20	Route_cluster_8	0.551
X21	County_cluster_1	0.436
X22	County_cluster_2	-0.893
X23	County_cluster_3	-1.325
X24	County_cluster_4	-0.317
	Theta (overdispersion)	2.755
	AIC	18860.000
	Log-likelihood	-9403.837
	RMSE	0.732
	MAD	0.542
	MAPE	164.056

Considering the statistically significant 39 routes and 19 counties, it is difficult to incorporate all of these into the model. Hence, the clustering approach was adopted as described in Section 4.7.2, and routes were clustered into eight groups and counties into four, as shown in tables 4.66 and 4.67, respectively.

Table 4.66 Route Cluster in FSV Urban Signalized Intersection SPF

1	2	3	4	5	6	7	8
183	2	137	83	68	121	164	273
61	107	232	395	221	16	185	105
4		92	62	123		29	18
54		118	225			79	25
		218	38			112	39
		22					101
		131					184
		213					138
							72
							237

Table 4.67 County Cluster in FSV Urban Signalized Intersection SPF

1	2	3	4
SAC	HUM	IMP	SON
NAP	SHA	ED	PLA
RIV	YUB	SBD	SM
	SF		BUT
			FRE
			SCL
			SB
			LA
			KER

Urban Un-Signalized Intersection SPF

Hence, Urban Un-Signalized Intersection SPF based on FSV is selected and is shown in Eqn. 4.28. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Urban Un-Signalized Intersection SPF as shown in tables 4.68, 4.69, and 4.70.

$$Predicted\ Crashes\ (Np) = exp\,(-6.575 + 0.346 * X1 + 0.161 * X2 - 0.684 * X3 - 0.198 * X4 - 0.453 * X5 - 0.304 * X6 - 0.163 * X7 + 0.429 * X8 + 0.264 * X9 + 1.129 * X10 + 1.738 * X11 + 0.624 * X12 + 0.099 * X13 + 0.160 * X14 + 0.120 * X15 - 1.274 * X16 + 0.886 * X17 + 3.464 * X18 + 0.576 * X19 - 0.633 * X20 - 1.966 * X21 - 0.260 * X22 + 0.718 * X23 + 1.539 * X24 - 0.389 * X25 - 1.189 * X26 - 0.430 * X27 - 2.273 * X28 + 0.347 * X29) \quad (4.28)$$

Table 4.68 Significant Variables - FSV Urban Un-Signalized Intersection SPF

Code	FSV	
	Variables	**Estimate**
	Intercept	-6.575
X1	Mainline_adt	0.346
X2	Cross street_adt	0.161
X3	Design: IM	-0.684
X4	Design: IS	-0.198
X5	Design: IT	-0.453
X6	Design: IY	-0.304
X7	Light: Y	-0.163
X8	Main_flow: P	0.429
X9	Control: B	0.264
X10	Control: E	1.129
X11	Control: F	1.738
X12	Control: Z	0.624
X13	Year_3	0.099
X14	Year_4	0.160
X15	Year_5	0.120
X16	Route_cluster_1	-1.274
X17	Route_cluster_2	0.886
X18	Route_cluster_3	3.464
X19	Route_cluster_4	0.576
X20	Route_cluster_5	-0.633
X21	Route_cluster_6	-1.966
X22	Route_cluster_7	-0.260
X23	Route_cluster_8	0.718
X24	Route_cluster_9	1.539
X25	Route_cluster_10	0.389
X26	County_cluster_1	-1.189
X27	County_cluster_2	-0.430
X28	County_cluster_3	-2.273
X29	County_cluster_4	0.347
	Theta (overdispersion)	1.806
	AIC	25022.000
	Log-likelihood	-12479.790
	RMSE	0.416
	MAD	0.248
	MAPE	191.720

Considering the statistically significant 56 routes and 22 counties, it is difficult to incorporate all of these into the model. Hence, the clustering approach was adopted as described in Section 4.7.2, and routes were clustered into ten groups and counties into four, as shown in tables 4.69 and 4.70 respectively.

Table 4.69 Route Cluster in FSV Urban Un-Signalized Intersection SPF

1	2	3	4	5	6	7	8	9	10
2	395	199	192	3	210	123	180	190	116
233	237		184	32	243	162	135	262	39
66	119		173	185	213		4	45	49
269	165		189	99	112		198		120
	60		71	183	218		156		62
	63		75		95		126		83
			94				50		108
			101				59		
							65		
							67		
							133		
							154		
							88		
							204		

Table 4.70 County Cluster in FSV Urban Un-Signalized Intersection SPF

1	2	3	4
ED	SB	DN	LAK
COL	KER		SCR
IMP	TUL		BUT
SF	NAP		LA
FRE	MER		SUT
	HUM		ORA
	LAS		RIV
	SBD		
	MEN		

In general, the variables' coefficients, which have positive and negative impacts in an intersection crash based on model estimates are as follows:

- Positive Coefficients
    - ADT
    - Cross street lanes amount
    - Control condition (majority)
    - Highway group
    - Main street flow

- Negative Coefficients
    - Design type
    - Cross street flow

Appendix B provides the model outputs of total collision Intersection SPFs developed as part of this project.

4.8.3. Ramp SPF

Ramp infrastructure and collision data for the five-year period between 2013 and 2017 was used in the final analysis. Ramp lengths are not considered in this project due to their unavailability. Table 4.71 provides district ramp distribution. Districts 4 and 7 account for the major portion of ramps. Explanatory variables considered in the development of ramp SPF are shown in Table 4.72, and the summary statistics of SPF category are shown in Table 4.73.

Table 4.71 District Ramp Distribution

District	OFF Ramp	ON Ramp	OTHERS	Total
1	169	166	5	340
2	189	191	4	384
3	536	601	4	1,141
4	1,386	1,425	43	2,854
5	384	377	6	767
6	491	513	4	1,008
7	1,589	1,652	67	3,308
8	604	621	3	1,228
9	59	60	0	119
10	262	254	8	524
11	788	758	8	1,554
12	426	440	8	874
Total	6,883	7,058	160	14,101

Table 4.72 Explanatory Variables for Ramp SPF

Explanatory Variables	
ADT	Highway group
Design description	County
Ramp On/Off	Route
Population group	Year

Table 4.73 Summary Statistics of Injury-based Ramp SPF Category

SPF Category	Mean	Std. Dev.	Min.	Max.
Total	7.320	9.847	0	195
FSV	0.864	1.487	0	42
FS	0.160	0.474	0	15

4.8.3.1. Summary of Ramp SPF

After processing the ramp infrastructure and collision data through each of the SPF development steps, Type 1 and 2 SPFs were developed with Total and FSV collisions. FS-based SPFs fail to provide good results due to limited observations. Furthermore, based on the performance measures, Type 2 FSV SPFs perform better than the Total collision SPFs.

Three main points observed from the model outputs are as follows:

- Sites based on total collisions are likely to be influenced by PDOs and complaint of pain
- Fatal + severe collisions may yield few collisions for investigators to recommend countermeasures
- Performance of sites identified using FSV collisions can provide higher resolution for investigation purposes while limiting influence of PDOs

Hence, Ramp SPF based on FSV is selected and is shown in Eqn. 4.29. Furthermore, the model estimates including goodness-of-fit measures, clustering results of route and county for the Ramp FSV SPF are shown in tables 4.74, 4.75, and 4.76.

$$\begin{aligned} Predicted\ Crashes\ (Np) &= exp\,(-6.034\ +\ 0.504 * X1\ - 1.178 * X2 - 1.021 * X3 - 1.059 * X4 - 0.953 * X5 \\ &- 1.165 * X6 - 0.912 * X7 - 1.097 * X8 - 1.047 * X9 - 2.161 * X10 - 1.323 * X11 \\ &- 2.228 * X12 - 1.705 * X13 - 0.977 * X14 - 0.437 * X15 - 0.785 * X16 + 0.29 \\ &* X17 - 0.154 * X18 - 0.111 * X19 - 0.06 * X20 + 0.487 * X21 + 0.666 * X22 \\ &+ 0.373 * X23 + 1.441 * X24 + 0.726 * X25 + 0.603 * X26 + 0.857 * X27 + 1.149 \\ &* X28 + 0.936 * X29 + 1.88 * X30 + 0.274 * X31 + 0.424 * X32 + 0.589 * X33 \end{aligned} \quad (4.29)$$

Table 4.74 Significant Variables - FSV Ramp SPF

Code	Variables	Estimate
	(Intercept)	-6.034
X1	Ramp_adt	0.504
X2	Buttonhook Ramp	-1.178
X3	Collector Road	-1.021
X4	Diamond Type Ramp	-1.059
X5	Direct or Semi-direct Connector(Left)	-0.953
X6	Direct or Semi-direct Connector(Right)	-1.165
X7	Loop-with Left Turn	-0.912
X8	Loop-without Left Turn	-1.097
X9	Other-Ramp	-1.047
X10	Rest Area, Vista Point, Truck Scale	-2.161
X11	Scissors	-1.323
X12	Slip Ramp	-2.228
X13	Split Ramp	-1.705
X14	Two-way Ramp Segment	-0.977
X15	ON Ramp	-0.437
X16	OTH Ramp	-0.785
X17	Urban	0.290
X18	Year 1	-0.154
X19	Year 2	-0.111
X20	Year 3	-0.060
X21	Route_cluster_1	0.487
X22	Route_cluster_2	0.666
X23	Route_cluster_3	0.373
X24	Route_cluster_4	1.441
X25	Route_cluster_5	0.726
X26	Route_cluster_6	0.603
X27	Route_cluster_7	0.857
X28	Route_cluster_8	1.149
X29	Route_cluster_9	0.936
X30	Route_cluster_10	1.880
X31	County_cluster_1	0.274
X32	County_cluster_2	0.424
X33	County_cluster_3	0.589
	Theta (overdispersion)	1.3426
	AIC	67448
	Log-likelihood	-33689.6
	RMSE	0.5107
	MAD	0.0023
	MAPE	199.5419

From the model estimates, traffic volume, population group, routes and county positively contribute to the ramp crashes, while ramp design type and on/off ramp results in a negative impact on crashes. Considering the statistically significant 74 routes and 17 counties, it is difficult to incorporate all of these into the model. Hence, the clustering approach was adopted as described in Section 4.7.2. Routes were clustered into ten groups and counties into three, as shown in tables 4.75 and 4.76, respectively.

Table 4.75 Route Cluster in FSV Ramp SPF

1	2	3	4	5	6	7	8	9	10
17	15	1	90	275	41	2	4	37	79
71	52	56	244	5	58	57	14	44	190
92	67			8	78	87	80	50	
94	73			10	134	91	160	51	
125	84			12	238	99		65	
580	85			22	780	113		105	
	126			23		168		110	
	178			55		280		120	
	680			60		380		170	
				101		405		180	
				118		605		205	
				163				215	
				198				242	
				210				505	
				237				880	
				710					
				805					

Table 4.76 County Cluster in FSV Ramp SPF

1	2	3
ALA	LA	SBD
FRE	ORA	SBT
RIV	SAC	SD
SCL	SCR	
SHA	SF	
SJ	SON	
	STA	
	VEN	

Appendix B provides the model outputs of Total collision Ramp SPFs developed as part of this project.

4.9. Potential for Safety Improvement

Potential for safety improvement estimates the degree to which the long-term crash frequency could be reduced at a particular site. This helps to determine how much worse a given site is relative to sites with similar characteristics. This can be achieved through the Empirical Bayes method—a weighted average of the site's observed crash count and crashes expected at similar sites using a safety performance function.

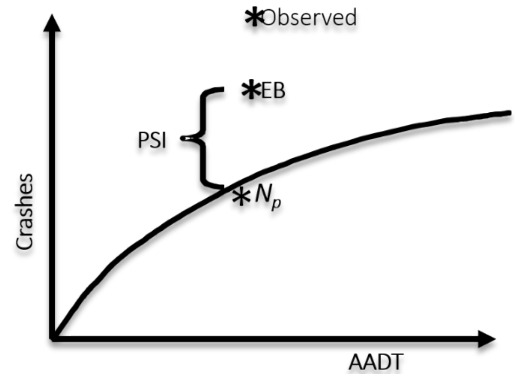

Figure 4.2 Potential for Safety Improvement of a Site

PSI is estimated as the difference between estimated crashes based on the Empirical Bayes (EB) approach (which takes the actual crash observation into account) and predicted using SPF (N_p) as shown in Figure 4.2.

4.9.1. Empirical Bayes Method

The Empirical Bayes method is a method used to combine observed crash frequency data for a given site with predicted crash frequency data from many similar sites, as a means to estimate its expected crash frequency. The EB method is only applicable when both predicted and observed crash frequencies are available for the specific roadway network conditions for which the estimate is being made (HSM 2010). Expected crash estimation using the EB method is shown in Eqn. 4.30.

EB estimation:

Expected crash,

$$N_{exp} = w * \sum_{y=i}^{Y=n} \mu_y + (1-w) * \sum_{y=i}^{Y=n} N_{obs,y} \tag{4.30}$$

Where,

N_{exp} - expected average crashes frequency for the study period;
w - weighted adjustment to be placed on the SPF prediction;
μ - predicted average crash frequency predicted using an SPF for the study period under the given conditions;
N_{obs} = observed crash frequency at the site over the study period.

In this project, the most recent five years of data from 2013 through 2017 were used. Hence the equation will be as shown in Eqn. 4.31:

Expected crash,

$$N_{exp} = w * \sum_{y=2013}^{Y=2017} \mu_5 + (1 - w) * \sum_{y=2013}^{Y=2017} N_{obs,5}$$

(4.31)

The EB Method uses a weight factor (as seen in Eqn. 4.32), which is a function of the SPF overdispersion parameter, to combine the two estimates into a weighted average. The weighted adjustment is therefore dependent only on the variance of the SPF and is not dependent on the validity of the observed crash data.

Weight factor;

$$w = \frac{1}{1 + \frac{\sum_{y=i}^{Y=n} \mu_y}{\phi}}$$

(4.32)

Where, 'y' is the years of analysis, which is five years (2013 through 2017) in this project so 'n' is 5 and 'μ_y' is the sum of predicted crashes. In this project it is the sum of predicted crashes for the five-year period between 2013 and 2017. 'ϕ' is the overdispersion from the model estimate.

In this project, weight factor was estimated as shown in Eqn. 4.33.

Weight factor,

$$w = \frac{1}{1 + \frac{\sum_{y=1}^{Y=5} \mu_5}{\phi}}$$

(4.33)

4.10. Network Screening Methods

Network screening is the first step in the site safety improvement process. The output of network screening is a list of sites that are ranked by priority for safety investigation. This project also considers the two network screening methods commonly used—the sliding window and peak searching methods. These two methods are included in the tool for network screening of Caltrans State Highway System and for performance evaluation purpose if the agency requires. The sliding window and peak searching approaches used for safety evaluation of highway segment are explained in the subsequent sub-sections.

4.10.1. Sliding Window

In the sliding window method, a window of a specified length is conceptually moved along the road segment from beginning to end in increments of a specified size. The performance measure chosen to screen the segment is applied to each position of the window, and the results of the analysis are recorded for each window. Figure 4.3 shows a schematic diagram of the sliding window method of network screening.

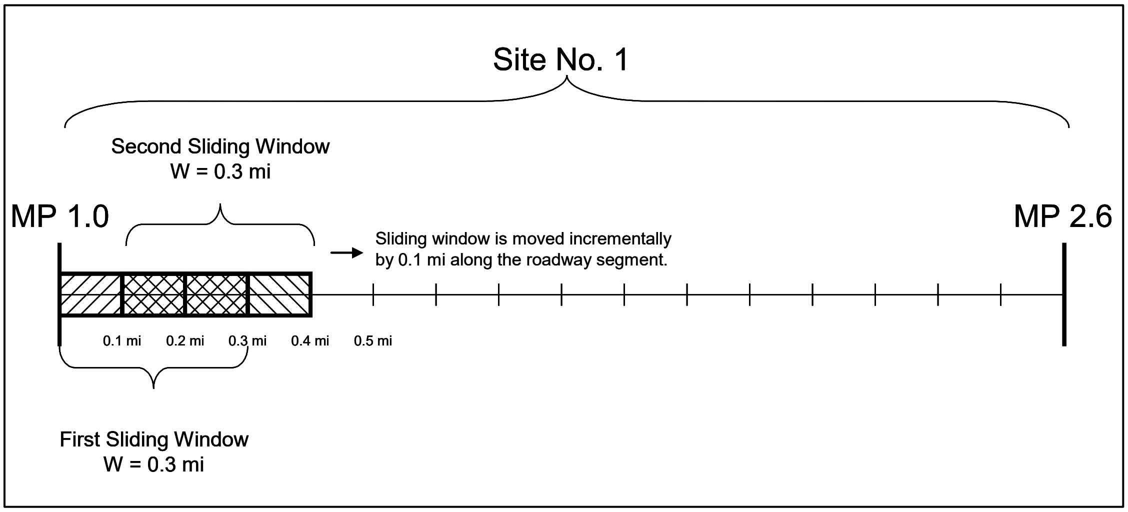

Figure 4.3 Schematic diagram of the Sliding Window approach

A window pertains to a given segment if at least some portion of the window is within the boundaries of the segment. From all the windows that pertain to a given segment, the window that shows the most potential for reduction in crash frequency from among all of those in the entire segment is identified and is used to represent the potential for reduction in crash frequency of the entire segment. After all segments are ranked according to the respective highest sub-segment value, those segments with the greatest potential for reduction in crash frequency or severity are studied in detail to identify potential countermeasures (HSM 2010).

4.10.2. Peak Searching

The peak searching method is used to identify the segments that are most likely to benefit from a safety improvement within a homogeneous section. Based on Highway Safety Manual 2010, using the peak searching method, each individual roadway segment is subdivided into windows of similar length, potentially growing incrementally in length until the length of the window equals the length of the entire roadway segment. The windows do not span multiple roadway segments. For each window, the chosen performance measure is calculated. Based upon the statistical precision of the performance measure, the window with the maximum value of the performance measure within a roadway segment is used to rank the potential for reduction in crashes of that site (i.e., entire roadway segment) relative to the other sites being screened. Figure 4.4 shows a schematic diagram of the peak searching method of network screening.

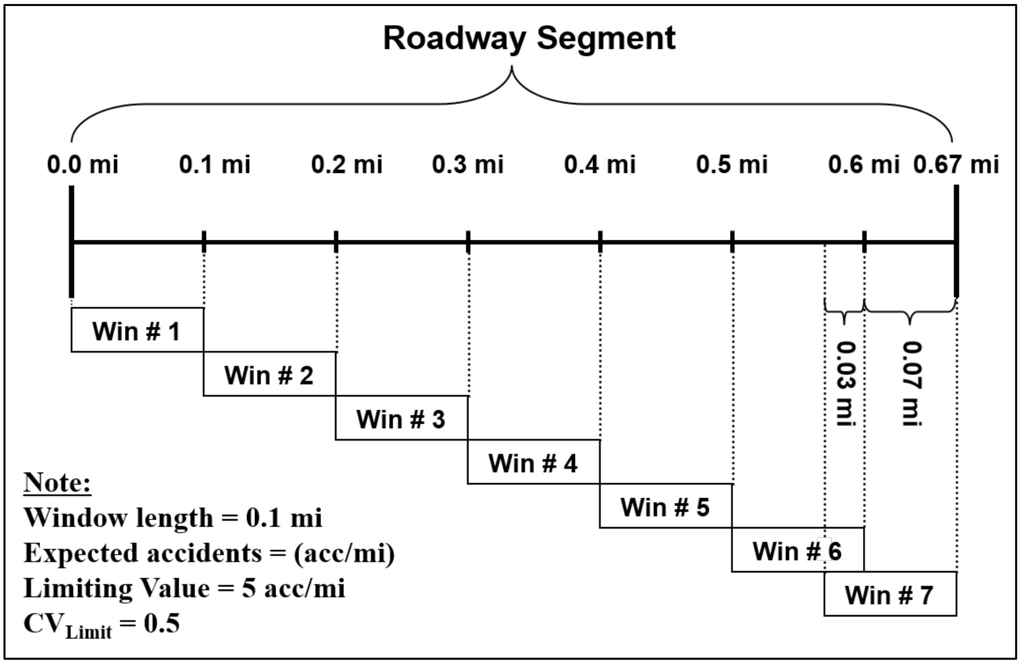

Figure 4.4 Schematic diagram of the Peak Searching approach

4.11. SPF Implementation Challenges

Implications of using different severity levels for screening

List of top HCCLs based on Total, FSV and FS may differ based on:

- Limited overlap (same sites may not be present across lists)
- Number and type of collisions available for investigation analysis
- Spatial distribution of the sites identified in each list

4.12. Chapter Summary

This chapter summarizes the development of California-specific SPFs and their application through network screening by generating a list of potential sites for safety improvements. The next chapter will describe the stages in the MS Excel macro spreadsheet tool developed as part of the implementation of SPFs for identifying high crash concentration locations.

Chapter 5. SPF Implementation - Design and Development of the SPF Tool

This chapter describes the design and development of the MS Excel Macro Spreadsheet tool for network screening. The developed tool incorporates all of Caltrans' reviewed Safety Performance Functions (SPFs) for identifying high crash concentration locations along the State Highway System. This chapter also explains the different functionalities within this spreadsheet tool for safety investigation.

5.1. Desired Functionality of the Excel Macro Spreadsheet Tool

The first and foremost part of the development of the tool is to define and document its potential capabilities in a way that allows the research team to develop a forward-compatible tool with the flexibility to accommodate future enhancements that cannot be implemented within this project due to many reasons including data and SPF limitations. Achieving this was accomplished by interviewing the safety experts from Caltrans Division of Traffic Operations.

Based on these interviews and interactions, the functionality developed after multiple iterations includes three stages—data preparation, analysis, and reporting. Figure 5.1 shows the flow chart of the functionality of the tool. First, the user will import the required data—three infrastructure data files and one collision data file—extracted from the TASAS-TSN based on an analysis period into the Stage A-Data Preparation. The output of Stage A is a summary of the input raw data file for cross reference, and a pre-processed file which will be used as an input in Stage B-Data Analysis. During this stage, the analysis of the data will be carried out based on the requirement and an Analysis Results File will be produced. This file can be used for generating reports based on the need later in Stage C- Reporting.

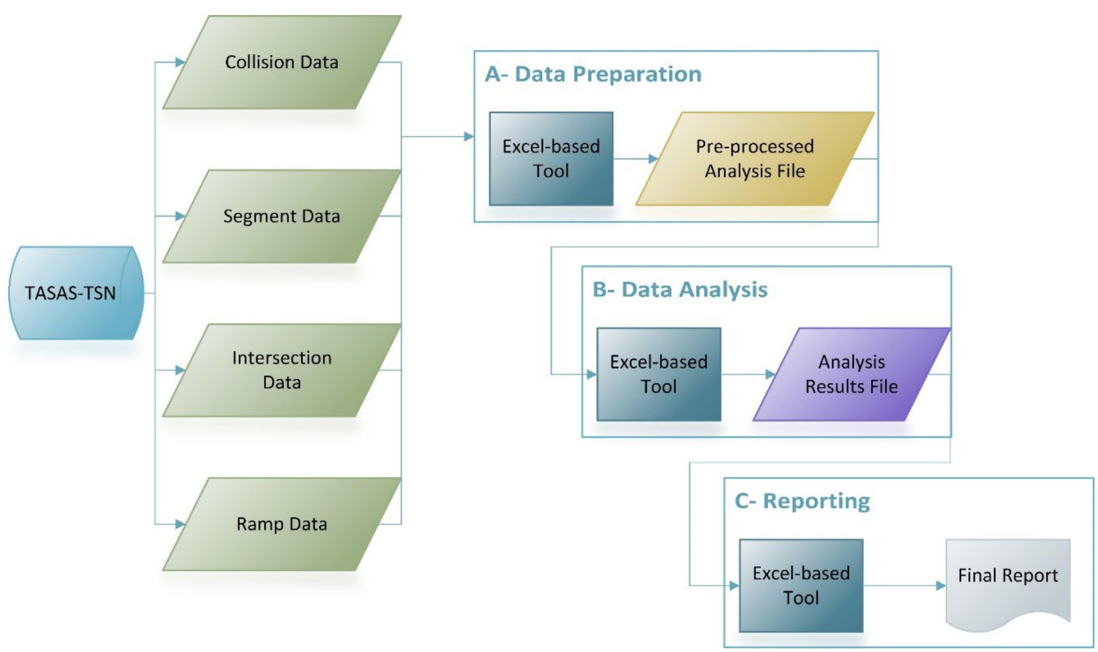

Figure 5.1 Flowchart of SPF Tool

5.2. Development of the SPF Tool

After establishing the desired functionality, the MS Excel Macro Spreadsheet tool was developed. For the data process from Excel, Microsoft Access data objects within Excel were used, which allows creation of temporary files for data manipulation, and is approximately ten times faster than Excel.

In addition, this section will provide detailed descriptions of the different stages of the network screening using SPF Tool. The user interface of the SPF tool is shown in Figure 5.2. UI provides two buttons—'NEW' and 'OPEN'—the first helps to create a new analysis, while the second enables the ability to re-run the analysis already started. Additionally, the UI of the tool helps to identify the version number and last update date by providing both.

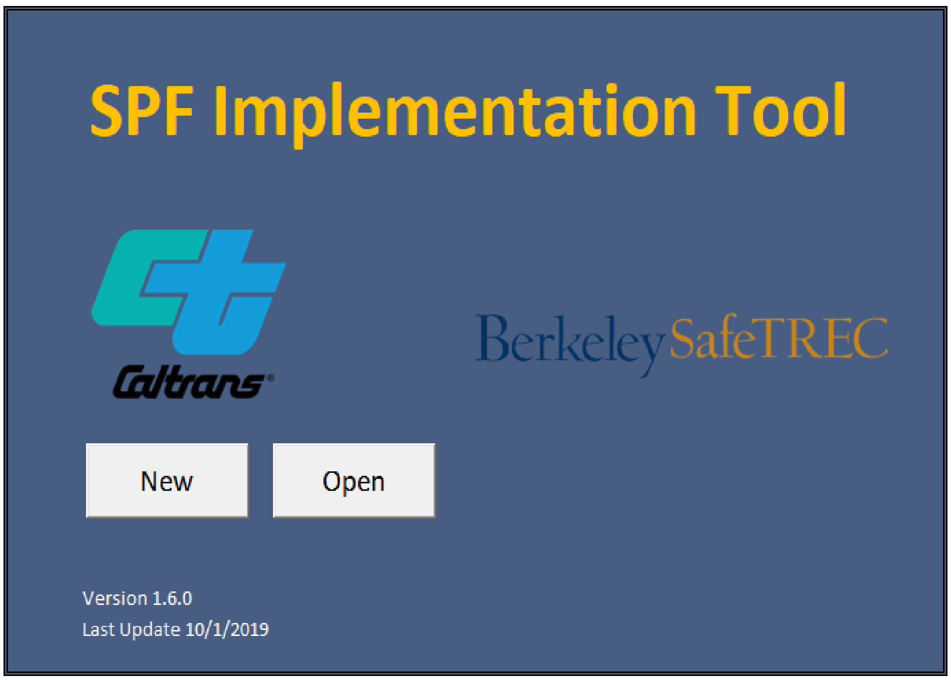

Figure 5.2 User interface of the SPF tool

As described in Section 5.1, this spreadsheet tool involves three stages as follows:

I. Data Input
II. Data Analysis
III. Report Generation

The process involved in each stage will be described in subsequent sub-sections.

5.2.1. Data Input

As seen in Figure 5.1, data input is the first stage in the SPF tool. The data required for safety analysis includes all three facility type infrastructure data and the collision data recommended in Chapter 2 and Chapter 4. The raw data should match the analysis period selected for identifying HCCLs. This is a primary key for the network screening. Any missing data input leads to developing an incorrect list of high collision concentration locations for investigations. The data structure required for the analysis should be in CSV format. The tool is made compatible only with this format and all other files will be excluded due to difficulty of processing within the Tool. A screenshot of this stage is shown in Figure 5.2.

Four data files from TASAS-TSN in CSV format for the selected analysis period:

- 3 infrastructure data files—segment, intersection and ramp
- 1 collision data file

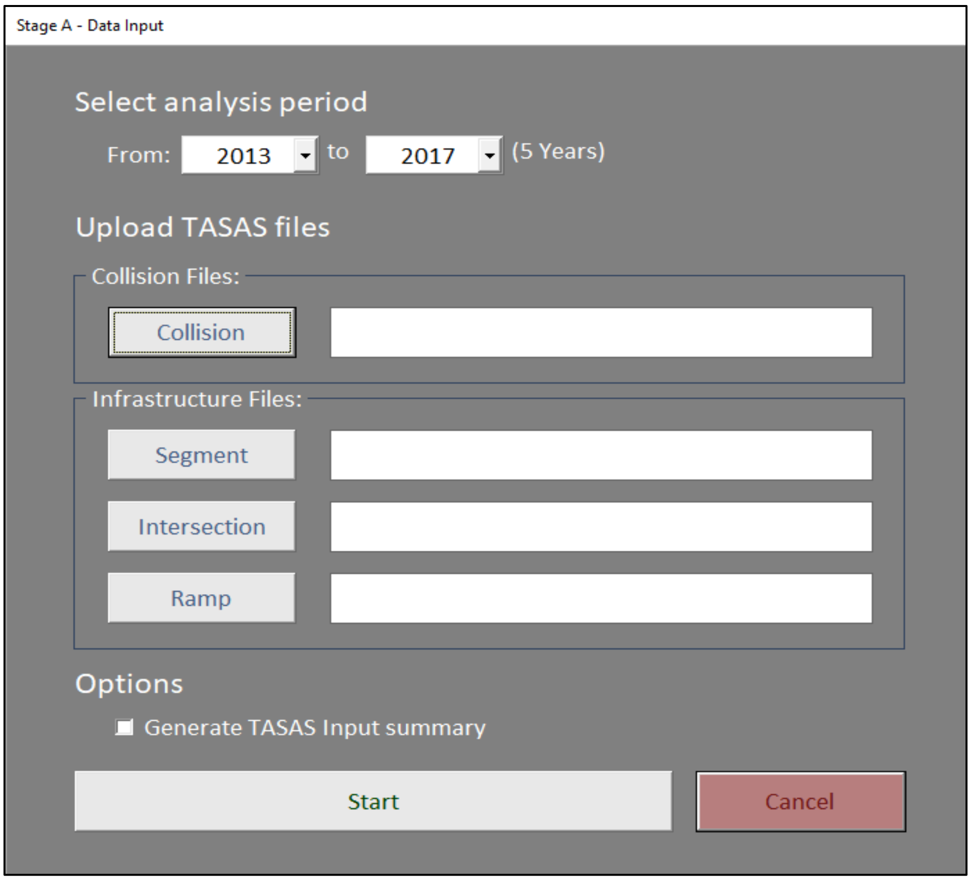

Figure 5.3 Data input module in the tool

To initiate a new analysis, the user must click the 'New' button on the tool interface as shown in Figure 5.2. This will take the user to the Stage A – Data preparation, during which required collision and infrastructure data, including the highway segment, intersection, and ramp data can be imported into the tool as shown in Figure 5.3. This stage also allows the user to generate a summary of TASAS data imported, by checking the appropriate box. After importing the necessary data, the data preparation for the analysis will be processed by clicking the 'START' button at the bottom of the page. If no errors are encountered, the user will automatically obtain a two main outputs—a Preprocessed File for the Stage B which will be stored within the tool, and a Summary Report of the imported data. Table 5.1 shows the format of the summary table of imported data.

Table 5.1 Format of the summary table of imported data (2013-2017)

	# collisions					# observations		
DISTRICTS	2013	2014	2015	2016	2017	HIGHWAY SEG	INTERSECTION	RAMP
DISTRICT 1								
DISTRICT 2								
DISTRICT 3								
DISTRICT 4								
DISTRICT 5								
DISTRICT 6								
DISTRICT 7								
DISTRICT 8								
DISTRICT 9								
DISTRICT 10								
DISTRICT 11								
DISTRICT 12								
TOTAL								

Steps involved in this stage are as follows:

1. Identify unique facilities based on the active segments
2. Standardize updates based on the years of analysis
3. Estimate intersection influence distance based on the override length in the intersection data file
4. Merge infrastructure and collision data files based on the post mile

Output from this stage is the summary tables for the imported infrastructure and collision data and the preprocessed file for the Stage B.

5.2.2. Data Analysis

This is the second of three stages in the tool. After performing Stage A, the tool will take the user to Stage B to perform analysis based on need with preprocessed file stored in the tool from the Stage A. The user interface, as shown in Figure 5.4, will provide information about the preprocessed data file, and allow the user to choose the appropriate PSI threshold, detection method and window length for analysis. Initiate the query by clicking 'Generate Report,' which will result in an Analysis Results file that can be used in Stage C-Reporting.

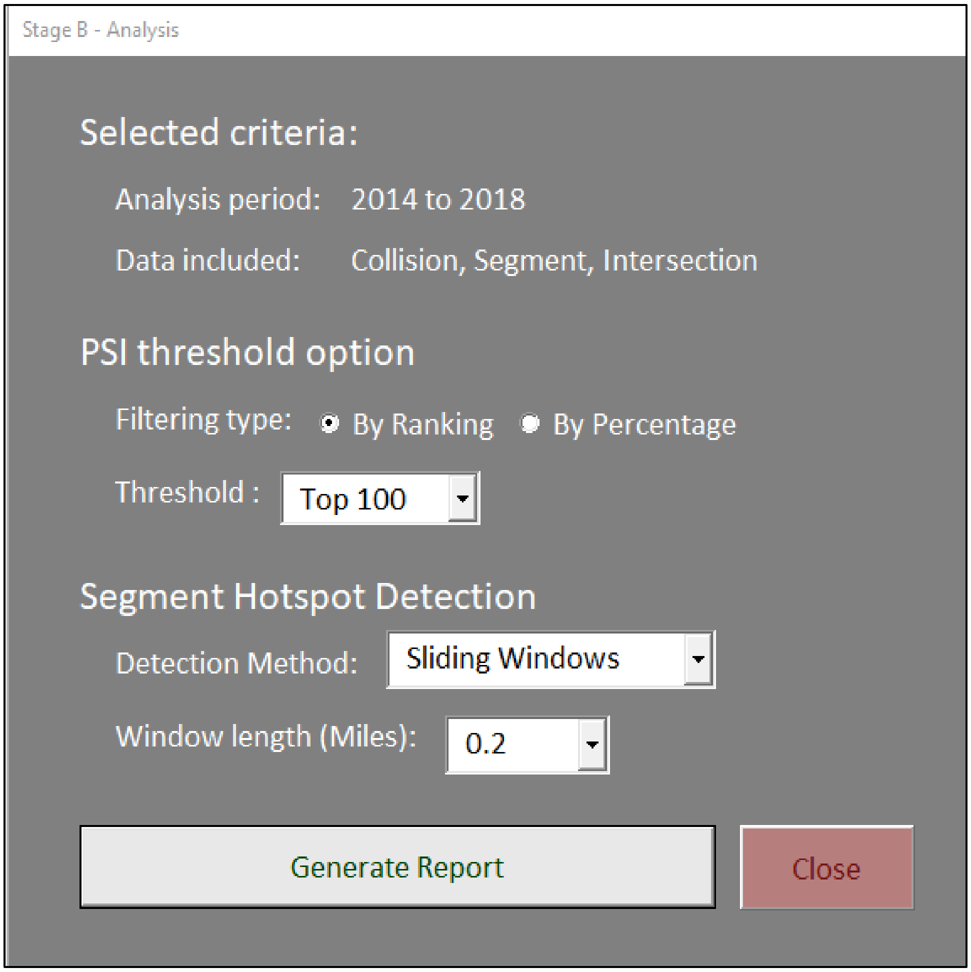

Figure 5.4 Analysis Module within the SPF tool

In this stage, following steps were involved:

- Select the potential for safety improvement (PSI) threshold
    - By ranking (100, 200, 300,400 & 500)
    - By Percentage (1% & 5%)
- Network Screening (Select one detection method and is applicable in the case of segment safety analysis only)
    - Sliding Window
    - Peak Searching
    - Select window length (0.1 – 1.0, with an increment of 0.1)

The stage will take the user to the Excel sheet with a list of locations based on potential for safety improvement (PSI).

5.2.3. Report Generation

This is the final stage in the tool which enables the user to generate a report based on the output from Stage B–an Analysis Results File. This stage allows the user to choose the information required in the report. The tool enables to the user to save or print the report, which can be generated using different options—Traffic Investigation Report Tracking System (TIRTS) format, district report, or statewide/all districts together report, and in two different formats—CSV/Excel and PDF, as shown in Figure 5.5. The greatest advantage of this stage is that the user can generate reports based on the intended purpose, and is able to easily share them with others.

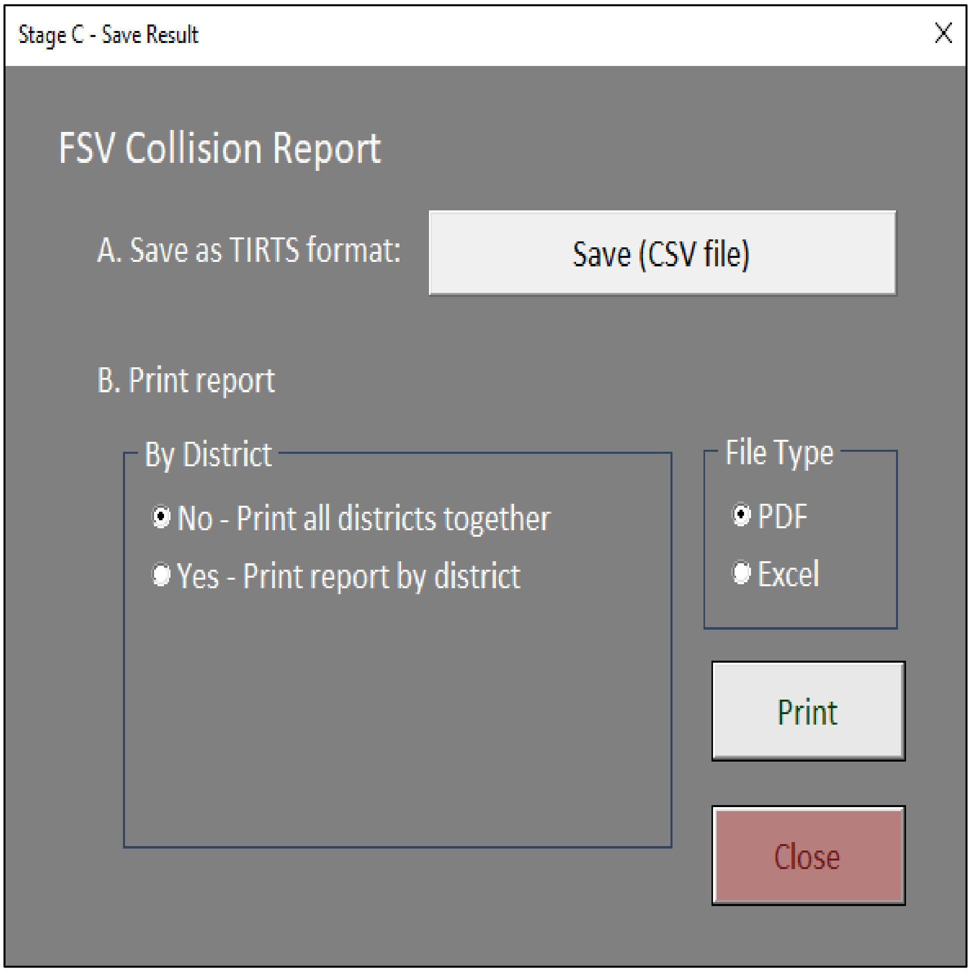

Figure 5.5 Reporting Module of the SPF tool

<u>List of locations based on PSI Threshold</u>:

- Format of the report: PDF
- Application: List differs based on the PSI Threshold
- Advantage: District or State-wide report (sample report shown in Figure 5.6)

10/14/2019
1:34 PM

California Department of Transportation
POTENTIAL INVESTIGATION LOCATIONS
TOTAL Collision Report from 2013 to 2017

District	Facility Type	Location Description	Main Line Avg ADT	X Street Avg ADT	Actual Coll.	SPF Type	Pred. Coll. (SPF)	Exp. Coll. (EB)	PSI (EB-SPF)
5	Intx	17 SCR 10.482 TRINITY MISSION PVT RD	58,198	71	111	I1	23.36	105.65	82.29
7	Ramp	110 LA 25.488 110/NB OFF TO NB 5/BARCLAY	40,069		105	Ramp	8.33	88.90	80.56
7	Ramp	10 LA 14.13 010/SEG EB VERMNT TO SB 110	68,449		71	Ramp	12.26	63.98	51.71
7	Ramp	710 LA 24.16 710/SB ON FR RTE 60	39,732		70	Ramp	9.18	60.66	51.48
3	Ramp	50 SAC R12.205 050/EB OFF TO SUNRISE	23,040		88	Ramp	44.07	86.40	42.33
8	Ramp	210 SBD R24.052 210/EB OFF TO WATERMAN/18	1,001		60	Ramp	5.68	47.67	42.00
7	Ramp	405 LA 29.128 405/NB OFF TO RTE 10	48,529		57	Ramp	9.35	49.80	40.44
12	Ramp	22 ORA R10.022 022/SEG EB OFF TO SB5/NB5/NB57	1,001		57	Ramp	6.38	46.52	40.14
7	Ramp	2 LA 15.387 002/SEG WB OFF TO SB 5/RIVS	39,500		70	Ramp	27.95	67.64	39.68
7	Ramp	91 LA R17.42 091/WB TO RTE 605	52,916		56	Ramp	9.86	49.33	39.47
4	Ramp	80 SF 5.044L 080/WB OFF TO 5TH/HARRISON	19,700		53	Ramp	12.66	48.31	35.65
11	Ramp	163 SD 3.643 163/NB OFF TO EB RTE 8	21,360		59	Ramp	22.64	56.51	33.86
8	Intx	395 SBD R5.613 PHELAN RD LT/MAIN ST RT	26,510	101	59	I2	21.55	55.25	33.70

Figure 5.6 Sample statewide list generated based on PSI

<u>TIRTS format:</u>

- Format of the report: CSV
- Application: To upload in the TIRTS
- Advantage: Compatible with existing TIRTS structure (sample report shown in Figure 5.7)

District	Facility Type	Begin County	Begin Route Three digits (e.g. 005, 163)	Begin Route Suffix	Begin Postmile Prefix	Begin Postmile (Numbers Only)	Begin Postmile Suffix	Main County	Main Route Three digits (e.g. 005, 163)	Main Route Suffix	Main Postmile Prefix	Main Postmile (Numbers Only)	Main Postmile Suffix	End County	End Route Three digits (e.g. 005, 163)	End Route Suffix	End Postmile Prefix	End Postmile (Numbers Only)	End Postmile Suffix	Highway Group	Population Group	ADT	SPFI_XSTREET_ADT	SPFR_DESIGN_DESC	SPFR_ON_OFF_CODE	SPFI_DESIGN_CODE	SPFI_CONTROL_CODE	SPFI_LIGHTED_IND	SPFI_MAIN_FLOW_CODE	SPFI_CROSS_FLOW_CODE	SPFI_MAIN_LANES_AMT	SPFI_CROSS_LANES_AMT	SPFH_LENGTH_MILES_AMT	SPFH_TERRAIN_CODE	SPFH_DESIGN_SPEED	SPFH_RT_SURF_TYPE_CODE	SPFH_MEDIAN_BARRIER_CODE	SPFH_MEDIAN_TYPE_CODE	SPFH_MEDIAN_WIDTH	SPFH_LT_SURF_TYPE_CODE	Investigation Type	Initiation Comments
10	H	TUO	108		R	9.36								TUO	108		R	9.54		D	B	8078											0.22	M	65	H	Z	J	22	H		
01	H	DN	101			8.53								DN	101			8.57		U	R	5467											0.28	R	50	H	Z	B	0	H		
06	H	TUL	216			8.15								TUL	216			8.49		U	R	4212											1.55	F	60	H	Z	B	0	H		
02	I	LAS	299			12.10		LAS	299			12.15		LAS	299			12.20		U	R	1512	11			IT	A	N	P	P	2.00	2.00										
04	I	ALA	238			14.24		ALA	238			14.29		ALA	238		R	14.52		D	B	32000	18001			IT	P	Y	P	P	6.00	2.00										
05	I	SB	154		R	3.10		SB	154		R	3.15		SB	154		R	3.20		U	B	12500	1350			IF	B	N	P	P	2.00	2.00										
07	R							LA	110			25.49								D	U	40211		Split Ramp	OFF																	
12	R							ORA	605		R	1.22								D	U	9200		Loop - without Left Turn	OFF																	
04	R							SCL	101			48.71								D	U	3150		Diamond Type	ON																	

Figure 5.7 Sample TIRTS format

5.3. Summary

This chapter describes the design and development of the SPF tool for network screening to identify HCCLs. The spreadsheet tool was developed using MS Excel Macro based on the required functionality from the Caltrans' safety experts. This chapter also describes in detail the application of the tool and its outputs. The following chapter provides the summary and conclusion of the project, and the key take-aways.

Chapter 6. Summary and Conclusions

This chapter summarizes the project with the methodology adopted for the development of California-specific Safety Performance Functions and final output—MS Excel Macro Spreadsheet tool. The chapter concludes with the key takeaways from this project.

6.1. Summary

As part of the effort to improve highway safety, the California Department of Transportation is developing new safety network screening practices. This effort encompasses two main components: (i) develop state-of-the-practice Safety Performance Functions (SPFs); and (ii) develop a tool to implement a network screening method that utilizes these SPFs. Accomplishing both components will serve as the first phase towards implementation of such models. SPFs are statistical models used for high collision concentration location (HCCL) identification procedures, as described in the Highway Safety Manual. The SPFs that are currently described in the HSM are jurisdiction-specific and may not apply to jurisdictions without calibration. To overcome such challenges, California-specific SPFs should be developed for all the three facilities based on data availability in the Traffic Accident Surveillance and Analysis System –Transportation System Network (TASAS-TSN) for the State highway system.

To better understand the current system in place—Table C—and the interactions and workflow between and among these entities and the sequence of work activities, we will develop process maps. While the necessary and specific details of the individuals who are routinely involved in the generation of existing Table C network screening process within Caltrans were available, there was limited documentation of the various components involved in the decision-making process. Review of report titled 'Summary Report of Task Force's Findings and Recommendations' by the Caltrans task force in 2002, assisted in providing a theoretical background behind the algorithms used to process the collision data, but was limited in terms of identifying the key activities involved in the process. The State Highway Safety Improvement Program Guideline describes the steps following HCCL identification by traffic safety engineers to assess these high collision concentration locations for potential improvement (HSIP, Caltrans 2017).

Based on the SPF data needs project and other related Caltrans projects, a thorough data exploration was implemented, which included assessment of data availability in the TASAS-TSN. This helped to identify the suitability of the most recent five years of data (2013-2017) for infrastructure and collision data for developing SPFs. Data was subject to a cleaning process, which included checks for update consistency, attention to whether the facility type was open/closed during the analysis period, and other considerations. Later, segmentation procedures, which are groupings of homogeneous segments, were incorporated. Trimming intersection influence distances was also conducted. After data cleaning and segmentation, infrastructure data was then merged with the collision data including severity level. SPFs were developed for all the three highway facility types—segments, intersections, and ramps. The developed SPFs will then be incorporated into an Excel macro based tool toward applying the desirable network screening methods to identify high crash concentration locations. The tool is expected have the ability to generate reports based on the needs of the Caltrans Traffic Safety Investigation team, and be incorporated into the spreadsheet tool for network screening. The SPF tool enhancement project was effective in enhancing the SPF tool developed to be compatible with Transportation System Network Replacement project.

6.2. Conclusions

This project was successful in achieving its goal—to develop the first version of an implementable tool utilizing best-practice safety performance evaluation procedures for improving highway safety on the California State Highway System—a functional tool that can conduct SPF-based network screening analyses for identifying HCCLs. More specifically, the outcome includes the development of seventeen segment SPFs, three intersection SPFs, and one ramp SPF, as described in Chapter 4. SPFs were developed based on total collisions, and combination of fatal, severe, and visible injury levels (FSV). Based on the performance measures, Type 2 SPFs for FSV were recommended over Type 2 SPFs for all collisions. An MS excel macro spreadsheet tool was then designed and developed incorporating all of the Caltrans-reviewed SPFs. For the network screening process of a highway segment, the two most common approaches—sliding window and peak searching—were incorporated in the tool. In addition, the tool will generate the output in two different formats: (i) Traffic Investigation Report Tracking System (TIRTS), and (ii) state-wide/district-wise potential site for investigation list format. Furthermore, the SPF tool has the capability to be updated with more advanced California-specific Safety Performance Functions.

6.3. Project Takeaway

This research provides Caltrans with a testable version of a SPF-based highway safety assessment procedure using existing infrastructure data. This procedure brings Caltrans closer to implementing more efficient resource allocation for identifying HCCLs. The developed Excel spreadsheet tool is simple and easy to operate. In addition, the research also provided guidelines for incorporating additional SPFs and re-calibrating existing SPFs for network screening based on the availability of geometric characteristics of the roadway. Finally, various report generation options enable Caltrans experts to create reports based on the described specifications. As part of this project, the limited documentation of existing practice of safety analysis—Table C—coupled with Caltrans' desire to evaluate the value of transitioning to other network screening methods, resulted in an effort to identify the entities that contribute to, or are a part of, the process, and the entire process was mapped through the process mapping technique—relationship maps, cross-functional maps and flowchart maps. Both the tool and the maps can be used to optimize resource allocation across different highway safety related projects. In the future, Caltrans will be able to use the SPF tool to identify high collision concentrations locations through network screening which will ultimately result in the reduction of traffic-related fatalities and injuries in California.

References

Highway Safety Manual 2010, AASHTO, 1st edition

Methods for identifying high collision concentrations for Identifying potential safety improvements: development of safety Performance functions for California, Report submitted to Caltrans by Institute of Transportation Studies, University of California Berkeley

Report on Table C, Caltrans 2002

State Highway Safety Improvement Program, Caltrans 2017

SPF Data Need Project, Report submitted to Caltrans by Safe Transportation Research and Education Center, University of California Berkeley

Appendix A – Data Structure used for SPF Development

I.Infrastructure Data

This section provides infrastructure data structure for all the three facility types—highway, intersection and ramp—for safety screening.

Table 6.1: Highway Infrastructure Data Structure for Safety Analysis

Field Name	Description
SPFH_DISTRICT	District number segment belongs
SPFH_COUNTY	County code
SPFH_ROUTE	Route number
SPFH_RTE_SFX	Route suffix
SPFH_PM_PFX	Post mile prefix
SPFH_BEGIN_PM	Begin post mile
SPFH_END_PM	End post mile
SPFH_PM_SFX	Post mile suffix
SPFH_HIGHWAY_GROUP_CODE	Highway group
SPFH_LENGTH_MILES_AMT	Length of segment in miles
SPFH_ADT_AMT	Average daily traffic
SPFH_POPULATION_CODE	Population
SPFH_TERRAIN_CODE	Terrain
SPFH_DESIGN_SPEED	Design speed
SPFH_BREAK_DESC	Break description (End/Begin of District/County/ Route)
SPFH_EQUATE_CODE	Equate code
SPFH_RT_LANES_AMT	Right side - Number of lanes
SPFH_RT_SURF_TYPE_CODE	Right side - Surface type
SPFH_RT_TRAV_WAY_WIDTH	Right side - Travel way width
SPFH_RT_I_SHD_TOT_WIDTH	Right side - Inner Shoulder total width
SPFH_RT_O_SHD_TOT_WIDTH	Right side - Outer Shoulder total width
SPFH_MEDIAN_BARRIER_CODE	Median barrier
SPFH_MEDIAN_TYPE_CODE	Type of median
SPFH_MEDIAN_WIDTH	Width of median
SPFH_MEDIAN_WIDTH_VAR_CODE	Median width code
SPFH_LT_LANES_AMT	Left side - Number of lanes

SPFH_LT_SURF_TYPE_CODE	Left side - Surface type
SPFH_LT_TRAV_WAY_WIDTH	Left side - Travel way width
SPFH_LT_I_SHD_TOT_WIDTH	Left side - Inner Shoulder total width
SPFH_LT_O_SHD_TOT_WIDTH	Left side - Outer Shoulder total width
SPFH_R_ODOMETER_BEGIN	Right side - Begin Odometer reading
SPFH_R_ODOMETER_END	Right side - End Odometer reading
SPFH_L_ODOMETER_BEGIN	Left side - Begin Odometer reading
SPFH_L_ODOMETER_END	Left side - End Odometer reading
SPFH_BEGIN_DATE	Begin date (Depends on the analysis period)
SPFH_END_DATE	End date (Depends on the analysis period)
SPFH_EXTRACT_DATE	Data extraction date
SPFH_SEG_ORDER_ID	Segment Order Id
SPFH_BEGIN_OFFSET_AMT	Begin Offset
SPFH_RATE_GROUP	Highway Rate Group
SPFH_RATE_GROUP_DESC	Highway Rate Group Description
SPFH_ACCESS_CODE	Highway Access
SPFH_ACCESS_CODE_DESC	Highway Access Description
SPFH_LANDMARK_SHORT_DESC	Highway Landmark

Table 6.2: Intersection Infrastructure Data Structure for Safety Analysis

Field Names	Description
SPFI_DISTRICT	District
SPFI_COUNTY	County
SPFI_ROUTE	Route
SPFI_RTE_SFX	Route suffix
SPFI_MAIN_BEGIN_PM_PFX	Mainline Begin Post mile Prefix
SPFI_MAIN_BEGIN_PM	Mainline Begin Post mile (Buffer)
SPFI_MAIN_BEGIN_PM_SFX	Mainline Begin Post mile Suffix
SPFI_MAIN_PM_PFX	Mainline Post mile Prefix
SPFI_MAIN_PM	Mainline Post mile
SPFI_MAIN_PM_SFX	Mainline Post mile Suffix
SPFI_MAIN_END_PM_PFX	Mainline End Post mile Prefix
SPFI_MAIN_END_PM	Mainline End Post mile (Buffer)
SPFI_MAIN_END_PM_SFX	Mainline End Post mile Suffix
SPFI_HIGHWAY_GROUP	Highway Group
SPFI_CITY_CODE	City
SPFI_POPULATION_GROUP	Population
SPFI_DESIGN_CODE	Intersection Design
SPFI_DESIGN_DESC	Intersection Design Description
SPFI_DESIGN_DATE	Date of Design
SPFI_LIGHTED_IND	Presence of Light Condition at Intersection
SPFI_LIGHTED_BEGIN_DATE	Begin date of Light Condition at
SPFI_MAIN_SIGNAL_MAST_ARM_IND	Presence of Mainline Mast Arm Signal
SPFI_MAIN_LEFT_CHANNEL_CODE	Presence of Mainline Left Channel
SPFI_MAIN_RIGHT_CHANNEL_CODE	Presence of Mainline Right Channel
SPFI_MAIN_FLOW_CODE	Mainline Flow description
SPFI_CROSS_SIGNAL_MAST_ARM_IND	Presence of Cross street Mast Arm Signal
SPFI_CROSS_LEFT_CHANNEL_CODE	Presence of Cross street Left Channel
SPFI_CROSS_RIGHT_CHANNEL_CODE	Presence of Cross street Right Channel
SPFI_CROSS_FLOW_CODE	Cross street Flow description
SPFI_CONTROL_CODE	Intersection Control Condition
SPFI_CONTROL_DESC	Intersection Control Condition Description
SPFI_CONTROL_DATE	Intersection Control Condition Begin date
SPFI_MAIN_LANES_AMT	Mainline - Number of lanes

SPFI_MAIN_OVERRIDE_LENGTH_AMT	Mainline - Override length (Buffer)
SPFI_CROSS_LANES_AMT	Cross street - Number of lanes
SPFI_CROSS_OVERRIDE_LENGTH_AMT	Cross street - Override length
SPFI_MAINLINE_ADT	Mainline - Average Daily Traffic
SPFI_X_ROUTE	Cross street route number
SPFI_X_RTE_SFX	Cross street route number suffix
SPFI_X_BEGIN_PM_PFX	Cross street begin post mile prefix
SPFI_X_BEGIN_PM	Cross street post mile
SPFI_X_BEGIN_PM_SFX	Cross street begin post mile suffix
SPFI_X_PM_PFX	Cross street post mile prefix
SPFI_X_PM	Cross street post mile
SPFI_X_PM_SFX	Cross street post mile suffix
SPFI_X_END_PM_PFX	Cross street end post mile prefix
SPFI_X_END_PM	Cross street post mile
SPFI_X_END_PM_SFX	Cross street begin post mile suffix
SPFI_XSTREET_ADT	Cross street - Average Daily Traffic
SPFI_R_BEIN_ODOMETER	Mainline Right-side begin odometer
SPFI_R_ODOMETER	Mainline Right-side odometer
SPFI_R_END_ODOMETER	Mainline Right-side end odometer
SPFI_L_BEGIN_ODOMETER	Mainline Left-side begin odometer
SPFI_L_ODOMETER	Mainline Left-side odometer
SPFI_L_END_ODOMETER	Mainline Left-side end odometer
SPFI_X_R_BEIN_ODOMETER	Cross street Right-side begin odometer
SPFI_X_R_ODOMETER	Cross street Right-side odometer
SPFI_X_R_END_ODOMETER	Cross street Right-side end odometer
SPFI_X_L_BEGIN_ODOMETER	Cross street Left-side begin odometer
SPFI_X_L_ODOMETER	Cross street Left-side odometer
SPFI_X_L_END_ODOMETER	Cross street Left-side end odometer
SPFI_SKEW_ANGLE	Intersection skew angle
SPFI_MAIN_LANE_FUN_CLASS	Main lane functional class
SPFI_MAIN_LANE_WIDTH	Width of main lane
SPFI_CROSS_STREET_WIDTH	Width of cross street
SPFI_BEGIN_DATE	Begin date of intersection update
SPFI_END_DATE	End date of intersection update
SPFI_EXTRACT_DATE	Data extraction date

SPFI_SEG_ORDER_ID	Mainline segment order Id
SPFI_X_SEG_ORDER_ID	Cross street segment order Id
SPFI_RATE_GROUP	Intersection rate group
SPFI_RATE_GROUP_DESC	Intersection rate group description
SPFI_INTERSECTION_NAME	Name of the intersection

Table 6.3: Ramp Infrastructure Data Structure for Safety Analysis

Field Names	**Description**
SPFR_DISTRICT	District
SPFR_COUNTY	County
SPFR_ROUTE	Route
SPFR_RTE_SFX	Route suffix
SPFR_PM_PFX	Post Mile prefix
SPFR_PM	Post Mile
SPFR_PM_SFX	Post Mile Suffix
SPFR_DESIGN_DESC	Ramp Design Description
SPFR_ON_OFF_CODE	ON/OFF Ramp
SPFR_CITY_CODE	City
SPFR_ADT	Ramp Average Daily Traffic
SPFR_POP_GROUP	Population group
SPFR_HIGHWAY_GROUP	Highway group
SPFR_R_ODOMETER	Right-side Odometer
SPFR_L_ODOMETER	Left-side Odometer
SPFR_RAMP_LENGTH	Length of ramp
SPFR_RAMP_LANES_AMT	Number of lanes in ramp
SPFR_RAMP_LANE_WIDTH	Width of lane
SPFR_ORDER_ID	Order Id
SPFR_RATE_GROUP	Rate group code
SPFR_RATE_GROUP_DESC	Rate group description
SPFR_BEGIN_DATE	Begin date of update
SPFR_END_DATE	End date of update
SPFR_EXTRACT_DATE	Data extraction date
SPFR_RAMP_DESCRIPTION	Ramp description

II.Collision Data

Table 6.4: Collision Data Structure for Safety Analysis

Field Names	**Description**
ACCIDENT_YEAR	Year accident occurred
ACCIDENT_NUMBER	Accident number
DISTRICT	District accident occurred
COUNTY	County code within district
COUNTY_NAME	County name within district
CITY	City code within county
CITY_NAME	City name within county
ROUTE_NAME	Route name within the county
ROUTE_SUFFIX	Route suffix
PM_PREFIX	Prefix to the post mile
POSTMILE	Post mile
PM_SUFFIX	Suffix to the post mile
FILE_TYPE	Facility type – highway/intersection/ramp
ACCIDENT_DATE	Accident date
ACCIDENT_TIME	Accident time
COMMON_ACCIDENT_NUMBER	Combination of jurisdiction, badge id, date &time
PRIMARY_COLL_FACTOR	Primary collision factor
SEVERITY_LEVEL	Level of severity of accident

Appendix B –Total Collision SPFs Developed

This appendix shows the statistically significant variables in total collision SPFs, and the corresponding route and county clustering results.

1. Segment SPF

Class – 1 Rural 2 Lane Non-Freeway Undivided

Variables	Estimate
Intercept	-8.012
ADT	1.026
Segment length	0.922
Terrain: F	-0.449
Terrain: R	-0.491
Right surface type: H	-0.204
Right surface type: B	-0.264
Right surface type: E	-2.922
Left surface type: P	-0.602
Median type: A	1.037
Route_cluster_1	-0.613
Route_cluster_2	0.630
Route_cluster_3	-0.167
Route_cluster_4	0.180
Route_cluster_5	0.420
Route_cluster_6	1.245
County_cluster_1	-0.689
County_cluster_2	0.112
County_cluster_3	-0.339
County_cluster_4	-0.141
County_cluster_5	0.457
Theta (overdispersion)	1.721
AIC	493490
Log likelihood	-246721.00

1. Route Clusters					
1	2	3	4	5	6
299	36	254	29	169	65
44	96	186	1	199	156
121	175	263	104	59	183
58	26	173	120	78	229
18	174		111	79	46
38	92		154	37	145
	155		41	130	203
	178		269	150	
	2		118	227	
			138	184	
			62		

1. County Clusters				
1	2	3	4	5
IMP	SD	MER	ALP	HUM
SHA	GLE	SJ	SCR	MEN
SIS	NEV	STA		PLU
TEH	YOL	TUO		NAP
KIN	SM	ORA		SCL
INY	KER	LAS		
		MRN		
		SLO		
		SBD		

Class – 2 Rural 2 Lane Non-Freeway Divided

Variables	Estimate
Intercept	-10.298
ADT	0.980
Segment length	1.593
Median type: F	-0.583
Year 2	-0.137
Route_Cluster_1	-0.383
Route_Cluster_2	0.575
County_Cluster_1	-0.263
County_Cluster_2	0.545
County_Cluster_3	0.103
Theta(overdispersion)	0.6037
AIC	564216
Log likelihood	-282097.17

2. Route Clusters	
1	**2**
1	96
140	88
120	26
65	165
116	37
166	
46	
137	
62	

2. County Clusters		
1	**2**	**3**
AMA	DN	LAK
SJ	MEN	LAS
SM	CAL	NAP
MAD	MPA	SON
		MON
		SB
		VEN
		FRE
		RIV
		MNO

Class – 3 Rural 3+ Lane Non-Freeway Undivided

Variables	Estimate
Intercept	-10.193
ADT	0.927
Segment length	1.585
Design speed	0.008
Terrain: R	-0.079
Right surface type: H	-0.086
Year 1	-0.184
Year 2	-0.207
Year 3	-0.115
Route_Cluster_1	0.376
Route_Cluster_2	-0.306
County_Cluster_1	-0.501
County_Cluster_2	0.325
County_Cluster_3	0.027
Theta (Overdispersion)	0.6089
AIC	563165
Log likelihood	-281581.85

3. Route Clusters	
1	**2**
299	12
36	120
175	108
20	395
88	97
89	14
49	
79	
94	
3	
92	
121	
168	

3. County Clusters		
1	**2**	**3**
IMP	DN	BUT
SUT	HUM	LAS
SM	MEN	SCL
	MPA	SON
	LA	SB
	PLU	SLO
	SAC	VEN
	NAP	RIV
	MNO	SBD

Class – 4 Rural 3+ Lane Non-Freeway Divided

Variables	Estimate
Intercept	-8.382
ADT	0.858
Segment length	1.411
Terrain: F	-0.185
Right surface type: B	-0.537
Left surface type: M	0.253
Left surface type: C	0.201
Median barrier: Z	-0.616
Median barrier: G	-0.607
Median type: F	-0.556
Median type: H	-0.140
Median type: G	-0.991
Median type: J	-0.212
Year 1	-0.182
Year 2	-0.204
Year 3	-0.116
Route_Cluster_1	0.247
Route_Cluster_2	-0.200
County_Cluster_1	-0.203
County_Cluster_2	0.210
County_Cluster_3	0.451
Theta (Overdispersion)	0.6751
AIC	553715
Log likelihood	-276835.56

4. Route Clusters	
1	2
88	1
49	33
67	50
76	43
79	62
70	
17	
129	

4. County Clusters		
1	2	3
COL	LAK	DN
SM	CAL	HUM
SOL	MPA	MEN
SLO	LA	TRI
KIN	VEN	SAC
MAD	FRE	NAP
SBD	RIV	
INY	MNO	

Class – 5 Rural 2-4 Lane Freeway Undivided

Variables	Estimate
Intercept	-10.199
ADT	1.049
Segment length	1.514
Terrain: F	-0.716
Terrain: R	-0.609
Right surface type: H	-0.312
Right surface type: B	-0.82
Year 2	-0.132
Route_Cluster_1	-0.027
Route_Cluster_2	-1.276
County_Cluster_1	-0.176
County_Cluster_2	0.27
County_Cluster_3	0.565
Theta (Overdispersion)	0.6446
AIC	558286
Log likelihood	-279129.21

5. Route Clusters	
1	**2**
36	124
29	83
4	
120	
108	
32	
395	
97	
178	

5. County Clusters		
1	**2**	**3**
SD	LAK	DN
SHA	MEN	HUM
SIS	CAL	ALP
TEH	MPA	MOD
NEV	TUO	MNO
CC	LAS	
SLO	PLU	
KIN	TRI	
TUL	SAC	
SBD	NAP	
	SCR	
	FRE	

Class – 6 Rural 2-4 Lane Freeway Divided

Variables	Estimate
Intercept	-9.896
ADT	0.911
Segment length	1.409
Design speed	0.01
Terrain: F	-0.675
Terrain: R	-0.585
Right surface type: H	0.255
Right surface type: M	0.578
Right surface type: F	0.383
Right surface type: C	-0.738
Left surface type: H	0.05
Left surface type: C	0.268
Median barrier: G	-0.101
Median barrier: N	0.072
Median barrier: J	0.142
Median barrier: R	0.061
Median barrier: D	0.327
Median type: H	0.074
Median type: G	-0.923
Median type: Q	-0.322
Median type: P	-1.092
Year 1	-0.189
Year 2	-0.206
Year 3	-0.117
Route_Cluster_1	0.698
Route_Cluster_2	0.132
Route_Cluster_3	-0.181
Route_Cluster_4	-0.683
County_Cluster_1	-0.319
County_Cluster_2	0.126
Theta (Overdispersion)	0.6788
AIC	553627
Log likelihood	-276782.63

6. Route Clusters			
1	**2**	**3**	**4**
36	101	299	108
37	29	1	7
178	580	152	86
215	395	5	125
	70	120	505
	80	132	14
	198	111	62
	58	8	
	168	15	
	10	113	
	18	40	
	60		

6. County Clusters	
1	**2**
IMP	MER
BUT	LA
LAS	SAC
COL	SCL
NEV	VEN
SOL	RIV
SLO	
KER	
TUL	
SBD	
INY	

Class – 7 Rural 5+ Lane Freeway

Variables	Estimate
Intercept	-8.154
ADT	0.878
Segment length	1.424
Terrain: F	-0.614
Terrain: R	-0.513
Left surface type: H	0.038
Left surface type: C	0.367
Median barrier: Z	-0.645
Median barrier: G	-0.436
Median barrier: H	-0.505
Median barrier: S	-0.333
Median type: H	-0.206
Median type: G	-0.959
Median type: K	-0.227
Median type: J	-0.23
Median type: Q	-0.641
Year 1	-0.185
Year 2	-0.205
Year 3	-0.119
Route_Cluster_1	-0.142
Route_Cluster_2	0.167
Route_Cluster_3	-0.575
County_Cluster_1	-0.21
County_Cluster_2	0.17
Theta (Overdispersion)	0.673
AIC	554258
Log likelihood	-277103.76

7. Route Clusters		
1	**2**	**3**
152	29	905
5	580	14
132	70	
205	80	
50	198	
65		
113		
126		

7. County Clusters	
1	2
IMP	HUM
SD	MEN
BUT	SCR
SIS	RIV
NEV	
CC	
SLO	
MAD	
SBD	

Class – 8 Urban 2 Lane Non-Freeway Undivided

Variables	Estimate
Intercept	-11.735
ADT	1.084
Segment length	1.437
Design speed	0.014
Terrain: F	-0.694
Terrain: R	-0.625
Right surface type: B	-0.566
Right surface type: M	0.221
Right surface type: C	0.274
Right surface type: O	0.808
Median type: B	0.565
Year 1	-0.14
Year 2	-0.164
Route_Cluster_1	-0.844
Route_Cluster_2	-0.339
Route_Cluster_3	-1.343
Route_Cluster_4	0.279
Route_Cluster_5	-0.581
Route_Cluster_6	0.891
Route_Cluster_7	-0.076
County_Cluster_1	0.197
County_Cluster_2	-0.179
Theta (Overdispersion)	0.6673
AIC	555584
Log likelihood	-277769.04

8. Route Clusters						
1	2	3	4	5	6	7
27	1	183	4	28	173	12
46	38		58	32	193	16
145	63		59	34	200	18
267	65		70	108		33
	68		79	111		74
	104		92	116		99
	113		94	142		120
	154		160	217		132
	156		162	233		138
			174			152
			178			395
			180			
			189			
			198			
			227			

8. County Clusters	
1	2
MEN	SJ
SON	SD
MON	ORA
	ED
	NEV
	CC
	MRN
	FRE
	KER
	KIN
	TUL

Class – 9 Urban 2 Lane Non-Freeway Divided

Variables	Estimate
Intercept	-10.385
ADT	1.017
Segment length	1.516
Median width	-0.004
Terrain: F	-0.725
Terrain: R	-0.582
Right surface type: M	0.323
Right surface type: C	0.574
Right surface type: O	0.953
Left surface type: H	0.251
Left surface type: M	0.206
Route_Cluster_1	-2.561
Route_Cluster_2	0.186
Route_Cluster_3	-0.369
County_Cluster_1	-0.245
County_Cluster_2	0.21
Theta (Overdispersion)	0.639
AIC	558709
Log likelihood	-279337.36

9. Route Clusters		
1	2	3
218	59	1
	94	12
	99	20
	180	32
		38
		68
		116
		183

9. County Clusters	
1	**2**
TUO	HUM
IMP	LAK
NEV	MEN
SLO	LA
MAD	SON
	MON
	VEN

Class – 10 Urban 3+ Lane Non-Freeway Undivided

Variables	Estimate
Intercept	-9.635
ADT	0.951
Segment length	1.521
Terrain: F	-0.739
Terrain: R	-0.615
Right surface type: C	0.513
Left surface type: H	0.119
Route_Cluster_1	-1.532
Route_Cluster_2	-0.292
Route_Cluster_3	0.127
County_Cluster_1	-0.204
County_Cluster_2	0.153
Theta (Overdispersion)	0.6313
AIC	559772
Log likelihood	-279873.03

10. Route Clusters		
1	2	3
61	1	36
72	23	101
123	28	237
	49	
	62	
	65	
	86	
	113	
	120	
	152	
	247	
	395	

10. County Clusters	
1	2
STA	LAK
SD	LA
BUT	ORA
SHA	ALA
SM	NAP
MAD	SON
TUL	MON
	VEN
	FRE
	RIV

Class – 11 Urban 3+ Lane Non-Freeway Divided

Variables	Estimate
Intercept	-9.361
ADT	0.907
Segment length	1.481
Terrain: F	-0.237
Right surface type: C	0.176
Left surface type: C	0.173
Median barrier: N	0.049
Median barrier: Y	-0.587
Median barrier: M	3.483
Median type: Q	-0.629
Year 1	-0.187
Year 2	-0.203
Year 3	-0.115
Route_Cluster_1	-0.102
Route_Cluster_2	-0.48
Route_Cluster_3	-1.134
Route_Cluster_4	0.404
Route_Cluster_5	-1.88
Route_Cluster_6	0.165
County_Cluster_1	-0.066
County_Cluster_2	-0.525
County_Cluster_3	0.326
Theta (Overdispersion)	0.6555
AIC	556667
Log likelihood	-278310.59

11. Route Clusters					
1	**2**	**3**	**4**	**5**	**6**
2	1	27	37	61	4
12	9	62	79	66	10
35	39	72	92	82	17
68	40	83	180	185	18
84	47	107	260	213	22
101	86	114			55
132	90	131			88
133	108	164			99
138	129	219			110
280	135				121
395	142				
	152				
	166				
	184				
	204				
	246				
	255				

11. County Clusters		
1	**2**	**3**
SJ	IMP	DN
SD	BUT	HUM
ORA	SIS	MEN
TEH	SUT	LA
ED		NAP
PLA		SCL
YUB		SF
CC		SCR
SM		
SB		
SLO		
VEN		
SBD		

Class – 12 Urban 2-7 Lane Freeway Undivided

Variables	Estimate
Intercept	-7.741
ADT	0.791
Segment length	1.478
Right surface type: H	-0.254
Right surface type: B	-0.921
Median barrier: Z	-0.677
Year 3	-0.033
Route_Cluster_1	-0.11
County_Cluster_1	0.129
Theta (Overdispersion)	0.6299
AIC	559987
Log likelihood	-279983.33

12. Route Cluster
1
1
70
18

12. County Cluster
1
DN
HUM
LAK
MEN
SD
NEV
RIV

Class – 13 Urban 2-7 Lane Freeway Divided

Variables	Estimate
Intercept	-8.255
ADT	0.823
Segment length	1.366
Terrain: F	-0.655
Terrain: R	-0.479
Right surface type: H	0.320
Right surface type: C	0.406
Right surface type: O	1.024
Right surface type: F	-0.746
Left surface type: H	0.161
Left surface type: M	0.691
Left surface type: C	0.338
Median barrier: Z	-0.521
Median barrier: I	-0.270
Median barrier: F	0.288
Median barrier: S	-0.179
Median barrier: D	0.244
Median type: H	-0.094
Median type: G	-0.822
Median type: K	-0.063
Median type: Q	-0.569
Median type: S	-1.041
Year 1	-0.188
Year 2	-0.205
Year 3	-0.120
Route_Cluster_1	-0.415
Route_Cluster_2	0.325
Route_Cluster_3	0.186
Route_Cluster_4	-0.087
Route_Cluster_5	0.484
Route_Cluster_6	0.895
County_Cluster_1	0.239
County_Cluster_2	0.602
County_Cluster_3	-0.407
County_Cluster_4	-0.775
County_Cluster_5	-0.188
Theta (Overdispersion)	0.7073
AIC	550028
Log likelihood	-274976.87

13. Route Clusters					
1	**2**	**3**	**4**	**5**	**6**
73	4	101	20	67	880
13	580	299	1	94	92
14	76	49	33	605	180
	22	5	205	241	215
	70	99		91	
	80	78		51	
	237	8		242	
	87	44		37	
	198	50		10	
	41	84		105	
	118	680		110	
	210	17		710	
	60	58			

13. County Clusters				
1	**2**	**3**	**4**	**5**
DN	SF	SIS	IMP	TUO
HUM		TEH	SHA	SD
SAC		ED		BUT
SCR		GLE		NEV
		MRN		PLA
		SM		SUT
		SLO		YUB
		MAD		CC
				SB
				FRE
				KER
				TUL
				RIV

Class – 14 Urban 8+ Lane Freeway Divided

Variables	Estimate
Intercept	-9.407
ADT	0.972
Segment length	1.477
Median width	-0.004
Terrain: F	-0.729
Terrain: R	-0.581
Right surface type: H	-0.147
Right surface type: B	-0.660
Left surface type: H	-0.109
Left surface type: C	0.095
Left surface type: P	-0.751
Median barrier: G	-0.112
Median barrier: C	0.204
Median barrier: N	0.226
Median barrier: J	0.252
Median barrier: F	0.264
Median barrier: Q	0.191
Median type: H	0.161
Median type: J	0.078
Median type: R	0.320
Median type: T	-0.341
Median type: V	-0.599
Median type: U	-0.765
Median type: S	-0.624
Year 1	-0.185
Year 2	-0.202
Year 3	-0.114
Route_Cluster_1	-0.346
Route_Cluster_2	0.761
Route_Cluster_3	0.268
County_Cluster_1	0.421
County_Cluster_2	-0.105
Theta (Overdispersion)	0.6659
AIC	555836
Log likelihood	-277885.24

14. Route Clusters		
1	**2**	**3**
125	241	4
163	92	580
805	180	405
24	215	22
980		55
23		91
		51
		17
		10
		105
		110
		118

14. County Clusters	
1	**2**
SAC	ORA
SF	MRN
	SCL
	SM
	SON
	VEN
	RIV
	SBD

Class – 15 Rural L/R Alignment Freeway

Variables	Estimate
Intercept	-3.871
Segment length	1.238
Design speed	0.050
Terrain: F	0.113
Terrain: R	-0.093
Right surface type: B	-0.171
Left surface type	0.412
Left surface type: M	-0.123
Left surface type: C	0.927
Median barrier: K	0.236
Median barrier: R	0.436
Median barrier: S	0.477
Median type: H	1.477
Median type: K	0.722
Median type: J	0.715
Median type: Q	0.595
Year 1	-0.193
Year 2	-0.199
Route_Cluster_1	-1.419
Route_Cluster_2	0.233
County_Cluster_1	-1.218
County_Cluster_2	0.240
Theta (Overdispersion)	0.4127
AIC	595001
Log likelihood	-297477.50

15. Route Clusters	
1	**2**
395	120
113	580
138	8
40	70
	80

15. County Clusters	
1	2
SHA	SJ
SIS	LA
KER	BUT
INY	CC
	SBT
	RIV
	SBD

Class – 16 Urban L/R Alignment Freeway

Variables	Estimate
Intercept	-4.634
Segment length	1.219
Design speed	0.054
Median width	0.001
Terrain: F	0.263
Right surface type: H	0.487
Right surface type: B	0.291
Right surface type: C	0.969
Left surface type	-0.439
Left surface type: M	0.324
Left surface type: C	0.579
Median barrier: J	0.394
Median barrier: Q	0.715
Median type: M	-1.819
Year 1	-0.267
Year 2	-0.268
Year 3	-0.167
Route_Cluster_1	1.644
Route_Cluster_2	-0.813
Route_Cluster_3	0.541
County_Cluster_1	0.996
County_Cluster_2	0.187
Theta (Overdispersion)	0.3858
AIC	601447
Log likelihood	-300700.58

16. Route Clusters		
1	2	3
580	125	101
80	905	1
680	73	5
880	160	15
37		94
280		57
		91
		110
		118
		134

16. County Clusters	
1	2
SD	SM
LA	SOL
ORA	MON
SAC	
SF	
RIV	

Class – 17 L/R Alignment Non-Freeway

Variables	Estimate
Intercept	-7.382
ADT	0.738
Segment length	1.470
Terrain: R	-0.078
Right surface type: C	0.206
Left surface type: C	0.138
Left surface type: F	-0.473
Median barrier: Z	-0.744
Median barrier: G	-0.528
Median barrier: Q	-0.449
Route_Cluster_1	0.078
County_Cluster_1	-0.278
Theta (Overdispersion)	0.6337
AIC	559536
Log likelihood	-279754.85

17. Route Cluster
1
101

17. County Cluster
1
BUT SHA MNO

2. Intersection SPF

Rural Intersection SPF

Total collision	
Variables	**Estimate**
Intercept	-7.970
Mainline ADT	0.629
Cross street ADT	0.239
Design: IT	-0.304
Design: IY	-0.245
Cross street lanes amt	0.136
Highway group: L	0.808
Highway group: R	0.624
Highway group: U	0.127
Control: B	0.221
Control: C	0.778
Control: D	0.463
Control: E	0.664
Control: G	0.351
Control: M	1.060
Control: N	0.635
Control: P	0.871
Cross street flow: P	-0.178
Light: Y	0.062
Mainline lanes amt	0.056
Year_3	0.090
Year_4	0.199
Year_5	0.130
Route_cluster_1	-0.612
Route_cluster_2	0.570
Route_cluster_3	-0.334

Route_cluster_4	-1.083
Route_cluster_5	0.690
Route_cluster_6	0.341
Route_cluster_7	0.973
Route_cluster_8	0.438
Route_cluster_9	0.167
Route_cluster_10	1.343
County_cluster_1	-0.504
County_cluster_2	0.422
County_cluster_3	-0.188
County_cluster_4	-0.345
County_cluster_5	-0.924
Theta(overdispersion)	1.531
AIC	65700.000
Log-likelihood	-32810.964

Route Clusters									
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
6	38	395	211	246	178	371	223	166	270
227	41	168	203	43	160	23	26	12	17
115	121	86		76	18	34	27	20	
	79			94	180	39	174	33	
	118			138	74	184	97	65	
					78	129	137	70	
					113	330	269	88	
					201	60		165	
								99	

County Clusters				
1	**2**	**3**	**4**	**5**
SHA	STA	KER	INY	SIS
TEH	CC	RIV	PLU	SIE
DN	SM	IMP	LA	MOD
	SCL	SBD	SAC	
	MON	YOL	SUT	
	MER	TRI		
		BUT		
		COL		
		KIN		

Urban Signalized SPF

Total collision	
Variables	**Estimate**
(Intercept)	-3.545
Main ADT	0.313
Cross street ADT	0.102
Design: IM	-0.148
Design: IS	-0.301
Design: IT	-0.238
Design: IY	-0.447
Design: IZ	-0.173
Control: M	0.308
Control: N	0.350
Control: P	0.302
Cross street lanes amt	0.118
Main lanes amt	-0.028
Cross street flow: R	0.328
Cross street flow: Z	-0.599
Main flow: R	-0.435
Main flow: W	-0.287
Year_4	0.058
Route_cluster_1	1.007
Route_cluster_2	-2.411
Route_cluster_3	-0.867
Route_cluster_4	-1.597
Route_cluster_5	-0.241
Route_cluster_6	-0.596
Route_cluster_7	0.281
Route_cluster_8	1.784
Route_cluster_9	1.278
Route_cluster_10	0.566
County_cluster_1	0.885
County_cluster_2	0.312
County_cluster_3	-0.405
County_cluster_4	-3.738
County_cluster_5	-1.006
Theta(overdispersion)	2.751
AIC	46641.000
Log-likelihood	-23286.555

Route Clusters									
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
166	191	165	216	299	183	184	83	225	13
8	187	218	22	61	27	19	125	12	23
18	40	202	77	66	53	20	395	62	25
26	174	2	109	79	232	39		67	34
168		63	121		92	72		71	36
244		68	213		111	84		88	38
180		178			131	90		94	46
54		98			204	107		116	49
76		119				108			262
99		130				132			70
101		137				164			86
105		140							91
219									145
280									162
237									201
135									
138									

County Clusters				
1	**2**	**3**	**4**	**5**
MON	SLO	BUT	DN	SON
TUL	SCR	SJ		IMP
NAP	ORA	SD		ED
MER	CC			GLE
	LA			SBD
	KIN			FRE
	VEN			LAS
	RIV			YUB
	KER			SF
				HUM

Urban Un-signalized SPF

Total collision	
Variables	**Estimate**
Intercept	-6.358
Mainline ADT	0.466
Cross street ADT	0.166
Design: IM	-0.353
Design: IS	-0.195
Design: IT	-0.431
Design: IY	-0.433
Design: IZ	0.210
Mainlanes_amt	-0.028
Cross street lanes_amt	0.108
Cross street flow: P	0.300
Cross street flow: R	-0.654
Year_2	0.061
Year_3	0.101
Year_4	0.172
Year_5	0.076
Route_cluster_1	-1.280
Route_cluster_2	0.297
Route_cluster_3	0.659
Route_cluster_4	-2.732
Route_cluster_5	2.263
Route_cluster_6	0.943
Route_cluster_7	0.218
Route_cluster_8	0.469
Route_cluster_9	-0.791
Route_cluster_10	-0.455
County_cluster_1	0.750
County_cluster_2	-1.316
County_cluster_3	0.925
County_cluster_4	0.281
County_cluster_5	1.297
County_cluster_6	1.063
County_cluster_7	0.506
Theta(overdispersion)	1.6686
AIC	69273
Log-likelihood	-34602.49

Route Clusters									
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
2	39	4	210	199	395	18	46	28	20
3	94	17	245		262	43	54	32	23
22	107	19	243		60	62	67	269	27
216	120	237	193		71	108	219	63	29
213	126	59			86		90	66	53
65	129	83			88		101	201	68
114	132	115			133		118	89	78
121	142	119			135		138	92	111
130	155	154			166		164	95	282
233		156					200	137	218
187		173					180	267	232
191							184	178	128
							189		150
									174
									185

County Clusters						
1	**2**	**3**	**4**	**5**	**6**	**7**
RIV	IMP	TUO	KER	SIS	PLA	SOL
NEV	SF	SBT	ED	TUL	MAD	LA
CC	DN	KIN	SM		LAK	BUT
SUT		VEN	MRN			SLO
SCR		MON	STA			MER
		NAP	SB			ORA
						SON
						SD
						YOL
						SJ
						MEN
						SCL

3. Ramp SPF

Total Collision	
Variables	Estimate
(Intercept)	-9.846
Ramp ADT	0.604
Buttonhook Ramp	-0.951
Collector Road	-0.701
Diamond Type Ramp	-0.749
Direct or Semi-direct Connector(Left)	-0.917
Direct or Semi-direct Connector(Right)	-1.021
Loop-with Left Turn	-0.735
Loop-without Left Turn	-0.818
Other-Ramp	-0.934
Rest Area, Vista Point, Truck Scale	-1.108
Scissors	-1.239
Slip Ramp	-1.868
Split Ramp	-1.493
Two-way Ramp Segment	-0.902
ON Ramp	-0.406
OTH Ramp	-0.729
Urban	0.476
Year 1	-0.152
Year 2	-0.132
Year 3	-0.082
Route_cluster_1	4.173
Route_cluster_2	4.703
Route_cluster_3	3.637
Route_cluster_4	5.079
Route_cluster_5	4.521
Route_cluster_6	5.671
Route_cluster_7	5.267
Route_cluster_8	6.095
Route_cluster_9	4.863
Route_cluster_10	4.977
County_cluster_1	1.083
County_cluster_2	0.877
County_cluster_3	1.211
County_cluster_4	0.601
County_cluster_5	0.406
County_cluster_6	0.966
County_cluster_7	0.737
County_cluster_8	1.558
Theta (overdispersion)	1.6656
AIC	215835
Log-likelihood	-107870.67

Route Clusters									
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	2	13	14	17	79	4	244	5	8
20	29	40	37	73	205	22		7	10
24	49	68	41	90	275	50		12	59
33	52	82	57	92		65		15	60
47	58	261	67	118		78		23	91
56	85	330	80	120				44	94
75	87	980	105	132				51	99
84	101		113	135				54	110
86	178		125	138				55	134
103	237		170	190				70	160
108	280		180	204				71	168
126	380		605	238				163	198
133	505		710	299				210	215
152	680			580				217	242
154								395	405
156								780	805
241								880	
259									
905									

County Clusters							
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
SCR	MAD	SBD	CC	BUT	ALA	FRE	TUO
SON	NAP	VEN	MRN	COL	LA	KER	
	ORA		SD	HUM	RIV	MON	
	SB		SLO	KIN	SAC	SJ	
			SM	MEN	SBT		
			TEH	MER	SCL		
			TUL	NEV	SF		
			YOL	PLA	STA		
				SHA			
				SOL			
				SUT			

Appendix C – SPF Tool Version -1 Process Flowchart