

Quantifying and Mitigating Uncertainty in Crash Risk Prediction for Road Safety Analysis

by

Reza Aminghafouri

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Civil Engineering

Waterloo, Ontario, Canada, 2025

© Reza Aminghafouri 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner

Dr. Bhagwant Persaud
Professor

Supervisor(s)

Dr. Liping Fu
Professor

Internal Member

Dr. Bruce Hellinga
Professor

Internal Member

Dr. Christian Bachmann
Associate Professor

Internal external Member

Dr. Changbao Wu
Professor

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Reza Aminghafouri was the sole author of Chapters 1, 2, and 6, which were written under the supervision of Dr. Liping Fu and were not intended for publication.

This thesis consists in part of three manuscripts written for publication. Exceptions to sole authorship of material are as follows:

Research presented in Chapter 3

This research was conducted at the University of Waterloo by Reza Aminghafouri under the supervision of Dr. Liping Fu. The study conception and design were a collaborative effort between Reza Aminghafouri and Dr. Liping Fu. Reza Aminghafouri was responsible for data collection, processing, and management. The analysis and interpretation of results were conducted jointly by Reza Aminghafouri and Dr. Liping Fu. The manuscript was drafted by Reza Aminghafouri with contributions and revisions from Dr. Liping Fu. Both authors reviewed the results and approved the final version of the manuscript.

Aminghafouri, R., & Fu, L. (2024). A Joint Confidence Region Approach to Ranking Hotspot Locations Considering Uncertainty in Expected Risk Estimates. *Transportation Research Record*, 2678(1), 555-567. <https://doi.org/10.1177/03611981231171917>

Research presented in Chapter 4

This research was conducted at the University of Waterloo by Reza Aminghafouri under the supervision of Dr. Liping Fu. The study was conceptualized and designed by Reza Aminghafouri and Dr. Liping Fu. All data collection, which involved the deployment and management of LiDAR sensors, was carried out by Reza Aminghafouri. The analysis of the relationship between sample size and prediction reliability, as well as the interpretation of the results, were performed by Reza Aminghafouri and Dr. Liping Fu. The manuscript was drafted by Reza Aminghafouri with critical feedback and revisions provided by Dr. Liping Fu. Both authors reviewed the results and approved the final version of the manuscript.

Aminghafouri, R., & Fu, L. The Myth of Quick Conflict-Based Road Safety Analysis: Limits of Short-Term Conflict Data in Collision Risk Prediction. *Accident Analysis & Prevention*, (Under Review).

Research presented in Chapter 5

This research was conducted at the University of Waterloo by Reza Aminghafouri under the supervision of Dr. Liping Fu. Reza Aminghafouri and Dr. Liping Fu jointly contributed to the conception and design of the automated threshold selection framework. Reza Aminghafouri was responsible for collecting and preparing the traffic conflict data from the City of Toronto. The analysis, which included the simulation study and case study application, and the interpretation of results were conducted by Reza Aminghafouri and Dr. Liping Fu. The manuscript was drafted by Reza Aminghafouri, with revisions and intellectual input from Dr. Liping Fu. Both authors reviewed the results and approved the final version of the manuscript.

Aminghafouri, R., & Fu, L. An Automated Threshold Selection Framework for Conflict-Based Road Safety Analysis. *Transportation Research Record*, (Under Review).

As the lead author of these three chapters, I was responsible for conceptualizing the study design, carrying out data collection and analysis, and drafting and submitting the manuscripts. My co-author provided guidance throughout each step of the research and gave feedback on draft manuscripts.

Abstract

Road safety analysis is a cornerstone of traffic safety management programs like Vision Zero, which aim to eliminate fatalities and serious injuries on roadways. Central to road safety analysis is the ability to accurately predict crash risk; however, this task is challenged by significant uncertainty arising from the random nature of crashes (aleatoric uncertainty) and limitations in data and modeling (epistemic uncertainty). These uncertainties can lead to the misidentification of hazardous locations, resulting in false positives and negatives, and the inefficient allocation of limited safety resources.

While numerous statistical models exist for risk prediction, most traditional crash-based approaches provide simple point estimates, failing to formally quantify the inherent uncertainty in their predictions. Proactive conflict-based analysis has emerged as a promising alternative that avoids direct reliance on sparse crash data, but its application introduces new methodological challenges. The reliability of conflict-based predictions is not well understood, and key methodological choices, such as the duration of data collection and the selection of analytical thresholds for Extreme Value Theory (EVT) models, introduce significant, often unaddressed, uncertainty into the results.

To overcome these challenges, this thesis systematically develops and evaluates a framework to quantify, investigate, and reduce critical sources of uncertainty in road safety analysis. First, to quantify the impact of uncertainty on network screening, a frequentist approach is employed to establish a joint confidence region (CR) for hotspot rankings, moving beyond simple point estimates. This is achieved by first estimating the confidence interval (CI) of risk for each location using a hierarchical Full Bayesian (FB) model that considers both crash frequency and severity. Second, this research investigates a primary source of data uncertainty in conflict-based analysis by systematically assessing the relationship between sample size and prediction reliability using a unique, year-long LiDAR dataset and a Bayesian Peak-Over-Threshold (POT) EVT model. Third, to address methodological uncertainty in EVT, an automated and objective approach for threshold selection is developed and validated, comparing a Sequential Goodness-of-Fit Selection Method (SGFSM) with an Automatic L-moment Ratio Selection Method (ALRSM) to reduce analytical subjectivity.

The analysis demonstrates that explicitly accounting for uncertainty can lead to substantially different hotspot identifications, revealing that rankings based on point estimates alone may be unreliable. The sample size analysis reveals that the common practice of using short-term conflict data is inadequate for reliable collision predictions, a finding that challenges the validity of a significant portion of the existing literature on conflict-based safety analysis. Finally, the automated threshold selection approach, particularly the L-moment-based approach, proves to be a robust and objective method that improves the accuracy of crash risk estimation. Collectively, this research provides researchers and practitioners with an evidence-based methodology to understand, quantify, and mitigate key uncertainties in road safety analysis, fostering more reliable safety assessments and a more effective allocation of resources.

Acknowledgements

I am profoundly grateful to my supervisor, Professor Liping Fu, for his invaluable guidance, expertise, and unwavering support throughout my Ph.D. journey. His mentorship was instrumental in shaping this research and my development as a researcher. I am especially thankful for his encouragement in both my academic and professional career, and for his flexibility, which made it possible for me to pursue this doctorate while working.

I also extend my sincere thanks to my Ph.D. examining committee: Dr. Bruce Hellinga, Dr. Christian Bachmann, Dr. Changbao Wu, and my External Examiner, Dr. Bhagwant Persaud. Your insightful feedback and constructive questions significantly improved the quality of this thesis. I am also deeply appreciative of Dr. Liraz Fridman and Steven Anderson from the City of Guelph; their extraordinary understanding and support were essential to this endeavor.

This research would not have been possible without institutional support. I am grateful to the Department of Civil & Environmental Engineering and my colleagues in the Innovative Transportation System Solutions (ITSS) Lab at the University of Waterloo for the excellent research environment. Special thanks are due to the City of Toronto and the City of Kitchener for providing the essential data for this work. I also gratefully acknowledge the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the University of Waterloo.

Finally, I am deeply thankful to my family and friends for their unwavering love, patience, and support. Your encouragement has been my source of strength, and I would not have accomplished this without you.

Dedication

To my beloved wife.

This journey would have been impossible without her unwavering love and encouragement. Through every late night of research and every moment of doubt, she was my constant source of strength and my loudest cheerleader.

And to my dearest parents.

Their sacrifices and endless support, from my very first day of school to this final milestone, laid the foundation upon which this achievement is built. I would not be here today, and I would not be the person I am, if it were not for the love and encouragement they have given me since I was a little boy.

This is for all of you.

Table of Contents

Author's Declaration	iii
Statement of Contributions	iv
Abstract	vi
Acknowledgements	viii
Dedication	ix
List of Figures	xiii
List of Tables	xiv
List of Abbreviations	xv
Chapter 1 Introduction.....	1
1.1 Background	1
1.2 Problem Statement.....	2
1.3 Research Objectives	3
1.4 Thesis Structure	5
Chapter 2 Literature Review	6
2.1 Risk Prediction Methods	6
2.1.1 Crash-based Road Safety Analysis.....	6
2.1.2 Conflict-based Road Safety Analysis	11
2.2 Sources and Quantification of Uncertainty	19
2.2.1 Types of Uncertainty	19
2.2.2 Sources of Uncertainty in Road Safety Analysis.....	19
2.2.3 Uncertainty Quantification Methods	26
2.3 Impact of Uncertainty on Road Safety Analysis	31
2.4 Approaches to Reduce Uncertainty in Road Safety Analysis	33
2.4.1 Development of Improved Crash-based Risk Prediction Models	33
2.4.2 Development of Improved Conflict-based Risk Prediction Models.....	35

2.4.3 Fusion of Different Sources of Evidence	36
2.5 Summary	40
Chapter 3 A Joint Confidence Region Approach to Ranking Hotspot Locations	42
3.1 Literature Review	43
3.2 Methodology	46
3.2.1 Confidence Interval of Crash Risk at Each Location	46
3.2.2 Confidence Region for Ranking of the Hotspots.....	50
3.3 Case Study	52
3.3.1 Data.....	52
3.3.2 Results and Discussions	54
3.4 Conclusion.....	62
Chapter 4 Conflict-Based Road Safety Analysis: Limits of Short-Term Conflict Data.....	65
4.1 Literature Review	66
4.2 Methodology	73
4.2.1 Study Framework	73
4.2.2 Traffic Conflict Indicator.....	74
4.2.3 Extreme Value Theory Models.....	76
4.2.4 Data Collection and Processing.....	80
4.2.5 Model Evaluation	82
4.2.6 Analysis of Sample Size Effects.....	84
4.3 Results	85
4.3.1 Descriptive Data Summary.....	85
4.3.2 Threshold Selection:	87
4.3.3 Generalized Pareto Distribution Model Fit and Validation	91
4.3.4 Sample Size Effects on Collision Prediction Reliability	93
4.3.5 Identification of Minimum Sample Size	97

4.4 Discussion and Conclusion.....	99
Chapter 5 An Automated Threshold Selection Approach for Conflict-Based Road Safety Analysis	102
5.1 Literature Review	103
5.2 Methodology	107
5.2.1 Traffic Conflict Indicator.....	107
5.2.2 Extreme Value Theory Model	108
5.2.3 Methods for Threshold Selection	109
5.3 Simulation Study	119
5.3.1 Simulation Settings:.....	119
5.3.2 Simulation Main Results	121
5.4 Case Study	123
5.4.1 Data Collection and Preparation.....	123
5.4.2 Results and Discussions	128
5.5 Conclusion.....	134
Chapter 6 Conclusion	136
6.1 Summary of Major Studies and Findings	136
6.2 Summary of Major Contributions.....	137
6.3 Recommendations for Future Research.....	138
References	140

List of Figures

Figure 1-1 Research Framework	4
Figure 2-1 Hyden's Safety Pyramid.....	12
Figure 2-2 Simple graphical model of crash events depicting conditional independence.....	16
Figure 2-3 95% CI and PI for NB model from (Ash et al., 2021)	32
Figure 3-1 Hierarchical NB Model Structure	49
Figure 3-2 Trace Plot of the MCMC Model for each Parameter.....	56
Figure 3-3 Estimated Total Risk with 95% CI	62
Figure 4-1 Selected intersections and collected LiDAR data.....	74
Figure 4-2 Illustration of PET calculation.....	75
Figure 4-3 Negated PET Conflict Distribution for Each Study Intersection.....	87
Figure 4-4 MRL Plot for the 30-day sample at Strasburg Rd & Old Country Dr	88
Figure 4-5 Modified Scale and Shape Parameters Stability Plot.....	89
Figure 4-6 GP distribution Model Diagnostic Plots for the 30-day sample	92
Figure 4-7 Sample of MCMC Trace plot for 30-days sample days	94
Figure 4-8 Convergence of Annual Collision Prediction and 95% CI with Increasing Sample Size...	96
Figure 4-9 Incremental Percentage Change in Mean Prediction	97
Figure 4-10 Incremental Percentage Change in CI Width.....	98
Figure 5-1 Hybrid Density Function Example	121
Figure 5-2 SGFSM and ALRSM Threshold Mean Bias Error Comparison.....	122
Figure 5-3 SGFMS and ALRSM Threshold Root Mean Squared Error based on 1000 Simulations	123
Figure 5-4 Selected Intersections	124
Figure 5-5 Negated PET Conflict Distribution.....	127
Figure 5-6 MRL Plot with %95 CI.....	128
Figure 5-7 Shape Parameter Stability Plot	129
Figure 5-8 Modified Scale Parameter Stability Plot.....	129
Figure 5-9 Euclidean Distance by Threshold	130
Figure 5-10 Data Probability Density Function	131
Figure 5-11 Data Cumulative Density Function.....	131
Figure 5-12 P-P and Q-Q Plots.....	132

List of Tables

Table 3-1 Descriptive Summary of Motor Vehicle Crashes by Severity and Traffic Volumes	54
Table 3-2 Summary of Model Calibration Results.....	57
Table 3-3 Summary of Crash Database Variables.....	58
Table 3-4 Summary of Severity Model Calibration Results	58
Table 3-5 Joint CR for Ranking based on 95% CI	60
Table 4-1 Summary of sample sizes and prediction uncertainties from recent EVT studies	68
Table 4-2 Characteristics and LiDAR deployment periods of the four study intersections	80
Table 4-3 Summary of conflict and collision data for the four study intersections.....	86
Table 4-4 Optimum threshold and number of extreme conflicts for each sampling period	90
Table 4-5 Estimated GP distribution Model Parameters for Strasburg Rd & Old Country Dr	93
Table 4-6 Predicted Annual Collisions and 95% CI by Sample Size.....	95
Table 4-7 Optimal Sample Size and Corresponding Collision Prediction Reliability	99
Table 5-1 Review Studies on Traffic Conflict Analysis.....	103
Table 5-2 Descriptive Summary of Traffic Conflict Data.....	125
Table 5-3 Threshold Selection and Crash Estimation	133

List of Abbreviations

AADT	Annual Average Daily Traffic
ACT	Anticipated Collision Time
AD	Anderson-Darling
AIC	Akaike Information Criterion
ALRSM	Automatic L-Moment Ratio Selection Method
AV	Autonomous Vehicle
BM	Block Maxima
CDF	Cumulative Distribution Function
CI	Confidence Interval
CR	Confidence Region
CTTC	Cross Time-to-Collision
DIC	Deviance Information Criterion
DRAC	Deceleration Rate to Avoid Crash
EB	Empirical Bayesian
EPDO	Equivalent Property Damage Only
EVT	Extreme Value Theory
FB	Full Bayesian
GEV	Generalized Extreme Value
GP	Generalized Pareto
GPS	Global Positioning System
GT	Gap Time
HSM	Highway Safety Manual
LMRD	L-Moment Ratio Diagram
MBE	Mean Bias Error
MCMC	Markov Chain Monte Carlo
MCSE	Monte Carlo Standard Errors
MRL	Mean Residual Life
MTTC	Modified Time to Collision
NB	Negative Binomial
NUTS	U-Turn Sampler
ORTT	Obstructed Right Turn Time
PDF	Probability Density Function

PDO	Property Damage Only
PET	Post-Encroachment Time
PI	Prediction Interval
POT	Peak-Over-Threshold
PSD	Proportion Of Stopping Distance
PWM	Probability Weighted Moments
QR	Quantile Regression
RMSE	Root Mean Squared Error
ROC	Risk of Collision
RSI	Relative Severity Index
RSMP	Roadway Safety Management Process
RTM	Regression-to-the Mean
SD	Standard Deviation
SGFSM	Sequential Goodness-of-Fit Selection Method
SPFs	Safety Performance Functions
SSMs	Surrogate Safety Measures
TTC	Time-to-Collision
VKT	Vehicle Kilometers Traveled
VRUs	Vulnerable Road Users

Chapter 1

Introduction

1.1 Background

In many countries, road crashes lead to high economic and social costs. For example, in Canada, the number of fatalities and injuries caused by road crashes was approximately 1,600 and 102,000 per year, respectively; in 2020, they resulted in approximately \$20 billion in total social costs or 1.2% of Canada's gross domestic product (Canada, 2020). Authorities of all government levels, from municipal and provincial to federal, are therefore under increasing pressure to improve road safety and reduce road crashes by initiating various safety programs such as Vision Zero. Vision Zero is an innovative road safety policy that aims to eliminate traffic fatalities or serious injuries completely in a transportation system. Over the past few decades, many municipalities have adopted the increasingly popular Vision Zero road safety strategy. The number continues to grow with significant amounts of resources being invested in these programs every year, including the implementation of a variety of safety improvement projects (Parachute, 2022; Vision Zero Network, 2022).

Essential to all safety programs is the Roadway Safety Management Process (RSMP) recommended by the Highway Safety Manual (HSM). RSMP encompasses a multi-step road safety analysis approach, including network screening, diagnosis, countermeasure selection, and safety effectiveness evaluation. A critical component underpinning all of these steps is the ability to accurately predict the safety level, or crash risk, of individual facilities like intersections and road sections. Crash risk prediction is a significantly challenging task due to the inherent random nature of crash occurrence and the large uncertainty associated with the model structure and the effects of various potential influencing factors. There is therefore a need to address various uncertainties inherent in road safety analysis, which is the primary focus of this thesis.

1.2 Problem Statement

The ultimate goal of any road safety program is to improve the overall safety level of a road network by implementing appropriate and effective countermeasures. Given a normally constrained budget, it is paramount to accurately identify locations with the highest priority for safety intervention and develop the most cost-effective safety improvements. The underlying tasks, known as network screening or hotspot identification and countermeasure selection, relies on the accurate prediction of crash risk. However, the assessment and prediction of crash risk are subject to significant uncertainties, which can be broadly categorized as aleatoric (related to data randomness) and epistemic (related to model limitations). Failing to account for these uncertainties can lead to flawed decision-making, such as false negatives (dangerous locations being overlooked) or false positives (safe locations being erroneously treated). While past research has been devoted to improving risk prediction, most studies have focused on developing models that produce a single point estimate of risk, largely ignoring the impact of this underlying uncertainty on the final decisions.

This thesis addresses this critical gap by focusing on three key methodological challenges facing road safety researchers and practitioners in the field of road safety:

1. **Uncertainty in Hotspot Ranking:** The conventional approach to network screening involves ranking locations based on point estimates of expected risk and identifying those at the top of the list as hotspots. This practice is problematic because it fails to consider the inherent statistical uncertainty of the risk estimates. An intersection ranked 5th might, when uncertainty is considered, have a risk level statistically indistinguishable from the intersection ranked 10th. Relying on simple point estimates without considering their inherent uncertainty can lead to flawed and unstable rankings, undermining the core objective of network screening.
2. **Uncertainty in Conflict-Based Analysis:** As a proactive alternative to traditional reactive, crash-based analysis, the use of traffic conflicts, or near-misses, has become increasingly popular. This approach uses the higher frequency of conflicts to extrapolate collision risk, which is especially valuable for locations with limited crash history. However, the prevalence of studies using very short observation periods, often

just a few hours or days, has created a misleading impression that reliable safety analysis using conflict data can be performed quickly. In reality, these short-term datasets could yield predictions with unacceptably wide confidence interval (CI), rendering them practically useless for decision-making. Near-misses severe enough to be indicative of collision risk are themselves rare events, and it remains largely unknown how much data is truly required to achieve reliable and robust risk estimates.

3. **Subjectivity and Model Uncertainty in Conflict Analysis:** Beyond the challenge of data quantity, the analytical methods for conflict-based analysis have their own sources of epistemic (model) uncertainty. A dominant methodology is the application of Extreme Value Theory (EVT), particularly the Peak-over-Threshold (POT) approach, which models all conflict events exceeding a certain severity threshold. The selection of this threshold is a critical step that fundamentally affects the resulting risk estimate. To date, most road safety studies have relied on subjective graphical methods, such as Mean Residual Life (MRL) plots, to determine this threshold. This subjectivity introduces a significant source of uncertainty and potential bias into the modeling process, as different analysts could arrive at different thresholds and, consequently, different risk estimates from the same dataset.

In conclusion, a systematic approach is needed to quantify, manage, and mitigate these key sources of uncertainty to advance the reliability and practical utility of road safety analysis.

1.3 Research Objectives

The primary goal of this research is to develop, and empirically validate, a framework, demonstrated in **Figure 1-1**, for understanding, quantifying, and mitigating key uncertainties in road safety analysis to support more reliable and evidence-based decision-making.

The specific objectives of this thesis are as follows:

1. To conduct a comprehensive literature review to identify, categorize, and synthesize the various sources of uncertainty in road safety analysis, including aleatoric (data) and epistemic (model) uncertainties in both traditional crash-based and modern conflict-based safety studies.

2. To develop and evaluate a methodology for explicitly incorporating uncertainty into the network screening process. This research moves beyond simple point-based rankings by constructing a joint confidence region (CR) for the rank of each location, providing decision-makers with a more robust and statistically sound understanding of site risk and priority.
3. To systematically quantify the impact of data collection duration on the reliability of conflict-based safety analysis. This work challenges the "myth of quick analysis" by using a unique, long-term (up to one year) empirical dataset to establish evidence-based guidelines on the minimum data required to achieve robust and meaningful collision predictions, identifying the point of diminishing returns for data collection efforts.
4. To address a critical source of model uncertainty in conflict-based analysis by developing and validating two different automated, objective approaches for threshold selection in POT models. This objective aims to replace subjective graphical techniques with a repeatable and statistically rigorous procedure, thereby improving the accuracy and consistency of risk estimation using EVT.

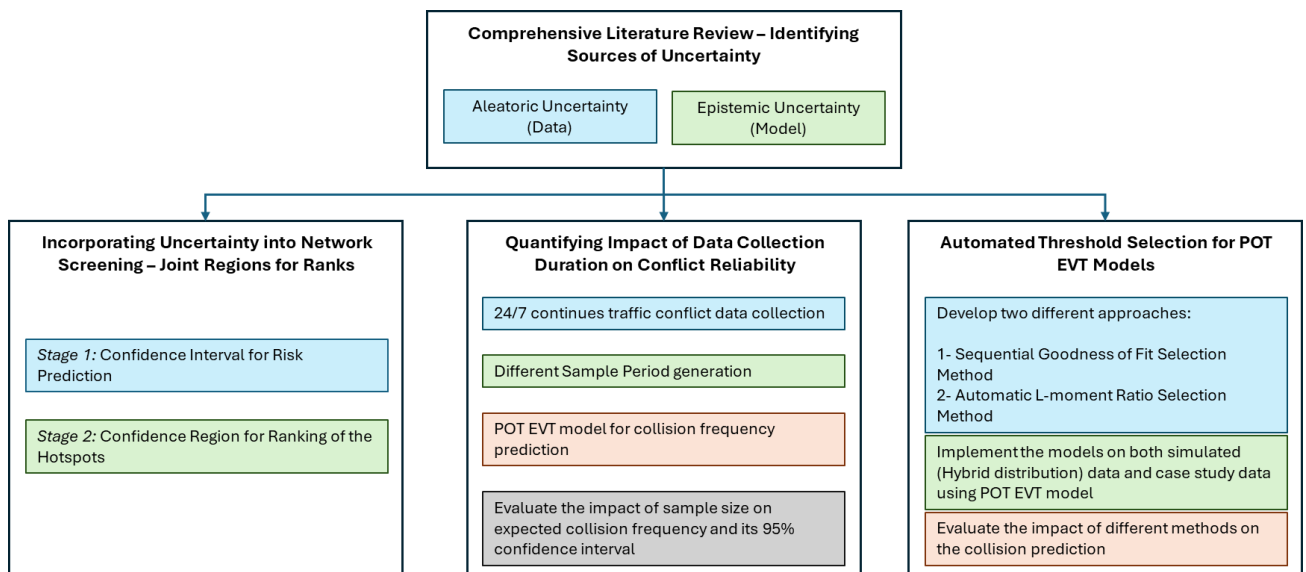


Figure 1-1 Research Framework

1.4 Thesis Structure

This thesis is presented in a manuscript-based format. Following this introduction, **Chapter 2** provides a comprehensive literature review that fulfills the first objective by systematically investigating the sources of uncertainty in road safety analysis. The subsequent chapters are presented as standalone manuscripts, each addressing one of the remaining objectives. **Chapter 3** introduces a novel method for ranking hotspots that explicitly accounts for uncertainty in risk estimates. **Chapter 4** presents automated threshold selection models to reduce model uncertainty in the application of EVT. **Chapter 5** investigates the critical relationship between data sample size and prediction reliability in conflict-based analysis. Finally, **Chapter 6** synthesizes the key contributions of this research, discusses its practical implications, and provides recommendations for future work.

Chapter 2

Literature Review

The problem of quantifying and reducing uncertainty in risk prediction has been studied extensively in the field of road safety study. This chapter provides a review of past efforts on the state-of-the-art models, and sources of uncertainty, along with methodologies on uncertainty quantification and reduction of crash risk prediction in the road safety analysis.

2.1 Risk Prediction Methods

As discussed in section 1.1, risk prediction is the first step of RSMP which can be done via various methods. Some methods rely on using crash data while others attempt to predict the risk through some surrogate safety measures (SSMs) data (i.e., traffic conflicts).

2.1.1 Crash-based Road Safety Analysis

A crash-based analysis evaluates a site's safety using its historical crash data. There are two common ways to estimate the risk of road entities using this method. The first approach is based on the number of crashes that have been observed (the Naïve approach), while the second approach makes use of a crash prediction model (model-based methods).

2.1.1.1 Naïve Approach

In the naïve methods, some risk measures, such as average crash frequency, crash rate, and weighted crash frequency according to crash severities, are first estimated directly using the crash history data observed over the road facilities of interest and then applied to estimate the crash risk of these facilities. The HSM gives a thorough overview of some of the most often-used methods (National Research Council, 2010).

1. The most basic criterion for identification is the **Average Crash Frequency**. Sites are listed in descending order of average observed crash frequency over time. In order to compare road segments with various lengths, the total number of crashes is divided by the segment length and time period. Consequently, this measure indicates the number of crashes per year per kilometer.

2. The **Crash Rate** is a statistic that compares the frequency of crashes normalized by the amount of traffic volume (i.e., exposure). The traffic volume on a road segment is measured in vehicle kilometers during the duration of the study period. This measure considers the crash risk of individual drivers on the road.
3. **Equivalent Property Damage Only (EPDO)** calculates a total frequency and severity score for each location by weighting crashes according to severity (death, injury, and property damage alone). The crash costs in the form of EPDO are used to determine the weighting factors.
4. **The Relative Severity Index (RSI)** technique assigns monetary crash costs to each crash category and sums the total cost of all crashes for each location. The average crash cost for each location is then compared to the total average crash cost of control group locations. As a result, the RSI determines whether a location has larger crash costs compared to the average of sites with similar attributes.

There are also more complicated naïve performance measures described in HSM (2010), which are less widely used in the literature, such as “Critical Rate” and “Excess Predicted Average Crash Frequency Using Method of Moments.”

1. The **Critical Rate** is another performance measure that compares the observed crash rate at each location to the critical crash rate of each site. The critical crash rate is a measure that is used to compare sites which have similar characteristics. Locations with higher observed crash rates than critical rates are identified for additional investigation. Generally, the critical crash rate depends on factors such as average crash rates at similar locations and traffic exposure.
2. The **Excess Predicted Average Crash Frequency Using the Method of Moments** performance measure modifies the observed average crash frequency for a site based on the variance in the crash data and the average crash frequency of the sample size locations. Then, the site's modified observed average crash frequency is compared with the average crash frequency of the reference population. The output of this comparison is the potential for safety improvement (PSI) index that can be used as a site's ranking criteria.

2.1.1.2 Safety Performance Functions

Safety performance functions (SPFs) are crash prediction models developed by fitting a statistical distribution on historical crash data and estimating the model parameters accordingly. Given specific features of interest (e.g., traffic volume), crash prediction models estimate a location's expected number of crashes. One advantage of implementing model-based approaches over naïve methods is that they do not rely solely on crash frequencies. They also consider factors like traffic exposure and road geometry related to the number of crashes. As a result, key factors that predict the risk level of a location can be identified.

The early efforts in crash frequency modelling were focused on predicting crash frequency using Poisson regression models (Gustavsson, 1969; Gustavsson & Svensson, 1976; Jovanis & Chang, 1986; Joshua & Garber, 1990; Miaou & Lum, 1993). Let Y_i be the number of crashes that may occur at a specific location i over a specified time period (i.e., five years). The Poisson model is based on the assumption that the crash frequency at location i follows a Poisson distribution with the mean of μ_i , that is as **Equation (2-1)**:

$$Y_i | \mu_i \sim \text{Poisson}(\mu_i) \quad (2-1)$$

The Poisson distribution gives the probability of a number of independent events occurring in a fixed time period. The mean μ_i is generally specified to be a function of covariates in the form of an exponential function as **Equation (2-2)**:

$$\mu_i = e^{\beta_0 + \sum \beta_i X_i} \quad (2-2)$$

where X_i is a vector of covariates and β_0 and β_i are the model intercept and a vector of covariates' coefficients, respectively.

While the Poisson model is quite simple to apply, it assumes the mean and variance of the distribution are the same and is, therefore, incapable of dealing with the overdispersion that is often found in crash data (Lord & Mannering, 2010). Overdispersion occurs when the variation of the crash counts is larger than the mean (Lord & Mannering, 2010). It is common to find

significant numbers of locations with zero crashes in over-dispersed datasets, which the conventional Poisson distribution cannot model well.

Several alternative models were subsequently developed to account for the overdispersion in crash data. In this regard, the most common model is the Negative Binomial (NB) model (Hauer et al., 1988; El-Basyouny & Sayed, 2006; Lord & Mannering, 2010; Srinivasan et al., 2010; Park et al., 2012; Connors et al., 2013; Ye et al., 2013). Mathematically, the NB model can be described as **Equation (2-3)**. Let Y_i be the number of crashes that may occur at a specific location i over a specified time period (i.e., five years), and X and β represent the vectors of site characteristics (e.g., traffic exposure, speed limit, number of lanes and legs) and their respective coefficients, respectively, with β_0 representing the model intercept. By assuming Y_i follows an NB distribution, the resulting NB model can be written:

$$Y_i \sim NB(\mu_i, \alpha) \quad (2-3)$$

where $\mu_i = e^{\beta_0 + \sum \beta_i X_i}$ is the mean and α is the overdispersion parameter.

In general, the NB model is an extension of the Poisson model and is also known as a Poisson-Gamma mixture model. When $\alpha \rightarrow 0$, the NB model converges to a Poisson model. In the NB model, the marginal mean and variance of a crash count are derived in such a way that the variance can be greater than the mean (Lawless, 1987; Hauer, 1992; Lord & Mannering, 2010).

Although the NB model can address the issue that arises from overdispersion, this model also has limitations since the error term can follow a distribution other than gamma (Hauer, 1997; Lord et al., 2005). Thus, many researchers have explored a range of other Poisson parameter distributions, resulting in a variety of additional forms of mixed-Poisson regression models (Lawless, 1987; Cameron & Trivedi, 2013). For instance, another type of mixed-Poisson distribution is the generalized inverse Gaussian distribution, which results in the Sichel model for overdispersed count data (Rigby et al., 2008). Other alternative mixed distributions are the Weibull and lognormal distributions, which result in the Poisson-Weibull and Poisson-Lognormal models, respectively. Both types of models can account for overdispersion. Cheng et al. (2013) investigated the Poisson-Weibull model using crash data, while others

investigated the Poisson-Lognormal model (Park & Lord, 2007; Aguero-Valverde & Jovanis, 2008; Lord & Miranda-Moreno, 2008; Lan & Persaud, 2012; Zhao et al., 2018).

In other attempts, the Poisson and NB regression models had been further extended to account for the excess of zeros which is expected in the crash count data and causes overdispersion. These modifications result in the zero-inflated Poisson and zero-inflated NB models, which have been utilized for a variety of purposes, including hotspot identification (Miaou, 1994; Shanker et al., 1997).

2.1.1.3 Empirical Bayesian Estimate

Although the aforementioned model-based approach using SPFs can address to a certain degree the limitation of the naïve method caused by the rarity of crash occurrences and low sample size, many issues remain such as regression-to-the-mean (RTM). Empirical Bayesian (EB) is a statistical technique that combines prior knowledge with empirical data to estimate unknown parameters and is proposed as an approach to overcome this source of uncertainty (Hauer et al., 2002).

EB method relies on Bayes' theorem, which is a mathematical formula (**Equation (2-4)**) that allows updating the beliefs (or probabilities) about an event based on new evidence.

$$P(\theta|D) = P(D|\theta) \times \frac{P(\theta)}{P(D)} \quad (2-4)$$

where, θ is the unknown parameter we are interested in estimating; D is the observed data; $P(\theta|D)$ is the posterior distribution of θ given D ; $P(D|\theta)$ is the likelihood function, which describes the probability of observing the data D given a particular value of θ ; $P(\theta)$ is the prior distribution, which represents the beliefs about θ before observing any data; and $P(D)$ is the marginal likelihood, which is the probability of observing the data D under all possible values of θ .

In general, the fundamental idea behind the EB method in road safety analysis is to estimate crash frequency at a given site by combining observed crashes (i.e., D) at the site and predicted average crash frequency from an SPF (i.e., θ).

These two safety measures are combined using a weight, which depends on the reliability of the SPF model. Mathematically, the expected annual number of crash frequencies from the EB method at site i , denoted as EB_i , is given as **Equation (2-5)**:

$$EB_i = w_i \bar{\mu}_i + (1 - w_i) \bar{y}_i \quad (2-5)$$

where, w_i is a weight given by $\frac{1}{1+\alpha.T.\bar{\mu}_i}$, T is the number of years of counted crashes, $\bar{y}_i$ is the observed average annual crash frequency at site i over the past T years, $\bar{\mu}_i$ is predicted average annual crash frequency at site i and is obtained from an SPF, and α is the dispersion parameter of the developed model (e.g., NB model), which is obtained as part of the calibration process.

The result of several studies shows that the EB method enhances the accuracy of the predictions. For example, Cheng and Washington (2005) evaluated three methods for identifying hotspots: simple ranking, CI, and EB. Using observed crash data properties, the authors generated hypothetical crash frequency distributions for various sites in order to conduct the evaluation. Also, simulations were run using a variety of factors to mimic real-world situations. The authors stated that when considering simulated data, sites with promise can be identified a priori, and false positives and false negatives can be calculated for method comparison. As a result, the EB method is shown to be significantly more accurate (i.e., less uncertainty) than the two other techniques.

2.1.2 Conflict-based Road Safety Analysis

Collecting crash data for road safety studies is essentially reactive, meaning we must wait for crashes to occur (and loss of life) to be able to identify hotspot locations. The relationship between road user interactions and their outcomes is effectively visualized using Hyden's safety pyramid (**Figure 2-1**). This conceptual model illustrates that for every fatal crash at the pyramid's peak, there is a much larger number of serious injuries, minor crashes, and an even broader base of traffic conflicts and undisturbed maneuvers at the bottom.

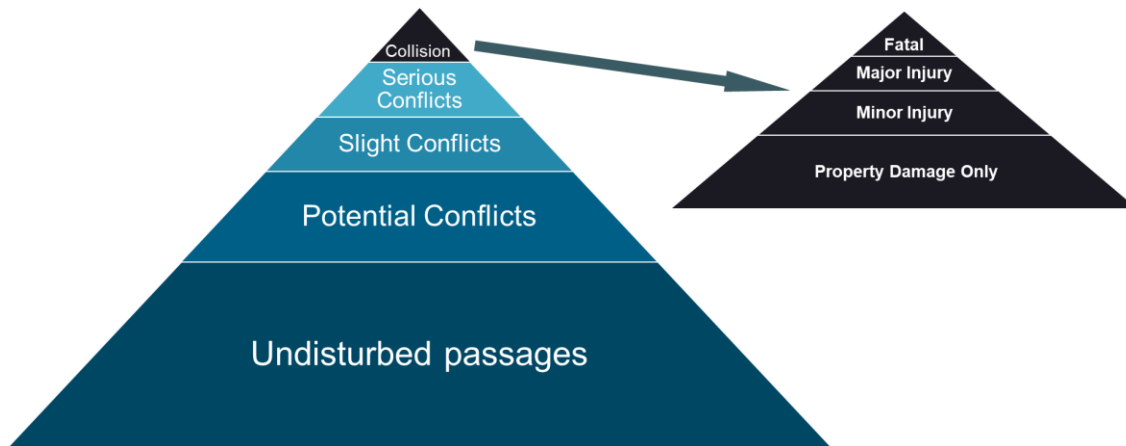


Figure 2-1 Hyden's Safety Pyramid

The pyramid highlights that severe crashes are relatively rare events, which makes statistical analysis challenging. This limitation as well as some other issues associated with crash data which will be discussed in detail in section 2.2.2 have led to the investigation of non-crash-based measures of safety, which are known as SSMs in the literature. Wu and Jovanis (2012) described some desirable criteria for selecting an appropriate SSM. In this regard, the crash surrogates should be able to be collected in a short period of time and be correlated with crash events.

Based on these criteria, traffic conflicts are considered an appropriate SSM for crashes. There are different definitions for traffic conflicts in the literature; in one of the earliest ones, Klebelsberg (1964) described a traffic conflict as a dangerous traffic interaction. Amundsen and Hyden (1977) defined a conflict as “... *an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged*”. In the literature, traffic conflicts are also known as “near-crashes”, “near misses”, “safety-critical events”, and “risky events”.

Various indicators have been developed to conduct road safety analysis based on traffic conflicts. Evasive actions and close proximity are the two main indicators of traffic conflicts. In general, the closer proximity of two vehicles in time and space, the higher the probability of a collision occurring. Similarly, a harsher evasive action taken by a road user suggests a more severe conflict.

There are various indicators in the literature based on proximity to a collision. Time-to-collision (TTC) and post-encroachment time (PET) are the most common indicators. The other proximity-based indicators can be categorized as a derivation of the TTC-based or PET-based indicators. TTC is the expected time for two vehicles to collide if they remain at their present speeds and on the same trajectories. The minimum observed TTC is taken as an indicator of a conflict's criticality. PET is defined as the "Time difference between the moment of an offending vehicle passes out of the area of a potential collision and the moment of arrival at the potential collision point by the conflicted vehicle possessing the right of way" (Allen et al., 1978).

Evasive-based indicators can better capture the severity of the conflicts than proximity-based ones. Generally, evasive actions can be measured based on factors such as acceleration, deceleration, and angular velocity of the vehicle's rotation (Zheng, Sayed, et al., 2021). For example, Cooper and Ferguson (1976) defined Deceleration Rate to Avoid a Crash (DRAC) indicator as the minimum required deceleration rate to avoid two vehicles colliding.

There are various methods for estimating the level of risk associated with any road entity based on conflict data. These methods can be divided into two main approaches: The first approach uses the traffic conflicts themselves (i.e., without considering the crash data) to model the safety level through conflict-frequency and conflict-severity models. The other approach is to find the correlation between crashes and traffic conflicts, which can be done through either crash-based or non-crash-based models (Zheng, Sayed, et al., 2021).

2.1.2.1 Conflict as a Surrogate Measure to Crash Data

In this approach, traffic conflicts of various types and levels of severity at locations of interest are first observed and models for these conflicts are then developed, linking to various factors such as traffic exposure and road characteristics.

Similar to crash-based analysis, different statistical methods could be applied to model conflict frequency as a function of various explanatory variables.

In the first attempt, Salman and Al-Maita (1995) conducted a study to examine how traffic conflicts are related to two different measures of traffic volume: the sum of volumes, and the square root of the product of volumes that caused conflicts. During the summer of 1992, a

comprehensive field study was carried out at 18 unsignalized, three-way junctions situated in Amman, Jordan. The study involved two observers being dispatched to each intersection to monitor and identify various traffic conflicts. Daily observations were made from 7:00 a.m. to 6:00 p.m. on typical weekdays throughout July and August of that year. They found that both measures had a strong correlation with traffic conflicts, but the square root of the product of volumes had a higher correlation than the sum of volumes.

In another study, Sayed and Zein (1999) describe how to use the traffic conflict technique to estimate traffic safety at intersections. This study has been conducted based on the traffic conflict frequency and severity for 94 signalized and unsignalized intersections throughout British Columbia. The study involved observing traffic conflicts at each intersection for two days, with 8 hours of observation per day. The severity of each conflict was assessed using two scores: the TTC score and the Risk of Collision (ROC) score. The ROC score is a subjective measure that takes into account the driver's perceived control over the situation, the severity of the evasive maneuver, and any limitations on response options. The overall severity score is the sum of the TTC and ROC scores, with the ROC score being independent of the TTC score. In addition, regression analysis was employed to develop predictive models linking the frequency of traffic conflicts with both traffic volume and collisions. The analysis findings indicate that, for both signalized and unsignalized intersections, there is a reasonable correlation between traffic volume and the average hourly conflict rate.

2.1.2.2 Conflict as an Explanatory Variable in Crash-based Models

In this method, the crash frequency is the dependent variable and traffic conflicts are an explanatory variable. Then, the relationship between the dependent and explanatory variables is estimated using various statistical methods.

In the first attempt, Hauer (1982) used a linear regression model to identify the relationship between crashes and conflicts: $\lambda = \pi * c$, where λ is the crash frequency at a location, c is the conflict frequency, and π is the conflict-to-crash ratio, which can be estimated based on historical crash and conflict counts for each location. Later, Hauer and Garder (1986) considered the severity level by extending the model to $\lambda = \sum \pi_i * c_i$, where c_i is the observed traffic conflict counts by severity level i and π_i is the conflict-to-crash ratio for severity level

i. Although linear regression is easy and straightforward to implement, it is unrealistic to assume that the conflict-to-crash ratios are constant for all locations and at all times within the different severity levels.

Generalized linear regression models are another type of model which could be used to overcome the abovementioned limitation of linear regression models. As explained earlier, Poisson, NB, and other types of mixed-Poisson models are commonly used in the literature to develop SPFs. The same models can also be employed to establish the relationship between crashes and traffic conflicts.

Several studies in literature tried to develop SPFs, including traffic conflicts as one of the explanatory variables. For example, El-Basyouny and Sayed (2013) used the NB modelling technique to predict crashes as a function of conflict counts. The conflict counts were derived from traffic volume and various geometry-related parameters using the Poisson-lognormal model.

Sacchi and Sayed (2016) proposed a two-phase nested modeling process for predicting traffic collisions by type at signalized intersections. In the first phase, a Poisson-gamma model was used to predict traffic conflicts from traffic volume and area type (urban or suburban). In the second phase, these predicted conflicts were used as the exposure variable in another Poisson-gamma model to predict collision frequency. The authors applied this methodology to data from 49 signalized intersections in British Columbia, using collision data from a three-year period (1995-1997) and conflict data collected by field observers in 1996. The study developed conflict-based SPFs for total, left-turn, and rear-end collisions and conflicts. The results show that conflict-based SPFs have better goodness-of-fit measures than traditional SPFs for all collision types except sideswipe, and can be used to estimate collision frequency by type more accurately, making them useful for safety diagnosis and evaluation at signalized intersections.

In another study, Stipancic et al. (2018) used Latent Gaussian Spatial Model to predict crash frequency to investigate the association between collisions and other possible safety indicators such as hard braking, traffic congestion, and speed variation. The study extracted SSMs from GPS data collected in Quebec. The results showed that incorporating spatial correlations and SSMs improved the model fit and supported the positive relationship between

certain SSMs and crash counts at the intersection level. This method has the potential to eliminate the reliance on crash data in road safety evaluation and monitoring.

2.1.2.3 Conflict as an Indicator to Predict Crashes

In this approach, which is also known as non-crash-based models, historical crash data are not used to estimate the models (but are used later to validate them). Non-crash-based models extrapolate the crash rate from observed traffic conflicts, and the extrapolation may be done using either causal inference or EVT (Zheng, Sayed, et al., 2021).

2.1.2.3.1 Causal models

The causal model does not rely on crashes as its basis. It suggests that the outcome of a crash denoted as " y " is dependent on both the initial condition " u " and the evasive action " x " as shown in **Figure 2-2**.

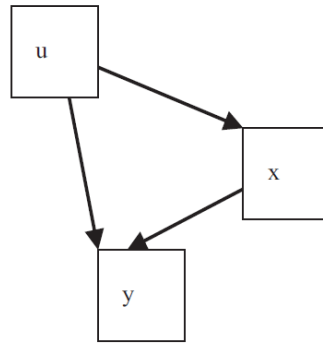


Figure 2-2 Simple graphical model of crash events depicting conditional independence (Davis et al., 2011)

Davis et al. (2011) constructed the causal model of conflicts and crashes in the form of **Equation (2-6)**:

$$P(y, x, u) = P(y|x, u)P(x|u)P(u) \quad (2-6)$$

where, y is the number of crashes, x is the number of evasive actions (i.e., conflict), $P(y, x, u)$ is the probability of crash outcome; $P(y|x, u)$ is a conditional probability for the evasive action and the initial conditions; $P(x|u)$ is a conditional probability for the evasive action; and $P(u)$ is a probability distribution over the values taken on the initial conditions.

Davis et al. (2011) propose that the likelihood of a crash occurring can be determined by adding up the probabilities of all potential actions that could result in a crash. By utilizing both causal and potential outcome models, a traffic conflict measurement can be created that evaluates the effectiveness of a driver's evasive maneuvers in preventing crashes. However, establishing the causal model requires a detailed analysis of an event and a complete knowledge of how evasive behaviors are carried out over various conflict severity levels. Given that, most traffic conflict approaches only obtain partial observation and a full understanding of the evasive action mechanism is lacking, such conditions are difficult to satisfy (Zheng, Sayed, et al., 2021).

2.1.2.3.2 Extreme Value Theory models

As discussed earlier, traffic conflicts, as measured by various conflict indicators, could vary over a wide spectrum in terms of the degree of severity, ranging from trivial encounters to near misses. In fact, road crashes could also be framed as a special type of traffic conflict with conflict indicators reaching their extreme levels, such as a TTC or PET equal to zero. This unifying view of crashes and conflicts has motivated the idea of applying EVT for modelling and predicting crashes based on traffic conflicts (Leadbetter et al., 1989; Songchitruksa & Tarko, 2006).

In general, there are two different approaches in EVT that can be applied for modelling the number of crashes from traffic conflicts. The first approach is called Block Maxima (BM) (or Minima) in which the observed traffic conflicts are grouped by a specific time interval (e.g., hour) into set blocks, and each block's maximum (in the case of traffic conflict measures such as PET and TTC, it should be minimum) is considered an extreme observation. The resulting extreme values (X) could be fit into a statistical distribution such as the generalized extreme value (GEV) distribution as shown in **Equation (2-7)** (Haan & Ferreira, 2006).

$$G(x) = \begin{cases} \exp \left[- \left(1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right)^{-\frac{1}{\xi}} \right], & \xi \neq 0 \\ \exp \left[- \exp \left\{ - \left(\frac{x - \mu}{\sigma} \right) \right\} \right], & \xi = 0 \end{cases} \quad (2-7)$$

where, μ is the location parameter ($-\infty < \mu < \infty$), σ is the scale parameter ($\sigma > 0$), and ξ is the shape parameter ($-\infty < \xi < \infty$).

The second method is the POT which involves identifying a threshold value (e.g., PET = 1.0 sec) and then labelling the observations that exceed (or fall below) this threshold as extreme values. With the POT data, the subsequent analysis could focus on either the frequency of extreme values (events) in a time period considered or the size of the exceedances. The first could be modelled as a count distribution, just like crashes while the threshold exceedance is commonly modelled as following the generalized Pareto (GP) distribution (**Equation (2-8)**):

$$G(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0 \end{cases} \quad (2-8)$$

When applying the two approaches for conflict analysis, conflict measures of interest (e.g., TTC, PET) are first selected and the best-fit distribution can then be determined following the standard statistical process. Once the distribution model is calibrated, the risk of a crash will be computed as **Equation (2-9)** (Zheng, Sayed, et al., 2021):

$$R = \Pr(Z \geq 0) = 1 - G(0) \quad (2-9)$$

where R is the crash risk probability, Z is the negated traffic conflict indicator, and $G(\cdot)$ is the EVT distribution calculated based on either the GEV distribution or the GP distribution. With the crash risk probability (i.e., R), the expected number of crashes can be estimated by assuming that the traffic conflict observation period t is representative of a long period T and calculating the estimated crashes N_T , as shown in **Equation (2-10)** (Zheng, Sayed, et al., 2021):

$$N_T = \frac{T}{t} R \quad (2-10)$$

where N_T is the expected number of crashes, t is the traffic conflict observation period, and T is the time period for predicting the number of crashes.

Campbell (1996) originally suggested using EVT for safety analysis, while Songchitruksa and Tarko (2006) provided concrete modelling approaches for demonstrating the association between traffic conflicts and collisions. Since then, the application of EVT for traffic conflict-based safety prediction has increased significantly in the literature.

2.2 Sources and Quantification of Uncertainty

2.2.1 Types of Uncertainty

In general, “uncertainty is a situation which involves unknown or imperfect information” (Knight Frank, 1921). Uncertainty comes from different sources, categorized into aleatoric and epistemic uncertainty (Li et al., 2012).

Aleatoric uncertainty arises due to the random nature of data which cannot be reduced. For instance, on a rainy day, the road surface can become slippery, and visibility can decrease, making it more difficult for drivers to maintain control of their vehicles. This increases the probability of collisions, but the precise timing, location, and severity of those collisions are unknown and are largely determined by the stochastic nature of the crash occurrence. This kind of uncertainty is also known as “external uncertainty”, “stochastic uncertainty”, “inherent uncertainty”, “irreducible uncertainty”, and “primary uncertainty” in the literature.

Epistemic uncertainty is another category that originates from the lack of knowledge. In other words, epistemic uncertainty happens due to the accuracy of a prediction model. Unlike aleatoric uncertainty, this type of uncertainty can be reduced by taking into account more knowledge and appropriate methodologies (Abdar et al., 2021). In the literature, epistemic uncertainty is also known as “knowledge or systematic uncertainty”, “internal uncertainty”, “informative uncertainty”, and “secondary uncertainty” (Kahneman & Tversky, 1982; Natke & Ben-Haim, 1997; Council, 2000).

2.2.2 Sources of Uncertainty in Road Safety Analysis

Uncertainty in any prediction model comes from various sources which can be divided into two main categories: 1) data uncertainty and 2) model uncertainty. Data uncertainty arises from noise, errors, or incompleteness in the data used for developing the model. Model uncertainty

arises from the assumptions, parameters, or structure of the model used to connect the input and output variables (Lazic & Williams, 2021).

2.2.2.1 Data Uncertainty

2.2.2.1.1 Uncertainty in Crash Data

As noted previously, risk prediction for road safety analysis is done based on either frequency of crashes or traffic conflicts. Thus, due to the random nature of crashes and conflicts, road safety analysis is subject to aleatoric uncertainty that cannot be reduced. Epistemic uncertainty in risk prediction could arise due to the quality issue of crash data that are often directly used either in calibrating models or as input to risk prediction. This source of uncertainty has been investigated in a few past studies. For example, Schlögl and Stütz (2019) analyzed both aspects to identify the sources of uncertainty in road safety analysis. Apart from the well-recognized issue of the under-reporting of crashes, the quality of the crash reports is often another area of concern.

On the one hand, it might not be feasible to determine the precise location of a crash due to the fact that a crash could occur over a certain distance along a road segment, ranging from some “crash-inducing” point to the final end position of the vehicles. On the other hand, additional uncertainty could be introduced when the police finally reached the crash scene. For example, in Austria, crash location is usually reported in two ways: 1) police officers manually set coordinates for the final end point of the vehicles within a web mapping service, or 2) police officers can provide a specification of the respective road kilometer based on driver location signs. This double reporting leads to spatial inconsistencies, which can be used to illustrate this aspect – a form of spatial uncertainty. In addition, the authors discussed the rounding effect of estimation of the time of crashes which is the temporal aspect of crash data uncertainty.

Both under-reporting and inaccurate crash reports could result in inaccurate risk models and predictions. However, little is known about the magnitude of this uncertainty and its propagation effect on the final estimated or predicted risk.

2.2.2.1.2 Uncertainty in Conflict Data

Although conflict data can solve some of the sources of uncertainties associated with crash data such as under-reporting and low-sample size, it still suffers from other sources of uncertainties. Zheng, Sayed, et al. (2021) provided a comprehensive literature review in terms of sources of uncertainties associated with conflict data.

Inconsistency in identifying conflicts is the first source of uncertainty; the detection of traffic conflicts is typically based on factors such as time and space proximity or evasive maneuvers. However, there are no universally accepted guidelines to aid in their identification, which leads to issues of inconsistency. When human observers identify traffic conflicts, inconsistency arises due to differences in observations both between and within observers. Similarly, when using quantitative measures to identify traffic conflicts, specific thresholds are necessary to differentiate between conflict and non-conflict events. However, these thresholds are often arbitrarily determined and have varied across different studies. The inconsistent identification of traffic conflicts poses challenges in cross-validating and generalizing research findings.

In addition, although traffic conflicts occur more frequently than crashes, observing them is time-consuming and requires significant resources. Typically, traffic conflicts are only observed for short periods of time due to resource constraints, such as several hours to several days. However, collecting data over short periods can result in biased estimations of safety levels due to the randomness of collected conflicts and the inability to account for long-term traffic variability. Therefore, using traffic conflicts observed over a short period to estimate the safety level over a longer period will likely result in biased estimations which is one of the sources of uncertainties.

2.2.2.1.3 Uncertainty in Traffic Volume or Exposure

Traffic volume or exposure of a road entity, one of the most critical pieces of input data in road safety analysis, is mostly obtained by field observation (e.g., permanent traffic counting stations or period traffic volume study) at some limited locations of a road network. Some data interpolation and traffic prediction process is required to estimate the traffic exposure at the road entities of interest (Schlögl and Stütz (2019)). Therefore, this process could lead to spatial uncertainty regarding the exact traffic volume at the location of interest. Also, Annual Average

Daily Traffic (AADT), which is used in most risk prediction models, is not always available due to limited temporal coverage of traffic counts, a source of temporal uncertainty in traffic volume data. Lastly, for predicting the crash risk of a road facility for a horizon year, future traffic exposure must first be forecasted, which could lead to another significant source of errors or uncertainty. Several studies have attempted to assess the impact of uncertainty in the traffic volume (i.e., AADT) in their analysis which will be discussed in section 2.3.

2.2.2.1.4 Uncertainty in Other Sources of Data

In addition to crash data and traffic volume, Schlögl and Stütz (2019) investigated data from other sources (including road geometry and condition data, and weather data) used in road safety analysis to find the spatial and temporal aspects of uncertainty. Regarding road geometry and conditions data, they showed no spatial uncertainty due to the high accuracy of the developed technologies. However, collecting those data in a snapshot of time can result in temporal uncertainties.

Weather data at a crash location is obtained from the nearest weather station; however, this proximity is the source of spatial uncertainty for these data. The temporal aspect of uncertainty in weather data is related to the temporal resolution of the data. For example, in most risk modelling, the risk is represented on a yearly basis, which means daily weather characteristics are difficult to be aggregated and considered in these modelling processes.

2.2.2.2 Model Uncertainty

Typically, the purpose of risk prediction models is to find the relationship between crashes and some explanatory variables by fitting a statistical distribution; however, these models also involve uncertainties. Modelling uncertainty occurs when this selected distribution does not perfectly represent the crash data. As the result, the predictions will suffer from epistemic uncertainty. Note that modelling uncertainty is different from the uncertainty due to the random nature of the crashes/conflicts data (i.e., aleatoric uncertainty).

2.2.2.2.1 Uncertainty in Model Assumption

One source of uncertainty in risk prediction models is the model assumptions. Models are simplifications of reality and rely on various assumptions about the factors that influence the crash occurrence and severity such as traffic flow, weather, road geometry, driver behavior, vehicle characteristics, and environmental complexity. However, these assumptions may not always hold true in reality, leading to uncertainty in the model's predictions.

For example, the transferability assumption of a risk prediction model in road safety analysis is the idea that a model developed for one data domain (such as a specific road segment, time period, or traffic condition) can be applied to another data domain with similar characteristics (Orsini et al., 2020). Farid et al. (2016) explored the transferability of SPFs (i.e., different NB models) for rural divided multilane highway segments across four states in the United States. Their study highlighted the importance of conducting transferability analysis before applying SPFs from one region to another and suggested using random effects models for calibration. They found that none of the SPFs were fully transferable without calibration, and they were resulted in prediction errors and biases. Therefore, it is important to calibrate or update the model parameters when transferring it to a new data domain.

In another study, Elvik (2010) identified some sources of uncertainty associated with the countermeasure effectiveness evaluation (i.e., risk prediction for the implemented countermeasures). Some of these sources can be implied that they will drive due to the assumptions considered during the model development such as:

1. Uncertainty about the definition of the target group of crashes influenced by a road safety countermeasure
2. Ignoring the random variation in the effects of countermeasures on crashes
3. Uncertainty in the assumptions of the effects of the exogenous factors on the effectiveness of a countermeasure
4. Uncertainty about the degree to which planned road safety countermeasures will be implemented
5. Uncertain monetary valuation of the benefits of road safety countermeasures

2.2.2.2.2 Uncertainty in Risk Modelling Parameters

Parameter uncertainty in road safety analysis refers to the variability or imprecision of the parameters that affect the performance or outcome of a prediction model. Parameter uncertainty can arise from different sources, such as input data, calibration methodology, and model structure.

For example, input data which is used to calibrate and validate the model parameters can be difficult to collect and may be subject to errors or biases. Insufficient, incomplete, or inaccurate data can lead to incorrect parameter values for variables in the model, which can affect the prediction results. For example, the crash-based models suffer from some limitations associated with the crash data, such as under-reporting, small sample size issues, and incapability to address the unobserved heterogeneity (Lord & Mannering, 2010; Savolainen et al., 2011; Mannering & Bhat, 2014). Also, since crashes are rare, observed crash data collected for only a short period of time are not the best measure of safety. Moreover, according to Mannering (2018) and Mannering et al. (2020), crash data may not accurately represent all drivers' behavior since riskier drivers are more likely to be involved in crashes. This limitation might lead to incorrect model parameter interpretations.

In addition, some studies have attempted to conduct a sensitivity analysis to figure out the impact of parameters' uncertainty in road safety analysis. For example, Lima Azevedo et al. (2015) used a multi-step sensitivity analysis framework to deal with a complex traffic simulation model with various traffic parameters. The authors attempted to measure uncertainty related to the parameters by using different methods including standard deviation (SD), coefficient of variance, CI, and error functions. The paper shows that assumptions on different parameters have different effects on SSMs. For example, the result of their study shows that the car-following sensitivity factor has a high impact on TTC, but a low impact on PET.

In another study, Zheng, Sayed, et al. (2021) discuss one source of uncertainty associated with conflict-based analysis. The authors mention that the required data for conflict-based analysis usually collect in small time intervals (e.g., a couple of hours) which can create a temporal correlation between multiple observations from the same road entity due to unobserved factors that remain constant over time. Additionally, traffic conflicts from different roadway entities

that are nearby can be spatially correlated due to shared unobserved factors such as weather or lighting conditions which can lead to spatial correlation. Therefore, these temporal and spatial correlations can result in uncertainty in the parameter estimates.

2.2.2.2.3 Uncertainty in Risk Prediction Methods

Uncertainty can also arise from the limitations associated with the risk prediction methods (section 2.1). For example, the Naïve approaches incorporate significant uncertainty in evaluating the safety level of the individual location, particularly when data are only accessible for a short period of time (e.g., 2 to 3 years or less). RTM is the main limitation of these kinds of measures resulting in uncertainty in crash risk prediction. Given the random nature of crashes, their frequency is expected to be lower following periods of higher levels, even if no safety improvement actions were taken. In addition, naïve techniques are limited in considering crash dispersion (Elvik, 2007) and heterogeneities associated with various site-specific characteristics such as traffic volume, which are other sources of uncertainty.

To overcome the uncertainties associated with naïve methods, several SPFs are developed (section 2.1.1.2); however, the main source of uncertainty, i.e., RTM biased, still exists in these kinds of models. Therefore, the EB technique was the next step toward risk prediction models to deal with uncertainty derived from the RTM issue. Although in all studies in the literature, it is obvious that the EB technique has a better risk prediction performance (i.e., less uncertainty) than conventional approaches, this technique also has some uncertainties. The main limitation of the EB approach is that the uncertainty in the predicted crash frequency can be only characterized by the overdispersion term, which is assumed to be constant across all the sites. Thus, after the SPF has been calibrated, all the point-estimated coefficients of the covariates in the model will be presumed to be true with no uncertainty. However, since the model covariate coefficients are often derived from crash data, they are inherently unpredictable. In some cases, these uncertainties result in predictions which do not represent the actual situation of the investigated location.

In addition, Zheng, Sayed, et al. (2021) discussed that in road safety analysis based on traffic conflicts, many studies rely on a single indicator to evaluate the severity of all traffic interactions, but this approach only captures a portion of traffic events. Consequently, traffic

conflicts defined by a single indicator will result in uncertainty since it cannot provide a complete picture of the safety of investigated locations (i.e., multidimensional issues).

In addition, conflict-based methods also suffer from sources of uncertainty. For example, when utilizing causal models (discussed in section 2.1.2.3.1), several assumptions need to be made, such as defining a set of initial conditions (U) and evasive actions (X). However, it is challenging to estimate the probability distribution for all potential evasive actions and initial conditions; so studies that use this definition tend to focus on a limited number of interactions and participants (Arun, Haque, et al., 2021b) which can be resulted in uncertainty in risk prediction.

2.2.3 Uncertainty Quantification Methods

All the above-discussed risk models focus on quantifying road safety based on a point estimate of the expected crash frequency at a specific location. While point estimates may be useful for many safety analysis tasks due to their simplistic representativeness of the overall risk, they do not contain any information on the variation or uncertainty associated with the risk (Casella & Berger, 2001). This section reviews the alternative approaches for quantifying the uncertainty in risk prediction.

2.2.3.1 Full Bayesian Approach

The Full Bayesian (FB) technique is a possible solution to the EB limitations. FB method is a statistical inference technique that uses Bayes' theorem to update the probability of a hypothesis based on all available data and prior knowledge. This method is characterized by its straightforward use of probability to quantify uncertainty in statistical data analysis. Therefore, since the uncertainty is considered in the model, a fitted distribution can show the estimated crash frequency (Gelman et al., 1995; Miaou & Lord, 2003). By estimating the posterior distribution of the parameters associated with individual parameters, FB represents a generalization of the EB technique, which assumes that the safety impact of different explanatory variables is fixed (i.e., no uncertainty).

The FB approach has three stages for calculating the predicted crash frequency at a certain traffic location (Huang et al., 2009):

Stage 1. Prior distributions are assigned to each model parameter (i.e., $Prior(\beta, v)$). The prior distributions are a measure of information about the effects of risk covariates (i.e., β) and dispersion parameters (i.e., v) on crash data.

Stage 2. Establishing a comprehensive probability model, which includes the cumulative probability of all coefficient covariates and the unobservable predicted number of crashes at certain locations as **Equation (2-11)**:

$$(M): Pr(y_{it}, \beta, v) = Pr(y_{it} | \beta, v) Pr(\beta, v) \quad (2-11)$$

where y_{it} is the crash frequency at i^{th} location in t^{th} time.

Stage 3. Based on all observed data (i.e., D), calculating the proper posterior distribution for all unknown coefficients and then combining it over the model covariates to achieve the marginal distribution for the estimated number of crashes at a specific location (**Equation (2-12)**).

$$\begin{aligned} & Pr(y_{it}^{est} | X_{it}, M, D, prior(\beta, v)) \quad (2-12) \\ &= \int Pr(y_{it}^{est} | X_{it}, V_{it}, M, D, prior(\beta, v), \beta, v) \\ &\times Pr(\beta, v | M, D, prior(\beta, v)) d(\beta, v) \end{aligned}$$

where, X_{it} is risk covariates at location i in time t , y_{it}^{est} is estimated crash frequency at location i in time t .

Markov chain Monte Carlo (MCMC), also known as Markov chain simulation, is a common method for establishing Bayesian inference through the random selection of parameters values from the fitted distribution and improving the accuracy by selecting new values considering the posterior distribution (Gelman et al., 1995).

With the advancement of statistical methods, many researchers have applied FB models for risk prediction of road entities (Miranda-Moreno et al., 2005; Miranda-Moreno et al., 2007; Huang et al., 2009; Miranda-Moreno et al., 2013; Wang et al., 2014).

Carriquiry and Pawlovich (2004) compare two Bayesian methods for analyzing traffic safety data: EB and FB. They explain the advantages and disadvantages of each method and provides examples of their applications. It concludes that FB is more flexible and realistic than EB. According to this study, FB methods could lead to dramatically low standard errors in predictions. In another study conducted by Persaud et al. (2010), it was determined that the difference between EB and FB methods was negligible and held no statistical importance. However, the research showed that the magnitude of uncertainties associated with the predictions varied significantly between the two approaches, and the FB method generated more precise estimates.

Aguero-Valverde and Jovanis (2009) applied a FB multivariate Poisson lognormal model to estimate the expected crash frequency for varying levels of crash severity. Their research assessed 6,353 rural two-lane road segments in Pennsylvania state, United States. The accuracy of these estimates was compared with traditional univariate Poisson lognormal estimates. The results indicate that the multivariate Poisson lognormal model provides a better fit and improves the precision of the crash-frequency estimates. The study then used these multivariate estimates in combination with cost data to develop the expected crash cost for each road segment, which was utilized to rank sites for safety improvements. The findings demonstrate that the top-ranked segments consistently had higher costs and excess costs than the univariate estimates. This is due to higher multivariate estimates of fatalities and major injuries caused by the random effects parameter. These higher estimated frequencies produce different rankings for the multivariate and univariate models.

In general, the FB method has numerous advantages in addition to providing a natural way to incorporate uncertainty. One of its primary benefits is that the method allows for the estimation of a full posterior distribution, which is more informative than point estimates, providing a range of plausible values for the parameters of interest. Additionally, the FB method can handle complex models and non-standard distributions, such as skewed or multimodal distributions, more efficiently than other methods. Furthermore, the approach has the capability to provide accurate predictions even in case of a low sample size.

2.2.3.2 Confidence and Prediction Intervals Approach

A Confidence Interval (CI) and a Prediction Interval (PI) are two different concepts used to quantify uncertainty. A CI estimates the plausible range for a population parameter (like the average), while a PI estimates the plausible range for a single future observation. In general, both are frequentist methods.

In road safety analysis based on PI, a normal or t-student distribution is approximated for the predicted crash frequency. Using this approximation, an interval is derived for the predicted crash frequency (Kim et al., 2022).

A few studies have tried to consider uncertainty by developing CIs and PIs for the estimates and predictions of road safety, respectively. In the first effort, Wood (2005) proposed a PI for the crash frequency at a new location (i.e., prediction response) as well as a CI for the expected crash frequency (i.e., Poisson mean) from the NB model. According to this study, an NB model comprises of three levels in its hierarchy. The first and lowest level is the mean response (μ_i); the second level is the Poisson parameter (λ_i); and lastly, the predicted response (y_i), which denotes the frequency of crashes at a particular site i , follows a Poisson distribution upon conditioning on the Poisson parameter (λ_i).

The author proved that by normal approximation assumption, $(1 - \alpha)\%$ CI for the mean response can be calculated based on **Equation (2-13)**:

$$\mu = \left[\frac{\hat{\mu}}{\exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\eta})}\right)}, \hat{\mu} * \exp\left(Z_{1-\frac{\alpha}{2}}\sqrt{Var(\hat{\eta})}\right) \right] \quad (2-13)$$

$$\eta = \log\left(\beta_0 + \sum_j \beta_j x_j\right)$$

where, η is the linear predictor, β_j is the j^{th} regression coefficient, x_j is the j^{th} covariate.

For the second level, the approximate $(1 - \alpha)\%$ CI for the Poisson parameter can be calculated as **Equation (2-14)**. Mathematically, it is feasible for the minimum value of CI to be less than

zero. However, from a physical perspective, the negative values of the Poisson parameter (i.e., collision rate) do not make sense which is considered in the lower bound of the CI.

$$\lambda = \left[\max \left\{ 0, \hat{\mu} - Z_{1-\frac{\alpha}{2}} \sqrt{Var(\lambda)} \right\}, \hat{\mu} + Z_{1-\frac{\alpha}{2}} \sqrt{Var(\lambda)} \right] \quad (2-14)$$

And in the last level, the predicted crash frequency considering $(1 - \alpha)\%$ PI can be obtained through **Equation (2-15)**:

$$y = \left[0, \left[\hat{\mu} + \sqrt{\alpha^{-1} - 1} \sqrt{\hat{\mu} + Var(\lambda)} \right] \right] \quad (2-15)$$

After utilizing these CI and PI on rear-end collision data from 392 legs of signalized intersections in New Zealand, an assessment was made on how traffic flow affects the level of uncertainty.

In another attempt, Ash et al. (2021) followed the initial study of Wood (2005) and derived formulas for PI and CI from different mixed-Poisson regression models, including the Poisson-Inverse-Gaussian, Sichel, Poisson-Lognormal, and Poisson-Weibull models. The analysis of these two studies will be further elaborated in section 2.3.

2.2.3.3 Quantile Regression Models

Another approach considering uncertainty in crash prediction is quantile regression (QR), which originated from econometric studies. In QR models, a response variable (i.e., dependent variable) is expressed as a conditional quantile (e.g., median) to a linear function of independent variables. In other words, despite the conventional regression models in which parameters are estimated based on ordinary least squares, the QR models estimate the conditional quantile of the least absolute error. Although several studies in finance and econometrics have been done based on the QR approach, only a few attempts have been made in the field of the road safety study.

Qin et al. (2010) identified Wisconsin's top 5% riskiest intersections using a 95% QR model. They analyzed the association between various factors and the 95th quantile of intersection crash data. The factors they considered included daily entering volume, number of legs and

lanes, and traffic control types. Three different ranking approaches were compared to the QR model: 1) ranking based only on observed crash frequency, 2) ranking based on average estimated crash frequency using the NB model, and 3) ranking based on the difference between observed crash frequency and 90% upper CI of NB model. The predicted values from the QR model were used to identify intersections with severe safety issues. The authors compared the results of QR with other methods and found that QR was preferable because it offered insights into the risk of crashes given specific location characteristics, unlike the count method which focused only on crash frequencies. QR also provided a more specific set of crash-prone locations compared to the CI limit method, which lacked specificity.

In another study, using the EPDO technique for considering the crash cost of different severity levels, Washington et al. (2014) developed a non-parametric QR model. They evaluated locations based on the output of the QR model and the observed EPDO. They found that the QR model pays more attention to locations with a higher crash severity level than the traditional point estimate-based approaches.

2.3 Impact of Uncertainty on Road Safety Analysis

As discussed earlier, uncertainty in collected data and models impacts the accuracy of predicted risk, and consequently, will affect the outcome of the road safety analysis. Several past studies have attempted to quantify the impact of uncertainty on the outcome of a road safety study; however, the majority of these studies focused only on evaluating the effect of uncertainty in the input data on crash risk estimates by performing a sensitivity analysis. For example, El-Basyouny and Sayed (2010) analyzed the impact of uncertainty in AADT on the predicted crash frequency. They developed two statistical models (i.e., NB model and measurement error NB model) on two different simulated datasets through the Monte Carlo approach. The result of their study shows that the traditional NB model underestimates the predicted crash frequency in the presence of high AADT.

Zarei and Hellinga (2022) propose a simulation-based approach to quantify the monetary benefits of improving AADT accuracy. The study found that crash data sets with higher sample mean, and dispersion parameter values are more sensitive to AADT error and would benefit

more from improving AADT accuracy. The authors illustrate the application of their proposed method using a real-world example.

In another study, Ding et al. (2021) investigated the relationship between traffic flow uncertainty and some of the drivers' behavioral characteristics factors on the rear-end crash risk. They used a hierarchical hybrid Bayesian network model to compare the predicted crash risk variation with and without traffic flow uncertainty. The result of the study shows the risk of rear-end crashes and drivers' car-following approach were found to be affected to a large extent by traffic flow conditions.

Ash et al. (2021) and Wood (2005) attempted to analyze the impact of changes in traffic flow on the CIs for the Poisson mean (μ), and Poisson parameter (m), as well as PI for the predicted crash frequency (y). In these studies, they developed various Poisson-based regression models such as NB, Sichel, Poisson-Lognormal, etc. For example, in **Figure 2-3**, the range of uncertainty in all three CIs and PI is exponentially expanded as the AADT is increased.

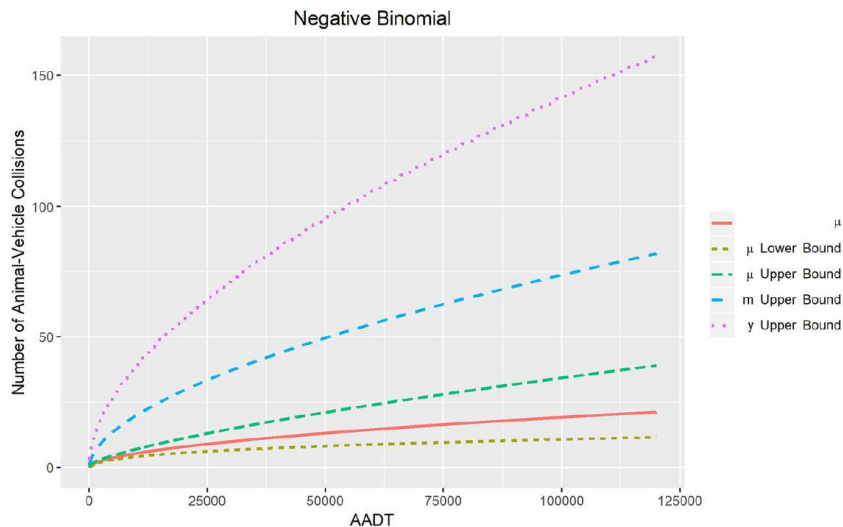


Figure 2-3 95% CI and PI for NB model from (Ash et al., 2021)

In another similar study, Geedipally and Lord (2008) evaluated the impact of varying dispersion parameters of the NB model on the width of PI for the predicted crash frequency. They concluded that considering different dispersion parameters will result in a narrower range of PI for the locations with higher traffic volume.

In summary, all of these past studies have focused on evaluating the impact of uncertainty, especially traffic exposure, on crash prediction. To the best of our knowledge, there is no study to quantify the impact of uncertainty on the final output of a road safety program (i.e., hot spot identification, countermeasure effectiveness evaluation, etc.).

2.4 Approaches to Reduce Uncertainty in Road Safety Analysis

As discussed earlier, the problem of risk prediction in road safety analysis is along with a variety of sources of uncertainty with some being inherent or unavoidable (i.e., aleatoric uncertainty) while others epistemic that can be reduced with improved data quality and risk modelling. As a result, significant efforts have been devoted in the past to reducing epistemic uncertainty with a specific focus on, developing more accurate risk prediction models or new risk measures and finding ways to maximize the use of available sources of evidence (i.e., different road safety measures), as discussed in the following section.

2.4.1 Development of Improved Crash-based Risk Prediction Models

The FB model, as discussed in section 2.2.3.1, is among the most flexible models available; however, the basic forms of this model are still subject to uncertainty arising from various sources. Thus, several past studies have attempted to improve FB models by incorporating other factors or models to address some of the uncertainties associated with previous ones. For example, Miaou and Song (2005) ranked locations using Bayesian generalized linear mixed models considering multivariate spatial random effects. This study states that spatial correlation considerably affects model goodness of fit and should be addressed during hotspot identification.

Another advantage of using the FB method is that risk can be estimated considering panel crash data (i.e., time series data). One significant challenge when studying panel data is the possibility of correlations between observations. In general, panel data possess two types of correlations which are another source of uncertainties in the prediction models: (1) temporal correlations between crash data at a particular location across time and (2) spatial correlations between crash data at several locations within a given geographical area over the same period of time.

Greene (2003) demonstrated when there are temporal and spatial correlations in the data, the standard statistical models which consider the unobserved errors as an independent term will lead to biased estimations in the parameters. Thus, some studies have attempted to consider temporal and spatial correlations in their crash prediction models. The fixed and random effect Poisson and NB models are the most often employed models (Shankar et al., 1998; Chin & Quddus, 2003; Ulfarsson & Shankar, 2003; Miaou & Song, 2005; Caliendo et al., 2007; Jiang et al., 2013). For example, Jiang et al. (2013) examined the predictive ability of random effect Poisson and random effect NB models. They concluded that the random-effect NB model, considering the temporal correlation of crash data, is better than the cross-sectional NB and Poisson models. In another study, using multivariate spatial and temporal models, Cheng et al. (2018) predicted both vehicle and active user crashes to conduct hotspot identification. The research demonstrated that considering temporal and spatial correlations between crash data leads to finding more accurate hotspot locations.

Unobserved heterogeneity associated with crash data is one of the uncertainty sources of crash prediction models. To fully address this issue of heterogeneity, Heydari et al. (2016) proposed a Bayesian nonparametric Dirichlet process mixture model and evaluated the performance of their approach by comparing with various types of models, including NB, random effects and traditional latent class models.

In another study, by considering heterogeneity in mean and variance of the crash data, Heydari et al. (2018) developed a Bayesian heteroskedastic random parameter Poisson lognormal model on the highway railway grade crossings crash data of eight major Canadian provinces. They showed the effectiveness of their model in comparison to several standard random effects and parameters models.

Additionally, the FB method's flexibility in explicitly describing hierarchical structures is another advantage in addressing uncertainty. Due to the data collecting and grouping processes, hierarchical patterns are prevalent in crash data. Various research has shown that considering spatial and temporal correlations using a hierarchical modelling method may improve the accuracy of the fitted models on the crash data (MacNab, 2003; Miaou & Song, 2005; Huang et al., 2008). Huang et al. (2009) developed a hierarchical FB framework for road safety analysis. To consider constant and serial temporal correlation, the authors used two

hierarchical FB Poisson Log-Normal models and compared them with the result of the NB model using EB adjustment. The road safety analysis in this study started with ranking the locations with respect to their average crash count estimated through each model by using three years of historical crash data. Then, the prone locations were identified by defining a percentage threshold for the top-ranked locations (e.g., top 5%). In the end, these prone locations were compared to "true" hotspots determined by a ten-year period of historical crash counts (1997–2006). The results indicate that the hierarchical FB models greatly outperform the usual EB technique even when it comes to hotspot identification considering only temporal correlation.

2.4.2 Development of Improved Conflict-based Risk Prediction Models

Due to the uncertainties and limitations associated with crash data discussed in section 2.2.2.1.1, the usage of traffic conflict data (i.e., near-miss) in road safety analysis has increased significantly in recent years. Although using conflict data instead of crash data can reduce some aspects of the uncertainty associated with crash data, such as under-reporting, and temporal and spatial uncertainties, these kinds of data still suffer from the uncertainty related to the modelling structure discussed in section 2.2.2.2.

To improve the accuracy of the risk prediction using conflict data, several more modelling techniques have also been developed. For example, Xie et al. (2019) created a hidden Markov model to estimate the frequency of conflicts over intervals of five minutes. In this study, unobserved heterogeneity is accounted in the model by the hidden states of traffic situations, which are associated with various safety levels. Moreover, some studies predicted traffic conflicts using artificial intelligence models; for example, Stylianou and Dimitriou (2018) used a Bayesian network model for conflict prediction. In another attempt, Formosa et al. (2020) predicted real-time traffic conflicts utilizing deep learning models.

In addition, in most of the early efforts, the EVT method for crash prediction using conflict data was mainly applied to univariate models for individual sites, whereby crashes were modelled as the extreme value of a single traffic conflict indicator of a specific location. Most of the studies based on the EVT approach are of this type (i.e., univariate EVT model) (Tarko, 2012; Lai et al., 2014; Farah & Azevedo, 2017; Tarko, 2018).

Although the occurrences of traffic conflicts are significantly higher than crashes, extreme conflicts are still rare. Thus, the number of conflicts, which are closer to crashes at each individual site, is usually limited, especially when the data are collected over a short period of time. To overcome this limitation, some researchers tried to develop an EVT model based on a single traffic conflict indicator collected from multiple sites. Hence, these kinds of models are called “joint-site univariate EVT models” as they use data from multiple locations. Three main advantages are associated with using the “joint-site univariate EVT model. First, by combining the traffic conflict extremes of several sites, it is possible to borrow data from other sites, significantly improving the accuracy of model predictions. In addition, considerations of covariates that may impact the stationarity of extremes can be used to obtain better model estimates and improve inferences for nonstationary extremes. Finally, additional factors that may cause unobserved heterogeneity can be accounted for, allowing fewer uncertain inferences about crashes (Zheng, Sayed, et al., 2021).

Based on the GEV distribution and the GP distribution, Zheng and Sayed (2019a) and Zheng and Sayed (2019b) constructed Bayesian hierarchical extreme value models to aggregate traffic conflicts of multiple locations for crash prediction. The structure of their model includes three layers: 1) a data layer in which observations are modelled with the GEV distribution and the GP distribution; 2) a process layer in which a latent Gaussian process underlying extremes is used to connect the parameters of the data layer to covariates; and 3) assigning the prior values to unknown parameters which represent the latent Gaussian process as a prior layer. The findings indicate that the hierarchical FB model based on GP distribution produces more precise crash estimates rather than the models based on GEV distribution. Additionally, the results show that the proposed hierarchical FB framework helps to overcome the issue of limited sample size by taking into account unobserved heterogeneity.

2.4.3 Fusion of Different Sources of Evidence

One source of epistemic uncertainty associated with conflict-based analysis is the multidimensional nature of conflicts (section 2.2.2.2.3). Thus, using only one indicator for analyzing safety levels will lead to unreliable predictions since it captures an incomplete understanding of the event. In the literature, there are various approaches to integrating multiple traffic conflict indicators to overcome the multi-dimensional issue:

2.3.3.1 Joint Activated Triggers of Multiple Conflict Indicators

In this approach, a conflict is defined as an occurrence that takes place when any combination of specified "triggers" is engaged. In this method, defining the severity of traffic incidents will be done by personal judgment or objective cluster approaches of a number of conflict indicators (Kaparias et al., 2010; Salamati et al., 2011; Tageldin & Sayed, 2019). For example, naturalistic driving studies can include several kinematic triggers, such as horizontal and vertical acceleration (Dingus et al., 2006). The triggers will be activated when the indicators exceed the defined thresholds. The triggered events are then evaluated by the experts to identify truly critical events.

This approach has several limitations, one of which is its subjectivity. In addition, the lack of standards in selecting the kinematic indicators and their thresholds is another drawback of this approach (Wu & Jovanis, 2012; Wu & Jovanis, 2013).

2.3.3.2 Mapping Multiple Traffic Conflict Indicators

The mapping approach is the other way to combine multiple traffic conflict indicators and crashes to a common index, and it can be done in two ways: 1) functional mapping, and 2) distribution mapping.

In functional mapping, closed-form functions are employed to convert the results of different crash surrogate indicators into a single safety index (Ismail et al., 2011; Ghanipour Machiani & Abbas, 2016). Numerous studies have simultaneously mapped conflict frequency and severity indicators using functional mapping (Zhou et al., 2011; Wang & Stamatiadis, 2014a; Zheng & Ismail, 2017).

In the latter method, i.e., distribution mapping, a statistical distribution (e.g., gamma distribution) is individually assigned to each traffic conflict indicator; after that, each surrogate measure's extremes (i.e., conflicts that exceed the thresholds) are jointly modelled to evaluate safety. In other words, by combining different statistical distributions, an integrated distribution function is extracted to evaluate the safety (Ismail et al., 2011). For example, Wang and Stamatiadis (2014b) combined TTC and Required Braking Rate into a single probabilistic function called Aggregated Crash Propensity Metric.

Copulas models are one of the most common approaches in developing joint distribution that has been used in various studies in the literature. For example, in one of the latest studies, Yang et al. (2021) developed a joint model for the crash count and two traffic conflict indicators including TTC and DRAC. The authors, at first, developed marginal NB distribution for each piece of evidence and then using a multivariate Gaussian copula model, they combined all three marginal models. They found that this approach allows for better consideration of the impact of potential unobserved factors on the crash count and conflict events.

2.3.3.3 Meta-analysis Approach

A meta-analysis is a statistical analysis that integrates the findings of multiple investigations. This method is used when several studies are addressing the same issue, each of which provides measurements that are expected to contain some degree of error. Consequently, the objective is to apply statistical methods to obtain a joint estimate to reduce the uncertainty derived from each individual piece of evidence (Rosenthal & DiMatteo, 2001). According to the literature on evidence-based medicine, meta-analysis results are the most reliable source of evidence. In addition to being used in the field of medical research, this method has also been applied extensively in the field of road safety to evaluate the effectiveness of various countermeasures (i.e., crash modification factors) or analyze the impact of different factors on the number of crashes. For example, Elvik (2013) performed a meta-analysis on 264 risk estimates to investigate the effects of eleven drugs on the number of crashes.

2.3.3.4 Multivariate Extreme Value Theory Approach

In this approach, GEV or GP distributions (section 2.1.2.3.2) are used to approximate the marginal distribution of the various traffic safety indicators (Zheng et al., 2018; Wang et al., 2019; Zheng & Sayed, 2019c, 2019d).

In the first attempt, Zheng et al. (2018) developed a bivariate GP distribution model to predict the crash frequency of entrance merging areas on freeways. To validate the developed method, Zheng and Sayed (2019c) used the same model on the collected data of four signalized intersections. In another study, Wang et al. (2019) employed a bivariate GP distribution model on traffic conflict data to predict crash frequency at signalized intersections.

Zheng and Sayed (2019d) used both GP and GEV distributions to develop a bivariate EVT framework on two signalized intersections in the city of Surrey, British Columbia. The authors compared the result of predicted crashes using univariate and bivariate EVT models to the observed historical crashes and discussed the strengths and weaknesses of each approach.

Arun, Haque, et al. (2021b) established a bivariate EVT model to predict the crash frequency by severity (i.e., severe crashes and non-sever crashes) using TTC and Modified Time to Collision (MTTC) as traffic conflict frequency indicators and Delta-V as an indicator for crash severity analysis. The models were developed on two four-legged signalized intersections in Brisbane, Australia. In general, the result of their study demonstrates that bivariate EVT models outperform the univariate models in terms of more precise estimates of crashes.

The above-mentioned bivariate models can be considered as “at-site multivariate EVT models” since the models are separately developed on each individual site (Zheng, Sayed, et al., 2021). A follow-up model of bivariate EVT is the “joint-site multivariate EVT model”, which uses a hierarchical FB structure to combine the various traffic conflict indicators of different sites. In this regard, Zheng and Sayed (2020a) developed a joint-site bivariate EVT model based on the PET and MTTC indicators of four signalized intersections. They showed that this kind of EVT approach could result in more accurate crash prediction.

While bivariate models have demonstrated significant enhancements when compared to univariate models, using only two traffic conflict indicators may not capture all the severity dimensions of a traffic event. Hence, a multivariate hierarchical Bayesian extreme value model with a high dimensionality that employs more than two conflict indicators is necessary (Fu et al., 2020). In this regard, few studies have attempted to predict the number of crashes by combining three conflict indicators.

As the first attempt, Fu et al. (2020) propose a method for estimating the multivariate hierarchical Bayesian extreme value model and apply it to estimate rear-end crashes at four signalized intersections in Surrey, British Columbia. The conflict indicators used were modified TTC, PET, and DRAC. To account for non-stationarity in conflict extremes, three covariates including traffic volume, shock wave area, and platoon ratio were considered. The results indicate that the trivariate hierarchical Bayesian extreme value model performs better

than both the bivariate and univariate hierarchical Bayesian extreme value models in terms of crash estimation accuracy (i.e., comparing the estimation and the historical collisions).

In another study, Fu and Sayed (2022b) propose a new multivariate method for evaluating real-time road safety using novel multivariate EVT models that integrate multiple conflict indicators and joint safety indices. The method was applied to evaluate cycle-level real-time safety at four signalized intersections in Surrey, British Columbia using three conflict indicators: modified TTC, PET, and DRAC. The authors developed four types of multivariate EVT models (i.e., tilted Dirichlet, pairwise beta, Husler-Reiss, and extremal-t) and found that the Husler-Reiss model had the best goodness-of-fit for both bivariate and trivariate frameworks. The results showed that multivariate models can identify higher numbers of crash-risk cycles than univariate models.

2.5 Summary

This chapter provided a comprehensive review of the literature on risk prediction methods in road safety analysis, examining the sources of uncertainty and the methods for their quantification and reduction.

The principal findings are as follows:

- **Sources of Uncertainty:** Risk prediction in road safety analysis is subject to numerous uncertainties stemming from various sources, including the quality and quantity of crash and conflict data, as well as model assumptions, parameters, and structures.
- **Gap in Existing Research:** The majority of past studies have focused on point estimates of risk, such as expected crash frequency. While a few studies have explicitly considered uncertainty in their predictions, very few have evaluated the impact of this uncertainty on key road safety analysis outcomes like network screening and safety effectiveness evaluation.
- **Challenges in Proactive Methods:** Conflict-based analysis is a promising proactive alternative to traditional crash-based methods. However, it introduces its own significant uncertainties, including inconsistent definitions of conflict indicators, challenges in data collection duration, and the often-unsubstantial correlation between observed conflicts and actual crashes.

- **Limitations of Conflict Indicators:** Most existing conflict-based analyses rely on a single indicator to measure risk, which provides only a partial view of a location's safety level (Ismail et al., 2011). While recent research has begun to incorporate multiple indicators, a clear understanding of how to best select and combine them is still lacking.

Among these identified gaps, the prevalent use of point estimates in crash prediction represents a fundamental flaw in current practice. This is particularly problematic for network screening, where ranking locations based on single values can lead to flawed and unstable hotspot identification, undermining a core objective of safety management programs.

To address this critical gap, this thesis first turns its attention to quantifying and incorporating uncertainty within the conventional crash-based network screening process. The following chapter introduces a novel method for ranking hotspot locations that moves beyond misleading point-based ranks by constructing a joint CR for each location. This approach provides a more robust and statistically sound understanding of site risk and priority. By tackling this foundational issue in road safety analysis, Chapter 3 sets the stage for the subsequent chapters, which will address the specific data and model uncertainties within proactive, conflict-based methods.

Chapter 3

A Joint Confidence Region Approach to Ranking Hotspot Locations Considering Uncertainty in Expected Risk Estimates

As established in the preceding chapters, road safety analysis is a critical process for improving road safety, yet it is fundamentally challenged by uncertainty arising from the random nature of crashes and limitations in modeling. The literature review highlighted that a primary component of road safety analysis, network screening, traditionally relies on point estimates of crash risk to rank and identify hazardous locations. This practice fails to account for the inherent statistical uncertainty in these predictions, which can lead to flawed rankings and an inefficient allocation of safety resources. An intersection ranked as the most dangerous based on a point estimate might not be statistically distinguishable from one ranked fifth, a nuance that conventional methods ignore.

To address this critical gap, this chapter introduces a novel method for ranking hotspot locations that explicitly incorporates statistical uncertainty. By moving beyond simple point-based rankings, this work provides decision-makers with a more robust and statistically sound understanding of site risk and priority. This is achieved by first estimating a CI for the crash risk at each location and then using these intervals to construct a joint CR for the overall ranking. This approach provides a range of plausible ranks for each location, reflecting the true uncertainty of the estimation process.

This chapter is based on the manuscript: Aminghafouri, R., & Fu, L. (2024). A Joint Confidence Region Approach to Ranking Hotspot Locations Considering Uncertainty in Expected Risk Estimates. *Transportation Research Record*, 2678(1), 555-567.

The remainder of this chapter reviews the literature and details the methodology for developing this approach, including the use of a hierarchical FB model to estimate crash frequency and severity. It then presents a case study on three-legged stop-controlled intersections in Kitchener, Ontario, to demonstrate how considering uncertainty can fundamentally alter hotspot identification and lead to more reliable decision-making.

3.1 Literature Review

SPFs, developed through fitting a statistical distribution on historical crash data, are a common approach for evaluating road entities' risk levels. Poisson distribution is one of the most used probabilistic models in road safety studies. While the Poisson model is quite simple to apply, it cannot deal with the overdispersion often found in crash data. Hence, several models have been developed to account for the overdispersion in crash data. The most often used model is the NB employed by several researchers to model crash data (Lord & Mannering, 2010).

The NB distribution is generally derived from the Poisson distribution by including a gamma-distributed error term in the Poisson parameter. However, some scholars argue that it is not needed to suppose that the Poisson parameter must only follow the gamma distribution (Lord et al., 2005). Thus, researchers have explored a range of other Poisson parameter distributions, resulting in a variety of additional forms of mixed-Poisson regression models such as Poisson-Weibull and Poisson-Lognormal models (Zhao et al., 2018).

All of the above-mentioned models are only focused on a point estimate of the predicted crash frequency (i.e., mean) at a specific location. While point estimates may be useful for prediction due to their simplicity, they do not provide information on the uncertainty in the predictions. In general, there are two main categories of uncertainty in statistics: aleatoric and epistemic (Hüllermeier & Waegeman, 2021). Aleatoric uncertainty (known as data uncertainty) arises due to the data distribution's stochastic nature. On the other hand, epistemic uncertainty (known as knowledge or systematic uncertainty) happens due to the accuracy of a prediction model (Abdar et al., 2021). According to Abdar et al. (2021), uncertainty quantification is important in any prediction models that are used for critical decision-making. In this regard, few studies have tried to consider uncertainty in the predictions by developing CI for the estimates of road safety.

In the first effort, Wood (2005) conducted a PI for the crash frequency at a new location (i.e., prediction response) as well as a CI for the expected crash frequency (i.e., Poisson mean) from the NB model. In another attempt, Ash et al. (2021) followed the initial study of Wood (2005) and derived formulas for PI and CI from different mixed-Poisson regression models, including the Poisson-Inverse-Gaussian, Sichel, Poisson-Lognormal, and Poisson-Weibull models.

Another approach to considering uncertainty in the crash prediction model is QR models, which originated from econometric studies. A few attempts have been made in road safety using the QR method. Qin et al. (2010) identified Wisconsin's top 5% riskiest intersections using a 95% QR model. In another study, using the EPDO technique for considering the crash cost of different severity levels, Washington et al. (2014) developed a non-parametric QR model. They evaluated locations based on the output of the QR model and the observed EPDO.

The FB technique is another possible solution to quantifying uncertainty in statistical data analysis. FB approach combines prior and posterior information, resulting in a fitted distribution that can provide the estimated crash frequency (Miaou & Lord, 2003). Several researchers have used FB models for hotspot identification (Miranda-Moreno et al., 2007; Huang et al., 2009; Miranda-Moreno et al., 2013). For example, Agüero-Valverde and Jovanis (2009) applied a FB approach by considering multivariate Poisson Log-Normal models to identify hotspots through crash severity estimation. The authors concluded that the multivariate FB model performs well for hotspot identification. Miaou and Song (2005) ranked locations by crash cost rate using Bayesian generalized linear mixed models considering multivariate spatial random effects. This study has analyzed the frequency of the crashes by three different severity levels, including fatal, major injury, and minor injury at the county level.

Unobserved heterogeneity associated with crash data is one of the uncertainty sources of crash prediction models. To fully address this issue of heterogeneity, Heydari et al. (2016) proposed a Bayesian nonparametric Dirichlet process mixture model and evaluated the performance of their approach by comparing with various types of models including NB, random effects, and traditional latent class models. In another study, by considering heterogeneity in mean and variance of the crash data, Heydari et al. (2018) developed a Bayesian heteroskedastic random parameter Poisson lognormal model on the highway railway grade crossings crash data of eight major Canadian provinces. They showed the effectiveness of their model in comparison to several standard random effects and parameters models.

The FB flexibility in explicitly describing hierarchical structures is another advantage. Due to the data collecting and grouping processes, hierarchical patterns are prevalent in crash data. Recent research has shown that using a hierarchical modelling method may improve the accuracy of the fitted models on the crash data (Miaou & Song, 2005). Huang et al. (2009)

developed a hierarchical FB model for road safety hotspot identification. The authors used two hierarchical FB Poisson Log-Normal models to consider constant and serial temporal correlation. The analysis in this study started with ranking the locations with respect to their average crash count estimated through each model using three years of historical crash data. Then, the crash prone locations were identified by defining a percentage threshold for the top-ranked locations (e.g., top 5%). The results indicate that the hierarchical FB models significantly outperform the common EB technique even when it comes to hotspot identification considering only temporal correlation.

So far, all the studies which used the FB approach for hotspot identification obtained the final predicted crash frequency by taking an average from all possible values corresponding to the fitted distribution. To the best of our knowledge, no study has considered uncertainty explicitly in ranking and prioritizing the locations for safety improvement.

Several studies in statistics and economics have considered uncertainty in the ranking. In general, uncertainty in the ranking of a population can be done through two main approaches, the Bayesian and the frequentist frameworks. In the Bayesian approach, the uncertainty in the ranking is evaluated by defining credible intervals for an estimator, and the rankings are then obtained using posterior probability (Vitelli et al., 2018). However, the frequentist method is based on constructing a CI to rank a population. Developing such confidence intervals can be done using either the bootstrap and other computational methods (Zhang et al., 2014) or differences in means (Al Mohamad et al., 2022). Most studies in the literature use a frequentist approach to produce confidence intervals for the individual ranks. However, recently, some studies have focused on developing a joint CR by considering simultaneous confidence intervals for the individual ranks.

For example, the paper of Klein et al. (2020) was the first study on conducting a method to create a joint CR for the ranking of the U.S. states based on the mean of travel time using the Bonferroni-corrected CI method. In another study, Rising (2021) followed the method of Klein et al. (2020) in generating a joint CR; however, they used a partial order set to develop the CI for the rank of U.S. states in terms of the mean of travel time. Their results showed that their set estimator could produce a shorter range for the same CR.

Mogstad et al. (2020) also conducted a similar analysis to the paper of Klein et al. (2020) to construct confidence sets for the joint ranking of a population. The only difference between their research is that they exploited joint confidence sets for the differences in the estimator value of certain population pairs instead of constructing simultaneous confidence sets for all population entities. They also implemented the method of this paper in another study to construct a 95% CR for the ranking of the academic journals and universities around the world based on considering uncertainty in their impact factor (Mogstad et al., 2022).

3.2 Methodology

The main objective of this study is to identify hotspots by considering uncertainty for the rankings. A two-step approach is proposed. In the first step, 95% CI for the crash risk of each location is estimated. Based on the calculated CI, a CR for the ranking of each location is then calculated in the second step. In the following, the detail of each step is discussed.

3.2.1 Confidence Interval of Crash Risk at Each Location

Consider a set of K sites, denoted by j ($j = 1, 2, \dots, K$), to be ranked by their collision risk. A 95% CI for the risk value of a given site can be calculated by estimating the lower and upper limit of crash frequency by severity. By considering the social cost of each severity level, the CI for risk can be calculated as **Equation (3-1)**:

$$95\% \text{ CI of } \hat{R}_i \in [L_{R_i}, U_{R_i}] \quad (3-1)$$

$$L_{R_i} = \sum_j L_{N_{ij}} \times C_j$$

$$U_{R_i} = \sum_j U_{N_{ij}} \times C_j$$

where, $\hat{R}_i$ is the estimated total risk at site i ; L_{R_i} and U_{R_i} are the lower and upper limit of estimated total risk at site i , respectively; j is an index representing a collision severity level (e.g., fatal, major, minor, etc.); $L_{N_{ij}}$ and $U_{N_{ij}}$ are the lower and upper limit of expected annual

crash frequency of severity level j at site i , respectively; and C_j is cost associated with the severity type j .

Then, a two-stage modelling approach is used to calculate crash frequencies for each severity level. First, we estimate the 95% CI for the total crash frequency (i.e., all severity levels) of a location using a hierarchical full Bayesian NB model. Accordingly, the crash frequency by severity at each site is derived by multiplying the probability of each severity type calculated by a multinomial logit model as **Equation (3-2)**:

$$L_{N_{ij}} = L_{\mu_i} \times P_{ij} \quad (3-2)$$

$$U_{N_{ij}} = U_{\mu_i} \times P_{ij}$$

where, L_{μ_i} and U_{μ_i} are respectively the lower and upper limit of 95% CI for the expected annual collision frequency at site i and P_{ij} is an estimate of the probability of severity type j for a given site i .

3.2.1.1 Full Bayesian Method for Predicting Crash Frequency ($L_{N_{ij}}, U_{N_{ij}}$)

FB analyses are characterized by their straightforward use of probability to quantify uncertainty in statistical data analysis. In FB, statistical distributions are fitted to the model parameters, coefficients, and estimated crash frequency. The final predicted crash frequency can be obtained by taking an average or any quantiles from all possible values corresponding to the fitted distribution (Miaou & Lord, 2003). This technique has three stages for calculating the predicted crash frequency at a specific traffic location (Huang et al., 2009):

Stage 1. Establishing a probability model for crash frequency: In road safety, one of the most popular statistical models in use is the NB regression model, which is obtained from the Poisson model by assuming a gamma distribution for the error term (i.e., ε_{it}). The NB regression model can be expressed as **Equation (3-3)**:

$$y_{it} \sim \text{Poisson}(\lambda_{it}) \quad (3-3)$$

$$\lambda_{it} = \text{Exp}(x'_{it}\beta + \varepsilon_{it}) = \mu_{it} \times \text{Exp}(\varepsilon_{it})$$

$$\varepsilon_{it} \sim \text{Gamma}(\phi, \theta)$$

where, y_{it} is the observed crash frequency of site i in time t ; λ_{it} is the mean predicted crash frequency of site i in time period t ; x'_{it} is a vector of intercept and explanatory variables (e.g., traffic volume); β is a vector of coefficients of explanatory variables; and ϕ and θ are the shape and scale parameters of the gamma distribution, respectively.

As it can be seen in **Equation (3-3)**, the error term changes across road locations and time. Therefore, the NB model can consider the overdispersion issue inherent in the standard Poisson model (Lord & Mannering, 2010). However, the NB model limitation is that it does not account for the overdispersion associated with unobserved heterogeneities. Hierarchical models are an alternative to usual statistical models to address this issue. Therefore, this thesis adopts a hierarchical NB model to develop the FB model.

Stage 2. Assigning prior distributions to each model parameter: The prior distributions are a measure of information about the effects of risk covariates and dispersion parameters on crash data. In the second stage, the parameters' prior distributions must be specified to obtain the FB estimates. The prior distributions can generally be categorized into informative (i.e., prior information about the priors is available) and uninformative priors. Furthermore, in the context of a hierarchical FB model, it is assumed that a prior distribution assigned to a parameter has an error term depending on another parameter with a different distribution. With this definition, we follow the hierarchical Bayes structure presented by Miranda-Moreno et al. (2007) for the error term (ε_{it}) of the NB model as **Equation (3-4)**:

$$\begin{aligned}\varepsilon_{it} &\sim \text{Gamma}(\phi, 1/\phi) \\ \phi &\sim \text{Gamma}(a, b)\end{aligned}\tag{3-4}$$

where a and b are the fixed hyper-parameter values, and the relation between the parameters can be shown in **Figure 3-1**.

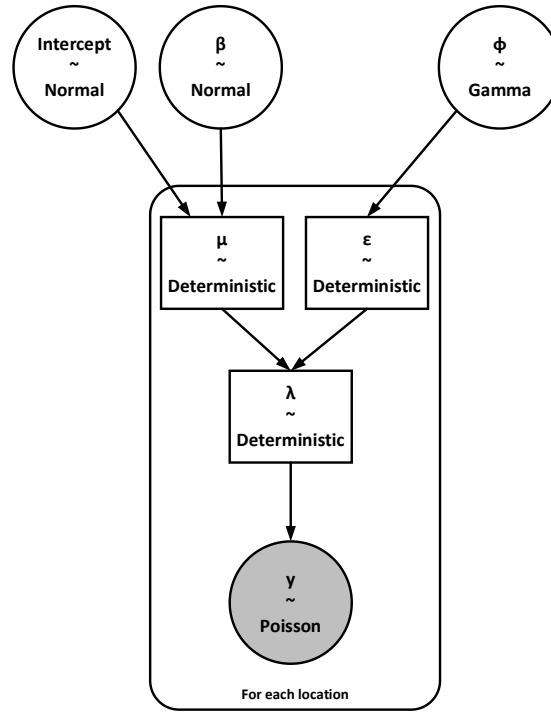


Figure 3-1 Hierarchical NB Model Structure

Stage 3. Calculating the proper posterior distribution for all unknowns' coefficients: MCMC sampling is a common method for establishing Bayesian inference through the random selection of parameter values from the fitted distribution and improving the accuracy by selecting new values considering the posterior distribution (Gelman et al., 1995). This method considers the prior distributions as an initial point and generates different chains of random points until the posterior distributions are converged using Bayesian theory. During the sampling, one part of the iterations will be considered burn-in, and the remaining ones will be used for parameter estimation.

Convergence is critical in FB models since it will guarantee that a posterior has been found for the parameters. An assessment of convergence is made by examining the MCMC trace plots for the model parameters as well as monitoring the ratio of Monte Carlo errors to SD of the estimates; these ratios should be less than 0.05 as a general rule (El-Basyouny & Sayed, 2009). This study uses the PyMC3 Package of Python to generate the samples. After posterior calculations, the marginal distribution for the estimated number of crashes at a 95% CI will be obtained for each specific location.

3.2.1.2 Multinomial Logit Model for Predicting Crash Severity (P_{ij})

A collision severity model is used to predict the possible consequence of a collision. In this study, five possible discrete outcomes are considered: fatal, major injury, minor injury, minimal injury or property damage only (PDO), each representing a specific level of severity. The probability associated with each severity type is modelled using a multinomial logit model.

While the severity outcomes are inherently ordered, the multinomial logit model was chosen over ordered alternatives (e.g., ordered probit) as it provides greater flexibility, allowing the model to capture how specific factors might influence one severity level (like a fatal crash) differently than another (like a minor injury), revealing more nuanced relationships. In a multinomial logit model, the probability that a collision results in a given severity type j at site i can be estimated as **Equation (3-5)**:

$$P_{ij} = \frac{e^{Z_{ij}}}{\sum_{j=1}^5 e^{Z_{ij}}} \quad (3-5)$$

where, i is an index representing a given site i ; j is an index representing severity type –with $j = 1, 2, 3, 4$ and 5 for fatality, major injury, minor injury, minimal and PDO, respectively; Z_{ij} is a measure representing the propensity for a collision at site i to experience severity type j , and; P_{ij} is an estimate of the probability of severity type j for a given site i .

The propensity value (Z) in **Equation (3-5)** is commonly expressed as a linear function of the site characteristics and attributes of the individual involved in the collision. The coefficients of the model are calibrated using the maximum likelihood method.

3.2.2 Confidence Region for Ranking of the Hotspots

As a first step towards identifying hotspots, it is necessary to rank the locations based on their total risk. Generally, the estimated total risk is ranked using point estimates such as the mean of the expected risk for each location (i.e., $\hat{R}_i$). As a result, the location with the highest mean of expected risk is ranked 1, the site with the second highest means is ranked 2, and so forth. Mathematically, the estimated overall ranking of the location i (denoted by $\hat{r}_i$) based on the mean of estimated total risk, $\hat{R}_i, \dots, \hat{R}_K$ can be defined as **Equation (3-6)**:

$$\hat{r}_i = 1 + \sum_{n: n \neq i} I(\hat{R}_i \leq \hat{R}_n), \quad \text{for } i = 1, \dots, K \quad (3-6)$$

where, $I(\hat{R}_i \leq \hat{R}_n)$ is a binary function that returns 1 if the condition is met and 0 if false. However, using the point estimate approach does not incorporate uncertainty in the ranking since the true values of the estimator (i.e., the total risk) are unknown. To address this issue, the frequentist approach presented by Klein et al. (2020) is used in this study. Based on the calculated 95% CI for the estimated total risk (i.e., $[L_{R_i}, U_{R_i}]$), the joint CR for the overall ranking of the locations is determined.

Once the CI has been calculated for each location, $i \in \{1, 2, \dots, K\}$, the following sets (Λ) can be defined as **Equation (3-7)**:

$$I_i = \{1, 2, \dots, K\} \setminus \{i\}, \quad (3-7)$$

$$\Lambda_{L_i} = \{j \in I_i : U_{R_j} \leq L_{R_i}\},$$

$$\Lambda_{R_i} = \{j \in I_i : U_{R_i} \leq L_{R_j}\},$$

$$\Lambda_{O_i} = \{j \in I_i : U_{R_j} > L_{R_i} \text{ and } U_{R_i} > L_{R_j}\} = I_i \setminus (\Lambda_{L_i} \cup \Lambda_{R_i})$$

where $\Lambda_{L_i}, \Lambda_{R_i}, \Lambda_{O_i}$ are “left”, “right”, and “overlap” sets, respectively. I_i is the set of all item indices from 1 to K, except for the index i itself for each $i \in \{1, 2, \dots, K\}$, and $j \in I_i$:

- 1) $j \in \Lambda_{L_i}$ if and only if $[L_{R_i}, U_{R_i}] \cap [L_{R_j}, U_{R_j}] = \emptyset$ and $[L_{R_j}, U_{R_j}]$ lies to the left of $[L_{R_i}, U_{R_i}]$;
- 2) $j \in \Lambda_{R_i}$ if and only if $[L_{R_i}, U_{R_i}] \cap [L_{R_j}, U_{R_j}] = \emptyset$ and $[L_{R_j}, U_{R_j}]$ lies to the right of $[L_{R_i}, U_{R_i}]$;
- 3) $j \in \Lambda_{O_i}$ if and only if $[L_{R_i}, U_{R_i}] \cap [L_{R_j}, U_{R_j}] \neq \emptyset$.

Based on the above definitions, these three sets are mutually exclusive, and the number of elements in each set can be denoted by $|\Lambda_{L_i}|$, $|\Lambda_{R_i}|$, and $|\Lambda_{O_i}|$. Therefore, based on the proof of Klein et al. (2020), the 95% CR for the overall ranking of each location (r_i) can be written as **Equation (3-8)**:

$$\{(r_1, \dots, r_i): r_i \in \{|A_{R_i}| + 1, |A_{R_i}| + 2, |A_{R_i}| + 3, \dots, |A_{R_i}| + |A_{O_i}| + 1\} \text{ for } i = 1, \dots, K\} \quad (3-8)$$

3.3 Case Study

The application of the above discussed methodology is demonstrated through a case study on identification of high-risk three-legged stop-controlled intersections in Kitchener, Ontario, Canada. The scope of this study is confined to intersections under the jurisdiction of the City of Kitchener, which primarily consist of unsignalized locations; signalized intersections fall under the purview of the Region of Waterloo and are therefore excluded from this analysis. Within this defined dataset, three-legged stop-controlled intersections were the most common type of intersection in Kitchener, and they had the highest number of crashes during the period of analysis. Hence, the study focused exclusively on these intersections. The following section discusses the details of data sources, risk modelling and results of the analysis.

3.3.1 Data

A motor vehicle crash dataset (i.e., excluding pedestrian-involved crashes) at unsignalized intersections consisting of 1,629 records that occurred from 2014 to 2018 in the city of Kitchener was the primary source of data for this case study. Note that this study only considers crashes that occurred at intersections. Collisions that took place shortly before or after the intersections (i.e., out of the buffer of 20 meters from the intersections) are not included in the analysis. It contains information related to each crash, such as crash date and time, location, driver and vehicle characteristics, severity level, and weather conditions. The unsignalized intersection inventory data was obtained from the City of Kitchener's data platform, from which data on all three-legged minor approach stop-controlled intersections (a total of 2,226) were extracted. In addition, the AADT for the intersecting roads for the period 2008 to 2020 was obtained from the City of Kitchener and the Open Data platform. Based on the analysis conducted by the City of Kitchener in their traffic and safety database management system (i.e., Traffic Engineering Software (TES) platform), an annual growth rate of 2% has been considered for the AADT. Therefore, for those with missing or incomplete AADT, the following approach was followed:

1. If a midblock had AADT for any year between 2008 to 2020, the AADT of the latest year was considered as a base of prediction for the other missing years using a 2% annual growth rate calculated by the City of Kitchener.
2. Otherwise, the estimated data from the Open Data platform were considered as the AADT of 2014 and then using the annual growth, the AADT of other years was predicted.
3. After following step 1 and 2, there were still 52 local street road segments (with two lanes) that no observations and estimation was available in the City of Kitchener data platform. Thus, it was tried to consider a rough estimation based on similar locations with available AADT. Therefore, for simplicity, the average of available estimated AADT from the Open Data platform for the local street with 2 lanes (i.e., AADT of 500) was assumed for those few road segments without any available data.

To obtain the traffic volume on both major and minor approaches at intersections, first, the average AADT of 2014 – 2018 for the crossing streets of each intersection was integrated into the intersection inventory database. Then, to find the AADT of the minor and major streets, the road type of each crossing street was indexed from the midblock database, and the total AADT of the two major streets was considered as the AADT of Major. In the same way, the AADT of the minor directions was found. In some cases, the road type of all the connected segments was the same (e.g., all the crossing streets were major or minor types). In these cases, the sum of the first two highest AADT was considered as the major AADT, and the remaining AADT was considered as the minor. **Table 3-1** presents a brief statistical description of motor vehicle crash data by severity and traffic volumes for all types of unsignalized intersections within the database.

Table 3-1 Descriptive Summary of Motor Vehicle Crashes by Severity and Traffic Volumes

Variable	One/Two-way Stopped Intersections				All-way Stopped Intersections			
	Min.	Max.	Mean	SD	Min.	Max.	Mean	SD
Total Crashes (5yrs.)	0	35	0.53	1.60	0	26	2.95	4.53
Fatal (5 yrs.)	0	1	3.80E-04	0.02	0	0	0	0
Major Injury (5 yrs.)	0	1	4.56E-03	0.07	0	1	0.012	0.11
Minor Injury (5 yrs.)	0	9	0.04	0.29	0	3	0.26	0.63
Minimal Injury (5 yrs.)	0	6	0.06	0.30	0	5	0.38	0.93
PDO (5 yrs.)	0	26	0.42	1.27	0	19	2.30	3.39
AADT _{major} (average 5 yrs.)	203	102647	5274	7358	1284	40000	7721	6443
AADT _{minor} (average 5 yrs.)	10	51324	842.50	1930	229	18303	2554	2482
Total Number of Intersections	2634				81			
Total Number of Intersections w/ Crashes (5 yrs.)	636				51			
Total Number of Intersections by Highest Crash Severity Type (5 yrs.)	Fatal: 1				Fatal: 0			
	Major Injury: 12				Major Injury: 1			
	Minor Injury: 88				Minor Injury: 15			
	Minimal Injury: 101				Minimal Injury: 8			
	PDO: 434				PDO: 27			
Total Crashes (5 yrs.)	1392				237			

3.3.2 Results and Discussions

3.3.2.1 Crash Frequency Model

Due to lack of full information about the various characteristics of the intersections, the traffic-exposure-only functional form of **Equation (3-9)** was used to predict the mean of crashes through a hierarchal full Bayesian NB model.

$$\mu_i = \beta_0 F_{major_i}^{\beta_1} F_{minor_i}^{\beta_2} \quad (3-9)$$

where F_{major_i} and F_{minor_i} are the AADT of the major and minor approaches of i^{th} intersection, respectively and β 's are the coefficients of the prediction model. Although this traffic-exposure-only model could lead to biased results since they ignore the effect of the other variables on the prediction, they are still one of the most commonly used models in road safety analysis as they can be recalibrated easily (Geedipally & Lord, 2008). In addition, since this model is calibrated for only three-legged minor approach stop-controlled intersections, both the number of legs and traffic control type of the intersections are implicitly considered as the other variables with potential impact on the estimated crash frequency.

For model calibration, Heydari et al. (2014) and Miranda-Moreno et al. (2013) showed that using an informative prior distribution for the dispersion parameter can result in a more accurate posterior. Therefore, the NB model was first calibrated using the maximum likelihood method, which resulted in a dispersion parameter of $\hat{\phi} = 0.57$. The first moment of a gamma distribution (i.e., mean) was calculated by dividing the shape parameter by the scale parameter (i.e., a/b). By assuming the non-informative value of 0.01 for the scale hyper-parameter of ϕ (i.e., b), the shape hyper-parameter (i.e., a) was calculated as $\hat{\phi} \times b = 0.0057$. So, the prior distribution for the error term of the hierarchical NB was defined as $\phi \sim Gamma(0.0057, 001)$. Moreover, we considered non-informative priors for the intercept and coefficient parameters as **Equation (3-10)**:

$$\beta_k \sim Normal(0,100); \quad for \ k = 0,1,2 \quad (3-10)$$

Once the prior distributions were defined, the posterior distributions were calculated through MCMC sampling method using PyMC3 package in Python. For each parameter, four different chains consisting of 3000 simulations were considered; the first 2000 samples were used as burn-in iterations. Having tested various MCMC algorithms, such as Metropolis-Hastings, Gibbs sampler, Hamiltonian Monte Carlo, and no U-turn sampler (NUTS), we found that NUTS performed better on our model. As shown in **Figure 3-2**, the MCMC model converged on all parameters, which represents the accuracy of the model.

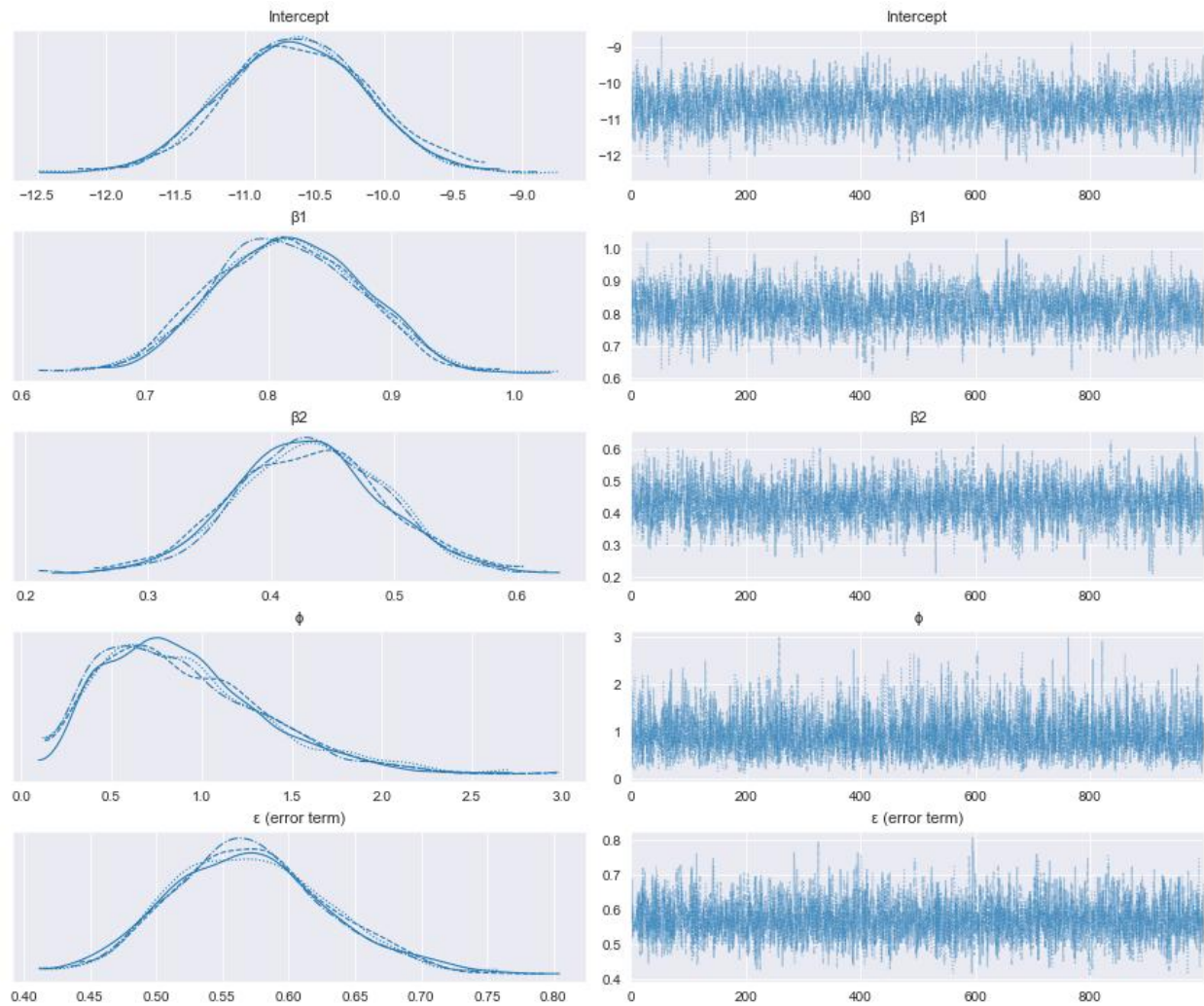


Figure 3-2 Trace Plot of the MCMC Model for each Parameter

As a result, the expected posterior values of all parameters were obtained in the 95% CI. **Table 3-2** shows that none of the regression coefficients contained 0 in their 95% CI; this implies that all parameters were statistically significant in the model. Additionally, the results indicate that the ratio of Monte Carlo standard errors (MCSE_Mean) to SD of the estimates for all parameters was less than 0.05, which confirms the convergence of posterior distributions.

Table 3-2 Summary of Model Calibration Results

Parameters	Mean	SD	95% CI	MCSE_Mean n	MCSE_SD D	MCSE_Mean/SD
Intercept	-10.638	0.503	(-11.641, -9.672)	0.011	0.008	0.022
β_1	0.815	0.059	(0.707, 0.927)	0.001	0.001	0.017
β_2	0.432	0.060	(0.319, 0.558)	0.001	0.001	0.017
ϕ	0.899	0.453	(0.164, 1.775)	0.008	0.005	0.018
ε (error term)	0.571	0.059	(0.462, 0.689)	0.001	0.001	0.017

3.3.2.2 Crash Severity Model

In this study, the crash database contained information about the number of fatalities, major injuries, minor injuries, and minimal injuries, if any, for each crash that occurred at a given intersection. However, due to the low number of fatal motor vehicle crashes (i.e., only one fatality), it was difficult to develop a separate severity model for fatalities. As a result, the crash severities were classified into four levels:

- Fatal&Major: if at least a person died or had a major injury, otherwise
- Minor injury: if at least a person had a minor injury, otherwise
- Minimal injury: if at least a person had a minimal injury, otherwise
- No injury, i.e., PDO

Based on the above classification, each crash has been categorized based on the highest level of severity with which was associated. For example, if a crash results in all fatal, minor, and PDO injuries, this crash is labelled as Fatal&Major injury which is the highest level of severity.

To incorporate site-specific characteristics, the crash dataset was combined with the intersection inventory database. **Table 3-3** presents a summary of the final dataset. Out of 1,629 motor vehicle crashes at unsignalized intersections during 2014 – 2018, there were 14 Fatal&Major crashes (1 fatal and 13 major), 133 Minor injuries, 193 Minimal injuries and 1,289 No injuries.

Table 3-3 Summary of Crash Database Variables

Continuous Variables	Statistic for Continuous Variables				Categorical Variables (number of crashes per 5 yrs.)	
	Min	Max	Mean	SD		
AADT _{minor} (average 5 yrs.)	10	27537	3695	4205.33	Traffic Control	Severity
AADT _{major} (average 5 yrs.)	414	50251	14676	10322.10	All-way Stop: 237	Fatal: 1
No of Legs	3	5	3.553	0.51	Two-way Stop: 1392	Major: 13
						Minor: 133
						Minimal: 193
						PDO: 1289

A multinomial logit model was calibrated based on crashes during the five years at all unsignalized intersections with any number of legs to consider more sample in calibration process. Then, this calibrated model was used to estimate the probability associated with each collision severity at three-legged one-way stop-controlled intersections. the model. **Table 3-4** illustrates the model coefficients associated with each variable considered in the model. The model had three statistically significant variables: total traffic volume, number of legs, and traffic control type. Both all-way and one-way stop-controlled (minor approach stop-controlled in three-legged intersections) types were positively correlated with the probability of fatal, major, and minor injury levels.

Table 3-4 Summary of Severity Model Calibration Results

Severity Type	Variables	Coefficients	std. error	Z Value	P Value
PDO	(Base Type)	0			
Fatal and Major	Intercept	-7.560	3.62	-2.09	0.04
	Ln (Total AADT)	-0.399	0.32	-1.23	0.22
	Number of legs	1.430	0.54	2.63	0.01
	Traffic Control: Two-way stop control	1.717	1.10	1.56	0.01
	Traffic Control: All-way stop control	0.395	1.43	0.28	0.07
Minor	Intercept	-4.958	1.20	-4.14	3.48E-05
	Ln (Total AADT)	0.189	0.11	1.74	0.08
	Number of legs	0.228	0.16	1.44	0.15
	Traffic Control: Two-way stop control	0.167	0.21	0.78	0.04
	Traffic Control: All-way stop Control	0.031	0.28	0.11	0.09
Minimal	Intercept	-3.894	0.99	-3.94	8.09E-05

Ln (Total AADT)	0.137	0.09	1.53	0.13
Number of legs	0.258	0.13	1.96	0.05
Traffic Control: Two-way stop control	-0.147	0.17	-0.85	0.04
Traffic Control: All-way stop control	-0.254	0.23	-1.10	0.02
Log-Likelihood	-1537.5			
McFadden R2	0.010548			
Likelihood ratio test: chisq	32.781 (p.value = 0.0010472)			

3.3.2.3 Confidence Region for the Ranking of the Intersections based on Total Risk

Each severity type was converted to a dollar value by assuming a nominal cost. de Leur and De (2018) calculated the equivalent monetary of the average social cost of crashes by severity in 2018 as follows:

- Fatal Crash: \$2,389,953
- Major Injury Crash: \$376,771
- Minor Injury Crash: \$50,546
- Minimal Injury Crash: \$32,041
- PDO: \$14,391

Note that since the minimal injury cost has not been calculated in this study, we used the social cost defined in Vodden et al. (2007) for this severity level. In this regard, we first calculated the growth rate of PDO cost between these two studies (i.e., 1.79) and then by multiplying the cost of minimal injury level in Vodden et al. (2007), i.e., \$17,900, by the calculated growth rate, we estimated the cost of minimal injury crashes.

By multiplying the lower and upper limits of the expected annual number of crashes (i.e., $L_{N_{ij}}$ and $U_{N_{ij}}$) by the monetized probability of each severity type, the 95% CI for the risk value of each intersection was calculated. Finally, the 95% CR for each intersection was obtained by using **Equation (3-8)**. The result of the confidence region for the top 30 ranked crash hotspots based on the point estimate of predicted risk (i.e., mean) is presented in **Table 3-5**.

Table 3-5 Joint CR for Ranking based on 95% CI

Mean Rank	ID	Mean of Risk	SD of Risk	95% CI of Risk	95% CR for Rank
		(thousand dollars)			
1	2743	146.64	38.49	(81.15, 224.7)	{1, 2}
2	2872	101.68	24.14	(59.59, 149.39)	{1, ..., 4}
3	1659	55.77	11.80	(35.14, 79.7)	{2, ..., 5}
4	2023	45.90	9.30	(28.68, 63.98)	{2, ..., 7}
5	1918	32.90	5.10	(23.75, 43.41)	{3, ..., 9}
6	1891	23.12	3.62	(16.51, 30.33)	{4, ..., 25}
7	2329	22.21	3.60	(15.76, 29.53)	{4, ..., 25}
8	1046	20.32	2.70	(15.27, 25.7)	{5, ..., 29}
9	2817	18.93	2.42	(14.6, 23.98)	{5, ..., 33}
10	3205	17.86	2.29	(13.62, 22.47)	{6, ..., 43}
11	1951	17.85	2.58	(13.29, 23.12)	{6, ..., 46}
12	3192	16.64	2.55	(12.15, 21.88)	{6, ..., 53}
13	1489	16.43	2.00	(12.46, 20.15)	{6, ..., 52}
14	2007	16.41	1.98	(12.6, 20.17)	{6, ..., 51}
15	2138	16.20	2.17	(12.19, 20.48)	{6, ..., 53}
16	2426	15.70	1.87	(12.19, 19.43)	{6, ..., 53}
17	3122	15.11	1.75	(11.74, 18.49)	{6, ..., 56}
18	2923	14.73	1.81	(11.27, 18.24)	{6, ..., 63}
19	2422	14.47	2.18	(10.63, 18.95)	{6, ..., 75}
20	2981	14.45	1.67	(11.25, 17.75)	{6, ..., 63}
21	2735	14.36	2.17	(10.54, 18.8)	{6, ..., 77}
22	1549	14.16	1.83	(10.65, 17.64)	{6, ..., 75}
23	1422	13.77	1.67	(10.73, 17.18)	{6, ..., 75}
24	2608	13.55	1.94	(10.08, 17.49)	{6, ..., 81}
25	1054	13.22	1.79	(10.02, 16.88)	{6, ..., 82}
26	1295	12.65	1.54	(9.77, 15.68)	{8, ..., 85}
27	3073	12.52	1.39	(9.94, 15.32)	{8, ..., 82}
28	1365	12.19	1.60	(9.26, 15.4)	{8, ..., 89}
29	1221	12.10	1.52	(9.19, 15.08)	{9, ..., 89}
30	2983	12.02	1.43	(9.35, 14.81)	{9, ..., 88}

As an example, the results shown in **Table 3-5** imply that with the probability of 95%, the estimated annual risk for intersection #2872, will be in the range of \$59,590 to \$149,390. Moreover, it can be concluded that considering the mean of the estimated risk, this intersection will be selected as the second riskiest location since its expected annual total risk is \$101,680, which is \$45,910 more than the third-ranked intersection (i.e., #1659). However, if the uncertainty in the expected total risk is taken into consideration, it could be ranked as the first, second, third, or even fourth riskiest location.

This approach provides a crucial understanding that more locations may be risky than a simple point estimate suggests. When moving to the next step of the road safety management process, such as detailed site evaluations, starting from the broader shortlist identified by the proposed method, the practitioners would have the opportunity to incorporate additional information that is not considered in the crash prediction model or their own knowledge of the sites in the process of finalizing the shortlist. It allows them to consider locations that may not seem risky based on their mean value but whose confidence interval shows they could be just as hazardous as higher-ranked sites. This prevents potentially dangerous locations from being overlooked simply because their average risk score fell below an arbitrary cutoff.

In addition, if the goal is to identify a list of hotspots based on a certain risk threshold value, the CI of the estimated risk at the individual locations would provide more complete information for decision-making. In this case, CI provides some underlying information about the safety of the location, which would not be captured if only the mean of the total risk is taken into account. As an example, suppose that the decision-makers wish to identify the hotspot locations with risk values exceeding \$20,000 annually. If only the mean estimated risk value is used as in the conventional approach, eight intersections will be identified as hotspots. As shown in **Figure 3-3**, however, when the 95% CI is taken into account, 15 intersections will be categorized as hotspots.

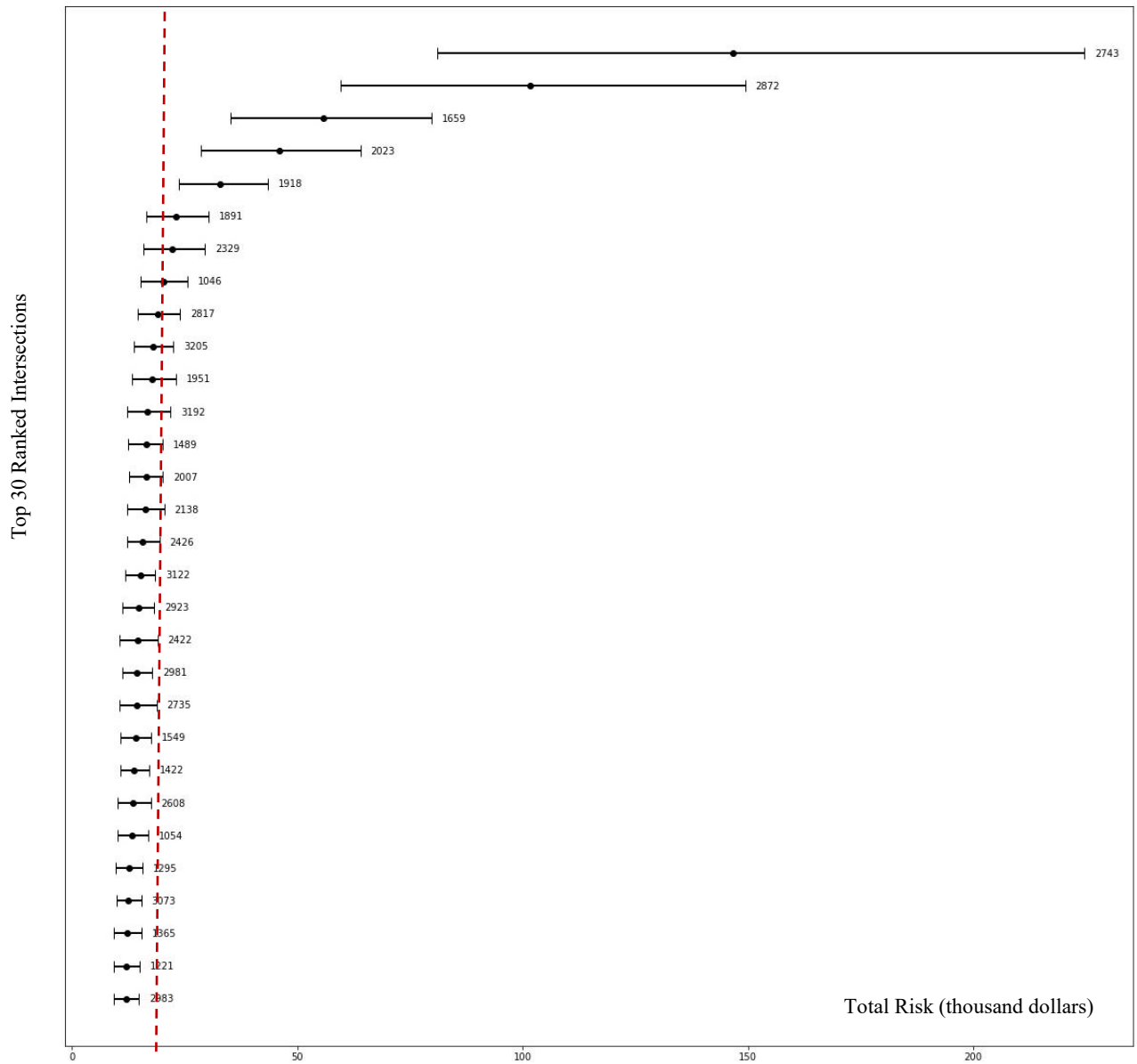


Figure 3-3 Estimated Total Risk with 95% CI

3.4 Conclusion

This study employed a frequentist approach to finding a joint CR for ranking locations of interest by risk and identifying crash hotspots. A case study on ranking three-legged stop-controlled intersections in Kitchener, Ontario, Canada, was conducted to show the benefits of the proposed approach. The crash frequency and severity of a given crash were estimated using a hierarchical full Bayesian NB Model and a multinomial Logit model respectively. For each

location, the CI of risk is obtained on the basis of the estimated 95% CI of the crash frequency and the predicted probability of each severity level. Based on the CR, individual intersections can be ranked accordingly with the additional consideration of uncertainty. Specifically, when uncertainty was considered, an intersection could have varied ranking positions with different probabilities since the true value of its estimated crash risk follows a distribution. As a result, consideration of uncertainty in the road safety analysis may help identify more accurate hotspots for applying safety action items.

To our knowledge, there have been few studies on integrating uncertainty into the network screening process; therefore, this study has provided some insightful information for future researchers and practitioners who wish to incorporate uncertainty into the analysis. In this regard, the following research direction can be suggested for future studies:

- This study developed a Bayesian model based on AADT for intersection crash predictions. However, additional predictors could be included to enhance accuracy. In addition, future studies may consider developing a Bayesian model to account for uncertainty in crash severity assessment and incorporating uncertainty introduced by the predictors.
- This study used a simple method to fill in missing AADT data of road segments. While the estimated AADT had an acceptable level of accuracy, future studies could consider a better method, such as estimating traffic volumes based on available data in other crossed road segments, for more accurate prediction models.
- In addition to the uncertainty associated with the modelling part, there are various sources of uncertainty resulting from the data. A wide range of variations in the estimated traffic volume (i.e., AADT) is one of these sources which can affect the result of risk estimations. However, for simplicity, the impact of traffic volume uncertainty has not been investigated in this study and can be considered in future research.
- In this study, we employed a frequentist model to determine a joint CR for the rankings of the locations. However, this CR can also be determined using the Bayesian method.
- In addition, conditional probability theory can also be employed to calculate the probability of getting each rank within the defined CR.

- The application of the proposed methodology was demonstrated by applying on three-legged minor approach stop-controlled intersections to identify the hotspot with consideration of uncertainty. However, this method can be applied to all other types of roadway entities which can be considered in future studies to compare the result of the analysis.

Chapter 3 demonstrated the significant impact of uncertainty on traditional, crash-based road safety analysis by proposing a robust method for hotspot ranking that accounts for the statistical variability in risk estimates. While this approach enhances the reliability of reactive safety programs, the literature review also points to proactive, conflict-based analysis as a promising alternative. This proactive paradigm allows for safety assessment without relying on sparse and often under-reported crash data.

However, the practical application of this approach is hampered by a fundamental question of data uncertainty: how much conflict data is actually required to produce reliable collision predictions? The prevalence of studies using very short-term observations has fostered a "myth of quick analysis," often resulting in predictions with unacceptably wide CI that limit their practical use. Before refining the analytical models themselves, it is crucial to first understand the foundational data requirements for any conflict-based study to be meaningful.

Therefore, the following chapter addresses this critical data uncertainty by systematically quantifying the relationship between the duration of data collection and the reliability of collision estimates, using established conflict analysis techniques.

Chapter 4

Conflict-Based Road Safety Analysis: Limits of Short-Term Conflict Data in Collision Risk Prediction

As established in the preceding chapters, proactive, conflict-based safety analysis using EVT presents a powerful alternative to traditional, reactive methods that are often hindered by collision data scarcity and reporting inconsistencies. By extrapolating from more frequent traffic conflicts to rare collisions, EVT provides a robust statistical method for proactive safety assessment. However, despite its theoretical promise, the practical implementation of EVT faces a fundamental methodological challenge: determining the optimal amount of data needed to achieve reliable collision frequency predictions.

It is crucial to distinguish between accuracy and reliability in this context. While accuracy focuses on how closely predictions match historical collision data, reliability concerns the precision of these predictions, specifically, the width of the CI surrounding the estimates. Most of the existing conflict-based studies relied on data collected over extremely short periods, typically only a few hours or days. This body of literature has fostered the belief among practitioners that robust safety assessments can be performed quickly with a minimum data collection effort. In reality, because severe conflicts remain rare events, collision predictions based on data from these limited observation durations could result in unacceptably wide CI with little practical value.

This chapter directly addresses the problem of the relationship between sample size and the reliability of collision predictions derived from conflict data. It seeks to provide an evidence-based answer to a critical question for practitioners: how much conflict data is required to move from highly uncertain estimates to statistically robust and meaningful collision frequency predictions?

This chapter is based on the manuscript: Aminghafouri, R., & Fu, L. (2025). The Myth of Quick Conflict-Based Road Safety Analysis: Limits of Short-Term Conflict Data in Collision Risk Prediction. *Accident Analysis & Prevention*, (Under Review).

Using a unique, year-long LiDAR dataset collected continuously at four unsignalized intersections, this chapter evaluates how collision estimates, and their 95% CI converge as the sample size increases from two days to one year. The analysis employs a POT approach within a Bayesian framework to provide a methodology for determining the necessary data collection duration for achieving a desired level of precision in collision frequency estimation. By identifying the optimal balance between data collection effort and prediction reliability, this research aims to enhance the practical utility of EVT-based conflict analysis for more efficient and effective crash risk estimation.

4.1 Literature Review

Traffic conflict techniques using EVT have emerged as a promising approach for proactive safety assessment. However, a fundamental methodological question persists: how much conflict data is necessary to achieve reliable collision predictions? This review examines the current state of knowledge regarding sample size requirements for EVT-based conflict analysis, highlighting the critical research gap that this study addresses.

The application of EVT to traffic safety analysis has evolved significantly since its introduction by Songchitruksa and Tarko (2006), who established the theoretical method for extrapolating from traffic conflicts to collision probability. Two primary approaches have been established: the BM approach, which models the distribution of maximum values within fixed time intervals using the GEV distribution, and the POT approach, which models all observations exceeding a predetermined threshold using the GP distribution (Stuart & Coles, 2001).

Songchitruksa and Tarko (2006), highlight the challenges associated with the short data collection periods often employed in traffic conflict studies. Their findings suggest that an observation period of at least 3 to 6 weeks is necessary to obtain safety estimates that are statistically reliable and comparable to those derived from traditional crash data analysis. Using simulations, they found that a minimum of 30 to 50 days of observation is required to achieve CI around collision predictions that parallel those obtained from 4 years of crash history.

A comprehensive review of contemporary literature reveals profound inconsistencies in sampling methodologies. Most existing studies rely on extremely limited observation periods, ranging from mere hours (Zheng et al., 2018; Wang et al., 2019; Zheng et al., 2019; Wen et

al., 2024) to a maximum of 24 hours (Arun, Haque, et al., 2021a; Arun, Haque, Bhaskar, et al., 2022; Arun, Haque, Washington, et al., 2022; Hussain et al., 2022). This methodological variability introduces substantial uncertainty in collision risk estimations, with studies based on smaller conflict samples consistently reporting exceptionally wide CI.

The challenges of limited sampling are starkly illustrated across multiple recent studies. Wen et al. (2024) investigated urban expressway diversion areas using just 2 hours of observation, generating 1,331 traffic conflicts. Despite the seemingly robust sample, their analysis produced 95% CI for crash frequency ranging from 0 to 273, demonstrating significant variability inherent in limited sampling approaches. In a similar approach, Yue et al. (2024) examined signalized intersections over 11 hours, revealing crash frequency CI spanning 0 to 59 across different thresholds. Pedestrian crash risk research further illuminates these methodological limitations. Hussain et al. (2024a) employed the BM approach and produced 95% CI ranging from 0 to 120. These findings are not exceptions but represent a systematic issue pervading current EVT-based traffic safety research.

Although some researchers have attempted to address sampling limitation through developing advanced modeling approaches, they have not fully resolved these methodological challenges. Fu et al. (2020) developed a high-dimensional multivariate Bayesian hierarchical extreme value model for traffic crash estimation by incorporating multiple traffic conflict indicators such as MTTC, PET, and DRAC. The study collected traffic conflict data over 24 hours of observation, identifying 7,998 traffic conflicts across signalized intersections in Surrey, British Columbia. Their analysis revealed significant uncertainty in crash predictions, with CI ranging from 0 to 661 for a single intersection, as an example. In a subsequent study, Fu and Sayed (2021b) further explored advanced modeling techniques by developing a multivariate Bayesian hierarchical Gaussian copula model. Analyzing traffic conflicts from signalized intersections in Surrey, British Columbia, over 24 hours, they identified 7,998 traffic conflicts using similar conflict indicators. Despite the sophisticated approach, the study still encountered substantial estimation uncertainties, with CI spanning from 0 to approximately 1,595. This continued variability underscored the persistent challenges in accurately predicting traffic crash risks through advanced statistical methodologies.

These studies are not isolated instances but representative of a broader methodological challenge. The systematic uncertainty in crash risk estimation appears pervasive across recent EVT-based traffic safety research. **Table 4-1** provides a comprehensive summary of studies from 2021 onwards, highlighting the sample sizes used and the resulting prediction uncertainties in terms of the 95% CI.

Table 4-1 Summary of sample sizes and prediction uncertainties from recent EVT studies (2021 onwards)

Study	Video Data Collection Duration	Total Conflict Sample Size	Conflict Indicators	Model Used	Annual Estimated Crashes	95% CI
Arun, Haque, et al. (2021a)	24 hours (over 2 days)	3,829 rear-end conflicts	TTC, MTTC, and Delta-V	Bivariate POT EVT Model with Logistic dependence structure	3.515	[0.17, 11.07]
Fu and Sayed (2021a)	4 hours	535 events with DRAC > 0 m/s ²	DRAC	Multisite Bayesian Hierarchical Extreme Value Model	5.6	[0.0, 479.6]
Fu et al. (2021)	88 hours (over 2 days)	1,515 conflicts (1,108 rear-end + 407 side-impact)	TTC	Multi-type Bayesian hierarchical EVT model	3.6	[0.1, 219.1]
Fu and Sayed (2021b)	6 hours	1,384 PET, 1,304 DRAC	PET, DRAC	Bivariate Bayesian hierarchical Gaussian copula	6.9	[0.0, 1595.3]
Fu and Sayed (2021c)	6 hours	2,065 conflicts	TTC, MTTC, PET, DRAC	Random Parameters Bayesian Hierarchical Univariate EVT Model	6.8	[0.0, 471.1]
Gastaldi et al. (2021)	230 maneuvers performed by 23 drivers in a simulator	223 valid maneuvers	PET and TTC	Bivariate EVT: Component-wise Maxima and Excesses Over Threshold	8.79	[4.20; 13.52]
Guo et al. (2021)	36 hours	Not explicitly stated	TTC	GP distribution EVT model	1.14	[0, 29.07]

Zheng, Huang, et al. (2021)	~ 3 hours per segment	1,588 lane change conflicts	PET	Bayesian Hierarchical EVT Model (BHEVM) with site-level factors	60.8	[0, 2,447.1]
Alozi and Hussein (2022)	5.5 hours AV ¹ data	726 AV-pedestrian conflicts	TTC and PET	EVT with POT method	3.735 collisions per million VKT ⁷	[0.049, 14.558]
Hussain et al. (2022)	48 hours	3,829 conflicts	TTC, MTTC, DRAC	Hybrid method (machine learning + EVT): Univariate/Bivariate GEV/GP models	425.9 crashes (MTTC-based univariate GEV model)	~200–600 crashes (BM model)
Arun, Haque, Bhaskar, et al. (2022)	24 hours	2,839 rear-end conflicts	MTTC, Delta-V	Bivariate Logistic POT EVT	1.203	[0, 146]
Arun, Haque, Washington, et al. (2022)	48 hours	12,471 rear-end conflicts	MTTC, DRAC, PSD ² , Delta-V	Multivariate EVT with Gumbel-Hougaard Copulas	6.160	[0.66, 21.76]
Fu and Sayed (2022b)	6 hours	7,998 rear-end conflicts	MTTC, PET, DRAC	Husler-Reiss multivariate EVT model	10.0	N/A
Fu and Sayed (2022a)	6 hours	179 cycles	MTTC	Bayesian Dynamic EVT Model	7.4	[3.6, 9.5]
Fu and Sayed (2022c)	6 hours	7,998 rear-end conflicts	MTTC, TTC, PET, DRAC	Random-Parameter Bayesian Hierarchical EVT Model	7.0	[0.0, 471.1]
Fu and Sayed (2023a)	6 hours	1,074 conflicts	MTTC	Third-order dynamic linear Bayesian hierarchical POT model	7.7	[0.0, 285.7]
Ali, Washington, et al. (2023)	96 hours	2,915 rear-end conflicts	MTTC	Bayesian GEV model	Stationary model: 6 Non-stationary model: 9	Stationary: [0, 42.51], Non-stationary: [0, 20.08]
Hussain, Ali, Li and Boyle (2023)	4 days	632 signal cycles	MTTC	Non-stationary GEV model + ARIMAX	5.99	[0, 19.01]

Kar et al. (2023)	2 hours	1,113 sideswipe conflicts	ACT ³	Non-stationary GP distribution model	2	[0, 28]
Kumar and Mudgal (2023)	3.5 hours	214 conflicts	TTC and Lateral Gap	Bivariate EVT Model (Logistic distribution)	0.57	[0.55, 0.57]
Nassereddine et al. (2023)	9 hours	109 PET, 119 ORTT ⁴	PET, ORTT	Bivariate Bayesian EVT model	0.73	[0, 15.81]
M. M. Howlader, Y. Ali, et al. (2024)	654 hours	7,012 conflicts	PET and Delta-V	Non-stationary bivariate GEV models	0.839	[0.0, 8.1]
M. M. Howlader, A. Bhaskar, et al. (2024)	144 hours	PET: 612; T _{2min} : 415; GT ⁵ : 301; Delta-V: 658	PET, GT, T _{2min} , Delta-V	PET-GT bivariate EVT model	2.67	[0.0, 139.59]
M. Howlader et al. (2024)	45.33 hours	1,926 conflicts	PET for right-turn conflicts, MTTC for rear-end conflicts	Bayesian non-stationary bivariate GEV model	1.773	[0.0, 9.836]
Hussain et al. (2024a)	48 hours	1,217 pedestrian interactions combined from 3 intersections	PET	Hybrid Bayesian Models (AI & GEV/GP components)	5.88	[0, 23.42]
Kumar and Mudgal (2024)	4.5 hours	110 exceedances conflicts	TTC, Lateral Gap	GP distribution Model	445	[360, 532]
Niu et al. (2024)	24 hours	884 MTTC, 1,320 PET	MTTC, PET	Bivariate POT EVT Model	4.364	[0.0, 21.02]
Wen et al. (2024)	2 hours	382 conflicts	TTC, CTTC ⁶	Bayesian Hierarchical EVT	1.078	[0.0, 177.39]
Yue et al. (2024)	4 hours	601 conflicts	MTTC	Bayesian Hybrid Gamma-GP distribution Model	4.2	[0.7, 12.0]

1) Autonomous Vehicle; 2) proportion of Stopping distance; 3) Anticipated Collision Time; 4) Obstructed Right Turn Time; 5) Gap Time; 6) Cross Time-to-Collision; 7) Vehicle Kilometers Traveled

As shown in the table, even recent studies reveal significant differences in data collection methods and reliability of collision predictions, highlighting the need for standard guidelines on sample size. This uncertainty demonstrates the limitations of short observation periods and emphasizes that longer data collection times are necessary to improve the accuracy of pedestrian crash predictions.

Conversely, studies with larger sample sizes generally achieve narrower CI. M. M. Howlader, A. Bhaskar, et al. (2024) developed a bivariate, non-stationary extreme value model for estimating opposing-through crash frequencies by collecting traffic conflict data over 144 hours across multiple days. They achieved 95% CI for annual collisions ranging from 0 to 11.24, representing a more reliable prediction. However, this improvement can come at the cost of extended data collection periods and corresponding resource investments. The fundamental question of whether this additional expenditure of resources is justified by the marginal improvement in prediction reliability remains unanswered.

The relationship between sample size and prediction reliability appears to differ between EVT modeling approaches. Multiple studies suggest that POT models exhibit more stable performance with smaller samples compared to BM models (Zheng et al., 2014; Zheng & Sayed, 2019c). This difference likely stems from the POT approach's more efficient use of extreme data, as it considers all observations exceeding a threshold rather than just BM. However, a systematic comparison of sample size requirements between these approaches under comparable conditions has yet to be conducted.

A limited number of studies have explicitly examined the relationship between sample size and prediction reliability in traffic conflict analysis. Tarko (2012) conducted a simulation study indicating that approximately 2,500 risk events were needed for stable parameter estimates in GP distribution models. More recently, Fu and Sayed (2023c) systematically investigated sample size requirements for rear-end conflicts using the BM approach. Through bootstrapping analysis, they determined that 90 and 42 traffic conflicts were adequate for stationary and non-stationary models, respectively. The lower requirement for non-stationary models was attributed to the inclusion of informative covariates that reduced unexplained variability. Similarly, Nazir et al. (2024) examined sample size effects on both BM and POT models for pedestrian conflicts. They collected the data for 144 hours from three different intersections in

Brisbane, Australia, which resulted in 1,537 conflicts in total. Their findings show that POT models could achieve acceptable reliability with approximately 30% of the total collected conflicts, while BM models required at least 60% conflicts distributed across a minimum of 831 signal cycles.

EVT modeling depends critically on sample size, with both GEV and GP distribution models requiring comprehensive datasets for reliable estimation. The mathematical foundations of these models, rooted in asymptotic theory, demonstrate that distribution characteristics, particularly tail heaviness, significantly influence data collection requirements. More complex models incorporating temporal variations demand larger samples to ensure accurate parameter estimation (Stuart & Coles, 2001). Practical challenges in conflict data collection extend beyond theoretical considerations. Safety practitioners face a fundamental dilemma: prolonged data gathering delays potential interventions, while abbreviated collection risks unreliable predictions. The resource constraints are substantial, with sensor deployment involving considerable financial investment, technical expertise, and operational limitations.

The optimization of safety assessment resources represents a critical research concern. Collecting excessive data proves as problematic as gathering insufficient information. The potential to achieve reliable predictions through more targeted data collection could substantially accelerate safety interventions and expand assessment capabilities across multiple locations. Current literature reveals significant gaps in understanding the relationship between sample size and prediction reliability. Existing studies have provided isolated insights into minimum sample requirements but have not systematically explored how prediction accuracy improves with incremental data increases. This knowledge deficit leaves practitioners without clear guidance, forcing them to navigate between the risks of under-sampling and the inefficiencies of over-collection.

This study directly addresses these challenges by systematically quantifying the relationship between sample size and prediction reliability in EVT models. By rigorously evaluating how data collection periods impact prediction CI, the study aims to provide practical, actionable guidance for traffic safety practitioners. The proposed methodology establishes an approach for analyzing prediction reliability across diverse traffic environments, promising a more efficient approach to proactive safety assessment.

4.2 Methodology

4.2.1 Study Framework

This study employs a systematic approach to evaluate the effects of sample size on the reliability of collision predictions derived from traffic conflicts using EVT. The framework consists of five main components; first, conflict data is collected continuously using LiDAR sensors installed at four intersections in Kitchener, Canada (**Figure 4-1**). The sensors capture vehicle trajectories, which are processed to identify traffic conflicts characterized by PET values. Second, different sampling periods are constructed from the collected data, ranging from 2 days to 365 days, to systematically investigate the effect of sample size on prediction reliability. Third, for each sampling period, appropriate thresholds are determined using MRL plots and parameter stability plots to identify extreme conflict events following the POT approach. Fourth, Bayesian stationary EVT models are developed to estimate collision risks from the identified extreme events. Finally, the effect of sample size on prediction reliability is assessed by analyzing how the mean collision predictions and their associated CI evolve with increasing sample size.

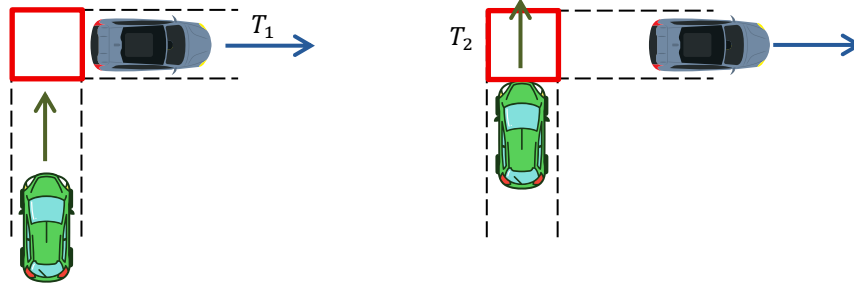
This methodology allows for a comprehensive examination of the relationship between data collection duration and prediction reliability, ultimately enabling the identification of the minimum adequate data collection period required for reliable collision prediction.



Figure 4-1 Selected intersections and collected LiDAR data

4.2.2 Traffic Conflict Indicator

This study utilizes PET as the primary SSM for identifying and quantifying traffic conflicts. PET is defined as the time difference between the moment when the first road user leaves a potential conflict point and the moment when the second road user arrives at the same point (Allen et al., 1978). The conceptualization of PET is illustrated in **Figure 4-2**, which visualizes the temporal dynamics of potential traffic conflicts.



T_1 = the time when the first vehicle leaves the conflict area

T_2 = the time when the second vehicle arrives the conflict area

Figure 4-2 Illustration of PET calculation

The PET calculation is represented by **Equation (4-1)**, which captures the absolute time difference between the maximum departure time (T_1) and minimum arrival time (T_2).

$$PET = ||\max\{T_1\} - \min\{T_2\}|| \quad (4-1)$$

PET's significance in safety analysis lies in its ability to quantify collision risk through temporal proximity. Lower PET values indicate a higher potential for collision, with values approaching zero representing critical near-miss scenarios. For this study, a threshold of 4 seconds was established to define conflicts, consistent with previous research suggesting that interactions with PET values below this threshold have significant safety implications (Zheng & Sayed, 2019c).

The calculation of PET in this study is performed automatically by the BlueCity safety platform using trajectory data captured by LiDAR sensors. The platform identifies potential conflict points by analyzing the trajectories of all road users and calculates PET values for each interaction. This automated approach ensures consistency in conflict identification and measurement across all study locations and time periods.

4.2.3 Extreme Value Theory Models

4.2.3.1 Peak Over Threshold Approach

This study employs the POT approach, which models all observations exceeding a predetermined threshold as extreme events using the GP distribution. The POT approach is particularly appropriate for this application as it makes more efficient use of available conflict data compared to the BM approach, especially when dealing with varying sample sizes (Zheng et al., 2014; Zheng & Sayed, 2019c).

Let $z_1, z_2, \dots, z_n$ be a set of independent and identically distributed random observations representing negated PET values from traffic conflicts. In this study, to align with the EVT convention where extreme values are modeled as maxima rather than minima, we transform the PET values by negation (i.e., using $z_i = -PET_i$). This transformation ensures that more severe conflicts (smaller PET values) correspond to larger values of z_i , consistent with the EVT framework.

Defining a threshold u , observations exceeding this threshold ($z_i > u$) are considered extreme events. The distribution of threshold exceedances $y = z_i - u$, conditioned on $z_i > u$, follows a GP distribution as **Equation (4-2)**:

$$G(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma_u}\right), & \xi = 0 \end{cases} \quad (4-2)$$

where $\sigma_u > 0$ is the scale parameter at threshold u and $-\infty < \xi < \infty$ is the shape parameter. The shape parameter determines the tail behavior of the distribution: $\xi > 0$ indicates a heavy-tailed distribution, $\xi = 0$ corresponds to an exponential distribution, and $\xi < 0$ represents a distribution with a finite upper bound.

The probability of a collision (crash risk) at a specific location can be calculated from the fitted GP distribution as **Equation (4-3)**:

$$RC = Pr(z_i \geq 0) = 1 - G(0) = \left(1 + \frac{\xi u}{\sigma_u}\right)^{-\frac{1}{\xi}} \quad (4-3)$$

4.2.3.2 Threshold Selection

The selection of an appropriate threshold is critical in the POT approach. Too low a threshold violates the asymptotic basis of the model, while too high a threshold reduces the number of exceedances, increasing estimation variance. This study employs two complementary graphical methods for threshold selection:

1. MRL Plot: The MRL plot displays the mean of exceedances over a threshold against the threshold value. A threshold u is appropriate if the mean excess function $e(u)$ is approximately linear beyond that point. Mathematically, the mean excess function is defined as **Equation (4-4)**:

$$e(u) = E(x - u | x > u) = \frac{\sigma_u + \xi u}{1 - \xi}, \xi < 1 \quad (4-4)$$

where u , ξ , and σ_u denote the threshold, the shape parameter, and the scale parameter of the fitted GP distribution, respectively.

Therefore, if the exceedances follow a GP distribution, the MRL plot should be approximately linear beyond the appropriate threshold.

2. Parameter Stability Plot: A threshold stability plot is employed as an additional graphical tool to identify a suitable threshold for modeling extreme events with the GP distribution. Through this plot, the estimates of the shape and scale parameters are examined across various thresholds or numbers of exceedances in search of a stable region. The principle underlying this approach is that if the GP distribution is appropriate for modeling values exceeding a threshold u_0 , it should also be suitable for values exceeding any threshold $u > u_0$, and the shape parameter should remain consistent. In this context, the dataset for the plot is obtained using **Equation (4-5)**:

$$\sigma_u = \sigma_{u_0} + \xi(u - u_0) \quad (4-5)$$

For each sample period in this study, both methods are applied to identify an appropriate threshold range, and the lowest value within this range is selected as the final threshold to maximize the number of extreme events while maintaining the validity of the GP distribution model.

4.2.3.3 Stationary Bayesian Model Specification

This study employs a stationary Bayesian GP distribution model, which assumes that the distribution parameters remain constant over time. This approach is appropriate for this analysis as our primary focus is on the effect of sample size rather than temporal and spatial variations in conflict patterns. The stationary model is specified as **Equation (4-6)**:

$$y_j \sim GPD(\sigma_j, \xi_j) \quad (4-6)$$

where y_j is the sample over threshold at site j . The parameters σ_j and ξ_j are the scale and shape parameters of the GP distribution for site j , respectively.

4.2.3.4 Bayesian Estimation

This study employs Bayesian inference to estimate model parameters, which is particularly advantageous for small samples and when incorporating parameter uncertainty in subsequent crash risk calculations. The posterior distribution of model parameters is expressed as **Equation (4-7)**:

$$p(\theta | y) \propto p(y | \theta)p(\theta) \quad (4-7)$$

where θ represents the model parameters, $p(y | \theta)$ is the likelihood function, and $p(\theta)$ is the prior distribution of parameters.

Because information on the variability of GP distribution parameters is limited, uninformative priors are assigned to the scale parameter. In this regard, a normal distribution with zero mean

and large variance ($\log(\sigma_j) \sim N(0, 10^6)$) is adopted for the scale parameter. It has been reported in the literature that improperly defined priors for the shape parameter in extreme value distributions can result in convergence issues (Zheng & Sayed, 2020a). To address this, prior information from previous studies (Yue et al., 2024) is utilized, wherein the shape parameter is reported to range between -1.0 and 1.0. Therefore, a uniform distribution is specified for the shape parameter in POT models ($\xi_j \sim Unif(-1.0, 1.0)$).

To obtain the posterior distributions of the model parameters, MCMC simulation is employed. Three parallel chains with different initial values are run for 20,000 iterations each, with the first 2,000 iterations discarded as burn-in. Convergence is assessed using the Gelman-Rubin statistic (with values below 1.1 indicating convergence) and visual inspection of trace plots. Also, the Deviance Information Criterion (DIC) is employed to evaluate the goodness of fit for Bayesian models (Spiegelhalter et al., 2002).

4.2.3.5 Collision Frequency and Confidence Intervals Estimation

The POT model gives the probability that an extreme conflict (i.e., a PET exceeding the threshold u) will culminate in a collision, RC , already defined in **Equation (4-3)**. To turn this probability into an expected number of collisions, the model must also consider how many exceedances actually occur within the prediction horizon.

Let n_u be the number of exceedances recorded during the conflict observation period (i.e., T); the corresponding exceedance rate is $\lambda_u = \frac{n_u}{T}$. Over a forecast period of $\tilde{T}$ (365 days for annual estimates in this study), we therefore expect $\lambda_u \times \tilde{T}$ exceedances. Therefore, assuming stationarity, the estimated collision frequency over a period $\tilde{T}$ is calculated as **Equation (4-8)**:

$$N = \lambda_u \times \tilde{T} \times RC = n_u \times \frac{\tilde{T}}{T} \times RC \quad (4-8)$$

To account for collision estimation uncertainty, the posterior distributions of the GP distribution parameters are used to generate a distribution of collision frequency estimates. The mean of posterior parameter distributions is used to estimate the expected collision frequency,

while the 2.5th and 97.5th percentiles of the posterior parameter distributions are used to calculate the 95% CI of the collision estimates.

In addition, the CI for observed collisions are calculated using a 95% Poisson CI (Songchitruksa & Tarko, 2006), represented as $[\lambda: \frac{1}{2N} \chi^2_{y0,0.975} \leq \lambda \leq \frac{1}{2N} \chi^2_{2(y0+1),0.025}]$, where λ is the true mean annual collisions and N represents number of years.

4.2.4 Data Collection and Processing

This study examines four unsignalized four-leg intersections in the City of Kitchener, Canada. Three of the sites operate under two-way stop control and one under all-way stop control, giving a realistic cross-section of the control strategies most frequently encountered on the City's collector and local road network.

Table 4-2 summarizes the geometric and operational characteristics as well as the LiDAR deployment periods for each site.

Table 4-2 Characteristics and LiDAR deployment periods of the four study intersections

Intersection	Major Road	Minor Road	Control Type	Data Collection Period
	AADT (average 2014-2020)	AADT (average 2014-2020)		
Charles St E & Sydney St S	13,555	4,295	two-way stop control	1 May 2023 – 31 Oct 2023 and 1 Sep 2024–30 Nov 2024
Spadina Rd E & Queens Blvd	16,668	4,519	two-way stop control	1 May 2023 – 31 Jul 2024
Strasburg Rd & Old Country Dr	21,867	6,340	all-way stop control	1 May 2023–31 Dec 2024
Wabanaki Dr & Wilson Ave	17,892	2,102	two-way stop control	1 Jun 2023–29 Feb 2024

Although all sites are unsignalized, the combination of control type (two- vs all-way stop), traffic volumes, and approach geometry provides sufficient heterogeneity to investigate how sample-size requirements vary across typical suburban intersection environments.

Traffic conflict data was collected using LiDAR sensors installed on hydro poles at each intersection. These sensors provided 24/7 monitoring during the periods listed in **Table 4-2**. It should be noted that despite the continuous monitoring setup, some data gaps occurred due to technical issues with the sensors, particularly during adverse weather conditions. Unlike traditional video-based conflict analysis, LiDAR sensors offer several advantages including privacy preservation, operation in varying lighting and weather conditions, and precise spatial measurement capabilities.

The LiDAR sensors captured 3-D point clouds representing all road users within their detection range. These point clouds were then processed using the BlueCity safety platform to extract road user trajectories and identify conflicts (Lesani et al., 2020). For this study, a conflict was defined as an interaction between two road users with a PET value less than 4 seconds. Also, consistent with the validation dataset, only vehicle-vehicle conflicts were retained.

Historical vehicle-vehicle collision data for the four intersections were obtained from the City of Kitchener for 2018–2022. Each recorded crash provided crucial details such as time, location, severity, road user information, and type of collision. These collisions act as the benchmark for assessing the accuracy of collision predictions derived from traffic conflicts. Result section presents a summary of the detected conflicts and historical collisions at each intersection.

To investigate the effect of sample size on prediction reliability, multiple sampling periods were created from the collected conflict data. For each intersection, we created multiple datasets with progressively increasing durations:

1. Initial Period: Starting from May 1, 2023, datasets of 2, 3, 4, 5, 6, and 7 days were formed to examine short-term data collection effects.
2. Weekly Increments: After the first week, the sample size was increased by weekly increments (14, 21, and 28 days) to evaluate moderate-term data collection effects.

3. Monthly Increments: Beyond one month (30 days), sample sizes were increased by monthly increments (60, 90, 120, 180, 270, 300, 330, and 365 days) to assess long-term data collection effects where data availability permitted.

This incremental approach allows for a comprehensive analysis of how prediction reliability evolves with increasing data collection duration, from very short periods (a few days) that are common in existing studies to extended periods (up to a full year) that to the best of our knowledge, have not been implemented in any practice. Note that, for each sampling period and intersection, the appropriate threshold for the POT approach was determined using the MRL and parameter-stability plots described in Section 4.2.3.2.

4.2.5 Model Evaluation

For each sample period, the adequacy of the fitted GP distribution model was assessed using both visual diagnostics and formal statistical tests. These evaluation approaches are essential to ensure that the GP distribution model appropriately represents the extreme conflict data, which is a prerequisite for reliable collision prediction.

4.2.5.1 Visual Diagnostics

Three complementary visual diagnostic tools were employed to assess model fit:

1. Probability-Probability (P-P) Plots: These plots compare the empirical cumulative distribution function (CDF) with the fitted GP distribution CDF. For a perfect fit, the points should fall along the diagonal line $y = x$. Systematic deviations from this line indicate model misspecification. The P-P plot is constructed by:
 - Computing the empirical CDF of the observed exceedances: $F_{emp}(y_i) = \frac{i}{n+1}$ for ordered observations $y(1), y(2), \dots, y(n)$
 - Computing the theoretical CDF using the fitted GP distribution parameters: $F_{theo}(y_i) = G(y_i; \hat{\sigma}, \hat{\xi})$; where $G(y_i; \hat{\sigma}, \hat{\xi})$ is the fitted GP distribution with scale and shape parameters of $\hat{\sigma}$ and $\hat{\xi}$, respectively.
 - Plotting the pairs $(F_{theo}(y_i), F_{emp}(y_i))$ for all observations $i = 1, 2, \dots, n$.

2. Quantile-Quantile (Q-Q) Plots: These plots compare empirical and theoretical quantiles. For a perfect fit, the points should fall along the diagonal line $y = x$. The Q-Q plot is more sensitive to discrepancies in the tail of the distribution than the P-P plot.

It is constructed by:

- Computing the empirical quantiles: $y(i)$ for observations $i = 1, 2, \dots, n$.
 - Computing the theoretical quantiles: $G^{-1}(F_{emp}(y_i); \hat{\sigma}, \hat{\xi})$
 - Plotting the pairs $(G^{-1}(F_{emp}(y_i); \hat{\sigma}, \hat{\xi}), y(i))$ for all observations $i = 1, 2, \dots, n$
3. Density Function Comparison: This involves comparing the histogram of exceedances with the fitted GP distribution density function. It provides a direct visual assessment of how well the fitted model represents the empirical distribution.

4.2.5.2 Formal Statistical Test

The Anderson-Darling (AD) test was employed as a formal statistical test of model fit. This test is particularly appropriate for evaluating the fit of extreme value distributions because it gives more weight to discrepancies in the tails of the distribution compared to other goodness-of-fit tests like the Kolmogorov-Smirnov test.

The AD test statistic is calculated as **Equation (4-9)**:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\ln(F(y_i)) + \ln(1 - F(y_{n+1-i}))] \quad (4-9)$$

where $F(y)$ is the CDF of the specified GP distribution with estimated parameters, and y_i are the ordered observations.

The null hypothesis is that the data follows the specified GP distribution. A smaller test statistic indicates better fit, and the test statistic is compared to critical values to determine statistical significance. For this study, a significance level of $\alpha = 0.05$ was used, with the null hypothesis rejected if the p-value was less than 0.05, indicating an inadequate fit of the GP distribution model.

4.2.6 Analysis of Sample Size Effects

4.2.6.1 Prediction Reliability Assessment

For each sample period, we evaluate the reliability of collision predictions by examining both the point estimate (i.e., mean) and the associated uncertainty (i.e., CI):

1. *Mean Collision Prediction:* The expected annual collision frequency for each intersection, calculated as described in Section 4.2.3.5.
2. *Prediction Uncertainty:* The width of the 95% CI for the annual collision prediction, which provides a direct measure of prediction reliability. Narrower CI indicate more reliable predictions.
3. *Magnitude Comparison with Historical Data:* The predicted annual collision frequencies will be compared to observed historical collision data for a high-level sense of magnitude. It is crucial to emphasize that this is not a formal accuracy assessment, as the historical data does not represent a stable "ground truth." More fundamentally, the extreme rarity of crashes means that even several years of historical data provide only a statistical estimate, not the true long-term average. This inherent uncertainty, the fact that the ground truth is effectively unknown, combined with differences in data sources (e.g., sensor detection versus police reports), means a direct one-to-one validation is inappropriate. This is a primary reason why statisticians often rely on simulation-based validation where the ground truth is known by design.

4.2.6.2 Convergence Analysis

To identify the minimum adequate sample size for reliable collision prediction, we analyze how prediction reliability changes with increasing sample size by calculating:

1. *Incremental Change in Mean Prediction:* The percentage change in mean collision frequency prediction between consecutive sample periods as **Equation (4-10)**:

$$\Delta\mu_i = \frac{|\mu_i - \mu_{i-1}|}{\mu_{i-1}} \times 100\% \quad (4-10)$$

where μ_i is the mean collision prediction for sample period i .

2. Incremental Change in CI Width: The percentage change in the width of the 95% CI between consecutive sample periods as **Equation (4-11)**:

$$\Delta CI_i = \frac{|CI_i - CI_{i-1}|}{CI_{i-1}} \times 100\% \quad (4-11)$$

where CI_i is the width of the CI for sample period i .

The optimal sample size is determined as the point where both the mean prediction and CI width stabilize, defined as the sample period beyond which additional data provides only marginal improvements in prediction reliability. There are no standards on the definition of the marginal improvement; however, in this study it is considered less than 5% change to keep it relatively small.

This approach quantifies the diminishing returns of extended data collection periods and identifies the most efficient resource allocation for traffic conflict studies. By determining the point at which additional data collection yields minimal improvements in prediction reliability, this study provides practical guidance on the minimum adequate data collection period for reliable collision prediction using traffic conflicts.

4.3 Results

This section presents the findings of the sample size analysis. It begins with a descriptive summary of the collected traffic conflict and historical collision data. Following this, the process of threshold selection and model validation is detailed. The core of the section then focuses on the primary analysis: quantifying the effect of sample size on the reliability of collision predictions and identifying the point of diminishing returns for data collection.

4.3.1 Descriptive Data Summary

The analysis is based on data from four unsignalized intersections in the City of Kitchener, Canada. Traffic conflict data, characterized by PET, were collected continuously using LiDAR

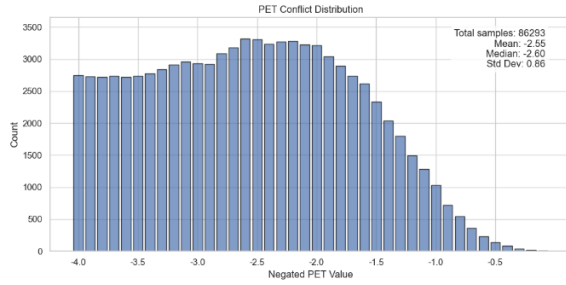
sensors, although the total observation period varied by site due to logistical constraints. Historical collision data for the same locations were obtained for the period of 2018–2022.

A summary of the collected data for each intersection is provided in **Table 4-3**. This includes the total data collection duration, the number of observed historical crashes, and descriptive statistics for the raw PET conflict data. The "Strasburg Rd & Old Country Dr" intersection yielded the most conflicts and the most complete dataset over a one-year period. Consequently, this intersection will be used as the primary example to illustrate the detailed methodology in the following sections, with summary results provided for the other locations.

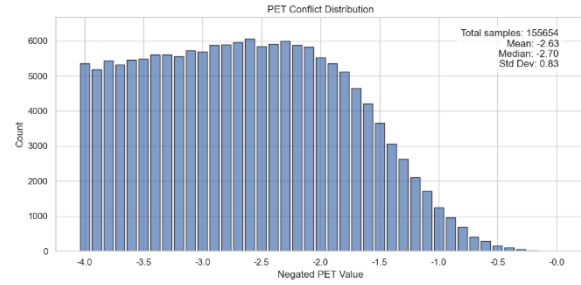
Table 4-3 Summary of conflict and collision data for the four study intersections

#	Intersection name	No. of crashes (2018-22)	Data Collection duration (days)	No. of conflicts (< 4 sec)	PET indicator (sec)		
					Mean	Median	SD
1	Charles St E & Sydney St S	20	306	86,293	2.55	2.60	0.86
2	Spadina Rd E & Queens Blvd	7	462	155,654	2.63	2.70	0.83
3	Strasburg Rd & Old Country Dr	13	543	718,860	2.48	2.5	0.78
4	Wabanaki Dr & Wilson Ave	25	258	41,290	2.77	2.80	0.81

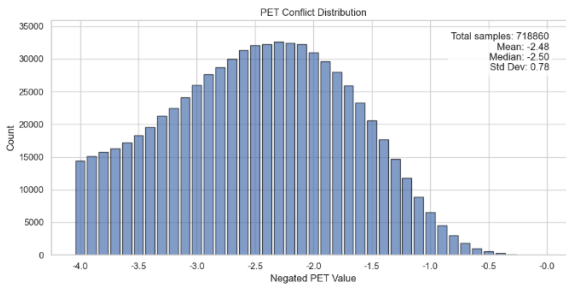
Figure 4-3 shows the frequency distribution of negated PET values for each intersection. As required by the EVT framework, which models maxima, PET values were negated so that more severe conflicts (smaller PETs) correspond to larger numerical values. The distributions illustrate the high frequency of less severe conflicts and the characteristic long tail of more extreme events that are the focus of the POT analysis.



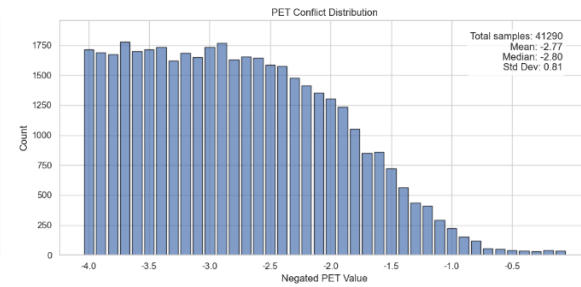
1) Charles St E & Sydney St



2) Spadina Rd E & Queens



3) Strasburg Rd & Old



4) Wabanaki Dr & Wilson

Figure 4-3 Negated PET Conflict Distribution for Each Study Intersection

4.3.2 Threshold Selection:

The selection of an appropriate threshold is a critical step in the POT approach. For each sample period at each intersection, an optimal threshold was identified using the graphical methods outlined in the methodology. To illustrate this process, the analysis for the 30-day sample period at the Strasburg Rd & Old Country Dr intersection is presented.

The MRL plot, shown in **Figure 4-4**, displays the mean of the exceedances above a given threshold. An appropriate threshold should be located at the point where the plot becomes approximately linear. For this dataset, the plot suggests an optimal threshold for negated PET values lies within the range of $[-1.5, -1.2]$.

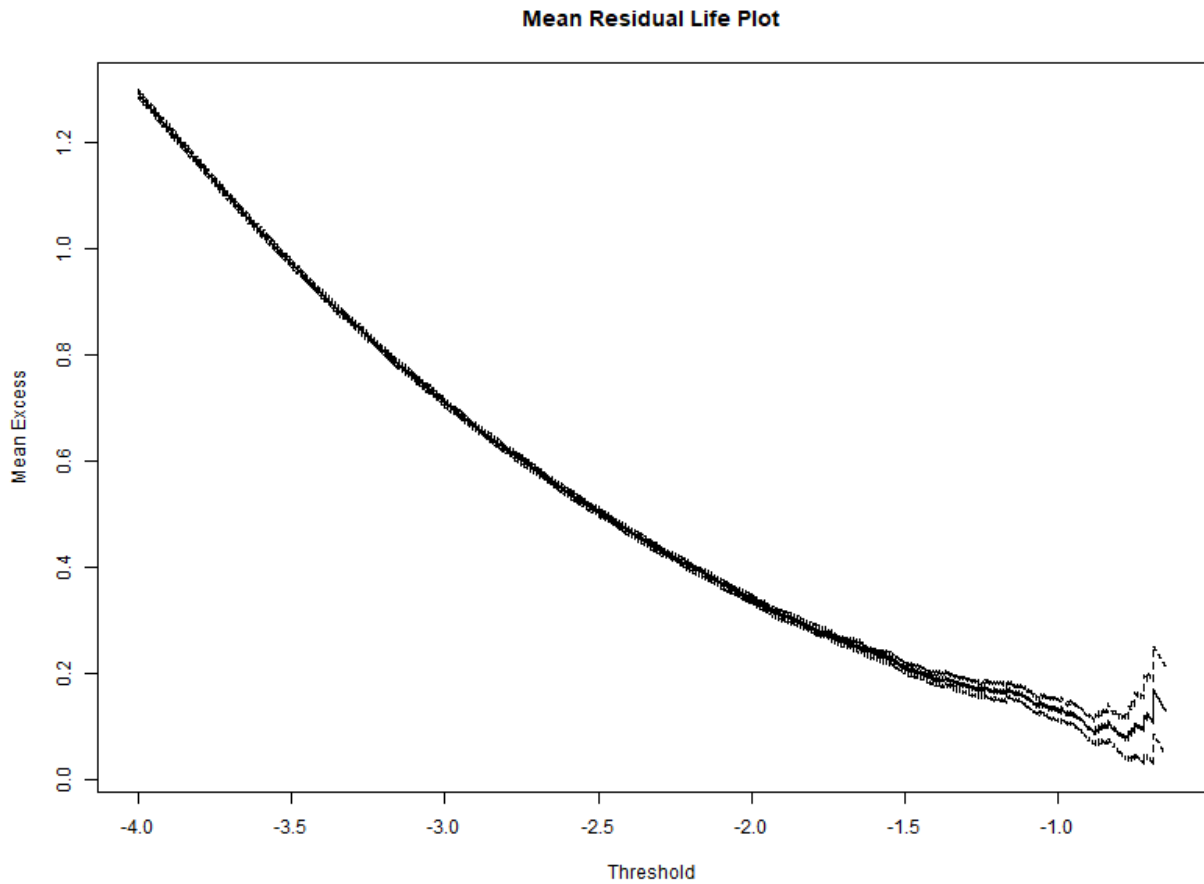


Figure 4-4 MRL Plot for the 30-day sample at Strasburg Rd & Old Country Dr

This finding was complemented by the parameter stability plot (**Figure 4-5**), which examines the stability of the GP distribution shape (ξ) and modified scale (σ_u) parameters across a range of potential thresholds. The plot shows that the shape parameter remains relatively constant for thresholds above $[-1.4]$, while the scale parameter is stable in the range of $[-1.4, -1.3]$.

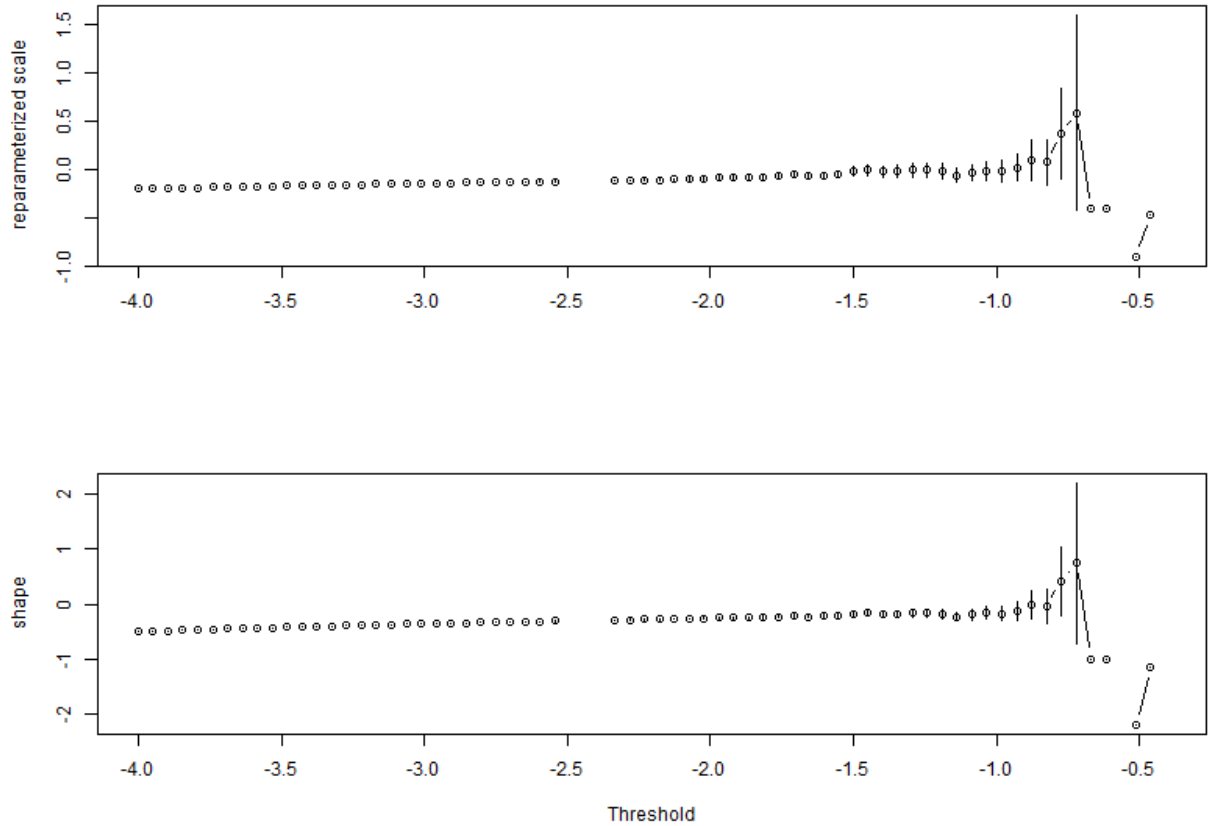


Figure 4-5 Modified Scale and Shape Parameters Stability Plot

By combining the insights from these plots, a stable threshold range of $[-1.4, -1.3]$ was identified. To select the final value, the GP distribution model was fitted for multiple thresholds within this range, and the value of -1.35 was chosen as it yielded the lowest DIC value, indicating the best model fit.

This process was repeated for every sample period at all four intersections. **Table 4-4** summarizes the selected optimal threshold and the corresponding number of extreme conflict events (i.e., PET over the threshold) for each sampling scenario at all intersections. As expected, the number of extreme events generally increases with the sample period duration, though this relationship is not perfectly linear due to the stochastic nature of traffic conflicts.

Table 4-4 Optimum threshold and number of extreme conflicts for each sampling period at all intersections

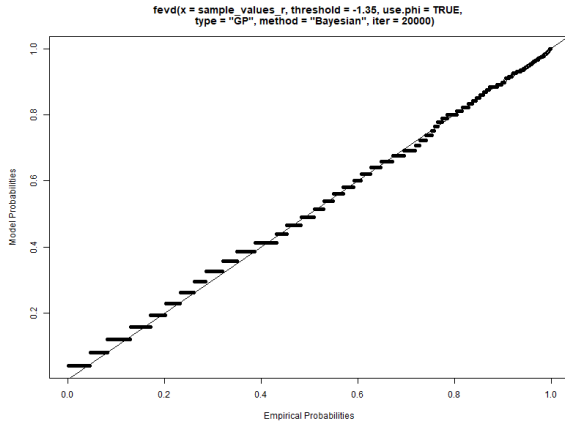
Sample Period	Charles St E & Sydney St S		Spadina Rd E & Queens Blvd		Strasburg Rd & Old Country Dr		Wabanaki Dr & Wilson Ave	
	Threshold	# of Extreme Conflicts	Threshold	# of Extreme Conflicts	Threshold	# of Extreme Conflicts	Threshold	# of Extreme Conflicts
2 days	-2.93	468	-3.2	441	-1.60	137	-3.70	311
3 days	-2.65	531	-3.1	648	-1.60	242	-3.70	417
4 days	-2.35	554	-2.84	756	-1.57	287	-3.50	586
5 days	-2.25	615	-2.48	742	-1.55	344	-3.30	648
6 days	-2.10	563	-2.21	641	-1.53	353	-3.10	755
7 days	-1.96	498	-2.06	574	-1.50	342	-2.90	766
14 days	-1.74	756	-1.81	747	-1.45	697	-2.20	815
21 days	-1.52	736	-1.68	853	-1.40	918	-1.86	691
30 days	-1.36	882	-1.49	808	-1.35	1074	-1.72	795
60 days	-1.20	998	-1.11	2066	-1.25	1213	-1.66	1506
90 days	-1.14	1090	-0.82	1100	-1.20	1282	-1.43	1283
120 days	-1.12	1255	-0.65	570	-1.18	1623	-1.18	738
150 days	-1.10	1454	-0.61	495	-1.15	1894	-1.09	584
180 days	-1.06	1430	-0.6	520	-1.14	2195	-1.06	588

210 days	-0.88	695	-0.59	528	-1.13	2441	-1.04	691
240 days	-0.73	571	-0.58	505	-1.12	2600	-1.03	813
270 days	-0.69	637	-0.57	484	-1.10	2582	N/A	N/A
300 days	-0.64	639	-0.57	496	-1.10	2897	N/A	N/A
330 days	N/A	N/A	-0.56	479	-1.00	1804	N/A	N/A
365 days	N/A	N/A	-0.56	490	-0.95	1544	N/A	N/A

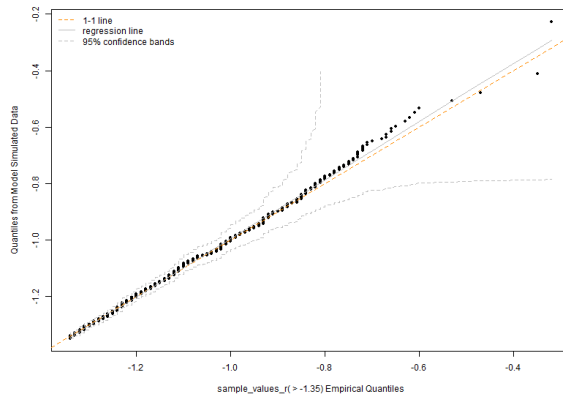
4.3.3 Generalized Pareto Distribution Model Fit and Validation

Before analyzing the sample size effects, the adequacy of the fitted GP distribution model was assessed for each sample period using both visual diagnostics and the AD formal statistical test. For illustration,

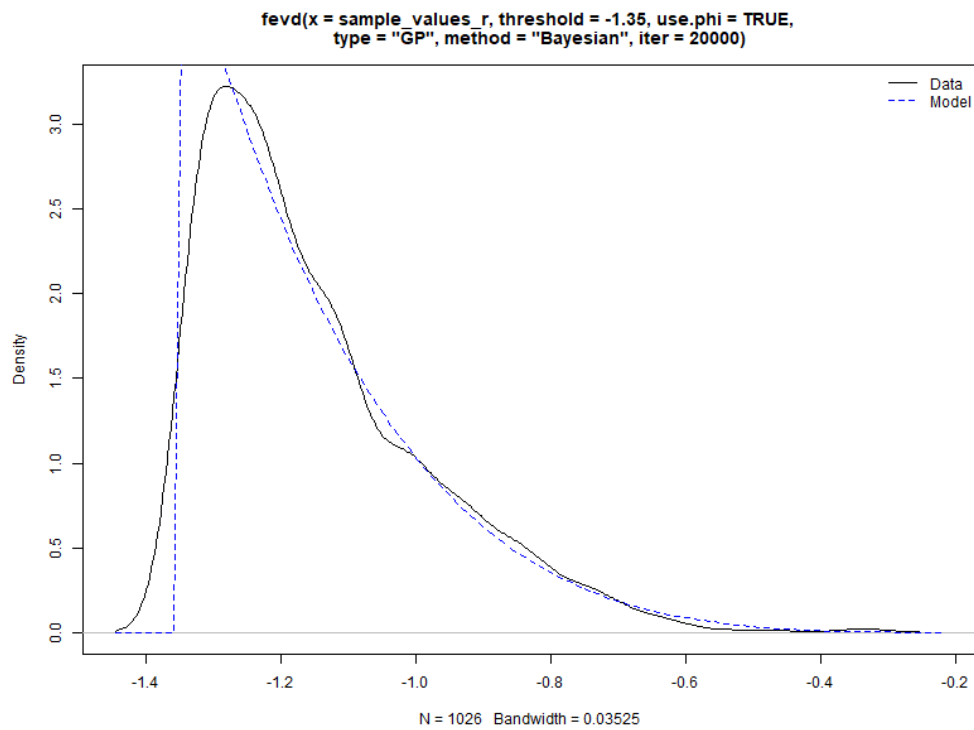
Figure 4-6 presents the diagnostic plots for the 30-day sample at the Strasburg Rd & Old Country Dr intersection. The P-P and Q-Q plots show points falling closely along the diagonal line, indicating a good fit between the empirical data and the theoretical GP distribution model. The density plot further confirms that the fitted GP distribution density function accurately represents the histogram of the observed extreme conflicts.



1) Probability-Probability (P-P) Plot



2) Quantile-Quantile (Q-Q) Plot



3) Density Plot

Figure 4-6 GP distribution Model Diagnostic Plots for the 30-day sample at Strasburg Rd & Old Country Dr

The AD goodness-of-fit test was then performed. For the 30-day sample at Strasburg Rd & Old Country Dr, as an example, the test statistic was 4.6395, with a corresponding p-value of 0.069. Since the p-value is greater than the significance level of 0.05, the null hypothesis that

the data follows a GP distribution cannot be rejected, providing statistical support for the model's adequacy. This validation process was conducted for all sample periods and all intersection, confirming that the GP distribution model was appropriate for subsequent collision prediction in all cases.

4.3.4 Sample Size Effects on Collision Prediction Reliability

This section details the core findings regarding the relationship between sample size and collision prediction reliability. The analysis focuses on how the mean annual collision estimates and their 95% CI evolve with increasing data collection duration.

Table 4-5 Estimated GP distribution Model Parameters for Strasburg Rd & Old Country Dr

Days of Data	Shape	95% CI for Shape	Scale	95% CI for Scale
2	-0.24411	[-0.37461, -0.07039]	0.31026	[0.242306, 0.382502]
3	-0.21413	[-0.32433, -0.08293]	0.299823	[0.249501, 0.356692]
4	-0.26177	[-0.36617, -0.13728]	0.322692	[0.272046, 0.377495]
5	-0.22923	[-0.32302, -0.11896]	0.299321	[0.256925, 0.344032]
6	-0.22991	[-0.3263, -0.11972]	0.29895	[0.256454, 0.345459]
7	-0.23082	[-0.32363, -0.12507]	0.287986	[0.247961, 0.330634]
14	-0.15736	[-0.21565, -0.08822]	0.246652	[0.222704, 0.272523]
21	-0.18001	[-0.22563, -0.1214]	0.241825	[0.221181, 0.26345]
30	-0.18361	[-0.22407, -0.13536]	0.236028	[0.21814, 0.254136]
60	-0.1686	[-0.21278, -0.11717]	0.213033	[0.197065, 0.229027]
90	-0.185	[-0.22423, -0.14178]	0.209427	[0.195211, 0.225092]
120	-0.18354	[-0.21905, -0.14378]	0.203193	[0.190776, 0.216558]
150	-0.17113	[-0.20197, -0.1345]	0.19549	[0.183466, 0.206779]
180	-0.1655	[-0.19565, -0.132]	0.190955	[0.181024, 0.200786]
210	-0.16566	[-0.19549, -0.13523]	0.189959	[0.180339, 0.200757]
240	-0.17138	[-0.19652, -0.14107]	0.190193	[0.180831, 0.19943]
270	-0.17758	[-0.20347, -0.14794]	0.190899	[0.182093, 0.199819]
300	-0.17446	[-0.20028, -0.1433]	0.192098	[0.183536, 0.201185]
330	-0.16398	[-0.2013, -0.12278]	0.176396	[0.16558, 0.187236]
365	-0.123	[-0.16933, -0.07091]	0.156878	[0.145917, 0.16827]

Using the Bayesian approach, the posterior distributions for the GP distribution shape and scale parameters were estimated for each sample period. As an example, **Table 4-5** shows the estimated parameters for the Strasburg Rd & Old Country Dr intersection. A clear trend emerges: as the sample size increases, the 95% CI for both the shape (ξ) and scale (σ_u) parameters become progressively narrower, indicating greater precision in parameter estimation. The convergence of the MCMC simulation was confirmed for all models by visually inspecting trace plots and ensuring the ratio of Monte Carlo Standard Error to the posterior SD was below 0.05. **Figure 4-7** shows one example of MCMC trace plots for the 30 days sample periods at Strasburg Rd & Old Country Dr.

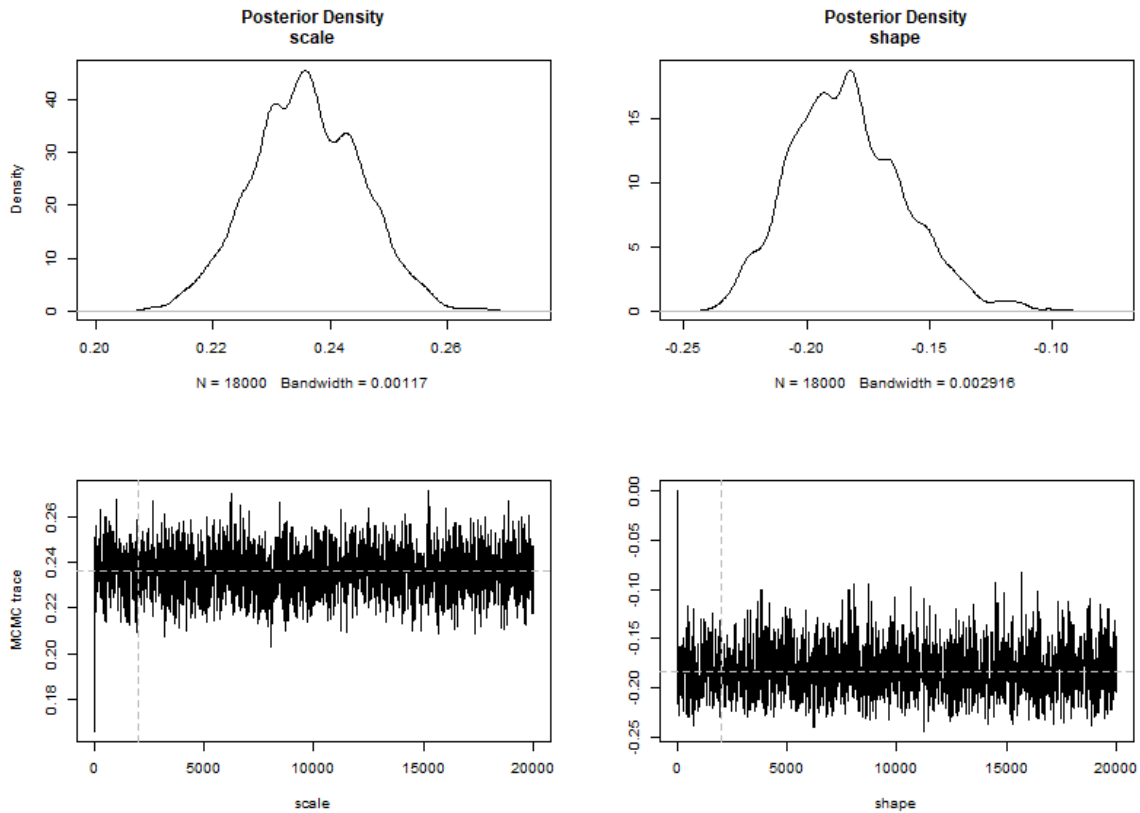


Figure 4-7 Sample of MCMC Trace plot for 30-days sample days at Strasburg Rd & Old Country Dr

The estimated GP distribution parameters were then used to predict the annual collision frequency and its 95% CI for each sample period. **Table 4-6** presents these results for all four study intersections.

Table 4-6 Predicted Annual Collisions and 95% CI by Sample Size

Sample Period	Charles St E & Sydney St S		Spadina Rd E & Queens Blvd		Strasburg Rd & Old Country Dr		Wabanaki Dr & Wilson Ave	
	Expected Collision	95% CI	Expected Collision	95% CI	Expected Collision (95% CI)	95% CI	Expected Collision	95% CI
2 days	0.04	[0, 1308.90]	37.50	[0, 1053.12]	0.18	[0, 721.87]	8.96	[0, 1491.77]
3 days	0.00	[0, 659.79]	21.35	[0, 616.91]	3.34	[0, 570.08]	2.79	[0, 1054.87]
4 days	0.00	[0, 472.12]	13.09	[0, 430.78]	0.30	[0, 462.17]	10.53	[0, 623.50]
5 days	0.01	[0, 326.56]	8.22	[0, 246.27]	2.35	[0, 354.21]	9.62	[0, 454.30]
6 days	0.01	[0, 302.45]	7.77	[0, 202.18]	3.10	[0, 338.27]	9.46	[0, 402.56]
7 days	0.10	[0, 223.08]	7.06	[0, 160.96]	1.72	[0, 241.51]	7.90	[0, 311.64]
14 days	22.62	[0, 250.32]	3.28	[0, 74.49]	15.87	[0, 163.92]	16.90	[0, 206.04]
21 days	23.79	[0, 203.31]	1.45	[0, 34.26]	8.17	[0, 98.45]	25.32	[0, 174.16]
30 days	16.17	[0, 129.03]	1.59	[0, 33.86]	8.69	[0.00, 71.38]	33.72	[0.73, 168.20]
60 days	14.95	[0, 89.95]	6.75	[0.01, 44.95]	9.79	[0.06, 56.59]	35.31	[2.22, 126.31]
90 days	15.93	[0.24, 67.94]	12.16	[0.15, 57.92]	6.86	[0.07, 37.04]	44.50	[4.47, 129.76]
120 days	17.98	[1.62, 58.68]	10.36	[0.08, 48.66]	6.43	[0.16, 30.89]	38.99	[1.71, 121.96]
150 days	20.82	[2.38, 62.78]	13.47	[0.34, 49.29]	8.94	[0.81, 29.02]	28.61	[0, 106.12]
180 days	20.80	[2.80, 58.50]	12.44	[0.21, 46.32]	9.07	[1.20, 28.55]	22.05	[0, 86.82]
210 days	19.21	[1.63, 55.54]	11.21	[0.24, 40.70]	9.41	[1.40, 28.46]	22.59	[0.03, 80.53]
240 days	24.14	[3.21, 60.85]	9.85	[0.21, 35.37]	8.82	[1.68, 25.43]	21.67	[0.1, 75.28]

270 days	29.71	[5.09, 66.82]	8.95	[0.12, 32.69]	9.27	[1.95, 25.05]	N/A	N/A
300 days	33.56	[7.95, 70.08]	8.02	[0.14, 28.72]	10.79	[2.61, 28.50]	N/A	N/A
330 days	N/A	N/A	7.28	[0.12, 25.92]	13.77	[3.02, 34.00]	N/A	N/A
365 days	N/A	N/A	6.81	[0.13, 23.81]	15.71	[3.81, 37.48]	N/A	N/A

The results in **Table 4-6** reveal a critical finding: while the mean collision predictions stabilize relatively quickly, the CI for these predictions are unacceptably wide for short data collection periods. For example, at the Strasburg Rd & Old Country Dr intersection, the 95% CI width for a 2-day sample is over 720 collisions, rendering the estimate impractical for decision-making. As the sample size increases, the CI width shrinks dramatically, demonstrating a significant improvement in prediction reliability. To better visualize this convergence, **Figure 4-8** plots the mean annual collision prediction and its 95% CI against the sample size for the Strasburg Rd & Old Country Dr intersection.

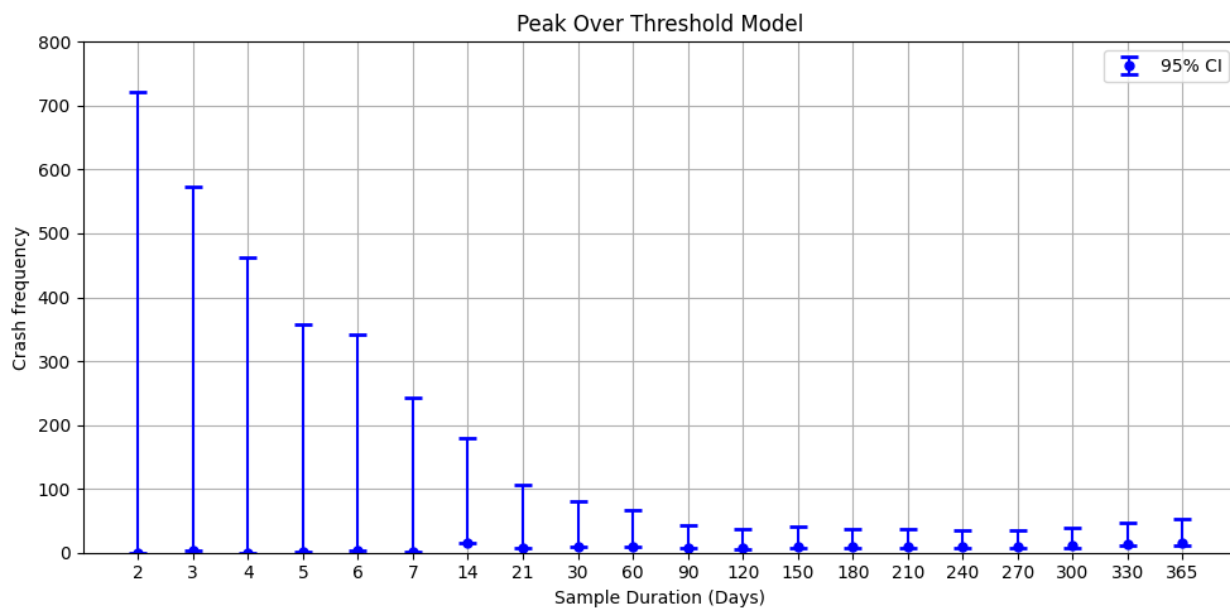


Figure 4-8 Convergence of Annual Collision Prediction and 95% CI with Increasing Sample Size for Strasburg Rd & Old Country Dr

4.3.5 Identification of Minimum Sample Size

To objectively identify the point of diminishing returns, the incremental change in both the mean prediction and the CI width was calculated between consecutive sample periods, as per the methodology. The optimal sample size is defined as the point where these metrics stabilize, indicating that further data collection provides only marginal improvements in reliability. A stability threshold of less than 5% incremental change was used.

Figure 4-9 and **Figure 4-10** show the percentage change in the mean prediction and CI width for the Strasburg Rd & Old Country Dr intersection, respectively. The CI width shows large reductions for initial increases in sample size, but the rate of improvement slows considerably after a certain point.

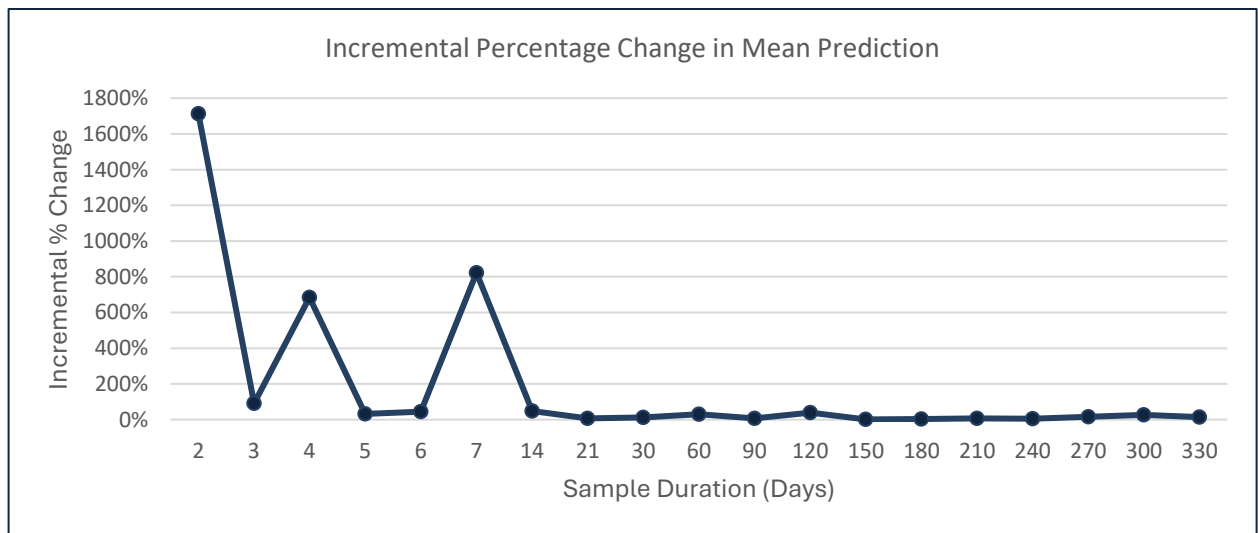


Figure 4-9 Incremental Percentage Change in Mean Prediction for Strasburg Rd & Old Country Dr

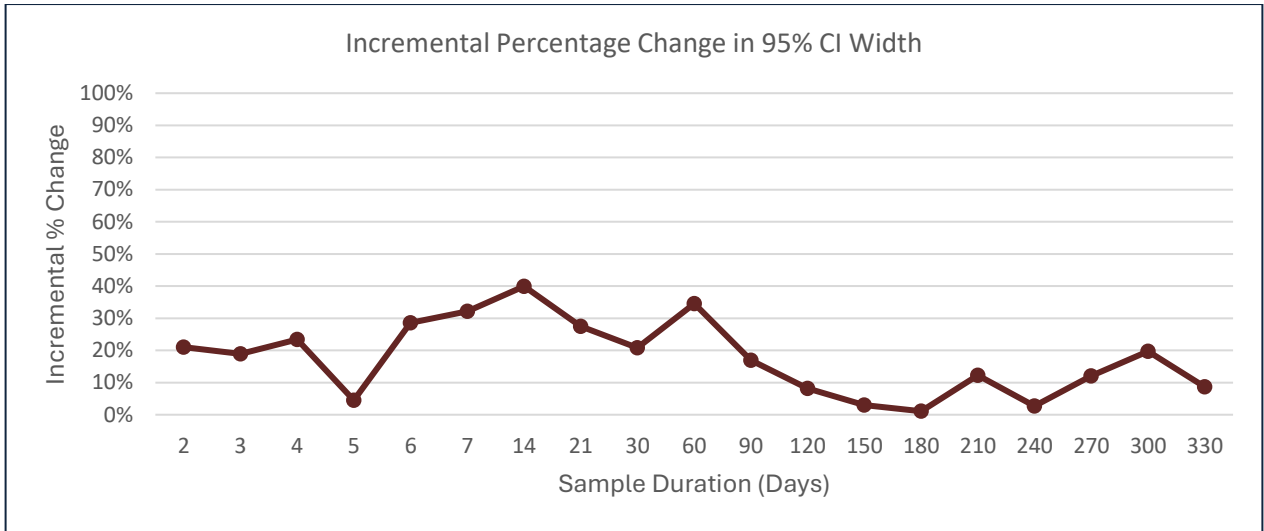


Figure 4-10 Incremental Percentage Change in CI Width for Strasburg Rd & Old Country Dr

Based on this analysis, the point of convergence for the Strasburg Rd & Old Country Dr intersection was identified. The incremental change in both the mean prediction and the CI width falls below 5% after 180 days of data collection for the first time. Beyond this point, the marginal benefit of collecting more data diminishes significantly. At this duration, the predicted annual collisions were 9.07 with a 95% CI of [1.20, 28.55]. While still uncertain, this is a vast improvement over the results from shorter periods and provides a practically useful estimate for safety practitioners. For context, the historical average at this location was 2.6 collisions per year (13 crashes over 5 years). The predicted value of 9.07 is within a plausible range for this type of intersection, but this comparison should be viewed only as an order-of-magnitude check, not a measure of accuracy. The discrepancy is not an indication of model error. The historical average itself is simply a statistical estimate of a rare event process over a limited time frame, meaning the true long-term average (the 'ground truth') remains unknown. Therefore, a direct validation against this number is statistically inappropriate.

Repeating this analysis for the other intersections revealed variations in the required sample size. For instance, the lower volume "Wabanaki Dr & Wilson Ave" intersection required 240 days to reach stability, likely due to a lower frequency of extreme conflict events. These findings suggest that the optimal data collection duration is site-specific and influenced by traffic characteristics. A summary of the identified optimal sample sizes and corresponding collision predictions for all sites is provided in **Table 4-7**.

Table 4-7 Optimal Sample Size and Corresponding Collision Prediction Reliability for All Intersections

Intersection Name	Minimum Sample Size (Days)	Predicted Annual Collisions	95% CI	CI Width	Historical Annual Collisions
Charles St E & Sydney St S	210	19.21	[1.63, 55.54]	53.91	4
Spadina Rd E & Queens Blvd	150	13.47	[0.34, 49.29]	48.94	1.75
Strasburg Rd & Old Country Dr	180	9.07	[1.20, 28.55]	27.35	2.6
Wabanaki Dr & Wilson Ave	240	21.67	[0.1, 75.28]	75.18	5

4.4 Discussion and Conclusion

This study's findings provide a quantitative answer to a critical methodological question in traffic safety: how reliable is the collision prediction based on conflict data and how much conflict data is needed for reliable collision prediction using EVT? The results show that the reliability of collision estimates is profoundly dependent on the sample size. While mean collision predictions may begin to stabilize with smaller datasets, the associated CI is impractically wide for short observation periods, confirming the concerns raised about studies that rely on limited samples. A key outcome of this research is the identification of a point of diminishing returns, where the improvement in prediction reliability no longer justifies the cost of more data collection. For a typical suburban intersection in our study (i.e., Strasburg Rd & Old Country Dr), as an example, this point was reached after approximately 180 days of continuous monitoring, suggesting that the common practice of collecting data for only a few hours or days is inadequate for producing reliable safety estimates. This inadequacy stems from a fundamental reality: near-misses and severe conflicts, while more frequent than collisions, are still rare events that demand extensive observation periods to be captured with statistical confidence.

Our empirical results both support and expand upon insights from the literature. For example, simulations by Songchitruksa and Tarko (2006) suggested a need for 30 to 50 days of observation; our findings indicate a longer period is necessary, likely due to the greater variability of real-world traffic conditions compared to simulations. The study's primary contribution is its systematic, long-term empirical basis, which moves beyond the theoretical or short-term analyses that have characterized much of the previous work.

The practical implications for traffic safety practitioners are substantial. This research provides an evidence-based framework that can guide decisions on resource allocation for data collection, helping practitioners avoid the dual risks of under-sampling (which yields unusable results) and over-sampling (which wastes resources). By establishing a methodology for achieving reliable estimates, this work enhances the utility of proactive safety analysis, particularly for locations with little to no crash history.

While this study provides a crucial methodological baseline, the findings should be considered in light of its limitations, which also point toward important avenues for future research.

- Scope and Generalizability: The analysis was based on four unsignalized intersections in one city. Future work should expand this analysis to a larger and more diverse set of locations, including signalized intersections and freeway segments, to develop more broadly applicable guidelines.
- Road User and Conflict Types: This study focused exclusively on vehicle-vehicle conflicts using PET as the indicator. Future research should investigate the sample size requirements for other road users, such as pedestrians and cyclists, and for other widely used conflict indicators like TTC, speed, and DRAC.
- Model Complexity: To clearly demonstrate the impact of the sample period, a simple stationary EVT model was intentionally used. However, more complex models could be explored. Future studies should investigate sample size requirements when using non-stationary models, which incorporate covariates like traffic volume or weather and may reduce unexplained variability. A systematic comparison between the POT approach used here and the BM approach would also be valuable to determine if the model choice impacts data requirements.

In conclusion, this research confirms that achieving reliable, proactive safety estimates requires a significant data collection effort, likely on the order of several months. By addressing the current limitations, future work can continue to refine these guidelines, moving the field of traffic safety toward a more robust and evidence-based paradigm.

Chapter 4 addressed a fundamental source of data uncertainty in conflict-based road safety analysis. It established that reliable collision predictions require extensive, long-term data collection, thereby debunking the "myth of quick analysis" and providing an evidence-based methodology for determining adequate sample size.

As noted in the chapter's limitations, the analysis intentionally employed a simple, stationary EVT model to clearly demonstrate the impact of the sample period. A critical component of this standard methodology is the selection of a severity threshold for the POT approach. This step, which was performed using conventional graphical methods in the preceding analysis, relies on subjective interpretation and introduces a significant source of epistemic (model) uncertainty into the results.

Having established the data quantity required for reliable analysis, the focus now shifts to refining the analytical methodology itself. To enhance the objectivity and consistency of conflict-based analysis, the following chapter develops and validates an automated approach for threshold selection, aiming to replace subjective graphical techniques with a repeatable and statistically rigorous procedure.

Chapter 5

An Automated Threshold Selection Approach for Conflict-Based Road Safety Analysis

The preceding chapter established the critical need for long-term data collection to address data uncertainty in conflict-based safety analysis, demonstrating that reliable collision predictions require an extensive observational period. Having established the foundational data requirements, the focus now shifts to refining the analytical methodology itself. A primary approach for assessing crash risk from conflict data is EVT, particularly the POT method, which models all conflict events exceeding a certain severity threshold.

While the POT approach is effective, its reliability is often compromised by a significant source of model uncertainty: the selection of an appropriate threshold. This step is crucial, as a threshold set too low can lead to inaccurate results, while one set too high can result in high variance due to limited data. To date, most road safety studies have relied on subjective graphical methods, such as MRL plots, to determine this threshold. This practice introduces potential inconsistency, as different analysts could arrive at different conclusions from the same dataset.

This chapter directly confronts this methodological challenge by proposing and validating an automated and objective approach for threshold selection in POT models. The primary aim is to replace subjective interpretation with a repeatable and statistically rigorous procedure, thereby improving the accuracy and consistency of crash risk estimation and addressing a critical gap in the application of EVT for road safety analysis.

This chapter is based on the manuscript: Aminghafouri, R., & Fu, L. (2025). An Automated Threshold Selection Framework for Conflict-Based Road Safety Analysis. Transportation Research Record, (Under Review).

The approach, which combines statistical measures of goodness-of-fit with L-moment ratio analysis, is first assessed through an extensive simulation study and then demonstrated using a real-world case study with conflict data from ten signalized intersections. The results show its effectiveness in improving crash prediction accuracy, providing a crucial tool for more

reliable proactive safety analysis and enhancing the foundation of conflict-based safety management.

5.1 Literature Review

The use of the EVT technique for traffic safety evaluation has increased significantly in recent years. The basic idea behind using EVT in safety analysis is to estimate rarely occurring crashes using multiple traffic conflict measures that are more common and easily observed. Several studies emphasized various advantages of using the EVT technique, such as the lack of a need for crash data throughout the model estimation phase. These advantages make the EVT approach a promising and flexible tool for traffic safety analysis (Ali, Haque, et al., 2023a; Hussain, Ali, Li, & Haque, 2023; Ankunda et al., 2024; Singh et al., 2024).

Table 5-1 Review Studies on Traffic Conflict Analysis

Num.	Authors (Date)	Main discussed topics
1	Lai et al. (2014)	- Define traffic conflicts and explain conceptual issues - Review major proximity-based measures for traffic conflict - Explore methods for collecting traffic conflict data - Evaluate the validity of traffic conflict techniques - Analyze the relationship between conflict and crash
2	Johnsson et al. (2018)	- Overview different traffic conflict measures - Examine traffic conflict measures for vulnerable road users - Assess traffic conflict measures' validity and reliability
3	Zheng, Sayed, et al. (2021)	- Discuss traffic conflict-crash theoretical concepts - Review conflict quantification measures - Evaluate conflict analysis statistical models - Analyze conflict-crash relationship models
4	Arun, Haque, et al. (2021b)	- Define crash surrogacy and identify good traffic conflict measures - Examine crash dimensions captured by traffic conflict measures - Discuss traffic conflict observation methods - Evaluate traffic conflict measures and applications - Explore conflict identification thresholds - Investigate crash-conflict validation methods
5	Arun, Haque, Washington, et al. (2021)	- Categorize conflict measures into six families - Analyze study context impact on measure selection - Examine conflict measures for vulnerable road users - Investigate context-specific measure thresholds - Demonstrate traffic environment's influence on measure selection
6	Ali, Haque, et al. (2023a)	- Review EVT in conflict/crash risk analysis - Propose framework to categorize extreme value models - Compare BM and POT sampling - Discuss uni-, bi-, and multivariate extreme value models - Examine methods for capturing unobserved heterogeneity - Evaluate conflict measures in extreme value models - Analyze data sources and processing for applications

This section reviews existing literature that has applied EVT models to predict crash risk in road safety studies, focusing specifically on the various threshold selection techniques used within the POT approach. However, for readers seeking a more comprehensive overview of traffic conflict safety studies, **Table 5-1** presents a list of review studies that offer additional insights and information.

One of the pioneering works was conducted by Campbell (1996), who applied the EVT approach in road safety analysis. Building on this foundation, some other studies further utilized EVT to establish relationships between non-crash events and crashes. Songchitruksa and Tarko (2006), for example, used the EVT approach to predict right-angle collisions at signalized intersections; their research found a significant connection between expected crashes and actual crash data. In a recent study conducted by Zheng and Sayed (2019a), the EVT approach was utilized in before-after safety studies, and their findings indicated that this method yields reliable and robust evaluation results.

In addition to crash estimation, researchers also compared different sampling approaches (i.e., BM and POT methods) for specific crash-related scenarios. In recent years, researchers have utilized the BM approach to develop various EVT models (e.g., univariate, multivariate, stationary, and non-stationary) based on conflict analysis for road safety assessment. For example, using BM approach, Zheng and Sayed (2020b) developed a Bayesian hierarchical GEV model for real-time crash risk prediction at signalized intersections. Their approach uses EVT to combine data from multiple intersections and dynamic traffic parameters, enabling cycle-level risk quantification and identification of crash-prone conditions in real-time. In another study, Fu and Sayed (2022a) proposed an integrated Bayesian dynamic linear model with BM EVT for real-time safety analysis. Their method allows for time-varying parameters and accounts for temporal variability by recursively updating model parameters. Fu and Sayed (2022b) proposed a multivariate approach for real-time safety evaluation using EVT and multiple conflict indicators. Their study employed the BM method, treating cycle-level maximums of conflict indicators as extremes. Ali, Haque, et al. (2023b) applied a BM approach in their real-time vehicle-pedestrian crash risk estimation framework for signalized intersections. This BM approach was incorporated into a Bayesian model to estimate crash risk, with covariates introduced in the location parameter to capture heterogeneity.

Various research studies indicate that the POT approach surpasses the BM method in several important aspects. These include making better use of available data, providing more accurate predictions, and offering more dependable forecasts. Zheng et al. (2014), for example, evaluated the BM and POT approaches for assessing lane-changing maneuver-related collisions on highways. Their findings demonstrated the superiority of the POT approach in terms of data utilization, estimated accuracy, and reliability. In another research, Zheng and Sayed (2019c) used the POT technique to analyze four distinct traffic conflict indicators in terms of collision estimation. Their research confirmed the POT approach's reliability as a technique for predicting collisions based on traffic conflict data. Ali et al. (2022) applied the POT approach with GP distribution to assess crash risk during mandatory lane-changing in connected vehicle environments. Their study demonstrated that the POT approach outperformed the BM method in estimating crash risk for connected environments, providing tighter CI and better predictive capability. In another study, Fu and Sayed (2023b) proposed a dynamic Bayesian hierarchical POT modeling approach to estimate real-time crash risk based on traffic conflicts. The authors compared five dynamic updating approaches and applied the model to estimate cycle-level crash risk at signalized intersections. Their results demonstrated that dynamic models outperformed static models in terms of statistical fit and predictive performance.

In addition, some researchers have explored alternative methodologies, including machine learning algorithms, to enhance the sampling procedure in the POT approach. For example, Hussain et al. (2022) introduced an innovative approach that merges machine learning-based anomaly detection techniques with EVT to pinpoint severe traffic conflicts. The author stated this method serves as an alternative to traditional sampling techniques such as BM and POT, yielding more accurate crash predictions than conventional extreme value models.

One of the primary challenges in the POT approach is selecting an appropriate threshold to ensure that the fitted distribution of extreme cases optimally follows the GP distribution. To the best of the authors' knowledge, all studies within the EVT framework in the domain of road safety analysis that require threshold identification continue to rely on graphical techniques, such as MRL plots and threshold stability plots, for determining the threshold of extreme events (Zheng et al., 2015; Guo et al., 2020; Arun, Haque, et al., 2021a; Ali et al., 2022; Hussain et al., 2024b).

It is crucial to note that these graphical threshold selection methods are not definitive and involve a degree of subjectivity. Consequently, the chosen thresholds may vary based on individual interpretations, even within the same study, potentially introducing subjectivity and thus uncertainty into the results. This variability in threshold selection underscores the need for more robust and consistent methods in extreme event analysis (Beirlant et al., 2006).

In this regard, there is only few studies in the literature that has attempted to evaluate the impact of threshold selection on crash prediction in EVT models. Fu and Sayed (2022b) investigated the transferability of multivariate POT models for estimating crash frequency by severity at signalized intersections. They proposed two transferability approaches: an uncalibrated approach and a threshold calibration approach. The study found that the threshold calibration approach provided the most accurate and precise predictions of crash frequency by severity. The authors concluded that for POT models, the threshold parameter is the most critical, and its calibration significantly improves the performance of transferred base models.

Although using more effective techniques for threshold selection in the topic of EVT model application in road safety is a limitation, in the realm of statistics, researchers have utilized a variety of techniques to determine the optimal threshold for EVT models. While graphical methods have been popular, there has been a growing interest in developing automated selection procedures to overcome the inherent subjectivity of visual tools and facilitate the analysis of large datasets.

One such approach, proposed by Thompson et al. (2009), relies on the approximate normal distribution of the differences in parameter estimates at different thresholds. Another method, introduced by Langousis et al. (2016), involves fitting linear models to points above each observation in the standard MRL plot using the weighted least squares technique. The threshold is then chosen as the lowest level that corresponds to a local minimum of the weighted mean squared error of the linear fit, and the GP distribution parameters are estimated based on this threshold.

Some approaches are rooted in assessing the goodness-of-fit of the fitted distribution to the tail data (Scarrott & MacDonald, 2012; Bader et al., 2018; Bladt et al., 2020; Schneider et al., 2021). Bader et al. (2018) developed an automated approach to threshold selection that integrates the research of Choulakian and Stephens (2001) on goodness-of-fit testing under the

GP distribution assumption with the work of G'Sell et al. (2016) on stopping rules for controlling the false discovery rate in ordered hypotheses. This method combines the strengths of both approaches, utilizing goodness-of-fit testing to assess the appropriateness of the GP distribution model at various thresholds and employing stopping rules to manage the risk of false discoveries when considering multiple hypotheses in a specific order. Another suggestion put forth by other studies involves utilizing a combination of a GP distribution for the tail and another distribution for the "bulk," connected at the threshold (MacDonald et al., 2011; Naveau et al., 2016).

Silva Lomba and Fraga Alves (2020) have proposed an innovative, automatic, and computationally efficient approach for selecting a single threshold within the POT generalized Pareto (POT-GP) distribution framework. While inspired by a visual diagnostics tool from the well-established theory of L-moments, their method is not restricted to this tool alone. The approach leverages an alternative characterization of the GP distribution based on its skewness and kurtosis properties, offering a novel perspective on threshold selection in the POT-GP distribution methodology.

In general, the presence of uncertainty in statistical models is a natural attribute arising from various factors. Consequently, this uncertainty can occasionally result in incorrect negative or positive outcomes (Aminghafouri & Fu, 2023). This study addresses the uncertainty caused by threshold selection by considering the threshold as a parameter to estimate.

5.2 Methodology

5.2.1 Traffic Conflict Indicator

The primary indicators of traffic conflicts are evasive actions and close proximity. Various indicators have been identified in the literature, such as TTC, PET, and DRAC. In this research, one commonly used measure, PET, is utilized as the conflict measure to estimate the frequency of crashes at a specific location using EVT. Note that the proposed methodology is equally applicable with other conflict measures.

PET is defined as the "Time difference between the moment an offending vehicle exits the potential collision area and the moment the conflicted vehicle with the right of way arrives at

the potential collision point" (Allen et al., 1978). To determine PET, observed position and velocity values of the road users at each time point (t) should be used (Ismail et al., 2009; Asmussen, 2012). PET is defined in **Equation (5-1)**:

$$PET = ||\max\{T_1\} - \min\{T_0\}|| \quad (5-1)$$

PET is determined by evaluating the instances when road users occupy points in the set P_0 , which represents the points where the observed trajectories of road users intersect. Meanwhile, T_0 denotes the corresponding times at which these instances occur. Additionally, T_1 corresponds to the times when the observed positions of vehicles pass through the conflict point, as defined by Ismail et al. (2009). Essentially, PET is a measure that quantifies the time interval between two road users passing through a potential conflict point. The probability of a collision occurring between two vehicles increases with a decrease in both temporal and spatial proximity. Consequently, a lower PET value indicates a higher likelihood of a collision taking place.

5.2.2 Extreme Value Theory Model

This study utilizes the POT approach of the EVT model to estimate crash frequency. The process involves first identifying a specific threshold value, for example, $PET = 1.0$ sec, and then categorizing the observations that surpass or fall below this threshold as extreme values. By using the POT data, the subsequent analysis can focus on two aspects: either the frequency of extreme events during a specific time period or the magnitude of the exceedances. The former can be modeled as a count distribution, similar to how crashes are typically analyzed, while the latter, which deals with the threshold exceedances, is commonly modeled using the GP distribution (i.e., $G(y)$) (**Equation (5-2)**).

$$G(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma_u}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma_u}\right), & \xi = 0 \end{cases} \quad (5-2)$$

Where y is random variable, ξ is shape parameter, and σ_u is the scale parameter at threshold u . For conflict analysis, a conflict measures of interest (i.e., PET) is first selected and the best-fit distribution can then be determined following the standard statistical process. Once the distribution model is calibrated, the risk of a crash will be computed as **Equation (5-3)** (Zheng, Sayed, et al., 2021):

$$R = \Pr(Z \geq 0) = 1 - G(0) \quad (5-3)$$

where R is the crash risk probability, Z is the negated PET, and $G(\cdot)$ is the EVT distribution calculated based on either the GEV distribution or the GP distribution. With the crash risk probability (i.e., R), the expected number of crashes can be estimated by assuming that the traffic conflict observation period t is representative of a long period T and calculating the estimated crashes N_T , as shown in **Equation (5-4)** (Zheng, Sayed, et al., 2021):

$$N_T = \frac{T}{t} R \quad (5-4)$$

where N_T is the expected number of crashes, t is the traffic conflict observation period, and T is the time period for predicting the number of crashes.

5.2.3 Methods for Threshold Selection

As previously mentioned, the literature presents two primary methods for selecting the threshold to construct the EVT model using the POT approach: 1) Graphic Methods; 2) Automated Threshold Selection Methods.

5.2.3.1 Graphic Method

1. MRL Plot:

The MRL plot is a widely used technique in various fields for estimating the threshold in EVT. This approach involves selecting a threshold based on the mean of exceedances plotted against the threshold value (Davison & Smith, 1990). In this regard, by fitting the GP distribution

model on the data over the threshold, the MRL plot, i.e., , the mean of exceedances ($e_n(u)$) as a function of u , can be drawn based on the **Equation (5-5)**:

$$e_n(u) = E(x - u | x > u) = \frac{\sigma_u + \xi_u}{1 - \xi}, \xi < 1 \quad (5-5)$$

where u , ξ , and σ_u denote the threshold, the shape parameter, and the scale parameter, respectively. Typically, the optimal threshold can be identified as the value of u where the plot forms is almost linear (Yang et al., 2018).

2. Threshold Stability Plot:

A threshold stability plot is another graphical tool that helps to choose a suitable threshold for modelling extreme events using GP distribution. It plots the estimates of the shape and scale parameters of the GP distribution against different thresholds or the number of exceedances and looks for a region where they are stable. The logic behind this method is that if a GP distribution is appropriate for modelling values exceeding a threshold u_0 , then it should also be appropriate for values exceeding any threshold $u > u_0$, and the two distributions will have the same shape parameter. In this regard, the set of data for the plot can be calculated by the following **Equation (5-6)**:

$$\sigma_u = \sigma_{u_0} + \xi(u - u_0) \quad (5-6)$$

where σ_u and ξ denote the scale parameter, and shape parameter of the fitted GP distribution on the data over the threshold u , respectively (Zheng et al., 2016).

By using both approaches, it is possible to identify two sets of thresholds: one where the MRL plot appears almost linear, and another where both the shape parameter and modified scale parameter remain almost constant. The ultimate threshold range is derived by identifying the overlapping area between these two ranges. The selection process takes into consideration lower Akaike information criterion (AIC) values and higher prediction accuracy specific to their corresponding GP distribution models (Zheng et al., 2016).

5.2.3.2 Automated Threshold Selection Method

Two automated threshold selection approaches are examined in this study, which are based on a series of goodness-of-fit tests with error control from Bader et al. (2018) as well as l-moment approach from Silva Lomba and Fraga Alves (2020), as detailed in the following section.

1. Sequential Goodness-of-fit Selection Method (SGFSM):

We consider a scenario where we have a random sample, $X_1, \dots, X_n$, of size n . If u is sufficiently high, we define exceedances, $Y_i = X_i - u$, for all i such that $X_i > u$. These exceedances are assumed to approximately follow a GP distribution. The SGFSM aims to identify the lowest threshold for which the GP distribution fits the exceedances satisfactorily. This approach is chosen to maximize the amount of data used in the analysis while still ensuring a good fit to the extreme value distribution. By using the lowest acceptable threshold, we can include more events in our analysis, potentially capturing a wider range of traffic conflicts and increasing the statistical power of our model, all while maintaining the statistical validity required for extreme value analysis. To achieve this, we adopt the two-stage approach proposed by Bader et al. (2018). This approach involves sequential goodness-of-fit testing with adjustments for multiple testing. It should be noted again that, in the context of traffic conflict analysis, the negated traffic conflict measures (e.g., $-PET$) should be considered to fit into the convention EVT problem definition (POT or exceedance events).

In the first stage, we perform model specification testing for the GP distribution on the exceedances at different candidate thresholds in increasing order. The second stage deals with multiple testing in this ordered setting, which will be handled using the stopping rules proposed by G'Sell et al. (2016).

Let consider a set of candidate thresholds, $u_1 < \dots < u_l$; for each threshold, we have n_i exceedances, $i = 1, \dots, l$. The sequence of null hypotheses can be formulated as follows:

$H_0^{(i)}$: *The distribution of the n_i exceedances above u_i follows the GP distribution*

There are various tests available for each fixed u_i to test $H_0^{(i)}$. We can automate the procedure by starting at the lowest threshold, u_1 , and proceeding until a threshold u_i provides an acceptance of $H_0^{(i)}$. However, the issue is that if the statistical test is not sufficiently powerful,

it may erroneously accept the null hypothesis at a low threshold due to chance. As another way, if we start at the highest threshold, i.e., u_l , and descend until a rejection occurs, the type I error rate would increase.

Addressing the multiple testing issue in this ordered setting is challenging. To tackle this problem, we adopt the ForwardStop rule proposed by G'Sell et al. (2016) for the sequential testing of ordered null hypotheses $H_0^{(1)}, \dots, H_0^{(l)}$. The rule involves transforming the sequence of p-values (i.e., p_i) associated with the hypotheses into a monotone sequence (i.e., Z_i) using **Equations (5-7), (5-8), and (5-9)** and applying the method of Benjamini and Hochberg (1995) on the monotone sequence. The rejection rule is then determined by finding a cut-off $\hat{k}$ such that $H_0^{(1)}, \dots, H_0^{(k)}$ are rejected (i.e., **Equation (5-10)**). If no such $\hat{k}$ exists, no rejection is made. The ForwardStop rule controls the false discovery rate at a prespecified level α , assuming independence among the tests. Thus, for each $H_0^{(i)}$, ForwardStop involves computing a modified average of the preceding and current p-values:

$$Y_i = -\log(1 - p_i) \quad (5-7)$$

$$Z_i = \sum_{j=1}^i \frac{Y_j}{l - j + 1} \quad (5-8)$$

$$q_i = 1 - \exp(-Z_i) \quad (5-9)$$

$$\hat{k} = \max \left\{ k \in \{1, \dots, l\} : q_k \leq \frac{\alpha k}{l} \right\} \quad (5-10)$$

In the context of this study, stopping at a specific threshold k indicates that the goodness-of-fit of the GP distribution is rejected for the exceedances at the initial k thresholds $\{u_1, \dots, u_k\}$. Consequently, the first k null hypotheses $\{H_0^{(1)}, \dots, H_0^{(k)}\}$ are not accepted. By combining the ForwardStop method with sequential hypothesis testing, an automated threshold selection procedure is obtained. This approach requires only a predetermined set of thresholds and the desired level of error control, denoted by α . To prevent subjectivity related to selecting a

predetermined set of thresholds, our proposed method considers all potential thresholds greater than a reasonably small value.

The automated process described can be applied using any suitable test for each hypothesis $H_0^{(i)}$ associated with the threshold u_i . In this study, the AD test, originally proposed by Stephens (1974), is employed. The AD test is designed to assess whether a given data sample is drawn from a population following a specified distribution. It can be seen as a modified version of the Kolmogorov-Smirnov (K-S) test, with a greater emphasis on the tails of the distribution. Let $\hat{\theta}_n$ denote the maximum likelihood estimator of θ under the hypothesis $H_0^{(i)}$, which is based on the observed exceedances. The probability integral transformation is then applied using $\hat{\theta}_n$, represented as $z_i = F(y_i | \hat{\theta}_n)$ according to **Equation (5-2)**, for the order statistics of the exceedances $y_1 < \dots < y_n$. The AD statistic is subsequently calculated using **Equation (5-11)**:

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^k (2i - 1) [\log(z_i) + \log(1 - z_{n+1-i})] \quad (5-11)$$

In the AD test, the asymptotic distributions of the statistic A_n^2 are expressed as a summation of weighted chi-squared variables, each with one degree of freedom. The weights for these chi-squared variables are calculated from the eigenvalues of an integral equation, as proposed by Choulakian and Stephens (2001). Notably, these distributions are entirely determined by the estimate of shape parameter, i.e., ξ . The tests are frequently applied by referring to a table of upper tail percentiles of asymptotic distributions or by using the bootstrap method. In either scenario, the p-values are reduced by a lower bound, which might weaken the ForwardStop rule described in **Equation (5-10)**.

To address these limitations and enable the application of these tests in an automated sequential setting, more accurate p-values are needed. Two solutions are proposed by Bader et al. (2018) to enhance the existing table in Choulakian and Stephens (2001). First, for ξ values in the range of $(-0.5, 1)$, which is applicable for most cases, the table is expanded to a much finer resolution using a pre-set Monte Carlo method. To approximate the asymptotic distributions of A_n^2 , the authors have generated 2,000,000 replicates for each ξ value within the range of

-0.5 to 1 (with an increment of 0.01). The replicates are based on a sample size of $n = 1000$. For each ξ value, a table of upper percentiles ranging from 0.999 to 0.001 is created and stored for future reference. Subsequently, when the estimate $\hat{\xi}_n$ and the test statistic are within the range covered by the table, the p-value is computed using interpolation on a logarithmic scale. This involves finding the distance between the two closest critical values in the table. The interpolation method significantly improves the accuracy of p-values and enhances the reliability of the statistical tests based on A_n^2 , particularly in automated sequential settings.

The second approach to resolving the issue of observed test statistics surpassing the tabulated values, which implies a p-value below 0.001 , is based on the authors' observation that the tails of the asymptotic distributions follow an exponential pattern. This pattern can be confirmed by examining the tail values provided in the table. The authors suggest conducting a linear regression analysis between the upper tail percentiles listed in the table, spanning from 0.05 to 0.001 , and the corresponding $-\log(\text{p-value})$ for a given $\hat{\xi}_n$. This method allows for the estimation of p-values even when the observed test statistics exceed the tabulated values, by extrapolating the exponential behavior of the distribution tails. By extrapolating this linear model, it becomes possible to estimate the p-value for observed statistics beyond the table's range. This estimation of extremely small p-values is valuable for mitigating power loss in the stopping rules.

2. Automatic L-moment Ratio Selection Method (ALRSM):

Drawing inspiration from the Automatic L-moment Ratio Selection Method (ALRSM) proposed by Silva Lomba and Fraga Alves (2020), we present an automated threshold selection approach tailored for road safety applications. The method builds upon the visual methodology, the L-moment ratio diagram (LMRD). The proposed ALRSM introduces several modifications to the LMRD procedure to make it fully automatic and easily applicable by practitioners without the need for subjective judgment.

L-moment Theory: Greenwood et al. (1979) introduced probability weighted moments (PWM) as an alternative approach to traditional moments for characterizing probability distributions, estimating parameters, and conducting hypothesis tests. However, the interpretation of these quantities in terms of distributional properties proved challenging. Hosking et al. (1985)

discovered that by combining PWM in specific linear combinations, known as L-moments, this issue could be resolved, and the resulting measures could be interpreted as location, scale, and shape parameters of probability distributions while preserving the desirable properties of PWM.

L-moments offer a superior approach to PWM-based analysis, providing enhanced accuracy and interpretability, and facilitating the description, identification, and parameter estimation of distributions. The existence and uniqueness of the complete set of L-moments for a random variable X with distribution function F require only that $E|X| < \infty$, encompassing a broader range of probability distributions compared to conventional moments (Hosking, 2006). The first four L-moments can be expressed as linear combinations of specific PWM, denoted as $a_r = M_{1,0,r} = E[X\{1 - F(X)\}^r]$ as **Equation (5-12)**:

$$\begin{aligned}\lambda_1 &= \alpha_0 && \text{L-location (expected value)} && (5-12) \\ \lambda_2 &= \alpha_0 - 2a_1 && \text{L-scale} \\ \lambda_3 &= \alpha_0 - 6a_1 + 6a_2 && \text{third L-moment} \\ \lambda_4 &= \alpha_0 - 12a_1 + 30a_2 - 20a_3 && \text{fourth L-moment}\end{aligned}$$

And dimensionless versions, called L-moment ratios, can be defined as $\tau_r = \lambda_r / \lambda_2$, $r = 3, 4, \dots$, providing measures of distributional shape independent of scale. These L-moment ratios, namely L-skewness (τ_3) and L-kurtosis (τ_4), satisfy the condition $|\tau_r| < 1$.

The general upper bound for L-kurtosis in terms of L-skewness, applicable to any distribution with a finite expected value, is **Equation (5-13)**:

$$\frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1 \quad (5-13)$$

Hosking et al. (1985) has computed these quantities for several important distributions, including the GP distribution, which is defined for $\xi < 1$ with the L-moments in **Equation (5-14)**:

$$\lambda_1 = \frac{\sigma_u}{1 - \xi} \quad (5-14)$$

$$\lambda_2 = \frac{\sigma_u}{(1 - \xi)(2 - \xi)}$$

$$\tau_3 = \frac{1 + \xi}{3 - \xi}$$

$$\tau_4 = \tau_3 \frac{1 + 5\tau_3}{5 + \tau_3}$$

The relationship between L-skewness and L-kurtosis for the GP distribution suggests an alternative method for directly estimating its parameters (Hosking & Wallis, 1987). Furthermore, the L-moments for a location-shifted GP distribution, $X^* = \mu + X$, remain unchanged, except for the L-location, which becomes $\lambda_1^* = \mu + \lambda_1$.

Estimating L-moments involves simple linear combinations of ordered sample elements, with advantages over traditional moments outlined by Hosking and Wallis (1997). These advantages include reduced bias and increased robustness against sampling variability or measurement errors, particularly in extreme data values. The unbiased estimators for the PWM, a_r , are given by **Equation (5-15)**:

$$a_r = \frac{1}{n} \sum_{i=1}^n \binom{n-i}{r} x_{i:n} \binom{n-1}{r}^{-1}, \quad r = 0, 1, \dots, n-1 \quad (5-15)$$

The corresponding unbiased estimators for the first four L-moments are as **Equation (5-16)**:

$$l_1 = \alpha_0 \quad \text{sample mean} \quad (5-16)$$

$$l_2 = \alpha_0 - 2a_1 \quad \text{sample L-scale}$$

$$l_3 = \alpha_0 - 6a_1 + 6a_2 \quad \text{third L-statistic}$$

$$l_4 = \alpha_0 - 12a_1 + 30a_2 - 20a_3 \quad \text{fourth L-statistic}$$

These estimators, termed L-statistics, are discussed further by Hosking and Balakrishnan (2015) in terms of their uniqueness and computational methods. The first two L-moment ratios, t_3 and t_4 , are asymptotically unbiased and are calculated as **Equation (5-17)**:

$$t_3 = \frac{l_3}{l_2} \quad \text{sample L-skewness} \quad (5-17)$$

$$t_4 = \frac{l_4}{l_2} \quad \text{sample L-kurtosis}$$

These ratios are not constrained algebraically by sample size, allowing them to occasionally challenge theoretical bounds, although this is rare in practice. If the underlying distribution has finite variance, the asymptotic normality of these estimators can be demonstrated, along with their asymptotic bias and variance calculations. These estimators often approach their theoretical performance even in relatively small samples, as noted by Hosking et al. (1985).

Automated selection method: The LMRD is a robust and visually intuitive tool that enables researchers to compare the characteristics of various distributions by plotting their L-moment ratios on a single graph. This innovative approach offers a comprehensive overview of the similarities and differences among the studied distributions, facilitating the identification of key patterns and trends. By meticulously analyzing the positions and arrangements of the plotted points, researchers can gain valuable insights into the intrinsic properties of each distribution and make well-informed decisions regarding the most suitable model for their specific dataset. The LMRD proves to be an indispensable asset in statistical analysis, streamlining the process of evaluating and selecting the optimal distribution for a given context, thereby enhancing the accuracy and reliability of the results. For a more in-depth understanding of the creation and interpretation of LMRDs, you can refer to the seminal work by Hosking and Wallis (1997), which provides a comprehensive and authoritative guide to this powerful analytical tool.

In the domain of threshold diagnostics, the LMRD has found extensive application and has proven to be particularly effective. The `lmomplot` function, a key component of the R package POT described by Ribatet (2011), offers a sophisticated visual implementation of this methodology. The function takes an input data array x and computes the estimates $(t_{3,u}, t_{4,u})$ for the sample of excesses above a range of potential thresholds. By default, the function

evaluates a minimum of 50 candidate thresholds that are equally distributed across the span of the original data x . This approach ensures that the analysis is thorough and representative of the entire dataset, allowing for a comprehensive assessment of the excesses and their associated estimates. The function automatically selects the candidate thresholds based on the input data, eliminating the need for manual specification and facilitating a systematic exploration of the threshold space. The code also incorporates an advanced identify option, allowing users to directly identify the threshold associated with each estimate point within the plot, thereby facilitating the interpretation of the results. The user is then tasked with visually determining which L-statistics pair most closely aligns with the GP distribution theoretical line, defined by **Equation (5-14)**, with the objective of selecting a threshold for which the behavior of these estimators for the excesses' sample closely approximates the expected behavior for the GP distribution. However, as astutely noted by the author, this technique often yielded suboptimal results when applied to real-world data, highlighting the need for further refinement and automation.

To address this issue and enhance the practicality and reliability of the methodology, we propose an innovative enhancement to the procedure, aiming to automate the process and facilitate its application by practitioners to any dataset without the need for subjective judgment. These modifications are designed to improve the accuracy and consistency of the results, while also reducing the potential for human error and bias. Given a sample $x_1; \dots; x_n$ of size n , the ALRSM operates as follows:

1. First, a set of candidate thresholds $(\{u_i\}_{i=1}^I)$ is defined, representing the range of potential threshold values to be considered in the analysis.
2. For each candidate threshold u_i , the sample L-skewness (t_{3,u_i}) and L-kurtosis (t_{4,u_i}) are computed for the excesses over that specific threshold. Then, the minimum Euclidean distance d_{u_i} between each point (t_{3,u_i}, t_{4,u_i}) and the GP distribution theoretical curve based on **Equation (5-14)** is determined using **Equation (5-18)**:

$$d_{u_i} = \min_{\tau_3} \sqrt{(t_{3,u_i} - \tau_3)^2 + (t_{4,u_i} - \tau_4)^2}, \quad \text{for } i = 1, \dots, I \quad (5-18)$$

3. The optimal threshold u^* is automatically determined by selecting the level at which the corresponding L-statistics are closest to the theoretical curve of the GP distribution. This is achieved by finding the threshold u_i that minimizes the distance d_{u_i} across all candidate thresholds as **Equation (5-19)**:

$$u^* = \arg \min_{u_i, 1 \leq i \leq I} \{d_{u_i}\} \quad (5-19)$$

The main differences between the ALRSM and existing LMRD approach lie in how the candidate thresholds are defined and how the closeness of estimates to theoretical values is evaluated mathematically, thus enabling an automatic selection process.

To ensure the accuracy of the approximation to the GP distribution, the ALRSM concentrates on thresholds that exceed the 25th percentile of the sample. This approach prevents the selection of levels that are frequently too low for a satisfactory approximation. Additionally, the method does not consider extremely high levels that surpass the 95th percentile of the sample. By excluding these levels, the ALRSM guarantees that an adequate number of excesses are accessible, allowing for a reliable estimation of the L-statistics and preserving the integrity of the outcomes.

5.3 Simulation Study

5.3.1 Simulation Settings:

In order to evaluate the effectiveness and performance of the proposed automated threshold selection method, a comprehensive simulation study was carried out. The compared methods are based on the same principle: choosing the most suitable threshold from a set of candidate thresholds $u_1 \leq \dots \leq u_I$ for a POT-GP distribution analysis.

To evaluate the effectiveness of the SGFSM and ALRSM techniques, we mainly used random samples taken from a Hybrid distribution, which was first described by Northrop and Coleman (2014). The key benefit of employing this distribution is that we know the exact threshold value that should be returned by the method, since the GP distribution assumption is valid

beyond u . This characteristic is especially helpful for confirming the validity of the suggested approach.

The probability density function (PDF) of the $Hybrid(u, \xi)$ distribution is defined continuously for real numbers greater than zero. It consists of a constant function up to the actual threshold u , which falls between 0 and 1, adhering to a $Uniform(0,1)$ density until it reaches the $(u \times 100\%)^{th}$ percentile. Beyond this threshold, the distribution follows a GP distribution density with a shape parameter ξ ranging from -1 to 1 , and a scale parameter σ_u equal to $1 - u$. The PDF is given by **Equation (5-20)**:

$$Hybrid(x|u, \xi) = \begin{cases} 1, & 0 < x < u \\ \left(1 + \xi \frac{x - u}{1 - u}\right)^{-\frac{1}{\xi} - 1}, & u \leq x < x^{**} \end{cases} \quad (5-20)$$

where x^{**} represents the distribution's right endpoint, defined as **Equation (5-21)**:

$$\begin{aligned} x^{**} &:= \sup \left\{ x: \int_0^x Hybrid(t|u, \xi) dt < 1 \right\} \\ &= \begin{cases} \frac{(1 + \xi)u - 1}{\xi}, & for \xi < 0 \\ +\infty, & for \xi > 0 \end{cases} \end{aligned} \quad (5-21)$$

As an example, **Figure 5-1** shows the PDF for a Hybrid distribution. In this example, the shape parameter (ξ) is set to 0.2, and the actual threshold (u) is 0.75.

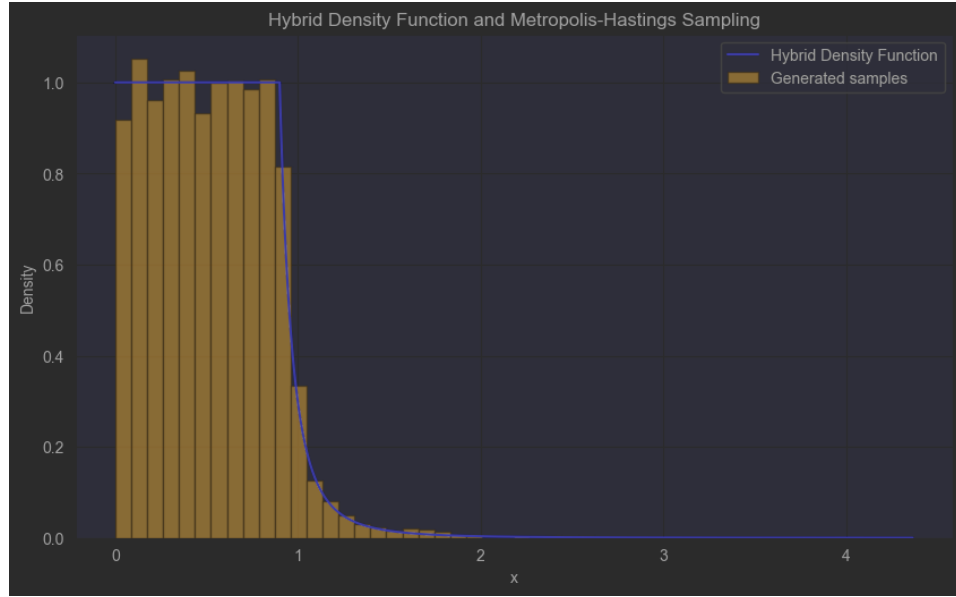


Figure 5-1 Hybrid Density Function Example

Both proposed selection methods were applied to each sample. After choosing the "optimal" level u , a standard POT analysis was performed. Maximum likelihood estimation was used to estimate the parameters of the GP distribution fit to the excess samples across both methodologies, ensuring that comparisons were solely based on the threshold selection process and not the parameter estimation method.

To assess the effectiveness of the methods, mean bias and root mean squared error (RMSE) for the selected threshold and estimated entities were calculated from a set of 1000 samples generated for each combination of the following scenarios:

- GP distribution tail shape parameter: $\xi = 0.2$; $\xi = 0.5$
- True threshold: $u = 0.75$; $u = 0.9$
- Sample size: $n = 1750$; $n = 2500$; $n = 5000$; $n = 10000$

The candidate threshold was selected between the 25% and 95% quantiles, with an increment of 1%.

5.3.2 Simulation Main Results

In this section, we focus on the results from scenarios related to the true threshold $u = 0.90$ and $\xi = 0.5$ for the sake of presentation clarity. The simulation study results are summarized

in **Figure 5-2** and **Figure 5-3**, with blue representing the SGFSM and red representing the ALRSM. The performance of these methods is assessed and compared based on the mean bias error (MBE) and RMSE of threshold selection relative to their true values.

Figure 5-2 reveals that the ALRSM outperforms the SGFSM in terms of mean bias error for smaller sample sizes. However, as the sample size increases to 10000, the SGFSM demonstrates better performance and accurately identifies the threshold equal to the true value. The result of **Figure 5-2** indicates that the ALRSM is typically the better option, as the highest average deviation is around 0.13 in absolute terms. Furthermore, it is evident that a reduction in sample size has a lesser impact on the results from ALRSM, while it significantly increases the mean bias of the threshold for the SGFSM (MBE error of 0.4 in absolute value for a sample size of 1750).

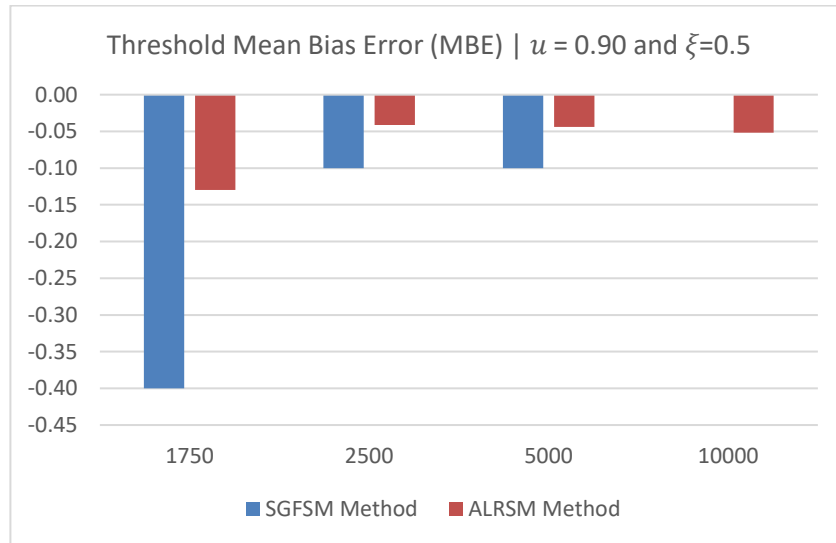


Figure 5-2 SGFSM and ALRSM Threshold Mean Bias Error Comparison

Figure 5-3 illustrates the RMSE of the threshold choice. In terms of this indicator, the ALRSM method also demonstrates more consistent performance, yielding lower RMSE values across all sample sizes.

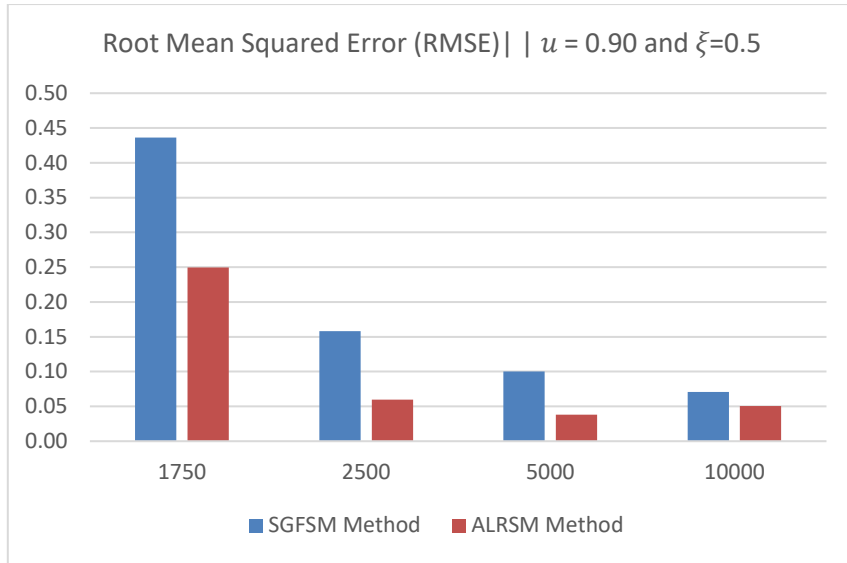


Figure 5-3 SGFMS and ALRSM Threshold Root Mean Squared Error based on 1000 Simulations

Overall, the simulation study results demonstrate the effectiveness of both proposed automated threshold selection methods. A qualitative assessment, based on a visual inspection of the Root Mean Square Error (RMSE) and Mean Bias Error (MBE) for each method, showed that the estimated thresholds were consistently close to the true threshold value. By confirming the validity of the threshold selection methods, we can apply them to real-world road safety cases to predict the number of collisions from collected conflict data using the EVT method.

5.4 Case Study

The proposed methodology is demonstrated through a case study that involves estimating crash frequency using the POT approach of EVT model on traffic conflict data.

5.4.1 Data Collection and Preparation

A comprehensive video data collection study was conducted at ten signalized intersections situated in Toronto, Ontario, Canada, from November 11th to 22nd, 2019, as shown in **Figure 5-4**. The data collection period spanned from 5 am to 10 pm, encompassing 17 hours per day, covering a total of 1,139 hours. These intersections were carefully chosen in collaboration with the City of Toronto officials, ensuring that they shared similar geometric characteristics, traffic patterns, and environmental conditions. Notably, all selected intersections in this research were

signalized with four legs, had a consistent posted speed limit of 50 km/h, and consisted of five lanes in each leg.



Figure 5-4 Selected Intersections

The video data underwent processing using the TrafxSAFE commercial software, provided by Transoft. This enabled the extraction of road user trajectories and conflict data, facilitating conflict analyses. The video data processing procedure comprised four essential steps. First, it involved detecting and tracking the various road users visible in the videos. Next, these detected road users were classified into distinct types, namely, vehicles, pedestrians, and cyclists. Subsequently, interactive road user pairs that engaged in traffic conflicts were carefully selected. Finally, traffic conflict measures were calculated for the extracted traffic conflicts (Shangguan et al., 2023).

Notably, the software recorded essential data for each conflict, encompassing details about the conflicting traffic, such as vehicle type and pedestrian information, along with the PET and speeds involved. To ensure comprehensive coverage of conflicts, the camera locations adhered to specific requirements, including video resolution, field of view, and distance from the intersection.

While other crash types may be more frequent, the focus of this research was specifically on traffic conflicts between right-turning vehicles and pedestrians and cyclists at crosswalks. This

specific focus aligns with the objective of the partner, the City of Toronto, project under the Vision Zero program. Their primary goal was to evaluate the safety of cyclists and pedestrians at the signalized intersections. Hence, the study specifically recorded conflicts where right-turning Vulnerable Road Users (VRUs) had a PET of less than 10 seconds.

Table 5-2 shows the descriptive data for each intersection of the conflict measures used in this study. The crash data shown in the table were critical for validating the crash prediction models created in this study. The study intersections' five-year crash statistics (2017 – 2021) were obtained through the City of Toronto's open data portal. For each recorded crash, the crash data included critical information characteristics such as time, location, severity, road user details, and collision type.

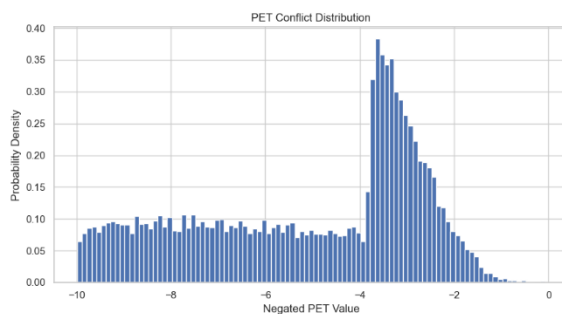
It was critical to filter the crash data to match the parameters under which the conflicts were seen in order to make appropriate comparisons with the crashes predicted by the POT models. Consequently, the crash data were filtered to include only cases that met the following criteria: pedestrian and cyclist crashes occurring between 5 am and 10 pm as a result of right turn movements. This refined dataset allowed for precise assessments and evaluations of the predictive models.

Table 5-2 Descriptive Summary of Traffic Conflict Data

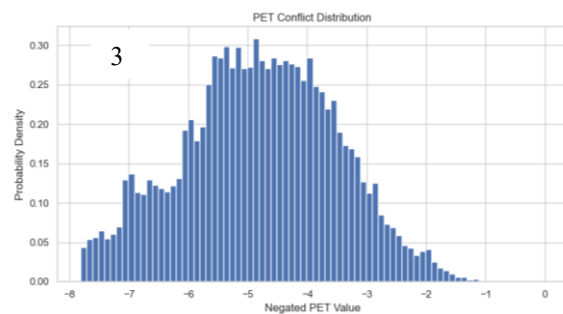
Num.	Intersection name	Data collection	Total duration (hrs)	No. of VRU crashes (2017-21)	No. of VRU conflicts	PET indicator				
						Mean	Median	Std. Dev.	Min.	Max.
1	Bloor St W & Avenue Rd	11/11/2019	136	3	15560	5.11	4.42	2.40	0.1	9.98
		– 22/11/2019								
2	Bloor St W & Jane St	11/11/2019	136	2	12000	4.82	4.8	1.30	0.1	7.81
		– 18/11/2019								
3	Bloor St W & Keele St	11/11/2019	102	2	6804	5.18	5.14	1.57	0	9.98
		– 18/11/2019								

4	Danforth Ave & Pape Ave	11/11/2019	–	102	1	14196	5.55	5.31	1.95	0.13	9.98
		16/11/2019									
5	Eglinton Ave W & Martin Grove Rd	11/11/2019	–	102	2	3134	6.20	6.21	2.16	0.1	9.98
		16/11/2019									
6	King St W & Strachan Ave	13/11/2019	–	119	2	11095	5.30	5.04	1.97	0.07	9.98
		20/11/2019									
7	Lawrence Ave E & Don Mills Rd	11/11/2019	–	102	1	9207	5.69	5.57	1.94	0.3	9.98
		16/11/2019									
8	Sheppard Ave E & Victoria Park Ave	11/11/2019	–	102	3	12790	5.46	5.47	1.77	0.03	9.98
		16/11/2019									
9	Spadina Ave & Dundas St W	11/11/2019	–	136	0	8979	4.67	4.24	1.94	0.03	9.98
		18/11/2019									
10	Yonge St & Finch Ave W	11/11/2019	–	102	1	11623	4.91	4.37	1.97	0.03	9.98
		16/11/2019									

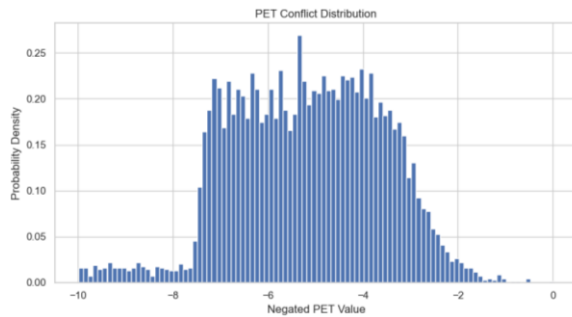
To gain a deeper insight into the distribution of the collected conflict data (negated PET), we have created histogram plots for each location, shown in **Figure 5-5**.



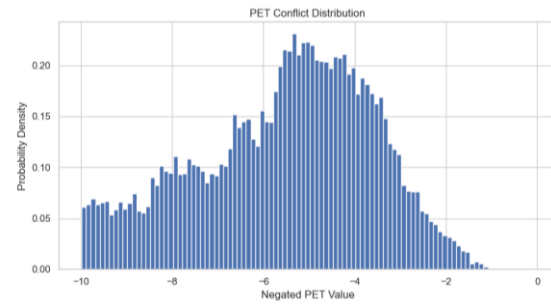
a) Bloor St W & Avenue Rd



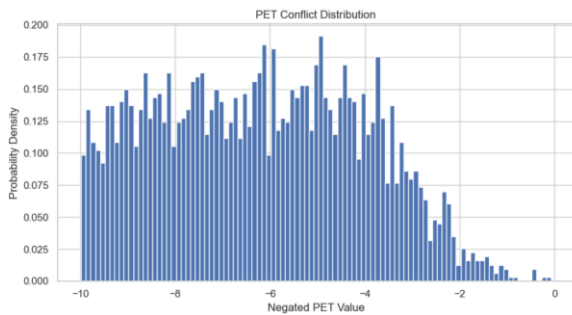
b) Bloor St W & Jane St



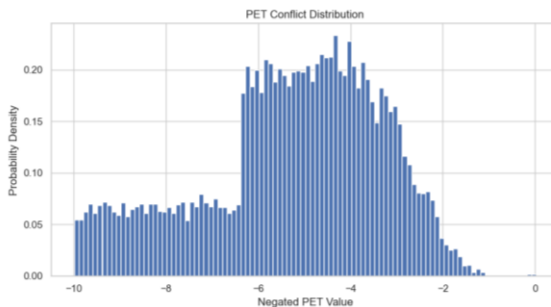
c) Bloor St W & Keele St



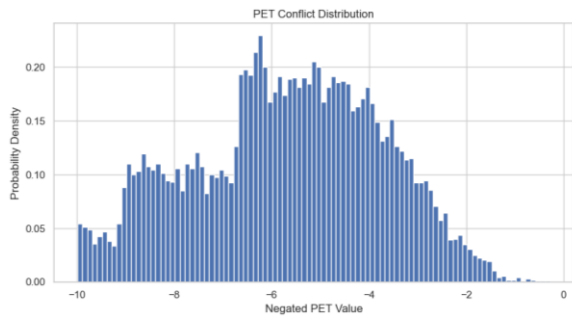
d) Danforth Ave & Pape Ave



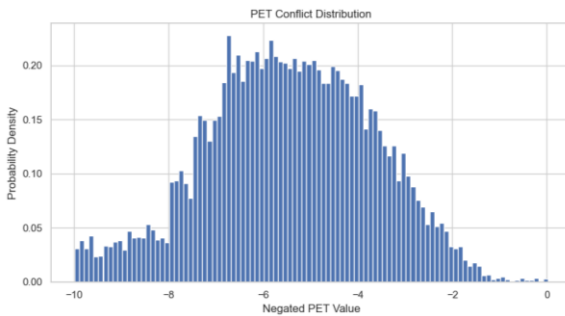
e) Eglinton Ave W & Martin Grove Rd



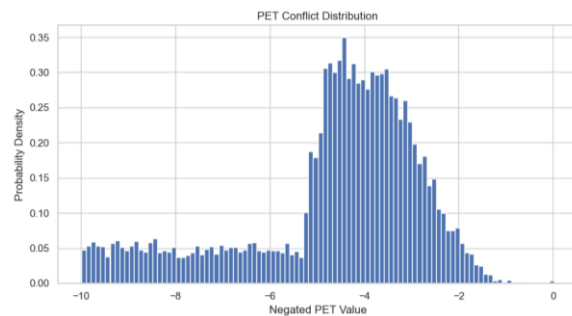
f) King St W & Strachan Ave



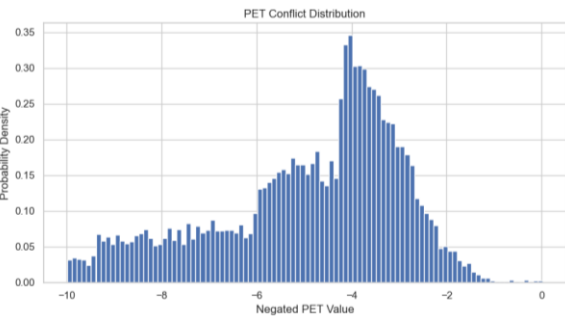
g) Lawrence Ave E & Don Mills Rd



h) Sheppard Ave E & Victoria Park Ave



i) Spadina Ave & Dundas St W



j) Yonge St & Finch Ave W

Figure 5-5 Negated PET Conflict Distribution

5.4.2 Results and Discussions

The threshold determination process involves the utilization of both a graphical method and the suggested automated threshold selection approach. Subsequently, the outcomes derived from these distinct approaches are employed to estimate the number of crashes. A comparative analysis between the estimated crashes and the actual historical crash data is then conducted to assess the performance of each threshold selection method.

To illustrate the threshold identification process, we present a detailed account of the steps undertaken for one of the intersections, i.e., Bloor St W & Avenue Rd, employing both the graphical method and the automated approach. This procedure is subsequently replicated for other selected locations. However, for brevity, only the results of this comprehensive analysis are presented in **Table 5-3**.

5.4.2.1 Graphical Method

Based on the analyses of the MRL plot and the Threshold Stability plot, the appropriate range of thresholds for each intersection is determined. In particular, **Figure 5-6** displays the MRL plot specifically for the Bloor St W & Avenue Rd intersection. According to this plot, the optimal threshold for the negated PET values is projected to lie within the range of $(-3.2, -1.6)$.

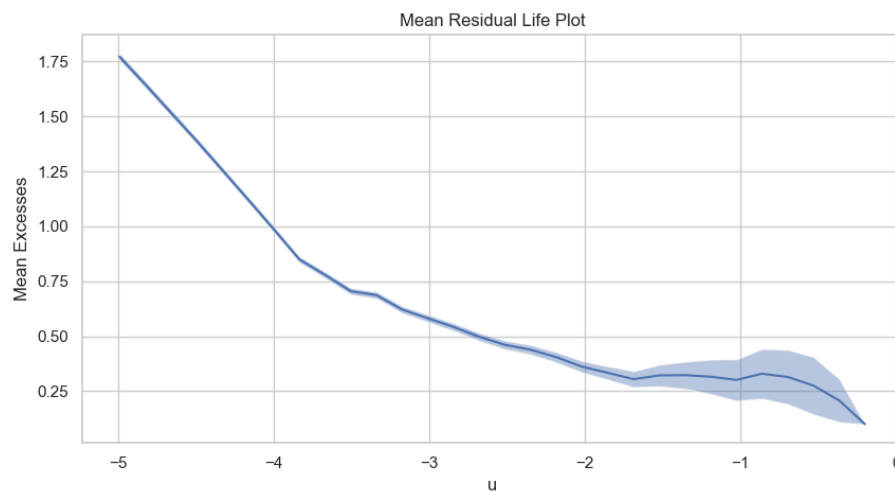


Figure 5-6 MRL Plot with %95 CI

Through the application of the threshold stability plot approach (**Figure 5-7** and **Figure 5-8**), it is observed that the shape parameter remains constant within the range of $(-3, -2)$, while the modified scale parameter exhibits relative constancy within the range of $(-2.7, -2.1)$.

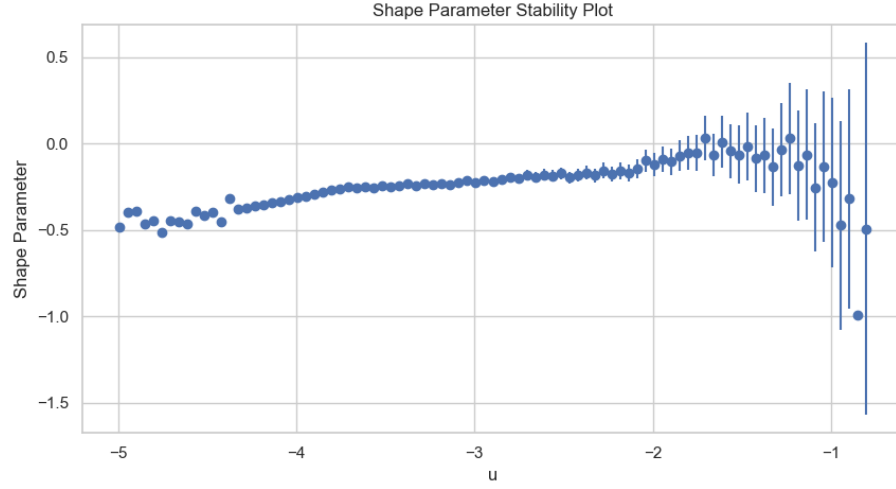


Figure 5-7 Shape Parameter Stability Plot

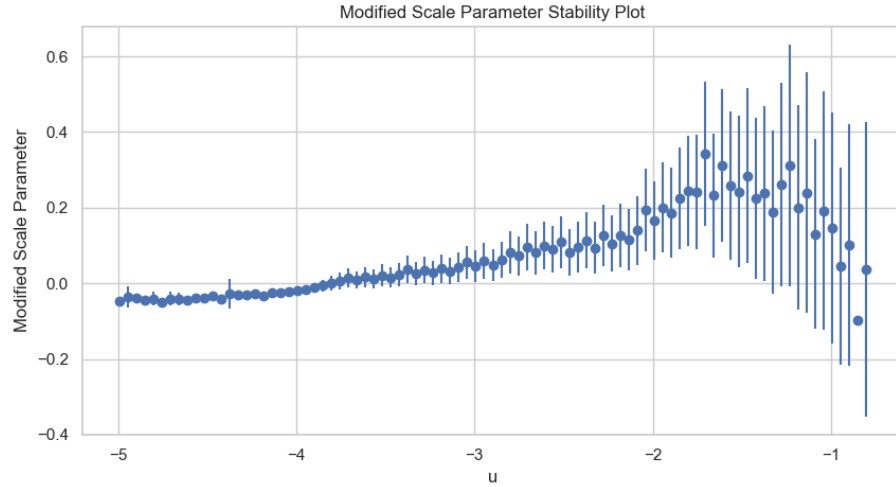


Figure 5-8 Modified Scale Parameter Stability Plot

By intersecting these three ranges, it is deduced that the optimal threshold based on the graphical approach should fall somewhere between -2.7 and -2.1 . To identify the precise optimal threshold within this final range, the GP distribution model is fitted on different thresholds ranging from $(-2.7, -2.1)$. After conducting these fittings, it is found that the

smallest AIC value is attained at threshold value of -2.11 which provides the best fit for the GP distribution model.

5.4.2.2 Sequential Goodness-of-fit Selection Method (SGFSM)

Based on the proposed SGFSM method, the calculation of the AD test is essential for various thresholds which is done using **Equation (5-11)** across different thresholds. Subsequently, in accordance with the outlined methodology, the corresponding p-values for each AD test result are computed.

The selection of the optimal threshold for implementing the POT model of EVT model is accomplished by adopting the ForwardStop selection (i.e., **Equation (5-10)**). Consequently, the threshold that leads to the acceptance of the null hypothesis is identified as the most suitable threshold according to the automated approach. Therefore, for this specific location, the best threshold determined through the automated approach is -1.84 .

5.4.2.3 Automatic L-moment Ratio Selection Method (ALRSM)

As illustrated in **Figure 5-9**, by implementing the ALRSM approach, the minimum Euclidean distance is associated with a threshold value of -1.47 , resulting in a Euclidean distance equal to 0.000202. This finding demonstrates the effectiveness of the ALRSM approach in identifying the optimal threshold value that minimizes the discrepancy between the predicted and observed conflict frequencies.

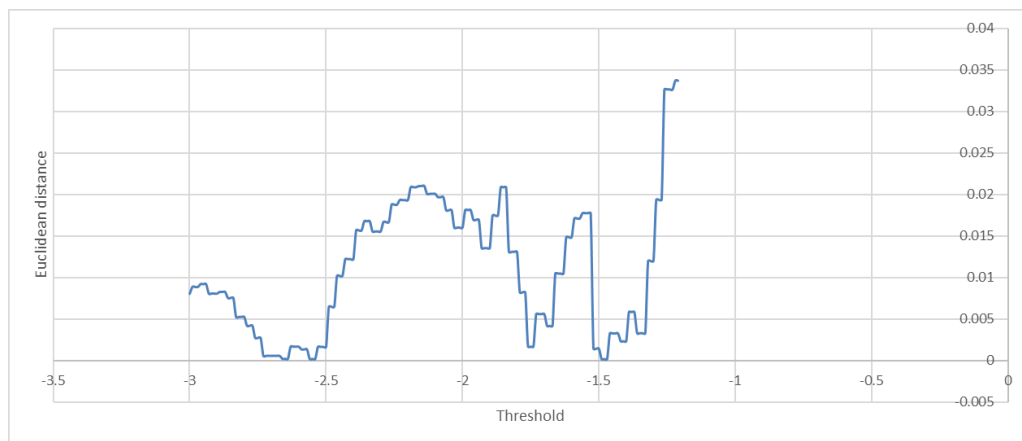


Figure 5-9 Euclidean Distance by Threshold

The validation of the chosen threshold, based on automated threshold selection approaches, can be examined through the plotting of the PDF and CDF of the fitted GP distribution model against the empirical data from the tail. For example, as shown in **Figure 5-10** and **Figure 5-11**, selecting the threshold as -1.47 according to ALRSM results in a well-fitted theoretical PDF that appropriately aligns with the historical data, confirming that the tail data adheres to a GP distribution with the calibrated parameters (i.e., $shape = -0.1239$, and $scale = 0.3752$).

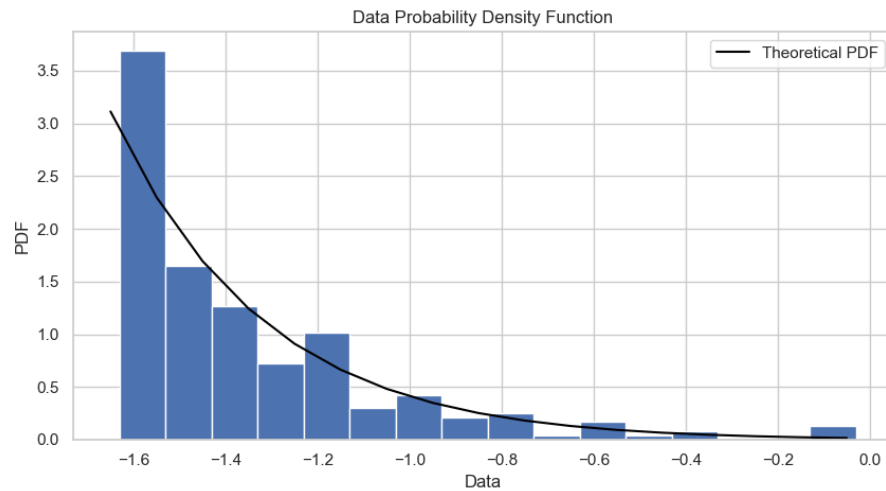


Figure 5-10 Data Probability Density Function

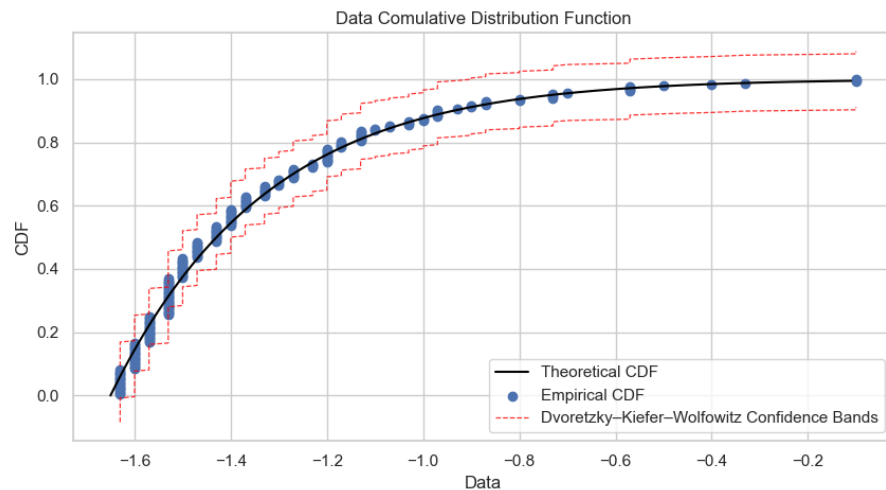


Figure 5-11 Data Cumulative Density Function

Furthermore, an additional means of confirming the accuracy of the fitted model involves employing Probability-Probability (P-P) plots and Quantile-Quantile (Q-Q) plots. By utilizing both P-P and Q-Q plots, we can verify the precision of the calibrated model for extreme conflict events in our study at the specific intersection (i.e., Bloor St W & Avenue Rd.). Based on **Figure 5-12**, these validation techniques further reinforce the reliability of the GP distribution model in accurately capturing extreme conflict events at the given location.

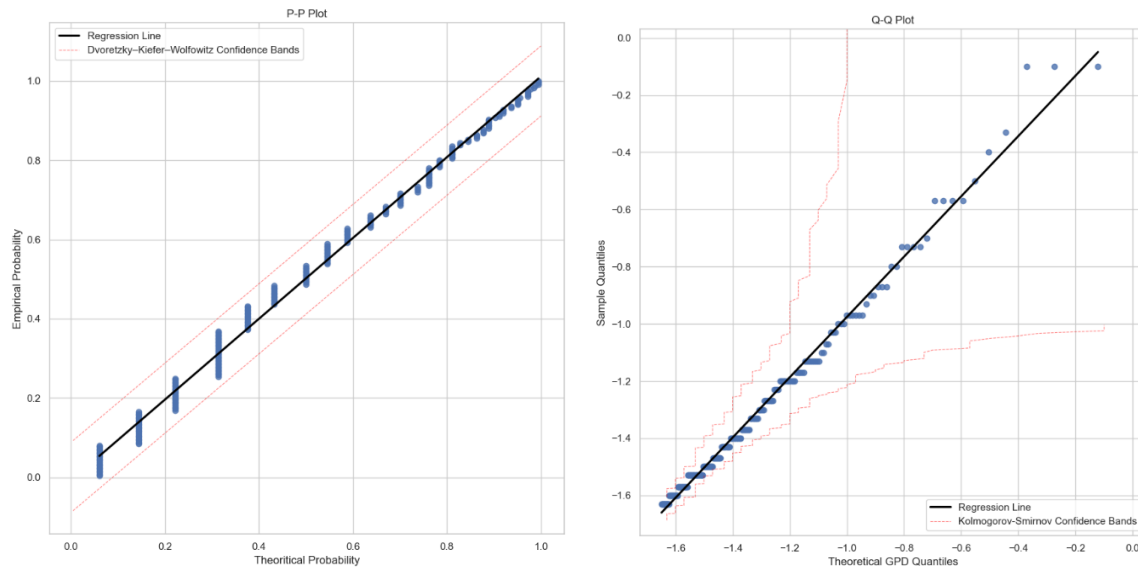


Figure 5-12 P-P and Q-Q Plots

5.4.2.4 Crash Frequency Estimation

Based on **Equation (5-4)**, the estimation for the number of crashes over a duration of 5 years is calculated for each intersection. The results of these estimations are presented in **Table 5-3**.

Table 5-3 Threshold Selection and Crash Estimation

Intersection	Observed crashes per year	95% CI Observed Crash	Graphical method								Sequential Goodness-of-fit Selection Method (SGFSM)								Automatic L-moment Ratio Selection Method (ALRSM)							
			Threshold	Sample size	Fitted GP distribution			Crash est. per year	95% CI Crash Estimate	Threshold	Sample size	Fitted GP distribution			Crash est. per year	95% CI Crash Estimate	Threshold	Sample size	Fitted GP distribution			Crash est. per year	95% CI Crash Estimate			
					Shape	Scale	AIC					Shape	Scale	AIC					Shape	Scale	AIC					
Bloor St W & Avenue Rd	0.4	[0, 1.63]	-2.11	747	-0.13	0.43	44.78	0.03	[0, 0.29]	-1.84	372	-0.13	0.41	-22.59	0.06	[0, 0.69]	-1.47	122	-0.12	0.37	-21.39	0.30	[0, 3.41]			
Bloor St W & Jane St	0.6	[0, 2.11]	-2.48	398	-0.12	0.46	85.26	0.009	[0, 0.21]	-2.54	418	-0.15	0.52	157.46	0.005	[0, 0.17]	-2.86	740	-0.17	0.59	442.40	0.002	[0, 0.07]			
Bloor St W & Keele St	0.6	[0, 2.11]	-2.32	138	-0.09	0.52	74.12	0.27	[0, 3.43]	-2.9	404	-0.11	0.58	291.94	0.07	[0, 0.86]	-2.91	428	-0.07	0.54	279.61	0.11	[0, 0.97]			
Danforth Ave & Pape Ave	0.2	[0, 1.07]	-2.21	321	-0.17	0.43	1.24	8.9E-04	[0, 0.16]	-2.94	947	-0.25	0.73	826.24	1.97E-09	[0, 0.02]	-2.67	663	-0.23	0.62	396.06	8.42E-06	[0, 0.04]			
Eglinton Ave W & Martin Grove Rd	0.8	[0, 2.55]	-1.78	43	-0.19	0.58	26.58	0.82	[0, 13.69]	-2.67	147	-0.24	0.83	172.02	0.18	[0, 3.81]	-2.26	82	-0.22	0.73	78.93	0.42	[0, 7.82]			
King St W & Strachan Ave	0	[0, 0]	-2.41	470	-0.02	0.38	19.76	0.08	[0, 0.59]	-2.14	215	-0.07	0.42	33.14	0.15	[0, 1.26]	-1.74	78	0.15	0.28	-13.37	0.95	[0, 5.30]			
Lawrence Ave E & Don Mills Rd	0.4	[0, 1.63]	-2.21	227	-0.15	0.49	69.58	0.05	[0, 1.49]	-2.54	360	-0.24	0.66	254.04	0.001	[0, 0.56]	-2.76	524	-0.22	0.67	414.92	0.004	[0, 0.41]			
Sheppard Ave E & Victoria Park Ave	1.2	[0, 3.34]	-2.36	437	-0.08	0.56	299.33	0.49	[0.002, 2.2]	-1.84	147	-0.18	0.62	106.03	1.27	[0, 8.26]	-2.1	266	-0.08	0.54	168.78	0.85	[0, 4.05]			
Spadina Ave & Dundas St W	0.6	[0, 2.11]	-1.99	218	-0.02	0.35	-24.7	0.15	[0, 1.23]	-1.86	151	0.009	0.33	-24.26	0.27	[0, 2.04]	-1.69	90	0.07	0.31	-16.22	0.60	[0, 4.05]			
Yonge St & Finch Ave W	0	[0, 0]	-1.93	238	0.01	0.41	63.11	0.90	[0.003, 3.8]	-2.14	346	-0.09	0.51	167.16	0.42	[0.0001, 2.2]	-1.87	207	0.03	0.40	51.40	1.11	[0.002, 4.64]			
Total Mean Squared Error (MSE)			0.40							0.57							0.02									

In this case study, the small sample size for each intersection is expected to result in less accurate predictions of collisions using the EVT model. However, despite the limited sample size, the analysis demonstrates that the ALRSM approach generally exhibits superior performance in estimating the number of crashes, as evidenced by the lower mean squared error (i.e., 0.020). This observation highlights the effectiveness and reliability of the automated threshold selection method in accurately predicting crash occurrences for the majority of the intersections under study.

5.5 Conclusion

This study introduces a novel approach for automating threshold selection – a step that is required in all conflict-based road safety studies. With the most commonly adopted POT approach of the EVT in these studies, two automated threshold selection methods are examined: 1) sequential goodness-of-fit tests to assess the appropriateness of fitting extreme traffic events to a Pareto distribution; 2) l-moment ratio analysis to find the minimum Euclidian distance between predicted l-statistics and theoretical l-moments of the data. The proposed threshold selection methods were evaluated through an extensive simulation analysis and a case study using traffic conflict data from 10 intersections in Toronto, showing their performance differences as well as advantages over the conventional approaches.

The case study has also shown a significant difference in crash estimations resulting from slight variations in the selected thresholds. This finding underscores the importance of employing a reliable threshold selection methodology to achieve precise estimations in EVT models. Consequently, future researchers can benefit from this approach to enhance the accuracy of their studies. In light of these findings, the following research directions can be suggested for future studies:

- While the SGFSM in this study primarily focuses on the AD test, future research can delve into other available statistical tests to evaluate their performance in threshold selection and compare their results against the proposed methodology.
- Expanding the current model to consider crash severity can provide a more comprehensive analysis of traffic safety, enabling researchers to estimate crash

frequency and severity levels jointly, which would be valuable for decision-making and safety planning.

- To address the uncertainty associated with selected thresholds, a Bayesian framework can be developed to incorporate and quantify uncertainties in the threshold selection process, providing a more robust and comprehensive analysis.

By exploring these research directions, the field of road safety study can benefit from enhanced methodologies that offer more accurate and reliable estimations of crash risk, ultimately contributing to improved safety measures and decision-making processes.

The preceding three chapters have systematically addressed critical sources of uncertainty in road safety analysis. Chapter 3 began by quantifying the impact of uncertainty in traditional crash-based network screening. Chapter 4 then pivoted to proactive, conflict-based analysis to tackle a fundamental data uncertainty problem, establishing the extensive data requirements needed for reliable predictions. Finally, this chapter addressed a key source of model uncertainty by developing an automated and objective approach for threshold selection in EVT models.

Together, these chapters form a cohesive framework (see **Figure 1-1**) for understanding, quantifying, and mitigating key uncertainties in road safety analysis. The final chapter will now synthesize the key findings and major contributions from this entire body of work. It will discuss their collective practical implications for road safety practitioners and conclude by outlining promising directions for future research that build upon the foundations established in this thesis.

Chapter 6

Conclusion

This research introduces a comprehensive framework to systematically quantify, analyze, and reduce critical sources of uncertainty in road safety analysis. The work presented in this thesis addresses fundamental challenges in both traditional crash-based and modern conflict-based safety studies, with the overarching goal of supporting more reliable and evidence-based decision-making.

6.1 Summary of Major Studies and Findings

The key findings across the chapters of this thesis are summarized as follows:

- In **Chapter 2**, a comprehensive literature review was conducted to systematically synthesize major studies on the sources of uncertainty in road safety analysis. The review confirmed that risk prediction is subject to numerous uncertainties stemming from data quality, model assumptions, and analytical methods. Key gaps in the existing research were identified, including a prevalent reliance on point estimates for risk assessment which ignores statistical uncertainty, and the unaddressed data and model uncertainties in proactive, conflict-based analysis. This review identified the critical problems that the subsequent research chapters aimed to solve.
- In **Chapter 3**, a novel method was developed to explicitly incorporate uncertainty into the network screening process. By employing a frequentist approach to establish a joint CR for hotspot rankings, the study moved beyond traditional point estimates of risk. The analysis, based on a hierarchical FB model, demonstrated that explicitly accounting for uncertainty can fundamentally alter hotspot identification. The results showed that an intersection's rank is not a single, fixed number but a range of plausible positions, confirming that rankings based on simple point estimates can be statistically unstable and potentially misleading.
- In **Chapter 4**, the research focused on a primary source of data uncertainty in proactive, conflict-based analysis: the impact of sample size on prediction reliability. This study challenged the "myth of quick analysis" by using a unique, year-long LiDAR dataset

to systematically evaluate how collision estimates, and their CI converge with increasing data collection duration. The findings compellingly show that the common practice of using short-term conflict data is inadequate for reliable safety analysis. Several months of continuous monitoring were required to achieve collision predictions with acceptably narrow confidence intervals, establishing an evidence-based methodology for determining the necessary data collection duration.

- In **Chapter 5**, the focus shifted to addressing methodological (epistemic) uncertainty in the application of EVT. An automated and objective approach for threshold selection in POT models was developed and validated. The study compared a SGFSM with an ALRSM, demonstrating through simulations and a real-world case study that an automated approach can significantly reduce analytical subjectivity. The L-moment-based method, in particular, proved to be a robust and objective technique that improves the accuracy of crash risk estimation compared to conventional graphical methods.

6.2 Summary of Major Contributions

This thesis makes several key contributions to the field of road safety analysis by providing a systematic framework for managing uncertainty. The major contributions, which directly align with the research objectives, are:

1. *Established the Theoretical and Methodological Gaps in road safety analysis*: This thesis began by conducting a comprehensive literature review that identified and synthesized the critical sources of uncertainty in both traditional and proactive RSA. This review established the foundational argument for this research by highlighting the prevalent reliance on point estimates and identifying unaddressed data and model uncertainties, thereby clearly defining the gaps that the subsequent chapters aimed to fill.
2. *Developed a Methodology for Incorporating Uncertainty in Network Screening*: This research addressed a critical gap in traditional road safety analysis by creating a method that moves beyond simple point-based rankings. By constructing a joint CR for the rank of each location, this work provides decision-makers with a more robust and

statistically sound understanding of site risk and priority, leading to more reliable hotspot identification.

3. *Quantified the Impact of Data Collection Duration on Prediction Reliability:* This thesis provides the first long-term, empirical investigation into the sample size requirements for conflict-based safety analysis. By challenging the "myth of quick analysis," this work established evidence-based guidelines on the minimum data required to achieve robust collision predictions. This contribution provides practitioners with a clear methodology for balancing the need for reliable results with practical resource constraints.
4. *Developed an Automated Approach for Objective Threshold Selection in EVT Models:* To address a critical source of model uncertainty in proactive safety analysis, this research developed and validated an automated, objective approach for threshold selection in POT models. This work replaces subjective graphical techniques with a repeatable and statistically rigorous procedure, thereby improving the accuracy and consistency of risk estimation using EVT.

Collectively, these contributions form a cohesive framework for understanding, quantifying, and mitigating key uncertainties in road safety analysis. This research provides practitioners with an evidence-based methodology to foster more reliable safety assessments and a more effective allocation of limited safety resources.

6.3 Recommendations for Future Research

Although this research provides significant advancements in road safety analysis, several important areas remain to be explored in future studies. Building on the findings of this thesis and the gaps identified in the literature review, the following recommendations are made:

- *Investigate the Integration of Multiple Conflict Indicators:* As identified in Chapter 2, a key limitation of many conflict-based studies is the reliance on a single conflict indicator, which captures only a partial view of a location's safety. Now that a more robust framework for data collection and threshold selection has been established,

future research should focus on evaluating the uncertainty of the current research efforts for integrating multiple indicators (e.g., TTC, PET, and DRAC).

- *Expand the Scope and Generalizability of the Findings*: The analyses in this thesis were based on specific types of road facilities. Future work should expand these investigations to a larger and more diverse set of locations, including signalized intersections and freeway segments, to develop more broadly applicable guidelines for uncertainty management.
- *Develop and Assess More Complex Predictive Models*: To clearly demonstrate the impact of the core uncertainties, this thesis intentionally used relatively simple stationary models. Future research should explore more complex models, such as non-stationary EVT models that incorporate covariates like traffic volume and weather, which may alter data requirements and improve prediction accuracy.
- *Address Other Sources of Data and Model Uncertainty*: This thesis focused on uncertainty in ranking, sample size, and threshold selection. Future work could expand this to quantify other uncertainties identified in the literature, such as those arising from crash data quality (e.g., under-reporting), traffic volume estimates (AADT), and the spatial or temporal correlations between sites.
- *Develop a Bayesian Framework for Threshold Selection*: As noted in Chapter 5, a Bayesian framework could be developed to formally incorporate and quantify the uncertainty associated with the threshold selection process itself, providing a more robust and comprehensive analysis of model uncertainty.

By exploring these research directions, the field of road safety can continue to build upon the foundations established in this thesis, moving toward enhanced methodologies that offer more accurate and reliable estimations of crash risk and ultimately contribute to improved safety measures and decision-making processes.

References

- Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., & Acharya, U. R. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76, 243-297.
- Aguero-Valverde, J., & Jovanis, P. P. (2008). Analysis of road crash frequency with spatial models. *Transportation Research Record*, 2061(1), 55-63.
- Aguero-Valverde, J., & Jovanis, P. P. (2009). Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. *Transportation Research Record*, 2136(1), 82-91.
- Al Mohamad, D., Goeman, J. J., & van Zwet, E. W. (2022). Simultaneous confidence intervals for ranks with application to ranking institutions. *Biometrics*, 78(1), 238-247.
- Ali, Y., Haque, M. M., & Mannering, F. (2023a). Assessing traffic conflict/crash relationships with extreme value theory: Recent developments and future directions for connected and autonomous vehicle and highway safety research. *Analytic Methods in Accident Research*, 100276.
- Ali, Y., Haque, M. M., & Mannering, F. (2023b). A Bayesian generalised extreme value model to estimate real-time pedestrian crash risks at signalised intersections using artificial intelligence-based video analytics. *Analytic Methods in Accident Research*, 38, 100264.
- Ali, Y., Haque, M. M., & Zheng, Z. (2022). An Extreme Value Theory approach to estimate crash risk during mandatory lane-changing in a connected environment. *Analytic Methods in Accident Research*, 33, 100193. <https://doi.org/10.1016/j.amar.2021.100193>
- Ali, Y., Washington, S., & Haque, M. M. (2023). Estimating real-time crash risk at signalized intersections: A Bayesian Generalized Extreme Value approach. *Safety Science*, 164, 106181. <https://doi.org/10.1016/j.ssci.2023.106181>
- Allen, B. L., Shin, B. T., & Cooper, P. J. (1978). *Analysis of traffic conflicts and collisions* (0361-1981).
- Alozi, A. R., & Hussein, M. (2022). Evaluating the safety of autonomous vehicle-pedestrian interactions: An extreme value theory approach. *Analytic Methods in Accident Research*, 100230-100230. <https://doi.org/10.1016/j.amar.2022.100230>
- Aminghafouri, R., & Fu, L. (2023). Joint Confidence Region Approach to Ranking Hotspot Locations Considering Uncertainty in Expected Risk Estimates. *Transportation research record*, 03611981231171917.
- Amundsen, F., & Hyden, C. (1977). Proceedings of first workshop on traffic conflicts. *Oslo, TTI, Oslo, Norway and LTH Lund, Sweden*, 78.
- Ankunda, A., Ali, Y., & Mohanty, M. (2024). Pedestrian crash risk analysis using extreme value models: new insights and evidence. *Accident Analysis & Prevention*, 203, 107633.
- Arun, A., Haque, M. M., Bhaskar, A., & Washington, S. (2022). Transferability of multivariate extreme value models for safety assessment by applying artificial intelligence-based video analytics. *Accident Analysis & Prevention*, 170, 106644. <https://doi.org/10.1016/j.aap.2022.106644>
- Arun, A., Haque, M. M., Bhaskar, A., Washington, S., & Sayed, T. (2021a). A bivariate extreme value model for estimating crash frequency by severity using traffic conflicts. *Analytic Methods in Accident Research*, 32, 100180. <https://doi.org/10.1016/j.amar.2021.100180>
- Arun, A., Haque, M. M., Bhaskar, A., Washington, S., & Sayed, T. (2021b). A systematic mapping review of surrogate safety assessment using traffic conflict techniques. *Accident Analysis & Prevention*, 153, 106016. <https://doi.org/10.1016/j.aap.2021.106016>
- Arun, A., Haque, M. M., Washington, S., Sayed, T., & Mannering, F. (2021). A systematic review of traffic conflict-based safety measures with a focus on application context. *Analytic Methods in Accident Research*, 32, 100185. <https://doi.org/10.1016/j.amar.2021.100185>
- Arun, A., Haque, M. M., Washington, S., Sayed, T., & Mannering, F. (2022). How many are enough?: Investigating the effectiveness of multiple conflict indicators for crash frequency-by-severity

- estimation by automated traffic conflict analysis. *Transportation Research Part C: Emerging Technologies*, 138, 103653. <https://doi.org/10.1016/j.trc.2022.103653>
- Ash, J. E., Zou, Y., Lord, D., & Wang, Y. (2021). Comparison of confidence and prediction intervals for different mixed-Poisson regression models. *Journal of Transportation Safety & Security*, 13(3), 357-379.
- Asmussen, E. (2012). *International calibration study of traffic conflict techniques* (Vol. 5). Springer Science & Business Media.
- Bader, B., Yan, J., & Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate.
- Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. L. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289-300.
- Bladt, M., Albrecher, H., & Beirlant, J. (2020). Threshold selection and trimming in extremes. *Extremes*, 23, 629-665.
- Caliendo, C., Guida, M., & Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis & Prevention*, 39(4), 657-670.
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53). Cambridge university press.
- Campbell, K. L. (1996). *A bridging analysis for estimating the benefits of active safety technologies*. Canada, T. (2020). *Canadian Motor Vehicle Traffic Collision Statistics: 2020*. <https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2020>
- Carriquiry, A. L., & Pawlovich, M. (2004). From empirical Bayes to full Bayes: methods for analyzing traffic safety data.
- Casella, G., & Berger, R. (2001). Hypothesis testing in statistics.
- Cheng, L., Geedipally, S. R., & Lord, D. (2013). The Poisson–Weibull generalized linear model for analyzing motor vehicle crash data. *Safety science*, 54, 38-42.
- Cheng, W., Gill, G. S., Ensich, J. L., Kwong, J., & Jia, X. (2018). Multimodal crash frequency modeling: Multivariate space-time models with alternate spatiotemporal interactions. *Accident Analysis & Prevention*, 113, 159-170.
- Cheng, W., & Washington, S. P. (2005). Experimental evaluation of hotspot identification methods. *Accident Analysis & Prevention*, 37(5), 870-881.
- Chin, H. C., & Quddus, M. A. (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis & Prevention*, 35(2), 253-259.
- Choulakian, V., & Stephens, M. A. (2001). Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, 43(4), 478-484.
- Connors, R. D., Maher, M., Wood, A., Mountain, L., & Ropkins, K. (2013). Methodology for fitting and updating predictive accident models with trend. *Accident Analysis & Prevention*, 56, 82-94.
- Cooper, D. F., & Ferguson, N. (1976). Traffic studies at T-Junctions. 2. A conflict simulation Record. *Traffic Engineering & Control*, 17(Analytic).
- Council, N. R. (2000). *Risk analysis and uncertainty in flood damage reduction studies*. National Academies Press.
- Davis, G. A., Hourdos, J., Xiong, H., & Chatterjee, I. (2011). Outline for a causal model of traffic conflicts and crashes. *Accident Analysis & Prevention*, 43(6), 1907-1919. <https://doi.org/https://doi.org/10.1016/j.aap.2011.05.001>
- Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3), 393-425.

- de Leur, P., & De, P. (2018). *Collision Cost Study Update FINAL Report*.
- Ding, N., Lu, L., & Jiao, N. (2021). Rear-End Crash Risk Analysis considering Drivers' Visual Perception and Traffic Flow Uncertainty: A Hierarchical Hybrid Bayesian Network Approach. *Discrete Dynamics in Nature and Society*, 2021, 7028660. <https://doi.org/10.1155/2021/7028660>
- Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., & Gupta, S. (2006). *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment*.
- El-Basyouny, K., & Sayed, T. (2006). Comparison of two negative binomial regression techniques in developing accident prediction models. *Transportation Research Record*, 1950(1), 9-16.
- El-Basyouny, K., & Sayed, T. (2009). Collision prediction models using multivariate Poisson-lognormal regression. *Accident Analysis & Prevention*, 41(4), 820-828. <https://doi.org/https://doi.org/10.1016/j.aap.2009.04.005>
- El-Basyouny, K., & Sayed, T. (2010). Safety performance functions with measurement errors in traffic volume. *Safety Science*, 48(10), 1339-1344. <https://doi.org/https://doi.org/10.1016/j.ssci.2010.05.005>
- El-Basyouny, K., & Sayed, T. (2013). Safety performance functions using traffic conflicts. *Safety Science*, 51(1), 160-164. <https://doi.org/https://doi.org/10.1016/j.ssci.2012.04.015>
- Elvik, R. (2007). *State-of-the-art approaches to road accident black spot management and safety analysis of road networks*. Transportøkonomisk institutt Oslo.
- Elvik, R. (2010). Sources of uncertainty in estimated benefits of road safety programmes. *Accident Analysis & Prevention*, 42(6), 2171-2178.
- Elvik, R. (2013). Risk of road accident associated with the use of drugs: A systematic review and meta-analysis of evidence from epidemiological studies. *Accident Analysis & Prevention*, 60, 254-267. <https://doi.org/https://doi.org/10.1016/j.aap.2012.06.017>
- Farah, H., & Azevedo, C. L. (2017). Safety analysis of passing maneuvers using extreme value theory. *IATSS Research*, 41(1), 12-21. <https://doi.org/10.1016/j.iatssr.2016.07.001>
- Farid, A., Abdel-Aty, M., Lee, J., Eluru, N., & Wang, J.-H. (2016). Exploring the transferability of safety performance functions. *Accident Analysis & Prevention*, 94, 143-152. <https://doi.org/https://doi.org/10.1016/j.aap.2016.04.031>
- Formosa, N., Quddus, M., Ison, S., Abdel-Aty, M., & Yuan, J. (2020). Predicting real-time traffic conflicts using deep learning. *Accident Analysis & Prevention*, 136, 105429. <https://doi.org/https://doi.org/10.1016/j.aap.2019.105429>
- Fu, C., & Sayed, T. (2021a). Comparison of threshold determination methods for the deceleration rate to avoid a crash (DRAC)-based crash estimation. *Accident Analysis & Prevention*, 153, 106051. <https://doi.org/10.1016/j.aap.2021.106051>
- Fu, C., & Sayed, T. (2021b). Multivariate Bayesian hierarchical Gaussian copula modeling of the non-stationary traffic conflict extremes for crash estimation. *Analytic Methods in Accident Research*, 29, 100154. <https://doi.org/10.1016/j.amar.2020.100154>
- Fu, C., & Sayed, T. (2021c). Random parameters Bayesian hierarchical modeling of traffic conflict extremes for crash estimation. *Accident Analysis & Prevention*, 157, 106159. <https://doi.org/10.1016/j.aap.2021.106159>
- Fu, C., & Sayed, T. (2022a). Bayesian dynamic extreme value modeling for conflict-based real-time safety analysis. *Analytic Methods in Accident Research*, 34, 100204. <https://doi.org/10.1016/j.amar.2021.100204>
- Fu, C., & Sayed, T. (2022b). A multivariate method for evaluating safety from conflict extremes in real time. *Analytic Methods in Accident Research*, 36, 100244. <https://doi.org/10.1016/j.amar.2022.100244>
- Fu, C., & Sayed, T. (2022c). Random-Parameter Bayesian Hierarchical Extreme Value Modeling Approach with Heterogeneity in Means and Variances for Traffic Conflict-Based Crash

- Estimation. *Journal of Transportation Engineering, Part A: Systems*, 148(9), 04022056. <https://doi.org/10.1061/JTEPBS.0000717>
- Fu, C., & Sayed, T. (2023a). Dynamic Bayesian hierarchical peak over threshold modeling for real-time crash-risk estimation from conflict extremes. 100304-100304. <https://doi.org/10.1016/j.amar.2023.100304>
- Fu, C., & Sayed, T. (2023b). Dynamic Bayesian hierarchical peak over threshold modeling for real-time crash-risk estimation from conflict extremes. *Analytic Methods in Accident Research*, 40, 100304.
- Fu, C., & Sayed, T. (2023c). Identification of adequate sample size for conflict-based crash risk evaluation: An investigation using Bayesian hierarchical extreme value theory models. *Analytic Methods in Accident Research*, 39, 100281. <https://doi.org/10.1016/j.amar.2023.100281>
- Fu, C., Sayed, T., & Zheng, L. (2020). Multivariate Bayesian hierarchical modeling of the non-stationary traffic conflict extremes for crash estimation. *Analytic Methods in Accident Research*, 28, 100135. <https://doi.org/10.1016/j.amar.2020.100135>
- Fu, C., Sayed, T., & Zheng, L. (2021). Multi-type Bayesian hierarchical modeling of traffic conflict extremes for crash estimation. *Accident Analysis & Prevention*, 160, 106309. <https://doi.org/10.1016/j.aap.2021.106309>
- G'Sell, M. G., Wager, S., Chouldechova, A., & Tibshirani, R. (2016). Sequential selection procedures and false discovery rate control. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(2), 423-444.
- Gastaldi, M., Orsini, F., Gecchele, G., & Rossi, R. (2021). Safety analysis of unsignalized intersections: a bivariate extreme value approach. *Transportation Letters*, 13(3), 209-218. <https://doi.org/10.1080/19427867.2020.1861503>
- Geedipally, S. R., & Lord, D. (2008). Effects of varying dispersion parameter of Poisson–gamma models on estimation of confidence intervals of crash prediction models. *Transportation research record*, 2061(1), 46-54.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.
- Ghanipoor Machiani, S., & Abbas, M. (2016). Safety surrogate histograms (SSH): A novel real-time safety assessment of dilemma zone related conflicts at signalized intersections. *Accident Analysis & Prevention*, 96, 361-370. <https://doi.org/https://doi.org/10.1016/j.aap.2015.04.024>
- Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5), 1049-1054.
- Guo, Y., Sayed, T., & Zheng, L. (2020). A hierarchical bayesian peak over threshold approach for conflict-based before-after safety evaluation of leading pedestrian intervals. *Accident Analysis & Prevention*, 147, 105772.
- Guo, Y., Sayed, T., Zheng, L., & Essa, M. (2021). An extreme value theory based approach for calibration of microsimulation models for safety analysis. *Simulation Modelling Practice and Theory*, 106, 102172. <https://doi.org/10.1016/j.simpat.2020.102172>
- Gustavsson, J. (1969). On the use of regression models in the study of road accidents. *Accident Analysis & Prevention*.
- Gustavsson, J., & Svensson, Å. (1976). A Poisson regression model applied to classes of road accidents with small frequencies. *Scandinavian Journal of Statistics*, 49-60.
- Haan, L., & Ferreira, A. (2006). *Extreme value theory: an introduction* (Vol. 3). Springer.
- Hauer, E. (1982). Traffic conflicts and exposure. *Accident Analysis & Prevention*, 14(5), 359-364. [https://doi.org/https://doi.org/10.1016/0001-4575\(82\)90014-8](https://doi.org/https://doi.org/10.1016/0001-4575(82)90014-8)
- Hauer, E. (1992). Empirical Bayes approach to the estimation of “unsafety”: the multivariate regression method. *Accident Analysis & Prevention*, 24(5), 457-477.

- Hauer, E. (1997). *Observational before/after studies in road safety. estimating the effect of highway and traffic engineering measures on road safety.*
- Hauer, E., & Garder, P. (1986). Research into the validity of the traffic conflicts technique. *Accident Analysis & Prevention*, 18(6), 471-481. [https://doi.org/https://doi.org/10.1016/0001-4575\(86\)90020-5](https://doi.org/https://doi.org/10.1016/0001-4575(86)90020-5)
- Hauer, E., Harwood, D. W., Council, F. M., & Griffith, M. S. (2002). Estimating safety by the empirical Bayes method: a tutorial. *Transportation research record*, 1784(1), 126-131.
- Hauer, E., Ng, J. C., & Lovell, J. (1988). Estimation of safety at signalized intersections. *Transportation Research Record*, 1185, 48-61.
- Heydari, S., Fu, L., Joseph, L., & Miranda-Moreno, L. F. (2016). Bayesian nonparametric modeling in transportation safety studies: Applications in univariate and multivariate settings. *Analytic Methods in Accident Research*, 12, 18-34. <https://doi.org/https://doi.org/10.1016/j.amar.2016.09.001>
- Heydari, S., Fu, L., Thakali, L., & Joseph, L. (2018). Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: Applications to grade crossing safety analysis. *Analytic Methods in Accident Research*, 19, 33-48. <https://doi.org/https://doi.org/10.1016/j.amar.2018.06.003>
- Heydari, S., Miranda-Moreno, L. F., Lord, D., & Fu, L. (2014). Bayesian methodology to estimate and update safety performance functions under limited data conditions: A sensitivity analysis. *Accident Analysis & Prevention*, 64, 41-51. <https://doi.org/https://doi.org/10.1016/j.aap.2013.11.001>
- Hosking, J. R. M. (2006). On the characterization of distributions by their L-moments. *Journal of Statistical Planning and Inference*, 136(1), 193-198. <https://doi.org/https://doi.org/10.1016/j.jspi.2004.06.004>
- Hosking, J. R. M., & Balakrishnan, N. (2015). A uniqueness result for L-estimators, with applications to L-moments. *Statistical Methodology*, 24, 69-80. <https://doi.org/https://doi.org/10.1016/j.stamet.2014.08.002>
- Hosking, J. R. M., & Wallis, J. R. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29(3), 339-349. <https://doi.org/10.1080/00401706.1987.10488243>
- Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis.*
- Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics*, 27(3), 251-261. <https://doi.org/10.1080/00401706.1985.10488049>
- Howlader, M., Mannering, F., & Haque, M. (2024). Estimating crash risk and injury severity considering multiple traffic conflict and crash types: A bivariate extreme value approach. *Analytic Methods in Accident Research*. <https://doi.org/10.1016/j.amar.2024.100331>
- Howlader, M. M., Ali, Y., Burbridge, A., & Haque, M. M. (2024). Before-after safety evaluation of part-time protected right-turn signals: An extreme value theory approach by applying artificial intelligence-based video analytics. *Accident Analysis & Prevention*, 194, 107341. <https://doi.org/10.1016/j.aap.2023.107341>
- Howlader, M. M., Bhaskar, A., Yasmin, S., & Haque, M. M. (2024). A bivariate, non-stationary extreme value model for estimating opposing-through crash frequency by severity by applying artificial intelligence-based video analytics. *Transportation Research Part C: Emerging Technologies*, 160, 104509. <https://doi.org/10.1016/j.trc.2024.104509>
- Huang, H., Chin, H. C., & Haque, M. (2008). Bayesian hierarchical analysis on crash prediction models. TRB 87th Annual Meeting Compendium of Papers DVD,
- Huang, H., Chin, H. C., & Haque, M. M. (2009). Empirical evaluation of alternative approaches in identifying crash hot spots: Naive ranking, empirical bayes, full bayes methods. *Transportation Research Record*, 2103(1), 32-41.

- Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3), 457-506.
- Hussain, F., Ali, Y., Li, Y., & Boyle, L. N. (2023). Real-time crash risk forecasting using Artificial-Intelligence based video analytics: A unified framework of generalised extreme value theory and autoregressive integrated moving average model. 40, 100302-100302. <https://doi.org/10.1016/j.amar.2023.100302>
- Hussain, F., Ali, Y., Li, Y., & Haque, M. (2024a). Revisiting the hybrid approach of anomaly detection and extreme value theory for estimating pedestrian crashes using traffic conflicts obtained from artificial intelligence-based video analytics. *Accident Analysis and Prevention*. <https://doi.org/10.1016/j.aap.2024.107517>
- Hussain, F., Ali, Y., Li, Y., & Haque, M. M. (2023). Real-time crash risk forecasting using Artificial-Intelligence based video analytics: A unified framework of generalised extreme value theory and autoregressive integrated moving average model. *Analytic Methods in Accident Research*, 40, 100302.
- Hussain, F., Ali, Y., Li, Y., & Haque, M. M. (2024b). Revisiting the hybrid approach of anomaly detection and extreme value theory for estimating pedestrian crashes using traffic conflicts obtained from artificial intelligence-based video analytics. *Accident Analysis & Prevention*, 199, 107517.
- Hussain, F., Li, Y., Arun, A., & Haque, M. M. (2022). A hybrid modelling framework of machine learning and extreme value theory for crash risk estimation using traffic conflicts. *Analytic Methods in Accident Research*, 36, 100248. <https://doi.org/10.1016/j.amar.2022.100248>
- Ismail, K., Sayed, T., & Saunier, N. (2011). Methodologies for Aggregating Indicators of Traffic Conflict. *Transportation research record*, 2237(1), 10-19. <https://doi.org/10.3141/2237-02>
- Ismail, K., Sayed, T., Saunier, N., & Lim, C. (2009). Automated analysis of pedestrian-vehicle conflicts using video data. *Transportation research record*, 2140(1), 44-54.
- Jiang, X., Huang, B., Zaretski, R. L., Richards, S., & Yan, X. (2013). Estimating safety effects of pavement management factors utilizing Bayesian random effect models. *Traffic injury prevention*, 14(7), 766-775.
- Johnsson, C., Laureshyn, A., & De Ceunynck, T. (2018). In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. *Transport Reviews*, 38(6), 765-785. <https://doi.org/10.1080/01441647.2018.1442888>
- Joshua, S. C., & Garber, N. J. (1990). Estimating truck accident rate and involvements using linear and Poisson regression models. *Transportation planning and Technology*, 15(1), 41-58.
- Jovanis, P. P., & Chang, H.-L. (1986). Modeling the relationship of accidents to miles traveled. *Transportation Research Record*, 1068, 42-51.
- Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, 11(2), 143-157.
- Kaparias, I., Bell, M. G. H., Greensted, J., Cheng, S., Miri, A., Taylor, C., & Mount, B. (2010). Development and Implementation of a Vehicle-Pedestrian Conflict Analysis Method: Adaptation of a Vehicle-Vehicle Technique. *Transportation research record*, 2198(1), 75-82. <https://doi.org/10.3141/2198-09>
- Kar, P., Venthuruthiyil, S. P., & Chunchu, M. (2023). Non-stationary crash risk modelling of powered two-wheelers using extreme value analysis of surrogate crash events. *Accident; Analysis and Prevention*, 183, 106973. <https://doi.org/10.1016/j.aap.2023.106973>
- Kim, T., Lieberman, B., Luta, G., & Peña, E. A. (2022). Prediction intervals for Poisson-based regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(5), e1568.
- Kleibelsberg, D. (1964). Derzeitiger Stand der Verhaltensanalyse des Kraftfahrens, Zrbeit und Leistung. *Ablt. Arbeitswissenschaft Soziale Betriebspraxis*, 18, 33-37. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84856078522&partnerID=40&md5=7ebe53771c832235f8b77f94cd0d4811>

- Klein, M., Wright, T., & Wieczorek, J. (2020). A joint confidence region for an overall ranking of populations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(3), 589-606.
- Knight Frank, H. (1921). Risk, uncertainty and profit. *книга*.
- Kumar, A., & Mudgal, A. (2023). Effect of vehicle size on crash risk in a heterogeneous traffic scenario: a bivariate extreme value approach. *Transportation Letters*, 1-16. <https://doi.org/10.1080/19427867.2023.2171843>
- Kumar, A., & Mudgal, A. (2024). Defining Traffic Conflict in Nonlane-Based Traffic Conditions: An Extreme Value Approach. *Journal of Transportation Engineering, Part A: Systems*, 150(9), 04024050. <https://doi.org/10.1061/JTEPBS.TEENG-8037>
- Lai, Z., Karim, I., & Xianghai, M. (2014). Traffic conflict techniques for road safety analysis: open questions and some insights. *Canadian Journal of Civil Engineering*, 41(7), 633-641. <https://doi.org/10.1139/cjce-2013-0558>
- Lan, B., & Persaud, B. (2012). Evaluation of multivariate Poisson log normal Bayesian methods for before-after road safety evaluations. *Journal of Transportation Safety & Security*, 4(3), 193-210.
- Langousis, A., Mamalakis, A., Puliga, M., & Deidda, R. (2016). Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resources Research*, 52(4), 2659-2681.
- Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, 209-225.
- Lazic, S. E., & Williams, D. P. (2021). Quantifying sources of uncertainty in drug discovery predictions with probabilistic models. *Artificial Intelligence in the Life Sciences*, 1, 100004. <https://doi.org/https://doi.org/10.1016/j.ailsci.2021.100004>
- Leadbetter, M. R., Lindgren, G., & Rootzén, H. (1989). Extremes and related properties of random sequences and processes.
- Lesani, A., Nateghinia, E., & Miranda-Moreno, L. F. (2020). Development and evaluation of a real-time pedestrian counting system for high-volume conditions based on 2D LiDAR. *Transportation Research Part C: Emerging Technologies*, 114, 20-35.
- Li, Y., Chen, J., & Feng, L. (2012). Dealing with uncertainty: A survey of theories and practices. *IEEE Transactions on Knowledge and Data Engineering*, 25(11), 2463-2482.
- Lima Azevedo, C., Ciuffo, B., Cardoso, J. L., & Ben-Akiva, M. E. (2015). Dealing with uncertainty in detailed calibration of traffic simulation models for safety assessment. *Transportation Research Part C: Emerging Technologies*, 58, 395-412. <https://doi.org/https://doi.org/10.1016/j.trc.2015.01.029>
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation research part A: policy and practice*, 44(5), 291-305.
- Lord, D., & Miranda-Moreno, L. F. (2008). Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: A Bayesian perspective. *Safety science*, 46(5), 751-770.
- Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention*, 37(1), 35-46.
- MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., & Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, 55(6), 2137-2157.
- MacNab, Y. C. (2003). A Bayesian hierarchical model for accident and injury surveillance. *Accident Analysis & Prevention*, 35(1), 91-102.
- Mannering, F. (2018). Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research*, 17, 1-13.

- Mannering, F., Bhat, C. R., Shankar, V., & Abdel-Aty, M. (2020). Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. *Analytic Methods in Accident Research*, 25, 100113.
- Mannering, F. L., & Bhat, C. R. (2014). Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research*, 1, 1-22.
- Miaou, S.-P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis & Prevention*, 26(4), 471-482.
- Miaou, S.-P., & Lord, D. (2003). Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes methods. *Transportation Research Record*, 1840(1), 31-40.
- Miaou, S.-P., & Lum, H. (1993). Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention*, 25(6), 689-709.
- Miaou, S.-P., & Song, J. J. (2005). Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. *Accident Analysis & Prevention*, 37(4), 699-720.
- Miranda-Moreno, L. F., Fu, L., Saccomanno, F. F., & Labbe, A. (2005). Alternative risk models for ranking locations for safety improvement. *Transportation Research Record*, 1908(1), 1-8.
- Miranda-Moreno, L. F., Heydari, S., Lord, D., & Fu, L. (2013). Bayesian road safety analysis: Incorporation of past evidence and effect of hyper-prior choice. *Journal of safety research*, 46, 31-40.
- Miranda-Moreno, L. F., Labbe, A., & Fu, L. (2007). Bayesian multiple testing procedures for hotspot identification. *Accident Analysis & Prevention*, 39(6), 1192-1201.
- Mogstad, M., Romano, J., Shaikh, A., & Wilhelm, D. (2022). Statistical uncertainty in the ranking of journals and universities. AEA Papers and Proceedings,
- Mogstad, M., Romano, J. P., Shaikh, A., & Wilhelm, D. (2020). *Inference for ranks with applications to mobility across neighborhoods and academic achievement across countries*.
- Nassereddine, H., Santiago-Chaparro, K. R., & Noyce, D. A. (2023). Safety Estimation of Vehicle-Pedestrian Interactions Using Extreme Value Theory at Intersections with and without Right-Turn Flashing Yellow Arrow Indication. <https://doi.org/10.1061/9780784484883.021>
- National Research Council, T. R. B. (2010). *Highway safety manual* (Vol. 1). AASHTO.
- Natke, H., & Ben-Haim, Y. (1997). Uncertainty-A Discussion From Various Points of View. *MATHEMATICAL RESEARCH*, 99, 267-276.
- Naveau, P., Huser, R., Ribereau, P., & Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4), 2753-2769.
- Nazir, F., Ali, Y., & Haque, M. M. (2024). Effects of sample size on pedestrian crash risk estimation from traffic conflicts using extreme value models. *Analytic Methods in Accident Research*, 44, 100353. <https://doi.org/10.1016/j.amar.2024.100353>
- Niu, D., Sayed, T., Fu, C., & Mannering, F. (2024). A cross-comparison of different extreme value modeling techniques for traffic conflict-based crash risk estimation. *Analytic Methods in Accident Research*, 44, 100352. <https://doi.org/10.1016/j.amar.2024.100352>
- Northrop, P. J., & Coleman, C. L. (2014). Improved threshold diagnostic plots for extreme value analyses. *Extremes*, 17, 289-303.
- Orsini, F., Gecchele, G., Gastaldi, M., & Rossi, R. (2020). Transferability and seasonality in extreme value theory applications to road safety: a case study in an Italian motorway. *Advances in transportation studies*.
- Parachute. (2022). *Vision Zero map*. <https://parachute.ca/en/program/vision-zero/vision-zero-map/#!adopted-locations>
- Park, E. S., Carlson, P. J., Porter, R. J., & Andersen, C. K. (2012). Safety effects of wider edge lines on rural, two-lane highways. *Accident Analysis & Prevention*, 48, 317-325.
- Park, E. S., & Lord, D. (2007). Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transportation Research Record*, 2019(1), 1-6.

- Persaud, B., Lan, B., Lyon, C., & Bhim, R. (2010). Comparison of empirical Bayes and full Bayes approaches for before–after road safety evaluations. *Accident Analysis & Prevention*, 42(1), 38-43.
- Qin, X., Ng, M., & Reyes, P. E. (2010). Identifying crash-prone locations with quantile regression. *Accident Analysis & Prevention*, 42(6), 1531-1537.
- Ribatet, M. (2011). A user's guide to the POT package (version 1.4). <http://cran.r-project.org/package=POT>
- Rigby, R. A., Stasinopoulos, D. M., & Akantziliotou, C. (2008). A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. *Computational Statistics & Data Analysis*, 53(2), 381-393.
- Rising, J. (2021). Uncertainty in Ranking. *arXiv preprint arXiv:2107.03459*.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual review of psychology*, 52(1), 59-82.
- Sacchi, E., & Sayed, T. (2016). Conflict-based safety performance functions for predicting traffic collisions by type. *Transportation research record*, 2583(1), 50-55.
- Salamati, K., Schroeder, B., Roupail, N. M., Cunningham, C., Long, R., & Barlow, J. (2011). Development and Implementation of Conflict-Based Assessment of Pedestrian Safety to Evaluate Accessibility of Complex Intersections. *Transportation research record*, 2264(1), 148-155. <https://doi.org/10.3141/2264-17>
- Salman, N. K., & Al-Maita, K. J. (1995). Safety evaluation at three-leg, unsignalized intersections by traffic conflict technique. *Transportation research record*, 1485(28), 177-185.
- Savolainen, P. T., Mannering, F. L., Lord, D., & Quddus, M. A. (2011). The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. *Accident Analysis & Prevention*, 43(5), 1666-1676.
- Sayed, T., & Zein, S. (1999). Traffic conflict standards for intersections. *Transportation planning and technology*, 22(4), 309-323.
- Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical journal*, 10(1), 33–60-33–60.
- Schlögl, M., & Stütz, R. (2019). Methodological considerations with data uncertainty in road safety analysis. *Accident Analysis & Prevention*, 130, 136-150.
- Schneider, L. F., Krajina, A., & Krivobokova, T. (2021). Threshold selection in univariate extreme value analysis. *Extremes*, 24, 881-913.
- Shangguan, Q., Keung, J., Fu, L., Samara, L., Wang, J., & Fu, T. (2023). Do Traffic Countermeasures Improve the Safety of Vulnerable Road Users at Signalized Intersections? A Combination of Case-Control and Cross-Sectional Studies Using Video-Based Traffic Conflicts. *Transportation research record*, 03611981231172748.
- Shankar, V. N., Albin, R. B., Milton, J. C., & Mannering, F. L. (1998). Evaluating median crossover likelihoods with clustered accident counts: An empirical inquiry using the random effects negative binomial model. *Transportation Research Record*, 1635(1), 44-48.
- Shanker, V., Milton, J., & Mannering, F. (1997). Modeling accident frequencies as zero-altered probability process. *Accident Analysis and Prevention*, 29, 829-837.
- Silva Lomba, J., & Fraga Alves, M. I. (2020). L-moments for automatic threshold selection in extreme value analysis. *Stochastic environmental research and risk assessment*, 34(3), 465-491.
- Singh, S., Ali, Y., & Haque, M. M. (2024). A Bayesian extreme value theory modelling framework to assess corridor-wide pedestrian safety using autonomous vehicle sensor data. *Accident Analysis & Prevention*, 195, 107416.
- Songchitruksa, P., & Tarko, A. P. (2006). The extreme value theory approach to safety estimation. *Accident Analysis & Prevention*, 38(4), 811-822. <https://doi.org/10.1016/j.aap.2006.02.003>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4), 583-639.

- Srinivasan, R., Baek, J., & Council, F. (2010). Safety evaluation of transverse rumble strips on approaches to stop-controlled intersections in rural areas. *Journal of Transportation Safety & Security*, 2(3), 261-278.
- Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347), 730-737.
- Stipancic, J., Miranda-Moreno, L., Saunier, N., & Labbe, A. (2018). Surrogate safety and network screening: Modelling crash frequency using GPS travel data and latent Gaussian Spatial Models. *Accident Analysis & Prevention*, 120, 174-187. <https://doi.org/https://doi.org/10.1016/j.aap.2018.07.013>
- Stuart, C., & Coles, S. (2001). An Introduction to Statistical Modeling of Extreme Values. <https://doi.org/10.1198/jasa.2002.s232>
- Stylianou, K., & Dimitriou, L. (2018). Analysis of rear-end conflicts in urban networks using Bayesian networks. *Transportation research record*, 2672(38), 302-312.
- Tageldin, A., & Sayed, T. (2019). Models to evaluate the severity of pedestrian-vehicle conflicts in five cities. *Transportmetrica A Transport Science*, 15(2), 354-375. <https://doi.org/https://doi.org/10.1080/23249935.2018.1477853>
- Tarko, A. P. (2012). Use of crash surrogates and exceedance statistics to estimate road safety. *Accident Analysis & Prevention*, 45, 230-240. <https://doi.org/10.1016/j.aap.2011.07.008>
- Tarko, A. P. (2018). Estimating the expected number of crashes with traffic conflicts and the Lomax Distribution – A theoretical and numerical exploration. *Accident Analysis & Prevention*, 113, 63-73. <https://doi.org/https://doi.org/10.1016/j.aap.2018.01.008>
- Thompson, P., Cai, Y., Reeve, D., & Stander, J. (2009). Automated threshold selection methods for extreme wave analysis. *Coastal Engineering*, 56(10), 1013-1021. <https://doi.org/https://doi.org/10.1016/j.coastaleng.2009.06.003>
- Ulfarsson, G. F., & Shankar, V. N. (2003). Accident count model based on multiyear cross-sectional roadway data with serial correlation. *Transportation Research Record*, 1840(1), 193-197.
- Vision Zero Network. (2022). *Vision Zero Communities*. <https://visionzeronetwork.org/resources/vision-zero-communities/>
- Vitelli, V., Sørensen, Ø., Crispino, M., Frigessi Di Rattalma, A., & Arjas, E. (2018). Probabilistic preference learning with the Mallows rank model. *Journal of Machine Learning Research*, 18(158), 1-49.
- Vodden, K., Smith, D., Eaton, F., & Mayhew, D. (2007). Analysis and estimation of the social cost of motor vehicle collisions in Ontario. *Transport Canada*.
- Wang, C., & Stamatiadis, N. (2014a). Derivation of a New Surrogate Measure of Crash Severity. *Transportation research record*, 2432(1), 37-45. <https://doi.org/10.3141/2432-05>
- Wang, C., & Stamatiadis, N. (2014b). Evaluation of a simulation-based surrogate safety metric. *Accident Analysis & Prevention*, 71, 82-92. <https://doi.org/https://doi.org/10.1016/j.aap.2014.05.004>
- Wang, C., Xu, C., & Dai, Y. (2019). A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. *Accident Analysis & Prevention*, 123, 365-373. <https://doi.org/10.1016/j.aap.2018.12.013>
- Wang, X., Xie, K., Abdel-Aty, M., Chen, X., & Tremont, P. J. (2014). Systematic approach to hazardous-intersection identification and countermeasure development. *Journal of transportation engineering*, 140(6), 04014022.
- Washington, S., Haque, M. M., Oh, J., & Lee, D. (2014). Applying quantile regression for modeling equivalent property damage only crashes to identify accident blackspots. *Accident Analysis & Prevention*, 66, 136-146.
- Wen, J., Lyu, N., & Liu, C. (2024). Safety Evaluation and Influential Factor Analysis of Urban Expressway Diversion Areas Based on Non-Stationary Conflict Extremes. *Transportation Research Record*. <https://doi.org/10.1177/03611981241252836>

- Wood, G. (2005). Confidence and prediction intervals for generalised linear accident models. *Accident Analysis & Prevention*, 37(2), 267-273.
- Wu, K.-F., & Jovanis, P. P. (2012). Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data. *Accident Analysis & Prevention*, 45, 507-516.
- Wu, K.-F., & Jovanis, P. P. (2013). Defining and screening crash surrogate events using naturalistic driving data. *Accident Analysis & Prevention*, 61, 10-22. <https://doi.org/https://doi.org/10.1016/j.aap.2012.10.004>
- Xie, K., Yang, D., Ozbay, K., & Yang, H. (2019). Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure. *Accid Anal Prev*, 125, 311-319. <https://doi.org/10.1016/j.aap.2018.07.002>
- Yang, D., Xie, K., Ozbay, K., Zhao, Z., & Yang, H. (2021). Copula-based joint modeling of crash count and conflict risk measures with accommodation of mixed count-continuous margins. *Analytic Methods in Accident Research*, 31, 100162.
- Yang, X., Zhang, J., & Ren, W.-X. (2018). Threshold selection for extreme value estimation of vehicle load effect on bridges. *International Journal of Distributed Sensor Networks*, 14(2), 1550147718757698. <https://doi.org/10.1177/1550147718757698>
- Ye, X., Pendyala, R. M., Shankar, V., & Konduri, K. C. (2013). A simultaneous equations model of crash frequency by severity level for freeway sections. *Accident Analysis & Prevention*, 57, 140-149.
- Yue, Q., Guo, Y., Sayed, T., Liu, P., Zheng, L., & Lyu, H. (2024). Bayesian hybrid gamma-GPD model for extreme traffic conflict threshold determination in the peak over threshold approach. *Accident Analysis & Prevention*, 206, 107717. <https://doi.org/10.1016/j.aap.2024.107717>
- Zarei, M., & Hellinga, B. (2022). Method for Estimating the Monetary Benefit of Improving Annual Average Daily Traffic Accuracy in the Context of Road Safety Network Screening. *Transportation research record*, 0(0), 03611981221115720. <https://doi.org/10.1177/03611981221115720>
- Zhang, S., Luo, J., Zhu, L., Stinchcomb, D. G., Campbell, D., Carter, G., Gilkeson, S., & Feuer, E. J. (2014). Confidence intervals for ranks of age-adjusted rates across states or counties. *Statistics in Medicine*, 33(11), 1853-1866.
- Zhao, M., Liu, C., Li, W., & Sharma, A. (2018). Multivariate Poisson-lognormal model for analysis of crashes on urban signalized intersections approach. *Journal of Transportation Safety & Security*, 10(3), 251-265.
- Zheng, L., Huang, Y., Sayed, T., & Wen, C. (2021). Validating the Bayesian hierarchical extreme value model for traffic conflict-based crash estimation on freeway segments with site-level factors. *Accident Analysis & Prevention*, 159, 106269. <https://doi.org/10.1016/j.aap.2021.106269>
- Zheng, L., & Ismail, K. (2017). A generalized exponential link function to map a conflict indicator into severity index within safety continuum framework. *Accident Analysis & Prevention*, 102, 23-30. <https://doi.org/https://doi.org/10.1016/j.aap.2017.02.013>
- Zheng, L., Ismail, K., & Meng, X. (2014). Freeway safety estimation using extreme value theory approaches: A comparative study. *Accident Analysis & Prevention*, 62, 32-41. <https://doi.org/10.1016/j.aap.2013.09.006>
- Zheng, L., Ismail, K., & Meng, X. (2015). Evaluation of peak over threshold approach for road safety estimation. *Journal of Transportation Safety & Security*, 7(1), 76-90.
- Zheng, L., Ismail, K., & Meng, X. (2016). Investigating the heterogeneity of postencroachment time thresholds determined by peak over threshold approach [Article]. *Transportation research record*, 2601, 17-23. <https://doi.org/10.3141/2601-03>
- Zheng, L., Ismail, K., Sayed, T., & Fatema, T. (2018). Bivariate extreme value modeling for road safety estimation. *Accident Analysis & Prevention*, 120, 83-91. <https://doi.org/10.1016/j.aap.2018.08.004>

- Zheng, L., & Sayed, T. (2019a). Application of Extreme Value Theory for Before-After Road Safety Analysis. *Transportation Research Record*, 2673(4), 1001-1010. <https://doi.org/10.1177/0361198119841555>
- Zheng, L., & Sayed, T. (2019b). Bayesian hierarchical modeling of traffic conflict extremes for crash estimation: A non-stationary peak over threshold approach. *Analytic Methods in Accident Research*, 24, 100106. <https://doi.org/10.1016/j.amar.2019.100106>
- Zheng, L., & Sayed, T. (2019c). Comparison of Traffic Conflict Indicators for Crash Estimation using Peak Over Threshold Approach. *Transportation Research Record*, 2673(5), 493-502. <https://doi.org/10.1177/0361198119841556>
- Zheng, L., & Sayed, T. (2019d). From univariate to bivariate extreme value models: Approaches to integrate traffic conflict indicators for crash estimation. *Transportation Research Part C: Emerging Technologies*, 103, 211-225. <https://doi.org/10.1016/j.trc.2019.04.015>
- Zheng, L., & Sayed, T. (2020a). A bivariate Bayesian hierarchical extreme value model for traffic conflict-based crash estimation. *Analytic Methods in Accident Research*, 25, 100111. <https://doi.org/10.1016/j.amar.2020.100111>
- Zheng, L., & Sayed, T. (2020b). A novel approach for real time crash prediction at signalized intersections. *Transportation Research Part C: Emerging Technologies*, 117, 102683. <https://doi.org/10.1016/j.trc.2020.102683>
- Zheng, L., Sayed, T., & Essa, M. (2019). Validating the bivariate extreme value modeling approach for road safety estimation with different traffic conflict indicators. *Accident Analysis & Prevention*, 123, 314-323. <https://doi.org/10.1016/j.aap.2018.12.007>
- Zheng, L., Sayed, T., & Mannering, F. (2021). Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. *Analytic methods in accident research*, 29, 100142.
- Zhou, S., Sun, J., An, X., & Li, K. (2011). The development of a conflict hazardous assessment model for evaluating urban intersection safety. *Transport*, 26(2), 216-223.