IMPROVED GUIDELINES FOR RECALIBRATION OF PREDICTIVE MODELS OVER TIME BASED ON MODEL UNCERTAINTY

A Thesis

by

BHARADWAJ BOMMANAYAKANAHALLI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Dominique Lord
Committee Members,	Yunlong Zhang
	Daren B. H. Cline
Head of Department,	Robin Autenrieth

May 2018

Major Subject: Civil Engineering

ABSTRACT

The Highway Safety Manual (HSM) summarizes the safety performance functions (SPFs) of various facility types. The primary use of SPFs is to estimate the safety performance (i.e., the number of crashes by severity level) of different facilities based on geometric and traffic variables. The SPFs were developed using the negative binomial (NB) regression model based on crash data obtained from a selected number of states and cities in the United States and Canada. Applied directly to the local jurisdictions, SPFs may yield biased or incorrect results. Therefore, calibration of the SPFs or predictive models is an important step before applying them to local jurisdictions. Moreover, it is also necessary to recalibrate SPFs over time to account for variations in factors that cannot be accounted for directly in SPFs, such as changes in driver behavior, crash-reporting thresholds, etc. The calibration factor (for a specific facility type) is defined as the ratio of the observed number of crashes to the predicted number of crashes. The HSM recommends that SPFs be recalibrated every 2 to 3 years. However, these guidelines are not based on sound research or reliable criteria. The lack of appropriate guidelines can lead to two types of errors: recalibrating of the models when it is not needed, and not recalibrating them when such a need arises.

The aim of this thesis is to develop guidelines regarding when or how often SPFs should be recalibrated. To this end, two methodologies were created related to the variance or uncertainty associated with the SPFs, and the guidelines were developed using statistical principles. These guidelines are that SPFs should be recalibrated when (i) the total number of crashes that occur in a network of similar types of facilities falls beyond the prediction intervals

of the predicted or estimated total number of crashes in that same network; or (ii) the calibration factor developed in a specific year is statistically significantly different than 1 (based on coefficient of variation (CV) of the SPF and the Calibration Factor C).

Both approaches were tested on several intersections and segment datasets from Michigan and Toronto. The results show that both approaches are feasible and could provide safety analysts with better and more reliable guidelines regarding when SPFs should be recalibrated. However, the methodologies developed in this thesis cannot be applied to the SPFs developed in the HSM since the information needed to evaluate the variance of SPFs is not available in this manual. The results of both the methodologies were compared to the results of a methodology recently proposed in the literature that can be applied to HSM SPFs and uses a fixed threshold value of C-factor error estimate (say 10%). This study indicated that the 10% error is a reasonable value to use for re-calibrating models.

The shortcomings of these methodologies include the need to develop a new SPF (which is time-consuming and work-intensive process) and to collect extensive data every year. When data is available every year, the practitioner might as well estimate a new calibration factor every year instead of needing to know the frequency of recalibration or use an approximate method (C-proxy). Future research in this area should focus on identifying the minimum data requirements for both methodologies proposed in this thesis.

DEDICATION

This thesis is dedicated to all my teachers and the people who accepted me for who I am

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Dominique Lord who came up with this research idea and helped me a lot over the course of the research. I would also like to thank Dr. Yunlong Zhang and Dr. Daren Cline for serving as the members of my thesis committee. Moreover, I would like to thank Mr. Mohammad Ali Shirazi who spent a lot of his time in clarifying my doubts and helped me in the process of understanding the problem statement better.

The Transportation Engineering Graduate Program at Texas A&M has been one of the best experiences in my life. The term papers and design projects I did as part of my coursework improved my critical thinking abilities and reinforced my interest in this specialization. I would like to specially thank Dr. Gene Hawkins and Dr. Mark Burris under whom I took some of the best courses in the transportation program.

I am immensely grateful to Mr. Adam Pike who offered me the position of Graduate Research Assistant in the Signs and Markings program of Texas A&M Transportation Institute. Under his supervision, I worked on several projects related to pavement markings and developed a good understanding of research in that area. The paper that we both wrote “Development of a pavement marking life cycle cost tool” has been selected for publication in the Transportation Research Record (TRR). I also worked as a Graduate Student Worker under Dr. Subasish Das in the Roadway Safety division. Under his supervision I developed skills in crash data reduction and modeling. I received exposure to the various kinds of safety projects during this time.

I am very grateful to my parents for supporting my decision to pursue higher studies in the United States. I also would like to thank almighty God for blessing me with whatever I have in my life today.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Dr. Dominique Lord and Dr. Yunlong Zhang of the Department of Civil Engineering and Dr. Daren Cline of the Department of Statistics.

The Toronto dataset used for analysis in this thesis was provided by Dr. Dominique Lord and the Michigan dataset was provided by Dr. Srinivas Geedipally of the Texas A&M Transportation Institute. The procedure to evaluate the variance associated with the predictive models (of crashes) has been adopted from the work done by Wood published in 2005 in Accident Analysis and Prevention. The concept proposed by Dr. Ezra Hauer in the NCHRP report HR 20-7(332) published in 2014 was used for evaluating the coefficient of variation of the calibration factor.

All other work completed for the thesis was completed by the student independently.

Funding Sources

There are no outside funding contributions to acknowledge related to the research and compilation of this document.

NOMENCLATURE

AIC	Akaike Information Criterion
AADT	Annual Average Daily Traffic
BIC	Bayesian Information Criterion
CI	Confidence Interval
C-Factor	Calibration Factor
CMF	Crash Modification Factor
CV	Coefficient of variation
DOT	Department of Transportation
HSM	Highway Safety Manual
m	Safety at a site
MAD	Median Absolute Deviance
MSPE	Mean Square Prediction Error
NB	Negative Binomial
PI	Prediction Interval
SPF	Safety Performance Function
Y	Observed number of crashes at a site
μ	Poisson Mean

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

Page

CHAPTER I
INTRODUCTION

This chapter presents the problem statement and objectives of this study. This is followed by an overview of the thesis.

1.1 PROBLEM STATEMENT

Traffic safety analysts often examine different alternatives to the design of a roadway or intersection to evaluate the safety associated with each of the designs. Moreover, they are also interested in identifying crash hotspots and the effects of traffic and geometric variables on the number of crashes and their severity. These analysts should be able to predict the frequency and severity of crashes beforehand to estimate the safety effects in the design before the different projects are ranked based on safety or sent into the construction phase. To achieve this, crash prediction models must be developed. There are several challenges associated with developing predictive models. First, analysts need to appropriately identify all the variables (geometric, traffic, human factors, etc.) that could potentially affect the number of crashes on a specific facility. Moreover, in addition to data quality, the sample size of the data should be large enough to develop a reliable model. All these activities require significant time and effort.

The first edition of the Highway Safety Manual (HSM) (AASHTO, 2010) was created to enable different safety analyses to be conducted in a simple manner. The current edition of the HSM summarizes safety performance functions (SPFs) for rural two-lane roads and multilane highways, urban/suburban arterials, freeways, and interchanges. The SPFs in the HSM were developed using a negative binomial (NB) regression model with data collected from a selected number of cities and states in the United States and Canada (Bahar, 2014). The SPFs in the HSM refer to the safety performance of roadway segments or intersections for the base conditions (note: the base conditions vary by facility type and are described in more detail in the HSM). The base conditions usually refer to the most commonly used highway design and operational characteristics, such as 12-ft lanes and no-turning lanes at intersections. The effect of variations from the base conditions (like an 11-ft lane) is considered in the SPFs using crash modification factors (CMFs). The value of the CMFs is an indication of the effectiveness of a treatment. A value of CMF greater than 1 indicates an increase in crash frequency due to the treatment, whereas a value less than 1 indicates a decrease.

The SPFs described in the HSM cannot be directly used for a particular jurisdiction since the SPFs do not directly consider factors such as driver behavior, weather, crash-reporting thresholds, and animal-related crashes, which are expected to vary from one jurisdiction to the next. Hence, the calibration of these SPFs to a local jurisdiction is necessary to better predict the number of crashes in a particular jurisdiction. Although this might be challenging, practitioners can also develop a jurisdiction-specific model instead of recalibrating the SPFs documented in the HSM. Similarly, even within the same jurisdiction, predictive models should be frequently recalibrated (or a new model fitted every few years); this is necessary since the frequency of crashes is expected to change with changes in driver behavior (further affected by changes in the demographics of a region, driver awareness programs by the DOTs), vehicle characteristics (advanced warning systems, better braking system), and roadway characteristics, among others.

Appendix A of part C of the HSM describes a method for calibrating SPFs through the multiplication of a scalar calibration factor by the SPFs. The same calibration methodology can be used to recalibrate the models over time. The HSM recommends that SPFs be recalibrated every 2 to 3 years. However, these guidelines are not supported by sound research or reliable criteria. The HSM methodology recommends collecting data from 30 to 50 sites (which are randomly sampled from a population for a given facility) with a yearly total of at least 100 crashes. A practitioner adopting the HSM guidelines may experience one of the following unfavorable scenarios: 1) recalibrating the model and later realizing it was not necessary; 2) not recalibrating the model immediately after such a need arises. Research is needed in this area to give practitioners better knowledge regarding the frequency of recalibration.

This thesis aims to develop better guidelines for the recalibration of the models over time. The research uses the uncertainty associated with the predictive models and develops two methodologies to decide when recalibration is needed.

1.2 THESIS ORGANIZATION

The thesis is organized into six chapters. Chapter 1 discusses the need for calibration of SPFs and the issues with the HSM methodology of calibration. Chapter 2 describes the HSM methodology of calibration, and then summarizes the research studies related to the calibration of SPFs to the local conditions and the studies related to the calibration of models over time. This chapter also summarizes the issues and shortcomings that researchers have found regarding the HSM methodology. Chapter 3 describes the various tasks of the research, beginning with the introduction to theory behind the proposed methodologies of recalibration. The chapter also describes the characteristics of the datasets used in the study and the model forms used for segment and intersection safety performance functions. Next, Chapters 4 and 5 discuss the results of the analysis after applying the proposed methodologies and compares these results with those obtained by applying the methodology proposed in Shirazi et al. (2017). Finally, Chapter 6 summarizes the results of the study, discusses its limitations, and proposes future research options in this area.

CHAPTER II

LITERATURE REVIEW

This chapter provides background information on the calibration of SPFs, reviews studies on the calibration of SPFs to local conditions, and discusses the methodologies developed by researchers for the recalibration of models over time. First, Section 2.1 describes the HSM methodology for calibration of the SPFs. Section 2.2 then reviews the studies related to the calibration of SPFs to local jurisdictions. The section also describes the challenges faced by researchers in adopting the HSM methodology and the comparison of its performance to that of the methodologies developed by researchers. Subsequently, Section 2.3 reviews the studies on the recalibration of models over time and the methodologies developed by researchers regarding the frequency of recalibration. Finally, the chapter summarizes the limitations of the existing methodologies and issues that previous research has not addressed.

2.1 HSM METHODOLOGY FOR CALIBRATION OR RECALIBRATION OF SPFs

The first edition of the HSM (AASHTO, 2010) specifies a common procedure for the recalibration of SPFs of all types of facility. Lord et al. (2016) reworded the five steps from Appendix A of part C of the HSM in simpler language, as follows:

1. **Identifying the facility type:** The first step of the calibration procedure is to identify the facility type. It should be noted that a separate calibration factor must be developed for SPFs of each of the facility types. The first edition of the HSM has the SPFs for rural two-lane roads and multilane highways, urban/suburban arterials, freeways, and interchanges.

2. **Selection of sites to be used in the calibration:** The HSM recommends using a random sample of 30-50 sites for a facility type that experiences a total of at least 100 crashes per year. The random sampling ensures that there is no bias towards sites that have a very large or very small number of crashes. The practitioner may use a larger sample in the same procedure, if readily available.
3. **Obtaining data for the selected sites:** The next step in the calibration procedure is to obtain the data for the sites selected in the second step. The data needed to conduct the recalibration is of two types: required data and desirable data. The required data is needed to conduct the recalibration. It varies by the type of facility but includes the AADT and the geometrics of the roadway for all facility types. The desirable data is not needed to conduct the calibration, but using it improves the estimation of the calibration factor.
4. **Predicting crashes on the selected sites:** Based on steps 1, 2, and 3, the appropriate SPFs and CMFs should be applied and the number of crashes on each site should be predicted. The data points collected in the previous step are used as values of the variables in the SPF, or appropriate CMFs are evaluated and multiplied as scalar factors by the base SPF.

5. **Evaluating the calibration factor:** An estimate of the calibration factor can be obtained as the number of observed crashes divided by the number of predicted crashes:

$$C = \frac{\sum N_{obs}}{\sum N_{pre}} \quad (2.1)$$

where N_{obs} = Number of crashes observed on the selected sites

N_{pre} = Number of crashes prediced on the selected sites

The recalibrated SPF is obtained by multiplying the base model by the calibration factor (as a scalar multiplicative factor).

$$N_{pre} = N_{base} * CMF_1 * CMF_2 * CMF_3 * \dots\dots\dots * CMF_n * C \qquad (2.2)$$

The HSM methodology of recalibration is adopted from the research by Harwood et al. (2000). Theirs is the first report to discuss the issue of recalibration and its importance in accident prediction, and it provides an elaborate description of the data needed to this end. The HSM guidelines are an updated version of the guidelines specified in Harwood et al.'s study. The methodology described in Harwood et al. assumes that the base model form remains unchanged for the different jurisdictions, as do the coefficients of the explanatory variables. The calibration is done through multiplication of a scalar calibration factor.

Persaud et al. (2002) evaluated the methodology proposed by Harwood et al. to develop the calibration factors for signalized and unsignalized 3-legged and 4-legged intersections in Toronto. In their study, they first developed jurisdiction-specific models for the Toronto intersections. Then, models developed in Vancouver and California were calibrated for those Toronto intersections. The prediction capabilities of the jurisdiction-specific models (of Toronto) were compared to the recalibrated models. Persaud et al. (2002) indicate that Harwood et al.'s (2000) methodology considers a single calibration factor for the entire AADT range for a specific facility type, whereas the results of their own study indicate that developing a different calibration factor for the different AADT ranges may be more appropriate.

2.2 CALIBRATION OF SPFs TO LOCAL CONDITIONS

Several researchers have used the methodology described in the HSM to calibrate SPFs to their local jurisdictions, and have found several limitations to the HSM procedures. These will be elaborated on later in the thesis. Lord et al. (2016) reviewed several studies related to the calibration of SPFs. Oregon was among the first states to calibrate the HSM SPFs to the local conditions, and the involved researchers indicated the following issues they faced when adopting the HSM methodology of calibration (Xie et al., 2011):

- Pedestrian volumes at urban intersections were not available in any of their databases. Similarly, the researchers could not find the signal phasing plan for the minor roads, and the AADT were only available for a few of the minor roads. Therefore, the researchers had to develop a methodology to calculate the minor road AADT values.
- The researchers found that for certain facility types, the target of a total of 100 crashes per year could not be reached. In these cases, they used the entire available sample.
- A significant amount of time and effort was required to evaluate the calibration factor, and even to obtain the minimum sample size.

Despite these limitations, other researchers have often used the HSM methodology to compare its results with the jurisdiction-specific models or calibration methodologies they have developed. Brimley et al. (2012) used the HSM methodology and calibrated the SPFs for rural two-lane two-way roads in the state of Utah. They also developed region-specific SPFs based on the NB regression and two variations: their first two models were conventional models without data transformations, while the other two considered a log transformation of AADT. The authors evaluated these four model outputs along with the output of calibrated HSM SPF (calibrated to Utah conditions) to find the model of best fit, using Bayesian information criteria (BIC). The researchers observed that four variables were significant in all four developed models: AADT, speed limit, segment length, and Percentage of multi-unit trucks. In contrast, the HSM base SPFs only consider the AADT and segment length as significant variables, since these are the variables with a maximum effect on crash frequency and this data is easily available (or can be collected). Furthermore, the researchers also found that the region-specific model considering the log transformation of AADT and the four significant variables was the best fit model.

In another study, Mehta and Lou (2013) calibrated the HSM SPFs in the state of Alabama for four-lane divided highways and rural two-lane two-way roads. The researchers first evaluated the calibration factors using two methodologies: the HSM methodology and a methodology considering a NB regression model with a constant calibration factor. They observed that the HSM methodology performed better than the second one. Next, they developed four region-specific SPFs for the rural two-lane two-way roads. The first model used the same form as the HSM base SPF, with AADT raised to a power and segment length as the offset. The second model considered the log transformation of both AADT and segment length and an additional term containing the number of minor junctions in a segment as a variable. The third model considered the log transformation of both AADT and segment length in addition to lane width and speed limit as variables, along with a dummy variable for effect of the year on intercept. Finally, the fourth model was formed like the HSM SPF but additionally considered lane width and shoulder width as variables. The researchers used the log-likelihood (LL) value, Akaike information criterion (AIC), median absolute deviance (MAD), mean square prediction error (MSPE), and mean prediction bias (MPB) to compare the models. They found the third model to be the best fit for Alabama (even better than the calibrated HSM model). This study indicates the significance of the lane width, speed limit, and year, which are not considered directly in the base HSM model for this facility type (the effect of lane width is considered in the form of a CMF in HSM methodology).

Furthermore, Brown et al. (2014) calibrated the HSM SPFs in the state of Missouri. In their paper, they discuss some of the challenges they faced during the recalibration process, like the need to refer to several types of data sources, sample size requirements, and the tradeoff between minimum segment length and homogeneity of the sections.

In their study, Martinelli et al. (2009) developed the calibration factors for the SPFs of rural two-lane roads located outside the United States in the province of Arezzo, Italy. They considered two types of models: a full model, which considered all variables, and a base model, which incorporated the effect of variables in the form of CMFs. They further examined the effect of averaging the parameters instead of applying the model to every section. Thus, they studied a total of four models. Furthermore, they also investigated the following three ways of evaluating the calibration factors:

1. The observed number of crashes divided by the predicted number of crashes is the calibration factor (which is the HSM methodology).
2. The density of the observed number of crashes divided by the density of the predicted number of crashes is the calibration factor.
3. The weighted average (weights being the segment length) of the observed crashes divided by the weighted average of predicted number of crashes is the calibration factor.

Martinelli et al. (2009) thus evaluated a total of 12 calibration factors. They found that the model that considered the CMFs and used the stratified classes and used the third method (weighted average) was the best fit model.

2.3 RECALIBRATION OF THE SPFs OVER TIME

So far, this chapter has reviewed the work done by researchers to calibrate the SPFs in the HSM to their local jurisdictions. Even within the same jurisdiction, however, predictive models may need to be updated frequently because there is an expected change in driver behavior (due to changes in the demographics of the region, driver awareness programs by the DOTs, etc.), vehicle characteristics (advanced warning systems, better braking system, etc.), and roadway characteristics. Updating the predictive models can again be done through either scalar calibration (which is updated over time; this is the methodology described by the HSM) or refitting the model every few years, both of which are highly time-consuming tasks that require a great deal of effort. However, of the two methods, the evaluation of the scalar calibration is relatively easier. Researchers have examined both these approaches and developed different procedures to recalibrate the models over time.

In the first study, Connors et al. (2013) investigated the recalibration of the models over time through scalar recalibration and refitting of the models. The researchers considered five different goodness of fit (GOF) criteria: absolute value of mean error (AME), root mean square error (RMSE), root mean square relative error (RMSRE), scaled deviance (SD), and median absolute deviance (MAD). Finally, they found that the best scalar factor depends on the adopted GOF criteria. In other words, the same scalar factor does not satisfy all the GOF criteria.

In a subsequent study, Wood et al. (2013) first studied the impact of model complexity on its temporal transferability. They defined the complexity of a model based on the variables considered in that model. The most basic model considers only the traffic flow as a variable, while the more complex models differentiate between the types of accidents and relate each type to traffic flow and several other parameters. They found that it was easier to transfer the more complex models over time compared to the less complex ones. Furthermore, the authors also examined two model-updating methods: the first considers the same model form as the initial fit model but recalculates the parameters using the latest data, while the second method entails scalar calibration of the models. Wood et al. (2013) conclude that both these strategies are better than developing a new crash prediction model.

More recently, Shirazi et al. (2017) developed a methodology for recalibration over time that requires the collection of very little information. This methodology has the following data requirements:

- For segment models: total number of crashes on the entire network for the specific facility type, the mean value of ADT/AADT, and total segment length.
- For intersection models: total number of crashes, average traffic flow for both major and minor streets, and total number of intersections.

The methodology involves calculating a parameter known as C-proxy, which is evaluated as

$$\check{C} = \frac{N_{obs}^{T}}{e^{b_0+b_1*\ln(\bar{F})} * L^{T}} \ (segment\ models) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.3)$$

$$\check{C} = \frac{N_{obs}^{T}}{e^{b_0+b_1*ln(\overline{F_{major}})+b_2*ln(\overline{F_{minor}})} * n_{total}} \ (intersection\ models) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.4)$$

where

$\check{C} = C - proxy$

$N_{obs}^{T} = the\ total\ number\ of\ crashes\ in\ network$

$\bar{F} = the\ mean\ value\ of\ AADT\ on\ all\ sites\ chosen$

$L^{T} = the\ total\ combined\ length\ of\ all\ sites$

$\overline{F_{major}} = the\ mean\ value\ of\ the\ traffic\ flow\ on\ major\ streets$

$\overline{F_{minor}} = the\ mean\ value\ of\ the\ traffic\ flow\ on\ minor\ streets$

$n_t = total\ number\ of\ intersections\ in\ network$

Shirazi et al. (2017) recommended evaluating the C-proxy periodically and calculating the percentage change in it compared to the C-proxy evaluated in the reference year (the year in which the model was recalibrated). The practitioner has the choice of choosing a threshold within which this value should lie; the authors set a threshold of 10% and validated the methodology using datasets from Texas and Michigan.

In another study, Saha et al. (2017) used the Bayesian estimation technique to establish guidelines for the frequency of recalibrating models. The authors' primary hypothesis is to evaluate the variation in the calibration factors for different facility types computed once every year, once every 2 years, and once every 3 years to determine the frequency of updating. The results of their study indicate that when the variation between C-factors (evaluated considering the total crashes) is less than or equal to 0.01, the model for 4-legged signalized intersections should be recalibrated every year, and the models for other facilities should be recalibrated every 2 years. Their results further suggest that when the variation (evaluated considering the total crashes) is more than 0.01 and less than 0.05, the C-factors should be updated every year or every 2 years for 4-legged intersections, and every 3 years for other facility types. According to Saha et al. (2017), the limitations of their study include lack of transferability to other jurisdictions and the fact that the data used only comes from arterial urban and suburban roads in Florida.

2.4 CHAPTER SUMMARY

In summary, this chapter has presented a review of literature on the HSM methodology of calibration, research conducted on calibration of SPFs to local jurisdictions and recalibration of SPFs over time. A key issue in adopting the HSM methodology for calibration of SPFs is the difficulty in meeting the required sample size for certain facility types. Moreover, collecting the required data for the recalibration of the models is difficult due to poor quality of crash data and lack of data for certain necessary variables. Several data sources should be considered, and some assumptions need to be made for some variables. This requires much man power and the process is time consuming.

The following are some of the issues that previous research has not been able to address:

1. Lack of transferability of the results, i.e. the results obtained in one jurisdiction cannot be directly applied to other jurisdictions.
2. Lack of sound statistical criteria while choosing a threshold of acceptable error in the prediction of the C-factor.
3. The studies always assume point estimates for the values predicted by SPFs. However, variance (uncertainty) is associated with the SPFs themselves, which needs to be considered for the recalibration guidelines.

Therefore, the objective of this thesis is to develop guidelines for recalibration of SPFs by accounting for the uncertainty associated with the SPF. These guidelines are based on statistical criteria.

CHAPTER III

METHODOLOGY

This section describes the methodology used to accomplish the study objectives. Section 3.1 describes the background theory used to evaluate the variance associated with SPFs. The section also describes the two methodologies developed for recalibration in this thesis. Section 3.2 provides details on the datasets (intersection and segment) used to apply these methodologies. Subsequently, Section 3.3 describes the model forms for the intersection and the segment model used in this study. Finally, Section 3.4 briefly describes the contents of the final chapter of the thesis.

3.1 DEVELOPING A METHODOLOGY FOR RECALIBRATION

In this study, guidelines were developed for the recalibration of models by considering the uncertainty associated with SPFs. SPFs serve to predict the number of crashes based on variables such as traffic and roadway characteristics. Researchers have studied several different models to identify the best fit model for the crash data; a summary of all these models can be found in the works of Lord and Mannering (2010) and Mannering and Bhat (2014).

The general nature of crash data is that it displays a high degree of randomness and usually exhibits over-dispersion. The most commonly used or popular model that can handle over-dispersion is the NB model. This is the model used for the predictive methodology described in the HSM. Hence, this is also the model used in the present study.

The probability mass function (pmf) of the NB distribution is:

$$P(Y = y, \mu, \phi) = \frac{\Gamma(\phi + y)}{\Gamma(\phi)\Gamma(y + 1)} \left(\frac{\phi}{\mu + \phi}\right)^{\phi} \left(\frac{\mu}{\mu + \phi}\right)^{y} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

where

Y= the observed crash count;

μ = mean response; and

ϕ = inverse of the dispersion parameter of the NB distribution.

The NB model can be used to model the crash counts when a gamma distribution is assumed for the safety *m* between sites with similar flows along with the assumption of Poisson distribution (the mean of this distribution is m) for *Y* at a given site with safety *m*. The following section describes the procedure to evaluate the variance associated with the parameters of the NB model.

The two methodologies developed in the thesis use the uncertainty associated with the SPFs to develop the guidelines. The first methodology uses the PI of the safety (m) for the entire network and assess if the observed counts in a future year lie within the PI of the safety (m) in that year. If the observed counts are outside the PI, we need to recalibrate. The second methodology uses the CV of the SPF and the calibration factor and evaluates the calibration factor every year. The main assumption here is that the CV of the calibration factor should not be larger than the CV of the predictive model. If it is, then the uncertainty of C is larger than that of the model, which is implies that the calibration factor is unreliable or biased. The recalibration is recommended when the calibration factor in any year is significantly different than 1. The theory behind the methodologies is described in the subsequent sections.

3.1.1 FIRST METHOD - PI OF THE SAFETY FOR THE ENTIRE NETWORK

Wood (2005) developed a methodology to calculate the confidence interval (CI) and prediction interval (PI) for a class of models known as generalized linear models (GLMs) with Poisson or NB errors. The NB model results when the safety *m* between sites with similar flows along with the assumption of Poisson distribution (the mean of this distribution is m) for *y* at a given site with safety *m*. This author documented a methodology for evaluating the following:

1. True accident rate μ (CI)
2. NB model – PI for safety at a new site (*m*) and crash rate at a new site (*y*).
3. Poisson model – PI for crash rate at a new site(*y*).

The GLMs in Wood's study use a log link function, meaning that the logarithmic value of μ can be expressed as a linear function of the model parameters. For example, for a model predicting the crashes with a single flow F (Wood, 2005):

$$\mu = \beta_0 * F^{\beta_1} \quad (3.2)$$

$$\eta = \log\mu = \log\beta_0 + \beta_1 * log(F) = \beta_0' + \beta_1 * log(F) \quad (3.3)$$

Wood references the work of Dobson (1990) on the standard generalized model theory, which states that asymptotically, the estimates of β_0' and β_1 denoted by of b_0' and b_1 have a bivariate normal distribution.

$$\begin{bmatrix} b_0' \\ b_1 \end{bmatrix} \sim N\left(\begin{bmatrix} \beta_0' \\ \beta_1 \end{bmatrix}, I^{-1}\right) \quad (3.4)$$

The estimates are unbiased. The inverse of the information matrix *I* is the covariance matrix of the estimates. Based on the above theory, η is asymptotically normally distributed, hence μ is approximately log normally distributed. Thus, the 95% confidence interval for η can be given by $\hat{\eta} \pm 1.96 * \sqrt{Var(\hat{\eta})}$ and the 95% confidence interval for μ is given by $e^{\hat{\eta} \pm 1.96 * \sqrt{Var(\hat{\eta})}}$:

$$Var(\hat{\eta}) = Var(b_0' + b_1 * \log F) = Var(b_0') + 2 * \log F * Cov(b_0', b_1) + (log\, F)^2 * Var(b_1)$$

$$= I_{11}^{-1} + 2 * \log F * I_{12}^{-1} + (\log F)^2 * I_{22}^{-1} \quad \ldots\ldots (3.5)$$

$$\hat{\eta} = (1, \log F)(b_0', b_1)^T \ \ (T\ denotes\ transpose) \ldots\ldots (3.6)$$

$$Var(\hat{\eta}) = (1, \log F) I^{-1} (1, \log F)^T \ldots\ldots (3.7)$$

where I^{-1} is the inverse of information matrix i.e. covariance matrix.

It should be observed that to evaluate $Var(\hat{\eta})$ the covariance matrix needs to be obtained for the parameters. The statistical software used to develop the SPFs using NB regression like R, SAS have the capability of directly outputting the covariance matrix.

Wood (2005) uses the work of Maher and Summersgill (1996), who developed an approximate normal distribution for the lognormal distribution of $\hat{\mu}$ as

$$\hat{\mu} \sim N(\mu_0 = \mu, \sigma_0^2 = \mu^2 * Var(\hat{\eta})) \ldots\ldots (3.8)$$

Wood (2005) developed the CI and PI for Poisson model and NB model. However, in the present study, only the NB model is used. Hence, only the development of CI and PI for NB is discussed.

3.1.1.1 Prediction interval for the safety (m)

Wood's (2005) methodology uses the assumption of an approximate normal distribution of $\hat{\mu}$, as described earlier. A gamma distribution is assumed for m; thus, to calculate the prediction interval, the gamma distribution should be mixed with a normal distribution. The mean and variance of the resulting distribution are μ_0 and $\sigma_0^2 + (\sigma_0^2 + \mu_0^2)/\phi$ where ϕ is the inverse of the dispersion parameter. Again, assuming normality for the above distribution and substituting the values of mean and variance, the 95% PI for safety (m) is

$$\hat{\mu} \pm 1.96 * \sqrt{\hat{\mu}^2 * Var(\hat{\eta}) + \frac{\hat{\mu}^2 * Var(\hat{\eta}) + \hat{\mu}^2}{\phi}} \quad (3.9)$$

3.1.1.2 Prediction interval for number of accidents (y)

Wood's (2005) methodology uses the assumption of an approximate normal distribution of $\hat{\mu}$, as described earlier. The prediction interval of the Y is calculated using a mixture of NB and a normal distribution. The mean and variance of the resulting distribution are μ_0 and $\sigma_0^2 + \frac{\sigma_0^2 + \mu_0^2}{\phi} + \mu_0$. Wood (2005) uses Chebyshev's one-sided inequality (Feller, 1966) to evaluate the PI for y:

$$P(Y - \mu_Y \geq t\sigma_Y) \leq \frac{1}{1+t^2} \; for \; t > 0 \quad (3.10)$$

To evaluate the 95% confidence interval, the probability of Y exceeding the upper limit of CI should be less than 5%.

$$t = \sqrt{19}$$

$$P(Y \geq \mu_Y + \sqrt{19}\sigma_Y) \leq \frac{1}{20} \quad (3.11)$$

Lord (2008) developed a procedure for calculating the variance and CI of the product of the baseline prediction models and the CMF's. The conventional form of the predictive model can be described as

$$\mu_{final} = \mu_{baseline} * CMF_1 * CMF_2 * CMF_3 * \dots \dots \dots * CMF_n \quad (3.12)$$

where

$\mu_{final} = final\ predicted\ number\ of\ crashes\ per\ unit\ of\ time$

$\mu_{baseline} = Predicted\ number\ of\ crashes\ with\ an\ SPF$

$CMF_1 * CMF_2 \ldots * CMF_n = product\ of\ CMFs\ assumed\ to\ be\ independent$

Lord (2008) references the work of Ang and Tang (1975) on the multiplication of the independent random variables. Let "p" be the product of the independent random variables $y_1, y_2, y_3, \ldots, y_n$.

$$p = y_1 * y_2 * y_3 * \ldots * y_n$$

Mean of the product p

$$p = y_1 * y_2 * y_3 * \ldots * y_n$$

$$E[p] = E[y_1] * E[y_2] * E[y_3] * \ldots * E[y_n] \qquad (3.13)$$

$$\lambda_p = \lambda_{y1} * \lambda_{y2} * \lambda_{y3} * \ldots \ldots * \lambda_{yn}$$

Variance of the product p

$$p^2 = y_1^2 * y_2^2 * \ldots . * y_n^2$$

$$E[p^2] = E[y_1^2] * E[y_2^2] * \ldots . * E[y_n^2]$$

$$(\lambda_p^2 + v_p) = (\lambda_{y1}^2 + v_{y1}) * (\lambda_{y2}^2 + v_{y2}) * \ldots . . (\lambda_{yn}^2 + v_{yn})$$

$$v_p = (\lambda_{y1}^2 + v_{y1}) * (\lambda_{y2}^2 + v_{y2}) * \ldots . . (\lambda_{yn}^2 + v_{yn}) - \lambda_p^2 \qquad (3.14)$$

The variance of the parameters of the baseline model was evaluated earlier by Wood (2005), and the methodology to evaluate the variance with the CMFs is described in Lord (2008). In summary, the variance and the 95% PI of the parameters are given in Table 1.

Table 1. Summary of the estimated variance and 95% CI/PI for the quantities

Reprinted with permission from Elsevier (Lord, 2008)

Quantity	**Estimated Variance**	**95% CI/PI of quantities**
μ	$\hat{\mu}^2 * Var(\hat{\eta})$	$[\frac{\lambda_p}{e^{1.96*\sqrt{v_{p\mu}}}}, \lambda_p * e^{1.96*\sqrt{v_{p\mu}}}]$
m	$\hat{\mu}^2 * Var(\hat{\eta}) + \frac{\hat{\mu}^2 * Var(\hat{\eta}) + \hat{\mu}^2}{\phi}$	[Max $\{0, \lambda_p - 1.96 * \sqrt{v_{pm}}), \lambda_p + 1.96 * \sqrt{v_{pm}}]$
Y	$\hat{\mu}^2 * Var(\hat{\eta}) + \frac{\hat{\mu}^2 * Var(\hat{\eta}) + \hat{\mu}^2}{\phi} + \hat{\mu}$	$[0, \lambda_p + \sqrt{19 * v_{py}}]$

This methodology considers the entire network and evaluates the total number of observed crashes and predicted crashes (y). Similarly, it also evaluates the sum of the safety (m), the variance associated with the safety, and the variance of the predicted value y. The variance can be directly summed since the values of the parameters (m, y) at each site are independent of other sites.

The distribution of the sum of the parameters must be investigated to calculate the PIs of the sums. The central limit theorem (CLT) states that the mean and the sum of a random sample from an arbitrary distribution have an approximately normal distribution when the sample size is sufficiently large (Kwak and Kim, 2017). It can be represented as

$$Sum\ of\ the\ mean(x_i) \sim Normal(Sum(x_i), Sum(Var(x_i))) \ldots\ldots\ldots\ldots\ldots\ldots \quad (3.15)$$

3.1.1.3 Steps in the first method

Based on the above discussion, the first method is divided into the following four steps:

- Develop the SPF (flow-only model ignoring the effect of other traffic and geometric variables) using data from the initial period and obtain the variance-covariance matrix for this SPF.
- Apply the SPF to the entire network in the subsequent period and evaluate the variance of the safety (m) and the predicted number of crashes for each of the individual sites.
- Compute the total observed number of crashes for the entire network, the total predicted number of crashes (using the latest calibrated model), the sum of the variance of the safety (m), and the sum of the variance of the predicted mean for the entire network.
- Evaluate the 95% PI of the safety (m). The model should be recalibrated when the observed number of crashes is outside this PI. If it is not, use the same model in the subsequent analysis period.
- Repeat the same procedure for the subsequent time periods (using the model last recalibrated) to check the need for recalibration.

The flow chart shown in Figure 1 presented the first methodology described in the thesis.

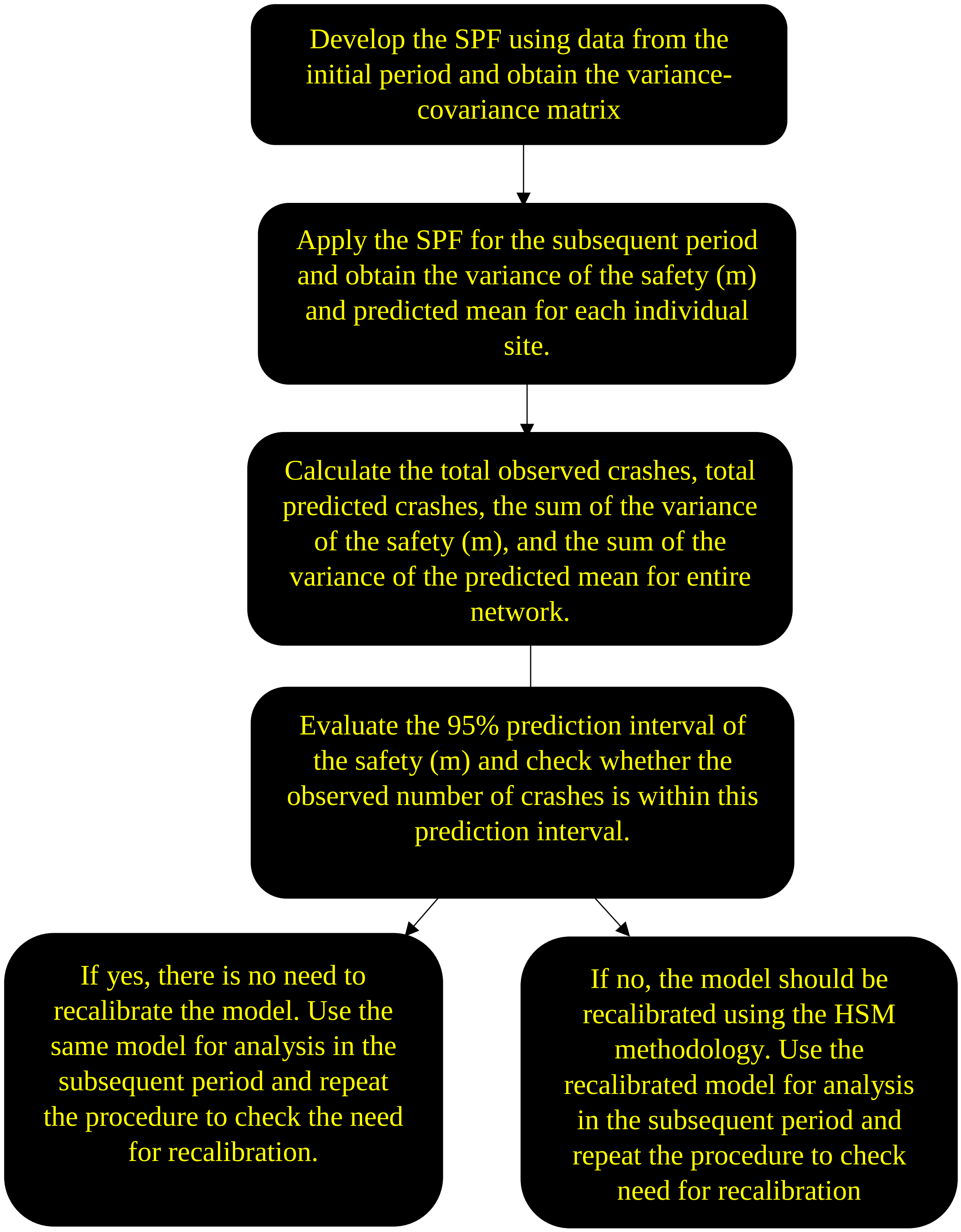

Figure 1. Flow chart for the first methodology proposed for recalibration

3.1.2 SECOND METHOD: STATISTICAL DIFFERENCE OF C-FACTOR ESTIMATE

The second method developed in this thesis requires an evaluation of the standard deviation of the C-factor estimate. The procedure to evaluate the coefficient of variation (CV) of the C-factor is adopted from the concept proposed by Dr. Hauer in the NCHRP report by Bahar (2014). The predicted value of the SPF can be expressed as a product of three factors:

$$N_{Pre} = N_{Base} * CMF * C$$

$$N = A * B * C$$

$$V\{N\} \cong (B \times C)^2 * V\{A\} + (A \times C)^2 * V\{B\} + (A \times B)^2 * V\{C\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.15)$$

Dividing by N^2

$$\frac{V\{N\}}{N^2} \cong \frac{V\{A\}}{A^2} + \frac{V\{B\}}{B^2} + \frac{V\{C\}}{C^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.16)$$

$$(cv\ \{N_{pre}\})^2 = (cv\ \{N_{base\ SPF}\})^2 + (cv\{CMFs\})^2 + (cv\{C\})^2 \ldots\ldots\ldots (3.17)$$

Flow-only models are developed in this methodology, and the effect of other traffic and geometric variables is ignored. Hence, the effect of the CMFs is ignored.

$$(cv\ \{N_{pre}\})^2 = (cv\ \{N_{base\ SPF}\})^2 + (cv\{C\})^2 \ldots\ldots\ldots (3.18)$$

In the current methodology, the above equation condenses to when the base model is made equal to the predicted model (since the method does not use CMFs).

$$\frac{Var(N_{pre})}{{N_{pre}}^2} = \frac{Var(C)}{C^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.19)$$

The variance and the standard deviation of the C-factor estimate can be evaluated using the above equation.

3.1.2.1 Steps in the second method

The second method is divided into the following five steps:

- Develop the SPF (flow-only model, ignoring the effect of other traffic and geometric variables) and calculate the C-factor for Year 1, called C_{ref}. If a good model is developed, this value is expected to be close to 1 (use the data of the entire network to evaluate the calibration factor).
- Adjust the estimates in Year 2 using C_{ref}. Calculate the variances of the model estimate. Assume that the entire variance in the model estimate is due to the variance of C. The variance of the model estimates is obtained using the methodology proposed by Wood (2005) and Lord (2008). Estimate the calibration factor in Year 2, called C_2.
- The CV of the C-factor in Year 2 is now known. Calculate the standard deviation in Year 2 as (CV of C-factor in Year 2) * (C-factor estimate of Year 2), called σ.
- Test the hypothesis that the C-factor in Year 2 is significantly different than 1 at a 5% level of significance:

$$H_0: C-factor=1$$

$$H_1: C-factor \neq 1$$

$$t = \frac{C_2 - 1}{\sigma}$$

- If the null hypothesis cannot be rejected, repeat the same procedure for Year 3 using the C-factor developed in Year 1 (C_{ref}) for Year 3.
- If the null hypothesis is rejected, the model needs to be recalibrated in Year 2. Use the C-factor developed in Year 2 (C_2) and multiply it by C_{ref} to analyze the Year 3 data.
- Repeat the same procedure for the subsequent time periods to check the need for recalibration.

The flowchart in Figure 2 indicates the second methodology described in the thesis.

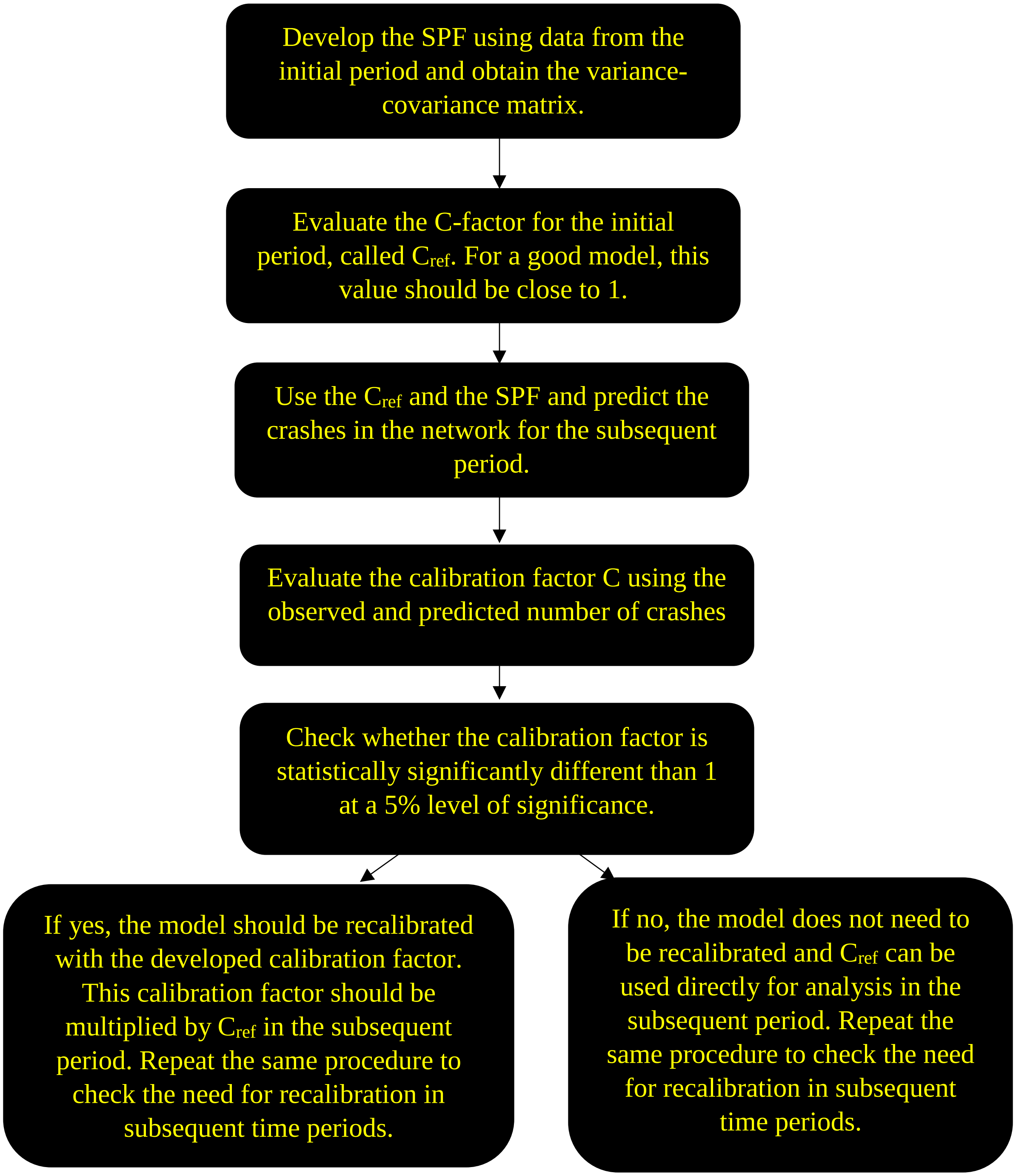

Figure 2. Flow chart for the second methodology proposed for recalibration

3.2 SELECTING THE DATASETS

The selection of the datasets is critical in this study, since the first step involves developing an SPF with the base condition data. Moreover, the data should be available over several years to test the different calibration strategies to determine which one works the best. The methodology developed in the present study was tested using the following datasets.

Intersection Data

- Toronto 4-legged intersections (6-year dataset)
- Michigan 3-legged intersections (5-year dataset)
- Michigan 4-legged intersections (5-year dataset)

Segment Data

- Michigan 2-lane undivided highways (5-year dataset)
- Michigan 4-lane undivided roads (5-year dataset)
- Michigan 4-lane divided roads (5-year dataset)

3.2.1 CHARACTERISTICS OF THE DATASETS

This section provides the summary statistics of the datasets used in the study

3.2.1.1 Toronto 4-legged intersections

Crash data and entering traffic flows were collected at 868 4-legged signalized intersections in Toronto for a period of 6 years. This dataset has been used extensively by other researchers (Lord, 2000; Miaou and Lord, 2003; Lord and Persaud 2004). The summary statistics for the Toronto 4-legged intersections are presented in Table 2.

Table 2. Summary statistics of the Toronto 4-legged intersections data

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Major AADT (Maximum)	71798	71257	71498	71450	72310	72178
Major AADT (Minimum)	5305	5294	5342	5369	5464	5469
Major AADT Average	27033	27014	27291	27460	27983	28045
Std. dev of Major AADT	10189.03	10187.73	10304.11	10384.74	10604.20	10654.25
Minor AADT (Maximum)	41306	41003	41150	41131	42012	42644
Minor AADT (Minimum)	52	52	52	52	53	53
Minor AADT Average	10581	10579	10694	10767	10979	11010
Std. dev of Minor AADT	8157.53	8172.70	8280.62	8358.04	8545.42	8594.44
Total number of intersections	868	868	868	868	868	868
Total number of crashes	8276	8141	8714	9818	10010	10030
Min. no. of crashes at a site	0	0	0	0	0	0
Max no. of crashes at a site	44	53	58	63	54	54
Avg. no. of crashes	9.53	9.38	10.04	11.31	11.53	11.56
Std. dev of crashes	7.82	7.68	7.95	9.67	10.09	10.01

3.2.1.2 Michigan 3-legged intersections

The crash data along with the flows has been collected at 3-legged signalized intersections in Michigan for a period of five years. Entering traffic flow and crash data were collected at 174 intersections for each of the five years. The summary statistics for Michigan 3-legged intersections are presented in Table 3.

Table 3. Summary statistics for Michigan 3-legged intersections data

	Year 1	Year 2	Year 3	Year 4	Year 5
Major AADT (Maximum)	61492	61662	61372	59046	62094
Major AADT (Minimum)	4652	4650	4625	4482	4391
Major AADT Average	19294	19318	19677	19272	19608
Std. dev of Major AADT	10196.28	10228.04	10269.17	9930.02	10349.57
Minor AADT (Maximum)	42413	42530	42330	40726	42828
Minor AADT (Minimum)	48	47	49	47	50
Minor AADT Average	4198	4200	4285	4218	4295
Std. dev of Minor AADT	5168.34	5180.83	5240.91	5107.57	5274.67
Total number of intersections	174	174	174	174	174
Total number of crashes	759	703	692	653	627
Min. no. of crashes at a site	0	0	0	0	0
Max no. of crashes at a site	30	20	21	22	27
Avg. no. of crashes	4.36	4.04	3.98	3.75	3.60
Std. dev of crashes	4.29	3.56	3.85	3.68	3.74

3.2.1.3 Michigan 4-legged intersections

The crash data along with the flows has been collected at 4-legged signalized intersections in Michigan for a period of five years. Entering traffic flow and crash data were collected at 349 intersections for each of the five years. The summary statistics for Michigan 4-legged intersections are presented in Table 4.

Table 4. Summary statistics for Michigan 4-legged intersections data

	Year 1	Year 2	Year 3	Year 4	Year 5
Major AADT (Maximum)	120082	118318	118771	116295	117627
Major AADT (Minimum)	4033	4087	4265	4114	4164
Major AADT Average	20889	20997	21445	21078	21380
Std. dev of Major AADT	15243.17	15013.86	15223.39	15027.13	15248.81
Minor AADT (Maximum)	66148	66135	69321	68508	68487
Minor AADT (Minimum)	88	92	94	94	99
Minor AADT Average	8781	8832	9034	8870	8992
Std. dev of Minor AADT	7915.43	7905.12	8104.12	7978.87	8086.98
Total number of intersections	349	349	349	349	349
Total number of crashes	2925	2872	2989	2965	2914
Min. no. of crashes at a site	0	0	0	0	0
Max no. of crashes at a site	61	53	51	46	44
Avg. no. of crashes	8.38	8.23	8.56	8.49	8.35
Std. dev of crashes	8.37	8.09	8.25	8.49	8.33

3.2.1.4 Michigan 2-lane undivided segments

The crash data along with the flows, segment length has been collected on 2-lane undivided segments in Michigan for a period of five years. The summary statistics for Michigan 2-lane undivided segments are presented in Table 5.

Table 5. Summary statistics for Michigan 2-lane undivided segments data

	Year 1	Year 2	Year 3	Year 4	Year 5
Segment AADT (Maximum)	30145	28494	25327	26158	26786
Segment AADT (Minimum)	326	318	234	237	253
Segment AADT Average	8483	8202	8322	8362	8486
Std. dev of Segment AADT	5066.95	4858.89	4860.47	4977.79	5055.11
Segment Length (Max in mil.)	5.63	5.63	5.63	5.63	5.63
Segment Length (Min in mil.)	0.013	0.013	0.013	0.013	0.013
Segment Length Avg. (in mil.)	0.948	0.945	0.943	0.941	0.941
Std. dev of Segment Length	0.83	0.83	0.84	0.86	0.86
Total number of segments	471	458	450	397	397
Total Segment Length (in mil.)	446.64	432.86	424.7	373.4	373.4
Total number of crashes	1960	1680	1561	1443	1558
Min. no. of crashes/ segment	0	0	0	0	0
Max no. of crashes/ segment	65	34	32	28	80
Avg. no. of crashes	4.16	3.67	3.47	3.63	3.92
Std. dev of crashes	6.02	4.53	4.35	4.79	6.74

3.2.1.5 Michigan 4-lane undivided segments

The crash data along with the flows, segment length has been collected on 4-lane undivided segments in Michigan for a period of five years. The summary statistics for Michigan 4-lane undivided segments are presented in Table 6.

Table 6. Summary statistics for Michigan 4-lane undivided segments data

	Year 1	Year 2	Year 3	Year 4	Year 5
Segment AADT (Maximum)	40830	40013	36142	36612	43824
Segment AADT (Minimum)	3981	3961	3700	3748	3849
Segment AADT Average	14157	13925	13922	14013	14117
Std. dev of Segment AADT	6004.43	5687.75	5738.18	6007.46	6162.67
Segment Length (Max in mil.)	5.247	5.247	5.247	5.247	5.247
Segment Length (Min in mil.)	0.009	0.009	0.009	0.009	0.009
Segment Length Avg. (in mil.)	0.707	0.705	0.712	0.700	0.700
Std. dev of Segment Length	0.59	0.59	0.60	0.56	0.56
Total number of segments	233	231	219	208	208
Total Segment Length (in mil.)	164.81	162.78	155.88	145.75	145.75
Total number of crashes	748	822	762	680	712
Min. no. of crashes/ segment	0	0	0	0	0
Max no. of crashes/ segment	25	26	28	25	30
Avg. no. of crashes	3.21	3.56	3.48	3.27	3.42
Std. dev of crashes	3.92	4.39	4.15	4.15	4.39

3.2.1.6 Michigan 4-lane divided segments

The crash data along with the flows, segment length has been collected on 4-lane divided segments in Michigan for a period of five years. The summary statistics for Michigan 4-lane divided segments are presented in Table 7.

Table 7. Summary statistics for Michigan 4-lane divided segments data

	Year 1	Year 2	Year 3	Year 4	Year 5
Segment AADT (Maximum)	35184	35536	35820	34064	34881
Segment AADT (Minimum)	1855	1873	1888	1964	2011
Segment AADT Average	10071	9898	9989	10043	10185
Std. dev of Segment AADT	5683.05	5670.59	5603.15	5602.36	5575.98
Segment Length (Max in mil.)	5.143	5.143	5.143	5.143	5.143
Segment Length (Min in mil.)	0.016	0.016	0.016	0.016	0.016
Segment Length Avg. (in mil.)	0.99	0.99	0.99	0.96	0.96
Std. dev of Segment Length	0.75	0.75	0.75	0.73	0.73
Total number of segments	373	372	371	347	347
Total Segment Length (in mil.)	369.49	368.82	367.95	334.63	334.63
Total number of crashes	1396	1354	1218	1334	1444
Min. no. of crashes/ segment	0	0	0	0	0
Max no. of crashes/ segment	68	64	46	70	90
Avg. no. of crashes	3.74	3.64	3.28	3.84	4.16
Std. dev of crashes	6.00	5.93	4.85	6.25	6.99

3.3 APPLYING THE CALIBRATION METHODOLOGY

The first step of both the methodologies involves developing an SPF with flow-only data, which may or may not reflect the base conditions. The NB regression model is used to this end. The SPF can be developed using any statistical package, such as R (2014), SAS (2008), etc. In this study, it is necessary to obtain the covariance matrix between the parameters to calculate the variance associated with the SPFs. Furthermore, the inverse of the dispersion parameter is needed to evaluate the variance. The two methodologies described earlier are applied for the recalibration of the models over time. The results of both the methodologies are also compared to those of the methodology described in the HSM and the one described by Shirazi et al. (2017).

3.3.1 MODEL FORM OF THE SPFS

The initial step of both the methodologies described in the thesis is to develop an SPF for the specific facility type using the NB regression model. The selection of an appropriate model form is important since the relationship between the number of crashes and the predictive variables should be appropriately captured. Moreover, the model form also affects the results of the prediction, and thus the analysis. Researchers have studied different model forms for intersections and segments in the past.

3.3.1.1 Intersection predictive model form

Miaou and Lord (2003) indicate that several model forms for intersection SPFs are possible even when very few covariates are considered. They state that when an appropriate model form must be chosen several factors should be considered engineering logics, exploratory data analysis, the flow values available, the covariate data, and the crash data. The authors reference the work of Turner and Nicholson, who classified the intersection model forms into three types. The Type 1 models relate the total number of crashes to the traffic volumes entering the intersection; the Type 2 models relate the crashes to two flows approaching from the major and minor roads, respectively; and the Type 3 models the number of crashes involving conflicting movements of vehicles. The Type 3 models require the maximum amount of data with detailed turning flows and number of crashes by movement group.

The most commonly used are the Type 2 models. The authors indicated that the Type 2 models follow the logic of no flow, no crashes, and they allow a nonlinear relationship between crashes and flows (this relationship has been found to be representative based on several studies in the past). The model form in the HSM is also a Type 2 model. The present study uses a Type 2 model with the model form shown in Equation 25.

$$\mu = \beta_0 * F_{Maj}^{\beta_1} * F_{Min}^{\beta_2} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . (3.20)$$

where F_{Maj}, F_{Min} are the entering flows (AADT) on major and minor streets, respectively.

In the present study, flow-only models are developed, ignoring the effects of the other traffic and geometric variables.

3.3.1.2 Segment predictive model form

The probability of a crash in a segment depends on the traffic flow and the length of the segment. The exploratory data analysis and engineering judgement based on several past studies indicate a nonlinear relationship between crashes and flow. For the selected functional form, the crash risk per unit length does not depend on the segment length, and the segment length (L) is therefore considered an offset in the modeling process. The model form is shown in Equation 26.

$$\mu = L * \beta_0 * Flow^{\beta_1} \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (3.21)$$

3.4 CHAPTER SUMMARY

This chapter has described the theory behind both the methodologies proposed for the recalibration procedure and described the steps that a practitioner or safety analyst needs to follow to apply both these methodologies. The chapter also presented a summary of the three intersection datasets and three segment datasets used to test the methodologies, and described the model form of the SPFs developed for intersections and segments.

CHAPTER IV

ANALYSIS RESULTS: FIRST METHOD

This chapter presents the results of applying the first method for recalibration to the datasets listed above. These results are compared to those of the methodology proposed in Shirazi et al. (2017). The chapter is divided into six sections. Section 4.1 describes the results of Toronto 4-legged intersections. Sections 4.2 and 4.3 then describe the results of Michigan 3-legged and 4-legged intersections. Finally, Sections 4.4, 4.5, and 4.6 describe the results of Michigan 2-lane undivided, 4-lane undivided, and 4-lane divided segments, respectively.

4.1 TORONTO 4-LEGGED INTERSECTIONS

The statistical software R was used to fit the safety performance function using the NB regression. The model was fit using the data of the first year, and was then recalibrated over time based on the two methodologies. The parameters of the model fit are presented in Table 8, and the variance-covariance matrix is presented in Table 9.

Table 8. Model output for the Toronto 4-legged intersections

Variable	Coefficients (std. error)
Intercept (ln β_0)	-8.84 (0.48)
$F_{Major}(\beta_1)$	0.51 (0.05)
$F_{Minor}(\beta_2)$	0.64 (0.02)
Inverse dispersion parameter (ϕ)	6.77 (0.62)
2*loglikelihood	-4873.968 (df = 4)
Akaike information criterion (AIC)	4881.968
Bayesian information criterion (BIC)	4901.033
Median absolute deviation (MAD)	1.007
Mean square prediction error (MSPE)	1.091

Table 9. Variance-covariance matrix for the Toronto 4-legged intersection model

	Intercept	Flow 1	Flow 2
Intercept	0.2346591000	-0.0210500408	-0.0021762592
Flow 1	-0.0210500410	0.0022846600	-0.0002422538
Flow 2	-0.0021762590	-0.0002422538	0.0005067881

Table 10 summarizes the PI for the sum of the safety (m) and variance over the period of 6 years using the first method. The decision to recalibrate models is based on the safety (m) PI. A 95% prediction interval is used throughout the study. Moreover, the methodology described by Shirazi et al. (2017) was applied to the dataset for comparison, and the results are also presented in Table 10. The graphs representing the results of both these methodologies are shown in Figures 3 and 4, respectively.

4.1.1 DISCUSSION OF THE RESULTS

The results using the first methodology indicate that the model needs to be recalibrated in both Year 3 and Year 4. In contrast, the results using Shirazi et al.'s (2017) methodology indicate that the model only needs to be calibrated in Year 4. It must be noted that the developed model has very little variance, and this is reflected in the results. The methodologies developed in this thesis consider the model variance, whereas Shirazi et al.'s does not. These contrasting results indicate the importance of considering the model variance in the recalibration of models over time. Moreover, it should be noted that a practitioner following the HSM methodology for recalibration would not have recalibrated the model in consecutive years, whereas the model variance indicates a need to do so in this case.

Table 10. Summary of the results for Toronto 4-legged intersections based on first methodology and Shirazi et al. (2017)

					Safety (m) PI (calibrated)							
Year	**Total Observed**	**Total Predicted**	**First Calibration**	**Second Calibration**	**LL**	**UL**	**C**	F_{Maj} **(avg)**	F_{Min} **(avg)**	**μ**	**C-proxy**	**Change in C-proxy**
1	8276	8254	-	-	8004.8	8503.2	1.00	27033	10581	10.110	0.94	0.00%
2	8141	8251	-	-	8001.8	8500.1	1.00	27014	10579	10.105	0.93	1.63%
3	**8714**	8353	8714	-	**8103.9**	**8602.3**	**1.04**	27291	10694	10.229	0.98	4.06%
4	**9818**	8418	8781	9818	**8521.2**	**9041.0**	**1.12**	27460	10767	10.306	1.10	**11.80%**
5	10010	8607	8979	10039	9748.5	10329.7	1.12	27983	10979	10.537	1.09	0.27%
6	10030	8634	9007	11214	9779.7	10360.9	1.12	28045	11010	10.568	1.09	0.35%

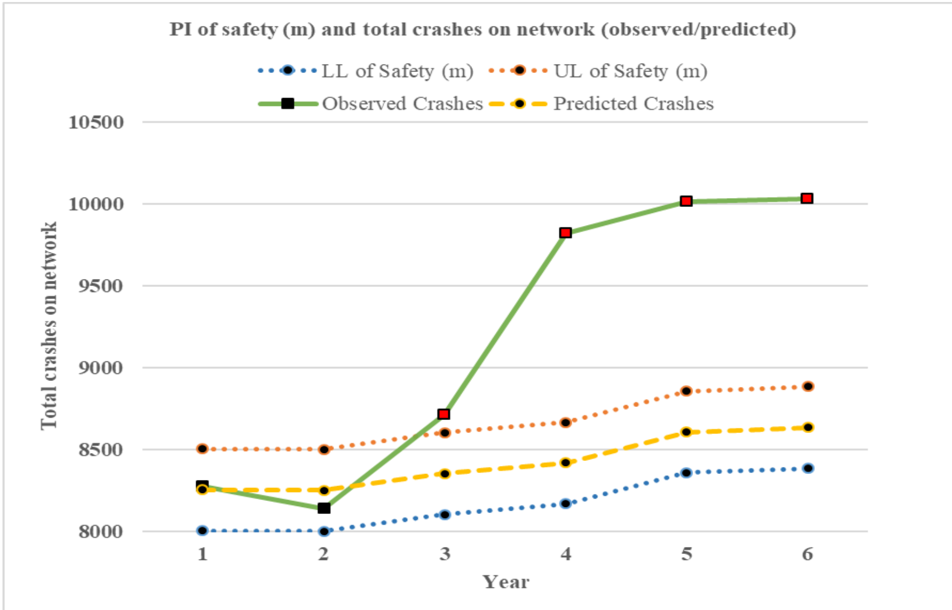

Figure 3. Illustration of first method for Toronto 4-legged intersections

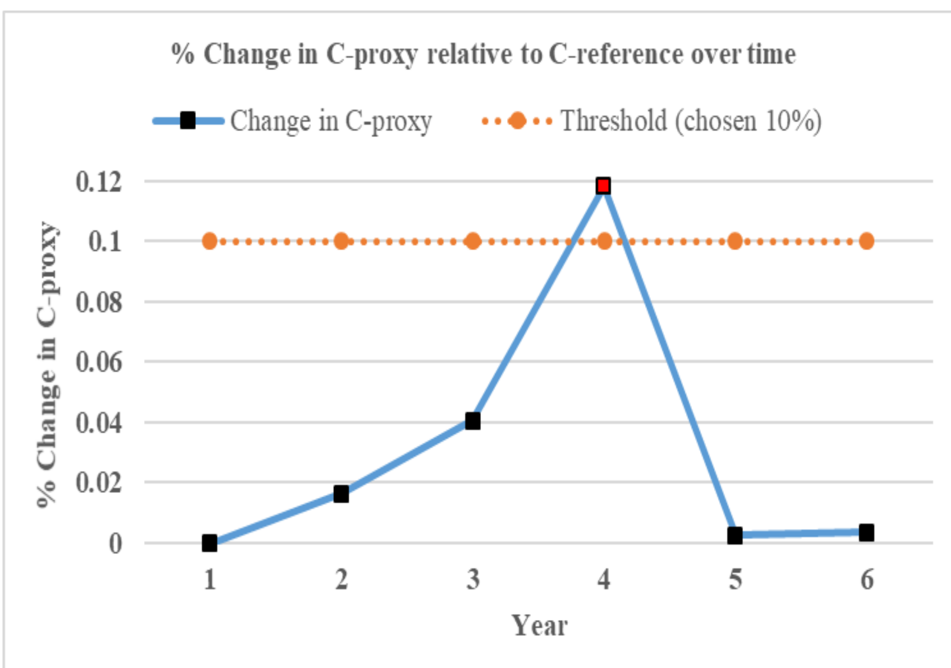

Figure 4. Illustration of Shirazi et al. (2017) method for Toronto 4-legged intersections

4.2 MICHIGAN 3-LEGGED INTERSECTIONS

The statistical software R was used to fit the safety performance function using the NB regression. The model was fit using the data of the first year, and was then recalibrated over time based on the two methodologies. The parameters of the model fit are presented in Table 11, while the variance-covariance matrix of the model is presented in Table 12.

Table 11. Model output for the Michigan 3-legged intersections

Variable	Coefficients (std. error)
Intercept (ln β_0)	-5.62 (1.40)
F_{Major} (β_1)	0.61 (0.14)
F_{Minor} (β_2)	0.14 (0.05)
Inverse dispersion parameter (ϕ)	1.696 (0.28)
2*loglikelihood	-572.523 (df = 4)
Akaike information criterion (AIC)	580.523
Bayesian information criterion (BIC)	593.159
Median absolute deviation (MAD)	0
Mean square prediction error (MSPE)	$1.038*10^{-7}$

Table 12. Variance-covariance matrix for the Michigan 3-legged intersection model

	Intercept	Flow 1	Flow 2
Intercept	1.9482083000	-0.1882637114	-0.0129243102
Flow 1	-0.1882637100	0.0199524509	-0.0009116704
Flow 2	-0.0129243100	-0.0009116704	0.0028165642

The above model was used to predict the number of crashes at each of the individual sites for each subsequent year. Table 13 shows the results of the analysis using the first methodology and that of Shirazi et al. (2017). Figure 5 and 6 correspond to the first methodology and to Shirazi et al.'s (2017), respectively.

4.2.1 DISCUSSION OF THE RESULTS

The results using the first methodology indicate a need to recalibrate the model in Year 4. In contrast, the methodology proposed by Shirazi et al. (2017) indicates a need for recalibration in Year 3. It should be noted that the observed number of crashes in the network is 692 in Year 3, which is close to the lower limit of the PI. Therefore, even if the analyst decides to recalibrate the model in Year 3, it should not drastically affect the predictions. In this case, it can be concluded that the results of both the methodologies are comparable, and therefore it is reasonable to use a 10% error.

Table 13. Summary of the results for Michigan 3-legged intersections based on first methodology and Shirazi et al. (2017)

				Safety (m) PI (calibrated)							
Year	**Observed**	**Total Predicted**	**First Calibration**	**LL**	**UL**	**C**	**F1 (avg)**	**F2 (avg)**	**μ**	**C-proxy**	**Change in C-proxy**
1	759	765	-	669.1	861.7	1.00	19294	4198	4.840	0.90	0.00%
2	703	766	-	669.6	862.3	1.00	19318	4200	4.844	0.83	7.45%
3	692	777	-	679.5	874.7	1.00	19678	4285	4.913	0.81	**10.17%**
4	**653**	766	653	**669.9**	**862.0**	**0.85**	19272	4218	4.840	0.78	4.21%
5	627	775	661	577.8	744.2	0.85	19608	4295	4.903	0.73	9.22%

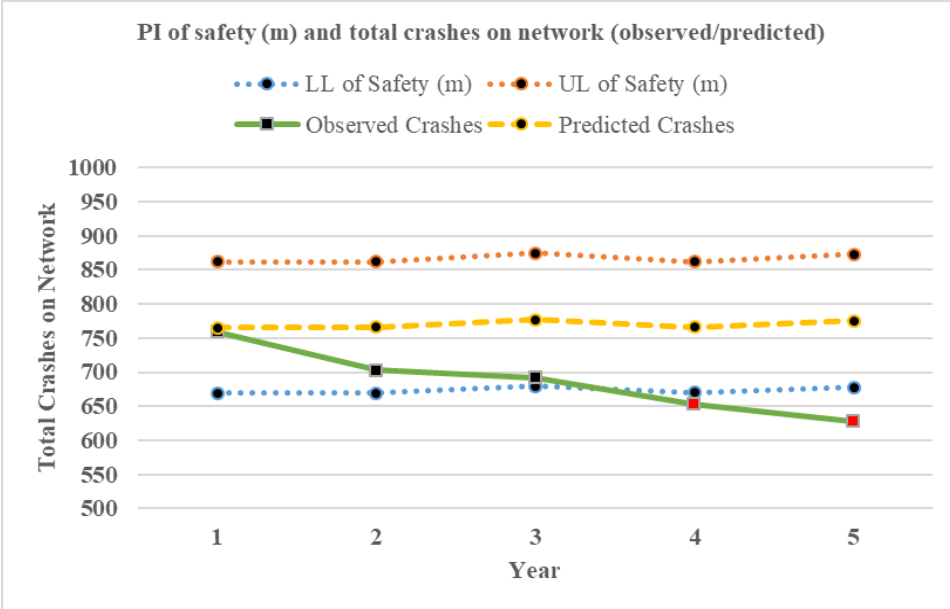

Figure 5. Illustration of first method for Michigan 3-legged intersections

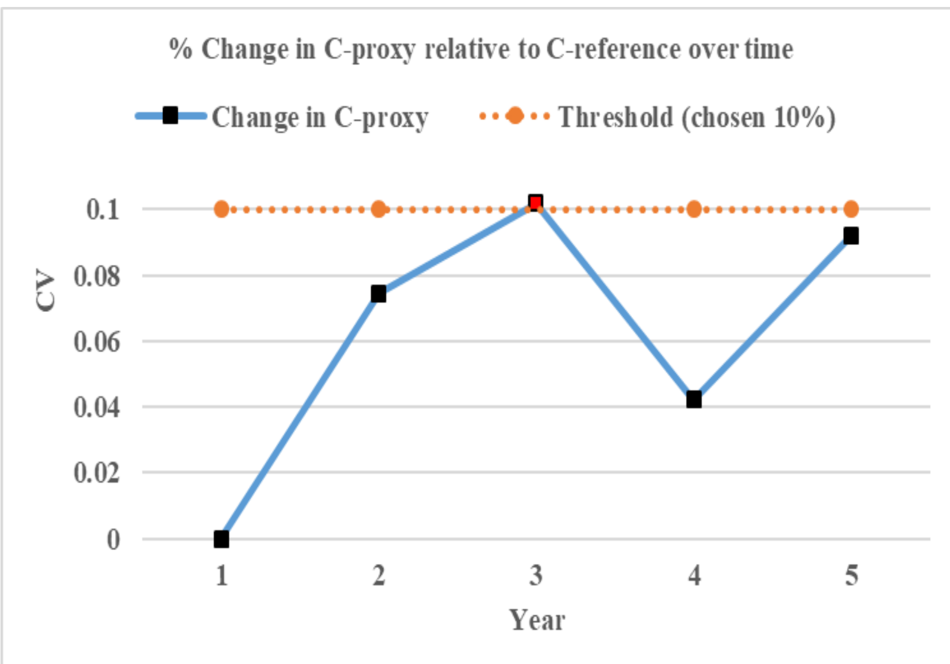

Figure 6. Illustration of Shirazi et al. (2017) method for Michigan 3-legged intersections

4.3 MICHIGAN 4-LEGGED INTERSECTIONS

The safety performance function was fit using the NB regression in the statistical software R. The model was fit using the first-year data, and was then recalibrated over time using the two methodologies. Table 14 presents the parameters of the model fit, while Table 15 presents the variance-covariance matrix of the model.

Table 14. Model output for the Michigan 4-legged intersections

Variable	Coefficients (std. error)
Intercept (ln β_0)	-7.42 (0.66)
$F_{Major}(\beta_1)$	0.75 (0.07)
$F_{Minor}(\beta_2)$	0.23 (0.04)
Inverse dispersion parameter (ϕ)	2.74 (0.295)
2*loglikelihood	-2028.62 (df = 4)
Akaike information criterion (AIC)	2036.62
Bayesian information criterion (BIC)	2052.04
Median absolute deviation (MAD)	0.976
Mean square prediction error (MSPE)	1.103

Table 15. Variance-covariance matrix for the Michigan 4-Legged intersection model

	Intercept	Flow 1	Flow 2
Intercept	0.436101710	-0.042030110	-0.002381334
Flow 1	-0.042030110	0.005249850	-0.001103465
Flow 2	-0.002381334	-0.001103465	0.001513798

Table 16 presents the results of the analysis using the first methodology and that of Shirazi et al. (2017). Furthermore, the graphs corresponding to the two methods are shown in Figures 7 and 8, respectively.

4.3.1 DISCUSSION OF THE RESULTS

Both methodologies indicate that there is no need to recalibrate the model in any of the 5 years. This suggests that the variables already included the SPF (i.e. AADT) can capture the variations in the observed number of crashes. It should also be noted that a practitioner adopting the HSM methodology will calibrate the model when this is not necessary.

Table 16. Summary of the results for Michigan 4-legged intersections based on first methodology and Shirazi et al. (2017)

			Safety (m) PI (calibrated)							
Year	Observed	Total Predicted	LL	UL	C	F1 (avg)	F2 (avg)	μ	C-proxy	Change in C-proxy
1	2925	2973	2738.2	3208.2	1.00	20889	8781	9.040	0.93	0.00%
2	2872	2988	2753.7	3223.1	1.00	20997	8832	9.088	0.91	2.33%
3	2989	3054	2814.3	3293.3	1.00	21455	9034	9.286	0.92	0.51%
4	2965	3000	2764.6	3235.7	1.00	21078	8870	9.123	0.93	0.44%
5	2914	3042	2802.9	3280.8	1.00	21380	8992	9.251	0.90	2.14%

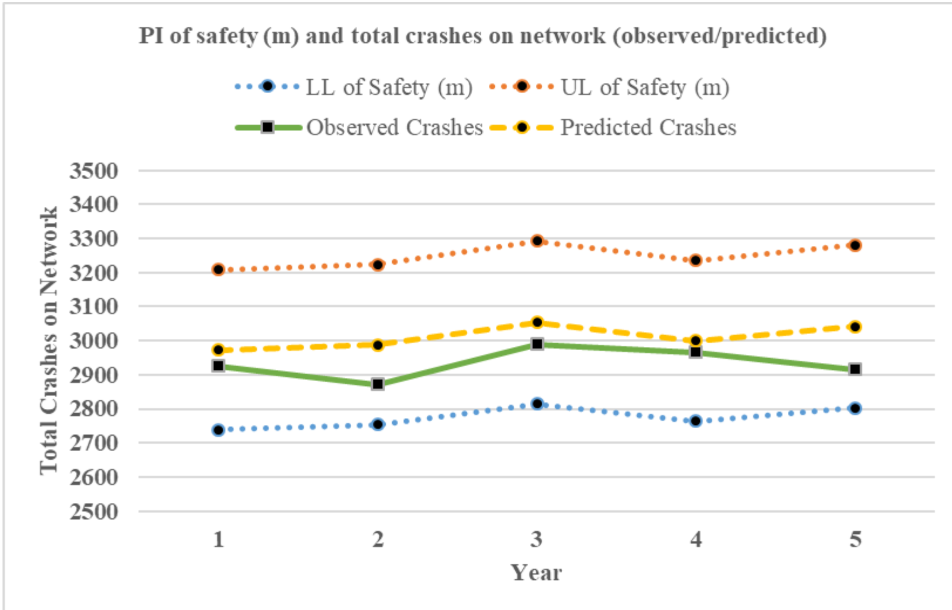

Figure 7. Illustration of first method for Michigan 4-legged intersections

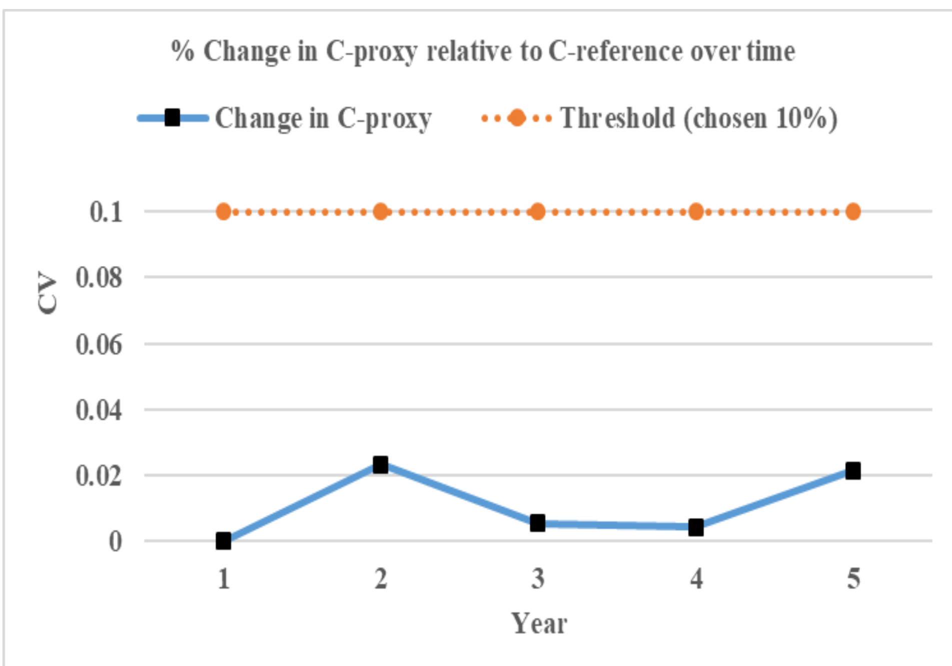

Figure 8. Illustration of Shirazi et al. (2017) method for Michigan 4-legged intersections

4.4 MICHIGAN 2-LANE UNDIVIDED SEGMENTS

The statistical software R was used to fit the safety performance function using the NB regression. The model was fit using the data from the first year, and it was then recalibrated over time based on the two methodologies. Table 17 presents the parameters of the model fit. The variance-covariance matrix of model is presented in Table 18.

Table 17. Model output for the Michigan 2-lane undivided segments

Variable	Coefficients (std. error)
Intercept (ln β_0)	- 4.96 (0.71)
Flow (β_1)	0.71 (0.07)
Inverse dispersion parameter (ϕ)	3.278 (0.43)
2*loglikelihood	-1974.074 (df = 3)
Akaike information criterion (AIC)	1980.074
Bayesian information criterion (BIC)	1992.538
Median absolute deviation (MAD)	1.036
Mean square prediction error (MSPE)	1.068

Table 18. Variance-covariance matrix for the Michigan 2-lane undivided segment model

	Intercept	Flow
Intercept	0.36946352	-0.040819281
Flow	-0.04081928	0.004527044

The results of the analysis based on the first methodology and based on Shirazi et al. (2017) are shown in Table 19. The graphs corresponding to the first methodology and Shirazi et al. (2017) are shown in Figures 9 and 10, respectively.

Table 19. Summary of the results for Michigan 2-lane undivided segments based on first methodology and Shirazi et al. (2017)

				Safety (m) PI (calibrated)							
Year	Observed	Total Predicted	First Calibration	LL	UL	C	AADT (avg)	μ	Segment Length	C-proxy	Change in C-proxy
1	1960	1932	-	1795.0	2069.4	1.00	8483	4.412	446.6	0.99	0.00%
2	**1680**	1825	1680	**1693.2**	**1957.2**	**0.92**	8202	4.308	432.9	0.90	8.95%
3	1561	1820	1675	1551.1	1798.5	0.92	8322	4.353	424.7	0.84	**15.16%**
4	1443	1619	1491	1370.3	1610.8	0.92	8362	4.367	373.4	0.88	5.34%
5	1558	1631	1502	1380.6	1622.4	0.92	8486	4.413	373.4	0.95	**12.55%**

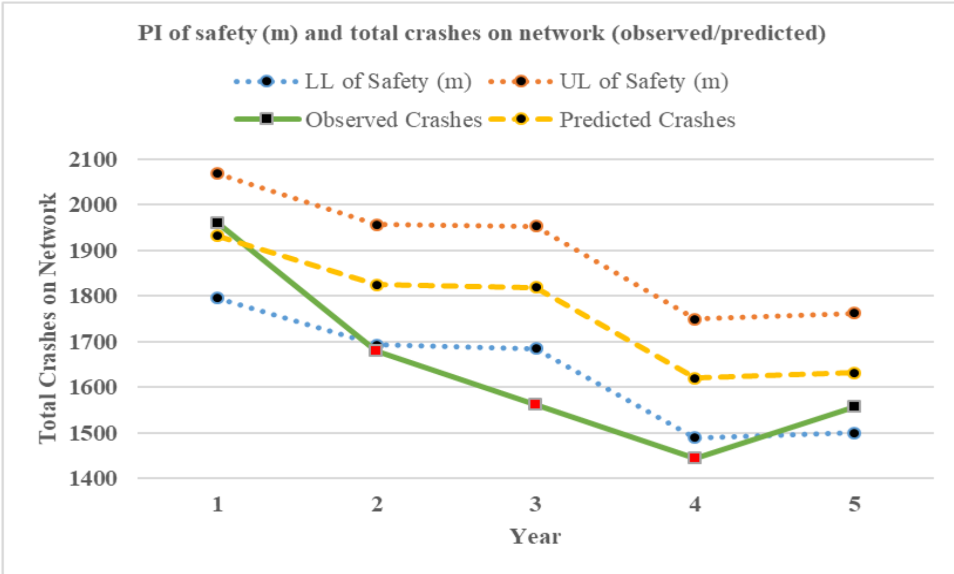

Figure 9. Illustration of first method for Michigan 2-lane undivided segments

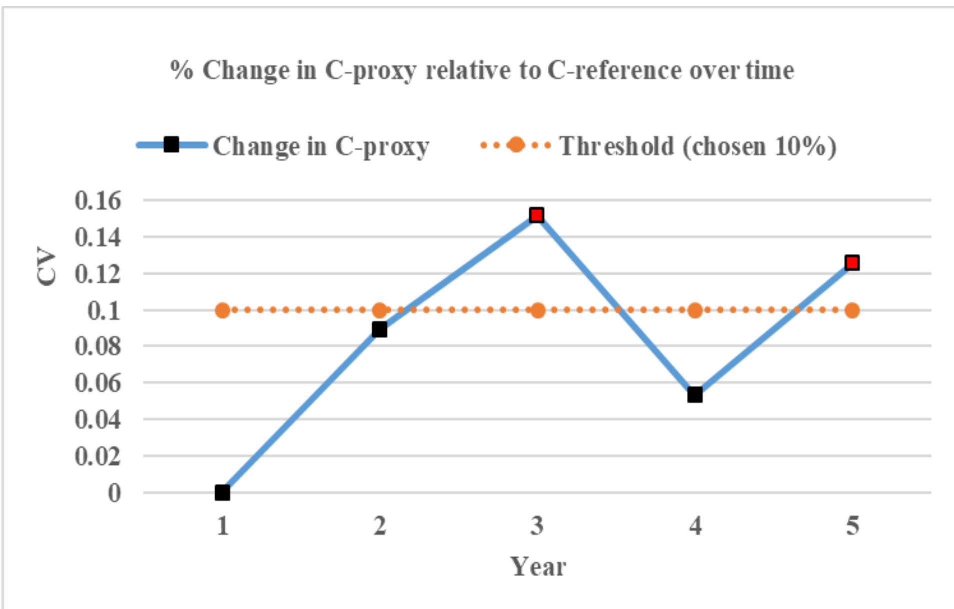

Figure 10. Illustration of Shirazi et al. (2017) method for Michigan 2-lane undivided segments

4.4.1 DISCUSSION OF THE RESULTS

The results of the first methodology indicate a need for model recalibration in Year 2, whereas Shirazi et al.'s (2017) methodology recommends recalibration in both Year 3 and Year 5. It should be noted that according to the latter method, in Year 2 the error is 8.95%, which is close to 10%. If the analyst recalibrates the model in Year 2, then there is no need for further recalibration in Years 3 and Years 5. This is another potential consideration concerning the error threshold Shirazi et al.'s (2017) methodology. The analyst should use his or her judgement to decide whether or not recalibration is necessary when the errors are close to 10%. Overall, it can be concluded that the results of both the methodologies are comparable for this dataset.

4.5 MICHIGAN 4-LANE UNDIVIDED SEGMENTS

The model was fitted (in R) using the first-year data, and it was then recalibrated over time based on the two methodologies. Table 20 presents the parameters of the model fit. The variance-covariance matrix of model is presented in Table 21.

Table 20. Model output for the Michigan 4-lane undivided segments

Variable	Coefficients (std. error)
Intercept (ln β_0)	- 10.80 (1.42)
Flow (β_1)	1.29 (0.15)
Inverse dispersion parameter (ϕ)	2.92 (0.61)
2*loglikelihood	-933.866 (df = 3)
Akaike information criterion (AIC)	939.865
Bayesian information criterion (BIC)	950.219
Median absolute deviation (MAD)	1.028
Mean square prediction error (MSPE)	1.117

Table 21. Variance-covariance matrix for the Michigan 4-lane undivided segment model

	Intercept	**Flow**
Intercept	2.0216386	-0.21051
Flow	-0.2105069	0.021955

The results of the analysis using the first methodology and Shirazi et al.'s (2017) are shown in Table 22. The graphs corresponding to these two methodologies are shown in Figures 11 and 12, respectively.

4.5.1 DISCUSSION OF THE RESULTS

The results of the analysis are consistent using both methodologies: the model needs to be recalibrated in Year 2. It can be concluded that a 10% error threshold is appropriate to use in the methodology proposed by Shirazi et al. (2017).

Table 22. Summary of the results for Michigan 4-lane undivided segments based on first methodology and Shirazi et al. (2017)

				Safety (m) PI (calibrated)							
Year	**Observed**	**Total Predicted**	**First Calibration**	**LL**	**UL**	**C**	**AADT (avg)**	**μ**	**Segment Length**	**C-proxy**	**Change in C-proxy**
1	748	765	-	684.8	845.3	1.00	14157	4.535	164.8	1.00	0.00%
2	**822**	734	822	**657.2**	**811.8**	**1.12**	13925	4.439	162.8	1.14	**13.67%**
3	762	703	787	701.2	872.4	1.12	13992	4.467	155.9	1.09	4.00%
4	680	663	742	660.6	823.1	1.12	14013	4.475	145.8	1.04	8.56%
5	712	675	755	671.3	838.8	1.12	14117	4.518	145.7	1.08	5.14%

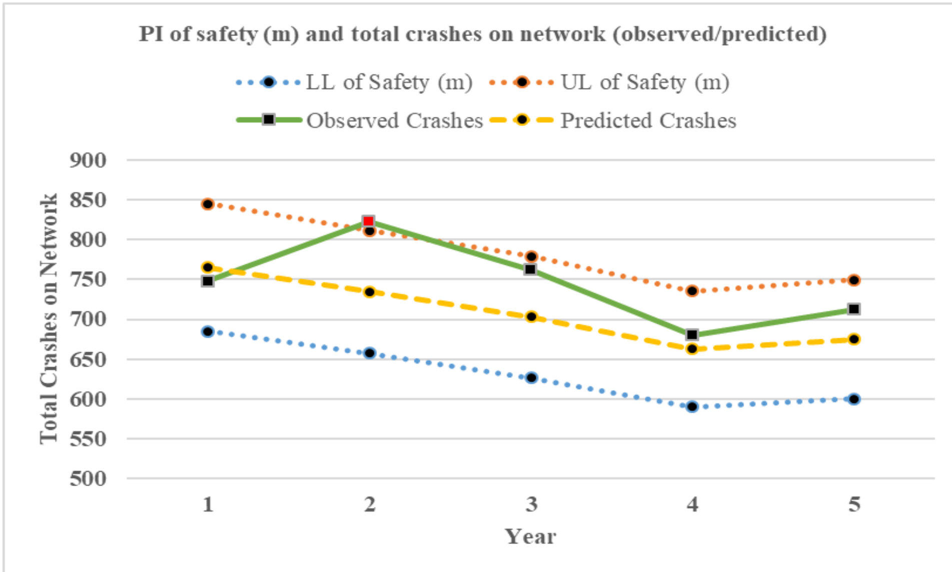

Figure 11. Illustration of first method for Michigan 4-lane undivided segments

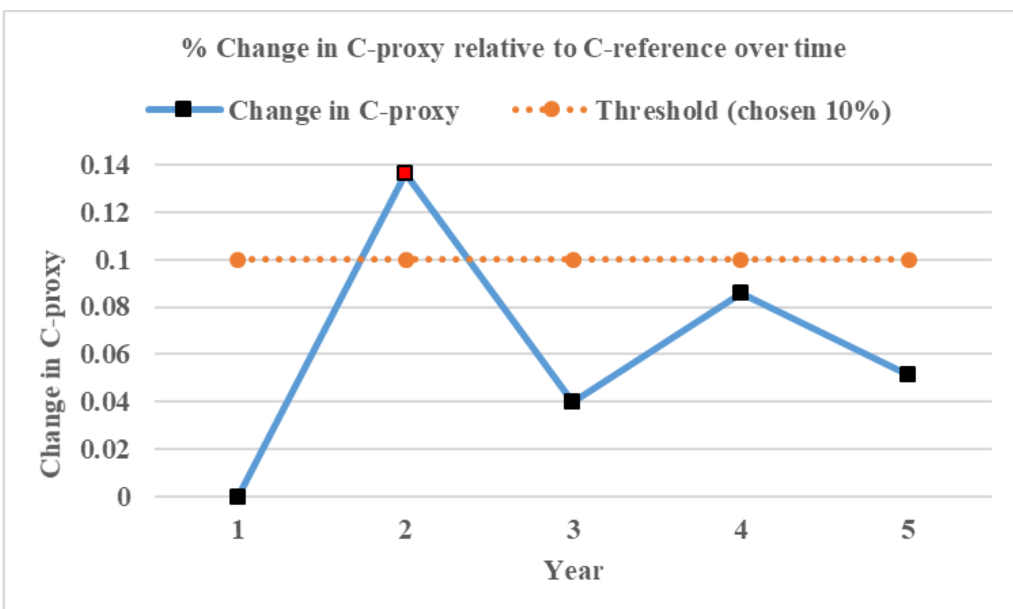

Figure 12. Illustration of Shirazi et al. (2017) method for Michigan 4-lane undivided segments

4.6 MICHIGAN 4-LANE DIVIDED SEGMENTS

The statistical software R was used to fit the safety performance function using the NB regression. The model was fit using the data from the first year, and it was then recalibrated over time based on the two methodologies. Table 23 presents the parameters of the model fit. The variance-covariance matrix of model is presented in Table 24.

Table 23. Model output for the Michigan 4-lane divided segments

Variable	Coefficients (std. error)
Intercept (ln β_0)	- 6.10 (0.80)
Flow (β_1)	0.797 (0.09)
Inverse dispersion parameter (ϕ)	2.82 (0.46)
2*loglikelihood	-1519.677 (df = 3)
Akaike information criterion (AIC)	1525.677
Bayesian information criterion (BIC)	1537.442
Median absolute deviation (MAD)	1.042
Mean square prediction error (MSPE)	1.133

Table 24. Variance-covariance matrix for the Michigan 4-lane divided segment model

	Intercept	Flow
Intercept	0.63957695	-0.069212709
Flow	-0.06921271	0.007514085

Table 25 shows the results of the analysis using the first methodology and that of Shirazi et al. (2017). The graphs corresponding to the two methodologies are shown in Figures 13 and 14, respectively.

4.6.1 DISCUSSION OF THE RESULTS

The first methodology demands a recalibration only in Year 5, while the methodology proposed in Shirazi et al. (2017) requires recalibration in both Year 3 and Year 4. It should be noted that even in the first methodology, the total observed number of crashes in the entire network is close to the lower limit of the PI of the safety (m). If the analyst decides to recalibrate the model in Year 3 using the first methodology, then he has to do so again in Year 4. Even for this specific dataset, the 10% threshold is reasonable.

Table 25. Summary of the results for Michigan 4-lane divided segments based on first methodology and Shirazi et al. (2017)

				Safety (m) PI (calibrated)							
Year	**Observed**	**Total Predicted**	**First Calibration**	**LL**	**UL**	**C**	**AADT (avg)**	**μ**	**Segment Length**	**C-proxy**	**Change in C-proxy**
1	1396	1343	-	1224.6	1462.0	1.00	10071	3.492	369.5	1.08	0.00%
2	1354	1325	-	1206.5	1442.6	1.00	9898	3.444	368.8	1.07	1.47%
3	1218	1337	-	1217.4	1456.0	1.00	9989	3.470	368.0	0.95	**11.66%**
4	1334	1242	-	1125.0	1358.9	1.00	10043	3.485	334.6	1.14	**20.43%**
5	**1444**	1253	1444	**1136.8**	**1369.9**	**1.15**	10185	3.524	334.6	1.22	7.42%

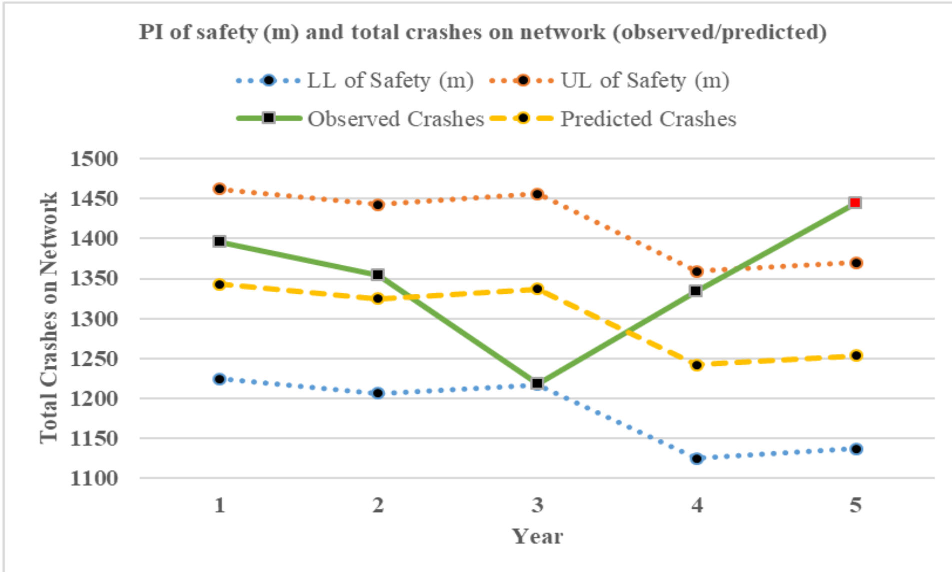

Figure 13. Illustration of first method for Michigan 4-lane divided segments

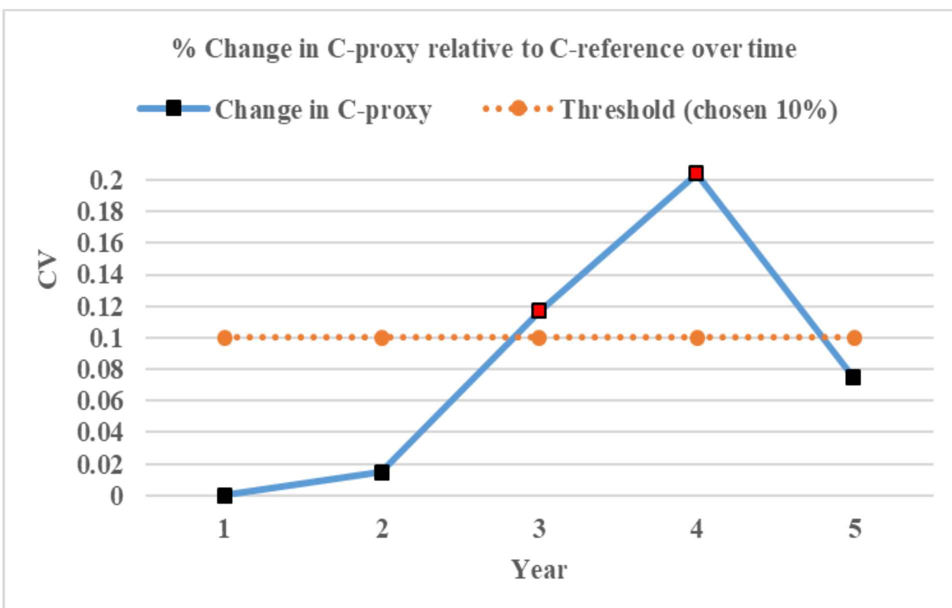

Figure 14. Illustration of Shirazi et al. (2017) method for Michigan 4-lane divided segments

CHAPTER V

ANALYSIS RESULTS: SECOND METHOD

This chapter presents the results of the second method of recalibration applied to the datasets listed earlier. These results are compared to the results of the first method and the one proposed by Shirazi et al. (2017). Sections 5.1 through 5.6 present the results for the Toronto 4-legged intersections, Michigan 3-legged and 4-legged intersections, Michigan 2-lane undivided, Michigan 4-lane undivided, and divided segments, respectively.

5.1 TORONTO 4-LEGGED INTERSECTIONS

The second methodology recommends recalibration when the calibration factor in a year (other than the year used to develop the model) is statistically significantly different than 1 (at a 5% level of significance). The results of this methodology are presented in Table 26.

The results demonstrate a need for recalibration in Years 3 and 4, which is consistent with the results of the first methodology.

5.2 MICHIGAN 3-LEGGED INTERSECTIONS

The second methodology recommends recalibration when the calibration factor in a year (other than the year used to develop the model) is statistically significantly different than 1 (at a 5% level of significance). Table 27 presents the results of this methodology**.**

The results indicate a need for recalibration in Year 4. This is in line with the results of the first methodology.

Table 26. Summary of the results for Toronto 4-legged intersections based on second methodology

Year	Y	Estimated	Var (η)	Var(m)	Var(y)	C	CV(m)	CV(y)	SD(m)	SD(y)	t(m)	t(y)	Comments
1	8276	8254.0	0.946	16161.79	24415.77	1.00	0.015	0.019	-	-	-	-	-
2	8141	8273.0	0.947	16258.80	24553.86	0.98	0.015	0.019	0.015	0.019	-1.052	-0.856	-
3	8714	8375.4	0.946	16691.81	25089.54	**1.04**	0.015	0.019	0.016	0.020	2.519	2.055	**Need to recalibrate**
4	9818	8781.1	0.945	18382.56	27543.09	**1.12**	0.015	0.019	0.017	0.021	6.840	5.588	**Need to recalibrate**
5	10010	10039.2	0.943	24079.81	35789.35	1.00	0.015	0.019	0.015	0.019	-0.188	-0.155	-
6	10030	10070.4	0.944	24284.49	36030.47	1.00	0.015	0.019	0.015	0.019	-0.260	-0.214	-

Table 27. Summary of the results for Michigan 3-legged intersections based on second methodology

Year	Y	Estimated	Var (η)	Var(m)	Var(y)	C	CV(m)	CV(y)	SD(m)	SD(y)	t(m)	t(y)	Comments
1	759	765.3	2.582	2413.47	3178.84	0.99	0.064	0.074	-	-	-	-	-
2	703	759.6	2.589	2417.29	3183.23	0.93	0.065	0.074	0.060	0.069	-1.244	-1.084	-
3	692	770.7	2.582	2480.87	3257.97	0.90	0.065	0.074	0.058	0.067	-1.759	-1.535	-
4	653	759.6	2.577	2401.35	3167.29	**0.86**	0.065	0.074	0.055	0.064	-2.531	-2.204	**Need to recalibrate**
5	627	661.0	2.595	2478.51	3253.80	0.95	0.075	0.086	0.071	0.082	-0.719	-0.628	-

5.3 MICHIGAN 4-LEGGED INTERSECTIONS

The second methodology recommends recalibration when the calibration factor in a year (other than the year used to develop the model) is statistically significantly different than 1 (at a 5% level of significance). The results using this methodology are shown in Table 28.

The results indicate no need for recalibration over the 5-year period. This result is consistent with the results of the first methodology and the methodology proposed by Shirazi et al. (2017).

5.4 MICHIGAN 2-LANE UNDIVIDED SEGMENTS

The second methodology recommends recalibration when the calibration factor in a year (other than the year used to develop the model) is statistically significantly different than 1 (at a 5% level of significance). Table 29 presents the results of this methodology.

The results of the analysis indicate a need for recalibration in Year 2, which is in line with the results obtained using the first methodology.

Table 28. Summary of the results for Michigan 4-legged intersections based on second methodology

Year	Y	Estimated	Var (η)	Var(m)	Var(y)	C	CV(m)	CV(y)	SD(m)	SD(y)	t(m)	t(y)	Comments
1	2925	2973.2	1.601	14375.60	17348.83	0.98	0.040	0.044	-	-	-	-	-
2	2872	2939.9	1.589	14108.27	17048.19	0.98	0.040	0.044	0.060	0.039	-0.585	-0.532	-
3	2989	3004.3	1.585	14692.06	17696.33	0.99	0.017	0.019	0.058	0.016	-0.308	-0.269	-
4	2965	2951.5	1.592	14212.09	17163.60	1.00	0.017	0.019	0.055	0.017	0.274	0.239	-
5	2914	2992.5	1.593	14617.87	17610.39	0.97	0.017	0.019	0.071	0.016	-1.620	-1.413	-

Table 29. Summary of the results for Michigan 2-lane undivided segments based on second methodology

Year	Y	Estimated	Var (η)	Var(m)	Var(y)	C	CV(m)	CV(y)	SD(m)	SD(y)	t(m)	t(y)	Comments
1	1960	1932.0	1.635	4900.74	6832.92	1.01	0.036	0.043	-	-	-	-	-
2	1680	1851.6	1.555	4667.64	6546.09	**0.91**	0.037	0.044	0.033	0.040	-2.769	-2.338	**Need to recalibrate**
3	1561	1674.8	1.537	3984.31	5525.93	0.93	0.038	0.044	0.035	0.041	-1.935	-1.643	-
4	1443	1490.6	1.381	3765.77	5137.79	0.97	0.041	0.048	0.040	0.047	-0.801	-0.686	-
5	1558	1501.5	1.377	3806.76	5188.86	1.04	0.041	0.048	0.043	0.050	0.882	0.756	-

5.5 MICHIGAN 4-LANE UNDIVIDED SEGMENTS

The second methodology recommends recalibration when the calibration factor in a year (other than the year used to develop the model) is statistically significantly different than 1 (at a 5% level of significance). The results of this methodology are shown in Table 30.

The results indicate the need to recalibrate in Year 2, which is consistent with both the first methodology and that of Shirazi et al. (2017).

5.6 MICHIGAN 4-LANE DIVIDED SEGMENTS

The second methodology recommends recalibration when the calibration factor in a year (other than the year used to develop the model) is statistically significantly different than 1 (at a 5% level of significance). Table 31 presents the results of this methodology.

The results show a need to recalibrate the model in both Year 3 and Year 4. This is in line with the results using Shirazi et al.'s (2017) methodology. In contrast, the results of the first methodology indicate a need to recalibrate in Year 5. As described in the previous chapter, the observed number of crashes in Year 3 were close to the lower limit of the predicted number of crashes. If the model was recalibrated in Year 3, then it would again have to be recalibrated in Year 4. This example emphasizes the need to exercise judgement while making the recalibration decision.

Table 30. Summary of the results for Michigan 4-lane undivided segments based on second methodology

Year	Y	Estimated	Var (η)	Var(m)	Var(y)	C	CV(m)	CV(y)	SD(m)	SD(y)	t(m)	t(y)	Comments
1	748	765.0	1.767	1677.48	2442.51	0.98	0.054	0.065	-	-	-	-	-
2	822	718.1	1.699	1488.36	2190.51	**1.14**	0.054	0.065	0.061	0.075	2.352	1.939	**Need to recalibrate**
3	762	786.8	1.635	1907.72	2788.27	0.97	0.056	0.067	0.054	0.065	-0.586	-0.485	-
4	680	741.8	1.582	1716.86	2547.08	0.92	0.056	0.068	0.051	0.062	-1.628	-1.337	-
5	712	755.0	1.582	1825.20	2670.19	0.94	0.057	0.068	0.053	0.065	-1.068	-0.883	-

Table 31. Summary of the results for Michigan 4-lane divided segments based on second methodology

Year	Y	Estimated	Var (η)	Var(m)	Var(y)	C	CV(m)	CV(y)	SD(m)	SD(y)	t(m)	t(y)	Comments
1	1396	1343.3	1.606	3669.50	5012.79	1.04	0.045	0.053	-	-	-	-	-
2	1354	1376.5	1.654	3916.44	5346.98	0.98	0.045	0.053	0.045	0.052	-0.366	-0.313	-
3	1218	1389.1	1.625	4000.75	5444.40	**0.88**	0.046	0.053	0.040	0.047	-3.086	-2.645	**Need to recalibrate**
4	1334	1131.7	1.567	2954.76	3985.97	**1.18**	0.048	0.056	0.057	0.066	3.157	2.718	**Need to recalibrate**
5	1444	1346.2	1.505	4076.37	5522.36	1.07	0.047	0.055	0.051	0.059	1.428	1.227	-

CHAPTER VI
DISCUSSION AND CONCLUSIONS

In this chapter, Section 6.1 summarizes the problem statement, the developed methodologies, and the findings of the study. Section 6.2 then describes the limitations of the study, and Section 6.3 suggests options for future research on this topic.

6.1 SUMMARY OF THE RESEARCH

The methodology of calibration described in the HSM is not based on sound research, and several researchers who have calibrated the HSM SPFs to local jurisdictions have indicated limitations of the methodology. The recommendation of recalibrating the model every 2 to 3 years is also not supported by any evidence. Moreover, few studies have examined the recalibration of SPFs over time, and those that have suffer from certain issues, such as lack of transferability of results between jurisdictions, lack of statistical criteria for an acceptable error threshold, etc. Moreover, these studies usually consider a point estimate for the predicted crashes, but there is a variance associated with the SPF that needs to be considered for the recalibration of the models over time. Therefore, the present thesis developed guidelines for recalibration of the models over time by accounting for the model certainty. The two developed methodologies are summarized below.

1. **First methodology:** Recalibrate the SPF when the observed number of crashes on the entire network (for a facility type) is outside the 95% PI of the safety (m) of the entire network of that facility type.

The steps to apply the first methodology are as follows:

- Develop the SPF (flow-only model ignoring the effect of other traffic and geometric variables) using data from the initial period and obtain the variance-covariance matrix for this SPF.
- Apply the SPF to the entire network in the subsequent period and evaluate the variance of the safety (m) and the predicted mean for each of the individual sites.
- Compute the total observed number of crashes for the entire network, the total predicted number of crashes (using the latest calibrated model), the sum of the variance of the safety (m), and the sum of the variance of the predicted mean for the entire network.
- Evaluate the 95% PI of the safety (m). The model should be recalibrated when the observed number of crashes is outside this PI. If it is not, use the same model in the subsequent analysis period.
- Repeat the same procedure for the subsequent time periods (using the model last recalibrated) to check the need for recalibration.

2. **Second methodology:** Use the C-factor developed in year "t" (using all the available site data) and predict the number of crashes for the network in year "t+1". Evaluate the C-factor in year "t+1" (using all the available site data) and check whether it is significantly different than 1 at a 5% level of significance.

The steps to apply the second methodology are as follows:

- Develop the SPF (flow-only model, ignoring the effect of other traffic and geometric variables) and calculate the C-factor for Year 1, called C_{ref}. If a good model is developed, this value is expected to be close to 1 (use the data of the entire network to evaluate the calibration factor).

- Adjust the estimates in Year 2 using C_{ref}. Calculate the variances of the model estimate. Assume that the entire variance in the model estimate is due to the variance of C. The variance of the model estimates is obtained using the methodology proposed by Wood (2005) and Lord (2008). Estimate the calibration factor in Year 2, called C_2.
- The CV of the C-factor in Year 2 is now known. Calculate the standard deviation in Year 2 as (CV of C-factor in Year 2) * (C-factor estimate of Year 2), called σ.
- Test the hypothesis that the C-factor in Year 2 is significantly different than 1 at a 5% level of significance:

$$H_0: C-factor=1$$

$$H_1: C-factor \neq 1$$

$$t = \frac{C_2 - 1}{\sigma}$$

- If the null hypothesis cannot be rejected, repeat the same procedure for Year 3 using the C-factor developed in Year 1 (C_{ref}) for Year 3.
- If the null hypothesis is rejected, the model needs to be recalibrated in Year 2. Use the C-factor developed in Year 2 (C_2) and multiply it by C_{ref} to analyze the Year 3 data.
- Repeat the same procedure for the subsequent time periods to check the need for recalibration.

In their current form, neither methodology can be directly applied to the SPFs specified in the HSM, since the data required to evaluate the variance of the SPFs is not currently available in the HSM. The results of this study were compared to those using the methodology developed by Shirazi et al. (2017) to check whether the 10% error threshold recommended in their study is reasonable. Another advantage of Shirazi et al.'s (2017) methodology is that it can be applied to the HSM SPFs.

The methodologies were applied to several intersection and segment datasets from Toronto and Michigan. The results indicate that these methodologies can give the practitioner better guidance regarding the frequency of recalibration. Moreover, the comparison of the results with those obtained using Shirazi et al.'s (2017) methodology indicate that the 10% error threshold is reasonable to use for the analysis.

6.2 LIMITATIONS OF THE METHODOLOGIES

The following are some of the limitations of the proposed methodologies:

- Developing an SPF requires several data points of good quality and some prior experience in developing SPFs. It is a time-consuming and work-intensive process.
- Both these methodologies (in their current form) require extensive data collection in each year. The practitioner might as well use the HSM methodology for recalibration and estimate a new calibration factor every year. The main intention in developing these guidelines is to reduce the frequency of data collection efforts by the practitioner.

6.3 RECOMMENDATIONS FOR FUTURE RESEARCH

Future research should focus on determining the minimum number (and potentially the nature) of data points required to apply the proposed methodologies to obtain the same output regarding the recalibration. The data requirements of methodologies are intensive and, with the availability of such data, a separate calibration factor could be estimated every year; hence, there would be no need to identify the recalibration year. Another potential area of research is to simulate the crash data for certain scenarios and identify the sample size requirements. Furthermore, the use of simulation will allow to properly identify false positives and false negatives.

REFERENCES

American Association of State Highways and Transportation Officials (AASHTO) (2010), *Highway Safety Manual (1st Edition)*. Transportation Research Board of the National Academies, Washington, D.C.

Ang, A. H.-S., and W.H. Tang (1975). *Probability Concepts in Engineering Planning and Design: Basic Principles (Vol. 1).* John Wiley & Sons, New York, N.Y.

Bahar, G. (2014). *User's Guide to Develop Highway Safety Manual Safety Performance Function Calibration Factors*. National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C.

Brimley, B. K., M. Saito, and G. G. Schultz (2012). "Calibration of Highway Safety Manual Safety Performance Function." *Transportation Research Record 2279: Journal of the Transportation Research Board*, pp. 82–89.

Brown, H., C. Sun, and P. Edara (2014). "Nuts and Bolts of Statewide HSM Calibration." *In Transportation Research Board 93rd Annual Meeting* (No. 14-2587), Washington, D.C.

Connors, R. D., M. Maher, A. Wood, L. Mountain, and K. Ropkins (2013). "Methodology for Fitting and Updating Predictive Accident Models with Trend." *Accident Analysis & Prevention*, 56, pp. 82–94.

Dobson, A.J (1990). *An introduction to generalized linear models*. CRC press.

Feller, W. (1966). *An Introduction to Probability Theory and its Applications, Vol. 2.* Wiley, New York.

Harwood, D. W., F.M. Council, E. Hauer, W. E. Hughes, & A. Vogt (2000). *Prediction of the expected safety performance of rural two-lane highways* (No. FHWA-RD-99-207, MRI 4584-09, Technical Report). United States. Federal Highway Administration.

Kwak, S. G., & J. H. Kim (2017). "Central limit theorem: the cornerstone of modern statistics." *Korean Journal of Anesthesiology*, Vol. 70(2), pp. 144–156. http://doi.org/10.4097/kjae.2017.70.2.144

Lord, D., and B. Persaud (2000). "Accident Prediction Models with and Without Trend: Application of the Generalized Estimating Equations Procedure." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1717, pp. 102–108.

Lord, D., and B. Persaud (2004). "Estimating the safety performance of urban road transportation networks." *Accident Analysis & Prevention*, Vol. 36, No. 4, pp. 609-620.

Lord, D. (2008). "Methodology for estimating the variance and confidence intervals for the estimate of the product of baseline models and AMFs." *Accident Analysis & Prevention*, *40*(3), pp.1013-1017.

Lord, D. and F. Mannering (2010). "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives." *Transportation Research Part A: Policy and Practice*, *44*(5), pp.291-305.

Lord, D., S. Geedipally, and M. Shirazi (2016). *Improved Guidelines for Estimating the Highway Safety Manual Calibration Factors* (No. ATLAS-2015-10).

Mannering, F. L., and C. R. Bhat (2014). "Analytic methods in accident research: Methodological frontier and future directions." *Analytic Methods in Accident Research*, Vol. 1, pp. 1–22.

Martinelli, F., F. La Torre and P. Vadi (2009). "Calibration of the Highway Safety Manual's Accident Prediction Model for Italian Secondary Road Network." *Transportation Research Record: Journal of the Transportation Research Board*, 2103(1), pp. 1–9.

Mehta, G., and Y. Lou (2013). "Calibration and Development of Safety Performance Functions for Alabama." *Transportation Research Record 2398: Journal of the Transportation Research Board*, pp.75–82.

Miaou, S.-P., and D. Lord (2003). "Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods." *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1840, pp. 31–40.

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL http://www.r-project.org/

Saha, D., P. Alluri and A. Gan (2017). "A Bayesian procedure for evaluating the frequency of calibration factor updates in highway safety manual (HSM) applications." *Accident Analysis & Prevention*, *98*, pp.74-86.

SAS Institute. *SAS/STAT 9.1 User's Guide the Reg Procedure:(Book Excerpt)*. SAS Institute, 2008.

Shirazi, M., S.R. Geedipally, and D. Lord (2017). "A procedure to determine when safety performance functions should be recalibrated." *Journal of Transportation Safety & Security*, *9*(4), pp.457-469.

Wood, A. G., L. J. Mountain, R. D. Connors, M. J. Maher and K. Ropkins (2013). "Updating Outdated Predictive Accident Models." *Accident Analysis & Prevention*, 55, pp: 54–66.

Wood, G.R. (2005). "Confidence and prediction intervals for generalised linear accident models." *Accident Analysis & Prevention*, *37*(2), pp.267-273.

Xie, F., K. Gladhill, K. K. Dixon, and C.M. Monsere (2011). "Calibration of Highway Safety Manual Predictive Models for Oregon State Highways." Transportation Research Record 2241: Journal of the Transportation Research Board, pp. 19–28.