Development of Crash Prediction Models for Transportation Planning Analysis

A Thesis

Presented to the

Graduate Faculty of the

University of Louisiana at Lafayette

In Partial Fulfillment of the

Requirements of the Degree

Master of Science

Nicholas Broussard

Fall 2015

ProQuest Number: 10002446

All rights reserved

INFORMATION TO ALL USERS The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.

ProQuest 10002446

Published by ProQuest LLC (2016). Copyright of the Dissertation is held by the Author.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code Microform Edition © ProQuest LLC.

> ProQuest LLC. 789 East Eisenhower Parkway P.O. Box 1346 Ann Arbor, MI 48106 - 1346

© Nicholas Broussard

2015

All Rights Reserved

Development of Crash Prediction Models for Transportation Planning Analysis

Nicholas Broussard

APPROVED:

Xiaoduan Sun, Chair Kenneth McManis
Professor of Civil Engineering Head and Professor

Mohammad J. Khattak Mary Farmer-Kaiser Professor of Civil Engineering Dean of the Graduate School

Head and Professor of Civil Engineering

Acknowledgments

I would like to give the highest of thanks to my wife, Sheila Broussard, my parents, and my brother for being a constant source of encouragement, love, and guidance in my life. Without them, I wouldn't have gotten as far as I have, driven myself to the heights I have reached, or made many of the decisions I have made.

I would also like to thank my peers Chuck LeBoeuf, Rene Broussard, and Subasish Das, for the time they took out of their own busy lives to help me make this project go from daunting to attainable with their selfless help.

Finally, I wish to thank Dr. Xiaoduan Sun for the immeasurable amount of knowledge that she has imparted to me over the course of the last decade. Her guidance and wisdom allowed me to continue on the path of becoming a Civil Engineer.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

Transportation planning is the act of evaluating the existing transportation system of an area, projecting its future growth, analyzing its current and future deficiencies, and selecting transportation projects with consideration of expected available funding to best suit the future needs of the area. The Transportation Bill, *Moving Ahead for Progress in the 21st Century* (MAP-21), provides the funding for roadway projects across the country and also provides the basic guidance for what a Metropolitan Transportation Plan (MTP) should encompass. Most transportation planning efforts are used for either the development of the MTP or the analysis of an upcoming project within an area for the purposes of Stage 0 analysis.

MAP-21 continued the requirements of the previous transportation bill; *Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users* (SAFETEA-LU). One of the most important features of the bill became known as the "eight SAFETEA-LU factors," which have been carried into MAP-21. One of these factors states that the transportation planning must increase the transportation system's safety for all users. This means that any transportation improvement plan must make an effort to improve the safety of the area or have a minimal impact upon it. This particular factor is difficult to analyze in planning analysis because quantitative analysis is not readily available and the term "safety" can be relative to a particular area and even to individuals.

The most common way that safety has been interpreted for the purposes of transportation planning also goes hand-in-hand with an engineering perspective. This is due to the fact that

safety is often thought of in terms of random and unintentional acts on our transportation system, which to many people means traffic crashes. Using traffic crashes as the definition of safety means that the most measureable metric available for quantitative analysis is the average crash frequency and the crash rate upon a particular roadway or intersections. In order to meet the safety requirement for MAP-21, transportation projects should produce no additional crashes and reduce them where possible.

The analysis of the base year conditions for safety can easily be done with the use of geographic information systems (GIS) by developing crash records with latitude and longitude data, importing the records to create a point layer, and begin spatial analysis with those points and a known roadway network. Using this information, crash frequencies and crash rates for roadway segments and intersections can be developed. Though segment and intersection analyses can be performed, the safety benefit for the entire project's length (encompassing both segments and intersections) is used to judge its effectiveness in meeting the safety requirement of MAP-21.

For transportation planning, the safety benefit for the future transportation network is currently often determined by engineering judgment, or that of an experienced transportation planner, due to the relative lack of quantitative analysis available for crashes. One of the few quantitative means available is that of the models developed from the *Highway Safety Manual* (HSM). However there are several issues that make this means of analysis largely inconvenient and incompatible with transportation planning.

2

The HSM models require a lot of input information in order to be used properly. Often, the input required is not normally used at the planning level and is not readily obtainable. The planning level information that is collected and can be used in crash analysis are:

- Segment length
- Average daily traffic
- Number of lanes
- Roadway configuration (divided, undivided, center turn lane, etc.)
- Posted speed

The HSM models additionally require the following data (dependent upon the type of model used):

- Lane width
- Shoulder width
- Shoulder type
- Median width
- Side slopes
- Lighting
- Auto speed enforcement
- On-street parking
- Proportion of curb length and on-street parking
- Amounts of driveways by various land use types
- Fixed object along roadside density
- Offset of fixed objects along roadside

Furthermore, when the data mentioned above can be found to perform analysis using the HSM methods, it is very time-consuming and either cannot be done within the time frame available to complete a transportation planning project (usually one year from start to finish), or would cost too many hours and money for the project to be completed within budget.

The HSM methodology is also known to need calibration factors developed at the local level. This calibration factor takes a lot of time to develop, and must be developed for each type of model that the HSM offers, usually in conjunction with the state Department of Transportation. These issues lead many to believe that the use of the HSM methodology in transportation planning to be an inadequate approach and considered "too umbrella" to be used properly.

1.2 Objectives

As previously mentioned, the HSM methodology is considered to be largely incompatible with transportation planning efforts due to time constraints and data availability. While the use of engineering judgment or those of experienced planners can be used for the safety element of a planning project, there is an increased need for quantitative analysis due to the need to show progress through performance measures as per MAP-21 requirements. To that end, the objective of this project is to establish a crash prediction model for both urban and rural roadway segments with data that is readily available from the state Department of Transportation, local government with jurisdiction of the transportation project(s), or another reliable source in order to determine the impact of a transportation project upon the safety of the roadway network.

4

Chapter 2: Literature Review

There are three ways that model development can be approached for developing a crash prediction model: the HSM, mathematical model development, and machine learning algorithms. The use of the mathematical model is a more traditional method that provides the end user with an equation to calculate the desired results (crash prediction, in this case). The drawback of the mathematical models is that they can only run at an aggregated level when developing the equation and will lack the accuracy needed to operate at a small level, meaning that some interpretation of the results may be necessary. Modeling that uses machine learning algorithms "learns" from the available data and determines how to perform the given task(s) by generalizing from example and mimicking the expected results (1). There are several machine learning algorithms types, which include clustering, support vector machines (SVR), fuzzy algorithms, and kriging methods. The drawback to using the machine learning algorithms is that they do not provide an equation, but rather a "black box" that does not provide the end user an equation that can show how the result was obtained.

2.1 Mathematical Model Development

2.1.1 Analysis of Current Models

The first model that was made available to engineers and planners for the purposes of crash analysis was the *Highway Safety Manual* (2). The book, created by the American Association of State Highway and Transportation Officials (AASHTO), discusses crash safety and analysis in detail based on 30 years of previously conducted research on safety modeling from all over the United States. Additionally, the HSM provides an explanation of the common factors in traffic crashes, develops crash modification factors based on the roadway

conditions, and provides a methodology to calculate crashes based on the given conditions. The models themselves are provided to the public for free in Microsoft Excel format by AASHTO (3) in order to assist in efforts to conduct safety analysis for roadway projects.

As previously discussed, the data contained in the HSM models is considered a very broad approach and the need to develop calibration factors as well as the large amount of data collection means that other modeling means would need to be used for transportation planning. In addition, the HSM models do not have the capacity to do crash analysis for anything greater than a four-lane roadway with a center turn lane, or two-way left turn lane (TWLTL), limiting the usefulness of the model in urban areas that wish for a project to widen a roadway to six lanes or beyond. The HSM also does not have a fully developed interstate model, though the draft has been published.

The next model that was made available is the Interactive Highway Safety Design Model (IHSDM). "The IHSDM **is** a suite of software analysis tools used to evaluate the safety and operational effects of geometric design decisions on highways" (4). It is intended as a support tool for decision-making and estimates expected safety of roadway conditions and designs. The suite was created with the expectation of being used by highway project managers, designers, and traffic and safety reviewers in state and local highway agencies, as well as engineering consultants and firms.

The IHSDM was developed using the HSM and SafetyAnalyst, which is now available as AASHTOWare, and is free for anyone to use. The IHSDM is administered by the Federal Highway Administration. Due to the fact that the IHSDM makes use of the HSM, the Crash

6

Prediction Module uses the HSM models for evaluating rural 2-lane highways, rural multilane highways, and urban/suburban arterials, as well as the draft interstate model. This unfortunately means that it cannot be used for planning-level analysis because it is effectively the HSM with additional software features.

Several states and cities have made the attempt to develop crash prediction models using the HSM and develop the necessary calibration factors for their state or local areas. Others have developed their own crash prediction models (safety performance functions) with their own available data. Table 2-1 shows the status of the current efforts for these models as of July 2014.

Table 2-1: Status of State/Local Crash Prediction Models

Source: The National Academies of Sciences, Engineering, and Medicine; Transportation Research Board

2.1.2 Alternative Model Development

While some states now have their own safety performance functions or HSM calibration factors, many more either have not begun to develop them, or have begun to look at their own alternatives. The use of these alternatives was the starting point for the development of this project's methodology as they are better aligned with the data and methodologies of transportation planning. However, they will only be covered in brief detail as none of these methodologies were used in the final product.

Efforts in Oregon showed that mathematical modeling for a state was possible, even with planning level analysis data, by Dixon and Avelar (5). Their research showed that the models

could be analyzed by collecting data on annual average daily traffic (AADT), driveway location, driveway width, driveway type, number of lanes, median configuration, posted speed, and traffic control. Of these variables, all but the driveway data and traffic control are commonly used in transportation planning models. Further analysis of the Dixon and Avelar models reinforced that such modeling could be used for transportation planning purposes but would have to be heavily modified beyond an acceptable level because of the need to remove the driveway data due to time constraints, which itself echoes one of the HSM issues.

The work that was done by Bonneson, Zimmerman, and Fitzpatrick (6) for the Texas Department of Transportation also yielded results that could be applied to the transportation planning level. Furthermore, their research provided the opportunity to choose between several types of equations for mathematical modeling analysis. This project was started because the Department of Transportation was looking to incorporate quantitative safety analysis into the design process at an earlier stage than is usually done. The project objective was to develop safety guidelines and evaluation tools and implement them in the project design process. The results of the project were used to develop the guidelines presented in the *Interim Roadway Safety Design Workbook*.

Their work only looks at the design factors and their relationship to safety. In a way, this is somewhat the same approach that is used in the HSM. However, their methodology does not use the HSM and actually uses models from several different sources. The models developed in their report were capable of handling several different land types as well as more than 4 lanes and were considered to be a good starting point for planning level analysis. After

analyzing the methodologies used in their report, a modification of the Hadi models that they used was selected for the starting point of the model being developed in this thesis.

Unfortunately, while the modeling process was successful in obtaining results and producing an equation, these results were considered to have too much variation within the dataset and produced results that were deemed in excess of a reasonable deviation in the data. This led to the decision to use a vector analysis methodology for model development.

2.2 Machine Learning Approach- Vector Regression

A study by M. Castero-Neto, Y. Jeong, M. Jeong, and L. Han (7) used support vector regression for the purposes of predicting AADT for the state of Tennessee. Their research evaluated the performance of a modified version of the support vector machine for regression (SVR) in order to forecast AADT for one year without the use of any external variables. The SVR methodology is becoming more commonly used due to its general performance and lack of local minima. The SVR models, and their results, are highly dependent upon on the settings of the type of kernel used, the value of C , and the value of ε for the ε -insensitive loss function. This information was used to guide the work of LeBouef and Sun (8) as they developed a model using the SVR methodology in order to estimate the AADT on roadways in Louisiana that are not maintained by the Louisiana Department of Transportation and Development (LADOTD). It is believed that through their work using the SVR modeling approach that was developed by LeBouef and Sun, but tailored to planning level data that can be used for crash analysis, a reasonable crash prediction model can be developed for any state, urbanized area, or municipality.

10

Chapter 3: Methodology

This chapter introduces the safety modeling techniques starting with the discussion on the data.

3.1 Crash Data

3.1.1 Crash Records

The development of any crash prediction model must begin with the actual crash records that are recorded by the various agencies dealing with these events: the local law enforcement forces that respond to the crashes. Like every state in the U.S., Louisiana has a standardized crash record form that is used to gather data on traffic crashes, including the information on driver, vehicle, and crash characteristics as shown in Appendix A. These crash records are provided to the LADOTD for the purposes of safety analysis in order to conduct traffic safety studies or for the identification of locations that are in need of improvement. All crashes that are handled by the Louisiana State Police are provided to the LADOTD while local governments are encouraged to provide their local crash data to the LADOTD and often do.

The development of this model uses three years of crash data (from 2011-2013) in order to use the most-up to-date data that is readily available from the LADOTD. The use of three years of data also allows for the establishment of more recent traffic trends while avoiding a regression to the mean (a statistical event that makes natural variation between samples look like real change) bias. The data received from the LADOTD contains data on crash number, highway number, AADT, control section, functional classification, highway classification, logmiles, latitude and longitude, milepost, and more. This allows for a large set of data and

potential variables, though only a few are necessary. The data used for the development of the model is detailed in Section 3.2.

3.1.2 Roadway Data

To link a roadway safety performance to its attributes (geometric features and traffic condition), roadway data must also be obtained for the research. Thus, in addition to the crash records themselves, a state-maintained roadways database by the LADOTD was also used. This data presented by a GIS layer was created in 2012 from the GIS division at the LADOTD headquarters, which provides a comprehensive data set on the roadways, including many of the attributes in the crash records. Additionally, the layer can be used to cross-check roadway attributes that are listed in the crash records or even provide supplemental data for model development.

3.2 Data Processing

To model roadway segment safety performance, all individual crash data must be populated to each segment. This means that the individual crash records need to be aggregated to the segment level. The task was accomplished using several programs.

The crash data provided by the LADOTD has all of the crashes in a given year compiled into one file. However, the record is known to change from year to year with column names changing and the addition or removal of other columns also being common. In order to address this, the known desired variables from engineering and planning experience were selected after the files were properly converted. Using the Microsoft Office software suite, the data files were converted into a usable format (Microsoft Excel) and each individual

record had a "FLAG" field placed at the end to assist with combining the records and calculating their pertinent data.

To avoid exceeding the Microsoft Access file limitation and improve the speed of the data processing, the original dataset was first cleaned up by eliminating unnecessary columns. Additionally, the three LADOTD files that were received by year were combined into one file in order to obtain a comprehensive set that accounts for all of the crashes under given roadway conditions. After combining the year data into one set and eliminating the unnecessary columns, a spreadsheet containing the data in Table 3-1 remained.

Column Name	Description
CONTROL_SECTION	LADOTD Control Section number where the crash occurred at
LOGMILE_FROM	LADOTD Logmile where the segment begins at
LOGMILE_TO	LADOTD Logmile where the segment end at
LENGTH	Length of the segment
ADT	Average daily traffic on the segment
FUNCTIONAL_CLASS	Highway function classification of the segment
HIGHWAY_CLASS	Description of the type of highway the segment is located on
MEDIAN_WIDTH	Width of the median, if any, on the segment
NUM LANES	Number of lanes in both directions of the segment
PAVEMENT_WIDTH	Total pavement width of the segment
INTERSECTION	A marker by LADOTD the shows if a crash was intersection related or not.

Table 3-1: List of Initial LADOTD Variables for Modeling

Source: LADOTD, 2015

The spreadsheet containing this data, along with a FLAG field in which all values were equal to one (1) was imported into Microsoft Access. There were over 450,000 crash records imported for the model analysis. Using a query code, the crash data was aggregated to the segment level based on crashes that had the same:

- Control section
- Logmile points
- Length
- Functional classification
- \bullet Highway class
- Median width
- Number of lanes
- Pavement width
- Intersection involvement

The average ADT over three years was used. Without averaging this value each segment would have multiple records, with one for each ADT, and negating the reason for using multiple years of data. This process provides data for a segment with specific conditions and an average ADT. The total crashes are a summation of the FLAG field, since each crash record had $FLAG=1$.

The result of the data aggregation provides over 15,400 segments for analysis purposes. However, these data segments contain data on the local crashes (identified by a lack of control section and logmile data as those features are unique to LADOTD roadways) and

intersection crashes, which are not necessary to this study. Planning level safety analysis for an MTP focuses only on the roadway network that is used in the travel demand model and that vast bulk of these models use the functionally classified roadways, using local roads only for connectivity purposes in the model. This meant that the segment crash records for the non-functionally-classified local roadways needed to be removed. Additionally, because this study focuses on developing a functionally-classified segment crash prediction model, any records that had INTERSECTION=1 were also removed (nearly 6,500 records).

In order to avoid the unintentional removal of any functionally classified roadways, segments that had missing control section or logmile data were analyzed using the TransCAD software and the provided LADOTD GIS layer. Using the developed data and the joining feature from the TransCAD software, reasonable queries were created to find matching data between the control sections or logmiles. The missing data was then copied from the LADOTD GIS layer to complete the data where possible. Those records that could not be completed (23 segments) were removed to avoid erroneous data.

During the process of correcting the missing logmile and/or control section data, supplemental data from the GIS layer was brought in. This data was used to correct missing highway class data as well as determine what each highway class code meant. Furthermore, additional data for the median widths and median types were added to the segment data where possible. Finally, where necessary, the median width data originally obtained from the crash records was adjusted using judgment based on experience with roadway classes and the supplemental DOTD median data. This was done to create reasonable numbers for values

that were incorrect based on roadway or median type (e.x., a rural 2-lane highway wouldn't have a median width of 299 feet and is almost always an undivided highway with 0 median width).

3.3 Model Development

3.3.1 Initial Variables

Following the aggregation of the data and the adjustments for missing or incomplete data, an initial selection of variables was chosen for modeling. The variables chosen were:

- Length
- Average ADT
- Median Width
- Number of Lanes
- Pavement Width
- Average Crashes Per Year (a.k.a. Crash Frequency)

3.3.2 Exploratory Data Analysis

Following the selection of the initial model variables the vector regression model was used in a "trainset" mode to allow the model to learn the patterns within the data. Trainset is where the model effectively develops the crash prediction model to be used in application based upon the trends in the data sets provided. In addition to analyzing the results of the model performance and its relationship between the observed values and actual values, an exploratory data analysis (EDA) is needed to determine if the variables being used are appropriate for the model and show the relationships between the chosen values.

For modeling purposes, the chosen independent variables should have a strong and positive relationship with the dependent variable (the crash frequency) where possible, while having little to no relationship with the other independent variables. If the model uses independent variables that have a relationship with one another the results provided will be influenced and skewed towards the variables with the relationship due to a type of redundancy within the data. Figure 3-1 through Figure 3-6 display the matrices provided by the model for the EDA to analyze the relationships between variables. A larger circle with a deeper color represents a stronger relationship between the variables and the color denotes whether the relationship is positive (blue) or negative (red). Note that the rural and urban 2-lane models do not use the number of lanes as a variable since the value is constant in those two models.

Figure 3-1: Rural 2-Lane Model Variable Relationships EDA

17

Figure 3-2: Rural Multi-Lane Model Variable Relationships EDA

Figure 3-3: Rural Interstate Model Variable Relationships EDA

Figure 3-4: Urban 2-Lane Model Variable Relationships EDA

Figure 3-5: Urban Multi-Lane Model Variable Relationships EDA

Figure 3-6: Urban Interstate Model Variable Relationships EDA

3.3.3 Variable Revisions

Analysis of the EDA results and knowledge of crash factors revealed that the usage of pavement width in the model is inappropriate for modeling purposes. The ADT of a roadway is influenced by several factors which include the number of lanes and the pavement width, both of which are displayed in the EDA. Even though the number of lanes has a relationship to the ADT, the variable is still important to this model as the number of lanes also introduces weaving and other driver behaviors that affect crashes. However, the total pavement width is a result of the number of lanes and as pavement width and the number of lanes increases so does the ADT. This requires the removal of the pavement width as a model variable based on the requirements mentioned in Section 3.3.2. It is known, however, that the average lane width can have an impact upon crash frequencies and is independent of the

ADT. The pavement width data was used to derive the average lane width data for the next model run.

In addition to the changes based on the pavement width, the median width data also needed to be changed in the model runs. The median width has a positive relationship to the crash frequency but the strength of the variable in the 2-lane and multi-lane models did not have the desired effect. The use of median width in the interstate models is necessary since all interstates are required by design standards to be divided in some manner and the median width has the proper strength of relationship. In order to model the impact of the median, which is known to be a factor in crash safety, upon the 2-lane and multi-lane highways while removing the issue of the widths having a weak relationship, the 2-lane and multi-lane models were tested with a variable that indicates whether the segment has a median presence (MED_PRESENCE) or not. A second EDA was performed after the variable revisions and the results are shown in Figures 3-7 through 3-12.

Figure 3-7: Rural 2-Lane Model Variable Relationships EDA- Revised

Figure 3-8: Rural Multi-Lane Model Variable Relationships EDA- Revised

Figure 3-9: Rural Interstate Model Variable Relationships EDA- Revised

Figure 3-10: Urban 2-Lane Model Variable Relationships EDA- Revised

Figure 3-11: Urban Multi-Lane Model Variable Relationships EDA- Revised

Figure 3-12: Urban Interstate Model Variable Relationships EDA- Revised

3.3.4 Final Variables

The revised variables provide the desired results for the variable relationships so that independent variables used to calculate the crash frequencies are not dependent upon one another. Of note is the relationship of the ADT to the average lane width in the 2-lane models. Like the ADT to number of lanes relationship, the average lane width is important on two lane roads due to its impact upon driver behavior and is still necessary despite the relationship. Wider roadways often attract more drivers due to the comfort of more "forgiving" roadways for when they make errors or need to move across the pavement to avoid roadway debris, animals, etc. These greater widths therefore affect the safety of the roadway and reduce crashes due to their more forgiving nature.

Following the second EDA, the model was again run in the trainset mode for the model to learn the patterns of the chosen variables and the crash frequency. The results of the trainset reveal whether the model is producing sufficiently acceptable results to move on to model validation and then into crash prediction for future years. Figures 3-13 through 3-18 show a plot of the trainset's predicted versus observed values.

Figure 3-13: Rural 2-Lane Model Trainset Results

Figure 3-14: Rural Multi-Lane Model Trainset Results

Figure 3-15: Rural Interstate Model Trainset Results

Figure 3-17: Urban Multi-Lane Model Trainset Results

Figure 3-18: Urban Interstate Model Trainset Results

The model results show that the SVR models are capable of predicting the crashes close to the observed values and have an acceptable amount of variation. Additionally, the models are not predicting values in a manner that is either consistently higher or consistently lower than

the observed crashes. Based upon the EDA for these new variables and the results shown in the plots of Figures 3-13 through 3-18, the new variables are acceptable to proceed with model validation and application.

The final variables for the crash prediction models are:

2-lane and multi-lane models:

- LENGTH
- AVG_ADT
- MED_PRESENCE
- NUM_LANES (Multi-lane only)
- AVG_LANE_WIDTH
- AVG_CRASH

Interstate models:

- LENGTH
- AVG_ADT
- MEDIAN_WIDTH
- NUM_LANES
- AVG_LANE_WIDTH
- AVG_CRASH

Chapter 4: Results and Application

4.1 Model Parameters

The chosen SVR models used in this study are dependent upon the kernel type, value of the penalty for excess deviation during training (*C, Gamma*), and error-term value (ε, Epsilon) for the ϵ -insensitive loss function. (7, 8) The number of support vectors to be used in modeling is determined before running the SVR analysis. The models used in this thesis are run using an open-source software programming language, R, to predict the average yearly crash frequency. The model parameters are as follows:

- SVM-Type, eps-regression
- SVM-Kernel, radial
- Cost, a value of 100 in the study
- Gamma, a value of 1
- Epsilon, a value of 0.1

However, the initial estimated values are not shown in the script window in R and the results need further analysis (both visual and mathematical) to be validated.

4.2 Model Validation

Validating results from the models is of paramount importance. Model validation is defined as the process of determining the degree to which a model is an accurate representation of the real world results. It is accomplished through the comparison of predictions from a model to the observed data. For this purpose, 25 percent of the data being tested are purposely reserved as testset for the model validation while 75 percent of the data, called trainset, were used in model development.

Validation of the crash predictions models is done at the individual model level using the testset data feature of the model. The use of the trainset feature is for the model to learn the trends and patterns that exist in the current data. The use of the testset feature allows the model to forecast crash frequencies using a given set of conditions based on the knowledge the model obtained from the trainset. Testset is the actual application of the crash prediction model developed during the trainset phase in order to predict the crash frequency of a roadway segment based on the given attributes.

It is important to note that the trainset values are often considerably better than the testset values. The trainset is specifically intended to match the trends and represents the best fit possible for the model. Because the testset is an application, it results in estimates that are considerably less accurate but still capable of being used for the intended purposes. Transportation planning as a field is built upon the knowledge that results obtained from the models used in the process are estimates based on the available knowledge and trends. This means that the testset application of the model is still in line with the methodologies used in transportation planning.

In order to validate the model, an analysis of the R^2 of the data plots (observed vs predicted), and RMSE was conducted. The use of RMSE was chosen due to the fact that a raw aggregate sum and percent deviation comparison can be misleading. This is due to the fact that the total sums of the observed and predicted crashes can be very close, but individual segments can have a high amount of variation between them, resulting in what appears to be a good overall model performance but a weak performance at the segment level. The RMSE is a

representation of the standard deviation of the differences between the observed field values and predicted values within the model sample.

Table 4-1 displays the model validation statistics. Based upon the values shown in Table 4-1, the testset values show that the model performs at an acceptable level. For those models in which the R^2 is not particularly favorable the RMSE values show that the particular model still produces acceptable results as these roadways often experience high crash frequencies.

	Trainset		Testset	
	R-Squared	RMSE	R-Squared	RMSE
Rural 2-Lane	0.59	2.29	0.515	2.63
Rural Multi-Lane	0.83	1.97	0.498	3.59
Rural Interstate	0.85	6.53	0.167	20.05
Urban 2-Lane	0.66		0.445	8.63
Urban Multi-Lane	0.86		0.157	19,70
Urban Interstate	0.67	18.2	0.159	43.86

Table 4-1: Base Model Validation Statistics

4.3 Model Application

The application of the model for the use of crash prediction comes from the testset feature built into the model's code. This allows for the transportation planning aspect of these models to be used as the forecast data input from the travel demand models used in the MTP process feeds directly into the SVR models, allowing for the crash frequencies of future years to be predicted.

The steps involved in the model application are as follows:

- 1.) The data for the roadways, using the variables described in Chapter 3, is collected for
	- each segment in the roadway network under analysis.

- 2.) Data from the statewide dataset used in the project, or that obtained from the local area in question (usually the MPO or Parish level) with the known crash frequencies, is used in trainset mode for the model to learn the base data conditions.
- 3.) The model trainset values are calibrated and validated as necessary.
- 4.) Using the same roadway data as collected in Step 1 but updated to the forecast year values (obtained from a travel demand model), the model is run in testset mode to determine the predicted crash frequencies per segment.
- 5.) Step 4 is then repeated, but using forecast year values that reflect a transportation planning test project under consideration instead, once more determining the crash frequencies per segment.
- 6.) The crash frequencies of Steps 4 and 5 are then compared for the given corridor of a test project and those nearby to determine the change in crash frequencies of the corridors and consequently if the test project makes the roadways safer or not.

Of note is that in Steps 4 and 5 the data used in the model runs relies upon information that may not be in the base data. When a roadway is widened or newly built, there are unknowns in the design of new medians, lane widths, and other factors. Much of this data necessary for the crash prediction models developed in this study is already determined during the transportation planning process (number of lanes, ADT, segment length, divided/undivided). For data which is not normally determined in the planning process (average lane width, median presence, median width), the use of the LADOTD roadway design guidelines (Appendix B) is used for new roadways to determine this data, while widened roadways use

both the LADOTD roadway design guidelines and engineering judgement based on the roadway location.

Step 6 as described above is what would allow for engineers and transportation planners to have a metric for the justification of ranking a project in an MTP update based upon the safety impacts it is projected to have. A decrease in the crashes along the analyzed corridors means that the project increases safety and therefore would receive a higher score in the transportation planning process as it helps to achieve one of the eight planning requirements of MAP-21. A case study in how the developed crash prediction models of this thesis is displayed in Appendix C.

Chapter 5: Conclusion

Urban transportation planning and metropolitan transportation plans are complex, regulated, and vitally important undertakings that impact the long range growth and health of an area. As such, the Federal government ensures that where possible the planning processes are adequately defined and the desired outcomes are well known. These outcomes have come even further to the fore with each new transportation bill, and MAP-21 has pushed transportation planning into an era of further numerical proof of the effects of the MTP documents that MPOs and consultants put out for a region.

Historically the safety element scoring in these transportation planning documents has relied solely upon the judgment and knowledge of experienced professionals. While these professionals know how the safety of roadways is likely to change from the projects involved in an MTP due to their long-used experience, the advent of the performance measures demanded by MAP-21 means that developing numerical means of displaying the safety change has become ever more important.

Few states have the means with which to show how safety changes on their roadways as the development of a program or model can be time-consuming and costly. This often means that planning level models are even more disadvantaged as most planning projects last only about a year to a year and a half. The development of a planning level crash prediction model, however, is now almost a must to meet the MAP-21 requirements.

While the use of HSM modeling is available to transportation planning efforts, they are cumbersome and time consuming. The data involved in properly developing HSM models requires more time to obtain than is often practical for the timeframe of an MTP update and most of the data is not available at the planning level. The models developed in this project require less time and less data, as well as making use of data that already used at the planning level or easily obtainable.

The models developed through the research in this project have shown that such models can be created using statewide crash data provided from an agency that maintains such information. Crash data from the LADOTD was used to create a database of crashes by roadway segment (using control section and logmile data) and its respective data for ADT, length, number of lanes, and other information. Using this database, six models were created for various roadway types and, using SVR modeling, a validated base model was created that emulated the trends of average crash frequencies on roadway segments for both rural and urban roadways.

These models can be used in order to predict the future crash frequencies on roadways based upon their expected future conditions and the impacts that test projects for an MTP process will have upon these roadways. Using the predicted crash frequencies, the relative change between the two model runs for given segments will show the change in crashes, and therefore, safety, on a roadway. This will allow transportation planners to use mathematical data to rank and score transportation projects based on safety and satisfy the MAP-21 requirements.

References

- 1. Domingos, P. *A Few Useful Things to Know about Machine Learning.* University of Washington, Seattle. 2012.
- 2. Bonneson, J. *Highway Safety Manual*. American Assosciation of State Highway and Transportation Officials, Washington, DC. 2010.
- 3. Tool Descriptions. American Assosciation of State Highway and Transportation Officials, Washington, DC. http://www.highwaysafetymanual.org/Pages/tools_sub.aspx#4. Accessed August 1, 2015.
- 4. Interactive Highway Safety Design Model: Overview. Federal Highway Administration, Washington, DC. http://www.fhwa.dot.gov/research/tfhrc/projects/safety/comprehensive/ihsdm/. Accessed August 1, 2015.
- 5. Dixon and Avelar. *Validation Technique Applied to Oregon Safety Performance Function Arterial Segment Models*. Texas A&M Transportation Institute, College Station, TX. 2014.
- 6. Bonneson, Zimmerman, and Fitzpartrick. *Roadway Safety Design Synthesis*. Texas A&M Transportation Institute, College Station, TX. 2005.
- 7. Castero-Neto, M., Jeong, Y., Jeong, M. K., and Han, L. D. AADT Prediction Using Support Vector Regression with Data-Dependent Parameters. Research Report. University of Tennessee, Knoxville, 2008.
- 8. LeBoeuf and Sun. *Estimating Annual Average Daily Traffic for Non-State Roads in Louisiana*. University of Louisiana at Lafayette, Lafayette, LA. 2014.

Appendix A: Sample Police Crash Report

REFER TO EACH BY VEHICLE NUMBER

officer's narrative: describe any unusual circumstances associated with crash, including officer's observations and opinions.
Include witness names, addresses, phone numbers, etc. IF NECESSARY. INDICATE DAMAGE TO PUBLIC OR PRIVATE PROPERTY (WITH OWNER'S NAME & ADDRESS) AT THE END OF THE NARRATIVE.

المنسارة الاستشارات

DPSSP 3106 (REV. JAN. 2005)

INVESTIGATING OFFICER'S INITIALS

PAGE # **COMPUTER NUMBER**

officer's narrative: describe any unusual circumstances associated with crash, including officer's observations and opinions.
Include witness names, addresses, phone numbers, etc.

IF NECESSARY, INDICATE DAMAGE TO PUBLIC OR PRIVATE PROPERTY (WITH OWNER'S NAME & ADDRESS) AT THE END OF THE NARRATIVE.

REFER TO EACH BY VEHICLE NUMBER

DPSSP 3110 (REV. JAN. 2005)

INVESTIGATING OFFICER'S INITIALS

INVESTIGATING OFFICER'S INITIALS

Appendix B: LADOTD Minimum Design Guidelines

LOUISIANA DEPARTMENT OF TRANSPORTATION AND DEVELOPMENT **Minimum Design Guidelines for Freeways**

Approved 4 le Chief Engineer

 $\frac{12.4.09}{\text{Date}}$

Minimum Freeway Dasign Guidelines - Sheet 1 of 2

Footnotes for Minimum Design Guidelines for Freeways

- 1. These guidelines may be used in urban areas.
- 2. Level of Service C can be used in urban areas.
- 3. Level of Service D can be used in heavily developed urban areas.
- 4. Four feet to be paved, 10 feet to be paved on 6 lane facilities, 12 feet to be paved on 6 lane facilities with truck DDHV greater than 250.
- 5. Twelve feet paved when truck DDHV is greater than 250.
- 6. For larger medians two harriers may be required. The maximum offset of 15 feet from barrier to edge of travel lane shall not be exceeded.
- 7. In Districts 04 and 05, where ice is more frequent, superelevation should not exceed 8 percent from the $e_{\text{max}} = 10\%$ table.
- 8. It may be necessary to increase the radius of the curve and/or increase the shoulder width (maximum of 12 feet) to provide adequate stopping sight distance on structure.
- 9. Grades 1 percent higher may be used in urban areas.
- 10. An additional 6 inches should be added for additional future surfacing. Seventeen feet is required for trusses and pedestrian overpasses.
- 11. As needed for urban projects: 300 feet to 330 feet for rural projects depending on median width.

12. Twenty-five feet shall generally be provided in accordance with EDSM II.1.1.1.

13. LRFD for bridge design.

General Note:

DOTD pavement preservation minimum design guidelines or 3R minimum design guidelines (separate sheets) shall be applicable to those projects for which the primary purpose is to improve the riding surface.

Minimum Freeway Design Guidelines - Sheet 2 of 2

LOUISIANA DEPARTMENT OF TRANSPORTATION AND DEVELOPMENT Minimum Design Guidelines for Rural Arterial Roads

State law requires that the state highway system conform to these guidelines.

Approved Chief Engineer

 $\frac{1}{2}$ $\frac{Z-4}{2}$ - 09

Minimum Rural Arterial Road Design Guidelines - Sheet 1 of 2

Footnotes for Minimum Design Guidelines for Rural Arterial Roads

- 1. The design speed may not be less than the current posted speed of the overall route.
- 2. Consider using RA-3 criteria (except Item No. 2) for roadways that will be widened in the future.
- 3. Consider increasing to a 4-lane facility if design volume is greater than 6,000 vehicles per day and six lanes if design volume is greater than 25,000 vehicles per day. If more than two lanes are to be provided, outside shoulders should be paved.
- 4. Twelve feet required when design ADT is 1.500 or greater.
- 5. Six foot shoulders are allowed if design volume is between 400 to 2,000 vehicles per day. Four foot shoulders are allowed if design volume is less than 400 vehicles per day.
- 6. Eight to ten feet to be provided on six lane facilities.
- 7. Consider using 10 foot outside shoulders where trucks are greater than 10 percent or if large agricultural vehicles use the roadway.
- 8. For ADT 5,000 or greater, the full shoulder width shall be paved.
- 9. In Districts 04 and 05, where ice is more frequent, superelevation should not exceed 8 percent from the $e_{\text{max}} = 10$ percent table.
- 10. It may be necessary to increase the radius of the curve and/or increase the shoulder width (maximum of 12 feet) to provide adequate stopping sight distance on structure.
- 11. Grades 1 percent higher are permissible in rolling terrain.
- 12. An additional 6 inches should be added for additional future surfacing.
- 13. On multilane facilities, use 32 feet.
- 14. LRFD for bridge design.

General Note:

DOTD pavement preservation minimum design guidelines or 3R minimum design guidelines (separate sheets) shall be applicable to those projects for which the primary purpose is to improve the riding surface.

Minimum Rural Arterial Road Design Guidelines - Sheet 2 of 2

LOUISIANA DEPARTMENT OF TRANSPORTATION AND DEVELOPMENT Minimum Design Guidelines for Urban Arterial Roads and Streets

State law requires that the state highway system conform to these guidelines.

ApprovedCle $12 - \frac{9 - 99}{\text{Date}}$ Jer The Chief Engineer (Chief Engineer Chief Engineer Chief Engineer Chief Engineer Chief Engineer (Chief Engineer Chief Engineer Sched 1 of 2

Footnotes for Minimum Design Guidelines for Urban Arterial Roads and Streets

- 1- Level of service D allowable in heavily developed urban areas.
- 2- Curb may be used in place of shoulders on UA-1 and UA-2 facilities. If used on UA-3, UA-4, or UA-5 facilities, curb should be placed at the edge of shoulder. For design speeds greater than 45 mph, curb will not be placed in front of guardrail.
- 3- With Chief Engineer's approval, curb offsets may be eliminated and the minimum median width can be reduced to 4 feet. On principal arterials, particularly at intersections, the upper limit should be considered.
- 4- Cannot be used on multilane roadways (with four or more through lanes) without the Chief Engineer's approval.
- 5- Sidewalks must be separated from the shoulder and should be placed as near the right of way line as possible. On high speed facilities, they should preferably be placed outside the minimum clear zone shown in item 18.
- 6- It may be necessary to increase the radius of the curve and/or increase the shoulder width (maximum of 12 feet) to provide adequate stopping sight distance on structure.
- 7- The following radii apply at divisional islands. The radius selected must match the design speed of the road. These radii also apply to the other guidelines where divisional islands are mentioned.

- 8- An additional 6 inches should be added for additional future surfacing.
- 9- Applies to facilities with shoulders. Refer to the Roadside Design Guide when 1:3 fore slopes are used or for slopes flatter than 1:4.
- 10- The distance may be reduced by 6 feet if 1:6 slopes are used. For outside shoulders wider than 8 feet, further reduction should be proportional to the added shoulder width.
- 11-If outside shoulders and curb are used, refer to the Roadside Design Guide.
- 12- Where left turn lanes are provided or where the median is less than 6 feet in width, the minimum clearance will be 1.5 feet from back of curb. For median slopes steeper than 1:6, refer to the Roadside Design Guide for the desirable clear zone.
- 13-LRFD for bridge design.
- 14-Refer to EDSM II.3.1.4 when sidewalks will be provided and for guardrail requirements.

General Note:

DOTD pavement preservation minimum design guidelines or 3R minimum design guidelines (separate sheets) shall be applicable to those projects for which the primary purpose is to improve the riding surface.

Minimum Urban Arterial Road and Street Design Guidelines - Sheet 2 of 2

LOUISIANA DEPARTMENT OF TRANSPORTATION AND DEVELOPMENT Minimum Design Guidelines for Suburban Arterial Roads and Streets

State law requires that the state highway system conform to these guidelines.

Approved (1 1 Chief Engineer

 $12 - 4.09$ Date

Minimum Suburban Arterial Road and Street Design Guidelines - Sheet 1 of 2

Footnotes for Minimum Design Guidelines for Suburban Arterial Roads and Streets

- 1. These guidelines may be used only on a rural roadway section that adjoins a roadway section currently classified as urban. The classification selected should be based on the posted speed.
- 2. If curb is used, it shall be placed at the edge of shoulder on two lane facilities and 1 fcot beyond the edge of the shoulders on multilane facilities. However, see EDSM IL2.1.7. Curb will not be placed in front of guardrail.
- 3. Sidewalks must be separated from the shoulder and should be placed as near the right of way line as possible. They should desirably be placed outside the minimum clear zone shown in Item 18.
- 4. It may be necessary to increase the radius of the curve and/or increase the shoulder width (maximum of 12 feet) to provide adequate stopping sight distance on structure.
- 5. Different radii apply at divisional islands. See Footnote 7 for "Minimum Design Guidelines for Urban Arterial Roads and Streets.
- 6. Grades 1 percent higher are permissible in rolling terrain.
- 7. An additional 6 inches should be added for additional future surfacing.
- 8. Use the larger value when 1:4 fore slopes are used.
- 9. LRFD for bridge design
- 10. For roadways with shoulders and curbs, consider widening each bridge 8 feet to allow for a future lane and 4 foot offsets to bridge rail.

General Note:

DOTD pavement preservation minimum design guidelines or 3R minimum design guidelines (separate sheets) shall be applicable to those projects for which the primary purpose is to improve the riding surface.

Minimum Suburban Arterial Road and Street Design Guidelines - Sheet 2 of 2

LOUISIANA DEPARTMENT OF TRANSPORTATION AND DEVELOPMENT Minimum Design Guidelines for Rural Collector Roads

Chief Engineer Approved(人

 $12 - 4 - 8$ Date

Minimum Rural Collector Road Design Guidelines - Sheet 1 of 2

Footnotes for Minimum Design Guidelines for Rural Collector Roads

- 1- Current traffic may be used to determine the appropriate classification.
- 2- The design speed may not be less than the current posted speed of the overall route.
- 3- For rolling terrain, limited passing sight distance and high percentage of trucks, further analysis should be made to determine if additional lanes are required when ADT is above 7,000.
- 4- For design speeds greater than 50 mph and ADT greater than 1,500 use 12-foot lanes.
- 5- Where bicycle activity is observed, a 4-foot shoulder should be provided.
- 6- For ADT greater than 1,500 use 6 foot shoulders.
- 7- For ADT of 5,000 or greater, a minimum of 4 feot must be paved.
- 8- 1:3 back slopes are allowed where right-of-way restrictions dictate.
- 9- In Districts 04 and 05, where ice is more frequent, superelevation should not exceed 8 percent from the $e_{max} = 10%$ table.
- 10- It may be necessary to increase the radius of the curve and/or increase the shoulder width (maximum of 12 feet) to provide adequate stopping sight distance on structure.
- 11- Radius based on 40 mph. Radii for 50 mph and 60 mph are shown under the RC-2 and RC-3 classifications respectively.
- 12- Radius based on 50 mph. The radius for 60 mph is shown under the RC-3 classification.
- 13- Where the roadway dips to pass under a structure, a higher vertical clearance may be necessary. An additional 6 inches should be added for additional future surfacing.
- 14- The lower value is based on a 40 mph design speed, the middle value for 50 mph and the upper value for 60 mph.
- 15-LRFD for bridge design.

General Note:

DOTD pavement preservation minimum design guidelines or 3R minimum design guidelines. (separate sheets) shall be applicable to those projects for which the primary purpose is to improve the riding surface.

Minimum Rural Collector Road Design Guidelines - Sheet 2 of 2

LOUISIANA DEPARTMENT OF TRANSPORTATION AND DEVELOPMENT Minimum Design Guidelines for Urban and Suburban Collector Roads and Streets

State law requires that the state highway system conform to these guidelines.

Approved CLL to Fee Chief Engineer $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$

Minimum Urban and Suburban Collector Road and Street Design Guidelines - Sheet 1 of 2

Footnotes for Minimum Design Guidelines for Urban and Suburban Collector Roads and Streets

- 1- These guidelines may be used only on a rural roadway section that adjoins a roadway section currently classified as urban. The classification selected should be based on the posted speed.
- 2- For ADT less than 2,000 refer to Exhibit 6-5 on page 425 in the '2004 AASHTO Policy on Geometric Design of Highways and Streets'.
- 3- Applicable to depressed medians only.
- 4- Curb may be used instead of shoulder. Where bicycle activity is observed, a bike lane should be considered.
- 5- If curb will not be used, shoulder widths may be reduced, see Footnote 2. When curb is used on mainline facilities, it shall be placed at the edge of shoulder. When curb is used on 2-lane facilities, 8 foot shoulders will be required if a future center turn lane will be added. Curb will not be placed in front of guardrail.
- 6- Seven and 8-foot widths are limited to residential areas for 30 and 40 mph respectively.
- 7- Cannot be used on multilane roadways (with four or more through lanes) without Chief Engineer's approval.
- 8- If shoulders are used, sidewalks should be separated from shoulder.
- 9. Where shoulders are used, 1:4 minimum fore slopes are required through the limits of minimum clear zone.
- 10-1:2 back slopes are allowed where right of way restrictions dictate.
- 11-It may be necessary to increase the radius of the curve and/or increase the shoulder width (maximum of 12 feet) to provide adequate stopping sight distance on structure.
- 12- Different radii apply at divisional islands. See Footnote 7 for "Minimum Design Guidelines for Urban Arterial Roads and Streets".
- 13- Where the roadway dips to pass under a structure, a higher vertical clearance may be necessary. An additional 6 inches should be added for additional future surfacing.
- 14- The higher value is applicable to roadways with an ADT greater than 6,000.
- 15- These values apply to roadways with 8-foot shoulders. For outside shoulders less than 8 feet, further increase should be proportional to the reduced shoulder width.
- 16-LRFD for bridge design.
- 17-Refer to EDSM II.3.1.4 when sidewalks will be provided and for guardrail requirements.

General Note:

DOTD pavement prescrvation minimum design guidelines or 3R minimum design guidelines (separate sheets) shall be applicable to those projects for which the primary purpose is to improve the riding surface.

Minimum Urban and Suburban Collector Road and Street Design Guidelines - Sheet 2 of 2

LOUISIANA DEPARTMENT OF TRANSPORTATION AND DEVELOPMENT Minimum Design Guidelines for Local Roads and Streets

State law requires that the state highway system conform to these guidelines.

Chief Engineer Approved LL

 $12 - 4 - 00$ Date

Minimum Local Road and Street Design Guldelines - Sheet 1 of 2

Footnotes for Minimum Design Guidelines for Local Roads and Streets

- 1- The design speed may not be less than the current posted speed of the overall route.
- 2- For ADT greater than 2,000, use 12-foot lane widths.
- 3- Lane widths in residential areas may be reduced to 9 feet if necessary. Twelve foot lane widths are preferred in industrial areas.
- 4- Where bicycle activity is prevalent, a paved 4-foot shoulder should be provided.
- 5- For ADT less than 1,500, the minimum shoulder width may be reduced to 4 feet if necessary. For ADT 1.500 to 2.000, use 6 foot shoulders. For ADT over 2,000, use 8-foot shoulders.
- 6- Select the shoulder width that corresponds to the ADT shown in the rural local road guidelines.
- 7- The value shown should be provided on new roadways. A lesser value may be used on existing roads depending on soil stability, right-of-way constraints, the safety record of the road, and the size vehicles using the road. Guidance is available in the AASHTO publication titled 'Guidelines for Geometric Design of Very Low Volume Local Roads (ADT < 400)'.
- 8- In Districts 04 and 05, where ice is more frequent, superelevation should not exceed 8 percent from the emax $= 10\%$ table.
- 9- It may be necessary to increase the radius of the curve and/or increase the shoulder width (maximum of 12 feet) to provide adequate stopping sight distance on structure.
- 10- On roadways with an ADT < 400, a sharper radius may be used on fully superelevated roadways if necessary. For specific values refer to the AASHTO publication titled 'Guidelines for Geometric Design of Very Low Volume Local Roads (ADT < 400)'. Different radii apply at divisional islands.
- 11- Grades 2 percent higher may be used in roral rolling terrain.
- 12- Varies from 14 feet to 28 feet. Refer to the Roadside Design Guide for the applicable value. For spot replacement projects refer to footnote 7.
- 13- LRFD for bridge design.
- 14- Fcr ADT greater than 2,000, use roadway width.
- 15- Refer to EDSM II.3.1.4 when sidewalks will be provided and for guardrail requirements.
- 16- When shoulders are provided, the minimum bridge width shall be the larger of that shown or the roadway width.

General Local Road Notes:

These guidelines shall not apply to:

- a. Dead end roads (open at one end only).
- b. Roads that are dependent on dead end roads for access.

Urban guidelines may be applied to any street for which curb is to be used and the posted speed is less than 50 mph, or any street for which a posted speed of 30 mph or less would be appropriate.

On spot replacement projects the existing geometry and superelevation may remain providing there are no safety problems.

The appropriate local governing body is authorized to make design exceptions for specific items listed in these guidelines, with proper engineering justification.

General Note:

DOTD pavement preservation minimum design guidelines or 3R minimum design guidelines (separate sheets) shall be applicable to those projects for which the primary purpose is to improve the riding surface.

Minimum Local Road and Street Design Guidelines - Sheet 2 of 2

Appendix C: Sample Case Study of Model Application C.1 Introduction

While the developed model of this thesis meets the steps necessary to be developed in the same manner as an MTP's Travel Demand Model (TDM), an example of the application of the model provides proof of its viability to transportation planning. Following the development of a TDM, the study team for a planning project uses it to forecast future travel patterns. This is done for both the network without any additional projects, and for a network with additional projects. The congestion relief of the new projects is often used as part of ranking test projects via a quantitative mean. The application of this thesis's models functions in the same way and two examples will be displayed.

C.2 Background and Study Area

This case study looks at the Lake Charles MTP 2040 that was recently conducted for the Imperial Calcasieu Regional Planning & Development Commission (IMCAL) by Neel-Schaffer, Inc. (NSI). IMCAL selected NSI for the model development and transportation planning process. The TDM created by NSI was made available for this thesis as well as the crash data that was received from the LADOTD for the safety element. This data will be used to train the thesis models to the Lake Charles local crash trends. While the entirety of the study area chosen for the MTP (Figure C-1) could be used, only two corridors and their parallel routes (Figure C-2) are analyzed for this case study for the purposes of brevity. Figure C-3 shows whether a roadway segment was analyzed under the conditions of a new roadway test or a roadway widening test.

Figure C-1: Lake Charles MTP 2040 Study Area

Source: IMCAL, NSI, TDM

Source: IMCAL, NSI, TDM

Figure C-3: Corridors by Test Project Type

The new roadway project analyzes three corridors that are northwest of the city of Sulphur, and is based upon a test project that adds a new north-south route. This project is meant to ease travel on US 90 and W Houston River Rd between WPA Rd and Claiborne St, which are currently existing roadways. WPA Rd and the new roadway are in rural areas while Claiborne St is in a rural area.

The widening project analyzes the impact upon Coleman St, Kirkman Rd, and Enterprise Blvd in Lake Charles based upon widening Kirkman Rd from two lanes to four lanes. All of the streets in the widening analysis are urban roadways. This means that four of the six models developed in the thesis will be tested under planning level conditions.

Source: IMCAL, NSI, TDM

C.3 Model Data

The TDM created by Neel-Schaffer provides much of the data necessary for the crash prediction models. The segments in the TDM contain the

- Length (LENGTH)
- Functional classification (used to develop the CLASS_DESC)
- Model classification (used to develop the MED_PRESENCE)
- Number of lanes (NUM_LANES)
- Traffic volumes (AVG ADT)

In addition to the TDM data by segment, further data needed to be collected. While the HSM methodologies require a considerable amount of data at this stage (see Chapter 2), the developed crash prediction models for the thesis require considerably less data and were easy to obtain. The median widths (MEDIAN_WIDTH) and average lane widths (AVG_LANE_WIDTH) were gathered using GIS and an overlay of the test project segments with an aerial map.

This data was gathered for the model base year, 2013; model horizon year, 2040; and model horizon year with the MTP test projects. Where necessary, such as for the information on the new roadways or a new median created on the widened roadway itself, the data in Appendix B was used to develop the median presence, median width, and average lane width data. Figure C-4 displays a screenshot of the data obtained in the collection process for the base year. This data was also gathered for the two different scenarios to be tested.

Figure C-4: Sample Segment Data

Source: NSI, TDM

C.4 Model Results

The crash prediction models were used for the segments in their respective classes to obtain the predicted crash frequencies of each given scenario. Using the trainset methodology that was described in Chapter 3, the model learned the trends of the Lake Charles area using the provided base year data that was collected. While the statewide set that was developed in the thesis could have been used, the Lake Charles specific data was used to account for local influences upon driver behavior.

After the model learned the trends of the crash data in trainset mode, the testset mode of the models was used for prediction purposes to estimate the crash frequencies in the horizon year without test projects (EC for the existing + committed network in the MTP), and with the test projects (MTP). Table C-1 displays the results of the crash forecasting at the segment level obtained from the model. Table C-2 displays the overall results by test type.

TDM ID	EC_CRASHES	MTP_CRASHES	DIFFERENCE	TEST
1247	1.42	1.23	-0.19	Widening
1264	1.58	1.65	0.06	Widening
1265	1.71	1.40	-0.31	Widening
1266	1.52	1.31	-0.21	Widening
1267	3.43	3.99	0.56	Widening
1269	1.58	1.71	0.12	Widening
1274	0.75	0.86	0.11	Widening
1275	0.73	0.88	0.15	Widening
1284	0.07	2.44	2.37	Widening
1285	1.40	1.21	-0.19	Widening
1287	1.39	3.34	1.95	Widening
1288	1.42	4.05	2.62	Widening
1290	1.57	3.06	1.50	Widening
1291	1.58	3.16	1.58	Widening
1292	1.57	3.09	1.52	Widening
1293	1.49	0.24	-1.25	Widening
1294	1.36	2.77	1.41	Widening
1300	1.54	3.10	1.56	Widening
1301	1.57	3.15	1.58	Widening
1302	1.44	3.34	1.90	Widening
1303	0.99	8.15	7.16	Widening
1313	2.56	15.32	12.76	Widening
1319	1.67	1.34	-0.33	Widening
1321	1.40	2.68	1.28	Widening
1323	2.84	2.90	0.06	Widening
1324	4.40	4.49	0.09	Widening
1326	0.01	$0.02\,$	0.01	Widening
1328	5.04	4.19	-0.85	Widening
1329	2.62	2.69	0.07	Widening
1331	4.87	3.70	-1.17	Widening
1332	1.20	1.51	0.31	Widening
1333	1.12	2.81	1.69	Widening
1334	24.17	24.20	0.03	Widening
1335	2.74	2.83	0.10	Widening
1336	3.81	3.85	0.04	Widening
1342	2.43	2.50	0.08	Widening
1346	4.23	4.01	-0.22	Widening
1584	0.98	9.77	8.79	Widening
1914	0.19	2.79	2.60	Widening
1962	0.19	8.89	8.70	Widening
2018	2.72	2.52	-0.19	Widening

Table C-1: Comparison of EC and MTP Predicted Crashes

Table C-2: Case Study Model Results

As shown in Table C-2, the widening of a roadway increases the amount of crashes that occurs on the affected corridors, with an overall change of about 51 percent. The majority of these increased crashes are on the widened roadway as expected due to the higher ADT that the road now experiences, as well as newly introduced driver behaviors such as weaving. The new roadway actually reduces the amount of crashes per year for the analyzed segments by

about four percent. Figure C-5 displays the overall change that each roadway segments experiences.

Figure C-5: Crash Frequency Changes of Tested Segments

Source: TDM, Thesis Crash Prediction Model

Based upon the results in Table C-2 and Figure C-5, the models are predicting in a reasonable manner and provide engineers and transportation planners with a quantitative means to rank these two test projects.

The new roadway "shifts" traffic crashes away from the existing roadways and onto the new roadway, but providing a safety benefit by reducing the overall crashes between the corridors. Both roadways that parallel the new roadway experience decreased crash frequencies upon all segments. This decreased crash frequency for the affected corridors

means that the project would receive a higher safety ranking in the MTP project selection process.

The widening of Kirkman St. increases the crash frequencies due to the widened road's increased capacity and the amount of traffic it draws from other parts of the network. Almost the entirety of Kirkman Rd. experiences increased crash frequencies. This is expected due to its higher capacity, increased ADT, and different driver behaviors introduced from the increased lanes.

While the roads that parallel Kirkman St. do experience some segments with increased crashes, these are likely due to a realignment of travel patterns from the congestion relief offered by the widened roadway and are not indicative of model failure. Many segments experience reduced crash frequencies as a result of these changed roadway patterns as well. Further analysis shows that the crashes would increase on Coleman St. by six crashes per year while Enterprise Blvd would stay relatively unchanged, meaning that most of the increase crash frequency is upon Kirkman St., the widened roadway, itself. However, due to the overall change that the project creates, a 51 percent increase in crashes per year, the project would receive a lower safety ranking in the MTP selection process due to the increased risk it creates for the general public and users of the roadway.

C.5 Conclusion

This case study of the developed thesis crash prediction models shows that the model can be run at the transportation planning level with available data and obtain reasonable results. The created TDMs of the MTP process provide data that is easy to use and readily accessible. Obtaining the other necessary data is also quickly done and the results of the thesis models can be quickly obtained and interpreted. Using the model results, analysis showed that a new roadway results in reduced crashes while the widened roadways change in travel patterns resulted in a large increase of crashes. This results in one process that would have a higher safety score for project ranking and one with a lower safety score, respectively. Using the mathematical results obtained from these model runs and observing the difference for individual roadways and overall corridors will provide a quick, efficient, and quantitative means for meeting the necessary MAP-21 requirements.

Broussard, Nicholas P. Bachelor of Science, University of Louisiana at Lafayette, Fall 2009; Master of Science, University of Louisiana at Lafayette, Fall 2015 Major: Engineering, Civil Engineering option Title of Thesis: Development of Crash Prediction Models for Transportation Planning Analysis Thesis Director: Dr. Xiaoduan Sun Pages in Thesis: 83; Words in Abstract: 239

ABSTRACT

Transportation planning is a vital and necessary operation for a metropolitan area to grow. As such, and in order to receive Federal funding for transportation projects, metropolitan areas engage in transportation planning as regulated by MAP-21. One element of meeting MAP-21 requirements is addressing the safety of a region. With new requirements by MAP-21, MPOs must demonstrate some sort of performance measure showing changes in the various elements, making quantitative means of displaying these changes ever more important.

The goal of this project was to develop a model or set of models that could produce quantitative results as opposed to the traditional qualitative results gained from professional opinion. This allows for better decision-making for test project scoring in transportation plans and additionally in meeting MAP-21 requirements. Following a review of the current available methodologies and an inventory of other states' efforts to develop crash prediction models, mathematical modeling for Louisiana statewide crash prediction formulae were attempted. These results and the methodology were deemed unsuitable for the desired outcomes and the use of SVR modeling was explored.

The use of the SVR models described in this report produce acceptable results, have been validated for use in forecasting, and allow for the comparison of conditions between base data, future years, and future years with MTP test projects included. The results of these models provide transportation planners increased means to determine project rankings based on safety as well as satisfy MAP-21 requirements.

Biographical Sketch

Nicholas Broussard is the son of David and Linda Broussard. He was born on March 23, 1987 in Lafayette, Louisiana. Mr. Broussard received his Bachelor of Science in Civil Engineering from the University of Louisiana at Lafayette in the fall of 2009. He received his Master of Science in Engineering from the University of Louisiana at Lafayette in the fall of 2015. As an undergraduate, he was a member the Institute of Transportation Engineers, Chi Epsilon, and UL Lafayette Honors. Mr. Broussard recently received his Professional Engineer licensure and currently works for Neel-Schaffer, Inc.

