SUSTAINABLE TRANSPORT SAFETY: APPLYING UPDATED COMMUNITY-BASED MACRO-LEVEL COLLISION PREDICTION MODELS FOR KELOWNA USING MANUAL AND AUTOMATED TOOLS

by

Vipul Garg

B.Tech., National Institute of Technology Kurukshetra, India, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF APPLIED SCIENCE in THE COLLEGE OF GRADUATE STUDIES (Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA (Okanagan)

May 2019

© Vipul Garg, 2019
The following individuals certify that they have read, and recommend to the College of Graduate Studies for acceptance, a thesis/dissertation entitled:

SUSTAINABLE TRANSPORT SAFETY: APPLYING UPDATED COMMUNITY-BASED MACRO-LEVEL COLLISION PREDICTION MODELS FOR KELOWNA USING MANUAL AND AUTOMATED TOOL

submitted by Vipul Garg in partial fulfillment of the requirements for the degree of Master of Applied Science

Examinining Committee:

Dr. Gordon Lovegrove, School of Engineering, UBCO

Supervisor

Dr. Mahmudur Fatmi, School of Engineering, UBCO

Co-supervisor

Dr. Ahmed Idris, Arab Academy for Science, Technology and Maritime Transport

Supervisory Committee Member

Dr. Zheng Liu, School of Engineering, UBCO

Supervisory Committee Member

Dr. Stephen O’Leary, School of Engineering, UBCO

University Examiner
Trafﬁc-related injuries and fatalities are considered the ninth worst global epidemic, exerting an enormous social and economic burden across countries. Transportation planners and engineers are now proactively considering safety in the transportation planning process to improve the overall safety of the transportation network. Macro-level collision prediction models (CPMs) have been identiﬁed as a reliable planning-level decision-support tool to inform safety conscious planning. However, the success or failure of these models in one spatial environment depends on the quality of the dataset, appropriate model functional form, and time frame of model development and application. As road parameters, network design, regional socio-demographic and size of geographical aggregation unit changes over time, updating these models becomes necessary. This research reviewed the literature corresponding to the development and application of community-based macro-level CPMs. Based on the insights gained from the literature, community-based macro-level CPMs were updated for Kelowna using the negative-binomial (NB) and full Bayesian (FB) modelling techniques, and IHSPM model development module. The updated CPM results are in good agreement with past research that developed the original Kelowna models. One new and signiﬁcant result in updating the CPMs, due to the availability of new data, was the decreased collisions with the increased proportion of roundabouts, a result seen previously only in micro-level collision prediction models. The updated CPMs were then applied to present-day Kelowna neighbourhoods, to identify and rank collision-prone zones (CPZs). A series of macro-level collision modiﬁcation factors (CMFs) was developed to examine the impact of different exposure, socio-demographic, transportation demand management and network-related countermeasure strategies on the safety of a trafﬁc analysis zone. Moreover, as the development of these models is a cumbersome, semi-manual, and time-consuming process that requires a sound knowledge of spatial analysis, and data analysis tools, this thesis assessed an online road safety planning tool 'IHSPM' that automates the development and application of community-based macro-level CPMs models, developed to date at STS lab.
LAY SUMMARY

Community-based macro-level collision prediction models (CPMs) can be utilized to predict the safety performance of planned facilities, identify and rank collision-prone zones, and evaluate the effectiveness of community-level road safety countermeasures. This research updated the community-based macro-level CPMs for the city of Kelowna using the negative binomial (NB) and full Bayesian (FB) modelling techniques. The models were also developed using automated software to validate its performance. The developed models were then applied to identify and diagnose collision-prone neighbourhoods in Kelowna having collision potential significantly higher than normal.
# Table of Contents

TABLE OF CONTENTS

ABSTRACT ........................................................................................................ iii
LAY SUMMARY .................................................................................................. iv
TABLE OF CONTENTS ....................................................................................... v
LIST OF TABLES ............................................................................................... ix
LIST OF FIGURES ........................................................................................... x
LIST OF ABBREVIATIONS ................................................................................ xii
ACKNOWLEDGEMENTS .................................................................................. xiv
DEDICATION .................................................................................................... xv

CHAPTER 1: INTRODUCTION .......................................................................... 1
  1.1 Overview .................................................................................................. 1
  1.2 Road Safety Problem ............................................................................. 1
  1.3 Problem Statement ............................................................................... 2
  1.4 Research Objectives ............................................................................ 3
    1.4.1 Objective 1: Updating Kelowna Community-Based Macro-Level CPMs – Transferring Road Safety Planning Models Across Time .............................................. 3
    1.4.2 Objective 2: Assessment of Interactive High-Level Safety Planning Model (IHSPM) – Beta Test................................................................. 4
    1.4.3 Objective 3: Macro-Reactive Road Safety Applications – Identifying Collision Prone Zones (CPZs) and Defining Collision Modification Factors (CMFs) ......................... 4
  1.5 Thesis Structure .................................................................................... 4

CHAPTER 2: LITERATURE REVIEW ................................................................. 6
  2.1 Overview ............................................................................................... 6
  2.2 Approaches to Macro-Level Collision Prediction Modelling ............... 6
    2.2.1 Generalized Linear Modelling.......................................................... 8
# Table of Contents

2.2.2 Full Bayesian (FB) Modelling ................................................................................................. 10

2.2.3 Comparison between the Regression Techniques ................................................................. 12

2.2.4 Variable Selection and Supporting Evidence ......................................................................... 13

2.3 Macro-Reactive Applications ................................................................................................... 15

2.3.1 Black Spot Studies .................................................................................................................... 16

2.3.2 Before-and-After Studies ......................................................................................................... 18

2.5 Summary .................................................................................................................................... 19

CHAPTER 3: ASSESSMENT OF INTERACTIVE HIGH-LEVEL SAFETY PLANNING MODEL (IHSPM) .................................................................................................................. 20

3.1 Overview .................................................................................................................................... 20

3.2 Need for Automation .................................................................................................................... 20

3.2.1 Attempts made in Past towards Automation .......................................................................... 20

3.2.2 PLANSAFE ............................................................................................................................... 21

3.3 IHSPM ....................................................................................................................................... 23

3.3.1 Languages Used ....................................................................................................................... 23

3.3.2 Implementations ....................................................................................................................... 24

3.4 IHSPM Modules .......................................................................................................................... 24

3.4.1 Map Data Aggregation Module ............................................................................................. 24

3.4.2 Walkthrough Module .............................................................................................................. 27

3.4.3 CPM Development Module ................................................................................................... 33

3.5 IHSPM Beta Test ......................................................................................................................... 36

3.5.1 Methodology ........................................................................................................................... 36

3.5.2 Observations and Results ...................................................................................................... 37

3.6 Summary .................................................................................................................................... 41

CHAPTER 4: DATA EXTRACTION AND MODEL DEVELOPMENT ................................................. 42

4.1 Overview .................................................................................................................................... 42
Table of Contents

4.2 Data Extraction and Description ................................................................. 42
  4.2.1 Geographical Scope and Aggregation Units .......................................... 42
  4.2.2 Data Sources and Summary Statistics .................................................. 45
4.3 Model Development and Results .................................................................. 48
  4.3.1 Negative Binomial Models ................................................................... 49
  4.3.2 Full Bayesian (FB) Models .................................................................... 58
  4.3.3 Comparison of NB & FB, and NB & IHSPM models ................................. 64
4.4 Summary and Discussions ............................................................................ 67
CHAPTER 5: MACRO REACTIVE SAFETY APPLICATIONS ............................... 69
  5.1 Overview ................................................................................................. 69
  5.2 Black Spot Case Study ............................................................................. 69
    5.2.1 Identification and Ranking Using EB Method ....................................... 70
    5.2.2 Identification and Ranking Using FB Method ....................................... 75
    5.2.3 Comparison of Identified CPZs Using EB and FB Method ..................... 80
  5.3 Diagnosis and Remedies of the CPZs ......................................................... 81
    5.3.1 Collision Prone Zone 8 ....................................................................... 81
    5.3.2 Collision Prone Zone 5 ....................................................................... 85
    5.3.3 Collision Prone Zone 44 ..................................................................... 89
    5.3.4 Collision Prone Zone 50 ..................................................................... 92
  5.4 Examining the Impact of Planning Design Changes ..................................... 95
    5.4.1 Increased Neighbourhood Core Area (CORE) ...................................... 96
    5.4.2 Increased Proportion of Roundabouts (IRBP) ....................................... 97

CHAPTER 6: CONCLUSIONS, CONTRIBUTIONS & FUTURE RESEARCH ....... 102
  6.1 Overview .................................................................................................. 102
  6.2 Summary & Conclusions .......................................................................... 102
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.1 Updating Community-Based Macro-Level CPMs for Kelowna</td>
<td>102</td>
</tr>
<tr>
<td>6.2.2 Assessment of Interactive High-Level Safety Planning Model (IHSPM) – Beta Test</td>
<td>104</td>
</tr>
<tr>
<td>6.2.3 Macro-Reactive Road Safety Applications</td>
<td>105</td>
</tr>
<tr>
<td>6.3 Research Contributions</td>
<td>105</td>
</tr>
<tr>
<td>6.4 Recommendations for Future Research</td>
<td>107</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>109</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>114</td>
</tr>
<tr>
<td>Appendix A: R Script for Development of NB Models</td>
<td>114</td>
</tr>
<tr>
<td>Appendix B. WinBUGS Samples of Model File, and Data File Scripts</td>
<td>119</td>
</tr>
<tr>
<td>Appendix C. Components of Interactive High Level Safety Planning Model (IHSPM)</td>
<td>120</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1 Recommended Model Groupings (G. R. Lovegrove & Sayed, 2006) ......................... 15
Table 3.1 Recommended Model Groups Based on Planning Topic and Comparison .................. 29
Table 3.2 Data Extraction Efforts for Planning Analysis, based on Theme and Time Horizon... 31
Table 3.3 Beta Test User Feedback and Recommended Actions ........................................... 39
Table 4.1 Dependent and Exposure Variables Definition and Data Summary (n=183) ............. 47
Table 4.2 Socio-demographic Variables Definitions and Data Summary (n=183) ....................... 47
Table 4.3 TDM Variables Definition and Data Summary (n=183) ........................................ 48
Table 4.4 Network Variables Definition and Data Summary (n=183) ..................................... 48
Table 4.5 Model Groupings ..................................................................................................... 51
Table 4.6 Negative Binomial URBAN CPMs – Total Collisions ............................................. 52
Table 4.7 Negative Binomial RURAL CPMs – Total Collisions ............................................. 53
Table 4.8 IHSPM Developed CPMs ......................................................................................... 57
Table 4.9 Urban Measured Full Bayesian Models ...................................................................... 60
Table 4.10 Urban Modelled Full Bayesian Models ................................................................. 61
Table 4.11 Rural Measured Full Bayesian Models .................................................................... 62
Table 4.12 Rural Modelled Full Bayesian Models ................................................................. 63
Table 4.13 MAD, MSPE and MSE for NB, FB and IHSPM models ........................................... 65
Table 5.1 Top 20 Identified URBAN CPZs Using EB Method .................................................. 72
Table 5.2 Top 20 Identified RURAL CPZs Using EB Method ................................................. 73
Table 5.3 Top 20 Identified URBAN CPZs Using FB Method .................................................. 77
Table 5.4 Top 20 Identified RURAL CPZs Using FB Method .................................................. 78
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Data Aggregation Module</td>
<td>25</td>
</tr>
<tr>
<td>3.2</td>
<td>Data Aggregation - IHSPM User Interface</td>
<td>26</td>
</tr>
<tr>
<td>3.3</td>
<td>Flow Chart Depicting Model Selection Guidelines</td>
<td>27</td>
</tr>
<tr>
<td>3.4</td>
<td>Scope of Analysis</td>
<td>28</td>
</tr>
<tr>
<td>3.5</td>
<td>Geographic Land-use Type</td>
<td>28</td>
</tr>
<tr>
<td>3.6</td>
<td>Selection of Trigger Variables</td>
<td>29</td>
</tr>
<tr>
<td>3.7</td>
<td>Planning Time Horizon</td>
<td>30</td>
</tr>
<tr>
<td>3.8</td>
<td>Collision Type</td>
<td>32</td>
</tr>
<tr>
<td>3.9</td>
<td>Recommended Model Groups</td>
<td>32</td>
</tr>
<tr>
<td>3.10</td>
<td>Working with IHSPM – Flow Chart</td>
<td>34</td>
</tr>
<tr>
<td>3.11</td>
<td>Model Development Module</td>
<td>35</td>
</tr>
<tr>
<td>3.12</td>
<td>Responses to the Questionnaire</td>
<td>38</td>
</tr>
<tr>
<td>4.1</td>
<td>Study Area: City of Kelowna</td>
<td>44</td>
</tr>
<tr>
<td>5.1</td>
<td>Top 10 Identified <em>Urban</em> CPZs Using EB Method</td>
<td>74</td>
</tr>
<tr>
<td>5.2</td>
<td>Top 10 Identified <em>Rural</em> CPZs Using EB Method</td>
<td>74</td>
</tr>
<tr>
<td>5.3</td>
<td>Top 10 Identified <em>Urban</em> CPZs Using FB Method</td>
<td>79</td>
</tr>
<tr>
<td>5.4</td>
<td>Top 10 Identified <em>Rural</em> CPZs Using FB Method</td>
<td>79</td>
</tr>
<tr>
<td>5.5</td>
<td>Urban CPZ 8</td>
<td>82</td>
</tr>
<tr>
<td>5.6</td>
<td>Urban CPZ 5</td>
<td>86</td>
</tr>
<tr>
<td>5.7</td>
<td>Rural CPZ 44</td>
<td>90</td>
</tr>
<tr>
<td>5.8</td>
<td>Rural CPZ 50</td>
<td>93</td>
</tr>
<tr>
<td>5.9</td>
<td>Increased Neighbourhood Core Size</td>
<td>96</td>
</tr>
<tr>
<td>5.10</td>
<td>Sample Grid Network Pattern Designs</td>
<td>98</td>
</tr>
<tr>
<td>C.1</td>
<td>Login or Register to IHSPM</td>
<td>120</td>
</tr>
<tr>
<td>C.2</td>
<td>Dropdown Navigation</td>
<td>121</td>
</tr>
<tr>
<td>C.3</td>
<td>Profile Access and History Dropdown</td>
<td>121</td>
</tr>
<tr>
<td>C.4</td>
<td>Help Resources Dropdown</td>
<td>122</td>
</tr>
<tr>
<td>C.5</td>
<td>Inputting Variables and Parameter Estimates</td>
<td>126</td>
</tr>
<tr>
<td>C.6</td>
<td>Estimating Collision Frequency Using Developed Models</td>
<td>128</td>
</tr>
</tbody>
</table>
Figure C.7 Ranking of Identified CPZs ................................................................. 129
LIST OF ABBREVIATIONS

AIC  Akaike’s Information Criterion
AREA Zonal Area
ALKP Percentage of Arterial Lane Kilometres
AASHTO American Association of State Transportation Officials
BIC Bayesian Information Criterion
BC Province of British Columbia
CCR Collision Risk Ratio
CD Cook’s Distance
CMF Collision Modification Factor
CORE Zonal Core Area
CRP Core to Zonal Area
CRF Collision Reduction Factor
CPM Collision Prediction Model
CPZ Collision Prone Zone
DIC Deviance Information Criterion
DoF Degree of Freedom
DOT Department Of Transportation
EB Empirical Bayesian
EMPD Zonal Employment Density
FB Full Bayesian
GLM Generalized Linear Regression
GVRD Greater Vancouver Regional District
ICBC Insurance Corporation of British Columbia
IHSPM Interactive High-Level Safety Planning Model
INT Zonal Intersections
INTD Zonal Intersection Density
IALP Percentage of Arterial-Local Intersections
IRBPP Percentage of Roundabouts
LLKP Percentage of Local Lane Kilometres
List of Abbreviations

MAD               Mean Absolute Deviation
MCMC              Markov Chain Monte Carlo Method
MPO               Metropolitan Planning Organization
MLE               Maximum Likelihood Estimation
MSE               Mean Squared Error
MSPE              Mean Squared Prediction Error
NAS               National Academy of Sciences
NB                Negative Binomial
NET               Network
NCHRP             National Corporative Highway Research Program
OR                Odds Ratio
PCR               Potential Collision Reduction
POPD              Zonal Population Density
PLN               Poisson Lognormal
RDCO              Regional District of Central Okanagan
RSP               Road Safety Planning
RSIP              Road Safety Improvement Program
RTM               Regression to the Mean
SD                Socio-Demographics
SD                Model Scaled Deviance
SIGD              Zonal Signal Density
SPF               Safety Performance Function
SRS               Sustainable Road Safety
STS               Sustainable Transport Safety
TAZ               Traffic Analysis Zone
TDM               Transportation Demand Management
TLKM              Total Lane Kilometres
TRB               Transportation Research Board
VKT               Vehicle Kilometres Travelled
VC                Volume to Capacity Ratio
ACKNOWLEDGEMENTS

This thesis has only been possible with the gracious support of many people.

I am deeply grateful to my supervisor Dr. Gordon Lovegrove and co-supervisor Dr. Mahmudur Fatmi for their enthusiastic and expert guidance, constant support, and encouragement in my academic life. I also owe sincere thanks to my supervisory committee members: Dr. Ahmed Idris, Dr. Zheng Liu, and Dr. Stephen O’Leary, for the care with which they reviewed the original manuscript and for their crucial comments. I would also like to acknowledge the help of my professors at UBC, who impacted my graduate study through their coursework.

Furthermore, I want to extend my appreciation to Insurance Corporation of British Columbia, and the City of Kelowna for the dataset and trusted advice, without which the data for research contained in this thesis would not have been realized.

Also, I wish to thank my colleagues and friends in Canada and India for their constant support and encouragement. Last, but not least, I want to thank all my family members for their love and support. My deepest gratitude and appreciation to my parents and my brother, for their unconditional love and support. Their support and encouragement, in the end, have made this thesis possible.
To my parents and brother
CHAPTER 1: INTRODUCTION

1.1 Overview

This chapter sets the layout of this thesis. Section 1.2 reports the road safety problems and economic burden at both the global and national level. Section 1.3 describes the motivation for this research in the form of the thesis problem statement. Section 1.4 expands on the three primary research objectives of this thesis. Section 1.5 concludes this chapter and describes the structure of the remaining thesis.

1.2 Road Safety Problem

Around the world road systems are leading to fatal outcomes. Injuries due to road crashes are considered the ninth worst global epidemics. This exerting an enormous social and economic burden across society (WHO 2015). More than 1.2 million people die on roadways each year in traffic accidents. Up to 50 million people incur non-fatal injuries, encumbering 3% of global GDP (WHO, 2015). Hence, the United Nations announced the decade of action for road safety 2011-2020 to combat this global epidemic (UNGA, 2010).

Despite the falling trend in traffic-related fatalities in high-income countries in recent years, the absolute number of fatalities and the associated societal burden has still been significantly high (Elvik, 2010; WHO, 2015). In Canada for instance, around 120,000 people were injured, and 1,900 people lost their lives in traffic crashes in 2016 (Transport Canada, 2016). In the same year, the Insurance Corporation of British Columbia (ICBC) reported 64,000 casualty crashes, including 273 fatalities, in British Columbia, 10% more than 2015 (ICBC, 2017). So, with the present value of a human life estimated at $8,087,204 (AEL, 2018), these traffic fatalities and crashes pose an immense burden on the Canadian economy, costing the country billions of dollars each year.

Given similar road safety issues worldwide, countries have adopted a safe-system approach to road safety (Welle et al., 2018). Some of these evidence-based road safety policy programs include Vision Zero in Sweden (Belin, Tillgren, & Vedung, 2012), Sustainable Safety in Netherlands (Weijermars & Wegman, 2011), Safe System in Australia (Mooren, L., Grzebieta, R., Job, 2011), Safer Journeys in New Zealand (Ministry of Transport, 2010), and Towards Zero
Chapter 1: Introduction

deaths in U.S (The American Traffic Safety Services Association, 2008). Therefore, to inform these evidence-based relevant regional-level road safety policies with limited capital budget and time, reliable empirical techniques are essential to monitor progress in achieving road safety targets. These practical tools allow road safety authorities, planners and decision makers to have insight into the present and future status of road safety.

1.3 Problem Statement

With advancement in the concept of safety conscious planning, transportation planners and engineers are now proactively considering safety in the transportation planning process to improve the overall safety of the transportation network. Macro-level collision prediction models (CPMs) have been identified as a reliable tool to inform safety conscious planning and have gained considerable attention in the past decade. They evaluate the road safety risk for the entire area (region, city, neighborhood, or traffic zone) instead of individual facilities and goes well with transportation planning models. Moreover, as these models examine the association between zonal collisions and the number of explanatory traffic intensity, network, transportation demand management (TDM) and socio-demographic variables, they enable transportation planners to access the safety implications of planned land use and/or transportation options.

However, the development of macro-level CPMs is a cumbersome, semi-manual, and time-consuming process that requires a sound knowledge of spatial analysis tools (ArcGIS or QGIS), and data handling and analysis tools (SPSS, R, or Python). Most practitioners have neither the time, nor resources, nor expertise to carry out this development process. They will not use such models unless a tool is available that is simpler to use and demonstrates added value to their practice. Interactive-High Level Safety Planning Model (IHSPM) – web-based road safety planning (RSP) tool is one such effort made by the researchers of UBCO Sustainable Transport Safety (STS) lab to automate the development and application of macro-level collision prediction models and will be unveiled in this thesis.

Moreover, the success or failure of macro-level CPMs for one spatial environment depends on 1) the quality of the dataset, 2) appropriate functional model form and 3) time frame of model development and model application (Hadayeghi, 2009). Since, road parameters (road lane kilometres, speed limit), network design (the type of intersections, block size, connectivity),
regional socio-demographic and size of geographical aggregation unit (traffic analysis zones) changes over time, updating macro-level CPMs becomes necessary. Not only this, the updating process may identify new variables that improve model quality and road safety. Khondakar et al. (2010) developed the zonal-level CPMs for the City of Kelowna using 2003 data to examine the association between total collision frequency and the suite of variables belonging to exposure (EXP), socio-demographic (SD), transportation demand management (TDM) and network (NET) variable class. However, due to change in size of aggregation unit in Kelowna, from 372 traffic analysis zones (TAZs) in 2003 to 183 TAZs in 2014, the application of these models in present context will provide unexpected results. The level of aggregation unit affects the quality of aggregated variables which in turns affects the model quality. Therefore, updating macro-level CPMs becomes important to inform relevant future road safety policies. This thesis updates the macro-level CPMs for Kelowna using latest collision, SD and network data for 2014. The section below describes in detail the research objectives of this thesis.

1.4 Research Objectives

1.4.1 Objective 1: Updating Kelowna Community-Based Macro-Level CPMs – Transferring Road Safety Planning Models Across Time

The first research objective of this thesis is to develop a set of updated community-based macro-level CPMs stratified by land use (i.e., urban or rural), independent variable class (i.e., EXP, SD, TDM and NET), and data derivation (i.e., measured or modelled) using the data from 183 traffic analysis zones (TAZs) in Kelowna. Generalized Linear Modelling (GLM) with negative binomial (NB) error distribution and Full Bayesian (FB) modelling techniques are employed to develop the CPMs. The model results are compared to determine if FB methods are better than traditional NB macro-level CPMs, regarding ease of model development while looking at the data and computational needs.
1.4.2 Objective 2: Assessment of Interactive High-Level Safety Planning Model (IHSPM) – Beta Test

The second research objective is to assess IHSPM software developed to date by Sustainable Transport Safety Lab (STS) UBCO is to the practitioners. IHSPM is a web-based RSP tool that automates the development of macro-level CPMs for their application in macro-reactive and proactive road safety analysis. A beta test was conducted with the transportation professionals from the City of Kelowna and research students from STS lab to obtain their feedback on the overall usability of the tool and how this tool can ameliorate the current practice of road safety analysis used by city professionals. Moreover, the outstanding needs to improve IHSPM are identified based on professional feedback from the beta test and are discussed in the thesis.

1.4.3 Objective 3: Macro-Reactive Road Safety Applications – Identifying Collision Prone Zones (CPZs) and Defining Collision Modification Factors (CMFs)

The third objective of this research is to apply both NB and FB CPMs developed in Objective 1 to identify and rank hazardous collision-prone zones (CPZs) in Kelowna. The identified CPZs results from both the regression methods are then compared to identify the ‘most-practical’ and ‘easy-to-use’ identification and ranking methodology. Two identified urban, and two rural CPZs are diagnosed for the road safety problem and relevant evidence-based context specific safety countermeasures are suggested to improve the overall safety of the CPZ. Further, a series of macro-level collision modification factors (CMFs) are defined for various countermeasures to examine their impact on the safety of a traffic zone.

1.5 Thesis Structure

The remaining thesis is organized in chapters 2 through 6. Chapter 2 reviews the literature on regression techniques for the development and application of community-based macro-level collision prediction models. Chapter 3 discusses the need for automation of macro-level RSP tools and attempts made in the past towards automation, describes the working of IHSPM, the theory behind the tool, and its user manual. Chapter 4 presents the data extraction process, methodology and results of TAZ-level CPMs using NB and FB methods, and IHSPM Model Development module. The chapter also presents the beta test results of IHSPM including the users’ feedback obtained and recommended actions to improve outstanding needs. Chapter 5 presents the macro-
Chapter 1: Introduction

reactive road safety applications of community-based macro-level CPMs. Chapter 6 summarizes the research conclusions, contributions and limitations, and outlines the need for future research.
CHAPTER 2: LITERATURE REVIEW

2.1 Overview

This chapter reviews the literature corresponding to the development and application of community-based macro-level collision prediction models (CPMs). The literature search was conducted using Transportation Research International Documentation (TRID), Google Scholar, and Compendex Engineering Village databases. The search keywords used included road safety, road safety planning, macro-level collision prediction models, full Bayesian, generalized linear regression, empirical Bayes, black spot studies, PlanSAFE, automating, beta test, and long-range road safety planning. The selected documents included journal, conference publications, and white paper reports.

The remainder of the chapter is structured as follows: Section 2.2 documents different modelling approaches to macro-level road safety planning and explains the generalized linear modelling (GLM) and full Bayesian (FB) modelling techniques. The section also compares the NB and FB regression techniques regarding the computational requirements and ease of use and discusses the technique of variable selection and model stratification. Section 2.3 reviews the reactive applications of community-based macro-level CPMs including black spot studies and before-and-after studies. Section 2.5 summarizes this chapter.

2.2 Approaches to Macro-Level Collision Prediction Modelling

Around the world, many organizations have started road safety improvement programs (RSIPs) as a priority initiative to identify a suite of factors that contribute to traffic collisions related to road system components: the driver, the vehicle, and the road (G. R. Lovegrove & Sayed, 2006). Transportation Engineering programs primarily focus on the safety performance of the road system components through a reactive or proactive approach (G. Lovegrove & Sayed, 2006). The reactive approach predominantly focuses on the safety improvement of individual facilities in response to pre-existing collision histories. The proactive approach emphasizes evaluating and improving the safety performance of planned facilities to prevent road safety problems from occurring in the first place. Though using a proactive approach focuses on the safety improvement of planned projects, it complements the traditional reactive approach by helping to evaluate and retrofit existing facilities (i.e., macro-reactive applications), with significant safety improvements.
In the past, macro-level CPMs have been identified as reliable empirical tools to conduct macro-reactive and proactive community-level road safety analyses (LOVEGROVE, 2010; Sun & Lovegrove, 2013; Feng Wei & Lovegrove, 2013). Additionally, macro-level CPMs have been found to complement and enhance the benefits of micro-level CPMs in black spot programs (G. R. Lovegrove & Sayed, 2006). Micro-level CPMs are primarily used to evaluate the road safety risk at intersections and road segments. On the contrary, macro-level CPMs help in evaluating the road safety risk for an entire area (neighbourhood, city, region, or TAZ) instead of individual facilities (intersections, road-links), thereby facilitating the earlier detection of collision-prone locations in a neighbourhood. In addition, macro-level CPMs help in problem diagnosis and assessment of effective counter-measure strategies in the identified collision-prone zones (CPZs). Hence, through earlier detection and diagnosis of hazardous locations, macro-level CPMs facilitates more effective allocation of road safety funding.

Given the several benefits of macro-level CPMs in macro-reactive and proactive road safety applications, numerous attempts have been made in the past to develop a functional model form (El-Basyouny & Sayed, 2009; Hadayeghi, Shalaby, & Persaud, 2003; Levine, Kim, & Nitz, 1995; G. R. Lovegrove & Sayed, 2006). Sawalha & Sayed (2006) recommended two conditions to identify appropriate model form.

a) A model must not predict negative crash frequency
b) It must predict zero collision frequency for zero exposure (total lane kilometres or vehicle kilometres travelled).

Given the random component in the observed collision frequency at a location, long term mean collision frequency can never be predicted with absolute accuracy. As such, it can be estimated as an expected value using reasonable assumptions and statistical techniques. Statistical techniques screen out randomness in the observed historic mean collision frequency to provide an estimate of the expected value of the mean collision frequency with some degree of accuracy. Following this linear regression approach proposed by Levine et al. (1995) was found unsuccessful as the traffic crash data does not fit the standard assumptions of linear regression (i.e., normal error structure and constant error variance). However, a generalized linear modelling (GLM) technique
with negative binomial (NB) error structure has emerged in the past decade. The generalized regression overcomes the limitations associated with linear regression in modelling discrete, non-negative, and rare events like traffic collisions (Abdel-Aty, Siddiqui, Huang, & Wang, 2011; G. R. Lovegrove & Sayed, 2006; Naderan & Shahi, 2010; Sun & Lovegrove, 2013; Takyi, Oluwajana, & Park, 2018; Feng Wei & Lovegrove, 2013). Researchers have also used inflated generalized linear modelling techniques including Zero-inflated NB and Hurdle models to account for excess zeros in count data, generally in case of pedestrians and bicyclist collisions (Cai, Lee, Eluru, & Abdel-Aty, 2016). Recently, attempts have been made to develop CPMs using Full Bayesian (FB) techniques to account for spatial/temporal variation and heterogeneity in a collision dataset (El-Basyouny & Sayed, 2009; Osama & Sayed, 2017; Siddiqui, Abdel-Aty, & Choi, 2012). This study develops NB and FB models. Further description of these methods is provided below.

2.2.1 Generalized Linear Modelling

Generalized Linear Modelling is an extension of linear regression to fit models to the data that follow a probability distribution other than normal distribution such as Poisson, negative binomial, gamma distribution, etc. (Feng Wei & Lovegrove, 2013). Since traffic collisions are sporadic, discrete, and non-negative events, Poisson and NB distributions have been employed to model such events (G. Lovegrove & Sayed, 2006; Sawalha & Sayed, 2006; Takyi et al., 2018; V. F. Wei & Lovegrove, 2012). However, NB models are recommended over Poisson models because they account for over dispersion in crash data in comparison with Poisson models, for which the sample variance is equal to the sample mean (G. R. Lovegrove & Sayed, 2006).

In NB regression, given the expected collision frequency $\lambda_i$, observed collision count $y_i$ follows a negative binomial distribution. The probability mass function of NB distribution takes the following form:

$$P(Y_i = y_i | X_i) = \frac{\kappa^\kappa \lambda_i^{y_i} \Gamma(\kappa + y_i)}{\Gamma(y_i + 1) \Gamma(\kappa) (\lambda_i + \kappa)^{\kappa + y_i}}$$  \hspace{1cm} (2.1)

Where the expected collisions $\lambda_i$ for a zone are derived from the developed macro-level CPMs with covariates $X_i$ and shape parameter $\kappa$. Eq. 2.2 and 2.3 present the commonly used model forms (Hadayeghi et al., 2003; Ladrón de Guevara, Washington, & Oh, 2004; G. Lovegrove & Sayed, 2007). The model form presented in Eq. 2.2, proposed by Ladrón et al. (2004) doesn’t
account for the zero risk logic i.e., zero collision frequency for zero traffic exposure. However, the model form proposed by Lovegrove & Sayed (2006) as shown in equation 2.3 not only accounts for the zero risk logic but differentiates the effect of lead exposure variable \(Z\) and other explanatory variables \(X\) on the collision frequency. The two model forms are written as:

\[
E(\Lambda_i) = e^{\sum b_j X_j} 
\]

(2.2)

\[
E(\Lambda_i) = a_0(Z) b_0 e^{\sum b_j X_j} 
\]

(2.3)

Where \(E(\Lambda_i)\) is expected collision frequency for the \(i^{th}\) zone, \(a_0, b_0, b_j\) are model parameters, \(Z\) represents zonal traffic exposure having a dominant influence on collision count (e.g. vehicle kilometre travelled or total lane kilometres), and \(X_j\) represents other explanatory variables (e.g. socio-demographic, economic, and network variables).

Moreover, the model parameter estimates are based on the maximum likelihood estimation (MLE) technique (G. Lovegrove & Sayed, 2007; Sawalha & Sayed, 2006). Scaled Deviance (SD), Pearson \(\chi^2\), and \(\kappa\) are used to describe the overall goodness of fit at 95% level of confidence. Penalized measures, Akaike’s Information Criterion (AIC), and Bayesian Information Criterion (BIC) are also employed for model comparison and selection. Eq. 2.4 and 2.5 represent the formulation of SD and Pearson \(\chi^2\) measures:

\[
SD = 2 \sum_{i=1}^{n} \left[ y_i \ln \left( \frac{y_i}{E(\Lambda_i)} \right) - (y_i + \kappa) \ln \left( \frac{y_i + \kappa}{E(\Lambda_i) + \kappa} \right) \right] 
\]

(2.4)

\[
Pearson \chi^2 = \sum_{i=1}^{n} \frac{[y_i - E(\Lambda_i)]^2}{Var(y_i)} 
\]

(2.5)

\[
Var(y_i) = E(y_i) + \frac{E(y_i)^2}{\kappa} = \lambda_i + \frac{\lambda_i^2}{\kappa} 
\]

(2.6)

Where \(y_i\) represents the observed collision frequency, \(Var(y_i)\) shows the variance of \(y_i\) as per Eq. 2.6, \(E(\Lambda_i)\) represents predicted collision frequency, and \(\kappa\) is the model shape parameter. The models for which the value of SD and Pearson \(\chi^2\) are smaller than the standard \(\chi^2\) value at 95% confidence interval are considered a good fit. Models with smaller AIC and BIC values are also generally a good fit. No minimum value of \(\kappa\) has been recommended but based on models developed in the
past, the value usually exceeds 1.0 (G. Lovegrove & Sayed, 2007).

Moreover, Sawalha & Sayed (2006) discuss the detailed methodology to develop parsimonious and best fit NB models with outlier analysis. As police data contains many errors, the unusual or extreme observations influence the model equation. Cook’s Distance (CD), a measure that reflects how well the model fits the $i^{th}$ observation, is used to conduct outlier analysis (G. R. Lovegrove & Sayed, 2006; Feng Wei & Lovegrove, 2013). As data points with high CD have a significant impact on model parameter estimates, they are considered influential outliers if their stepwise removal from the model’s dataset results in significant drop in model’s SD.

However, parameter estimation using the classical maximum likelihood estimation (MLE) is aimed at finding a single optimum estimate using asymptotic normality based on large sample, at particular confidence interval around its mode. Bayesian parameter estimation, on the other hand, determines the posterior density for each parameter under consideration which provides a considerable interpretive advantage over classical MLE. Full Bayesian modelling technique is discussed in detail in the section below.

2.2.2 Full Bayesian (FB) Modelling

The full Bayesian (FB) technique to collision prediction modelling accounts for heterogeneity in crash data (Aguero-Valverde & Jovanis, 2006; El-Basyouny & Sayed, 2009; Hadayeghi, Shalaby, & Persaud, 2010). Bayesian modelling approach provides better interpretive advantage over MLE as the posterior estimates reflect the probabilities that the analyst is primarily interested in. The output from these models consists of global parameter estimates for the regression variables and individual coefficients for error terms. Collisions are assumed to follow Poisson distribution with parameter $\lambda_i$ as shown in Eq. 2.7.

$$Y_i | \lambda_i = Poisson(\lambda_i)$$ (2.7)

Where $\lambda_i$ denotes mean collision frequency for the $i^{th}$ zone. Gamma and lognormal distribution are commonly used to handle over dispersion in crash data as shown in Eq. 2.8 and 2.9.

$$Poisson gamma models: \lambda_i \sim Gamma\left(\kappa, \frac{\kappa}{\lambda_i}\right)$$ (2.8)
Chapter 2: Literature Review

Poisson lognormal models: $\lambda_i \sim N(Ln(\lambda_i), \sigma^2)$ (2.9)

Where $\kappa$ denotes shape parameter in NB models, and $\sigma^2$ represents extra-variation in Poisson lognormal (PLN) models. Eq. 2.10, and 2.11 shows the mean and variance of $y_i$ for NB models, and Eq. 2.12, and 2.13 for PLN models (F. Wei, 2012).

$$E(y_i) = \lambda_i$$ (2.10)

$$Var(y_i) = \lambda_i + \frac{\lambda_i^2}{\kappa}$$ (2.11)

$$E(y_i) = \lambda_i e^{0.5\sigma^2}$$ (2.12)

$$Var(y_i) = \lambda_i e^{0.5\sigma^2} (e^{\sigma^2} - 1)$$ (2.13)

The Bayesian framework is based on determining the prior distributions for all the unknown parameters to reflect to some degree of knowledge about the parameters before incorporating information from the data (El-Basyouny & Sayed, 2009; Hadayeghi et al., 2010). El-Basyouny & Sayed (2009, 2010) have recommended using normal distributions (with zero mean and large variance) for the regression parameters and gamma distribution ($\Gamma(\epsilon, \epsilon)$ or $\Gamma(1, \epsilon)$) for $\sigma^2$ and $\kappa$, where $\epsilon$ is a small number (0.01 or 0.001). The prior distributions after being combined with observed data are updated to obtain the posterior distribution for all parameters. The statistical inference (mean, standard deviation, and credible intervals) about the regression parameters and expected collisions is then obtained from the posterior distributions. In general, the Bayesian analysis follows three steps (Hadayeghi et al., 2010):

1) Determining appropriate prior distribution for all regression parameters.
2) Conditioning the prior distribution of unknown parameters on observed data.
3) Interpreting the model results (Hadayeghi et al., 2010)(Hadayeghi et al., 2010).

The Markov Chain Monte Carlo (MCMC) method is generally employed to obtain the posterior distribution for parameter estimates, unlike the maximum likelihood estimation used in the GLM technique (El-Basyouny & Sayed, 2009). In this technique, a train of random points is generated, whose distribution is converged to the target posterior distribution. Of the generated
sample points, a sub-sample monitors the model convergence and the remaining iterations are used for parameter estimation, and statistical inference. The samples are generated using the Gibbs sampling (El-Basyouny & Sayed, 2009). Gibbs sampling technique involves parameter ordering and sampling from the conditional distribution for each parameter given the current value of all the other parameters and repeatedly cycling through this updating process (Lynch, 2007).

However, Bayesian model comparisons and selection are based on the Deviance Information Criterion (DIC), a Bayesian generalization of Akaike’s Information Criterion (AIC). The lower the value of DIC, the better the model fits the data. However, it is difficult to determine what constitutes a significant difference in DIC value. El-Basyouny & Sayed (2009) recommended ruling out the model with higher DIC if the difference in DIC values of two models is more than 10. A difference of 5 to 10 is considered substantial, whereas, a difference of less than 5 implies models are not statistically different.

2.2.3 Comparison between the Regression Techniques

Numerous studies have compared the accuracy of GLMs and FB models (El-Basyouny & Sayed, 2009; Hadayeghi et al., 2010). Hadayeghi et al. (2010) found FB models to perform better than conventional NB models as they permit the analysis of the effects of explanatory variables as well as spatial effects in more significant detail. Moreover, El-Basyouny & Sayed (2009), Li et al. (2008) and Persaud & Lyon (2007) conducted the comparison of the development and application results of hierarchical FB and GLMs, and identified three benefits of FB models over GLMs. First, hierarchical FB models account for heterogeneity in crash data as well as spatial/temporal uncertainties and interaction among covariates. Second, FB models are more flexible in selecting distribution for crash counts (i.e., both Poisson-gamma and Poisson lognormal distribution can be used). Lastly, FB methods offer a more integrated procedure for macro-reactive applications in comparison to the GLM technique which requires separate steps.

However, the development of NB models is a relatively simple process in comparison to FB models. Various data analysis tools including SPSS and R offer generalized linear modelling packages for simple development of NB models with well structured and directly usable model outputs. Moreover, FB modelling tools like WinBUGS require advanced knowledge of probabilistic analysis. Given the software’s sophisticated user interface, FB model development
becomes a tedious and time-consuming process requiring manual data handling and model development. Furthermore, the outlier analysis in WinBUGS involves the use of other tools like SPSS to identify outliers, which inhibits its use as an independent tool.

Three measures are generally used to compare the accuracy of GLMs and FB models (Hadayeghi, Shalaby, Persaud, & Cheung, 2006). The first measure, mean absolute deviation (MAD) as shown in Eq. 2.14 measures average model misprediction. The other two measures: mean squared prediction error (MSPE), and mean squared error (MSE) as shown in Eq. 2.15 and 2.16, show potential over or under fitting of models when compared among themselves. For example, the value of MSE of a model being higher than the MSPE of another model indicates that the first model may have been over fitted to the estimation data and the observed relationships may not be correctly represented (Hadayeghi, 2009). MAD, MSPE, and MSE are formulated as follows:

\[
MAD = \frac{\sum_{i=1}^{n} |Y_i - y_i|}{n}
\]  
(2.14)

\[
MSPE = \frac{\sum_{i=1}^{n} (Y_i - y_i)^2}{n}
\]  
(2.15)

\[
MSE = \frac{\sum_{i=1}^{n} (Y_i - y_i)^2}{n - p}
\]  
(2.16)

Where, \(Y_i\) denotes predicted collisions at zone \(i\), \(y_i\) is the observed collisions at zone \(i\), \(n\) is the data sample size, and \(p\) is the number of parameters in a model.

2.2.4 Variable Selection and Supporting Evidence

Different patterns of zonal crash frequencies have been used as the response variable in neighbourhood-level CPM studies (Hadayeghi, Shalaby, & Persaud, 2007; G. R. Lovegrove & Sayed, 2006). Based on the crashes modelled in past research, the response variable can be grouped by collision severity (fatality collisions, injury collisions or property-damage-only collisions) and collision victim type (pedestrian crashes, bicycle crashes or vehicle crashes). The explanatory variables can be mainly grouped into the following variable classes: traffic exposure (EXP), roadway network (NET), socio-demographics (SD), and transportation demand management.
(TDM). Though the selection of appropriate explanatory variables is based on the quality and adequacy of the datasets, expenses involved in obtaining data, as well as variable definitions and relevance of variables used in RSP (G. R. Lovegrove & Sayed, 2006), the retention of a variable in a model depends on the logic (+/-) of parameter estimates, t-statistic values, and the effect of a variable on model’s SD. Given that, exposure variables used in past research include vehicle kilometre travelled (VKT) and total lane kilometres (TLKM) as these variables provide surrogates for road traffic exposure and have a strong positive association with the response variables (Hadayeghi et al., 2007; G. Lovegrove & Sayed, 2007; V. F. Wei & Lovegrove, 2012).

Moreover, of different socio-demographics related variables used in previous studies, collisions were found to be positively associated with zonal population density, job density, and residential unit density (G. R. Lovegrove & Sayed, 2006). For TDM related variables, collisions were found to be positively associated with total commuter density, zonal core area, and shortcut attractiveness (G. R. Lovegrove & Sayed, 2006). Network related variables including intersection density, number of intersections per lane kilometre, total arterial lane kilometres, signalized intersection density, and arterial-local intersections showed a positive association with collisions (G. R. Lovegrove & Sayed, 2006). Whereas, other network variables including local lane kilometres and three-way intersections exhibited a negative association with collisions (G. R. Lovegrove & Sayed, 2006).

However, to improve model performances, Lovegrove et al. (2006) recommended following model groupings stratified by land use (i.e., urban or rural), independent variable class (i.e., EXP, SD, TDM and NET), and data derivation (i.e., measured or modelled) as shown in Table 2.1. So, for each variable theme there are four model groups, two for each urban and rural TAZs. The modelled theme pertains to modelled exposure variable such as VKT and measured model theme pertains to measured exposure variable such as TLKM.
Chapter 2: Literature Review

Table 2.1 Recommended Model Groupings (G. R. Lovegrove & Sayed, 2006).

<table>
<thead>
<tr>
<th>Variable Theme</th>
<th>Land Use</th>
<th>Data Derivation</th>
<th>Group#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure</td>
<td>Urban</td>
<td>Modelled</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measured</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>Modelled</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measured</td>
<td>4</td>
</tr>
<tr>
<td>Socio-Demographics</td>
<td>Urban</td>
<td>Modelled</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measured</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>Modelled</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measured</td>
<td>8</td>
</tr>
<tr>
<td>TDM</td>
<td>Urban</td>
<td>Modelled</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measured</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>Modelled</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measured</td>
<td>12</td>
</tr>
<tr>
<td>Network</td>
<td>Urban</td>
<td>Modelled</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measured</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>Modelled</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measured</td>
<td>16</td>
</tr>
</tbody>
</table>

2.3 Macro-Reactive Applications

Research to date has demonstrated the potential use of NB community-based macro-level CPMs in reactive road safety applications (i.e., in black spot studies) and before-and-after studies (Hadayeghi, Shalaby, & Persaud, 2007; G. Lovegrove & Sayed, 2006, 2007). In macro-reactive studies, black spots are the collision-prone zones (CPZs), with collision frequency significantly above average collision frequency at 95% confidence. Black spot studies are employed to identify and rank CPZs to diagnose and remediate hazardous locations. On the other hand, before and after studies are used to evaluate the effectiveness of road safety countermeasures.

Numerous studies have presented macro-reactive applications of CPMs. Lovegrove & Sayed (2006) used NB models to identify, rank, diagnose, and remediate CPZs in the Greater Vancouver Regional District (GVRD). Furthermore, the study estimated the area-wide traffic calming modification factor for urban areas in GVRD and compared the road safety risk of four neighbourhood design patterns. Sun & Lovegrove (2013) used the macro-level CPMs application guidelines proposed by Lovegrove & Sayed (2006) to compare the level of road safety of Fused Grid network pattern with four other networks including grid, cul-de-sac, 3-way offset, and Dutch sustainable road safety (SRS) patterns. Hadayeghi, Shalaby, & Persaud (2007) used the NB models to develop a series of macro-level collision modification factors (CMFs) for total and severe.
collisions to examine the impact of each safety planning explanatory variable on zonal road safety. F Wei (2012) applied both NB and PLN models to identify, rank, diagnose, and remediate CPZs for the Regional District of Central Okanagan (RDCO).

Appropriate techniques to identify and rank CPZs depend on the functional form of the macro-level CPMs. Empirical Bayes (EB) has been identified as the most reliable technique to identify CPZs when using GLMs (Hauer, 1992; G. Lovegrove & Sayed, 2006; V. F. Wei & Lovegrove, 2012). On the other hand, F Wei (2012) discusses the CPZ identification methodology for FB models. The following section discusses the methodology recommended for black spot studies and before-and-after studies in detail.

### 2.3.1 Black Spot Studies

As black spots are usually identified based on the recorded high occurrence of collisions, the earlier detection technique, regression-to-mean (RTM) bias leads to ‘false’ labelling of hazardous locations (Hauer et al., 2002; G. Lovegrove & Sayed, 2006). Therefore, to reduce the bias, the EB method expressed as a combination of observed collision frequencies and estimated collision frequencies was suggested (Hauer, 1992). The EB method follows three steps to identify CPZs or black spots. The first step involves identifying collision priors using CPMs. For NB models, the prior distribution of predicted collisions is assumed to be a gamma distribution with shape parameter \( \kappa \) and scale parameter \( \frac{\kappa}{E(A)} \). In the second step, posterior collision estimates are obtained by combining the prior collision estimates from NB models with the locally observed collision data using Bayes’ theorem. Bayes’ theorem is a formula that describes how to update the probabilities of hypotheses when given evidence (Swinburne, 2008). The posterior distribution for each site is also assumed to follow a gamma distribution with the shape and scale parameter shown in Eq. 2.17-a and 2.17-b (Hauer et al., 1992, 2002).

\[
\alpha = \kappa + \text{count} \quad (2.17-a)
\]

\[
\beta = \frac{\kappa}{E(A_i)} + 1 \quad (2.17-b)
\]

The mean and variance of site specific EB collisions are represented in Eq. 2.18-a and 2.18-b:
Chapter 2: Literature Review

\[ EB_i = E(\Lambda|Y = \text{count}) = \frac{\alpha}{\beta} = \left[ \frac{E(\Lambda_i)}{\kappa + E(\Lambda_i)} \right] (\kappa + \text{count}) \]  
\[ (2.18-a) \]

\[ Var(EB_i) = Var(\Lambda|Y = \text{count}) = \frac{\alpha}{\beta^2} = \left[ \frac{E(\Lambda_i)}{\kappa + E(\Lambda_i)} \right]^2 (\kappa + \text{count}) \]  
\[ (2.18-b) \]

The final step involves the comparison of prior collision estimates derived from NB models with their corresponding EB estimate. If the value of EB for one zone exceeds its \( E(\Lambda) \) at 95% confidence, the zone is considered a CPZ.

Two ranking criteria: Potential Collision Reduction (PCR) and Collision Risk Ratio (CRR) were suggested by Lovegrove & Sayed (2007) to rank CPZs and the two measures are formulated as follows:

\[ PCR = EB - E(\Lambda) \]  
\[ (2.19) \]

\[ CRR = \frac{EB}{E(\Lambda)} \]  
\[ (2.20) \]

Identified and ranked CPZs may not agree with each other for different model groups of CPMs, as the ranks are summed across multiple CPMs to obtain the total ranking score of each zone. Zones with the highest-ranking scores are those in need of the most attention. Following this, CPZs are diagnosed and remedied based on the guidelines proposed by G. Lovegrove & Sayed (2007).

Moreover, F Wei (2012) proposed the identification and ranking methodology for FB models. As in the FB method, posterior distribution for traffic crashes are obtained directly from the FB models. The zone is identified as CPZ if the posterior mean of collisions in one zone exceed its normal estimate at 95% confidence. The posterior collision estimates are interpreted from fitted distributions and normal estimates are obtained by averaging the possible values relative to all posterior distributions. The criteria can be mathematically represented as:

\[ \lambda_i(5\%) > E(y_i) \]  
\[ (2.21) \]
Where $\lambda_i(5\%)$ represents the 5% posterior value of collisions for zone $i$, and $E(y_i)$ is the normal collision estimate for zone $i$. Similar to EB, PCR and CRR are also used to rank CPZs in FB method. The criteria are formulated as follows:

\[
P_{CR} = \lambda_i(\text{mean}) - E(y_i) \quad (2.22)
\]
\[
C_{RR} = \lambda_i(\text{mean}) - E(y_i) \quad (2.23)
\]

Where $\lambda_i(\text{mean})$ is the posterior mean value of collisions and $E(y_i)$ is the normal collision estimate for zone $i$. The application results in Wei (2012) revealed that the FB method agrees with the EB method in CPZ identification as a similarity of 97% was observed among the identified CPZs results using two methods.

### 2.3.2 Before-and-After Studies

Before and after studies are conducted to evaluate the effectiveness of different countermeasure strategies implemented on CPZ identification and problem diagnosis. The effectiveness of countermeasures is usually indicated by percentage reduction in collisions, also known as Collision Reduction Factor (CRF). CRF subtracted from unity (1-CRF) is referred to as collision modification factor (CMF) or odds ratio (OR). Hauer (1997), Sayed et al. (2001), and Lovegrove et al. (2006) have discussed the methodology used to estimate the CMF of any countermeasures based on the CPMs and EB method. The OR is expressed as:

\[
OR = \frac{A/C}{B/D} \quad (2.24)
\]

Where $A/C$ represents the collision frequencies in the before/after period of implementation of countermeasure in the comparison group, $B$ represents the EB collision frequency estimate in the after period with no treatment in the subject site, and $D$ represents the EB estimate in the after period after the implementation of countermeasures in the subject site. The comparison group is used to lower the time trend bias from the before to after period. However, the data for the comparison group is obtained from randomly selected sample sites. All other quantities except $B$ are observed collision frequencies, where $B$ is computed using the EB safety estimate of the subject site in the before period, and the collision frequency of the subject site estimated using CPM with exposure in the before-and-after period (Hauer, 1997).
2.5 Summary

This chapter presented the fundamentals of macro-level CPMs, different modelling forms proposed in the past and modelling techniques adopted. Negative binomial (NB) and full Bayesian (FB) techniques were found appropriate when modelling total or severe collisions. Further, the chapter discussed the macro-reactive applications including black-spot studies and before-after studies of community-level CPMs. EB and FB methodologies for identification of collision-prone zones were discussed. However, the need for automation of macro-level CPMs for ease of use by transportation planners in safety conscious planning is discussed in the next chapter.

The information from this chapter is used to update and compare the NB and FB macro-level CPMs for the City of Kelowna as presented in chapter 4. The developed models are applied to identify, rank, diagnose, and remedy collision-prone zones in the city and are presented in chapter 5. Chapter 3 presents the web-based macro-level RSP tool, IHSPM that fills the research gap by automating the development and application of macro-level CPMs.
CHAPTER 3: ASSESSMENT OF INTERACTIVE HIGH-LEVEL SAFETY PLANNING MODEL (IHSPM)

3.1 Overview

This chapter explains the Interactive High-Level Safety Planning Model (IHSPM), an online Road Safety Planning (RSP) tool developed by researchers at the Sustainable Transport Safety (STS) Research Lab UBCO. IHSPM automates the development and application of community-based macro-level CPMs. This chapter is structured as follows: section 3.2 explains the need for automation in the development of CPMs, and attempts made in past towards automation. Section 3.3 provides a brief overview of IHSPM and the software development architecture used in the back-end and front-end for the development of different components of the tool. Section 3.4 explains different IHSPM modules including Map Data Aggregation, Walkthrough and CPM Development. Section 3.5 presents the IHSPM beta test method and results, including outstanding needs and recommendations to improve it. Section 3.6 presents summary of this chapter.

3.2 Need for Automation

With advancement in the concept of safety conscious planning, transportation planners and engineers are now explicitly considering evidence-based safety measures in the transportation planning process. Many countries and jurisdictions are setting quantitative road safety targets, such as a 50% reduction in fatalities in ten years (Wegman et al., 2015). Macro-level CPMs have been identified as realistic tools in setting, monitoring, and evaluating such road safety targets. However, the development of macro-level CPMs is a cumbersome, semi-manual, and time-consuming process that requires a sound knowledge of spatial analysis (ArcGIS or QGIS) and data analysis tools (SPSS, R, or Python). Most practitioners have neither the time, resources, nor expertise to carry out this process. They will not use such models unless a tool is available that is simpler to use and demonstrates added value to their practice.

3.2.1 Attempts made in Past towards Automation

Several research projects have been undertaken to automate the development and application of macro-CPMs. The first project was launched under the auspices of the US National Academy of Sciences (NAS) in concert with the American Association of State Transportation
Officials (AASHTO) and a steering panel comprised of nominated TRB members. A National Cooperative Highway Research Project (NCHRP 8-44) was initiated around 2010, known as PLANSAFE (discussed in detail in section 3.2.2), at the University of Arizona with field trials in Florida. However, progress stalled at the initial validation tests, which suggested problems with the software coding approach that could not be identified (Washington, 2006). Most recently, NCHRP 17-81 was awarded in Fall 2017 to research data sources, survey RSP practitioner needs, and recommend a revised automation software approach for macro-level CPM development and application in RSP. NCHRP 17-81 is expected to take 18 to 24 months until results are available. Meanwhile another effort at software development for automated macro-level CPM development and application has been ongoing at the UBCO Sustainable Transport Safety (STS) Research Lab since 2012, based on original work of Lovegrove (2007).

3.2.2 PLANSAFE

PLANSAFE was developed as a planning-level safety prediction modelling software with intent to support regional and statewide safety planning activities. Though NCHRP 546 describes the complete software in detail, the software was developed with the motivation to provide MPOs and DOTs with a tool to develop macro-level (TAZ, neighborhood, or region) log linear crash frequency models (considering Poisson distribution) to comment on road safety of an area based on its traffic exposure, socio-demographic, and economic characteristics. However, software initial validation tests suggested some problems with the software functioning.

3.2.2.1 Expertise Required to use PLANSAFE

As described in NCHRP 546, PLANSAFE required its user to have expertise with GIS-based software such as ArcGIS to prepare the dataset necessary to develop crash frequency models and sound knowledge of statistics and statistical software tools such as LIMDEP and STATA to develop the CPMs and interpret the modelling results.

3.2.2.2 Beta Test and Results

Two Florida MPO’s (Metroplan Orlando and Volusia TPO) with DOT oversight were recruited by the team PLANSAFE to test the software. The MPO and DOT staff was trained by the team PLANSAFE on the operations of the tool to implement it for data preparation and software calibration. The primary objectives of the beta test were to refine and update the algorithm
in the software, test and refine the graphical user interface of the tool, form an initial practitioner user group and identify user issues, long-term maintenance requirements of the software and to prepare software for the wide range adoption and implementation based on the user feedback.

Based on the feedback obtained from MPO staff, several problems with the software coding were identified due to the following reasons. First, it was suggested that the models developed by PLANSAFE require some more explanatory variables to accurately explain traffic safety of an area, which were not available or were difficult to forecast. Second, the testing team identified inappropriate planning level mitigation measures provided by the tool based on analysis. Third, software exhibited some problems with the installation process for certain versions of Windows and ArcGIS.

Along with limitations described above, the log linear crash frequency models developed by software doesn’t accounts for the zero-risk logic (i.e., zero collisions for zero traffic exposure) and the Poisson distribution assumption of crash data doesn’t account for extra-variation in the crash dataset. This indicates that despite the technical issues with the software, the inappropriate model form and error distribution assumption questions the usability of the software regarding the model accuracy and use in RSP applications.

Interactive High-Level Safety Planning Model (IHSPM), a web-based tool that automates the development and application of community-based macro-level CPMs, developed to date at the Sustainable Transport Safety (STS) lab overcomes some of the limitations associated with PLANSAFE. The online user interface of the tool introduces easy accessibility and eliminates the need for specific system requirements to run the software as exhibited by PLANSAFE. Moreover, the tool overcomes the limitations identified in PLANSAFE regarding the inappropriate model form and lack of explanatory variables. IHSPM develops NB models that account for extra-variation in count data and the developed model form accounts for zero-risk logic. Moreover, more explanatory variable themes including exposure, SD, TDM, and network are considered for the model development than just exposure and socio-demographic variables as considered in PLANSAFE. However, the software is explained in detail in the following sections below.
3.3 IHSPM

IHSPM is a web-based tool that facilitates the automatic development and application of community-based macro-level CPMs in transport safety planning. The development of these CPMs is based on the methodology defined by Lovegrove & Sayed (2006), and Hadayeghi et al. (2007). The software requires the users to start by creating an account and uploading their data. Once registered, users can upload traffic analysis zones (TAZs), road network data, socio-demographic data, TDM data, traffic volume data, and crash data. IHSPM then creates an interactive map that provides a spatial representation of the geographical region under analysis. The user interface of the tool leads the user through a series of questions on the purpose and particulars of their desired analysis, to help them narrow down the list of variables and model groups on which CPM development proceeds. Once CPMs are developed from the data, using the IHSPM software, they can be applied to analyze communities at the user’s desired scope – a neighborhood or development site, a city quadrant, or an entire region. IHSPM is comprised of three modules: ‘Map Data Aggregation’, ‘Walkthrough’ and ‘CPM Development Module’, which are detailed in section 3.4.

The following software and statistical languages were used to develop different back-end and front-end components of IHSPM.

3.3.1 Languages Used

R (Team R.C., 2013), statistical language was used for all the calculations in the CPM development module. PHP (Nixon, 2014), server-side web-development language was used for most of the back-end coding of the website. HTML (Graham, 1995), markup language, was used to illustrate the syntactic elements of the website. All visible pages were created using either raw HTML or HTML generated JavaScript (Flanagan, 2006) or PHP code. JavaScript, client-side web development language was used for a portion of the backend of the website. JavaScript is used for processes which are better run on a user’s computer, as well as most processes which disallow refreshing the page. Much of the data aggregation module relies on intensive JavaScript processes. Lastly, SQL (Hacigümüş et al., 2002), relational language was used to communicate with the database to store and retrieve data.
Chapter 3: Assessment of IHSPM

3.3.2 Implementations

The R processes needed are run on the server after being called from a PHP shell execution command. The output is returned and parsed so that R communicates with the program via a series of text inputs and outputs, rather than running natively as a part of the website. The R code run requires the modules RMySQL, brew, and MASS. These packages are used to communicate with the database, interact with a web environment, and run various regressions, respectively.

PHP is implemented on the server using the Laravel framework. The design and implementation adhere strictly to the Model-View-Controller style design pattern. The code implements and actively utilizes many of Laravel’s built-in modules, as well as the add-on “LaravelExcel,” created and maintained by Maatwebsite. This additional module is used to process the commonly used data formats “.csv,” “.xls,” and “.xlsx,” which would be generated by Microsoft Excel.

HTML is implemented on the server using HTML5, adhering to modern standards. JavaScript is implemented both in raw form and using jQuery 1.12 to handle button actions, UI component behaviour, and most of the logic within the Data Aggregation page. The server uses MySQL to handle relational databases.

3.4 IHSPM Modules

IHSPM can be accessed through www.ihspm.ok.ubc.ca. Its three modules (Data Aggregation, Walkthrough, and Table Data) are discussed in the following subsections.

3.4.1 Map Data Aggregation Module

As the name suggests Map Data Aggregation module under Develop Models permits the aggregation of raw data to the desired aggregation scale. The spatial tab on homepage requires variable data input, in geo-referenced formats, including traffic analysis zones (TAZs) from transportation models, networks and road centerline from digital road atlas (DRA), crash data from insurance companies (e.g. ICBC in the province of BC, Canada), land use data (e.g. build footprints, zoning) from civic planners, and socio-demographics from federal census bureaus. Intermediate automated steps are required to post-process some of the data, as follows:
1. Census data requires translation from the census tract (CT) aggregation units into TAZ aggregation units. This often requires analysis of how CT boundaries overlap and/or fit into TAZ boundaries, coupled with the general land use and location of dwelling units in each TAZ; and,

2. Association and aggregation of modelled network links (and their associated input/output attributes – speed, volume, capacity, travel time) within each TAZ.

(Note: The uploaded data is stored on the UBCO server for data security. By uploading the dataset in IHSPM, the user permits the STS lab to keep a copy of the uploaded data and to use it for research purposes. However, any uploaded information is not shared with a third party. Users are highly encouraged not to use this software if they are not flexible in sharing their dataset.)

IHSPM has pre-set defaults that can automate and aggregate this process, but also allows the user to examine what it does and make adjustments to improve model accuracy. Figure 3.1 illustrates the data aggregation module. The output of the Data Aggregation module is a fully integrated, geo-referenced dataset containing all relevant information for each TAZ in a MS Excel spreadsheet format.

![Figure 3.1 Data Aggregation Module](image-url)
The **Data Aggregation** module accepts the input data shapefiles (TAZ, DRA and crash data) in ‘.zip.’ format. The input data is parsed using JavaScript and ArcGIS and displayed graphically on a map. Once all required data is uploaded, the program aggregates the data into a matrix. Additionally, total lane kilometers (TLKM) and number of crashes in each TAZ are calculated by the software and added as columns to the matrix. The complete dataset can be passed directly to the **CPM Development** module or downloaded as a CSV file. Figure 3.2 depicts the interface of this module.

![Data Aggregation - IHSPM User Interface](image)

**Figure 3.2 Data Aggregation - IHSPM User Interface**

Buttons (1), (2) and (3) are the data upload buttons for input data (shapefiles - TAZ, DRA, and Crash data respectively). Button (4) is a miscellaneous data upload button for specialized analysis. Button (5) compiles the input data. The compiled excel spreadsheet can be saved using button (6). Button (7) redirects users to the development module of CPMs using aggregated dataset.
3.4.2 Walkthrough Module

The *Walkthrough* module follows a six-step selection checklist to help users in selecting the appropriate models for safety applications based on the guidelines recommended by Lovegrove & Sayed (2007). The response to each selection determines the appropriate models, allowing for a multitude of differing analysis types. Once a response to each question is selected, the ‘Next’ button proceeds to the next question to refine the model selection. Figure 3.3 represents the flow chart for the model selection guidelines.

![Flowchart of model selection guidelines](image)

**Figure 3.3 Flow Chart Depicting Model Selection Guidelines**

### Step 1. Scope of Analysis

The first step is related to the scope of analysis. This scope can either be micro-level or macro-level. If the analysis is confined to a single intersection or road segment, the scope size is considered micro-level. In this case, micro-level analysis using automated software can be conducted via the Interactive Highway Safety Design Model (IHSDM), available at www.highwaysafetymanual.org. However, if the scope of analysis is at a neighbourhood, municipality, or regional level, it is considered macro-level, and models can be developed and implemented within the IHSPM application, available via the UBCO STS Research Lab web site. Further clarification regarding macro scope size is as follows: neighbourhood scope containing transport systems and land use planning; municipal scopes pertain to overall community plans; and regional scopes are used for land use, road design, and transportation demand management planning.
Further, depending on the time-space of evaluation, selection of appropriate models depends on the type of analysis – reactive (e.g. black spots) or proactive (e.g. land use plan, road plan) as shown in figure 3.4. If reactive, a Collision Modification Factor (CMF) can also be used depending on the availability of previously developed CPMs. If proactive, application is predominantly used for regional planning, neighbourhood planning, and model transference. Further, model groups are considered based on the predominant land use, relevant variables, planning time frame, and available data. For the beta test, the application has been trained for the macro-reactive analysis.

**Step 2. Geographic Land Use Type**

After defining the application scope in step 1, step 2 requires the user to define the predominant type of land use in each neighbourhood under evaluation. This can be either urban or rural, depending on the type of development.
Step 3. Selection of Trigger Variables

The third step is related to the selection of trigger variables (exposure, socio-demographics, transport demand management and network) relevant to the analysis based on the planning topic (land use only, road only or TDM only) and comparison type as shown in Table 3.1 (Lovegrove & Sayed, 2006). The comparison type depends on time scope of planning, i.e., whether it includes the comparison of change and its effects (Present vs. Future) or a variety of different alternatives for planning (Future vs. Future). Given the desired planning scope and comparison type, the tool identifies relevant model groups for development and safety application. Figure 3.6 shows the IHSPM view for this step.

Table 3.1 Recommended Model Groups Based on Planning Topic and Comparison

<table>
<thead>
<tr>
<th>Time Horizon</th>
<th>Planning Topic(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison Type</td>
<td>Land Use Only</td>
</tr>
<tr>
<td>Future vs. Present</td>
<td>S-D, TDM, Network</td>
</tr>
<tr>
<td>Future vs. Future</td>
<td>S-D</td>
</tr>
</tbody>
</table>

Figure 3.6 Selection of Trigger Variables
**Step 4. Time Horizon**

Having identified the trigger variables in step 3, step 4 asks user input on the time frame of planning, i.e., whether it is a short-term plan (< 5 years), medium term plan (5 to 15 years), or long-term plan (>15 years). This consideration identifies recommended exposure data type (modelled or measured) for model accuracy and recommends data forecasting techniques for planning time horizon as shown in Table 3.2. As modelled (VKT + VC) and measured (TLKM) exposure variables influence CPM estimates and the model accuracy for particular safety applications, the selection of exposure data type further depends on the scope of analysis (regional or neighbourhood level) and time horizon which has been discussed in detail by Lovegrove & Sayed (2006). Figure 3.7 shows the IHSPM view of this step.

![Figure 3.7 Planning Time Horizon](image)

**Figure 3.7 Planning Time Horizon**
### Table 3.2 Data Extraction Efforts for Planning Analysis, Based on Theme and Time Horizon

<table>
<thead>
<tr>
<th>Data Theme/Model Group</th>
<th>Desired Planning Time Horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short term</td>
</tr>
<tr>
<td></td>
<td>&lt; 5 Years</td>
</tr>
<tr>
<td><strong>Measured Exposure</strong></td>
<td>Aggregate manually if small area, or use GIS software</td>
</tr>
<tr>
<td>(TLKM)</td>
<td></td>
</tr>
<tr>
<td><strong>Modelled Exposure</strong></td>
<td>Only available through VISUM type models</td>
</tr>
<tr>
<td>(VKT)</td>
<td></td>
</tr>
<tr>
<td><strong>S-D (Measured)</strong></td>
<td>Extrapolate for most recent census, using historical trends</td>
</tr>
<tr>
<td>POPD, WKGD, FS, NHD, UMEMP</td>
<td></td>
</tr>
<tr>
<td><strong>TDM (Measured)</strong></td>
<td>From the census, and/or maps. Extract manually on zone by zone basis at all levels</td>
</tr>
<tr>
<td>CORE, CRP, SCC, SCVC, DRIVE, BIKE, TCM</td>
<td></td>
</tr>
<tr>
<td><strong>Network (Measured)</strong></td>
<td>Aggregate manually if small area, or use GIS software</td>
</tr>
<tr>
<td>INTD, SIGN, I3WP, IALP, LLKP, ALKP, IRBP</td>
<td></td>
</tr>
</tbody>
</table>

**Step 5. Collision Type**

This step is related to the type of response variable and scale of aggregation unit (TAZ or census tract). Possible collision selections include severe collisions, total collisions, pedestrian collisions, AM collisions, AM/PM collisions, bike collisions, or non-rush hour collisions. However, the beta release of this application is limited to the total collision, severe collisions, and others. In future releases, analyses with the less commonly analyzed collision types including bicycle, pedestrians and AM/PM may be supported.
Step 6. Recommended Model Group

Based on the response to the previous five steps, step 6 recommends model groups relevant to the analysis as shown in figure 3.9. Below each group heading are the variables that make-up each model, running in a left to the right manner with a Develop Model link at the rightmost side. When clicked, the user is directed to the Model Development module as discussed in the next section.
3.4.3 CPM Development Module

The model development module can be accessed by clicking on either Table data under drop-down navigation panel on Homepage (Figure C.2) or Develop Models tab on the Map Data Aggregation page (Figure 3.2) or Develop this Model tab on Step 6 of the Walkthrough process (Figure 3.9). However, the selection depends on the quality of dataset and the prior knowledge of CPMs. Figure 3.10 illustrates the selection criteria based on the quality of the dataset available with the users’ prior knowledge of CPMs.

The model development page shows the recently uploaded or aggregated dataset, in table format as shown in Figure 3.11. On this page, the user can select specific variables relevant to the model development based on the model groups recommended in step 6 of the Walkthrough process. Moreover, the user can also save the table data, upload new data (if they already have geo-referenced and aggregated the dataset), edit existing data, and view model results. The components of this module are discussed in Appendix C in detail.
Chapter 3: Assessment of IHSPM

Data Sets

Raw or Unprocessed Data

Already processed & Geo-referenced

Require

Data Aggregation

If no prior knowledge of CPMs

Compile/Save

If prior knowledge of CPMs

Develop Models

Walkthrough

Follow steps

IHSPM Menu

Table Data – Develop Models

A) Scope of Analysis
B) Reactive or Pro-active
C) Land Use type
D) Trigger Variables
E) Collision Type
F) Recommended Models

A) Associate links to TAZs
B) Associate Census Tracts to TAZ
C) Calculate SCC and CORE

1) Collision Data
2) Exposure
3) Socio-Demographics
4) Network
5) TDM

Figure 3.10 Working with IHSPM – Flow Chart

(Note: Colored boxes explain the steps involved in the associated white boxes.)
### Figure 3.11 Model Development Module.

<table>
<thead>
<tr>
<th>rownum</th>
<th>objectid</th>
<th>zone_numb</th>
<th>csduid</th>
<th>muni</th>
<th>nmed_2014</th>
<th>media_2014</th>
<th>pop_2014</th>
<th>popn0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td>1590</td>
<td>5935010</td>
<td>Kelowna</td>
<td>31524</td>
<td>28115</td>
<td>518</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>1650</td>
<td>5935010</td>
<td>Kelowna</td>
<td>84860</td>
<td>75682</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>3000</td>
<td>5935010</td>
<td>Kelowna</td>
<td>59549</td>
<td>53108</td>
<td>443</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>3150</td>
<td>5935010</td>
<td>Kelowna</td>
<td>40980</td>
<td>36548</td>
<td>646</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>3180</td>
<td>5935010</td>
<td>Kelowna</td>
<td>58467</td>
<td>52143</td>
<td>425</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>3500</td>
<td>5935010</td>
<td>Kelowna</td>
<td>77088</td>
<td>68750</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>25</td>
<td>2100</td>
<td>5935010</td>
<td>Kelowna</td>
<td>31689</td>
<td>28261</td>
<td>1585</td>
<td>86</td>
</tr>
</tbody>
</table>
Chapter 3: Assessment of IHSPM

3.5 IHSPM Beta Test

The initial round of beta testing of the tool was conducted to assess the ease of use of the tool for transportation planners, road safety practitioners and students. Beta test is a second face of software testing in which a sampling of the intended audience tries the product out. The beta test was conducted with the transportation professionals from the City of Kelowna and research students from the Sustainable Transport Safety (STS) research laboratory, UBC Okanagan. The following subsections discuss the beta test methodology, the user feedback obtained, and the outstanding needs identified to improve IHSPM.

3.5.1 Methodology

The current version of IHSPM is limited to the development of community-based macro-level CPM. For the beta test, users were asked to test the program by applying the ‘Map Data Aggregation,’ ‘Walkthrough’ and ‘CPM development’ modules as discussed in section 3.4. Users were provided with georeferenced TAZ, road centreline, and collision data for Kelowna for the year 2014 along with an instruction sheet to develop the macro-level CPMs using the three modules. For the macro-reactive application of the developed models, beta testers were provided with an automated excel spreadsheet to identify and rank collision prone zones as discussed in Appendix C.3. The beta test was designed in such a way that users get an idea about the novelty of each module and how to use them. Figure 3.10 explains the use of different modules depending on the quality of dataset and prior knowledge of CPMs. Finally, after the development and application of models, users were asked to compare their final models with the provided models to see if they were successful in using the system correctly. Erroneous results would imply that either there is a bug in the code, or the beta tester was unable to understand the program’s user interface fully. At the end of the beta test, beta testers were provided with a questionnaire (see section 3.5.2) to capture their feedback on overall system performance and usability - visual appeal, ease of use, result expectations, model outputs, and user-interface. The questionnaire would help researchers and coders in identifying the outstanding needs to improve the tool for better. Moreover, IHSPM was alpha tested at each stage of development to ensure the correct functioning of each module.
3.5.2 Observations and Results

In all, nine beta testers including four transportation professionals, and five graduate students participated in the beta test. Their response to the following questions was obtained on a five-level Likert scale with the following levels: 1) Strongly Disagree, 2) Disagree, 3) Neutral, 4) Agree, 5) Strongly Agree. These questions were formulated in such a way that they capture beta tester’s overall experience while using the tool. The questions asked were as follows:

1. Was the overall system easy to use?
2. Was the overall system easy to navigate?
3. Did the overall system give the expected results?
4. Was the overall system visually appealing?
5. Was the data loading time fair?
6. Were you able to upload all relevant data to create CPMs?
7. Did the map view present useful information about road design?
8. Was the map view easy to interpret?
9. Did the compiled spreadsheet contain all the variables you wanted to input?
10. Was the model generation module easy to use?
11. Were you able to understand the values in the generated models?

The user’s responses to these questions are presented in figure 3.12. More than 75% of the test users agreed that the overall system was easy to use, easy to navigate, visually appealing, generated relevant results, and had a user-friendly interface. Based on the identified collision-prone zones from developed models, the city professionals agreed that the tool would ameliorate the current practice of network screening in identifying the blackspots. Moreover, the easy identification of hazardous neighbourhood would help the city save money and time and would permit the efficient use of road safety funding. Professionals also found the ‘Walkthrough’ module very helpful in selecting appropriate models for specific planning scope, however, they recommended adding more tips for first-time users.
### Figure 3.12 Responses to the Questionnaire

The user feedback obtained from the beta test for the future improvement of IHSPM and recommended actions to improve the outstanding needs identified in the software are discussed in Table 3.3.
Table 3.3 Beta Test User Feedback and Recommended Actions

<table>
<thead>
<tr>
<th>S. No.</th>
<th>User Feedback</th>
<th>Recommended Actions Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>It was observed that the zone variable (TAZ ID) in the CPM development module should be fixed if it does not influence the model development process. Also, the urban/rural button on the model development page was identified to be non-functional for the provided dataset because of unavailability of a dichotomous variable, identifying urban and rural zones.</td>
<td>The zone ID flag on model development page can be fixed by integrating it with the land use flag. The zone variable can be made draggable only when users are developing models for different land use type (i.e., urban and rural). The urban and rural land use variables, in this case, will act as a filter in selecting the appropriate zones to be used in the model development process.</td>
</tr>
<tr>
<td>2.</td>
<td>‘Compile’ button on the ‘Map Data Aggregation’ page was identified to give insufficient visual feedback after process completion.</td>
<td>The ‘Compile’ button on the ‘Map Data Aggregation’ page can be improved to provide better feedback on process completion. For example, once user compiles the data using the ‘Compile’ button, a pop-up message saying ‘Dataset compiled and ready to be saved’ should show up or the button should change colour on process completion.</td>
</tr>
<tr>
<td>3.</td>
<td>Final spreadsheet after data aggregation was found to contain only variables from TAZ file and not the crash or network data files. Also, it was identified that the final spreadsheet should contain aggregated crashes stratified by collision severity and type to develop models for various kinds of total collisions, such as total severe collisions, total non-severe collisions etc.</td>
<td>The data aggregation modules can be augmented with codes to incorporate quantifiable and aggregated collision variables into the final spreadsheet such as the proportion of pedestrian and bicyclist crashes in a zone, the proportion of severe crashes, and the proportion of property damage only crashes. Moreover, crash aggregation code can be augmented to include crash aggregation by severity and type.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>4.</td>
<td>‘Walkthrough’ process was identified to require a help section that provides users’ more information on selections guidelines for proper CPM development and safety application.</td>
<td>Each step of the ‘Walkthrough’ process can be augmented with a help button (on the top right corner) that gives the user more information about the selection checklist and how to choose them. The information to be used under the help section has been already discussed in section 3.3.3.</td>
</tr>
<tr>
<td>5.</td>
<td>The modelled/measured exposure variable flag on the model development page was found to be unclear to some users when it comes to inputting complementary variables (vehicular congestion). It was also observed that the system should explicitly highlight the input space for complementary exposure variable when selecting the modelled exposure variable.</td>
<td>For the user to better understand the use of ‘Complementary Variable’ flag on the model development page, a note can be added below the ‘Modelled/Measured’ variable flag saying, ‘Please input complementary exposure variable, when selecting modelled exposure variable, such as zonal vehicular congestion (VC).’</td>
</tr>
<tr>
<td>6.</td>
<td>The sub-table (at the bottom right of the model development page) was a little misaligned which could be attributed to some technical glitch. Moreover, some users found the sub-table redundant when reading modelling results.</td>
<td>The sub-table on the model development page can either be removed as it does not implicate anything on the front-end or can be provided with a note saying, ‘Sample data points used for the model development process.’</td>
</tr>
<tr>
<td>7.</td>
<td>IHSPM was identified not to perform outlier analysis and provide other goodness of fit measures including model $\chi^2$ and SD, and t-statistics for explanatory variables.</td>
<td>The model development code can be fixed to account for the outlier analysis using CD and other goodness of fit measures. Appendix A can be used for reference.</td>
</tr>
</tbody>
</table>
3.6 Summary

This chapter explained the working of IHSPM, a road safety planning (RSP) tool that automates the development of community-based macro-level CPMs (in its current release). The chapter explains the novelty of different modules in the tool and how to use them based on the available quality of a dataset and prior knowledge of CPMs. The beta test results revealed that the overall system is easy to use, easy to navigate, visually appealing, generated relevant results, and has a user-friendly interface. Moreover, IHSPM was identified as a reliable tool by the beta testers to refine the current practice of network screening in identifying the black spots that would permit the efficient use of road safety funding. Several technical issues were identified during the beta test and were discussed in this chapter.
CHAPTER 4: DATA EXTRACTION AND MODEL DEVELOPMENT

4.1 Overview

This chapter describes the data extraction process and the development of community-based macro-level CPMs using the NB and FB modelling techniques. Section 4.2 describes the data extraction process used in this study. Section 4.3 discusses the model form and methodology for the manual development of NB and FB models, and compares modelling results obtained using them with those developed using IHSPM. Finally, section 4.4 presents the summary and discussions.

4.2 Data Extraction and Description

A systematic and automated data extraction process reduces errors and improves data quality such that regression models more robustly reflect underlying causal mechanisms. This section describes the data extraction process and the resulting data used in this research for the development of community-based macro-level CPMs. Section 4.2.1 describes the geographic scope of the area under investigation and the data aggregation process, and section 4.2.2 introduces variable data sources and statistical summary.

4.2.1 Geographical Scope and Aggregation Units

The data for this research comes from the City of Kelowna, located in the Regional District of Central Okanagan in the province of British Columbia (BC), Canada. With a total area of 211.85 square kilometres and a population of 127,380 in 2016, Kelowna is the largest city in BC’s interior and Okanagan Valley (Statistics Canada, 2016). A majority of the neighbourhoods in the city are car-dependent, where 86% of the residents commute to work by cars, trucks and vans, 3.87% use public transit, 3.89% of the residents walk to work, 2.65% bike to work, and the remaining 2.83% use other methods (Statistics Canada, 2016). Moreover, Kelowna is anticipated to be home to 50,000 new residents by 2030, given its rapid urban growth and economic development (Official Community Plan Year 2030). Figure 4.1 shows the city’s geographic area and its urban and rural land use.
In this research, all data used for the development of community-based macro-level CPMs were aggregated to TAZ level areal units. The use of TAZs as the aggregation unit in the model development process ensures proper synchronization between strategic community planning as well as road safety planning objectives. Based on the 2014 dataset obtained from the City of Kelowna, there are 183 TAZs. For this study, these zones were classified into urban and rural zones based on the visual examination in Google Maps (Regional Municipality of Peel). Zones with advanced urban amenities including: educational and business opportunities, industries, transportation facilities, parks and residences were identified as urban zones. Zones with primarily agricultural and/or forested land use in the TAZ core, and otherwise limited amenities were classified as rural zones. This way 131 TAZs were identified as urban zones and 52 TAZs were identified as rural zones as shown in Figure 4.1.

ArcGIS 10.5.1 (Overlay analysis- Help | ArcGIS Desktop) was utilized to aggregate data at the TAZ level areal units. The georeferenced collision data, traffic exposure data and network data were aggregated directly at TAZ level using the overlay tool in ArcMap that matches the rows of the join features (collisions, lane kilometres, intersection layers) to the target feature (traffic zones layer) based on their relative spatial locations. It was assumed that collision location could have contained horizontal error along and across the road, large enough to overlap with one or more zones. Fortunately, ICBC had accounted for this error by geocoding the claim locations as either mid-block or intersections along the road centerlines. Therefore, the claim locations were considered reasonably accurate for direct aggregation.

Moreover, individual shapefiles were extracted for different road classes including arterial, collector, and local roads all from the road centreline shapefiles using the road class variable. This was done to calculate the TAZ level arterial lane kilometres, collector lane kilometres and local lane kilometres. However, TAZ-level aggregated socio-demographic and economic information was extracted from the attribute table of the traffic zones shapefile obtained from the City of Kelowna for the year 2014. Also, the TDM variables were aggregated manually by visually examining each traffic zone as no package was identified in ArcMap to automate the process. All the data used in this study were aggregated at least twice and reviewed to ensure accurate data aggregation.
Traffic Analysis Zones Land Use

Legend
- Urban
- Rural

Figure 4.1 Study Area: City of Kelowna
4.2.2 Data Sources and Summary Statistics

Geographically referenced collision locations are necessary for identifying the number of collisions in each zone. Therefore, ICBC collision database for the years 2013-2015 was selected for the development of community-based macro-level CPMs. In addition to providing georeferenced collisions, the ICBC database contains police and self-reported collision claims, therefore it is considered complete. In this study, three-year TAZ level total collisions were regarded as the dependent variable in macro-level CPMs. As per the ICBC collision database, total collisions include fatality, injury and property-damage-only (PDO) collisions involving motor vehicles, which may or may not also include pedestrians, bikes and/or other vehicles.

Most of the number of independent variables used in the model development were identified by Lovegrove (2007), divided into four themes, including exposure (EXP), socio-demographic (SD), transportation demand management (TDM), and the road network (NET). The aggregated socio-demographics dataset was obtained from the City of Kelowna. The network variable shapefiles including road centreline, intersections, speed humps, signalized intersections were retrieved from the City of Kelowna open database. However, some network variables related to three-way intersections, roundabouts, and arterial-local intersections were derived from the road centreline shapefile using analysis toolbox in ArcGIS. Following steps were followed for the identification of three-way intersections:

1. Using the 'Road Centreline' shapefile, points were generated at the ends of each link using the ‘Generate Points along Line’ tool in ArcMap with ‘Point Placement’ field value 100%.

2. Conceptually, once the points are generated, three-way intersections are identified as those spots that have three points at a single location (X, Y). For that matter, X and Y coordinates for the generated points (in Step 1) were calculated using the ‘Calculate Geometry’ tool. Two separate fields were created for each X and Y coordinate in the attribute table using the ‘Add Field’ tool.

3. ‘Dissolve’ tool was then employed to aggregate generated points based on the calculated X and Y coordinates (in step 2) using the 'Count' field statistics.
4. Three-way intersections were identified as those coordinates (X and Y) that have 'Count' field value '3' in the attribute table (from step 3). Three-way intersections were then extracted using the ‘Selection by Attributes’ tool for 'Count' field value '3' in ArcMap. New shapefile was created for the selected points using the 'Create Layer for Selected Feature' tool and named ‘3-way intersections’.

In addition, modelled exposure variables – vehicle kilometre travelled (VKT), and vehicular congestion (VC) were derived from the Kelowna EMME/4 transportation planning (TP) model obtained from the city. On the premise that limited-access highways had no casual associations with zonal traffic patterns, all exposure data related to limited-access highways (i.e., speed > 60km/h) were excluded. Measured exposure variable, total lane kilometres (TLKM) was computed in ArcGIS using the road centreline shapefile. TDM variables including neighbourhood core size (Core), shortcut capacity (SCC) and shortcut attractiveness (SCVC) were computed manually following the methodology described by Lovegrove (2007). Moreover, as roundabouts reduce injury crashes by 70-90% (Elvik et al., 2009), a new variable, proportion of roundabouts (IRBP) (i.e., the number of roundabouts in a zone divided by the total number of intersections in a zone) was introduced.

Finally, twenty-seven variables were set as independent candidate variables. Generally, any variable having an association with collisions may be set as independent candidate variable; however, statistical techniques were employed to check for their statistical significance and retention in the final model. The variable data sources and statistical summary including their maximum, minimum and average values are shown in Table 4.1 – 4.4.
### Table 4.1 Dependent and Exposure Variables Definition and Data Summary (n=183)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Symbol</th>
<th>Source</th>
<th>Years</th>
<th>Zonal Min</th>
<th>Zonal Max</th>
<th>Zonal Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Collisions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total collisions</td>
<td>T3</td>
<td>ICBC</td>
<td>13-15</td>
<td>1</td>
<td>1338</td>
<td>107</td>
</tr>
<tr>
<td>Fatal collisions</td>
<td>F3</td>
<td>ICBC</td>
<td>13-15</td>
<td>0</td>
<td>2</td>
<td>0.065</td>
</tr>
<tr>
<td>Injury collisions</td>
<td>I3</td>
<td>ICBC</td>
<td>13-15</td>
<td>0</td>
<td>374</td>
<td>25</td>
</tr>
<tr>
<td>PDO</td>
<td>PDO</td>
<td>ICBC</td>
<td>13-15</td>
<td>0</td>
<td>964</td>
<td>81</td>
</tr>
<tr>
<td><strong>Exposure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle kilometer</td>
<td>VKT</td>
<td>TP^2</td>
<td>2014</td>
<td>4.04</td>
<td>5445.26</td>
<td>1075.08</td>
</tr>
<tr>
<td>congestion</td>
<td>VC</td>
<td>TP</td>
<td>2014</td>
<td>0</td>
<td>1.25</td>
<td>0.26</td>
</tr>
<tr>
<td>Total lane km</td>
<td>TLKM</td>
<td>CoK^3</td>
<td>2014</td>
<td>1.16</td>
<td>59.68</td>
<td>12.93</td>
</tr>
<tr>
<td>Arterial lane km</td>
<td>ALKM</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>53.61</td>
<td>6.65</td>
</tr>
<tr>
<td>Collector lane km</td>
<td>CLKM</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>23.37</td>
<td>2.98</td>
</tr>
<tr>
<td>Local lane km</td>
<td>LLKM</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>27.88</td>
<td>3.26</td>
</tr>
<tr>
<td>Zonal area (hectares)</td>
<td>AREA</td>
<td>CoK</td>
<td>2014</td>
<td>4.36</td>
<td>1083.41</td>
<td>117.85</td>
</tr>
</tbody>
</table>

Note:
1. ICBC: Insurance Corporation of British Columbia
2. TP: Transportation Planning Model of Kelowna (typical)
3. CoK: City of Kelowna (typical)

### Table 4.2 Socio-demographic Variables Definitions and Data Summary (n=183)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Symbol</th>
<th>Source</th>
<th>Years</th>
<th>Zonal Min</th>
<th>Zonal Max</th>
<th>Zonal Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>POP</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>5045</td>
<td>663</td>
</tr>
<tr>
<td>Population density (=POP/Area)</td>
<td>POPD</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>61.29</td>
<td>14.92</td>
</tr>
<tr>
<td>Employed residents</td>
<td>EMP</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>4437</td>
<td>367</td>
</tr>
<tr>
<td>Employed density (=EMP/Area)</td>
<td>EMPD</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>323.14</td>
<td>16.92</td>
</tr>
<tr>
<td>Unemployed residents</td>
<td>UNEMP</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>2340</td>
<td>204.21</td>
</tr>
<tr>
<td>Unemployed rate (=UNEMP/(UNEMP + EMP)) (%)</td>
<td>UNEMPP</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>100</td>
<td>50.78</td>
</tr>
</tbody>
</table>
Chapter 4: Data Extraction and Model Development

Table 4.3 TDM Variables Definition and Data Summary (n=183)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Symbol</th>
<th>Source</th>
<th>Years</th>
<th>Zonal Min</th>
<th>Zonal Max</th>
<th>Zonal Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core area (Hectares)</td>
<td>CORE</td>
<td>CoK</td>
<td>2014</td>
<td>4.4</td>
<td>1083.4</td>
<td>113.12</td>
</tr>
<tr>
<td>CORE/AREA</td>
<td>CRP</td>
<td>CoK</td>
<td>2014</td>
<td>50</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>Shortcut capacity</td>
<td>SCC</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>48.62</td>
<td>5.1</td>
</tr>
<tr>
<td>Shortcut attractiveness</td>
<td>SCVC</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>31.55</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Table 4.4 Network Variables Definition and Data Summary (n=183)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Symbol</th>
<th>Source</th>
<th>Years</th>
<th>Zonal Min</th>
<th>Zonal Max</th>
<th>Zonal Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of intersections</td>
<td>INT</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>97</td>
<td>11.86</td>
</tr>
<tr>
<td>Intersection density (INT/AREA)</td>
<td>INTD</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>1.3</td>
<td>0.24</td>
</tr>
<tr>
<td>No. of intersections/TLKM</td>
<td>INTKD</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>10.98</td>
<td>0.99</td>
</tr>
<tr>
<td>No. of signalized intersections</td>
<td>SIG</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>5</td>
<td>0.63</td>
</tr>
<tr>
<td>Signalized intersection density</td>
<td>SIGD</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>0.59</td>
<td>0.02</td>
</tr>
<tr>
<td>No. of 3-way intersections/INT (%)</td>
<td>I3WP</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>100</td>
<td>70.23</td>
</tr>
<tr>
<td>No. of arterial-local intersections/INT (%)</td>
<td>IALP</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>100</td>
<td>18.9</td>
</tr>
<tr>
<td>No. of roundabouts/INT (%)</td>
<td>IRBP</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>33.33</td>
<td>0.97</td>
</tr>
<tr>
<td>ALKM/TLKM (%)</td>
<td>ALKP</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>100</td>
<td>50.47</td>
</tr>
<tr>
<td>CLKM/TLKM (%)</td>
<td>CLKP</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>89.85</td>
<td>22.93</td>
</tr>
<tr>
<td>LLKM/TLKM (%)</td>
<td>LLKP</td>
<td>CoK</td>
<td>2014</td>
<td>0</td>
<td>100</td>
<td>26.36</td>
</tr>
</tbody>
</table>

4.3 Model Development and Results

Based on the regression methods discussed in section 2.2, NB generalized linear regression (GLM) and full Bayesian (FB) community-based macro-level CPMs were developed. To test IHSPM model development module, CPMs developed using IHSPM are presented and compared with manually developed NB models, which have been verified previously as correct.
4.3.1 Negative Binomial Models

4.3.1.1 Model Development

NB regression is the most commonly used development method for community-based macro-level CPMs. NB models were developed following the methodology described by Lovegrove & Sayed (2006). The following model form was used for the development of GLM models:

\[ E(\Lambda) = a_0(Z)^{b_0}e^{\sum b_i X_i} \]

Where, \( E(\Lambda) \) is the predicted collision frequency per TAZ, \( Z \) is the exposure variable (TLKM or VKT), \( X_i \) are the explanatory variables and \( a_0, b_0, b_i \) are the model parameter estimates. The above model form accounts for the zero-risk logic and differentiates the effect of lead exposure variable \( Z \) and other explanatory variables \( X \) on the collision frequency. The models were developed in accordance with the principle of parsimony, which calls for the right balance between the model complexity (number of independent variables in the model) and its explanatory power (model accuracy) (Sawalha & Sayed, 2006). For that matter, explanatory variables in each model were added in the forward stepwise manner, which involves starting with no variables in the model, testing the addition of each variable using the model comparison criterion (i.e., scaled deviance), adding the variables from the same variable theme to improve the model the most, and repeating the process until there is no significant improvement. The primary variable in each model is the lead exposure variable (i.e., TLKM or VKT because of its dominating influence on collisions). As volume/capacity ratio (VC) compliments VKT in the modelled exposure CPMs (Lovegrove, 2007) they were used in pair. The retention of other independent variables in each model depends on the following four criteria discussed by Sawalha & Sayed (2006). First, the logic (+/−) of each parameter estimate should be intuitively associated with collisions. Second, the t-statistic of each parameter should be significant at 95% confidence (i.e., \( t >1.96 \)). Third, the explanatory variables in each model should have little or no correlation among each other. Lastly, the addition of each independent variable in the model should result in a significant drop in the SD at 95% confidence level (i.e., \( >3.84 \)).
Further model refinement was carried out to improve the model fit via outlier analysis using Cook’s Distance (CD) following the methodology defined by Sawalha & Sayed (2006). CD value reflects how well the model fits the \( i^{th} \) observation. High CD value indicates points that are likely outliers. In stepwise progression, data points with the highest CD value were removed from the data sample if their removal results in the significant drop in SD at 95% confidence (i.e., \( >3.84 \)). However, the number of removed outliers should not exceed 5% of the total sample.

Moreover, Scaled Deviance (SD), Pearson \( \chi^2 \) and \( \kappa \) were used to describe the overall goodness of fit of NB models at a 95% level of confidence. Penalized measures, Akaike’s Information Criterion (AIC) and Bayesian Information Criterion (BIC) were also employed for model comparison and selection. The formulations of these measures have been discussed in chapter 2.

Package ‘MASS’ in the statistical programming language R (Ripley et al., 2013) was used for the development of NB models. After running models, basic model outputs such as over-dispersion parameter \( \kappa \), the degree of freedom, scaled deviance, parameter estimates, and their t-value were derived. Package ‘Stargazer’ (Hlavac, 2018) in R was used to produce well-structured and easily readable model outputs. The model code used for the development of urban and rural NB models in R is presented in appendix A.

4.3.1.2. Stratified NB Model Results

Model development was conducted in sixteen model groupings stratified by two land uses (131 urban traffic zones and 52 rural traffic zones), two exposure variable sources (modelled and measured) and four variable themes (EXP, SD, TDM and NET) as shown in table 4.5.

In each of the sixteen model groupings, CPMs were developed for total collision type. Table 4.6 and Table 4.7 presents the developed CPMs for the urban and rural TAZs that includes variables parameter estimates, t-statistics, and model goodness of fit measures (SD, Pearson \( \chi^2 \) and \( \kappa \)). Due to the statistical insignificance of none of the TDM variable including core size, shortcut capacity and shortcut attractiveness, model 7 (urban modelled TDM), and model 15 (rural modelled TDM) were excluded. All other independent variables in different model groups were significant at 95% confidence based on the value of t-statistics. In a few models, the intercept term is not statistically significant at 95% confidence (\( t < 1.96 \)), suggesting the intercept term leading
the CPM may not influence the model. However, this does not pose any difficulty as the insignificance is marginal (similar results were noted by Lovegrove & Sayed, 2006).

### Table 4.5 Model Groupings

<table>
<thead>
<tr>
<th>Land Use</th>
<th>Data Derivation</th>
<th>Themes</th>
<th>Group#</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Urban</strong></td>
<td></td>
<td>Exposure</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Socio-Demographics</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TDM</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Network</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Measured</td>
<td>Exposure</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Socio-Demographics</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TDM</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Network</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Modelled</td>
<td>Exposure</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Socio-Demographics</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TDM</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Network</td>
<td>12</td>
</tr>
<tr>
<td><strong>Rural</strong></td>
<td></td>
<td>Exposure</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Measured</td>
<td>Socio-Demographics</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TDM</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Network</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Modelled</td>
<td>Exposure</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Socio-Demographics</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TDM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Network</td>
<td></td>
</tr>
</tbody>
</table>
### Table 4.6 Negative Binomial *URBAN* CPMs – Total Collisions

<table>
<thead>
<tr>
<th>Model Group #</th>
<th>Model form</th>
<th>$\kappa$</th>
<th>DOF</th>
<th>Pearson $\chi^2$</th>
<th>SD</th>
<th>$\chi^2$ (_{0.005, \text{dof}})</th>
<th>AIC</th>
<th>BIC</th>
<th>t-Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Urban, Measured, Exposure</td>
<td>1.389</td>
<td>128</td>
<td>121</td>
<td>103</td>
<td>155</td>
<td>1,557</td>
<td>1,565</td>
<td>Constant = 18.3, TLKM = 5.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total Collisions/3yr = 4.03TLKM(^{0.504})</td>
</tr>
<tr>
<td>2</td>
<td>Urban, Measured, Socio-demographics</td>
<td>1.764</td>
<td>128</td>
<td>89</td>
<td>79</td>
<td>155</td>
<td>1,532</td>
<td>1,546</td>
<td>Constant = 10.3, TLKM = 7.6, EMPD = 7.96, POPD = 2.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total Collisions/3yr = 2.91TLKM(^{0.749}e^{0.016EMPD+0.010POPD})</td>
</tr>
<tr>
<td>3</td>
<td>Urban, Measured, TDM</td>
<td>1.427</td>
<td>129</td>
<td>114</td>
<td>100</td>
<td>156</td>
<td>1,573</td>
<td>1,584</td>
<td>Constant = 15.6, TLKM = 6.13, Core = -4.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total Collisions/3yr = 3.79TLKM(^{0.798}e^{-0.694\text{core}})</td>
</tr>
<tr>
<td>4</td>
<td>Urban, Measured, Network</td>
<td>1.77</td>
<td>129</td>
<td>89</td>
<td>80</td>
<td>157</td>
<td>1,555</td>
<td>1,575</td>
<td>Constant = 3.72, TLKM = 4.48, SIGD = 2.54, IRBP = -2.23, LLKP = 1.98, ALKP = 3.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total Collisions/3yr = 3.72TLKM(^{0.417}e^{2.257SIGD+0.014ALKP-0.035IRBP-0.011LLKP})</td>
</tr>
<tr>
<td>5</td>
<td>Urban, Modelled, Exposure</td>
<td>1.784</td>
<td>127</td>
<td>92</td>
<td>80</td>
<td>154</td>
<td>1,556</td>
<td>1,568</td>
<td>Constant = 0.53, VKT = 7, VC = 2.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total Collisions/3yr = 1.0VKT(^{0.659}e^{0.997VC})</td>
</tr>
<tr>
<td>6</td>
<td>Urban, Modelled, Socio-demographics</td>
<td>1.928</td>
<td>125</td>
<td>84</td>
<td>73</td>
<td>152</td>
<td>1,543</td>
<td>1,544</td>
<td>Constant = 1.32, VKT = 7.33, VC = 2.169, UNEMPP = -4.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total Collisions/3yr = 1.0VKT(^{0.698}e^{0.076VC-0.008UNEMPP})</td>
</tr>
<tr>
<td>8</td>
<td>Urban, Modelled, Network</td>
<td>1.764</td>
<td>127</td>
<td>95</td>
<td>127</td>
<td>154</td>
<td>1,551</td>
<td>1,566</td>
<td>Constant = 1.384, VKT = 6.27, VC = 2.47, IRBP = -1.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total Collisions/3yr = 1.0VKT(^{0.591}e^{1.082VC-0.030IRBP})</td>
</tr>
</tbody>
</table>
Table 4.7 Negative Binomial *RURAL* CPMs – Total Collisions

<table>
<thead>
<tr>
<th>Model Group #</th>
<th>Model form</th>
<th>( \kappa )</th>
<th>DOF</th>
<th>Pearson ( \chi^2 )</th>
<th>SD</th>
<th>( \chi^2 ) ( 0.05, \text{dof} )</th>
<th>AIC</th>
<th>BIC</th>
<th>t-Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Rural, Measured, Exposure</td>
<td>1.046</td>
<td>48</td>
<td>74</td>
<td>54</td>
<td>65</td>
<td>527</td>
<td>533</td>
<td>Constant = 3.78 TLKM = 1.972</td>
</tr>
<tr>
<td></td>
<td>Total Collisions/3yr = ( 2.871TLKM^{0.478} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Rural, Measured, Socio-demographics</td>
<td>1.43</td>
<td>48</td>
<td>42</td>
<td>40</td>
<td>65</td>
<td>526</td>
<td>534</td>
<td>Constant = 3.35 TLKM = 2.61 EMPD = 4.51</td>
</tr>
<tr>
<td></td>
<td>Total Collisions/3yr = ( 2.21TLKM^{0.601}e^{0.185EMPD} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Rural, Measured, TDM</td>
<td>1.597</td>
<td>49</td>
<td>37</td>
<td>37</td>
<td>66</td>
<td>534</td>
<td>542</td>
<td>Constant = 0.47 TLKM = 6.15 Core = -7.026</td>
</tr>
<tr>
<td></td>
<td>Total Collisions/3yr = ( 1.0TLKM^{1.878}e^{-0.577\text{core}} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Rural Measured, Network</td>
<td>2.85</td>
<td>46</td>
<td>18</td>
<td>21</td>
<td>63</td>
<td>495</td>
<td>506</td>
<td>Constant = 2.38 TLKM = 3.97 SIGD = 4.99 INTD = 4.45 IALP = 3.13</td>
</tr>
<tr>
<td></td>
<td>Total Collisions/3yr = ( 1.19TLKM^{0.672}e^{5.0346SIGD + 8.583INTD + 0.017IALP} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Rural, Modelled, Exposure</td>
<td>2.758</td>
<td>48</td>
<td>23</td>
<td>22</td>
<td>65</td>
<td>491</td>
<td>499</td>
<td>Constant = 3.416 VKT = 3.842 VC = 6</td>
</tr>
<tr>
<td></td>
<td>Total Collisions/3yr = ( 1.47VKT^{0.292}e^{4.103VC} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Rural, Modelled, Socio-demographics</td>
<td>4.534</td>
<td>47</td>
<td>15</td>
<td>15</td>
<td>64</td>
<td>486</td>
<td>498</td>
<td>Constant = 2.62 VKT = 5.5 VC = 4.64 EMPD = 4.88 POPD = 3.76</td>
</tr>
<tr>
<td></td>
<td>Total Collisions/3yr = ( 1.0VKT^{0.348}e^{-2.613VC + 0.119EMPD + 0.017POPD} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Rural, Modelled, Network</td>
<td>4.311</td>
<td>46</td>
<td>16</td>
<td>15</td>
<td>63</td>
<td>459</td>
<td>469</td>
<td>Constant = 2.51 VKT = 5.35 VC = 4.38 INTD = 3.59</td>
</tr>
<tr>
<td></td>
<td>Total Collisions/3yr = ( 1.0VKT^{0.352}e^{2.587VC + 5.853INTD} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

53
4.3.1.2.1 Urban CPMs

Seven urban CPMs were developed. The goodness of fit tests for all the urban CPMs indicate that the models fit the data well at 95% confidence as the value of SD and Pearson\(\chi^2\) are less than \(\chi^2\) distribution value with (n-p-1) degrees of freedom at 95% confidence level. The slight difference in the AIC and BIC values for measured and modelled CPMs indicates no evidence to select one model group over the other (regarding the model performance). However, the goodness of fit for measured CPMs is slightly lower than that for modelled CPMs. The parameter estimates for the measured lead exposure variable (TLKM) are in range 0.5-0.8, which is in good agreement with the range (0.5-0.9) observed by Lovegrove (2006). The comparison of parameter estimates for EXP, SD, TDM and NET CPMs with those developed by Khondakar et al. (2010) indicates a stronger association with the total collision frequency due to more significant value of parameter estimates. Also, the goodness of fit measures (SD and Pearson\(\chi^2\)) are smaller than 2003 models (Khondakar et al., 2010), indicating better model fit, which can be attributed to the better quality of dataset used for the model development.

4.3.1.2.2 Rural CPMs

Seven rural CPMs were developed. The goodness of fit tests for all the rural CPMs (except rural measured) indicate that the models fit the data well at 95% confidence as the value of SD and Pearson\(\chi^2\) are less than \(\chi^2\) distribution value with (n-p-1) degrees of freedom at 95% confidence level. However, the goodness of fit for measured CPMs is slightly lower than that for modelled CPMs. Moreover, on average, the \(\kappa\) values for the rural models is slightly higher than urban models which might be due to the smaller rural sample size (i.e., 52 rural zones versus 131 urban zones). This may also be the reason why goodness of fit of rural models is better than that for urban models. Moreover, the comparison of parameter estimates for EXP, SD, TDM and NET CPMs with those developed by Khondakar et al. (2010) exhibits a stronger association with the collision frequency due to more significant value of parameter estimates. Also, the values for goodness of fit measures are smaller than 2003 models (Khondakar et al. 2010), indicating better fit.
4.3.1.2.3 Statistical Associations

The statistical associations between the total collisions and independent variables are consistent across all model groups. However, these models exhibited better performance than those developed by Khondakar et al. (2010) regarding the significance of parameter estimates, which can be attributed to the better quality dataset. Two variables including the proportion of unemployment (UNEMPP) and vehicular congestion (VC) exhibited an inverse association with the total collision frequency in a few model groups developed by Khondakar et al. (2010).

Based on the models obtained using 2014 dataset, the relationships revealed that increased collisions are associated with an increase in the following explanatory variables:

- Exposure-related: vehicle kilometres travelled (VKT), total lane kilometres (TLKM), and average zonal congestion (VC).
- SD-related: employment density (EMPD), and population density (POPD)
- Network-related: signal density (SIGD), intersection density (INTD), proportion of arterial-local intersections (IALP) (only rural models), and proportion of arterial lane kilometres (ALKP) (only urban models)

The association with vehicle kilometre travelled (VKT), total lane kilometres (TLKM) and vehicular congestion (VC) appeared to confirm the intuitive expectations of higher probability of collisions with increased travel, and road kilometres. The positive association with signal density (SIGD) is reasonable as signalized intersections have high-speed and high-volume traffic, and left-turn conflict, which presents high collision risk. Moreover, the increased crashes with an increase in the proportion of arterial-local intersections (IALP) in rural zones is also rational and can be related to one of the two issues,

a) usual nature of rural intersections, which are often screened by foliage, rolling hills, or other rural features,

b) unexpected conflicts at the intersections between the low-speed traffic on local roads and high-speed through traffic on rural arterials.
The increased collisions with increased intersection density (INTD) agreed with the results from the previous studies (Lovegrove & Sayed, 2006; Wei & Lovegrove, 2013). This association suggest that closer attention is needed when planning neo-traditional communities with grid street patterns that may encourage higher intersection densities. The association of increased collisions with the increased population density (POPD) and employment density (EMPD) seemed intuitive as well because greater population and employment densities lead to more commuters, more exposure and more collisions. Similar associations were observed by Lovegrove & Sayed (2006).

Several models revealed the inverse association between total collisions and neighbourhood core size (CORE), the proportion of roundabouts (IRBP), proportion of local lane kilometres (LLKP) and the unemployment rate (UNEMPP). The decreased collisions with the increase in local lane kilometres (LLKP) in urban zones is intuitive as local lane roads are low speed and low volume roads that indicate low collision risk. However, the association with increased rate of unemployment is difficult to explain but intuitively fewer residents with jobs on average results in fewer commuters, less exposure and fewer collisions. The decreased collisions with the increase in core neighbourhood size (CORE) confirmed the earlier findings of Lovegrove & Sayed (2006). Though decreased collisions at roundabouts have been seen previously in micro-level CPMs (i.e., looking at individual intersections) the decreased collisions with the increase in the proportion of roundabouts (IRBP) is observed for the first time in macro-level CPMs. This association is rational as roundabouts reduce vehicular speeds, and conflict points at intersections by eliminating left turns thereby reduces collision risks and severity.

4.3.1.3 IHSPM Model Results

To validate the IHSPM modelling code, models listed in Table 4.6 and 4.7 were recreated using the Development Module programmed into IHSPM (IHSPM models). For fairer comparison, the independent variables were set the same as those developed manually. IHSPM models were developed following the methodology discussed in chapter 3. Since the current version of IHSPM cannot perform outlier analysis, the models were developed with the original dataset. Table 4.8 contains the community-based macro-level CPMs for urban and rural TAZs developed using IHSPM software. Overall, the results show good agreement between IHSPM and manually developed models, suggesting that the IHSPM software works as intended. The slight differences in the parameter estimates among models developed manually and the ones developed
using IHSPM can be attributed to the presence of outliers in the dataset used for model development in IHSPM.

### Table 4.8 IHSPM Developed CPMs

<table>
<thead>
<tr>
<th>Model #</th>
<th>Equation</th>
<th>κ</th>
<th>Residual Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CPMs for Urban TAZs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>( T_3 = 3.771(TLM)^{0.488} )</td>
<td>1.16</td>
<td>147.73</td>
</tr>
<tr>
<td>2</td>
<td>( T_3 = 2.817(TLM)^{0.683} e^{0.008POP+0.013EMP} )</td>
<td>1.40</td>
<td>144.19</td>
</tr>
<tr>
<td>3</td>
<td>( T_3 = 3.606(TLM)^{0.598} e^{-0.143CORE} )</td>
<td>1.19</td>
<td>147.22</td>
</tr>
<tr>
<td>4</td>
<td>( T_3 = 3.460(TLM)^{0.377} e^{2.053SIGD−0.032IRBP−0.013LLKP+0.013ALKP} )</td>
<td>1.71</td>
<td>140.13</td>
</tr>
<tr>
<td>5</td>
<td>( T_3 = 0.69(VKT)^{0.608} e^{0.952VC} )</td>
<td>1.71</td>
<td>142.72</td>
</tr>
<tr>
<td>6</td>
<td>( T_3 = 0.585(VKT)^{0.719} e^{0.016VC−0.008UNEMP} )</td>
<td>1.90</td>
<td>141.66</td>
</tr>
<tr>
<td>8</td>
<td>( T_3 = 0.79(VKT)^{0.591} e^{1.082VC−0.03IRBP} )</td>
<td>1.76</td>
<td>140.191</td>
</tr>
<tr>
<td>9</td>
<td>( T_3 = 3.6936(TLM)^{0.265} )</td>
<td>0.86</td>
<td>60.45</td>
</tr>
<tr>
<td>10</td>
<td>( T_3 = 3.079(TLM)^{0.351} e^{0.173EMP} )</td>
<td>1.14</td>
<td>58.776</td>
</tr>
<tr>
<td>11</td>
<td>( T_3 = 0.344(TLM)^{1.878} e^{-0.577CORE} )</td>
<td>1.59</td>
<td>57.053</td>
</tr>
<tr>
<td>12</td>
<td>( T_3 = 1.493(TLM)^{0.538} e^{42.008SIGD+13.272INTD+0.014IALP} )</td>
<td>2.113</td>
<td>56.418</td>
</tr>
<tr>
<td>13</td>
<td>( T_3 = 1.231(VKT)^{0.346} e^{4.122VC} )</td>
<td>2.05</td>
<td>55.919</td>
</tr>
<tr>
<td>14</td>
<td>( T_3 = 0.955(VKT)^{0.348} e^{2.613VC+0.107POP+0.119EMP} )</td>
<td>4.53</td>
<td>57.093</td>
</tr>
<tr>
<td>16</td>
<td>( T_3 = 0.962(VKT)^{0.342} e^{2.913VC+8.458INTD} )</td>
<td>2.52</td>
<td>55.338</td>
</tr>
</tbody>
</table>

Also, the parameter estimates for all IHSPM model explanatory variables show the same logic (+/-) as that of NB models developed manually. This validates that the modelling code used in the back-end of IHSPM is correct. The detailed performance of IHSPM models and NB models (developed manually) are compared in section 4.3.3 in detail using three measures of misprediction, mean absolute deviation (MAD), mean squared prediction error (MSPE) and mean squared error (MSE).
Chapter 4: Data Extraction and Model Development

However, the IHSPM does not yet provide the following two goodness of fit measures, Pearson$\chi^2$, SD and the t-statistic value for the explanatory variable, which makes it difficult for the user to comment on the overall goodness of fit of the models and the statistical significance of explanatory variables. Therefore, the back-end R code script in IHSPM should be fixed to incorporate these model goodness of fit measures (Pearson$\chi^2$ and SD) and t-statistic for explanatory variables. This can be done by using the modelling code used in this study (refer Appendix A) to update the current back-end IHSPM code since language R has been used in both cases.

4.3.2 Full Bayesian (FB) Models

Though both Poisson-gamma (i.e., NB) and Poisson lognormal (PLN) models can account for extra-Poisson variations, PLN models were found to perform better than NB models (Kim et al. 2002). In this study, FB PLN models were developed following the methodology described by El-Basyouny and Sayed (2009). The model form used for the model development is shown below:

$$E(y) = e^{\ln(E(\lambda))+0.5\sigma^2} = E(\lambda)e^{0.5\sigma^2} = a_0(TLKM)^{a_1}e^{b_jX_j+0.5\sigma^2}$$ (4.2)

Where $\sigma$ is the parameter to account for extra variation in the PLN distribution.

The maximum likelihood estimation (MLE) technique in case of NB models finds a point estimate for parameters that maximize the likelihood which may sometimes over fit the data and does not generalize well. Bayesian estimation, on the other hand, comes up with a distribution of possible parameters using Bayes’ rule. As discussed in Section 2.2.2, the first step in the development of FB models requires determining the prior distributions of the model parameters. The second step is estimating the posterior distributions for the model parameters by combining prior distributions with the observed data. In this research, as recommended by El-Basyouny & Sayed (2009), the normal distribution of the form N(0,1000) was used as the prior distributions for all the regression parameters, and a gamma distribution of the form Gamma (0, 0.1) was used as the prior distribution for extra variation parameter. The estimation of the posterior distribution for all parameter estimates was based on the Markov Chain Monte Carlo (MCMC) method. Deviance Information Criteria (DIC) was used for model comparison and selection. For the unbiased comparison of the FB and NB models, the same variables as that of NB models were used in FB.
WinBUGS 14 (Lunn et al. 2000) was used for the development of PLN models. The models were developed with the following seven steps:

1. The model file and data file were created (as shown in Appendix B).
2. Using the check model button in Model Specification Tool in WinBUGS, logic check was done for the model code syntax.
3. Using the load data button and compile button in the Model Specification Tool, data file (created in step 1) was loaded and the model was compiled.
4. The initial values of the parameter estimates were generated using the gen inits button to initialize the model.
5. Parameter estimates were added to the Sample Monitor Tool to monitor their posterior summary statistics and plots.
6. Using the Inference tool, Posterior distributions for all parameter estimates were simulated for specified number of iterations and posterior summaries were recorded from the Sample Monitor Tool.
7. The DIC value for each model group was recorded using the Inference tool for model comparison and selection.

Table 4.9 – 4.12 shows the developed PLN models for urban and rural TAZs with the same variables and model groupings as the NB models developed in section 4.3.1.2. As PLN models require transference into the link function in WinBUGS, the constant estimates generated are the link function estimates. The constant estimates for PLN models are then obtained by taking the antilog of the constant estimate obtained from WinBUGS.

The variable significance in FB models is tested using the confidence intervals. If the parameter estimate for a variable has the same logic (i.e., -/+ from its 2.5% to 97.5% confidence intervals, the variable is considered significant at 95% confidence; otherwise, it is not. As shown in Table 4.9 – 4.12, some variables that are significant in NB models are not statistically significant.
in FB models at 95% confidence. However, parameter estimates for all variables except local lane kilometres (LLKP) and the unemployment rate (UNEMPP) show the same logic as that of NB models. To check if FB models perform better than NB models, the two modelling techniques are compared in section 4.3.3.

Table 4.9 Urban Measured Full Bayesian Models

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (Exposure)</th>
<th>Model 2 (SD)</th>
<th>Model 3 (TDM)</th>
<th>Model 4 (Network)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Intercept</td>
<td>(0.064, 0.27)¹</td>
<td>(0.021, 0.0329)</td>
<td>(0.0354, 0.0249)</td>
<td>(0.0178, 0.0276)</td>
</tr>
<tr>
<td></td>
<td>(0.0189, 0.200)²</td>
<td>(-0.0329, 0.0908)</td>
<td>(-0.0097, 0.0904)</td>
<td>(-0.0396, 0.0698)</td>
</tr>
<tr>
<td>Log TLKM</td>
<td>(0.195, 0.0036)</td>
<td>(0.147, 0.0047)</td>
<td>(0.225, 0.0029)</td>
<td>(0.048, 0.0061)</td>
</tr>
<tr>
<td></td>
<td>(0.1883, 0.2005)</td>
<td>(0.1394, 0.1565)</td>
<td>(0.218, 0.229)</td>
<td>(0.0369, 0.0575)</td>
</tr>
<tr>
<td>POPD</td>
<td>(0.1058, 0.0028)</td>
<td>(0.102, 0.1122)</td>
<td>(0.1058, 0.0028)</td>
<td>(0.102, 0.1122)</td>
</tr>
<tr>
<td>EMPD</td>
<td>(0.0245, 7.03E-4)</td>
<td>(0.0235, 0.02613)</td>
<td>(0.0245, 7.03E-4)</td>
<td>(0.0235, 0.02613)</td>
</tr>
<tr>
<td>Core</td>
<td>(-6.17E-4, 0.019)</td>
<td>(-6.17E-4, 0.019)</td>
<td>(-6.17E-4, 0.019)</td>
<td>(-6.17E-4, 0.019)</td>
</tr>
<tr>
<td>SIGD</td>
<td></td>
<td>(0.0016, 0.0319)</td>
<td>(0.0016, 0.0319)</td>
<td>(0.0016, 0.0319)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.0616, 0.0638)</td>
<td>(-0.0616, 0.0638)</td>
<td>(-0.0616, 0.0638)</td>
</tr>
<tr>
<td>IRBP</td>
<td></td>
<td>(-0.021, 0.0188)</td>
<td>(-0.021, 0.0188)</td>
<td>(-0.021, 0.0188)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(-0.0555, 0.0132)</td>
<td>(-0.0555, 0.0132)</td>
<td>(-0.0555, 0.0132)</td>
</tr>
<tr>
<td>LLKP</td>
<td></td>
<td>(0.0505, 0.00465)</td>
<td>(0.0505, 0.00465)</td>
<td>(0.0505, 0.00465)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0408, 0.0576)</td>
<td>(0.0408, 0.0576)</td>
<td>(0.0408, 0.0576)</td>
</tr>
<tr>
<td>ALKP</td>
<td></td>
<td>(0.0545, 0.0010)</td>
<td>(0.0545, 0.0010)</td>
<td>(0.0545, 0.0010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0527, 0.0563)</td>
<td>(0.0527, 0.0563)</td>
<td>(0.0527, 0.0563)</td>
</tr>
<tr>
<td>σ</td>
<td>(3.098, 0.1906)</td>
<td>(1.726, 0.1102)</td>
<td>(3.0265, 0.191)</td>
<td>(3.0265, 0.191)</td>
</tr>
<tr>
<td></td>
<td>(2.735, 3.516)</td>
<td>(1.533, 1.964)</td>
<td>(2.693, 3.437)</td>
<td>(2.693, 3.437)</td>
</tr>
<tr>
<td>DIC</td>
<td>1,123.23</td>
<td>1,122.55</td>
<td>1,125.02</td>
<td>1,120.60</td>
</tr>
</tbody>
</table>

Note: 1. (Mean, Standard deviation)
   2. (2.5% confidence interval, 97.5% confidence interval)
### Table 4.10 Urban Modelled Full Bayesian Models

**Dependent Variable:** Total Crash Frequency, $T_3$

<table>
<thead>
<tr>
<th></th>
<th>Model 5 (Exposure)</th>
<th>Model 6 (S-D)</th>
<th>Model 7 (TDM)</th>
<th>Model 8 (Network)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Intercept</td>
<td>$(0.098, 0.0299)^1$</td>
<td>$(0.084, 0.03058)$</td>
<td>$(0.0297, 0.146)$</td>
<td>$(0.1066, 0.0176)$</td>
</tr>
<tr>
<td></td>
<td>$(0.053, 0.1587)^2$</td>
<td>$(0.0297, 0.146)$</td>
<td></td>
<td>$(0.0655, 0.1385)$</td>
</tr>
<tr>
<td>Log VKT</td>
<td>$(0.5923, 0.0094)$</td>
<td>$(0.5656, 0.0249)$</td>
<td>$(0.5266, 0.6002)$</td>
<td>$(0.5734, 0.0003)$</td>
</tr>
<tr>
<td></td>
<td>$(0.5772, 0.6076)$</td>
<td>$(0.5656, 0.0249)$</td>
<td></td>
<td>$(0.5687, 0.5786)$</td>
</tr>
<tr>
<td>VC</td>
<td>$(0.0297, 0.0283)$</td>
<td>$(0.0269, 0.0292)$</td>
<td>$(0.0262, 0.0839)$</td>
<td>$(0.03, 0.02832)$</td>
</tr>
<tr>
<td></td>
<td>$(-0.0241, 0.085)$</td>
<td></td>
<td>Not Significant</td>
<td>$(-0.0139, 0.1015)$</td>
</tr>
<tr>
<td>UNEMPP</td>
<td>$(0.0272, 0.0319)$</td>
<td>$(0.0272, 0.0319)$</td>
<td>$(0.0357, 0.0908)$</td>
<td>$(0.03, 0.02832)$</td>
</tr>
<tr>
<td></td>
<td>$(-0.0357, 0.0908)$</td>
<td></td>
<td></td>
<td>$(-0.0139, 0.1015)$</td>
</tr>
<tr>
<td>IRBP</td>
<td>$(1.054, 0.079)$</td>
<td>$(1.182, 0.1371)$</td>
<td>$(0.9544, 1.454)$</td>
<td>$(1.14, 0.0759)$</td>
</tr>
<tr>
<td></td>
<td>$(0.909, 1.217)$</td>
<td>$(1.182, 0.1371)$</td>
<td>$(0.9544, 1.454)$</td>
<td>$(0.9988, 1.297)$</td>
</tr>
<tr>
<td>DIC</td>
<td>1,119.92</td>
<td>1,120.27</td>
<td>1,116.68</td>
<td></td>
</tr>
</tbody>
</table>

**Note:**
1. (Mean, Standard deviation)
2. (2.5% confidence interval, 97.5% confidence interval)
Table 4.11 Rural Measured Full Bayesian Models

<table>
<thead>
<tr>
<th></th>
<th>Model 9 (Exposure)</th>
<th>Model 10 (SD)</th>
<th>Model 11 (TDM)</th>
<th>Model 12 (Network)</th>
</tr>
</thead>
</table>
| Log Intercept    | (0.0119, 0.0287)
(-0.0356, 0.0789) | (0.0123, 0.0323)
(-0.0532, 0.0684) | (7.67E-3, 0.0248)
(-0.0426, 0.0533) | (0.0171, 0.0324)
(-0.0489, 0.0776) |
| Log TLKM         | (0.154, 0.00453)
(0.1454, 0.1629) | (0.143, 0.0031)
(0.1351, 0.1491) | (0.15, 0.0059)
(0.1387, 0.1597) | (0.111, 0.0094)
(0.0903, 0.1313) |
| EMPD             | (0.017, 0.0212)
(-0.0518, 0.0418) | (-0.0246, 0.0258)
(-0.0775, 0.0205) | (-0.0246, 0.0258)
(-0.0775, 0.0205) | (-0.0246, 0.0258)
(-0.0775, 0.0205) |
| Core             | (0.154, 0.00453)
(0.1454, 0.1629) | (0.143, 0.0031)
(0.1351, 0.1491) | (0.15, 0.0059)
(0.1387, 0.1597) | (0.111, 0.0094)
(0.0903, 0.1313) |
| SIGD             | (2.107, 0.2114)
(1.742, 2.562) | (2.107, 0.2095)
(1.74, 2.563) | (2.113, 0.2068)
(1.757, 2.572) | (2.113, 0.2068)
(1.757, 2.572) |
| INTD             | (2.107, 0.2114)
(1.742, 2.562) | (2.107, 0.2095)
(1.74, 2.563) | (2.113, 0.2068)
(1.757, 2.572) | (2.113, 0.2068)
(1.757, 2.572) |
| IALP             | (2.107, 0.2114)
(1.742, 2.562) | (2.107, 0.2095)
(1.74, 2.563) | (2.113, 0.2068)
(1.757, 2.572) | (2.113, 0.2068)
(1.757, 2.572) |
| SIGD             | (2.107, 0.2114)
(1.742, 2.562) | (2.107, 0.2095)
(1.74, 2.563) | (2.113, 0.2068)
(1.757, 2.572) | (2.113, 0.2068)
(1.757, 2.572) |
| INTD             | (2.107, 0.2114)
(1.742, 2.562) | (2.107, 0.2095)
(1.74, 2.563) | (2.113, 0.2068)
(1.757, 2.572) | (2.113, 0.2068)
(1.757, 2.572) |
| IALP             | (2.107, 0.2114)
(1.742, 2.562) | (2.107, 0.2095)
(1.74, 2.563) | (2.113, 0.2068)
(1.757, 2.572) | (2.113, 0.2068)
(1.757, 2.572) |
| σ                | (2.107, 0.2114)
(1.742, 2.562) | (2.107, 0.2095)
(1.74, 2.563) | (2.113, 0.2068)
(1.757, 2.572) | (2.113, 0.2068)
(1.757, 2.572) |

Note: 1. (Mean, Standard deviation)
2. (2.5% confidence interval, 97.5% confidence interval)
<table>
<thead>
<tr>
<th></th>
<th>Model 13 (Exposure)</th>
<th>Model 14 (SD)</th>
<th>Model 15 (TDM)</th>
<th>Model 16 (Network)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Intercept</td>
<td>(0.0183, 0.0317)(^1)</td>
<td>(0.0208, 0.032)</td>
<td>(-0.0388, 0.0787)</td>
<td>(0.0369, 0.0279)</td>
</tr>
<tr>
<td></td>
<td>(-0.0436, 0.0824)(^2)</td>
<td>(-0.0388, 0.0787)</td>
<td>(-0.0224, 0.0823)</td>
<td></td>
</tr>
<tr>
<td>Log VKT</td>
<td>(0.129, 0.0261)</td>
<td>(0.169, 0.0214)</td>
<td>(0.1149, 0.2037)</td>
<td>(0.1527, 0.0046)</td>
</tr>
<tr>
<td></td>
<td>(0.0792, 0.177)</td>
<td>(0.169, 0.0214)</td>
<td>(0.1426, 0.1603)</td>
<td></td>
</tr>
<tr>
<td>VC</td>
<td>(0.0029, 0.03016)</td>
<td>(0.00797, 0.03)</td>
<td>(-0.0503, 0.0649)</td>
<td>(0.0033, 0.0284)</td>
</tr>
<tr>
<td></td>
<td>(-0.0584, 0.0603)</td>
<td>(-0.0503, 0.0649)</td>
<td>(-0.0525, 0.0672)</td>
<td></td>
</tr>
<tr>
<td>POPD</td>
<td>(0.0627, 0.032)</td>
<td>(0.0095, 0.1394)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0095, 0.1394)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMPD</td>
<td>(0.0472, 0.0244)</td>
<td>(0.0191, 0.099)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0191, 0.099)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTD</td>
<td>(3.11, 0.343)</td>
<td>(2.649, 0.2877)</td>
<td></td>
<td>(0.0065, 0.0308)</td>
</tr>
<tr>
<td></td>
<td>(2.525, 3.86)</td>
<td>(2.148, 3.276)</td>
<td></td>
<td>(-0.0611, 0.0712)</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>(3.11, 0.343)</td>
<td>(2.649, 0.2877)</td>
<td></td>
<td>(2.939, 0.2875)</td>
</tr>
<tr>
<td></td>
<td>(2.525, 3.86)</td>
<td>(2.148, 3.276)</td>
<td></td>
<td>(2.46, 3.554)</td>
</tr>
<tr>
<td>DIC</td>
<td>394.77</td>
<td>393.94</td>
<td>392.93</td>
<td></td>
</tr>
</tbody>
</table>

**Note:**
1. (Mean, Standard deviation)
2. (2.5% confidence interval, 97.5% confidence interval)
4.3.3 Comparison of NB & FB, and NB & IHSPM models

Three measures including mean absolute deviation (MAD), mean squared prediction error (MSPE) and mean squared error (MSE) as recommended by Hadayeghi et al. (2006) were used to compare NB models with FB models and IHSPM models. The computational formulas for the three measures are presented in the equations below. The MAD measure provides an average magnitude of “misprediction” by a model where smaller values are preferred to larger values. Values of MAD close to zero suggests that the model, on average, predicts the observed data well. However, a comparison of MSPE and MSE provides potential over or underfitting of the models to the estimation data. For example, if the MSE value of a model is higher than the MSPE value of another, this indicates that the first model is overfitted to the estimation data and some observed relationships are not shown correctly in the model. The formulations of these three measures have been discussed in section 2.2.3 and are repeated here for convenience.

\[
MAD = \sum_{i=1}^{n} \frac{|Y_i - y_i|}{n} \tag{4.3}
\]

\[
MSPE = \sum_{i=1}^{n} \frac{(Y_i - y_i)^2}{n} \tag{4.4}
\]

\[
MSE = \sum_{i=1}^{n} \frac{(Y_i - y_i)^2}{n-p} \tag{4.5}
\]

Where \(Y_i\) denotes the predicted number of collisions, \(y_i\) denotes the observed number of collisions, \(n\) the sample data size, and \(p\) the number of model parameters. Table 4.13 shows the results of the value of three measures from the NB, FB and IHSPM models, stratified by land use (urban or rural), variable theme (EXP, SD, TDM, NET) and data derivation (measured or modelled).
### Table 4.13 MAD, MSPE and MSE for NB, FB and IHSPM models

<table>
<thead>
<tr>
<th>Measures</th>
<th>Regression Method</th>
<th>Exposure</th>
<th>SD</th>
<th>TDM</th>
<th>Network</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1 - Urban Measured Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>NB</td>
<td>115.36</td>
<td>104.27</td>
<td>105.4</td>
<td>111.45</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>136.99</td>
<td>427.68</td>
<td>119.66</td>
<td>145.17</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>116.52</td>
<td>107.92</td>
<td>115.29</td>
<td>114.11</td>
</tr>
<tr>
<td>MSPE</td>
<td>NB</td>
<td>40630.02</td>
<td>37300.07</td>
<td>37689.9</td>
<td>37849.12</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>30856.15</td>
<td>2048118.7</td>
<td>27585.2</td>
<td>39561.22</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>40953.76</td>
<td>37349</td>
<td>40530.6</td>
<td>39089.29</td>
</tr>
<tr>
<td>MSE</td>
<td>NB</td>
<td>40942.56</td>
<td>38174.29</td>
<td>38274.3</td>
<td>39351.07</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>31093.51</td>
<td>2096121.5</td>
<td>28012.9</td>
<td>41131.11</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>41268.79</td>
<td>37636.3</td>
<td>40842.3</td>
<td>39389.98</td>
</tr>
<tr>
<td><strong>Group 2 - Urban Modelled Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>NB</td>
<td>90.23</td>
<td>106.51</td>
<td>N/A</td>
<td>78.77</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>83.44</td>
<td>84.03</td>
<td>N/A</td>
<td>96.7</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>80.42</td>
<td>77.44</td>
<td>N/A</td>
<td>79.77</td>
</tr>
<tr>
<td>MSPE</td>
<td>NB</td>
<td>29537.05</td>
<td>35018.82</td>
<td>N/A</td>
<td>23384.65</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>19366.28</td>
<td>19673.85</td>
<td>N/A</td>
<td>31418.68</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>24515.49</td>
<td>21316.15</td>
<td>N/A</td>
<td>23762.76</td>
</tr>
<tr>
<td>MSE</td>
<td>NB</td>
<td>29994.99</td>
<td>35839.57</td>
<td>N/A</td>
<td>23932.73</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>19666.53</td>
<td>20134.95</td>
<td>N/A</td>
<td>32155.05</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>24895.58</td>
<td>21815.74</td>
<td>N/A</td>
<td>24319.7</td>
</tr>
<tr>
<td><strong>Group 3 - Rural Measured Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD</td>
<td>NB</td>
<td>44.88</td>
<td>37.71</td>
<td>51.03</td>
<td>32.41</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>43.01</td>
<td>43.11</td>
<td>43.44</td>
<td>44.85</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>47.34</td>
<td>42.89</td>
<td>36.57</td>
<td>33.64</td>
</tr>
<tr>
<td>MSPE</td>
<td>NB</td>
<td>7175.68</td>
<td>4812.19</td>
<td>7718.31</td>
<td>4118.36</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>6908.5</td>
<td>6880.27</td>
<td>6949.16</td>
<td>6565.66</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>7475.3</td>
<td>5889.66</td>
<td>5249.91</td>
<td>4475.86</td>
</tr>
<tr>
<td>MSE</td>
<td>NB</td>
<td>7316.38</td>
<td>5004.67</td>
<td>8027.04</td>
<td>4461.56</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>7043.96</td>
<td>7155.48</td>
<td>7227.13</td>
<td>7112.8</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>7621.87</td>
<td>6125.25</td>
<td>5459.91</td>
<td>4848.85</td>
</tr>
</tbody>
</table>
MAD values obtained from the FB and IHSPM models are compared with the NB models and mixed results are obtained.

a) For urban measured models, MAD values from the FB and IHSPM models are higher than the NB models indicating better performance of NB models over the other two. However, for urban modelled models, MAD values from the most of FB and IHSPM models are lower than the NB models indicating a better fit of FB and IHSPM models.

b) For rural measured models, MAD values from some FB and IHSPM models are lower than NB models and from some models higher than the NB models indicating mixed performances of the models. However, for rural modelled models, based on MAD values, NB models fit the observed data well over FB and IHSPM models.

The mixed results indicate that no firm conclusions can be achieved as of which models have better performance. In the future, hierarchical FB models accounting for spatial and temporal variations can be researched to improve currently developed FB models.

Moreover, mixed results regarding over or underfitting of the models to the estimation data are obtained when the MSE values from the FB and IHSPM models are compared with MSPE values from the NB models. The results indicate that the variability in the number of collisions is better captured by the FB models over NB models in case of urban TAZs and by NB models over

<table>
<thead>
<tr>
<th>Measures</th>
<th>Regression Method</th>
<th>Exposure</th>
<th>SD</th>
<th>TDM</th>
<th>Network</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MAD</strong></td>
<td>NB</td>
<td>28.17</td>
<td>25.71</td>
<td>N/A</td>
<td>29.94</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>240.51</td>
<td>80.76</td>
<td></td>
<td>152.38</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>28.86</td>
<td>25.72</td>
<td></td>
<td>25.9</td>
</tr>
<tr>
<td><strong>MSPE</strong></td>
<td>NB</td>
<td>3711.98</td>
<td>2339.31</td>
<td>N/A</td>
<td>3898.73</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>62614.45</td>
<td>7485.79</td>
<td></td>
<td>25527.28</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>3796.2</td>
<td>2339.46</td>
<td></td>
<td>3361.01</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td>NB</td>
<td>3860.46</td>
<td>2534.25</td>
<td>N/A</td>
<td>4137.43</td>
</tr>
<tr>
<td></td>
<td>FB</td>
<td>65119.03</td>
<td>8109.6</td>
<td></td>
<td>27090.17</td>
</tr>
<tr>
<td></td>
<td>IHSPM</td>
<td>3870.64</td>
<td>2433.04</td>
<td></td>
<td>3641.1</td>
</tr>
</tbody>
</table>
Chapter 4: Data Extraction and Model Development

FB models in case of rural TAZs. However, for most of the cases, NB models performed marginally better than the IHSPM models which can be attributed to the uncleaned dataset that was used for the development of models in IHSPM. The overall comparison of MSE and MSPE values for the NB and FB models indicates that no firm conclusion can be achieved regarding which models are the better fit.

4.4 Summary and Discussions

Twenty-eight community based macro-level safety planning models (including NB and FB models) were developed to examine the relationship between several types of transportation planning variables and total collision frequency. As it is believed that the model should be parsimonious (balance between the model complexity and its explanatory power) and sensitive enough to examine the impact of different policies affecting the safety of the neighborhood, separate models were developed for four variable themes. The parameter estimates for variables exhibited a direct association with total collision frequency including vehicle kilometres travelled, total lane kilometres, zonal vehicular congestion, population density, employment density, signalized intersection density, intersection density, the proportion of arterial lane kilometres, and proportion of arterial-local intersections. Parameter estimates for variables including local lane kilometres, the proportion of roundabouts, and core size exhibited an inverse association with total collision frequency. One new and significant result in updating the CPMs, due to the availability of new data, was the decreased total collisions associated with the increase in the proportion of roundabouts, a result seen previously only in micro-level collision prediction models (i.e., looking at individual intersections).

A relative comparison of the model performance of the NB, FB, and IHSPM models revealed mixed results. MAD values for different model groups revealed that FB performed better than NB models in some cases and NB models performed better than FB models in other cases. Similar results were seen when MAD values for IHSPM and NB models were compared. Moreover, comparison of MSE and MSPE values for different models revealed that variability in number of collisions was better captured by FB models in some cases and by NB models in other cases. Therefore, no firm conclusions can be derived regarding the performance of NB and FB models. However, variability in number of collisions was better captured by NB models over IHSPM models for most of the cases which can be attributed to the outliers in IHSPM dataset.
Overfitting in case of IHSPM models indicates the need for a module to conduct outlier analysis, which is a future work to refine models in IHSPM. Moreover, hierarchical FB models accounting for spatial and temporal variations can be researched in the future to improve the performance of currently developed FB models and to see if they perform better than NB models.
CHAPTER 5: MACRO REACTIVE SAFETY APPLICATIONS

5.1 Overview

This chapter presents macro-reactive applications of community-based macro-level collision prediction models. The NB, IHSPM, and FB models developed in chapter 4 were applied to identify, rank, diagnose, and remedy collision-prone zones (CPZs) in Kelowna. Additionally, the impact of two planning design changes (i.e., safety countermeasures) on the safety of a TAZ was examined. Section 5.2 presents the identification and ranking methodology and results of CPZs using empirical Bayes (EB) and full Bayesian (FB) techniques. Section 5.3 presents the diagnosis and remedy for two urban and two rural CPZ. Section 5.4 presents collision modification factors for two planning-level remedies. Section 5.5 provides a summary of this chapter.

5.2 Black Spot Case Study

The objective of black spot programs is to identify the hazardous locations that exhibit high collision risk, diagnose these locations to identify the problem and then to identify the countermeasures needed for a solution. These programs are vital to ensure the efficient use of road safety funding, i.e., spending resources to treat the locations with high collision risks. This case study presents the application of community-based macro-level NB, IHSPM and FB CPMs developed in chapter 4 to identify, rank, diagnose and remedy collision-prone zones (CPZs) in Kelowna, unlike traditional black spot programs which are limited to individual road intersections and road segments.

The following subsections present the Empirical Bayes (EB) and Full Bayesian (FB) techniques for the identification of the top twenty urban and rural CPZs in Kelowna, and diagnosis and remedy for two sample urban and two sample rural CPZs.
5.2.1 Identification and Ranking Using EB Method

Empirical Bayes (EB) method has been identified as a reliable statistical technique for CPZ identification as it reduces the regression-to-mean (RTM) selection bias. RTM errors occur when the observed collision frequency regresses to the long term mean value of collision frequency as time goes by.

As discussed in chapter 2, the EB method follows a three-step procedure to identify a location as collision-prone (Lovegrove & Sayed, 2007). The first step involves estimating the location-specific collision frequency mean of prior distribution $E(\Lambda)$ by entering location specific traits in the CPM equation. The second step involves combining the obtained prior estimate with observed collision history to obtain the posterior collision estimate or EB safety estimate. The EB collision mean and variance are defined as follows:

$$EB_i = E(\Lambda|Y = count) = \frac{\alpha}{\beta} = \frac{E(\Lambda_i)}{\kappa + E(\Lambda_i)}(\kappa + count) \quad (5.1)$$

$$Var(EB_i) = Var(\Lambda|Y = count) = \frac{\alpha}{\beta^2} = \left[\frac{E(\Lambda_i)}{\kappa + E(\Lambda_i)}\right]^2(\kappa + count) \quad (5.2)$$

The final step involves comparing the location’s safety estimate, EB, to its prior collision estimate at 95% confidence (i.e. $\delta=0.95$), then the zone is identified as a CPZ if the EB safety estimate exceeds the specified norm. Mathematically, a location can be identified as collision-prone if the following condition is met:

$$1 - \int_0^{E(\Lambda)} \frac{\kappa^{\kappa+count}}{\Gamma(\kappa + count)} \frac{1}{\lambda^{\kappa+count} \lambda^\kappa e^{-\frac{\kappa}{E(\Lambda)}\lambda - 1}} d\lambda \geq \delta \quad (5.3)$$

Once the locations are identified as CPZs, ranking becomes crucial to ensure the most severe zones are given the most attention. Two ranking criteria namely potential collision reduction (PCR= EB-\(E(\Lambda)\)) and collision risk ratio (CRR=EB/\(E(\Lambda)\)) are employed to rank identified CPZs.

Following the approach discussed above, NB and IHSPM models developed in chapter 4 (shown in table 4.6, 4.7 and 4.8) were utilized to identify the top twenty urban and rural CPZs in
Kelowna, which are presented in Table 5.1 and Table 5.2 respectively. The CPZs identified using NB and IHSPM models were found to be identical which can be attributed to their similar parameter estimates. This confirms that the models developed using IHSPM are accurate and relevant in macro-reactive applications. Figure 5.1 and Figure 5.2 show the geographic locations of the top-ranked urban and rural CPZs respectively obtained using the combined ranking across all model groups.
Table 5.1 Top 20 Identified *URBAN* CPZs Using EB Method

<table>
<thead>
<tr>
<th>Urban Measured Models – <em>TLKM</em> as lead exposure</th>
<th>Urban Modeled Models – <em>VKT</em> as lead exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>CPZ 1</td>
<td>CPZ 4</td>
</tr>
<tr>
<td>CPZ 2</td>
<td>CPZ 7</td>
</tr>
<tr>
<td>CPZ 3</td>
<td>CPZ 3</td>
</tr>
<tr>
<td>CPZ 4</td>
<td>CPZ 11</td>
</tr>
<tr>
<td>CPZ 5</td>
<td>CPZ 16</td>
</tr>
<tr>
<td>CPZ 6</td>
<td>CPZ 12</td>
</tr>
<tr>
<td>CPZ 7</td>
<td>CPZ 3</td>
</tr>
<tr>
<td>CPZ 8</td>
<td>CPZ 18</td>
</tr>
<tr>
<td>CPZ 9</td>
<td>CPZ 21</td>
</tr>
<tr>
<td>CPZ 10</td>
<td>CPZ 14</td>
</tr>
<tr>
<td>CPZ 11</td>
<td>CPZ 6</td>
</tr>
<tr>
<td>CPZ 12</td>
<td>CPZ 22</td>
</tr>
<tr>
<td>CPZ 13</td>
<td>CPZ 23</td>
</tr>
<tr>
<td>CPZ 14</td>
<td>CPZ 5</td>
</tr>
<tr>
<td>CPZ 15</td>
<td>CPZ 13</td>
</tr>
<tr>
<td>CPZ 16</td>
<td>CPZ 24</td>
</tr>
<tr>
<td>CPZ 17</td>
<td>CPZ 8</td>
</tr>
<tr>
<td>CPZ 18</td>
<td>CPZ 9</td>
</tr>
<tr>
<td>CPZ 19</td>
<td>CPZ 15</td>
</tr>
<tr>
<td>CPZ 20</td>
<td>CPZ 25</td>
</tr>
</tbody>
</table>

Note: 1. Rank = PCR + CRR
Where, \( PCR = EB - E(\lambda) \); \( CRR = \frac{EB}{E(\theta)} \)
### Table 5.2 Top 20 Identified *RURAL* CPZs Using EB Method

<table>
<thead>
<tr>
<th>Model 9</th>
<th>Model 10</th>
<th>Model 11</th>
<th>Model 12</th>
<th>Model 13</th>
<th>Model 14</th>
<th>Model 15</th>
<th>Model 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure</td>
<td>Rank</td>
<td>SD</td>
<td>Rank</td>
<td>TDM</td>
<td>Rank</td>
<td>Network</td>
<td>Rank</td>
</tr>
<tr>
<td>CPZ 40</td>
<td>1</td>
<td>CPZ 41</td>
<td>1</td>
<td>CPZ 41</td>
<td>1</td>
<td>CPZ 40</td>
<td>1</td>
</tr>
<tr>
<td>CPZ 41</td>
<td>2</td>
<td>CPZ 44</td>
<td>2</td>
<td>CPZ 40</td>
<td>2</td>
<td>CPZ 43</td>
<td>1</td>
</tr>
<tr>
<td>CPZ 42</td>
<td>3</td>
<td>CPZ 42</td>
<td>3</td>
<td>CPZ 45</td>
<td>3</td>
<td>CPZ 42</td>
<td>3</td>
</tr>
<tr>
<td>CPZ 43</td>
<td>4</td>
<td>CPZ 45</td>
<td>4</td>
<td>CPZ 44</td>
<td>4</td>
<td>CPZ 54</td>
<td>4</td>
</tr>
<tr>
<td>CPZ 44</td>
<td>5</td>
<td>CPZ 46</td>
<td>5</td>
<td>CPZ 46</td>
<td>4</td>
<td>CPZ 54</td>
<td>5</td>
</tr>
<tr>
<td>CPZ 45</td>
<td>6</td>
<td>CPZ 40</td>
<td>6</td>
<td>CPZ 45</td>
<td>6</td>
<td>CPZ 50</td>
<td>5</td>
</tr>
<tr>
<td>CPZ 46</td>
<td>7</td>
<td>CPZ 47</td>
<td>7</td>
<td>CPZ 54</td>
<td>7</td>
<td>CPZ 56</td>
<td>7</td>
</tr>
<tr>
<td>CPZ 47</td>
<td>8</td>
<td>CPZ 48</td>
<td>8</td>
<td>CPZ 48</td>
<td>8</td>
<td>CPZ 55</td>
<td>8</td>
</tr>
<tr>
<td>CPZ 48</td>
<td>9</td>
<td>CPZ 49</td>
<td>9</td>
<td>CPZ 55</td>
<td>9</td>
<td>CPZ 42</td>
<td>9</td>
</tr>
<tr>
<td>CPZ 49</td>
<td>10</td>
<td>CPZ 50</td>
<td>10</td>
<td>CPZ 49</td>
<td>10</td>
<td>CPZ 56</td>
<td>10</td>
</tr>
<tr>
<td>CPZ 50</td>
<td>11</td>
<td>CPZ 51</td>
<td>11</td>
<td>CPZ 47</td>
<td>11</td>
<td>CPZ 63</td>
<td>11</td>
</tr>
<tr>
<td>CPZ 51</td>
<td>12</td>
<td>CPZ 52</td>
<td>12</td>
<td>CPZ 44</td>
<td>12</td>
<td>CPZ 55</td>
<td>12</td>
</tr>
<tr>
<td>CPZ 52</td>
<td>13</td>
<td>CPZ 55</td>
<td>13</td>
<td>CPZ 56</td>
<td>13</td>
<td>CPZ 64</td>
<td>13</td>
</tr>
<tr>
<td>CPZ 53</td>
<td>14</td>
<td>CPZ 54</td>
<td>14</td>
<td>CPZ 60</td>
<td>14</td>
<td>CPZ 58</td>
<td>14</td>
</tr>
<tr>
<td>CPZ 54</td>
<td>15</td>
<td>CPZ 56</td>
<td>15</td>
<td>CPZ 51</td>
<td>15</td>
<td>CPZ 65</td>
<td>15</td>
</tr>
<tr>
<td>CPZ 55</td>
<td>15</td>
<td>CPZ 57</td>
<td>15</td>
<td>CPZ 58</td>
<td>15</td>
<td>CPZ 62</td>
<td>16</td>
</tr>
<tr>
<td>CPZ 56</td>
<td>17</td>
<td>CPZ 43</td>
<td>17</td>
<td>CPZ 61</td>
<td>17</td>
<td>CPZ 66</td>
<td>17</td>
</tr>
<tr>
<td>CPZ 57</td>
<td>18</td>
<td>CPZ 59</td>
<td>18</td>
<td>CPZ 62</td>
<td>18</td>
<td>CPZ 47</td>
<td>18</td>
</tr>
<tr>
<td>CPZ 58</td>
<td>19</td>
<td>CPZ 60</td>
<td>19</td>
<td>CPZ 52</td>
<td>19</td>
<td>CPZ 67</td>
<td>19</td>
</tr>
<tr>
<td>CPZ 59</td>
<td>19</td>
<td>CPZ 58</td>
<td>20</td>
<td>CPZ 57</td>
<td>19</td>
<td>CPZ 62</td>
<td>20</td>
</tr>
</tbody>
</table>

Note: 1. Rank = PCR + CRR
2. Where, \( PCR = EB - E(A) \); \( CRR = \frac{EB}{E(A)} \)
Chapter 5: Macro-Reactive Safety Applications

Figure 5.1 Top 10 Identified Urban CPZs Using EB Method

Figure 5.2 Top 10 Identified Rural CPZs Using EB Method
Several results related to identified urban and rural CPZs were observed. Fourteen different NB model groups identified different top-rank urban and rural CPZs. In all, 31 urban and 29 rural CPZs were identified among 20 CPZs across different model groups. Of the 31 urban CPZs and 29 rural CPZs, ten urban zones and nine rural zones were identified by all the urban and rural CPMs (developed in chapter 4) respectively. Additionally, most of the urban CPZs are identified along Highway 97, which can be attributed to high-speed traffic on arterials, the predominance of signalized intersections, high volume/capacity ratio and left-turn conflicts.

5.2.2 Identification and Ranking Using FB Method

Wei (2012) discussed the FB methodology for the identification and ranking of macro-level hot spots. The identification and ranking methodology is similar to the empirical Bayes method. In the case of EB, NB models are used to generate the prior distribution of collisions, which are then combined with observed collision data to estimate the posterior distribution of collisions. A zone is identified as collision-prone if the posterior collision mean exceeds the prior collision mean at 95% confidence. Similarly, the posterior distribution of collisions in the case of the FB method is directly obtained from the FB models. A zone is identified as collision-prone if its posterior mean of collisions exceeds its normal estimate at the 95% confidence level. The posterior collision estimate for each zone is obtained from the fitted distribution, while the normal collision estimate for each zone is obtained by averaging the possible values relative to all posterior parameter distributions. The macro-level CPZ identification using the FB method is mathematically presented as follows (Wei, 2012):

\[
\hat{\lambda}_{i}(5\%) > E(y_i) \tag{5.4}
\]

Where \(\hat{\lambda}_{i}(5\%)\) is the 5% posterior collision value for a zone obtained from the WinBUGS simulation results and \(E(y_i)\) is the normal collision estimate for a zone obtained from the following equation:

\[
E(y_i) = a_0(TLKM)^{a_1}e^{b_1}j^2 + 0.5\sigma^2 \tag{5.5}
\]

The parameter estimates \((a_0, a_1, b_1, \sigma^2)\) are the mean of the posterior distribution simulated from WinBUGS.
Similar to EB, potential collision reduction (PCR) and collision risk ratio (CRR) are employed as the potential measures to rank CPZs in the FB method. PCR and CRR are formulated as follows:

\[
PCR = \lambda_{i(\text{mean})} - E(y_i) \tag{5.6}
\]

\[
CRR = \lambda_{i(\text{mean})} - E(y_i) \tag{5.7}
\]

Where \(\lambda_{i(\text{mean})}\) is the posterior mean of collisions for a zone obtained from the simulation results in WinBUGS, and \(E(y_i)\) is the normal collisions estimate for a zone \(i\).

Following the approach discussed above, the top 20 urban and rural collision-prone zones (CPZs) based on the combined ranking (PCR and CRR) were identified and presented in Table 5.3 and Table 5.4 respectively. Figure 5.3 and Figure 5.4 show the geographic locations of the top ranked urban and rural CPZs in the city obtained using combined ranking across all model groups.
Chapter 5: Macro-Reactive Safety Applications

Table 5.3 Top 20 Identified **URBAN** CPZs Using FB Method

<table>
<thead>
<tr>
<th>Urban Measured Models</th>
<th>Urban Modeled Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposures</td>
<td>Rank</td>
</tr>
<tr>
<td>CPZ 1</td>
<td>1</td>
</tr>
<tr>
<td>CPZ 2</td>
<td>1</td>
</tr>
<tr>
<td>CPZ 3</td>
<td>3</td>
</tr>
<tr>
<td>CPZ 4</td>
<td>4</td>
</tr>
<tr>
<td>CPZ 6</td>
<td>4</td>
</tr>
<tr>
<td>CPZ 7</td>
<td>7</td>
</tr>
<tr>
<td>CPZ 28</td>
<td>7</td>
</tr>
<tr>
<td>CPZ 8</td>
<td>9</td>
</tr>
<tr>
<td>CPZ 9</td>
<td>10</td>
</tr>
<tr>
<td>CPZ 10</td>
<td>10</td>
</tr>
<tr>
<td>CPZ 11</td>
<td>12</td>
</tr>
<tr>
<td>CPZ 16</td>
<td>13</td>
</tr>
<tr>
<td>CPZ 15</td>
<td>14</td>
</tr>
<tr>
<td>CPZ 12</td>
<td>15</td>
</tr>
<tr>
<td>CPZ 13</td>
<td>16</td>
</tr>
<tr>
<td>CPZ 14</td>
<td>17</td>
</tr>
<tr>
<td>CPZ 32</td>
<td>18</td>
</tr>
<tr>
<td>CPZ 17</td>
<td>19</td>
</tr>
<tr>
<td>CPZ 18</td>
<td>20</td>
</tr>
</tbody>
</table>

Note: 1. Rank = PCR + CRR
Where, PCR = \( \lambda_i (\text{mean}) - E(y_i) \); CRR = \( \frac{\lambda_i (\text{mean})}{E(y_i)} \)
Table 5.4 Top 20 Identified *RURAL* CPZs Using FB Method

<table>
<thead>
<tr>
<th>Model 9</th>
<th>Model 10</th>
<th>Model 11</th>
<th>Model 12</th>
<th>Rural Measured Models</th>
<th>Exposure</th>
<th>Rank</th>
<th>SD</th>
<th>Rank</th>
<th>TDM</th>
<th>Rank</th>
<th>Network</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPZ 40</td>
<td>1</td>
<td>CPZ 40</td>
<td>1</td>
<td>CPZ 40</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 41</td>
<td>2</td>
<td>CPZ 41</td>
<td>1</td>
<td>CPZ 41</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 42</td>
<td>3</td>
<td>CPZ 42</td>
<td>3</td>
<td>CPZ 49</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 43</td>
<td>3</td>
<td>CPZ 43</td>
<td>4</td>
<td>CPZ 47</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 44</td>
<td>5</td>
<td>CPZ 44</td>
<td>5</td>
<td>CPZ 42</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 45</td>
<td>6</td>
<td>CPZ 45</td>
<td>6</td>
<td>CPZ 51</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 46</td>
<td>7</td>
<td>CPZ 46</td>
<td>7</td>
<td>CPZ 44</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 47</td>
<td>8</td>
<td>CPZ 47</td>
<td>8</td>
<td>CPZ 56</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 48</td>
<td>9</td>
<td>CPZ 48</td>
<td>9</td>
<td>CPZ 55</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 51</td>
<td>10</td>
<td>CPZ 51</td>
<td>10</td>
<td>CPZ 69</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 49</td>
<td>11</td>
<td>CPZ 50</td>
<td>10</td>
<td>CPZ 58</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 50</td>
<td>11</td>
<td>CPZ 51</td>
<td>12</td>
<td>CPZ 59</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 52</td>
<td>13</td>
<td>CPZ 52</td>
<td>13</td>
<td>CPZ 50</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 54</td>
<td>14</td>
<td>CPZ 69</td>
<td>13</td>
<td>CPZ 41</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 55</td>
<td>14</td>
<td>CPZ 55</td>
<td>15</td>
<td>CPZ 55</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 53</td>
<td>16</td>
<td>CPZ 53</td>
<td>16</td>
<td>CPZ 60</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 56</td>
<td>17</td>
<td>CPZ 56</td>
<td>17</td>
<td>CPZ 65</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 57</td>
<td>18</td>
<td>CPZ 57</td>
<td>18</td>
<td>CPZ 53</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 58</td>
<td>19</td>
<td>CPZ 57</td>
<td>19</td>
<td>CPZ 60</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 59</td>
<td>20</td>
<td>CPZ 60</td>
<td>20</td>
<td>CPZ 71</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 13</th>
<th>Model 14</th>
<th>Model 15</th>
<th>Model 16</th>
<th>Rural Modelled Models</th>
<th>Exposure</th>
<th>Rank</th>
<th>SD</th>
<th>Rank</th>
<th>TDM</th>
<th>Rank</th>
<th>Network</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPZ 40</td>
<td>1</td>
<td>CPZ 41</td>
<td>1</td>
<td>CPZ 40</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 41</td>
<td>2</td>
<td>CPZ 43</td>
<td>2</td>
<td>CPZ 40</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 43</td>
<td>3</td>
<td>CPZ 43</td>
<td>3</td>
<td>CPZ 43</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 42</td>
<td>2</td>
<td>CPZ 45</td>
<td>2</td>
<td>CPZ 45</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 41</td>
<td>4</td>
<td>CPZ 45</td>
<td>4</td>
<td>CPZ 45</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 43</td>
<td>5</td>
<td>CPZ 45</td>
<td>5</td>
<td>CPZ 45</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 40</td>
<td>6</td>
<td>CPZ 45</td>
<td>6</td>
<td>CPZ 45</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 46</td>
<td>7</td>
<td>CPZ 46</td>
<td>7</td>
<td>CPZ 46</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 45</td>
<td>8</td>
<td>CPZ 46</td>
<td>8</td>
<td>CPZ 46</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 45</td>
<td>9</td>
<td>CPZ 47</td>
<td>9</td>
<td>CPZ 47</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 47</td>
<td>10</td>
<td>CPZ 47</td>
<td>10</td>
<td>CPZ 47</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 48</td>
<td>11</td>
<td>CPZ 47</td>
<td>11</td>
<td>CPZ 47</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 48</td>
<td>12</td>
<td>CPZ 47</td>
<td>12</td>
<td>CPZ 47</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 47</td>
<td>13</td>
<td>CPZ 47</td>
<td>13</td>
<td>CPZ 47</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 47</td>
<td>14</td>
<td>CPZ 47</td>
<td>14</td>
<td>CPZ 47</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 47</td>
<td>15</td>
<td>CPZ 47</td>
<td>15</td>
<td>CPZ 47</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 47</td>
<td>16</td>
<td>CPZ 47</td>
<td>16</td>
<td>CPZ 47</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 47</td>
<td>17</td>
<td>CPZ 47</td>
<td>17</td>
<td>CPZ 47</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 47</td>
<td>18</td>
<td>CPZ 47</td>
<td>18</td>
<td>CPZ 47</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 47</td>
<td>19</td>
<td>CPZ 47</td>
<td>19</td>
<td>CPZ 47</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPZ 47</td>
<td>20</td>
<td>CPZ 47</td>
<td>20</td>
<td>CPZ 47</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: 1. Rank = PCR + CRR

Where, PCR = \( \lambda(\text{mean}) - E(y_i) \); CRR = \( \frac{\lambda(\text{mean})}{E(y_i)} \)
Figure 5.3 Top 10 Identified *Urban* CPZs Using FB Method

Figure 5.4 Top 10 Identified *Rural* CPZs Using FB Method
Several results related to identified urban and rural CPZs were observed. Fourteen different FB models identified different top-rank urban and rural CPZs. In all, 37 different urban and 27 rural CPZs were identified among the top twenty CPZs across different model groups. Of the 37 urban CPZs and 27 rural CPZs, seven urban zones and eleven rural zones were identified by all urban and rural FB CPMs (developed in chapter 4) respectively. Moreover, like EB results, most of the worst-ranked urban CPZs were observed along highway 97 as well.

Following identification and ranking of CPZs using EB and FB methods, a sample of four zones (two urban and two rural) identified as CPZs was carried forward for diagnosis and remediation (discussed in section 5.3).

5.2.3 Comparison of Identified CPZs Using EB and FB Method

The top-ranked urban and rural CPZs identified using EB and FB methods are compared in this section. Considering all top 20 urban and rural CPZs identified by different model groups, 28 urban CPZs and 25 rural CPZs were found similar in both NB and FB methods, exhibiting 90% and 86% similarity in model results respectively. This observation indicates that the EB and FB methods agree with each other in the identification of CPZs. Moreover, when the identified urban CPZs using the EB and FB methods were compared across different model groups, a similarity of 95% was observed among EXP CPMs, 77.5% among SD CPMs, 100% among TDM CPMs and 71% among NET CPMs. For rural CPZs, a similarity of 85% was observed among EXP CPMs, 88% among SD CPMs, 90% among TDM CPMs and 73% among NET CPMs. Despite low similarity among urban and rural NB and FB NET CPM results (around 70%), the identified CPZs results from the EB and FB methods are roughly consistent across all model groups and in general.

However, the similarity in application results obtained using EB and FB methods raises a question regarding the practicality of NB and FB modelling techniques, i.e., which one is more practical to use when it comes to development and application. In this research, it was observed that the development of NB models is a relatively simple process in comparison to FB models as various data analysis tools (such as SPSS and R) are available that offer generalized linear modelling (GLM) packages for simple and easy development of NB models with well structured and directly usable model outputs. On the other hand, FB modelling tools like WinBUGS require
advanced knowledge of probabilistic analysis. Given the software’s sophisticated user interface, FB model development becomes a tedious and time-consuming process requiring manual data handling and model development. However, the application of FB models is a comparatively easy process as posterior distribution of collisions is directly available from the FB models results, unlike NB models where posterior estimates are calculated using the EB method. Nevertheless, NB models seem more practical than FB models as their development process demands relatively less time and human effort.

5.3 Diagnosis and Remedies of the CPZs

After the identification and ranking of CPZs, zones are diagnosed for safety problem, and suitable zone-wide safety remedies are suggested. Lovegrove (2007) suggested over-represented collision patterns and trigger variables as the indicators for in-office diagnosis. Trigger variables are defined as the independent variables in models that trigger collision-prone identification, i.e. whose values are significantly different from the regional averages. In-office diagnosis is complemented with on-site analyses which is conducted to verify the findings and to supplement other undiscovered causes for safety problems in CPZs. Strategic zonal-safety remedy analyses follow this two-indicator problem diagnosis process. The potential safety remedies are categorized under four main themes (i.e., EXP, SD, TDM, and NET). Lovegrove (2007) recommended considering at least one remedy under each theme to improve zonal safety. For this study, two urban and two rural CPZs were chosen for safety problem identification and problem solutions and are discussed in detail in the following subsections. Both in-office and on-site analyses were conducted to identify the road safety problem in these TAZs.

5.3.1 Collision Prone Zone 8

CPZ 8 (as shown in figure 5.5) was identified as a CPZ by the NB and FB models in the EXP, SD, TDM and NET groups. Harvey Ave, Burtch Rd, Springfield Rd, and Spall Rd enclose this neighbourhood. Three clues for the road safety problems are identified by looking at the collision patterns, trigger variables, land use and roadmaps. First, the land use type in this zone is predominantly commercial development with large parking lots, surrounded by residential and commercial development in adjacent neighbourhoods. Large parking lots in the zones preclude the internal neighbourhood connectivity, and hidden driveways along arterials increase vehicular
conflict. Second, the CPZ ranking was triggered by variables including volume to capacity ratio (high), arterial-lane kilometres (high), the proportion of roundabouts (low), and employment density (high). Third, the collision pattern revealed over-represented arterial and collector collision locations. From these clues, it appeared that the safety problem in this neighbourhood might be due to the following two reasons: a) high traffic volume and high speeds along the arterial and collector roads, b) conflict between the traffic to/from parking lots and the high-speed through-traffic on the collectors and arterials.

Figure 5.5 Urban CPZ 8
The following potential safety remedies are suggested to solve these safety problems:

- **SD** – *Neighbourhood redesign (to improve the land use mix, walkability, and bike/pedestrians path continuity) with enough proportions of residential and commercial development to reduce zonal vehicular trip generations and destinations.*
  
  - Past research has shown that mixed-use areas with good pedestrian and bicycle facilities (CPSTF 2017) typically have 5-15% less vehicle travel (Guo et al., 2010; Ewing et al., 2011; Frank et al., 2011; Sperry et al., 2011).
  
  - Therefore, collision modification factors (CMFs) were formulated for the above countermeasure on an optimistic end considering a 15% reduction in VKT with improved land use mix (i.e., increased zonal population density (twice the previous value) and reduced employment density (40% less than previous value)).
  
  - Based on the obtained collision frequency ratios in the after \(E_1\) and before \(E_0\) period of countermeasure implementation (using the *urban measured* and *modelled SD CPMs*), it can be deduced that improving neighbourhood land use mix and pedestrian and bike facilities results in 11% (CMF: -0.11) to 40% (CMF: -0.40) reduction in total collisions. The collision frequency ratios can be presented as follows:

\[
\frac{E_1}{E_0} = \frac{2.91(TLMK)^{0.749}e^{0.016+0.6\cdot EMPD+0.10\cdot POPD}}{2.91(TLMK)^{0.749}e^{0.016\cdot EMPD+0.01\cdot POPD}} = 0.60 \quad (5.8)
\]

**Urban Measured SD CPM**

\[
\frac{E_1}{E_0} = \frac{(0.85 \cdot VKT)^{0.698}e^{0.076\cdot VC-0.008\cdot UNEMP\cdot PP}}{(VKT)^{0.698}e^{0.076\cdot VC-0.008\cdot UNEMP\cdot PP}} = 0.89 \quad (5.9)
\]

**Urban Modelled SD CPM**
Chapter 5: Macro-Reactive Safety Applications

- **TDM** – *Transit service improvement (e.g. dedicated bus lanes) on arterials to decrease driving mode split; parking fees during peak hours in parking lots to reduce parking volume and hence the conflict of the low-speed parking traffic with the high speed through traffic on collectors and arterials.*
  
  o Past research (Cervero et al., 2004; Todd, 2006; Frank et al., 2011; Gallivan et al., 2015; VTPI, 2008) has shown that improved quality of transit services and parking management strategies reduces automobile travel by 10-30%.
  
  o Therefore, CMF was formulated for the above countermeasure considering 15% reduction in VKT with improved transit services and parking fees implementation in the neighbourhood.
  
  o Based on the obtained collision frequency ratios in after ($E_1$) and before ($E_0$) periods as shown in equations below, it can be deduced that transit service improvement and parking fees implementation results in 20% (CMF: -0.20) reduction in total collisions. Since *urban modelled TDM CPM* was insignificant, the guided inference was made based on the *urban modelled EXP CPM*.

  **Urban Modelled EXP CPM**

  $\frac{E_1}{E_0} = \left( \frac{0.85 + VKT}{VKT} \right)^{0.659} e^{-0.15\cdot0.997\cdotVC} = 0.80$ \hspace{1cm} (5.9)

- **NET** – *Traffic calmed collectors near entry/exit of parking lots to remove the conflict of low-speed traffic to/from parking lots with the high-speed through-traffic on collectors; replacing signalized intersections (e.g. Springfield Rd @ Burtch Rd) with roundabouts to reduce speed and left turn conflicts.*
  
  o Parameter estimates for the proportion of roundabouts (IRBP) have shown an inverse association with the total collision frequency (parameter estimates of -0.035 obtained from *urban measured CPMs* and -0.03 from *urban modelled CPMs*).
  
  o CMFs were formulated considering the replacement of three intersections including Springfield Rd @ Burtch Rd, Burtch Rd @ Sutherland, and Harvey Ave @
Chapter 5: Macro-Reactive Safety Applications

Sutherland with roundabouts resulting in 25% increase in the proportion of roundabouts in the TAZ.

- Based on the obtained collision frequency ratios in the after \( E_1 \) and before \( E_0 \) periods (using the *urban measured* and *modelled NET* models) as shown below, it can be deduced that replacing three signalized intersections with roundabouts results in 53-58% (CMF: -0.53 to -0.58) reduction in total collisions in the TAZ.

**Urban Measured NET CPM**

\[
\frac{E_1}{E_0} = \exp^{\alpha_s(1 \text{RBPs} - 1 \text{RBPs})} = e^{-0.035 \times 25} = 0.42
\]

\[\text{(5.10)}\]

**Urban Modelled NET CPM**

\[
\frac{E_1}{E_0} = \exp^{\alpha_s(1 \text{RBPs} - 1 \text{RBPs})} = e^{-0.03 \times 25} = 0.47
\]

\[\text{(5.11)}\]

### 5.3.2 Collision Prone Zone 5

CPZ 5 (as shown in figure 5.6) was identified as a CPZ by the NB and FB models in the EXP, SD, TDM, and NET groups. K.L.O Rd, Lakeshore Rd, and Richter St enclose this neighbourhood. Three clues for the road safety problems were identified by looking at the collision patterns, trigger variables, land use, roadmaps, and through on-site analysis. First, the land use type in this zone is entirely commercial development with large parking lots. Second, the CPZ ranking was triggered by variables including volume to capacity ratio (high), signal density (high), the proportion of roundabouts (low), employment density (high), and arterial lane kilometres (high). Third, the review of collision locations revealed that all the collisions occurred on the perimeter arterials. From these clues, it appeared that the safety problem in this neighbourhood might be due to the following three reasons: a) high traffic volume and high speeds along the arterials, b) left turn conflict between the traffic to/from the parking lots and the high-speed through-traffic on the arterial roads, c) left turn conflicts at signalized intersections.
Figure 5.6 Urban CPZ 5
The following potential remedies are suggested to solve these safety problems:

- **EXP** – Access management measures (e.g. turn restrictions) on arterials to increase capacity and reduce congestion (volume to capacity ratio). Physically separated bike lanes to eliminate conflict with the vehicular traffic.
  
  - CMF was formulated considering 10% vehicular mode shift to bikes as a result of improved bike facilities.
  
  - Based on the collision frequency ratio in the after ($E_1$) and before ($E_0$) periods (using the urban modelled EXP), it can be deduced that automobile travel will be reduced as a result of physically separated bike lanes resulting in a 10% reduction in total collisions (CMF: -0.1).

  \[
  \frac{E_1}{E_0} = \left( \frac{0.90 \cdot VKT}{VKT} \right)^{0.659} e^{-0.1 \cdot 0.997 \cdot VC} = 0.90 \tag{5.12}
  \]

- **SD** – Improved land use mix to increase pedestrian access to nearby stores and to reduce vehicular trips generations and destinations.
  
  - CMFs were formulated considering a 15% reduction in VKT as a result of improved land use mix (i.e., increased zonal population density (twice the previous value) and reduced employment density (40% less than previous value)).
  
  - Based on the collision frequency ratio in the after ($E_1$) and before ($E_0$) periods, it can be deduced that improving land use mix results in an 11% (CMF: -0.11) to 44% (CMF: -0.44) reduction in total collisions in the neighbourhood.

  \[
  \frac{E_1}{E_0} = \frac{2.91(TLKM)^{0.749} e^{0.016 \cdot EMPD + 0.10 \cdot POPD}}{2.91(TLKM)^{0.749} e^{0.016 \cdot EMPD + 0.010 \cdot POPD}} = 0.56 \tag{5.13}
  \]
Chapter 5: Macro-Reactive Safety Applications

**Urban Modelled SD CPM**

\[
\frac{E_1}{E_0} = \frac{(0.85 \times VKT)^{0.698} e^{0.076V_c - 0.008UNEMPP}}{(VKT)^{0.698} e^{0.076V_c - 0.008UNEMPP}} = 0.89
\]  

- **TDM** – *Transit service improvement (e.g. improved frequency and more transit stops) on the arterials to reduce drive mode split; traffic calmed entry and exit to/from the parking lots.*
  - CMF was formulated considering 15% reduction in VKT as a result of improved transit service quality. Due to insignificant *urban modelled TDM CPM*, a guided inference regarding CMF was made using the *urban modelled EXP CPM*.
  - Based on the obtained collision frequency ratios in the after \((E_1)\) and before \((E_0)\) periods (as shown below), it can be deduced that improved transit services results in a 20% (CMF: -0.20) reduction in total collisions.

**Urban Modelled EXP CPM**

\[
\frac{E_1}{E_0} = \left(\frac{0.85 \times VKT}{VKT}\right)^{0.659} e^{-0.15 + 0.997V_c} = 0.80
\]  

- **NET** – *Replacing signalized intersections (e.g. K.L.O Rd @ Lakeshore Rd, and Lakeshore Rd @ Richter St) with roundabouts to remove left turn conflicts. Restricted left turns (i.e. right-in/right-out only) to/from the parking lots to reduce conflict between the traffic from the parking lot with the high-speed through-traffic on arterials.*
  - Replacing two signalized intersections with roundabouts results in 66.67% increase in the proportion of roundabouts.
  - Based on the obtained ratios of collision frequencies in the before and after periods, it can be deduced that replacing two-thirds of the intersections with roundabouts results in an 86-90% (CMFs: -0.86 to -0.90) reduction in total collisions.

88
Urban Measured NET CPM

\[ \frac{E_i}{E_0} = e^{a_2(\text{IRBP}_z - \text{IRBP}_d)} = e^{-0.035\times66.67} = 0.10 \]  \hspace{1cm} (5.16)

Urban Modelled NET CPM

\[ \frac{E_i}{E_0} = e^{a_2(\text{IRBP}_z - \text{IRBP}_d)} = e^{-0.03\times66.67} = 0.14 \]  \hspace{1cm} (5.17)

5.3.3 Collision Prone Zone 44

CPZ 44 (as shown in figure 5.7) was identified as a CPZ by the NB and FB models in the EXP, SD, TDM, and NET groups. K.L.O Rd and Benvoulin Rd enclose this zone. Three clues for the road safety problems are identified by looking at the collision patterns, trigger variables, land use, roadmaps, and through on-site analysis. First, the zone contains predominantly agricultural land with a small rural residential development. Second, the CPZ ranking was triggered by variables including volume to capacity ratio (high), signal density (high), intersection density (high), and arterial-local intersections (high). Third, review of collision locations revealed that the collisions occurred on either the zonal perimeter roads or the internal local roads in the small residential core. Taken together, these clues suggest that the safety problem in this neighbourhood might be due to the following two reasons: a) conflicts between the traffic to/from the local roads and the high speed through traffic on the arterial roads, b) shortcutting through-traffic in conflict with the low speed, local trips on the internal local roads.
The following potential safety remedies are suggested to solve these safety problems:

- **SD** – *Introducing a local ‘corner store’ within the zone to improve the land use mix that would reduce local/through conflicts, and traffic volume.*
  
  - CMF was formulated considering a 10% reduction in VKT and VC as a result of a new local store in the neighbourhood, with the population and employment density remaining constant.
  
  - Based on the collision frequencies in the before ($E_0$) and after ($E_1$) periods, it can be deduced that introducing a corner store in this neighbourhood would result in up to a 13% reduction in total collisions (CMF: -0.13).
Rural Modelled SD CPM:

\[ \frac{E_1}{E_0} = \left( \frac{0.90 \times VKT}{VKT} \right)^{0.348} e^{-0.1 \times 2.613 + V_C + 0.119 \times EMPD + 0.107 \times POPD} = 0.87 \quad (5.18) \]

- **TDM** – *Improving transit service quality and bicycle facilities (by introducing new bike lanes and widening existing bike lanes) to reduce drivers’ mode split*
  - CMF was formulated considering a 10% reduction in vehicle travel and vehicular congestion as a result of improved bicycle facilities and transit services. Guided inference regarding CMF was made using the *rural modelled EXP model*, as the rural *modelled TDM* was found insignificant during model development.
  - Based on the frequency ratios as shown below, it can be deduced that transit service improvement and new bike lanes would result in a 21% reduction in total collisions (CMF: -0.21).

Rural Modelled EXP CPM:

\[ \frac{E_1}{E_0} = \left( \frac{0.90 \times VKT}{VKT} \right)^{0.292} e^{-0.1 \times 4.103 + V_C} = 0.79 \quad (5.19) \]

- **Network** – *Converting existing full movement arterial-local intersections to a restricted right-in/right-out intersection with no left turns or crossings that maintains local access and highway capacity but discourages shortcutting. Also, traffic calming internal local roads with speed humps, closures/diverters and replacing arterial-arterial intersections (e.g. Benvoulin Rd @ Cooper Rd) with roundabouts to discourage high speed.*
  - During the on-site analysis, it was identified that the neighbourhood has already been traffic calmed with speed humps to discourage high speeds on local roads. This confirms that the road safety problems identified based on trigger variables are relevant and CPMs are a practical tool in the road safety planning process.
  - CMF was formulated for the decreased proportion of arterial-local intersections (IALP). Based on the collision frequencies in the before \((E_0)\) and after \((E_1)\) periods
of design change, it can be inferred that decreasing IALP by 10% results in 5% fewer collisions in the neighbourhood (CMF: -0.5). Since IALP is the only significant variable in rural measured NET CPM, it was used to calculate frequency ratios.

**Rural measured NET CPM:**

\[
\frac{E_1}{E_0} = e^{0.017(IALP_2-IALP_1)} = e^{-0.1*0.017*30} = 0.95
\]  

(5.20)

### 5.3.4 Collision Prone Zone 50

CPZ 50 (as shown in Figure 5.8) was identified as CPZ by the NB and FB models in the EXP, SD, TDM, and NET groups. This zone lies along Clifton Rd. Three clues for the road safety problems are identified by looking at the collision patterns, trigger variables, land use, and roadmaps. First, it contained predominantly agricultural land use with a small rural residential development amidst a cul-de-sac internal local road pattern. The cul-de-sac road patterns reduces the zonal connectivity. Second, the CPZ ranking was triggered by variables including volume to capacity ratio (high), signal density (high), intersection density (high), arterial-local intersections (high), and population density (high). Third, review of collision locations revealed that the collisions occurred either on the internal local roads or at the arterial-local junctions. These clues suggest that the safety problem in this neighbourhood might be due to following two reasons: a) conflict between the traffic to/from the local roads and the high speed through traffic on the arterial roads, b) conflict due to high-speed traffic on the internal local roads.
The following potential safety remedies are suggested to solve these safety problems:

- **SD** – *Improving the land use mix by introducing a corner store in the rural zone to meet the daily requirements of the residents to reduce the traffic volume.*
CMF was formulated considering a 10% reduction in VKT and VC as a result of a new local store in the neighbourhood, with the population and employment density remaining constant.

Based on the collision frequencies in the before ($E_0$) and after ($E_1$) periods, it was deduced that introducing a corner store in this neighbourhood would result in a 7% reduction in total collisions (CMF: -0.07).

**Rural Modelled SD CPM:**

$$\frac{E_1}{E_0} = \left( \frac{0.90 \cdot VKT}{VKT} \right)^{0.348} e^{-0.1 \cdot 2.613 \cdot VC + 0.119 \cdot EMPD + 0.107 \cdot POPD} = 0.93 \quad (5.21)$$

- **TDM** – *Introducing community shuttle transit to reduce vehicle volumes, drivers, VKT and vehicular congestion; widening bike lanes on arterial (Clifton Rd); enhanced traffic signs on local roads*

  CMF was formulated considering 10% reduction in vehicle travel as a result of improved bicycle facilities and transit services. A guided inference regarding the CMF was made using the *rural modelled EXP* model as the *rural modelled TDM* group was found insignificant during model development.

  Based on the frequency ratios in the after ($E_1$) and before ($E_0$) periods of countermeasure implementation, it can be deduced that transit service and bicycle facility improvements results in 8% reduction in total collisions (CMF: -0.08).

**Rural Modelled EXP CPM:**

$$\frac{E_1}{E_0} = \left( \frac{0.90 \cdot VKT}{VKT} \right)^{0.292} e^{-0.1 \cdot 4.103 \cdot VC} = 0.92 \quad (5.22)$$

- **Network** – *Increasing zonal connectivity (with good quality bike lanes and sidewalks on local roads) for pedestrians/bicycle routes to the corner store to reduce vehicular use; traffic calming internal local roads with speed humps; closing hidden driveways along arterial roads.*
Studies from the past have shown that speed humps reduce crashes by around 50% for a given amount of traffic (Lovegrove et al. 2006; Elvik et al., 2009; Welle et al., 2015).

CMFs were formulated by closing two hidden driveways along Clifton Rd (a 4.7% reduction in intersection density) resulting in a 7-9% reduction in total collisions (CMFs: -0.06 to -0.09).

### Rural Measured NET CPM

\[
\frac{E_1}{E_0} = e^{a_2(INTD_2-INTD_1)} = e^{-8.583*(0.234-0.223)} = 0.91
\]  

### Rural Modelled NET CPM

\[
\frac{E_1}{E_0} = e^{a_2(INTD_2-INTD_1)} = e^{-5.853*(0.234-0.223)} = 0.93
\]

### 5.4 Examining the Impact of Planning Design Changes

The objective of this section is to examine the impact of planning design changes (i.e., safety countermeasures) on the safety of TAZs. Two network design changes involving the increased neighbourhood core area (CORE) and the proportion of roundabouts (IRBP) were explored to examine the shift in the safety of a TAZ regarding the percentage reduction in the total collisions. Due to the unavailability of collision data in the ‘before’ and ‘after’ periods of the network design change, conventional before-after analysis proposed by Hauer (2006) (as discussed in chapter 2) for the development of collision modification factors (CMFs) was impractical. Therefore, guided inferences from the parameter estimates of the explanatory variables were made (like the ones discussed in section 5.3). Conceptually, it provides similar information to the conventional collision modification factors (CMFs) which are used to estimate the change in collision frequency because of network design or operational change. The following two design changes were examined.
5.4.1 Increased Neighbourhood Core Area (CORE)

As the neighbourhood core area (CORE) has an inverse association with the total collision frequency (as can be seen from the parameter estimates of -0.694 for urban zones and -0.577 for rural zones), the percentage change in total collision due to an increase in neighbourhood CORE by 100% (double the previous CORE) was examined. CORE, as defined by Minnen (1999) and Lovegrove (2007), is the largest portion of a traffic zone area not bisected by major roads and zonal boundaries.

![Figure 5.9 Increased Neighbourhood Core Size](chart)

Figure 5.9 Increased Neighbourhood Core Size

Therefore, to access the increased safety of a TAZ with the increased CORE, a neighbourhood uniformly bisected by two major roads (as shown in Figure 5.9 (a)) was considered. The neighbourhood core area was increased from 1.13 km\(^2\) in Figure 5.9(a) (average zonal value obtained from 2014 Kelowna data) to 2.26 km\(^2\) in Figure 5.9(b) by removing the major road passing horizontally through the traffic zone. Keeping other variables in the urban and rural TDM models constant, the ratio of collision frequencies in the 'after' (\(E_1\)) and 'before' (\(E_0\)) periods were calculated as follows:

\[
\frac{E_1}{E_0} = \frac{a_0(TLKM)^{a_1}e^{2a_2\text{CORE}}}{a_0(TLKM)^{a_1}e^{a_2\text{CORE}}} = e^{2a_2\text{CORE}-a_2\text{CORE}} = e^{a_2\text{CORE}} \tag{5.25}
\]

Inputting the values of \(a_2\) equal to -0.694 for the urban TAZ and -0.577 for the rural TAZ (obtained from their respective developed measured CPMs), and CORE, 1.13 km\(^2\) in the equation above, the following \(E_1/E_0\) ratios were obtained:
Chapter 5: Macro-Reactive Safety Applications

Using *urban TDM CPM*: \[
\frac{E_1}{E_0} = e^{-0.694 \times 1.13} = 0.46
\] (5.26)

Using *rural TDM CPM*: \[
\frac{E_1}{E_0} = e^{-0.577 \times 1.13} = 0.52
\] (5.27)

Therefore, based on the obtained ratios, it can be deduced that increasing neighbourhood CORE by 100% results in a 54% ((1-0.46)*100) reduction in collisions in *urban* TAZs and a 48% ((1-0.52)*100) reduction in *rural* TAZs. Therefore, the CMF for increased neighbourhood core area in the urban TAZ is -0.54 and in the rural TAZ is -0.48.

### 5.4.2 Increased Proportion of Roundabouts (IRBP)

Second, as the proportion of roundabouts (IRBP) has an inverse association with the total collision frequency (parameter estimates of -0.035 obtained from *urban measured* CPMs and -0.03 from *urban modelled* CPMs), the percentage reduction in collisions for various percentage increase in the proportion of roundabouts for a TAZ was examined. For that matter, four identical neighbourhoods with 80 m grid pattern network design and twenty intersections were considered as shown in Figure 5.10. Two signalized intersections in neighbourhood (a), four signalized intersections in neighbourhood (b), four signalized and two non-signalized intersections in neighbourhood (c), and four non-signalized intersections in neighbourhood (d) were replaced with roundabouts resulting in a 10%, 20%, 30% and 20% increase in the proportion of roundabouts (IRBP) in the neighbourhoods (a), (b), (c), and (d), and 50%, 100%, 100% and 0% reduction in signal density (SIGD) in neighbourhood (a), (b), (c), (d) respectively.
Following the similar logic that was discussed in previous sections, the ratios of collision frequencies in the 'after' (E₁) and 'before' (E₀) periods of design change in the four neighbourhoods are presented as follows:

**Neighbourhood (a)**

Using *urban-measured* NET CPM:

\[
\frac{E_1}{E_0} = e^{a_1(SIGD_2 - SIGD_1) + a_2(IRBP_2 - IRBP_1)} = e^{-2.257 + 0.5 + 0.52 - 0.035 \times 10} = 0.40
\] (5.28)

Using *urban-modelled* NET CPM:

\[
\frac{E_1}{E_0} = e^{a_2(IRBP_2 - IRBP_1)} = e^{-0.03 \times 10} = 0.74
\] (5.29)
Neighbourhood (b)

Using *urban-measured* NET CPM:

\[
\frac{E_1}{E_0} = e^{a_2(SIGD_2 - SIGD_1) + a_2(IRBP_2 - IRBP_1)} = e^{-2.257 \times 0.52 - 0.035 \times 20} = 0.15
\] (5.30)

Using *urban-modelled* NET CPM:

\[
\frac{E_1}{E_0} = e^{a_2(IRBP_2 - IRBP_1)} = e^{-0.03 \times 20} = 0.55
\] (5.31)

Neighbourhood (c)

Using *urban-measured* NET CPM:

\[
\frac{E_1}{E_0} = e^{a_2(SIGD_2 - SIGD_1) + a_2(IRBP_2 - IRBP_1)} = e^{-2.257 \times 0.52 - 0.035 \times 30} = 0.10
\] (5.32)

Using *urban-modelled* NET CPM:

\[
\frac{E_1}{E_0} = e^{a_2(IRBP_2 - IRBP_1)} = e^{-0.03 \times 30} = 0.40
\] (5.33)

Neighbourhood (d)

Using *urban-measured* NET CPM:

\[
\frac{E_1}{E_0} = e^{a_2(SIGD_2 - SIGD_1) + a_2(IRBP_2 - IRBP_1)} = e^{-2.257 \times 0.52 - 0.035 \times 20} = 0.50
\] (5.34)

Using *urban-modelled* NET CPM:

\[
\frac{E_1}{E_0} = e^{a_2(IRBP_2 - IRBP_1)} = e^{-0.03 \times 20} = 0.55
\] (5.35)
Therefore, from the obtained ratios, it can be deduced that increasing proportion of roundabouts (IRBP) in a neighbourhood (through signalized intersection to roundabout conversion) by 10% results in 26-60% fewer collisions (CMFs: -0.26 to -0.60); increasing by 20% results in 45-85% fewer collisions (CMFs: -0.45 to -0.85); increasing by 30% results in 60-90% fewer collisions (CMFs: -0.6 to -0.90). Moreover, increasing IRBP by 20% (through non-signalized intersection to roundabout conversion) results in 45-50% fewer collisions (CMFs: -0.45 to -0.50).

This indicates that roundabouts can be considered as a potential countermeasure in urban CPZs where collisions are triggered by high intersection density. Similar results were noted by various roundabout studies conducted in the past. Research conducted by the Insurance Institute of Highway Safety found a 39% reduction in total crashes following the conversion of 24 conventional intersections with roundabouts in the U.S (Persaud et al. 2000). Moreover, a before-after roundabouts study conducted in Maryland observed an 80% reduction in total collisions on single-lane roundabout conversions and 88% on two-lane roundabout conversions (Cunningham, 2007). Even though no CMF can be obtained for rural zones due to statistical insignificance of IRBP in rural CPMs, roundabouts can be implemented as a traffic calming measure in rural zones to remove the conflict of the low-speed local traffic with high-speed shortcutting traffic or high-speed through traffic on connected collectors or arterials.

5.5 Summary

Using the developed community-based macro-level CPMs for Kelowna, a macro-reactive application (black spot study) was conducted for Kelowna to identify and rank CPZs in the city. The macro-reactive black spot studies facilitate early identification of neighbourhood road safety problems. In this study, the EB method with NB models and the FB method with FB models were employed to identify CPZs in Kelowna. The identification results demonstrated roughly consistent results using the two methods. The identified urban and rural CPZs using the EB and FB methods exhibited a similarity of 90% and 86% respectively. Post identification, two urban and two rural zones were diagnosed and remediated for road safety problems. The potential zonal remedies were discussed for each zone under four themes including EXP, SD, TDM, and NET. Moreover, CMFs were formulated for two planning design changes to estimate the change in collision frequency as
a result of increased neighbourhood core area and the proportion of roundabouts. It was identified that increasing proportion of roundabouts (IRBP) in a neighbourhood (through signalized intersection to roundabout conversion) by 10% results in 26-60% fewer collisions (CMFs: -0.26 to -0.60); increasing by 20% results in 45-85% fewer collisions (CMFs: -0.45 to -0.85); increasing by 30% results in 60-90% fewer collisions (CMFs: -0.6 to -0.90). Moreover, increasing IRBP by 20% (through non-signalized intersection to roundabout conversion) results in 45-50% fewer collisions (CMFs: -0.45 to -0.50).
6.1 Overview

This chapter is divided into three sections. Section 6.2 presents a summary of the research results and conclusions. Section 6.3 describes the contributions of this research. Section 6.4 provides recommendations for future research.

6.2 Summary & Conclusions

Macro-level CPMs have been identified as reliable empirical tools to conduct community-based macro-level road safety analyses (Lovegrove, 2010; Sun & Lovegrove, 2013; Feng Wei & Lovegrove, 2013). These models can be utilized to predict the safety performance of planned facilities, identify and rank collision-prone zones, and evaluate the effectiveness of community-level road safety countermeasures. In this research study, community-based macro-level CPMs were developed for the city of Kelowna to investigate the relationships between the total collision frequency and suit of planning level explanatory variables. The developed models were applied to identify collision-prone zones (CPZs) in the city and to examine the impact of different planning design changes on the safety of a TAZ. Additionally, Interactive High-Level Safety Planning Model (IHSPM), a web-based tool that automates the development and application of community-based macro-level CPMs was unveiled and assessed. The results and conclusions of this study are summarized below.

6.2.1 Updating Community-Based Macro-Level CPMs for Kelowna

The first objective of this research study was to update the community-based macro-level CPMs (total collisions frequency) for Kelowna. As the model should be parsimonious and sensitive enough to examine the impact of different policies affecting the safety of a neighbourhood, models were stratified by land use type (urban or rural), variable theme (EXP, SD, TDM and NET), and data derivation (measured or modelled). Three modelling techniques were employed to develop TAZ level CPMs including the generalized linear modelling (GLM) with NB error distribution, full Bayesian (FB) Poisson-lognormal models, and IHSPM model development module. The model development methodologies were derived from previous studies. The ICBC’s collision data for the year 2013-2015, land use, socio-demographic, economic, and network data for the year
2014 (obtained from the City of Kelowna) were utilized to develop CPMs.

Based on model results obtained, consistent statistical associations were observed between the total collision frequency and planning level explanatory variables. Direct associations were identified between total collision frequency and vehicle kilometres travelled (VKT), total lane kilometres (TLKM), zonal vehicular congestion (VC), population density (POPD), employment density (EMPD), signalized intersection density (SIGD), intersection density (INTD), the proportion of arterial lane kilometres (ALKM), and proportion of arterial-local intersections (IALP). Inverse associations were observed between total collision frequency and variables including local lane kilometres (LLKM), the proportion of roundabouts (IRBP), and core size (Core). One new and significant result that was observed in updating the CPMs, due to the availability of new data, was the decreased total collisions with the increase in the proportion of roundabouts, a result seen previously only in micro-level collision prediction models (i.e. looking at individual intersections).

A relative comparison of the model performances of the NB, FB, and IHSPM models revealed mixed results. Mean absolute deviation (MAD) values for different model groups revealed that FB performed better than NB models in some cases and NB models performed better than FB models in other cases. Moreover, comparison of mean squared error (MSE) and mean squared prediction error (MSPE) values for different models revealed that variability in the number of collisions was better captured by FB models in some cases and by NB models in other cases. Therefore, no firm conclusions can be derived regarding the performance of NB and FB models. However, NB models better-captured variability in the number of collisions over IHSPM models for most of the cases which can be attributed to the outliers in the IHSPM dataset. The results obtained from IHSPM were comparable to NB models, given the slight difference in parameter estimates and same logic (+/-) for estimates as that of manually developed NB models.

Based on these model development results, three main conclusions were drawn. First, the NB and FB model results demonstrated that it is possible to quantify statistical associations between community-level collision frequency and community traits. Second, given the same model form and variables in NB and FB models, it is hard to comment on the competitiveness of these models based on MAD, MSE, and MSPE. Third, models from IHSPM agreed with the NB
models, which confirms the accuracy in the back-end modelling code of the tool.

6.2.2 Assessment of Interactive High-Level Safety Planning Model (IHSPM) – Beta Test

The second objective of this research study was to examine how ‘value-added’ the IHSPM software developed so far by the Sustainable Transport Safety Lab (STS) UBCO is to the practitioners. IHSPM is a web-based tool that facilitates the automatic development and application of community-based macro-level CPMs to be used in transportation safety planning. A beta test was conducted with the transportation professionals from the City of Kelowna and research students from the Sustainable Transport Safety (STS) research laboratory, UBC Okanagan to assess overall system performance and usability. Beta testers were asked to perform an exercise to test the usability of various modules of the tool including Map Data Aggregation, Walkthrough, and CPM Development. Post-beta test, users’ feedback was obtained on a five-point Likert scale to get an idea of the overall system performance and usability. More than 75% of the beta testers agreed that the system was easy to use, easy to navigate, visually appealing, generated relevant results, and had a user-friendly interface. The city professionals also agreed that IHSPM could refine the current practice of network screening in identifying black spots and would permit the efficient use of road safety funding.

However, several technical issues were identified during the beta test and the process of development of models in this study. It was identified that the final IHSPM model results do not yet provide the t-statistics for the explanatory variables and model goodness of fit measures for developed models (Pearson\(^2\), SD). Therefore, the back-end modelling code in IHSPM can be fixed to incorporate the missing goodness of fit measures and t-statistics for explanatory variables for the users to better interpret the developed models. Moreover, some problems were identified with the data aggregation module during the aggregation of road lane kilometres for the city of Medellin in Columbia which can be attributed to the inconsistent metadata of the Medellin’s road centreline shapefile in comparison to what the model is trained for. Therefore, in the future releases, the backend data aggregation code can be made more dynamic to improve its compatibility with different metadata types to ensure accurate working of the module.
6.2.3 Macro-Reactive Road Safety Applications

The third objective of this research study was to apply the developed community-based macro-level CPMs to identify, rank, diagnose and remedy CPZs in Kelowna and to examine the impact of two planning design changes on the safety of a TAZ. Empirical Bayes (EB) technique for NB models and Full Bayesian (FB) technique for PLN models were employed to identify CPZs in the city. Two ranking criteria namely potential collision reduction (PCR) and collision risk ratio (CRR) were used to rank the identified CPZs. The comparison of the top twenty identified CPZs using EB and FB methods, exhibited a similarity of 90% among urban CPZs and 86% among rural CPZs. Overall, the identified CPZs using the EB and FB methods were roughly consistent across all model groups. Post identification and ranking of CPZs, two urban and two rural CPZs were diagnosed and remediated. A series of collision modification factors (CMFs) were developed for different exposure, socio-demographic, transportation demand management or network related countermeasures in the four CPZs. Also, CMFs were developed for two planning design changes including the increased proportion of roundabouts (IRBP) in an urban TAZ. It was observed that increasing the proportion of roundabouts (IRBP) in an urban TAZ (through signalized intersection to roundabout conversion) by 10% results in 26-60% fewer collisions (CMFs: -0.26 to -0.60); increasing by 20% results in 45-85% fewer collisions (CMFs: -0.45 to -0.85); increasing by 30% results in 60-90% fewer collisions (CMFs: -0.6 to -0.90). Moreover, increasing IRBP by 20% (through non-signalized intersection to roundabout conversion) results in 45-50% fewer collisions (CMFs: -0.45 to -0.50). This indicated that roundabouts could be considered as a potential countermeasure in urban CPZs where collisions are triggered by high intersection density.

6.3 Research Contributions

This research study makes the following significant contributions to the community-based macro-level road safety studies:

1) As road parameters (road lane kilometres, speed limit), network design (the type of intersections, block size, connectivity), regional socio-demographic, and size of geographical aggregation unit (traffic analysis zones) changes over time, updating community-based macro-level CPMs becomes necessary to inform relevant future road safety policies. This research updated the community-based macro-level CPMs for
Kelowna using negative binomial (NB) and full Bayesian (FB) modelling techniques. The results are in good agreement with the past research that developed the original Kelowna models (Khondakar et al. 2010).

2) The updating process may identify new variables due to the availability of new data that improve model quality. One new and significant result in updating Kelowna CPMs was the decreased total collision frequency with an increase in the proportion of roundabouts (IRBP), a result seen previously only in micro-level collision prediction models (i.e. looking at individual intersections).

3) Besides the traditional NB regression method, FB regression was also explored. The macro-reactive application results of the FB method (based on FB models) were in good agreement with the EB method (based on NB models), indicating that the FB models are as practical as NB models in black spot studies. However, it was also observed that the development of NB models is a relatively simple process in comparison to FB models as various data analysis tools offer simple and easy development of NB models with well-structured and directly usable model outputs. On the other hand, FB modelling tools have a sophisticated user interface that makes development a tedious and time-consuming process requiring manual data handling and model development.

4) As the development of macro-level CPMs is a cumbersome, semi-manual, and time-consuming process that requires a sound knowledge of spatial analysis tools (ArcGIS or QGIS) and data handling and analysis tools (SPSS, R, or Python), this research unveiled and assessed the Interactive High-Level Safety Planning Model (IHSPM) developed to date at STS lab. A beta test was conducted to obtain professional feedback on the overall usability of the tool, and this research presented possible recommendations/actions to resolve the outstanding needs identified.
6.4 Recommendations for Future Research

In addition to proposed research contributions, several topics have been recommended for future road safety research related to model development, model use, and IHSPM.

1) The data aggregation process for response and explanatory variables in this study used manual and automated (GIS) techniques. For variables including core size, and shortcut capacity, a manual extraction was done, as no automated techniques were available. Moreover, the derivation of modelled variables like VKT and VC was semi-manual involving the use of multiple platforms like VISUM and ArcGIS. Therefore, to maximize data quality and model goodness of fit, automated techniques could be employed in the future to extract all data involving those variables extracted manually in this study. The improved data quality further reveals the underlying causal mechanisms, which can then be applied in a more refined road safety applications.

2) Community-based macro-level CPMs developed in this research were only updated for the total collisions type. However, in the future, CPMs can be updated for the different collision severities including fatal, injury, PDO, and other collision types including vehicular-vehicular, vehicular-bicycle, and vehicle-pedestrian.

3) The FB models developed in this research did not account for the spatial and temporal influences and covariates interaction effects. Therefore, hierarchical FB models accounting for temporal and spatial variations as well as covariate intersection effects can be researched in the future and compared with ones developed manually in this research to comment on their competitiveness. Moreover, Poisson-gamma (NB) FB models can also be researched and compared with Poisson Lognormal (PLN) FB models developed here to comment on their competitiveness.

4) Guided inferences regarding the impact of safety countermeasures on the safety of a TAZ were made based upon the parameter estimates of the explanatory variables as the before-after analysis was impractical. However, in the future, CMFs can be developed by using the observed dataset in the before and after periods to compare them with the ones proposed in this study.
5) IHSPM's CPM development module can be extended with a package to conduct outlier analysis that would help enhance the goodness of fit inference. The code can be fixed to account for the missing goodness of fit measures including Pearson chi-sq., SD, and t-statistics for the explanatory variables. Furthermore, multiple data aggregation modules including shortcut capacity, core size, and intersection variables can be added to the Map Data Aggregation module to automate the task and supplement the use of IHSPM as a single independent tool.

6) The development and testing of macro-reactive CPM application modules in IHSPM that automates the identification and ranking of collision-prone zones (CPZs) in black spot studies and develops a series of macro-level collision modification factors are future goals.

7) Only macro-reactive road safety applications were explored as case studies in this research. Future research should focus on proactive regional and neighbourhood safety planning applications as well.
REFERENCES


References


Appendices

APPENDICES

Appendix A: R Script for Development of NB Models

For NB models stratified by land use (urban or rural), data derivation (measured or modelled) and variable theme (exposure, SD, TDM and network).

# Installing packages

if(!require(pastecs)) {install.packages("pastecs"); require(pastecs)}
if(!require(foreign)) {install.packages("foreign"); require(foreign)}
if(!require(data.table)) {install.packages("data.table"); require(data.table)}
if(!require(dplyr)) {install.packages("dplyr"); require(dplyr)}
if(!require(plyr)) {install.packages("plyr"); require(plyr)}
if(!require(tidyverse)) {install.packages("tidyverse"); require(tidyverse)}
if(!require(ggplot2)) {install.packages("ggplot2"); require(ggplot2)}
if(!require(car)) {install.packages("car"); require(car)}
if(!require(GGally)) {install.packages("GGally"); require(GGally)}
if(!require(scales)) {install.packages("scales"); require(scales)}
if(!require(corrplot)) {install.packages("corrplot"); require(corrplot)}
if(!require(moments)) {install.packages("moments"); require(moments)}
if(!require(stats)) {install.packages("stats"); require(stats)}
if(!require(sjPlot)) {install.packages("sjPlot"); require(sjPlot)}
if(!require(sjmisc)) {install.packages("sjmisc"); require(sjmisc)}
if(!require(stargazer)) {install.packages("stargazer"); require(stargazer)}
if(!require(Gmisc)) {install.packages("Gmisc"); require(Gmisc)}
if(!require(broom)) {install.packages("broom"); require(broom)}
if(!require(pscl)) {install.packages("pscl"); require(pscl)}
if(!require(MASS)) {install.packages("MASS"); require(MASS)}
if(!require(boot)) {install.packages("boot"); require(boot)}
if(!require(useBIC)) {install.packages("AICcmodavg"); require(useBIC)}

# Loading and calling data

my.data = read.csv("Thesis_Data_Modeling.csv", header = T, sep = ",")
my.data.urban = subset(my.data, Land_Use == 1)
my.data.rural = subset(my.data, Land_Use == 0)

# Checking extra variance in crash data

mean_collisions = mean(my.data$T3_Collision)
variance_collisions = var(my.data$T3_Collision)
mean_collisions > variance_collisions
# Data cleaning - Urban TAZs

- **Urban, Measured, Exposure**
  ```
  my.data.urban1 = subset(my.data.urban, Zone_Number != 1810 & Zone_Number!=3160 & Zone_Number!=1800)
  ```

- **Urban, Measured, SD**
  ```
  my.data.urban2 = subset(my.data.urban, Zone_Number != 1650 & Zone_Number!=1810 & Zone_Number!=1660)
  ```

- **Urban, Measured, TDM**
  ```
  my.data.urban3 = subset(my.data.urban, Zone_Number != 1810 & Zone_Number!=3740)
  ```

- **Urban, Measured, Network**
  ```
  my.data.urban4 = my.data.urban
  ```

- **Urban, Modelled, Exposure**
  ```
  my.data.urban6 = subset(my.data.urban, Zone_Number != 1030)
  ```

- **Urban, Modelled, SD**
  ```
  my.data.urban7 = subset(my.data.urban, Zone_Number != 1020 & Zone_Number!= 1650)
  ```

- **Urban, Modelled, Network**
  ```
  my.data.urban8 = my.data.urban
  ```

- **Urban, Modelled, TDM**
  ```
  my.data.urban9 = my.data.urban
  ```

# Data cleaning - Rural TAZs

- **Rural, Measured, Exposure**
  ```
  my.data.rural1 <- subset(my.data.rural, Zone_Number != 3500 & Zone_Number!= 1710)
  ```

- **Rural, Measured, SD**
  ```
  my.data.rural2 <- subset(my.data.rural, Zone_Number != 3500)
  ```

- **Rural, Measured, TDM**
  ```
  my.data.rural3 <- my.data.rural
  ```

- **Rural, Measured, Network**
  ```
  my.data.rural4 <- subset(my.data.rural, Zone_Number != 2020)
  ```

- **Rural, Modelled, Exposure**
  ```
  my.data.rural6 <- subset(my.data.rural, Zone_Number != 1710)
  ```

- **Rural, Modelled, SD**
  ```
  my.data.rural7 <- my.data.rural
  ```

- **Rural, Modelled, Network**
  ```
  my.data.rural8 <- subset(my.data.rural, Zone_Number != 2020 & Zone_Number!= 1710)
  ```

- **Rural, Modelled, TDM**
  ```
  my.data.rural9 <- subset(my.data.rural, Zone_Number != 1710)
  ```

# Urban Measured NB model development and results - TLKM as lead exposure

```R
n1 = glm.nb(T3_Collisions ~ TLKM_new, data = my.data.urban1) #Exposure
n2 = glm.nb(T3_Collisions ~ TLKM_new + EMPD + POPD, data = my.data.urban2) #SD
```
n3 = glm.nb(T3_Collisions ~ TLKM_new + Core, data = my.data.urban3) #TDM
n4 = glm.nb(T3_Collisions ~ TLKM_new + SIGD + IRBP + LLKP + ALKP, data = my.data.urban4) #Network

stargazer(n1, n2, n3, n4,
   ci = T,
   ci.level = 0.95,
   t.auto=FALSE, p.auto=FALSE,
   column.labels = c("Exposure (1) ","Socio-Demographic (2) ","TDM (3)","Network (4)"),
   model.numbers = FALSE,
   dep.var.caption = "Dependent Variable: Total Crash Frequency",
   dep.var.labels.include = FALSE,
   type='text',
   style = "all",
   single.row=TRUE,
   no.space = TRUE,
   intercept.bottom = TRUE)

BIC(n1)
BIC(n2)
BIC(n3)
BIC(n4)

#Urban Modelled NB model development and results - VKT as lead exposure

q1 = glm.nb(T3_Collisions ~ VKT_new + VC, data = my.data.urban6) ##Exposure
q2 = glm.nb(T3_Collisions ~ VKT_new + VC + UNEMPP, data = my.data.urban7) #SD
q3 = glm.nb(T3_Collisions ~ VKT_new + VC + Core, data = my.data.urban9) #TDM
q4 = glm.nb(T3_Collisions ~ VKT_new + VC + IRBP, data = my.data.urban8) ##Network

stargazer(q1, q2, q3, q4,
   ci = T,
   ci.level = 0.95,
   t.auto=FALSE, p.auto=FALSE,
   column.labels = c("Exposure (5) ","Socio-Demographic (6) ","TDM (7)" ,"Network (8)"),
   model.numbers = FALSE,
   dep.var.caption = "Dependent Variable: Total Crash Frequency",
   dep.var.labels.include = FALSE,
   type='text',
   style = "all",
   single.row=TRUE,
   no.space = TRUE,
   intercept.bottom = TRUE)

BIC(q1)
BIC(q2)
BIC(q3)
Appendices

#Rural Measured NB model development and results - TLKM as lead exposure

c1 = glm.nb(T3_Collisions ~ TLKM_new, data = my.data.rural1) #Exposure
c2 = glm.nb(T3_Collisions ~ TLKM_new + EMPD , data = my.data.rural2) #SD
c3 = glm.nb(T3_Collisions ~ TLKM_new + Core, data = my.data.rural3) #TDM
c4 = glm.nb(T3_Collisions ~ TLKM_new + SIGD + INTD + IALP, data = my.data.rural4) #NW

stargazer( c1, c2, c3, c4,
  ci = T,
  ci.level = 0.95,
  t.auto=FALSE, p.auto=FALSE,
  column.labels = c("Exposure (9) ", "Socio-Demographic (10)", "TDM (11)","Network (12)"),
  model.numbers = FALSE,
  dep.var.caption = "Dependent Variable: Total Crash Frequency",
  dep.var.labels.include = FALSE,
  type="text",
  style = "all",
  single.row=TRUE,
  no.space = TRUE,
  intercept.bottom = TRUE)

BIC(c1)
BIC(c2)
BIC(c3)
BIC(c4)

#Rural Modelled NB model development and results - VKT as lead exposure

d1 = glm.nb(T3_Collisions ~ VKT_new + VC , data = my.data.rural6) #Exposure
d2 = glm.nb(T3_Collisions ~ VKT_new + VC + POPD + EMPD , data = my.data.rural7) #SD
d3 = glm.nb(T3_Collisions ~ VKT_new + SCVC + Core , data = my.data.rural9) #TDM
d4 = glm.nb(T3_Collisions ~ VKT_new + VC+ INTD, data = my.data.rural8) #Network

stargazer(d1, d2,d3, d4,
  ci = T,
  ci.level = 0.95,
  t.auto=FALSE, p.auto=FALSE,
  column.labels = c("Exposure (13) ","Socio-Demographic (14)","TDM (15)", "Network (16)"),
  model.numbers = FALSE,
  dep.var.caption = "Dependent Variable: Total Crash Frequency",
  dep.var.labels.include = FALSE,
Appendices

type='text',
style = "all",
single.row=TRUE,
no.space = TRUE,
intercept.bottom = TRUE)

BIC(d1)
BIC(d2)
BIC(d3)
BIC(d4)
Appendices

Appendix B. WinBUGS Samples of Model File, and Data File Scripts

For FB total CPM in urban, measured, exposure group

**Model File Sample** % in *.txt format

```math
model {
    for( i in 1 : N) {
        t[i] ~ dpois(lambda2[i])
        mu[i] ~ dnorm(0.0, tau)
        log(lambda[i]) <- a0 + a1* TLKM[i] + mu[i]
        lambda2[i] <- max(lambda[i], 0.0001)
    }
    a0 ~ dnorm(0.0,1000)
    a1 ~ dnorm(0.0,1000)
    tau ~ dgamma(1, 0.1);
    sigma <- 1/sqrt(tau);
}
```

**Data File Sample** % in *.txt format

```math
list(
    TLKM = c(9.22, 4.92, 5.71, 4.49, 3.96, 1.17, 5.91, 4.6, 2.72, 1.43, 2.38, 3.45, 5.99, 7.06, 12.75, 10.12, 2.34, 2.1, 2.73, 2.89, 6.04, 6.32, 4.39, 11.69, 6.45, 10.3, 18.85, 16.78, 4.16, 3.64, 6.82, 2.83, 3.74, 5.91, 6.26, 19.57, 3.97, 5.13, 4.32, 8.51, 4.74, 5.02, 5.88, 1.35, 3.42, 6.83, 4.52, 7.23, 8.2, 7.17, 14.27, 8.38, 9.25, 10.78, 11.09, 2.07, 8.87, 6.78, 8.91, 9.64, 17.7, 11.57, 18.6, 8.71, 7.82, 12.46, 13.94, 11, 7.77, 16.18, 20.2, 14.4, 12.01, 23.73, 29.72, 8.34, 16.6, 10.84, 24.46, 41.04, 24.9, 3.01, 3.71, 4.3, 3.23, 2.5, 4.49, 5.74, 3.9, 4.07, 9, 6.02, 8.73, 3.64, 5.57, 4.69, 7.67, 5.2, 10.36, 10.65, 14.95, 3.28, 10.89, 3.64, 12.35, 20.91, 10.43, 14.18, 11.88, 16, 17.01, 31.19, 57.23, 21.96, 28.28, 6.39, 19.31, 19.97, 13.11, 19.48, 29.47, 10.56, 11.62, 6.58, 7.18, 12.33, 17.58, 14.78, 24.6, 53.13, 10.37) , N=131)
)
Appendix C. Components of Interactive High Level Safety Planning Model (IHSPM)

C.1 Homepage

The following are the components of the IHSPM homepage:

1. **Login and Register** – Register and Login buttons on the homepage as shown in Figure C.1 allow the user to create a new account or log in using an existing account. The user, once logged in, can continue past work or save present work.

2. **Drop Down Navigation** – Once logged in, the drop-down menu on the homepage allows fast and easy access to the main components of the IHSPM application including Data Aggregation, Walkthrough, My Data Sets, User History, Home Page, Profile Access, and a link to Additional Research as shown in Figure C.2, C.3, and C.4.

3. **Map Data Aggregation** - Within the drop-down and fixed upper navigation bars, the Map Data Aggregation (as shown in Figure C.2) link provides direct access to the ArcGIS integrated data upload map. The page allows the user to upload data files relevant to the study area for visual inspection and data aggregation. Section 3.3.2 discusses Map Data Aggregation module in detail.

4. **Walkthrough** - The Walkthrough (as shown in Figure C.2) link within the drop-down and fixed top navigation bar provides direct access to a six-step procedure that leads the user in the selection of models relevant to the scope of their analysis and is discussed in detail in section 3.3.3.
Appendices

5. **Table Data** - The *Table Data* (as shown in Figure C.2) link on the upper navigation bar allows quick access to the previously uploaded data for expedited and selectable model development.

![Figure C.2 Dropdown Navigation](image)

6. **Profile Access** - This tab within the drop-down and fixed upper navigation bar (as shown in Figure C.3) provides users access to their profile to edit their personal information (i.e., name, email address, or profile picture). This drop-down menu also allows users to log out of their account.

![Figure C.3 Profile Access and History Dropdown](image)

7. **User History** - The *History* link within the drop-down navigation bar (as shown in Figure C.3) provides quick access to previously uploaded data sets. Here tables can be uploaded or previously uploaded tables can be renamed, edited, or deleted as per the user’s requirements.
Appendices

8. **Research Tools and Analytical Expertise** - As shown in Figure C.4, this button redirects users to the Sustainable Transportation Safety (STS) Research Laboratory webpage providing more information about the lab’s research areas and leading software expertise.

9. **Lead Researcher** – As shown in Figure C.4, this tab redirects users to Dr. Gord Lovegrove’s (principal investigator, STS lab) personal website, providing users with more information about him and his research team, ongoing projects in his lab, and opportunities for future collaboration.

10. **Send Us an Email** – As shown in Figure C.4, this link provides users with technical support on tool’s *Data Aggregation, Walkthrough* and *CPM Development* module, and any other application related assistance.

![Figure C.4 Help Resources Dropdown](image)
C.2 Components of CPM Development Module

The following are the components of the CPM development module:

C.2.1 Variable Flags and Model Development

The variable requirements for the model development are labelled with colour-coded flags on the top of the page. Demographic flag, selectable by checkbox, pertains to the type of geographic area: urban or rural. The next flag, ‘Select Collision Variable,’ asks for user input on the response variable. The following flag, ‘Select Exposure Variable’ asks for user input on the type of data derivation—measured (TLKM) or modelled (VKT + VC). Finally, the last flag asks for user input on explanatory variables stratified by three variable classes—socio-demographic, network, and TDM variables. The variables can be applied to the above flags by dragging the categorical headings to the respective flag position. This process is repeated for each flag. However, the selections can be cleared using the ‘Clear Selection’ button that appears after the variable assignment.

C.2.2 Table Options

Along the right side of the screen in Figure 3.11, Table Options contains different buttons to perform specific operations. These buttons include: Upload Data button to upload data, Edit Columns button to edit the column names, and Develop Model button to develop generalized linear models with NB error structure. The developed models are displayed at the bottom of this page. The Display History tab enables generated tables to be accessible within the division titled Other Generated Tables for quick selection of tables and the Display Models tab displays or hides the developed models.

C.2.3 Table Display

The current selection of table data is displayed with each variable category labelled at the top of each column. Beneath the table display, there are options to save the data as either an Excel or CSV file when clicked, a search query that allows quick locating of specific column names, and a scroll bar to navigate the dataset.
C.2.4 Model Display

At the bottom of the page, model results including the CPM formula, over-dispersion parameter, degrees of freedom, and residual deviance are displayed. At the top of the model display table there are buttons to close the display, an indication of the model being viewed within, and a button that displays or hides the source numbers from the dataset used to develop the models.
C.3 Macro-Reactive Applications – Using Excel Spreadsheet

Given the set of developed models, they can be applied for the macro-reactive applications in the identification and ranking of collision prone zones (CPZs). IHSPM comes with an excel spreadsheet that automates the process of identification and ranking of CPZs. This section explains the procedure for using the excel spreadsheet in identifying and ranking the CPZs using Empirical Bayes (EB) method, discussed in section 2.3.1.

The identification of CPZs using EB method is a three-step procedure. The first two steps are based on estimating the expected mean collision frequency, $E(\Lambda)$ of a zone under consideration and using the zone collision history ($\text{count}$) along with measured $E(\Lambda)$ to provide an EB safety estimate for a zone. Step three is based on comparing the zonal level of safety estimate $EB_i$ with the regional average $E(\Lambda)$. When there is a significant probability (usually not less than 0.95) that the $EB$ safety estimate exceeds the specified norm, the zone is identified as a CPZ.

Following steps are followed to identify if a zone is collision prone:

1. Add TAZ numbers and the model specific explanatory variables columns to the specified location in the Figure C.5 and input the parameter estimates ($a_0$, $a_1$, $b_0$...) and kappa $k$ value under each of the specific cells.
## Appendices

### Figure C.5 Inputting Variables and Parameter Estimates

| TAZ # | CPZ? | P(EB=1|E|@95%) | Weight | Calculated | ICBC | From CPM | E(λ) | CRI | Rank CRI | Rank PCR | CRRI-EV/E | TAZ Rank | CPM Prediction to compute E(λ) |
|-------|------|------------|--------|-----------|------|----------|------|------|---------|---------|------------|---------|---------------------------------|
| 1     | no   | 0          | 3.5    | 1.03231  | 3.9047 | 0.03129  | 2    | 46.43154 | "VALUE!" | "VALUE!" | "VALUE!" | "VALUE!" | 1      | 5.356 1592 | 69 |

*Kappa, K = 1.5*
2. Add a column for collision history, *count* under ‘ICBC’ and calculate the collision frequency using the CPM for each traffic zone as specified in Figure C.6.

3. After performing the first two steps, the spreadsheet automatically performs the remaining steps that involves calculating the EB safety estimate and using the gamma function in excel to calculate the probability to check if the EB estimate exceeds the specified norm and whether a zone is CPZ or not. The zones identified as CPZs are marked as ‘Yes’ in column C of the spreadsheet and remaining zones are marked as ‘No’. The EB safety estimate in the spreadsheet is calculated using equation C.1, and probability using equation C.3.

\[ EB_i = \text{weight} \times E(\Lambda) + (1 - \text{weight}) \times \text{count} \]  \hspace{1cm} (C.1)

\[ \text{weight} = \frac{\kappa}{\kappa + E(\Lambda)} \]  \hspace{1cm} (C.2)

\[ P(EB > E) = 1 - \text{GAMMADIST}(E, \alpha, \frac{1}{\beta}, \text{true}) \]  \hspace{1cm} (C.3)

\[ \alpha = \kappa + \text{count} \]  \hspace{1cm} (C.4)

\[ \beta = \frac{\kappa}{E(\Lambda)} + 1 \]  \hspace{1cm} (C.5)

4. Identified CPZs are automatically ranked in the spreadsheet based on two ranking criteria namely potential collision reduction (PCR) and collision risk ratio (CRR) as shown in Figure C.7. The PCR and CRR are estimated in the spreadsheet using the equations C.6 and C.7 respectively.

\[ PCR = EB - E(\Lambda) \]  \hspace{1cm} (C.6)

\[ CRR = \frac{EB}{E(\Lambda)} \]  \hspace{1cm} (C.7)
### Collision Prone Zone (CPZ) Search in Kelowna

**October 2017**

Kappa, K

Macro-CPM Group 1: Urban Modelled Exposure

| TAZ # | CPZ? | \( \text{P}(\text{EB}|\text{EP}) \) @ 95% | Alpha | Beta | EB Weight | Calculated | CBC | From CPM | Step 1B: Rank them | TAZ Data for Model (CPM) prediction to compute \( E(\lambda) \) |
|-------|------|----------------------------------|-------|------|-----------|------------|-----|----------|---------------------|--------------------------------------------------|
| 10    | 359  | yes                              | 36.5  | 1.1786 | 356.48    | 0.01658    | 361 | 88.949359 | 267.59           | 3.000777 1 3.000777 1 4 3 5.356 106 0 |
| 11    | 308  | yes                              | 483.5 | 1.00814 | 479.597  | 0.00807    | 482 | 184.32305 | 295.274          | 2.60194 4 6 9 5.356 153 118 |
| 12    | 290  | yes                              | 526.5 | 1.06667 | 523.049  | 0.00655    | 525 | 227.36869 | 295.681          | 2.30045 6 7 12 5.356 1950 45 |
| 13    | 280  | yes                              | 227.5 | 1.02025 | 222.984  | 0.01398    | 226 | 74.06791 | 148.916          | 5.01054 3 8 2 5.356 318 0 |
| 14    | 141  | yes                              | 87.5  | 1.06334 | 82.2879  | 0.00957    | 88  | 23.681669 | 58.6062          | 3.47475 2 10 0 5.356 139 45 |
| 15    | 159  | yes                              | 197.5 | 1.01942 | 193.737  | 0.01905    | 196 | 77.22899 | 166.508          | 6.05861 5 11 2 5.356 115 0 |
| 16    | 364  | yes                              | 383.5 | 1.00755 | 380.628  | 0.00749    | 382 | 109.19978 | 181.849          | 1.91483 7 11 11 5.356 345 11 |
| 17    | 178  | yes                              | 0.99999939 | 159.0  | 1.01501 | 157.141  | 0.01479 | 158 | 93.99023 | 57.211           | 9.15721 8 17 3 5.356 701 195 |
| 18    | 202  | yes                              | 0.999999014 | 260.5 | 1.00792 | 258.452  | 0.00786 | 259 | 189.3049 | 69.1472          | 7.13627 12 19 8 5.356 2783 274 |
| 19    | 140  | yes                              | 0.99968507 | 94.5  | 1.0327 | 92.3188  | 0.02316 | 93  | 63.27866 | 29.0331          | 11.145881 9 20 1 5.356 871 296 |
| 20    | 105  | yes                              | 0.99869846 | 157.5 | 1.01241 | 155.577  | 0.02215 | 156 | 120.3059 | 34.664          | 10.2867 14 24 4 5.356 353 0 |
| 21    | 347  | yes                              | 0.99790186 | 93.5  | 1.02249 | 91.44366 | 0.01299 | 92  | 66.70150 | 24.7422    | 13.130794 11 24 2 5.356 284 0 |
| 22    | 356  | yes                              | 0.997983485 | 83.5  | 1.05781 | 81.39967 | 0.01867 | 82  | 58.11315 | 23.7681          | 14.00037 10 24 1 5.356 115 0 |
| 23    | 223  | yes                              | 0.995751971 | 110.5 | 1.08101 | 108.546  | 0.01769 | 109 | 83.30673 | 25.236          | 12.13092 13 25 2 5.356 779 178 |
| 24    | 281  | yes                              | 0.981501375 | 82.5  | 1.03274 | 80.5877  | 0.02319 | 81  | 63.187385 | 17.3996       | 15.127536 15 30 1 5.356 800 0 |
| 25    | 139  | no                               | 0.000000035 | 1.03231 | 3.39047 | 0.001129 | 2   | 46.43156 | No CPZ | VALUE | No CPZ | VALUE | No CPZ | VALUE | 1 5.356 1592 69 |

**Figure C.6 Estimating Collision Frequency Using Developed Models**
### Figure C.7 Ranking of Identified CPZs

**Collision Prone Zone (CPZ) Search in Kelowna**

<table>
<thead>
<tr>
<th>TAZ #</th>
<th>CPM:</th>
<th>Step 1A: if (1-8)&gt;0.95, then it is a CPZ</th>
<th>Calculated</th>
<th>ICBC</th>
<th>From CPM</th>
<th>E(\lambda)</th>
<th>Step 1B: Rank them</th>
<th>TAZ Data for Model (CPM) prediction to compute E(\lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>357</td>
<td>Yes</td>
<td>362.5</td>
<td>1</td>
<td>1.01686</td>
<td>356.488</td>
<td>0.31658</td>
<td>Rank[10,K(10,K153,0)]</td>
</tr>
<tr>
<td>6</td>
<td>130</td>
<td>Yes</td>
<td>483.5</td>
<td>1</td>
<td>1.00141</td>
<td>479.597</td>
<td>0.00867</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>7</td>
<td>192</td>
<td>Yes</td>
<td>526.5</td>
<td>1</td>
<td>1.0066</td>
<td>523.049</td>
<td>0.00655</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>11</td>
<td>280</td>
<td>Yes</td>
<td>227.5</td>
<td>1</td>
<td>1.02025</td>
<td>222.984</td>
<td>0.01985</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>11</td>
<td>280</td>
<td>Yes</td>
<td>227.5</td>
<td>1</td>
<td>1.02025</td>
<td>222.984</td>
<td>0.01985</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>15</td>
<td>141</td>
<td>Yes</td>
<td>87.5</td>
<td>1</td>
<td>1.05334</td>
<td>82.289</td>
<td>0.05957</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>11</td>
<td>159</td>
<td>Yes</td>
<td>197.5</td>
<td>1</td>
<td>1.01942</td>
<td>193.737</td>
<td>0.01905</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>16</td>
<td>304</td>
<td>Yes</td>
<td>383.5</td>
<td>1</td>
<td>1.00755</td>
<td>380.628</td>
<td>0.00749</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>17</td>
<td>178</td>
<td>Yes</td>
<td>159.5</td>
<td>1</td>
<td>1.01501</td>
<td>157.141</td>
<td>0.01479</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>19</td>
<td>202</td>
<td>Yes</td>
<td>0.99999993</td>
<td>1</td>
<td>1.01051</td>
<td>157.141</td>
<td>0.01479</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>20</td>
<td>140</td>
<td>Yes</td>
<td>94.5</td>
<td>1</td>
<td>1.0237</td>
<td>92.3118</td>
<td>0.02316</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>24</td>
<td>247</td>
<td>Yes</td>
<td>0.99968946</td>
<td>1</td>
<td>1.01241</td>
<td>157.585</td>
<td>0.01225</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>25</td>
<td>356</td>
<td>Yes</td>
<td>83.5</td>
<td>1</td>
<td>1.0258</td>
<td>81.3866</td>
<td>0.02152</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>30</td>
<td>223</td>
<td>Yes</td>
<td>0.99575179</td>
<td>1</td>
<td>1.01801</td>
<td>108.546</td>
<td>0.01669</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>30</td>
<td>281</td>
<td>Yes</td>
<td>0.98150173</td>
<td>1</td>
<td>1.02374</td>
<td>80.587</td>
<td>0.02319</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>Yes</td>
<td>3.5</td>
<td>3</td>
<td>1.03231</td>
<td>3.59047</td>
<td>0.03129</td>
<td>Rank[10,K(10,K155,0)]</td>
</tr>
</tbody>
</table>
5. The spreadsheet uses the “Rank” function to calculate two ranks based on PCR and CRR values separately for each zone and adds them together to calculate the combined ranking. The zones with the highest-ranking scores are those in most need of attention.