VALIDATION OF MODELS FOR QUANTIFYING SAFETY PERFORMANCE OF DRIVEWAYS ON STATE HIGHWAYS

Final Report

Prepared for:

Oregon Department of Transportation
Salem, OR

Prepared by:

Texas A&M Transportation Institute
3135 TAMU
College Station, TX 77845

Primary Authors:

Karen Dixon, Ph.D., P.E.
Raul Avelar, Ph.D.

June 2014

Page Intentionally Blank

TABLE OF CONTENTS

1.0 INTRODUCTION 3

2.0 REVIEW OF AVAILABLE DATA 5
2.1 Crash Data 5
2.2 Driveway and Road Data 6
2.2.1 Driveway Data 6
2.2.2 Road Data 6
2.3 Summary 6

3.0 DATA COLLECTION 7
3.1 Sample Size and Associated Data Variables 7
3.2 Data Collection Methods 10
3.3 Final Validation Data Sets 10
3.4 Summary 13

4.0 VALIDATION ANALYSIS AND RESULTS 15
4.1 Overview of Validation Procedures 15
4.2 Urban Model Validation 16
4.2.1 Review of Initial Urban Model 16
4.2.2 Characteristics of the Urban Validation Data 18
4.2.3 Spatial Transferability of Urban Model 22
4.2.3.1 Direct Comparison of Predicted to Observed Crashes (Urban) 22
4.2.3.2 Expected Frequency Thresholds 23
4.2.4 Spatial-Temporal Transferability of Urban Model 26
4.2.5 Urban Model Coefficient Evaluation and Enhancement 27
4.2.5.1 Submodel Validation 27
4.2.5.2 Validation of Individual Predictor Coefficients 29
4.2.5.3 Validation of NB2 Theta Values 31
4.2.6 Enhanced Urban Model 31
4.3 Rural Model Validation 34
4.3.1 Review of Initial Rural Model 34
4.3.2 Characteristics of the Rural Validation Data 36
4.3.3 Spatial Transferability of Rural Model 39
4.3.3.1 Direct Comparison of Predicted to Observed Crashes (Rural) 39
4.3.3.2 Expected Frequency Thresholds 40
4.3.4 Spatial-Temporal Transferability of Rural Model 42
4.3.5 Rural Model Coefficient Evaluation and Enhancement 44
4.3.5.1 Submodel Validation 44
4.3.5.2 Validation of Individual Predictor Coefficients 45
4.3.5.3 Validation of NB2 Theta Values 47
4.3.6 Enhanced Rural Model 47
4.4 Summary of Results 50

5.0 CONCLUSIONS AND RECOMMENDATIONS **53**
5.1 Final Urban Model and Example Application 53
5.1.1 Use of the Urban Model 54
5.1.2 Example Use of the Urban Model 54
5.2 Final Rural Model and Example Application 56
5.2.1 Use of the Rural Model 56
5.2.2 Example Use of the Rural Model 59

6.0 REFERENCES **61**

7.0 APPENDIX **63**
7.1 Supplemental Tables 63
7.2 Process for Collecting Driveway Data using Google Earth 65
7.3 Spreadsheet User Instructions 73
7.4 EXample Driveway Cluster Calculations 79

List of Tables

Table 3.1: Data Variables Collected 9
Table 3.2: Field Data and Corresponding Collection Method 10
Table 3.3: Urban Validation Sites 11
Table 3.4: Rural Validation Sites 12
Table 4.1: 2012 Study Best Performing Urban Crash Prediction Model 16
Table 4.2: Table of Possible Cases (Effect of Roadway at Urban Environments) 18
Table 4.3: Summary Statistics for the Urban Validation Data 18
Table 4.4: Summary Statistics for the Original Urban Data 19
Table 4.5: Pearson Correlations for Urban Validation Data 20
Table 4.6: Pearson Correlations for Urban Original Data 21
Table 4.7: Multiplicative Factor Estimates for Urban Submodels in the Validation Data 28
Table 4.8: Urban Crash Prediction Model Estimated from the Validation Sample 29
Table 4.9: Summary Statistics for the Combined Urban Dataset (2004 to 2008) 31
Table 4.10: Enhanced Urban Crash Prediction Model 32
Table 4.11: Original Crash Prediction Model for Rural Environments 34
Table 4.12: Possible Cases of the Effect of Roadway at Rural Environments 35
Table 4.13: Summary Statistics for the Rural Validation Data 36
Table 4.14: Summary Statistics for the Original Rural Data 37
Table 4.15: Pearson Correlations for Rural Validation Data 37
Table 4.16: Pearson Correlations for Original Rural Data 38
Table 4.17: Multiplicative Factor Estimates for Rural Submodels in the Validation Data 44
Table 4.18: Rural Crash Prediction Model Estimated from the Validation Sample 46
Table 4.19: Summary Statistics for the Combined Rural Dataset (2004 to 2008) 47
Table 4.20: Enhanced Rural Crash Prediction Model 48
Table 5.1: Sample Input for Urban Example Problem from Redmond, Oregon 54
Table 5.2: Sample Input for Rural Example Problem for Corvallis-Newport, Oregon 59
Table 7.1: Abbreviations and Acronym Definitions 63
Table 7.2: Summary of Original Study Corridors 64

List of Figures

Figure 3.1: Urban Validation Sites 8
Figure 3.2: Rural Validation Sites 8
Figure 4.1: Original Model Predictions versus Observed Validation Crashes (Urban) 23
Figure 4.2: Expected and Observed Urban Crash Frequencies (Sites identified by their number) 24
Figure 4.3: Expected Site Frequencies for Predictions Below Ten Crashes (Sites #28 and #41 not shown) 25
Figure 4.4: Crash Distributions Associated with Time 26
Figure 4.5: Site Expected Frequencies for years 2004 through 2011 27
Figure 4.6: Statistical Deviations of Submodel Factors in the Urban Validation Data 28
Figure 4.7: Statistical Deviations of the Urban Model Coefficients 30
Figure 4.8: Fit of Enhanced Urban Model to Original and Validation Data 33
Figure 4.9: Original Model Predictions versus Observed Validation Crashes (Rural) 40
Figure 4.10: Expected and Observed Rural Crash Frequencies (Sites are identified by their assigned sample numbers) 41
Figure 4.11: Expected and Actual Crash Frequencies in Rural Dataset for predictions smaller than 10 crashes in 5 years (sites are identified by their assigned numbers) 42
Figure 4.12: Marginal Distributions of Crashes for 2004 through 2011 43
Figure 4.13: Site Expected Frequencies for years 2004 through 2011 43
Figure 4.14: Statistical Deviations of Submodel Factors in the Rural Validation Data 45
Figure 4.15: Confidence Intervals for the Statistical Deviations of the Rural Model Coefficients 46
Figure 4.16: Fit of Enhanced Rural Model to Original and Validation Data 50
Figure 5.1: Sample Site #23, Redmond, Oregon 55
Figure 5.2: Example Calculations for Rural Directional Driveway Clusters 58
Figure 5.3: Sample Site, Corvallis-Newport, Oregon 59
Figure 7.1: Driveways on only one side of road (Example 1) 79
Figure 7.2: Driveways on only one side of road (Example 2) 80
Figure 7.3: Driveways on only one side of road (Example 3) 80
Figure 7.4: Driveways on only one side of road (Example 4) 81
Figure 7.5: Driveways on only one side of road (Example 5) 81
Figure 7.6: Driveways on both sides of road (Example 1) 82
Figure 7.7: Driveways on both sides of road (Example 2) 83
Figure 7.8: Driveways on both sides of road (Example 3) 84
Figure 7.9: Driveways on both sides of road (Example 4) 85
Figure 7.10: Driveways on both sides of road (Example 5) 86
Figure 7.11: Driveways on both sides of road (Example 6) 87

EXECUTIVE SUMMARY

This *Final Report* documents the validation of urban and rural arterial safety performance functions (SPFs) developed for the Oregon Department of Transportation (ODOT) SPR 720 study titled *Quantifying Safety Performance of Driveways on State Highways.* For this previous effort, the research team determined that the crash reporting that indicated a driveway may have been involved in the crash was not a dependable variable, so they developed SPFs for all non-intersection-related arterial crashes (of which many were likely due to vehicle interactions at driveway locations). Due to a limited sample size in the original study, the Final Report for SPR 720 recommended sampling of additional study sites and validation of the original study models based on these new randomly sampled locations. The information in this report reviews the subsequent validation effort and the resulting recommendations.

In general the original models performed very well in response to the validation tests. The research team evaluated spatial transferability, spatial-temporal transferability, and individual coefficient stability and significance. The **urban model** performed well with the spatial transferability resulting in statistically equivalent values, the spatial-temporal transferability providing similar values but not statistically equivalent at the 95 percent level, and all but one of the model variables (titled "Other DW") determined to be statistically significant. The **rural model** also performed well as it was determined to provide statistically equivalent predictions for spatial transferability as well as for spatial-temporal transferability. In addition, the validation analysis for the individual coefficients found that only the "Four.Travel.Lanes" and the "Number.of.DW.Clusters" variables were not statistically equivalent at the 95 percent level. Ultimately, the research team developed enhanced models with the enriched data set so as to refine the original models and simplify their structure, where feasible.

Page Intentionally Blank

1.0 INTRODUCTION

The presence and distribution of driveways is known to affect the safety performance of roadways. Current literature quantifies the safety impact of different road elements by using Crash Modification Factors or Functions (CMFs). CMFs are part of a framework of currently accepted methodologies to assess safety, as in the Highway Safety Manual (HSM). Several existing CMFs account for the safety impact of driveways using driveway density as the primary measure. A recently completed project by these researchers outlines an analysis of the relationship between segment crashes and driveway characteristics in the State of Oregon. This study focused on both urban and rural principal arterial state highways (as designated by the Oregon highway classification system) and used crash data from the years 2004-2008, collected from a randomly selected statistical sample of ODOT's road inventories. The research team developed alternative models incorporating land use and spatial distribution of driveways, in place of the HSM CMF driveway density only approach.

At the conclusion of the previous study, the researchers included a recommendation in the final report that the models be validated to ensure they accurately represent Oregon arterial safety conditions. This report summarizes the resulting model validation study. Chapter 2.0 of this report presents a brief review of the available data. Chapter 3.0 next addresses the data collection effort including a list of new randomly selected study sites. Chapter 4.0 summarizes the validation process and findings. Chapter 5.0 then reviews the overall project findings and recommendations. Cited references are listed in Chapter 6.0. Finally, the Appendix of this report (Chapter 7.0) includes tables that detail the abbreviations and acronyms used in this report. Also in Chapter 7.0 is a summary of the study locations included in the original 2012 study as well as user instructions for the companion spreadsheet tool. This research project did not include a literature review since this effort is an extension of a previous project for which the project team completed an applicable literature review.

Page Intentionally Blank

2.0 REVIEW OF AVAILABLE DATA

The Oregon Department of Transportation (ODOT) maintains databases containing significant information related to state-maintained roadways across Oregon. Since the effort summarized in this report addresses the validation of models previously developed for ODOT, the research team used consistent procedures for data collection and analysis. The following summary briefly discusses the data sources available for use in this research effort.

2.1 CRASH DATA

In order to understand the crash trends in Oregon, all crashes should be associated with a specific location on the Oregon roadway network. The analysis in this study is based on the same year of state highway network data (2008) previously used in the development of the initial models. The type of information used to locate the observed crashes includes route names, numbers, functional classification, etc., for every hundredth of a mile point in the State of Oregon. The initial model development used crash data that extended from the year 2004 through 2008. This data included information directly related to each crash and all vehicles and participants involved. In addition to the highway and crash data, staff at ODOT and Portland State University provided the team with supplemental geographic information. This data was primarily in the form of geographic information system (GIS) layers and included boundaries for cities, counties, census block groups, zoning, urban growth areas and climatic regions, among others. The research team also acquired GIS layers including the locations of all schools, hospitals, and liquor license locations in the State of Oregon.

As previously indicated, the initial model development project crash data included crashes that occurred for the five year period from 2004 to 2008. For the validation effort, the project team has extended the crash analysis for the more recent 2004 to 2011 years. Including subsequent years in the analysis further enables the research team to determine if the models are valid based on differences in geographic locations as well as time.

A wide variety of site characteristics are included in the original and subsequent data sets. In addition to determining the reported crash location on the highway network (to the nearest hundredth of a mile point), the data collection effort also evaluated the crash and its orientation to schools, hospitals and liquor sales locations, and all available geographic information. This robust data then permits evaluation of site features including those identified as critical in the HSM as well as extensive driveway characteristic information (collected separately). Due to errors or omissions in the crash data and changes in the roadway network, approximately five percent of all crashes could not be tied to the state highway network.

In the initial study, the research team determined that isolating crashes that are known to be driveway-related (according to the crash report) resulted in questionable results. Many mid-block angle crashes, for example, were not identified in the original crash report or subsequent crash database as driveway-related. Consequently, the Technical Advisory Committee (TAC) for that project agreed that focusing on major (principal) arterial mid-block crashes would be appropriate and should help to compensate for crash reporting errors for the driveway-related indicator. The TAC also recommended that rural and urban highways be investigated separately. Therefore, the initial safety performance models in this report include mid-block crashes (segment) on principal arterials for urban and rural locations.

2.2 DRIVEWAY AND ROAD DATA

Prior to developing the data collection plan for the original research effort, the project team completed a thorough review of all driveway and roadway data available through ODOT and other sources. This information primarily represented driveway data and road data.

2.2.1 Driveway Data

Currently, ODOT does not maintain a comprehensive database that includes information relating to all driveways on the state highway network. As a result, the data collection effort required acquisition of this data through the use of aerial photography and roadway video information. Chapter 3.0 reviews the approach for the current data collection effort. This effort addresses the collection of the same type of driveway data as acquired for the initial research project.

2.2.2 Road Data

As indicated previously, roadway data is available for the state highway network system. This information includes route names, numbers, functional classification, and general physical road characteristics. Since maintenance of a road characteristic database requires significant effort, it is likely that some inaccurate data may be included in the database. As a result, during the driveway data collection effort, the project team verified all associated roadway information for the study sample so as to ensure accurate data is included in the analyses.

2.3 SUMMARY

The primary objective for the collection of crash, driveway, and road data as required for this validation effort was the need to acquire data consistent with the format used for the initial 2012 study. Since the original study crash data extended from 2004 to 2008, data for the same five year period as well as data from 2009 to 2011 are included in the validation study data. Specific randomly selected data collection sites are further reviewed in Chapter 3.0 of this report.

3.0 DATA COLLECTION

As indicated in Chapter 2.0 , the project team developed a recommended data collection plan for each study location. Since the focus of this effort is to validate the models created in the initial project, data collection techniques mirrored those performed in the previous study. The research team based the sample size calculation, however, on observations from the previous study.

The following sections describe the validation data sample for urban and rural locations and the associated data collection efforts that occurred as part of this validation project.

3.1 SAMPLE SIZE AND ASSOCIATED DATA VARIABLES

In the initial research effort, the project team identified all of the known Oregon principal arterials located on the state highway system. This summary of locations then formed the basis for site selection in the validation effort. In order to verify each coefficient of the original model (with sufficient statistical power), the research team developed a sample size that included a minimum of 61 additional rural sites. This number is based on the size and standard error of the least significant continuous variable in the initial rural safety prediction model (segment length with a coefficient of -0.2864, standard error = 0.1259, and $p = 0.0156$). Research team members used a similar approach to determine the minimum urban model sample size. Based on these sample size estimation techniques, 61 additional rural sites and approximately 80 additional urban sites provide sufficient information for this extended analysis.

At rural locations, a segment with similar traffic volume and lane width designations will often extend over a considerable length of road. This can result in varying driveway density distributions. For example, assume a one mile section of rural road had eight driveways. If the entire mile is used for analysis, this will effectively assume that the eight driveways are evenly spaced along the length of road. Assume that you then further inspect the location and note that all eight driveways are located within a 0.25 mile section of road. By using the entire one mile analysis length, you will overlook the short section of dense driveway activity. Optimally, this location should be further segmented into a 0.75 mile section of road with no driveways and a 0.25 mile of road with eight driveways. This further segmentation will help account for locations with more heterogeneous driveway distributions. Following the additional segmentation of the rural sites to minimize these variations in driveway location spacing, the final validation sample size includes 114 rural sites and 80 urban sites. Figure 3.1 (urban) and Figure 3.2 (rural) each graphically depicts the geographic locations of these additional validation sites. The specific data variables collected for the original model development and for this validation effort are summarized in Table 3.1.

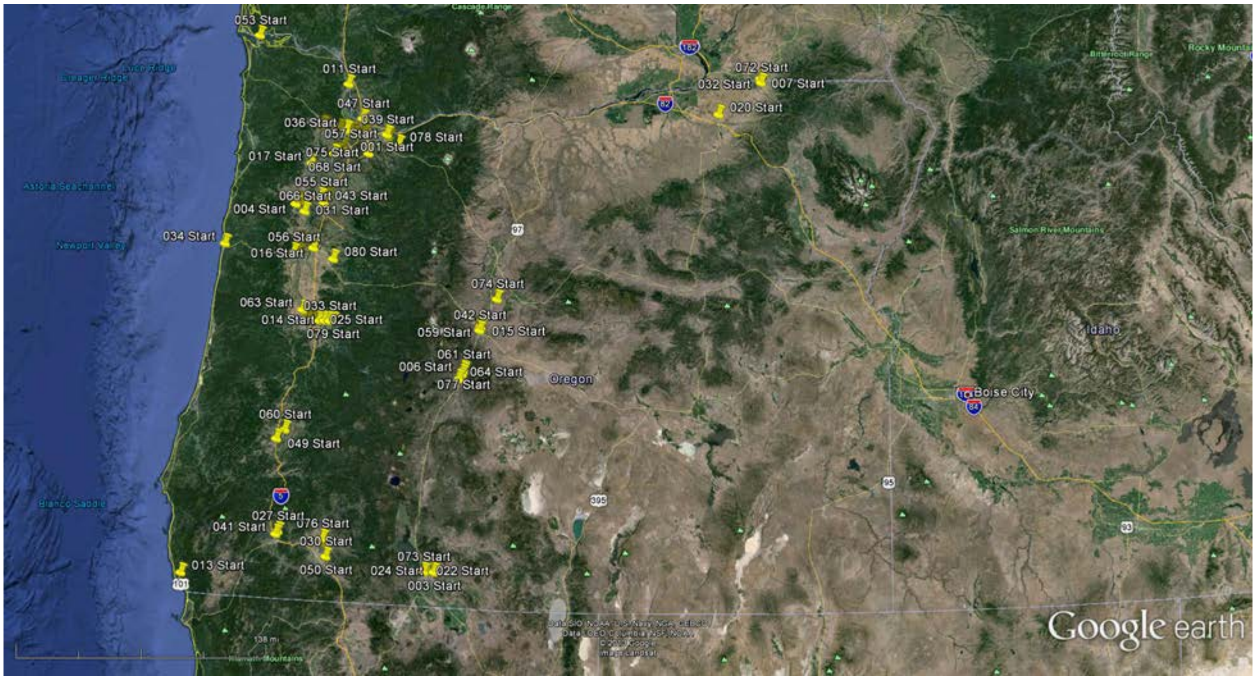


Figure 3.1: Urban Validation Sites

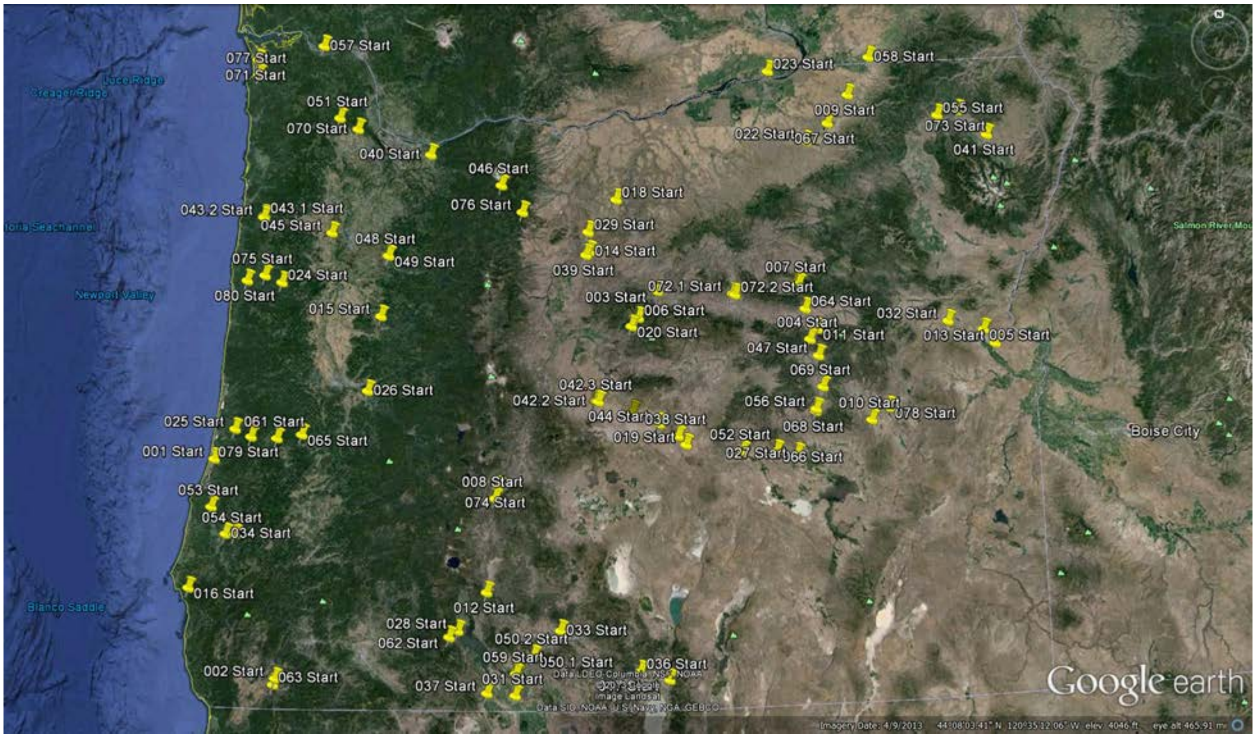


Figure 3.2: Rural Validation Sites

Table 3.1: Data Variables Collected

Variables Common to Roadway and Driveway		**Description**	
Highway name		Local name of highway	
Highway number		State highway number designation (may not correspond to state or federal route number)	
Roadway-Specific Data Items		**Driveway-Specific Data Items**	
Road width	To edge of outside travel lanes (ft) – edge stripe or lip of gutter	Road name	Name if available and different from highway name
Number of lanes	Total both directions	Rural/urban	Only one designation of either rural or urban can be assigned to a corridor
Median type	Raised, painted, two-way left-turn lane (TWLTL), grass, or none	Generic land use	• Residential • Commercial • Industrial • Institutional • Agricultural • Other (recreational, ports, etc. • Public road • Unknown
Speed limit	Posted value (mph)		
Average Annual Daily Traffic Volume (AADT)	Volume (vehicles per day)		
Bicycle lane	Present or Not Present		
Sidewalks	Present or Not Present		
On-street parking	Permitted or Prohibited		
Number of driveways in segment boundaries	Total both direction (identify side of road based on direction of travel)	Parking spots	Number of stalls visible from aerial inspection
		Dwelling units	Number of visible dwelling units
Segment length	Based on milepoints (miles)	Additional driveways into same facility	Number
Segment beginning break type	Intersection, change in cross-section, speed limit, urban or rural boundary, other	Driveway width	Driveway width at edge of outside travel lane (ft)
		Lanes	Number of lanes at driveway
Segment beginning additional information	Comment -- information pertaining to break type, old segment width, speed limit, other	Throat	Present or not present
		Throat width	Width if answer to "throat" was Present
Segment ending break type	Intersection, change in cross-section, speed limit, urban or rural boundary, other	Throat lanes	Number
Segment ending additional information	Comment -- information pertaining to break type, old segment width, speed limit, other	Distance from start of the study section	Measurement in feet

3.2 DATA COLLECTION METHODS

As indicated in the previous section, the project team acquired validation data for randomly selected principal arterial driveways on the Oregon state highway network. Due to the large geographic distribution of the candidate sites, it was not feasible for the project team to collect the data via site visits. The majority of the data collection effort, therefore, depended on the use of digital video logs and aerial photography via Google Earth. The data collection techniques for general data categories are shown in Table 3.2.

Table 3.2: Field Data and Corresponding Collection Method

Driveway Data to Collect	Collection Method
Traffic Volume (AADT)	ODOT Databases
Driveway Location	Google Earth
Driveway Width	Google Earth
Driveway Type (land use being served)	Google Earth/Video Log
Number of Lanes	Google Earth
Median Configuration	Google Earth/Video Log
Posted Speed	Google Earth/Video Log/ODOT Databases
Traffic Control	Google Earth/Video Log

These data collection methods are consistent with the approach used for the initial project. The actual technique used to acquire data from Google Earth is documented in Section 7.2.

3.3 FINAL VALIDATION DATA SETS

The original urban and rural data set from the 2012 study is included in Appendix 7.1. Table 3.3 and Table 3.4 depict the new urban and rural sites respectively that were randomly selected for the validation effort as described in this report. For locations where the randomly selected segment included a change in driveway density, the research team members further subdivided the segments, as previously described, so as to capture this site diversity. An example of one of these locations would be John Day-Burns (Highway number 48) where the segment extending from milepoint 18.31 to 20.31 was subdivided into three final subsections so as to maintain consistent driveway spacing along the study corridor.

Table 3.3: Urban Validation Sites

Hwy. Name	Hwy. No.	Begin MP	End MP	Hwy. Name	Hwy. No.	Begin MP	End MP
Mt. Hood	026	15.61	15.82	Redwood	025	0.98	1.11
Corvallis-Newport	033	52.81	53.12	The Dalles-California	004	0.93	1.16
Klamath Falls-Malin	050	5.20	5.31	Salem	072	6.31	6.66
Kings Valley	191	4.29	4.82	Pacific Highway East	081	4.54	4.64
Crater Lake	022	2.44	2.64	The Dalles-California	004	136.52	136.85
The Dalles-California	004	165.41	165.87	Clackamas-Boring	174	6.91	7.01
Oregon-Washington	008	31.31	31.64	Pacific Highway East	081	3.93	4.05
Pacific Highway West	091	125.52	125.65	Pacific Highway West	091	19.87	20.03
Pacific Highway East	081	45.66	45.80	Coos Bay-Roseburg	035	75.30	75.56
Pacific Highway West	091	16.02	16.21	Rogue Valley	063	11.50	11.67
Lower Columbia River	092	26.87	27.04	Albany-Junction City	058	4.32	4.81
Pacific Highway East	081	1.85	2.03	The Dalles-California	004	273.36	273.48
Oregon Coast	009	354.92	355.12	Nehalem	102	0.43	0.68
McKenzie	015	5.29	5.42	McKenzie-Bend	017	18.30	18.62
The Dalles-California	004	137.80	137.93	Salem-Dayton	150	18.37	18.50
Albany-Junction City	058	5.53	5.80	Corvallis-Newport	033	52.80	53.04
Pacific Highway West	091	36.39	36.67	Mt. Hood	026	16.35	16.67
Redwood	025	-1.54	-1.41	The Dalles-California	004	137.55	137.67
Mt. Hood	026	18.05	18.24	The Dalles-California	004	138.41	138.51
Pendleton	067	0.74	1.15	North Umpqua	138	1.96	2.17
Pacific Highway West	091	78.97	79.54	The Dalles-California	004	161.95	162.81
The Dalles-California	004	276.39	276.68	Pacific Highway West	091	122.85	122.96
Pacific Highway West	091	79.74	79.92	Pacific Highway West	091	116.36	116.64
Klamath Falls-Malin	050	4.56	4.73	The Dalles-California	004	167.98	168.49
McKenzie	015	2.32	2.48	Mt. Hood	026	0.35	0.94
Pacific Highway West	091	119.96	120.15	Pacific Highway West	091	62.55	62.79
Redwood	025	-1.29	-1.19	Beaverton-Hillsdale	040	2.22	2.33
Dallas-Rickreall	189	0.94	1.09	Pacific Highway West	091	22.26	22.36
Klamath Falls-Malin	050	-5.18	-5.05	Nehalem	102	90.47	90.59
Rogue Valley	063	11.19	11.67	Tualatin Valley	029	18.68	18.78
Pacific Highway	091	62.54	62.82	Tualatin Valley	029	18.32	18.44
Oregon-Washington	008	31.47	31.75	Oregon-Washington	008	26.59	26.81
McKenzie	015	7.97	8.12	Lake of the Woods	270	68.05	68.66
Oregon Coast	009	144.38	144.82	Ochoco	041	1.70	2.02
Dallas-Rickreall	189	1.60	1.70	Cascade Hwy South	160	3.92	4.02
Tualatin Valley	029	4.56	4.70	Crater Lake	022	4.73	5.20
Pacific Highway West	091	6.63	6.88	The Dalles-California	004	164.26	164.40
Pacific Highway West	091	82.70	82.87	Mt. Hood	026	24.47	24.57
Mt. Hood	026	16.35	16.67	McKenzie	015	9.52	9.78
Tualatin Valley	029	8.03	8.13	Santiam	016	15.32	15.46
			Total Length of Urban Segments:	**19.10 miles**			

Table 3.4: Rural Validation Sites

Hwy. Name	Hwy. No.	Begin MP	End MP	Hwy. Name	Hwy. No.	Begin MP	End MP
Oregon Coast	009	228.42	228.84	Williemette	018	7.54	7.82
Redwood	025	31.51	31.80	Central Oregon	007	121.62	121.89
Ochoco	041	54.39	54.66	Lake of the Woods	270	44.56	45.08
John Day-Burns	048	18.31	18.85	Lake of the Woods	270	45.08	45.48
John Day-Burns	048	18.85	19.83	Lake of the Woods	270	45.48	45.87
John Day-Burns	048	19.83	20.31	Sherman	042	63.97	64.82
John Day	005	255.93	256.15	Sherman	042	64.82	65.53
John Day	005	256.15	256.65	Sherman	042	65.53	65.97
John Day	005	256.65	257.10	Klamath Falls-Lakeview	020	12.61	13.81
John Day	005	257.10	257.94	Klamath Falls-Lakeview	020	13.81	14.47
Ochoco	041	37.66	38.27	Klamath Falls-Lakeview	020	14.47	14.57
Ochoco	041	38.27	39.06	Klamath Falls-Malin	050	12.40	13.62
Ochoco	041	39.06	39.15	John Day	005	238.07	239.03
Pendleton-John Day	028	104.95	105.42	John Day	005	239.03	240.03
The Dalles-California	004	197.32	197.50	Klamath Falls-Lakeview	020	43.01	43.40
The Dalles-California	004	197.50	198.01	Coos Bay-Roseburg	035	23.76	24.21
The Dalles-California	004	198.01	199.32	Pendleton-John Day	028	103.54	103.95
Oregon-Washington	008	9.47	10.65	Klamath Falls-Lakeview	020	85.06	85.67
Oregon-Washington	008	10.65	11.41	The Dalles-California	004	285.69	286.57
Central Oregon	007	161.76	162.34	The Dalles-California	004	286.57	286.59
Central Oregon	007	162.34	162.79	The Dalles-California	004	286.59	287.12
John Day-Burns	048	9.71	10.13	The Dalles-California	004	287.12	287.34
John Day-Burns	048	10.13	11.13	Central Oregon	007	66.32	67.20
The Dalles-California	004	237.86	238.43	Central Oregon	007	67.20	68.32
The Dalles-California	004	238.43	239.18	The Dalles-California	004	75.85	76.78
John Day	005	263.96	265.25	Mt. Hood	026	20.41	20.93
The Dalles-California	004	75.28	75.62	Wallowa Lake	010	60.07	60.69
Santiam	016	25.05	25.40	Central Oregon	007	26.85	27.56
Oregon Coast	009	307.33	307.66	Central Oregon	007	27.56	28.30
Fremont	019	148.36	148.65	Central Oregon	007	28.30	28.76
Sherman	042	45.61	46.60	Salmon River	039	8.24	9.04
Sherman	042	46.60	47.04	Salmon River	039	9.04	9.60
Sherman	042	47.04	47.23	Central Oregon	007	56.71	58.71
Central Oregon	007	71.40	72.79	Willamina-Salem	030	17.21	17.87
Central Oregon	007	72.79	73.39	Warm Springs	053	62.08	62.34
Ochoco	041	32.46	32.76	John Day-Burns	048	26.83	27.93
Ochoco	041	32.76	33.75	North Santiam	162	17.96	18.23
Ochoco	041	33.75	34.11	North Santiam	162	18.30	18.99
Central Oregon	007	43.92	45.92	Klamath Falls-Lakeview	020	23.77	24.09
Pendleton-John Day	028	23.97	24.38	Klamath Falls-Lakeview	020	24.09	24.41
Umatilla-Stanfield	054	0.70	0.98	Nehalem	102	48.41	48.74
Corvallis-Newport	033	31.23	31.53	Central Oregon	007	95.74	97.74
Umpqua	045	3.77	4.08	Coos Bay-Roseburg	035	7.76	8.69
Coos Bay-Roseburg	035	29.26	30.04	Lower Columbia River	092	65.50	66.36

Hwy. Name	Hwy. No.	Begin MP	End MP	Hwy. Name	Hwy. No.	Begin MP	End MP
Wallowa Lake	010	26.66	27.19	Klamath Falls-Lakeview	020	10.36	11.01
John Day-Burns	048	53.67	54.36	Klamath Falls-Lakeview	020	15.83	16.41
Oregon-Washington	008	33.53	33.90	Oregon Coast	009	18.92	19.17
Umpqua	045	30.51	30.92	Ochoco	041	95.14	96.12
Lake of the Woods	270	38.78	40.39	Ochoco	041	96.12	96.71
Redwood	025	34.83	35.15	Wallowa Lake	010	41.83	42.13
John Day	005	156.79	157.17	The Dalles-California	004	192.17	193.05
Umpqua	045	43.04	43.82	Corvallis-Newport	033	21.05	21.47
Central Oregon	007	112.45	114.05	Warm Springs	053	78.24	79.81
Pendleton-John Day	028	8.83	9.42	Sunset	047	1.40	1.93
John Day-Burns	048	54.61	55.37	Central Oregon	007	172.40	172.89
John Day-Burns	048	42.66	42.94	Umpqua	045	11.67	12.54
Sunset	047	57.87	58.28	Corvallis-Newport	033	4.44	4.91
Total Length of Rural Segments:				**73.50 miles**			

3.4 SUMMARY

As demonstrated in this chapter, a random selection of urban and rural validation sites met sample size requirements that were based on the original study safety performance model and associated significant variables. Data collection techniques included a variety of remote data collection methods that ultimately resulted in a robust data set of 80 urban segments collectively extending 19.10 miles and 114 rural segments that resulted in a combined length of 73.50 miles.

Page Intentionally Blank

4.0 VALIDATION ANALYSIS AND RESULTS

This chapter summarizes the data analysis steps performed by the research team to validate and ultimately enhance the previously proposed prediction models for urban and rural arterial environments (see SPR 720 report titled “Quantifying Safety Performance of Driveways on State Highways”). Section 4.1 provides an overview of the validation procedures. Sections 4.2 and 4.3 respectively address the corresponding urban and rural efforts.

Within each of the urban and rural model validation sections is a brief overview of the safety models developed in the 2012 study. These sections are then followed by the associated model assessment and refinement validation steps. Finally, Section 4.4 presents a summary of results associated with the urban and rural validation efforts.

4.1 OVERVIEW OF VALIDATION PROCEDURES

Several candidate validation tests are available for the assessment of a statistical model. In general, these tests evaluate specific features of the models contrasted against an independent dataset. For this project, the research team was tasked with validating previously developed safety performance functions (SPFs). The research team concentrated the analysis on three aspects of the validation effort:

- Model spatial transferability: reviews the validity of model results for the same years in the original analysis (2004 through 2008) but at a new set of sites;
- Model spatial-temporal transferability: verifies the predicting power of the model for a different time period (2009 through 2011) at the new set of sites; and
- Equivalent model coefficients: analyzes and verifies the accuracy of original model coefficient values.

The first two analysis types can be assessed with a direct comparison of the model crash predictions to the new independent validation sample of sites and their associated crash data. The transferability of the model is then satisfactorily validated if the differences between the observed and predicted crashes do not exceed the theoretical thresholds imposed by the model.

For the final test of equivalent coefficients, the model values obtained in the original study are compared to the corresponding estimates obtained from the new independent data set. This analysis is the basis to determine the need to adjust any of the required coefficients during the final phase of model evaluation. In most cases, a larger data set will result in more reliable coefficients so some refinement of the initial models is usually expected during this final validation step if the original data and the new independent data are both applied to the model development task. This analysis step will also help to determine the reliability of the original model variables when applied to additional locations.

4.2 URBAN MODEL VALIDATION

The validation effort for the urban model is presented in this section of the report. First, the initial model is briefly reviewed. The urban validation data set is then assessed and incorporated into the individual model validation steps. The section concludes with a review of an enhanced urban model.

4.2.1 Review of Initial Urban Model

In 2012, members of the research team developed models to use for quantifying safety performance at urban arterial locations. This research effort, ODOT SPR 720, resulted in the base model shown in Equation 1.

Equation 1. General Estimation of Predicted Number of Crashes (1)

Predicted Number of Crashes = (Baseline Exposure Values) x *(Effect from Roadway)* x *(Effect from Roadside / Driveways)*

For urban arterial environments, the best performing crash prediction model determined in the 2012 study is depicted in Table 4.1.

Table 4.1: 2012 Study Best Performing Urban Crash Prediction Model

Term	Estimate	Standard Error	z-value	p-value	Significance[1]
(Intercept)	-12.891	2.380	-5.417	6.06E-08	***
LnAADT	1.686	0.253	6.670	2.57E-11	***
LnSegmentLength	0.358	0.159	2.244	0.0248	***
Speed.Lim.over.35	-0.469	0.215	-2.178	0.0294	*
MedianTWLTL	-0.898	0.339	-2.652	0.0080	**
Four.Travel.Lanes	-1.631	0.376	-4.335	1.46E-05	***
MedianTWLTL:Four.Travel.Lanes	1.098	0.445	2.465	0.0137	*
Com.andInd.DW	0.058	0.021	2.808	0.0050	**
Other.DW	-0.131	0.033	-3.972	7.14E-05	***

[1]*Significance values are as follows:*
º p<0.1; * p < 0.05; ** p < 0.01; and *** p < 0.001

NB2 Theta: 6.43 (Standard Error: 3.59), AIC: 209.66

The three key components depicted in Equation 1 are further defined by Equations 2, 3, and 4 as follows:

Equation 2. Baseline from Exposure at Urban Environments (2)

$$Baseline\ Exposure\ Values = (2.521 \times 10^{-6}) \times (AADT^{1.686}) \times (Segment\ Length^{0.358})$$

Where:

AADT = Annual Average Daily Traffic (vehicles per day), and
Segment Length = study corridor length (miles).

Note: The coefficient 2.521×10^{-6} is calculated as $e^{-12.891}$ (with the -12.891 value associated with the intercept in Table 4.1).

Equation 3. Effect of Roadway on Crashes at Urban Environments (3)

$$Effect\ from\ Roadway = exp\ [1.098 \times MedianTWLTL{:}Four.Travel.Lanes - (0.898 \times MedianTWLTL) - (1.631 \times Four.Travel.Lanes) - (0.469 \times Speed.Limit.over.35)]$$

Where:

MedianTWLTL = 1 if a two-way left-turn lane is present (0 value if not),
Four.Travel.Lanes = 1 if segment has 4 through lanes (2 lanes in each direction) or a value of zero if the segment has only 2 lanes (1 lane in each direction)
Speed.Limit.over.35 = 1 if the speed limit is greater than 35 mph and zero if the speed limit is 35 mph or less.

Equation 4. Effect of Roadside Elements on Crashes at Urban Environments (4)

$$Effect\ from\ Roadside\ \&\ Driveways = exp\ [0.058 \times (Com.and.Ind.DW - 2.259 \times Other.DW)]$$

Where:

Com.and.Ind.DW = number of commercial plus industrial driveways
Other.DW = number of driveways that are not commercial or industrial (Note: *Com.and.Ind.DW* + *Other.DW* = Total Driveways).

Note: The coefficient for Other.DW is derived as (-0.131 / 0.058 = -2.259) where -0.131 is the estimate for the Other.DW and 0.058 is the estimate for the Com.andInd.DW as shown in Table 4.1.

Equation 3 appears to be quite complex as it includes a large number of variables and input factors associated with the equation; however, these values can be easily determined and incorporated. Table 4.2 directly summarizes the results of Equation 3 for the various input values. A user can use this table in lieu of Equation 3 as a simplified way to determine input values representing the site features.

Table 4.2: Table of Possible Cases (Effect of Roadway at Urban Environments)

Median Type / Number of Lanes	Case 1: Speed Limits up to 35 mph		Case 2: Speed Limits above 35 mph	
	Two Travel Lanes	*Four Travel Lanes*	*Two Travel Lanes*	*Four Travel Lanes*
TWLTL Median	0.4074	0.2391	0.2549	0.1496
No Median or Other Median Types	1.0000	0.1957	0.6256	0.1225

The initial urban modeling effort included 40 urban locations with corridor lengths that ranged from 0.1 miles up to just over 1.0 miles in length. Since the original project budget limited the number of sites that could be studied, the urban validation effort, reviewed in the following sections, expanded the available data. As part of this validation, the research team first assessed the quality of the original models and then determined whether a refinement of the models might be appropriate. This effort is summarized in the following sections.

4.2.2 Characteristics of the Urban Validation Data

Initially, the project team collected urban segment data for 76 additional sites. Eight of the sites were removed because their cross-sections had three or six lanes, a situation that was excluded in the original study due to available sites. Team members removed one additional site because it was located on a bridge over the Willamette River in Portland. Data collection activities included acquiring associated 2004 through 2011 crash and exposure information for the sites. Where traffic volumes were not available, the research team estimated AADT based on the traffic volume trends observed for years 2005 through 2007. Table 4.3 summarizes the validation data characteristics for the urban sites.

Table 4.3: Summary Statistics for the Urban Validation Data

Variable	**Mean**	**Standard Deviation**	**Min**	**Max**	**Total**
Non-intersection Crashes	6.31	9.40	0	63	448
AADT (veh/day)	18,313	11,371	3045	54,122	--
Segment Length (mi)	0.23	0.15	0.10	0.86	16.2
Speed Limit (mph)	44.79	8.64	25	55	--
Number of Lanes	3.10	0.97	2	4	--
Number of Driveways	4.80	5.65	0	24	341
Number of Commercial driveways	2.24	3.70	0	15	159
Number of Industrial Driveways	0.27	1.04	0	6	19
Number of Other types of Driveways	2.30	3.94	0	24	163

The crash information included data that extended from years 2004 through 2011. A total of 448 non-intersection crashes occurred at the validation sites. Table 4.3 demonstrates a variety of urban data road features. The combined segment lengths collectively extended over 16.2 miles. There were 341 total driveways and roughly 47 percent of them were associated with commercial land use.

With the exception of the AADT ranges and the average number of industrial and commercial driveways, the summary statistics of the validation dataset are very similar to those of the original dataset (shown in Table 4.4). This is expected since both sets were collected using probability sampling, thus both should be representative of the prevailing conditions in Oregon.

Table 4.4: Summary Statistics for the Original Urban Data

Variable	Mean	Standard Deviation	Min	Max	Total
Non-intersection Crashes	7.59	8.46	0	41	296
AADT (veh/day)	18,061	10,187	1520	36,900	--
Segment Length (mi)	0.31	0.26	0.10	1.25	12.15
Speed Limit (mph)	44.49	8.72	25	55	--
Number of Lanes	3.62	1.31	2	5	--
Number of Driveways	5.72	5.68	0	23	223
Number of Commercial driveways	0.38	1.31	0	7	15
Number of Industrial Driveways	3.23	4.16	0	14	126
Number of Other types of Driveways	2.10	3.69	0	16	82

The upper and lower AADT thresholds were higher for the validation data than for the corresponding initial data set. This difference made model validation a unique challenge (as the data needs to be similar), but it also resulted in a richer database with a wider range of traffic volumes that could ultimately strengthen an enhanced model.

The composition of driveways is also somewhat different for the two samples (original and validation). These differences could indicate that larger sample sizes are needed for the complex urban environment so as to truly identify a representative probability sample of the entire population. Re-estimating the urban model, based on a pooled sample, is one promising way to deal with this issue.

The evaluation of Pearson Correlation values is a useful way to identify candidate variables that are influenced by other features. These values are often shown in a matrix as depicted in Table 4.5. Each cell in the matrix indicates how strongly one feature may be correlated to another. An exact correlation would have a value of 1.0. As shown in Table 4.5 (validation data) and Table 4.6 (original data), a row and a column with the same name would then have a value of 1.0 (as an item would be 100 percent correlated to that same variable). As a general rule, Pearson correlation values of 0.5 or greater indicate that features are strongly associated and, if retained in the model, must be uniquely addressed using an additional interaction term or a similar approach.

In some cases, an observed correlation is directly associated with roadway environment characteristics. For example, an increase in the traffic volume (represented by the AADT variable) can be expected to eventually be associated with a larger number of lanes (correlation values of 0.44 for the validation data and 0.53 for the original data support this expectation). This correlation, however, is not surprising since traffic volume is one of the key variables used to identify a need for constructing additional travel lanes.

Table 4.5: Pearson Correlations for Urban Validation Data

	Crashes	AADT	Segment Length	Speed Limit	Number of Lanes	Total Number of Driveways	Commercial Driveways	Industrial Driveways	Other Driveway Types
Crashes	1.00								
AADT	0.12	1.00							
Segment Length	-0.01	-0.30	1.00						
Speed Limit	-0.16	-0.13	0.30	1.00					
Number of Lanes	-0.03	0.44	-0.20	0.06	1.00				
Total Number of Driveways	0.31	-0.20	0.21	-0.49	-0.17	1.00			
Commercial Driveways	0.20	-0.02	-0.13	-0.52	0.05	0.70	1.00		
Industrial Driveways	0.07	0.01	0.25	-0.07	0.03	0.33	0.28	1.00	
Other Driveway Types	0.24	-0.28	0.36	-0.20	-0.29	0.69	0.00	-0.05	1.00

Table 4.6: Pearson Correlations for Urban Original Data

	Crashes	AADT	Segment Length	Speed Limit	Number of Lanes	Total Number of Driveways	Commercial Driveways	Industrial Driveways	Other Driveway Types
Crashes	1.00								
AADT	0.47	1.00							
Segment Length	0.10	-0.01	1.00						
Speed Limit	-0.31	-0.22	0.40	1.00					
Number of Lanes	0.10	0.53	-0.28	-0.28	1.00				
Total Number of Driveways	0.18	0.13	0.15	-0.20	0.12	1.00			
Commercial Driveways	0.27	0.36	0.26	0.04	0.20	0.30	1.00		
Industrial Driveways	0.37	0.13	-0.05	-0.44	0.21	0.73	0.13	1.00	
Other Driveway Types	-0.24	-0.07	0.20	0.17	-0.13	0.61	-0.05	-0.05	1.00

The Pearson's correlation values can also indicate relationship trends based on a positive or negative value. For example, there is a moderate and not surprising negative correlation between driveway variables and speed limit: roads with higher speed limits generally have fewer driveways. This relationship is clear by noting a negative sign for all driveway types shown in the speed limit column (see Table 4.5). This small negative correlation is also apparent for three of the five driveway types in the original urban data (see Table 4.6).

There are a few noticeable differences between the correlations associated with the original and the validation data. These are summarized as follows:

1. In the original data set, the speed limit was moderately associated with industrial driveways (correlation of -0.44), but this moderate negative association is greatly reduced in the validation data set (correlation of -0.07).

2. The linear association between speed limit and the total number of driveways (commercial driveways in particular) changed from -0.20 and +0.04 for the original data set to -0.49 and -0.52 for the validation data.

3. The correlation between speed limit and "other" driveway types roughly maintained its magnitude but the sign changed from +0.17 to -0.20.

4. The total number of driveways in the validation data set was strongly associated with commercial and "other" driveway types (correlations of +0.70 and +0.69 respectively), a clear shift from the original dataset where the total number of driveways was mostly comprised of industrial and "other" driveway types (correlations of +0.73 and +0.61 respectively).

The research team then proceeded to perform the various validation analyses that were introduced in Section 4.1. These assessments are reviewed in the following sections.

4.2.3 Spatial Transferability of Urban Model

The urban model spatial transferability analysis contrasted crash predictions obtained using the original urban model to the corresponding observed crashes at the validation sites. For this assessment, sites with AADTs larger than 37,000 vehicles per day were excluded as they exceeded the boundary conditions of the original model.

4.2.3.1 Direct Comparison of Predicted to Observed Crashes (Urban)

The researchers compared the distribution of validation site crashes from 2004 to 2008 to the theoretical distribution (prediction) obtained using the original urban model. Figure 4.1 shows these observed versus predicted crash distributions for the validation site locations. Generally, the original model tends to under predict crashes for the lower crash counts (i.e. more of the validation sites had zero, one, or two crashes than predicted by the model). Consequently, the model over predicts sites with large crash counts (sites with more than twelve crashes). A visual inspection and comparison, therefore, identifies these distinct differences. To assess if these differences are statistically significant at a 95 percent level, however, the p-value for the goodness of fit should have a value less than 0.05. The goodness of fit p-value for the urban original versus validation data has a value of 0.0828. This means that the crash frequencies predicted using the original models **are not statistically different** than the observed crash frequencies at the validation sites.

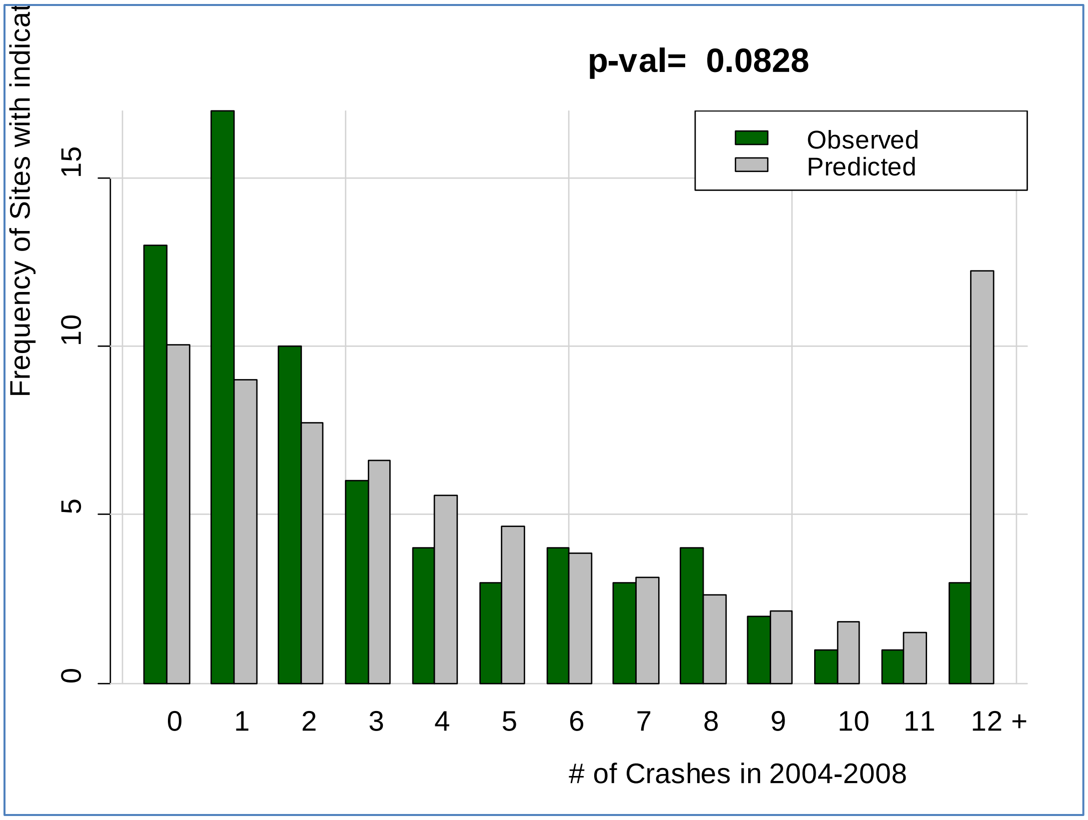


Figure 4.1: Original Model Predictions versus Observed Validation Crashes (Urban)

4.2.3.2 Expected Frequency Thresholds

As observed in Section 4.2.3.1, the number of crashes predicted contrasted to the observed crashes at the validation sites exhibit somewhat different but not statistically different frequency values. As the number of crashes increases, it is reasonable to expect the variability (or spread) of the crash frequency to also increase. One way of further assessing goodness of fit is to develop a comparison of crashes with an increasingly larger variation at higher crash frequencies. The research team constructed the plot shown in Figure 4.2 to graphically compare observed crashes versus predicted crashes while also identifying the region that should contain a predetermined percentage of the sites in the validation sample. For a predetermined percentage of 98 percent (i.e. virtually the entire sample), this region is referred to as the 98 percent *Expected Frequency Zone* in Figure 4.2. There are several of the original model values that do not occur within this 98 percent region. In fact, the actual percentage of sites with values located in the *Expected Frequency Zone* is 86 percent, clearly fewer sites than the 98 percent predicted by the original model.

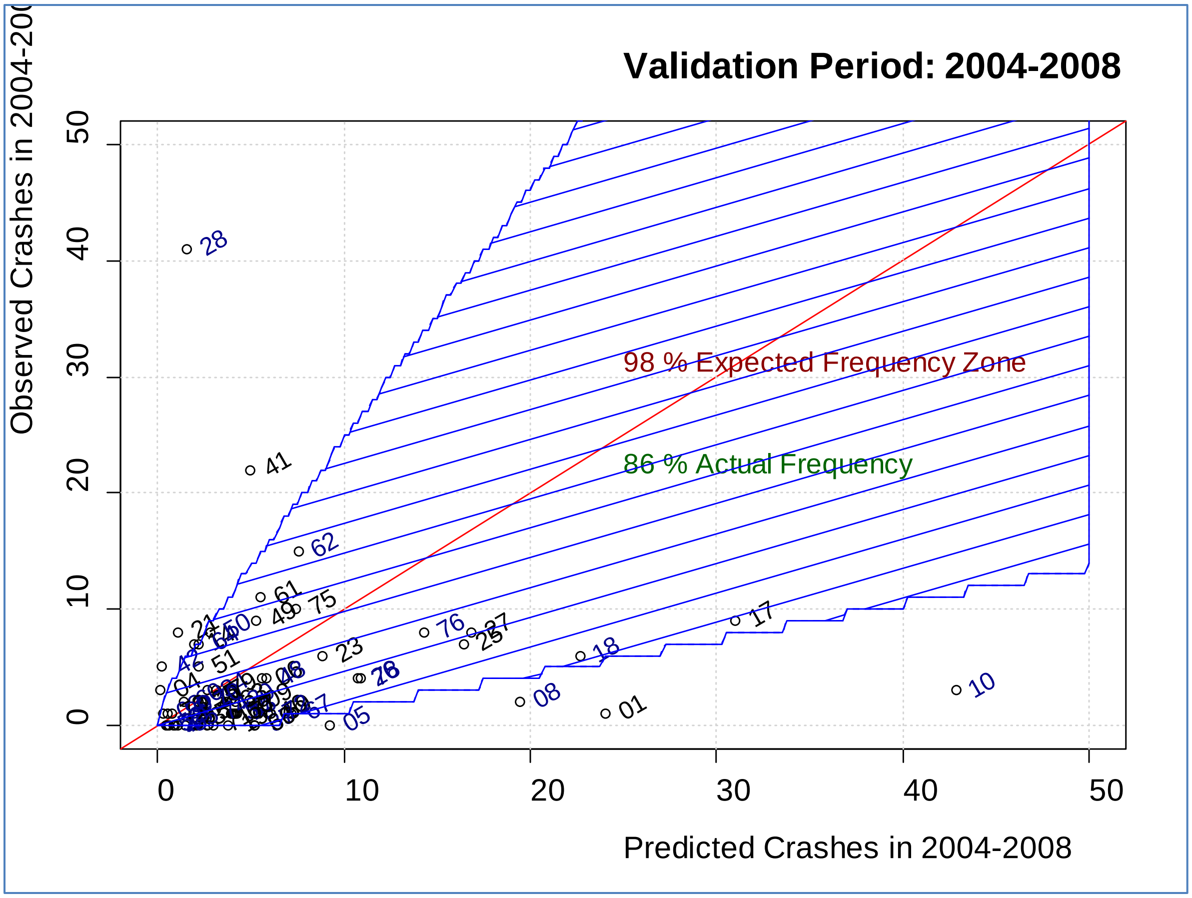


Figure 4.2: Expected and Observed Urban Crash Frequencies (Sites identified by their number)

In Figure 4.2, the diagonal line represents the point at which the observed number of crashes is equal to the predicted number of crashes from the original urban model. The 98 percent *Expected Frequency Zone* is then symmetrically computed about this diagonal line. Ideally, all site data should be located in the shaded regions with about the same number of points located above the line as below the line. It appears that crash frequencies greater than ten crashes for a five year period are not symmetrically distributed and are more likely to be located beyond the *Expected Frequency Zone* limits.

Site 28 clearly shows the largest vertical distance to the boundary of the *Expected Frequency Zone*. There were 41 observed crashes at this site, but the model predicts only 1.56 crashes for a five year period. Upon closer inspection, this site seems to have attributes better addressed by the rural model: it is located at the edge of Dallas where conditions are suburban and closely resembling rural. The speed limit at this site is 45 mph and the driveways are primarily residential driveways clustered together. Of the 14 driveways at this site, 12 appear to be residential and so would be classified as “other” (not commercial or industrial) by the urban model. By contrast, Table 4.4 shows that roughly three out of five of the driveways tended to be commercial or industrial for the original urban study sites. As noted in Section 4.2.2, the differences in driveway types are not unexpected when assessing model spatial transferability to the validation site locations.

Since the inclusion of Site 28 does not change the general trend of the data, the researchers elected to retain this outlier site in the validation data set. Even when including this site, the analysis indicates that the predicting power of the urban model is generally suitable. The predictive model, however, does tend to deviate from the observed crash frequencies more often than expected (i.e. the model prediction deviates from observed crash frequencies for 14 percent of the validation sites as opposed to the anticipated two percent).

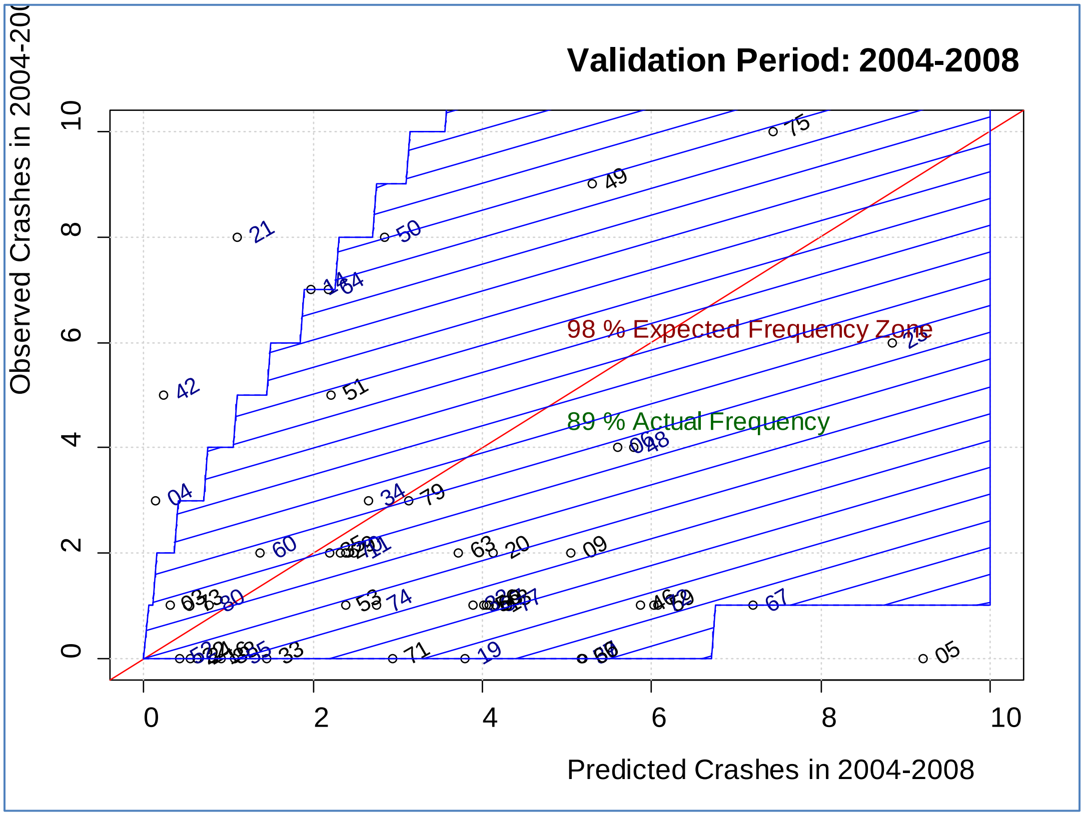


Figure 4.3: Expected Site Frequencies for Predictions Below Ten Crashes (Sites #28 and #41 not shown)

Since the trend to over predict crashes appears to be associated with sites with crash predictions greater than ten crashes for five years, the lower predicted site frequencies can be more closely considered for the range of fewer than ten predicted crashes (see Figure 4.3 – this figure is an enlargement of the Figure 4.2 lower left corner). The model seems to have more balanced predictions in this range. Upon this more detailed inspection, it is clear that the percentage of sites outside the *Expected Frequency Zone* is also larger than it should be (11 percent compared to the expected two percent).

4.2.4 Spatial-Temporal Transferability of Urban Model

The spatial analysis presented in Section 4.2.3 can be extended to the two time-based options shown in the following scenarios:

1. Evaluation for crashes that only occurred from 2009 to 2011, and

2. Evaluation of all crash data from 2004 through 2011.

Figure 4.4 shows the distributions corresponding to these two scenarios.

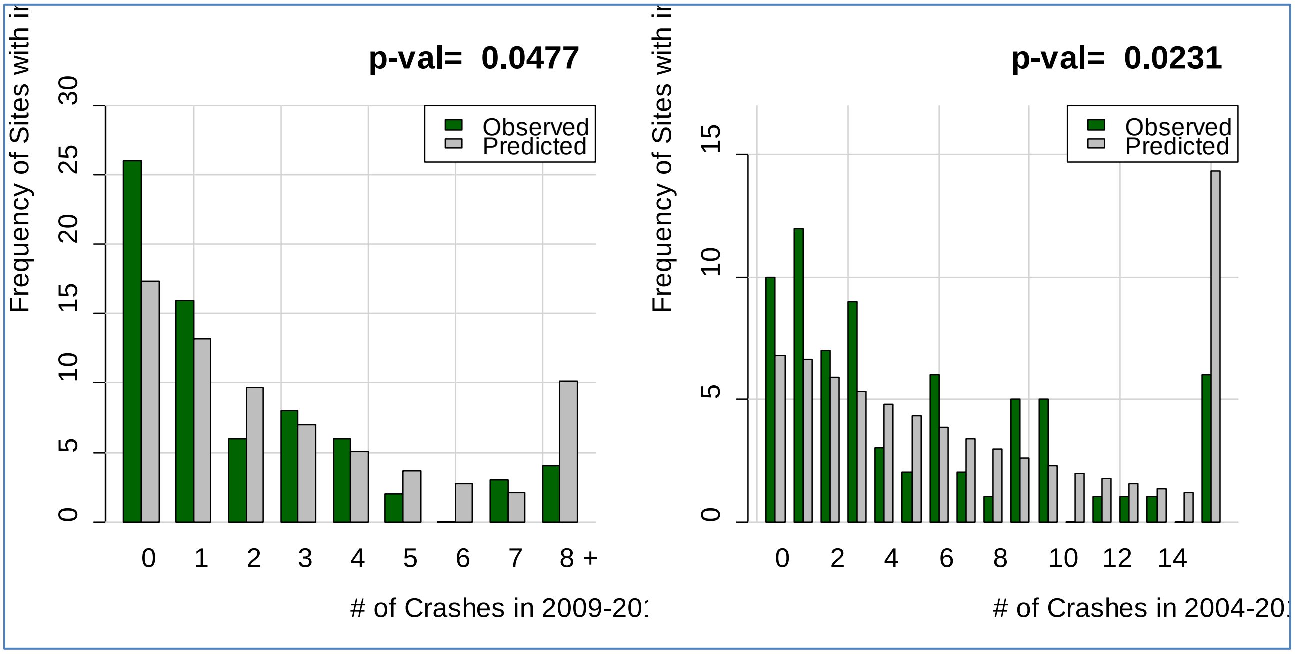


Figure 4.4: Crash Distributions Associated with Time

The model predictive power for the years 2009 through 2011 is similar to the original period of study but with a p-value for the fit that is borderline significant for years 2009-2011 and clearly significant for years 2004-2011. The research team then concluded that time based (or temporal) transferability for the original urban model appears to be limited, yet the model prediction trends appear to be consistent with the observed crashes.

Next, the research teams performed an expected frequency comparison for the two scenarios as shown in Figure 4.5. The patterns observed in the previous section remain: the model agrees with the actual crash frequencies for 86 percent of the sites during the years 2009 through 2011, but only for 80 percent for the period 2004 through 2011.

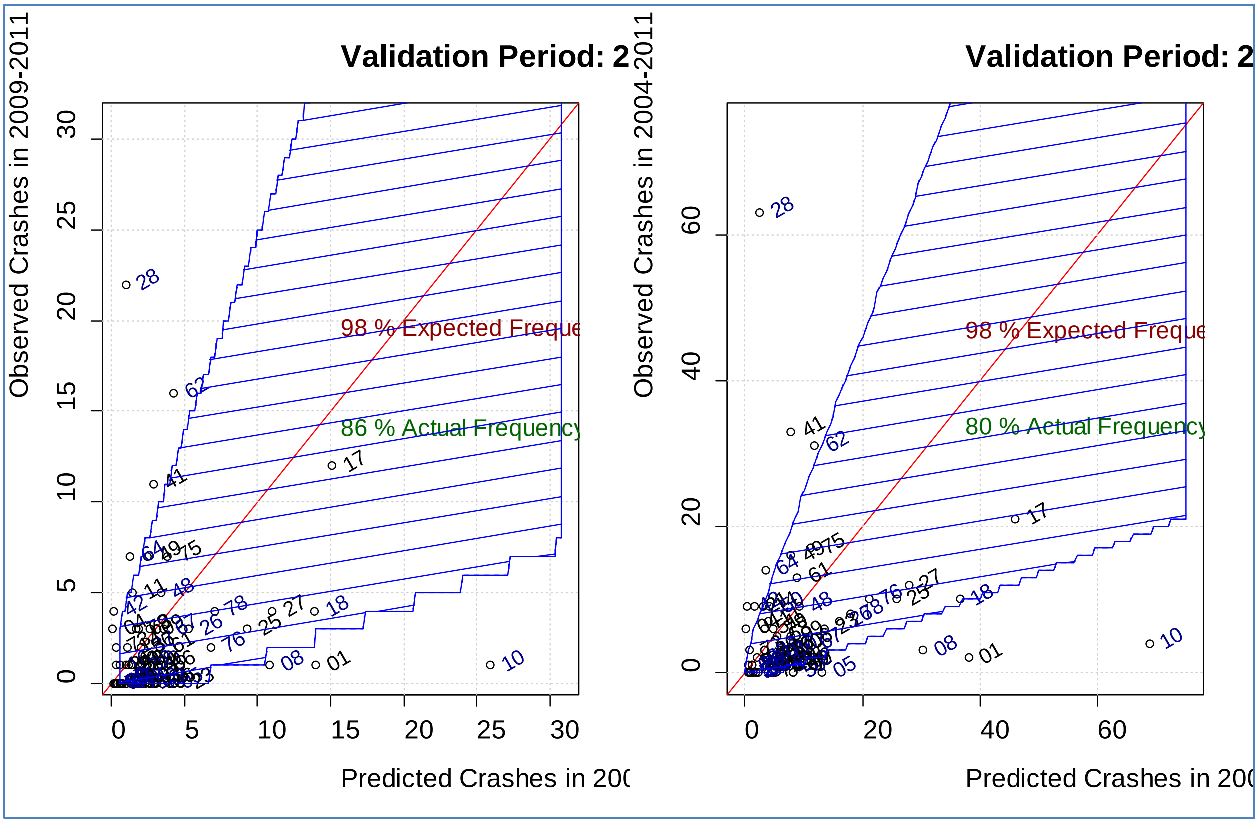


Figure 4.5: Site Expected Frequencies for years 2004 through 2011

These analyses verified that the original urban model has a **limited, but similar, temporal predictive power** for years that are different than the original study period. The validation site data seems to be more widely distributed (or dispersed) than for the original dataset.

4.2.5 Urban Model Coefficient Evaluation and Enhancement

The third metric for the urban model validation consists of comparing the specific coefficients estimated in the original model to similar coefficients that are based on the validation data for the same time period. This initial assessment only includes sites where the data characteristics were within thresholds consistent with the road characteristic values observed for the original study sites. Section 4.2.5.1 addresses submodel validation followed by a more detailed validation and model enhancement evaluation in subsequent sections.

4.2.5.1 Submodel Validation

Using validation site data for 2004 to 2008, the research team estimated multiplicative factors for the submodels associated with Equation 1 in the original analysis (i.e. Equations 2, 3, and 4 in this report but generated for the original study). As shown in Table 4.7, two of the multiplicative factors (baseline and roadway submodels) are statistically significant. The roadside submodel factor is slightly insignificant (p-value slightly larger than 0.10).

Table 4.7: Multiplicative Factor Estimates for Urban Submodels in the Validation Data

Term	Estimate	Std. Error	z-value	p-value	Significance[1]
(Intercept)	0.337	0.511	0.659	0.510	
Baseline	0.526	0.181	2.913	3.58×10^{-03}	**
Roadway	0.439	0.215	2.039	0.0415	*
Roadside	-0.376	0.250	-1.506	0.1321	

[1]*Significance values are as follows:*
° p<0.1; * p < 0.05; ** p < 0.01; and *** p < 0.001

If the validation data models perfectly matched the original urban model, the estimated values would be identical or have a multiplicative calibration value of one. After estimating factors for all submodels, the intercept is expected to equal zero. The value of 0.337 has a large p-value indicating that this estimate is not the zero value as expected. In order to further statistically quantify the differences, a statistical comparison of the individual variables resulted in the graphic representation depicted in Figure 4.6. For this figure, the plotted intervals represent the statistical boundaries (confidence intervals) for the difference between the submodel and expected values. As shown, all but one of the 95 percent confidence intervals contains the value zero. The roadside coefficient deviates significantly from its expected value, a result that identifies the roadside part of the model as the source of the observed differences between the validation dataset and the urban model. This result is very likely influenced by the differences in sample composition identified earlier in Section 4.2.2.

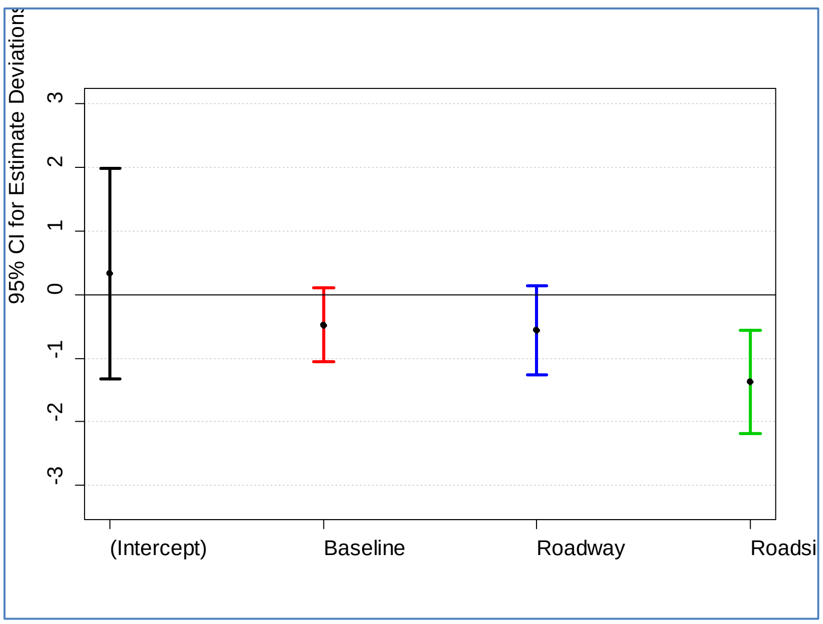


Figure 4.6: Statistical Deviations of Submodel Factors in the Urban Validation Data

A global p-value of 1.701×10^{-06} for the simultaneous comparison is highly significant therefore rejecting the hypothesis that the two composite models are statistically equivalent (p-value based on a Hotelling's T^2 test, with a statistic of 43.940 of four with 64 degrees of freedom).

4.2.5.2 Validation of Individual Predictor Coefficients

The research team followed a similar procedure to that described in Section 4.2.5.1 to simultaneously compare the individual coefficients obtained from the validation dataset (shown in Table 4.8) to those coefficients originally obtained for the urban model (shown in Table 4.1).

Table 4.8: Urban Crash Prediction Model Estimated from the Validation Sample

Term	**Estimate**	**Standard Error**	**z-value**	**p-value**	**Significance[1]**
(Intercept)	-6.825	2.799	-2.438	1.48E-02	*
LnAADT	0.826	0.296	2.795	5.18E-03	**
LnSegmentLength	0.033	0.315	0.104	0.9169	
Speed.Lim.over.35	0.351	0.494	0.711	0.4768	
MedianTWLTL	-0.789	0.471	-1.673	0.0943	º
Four.Travel.Lanes	-0.457	0.421	-1.086	0.2774	
MedianTWLTL:Four.Travel.Lanes	0.465	0.670	0.694	0.4877	
Com.andInd.DW	0.047	0.042	1.109	0.2673	*
Other.DW	0.079	0.039	2.000	0.0455	

[1]*Significance values are as follows:*
º $p<0.1$; * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$

NB2 Theta: 1.024 (Standard Error: 0.248)

Figure 4.7 graphically depicts the confidence intervals for the individual variables. As expected, the values for the original versus the validation model coefficients appear somewhat different. This observation is not surprising based on the previous statistical tests and the known differences in the original versus validation data set.

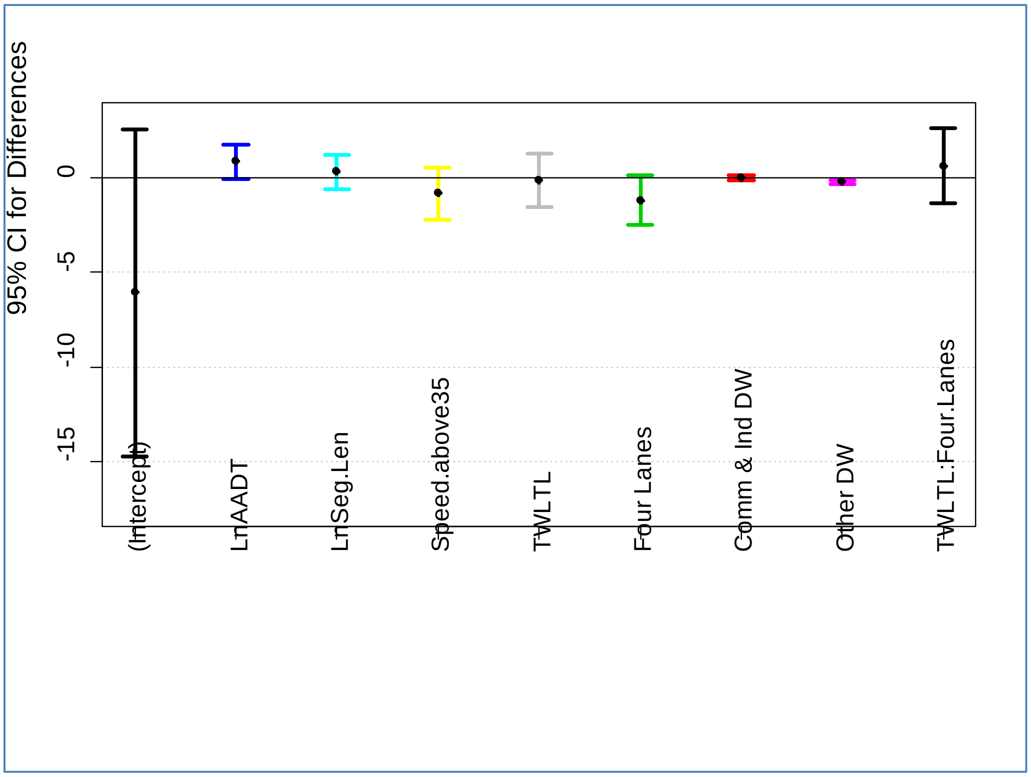


Figure 4.7: Statistical Deviations of the Urban Model Coefficients

A formal assessment of the differences confidently rejects the hypothesis that they are statistically equivalent (p-value of 0.0204 for a Hotelling's T^2 statistic of 22.789 on 9 and 93 degrees of freedom). Graphic inspection of the confidence intervals, however, clearly identifies the term for Other.DW as the main variable that was responsible for the large Hotelling's T^2 statistic as the confidence interval for Other.DW did not include the value zero. In addition to finding this coefficient statistically different for the two datasets, their signs are opposite. This observation suggests that the original model variable for "other" driveways may not adequately represent its expected influence on crashes.

Upon additional assessment, the exclusion of the "other" driveway variable from the simultaneous comparison then results in a statistically insignificant conclusion. This observation then suggests that the other coefficients have a generally acceptable equivalence (p-value of 0.1239 for a Hotelling's T^2 statistic of 15.731 for 9 with 93 degrees of freedom computed for a reduced set of coefficients).

Another important observation is that because the statistical significance of each of the new coefficient estimates is less than previously noted in the original model, the originally proposed model structure may no longer be appropriate to represent prevailing crash conditions at urban sites. This observation may suggest that the original sample did not include a robust enough representation of all possible conditions necessary to fully capture the complex urban condition.

Based on the results of this evaluation, the research team determined that the urban model would benefit from a re-estimation that included the combined data sets (i.e. the pool of the original study and the validation datasets). This enhanced model is presented in Section 4.2.6.

4.2.5.3 Validation of NB2 Theta Values

The negative binomial model also results in an additional parameter known as the theta value (referred to as the NB2 Theta by some statistical programs). This fitted value represents how well the final model converges with the underlying data (simply put, does the model accurately represent observed conditions?). The inverse of this theta value is often further used as an indication of how well the model represents the data that is known to be overdispersed and can then be used as an input into additional evaluations such as the Empirical Bayes site-specific approach. Table 4.8 (validation) and Table 4.1 (original model) show that these coefficients are 1.02 and 6.43, respectively. A formal comparison between the two parameters demonstrates that their differences are statistically significant (p-value of 1.448×10^{-22} for a 13.456 t-statistic on 83 degrees of freedom, for the difference of two independent samples).

This result indicates that the validation dataset is significantly more dispersed than the original dataset (i.e. a smaller NB2 parameter for the validation dataset).

4.2.6 Enhanced Urban Model

The research team proceeded to develop an enhanced urban model using all of the 2004 to 2008 crash data available from the original study and the current validation effort. Table 4.9 shows summary statistics for this combined pool of 110 sites.

Table 4.9: Summary Statistics for the Combined Urban Dataset (2004 to 2008)

Variable	**Mean**	**Standard Deviation**	**Min**	**Max**	**Total**
Non-intersection Crashes	5.24	7.12	0	41	576
AADT (veh/day)	18,506	11,035	1520	52,716	--
Segment Length (mi)	0.26	0.20	0.10	1.25	28.35
Speed Limit (mph)	44.68	8.63	25	55	--
Number of Driveways	5.13	5.65	0	24	564
Number of Commercial driveways	2.59	3.88	0	15	285
Number of Industrial Driveways	0.31	1.14	0	7	34
Number of Other types of Driveways	2.23	3.84	0	24	245

As previously observed during the model validation efforts, validation Site 28 does not closely resemble urban conditions and will likely influence (skew) values for an enhanced model. The research team therefore elected to remove this outlier site from the enhanced urban model data set. The resulting enhanced urban model is represented in Table 4.10. Note that the “other” driveway category is no longer significant and so it has been removed from the model as part of the development process.

Table 4.10: Enhanced Urban Crash Prediction Model

Term	**Estimate**	**Standard Error**	**z-value**	**p-value**	**Significance[1]**
(Intercept)	-7.7522	1.7401	-4.4550	8.39E-06	***
LnAADT	1.0439	0.1877	5.5623	2.66E-08	***
LnSegmentLength	0.4534	0.1656	2.7373	0.0062	**
MedianTWLTL	-0.6756	0.3201	-2.1105	0.0348	*
Four.Travel.Lanes	-0.7035	0.3053	-2.3041	0.0212	*
MedianTWLTL:Four.Travel.Lanes	0.8642	0.4223	2.0464	0.0407	*
No.Commercial.DW	0.1022	0.0286	3.5709	0.0004	*
No.Commercial.DW:Sp.Lim.over.35	-0.0887	0.0414	-2.1401	0.0323	***

[1]*Significance values are as follows:*
º $p<0.1$; * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$

NB2 Theta: 1.457 (Standard Error: 0.286)

This enhanced urban model simplifies the way to control for roadside characteristics and land use. Only the driveways associated with commercial and industrial land proved statistically meaningful. The roadside effect was also determined to be dependent on the speed limit value.

The enhanced urban model shown in Table 4.10 is still represented by the three submodels that collectively make up Equation 1 where:

Predicted Number of Crashes = (Baseline Exposure Values) x *(Effect from Roadway)* x *(Effect from Roadside / Driveways).*

The submodels for the enhanced approach are then represented by the following Equations 5, 6, and 7.

Equation 5. Enhanced Baseline Submodel for Urban Environments (5)

$$\text{Baseline Exposure Values} = (4.298 \times 10^{-4}) \times (AADT^{1.044}) \times (\text{Segment Length}^{0.453})$$

Where:

AADT = Annual Average Daily Traffic (vehicles per day), and
Segment Length = study corridor length (miles).

Note: The 4.298×10^{-4} value is equivalent to $e^{-7.7522}$ (the value shown in Table 4.10).

Equation 6. Enhanced Roadway Submodel for Urban Environments (6)

$$\text{Effect from Roadway} = \exp\left[(-0.676 \times MedianTWLTL) - (0.704 \times Four.Travel.Lanes) + (0.864 \times MedianTWLTL \times Four.Travel.Lanes)\right]$$

Where:

MedianTWLTL = one if a two-way left-turn lane is present (zero value if not),
Four.Travel.Lanes = one if segment has four through lanes (two lanes in each direction) or is equal to zero if the segment has only two lanes (one lane in each direction).

<u>Equation 7. Enhanced Roadside Submodel for Urban Environments</u> (7)

Effect from Roadside & Driveways = exp [(0.102 x *Com.and.Ind.DW) – (0.089* x *Com.and.Ind.DW* x *Sp.Lim.over35]*

Where:

Com.and.Ind.DW = number of commercial plus industrial driveways, and
Speed.Limit.over.35 = one if the speed limit is greater than 35 mph and zero if the speed limit is 35 mph or less.

For roads posted at 25, 30 or 35 mph, the multiplicative effect of the roadside submodel is associated with a 10.8 percent increase in crashes per additional commercial driveway (multiplicative factor of 1.10761= $e^{0.1022}$). The effect is minor for roads posted between 40 and 55 mph. It is estimated that the number of crashes increases by 1.4 percent per additional commercial driveway at these locations (multiplicative factor of 1.0136 = $e^{(0.1022-0.0887)}$).

The researchers finalized the enhanced urban model by verifying its fit to each of the two subsamples included in the new pool of sites. This is because the examination of the validation dataset and the subsequent analyses raised concerns about important differences between those subsets of data. Figure 4.8 shows the results of these comparisons.

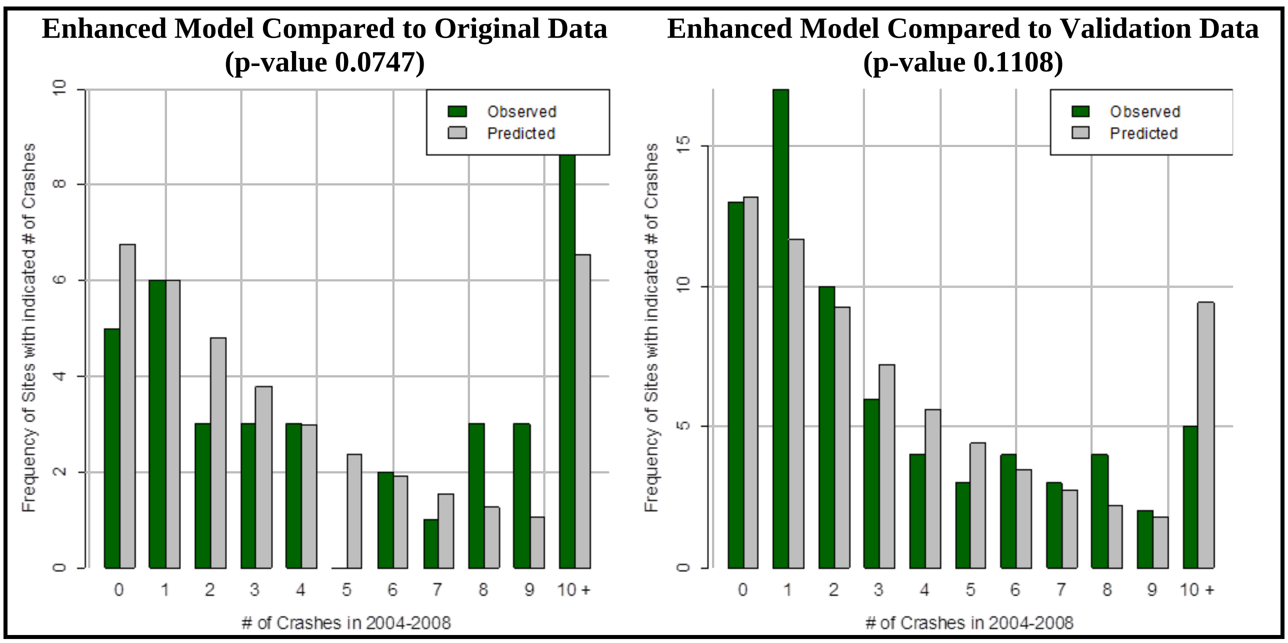


Figure 4.8: Fit of Enhanced Urban Model to Original and Validation Data

Figure 4.8 demonstrates that the enhanced model has reasonable predicting power over each of the subsamples. This is indicated by the fact that both p-values are larger than 0.05, which means that the differences between predicted versus observed crashes are not statistically significant at the 95 percent level.

4.3 RURAL MODEL VALIDATION

The procedures used for validation of the rural model are the same as those applied to the urban validation effort. Using similar sample size techniques, the research team determined that in order to verify each coefficient of the rural model with sufficient statistical power, the analysis required a sample of at least 61 new rural sites. This estimate is based on the magnitude and Standard error of the least significant coefficient in the rural model (the coefficient for the LnTotal.DW variable). The following sections review the original rural model and subsequent validation effort.

4.3.1 Review of Initial Rural Model

In 2012, members of the research team developed rural arterial crash prediction models to use for quantifying safety performance. This research effort, ODOT SPR 720, resulted in the model depicted in Table 4.11.

Table 4.11: Original Crash Prediction Model for Rural Environments

Term	Estimate	Standard Error	z-value	p-value	Significance[1]
(Intercept)	-5.6787	1.1412	-4.976	6.49E-07	***
LnAADT	0.7825	0.1429	5.476	4.35E-08	***
LnSegmentLength	0.2864	0.1259	2.276	0.02287	*
Four.Travel.Lanes	0.7862	0.3358	2.341	0.01922	*
Proportion.Ind.DW	1.2918	0.6077	2.126	0.03353	*
Number.of.DW.Clusters	0.1048	0.0347	3.021	0.00252	**
LnTotal.DW	-0.2864	0.1259	2.276	0.02287	*

[1]*Significance values are as follows:*
º p<0.1; * p < 0.05; ** p < 0.01; and *** p < 0.001

NB2 Theta: 5.5633 (Std.Err.: 4.04)

AIC: 280.7

The companion equation format is once again represented by the general Equation 1 defined as:

Predicted Number of Crashes = (Baseline Exposure Values) x *(Effect from Roadway)* x *(Effect from Roadside / Driveways)*

The three key components of Equation 1 for the rural model are then further defined by Equations 8, 9, and 10.

Equation 8. Baseline from Exposure at Rural Environments (8)

$$\textit{Baseline Exposure Values} = (3.418 \times 10^{-3}) \times (AADT^{\,0.7825}) \times (\textit{Segment Length}^{\,0.2864})$$

Where:

AADT = Annual Average Daily Traffic (vehicles per day), and
Segment Length = study corridor length (miles)

Note: The value 3.418×10^{-3} is equivalent to $e^{-5.6787}$ (the intercept value shown in Table 4.11).

Equation 9. Effect of Roadway on Crashes at Rural Environments (9)

$$\textit{Effect from Roadway} = \exp\,[0.7862 \times \textit{Four.Travel.Lanes}]$$

Where:

Four.Travel.Lanes = one if segment has four through lanes (two lanes in each direction) or a value of zero if the segment has only two lanes (one lane in each direction)

Table 4.12: Possible Cases of the Effect of Roadway at Rural Environments

Two Travel Lanes	*Four Travel Lanes*
1.0000	2.1950

Equation 10. Effect of Roadside Elements on Crashes at Rural Environments (10)

$$\textit{Roadside.effect} = \exp[(1.2918 \times \textit{Prop.of.Ind.DW}) + (0.1048 \times \textit{Total.\#.Clusters})] \;/\; (\textit{Total.\#.Driveways} + 0.5)^{0.2864}$$

Where:

Prop.of.Ind.DW = proportion of industrial driveways (number of industrial driveways divided by the total number of driveways),
Total.#.Clusters = number of directional driveway clusters with a 1.5 second travel time (see appendix Section 7.4 for example calculations), and
Total.#.Driveways = number of individual driveways (all land uses) located in the study corridor.

4.3.2 Characteristics of the Rural Validation Data

The research team collected road characteristic and crash data for 114 additional rural sites (initial site selection of 80 but the sites required additional subdivision to minimize the influence of heterogeneous driveway distributions). Eight sites were removed because their speed limits could not be determined. Two additional sites were removed because they had cross-sections with three lanes, a situation not present in the original model. Finally, AADT information was not available at one additional site; however, the research team estimated this value based on traffic volume trends for years 2005 through 2007. Table 4.13 shows the rural validation data characteristics.

Table 4.13: Summary Statistics for the Rural Validation Data

Variable (n=114)	**Mean**	**Standard Deviation**	**Min**	**Max**	**Total**
Non-intersection Crashes	3.00	3.51	0	16	312
AADT (veh/day)	3848	5578	324	37,669	--
Segment Length (mi)	0.65	0.40	0.02	2.00	67.14
Speed Limit (mph)	54.25	3.15	35	55	--
Number of Lanes	2.15	0.54	2	4	--
Number of Driveways	2.23	3.67	0	25	232
Number of Commercial driveways	0.19	1.39	0	14	20
Number of Industrial Driveways	0.13	0.61	0	5	14
Number of Other types of Driveways	1.90	2.66	0	13	198
Number of driveway clusters	2.02	3.04	0	16	210

In a manner similar to the urban analysis, crash data for years 2004 through 2011 is included in order to perform both the spatial and spatial-temporal transferability validations. A total of 312 non-intersection crashes were identified at the validation sites. The segment lengths extended over 67.14 miles. There are 232 total driveways, with roughly 15 percent used for commercial or industrial purposes. The number of driveway clusters (as defined in the development of the rural model) is very similar to the total number of driveways, a condition that suggests that driveway clustering is minimal (i.e. travel times at the speed limit between driveways tends to exceed 1.5 seconds).

Except for the range of AADTs and a notable increase in the average number of crashes, the summary statistics of the validation dataset are very similar to those of the original data (shown in Table 4.14). This is expected, since both sets were collected using a probability sampling procedure, thus both samples should be representative of prevailing conditions in Oregon.

Table 4.14: Summary Statistics for the Original Rural Data

Variable	**Mean**	**Std.Dev**	**Min**	**Max**	**Total**
Non-intersection Crashes	1.98	2.35	0	11	162
AADT (veh/day)	3129	2327	294	9932	--
Segment Length (mi)	0.66	0.36	0.1	2.00	54.09
Speed Limit (mph)	54.94	0.55	50	55	--
Number of Lanes	2.13	0.47	2	4	--
Number of Driveways	2.55	4.03	0	26	209
Number of Commercial driveways	0.05	0.27	0	2	4
Number of Industrial Driveways	0.13	0.64	0	5	11
Number of Other types of Driveways	2.37	3.68	0	23	194
Number of driveway clusters	2.30	3.29	0	18	189

Table 4.15 shows the Pearson correlation values corresponding to the possible pairs of variables in Table 4.13 (validation data). Refer to Section 4.2.2 for a review of how to interpret Pearson correlation values.

Table 4.15: Pearson Correlations for Rural Validation Data

	Crashs	AADT	Segment Length	Speed Limit	Number of Lanes	Total Number of Driveways	Commercial Driveways	Industrial Driveways	Other Driveway Types	Number of Driveway Clusters
Crashes	1.00									
AADT	0.41	1.00								
Segment Length	0.30	-0.13	1.00							
Speed Limit	-0.27	-0.18	0.05	1.00						
Number of Lanes	0.14	0.75	-0.15	0.01	1.00					
Total Number of Driveways	0.29	0.15	0.01	-0.25	0.18	1.00				
Commercial Driveways	0.25	0.21	-0.08	-0.18	0.35	0.68	1.00			
Industrial Driveways	0.25	0.20	0.05	-0.33	0.23	0.71	0.82	1.00		
Other Driveway Types	0.21	0.06	0.04	-0.17	0.01	0.86	0.22	0.32	1.00	
Number of Driveway Clusters	0.26	0.09	0.06	-0.24	0.09	0.97	0.53	0.61	0.92	1.00

Table 4.15 shows a moderate correlation between crashes and the driveway related variables. There are high correlations between the driveway related variables (as expected) since more driveways would be associated with more driveway clusters. Similarly, Table 4.16 shows the correlation matrix for the data used to develop the original rural model.

Table 4.16: Pearson Correlations for Original Rural Data

	Crashes	AADT	Segment Length	Speed Limit	Number of Lanes	Total Number of Driveways	Commercial Driveways	Industrial Driveways	Other Driveway Types	Number of Driveway Clusters
Crashes	1.00									
AADT	0.53	1.00								
Segment Length	0.07	-0.16	1.00							
Speed Limit	0.05	-0.03	0.06	1.00						
Number of Lanes	0.38	0.32	-0.07	0.03	1.00					
Total Number of Driveways	0.24	0.16	0.25	-0.10	-0.15					
Commercial Driveways	0.32	0.10	0.07	0.02	-0.05	0.54	1.00			
Industrial Driveways	-0.07	-0.04	-0.04	-0.32	-0.06	0.40	0.10	1.00		
Other Driveway Types	0.25	0.17	0.28	-0.05	-0.15	0.99	0.50	0.26	1.00	
Number of Driveway Clusters	0.22	0.15	0.29	-0.09	-0.16	0.98	0.44	0.35	0.98	1.00

There are a few noticeable differences between the Pearson correlation values in the original and the new data that warrant consideration:

1. In the original rural dataset, the speed limit does not seem to be associated with crashes (correlation of +0.05), but it is moderately associated for the validation rural dataset (correlation of -0.27).

2. The linear association between the number of lanes and crashes increased from +0.32 in the original dataset to +0.75 in the validation dataset.

3. The number of lanes has a mild negative correlation to the number of driveways in the original dataset (-0.15) but the sign for the Pearson correlation value changed by a similar magnitude in the validation dataset (+0.18).

4. The number of driveways were minimally associated with the segment length in the original dataset (correlation of +0.25), but these two variables are independent in the new dataset (correlation of +0.01).

The research team verified that these changes most likely reflect different sample compositions: the original dataset included sites posted at 50 or 55 mph and AADTs up to 9932 vehicles per day only. In contrast, the validation sites' speed limits ranged from 35 to 55 with AADT values as high as 37,670.

Because of the extended AADT range, the validation dataset offers the opportunity to re-estimate the rural model to better account for a wider range of traffic values. An expanded range of speed limits may also suggest a wider range of geometric and access management characteristics. Upon inspection, five of the sites had speed limits below 50 mph and road conditions more closely associated with urban conditions.

The fourth observation noted above suggests that the validation data may offer a chance to better distinguish between the effects of number of driveways and segment length. These two effects seemed to influence each other in the original analysis, so the research team elected to fix their negative correlation when estimating the original rural model (2012, pp. 51-52). This fixed negative correlation in the original model proved later to increase the complexity of using the model. Therefore, a simpler model may be feasible when developing an enhanced model since the previously noted correlations between driveways and segment length are no longer present in the validation dataset.

4.3.3 Spatial Transferability of Rural Model

This analysis entails the comparison of the rural model predictions and the corresponding actual crashes at sites in the validation dataset.

4.3.3.1 Direct Comparison of Predicted to Observed Crashes (Rural)

The research team compared the distribution of the 2004 to 2008 crashes in the validation dataset to the theoretical distribution of predicted crashes obtained from the original rural model as shown in Figure 4.9. The distribution does not have significant differences between the two distributions. This visual inspection indicates the original rural model cleanly fits the observed crashes for the validation sites.

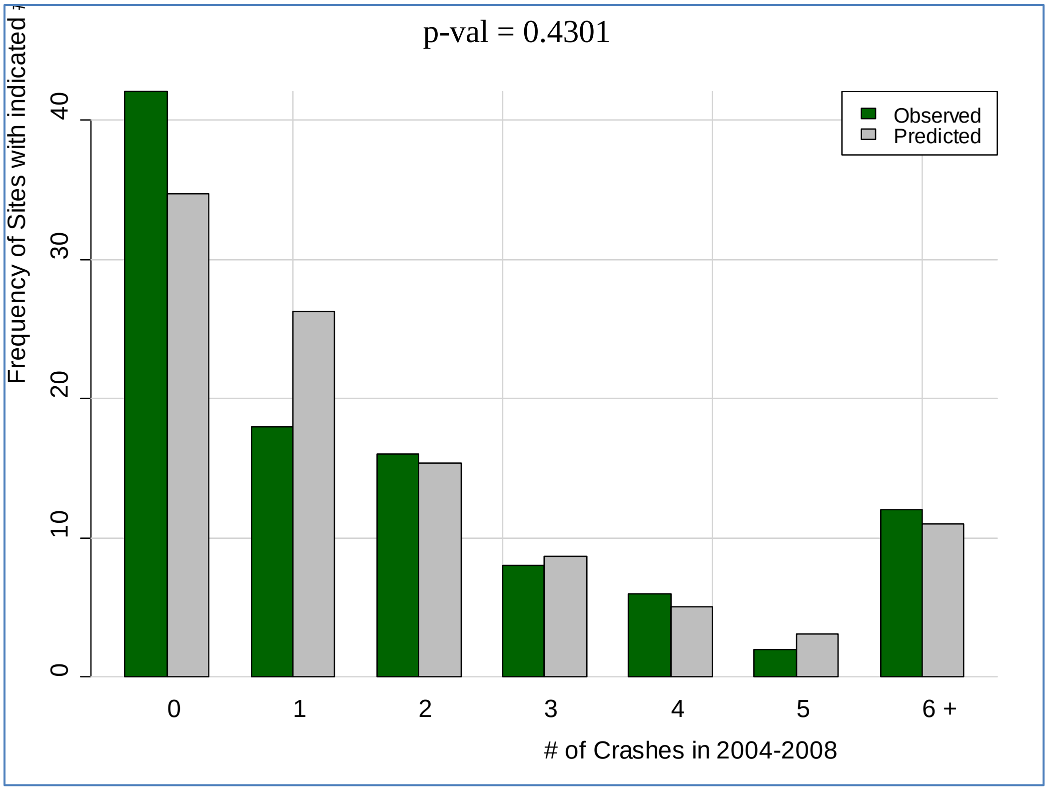


Figure 4.9: Original Model Predictions versus Observed Validation Crashes (Rural)

To verify that the differences between the two distributions are within the expected boundaries of prediction, the research team performed a chi-squared test to analytically compare the two series shown in the figure. The resulting p-value was 0.4301 for a 4.885 chi-squared statistic on five degrees of freedom. This test supports the hypothesis that the original model is valid, since the p-value is rather large (substantially greater than the widely accepted p-value significance level of 0.05). This means that the predicted crash frequency values developed using the original rural model are **not statistically different** than the observed crash frequencies of the validation sites.

4.3.3.2 Expected Frequency Thresholds

As introduced in Section 4.2.3.2 of the urban validation, the use of an Expected Frequency Zone, as shown in Figure 4.10, can help to demonstrate how well the original rural model performs at locations with increased crash frequency (and a wider variability).

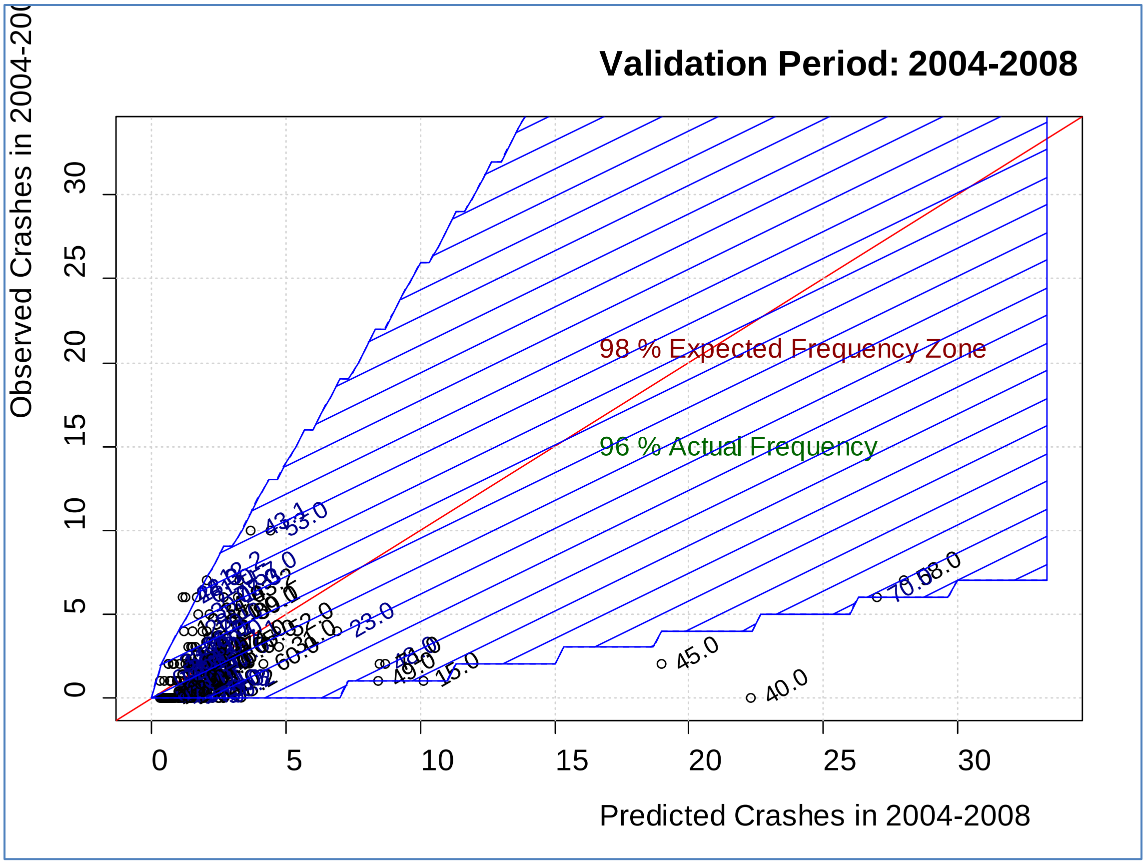


Figure 4.10: Expected and Observed Rural Crash Frequencies (Sites are identified by their assigned sample numbers)

Figure 4.10 supports the conclusion that the original rural model is valid beyond the original dataset, since the percentage of sites within the *Expected Frequency Zone* for 98 percent of the sites is actually 96 percent, a value very close to the prediction.

It is noticeable that when the original rural model predicts crash frequencies greater than 11 crashes, only half of the sites are contained in the *Expected Frequency Zone* (two out of four sites). It is also apparent that all the crash frequency values in this region greater than 11 predicted crashes are below the diagonal line. This observation would suggest that the original model tends to over-predict in this higher crash frequency region.

In particular, the model predicted 22.3 and 19.0 at Sites #40 and #45 respectively, but their actual crashes were zero and two. These sites both have speed limits of 55 mph, cross-sections with four lanes, and lengths of approximately one-half mile. The research team verified that the large AADTs seem to be responsible for the large predictions. Table 4.14 further shows that the range of crashes for the original dataset extended from zero to eleven. Therefore, crash predictions outside those boundaries are deemed unreliable. Regardless, the validation analysis provides convincing evidence of the good predicting power for the original rural model.

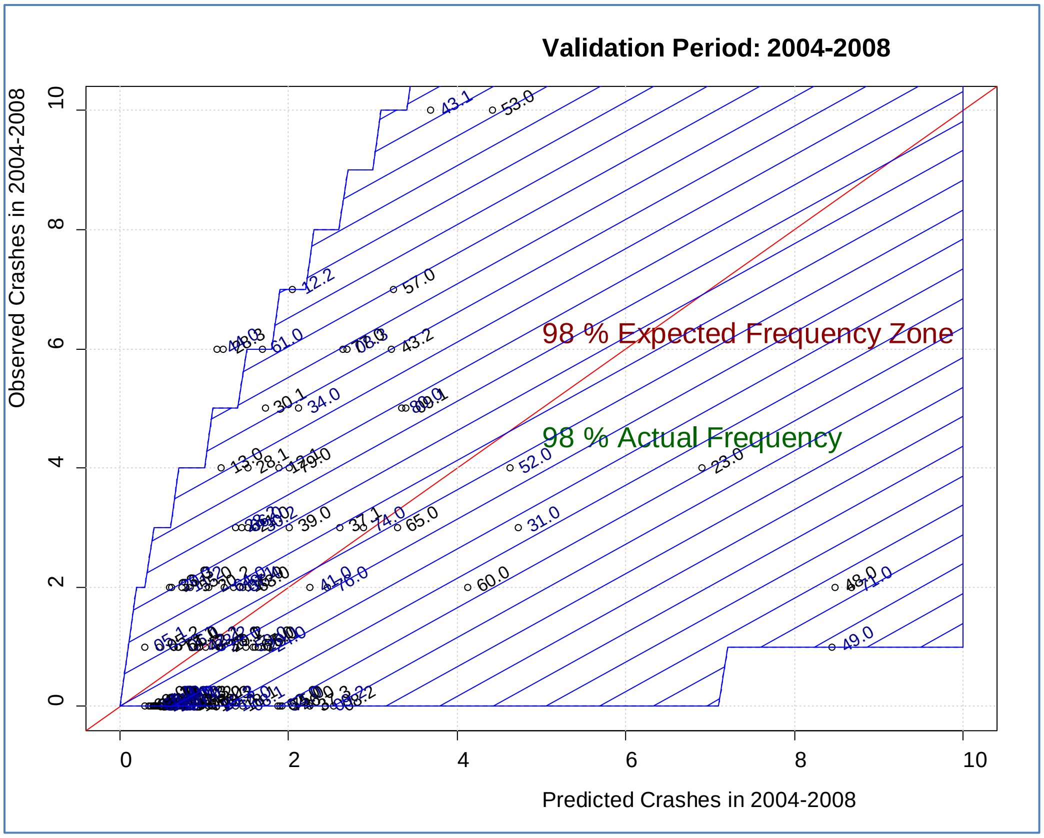


Figure 4.11: Expected and Actual Crash Frequencies in Rural Dataset for predictions smaller than 10 crashes in 5 years (sites are identified by their assigned numbers)

Figure 4.11 verifies that the original rural model performed exceptionally well for locations with less than ten crashes in the five year period. For these locations, the observed crashes were distributed evenly above and below the diagonal line (another strong indication that the original rural model performs well).

4.3.4 Spatial-Temporal Transferability of Rural Model

The spatial analysis presented in Section 4.3.3 can be applied to the assessment of the following two time-based scenarios:

1. Evaluation for crashes that only occurred from 2009 until 2011, and

2. Evaluation of all crash data from 2004 through 2011.

Figure 4.12 shows the observed versus predicted crash frequency distributions that correspond to these two time-based scenarios. Similarly, Figure 4.13 shows the associated graphic representation of the *Expected Frequency Zone* for 98 percent.

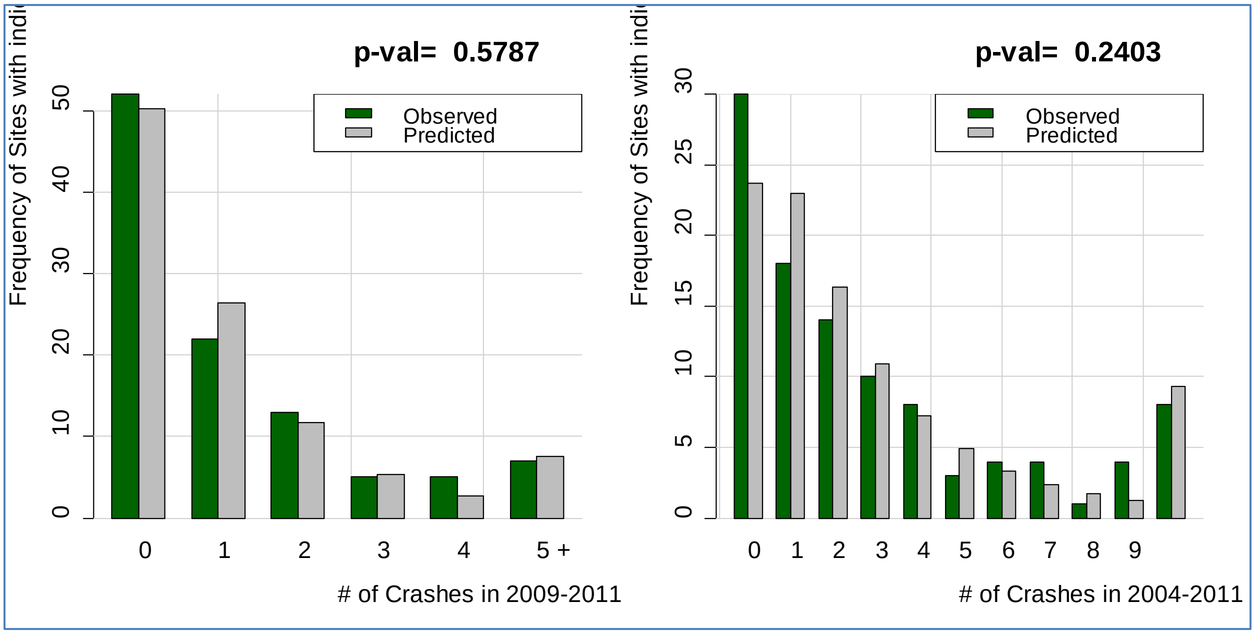


Figure 4.12: Marginal Distributions of Crashes for 2004 through 2011

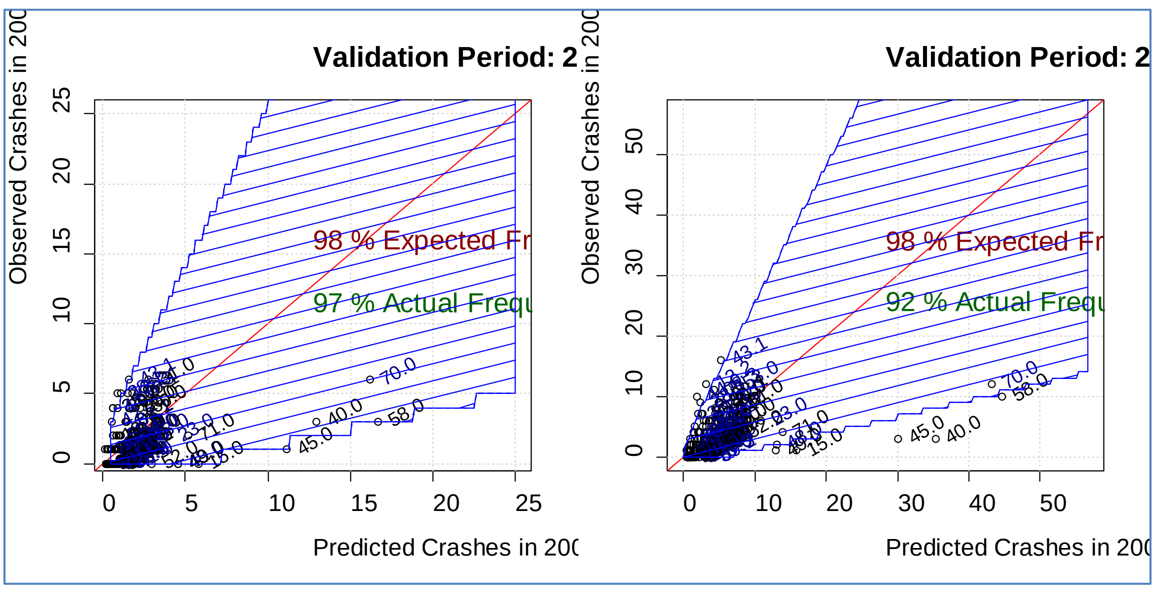


Figure 4.13: Site Expected Frequencies for years 2004 through 2011

This analysis verified that the original rural model is statistically valid over time and location. The higher crash frequency issue noted in Section 4.3.3.2 is once again noticeable when the predictions exceed 11 crashes for five years (corresponding to 6.6 crashes for three years and 17.6 crashes for eight years).

4.3.5 Rural Model Coefficient Evaluation and Enhancement

Detailed rural model validation included the following three procedures:

- Submodel validation,
- Validation of individual predictor coefficients, and
- Validation of NB2 coefficients.

For this assessment, the project team excluded eight sites with AADTs exceeding 10,000 vehicles per day and four additional sites with speed limits ranging from 35 to 45 mph since the original model prediction did not include these ranges of AADTs and speed limits. In addition, this section introduces the enhanced rural model developed using the combined original and validation data.

4.3.5.1 Submodel Validation

Based on validation data from 2004 to 2008, the research team used the maximum likelihood estimation technique to estimate multiplicative factors that correspond to the submodels in the original rural analysis (i.e. equations 8, 9, and 10 in this report). The multiplicative factors for the baseline and roadside submodels are statistically significant, as shown in Table 4.17. The roadway submodel factor is moderately significant (p-value between 0.05 and 0.10).

Table 4.17: Multiplicative Factor Estimates for Rural Submodels in the Validation Data

Term	Estimate	Std. Error	z value	p-value	Significance[1]
(Intercept)	-0.0146	0.1460	-0.1003	0.9201	
Baseline	1.3837	0.2036	6.7962	0.0000	***
Roadway	-1.7860	0.9527	-1.8746	0.0608	º
Roadside	0.8913	0.4091	2.1786	0.0294	*

[1]*Significance values are as follows:*
º p<0.1; * p < 0.05; ** p < 0.01; and *** p < 0.001

Except for the intercept (whose expected value is zero), a value of one is expected for all the factors in Table 4.17. By simple inspection, it appears that the rural roadway submodel may significantly deviate from 1.0. To formally assess how these estimates compare to their expected values, the research team performed a statistical comparison for the individual models. Figure 4.14 graphically shows the results of this comparison. Each 95 percent confidence interval includes a value of zero, indicating that the estimates in Table 4.17 are equivalent to their theoretically expected values.

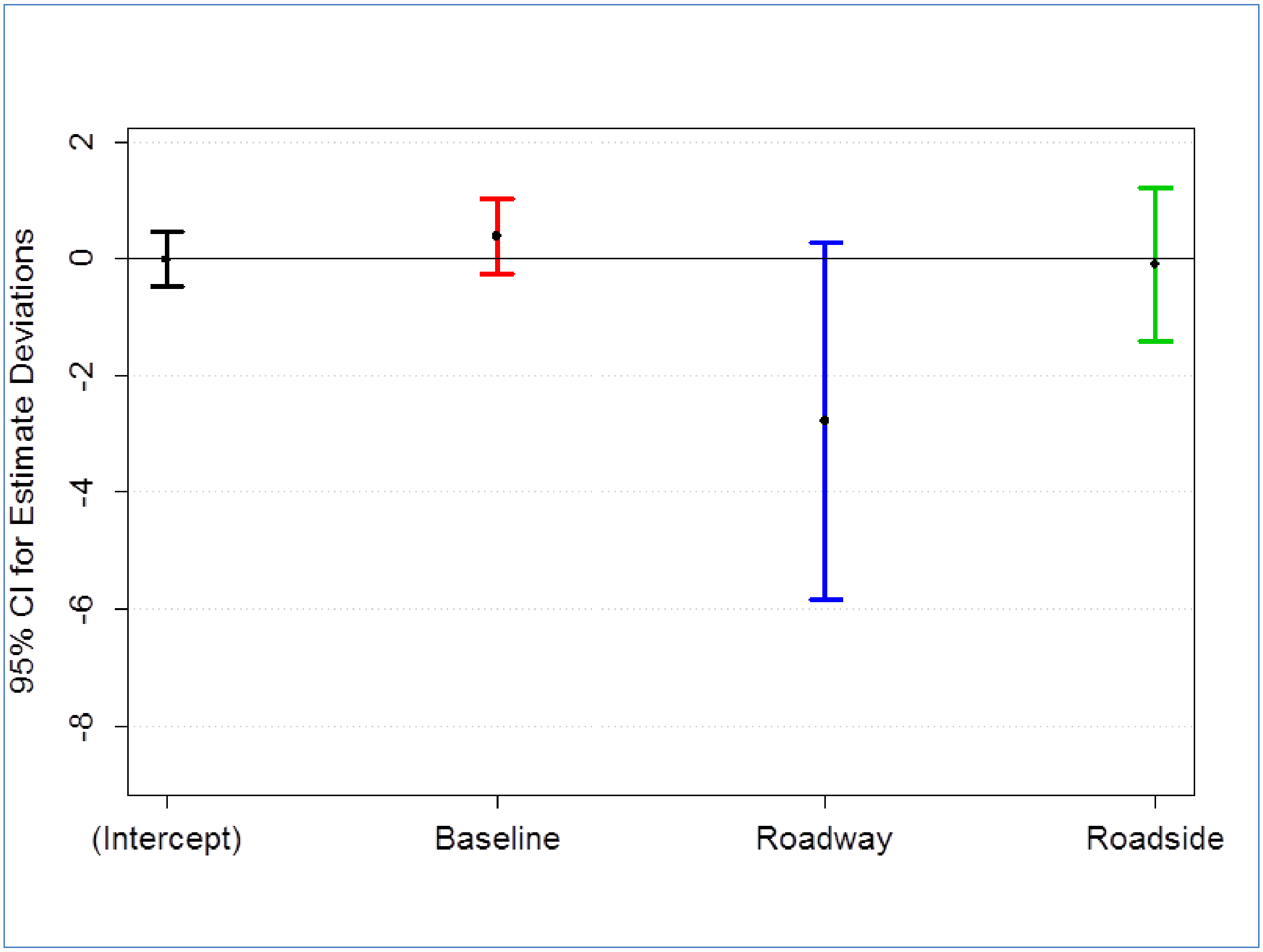


Figure 4.14: Statistical Deviations of Submodel Factors in the Rural Validation Data

A global p-value of 0.0478 for the simultaneous comparisons is marginally significant (p-value based on a Hotelling's T^2 test, with a statistic of 13.1582 on 4 and 92 degrees of freedom). This p-value, therefore, rejects the hypothesis that the two models are statistically equivalent based on assessing multiplicative factors. Additional inspection of individual model variables (see next section) will help identify the specific characteristic that is responsible for this difference.

4.3.5.2 Validation of Individual Predictor Coefficients

Estimation of the individual coefficients can occur by simultaneously comparing the coefficients obtained from the validation dataset (shown in Table 4.18) to those coefficients originally obtained for the rural model (shown in Table 4.11).

Table 4.18: Rural Crash Prediction Model Estimated from the Validation Sample

Term	Estimate	Standard Error	z-value	p-value	Significance[1]
(Intercept)	-8.2279	1.3193	-6.2365	4.47E-10	***
LnAADT	1.0712	0.1652	6.4850	8.88E-11	***
LnSegmentLength	0.3896	0.1141	3.4158	6.36E-04	***
Four.Travel.Lanes	-1.3226	0.7607	-1.7386	0.0821	º
Proportion.Ind.DW	1.0071	0.6125	1.6441	0.1001	
Number.of.DW.Clusters	0.2697	0.0512	5.2635	1.41E-07	***
LnTotal.DW	-0.3896	0.1141	3.4158	6.36E-04	***

[1]*Significance values are as follows:*
º $p<0.1$; * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$

NB2 Theta: 2.73 (Std. Err.: 1.21)

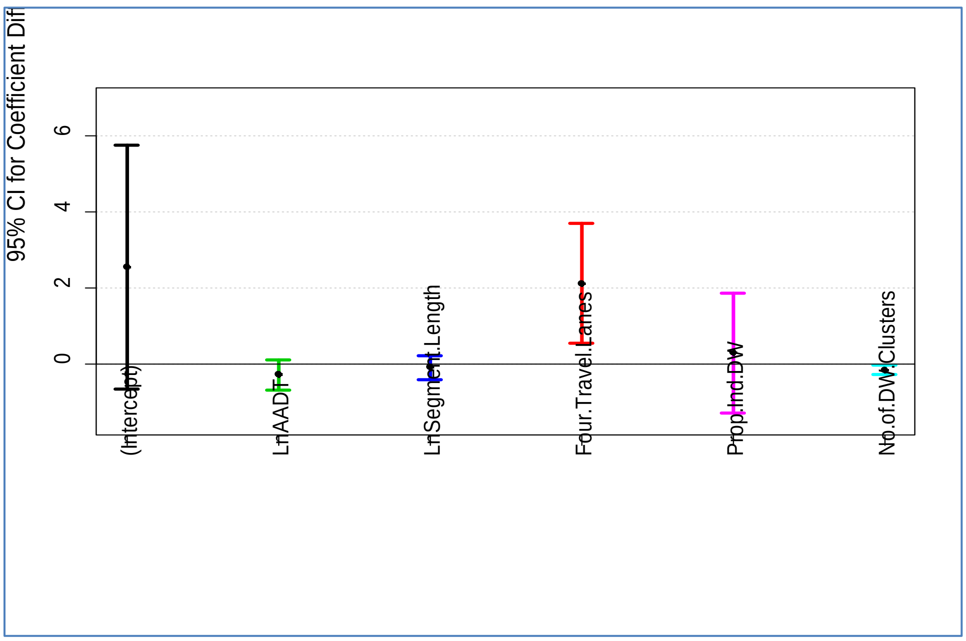


Figure 4.15: Confidence Intervals for the Statistical Deviations of the Rural Model Coefficients

A comparison of this set of coefficients to the original model coefficients rejects the hypothesis that they are statistically equivalent (p-value of 0.0146 for a Hotelling's T^2 statistic of 16.916 on six and 167 degrees of freedom). Graphically inspecting the plot of the 95 percent confidence intervals (see Figure 4.15) demonstrates that the differences between coefficients for the variables "Four.Travel.Lanes" and "Number of.DW.Clusters" are likely responsible for the large Hotelling's T^2 statistic. Upon removal of these two variables, the test results are then statistically insignificant (p-value of 0.6746 for a Hotelling's T^2 statistic of 4.1356 on 6 and 167 degrees of freedom computed from a reduced set of coefficients). This analysis, therefore, suggests that the set of coefficients obtained in the original rural model are statistically equivalent to the estimates in the validation dataset, except for the coefficients corresponding to the variables "Four.Travel.Lanes" and "Number.of.DW.Clusters".

4.3.5.3 Validation of NB2 Theta Values

The last step of this validation analysis requires an evaluation of the theta value (previously discussed in Section 4.2.5.3). From Table 4.11 (Original Rural Model) and Table 4.18 (Validation Rural Model), the theta coefficients are clearly different (values of 5.56 [original model] versus 2.73 [validation model]). A formal comparison of these values determined that the difference is statistically significant (p-value of 9.169×10^{-19} for a 10.023 t-statistic on 165 degrees of freedom).

This result indicates that, after accounting for all the predictor variables, the validation dataset is significantly more disperse than the original dataset (i.e. a smaller NB2 theta parameter from the validation dataset than from the original dataset). The research team concluded, based on this observation, that this parameter would also benefit from a new model estimation based on a pooled dataset.

4.3.6 Enhanced Rural Model

The research team next developed an enhanced rural model using data from the original study in combination with the data from the current validation effort. Only sites posted at 50 and 55 mph were used in this analysis due to site specific issues associated with the lower speed limit locations. The five sites with lower posted speed limits occurred at the edges of towns in regions where the road appeared to be transitioning from lower to higher speeds. In addition, two of the sites had advisory curve speed signs suggesting that the curve geometry is probably an important influence on safety for those specific locations (sharp curve geometry is not a variable unique to the other study sites). Table 4.19 summarizes the pooled data available for development of the enhanced model.

Table 4.19: Summary Statistics for the Combined Rural Dataset (2004 to 2008)

Variable	**Mean**	**Std.Dev**	**Min**	**Max**	**Total**
Non-intersection Crashes	1.82	2.25	0	11	330
AADT (veh/day)	3475	4420	294	37,653	--
Segment Length (mi)	0.65	0.38	0.02	2.00	118.09
Speed Limit (mph)	54.93	0.55	50	55	--
Number of Driveways	2.30	3.76	0	26	416
Number of Commercial driveways	0.16	1.14	0	14	29
Number of Industrial Driveways	0.08	0.45	0	5	14
Number of driveway clusters	2.08	3.08	0	18	376

The research team performed a model selection procedure and developed an enhanced model as shown in Table 4.20.

Table 4.20: Enhanced Rural Crash Prediction Model

Term	**Estimate**	**Standard Error**	**z-value**	**p-value**	**Significance[1]**
(Intercept)	-5.5213	0.6975	-7.9154	2.47E-15	***
LnAADT	0.7947	0.0877	9.0628	1.27E-19	***
LnSegmentLength	0.7333	0.1517	4.8352	0.0000	***
Proportion.Ind.DW	0.7558	0.4739	1.5949	0.1107	
Number.of.DW.Clusters	0.0457	0.0210	2.1790	0.0293	*

[1]*Significance values are as follows:*
º p<0.1; * p < 0.05; ** p < 0.01; and *** p < 0.001

NB2 Theta: 2.576 (Std. Err.: 0.745)

This new model greatly simplifies the procedure to control for roadside characteristics and land use. It also no longer includes a roadway effect term as was present in the original model. The only driveways that proved critical for the enhanced rural model are driveways at industrial sites and the coefficient for this variable (known as Proportion.Ind.DW) is, at best, borderline significant. The research team elected to preserve this coefficient, however, because of its slight but meaningful contribution to reducing the AIC statistic during model selection (this AIC reduction improves the model fit). During development of the enhanced model, the researchers observed that total number of driveways and the number of driveway clusters tended to not appear in the model at the same time. This observation is not unexpected because of the strong correlation between these variables. Consequently, the final recommended model no longer includes a total driveway variable.

The enhanced rural model shown in Table 4.20 is now only represented by two submodels that collectively contribute to the prediction of the number of crashes in five years as shown in Equation 11.

Equation 11. Enhanced Baseline Submodel for Rural Environments (11)

Predicted Number of Crashes = (Baseline Exposure Values) x *(Effect from Roadside / Driveways).*

The submodels for the enhanced approach are then represented by the following Equations 12 and 13.

Equation 12. Enhanced Baseline Submodel for Rural Environments (12)

$$\text{Baseline Exposure Values} = (4.0 \times 10^{-3}) \times (AADT^{0.7947}) \times (\text{Segment Length}^{0.7333})$$

Where:

AADT = Annual Average Daily Traffic (vehicles per day), and
Segment Length = study corridor length (miles).

Note: The 4.0×10^{-3} value is equivalent to $e^{-5.5213}$ (the value shown in Table 4.20).

Equation 13. Enhanced Roadside Submodel for Rural Environments (13)

$$\text{Effect from Roadside \& Driveways} = \exp\left[(0.7558 \times Prop.Ind.DW) + (0.0457 \times Total.\#.Clusters)\right]$$

Where:

Prop.of.Ind.DW = proportion of industrial driveways (number of industrial driveways divided by the total number of driveways), and
Total.#.Clusters = number of directional driveway clusters with a 1.5 second travel time.

The multiplicative effect of the roadside submodel can be reduced to two components: a 4.7 percent increase in crashes per each additional cluster of driveways (i.e. $1.0467 = e^{0.0457}$) and a 0.76 percent increase in crashes per percentage point increase of driveways associated with industrial land use (i.e. $1.007587 = e^{(0.75581/100)}$).

Figure 4.16 shows the comparison of the enhanced model to the two individual data subsets that collectively made up the pooled data (i.e., the original data and the validation data). The large p-value of 0.2718 for the validation dataset fit indicates that the model predicts crashes fairly well for this dataset. In contrast, the low p-value for the original dataset indicates a reduced prediction power. By inspection of the plot, it is evident that this small p-value is likely due to an unusual set of sites with four observed crashes in five years. This frequency is predicted at 4.67 but instead there are ten sites with that frequency. This unexpected trend in the observed crash data was also a consideration with the development of the original model. The remaining predictions are reasonably consistent to the actual frequencies. The research team believes, therefore, that the unexpected p-value for the original data comparison does not indicate a general prediction problem but rather is an artifact of an unexpected trend in the site data.

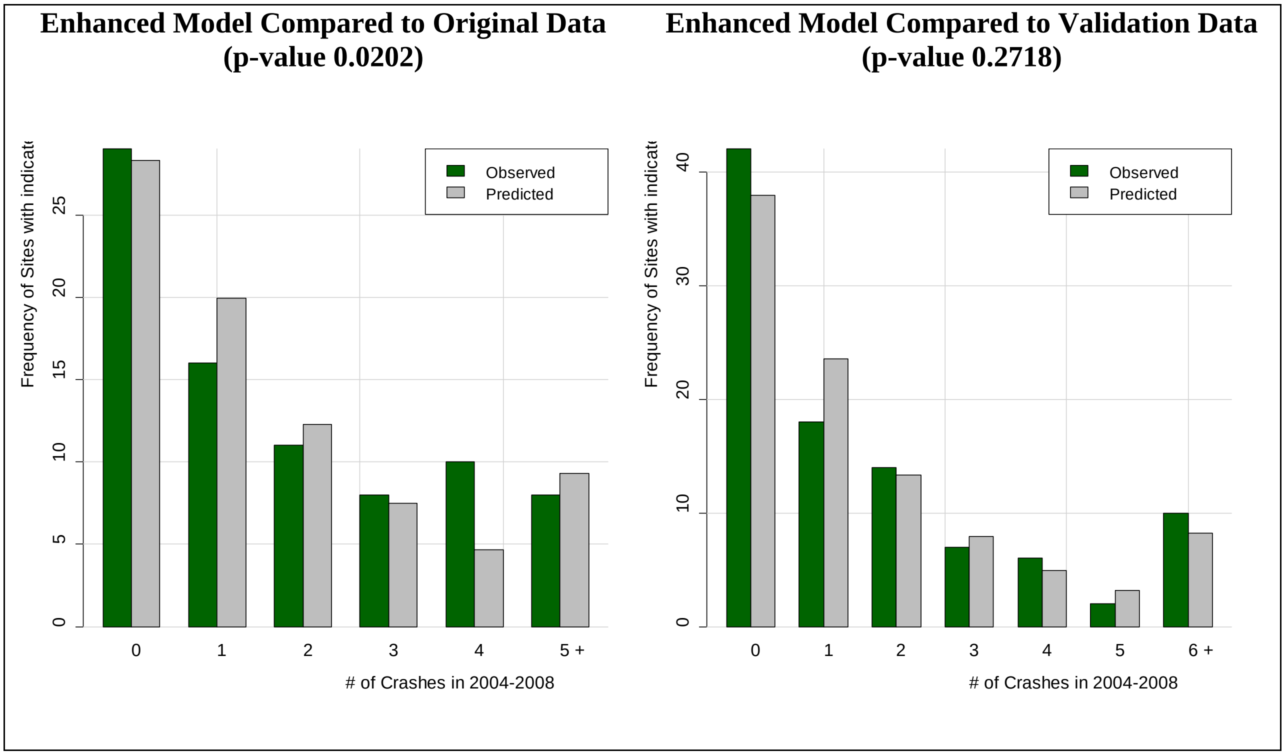


Figure 4.16: Fit of Enhanced Rural Model to Original and Validation Data

4.4 SUMMARY OF RESULTS

In general, the original models for urban and rural arterials adequately predicted crashes for similar Oregon locations. Specific findings are summarized as follows:

<u>Urban Model</u>

- Spatial transferability was determined to be statistically valid when comparing the original model to predicted crashes at the validation sites.
- Spatial-temporal transferability of the urban model had limited but similar time-based predictive powers, but the statistical validity was not as strong for spatial-temporal transferability as it was for spatial transferability alone.
- The *Baseline Exposure* and the *Roadway Effect* were determined to be statistically similar when compared to the validation data. The *Roadside & Driveway Effect* in Equation 1, however, demonstrated notable differences from those observed for the validation sites.
- The individual predictor coefficients were statistically equivalent except for the “Other DW” variable.
- Development of an enhanced urban model enabled a more refined roadside & driveway sub model that better predicts the overall urban arterial crashes.

<u>Rural Model</u>

- Spatial and temporal transferability for the rural model were statistically valid for sites with AADT values less than or equal to 10,000 vehicles per day or speed limits below 50 mph (i.e. site features for the original model).
- The *Baseline Exposure, Roadway Effects,* and *Roadside& Driveway Effects* in Equation 1 were determined to be statistically equivalent for the validation sites.
- The individual predictor coefficients were statistically equivalent except for the "Four.Travel.Lanes" and "Number.of.DW.Clusters" variables.
- Development of an enhanced rural model facilitated a more streamlined technique that no longer requires the *Roadway Effects* submodel.

Page Intentionally Blank

5.0 CONCLUSIONS AND RECOMMENDATIONS

This report summarized the validation effort for urban and rural arterial segment safety performance models developed as part of SPR 720 (report titled *Quantifying Safety Performance of Driveways on State Highways*). The project team randomly selected rural and urban validation sites, acquired physical site characteristic information, and located associated crashes. Using this new site information, the research team then evaluated how well the original models predicted crashes at these new locations. The original models performed well and could be confidently used; however, the larger data set also provided an opportunity to re-evaluate some of the variables that were borderline significant and identify opportunities to simplify the equations. Included with this effort was an update to the "smart spreadsheet" that provides a tool to easily apply the newly enhanced models. A user guide for this tool is included in the appendix of this report. The project team recommends that ODOT consider using the enhanced models. Section 5.1 summarizes the enhanced (final) urban model and demonstrates the use of the model with a sample problem. Similarly, Section 5.2 reviews the enhanced rural model and sample application.

5.1 FINAL URBAN MODEL AND EXAMPLE APPLICATION

As previously noted, the original urban model performed well when validated for spatial transferability, but this performance was not as strong when this assessment was expanded to spatial-temporal validation. Consequently, the project team recommends that ODOT use the following enhanced model (the equation shown combines equations 1, 5, 6, and 7 (introduced in Section 4.2.6).

Equation 14. Enhanced Urban Model (14)

$$\begin{aligned} &Crashes\ for\ 5\ years \\ &= (4.298\times\ 10^{-4})\times(AADT^{1.044})\times(Segment\ Length^{0.453}) \\ &\times\left(e^{(-0.676\times MedianTWLTL)-(0.704\times Four.Travel.Lanes)+(0.864\times MedianTWLTL\times Four.Travel.Lanes)}\right) \\ &\times\left(e^{(0.102\times Com.and.Ind.DW)-(0.089\times Com.and.Ind.DW\times Sp.Lim.over35)}\right) \end{aligned}$$

Where:

AADT = Annual Average Daily Traffic (vehicles per day), and
Segment Length = study corridor length (miles).
MedianTWLTL = one if a two-way left-turn lane is present (zero value if not),
Four.Travel.Lanes = one if segment has four through lanes (two lanes in each direction) or is equal to zero if the segment has only two lanes (one lane in each direction).
Com.and.Ind.DW = number of commercial plus industrial driveways, and
Speed.Limit.over.35 = one if the speed limit is greater than 35 mph and zero if the speed limit is 35 mph or less.

5.1.1 Use of the Urban Model

To demonstrate application of this enhanced model, the user must first locate the required site information for the urban arterial location. The following information is needed prior to evaluation:

- Length of the road segment to analyze (in miles),
- AADT for the segment (vehicles per day),
- Speed limit for the road segment,
- Cross-section information including the number of travel lanes and presence of TWLTL median, and
- Total number of driveways dedicated to commercial and industrial land uses (note that total driveways or the number of driveways dedicated to other lane uses with low volumes are no longer required).

The predictive procedure can then be performed by applying Equation 14. This equation consists of the following components:

1. *Effect of Exposure Factor* based on Equation 5.
2. *Roadway Effect Factor* from Equation 6.
3. *Roadside & Driveway Effect* using Equation 7.
4. Multiply the results as demonstrated in Equation 1 to obtain the expected number of crashes for the study segment during a five year period.

5.1.2 Example Use of the Urban Model

This section demonstrates how to use the methodology outlined in the previous section. This sample problem is the same one shown in SPR 720 for urban arterials but has been updated for the enhanced model. This study site, located in Redmond, Oregon, is illustrated in Figure 5.1.

The required information from this site is identified in Table 5.1.

Table 5.1: Sample Input for Urban Example Problem from Redmond, Oregon

Urban Segment Features	Characteristics
Segment length	0.12 miles
AADT	24,800 vpd
Speed limit	45
Number of travel lanes	4
TWLTL median	Yes
Total commercial and industrial driveways	7

Figure 5.1: Sample Site #23, Redmond, Oregon

Step 1: Compute the *Effect of Exposure Factor* using Equation 5.

Baseline Exposure Values = $(4.298 \times 10^{-4}) \times (AADT^{1.044}) \times (Segment\ Length^{0.453})$

Baseline Exposure Values = $(4.298 \times 10^{-4}) \times (24{,}800^{1.044}) \times (0.12^{0.453}) = 6.37$

Step 2: Calculate the *Roadway Effect Factor* using Equation 6.

Since this segment has a TWLTL median and four travel lanes, the adjustment factor should be 0.60 as calculated below.

Effect from Roadway = $\exp[(-0.676 \times MedianTWLTL) - (0.704 \times Four.Travel.Lanes) + (0.864 \times MedianTWLTL \times Four.Travel.Lanes)]$

Effect from Roadway = $\exp[(-0.676 \times 1) - (0.704 \times 1) + (0.864 \times 1 \times 1)] = 0.60$

Step 3: Compute the Roadside & Driveway Effect using Equation 7.

Since this segment has a speed limit of 45 mph and 7 commercial and industrial driveways, the adjustment factor should be 1.10 as calculated below.

Effect from Roadside & Driveways = $\exp[(0.102 \times Com.and.Ind.DW) - (0.089 \times Com.and.Ind.DW \times Sp.Lim.over35]$

Effect from Roadside & Driveways = $\exp[(0.102 \times 7) - (0.089 \times 7 \times 1] = 1.10$

Step 4: Obtain the predicted number of crashes for a 5 year period for the segment by multiplying all of the above results.

Predicted Number of Crashes = (Baseline Exposure Values) x *(Effect from Roadway)* x *(Effect from Roadside & Driveways)*

Predicted Number of Crashes = 6.37 x *0.60* x *1.10 = 4.2 predicted crashes in 5 years*

Example problem conclusion:
Based on exposure, roadway, and roadside characteristics we can predict that over a period of 5 years approximately 5 (rounded from 4.2) non-intersection crashes will occur for the 0.12 mile long segment (or approximately one crash per year).

5.2 FINAL RURAL MODEL AND EXAMPLE APPLICATION

As reviewed in this report, the original rural model was statistically valid for spatial transferability as well as spatial-temporal transferability; however, the enriched data did provide an opportunity to enhance the model and simplify the overall equation structure. Consequently, the project team recommends that ODOT use the following enhanced rural arterial model (the equation shown combines equations 11, 12, and 13 introduced in Section 4.3.6).

Equation 15. Enhanced Rural Model (15)

$$Crashes\ for\ 5\ years = (4.0 \times 10^{-3}) \times (AADT^{0.7947}) \times (Segment\ Length^{0.7333}) \times \left(e^{((0.7558 \times Prop.Ind.DW) + (0.0457 \times Total.\#.Clusters))}\right)$$

Where:

AADT = Annual Average Daily Traffic (vehicles per day),
Segment Length = study corridor length (miles),
Prop.of.Ind.DW = proportion of industrial driveways (number of industrial driveways divided by the total number of driveways), and
Total.#.Clusters = number of directional driveway clusters with a 1.5 second travel time.

5.2.1 Use of the Rural Model

The information needed to use the enhanced rural arterial model is summarized as follows:

- Length of the road segment to analyze (in miles),

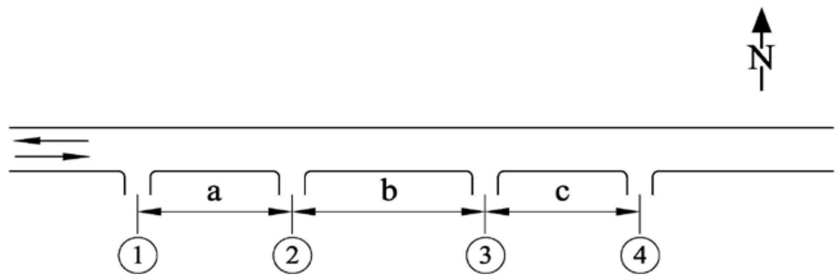
- AADT for the segment,
- Speed limit (model restricted to facilities with speed limits of 50 or 55 mph only),
- Total number of driveways in the segment, regardless of land use type,
- Total number of driveways dedicated to Industrial land use, and
- Total number of clusters of closely located driveways. A ‘cluster of closely located driveways’ is defined as the set of driveways such that the distance between two consecutive driveways on one side of the street can be traveled in 1.5 seconds or less. This distance is 121 feet and 110 feet for roads with speed limits of 55 mph and 50 mph respectively. Figure 5.2 demonstrates how to calculate driveway clusters for a variety of conditions (this is a graphic originally included in the SPR 720 Final Report).

The predictive procedure for rural arterial non-intersection crashes can then be performed by applying Equation 15. This equation consists of the following components:

1. *Effect of Exposure Factor* based on Equation 12.
2. *Roadside & Driveway Effect* using Equation 13.
3. Multiply the results as demonstrated in Equation 11 to obtain the expected number of crashes for the study segment during a five year period.

Note that the Roadway Effect is no longer required for the rural model (as it was required in the original model).

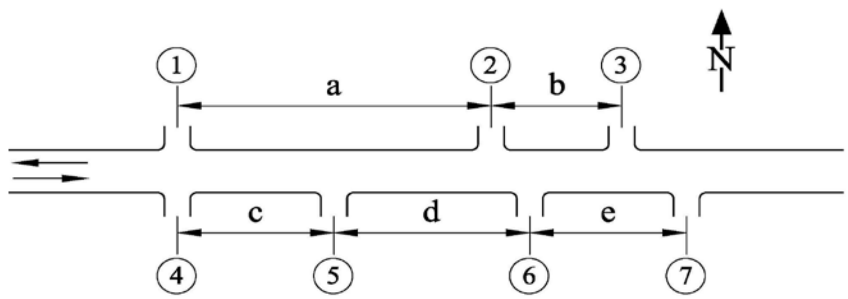
Case I. Driveways on Only One Side of the Road (based on 1.5 second spacing)



Example Calculation of Directional Clusters for Various Spacings:

a (ft)	b (ft)	c (ft)	# Drives	50 mph Speed Limit			55 mph Speed Limit		
				# Cluster	Clusters Noted	Comment	# Cluster	Clusters Noted	Comment
140	200	160	4	4	1, 2, 3, 4	All > 110’	4	1, 2, 3, 4	All > 121’
115	200	160	4	4	1, 2, 3, 4	All > 110’	3	1-2, 3, 4	a < 121’
80	115	125	4	3	1-2, 3, 4	a < 110’	2	1-2-3, 4	a & b < 110’
80	105	115	4	2	1-2-3, 4	a & b < 110’	1	1-2-3-4	All < 121’
80	105	90	4	1	1-2-3-4	All < 110’	1	1-2-3-4	All < 121’

Case II. Driveways on Both Sides of the Road (based on 1.5 second spacing)



Example Calculation of Directional Clusters for Various Spacings:

a (ft)	b (ft)	c (ft)	d (ft)	e (ft)	# Drives	50 mph Speed Limit			55 mph Speed Limit		
						# Cluster	Clusters Noted	Comment	# Cluster	Clusters Noted	Comment
200	125	130	125	150	7	7	WB: 1,2,3 EB: 4,5,6,7	All > 110’	7	WB: 1,2,3 EB: 4,5,6,7	All > 121’
200	115	130	125	150	7	7	WB: 1,2,3 EB: 4,5,6,7	All > 110’	6	WB: 1,2-3 EB: 4,5,6,7	b < 121’
200	105	120	125	150	7	6	WB: 1,2-3 EB: 4,5,6,7	b < 110’	5	WB: 1,2-3 EB: 4-5,6,7	b & c < 121’
200	105	105	105	150	7	4	WB: 1,2-3 EB: 4-5-6,7	b, c, & d < 110’	4	WB: 1,2-3 EB: 4-5-6,7	b, c, & d < 121’
120	90	90	95	105	7	3	WB: 1,2-3 EB:4-5-6-7	b, c, d, & e < 110’	2	WB: 1-2-3 EB:4-5-6-7	All < 121’
105	90	90	95	105	7	2	WB: 1-2-3 EB:4-5-6-7	All < 110’	2	WB: 1-2-3 EB:4-5-6-7	All < 121’

Figure 5.2: Example Calculations for Rural Directional Driveway Clusters

5.2.2 Example Use of the Rural Model

This section includes an example problem that demonstrates the use of the enhanced rural arterial model. For this demonstration, the study site is located on US 20, between Corvallis and Newport, as illustrated in Figure 5.3. The required information for this site is depicted in Table 5.2.

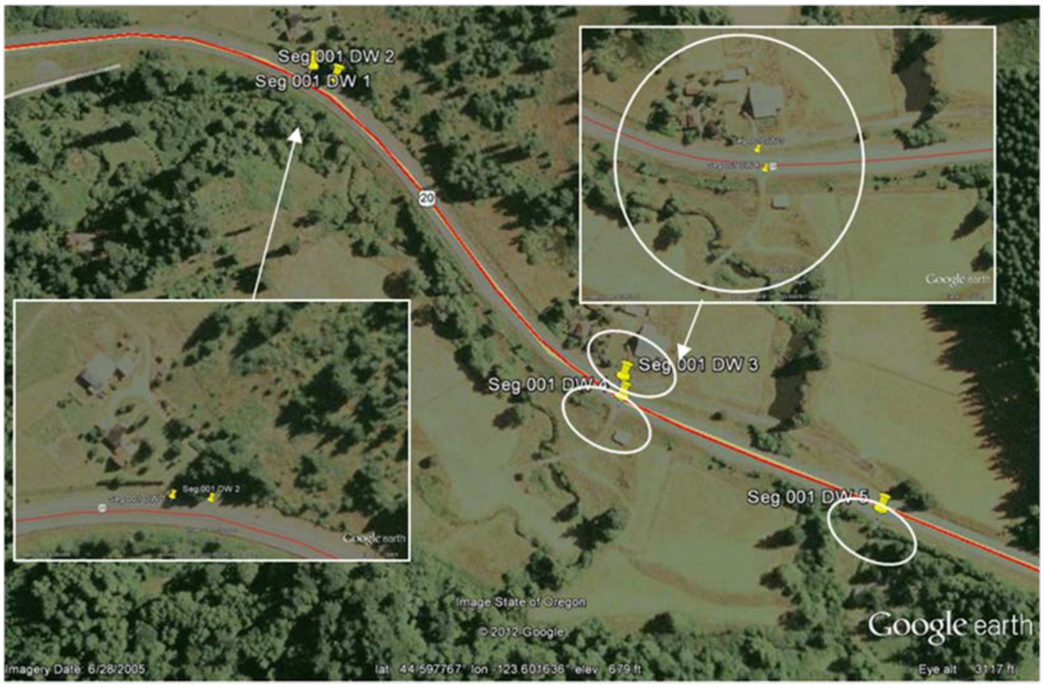


Figure 5.3: Sample Site, Corvallis-Newport, Oregon

Table 5.2: Sample Input for Rural Example Problem for Corvallis-Newport, Oregon

Rural Segment Features	Characteristics
Segment length (MP 33.78 to 34.34)	0.56 miles
AADT	4,940 vpd
Speed limit	55
Total driveways in segment	5
Proportion of industrial driveways	0.00
Number of clusters of closely located driveways (such that the maximum distance between two driveways in a cluster is 121 ft for the 55 mph speed of this road)	4

Since there are no industrial driveways in this segment, the proportion of industrial driveways is then: 0 ÷ 5 = 0.00.

As shown in Figure 5.3, there are 5 different driveways within the segment. While driveways DW1 and DW2 constitute a cluster because they are both on the same side of the road and located approximately 75 feet from each other, driveways DW4 and DW5 are not combined into a driveway cluster. This is because they are on opposite sides of the road. So, except for driveways DW1 and DW2, each driveway in this segment is at least 122 feet from each neighbor driveway at the same side of the road. The number of clusters is then 4 (i.e., DW1+DW2, DW3, DW4, and DW5).

The predicted number of non-intersection crashes during a 5 year period can then be calculated using the procedure below.

Step 1: Compute the *Effect of Exposure Factor* using Equation 12

Baseline Exposure Values $= (4.0 \times 10^{-3}) \times (AADT^{0.7947}) \times (\text{Segment Length}^{0.7333})$

Baseline Exposure Values $= (4.0 \times 10^{-3}) \times (4940^{0.7947}) \times (0.56^{0.7333}) = 2.25$

Step 2: Compute the *Effect of Roadside & Driveways* using Equation 13

Effect from Roadside & Driveways $= \exp[(0.7558 \times \text{Prop.Ind.DW}) + (0.0457 \times \text{Total.\#.Clusters}]$

Effect from Roadside & Driveways $= \exp[(0.7558 \times 0.0) + (0.0457 \times 4)] = 1.20$

Step 3: Obtain the predicted number of crashes in 5 years for the segment by multiplying the values obtained in Steps 1 and 2

Predicted Number of Crashes = (Baseline Exposure Values) x *(Effect from Roadside & Driveways)*

Predicted Number of Crashes $= 2.25 \times 1.20 = 2.7$ *predicted crashes in 5 years*

Example problem conclusion:

Based on exposure and roadside characteristics we can predict that over a period of 5 years approximately 3 (rounded from 2.7) non-intersection crashes will occur.

6.0 REFERENCES

American Association of State Highway and Transportation Officials. (2010). *Highway Safety Manual*. Washington, D.C.

Dixon, K., R. Avelar, L. Brown, M. Mecham, and I. vanSchalkwyk. (2012). *Quantifying Safety Performance of Driveways on State Highways.* Report No. FHWA-OR-RD-13-02. Oregon Department of Transportation, Salem, Oregon.

Page Intentionally Blank

7.0 APPENDIX

This appendix contains supplemental tables as well as a summary of the proposed process for collecting driveway data.

7.1 SUPPLEMENTAL TABLES

Table 7.1: Abbreviations and Acronym Definitions

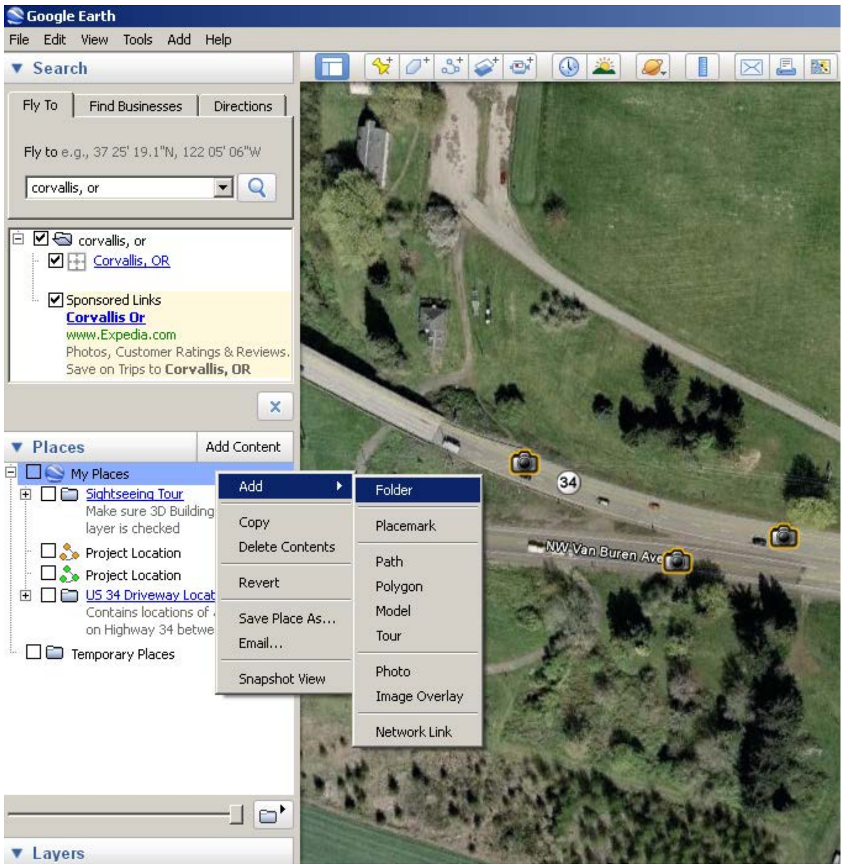
Acronym	Definition
AASHTO	American Association of State Highway and Transportation Officials
AADT	Average Annual Daily Traffic
CMF	Crash Modification Factor (or Function)
GIS	Geographic Information System
HSM	Highway Safety Manual
ODOT	Oregon Department of Transportation
SPF	Safety Performance Function
TAC	Technical Advisory Committee
TWLTL	Two-Way Left-Turn Lane

Table 7.2: Summary of Original Study Corridors

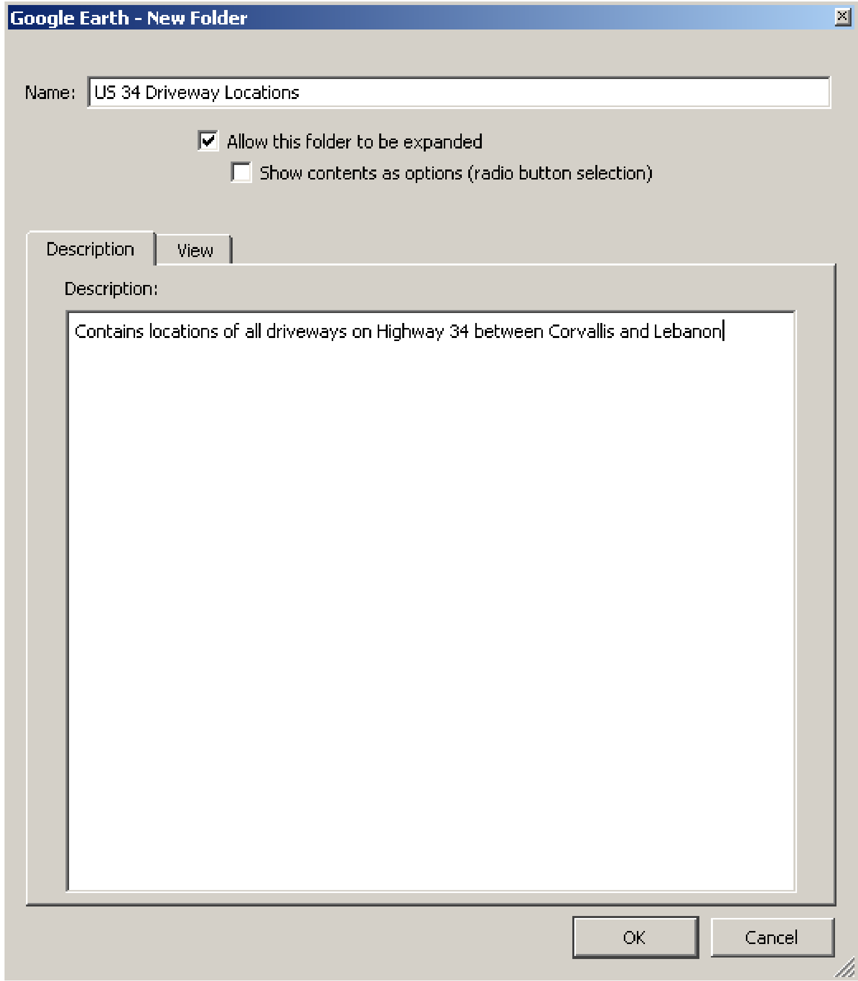
Urban Study Corridors				*Rural Study Corridors*			
Hwy. Name	**Hwy. No.**	**Begin MP**	**End MP**	**Hwy. Name**	**Hwy. No.**	**Begin MP**	**End MP**
Ochoco	041	20.09	20.22	Corvallis-Newport	033	33.46	34.34
Pacific Highway West	091	2.71	3.07	Willamina-Salem	030	4.18	4.65
Rogue Valley	063	14.68	14.89	Florence-Eugene	062	5.79	7.79
Pacific Highway East	081	42.78	43.19	Redwood	025	12.81	14.16
Olds Ferry-Ontario	455	29.55	29.79	Coos Bay-Roseburg	035	22.58	23.22
Pacific Highway West	091	75.7	75.92	Santiam	016	39.92	41.36
Clackamas-Boring	174	3.78	4.2	Santiam	016	73.92	74.58
Lower Columbia River	092	5.3	6.22	Florence-Eugene	062	23.33	24
La Grande-Baker	066	52.76	53.14	John Day	005	133.85	136.85
Klamath Falls-Malin	050	0.08	0.23	Ochoco	041	34.83	36.83
Clackamas-Boring	174	1.88	2.03	Central Oregon	007	87.44	89.44
Albany-Junction City	058	1.45	1.71	Klamath Falls-Lakeview	020	35.87	37.87
Lower Columbia River	092	27.88	28.18	Lake of the Woods	270	9.62	11.62
Albany-Junction City	058	0.75	0.88	Klamath Falls-Lakeview	020	41.74	43.74
Northeast Portland	123	13.04	13.26	Central Oregon	007	200.73	202.73
Pacific Highway West	091	13.32	19.43	Central Oregon	007	177.93	178.69
The Dalles-California	004	122.84	123.19	Central Oregon	007	241.83	243.3
Pacific Highway East	081	8.1	8.21	Sunset	047	27.75	28.52
The Dalles-California	004	92.46	92.58	Santiam	016	84.51	85.47
Clackamas	171	6.21	6.57	John Day-Burns	048	51.31	52.21
Tualatin Valley	029	11.72	11.96	Coos Bay-Roseburg	035	38.69	39.03
Umatilla-Stanfield	054	6.62	6.94	The Dalles-California	004	192.35	193.04
The Dalles-California	004	120.28	120.4	Sunset	047	10.17	11.82
Warm Springs	053	115.86	116.15	North Santiam	162	54.07	54.54
Jacksonville	272	0.96	1.15	Clear Lake-Belknap Springs	215	3.34	4.61
Crater Lake	022	3.26	3.56	Central Oregon	007	147.94	149.94
Albany-Junction City	058	1.15	1.25	John Day-Burns	048	36.79	38.79
Lower Columbia River	092	48.13	48.38	Central Oregon	007	224.42	226.42
Northeast Portland	123	14.20	14.30	Oregon Coast	009	247.54	248.97
South Klamath Falls	424	0.64	1.56	North Santiam	162	80.87	81.51
McKenzie	015	4.30	4.41	McKenzie	015	20.3	21.36
The Dalles-California	004	92.58	92.68	Sunset	047	4.7	6.7
Pacific Highway West	091	85.55	85.84	Oregon-Washington	008	9.07	11.05
Pacific Highway West	091	16.66	16.96	Florence-Eugene	062	28.02	29.66
Lake of the Woods	270	1.03	2.28	Lakeview-Burns	049	78.18	79.28
The Dalles-California	004	166.78	167.26	Oregon Coast	009	25.72	26.21
The Dalles-California	004	132.19	133.07	Central Oregon	007	136.31	138.3
The Dallas-Rickreall	189	0.00	0.22	John Day	005	271.51	272.11
Salmon River	039	44.61	45.76	Central Oregon	007	110.43	112.43
Lower Columbia River	092	28.25	28.37	Umpqua	045	53.91	54.68

7.2 Process for Collecting Driveway Data using Google Earth

1. Open Google Earth
2. In the navigation side bar on the left, there should be a "Places" menu.
3. Highlight the "My Places" line (the top-most line).
4. Right-click on the "My Places" line and add a new folder (Add->Folder).

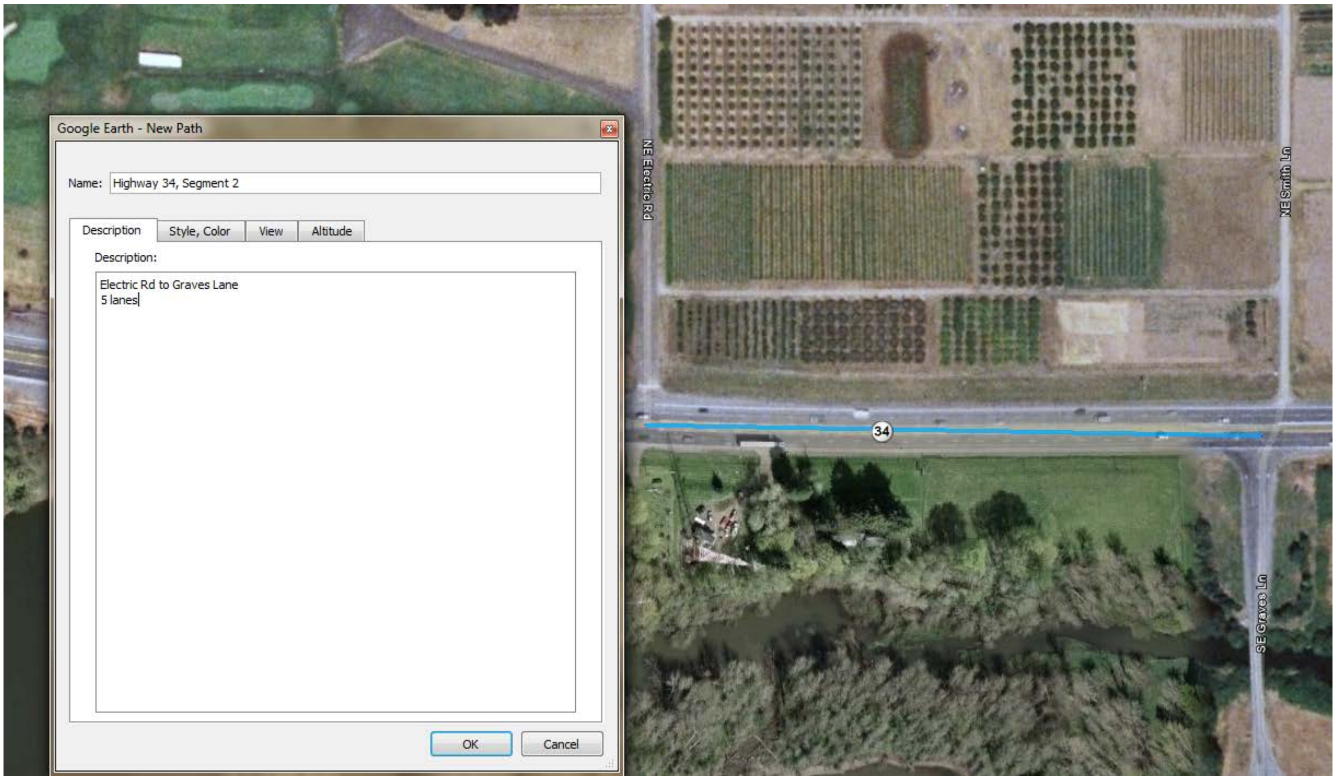


5. Name the folder and provide a brief description of the driveways to be collected.



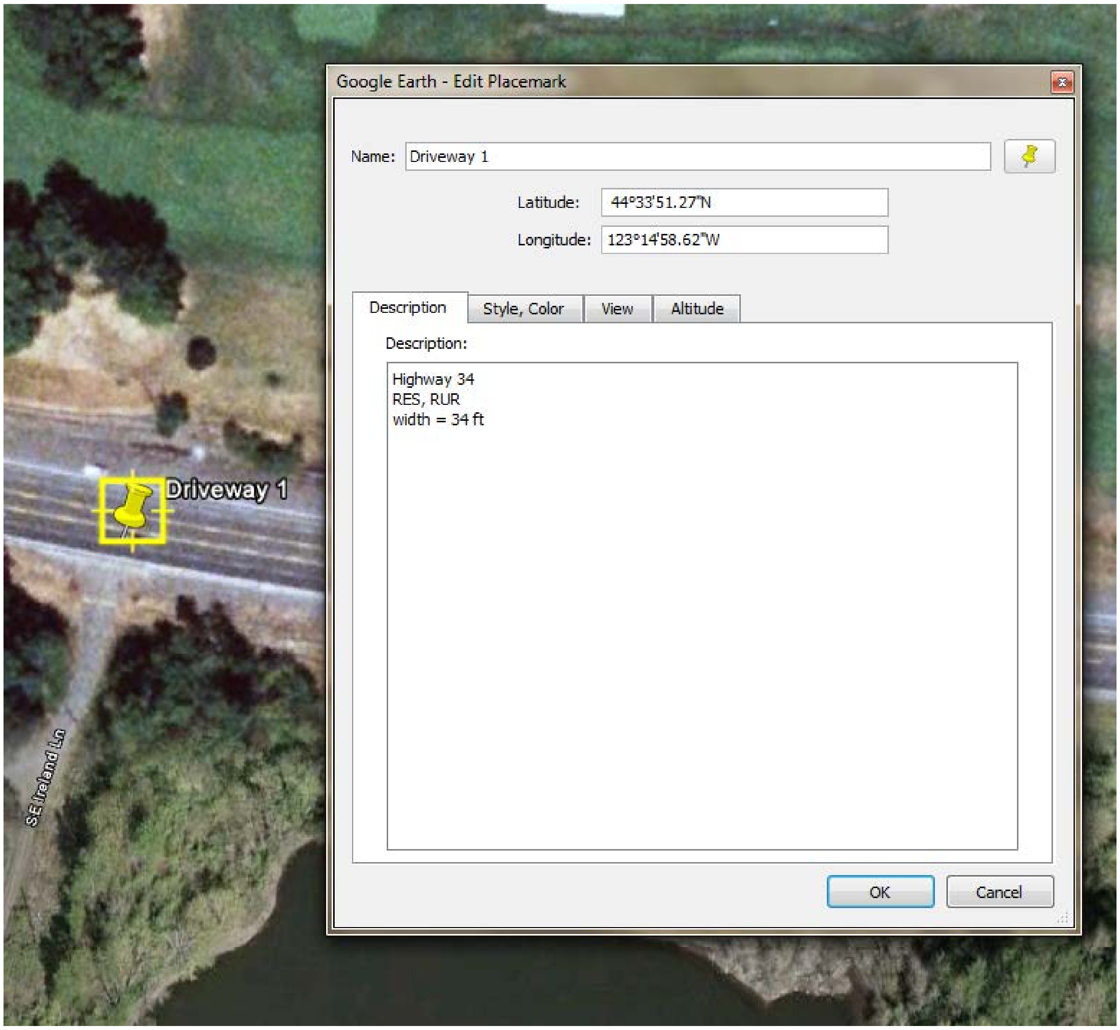
The newly created folder will appear at the bottom of the "My Places" list.

6. Locate the desired Highway/starting point on the map. For this example, we started in Corvallis.
7. The first thing to add to the map is a path of the current segment. The path will define the endpoints of the homogeneous segment, and will typically have a length less than two miles.
8. Make sure the newly created folder is selected (highlighted) in the "Places" menu.
9. Press CTRL+SHIFT+T or press the path icon in the menu bar at the top of the screen. This will open the path dialog box. With the dialog box open, draw a path along the roadway between two pre-set locations, usually designated by an intersection or driveway. Name the segment with the highway name and segment number. In the description area, include the endpoint locations and number of lanes.



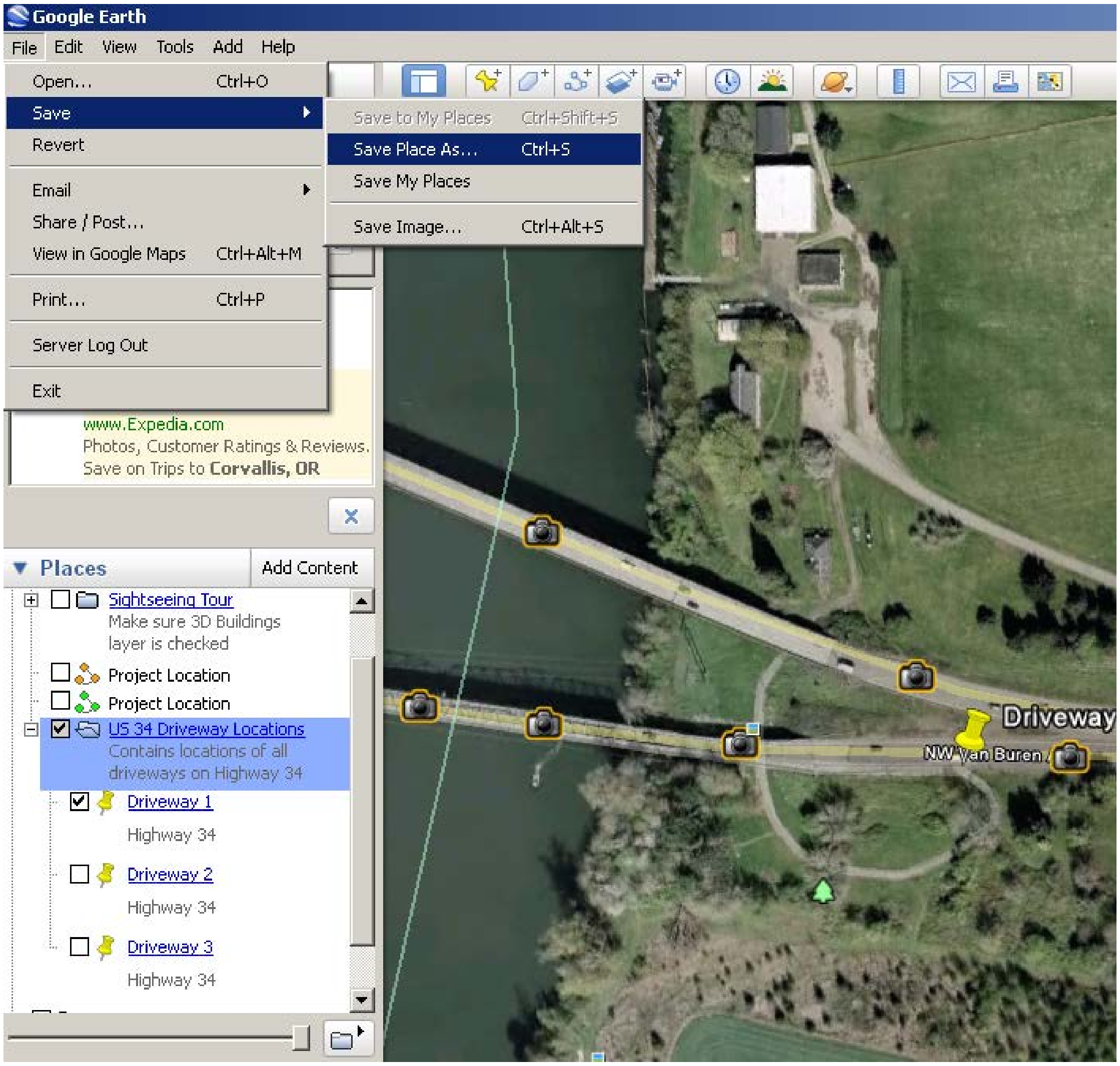
The newly created path will appear on the map and will also be listed under the folder created in the “My Places” menu.

10. Then, locate the first driveway to be tagged.
11. Press CTRL+SHIFT+P or press the push-pin icon in the menu at the top of the screen. This will insert a push-pin icon (place mark) onto the map and open up a new dialog box.
12. Move the push-pin icon to the desired location, and then enter descriptive information into the dialog box. In this example, we have placed the icon at the center of the driveway at approximately the fog line. We’ve named the place mark “Driveway 1” and denoted the highway number in the description field. Visually inspect the type of development served by the driveway, and denote the land use as residential (RES), commercial (COM), rural (RUR), or industrial (IND). Lastly, measure the width of the driveway using the ruler tool and enter the width into the description box as well.

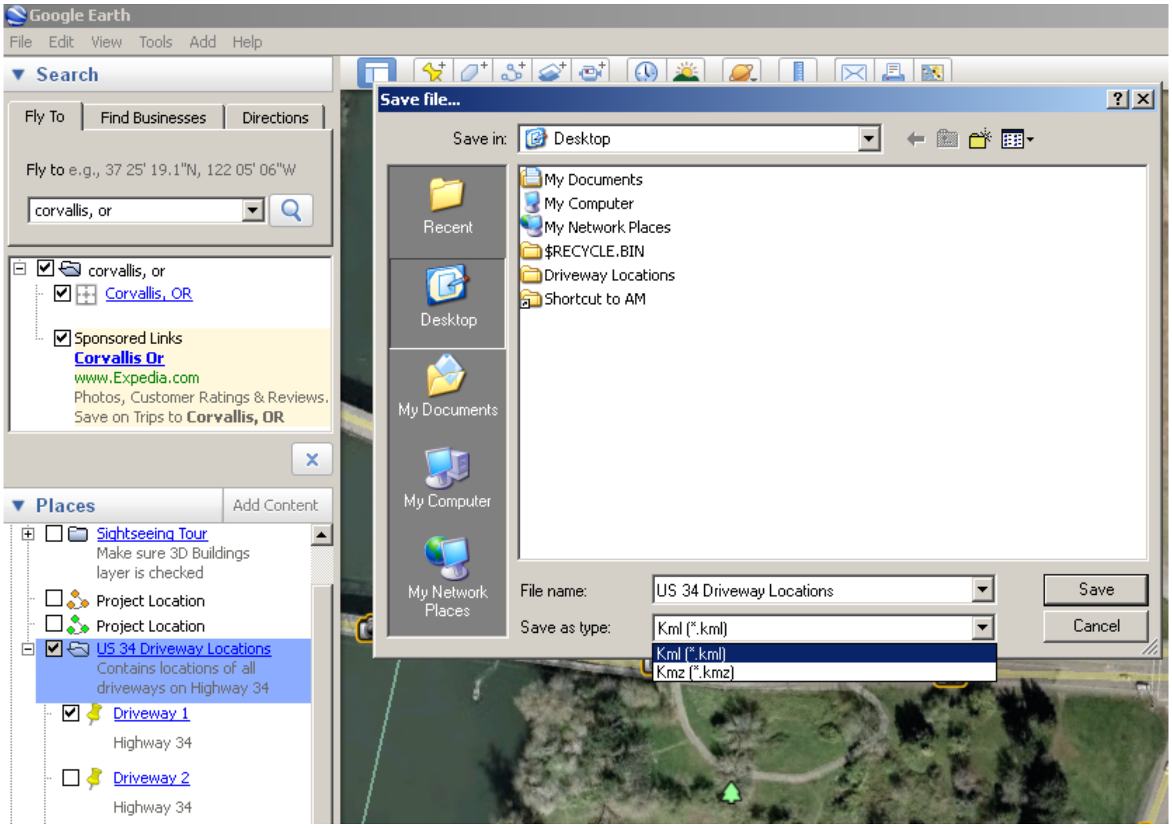


The newly created place mark will appear on the map and will also be listed under the folder created in the "My Places" menu.

13. Repeat steps 6 through 10 for all driveways along the same route/segment.
14. After all driveways have been located and marked, once again ensure that the main folder containing all place marks and the segment path is selected in the "My Places" menu.
15. Then, save the place marks in a .kml format. File->Save->Save Place As.



Make sure that you select the “Kml” file type, NOT the “Kmz” file type!



16. Next, locate the file wherever you saved it. Manually change the file extension from ".kml" to ".xml". This will allow the file to be opened in Excel.
17. Lastly, open the file from within Excel. (Simply double-clicking the file will open it in an html/web browser format.)

The Excel spreadsheet contains all information for each place mark. The first set of rows contains display information for the push-pin icons, and can be ignored. The relevant information (coordinates, location name, and location description) can be found in columns L-Z.

Repeat these steps for each highway and segment.

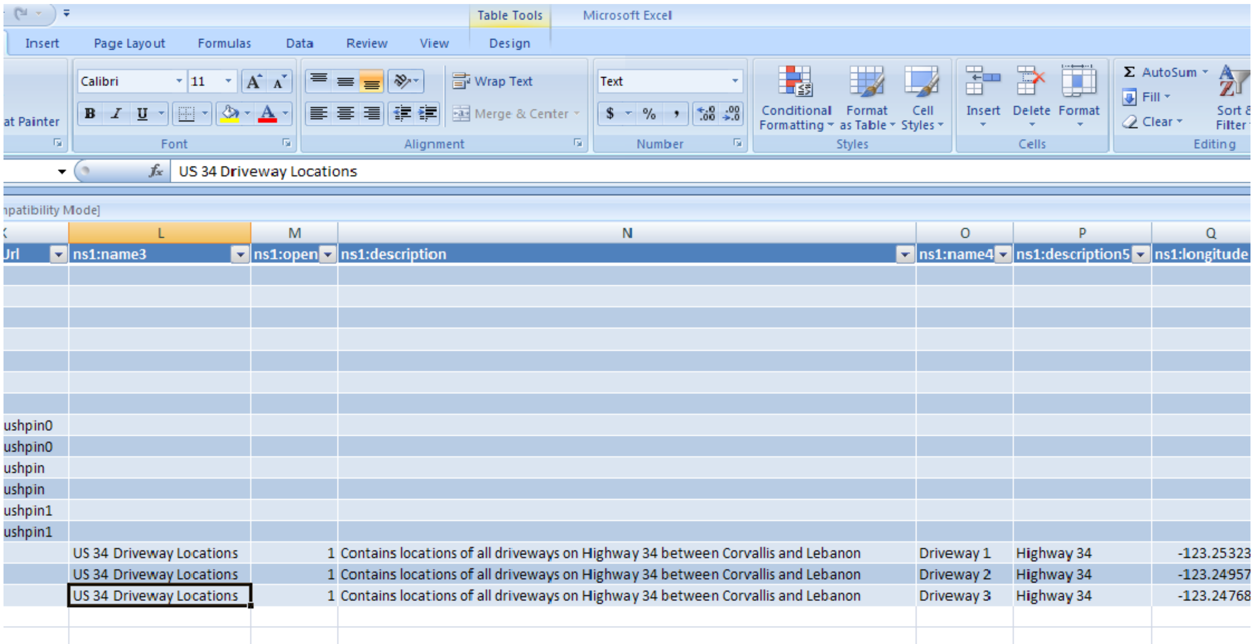


Table Tools
Microsoft Excel
Insert
Page Layout
Formulas
Data
Review
View
Design
at Painter
Calibri
11
Wrap Text
Merge & Center
Text
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
AutoSum
Fill
Clear
Sort & Filter
Font
Alignment
Number
Styles
Cells
Editing
US 34 Driveway Locations
npatibility Mode]
L
M
N
O
P
Q
Url
ns1:name3
ns1:open
ns1:description
ns1:name4
ns1:description5
ns1:longitude
ushpin0
ushpin0
ushpin
ushpin
ushpin1
ushpin1
US 34 Driveway Locations 1 Contains locations of all driveways on Highway 34 between Corvallis and Lebanon Driveway 1 Highway 34 -123.25323
US 34 Driveway Locations 1 Contains locations of all driveways on Highway 34 between Corvallis and Lebanon Driveway 2 Highway 34 -123.24957
US 34 Driveway Locations 1 Contains locations of all driveways on Highway 34 between Corvallis and Lebanon Driveway 3 Highway 34 -123.24768

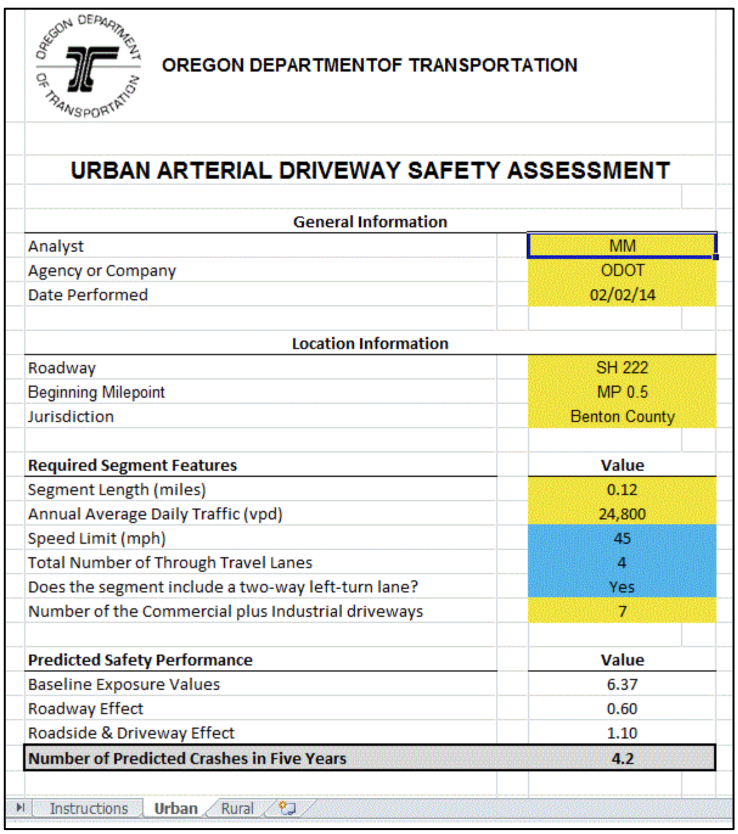
Page Intentionally Blank

7.3 SPREADSHEET USER INSTRUCTIONS

As a companion to the validation report, the project team developed a spreadsheet tool (named ODOT Driveway Safety Models.xls) to automate the calculation of the predicted number of non-intersection segment crashes on urban and rural arterials in Oregon. As noted in the body of the report (see Table 4.9 and Table 4.19), the urban and rural database included minimum and maximum values. The resulting models, therefore, should include site characteristics that are consistent with these thresholds. The spreadsheet tool, therefore, includes a series of error checks to confirm that the user does not input values beyond the limits of the model (which would result in extrapolating values for site conditions not represented in the study). The tool includes four worksheets; however, one of them (named Construction) is hidden and includes the content for pull-down menus. The three remaining worksheets are titled:

- Instructions,
- Urban, and
- Rural.

The following figure shows the urban worksheet and demonstrates these three tabs at the base of the drawing. To navigate between worksheets, the user should simply select the tab of choice.


OREGON DEPARTMENTOF TRANSPORTATION

URBAN ARTERIAL DRIVEWAY SAFETY ASSESSMENT

General Information	
Analyst	MM
Agency or Company	ODOT
Date Performed	02/02/14

Location Information	
Roadway	SH 222
Beginning Milepoint	MP 0.5
Jurisdiction	Benton County

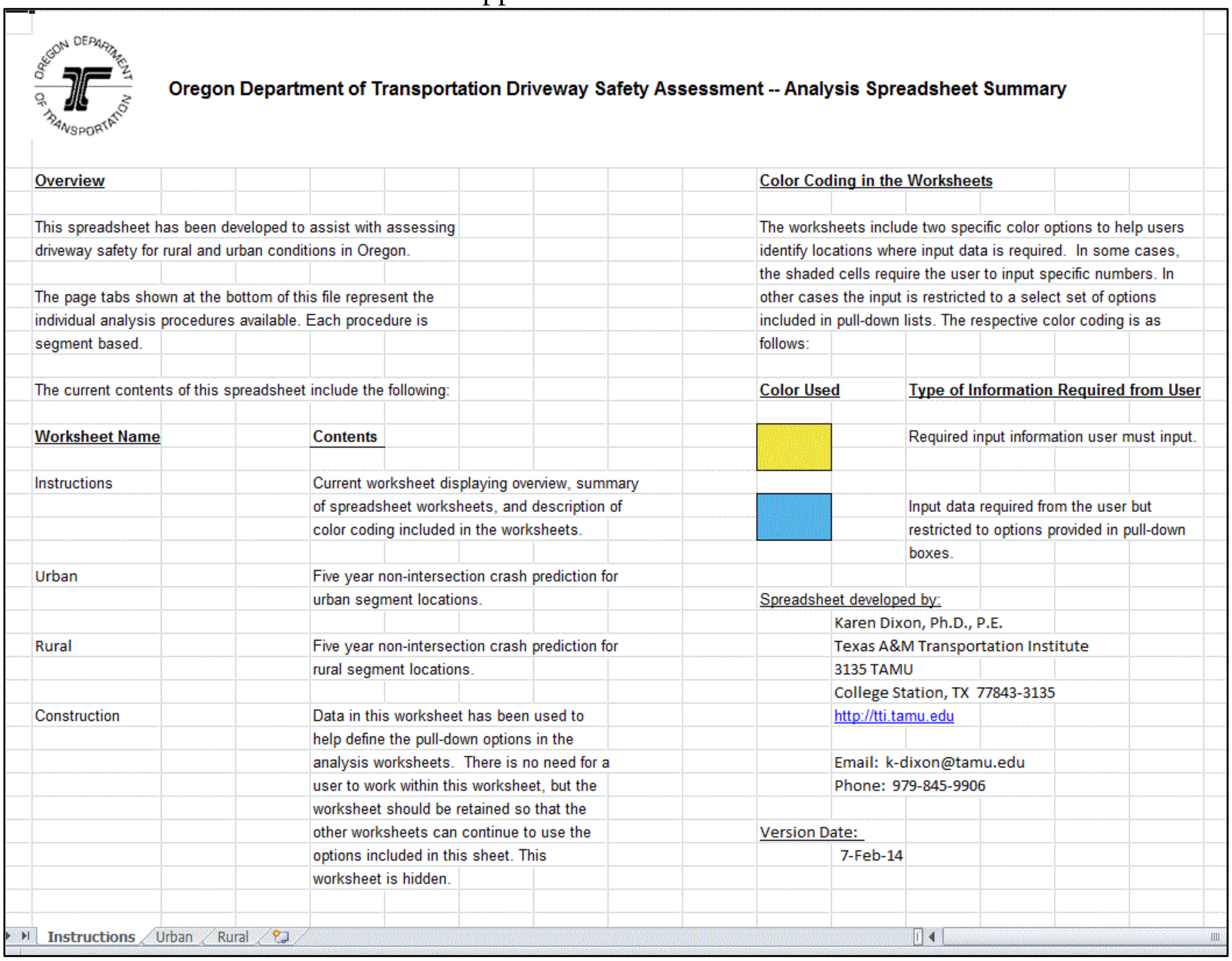
Required Segment Features	Value
Segment Length (miles)	0.12
Annual Average Daily Traffic (vpd)	24,800
Speed Limit (mph)	45
Total Number of Through Travel Lanes	4
Does the segment include a two-way left-turn lane?	Yes
Number of the Commercial plus Industrial driveways	7

Predicted Safety Performance	Value
Baseline Exposure Values	6.37
Roadway Effect	0.60
Roadside & Driveway Effect	1.10
Number of Predicted Crashes in Five Years	**4.2**

Instructions

To begin using the spreadsheet tool, select the Instructions worksheet tab:

The Instructions worksheet will then appear as shown below:



Oregon Department of Transportation Driveway Safety Assessment -- Analysis Spreadsheet Summary

Overview

This spreadsheet has been developed to assist with assessing driveway safety for rural and urban conditions in Oregon.

The page tabs shown at the bottom of this file represent the individual analysis procedures available. Each procedure is segment based.

The current contents of this spreadsheet include the following:

Worksheet Name	Contents
Instructions	Current worksheet displaying overview, summary of spreadsheet worksheets, and description of color coding included in the worksheets.
Urban	Five year non-intersection crash prediction for urban segment locations.
Rural	Five year non-intersection crash prediction for rural segment locations.
Construction	Data in this worksheet has been used to help define the pull-down options in the analysis worksheets. There is no need for a user to work within this worksheet, but the worksheet should be retained so that the other worksheets can continue to use the options included in this sheet. This worksheet is hidden.

Color Coding in the Worksheets

The worksheets include two specific color options to help users identify locations where input data is required. In some cases, the shaded cells require the user to input specific numbers. In other cases the input is restricted to a select set of options included in pull-down lists. The respective color coding is as follows:

Color Used	Type of Information Required from User
	Required input information user must input.
	Input data required from the user but restricted to options provided in pull-down boxes.

Spreadsheet developed by:
Karen Dixon, Ph.D., P.E.
Texas A&M Transportation Institute
3135 TAMU
College Station, TX 77843-3135
http://tti.tamu.edu

Email: k-dixon@tamu.edu
Phone: 979-845-9906

Version Date:
7-Feb-14

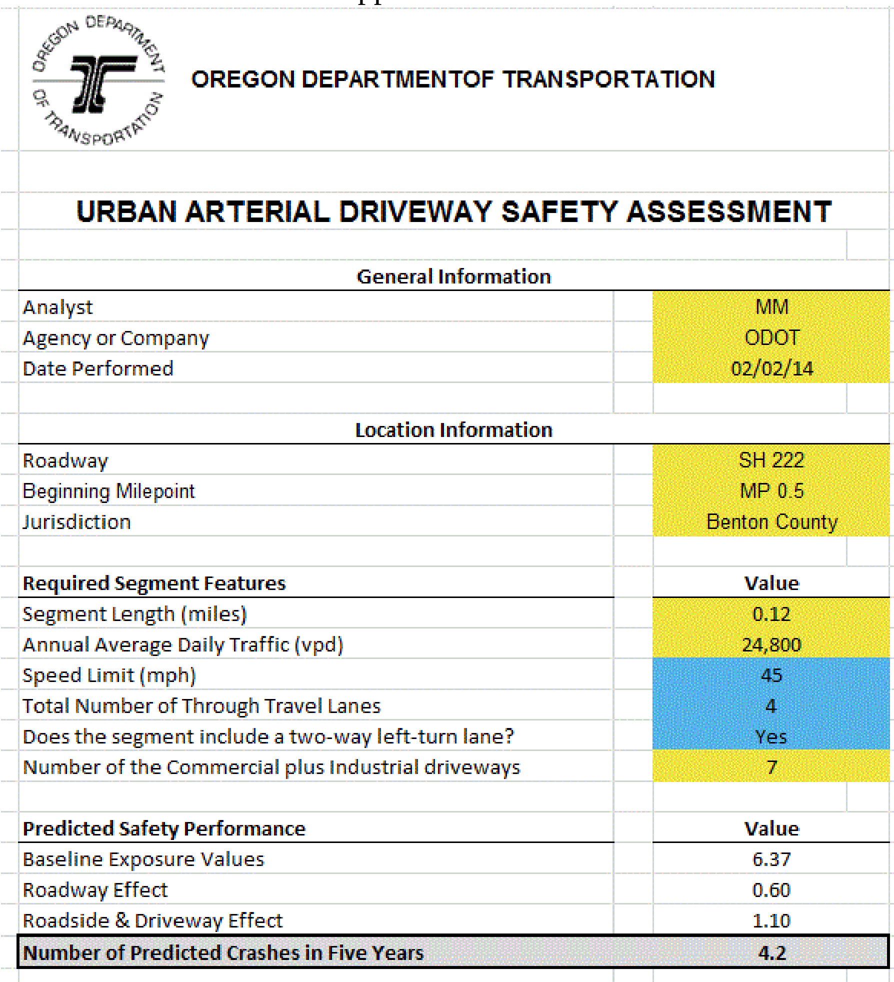
Instructions Urban Rural

This worksheet provides information to introduce the user to the format of the spreadsheet tool, explain the use of color coding, and provide contact information in the event the user has questions about the spreadsheet. Once the user is has reviewed the instruction page, he or she can continue to the appropriate predictive model. For the purposes of these instructions, the Urban worksheet is first reviewed followed by the Rural worksheet; however, there is no particular order that is required.

Urban Worksheet

To begin the working with the Urban worksheet, select the Urban tab as follows:

The Urban worksheet will appear and should look similar to the following:



OREGON DEPARTMENTOF TRANSPORTATION

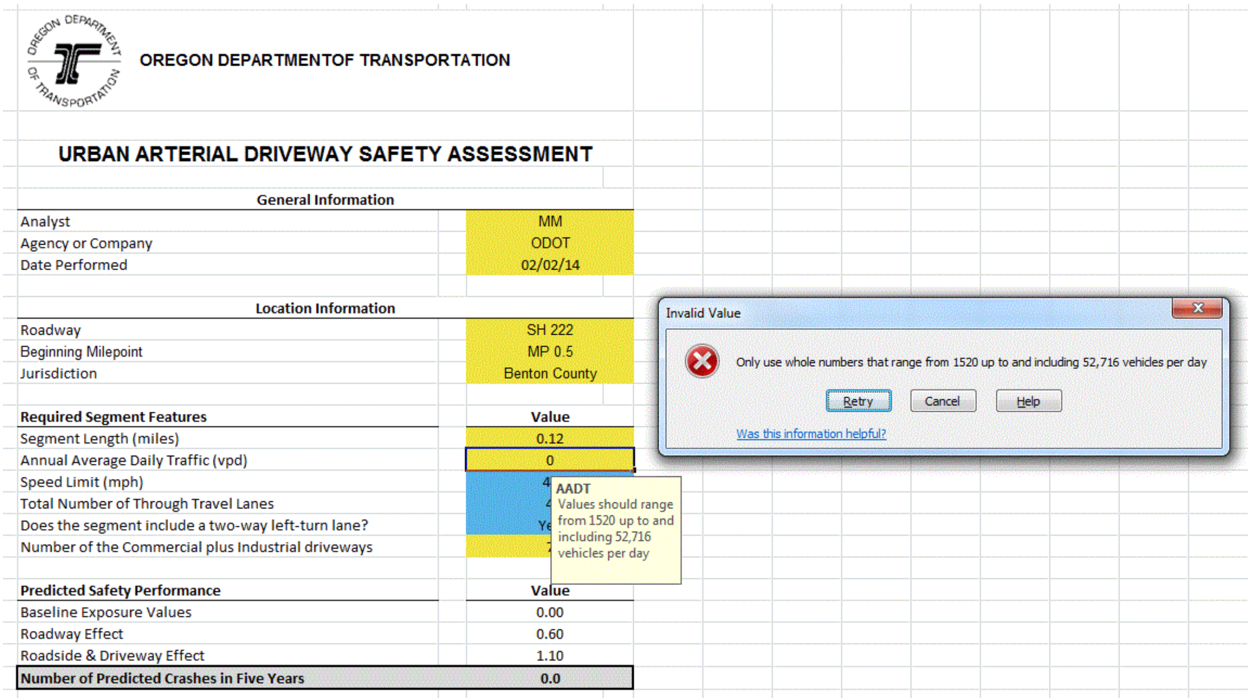
URBAN ARTERIAL DRIVEWAY SAFETY ASSESSMENT

General Information	
Analyst	MM
Agency or Company	ODOT
Date Performed	02/02/14
Location Information	
Roadway	SH 222
Beginning Milepoint	MP 0.5
Jurisdiction	Benton County
Required Segment Features	**Value**
Segment Length (miles)	0.12
Annual Average Daily Traffic (vpd)	24,800
Speed Limit (mph)	45
Total Number of Through Travel Lanes	4
Does the segment include a two-way left-turn lane?	Yes
Number of the Commercial plus Industrial driveways	7
Predicted Safety Performance	**Value**
Baseline Exposure Values	6.37
Roadway Effect	0.60
Roadside & Driveway Effect	1.10
Number of Predicted Crashes in Five Years	**4.2**

Input the information included in the following sections:

- General Information
- Location Information
- Required Segment Features.

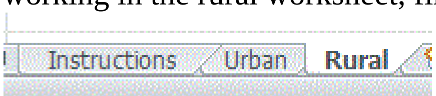
Note that the input boxes that are colored in yellow require manual entry of the values and the input boxes that are colored in blue have drop-down boxes associated with them that will provide menu choices. To reduce the likelihood of a user inserting content that is outside the ranges of the analysis, a series of error checks are included. The following figure depicts an error message that occurred when the user attempted to input an AADT value of zero.



The values included in the bottom section of the worksheet that is titled "Predicted Safety Performance" are then calculated based on the information the user has input into the worksheet. As noted in the above figure, the error temporarily resulted in a value of zero; however, once the error is corrected and additional information is input the Number of Predicted Crashes in Five Years can be calculated.

<u>Rural Worksheet</u>

The rural includes features similar to those observed for the Urban Worksheet. To begin working in the rural worksheet, first select the Rural tab as shown below:



The general information input section of the Rural worksheet looks very similar to that of the Urban except that the input information is specific to that required for use with the enhance rural model. The following figure shows a snapshot of the rural worksheet:

OREGON DEPARTMENTOF TRANSPORTATION

RURAL ARTERIAL DRIVEWAY SAFETY ASSESSMENT

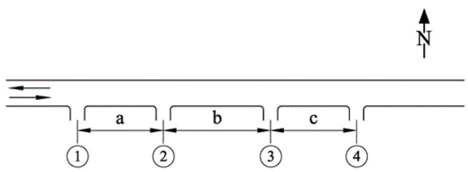
General Information	
Analyst	MM
Agency or Company	ODOT
Date Performed	02/02/14
Location Information	
Roadway	SH 123
Beginning Milepoint	MP 2.1
Jurisdiction	Marion County
Required Segment Features	**Value**
Segment Length (miles)	0.56
Annual Average Daily Traffic (vpd)	4,940
Speed Limit (mph) *[Note: Only 50 or 55 mph]*	55
Total Driveways in Segment	5
Number of the Driveways that are Industrial	0
Total Number of Driveway Clusters (see Figure A)	4
Predicted Safety Performance	**Value**
Baseline Exposure Values	2.25
Roadside & Driveway Effect	1.20
Number of Predicted Crashes in Five Years	**2.7**

In the "Required Segment Features" input section, the user is asked to input the total number of driveway clusters. This item is described in further detail in the body of the report, but is basically calculated based on a 1.5 second threshold. A table appears to the right of the rural input information to help the user understand how to compute these driveway clusters. The figure below shows the table that represents sites where driveways are only on one side of the road as well as locations where driveways are located on both sides of the road.

Figure A. Determining the Number of Directional Driveway Clusters that are Present on Rural Segments

Case I. Driveways on Only One Side of the Road (based on 1.5 second spacing)

Note: If speed limit = 55 mph, a 1.5 second cluster is 121 feet

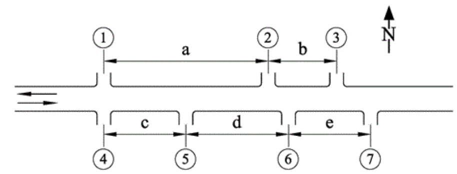


Example Calculation of Directional Clusters for Various Spacings:

a (ft)	b (ft)	c (ft)	Number Driveways	Number Clusters	Clusters as Noted	Comment
140	200	160	4	4	1, 2, 3, 4	All > 121
115	200	160	4	3	1-2, 3, 4	Only a < 121
80	115	125	4	2	1-2-3, 4	a & b < 121
80	105	115	4	1	1-2-3-4	All < 121
80	105	90	4	1	1-2-3-4	All < 121

Case II. Driveways on Both Sides of the Road (based on 1.5 second spacing)

Note: If speed limit = 55 mph, a 1.5 second cluster is 121 feet



Example Calculation of Directional Clusters for Various Spacings:

a (ft)	b (ft)	c (ft)	d (ft)	e (ft)	Number Driveways	Number Clusters	Clusters as Noted	Comment
200	125	130	125	150	7	7	WB: 1, 2, 3; EB: 4, 5, 6, 7	All > 121
200	115	130	125	150	7	6	WB: 1, 2-3; EB: 4, 5, 6, 7	Only b < 121
200	105	120	125	150	7	5	WB: 1, 2-3; EB: 4-5, 6, 7	b & c < 121
200	105	105	105	150	7	4	WB: 1, 2-3; EB: 4-5-6, 7	b, c, & d < 121
120	90	90	95	105	7	2	WB: 1-2-3; EB: 4-5-6-7	All < 121
105	90	90	95	105	7	2	WB: 1-2-3; EB: 4-5-6-7	All < 121

When the user changes the speed limit in the input box, the numbers in this example table will change accordingly.

The rural worksheet includes error checks as well as color coded input in a manner consistent with that shown in the Urban worksheet.

The content included in the "Predicted Safety Performance" section is automatically calculated based on the input information from the user.

7.4 EXAMPLE DRIVEWAY CLUSTER CALCULATIONS

The calculation of the number of driveway clusters is supplemented in this section by example schematics for similar road and driveway configurations but differing speed limits. First, Figure 7.1 through Figure 7.5 each demonstrates how the number of driveway clusters can be determined when driveways are only located on one side of the road. Similarly, Figure 7.6 through Figure 7.11 each depicts the example calculations for driveways located on both sides of the road.

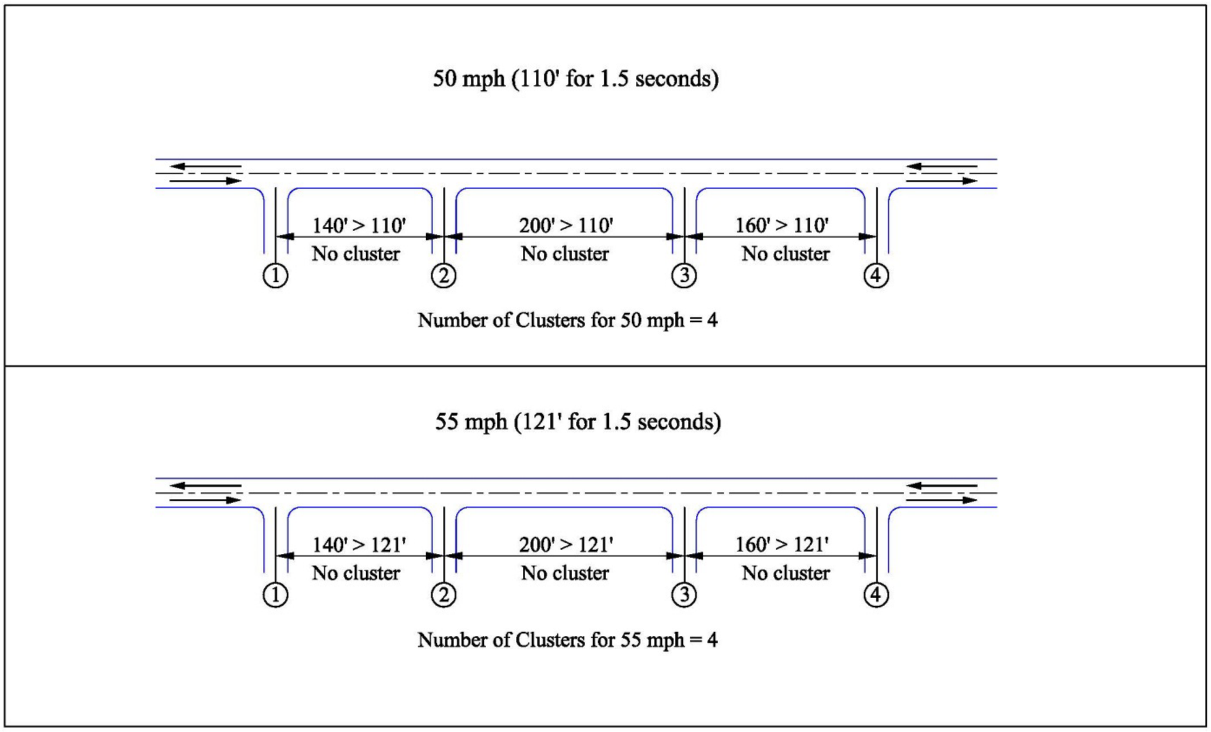


Figure 7.1: Driveways on only one side of road (Example 1)

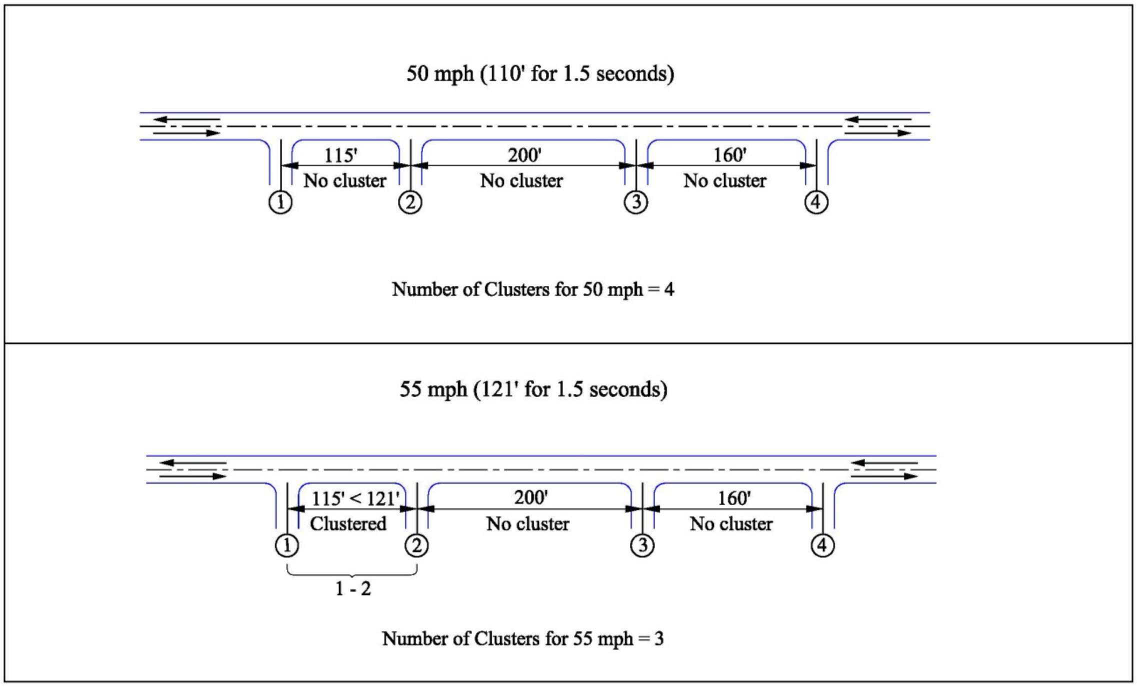


Figure 7.2: Driveways on only one side of road (Example 2)

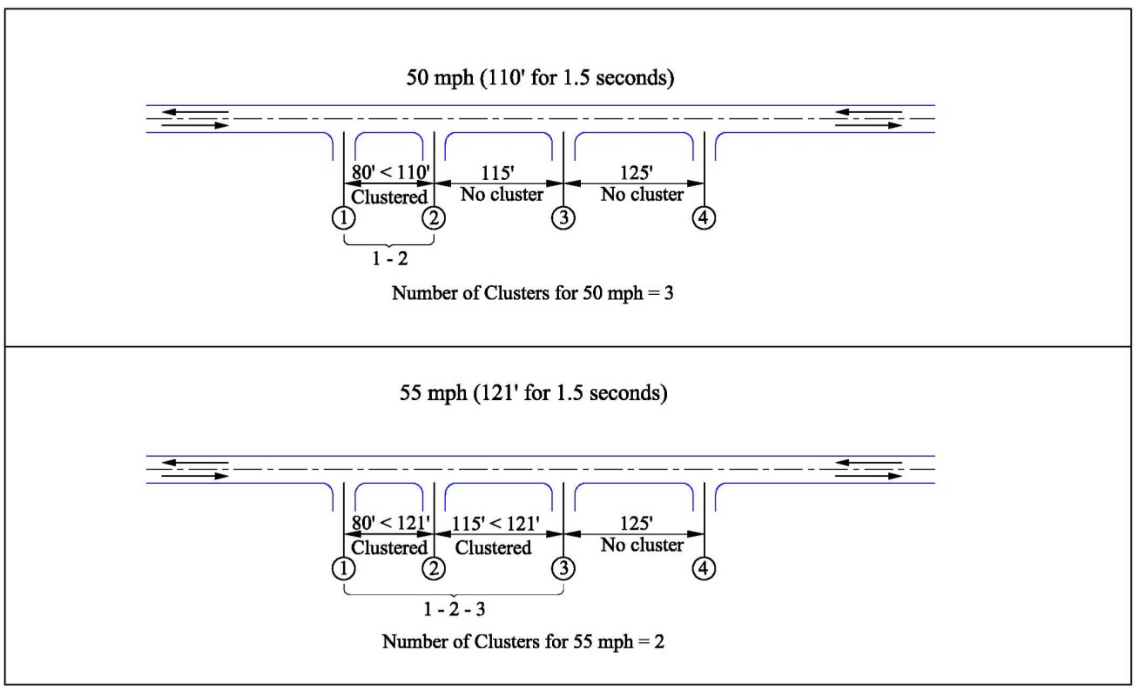


Figure 7.3: Driveways on only one side of road (Example 3)

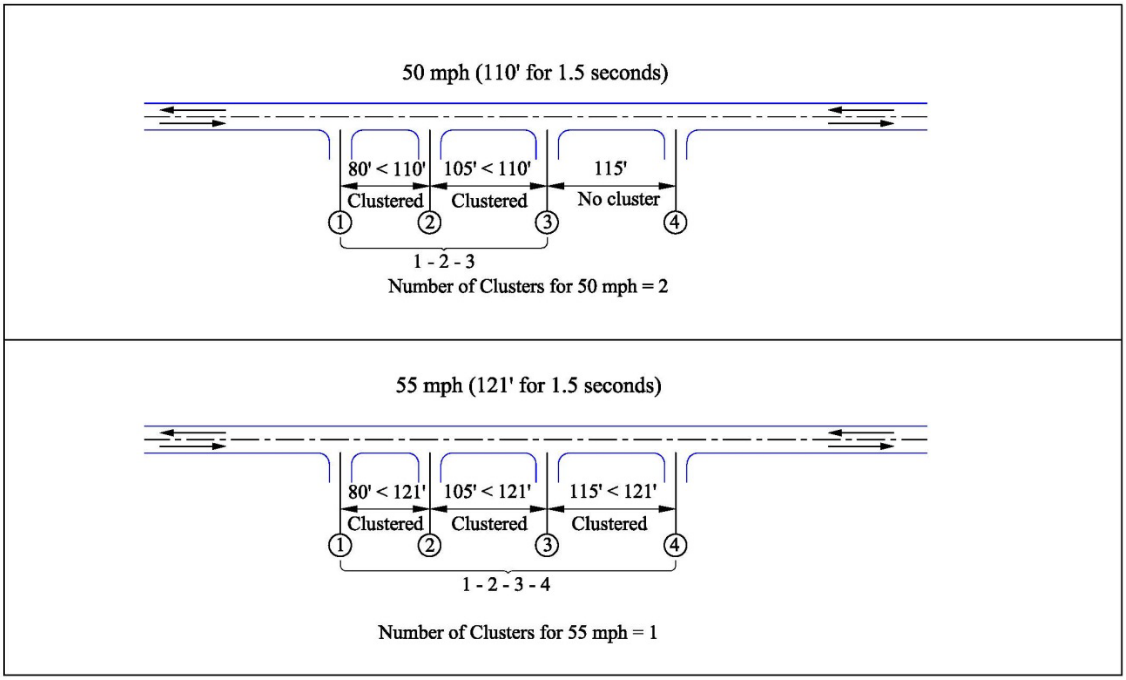


Figure 7.4: Driveways on only one side of road (Example 4)

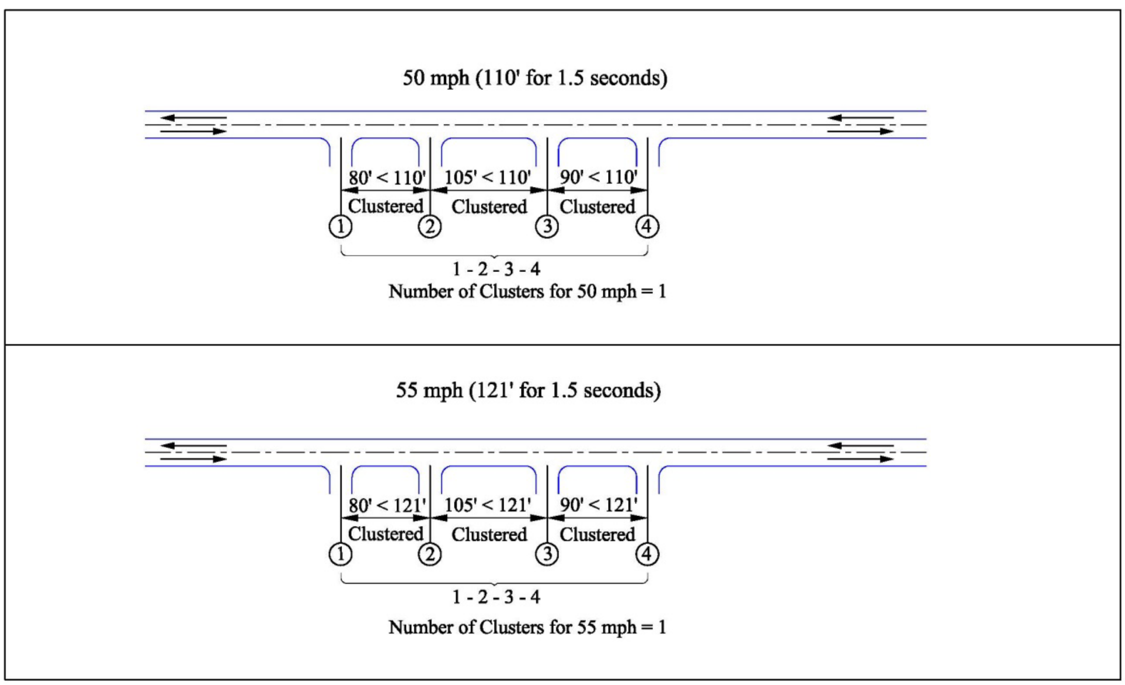


Figure 7.5: Driveways on only one side of road (Example 5)

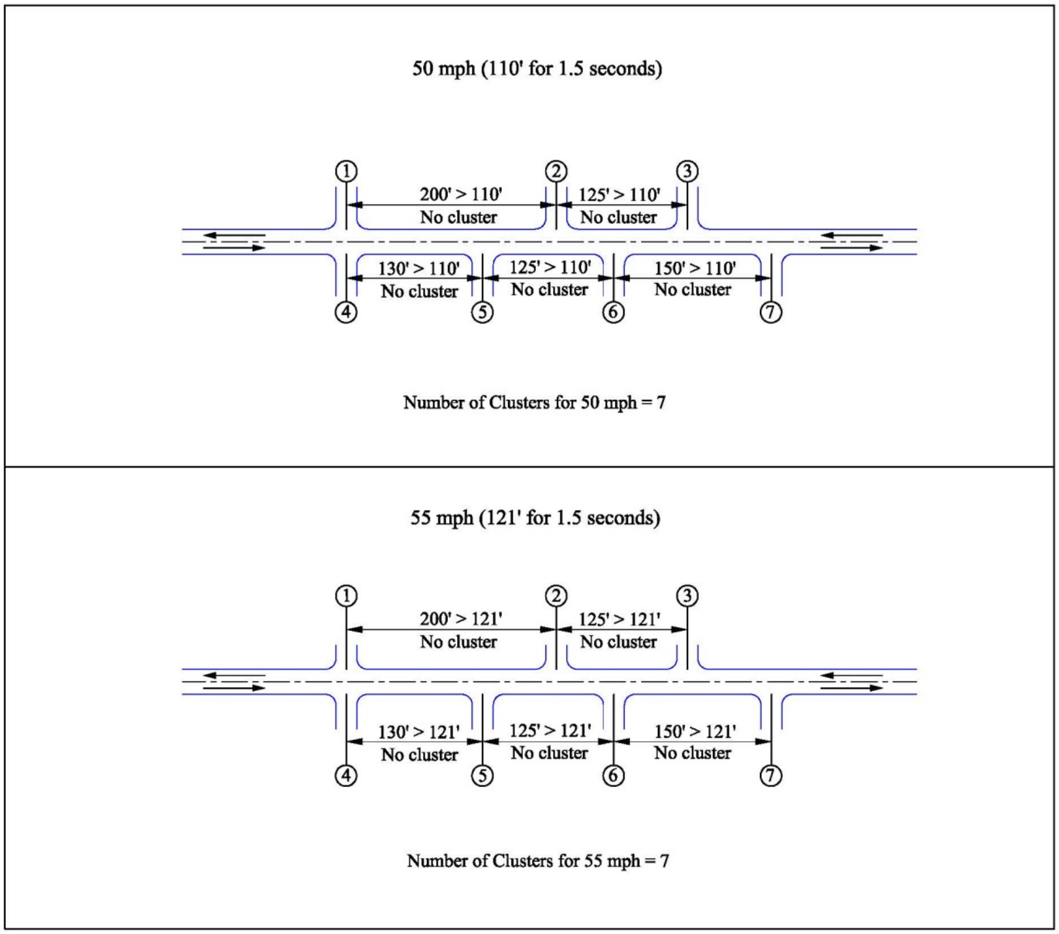


Figure 7.6: Driveways on both sides of road (Example 1)

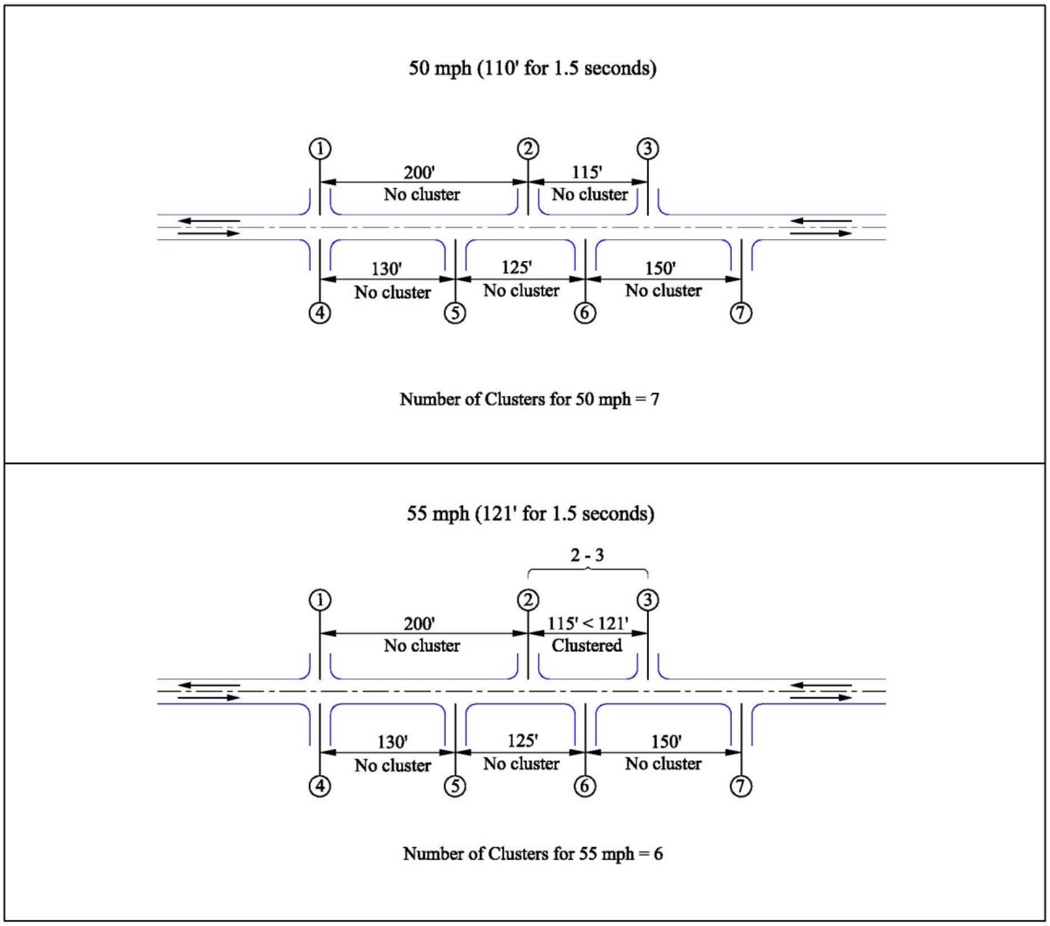


Figure 7.7: Driveways on both sides of road (Example 2)

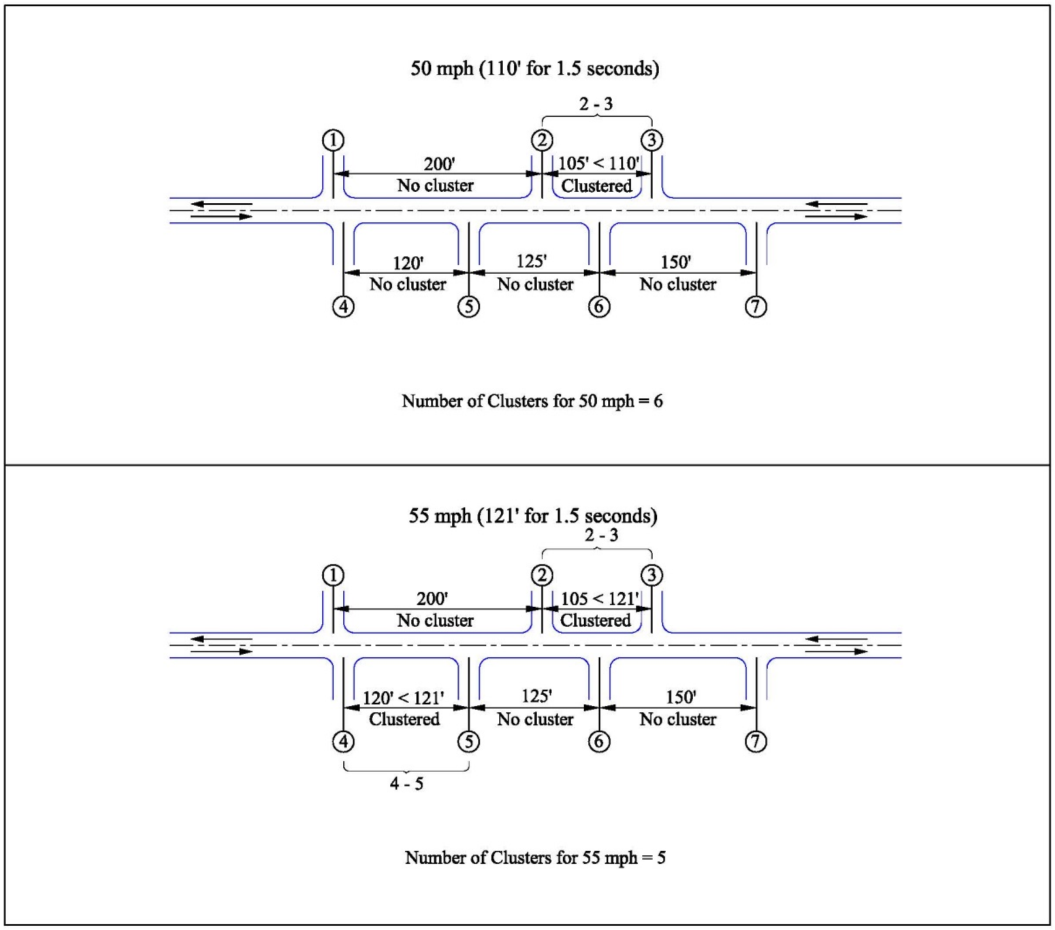


Figure 7.8: Driveways on both sides of road (Example 3)

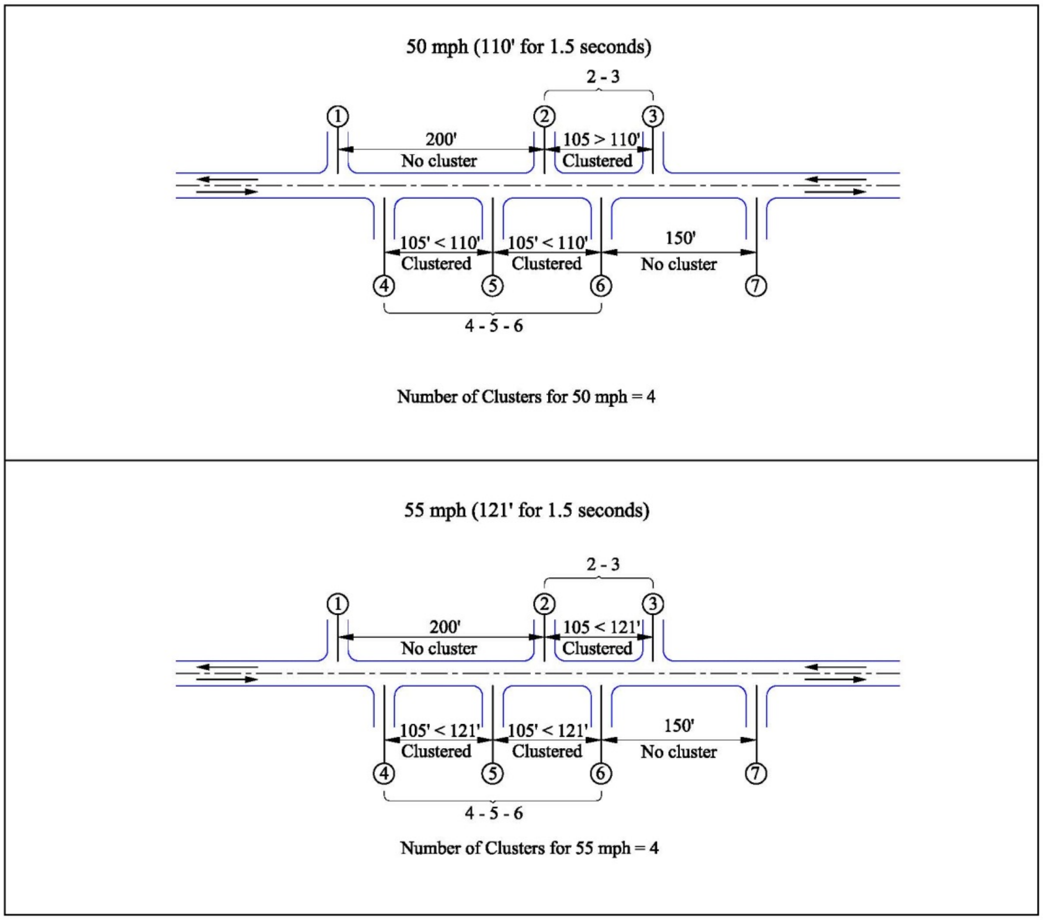


Figure 7.9: Driveways on both sides of road (Example 4)

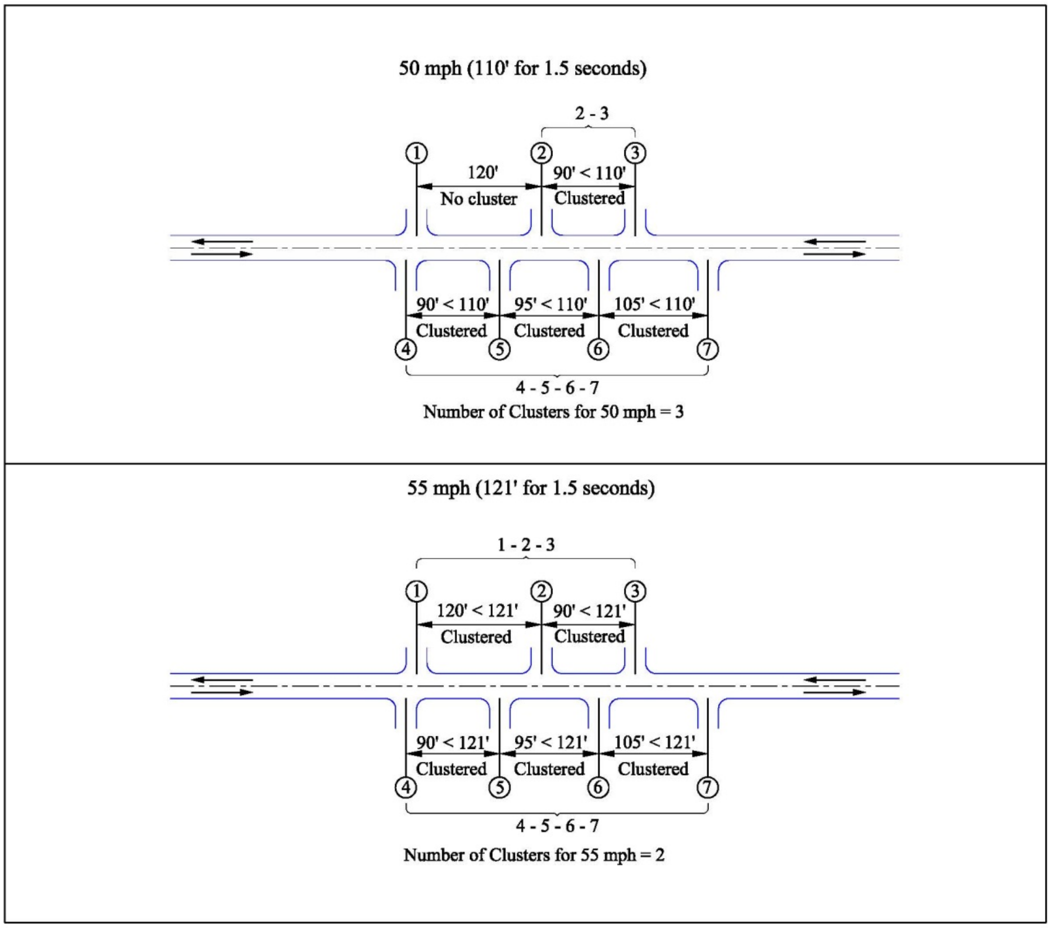


Figure 7.10: Driveways on both sides of road (Example 5)

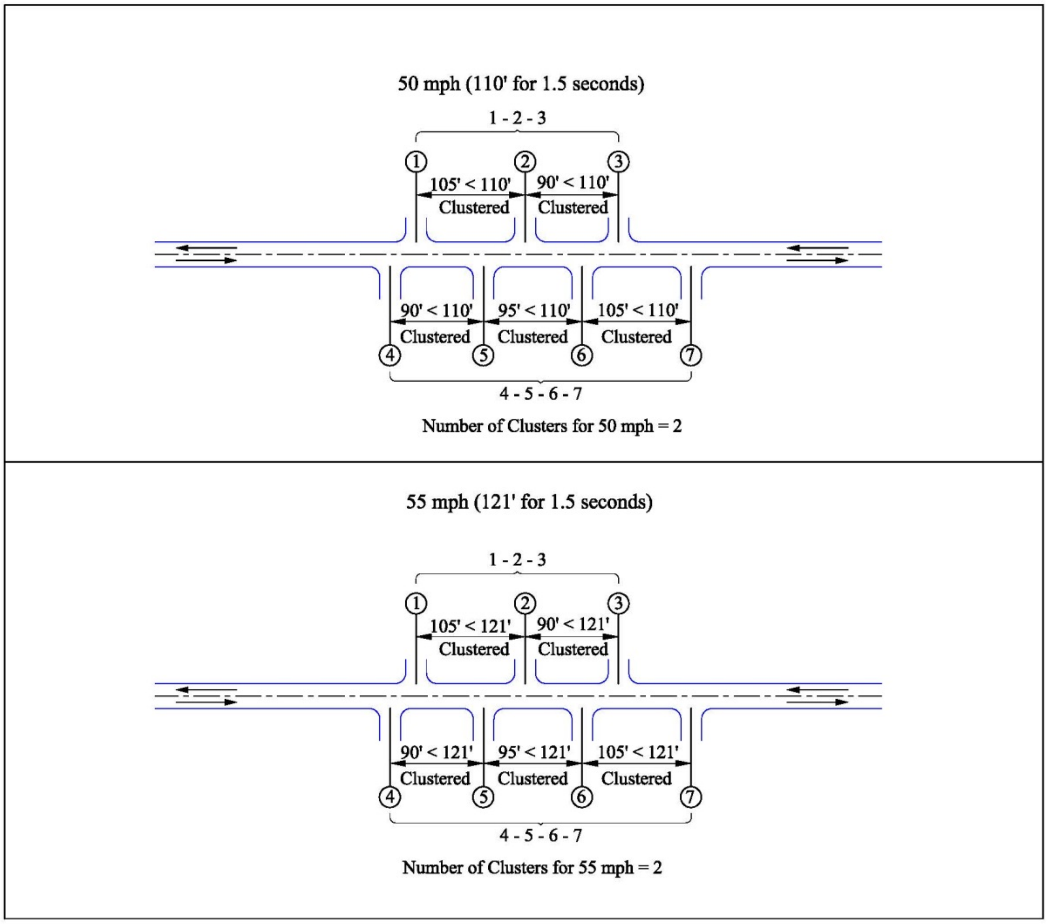


Figure 7.11: Driveways on both sides of road (Example 6)