

**DEVELOPING MACRO-LEVEL COLLISION PREDICTION MODELS TO ENHANCE
TRADITIONAL ROAD SAFETY IMPROVEMENT PROGRAMS AND EVALUATE
BICYCLE SAFETY IN THE CITY OF VANCOUVER**

by

Bianca Popescu

B.ES., The University of Waterloo, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Civil Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

July 2016

© Bianca Popescu, 2016

Abstract

To encourage greener cities while reducing transportation impacts such as climate change, traffic congestion, and road safety issues, governments have been investing in sustainable transportation modes such as cycling. A safe and comfortable cycling environment is critical to encourage bicycle trips, since cyclists are subject to greater safety risks and represent the highest share of severe and fatal road collisions. Traditionally, engineering approaches have addressed road safety in reaction to existing collision histories. For bicycle collisions, which are rare events, a proactive approach is more appropriate. This study described the development of bicycle related macro-level (i.e. neighbourhood or traffic analysis zone level) Collision Prediction Models (CPMs) and tested the models as empirical tools for bicycle road safety evaluation and planning. This study was unique in its usage of the bicycle exposure variable represented by Bicycle Kilometers Travelled (BKT) as a lead exposure variable in the models. The macro-level CPMs that were developed for bicycle-vehicle collisions were applied to a case study of the City of Vancouver at the zonal level. The objectives of the study were to: (1) identify bicycle data safety indicators, (2) develop bicycle macro-level CPMs using generalized linear regression modeling (GLM), (3) demonstrate model use by applying them to a case study of the City of Vancouver through a macro-reactive road safety application, and (4) identify potential safety countermeasures for the highest ranked Collision Prone Zones (CPZs). The models were effective in enhancing traditional road safety initiatives and identifying and ranking dangerous CPZs in the City of Vancouver. The top three collision prone areas were then brought forward for diagnosis and remedy analysis. This case study effectively demonstrated the use of the models to proactively enhance bicycle safety.

Preface

This thesis did not require ethics approval and has not been previously published. I designed the scope of this thesis with direction from my thesis supervisor, Dr. Tarek Sayed. I performed the literature review, data preparation and analysis primarily on my own, with some advice from my supervisor. Throughout the analysis, I received some guidance through discussions with Clark Lim, Ahmed Osama and Emanuele Sachi, who are all currently PhD students at the University of British Columbia.

Table of Contents

Abstract.....	ii
Preface.....	iii
Table of Contents	iv
List of Tables	viii
List of Figures.....	ix
Glossary	xi
Chapter 1: Introduction	1
1.1 Challenges.....	1
1.2 Motivations	2
1.2.1 Environmental Concerns.....	3
1.2.2 Public Health Benefits	4
1.2.3 Proactive Road Safety Initiatives.....	4
1.3 Problem Statement.....	6
1.4 Research Objectives.....	7
1.5 Thesis Structure	8
Chapter 2: Literature Review.....	9
2.1 Considerations of Transportation Safety in Planning	9
2.1.1 Safety in Long-Range Transportation Planning	9
2.1.2 Lack of Proactive Safety Considerations in Transportation Planning	12
2.1.3 Proactive Transportation Planning Bicycle Safety Considerations	13
2.2 Road Safety Improvement Programs	14

2.3	Macro-Level Collision Prediction Models.....	18
2.3.1	Previous Research on Macro-Level Collision Prediction Models	18
2.3.2	Previous Research on Systematically Applying CPMs to the Planning Process.....	21
2.4	Developing Bicycle- Related Macro-Level CPMs	23
2.4.1	Identifying Bicycle Safety Indicators	23
2.4.2	Macro-Level Bicycle Collision Prediction Models Using Generalized Linear Modelling	25
2.4.3	Macro-Level Bicycle Collision Prediction Models Using Other Statistical Techniques	27
2.5	Data Assembly	29
2.6	Model Development.....	33
2.6.1	Model Form	35
2.6.2	Error Structure	35
2.6.3	Selection of Explanatory Variables	36
2.6.4	Evaluation of Goodness of Fit	37
2.6.5	Outlier Analysis	38
Chapter 3: Methodology.....	40	
3.1	Data	40
3.1.1	Geographic Scope	40
3.1.2	Aggregation.....	41
3.1.3	Sources	42
3.1.4	Quality Issues.....	47
3.1.4.1	Bicycle Collision Data	48

3.1.4.2	Boundary Effects	48
3.1.4.3	Multiple Sources	49
3.2	Model Development.....	49
3.2.1	Selection of Explanatory Variables	50
3.2.2	Addition of Variables to Bicycle Collision Prediction Models	51
3.2.3	Application of CPMs to Macro-Reactive Black Spot Programs	53
3.2.3.1	Identification and Ranking.....	54
3.2.3.2	Diagnosis.....	55
3.2.3.3	Remedy	57
Chapter 4: Results and Discussion	58	
4.1	Approach and Results	58
4.1.1	Development and Selection of Models	58
4.1.2	Identification and Ranking.....	60
4.1.3	Diagnosis and Remedy	62
4.1.3.1	Collision Prone Zones 3160 & 3200.....	67
4.1.3.2	Collision Prone Zones 3460 & 3640.....	71
4.1.3.3	Collision Prone Zone 2290	74
Chapter 5: Conclusion	78	
5.1	Summary and Conclusions	78
5.1.1	Bicycle Macro-Level CPM Development	78
5.1.2	Macro-Reactive Road Safety Application	80
5.2	Contributions.....	81
5.3	Recommendations for Future Research	82

References	84
Appendices.....	94
Appendix A.....	94

List of Tables

Table 1: Additional variables influencing bicycle use according to Wei, Alam & Lovegrove [21] [38]	26
Table 2: Potential included variables for macro-level CPMs	31
Table 3: Explanatory variables and summary statistics.....	47
Table 4: Regional statistics in the City of Vancouver for included variables	56
Table 5: Collision Prediction Models and their goodness of fit summary statistics.....	59
Table 6: A list of the top 7% of CPZs rankings (top ten)	61
Table 7: Bicycle safety countermeasures.....	66

List of Figures

Figure 1: Methodology outline	40
Figure 2: City of Vancouver bicycle routes	41
Figure 3: City of Vancouver Traffic Analysis Zones	42
Figure 4: City of Vancouver intersections	44
Figure 5: Annual Average Daily Traffic (AADT) in the City of Vancouver	45
Figure 6: ICBC data on bicycle-vehicle collisions from 2009-2011 for the City of Vancouver..	46
Figure 7: Comparison between micro- and macro-reactive Black Spot Programs [38] [40]	54
Figure 8: Top ten CPZs for the exposure model using BKT	62
Figure 9: New cycling infrastructure built in the City of Vancouver [86]	65
Figure 10: Collision Prone Zones 3160 (north) & 3200 (south) showing bike routes and bicycle-vehicle collisions.....	70
Figure 11: Collision Prone Zones 3160 (north) & 3200 (south) showing topography (1 m pink contour lines and 10 m purple contour lines).....	70
Figure 12: Collision Prone Zones 3460 (north) & 3640 (south) showing bike routes and bicycle-vehicle collisions.....	73
Figure 13: Collision Prone Zones 3460 (north) & 3640 (south) showing topography (1 m pink contour lines and 10 m purple contour lines).....	74
Figure 14: Collision Prone Zone 2290 showing bike routes and bicycle-vehicle collisions	77
Figure 15: Collision Prone Zone 2290 showing topography (1 m pink contour lines and 10 m purple contour lines)	77
Figure 16: Top ten CPZs for the reference exposure model (BKT only)	94

Figure 17: Top ten CPZs for an exposure model (AAAKM, PMP)	95
Figure 18: Top ten CPZs for a TDM model (TCM)	96
Figure 19: Top ten CPZs for a socio-demographic model (COM, EMP).....	97
Figure 20: Top ten CPZs for a socio-demographic model (EMP, HHD)	98
Figure 21: Top ten CPZs for a socio-demographic model (POP, INTD)	99
Figure 22: Top ten CPZs for a network model (SIGD, BS)	100
Figure 23: Top ten CPZs for a network model (PMP, AKM)	101
Figure 24: Top ten CPZs for a network model (INTD, BS)	102
Figure 25: Top ten CPZs for a network model (RPKM, INTD).....	103

Glossary

α	Alpha, shape parameter
AAA	Bicycle infrastructure for users of all ages and abilities
AAAKM	Total "AAA" Bicycle Kilometers
AADT	Annual Average Daily Traffic
a_1	Model parameter
AKM	Total Arterial Kilometers
a_o	Model parameter
β	Beta, scale parameter
b_i	Model parameter
BKT	Bicycle Kilometers Traveled
Black Spots	Collision prone areas or locations
BS	Bus Stops
CD	Cook's Distance
COM	Commercial Land Use
COV	City of Vancouver
CPM	Collision Prediction Model
CPZ	Collision Prone Zone
DRA	Digital Road Atlas
$E(\Lambda)$	Predicted Collision Frequency
EB	EB Safety Estimate
EMME/2	Transportation forecasting software
EMP	Employment
GIS	Graphical Information Systems
GLM	Generalized Linear Regression Modeling
GWR	Geographically Weighted Regression
HHD	Household Density
ICBC	Insurance Corporation of British Columbia
INTD	Intersection Density
κ	Overdispersion parameter
Macro-level	Area-wide (i.e. zonal)
Micro-level	Single location (i.e. intersection)
n	Number of observations
NB	Negative Binomial
σ_d	Dispersion parameter
p	Number of parameters
PCR	Potential Collision Reduction
PDO	Property Damage Only
PMP	Percentage of Park Area
POP	Population

PR_i	Pearson Residual
RPKM	Percentage of Residential Street Kilometers
RSIP	Road Safety Improvement Programs
SAS	Statistical software used for CPMs
SD	Scaled Deviance
SIGD	Signal Density
TAZ	Traffic Analysis Zone
TB3	Total bicycle collisions over 3 years
TCM	Total Commuters
TDM	Transportation Demand Management
TLKM	Total Lane Kilometers
Trigger Variables	Variable hypothesized to contribute to classification of CPZ
VC	Neighbourhood Congestion
VKT	Vehicle Kilometers Traveled
VRU	Vulnerable Road Users
X^2	Pearson Chi Squared
X_i	Explanatory variable
y_i	Observed mean collision frequency
Z	External exposure variable

Chapter 1: Introduction

The transportation-related challenges of climate change, traffic congestion, public health and road safety are impacting cities worldwide. To encourage greener cities, governments are prepared to expand sustainable transportation systems for walking, bicycling, public transit, and car sharing to enable a better balance between private motorized transport and sustainable modes of transport. In contrast to motor vehicle travel, the public health benefits of active transportation, such as cycling and walking, are significant. In the case of bicycle transport, some key advantages include: energy efficiency, low cost, health benefits and zero emissions, as well as an effective use of road space and parking. In addition, bicycles are the most efficient and fast mode for the short- and medium-distance trips common in urban areas. To encourage active transportation, policies need to seriously consider the safety of these road users. Sustainable transportation planning and design that enables less dependence on single occupancy vehicles and promotes the safe use of bicycles could be an effective way to achieve a sustained reduction in road collision risk, frequency, and severity for cyclists. To emphasize the importance of bicycle safety research, this introduction will cover the challenges and motivations behind this important topic.

1.1 Challenges

The social and economic burdens due to road collisions are recognized as a global problem. According to the World Health Organization's global status report on road safety, road traffic injuries are the eighth leading cause of death globally, remaining unacceptably high at 1.24 million road traffic deaths per year [1]. Current global trends suggest that by 2030 road traffic deaths will become the fifth leading cause of death unless urgent action is taken [2].

Canada is a highly developed country, but despite significant investment in road infrastructure there are still concerns about road collisions. In Canada, the total number of reported motor vehicle fatalities and injuries in 2013 were 1,923 and 165,305, respectively [3]. In addition to the human cost, collisions put a large economic burden on society. Transport Canada estimates the annual cost of total collisions at \$62.7 billion for the Canadian economy [4].

Across Canada, cycling is growing in popularity as a daily commuting option, providing a convenient and affordable alternative to the congested roads and crowded transit systems in urban areas [5]. However, inadequate infrastructure and unsafe environments present a threat to meeting sustainable transportation system goals. Vulnerable road users (VRUs), such as pedestrians and cyclists, are subject to greater safety risks and represent the highest share of severe and fatal road collisions [6]. In British Columbia in 2013, 100% of the 1,500 reported automobile collisions involving bicycles resulted in an injury or fatality for the cyclist [7]. This thesis focuses on bicycle collisions with motor vehicles, since studies show that such collisions are more severe to the cyclist [8] [9]. The safety of cyclists is critical since bicycle travel has a higher per-mile casualty rate than car travel, yet poses minimal risk to other road users [10]. In addition, a study by the Ontario Coroner found that the majority of cycling deaths are preventable [11], making this an important topic of interest.

1.2 Motivations

Safety is key to increasing cycling mode share and is a critical indicator of the performance of sustainable transportation plans and cycling facilities [5]. Advancing sustainability and safety in active transportation, specifically cycling, are the motivations behind this thesis. The following sections outline the importance of this topic.

1.2.1 Environmental Concerns

The growing interest in developing multimodal transportation networks is driven by the increasing awareness of environmental issues, both globally and locally. Significant effort from urban planners, engineers and decision makers is required to build a multimodal transportation system that is not dominated by private vehicles. This is important because despite improvements in vehicle efficiency and technology, it is simply not sustainable to continue to plan urban environments solely for the private automobile. Strategies to reduce vehicle travel through promoting sustainable transportation modes help to conserve energy and reduce pollution emissions, as well as provide efficiency benefits such as the potential to reduce congestion, roadway and parking costs, vehicle collisions and sprawl [12].

A safe and comfortable cycling environment is critical to encourage bicycle trips, therefore allowing urban areas to reduce their carbon footprint, reduce their dependence on fossil fuels and lower greenhouse gas emissions [6]. This is because increasing the number of sustainable transportation users, such as cyclists, would not contribute to air pollution. In British Columbia, the transportation sector is the leading contributor to greenhouse gas emissions, with passenger vehicles contributing 39% of the transportation-related greenhouse gas emissions [13]. This is an important statistic, because increasing the mode share of active transportation is assumed to decrease the amount of people travelling by motorized vehicles. Transportation planning for bicycle facilities and improving bicycle safety are emerging areas of research that have yet to be developed to a level that matches research on vehicular traffic.

1.2.2 Public Health Benefits

Bicycling as a form of active transportation and exercise has the capacity to benefit public health by reducing the problems associated with heart disease, obesity, high blood pressure, type 2 diabetes, osteoporosis and depression. When cycling is available as a daily travel mode, substantial public health benefits occur, potentially decreasing spending on health care by over \$500 per year per person according to a British Columbia study [13]. The correlation between active transportation and obesity, for example, has been shown in countries around the world, with nations having a greater mode share of pedestrians, cyclists and transit users also having lower obesity rates [14]. Previous research has found that substantial health benefits received from cycling can offset the increase in collision risk, because longevity is found to increase with active transportation [15]. Therefore, to benefit public health, it is imperative to improve road safety for cyclists to decrease risks and encourage cycling for all ages and levels.

1.2.3 Proactive Road Safety Initiatives

Road safety improvement programs can be divided into two categories: reactive (i.e. responding to road safety problems), and proactive (i.e. taking actions to prevent the emergence of road safety problems). The traditional reactive road safety improvement programs (RSIPs) focus on the identification, diagnosis and remedy of collision-prone locations, or “black spots,” in reaction to observed collision history. Although RSIPs are important and have been proven to be successful, this reactive program requires a significant collision history to exist before improvement action is taken. These collision-based approaches rely on data typically drawn from collision records, police reports, and insurance claims [16]. The traditional reactive approach to safety based primarily on collision data has been challenged on several accounts:

- Attribution: it can be difficult to precisely know the cause and exact situation and location from information obtained by police and insurance reports, and therefore the definitive reason of the collision;
- Data volume: despite the social and economic burden of road collisions, they are infrequent rare events, making it challenging to draw statistically stable and significant inferences from actual collision data;
- Data quality and availability: the process of reporting collisions is based on subjective observations and witness accounts, and records can be incomplete and lack details, as well as non-injurious or non-property-damaging incidents can go unreported; and,
- Ethical concerns: paradoxical situation in which the safety analyst strives to observe events that they are attempting to prevent [17].

Using reactive road safety analysis, there is little ability to predict and prevent road safety issues before their induced costs. Therefore, there is a need to take a proactive approach to address safety before problems emerge, to complement traditional reactive methods [18] [19]. A proactive approach would be one that addresses road safety as an important factor in the evaluation of transportation projects at the planning stage [20], before collisions occur. It is particularly important to study the safety of VRUs, as the number of collisions observed is small but the severity of the collisions on pedestrians and bicyclists is high.

Typically, when implementing transportation investments, decision-making begins at the planning level, when current and future transportation needs are defined by planning scenarios. The modeled scenario results are quantified and weighted against economic, environmental and societal impacts. Transportation safety is an important element that should be considered at the planning level. While economic, environmental or social considerations can be quantified based

on a large amount of science and empirical data, the estimation of safety in transportation planning is still under development [19] [21] [22]. Until recently, practitioners lacked the empirical tools necessary to evaluate road safety proactively. The challenge is for planners, practitioners and decision makers to plan and design innovative, sustainable and safe transportation systems. To improve the safety of VRUs, there is a need for empirical tools to enable the evaluation of the bicycle safety proactively before collisions occur.

1.3 Problem Statement

The study of bicycle safety is important due to the physical vulnerability of cyclists and the key role that this mode of transportation plays in building sustainable, efficient and healthy communities. Challenges facing the study of cycling safety include the fundamental problems of reliance on the collision data that has quantity and quality issues and the high cost of bicycle collisions. In the City of Vancouver (COV), for example, VRUs (pedestrians, cyclists and motorcyclists) accounted for approximately 3% of road collisions between 2007 and 2013, however these users accounted for approximately 80% of fatalities during this time period. Looking towards the future, the city has a target for zero traffic-related fatalities [23]. This thesis will apply macro-level CPMs to the COV as a case study to proactively estimate bicycle safety. The development of these bicycle-related models is feasible and effective due to the city's growing cycling mode share, expanding bicycle infrastructure and rich databank of bicycle exposure data. Cycling accounted for approximately 4.4% of trips to work in the COV in 2011 [24], the base year of this study, and the number of people choosing to cycle every year is growing.

A recent City of Vancouver Cycling Safety Study performed an analysis based on reported cycling collisions, and identified the following safety issues in the city: ‘doorings’ from on-street vehicle parking, mid-block conflict zones, ‘right hook’ collisions with right turning vehicles, ‘left cross’ conflicts with left turning vehicles, two-way stop unsignalized intersections, streets without designated bikeway infrastructure, the PM peak period, and poor lightings and weather [23]. This study will build upon the City of Vancouver Cycling Safety Study [23] by developing macro-level Collision Prediction Models (CPMs) to proactively evaluate bicycle safety. CPMs are mathematical models that relate the collision frequency to exposure, geometric and socio-demographic characteristics, in order to estimate the expected collision frequency for a location and to facilitate safety planning. This study uses CPMs, developed using the generalized linear modelling approach, to overcome the shortcomings associated with traditional linear regression [25]. This modelling technique that predicts bicycle-vehicle collisions at the macro-level will produce models that can be used as a reliable empirical based decision tool for planners and engineers.

1.4 Research Objectives

Several studies have previously developed macro-level CPMs. However, little research has been undertaken to develop macro-level CPMs for bicycle collisions. The few studies that have developed macro-level CPMs for bicycle collisions have used proxy bicycle exposure measures such as bicycle lane kilometers and Vehicle Kilometers Travelled (VKT). This study is unique in its usage of the actual bicycle exposure variable represented by Bicycle Kilometers Travelled (BKT) as a variable in the models. Following methodologies previously developed in macro-level CPMs research, the objectives of this study are:

1. To identify data needs and collect exposure, infrastructure and socio-demographic variables that would influence bicycle safety to input into the statistical model;
2. To develop bicycle macro-level CPMs using generalized linear regression modeling (GLM);
3. To demonstrate the use of the bicycle macro-level CPMs by applying the models to a case study of the COV to proactively evaluate the safety of bicycle use, through identifying, ranking and diagnosing collision prone zones (CPZs) to evaluating the level of bicycle safety to inform future COV bicycle transportation plans; and,
4. To identify potential safety countermeasures for the highest ranked CPZs that could be applied to each area to improve bicycle safety.

1.5 Thesis Structure

This thesis is organized in five chapters. Chapter 1 introduces the challenges and motivations behind the topic of bicycle safety, as well as the research problem statement and objectives. Chapter 2 is a literature review covering the application of proactive road safety in transportation planning, the introduction of road safety improvement programs, the history of macro-level Collision Prediction Models (CPMs) and the process of model development. The literature review concentrates on previous bicycle safety studies, including the development of bicycle related macro-level CPMs. Chapter 3 describes the methodology of the thesis, including data extraction, model development and the application of the CPMs to macro-reactive Black Spot Programs. Chapter 4 presents the model results, as well as the identification, ranking, diagnosis and remedy of Collision Prone Zones (CPZs). Finally, Chapter 5 summarizes the research, conclusions, contributions and recommendations for future work.

Chapter 2: Literature Review

The purpose of this literature review is to provide a summary of the research foundation on which this thesis has been built. The representative review will focus on the consideration of safety in the transportation planning process, previous work on road safety improvement programs, the development and application of macro-level CPMs, the selection of active transportation indicators for developing bicycle-related macro-level CPMs, and finally the process of data assembly and model development.

2.1 Considerations of Transportation Safety in Planning

2.1.1 Safety in Long-Range Transportation Planning

Incorporating safety directly into the long-range transportation planning process can be an effective way to reduce the large societal and economic costs of collisions. The increasing need to consider road safety proactively within the transportation planning process has been driving road safety research throughout the past twenty-five years. The Netherlands' SWOV Institute for Safety Research first developed the “Sustainable Safety” approach in 1992, a policy vision with the objectives to proactively prevent severe collisions before they occur [26]. Van Schagen & Janssen (2000) proposed further work on the implementation of “Sustainable Safety” policies targeting the organizational and financial aspects [27]. This work on collision prevention was continued with “Advancing Sustainable Safety: National Road Safety Outlook for 2005-2020” (2006), which updated the existing research on “Sustainable Safety” and recommended implementation, innovation and reinforcement of research and development in the proactive safety planning field [28]. Weijermars et al. (2013) applied the “Sustainable Safety” vision to

cycling safety and states the importance of short, convenient and recognizable routes, and separation of bicycles from vehicles [29].

In the U.S., the Congress' Transportation Equity Act for the 21st Century (TEA-21, 1998) required safety and security to be incorporated into the planning process, the first time the U.S. federal government committed to including safety in transportation planning. While the Act showed a federal commitment to develop strong linkages between safety and planning, Berkovitz (2001) identified the inadequacy of existing highway safety tools to fulfill this Act's mandate. For example, a high-hazard location analysis may improve a dangerous location, but it will not correct safety problems for VRUs such as high speed and volume roadways with poorly designed and discontinuous facilities. The U.S. Department of Transportation recommended two national goals in "The National Bicycling and Walking Study" (1994): to double the number of trips made by cycling and walking, and to reduce the number of cyclists and pedestrians killed or injured in traffic collisions by 10%. Berkovitz (2001) stated the importance of planning for both goals simultaneously, or else efforts to achieve one goal can hinder the other [30]. Furthering this direction, researchers de Leur & Sayed (2003) developed a method to address road safety proactively in transportation planning before problems occur to complement the traditional reactive safety methods. Their goal was to develop a systematic framework for planners to use to effectively address road safety concerns when developing planning alternatives [19].

More recently in the U.S., in 2007, an update to the Statewide and Metropolitan Transportation Planning Final Rule (23 CFR 450) encouraged the strengthening of safety within the long-range transportation planning process. The Transportation Safety Planning Framework (2012) demonstrated how planners could integrate safety into every step of the traditional planning process and use safety as a decision factor in transportation plans [31]. The National

Cooperative Highway Research Program's Report 05-46 "Incorporating Safety into Long-Range Transportation Planning" (2006), provided transportation planners with comprehensive tools and strategies to consider safety in the planning process, such as an aggregate planning level safety prediction tool. The intent of the tool is to enable the forecasting of safety at the Traffic Analysis Zone (TAZ) level. The report found the safety prediction model appropriate in the following cases:

- Setting safety targets or performance measures;
- Understanding the safety impacts of large-scale projects (corridor level or higher) that may affect the Vehicle Kilometers Travelled (VKT), future growth, or other planning-related factors in the absence of targeted safety countermeasures; and,
- Comparing and contrasting growth scenarios in the absence of targeted safety countermeasures.

These analyses would be used to understand the different levels of safety investment required to meet regional safety performance targets. The report stated the model to be inappropriate in the following cases:

- Selecting land use and transportation investment strategies based on model results, as there are many different factors when considering an investment and safety countermeasures could be applied; and,
- Evaluating and selecting safety countermeasures, because the models predict but do not explain collisions [32].

2.1.2 Lack of Proactive Safety Considerations in Transportation Planning

Globally, the World Health Organization’s “World report on injury prevention” (2004) identified road crash injuries as predictable and preventable, and recommended the need to incorporate safety as a long-term goal into land-use and transportation planning [33]. Today, safety-conscious transportation planning is regarded as a vital addition to traditional reactive safety approaches to reduce road fatalities and injuries by supporting comprehensive, system-wide, multimodal, data-driven, and proactive transportation planning processes that integrate safety into transportation decision-making.

Popular methods of sustainable safety practices, such as road safety audits, overtly consider road safety during the design process. With these methods, there is the risk that planners will consider that safety concerns would be addressed in the design stage, and will not consider safety at the preceding planning stage. It is therefore essential to develop a framework to proactively evaluate safety at the planning stage. de Leur and Sayed proposed a framework to proactively evaluate road safety in the planning process (2003), by systematically considering safety within the planning environment for planners to understand the safety impacts of the plans they develop [19]. Gaines & Meyer (2008) surveyed mid-size metropolitan planning organizations in the U.S. and found that the majority had incorporated safety considerations within their long-range transportation plans, but some are more proactive in the quantitative analysis of safety outcomes than others [34]. Despite policy advancements, transportation safety is still not commonly considered proactively during the transportation planning processes, in part due to the lack of available empirical tools.

2.1.3 Proactive Transportation Planning Bicycle Safety Considerations

Transportation forecasting models, which help to inform decision makers on project prioritization in transportation planning, have traditionally excluded pedestrians and cyclists. To plan for future safe active transportation friendly communities, quantifying the use and potential demand of pedestrian and bicycling facilities is necessary. The Pedestrian and Bicycle Information Center (2015) summarized the state of practice of pedestrian and bicycle forecasting tools. The research suggested that modifying regional demand forecasting models to incorporate bicycle and pedestrian modes could provide consistency throughout the planning process. However, potential drawbacks include: Traffic Analysis Zones (TAZs) may be too large to capture internal trips, the number of trips may be less than the margin of error of model validation, and the models typically exclude non-motorized trips from the trip distribution and route selection steps. Alternative forecasting model tools include factor methods and sketch planning tools (using existing count data and elasticity assumptions for projections), aggregate demand models (regression models using existing activity and influencing attribute data) and Geographic Information Systems (GIS) and other spatial tools [35].

Traditional reactive road safety approaches require a significant observed bicycle collision history to exist before improvement action is taken. This is a problem for VRUs, because the number of observed collisions may be too small for significant statistical analysis but the severity on pedestrians and bicyclists is high. Due to the shortage of comprehensive bicycle exposure data available, such as bicycle counts and forecasted volumes, it has previously been difficult to proactively consider bicycle safety in the planning process, and previous research has not yet done so effectively. Developing an empirical method to systematically consider bicycle safety in transportation planning is an important topic of research to build onto existing

literature, because this is a sustainable and healthy but vulnerable transportation mode that is becoming increasingly popular.

2.2 Road Safety Improvement Programs

The objective of reactive Road Safety Improvement Programs (RSIPs), or black spot programs, is to identify and treat locations that are considered hazardous based on the analysis of collision, traffic and highway data. RSIPs involve the following purposes:

- 1) To identify hazardous *locations* and detect black spots;
- 2) To identify *problems* through location diagnosis; and,
- 3) To identify *solutions* and remedies by finding countermeasures to solve the problems [36].

The assumption of RSIPs is that road design typically plays a significant role in contributing to collision frequency. Sayed et al. (1995) found that road-related factors have caused about 32% of collisions based on analysis of collision data in British Columbia [37]. For that reason, improving the transportation engineering elements of black spots can significantly decrease a proportion of collisions. Properly identifying and ranking black spots for diagnosis and treatment is important to ensure that resources are only spent on areas with highest potential collision reduction (PCR).

A black spot is defined as a location or area that are found to have a significantly high collision potential compared to a group of similar locations, typically through the measure of collision frequency. The collision prediction models' (CPMs) use of an exposure variable corrects for possible frequency bias when comparing different locations with different traffic volumes (exposure levels). To ensure that only truly hazardous locations are identified as black

spots and avoid regression-to-the-mean (RTM) errors, a popular statistical technique used to reduce the selection bias is the Empirical Bayes (EB) technique. The EB technique defines the collision probability of the mean collision frequency of a given area as dependent on the observed mean collision frequency of the location and an objective prior distribution based on empirical data from a reference population or from using CPMs. CPMs have the advantage of analyzing collision frequency instead of collision rates and eliminate the need for a very large reference population [38]. In addition, using CPMs allows for location-specific prior distribution to be derived for each location from an imaginary reference population, of which the estimates of mean and variance have shown in previous research to be more reliable than a sample from real reference population [39]. It is therefore vital that the appropriate CPM is selected on the basis of traits examined and type of safety estimate [40].

After the development of CPMs (as described in subsequent sections 2.3, 2.4 and 2.5) and the selection of the appropriate model the first step is to estimate the safety of the prior distribution, $E(\Lambda)$, by inputting the values of each location's values into the CPM equation. The prior distribution variance, is calculated as:

$$Var E(\Lambda) = \frac{[E(\Lambda)]^2}{\kappa} \quad (2.1)$$

and assumes that the prior distribution follows a gamma distribution with the shape and scale parameters, alpha α and beta β , shown as:

$$\alpha = \kappa \quad (2.2)$$

$$\beta = \frac{\kappa}{E(\Lambda)} \quad (2.3)$$

Where:

$E(\Lambda)$: predicted collision frequency

κ : overdispersion parameter

α : alpha, shape parameter

β : beta, scale parameter

This model outputs allow for the development of location-specific estimates for $E(\Lambda)$ and $Var E(\Lambda)$. The following step is to gather local observed collision history data to refine the prior estimate provided by the CPM to define the posterior distribution. This posterior distribution represents how the mean collision frequency varies in a subpopulation of variables having similar traits in terms of traffic, geometry and collision history, and is gamma distributed with the shape and scale parameters α and β shown as:

$$\alpha = \kappa + count \quad (2.4)$$

$$\beta = \frac{\kappa + E(\Lambda)}{E(\Lambda)} \quad (2.5)$$

Where:

$count$: the location's or zone's collision history.

The mean of the posterior distribution is in other words the EB safety estimate for location i , EB_i . The mean, EB_i , and the variance, $Var(EB_i)$, of the posterior distribution are:

$$EB_i = E(\Lambda|Y = count) = \left[\frac{E(\Lambda_i)}{\kappa + E(\Lambda_i)} \right] (\kappa + count) \quad (2.6)$$

$$Var(EB_i) = Var(\Lambda|Y = count) = \left[\frac{E(\Lambda_i)}{\kappa + E(\Lambda_i)} \right]^2 (\kappa + count) \quad (2.7)$$

Where:

$$EB_i : \text{EB safety estimate for location, } i$$

The final step to identify black spots is to compare the results of the value of each location's safety, EB_i to the regional average or norm for locations with similar traits. Each location would be considered collision prone if there is a significant probability, σ , usually 0.95 or 0.99, that the EB safety estimate, EB_i exceeds the specified standard. The location is identified collision prone if the following condition is met:

$$1 - \int_0^{E(\Lambda)} f_{EB}(\lambda) d\lambda = \left[1 - \int_0^{E(\Lambda)} \frac{[\kappa/E(\Lambda) + 1]^{(\kappa+count)} \lambda^{(\kappa+count-1)} e^{-[\kappa/E(\Lambda)+1]\lambda}}{\Gamma(\kappa + count)} d\lambda \right] \geq \sigma \quad (2.8)$$

Where:

$$\sigma : \text{probability that the EB safety estimates exceed a specific value (usually 0.95 or 0.99)}$$

Once a location has been determined as collision prone, ranking must occur to ensure the locations in most need of treatment are looked at first. Ranking criteria has typically used the potential collision reduction (PCR), based on the difference between expected and observed collision frequency:

$$PCR_i = EB_i - E(\Lambda)_i \quad (2.9)$$

Following black spot identification and ranking for treatment, a safety diagnosis is performed. First, collision history is analyzed to identify an overrepresentation of clusters of specific collision types by comparing percentages of specific collision types to other similar locations.

Second, location specific traits are identified and analyzed to identify potential causes of overrepresented collision types. Once the safety issue has been identified, the next step is to generate a list of potential remedies to decide on the remedy for the specific location. The final choice of remedies to implement will involve engineering judgement and experience [38] [40].

2.3 Macro-Level Collision Prediction Models

2.3.1 Previous Research on Macro-Level Collision Prediction Models

Progress in road safety research has improved the empirical tools for proactive road safety management, including the development of macro-level collision prediction models (CPMs). Due to the interest in community-wide proactive safety planning, there has been a growing body of research produced on the development of macro-level CPMs to date. The first studies to develop CPMs as planning tools to predict and explain collisions by Ho & Guarnashelly (1998) and Lord, Persaud & Palmisano (2002) applied micro-level models determining the level of safety for single locations, such as intersections, based on an exposure variable of traffic volume obtained from regional transportation planning models, EMME/2 forecasts [41] [42]. Issues with micro-level CPMs were that they were sensitive to the exposure variable of traffic volume, data intensive, and applied to a singular locational level, making it difficult to predict safety accurately long-term on a larger scale. Further studies found macro-level CPMs as an effective alternate tool for macro-reactive use to identify, diagnose and remedy hazard locations [21] and proactive road safety planning [20]. Macro-level CPMs aggregate explanatory variables to the TAZ level, before relating them to collision occurrence. Successful macro-level CPMs require data assembly and model development, both processes that are explained in subsequent sections.

There are several important studies that apply macro-level CPMs to evaluate safety at the planning stage. Hadayeghi et al. (2003) developed macro-level accident prediction models to be used as a safety criterion in the evaluation of urban transportation systems. The models were developed for Toronto, Canada by applying data to the traffic zone level (463 total zones) with inputs from the regional EMME/2 transportation model. The models found that the number of accidents per zone increased as the VKT, road kilometers, total employment and household population and intersection density increased, and decreased with higher posted speeds and higher congestion. The results found the goodness-of-fit of the models produced mixed results, so the models were further developed into Geographically Weighted Regression (GWR) models to explore the relationship between zonal collisions and explanatory planning variables. However, since the GWR models resulted in inconsistent improvements the authors noted the significance of data quality to their results [43].

Ladron de Guevara et al. (2004) developed CPMs for 859 traffic zones in Tucson, Arizona that associated increased collisions with increased population density, employment density, intersection density and arterial and collector roads lane-miles. Due to the lack of availability of reliable traffic volume forecasts, this study used population density as the leading exposure variable instead of exposure variables typically used (i.e. VKT, AADT). The study found that high population density was related to high collision occurrence; zones with high employment, intersection density and arterial roads were associated with higher number of injury and property damage (PDO) collisions; and, high intersection density was associated with decreasing severity of collisions. The question of causation or correlation of the model variables to collisions has yet to be verified [44].

Lovegrove and Sayed (2006) developed macro-level CPMs calibrated for 577 zones the Greater Vancouver Regional District, now called Metro Vancouver, Canada, incorporating a higher quality of accident data due to the comprehensiveness of the Insurance Corporation of British Columbia (ICBC) data used. The models developed included the total and severe traffic collisions during the AM peak scenario for urban and rural land use. The principle exposure variable used in the models was either vehicle kilometers travelled (VKT) or total lane kilometers (TLKM). This study obtained the exposure variables from two sources: measured with GIS software (TLKM) or modeled from EMME/2 forecasts (VKT). The CPMs were developed using generalized linear regression to predict mean collision frequency based on statistical associations with both measured and modelled exposure variables, to compare results and provide practitioners without access to EMME/2 software an alternative method to estimate exposure variables. The developed models were grouped into four themes according to their explanatory variables:

- Exposure variables consisted of attributes describing the number of vehicles, roads and congestion in each zone (i.e. VKT, TLKM, average congestion, average speed).
- Sociodemographic variables consisted of variables describing residents, workers and land use (i.e. average family size, home density, population density).
- Network variables consisted of variables describing the road network in each zone (i.e. number of traffic signals, intersection density).
- Transportation demand management (TDM) variables consisted of variables describing the characteristics of travel demand in each zone (i.e. total commuters, commuter density, number of drivers commuting).

The study found that increased collisions were associated with the majority of explanatory variables, and decreased collisions were associated with higher family size, percentage of core residential area, number of three-way intersections and local road kilometers [21].

Most macro-level CPMs in previous studies have been developed for vehicle collisions only. Recently there have been models developed taking into account transit characteristics and transit collisions by Cheung et al. (2008) [45] and using transit physical and operational elements and transit network indicators as explanatory variables by Quintero et al. (2012) [46] [47]. In addition, studies developing macro-level CPMs to predict bicycle safety will be outlined in the following section.

2.3.2 Previous Research on Systematically Applying CPMs to the Planning Process

A key step after creating macro-level CPMs is to develop a systematic method to apply the CPMs to the planning environment. The models could be used to optimize the planning process by assessing the safety impacts of different planning alternatives. Chatterjee et al. (2003) summarized a number of crash predictions workshops in the U.S., with the objective to identify tools for assessing the safety impact of long-range transportation plans. The challenge was to develop simplified crash prediction models based on available variables that could be used for long-range forecasts, such as traffic volume. A major recommendation for considering quantified safety in long-range planning was the practice of comparing scenarios on a relative basis. Due to lack of data on bike and pedestrian collisions, the research stated a recommendation to analyze the impacts of alternative scenarios on the safety of bicyclists and pedestrians for policy makers to understand the safety differences for various levels of investment in alternate modes [48]. In

addition, Hirst et al. (2003) noted that the CPMs developed must be updated and recalibrated on a regular basis, because accident risks decline over time [49].

The macro-level CPMs developed in [21] for Metro Vancouver were applied to two road safety planning applications of (1) traffic calming and (2) neighbourhood planning road networks, by Lovegrove & Sayed (2006). While 47 CPMs were developed in [21], a six step selection process considering scope, task, land use type, relevant variable themes, collision theme and data quality was used to select the appropriate number of models needed for each safety application. Reducing the amount of data assembly effort through this process makes the models and guidelines reasonable for use by practitioners, providing them an empirical tool to complement traditional road safety improvement programs [22]. The planning-level safety prediction models developed in [43] for Toronto were further presented with illustrative applications to demonstrate how they could be used as decision support tools for planners to explicitly consider safety in the planning process. The study presented macro-level collision modification factors (CMFs) to illustrate how the models can be used to examine the impact of each planning variable on the safety of an urban zone [50].

Previously developed community-based macro-level CPMs were applied to evaluate the road safety of a transportation plan by Lovegrove, Lim, & Sayed (2010), with the objective to test model use in Metro Vancouver. The researchers found the exposure variable of VC (neighbourhood congestion) to be very influential in model estimates, significance and goodness-of-fit, and recommended that only VC-inclusive CPMs are used. In addition, due to large data extraction and assembly effort, it was recommended that the minimum number of models required to meet data quality and analytical needs are selected, and future research on process automation is needed [20]. The same previously developed community-based macro-level CPMs

were also used to calculate the road safety effects of mobility management strategies such as smart growth, congestion pricing and improved walking and cycling transportation options by Lovegrove and Litman (2008), concluding that these transportation demand management strategies have the potential to improve traffic safety [51].

2.4 Developing Bicycle- Related Macro-Level CPMs

2.4.1 Identifying Bicycle Safety Indicators

As discussed previously, there are limitations with using existing bicycle-vehicle collision data, due to the small nature of bicycle collision datasets. Littman et al. (2000) developed a guide which states bicycle crash data history needs to be evaluated by type of crash and contributing factors, pedestrian and cyclist demographics, location type (intersection midblock, driveways, etcetera), to identify potential problems. However, issues with relying on existing pedestrian and cyclist collision data include: locations with a high frequency of collisions indicate some (unknown) combination of high risk or heavy use, pedestrian and cycling collisions tend to be underreported, and a large number of collisions would have to actually occur to acquire sufficient collision data [52]. For these reasons, it would be beneficial to develop an empirical tool to predict collisions without waiting for collisions data to accumulate. To develop bicycle-related macro-level CPMs, it is necessary to understand bicycle-related indicators. These macro-level CPMs can support the economic justification for decision makers investing in bicycle infrastructure and can aid policy makers in promoting bicycling effectively by quantifying the economic trade-offs.

The Netherlands' SWOV Institute for Safety Research (2014) chose a set of indicators for lack of safety in cycling infrastructure based on a literature review and consultation with road

safety experts, with the objective to use these indicators to assess cycling infrastructure in practice. The study found factors that had the most impact on bicycle safety are: the volume of cyclists (exposure), the visibility at intersections, the density of intersections, the speed differences between road users, the lighting, the pavement surface, the width of the bicycle facility and the degree of separation from road users [53]. In addition, the National Highway Traffic Safety Administration (2014) found that most bicycle deaths occurred in urban areas and non-intersection locations and during the afternoon to evening rush hour [54].

To develop reliable empirical tools, first it is necessary to understand all the factors influencing bicycle use. Xing et al. (2010) classified all possible factors into individual factors defined by a cyclist's characteristics (gender, age, income and experience), social environment factors defined by transportation costs, bicycle culture and existing policies and physical environment factors defined by geography and infrastructure [55]. Previous studies have found that men are more likely than women to cycle [56], household income has a negative effect on bicycle use [57], and car ownership has a negative effect on bicycle use [58]. Population density is also an important predictor, as higher densities typically indicate urban environments with shorter destinations and mixed land-use [32], which are ideal environments for bicycle use. In terms of infrastructure, previous studies have found that good bicycle facilities (i.e. bike lanes, lighting and bicycle parking) promote bicycle mode shift, while a fragmented bicycle network reduces bicycle mode shift [59]. Reynolds et al (2009) performed a literature review on transportation infrastructure and bicycle safety, and concluded that purpose-built bicycle infrastructure reduces collision and injury frequency among cyclists, along with street lighting, paved surfaces and low grades [60].

2.4.2 Macro-Level Bicycle Collision Prediction Models Using Generalized Linear Modelling

Most macro-level CPMs have been developed for vehicle collisions only, and very few researchers have included bicycle variables in the analysis. Due to the increase of bicycle volume on-street and its effect on safety, researchers have recommended the importance of developing bicycle related macro-level CPMs [61]. Researchers who have studied this topic hypothesize that increasing the amount of bicycle infrastructure, which in turn increases bicycle use, will also improve road safety [62] [63]. Lovegrove (2007) developed community-based macro-level CPMs using negative binomial regression for Metro Vancouver and created a bicycle-vehicle collision model, finding increased collisions to be associated with increased bicycle mode share [38]. Wei, Alam and Lovegrove (2011) identified potential factors influencing bicycle use, based on a comprehensive literature review. The study then developed several new macro-level CPMs based on the bicycle-related indicators to estimate the road safety changes resulting from bicycle improvement at the macro-level. In addition to the previously used exposure, sociodemographic, transportation demand management and road network variables [21] [38], the study chose new variables representing factors influencing bicycle use and safety to be considered in new CPM development together with old variables, shown in Table 1.

Table 1: Additional variables influencing bicycle use according to Wei, Alam & Lovegrove [21] [38]

Variable	Abbreviation
Dependent Variables	
Total collisions (auto, bike, transit) over 3 years	T5
Bicycle-only collisions over 5 years	B5
Explanatory Variables	
Bicycle lane kilometers (leading variable)	BLKM
Bicycle kils traveled (leading variable)	BKT
Work-home trip distance (kilometers)	WHD
Share of population at age <=30 (%)	POP<30
Proportion of males/females	M/F
Mean number of vehicles per household (proxy of household income)	VHH
Average gasoline price	GAS
Average annual cost per capita spending on bicycle improvements	ACPB
Number of bicycle racking locations	NBR
Stop sign density for bicyclists	SSBD
Stop sign density for auto	SSD
Average annual precipitation (inches)	PRPT
Average annual temperature (centigrade)	TEPT
Average share of hills (=area of hills/zonal area)	HILL

The study used bicycle lane length as an exposure variable, and found it to have a significantly positive relationship with the dependent variable total collision frequency. Wei, Alam and Lovegrove (2011) identified several shortcomings, the most notable being the absence of a transportation planning model for bicycle use resulting in the omission of the leading exposure variable bicycle kilometers travelled and a lack of bicycle-related variable data, such as temperature and precipitation [63].

Wei and Lovegrove (2013) developed macro-level community-based CPMs for bicycle-vehicle collisions, using data from the Central Okanagan Regional District and using bicycle lane kilometers and total lane kilometers as lead exposure variables. In this study, the bicycle lane kilometers variable t-statistic did not meet the 95% level significance test and had to be removed.

The model results showed that an increase of bicycle-vehicle collisions was associated with an increase in total lane kilometers, bicycle lane kilometers, bus stops, traffic signals, intersection density and percentage of arterial-local intersections. Data issues in the study included the unavailability of EMME/2 exposure variables such as bicycle kilometers travelled, vehicle kilometers travelled and volume/congestion, and the relatively low mode share of bicycle use in the Central Okanagan Regional District. The results supported the study's hypothesis that North America's current low levels of bicycle use can expect to see an increase in bicycle collisions as bicycle mode share increases, but the authors continue to predict that as bicycle mode share increases beyond the yet unknown threshold level, a net safety improvement should be observed [62].

2.4.3 Macro-Level Bicycle Collision Prediction Models Using Other Statistical Techniques

There has been research on alternative modelling methods to the negative binomial regression model used to look at bicycle safety from the macro-level perspective. A least squares analysis was used to examine the relationship between number of pedestrians or cyclists and their collisions with motor vehicles by Jacobsen (2003). The research found the likelihood that a bicyclist will be struck by a motorist varies inversely with the amount of walking or bicycling, across communities of varying size, from specific intersections to cities and countries, and across time periods [64]. A binary logistics regression was used to model relationships between different types of collisions, including bicycle collisions, and independent variables in demographic, land use, and roadway accessibility by Kim et al. (2010). The bicycle collision model results suggested that demographic variables such as employment and income level were

significantly and positively associated with bicycle collisions and accessibility variables such as number of bus stops, bus route length and number of intersections were positively associated with bicycle collisions [65].

Further research on bicycle-vehicle collisions in Beijing from Yan et al. (2011) concentrated on the inter-relationships between irregular maneuvers, crash patterns and bicycle injury severity, using a binary logit model to preform bicyclist injury severity analysis. The study's results suggest the installation of median division between roadway and bikeway, improving lighting on road segments and reducing speeds on high bicycle volume roads would improve the safety of cyclists [66]. Research investigating the effects of spatial correlation using a Bayesian spatial framework to model pedestrian and bicycle collisions at the macro-level by Siddiqui, Abdel-Aty, & Choi (2011) recommended that spatial correlation should be considered when modelling pedestrian and cyclist collisions at the aggregate level [67]. Another study by Abdel-Aty et al. (2014) integrated trip and roadway characteristics to develop several models at the TAZ level, including bicycle-related crash models on the foundation of various estimator groups [68]. Yasmin & Eluru (2014) investigated the influence of built environment on bicycle safety at the TAZ level, using latent segmentation based count models from single state and dual state systems to formulate models for Montreal and Toronto [69]. A study by Kwon et al. (2015) developed a bicycle accident forecasting model using multiple regression analysis as well as performed a survey for the analysis of road safety issues and presented safety improvement measures [70]. Bicycle safety is increasingly becoming an important topic of interest worldwide, as more research is done on techniques to build models to predict and analyze safety impacts on cyclists.

2.5 Data Assembly

The acquisition and processing of the data required for model development is a major task for developing macro-level CPMs. Data needs at the planning level can range from socio-demographic to infrastructure to synthesized data from traffic forecast models. The quality and quantity of data is critical to the quality of the model predictions, and can vary substantially due to the large number of data sources typically required. For example, for traffic volume forecast data from regional transportation planning models (EMME/2), data quality, as a function of how road and transit network is defined, is more of an issue than data completeness [71]. Potential data extraction sources, such as regional or municipal governments and census, must be considered in order to confirm data costs, availability, quality, and predictability [38]. Previous research has identified four themes for variables, including:

- Exposure variables, consisting of data measured through GIS or modeled through EMME/2;
- Socio-demographic variables, typically consisting of data measured through census or extracted from EMME/2 model;
- Transportation demand management variables consisting of measured through census or modeled through EMME/2 data; and,
- Network variables, consisting of measured through GIS data [38] [62] [46] [20] [21].

In addition, collision variables are dependent parameters in the macro-level CPMs, and previous research has used total collisions, severe collisions or property damage only collisions as dependent variables [72]. Including new types of data sources and using up-to-date GIS data is important for the development of empirically sound and relevant models. Many previous studies, such as [46] [20] [21] [50], have used outdated data to develop macro-level CPMs [72].

With regards to acquiring bicycle exposure data, possibilities range from using annual average daily traffic (AADT) from bicycle counts [73] to using smartphones to collect bicycle data [74]. Big data from smartphones and other applications has the potential to influence future methods of collecting travel behaviour data suitable for mobility studies [75]. This thesis uses a method of quantifying the use of bicycle facilities by calculating exposure of total annual bicycle volume on a road, the AADT. Esawey, Lim, et al. (2013) expanded daily bicycle traffic volume to AADT, by developing daily adjust factors such as grouping daily factors by weekday/weekend, developing weather-specific factors and developing factors for different road classes. The COV has a rich catalogue of bicycle count data, which allowed this analysis to make use of daily bicycle volume data for the years 2010 and 2011. Due to monthly weather fluctuations that have an impact on bicycle travel, this research proposed to estimate monthly average daily bicycle traffic, as well as AADT. The study resulted in a comprehensive database of annual and monthly average daily bicycle traffic [73], which will be used further in this thesis to represent the bicycle exposure variable.

The following Table 2 represents variables used in previous macro-level CPMs development, as well as potential suggestions for future bicycle macro-level CPM development from other researchers [62] [72] and this study.

Table 2: Potential included variables for macro-level CPMs

Theme	Included Variables	Symbol	Derivation	Citations
Exposure	Vehicle Kilometres Traveled	VKT	Modeled	[20] [46] [21]
	Neighbourhood Congestion (volume/capacity)	VC	Modeled/ Measured	[20] [21]
	Total Lane Kilometres	TLKM	Modeled/ Measured	[20] [46] [62] [21]
	Total Neighbourhood Area	AREA	Measured	[21] [20] [62]
	Total Bicycle Lane Kilometres	TBLKM	Measured	[62]
	Total Transit and Vehicle Kilometres Travelled	TTVKT	Modelled	[46]
	Total Transit Kilometres Travelled	TKT	Modeled	[46]
	Average Posted Speed	ASP	Measured	[21]
	Bicycle Kilometres Traveled	BKT	Modeled	Suggested
	Average Neighbourhood Congestion	AVC	Measured	Suggested
Collisions	Total Collisions over X years	TX	Measured	[20] [21] [46]
	Severe Collisions (fatal and injury) over X years	SX	Measured	[46] [21] [20]
	Property Damage Collisions over X years	PDOX	Measured	[46]
	Total Bicycle/Vehicle Collisions	TBX	Measured	[62]
	Severe Bicycle/Vehicle Collisions	SBX	Measured	Suggested
	Bicycle Collision Frequency	BFX	Measured	Suggested
	Rush Hour Collisions	TRX	Measured	Suggested
Socio-demographic	Urban Zones	URB	Measured	[62]
	Rural Zones	RUR	Measured	[62]
	Population	POP	Measured	[62] [46] [21]
	Population Density	POPD	Measured	[62] [20] [21] [46]
	Population aged <30	POP30	Measured	[62]
	Male/Female Ratio	M/F	Measured	[62]
	Residents Working	WKG	Measured	[21]
	Jobs per Unit Area	WKGD	Measured	[20] [21]
	Home	NH	Measured	[62]
	Housing Units per Unit Area	NHD	Measured	[62] [20] [21] [21]
	Average Family Size	FS	Measured	[20] [21]
	Employment Level	EMP	Measured	[62] [20] [46]
	Employment Percentage	EMPP	Measured	[62] [20]
	Employment Density	EMPD	Measured	[62] [46]
	Participation in Labour Force	PARTP	Measured	[62]
	Unemployed Residents	UNEMP	Measured	[62] [21]
	Unemployed Rate	UNEMPR	Measured	[62]

Theme	Included Variables	Symbol	Derivation	Citations
Transportation Demand Management	Average Income	INCA	Measured	[62]
	Median Household Income	MEDI	Measured	Suggested
	Parking Costs	PC	Measured	Suggested
	Higher Education	HE	Measured	Suggested
	Land Use	LU	Measured	Suggested
	Percentage of Commercial Land Use	COMM	Measured	Suggested
	Percentage or Residential Land Use	RES	Measured	Suggested
	Percentage of High Density	HDLUP	Measured	Suggested
	Work-Home Trip Distance	WHTD	Measured	Suggested
Network	Core Area	CORE	Measured	[20] [21] [62]
	Core Percentage	CRP	Measured	[20] [21] [62]
	Total Commuters	TCM	Measured	[20] [21] [62] [46]
	Commuter Density	TCD	Measured	[21]
	Percentage of Transit Commuters	BUSP	Measured	[62] [46]
	Percentage of Bicycle Commuters	BIKEP	Measured	[62]
	Percentage of Pedestrian Commuters	PEDP	Measured	[62]
	Number of Drivers Commuting from Each Zone	DRIVE	Measured	[21]
	Percentage of Driving Commuters	DRIVEP	Measured	[20] [62]
	Percentage of Car Passenger Commuters	PASSP	Measured	[62]
	Shortcut Capacity Through Local Roads	SCC	Measured	[20] [21]
	Shortcut Capacity x Neighbourhood Congestion	SCVC	Measured	[20] [21]
Policy	Number of Signals	SIG	Measured	[21] [62]
	Signal Density	SIGD	Measured	[20] [21] [62]
	Number of Intersections	INT	Measured	[21] [62]
	Intersection Density	INTD	Measured	[20] [21] [62]
	Number of Intersection per Total Lane Km	INTKD	Measured	[21] [62]
	Percentage of Three-Way Intersections	I3WP	Measured	[20] [21] [62]
	Number of Arterial-Local Road Intersections	ALKM	Measured	[62]
	Number of Arterial Lane Kilometres	CLKM	Measured	[62]
	Number of Local Lane Kilometres	LLKM	Measured	[62]
	Percentage of Arterial-Local Road Intersections	IALP	Measured	[20] [21] [62]
	Percentage of Arterial Lane Kilometres	ALKP	Measured	[20] [21] [62]
	Percentage of Local Lane Kilometres	LLKP	Measured	[21] [62]
	Bus Only HOV Lanes	BUSHOV	Measured	[46]
	2+ HOV Lanes	TWOHOV	Measured	[46]

Theme	Included Variables	Symbol	Derivation	Citations
3+ HOV Lanes	THRHOV	Measured	[46]	
Bus Stops	BS	Measured	[62] [46]	
Bus Stop Density	BSD	Measured	[62] [46]	
Near Sided Stops	NS	Measured	[46]	
Far Sided Stops	FS	Measured	[46]	
Midblock Stops	MB	Measured	[46]	
Stops in a Parking Lane	PL	Measured	[46]	
Stops in a Through Lane	TL	Measured	[46]	
Stops in a Bus Bay	BB	Measured	[46]	
Stops in a Paved Shoulder	PS	Measured	[46]	
Stops in a Gravel Shoulder	GS	Measured	[46]	
Number of Routes	NR	Measured	[46]	
Total Route Length	L	Measured	[46]	
Net Route Length	NL	Measured	[46]	
Sum of Route Frequencies in Peak Period	FAM	Measured	[46]	
Number of Traffic Circle Intersections	INTTC	Measured	Suggested	
Percentage of Traffic Circle Intersections	INTTCP	Measured	Suggested	
Number of Alleyways	AW	Measured	Suggested	
Percentage of Alleyways	AWP	Measured	Suggested	
Number of Sidewalks	WALK	Measured	Suggested	
Sidewalk Density	WALKD	Measured	Suggested	
Percentage Topography Over Specific Grade	TOP	Measured	Suggested	
Number of On-street Parking	PARK	Measured	Suggested	
Percentage of On-street Parking	PARKD	Measured	Suggested	
Number of Bike Racks	BIKER	Measured	Suggested	
Percentage of Bike Racks	BIKERD	Measured	Suggested	
Number of Stop Signs	SSD	Measured	Suggested	
Percentage of Stop Signs	SSP	Measured	Suggested	
Average Annual Precipitation	PERP	Measured	Suggested	
Average Annual Temperature	TEMP	Measured	Suggested	

2.6 Model Development

Macro-level CPMs are valuable tools commonly used for road safety evaluations, detection and ranking of accident-prone locations and safety planning for neighbourhoods, cities, or regions. Because collisions are discrete, non-negative and rare events it is difficult to develop

suitable statistical models to explain them. Previously, models were developed using a linear relationship before further research by Miaou & Lum (1993) showed that the basic assumptions for linear regression (normal error structure, constant error variance and linearity) did not apply in the case of collision datasets [76].

Generalized linear regression (GLM) assuming a non-normal error structure distribution, usually a Poisson or a Negative Binomial (NB) error structure, has become the norm in recent literature, overcoming the limitations of linear regression models and producing better fit to the observed collision data [38]. Work to date by researchers developing and applying macro-level CPMs have used computer applications of GIS along with EMME/2 and traditional database and spreadsheet applications such that the datasets can be assembled for the creation of macro-level CPMs [71] [20] [21] [62].

The GLM approach is based on work by previous researchers, including Sayed & Rodriguez (1999). The assumption is Y is a random variable describing the number of collisions at a specific time period. Previous research has found the point probability function of Y is given by the NB distribution with an expected mean and variance of the of:

$$E(Y) = \mu; Var(Y) = \mu + \frac{\mu^2}{\kappa} \quad (2.10)$$

In contrast, the Poisson error structure has an equal mean and variance, giving it the advantage of being simpler. However this advantage can also be seen as a limitation. Since collision data has often been found to be overdispersed (with the variance being greater than the mean), the NB distribution is typically a more realistic assumption than the Poission distribution [25].

2.6.1 Model Form

Previous research [77] has noted that the model form should satisfy two conditions:

1. The model must produce logical results, for example at zero exposure collisions must equal to zero; and,
2. In order to use GLM, there must be a recognized link function that can linearize the model form to estimate coefficients during the GLM process [38] [46].

The model can be expressed mathematically based on the following equation form, from previous research [77] [21] [20]:

$$E(\Lambda) = a_0 Z^{a_1} e^{\sum b_i X_i} \quad (2.11)$$

Where:

$E(\Lambda)$: predicted collision frequency

a_0, a_1, b_i : model parameters

Z : external exposure variable

X_i : explanatory variables

2.6.2 Error Structure

The GLM approach for developing the models assumes an error structure that is best described by Poisson or NB distributions. To determine the adequate error structure, Poisson is usually assumed initially and the parameters of the distribution are estimated, with the dispersion parameter σ_d calculated as:

$$\sigma_d = \frac{\text{Pearson } X^2}{n - p} \quad (2.12)$$

Where n is the number of observations and p is the number of parameters in the model. The *Pearson X²* is expressed mathematically as:

$$\text{Pearson } X^2 = \sum_{i=1}^n \frac{(y_i - \bar{E}(\Lambda_i))^2}{Var(Y_i)} \quad (2.13)$$

Where:

y_i : observed mean collision frequency at location i over a specific time period

$Var(Y_i)$: variance of the observed mean collision frequency at location i

$\bar{E}(\Lambda_i)$: expected mean collision frequency for location i as obtained by the CPM

If the value of the dispersion parameter σ_d is equal to 1.0 or lower, the Poisson assumption is valid, and if it is greater than 1.0 than a NB error structure would provide a better fit, as the data has a greater dispersion than could be explained by using the Poisson distribution [77] [38] [46].

2.6.3 Selection of Explanatory Variables

While many factors can influence collision occurrence, not all of them are appropriate as explanatory variables in a CPM. Explanatory variables in the same model must be independent (not correlated). The recommended method by Sawalha & Sayed (2006) for adding independent variables is a forward stepwise procedure:

1. Variables are added one by one and their significance is tested, beginning with variables representing exposure;
2. Three tests must be performed to test the significance of each particular variable:
 - T-stat of the variable must be significant at the 95% confidence level (t-stat higher than 1.96)
 - The sign (i.e. +/-) of the variable must be logical

- The addition of a variable should cause a significant drop in the value of the scaled deviance (SD), at a decrease greater than $X^2_{0.05,1} = 3.84$ at a 95% confidence level [39]

Once the variable meets the criteria, it can remain in the model. The next variable is then added to the model and tested for significance, until there are no more variables left to evaluate [47].

2.6.4 Evaluation of Goodness of Fit

The goodness of fit measures how well the model predicts or fits the observed collision data. The typical quantitative methods of assessing goodness of fit are: the *Pearson X²* (described above), the scaled deviance (SD) and the shape parameter κ , and follows methodology set out by [77] [39]. If the error structure is Poisson distributed, then the SD is calculated as:

$$SD = 2 \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\bar{E}(Y_i)} \right) \quad (2.14)$$

If the error structure is NB distributed, the SD is calculated as:

$$SD = 2 \sum_{i=1}^n \left[y_i \ln \left(\frac{y_i}{\bar{E}(Y_i)} \right) - (y_i + \kappa) \ln \left(\frac{y_i + \kappa}{\bar{E}(Y_i + \kappa)} \right) \right] \quad (2.15)$$

A model with a good fit will have the SD and the *Pearson X²* less than the X^2 distribution value with $(n - p - 1)$ degrees of freedom at a 95% confidence level [38]. For the shape parameter κ , there is no minimum value recommended, but previous research of CPMs has found κ to be usually higher than 1.0 [78].

2.6.5 Outlier Analysis

When developing macro-level CPMs, the models may sometimes not meet one or more of the goodness of fit criteria. This could be due to extreme values, called outliers, which exist because of errors in the data collections or because data points can be atypical. The outlier analysis is a procedure to remove the outliers from the dataset to improve the fit of the models. The methodology is based on the Cook's Distance (CD) as described by [79] [39], in which the higher value for a given observation CD_i , the stronger the influence on the model.

$$CD_i = \frac{h_i}{p(1 - h_i)} (r_i^{PS'})^2 \quad (2.16)$$

Where:

h_i : leverage value

$r_i^{PS'}$: standardized residual of point i , calculated as

$$r_i^{PS'} = \frac{y_i - \hat{y}_i}{\sqrt{(1 - h_i)Var(y_i)}} = \frac{PR_i}{\sqrt{1 - h_i}}$$

PR_i : the Pearson Residual

p : the number of parameters

The Pearson Residual (PR_i) reflects how well or poorly the model fits the i^{th} observation, and h_i reflects how far the data is from the rest of the data points. The methodology for removing outliers involves:

1. Calculating CD for all points and sort the data in descending order according to the CD values;
2. Removing the point with the highest CD value;

3. Recalculating the CPM model but with the κ value from the previous model;
4. If the SD change is at least $X_{0.05,1}^2 = 3.84$ at a 95% confidence level then the model is re-estimated producing new parameters, including new κ and new CD_i for all remaining points; and,
5. Repetition of the procedure until the change in SD becomes less than $X_{0.05,1}^2 = 3.84$ at a 95% confidence level [38].

When the last outlier is removed from the dataset, the model parameters, including κ , are re-estimated for the final time [47].

Chapter 3: Methodology

This chapter describes the methodology used in this research. There are three main components: data extraction, macro-level CPM development, and application to macro-reactive black spot programs. Figure 1 below provides an outline of this methodology.

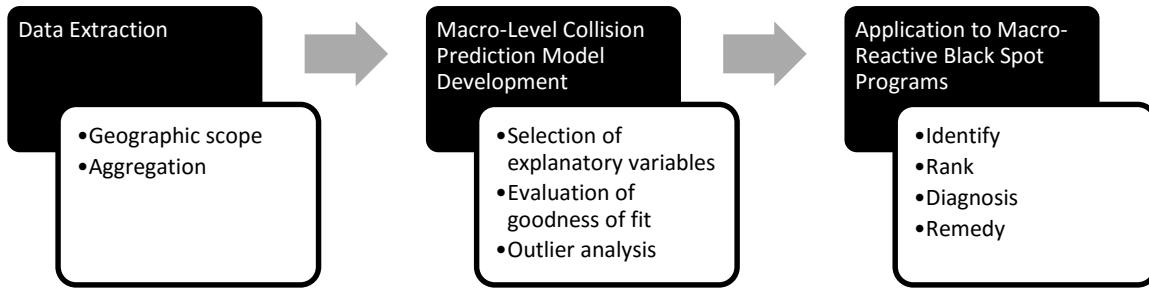


Figure 1: Methodology outline

3.1 Data

As described in Chapter 2, the quality and quantity of data is critical to developing accurate and reliable statistical models. Previous recommendations as described in the literature on the effective data extraction process have been followed to achieve best results for this thesis. The following sections describe the geographic scope of the data, the aggregation approach, data sources, quality issues and the final model development process.

3.1.1 Geographic Scope

The data extracted for model development is all within the geographic area of the City of Vancouver (COV), within the Province of British Columbia, Canada. The land area of the city is about 115 square kilometers. In 2011, the year of this study's focus, the COV population totaled about 600,000 residents, dwelling in almost 265,000 households. Land area in the COV is dense

and urban, with an average population density of 5,249 persons per square kilometer [80]. The COV's cycling network in 2011 was composed of about 240 kilometers of separated bike lanes, painted bike lanes, local street bikeways and shared use lanes; as shown below in Figure 2 [81]. In 2011, cycling accounted for approximately 4.4% of trips to work in the COV [24].

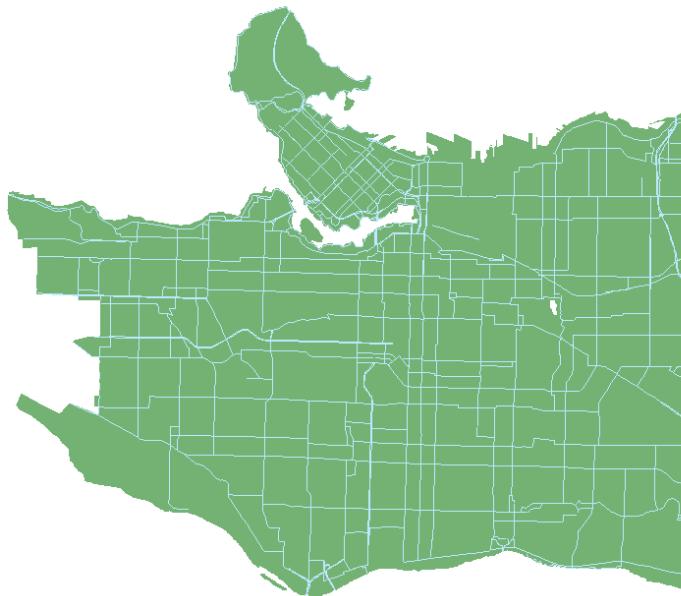


Figure 2: City of Vancouver bicycle routes

3.1.2 Aggregation

The aggregation units used for this thesis were based on 134 traffic analysis zones (TAZs) used in TransLink's EMME/2 transportation planning model. This macro-level aggregation was chosen after the consideration of the research objective, geographic scope, data availability and computational limits. The TAZs have been developed at the ideal size to keep population and employment densities at uniform levels and ensure adequate data in each zone for guaranteed goodness of fit. This level of aggregation was chosen because the TAZ boundaries

overlap with both census tracks and municipal boundaries, allowing for easy data integration of current and future demographic data and transportation demand [21]. The data was aggregated to TAZs to allow for the development of macro-level CPMs, which are traditionally performed at the zonal level. Using zones allows for the comparison of bicycle network and safety indicators between the various TAZs to identify CPZs. Figure 3 shows the level of aggregation, representing the city's 134 TAZs.

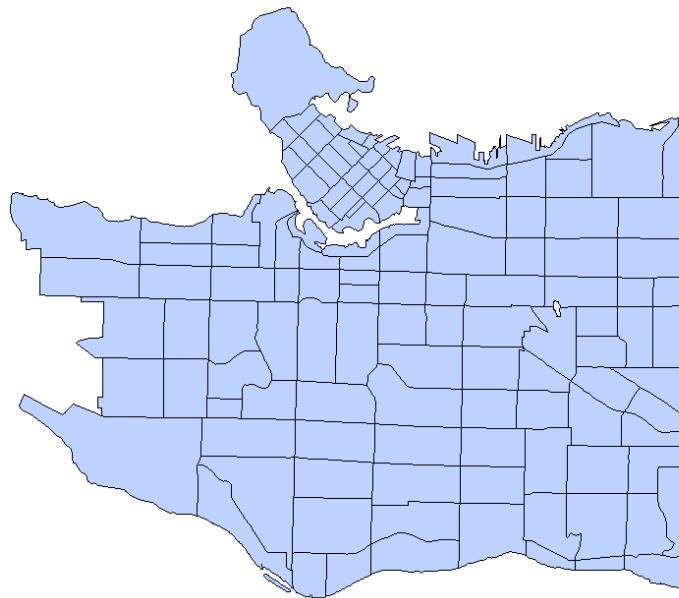


Figure 3: City of Vancouver Traffic Analysis Zones

3.1.3 Sources

The list of variables used in this thesis, along with summary statistics is presented in Table 3.1. The data was extracted and assembled from the following sources:

1. The Metro Vancouver regional transportation authority, TransLink, provided geocoded files of TAZs, as well as road, transit and bicycle networks from the year 2013. In

addition, TransLink provided EMME/2 transportation model outputs consisting of travel demand (vehicle and transit kilometers travelled, average zonal speed, average, zonal congestion and transportation mode split), as well as socio-demographic and land use data for the base year 2011.

2. City of Vancouver (COV) provided exposure and infrastructure variables such as number of signals from their most recent Open Data catalogue, accessed in late 2015. The exposure variable total “AAA” bicycle kilometers refers to the total kilometers of bicycle network that is classified as comfortable and safe for users of All Ages and Abilities (AAA). This classification is typically found for bicycle facilities that are classified as separated bike lanes or off-street paths.
3. The Digital Road Atlas (DRA) was used for intersection data for the base year 2011. The DRA provides a singular and authoritative source of road data for the province of British Columbia [82]. The following Figure 4 shows the large number of intersections in the COV accessed from the DRA.

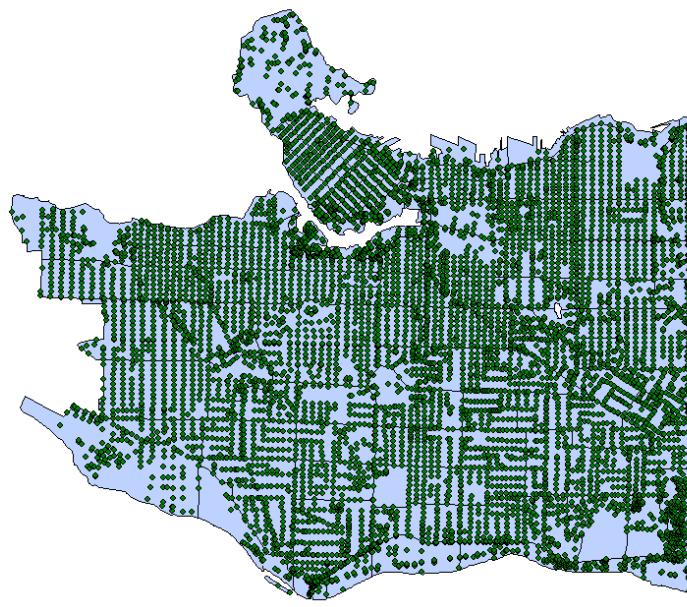


Figure 4: City of Vancouver intersections

4. AADT bicycle exposure data was acquired from the COV's bicycle count volume data for the years 2010 and 2011, which was a previously developed comprehensive database of bicycle volume data obtained from the expansion of temporary and permanent bicycle counts, from former research [73]. The bicycle network routes for which bicycle volume was collected are shown in the following Figure 5.

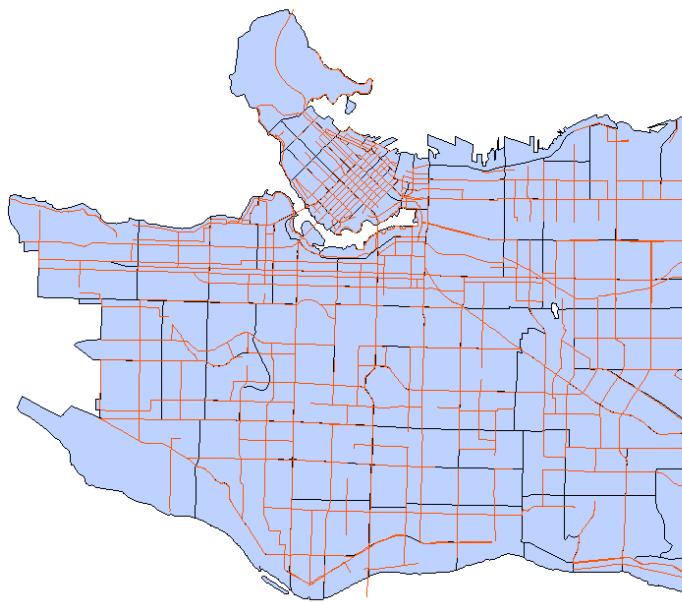


Figure 5: Annual Average Daily Traffic (AADT) in the City of Vancouver

5. The Insurance Corporation of British Columbia (ICBC), the province's public automobile insurance company delivered geocoded files of bicycle collision claims in the COV for the years of 2009, 2010 and 2011 shown in the following Figure 6. Three years of collision data was used to decrease the randomness bias and quantify the relatively uncommon bicycle collision data [21]. The availability of geocoded insurance claim data centralized from ICBC was considered a great advantage to overcoming any potential incomplete and unreported collision data problems [38].

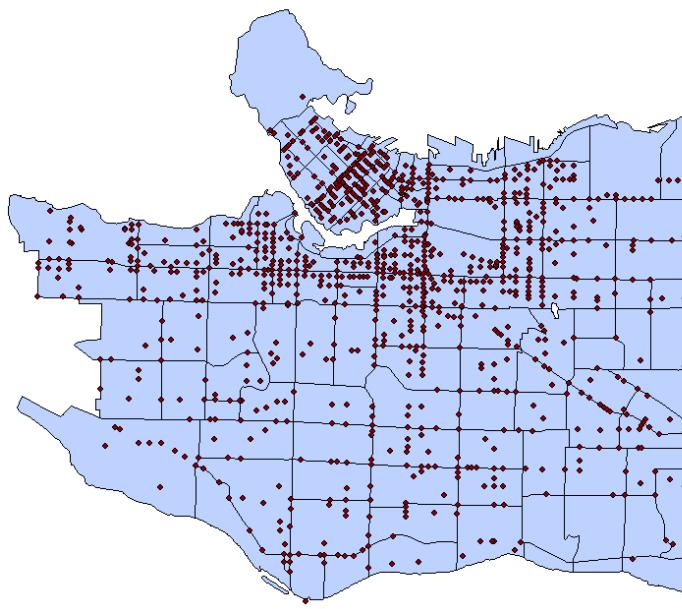


Figure 6: ICBC data on bicycle-vehicle collisions from 2009-2011 for the City of Vancouver

Potential explanatory variables were then grouped into themes of:

- Exposure variables related to bicycle and vehicle kilometers travelled, and therefore collision probability;
- Network variables related to the transportation road network; and,
- Socio-economic and commute preference variables.

Table 3 presents a list of explanatory variables and their definitions and summary statistics.

Table 3: Explanatory variables and summary statistics

Included Variables	Variable Symbol	Derivation	Units	Year	Source	Vancouver Total	Zonal	
							Average	Standard Deviation
Exposure								
Bicycle Kilometers Traveled	BKT	Modeled	km	2011	COV	140401.15	1047.77	2109.96
Total "AAA" Bicycle Kilometers	AAAKM	Measured	km	2015	COV	78.64	0.59	1.26
Collisions								
Total Bicycle Collisions over 3 years	TB3	Measured	cols.	2010-2013	ICBC	1704	12.72	13.49
Socio-demographic and Commute								
Population	POP	Modeled	pop	2011	Trans-Link	626867	4678.11	2648.33
Total Employment	EMP	Modeled	jobs	2011	Trans-Link	406185	3031.23	3337.34
Household Density	HHD	Modeled	hh/Ha	2011	Trans-Link	275899	2058.95	1225.99
Commercial Land Use	COM	Measured	m ²	2015	COV	4896105.71	36538.1	45152.6
Total Commuters	TCM	Modeled	comm.	2011	Trans-Link	123512.81	921.74	413.46
Network								
Signal Density	SIGD	Measured	sig/Ha	2015	COV	19.12	0.143	0.185
Intersection Density	INTD	Measured	int/Ha	2011	DRA	99.648	0.744	0.337
Bus Stops	BS	Measured	stops	2015	Trans-Link	1893	14.127	8.159
Percentage of Residential Street Kilometers	RPKM	Measured	%	2015	COV	-	56.444	30.635
Total Arterial Street Kilometers	AKM	Measured	km	2015	COV	285.294	2.129	1.395
Percentage of Park Area	PMP	Measured	%	2015	COV	-	7.695	12.75

3.1.4 Quality Issues

A few data issues were faced when extracting and preparing the data for model

development. These issues include the following.

3.1.4.1 Bicycle Collision Data

Despite the severity and societal burden of bicycle collisions, reported bicycle-vehicle collisions are rare events. During the three-year time period data was collected, approximately only 1.0% of all reported motor-vehicle collisions to ICBC involved cyclists [23], since incidents such as non-injurious or non-property damaging collisions may have gone unreported. In addition, the collision reporting process is typically based on the subjective observations by witnesses and police reports, which is self-reported by those involved [17]. The collision records therefore may be incomplete and lack accurate information, causing errors regarding the spatial location of the collision and their aggregation into TAZs. ICBC has attempted to solve this issue by geocoding collision claims as either mid-block or intersection on road centrelines, decreasing precision and addressing potential error by a “split the difference” geographic assumption [38].

3.1.4.2 Boundary Effects

When analyzing urban form, using aggregation to the traffic analysis zone can be problematic because boundaries tend to be on major transportation corridors and intersections. Due to the aggregation of the data into TAZs, assumptions had to be made concerning the influence of the aggregation grid choice on the location of collisions on a zone boundary. It has been observed in previous research that collisions located near zone boundaries may have influence in multiple zones [83]. In addition, the geographic scale and size of the TAZs can impact data homogeneity [38]. This thesis assumed that collision data geo-coded near zone boundaries has an influence on adjacent zones and proceeded with a method to assure that collisions are representative for each TAZ. The methodology to deal with the collision boundary

effects followed work by previous researchers to use exposure ratio methods to aggregate boundary date into adjacent zones [84]. Collisions were aggregated to each zone according to the BKT (leading exposure variable) ratio between adjacent zones.

3.1.4.3 Multiple Sources

This thesis integrated geocoded data from several years because the data was acquired from multiple sources. The highest quality vehicle exposure (VKT) and bicycle exposure (AADT) data available was for 2011, which was therefore determined as the base analysis year for this thesis. This was supported by the ICBC bicycle collision data existing for the years 2009, 2010 and 2011. Due to the unavailability of historical data, TransLink provided cycling infrastructure data for the year 2013. The most recent COV Open Data catalogue provided other road infrastructure data for the year 2015. The time discrepancy between 2011 to 2015 for vehicle infrastructure variables were assumed to be irrelevant for this thesis because the COV vehicle infrastructure has seen few changes in the past three years. In terms of bicycle infrastructure, the changes between 2011 and 2013 may cause a few inconsistencies between data sources. However, these bicycle infrastructure changes have been relatively few, and this thesis assumes that these differences due to several years of data can be considered negligible.

3.2 Model Development

After candidate variables have been chosen, and data has been extracted and aggregated to TAZs, the data preparation is then complete and ready for model development. The model development methodology followed the following procedures described below in the selection of explanatory variables and GLM regression.

3.2.1 Selection of Explanatory Variables

Before developing the models, the appropriate explanatory variables need to be chosen in a stepwise procedure, as described in research by Sawalha & Sayed (2006). In this technique each variable is added one by one, testing the change in goodness of fit as the model is constructed. As recommended in previous research, the first variable tested in each model was the exposure variable, because of its leading prediction influence since no collisions can occur without exposure [39]. In this thesis, the exposure variables used was bicycle kilometers travelled (BKT), for which the volume was represented by the bicycle Annual Average Daily Traffic (AADT). This basic model acts as a reference model, serving as a base for generating a new model containing additional variables. The next stage of the process involved developing a new reference model that contains the exposure variable and an additional explanatory variable that causes the maximum drop in scaled deviance, with remaining variables added step by step following the same method [39].

To choose additional candidate variables to build multiple CPMs, the reference (exposure only) CPM was used to identify Collision Prone Zones. CPZs are defined as zones that show a higher potential for collisions compared to a specified norm. Due to the randomness and small size of the bicycle collision dataset, statistical techniques accounting for randomness need to be applied to establish CPZs. Similar to previous techniques, an Empirical Bayes (EB) methodology using CPMs is used to identify the CPZs [6] [38] [40]. The EB refinement method can be used to identify CPZs using the process as described in Section 2.2. In addition to local zone collision history, the zonal $E(\Lambda)$ and $\text{Var}[E(\Lambda)]$ are calculated to provide a location-specific prior distribution of collisions. With this information the zonal EB safety estimate is calculated. The

additional candidate variables to be added to the reference model were systematically added from the list in Table 3 under Section 3.1.3. Variables were chosen from data categories including socio-demographic data, exposure data, TDM data and network infrastructure data. Once chosen, the decision to retain a variable in the model was based on the criteria as described in Section 2.6.3.

3.2.2 Addition of Variables to Bicycle Collision Prediction Models

The model form for the development of the Collision Prediction Models (CPMs) was chosen according to available data and previous research as described in Section 2.6. The model form that was used for the developed of CPMs was:

$$E(\Lambda) = a_o Z^{a_1} e^{\sum b_i X_i} \quad (3.1)$$

Where:

$E(\Lambda)$: predicted collision frequency

a_o, a_1, b_i : model parameters

Z : external exposure variable

X_i : explanatory variables

A log-linear transformation was then completed by using a logarithmic link function, using Generalized Linear Modelling (GLM), as shown in previous research [38], transforming Equation 3.1 into:

$$\ln[E(\Lambda)] = \ln(a_o) + a_1 \ln(Z) + \sum_{i=1}^n b_i X_i \quad (3.2)$$

The CPMs developed were based on the City of Vancouver's (COV) 134 TAZs, using the following data:

- $E(\Lambda)$ was represented by the number of bicycle collisions, $TB3$. The Insurance Corporation of British Columbia (ICBC) provided the data for the variable $TB3$ for three years (2009, 2010 and 2011) (see section 3.1.4). Automobile collisions involving bicycles were the collisions captured in this variable, as ICBC only reports collisions involving a motor vehicle. It would have been interesting to have collisions involving bicycles and other road users or collisions involving only one cyclist, but this data was not available because it is not reported to ICBC. As well, the number and severity of these collisions is likely to be small.
- The exposure variable Z , in this research, refers to bicycle kilometers travelled BKT . The exposure variables were obtained from COV AADT bicycle repository for the years 2010 and 2011 and the TransLink EMME/2 transportation model for the base year 2011 (see section 3.1.4). This variable is critical to the model, since at the level of zero exposure collisions must also remain zero.
- The explanatory variables X_i , consist of multiple bicycle-related indicators that are considered likely to influence collision occurrence. The list of potential explanatory variables tested in the CPMs were grouped into two categories:
 - 1) Socio-demographic: refers to population information, commuter characteristics and land use; and,
 - 2) Network: refers to citywide physical infrastructure characteristics.

Summary statistics of the explanatory variable data can be found in Table 3, Section 3.1.3.

The CPMs were developed following the GLM approach as suggested in previous research [77] [21] [20] [38] [47], and as described in Section 2.6 under Literature Review. The methodology considered the following:

- Assumed a Negative Binomial error structure.
- Procedure for selection of model variables was stepwise, developing models by adding one variable at a time and testing the goodness of fit for each added variable [39].
- Variables chosen were kept in the model based on three conditions:
 - i) The parameter t-statistic is significant ($t > 1.96$ at a confidence level of 95%);
 - ii) The addition of a new variable resulted in a drop of SD for the 95% confidence level (> 3.84); and,
 - iii) The variable showed a low correlation with either independent variables in the model.
- Once variables have been chosen, the model fit is assessed using the *Pearson X²* and the scaled deviance *SD*, described further in Section 2.5.
- Finally, model fit is improved by performing an outlier analysis based on Cook's Distance *CD* method, described further in Section 2.5.

The final CPMs were developed for the COV's 134 TAZs, for exposure, socio-demographic and network groups of indicators.

3.2.3 Application of CPMs to Macro-Reactive Black Spot Programs

Macro-reactive black spot analysis uses the individual TAZ as a unit of analysis instead of an intersection or road segment for micro-reactive analyses. Macro-reactive guidelines generally follow the traditional reactive methods, however there are some differences in methodology as indicated in Figure 7 by Lovegrove & Sayed (2007) and Lovegrove (2007) below [38] [40]:

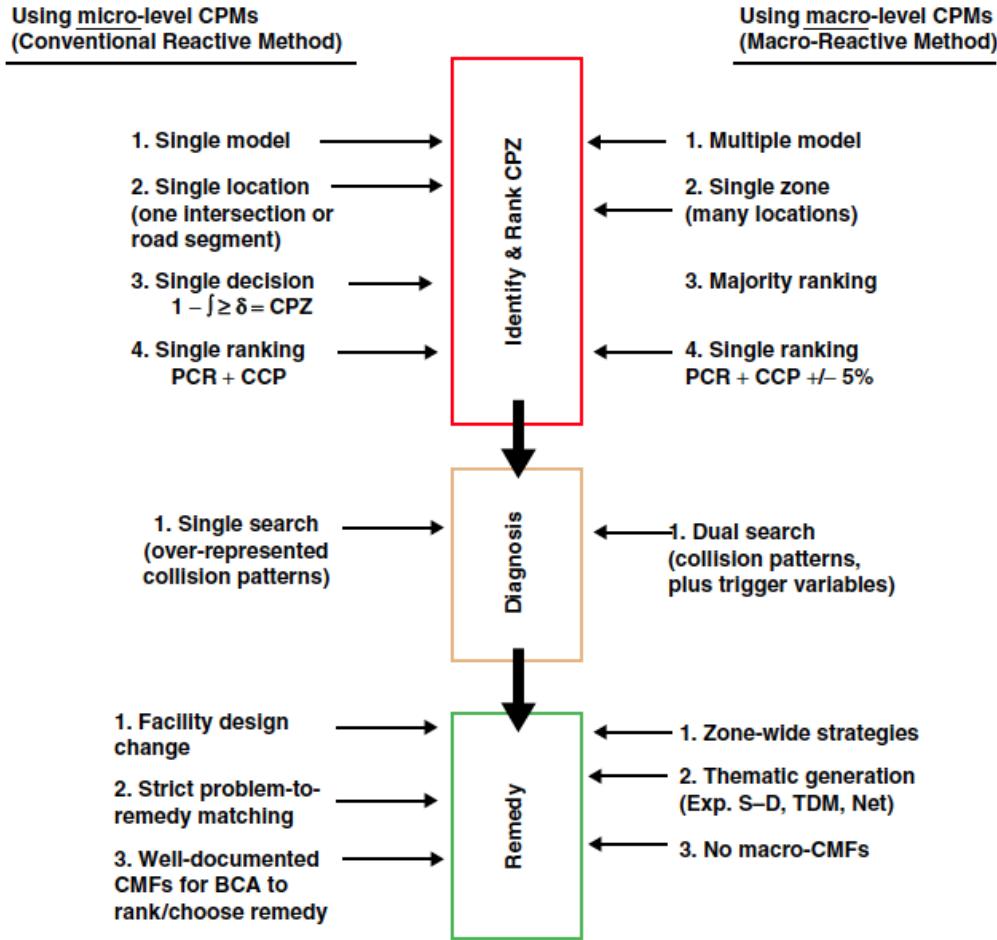


Figure 7: Comparison between micro- and macro-reactive Black Spot Programs [38] [40]

The CPMs were developed in this study with variables chosen to meet the needs of the macro-reactive black spot safety evaluation. The scope of this citywide bicycle safety study was to identify Collision Prone Zones (CPZs) and variables that may be hindering bicycle safety.

3.2.3.1 Identification and Ranking

The enhanced EB method using CPMs to identify black spots is based on research set out in Sawalha & Sayed (1999) [77], generally following the EB method described in Higle & Witkowski (1988) [85] with modifications to use CPMs and is described in detail in Section 2.2.

To identify and rank black spots with macro-level CPMs, there must be some adjustments made to the conventional reactive method.

1. The observed local collision history (count) is based on the zonal aggregate, providing the first observational clue on safety.
2. The zonal $E(\Lambda)$ and $Var E(\Lambda)$ are calculated to provide the location-specific prior clue to calculate the zonal EB safety estimate as described in Section 2.2.
3. Due to multiple CPMs, there will be several EB safety estimates for each zone resulting in a majority rule when determining a CPZ.

Finally, the calculated zonal $E(\Lambda)$ and EB safety estimates would differ for each zone resulting in differences in zonal rankings. This can be resolved by using a modified ranking approach, by summing each zones PCR rankings across all macro-level CPMs to develop a total ranking score for each CPZ, as described in Section 2.2. This score will denote which CPZs are most frequently ranked as the least safe and are in need for a diagnosis [40] [38].

3.2.3.2 Diagnosis

Once the CPZs have been identified and ranked for treatment, the diagnosis stage to find the cause of the safety problem is begun. Safety issues for CPZs can be diagnosed using a methodology similar to the conventional approach. As with this approach, diagnosis begins by first looking at an overrepresentation of collision patterns: clusters of particular collision types. Using macro-level CPMs method to determine CPZs, an additional indicator can be used: trigger variables from each model that are hypothesized to contribute to the identification of the zone as a CPZ. To identify trigger variables, the value of each variable in the top ranked CPZs is compared with the value of regional averages to understand which variables are triggering CPZ

identification. A regional average is the mean of the specific variable value for all zones in the study area. The regional COV statistics (average and standard deviation) for variables used in this study are shown in Table 3. The variable values that are found to be significantly different than the regional statistics are identified as trigger variables. This indicator can be used together with observations of collision patterns and site visits to understand the overall safety issues in each CPZ [40]. The identification of the zonal safety problem is an important step to realize potential suitable remedies.

Table 4: Regional statistics in the City of Vancouver for included variables

Included Variables	Variable Symbol	Zonal Regional Statistics	
		Average	Standard Deviation
Exposure			
Bicycle Kilometers Traveled	BKT	1047.77	2109.96
Total "AAA" Bicycle Kilometers	AAAKM	0.59	1.26
Collisions			
Total Bicycle Collisions over 3 years	TB3	12.72	13.49
Socio-demographic and Commute			
Population	POP	4678.11	2648.33
Total Employment	EMP	3031.23	3337.34
Household Density	HHD	2058.95	1225.99
Commercial Land Use	COM	36538.10	45152.60
Total Commuters	TCM	921.74	413.46
Network			
Signal Density	SIGD	0.143	0.185
Intersection Density	INTD	0.744	0.337
Bus Stops	BS	14.127	8.159
Percentage of Residential Street Kilometers	RPKM	56.444	30.635
Total Arterial Street Kilometers	AKM	2.129	1.395
Percentage of Park Area	PMP	7.695	12.750

3.2.3.3 Remedy

To match the safety issue with a suitable zone-wide safety remedy, a strategic zonal safety analysis must be conducted when using macro-level CPMs. Examples of possible remedies could be generated from considering variable theme and diagnosis results, whether the variable theme is network (i.e. number of signalized intersections could be associated with increased collisions), TDM (i.e. increased vehicle commuters could be associated with increased collisions), or socio-demographic (i.e. increased populations is associated with decreased collisions) [40]. After identifying the zonal safety problem using the methodology above, zonal characteristics were analyzed at a micro scale to identify possible remedies. This included factors such as the number of collisions, the quality of the bicycle infrastructure, the topography and the amount of arterials in each zone, for example. Finally, a list of potential bicycle safety countermeasures were analyzed and applied to each zone using engineering judgement.

Chapter 4: Results and Discussion

4.1 Approach and Results

4.1.1 Development and Selection of Models

There were 10 macro-level CPMs developed for the purpose of conducting a black spot case study for the City of Vancouver. The intention was to follow the methodology as described in Sections 2 and 3 to identify and then rank CPZs, diagnose safety issues and recommend remedies for CPZs in the COV. The collision prediction models developed used exposure, socio-demographic, TDM and network variables as indicators, along with Bicycle Kilometers Travelled (BKT) as the leading exposure variable. Table 5 presents the models that predict total bicycle collisions and their goodness of fit summary statistics. Most models presented showed explanatory variables as significant at a 95% confidence level, except for the explanatory variables of total “AAA” bicycle kilometers (AAAKM), total commuters (TCM) and percentage of residential street kilometers (RPKM), which were significant at a 90% confidence level. As expected, the models showed that increased collisions were positively associated with increased exposure variable BKT. This confirms the intuitive expectation that more bicycle exposure contributes to bicycle-vehicle collisions. However, the exponent of BKT is less than 1.0, indicating that the rate of increase of bicycle collisions reduces as more cyclists use the network. This confirms the safety in numbers concept, which states that an increase in people cycling will result in an increase in safety.

Using BKT as the leading exposure variable, the models were developed among the four themes of exposure, TDM, socio-demographic and network. The exposure variable total All Ages and Abilities (AAA) kilometers (AAAKM) was positively associated with collisions. The TDM and socio-demographic variables total commuters (TCM), total employment (EMP),

household density (HHD), commercial land use (COM), and population (POP) were also found to be positively associated with collisions. The network variables signal density (SIGD), intersection density (INTD), bus stops (BS), percentage of residential street kilometers (RPKM), and arterial street kilometers (AKM) were all found to be positively associated with collisions, while percentage of park area (PMP) was found to be negatively associated with collisions.

Table 5: Collision Prediction Models and their goodness of fit summary statistics

Model Form	k	df	SD	Pearson x^2	x2 0.05, df	p-value
Exposure						
0.1843BKT^0.6438	2.25	131	143	125	159	BKT <.0001
0.1425BKT^0.7167*exp(-0.1213AAAKM + -0.0224PMP)	2.80	129	143	146	157	BKT <.0001; AAAKM 0.0642; PMP 0.0007
Transportation Demand Management						
0.1566BKT^0.6248*exp(0.0003TCM)	2.34	130	143	127	158	BKT <.0001; TCM 0.0672
Socio-Demographic						
0.2745BKT^0.5099*exp(0.0001COM + 0.0001EMP)	2.99	129	144	121	157	BKT <.0001; COM <0.0001; EMP 0.0005
0.1461BKT^0.5358*exp(0.0001EMP + 0.0003HHD)	3.40	129	144	126	157	BKT <.0001; EMP <0.0001; HHD <0.0001
0.0816BKT^0.6336*exp(0.0001POP + 0.6203INTD)	2.57	129	141	124	157	BKT <.0001; INTD 0.0031; POP 0.001
Network						
0.1068BKT^0.5900*exp(1.6143SIGD + 0.0409BS)	3.12	129	142	136	157	BKT <.0001 SIGD <0.0001; BS <0.0001
0.1776BKT^0.6069*exp(-0.0263PMP + 0.1851AKM)	3.24	129	143	139	157	BKT <.0001; PMP <0.0001; AKM <0.0001
0.083BKT^0.6094*exp(0.6443INTD + 0.034BS)	2.74	129	142	118	157	BKT <.0001; INTD 0.0018; BS <0.0001
0.1835BKT^0.6342*exp(-0.0040RPKM + 0.3645INTD)	2.40	129	142	121	157	BKT <.0001; RPKM 0.0841; INTD 0.0879

Macro-reactive guidelines as described in Sections 2 and 3 were followed for the black spot case study due to the citywide scope of this thesis to evaluate 134 TAZs at a planning-level safety analysis. All variables that were chosen for the models were considered potential trigger variables for bicycle collisions from the four variable themes. Since the task was in part to

identify and rank CPZs, all model groups that were found to be significant using multiple variables were considered.

4.1.2 Identification and Ranking

The CPMs were used to first estimate the expected location specific collisions for each TAZ, $E(\Lambda)$. This clue, together with the observed zonal collision count over three years resulted in 10 EB safety estimates for each zone (one from each of the 10 models). Using the methodology described in Section 2.2 for the identification and ranking of black spots, each of the 10 CPMs identified a range from 7 to 20 collision prone zones at the 99% confidence level. The $E(\Lambda)$ was used as the reference group norm for comparison to the EB safety estimate to identify the Collision Prone Zones (CPZs) that have the highest potential collision reduction (PCR). The ranking of PCR was based on the difference between expected and observed collision frequency. Using a modified ranking technique described in Section 2.2, which considered the zonal ranking of all CPMs, the zones most frequently ranked with high PCR were further analyzed for diagnosis and remedy. A list of the top 7% of all top-ranked zones is shown in Table 6. TAZ 3420 was disqualified from the analysis due to a zero value for the exposure variable, BKT resulting in the inability to predict collisions and calculate $E(\Lambda)$.

Table 6: A list of the top 7% of CPZs rankings (top ten)

RANK	EXP BKT	EXP AAAKM PMP	TDM TCM	SOCD COM EMP	SOCD EMP HHD	SOCD INTD POP	NET-WORK SIGD BS	NET-WORK PMP AKM	NET-WORK INTD BS	NET-WORK RPKM INTD
1	3160	3160	3160	3640	3160	3160	3160	3160	3160	3160
2	3640	3640	3640	3200	3640	3460	3640	3640	3640	3640
3	3460	3460	3460	3460	3590	2290	3200	3200	2290	3460
4	2290	2290	2290	3160	3470	3590	3460	3460	3460	3590
5	3590	3590	3590	3590	2280	3470	3590	3590	3590	2290
6	3470	3070	3470	2090	3000	2160	3070	3100	3070	3470
7	2160	2090	2160	2290	3330	3000	2090	3070	2090	3170
8	3170	3000	3170	3480	2290	2150	3470	3000	3470	3000
9	3010	3470	3330	3170	2160	3330	3170	3470	3100	3010
10	3490	2160	3600	3470	3510	3640	3480	3170	3170	3490

The geographic locations of the top 10 ranked CPZs are shown for each of the ten CPMs is given in Appendix A, to spatially compare the similarities and differences between the models. The top ten zones were ranked by a gradient colour scheme: red for most severe top three zones, orange for the fourth to sixth severe zones, and yellow for the seventh to tenth severe zones. An example of the reference model, which used only BKT as the exposure explanatory variable, is shown in Figure 8.

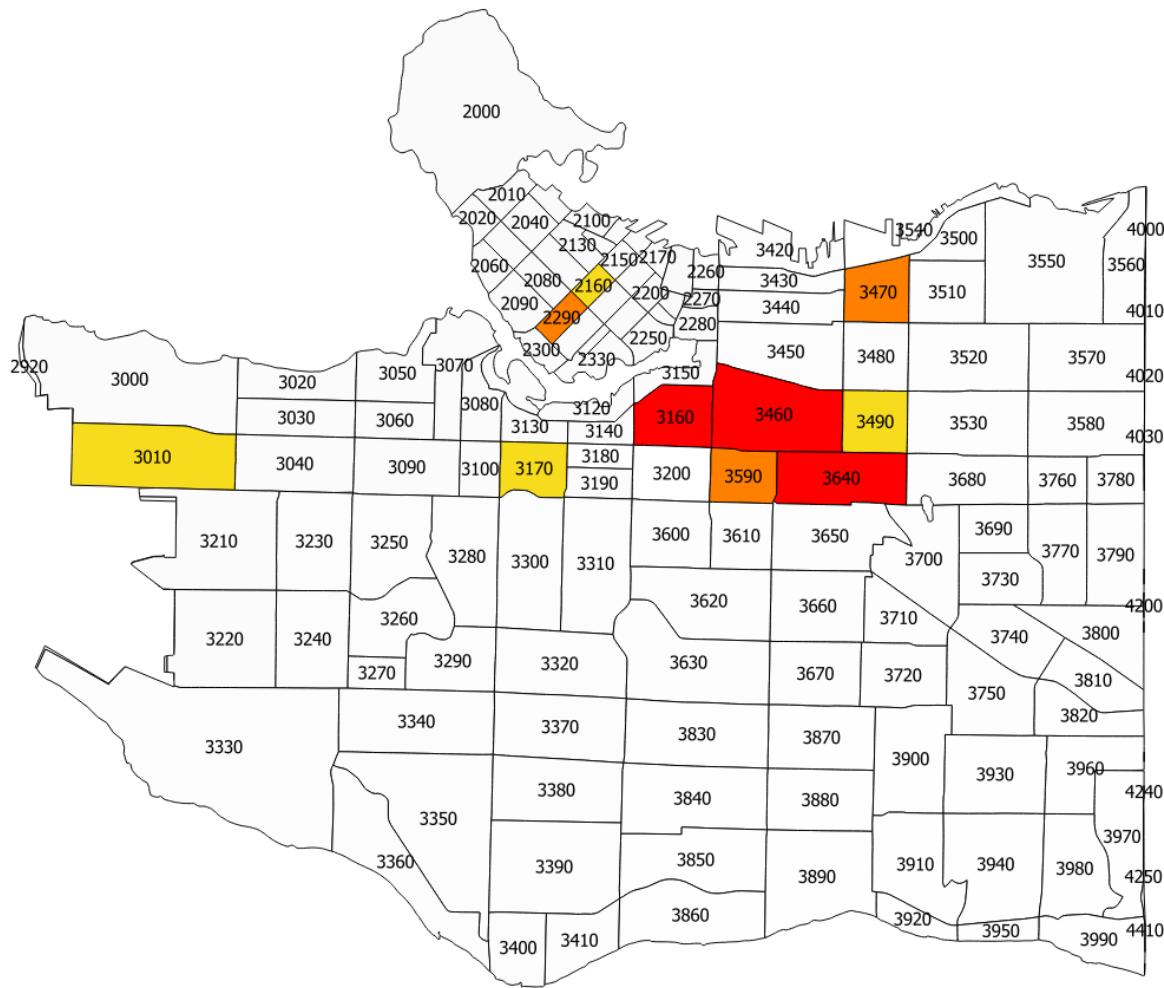


Figure 8: Top ten CPZs for the exposure model using BKT

The model groups showed consistency by providing relatively similar top CPZs identification and ranking, with a small variability. With this ranking technique, the three areas with CPZs that scored the worst were then carried forward for diagnosis to identify safety problems and potential solutions.

4.1.3 Diagnosis and Remedy

The top ten zones that were ranked collision prone were analyzed to identify the safety problems using two indicator techniques. First, average bicycle-vehicle collision frequencies for

each CPZ were compared to the regional average of 12.7 collisions over three years shown in Table 3.2. All top ten severe CPZs analyzed across the ten CPMs developed showed collisions frequencies higher than the regional average of 12.7 collisions per zone over three years. Second, the values of each CPM's top ranked CPZ variables were compared with their corresponding regional averages (also listed in Table 3) to understand which variables were triggering the collision prone ranking using a one sample student's t-test with $n - 1$ degrees of freedom.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (4.1)$$

Where:

t : t-statistic

\bar{x} : sample mean

s : sample standard deviation

μ_0 : specified value

n : sample size

These trigger variables, along with the collision frequencies identified the zonal safety issues. Across the models, the results show that collisions are associated with a large variability (defined as either low or high compared to the regional average) of the total All Ages and Abilities (AAA) kilometers (AAAKM), however the top most severe CPZs are associated with low AAAKM. Across the models, collisions are associated with a high variability of the TDM variable total commuters (TCM), as well as the socio-demographic variables of total employment (EMP), household density (HHD), percentage of commercial land use (COM), and population (POP). The most severe CPZs are found to be associated with high COM and EMP.

For infrastructure variables, collisions are associated with a large variability of intersection density (INTD), bus stops (BS), arterial street kilometers (AKM), and percentage of park area (PMP), but did not have an association with variability from the regional averages of the variables signal density (SIGD) and percentage of residential street kilometers (RPKM). The most severe CPZs found collisions to be associated with a low park area percentage, a high number of bus stops, a large amount of arterial streets and a high intersection density.

Subsequently, the top three collision prone areas were carried forward for diagnosis of safety problems and analysis of potential remedies. The three areas under analysis consisted of five CPZs: 3160, 3200, 3460, 3640 and 2290 (in the Mount Pleasant and Downtown neighbourhoods of the City of Vancouver). Due to data availability, the analysis base year used was 2011. Since 2011, the City of Vancouver has constructed additional cycling infrastructure such as local street bikeways, painted bike lanes, marked shared lanes and separated bike lanes (see Figure 9), as well as implemented traffic calming and spot improvements [86]. All infrastructure changes were taken into account in the detailed diagnosis and remedy for each CPZ.

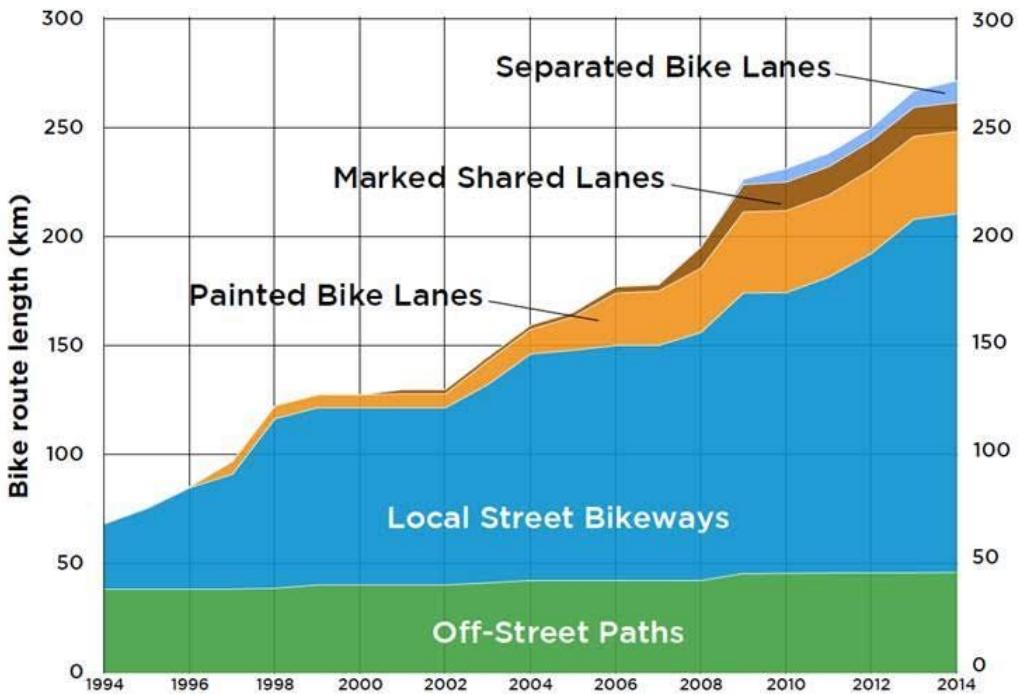


Figure 9: New cycling infrastructure built in the City of Vancouver [86]

The detailed diagnosis was followed by a strategic remedy analysis at the micro-level, including infrastructure treatments to improve bicycle safety. The following safety countermeasures that could be applied to the five diagnosed CPZs are described in Table 6. These infrastructure treatments are intended to remedy the trigger variables of arterial high volume roads and high intersection density. The remedies are intended for potentially high bicycle-vehicle conflict locations, such as intersections.

Table 7: Bicycle safety countermeasures

Remedy	Description	Photo
All Ages and Abilities ("AAA")	A general term referring to designing bicycle infrastructure for people of all ages and abilities, typically separated bicycle lanes for high vehicle volume streets and traffic calming local street bikeways for low volume streets.	
Curb Bulges and Traffic Calming	A traffic calming measures such as narrowing road at intersections to improve crossing safety for vulnerable road users (pedestrians and cyclists) or restricting vehicle access.	
Bicycle Refuges	Provides safety to bicycle crossings where a median continues through the intersection.	
Elephant's Feet	Bicycle crosswalk paint markings usually applied parallel to or part of a pedestrian crosswalk.	
Bike Boxes	Painted at an intersection to designate an area where cyclists may wait ahead of motor vehicles at a red signal to get into position for go before motor vehicles when the signal turns green.	
Bicycle Signals	A separate signal phase to allow bicycles to cross across high volume and speed traffic.	

Remedy	Description	Photo
Protected Intersections	An intersection design treatment so that both cyclists and pedestrians are separated and protected from vehicles. Turning phases are also protect bicycles and pedestrians from vehicles.	
Coloured Bicycle Lanes	Installed across high conflict vehicle-bicycle crossing zones to caution both drivers and cyclists.	

4.1.3.1 Collision Prone Zones 3160 & 3200

Collision prone zones 3160 and 3200 are two adjacent zones (see Figure 10 showing bike routes in green and bicycle-vehicle collisions as red dots) bordered by West 1st Avenue in the north, Cambie Street in the west, West 16th Avenue in the south and Main Street in the east. The area has multiple bike routes, represented in green, including local street bikeways on 5th Avenue, 10th Avenue, Columbia Street, Ontario Street and Yukon Street, marked shared lanes on Main Street, painted bike lanes on 7th Avenue and Yukon Street and separated bike lanes coming off of Cambie Street Bridge. Since 2011, there has been a change to CPZ 3160 with a painted bike lane on Yukon Street from 2nd Avenue to 10th Avenue installed in 2012.

The CPZ 3160 (north zone from West 1st Avenue to West Broadway) was identified as the top most severe collision prone zone using the EB safety estimate approach for the exposure, TDM, network and most socio-demographic model groups. CPZ 3200 (south zone from West Broadway to West 16th Avenue) was found to be in the top three severe collision prone zones for

one socio-demographic and two network model groups. From all the zones in the City of Vancouver, the CPZ 3160 had the highest number of bicycle vehicle collisions over three years, at 78, while CPZ 3200 had the third highest number of bicycle vehicle collisions over three years, at 65. CPZ 3160 showed a higher than average bicycle kilometer travelled, total commuters, commercial land use area, total employment, signal and intersection density, bus stop density and total arterial street kilometers; and showed a lower than average percentage of park area, total “AAA” bicycle kilometers and household density. Meanwhile, CPZ 3200 showed a higher than average commercial land use area, total employment and total arterial street kilometers; while also showing a lower than average percentage of park area.

There were a few clues to road safety problems in the two CPZs that could be gathered from observing the land use, trigger variables and bicycle-vehicle collisions spatial characteristics. The land use consisted primarily of light industrial in CPZ 3160 and medium density residential in CPZ 3200, with commercial uses throughout the two zones, specifically along the major transportation corridors of Cambie, West Broadway, Main Streets and West 2nd Avenue. These four arterial streets also had high vehicle, truck and transit traffic. All roads within these two zones were set in a grid pattern with high intersection density. The bicycle volume in both zones was also high due to high density and many bicycle routes. Supporting the observed zone attributes, the trigger variables for both zones were found to be: a higher than average total commercial land use area, employment and arterial street kilometers, along with lower than average percentage of park area. Additionally, CPZ 3160 had a higher than average signal and intersection density, as well as a high BKT with a low number of bicycle kilometers ranked “AAA”. The topography in the two CPZs was also found to be hilly, visualized with 1

and 10 meter contour lines in Figure 11. Finally, there was a high number of bicycle vehicle collisions observed in the two CPZs, primarily at intersections and arterials (see Figure 10).

These clues suggest that the safety issue may be a result of the land use type, coupled with high traffic volume particularly on the arterial routes and high bicycle volume on bicycle routes that are classified as painted bike lanes, shared lanes and local street bikeways. These types of bicycle routes are typically less safe for cyclists and they are not classified as “AAA”. There were many observed collisions along arterial routes or bicycle routes, areas that had the highest bicycle and vehicle volumes. Possible remedies suggested to solve the safety issues in these zones include:

- Update existing bicycle routes to be ranked “AAA,” by installing separated bike lanes or installing traffic calming measures where applicable, depending on vehicle volumes on the route. This includes installing future separated bicycle lanes on the busy arterial routes of Cambie, Broadway and Main Streets.
- Increase signage along all routes, and install more local street bikeways onto low roads with low vehicle volumes to raise drivers’ attention to cyclists.
- Continue to address intersection safety through bicycle infrastructure spot improvements such as curb bulges, bicycle refuges, elephant feet, bike boxes and bike signals, especially at high volume intersections. Protected intersections should be considered at high vehicle volume locations that also have high bicycle volumes, such as the Cambie Street Bridge & 6th Avenue intersection.

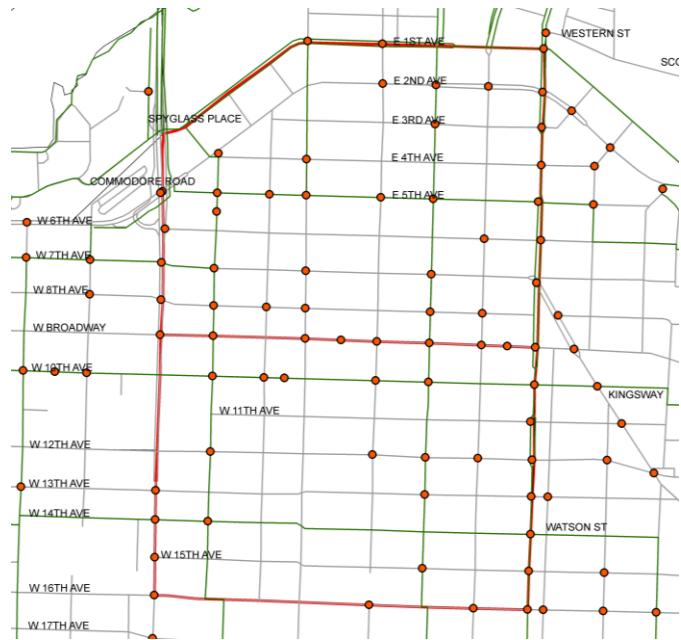


Figure 10: Collision Prone Zones 3160 (north) & 3200 (south) showing bike routes and bicycle-vehicle collisions



Figure 11: Collision Prone Zones 3160 (north) & 3200 (south) showing topography (1 m pink contour lines and 10 m purple contour lines)

4.1.3.2 Collision Prone Zones 3460 & 3640

Collision prone zones 3460 and 3640 are two adjacent zones (see Figure 12). CPZ 3460 is bordered by Terminal Avenue in the north, Main Street in the west, East Broadway in the south and Clark Drive in the east. CPZ 3640 is bordered by East Broadway in the north, Fraser Street in the west, East 16th Avenue in the south and Victoria Street in the east. The areas have multiple bike routes, represented in green, including off-street paths on Northern Street, local street bikeways on East 10th Avenue, East 5th/6th/7th Avenue, Windsor Street, Glen Drive and Woodland Drive, marked shared lanes on Main Street and 1st Avenue, and separated bike lanes on Great Northern Way, Victoria Drive and through Clark Park and China Creek South Park.

The CPZ 3460 (north zone from Terminal Avenue to East Broadway) was identified in the top three most severe collision prone zones using the EB safety estimate approach for the exposure, TDM, and most socio-demographic and one network model groups, and fourth most severe collision prone zone for all other network model groups. CPZ 3640 (south zone from East Broadway to East 16th Avenue) was found to be in the top two severe collision prone zones for all exposure, TDM and network model groups, and most socio-demographic model groups. From all the zones in the City of Vancouver, the CPZ 3460 had the fourth highest number of bicycle vehicle collisions over three years, at 45, while CPZ 3640 had the second highest number of bicycle vehicle collisions over three years, at 72. CPZ 3460 showed a higher than average total commuters, commercial land use area, total employment, total population, bus stop density and total arterial street kilometers; and showed a lower than average percentage of park area, total “AAA” bicycle kilometers, and intersection and signal density. Meanwhile, CPZ 3640 showed a higher than average bicycle kilometers travelled, commercial land use area, household density, total population, bus stop density and total arterial street kilometers; and showed a lower than

average total commuters, total employment, percentage of park area, percentage of residential street kilometers, total “AAA” bicycle kilometers, and intersection density.

There were a few clues to road safety problems in the two CPZs that could be gathered from observing the land use, trigger variables and bicycle-vehicle collisions values. The land use consisted primarily of light industrial, institutional (education), commercial and medium density residential in CPZ 3460 and medium density residential and commercial in CPZ 3460. There were also arterial streets (Terminal Avenue, Great Northern Way, East Broadway, East 12th Avenue, Clark Drive and Commercial Drive) that have high vehicle, truck and transit traffic. Most roads within these two zones are set in a grid pattern, but there are also physical barriers such as the industrial railway. Supporting the observed zone attributes, the trigger variables for both zones were found to be: a higher than average total commercial land use area, total population, bus stop density and arterial street kilometers, along with lower than average percentage of park area, total “AAA” bicycle kilometers, and intersection density. The topography in the two CPZs was also found to be hilly, visualized with 1 and 10 meter contour lines in Figure 13. In addition, there was a high number of bicycle vehicle collisions observed in the two CPZs, primarily clustered at intersections, along bicycle routes and along the arterial routes of Clark Drive and Commercial Drive (see Figure 12).

These clues suggest that the safety issue may be a result of land use type of commercial and industrial uses and high population density in the zones, coupled with high traffic volume on the arterial routes and high bicycle volume on bicycle routes resulting in more exposure to collision risk. High speeds along arterial high volume transportation corridors such as Clark Drive also alleviate the risk and severity of bicycle collisions. Possible remedies suggested to solve the safety issues in these zones include:

- Update existing bicycle routes to be ranked “AAA,” by installing separated bike lanes or installing traffic calming measures where applicable, depending on vehicle volumes on the route.
- Install new “AAA” bicycle infrastructure along high-volume corridors that attract cyclists due to commercial land uses such as on Main, Broadway, Clark and Commercial Drive.
- Continue to address intersection safety at high-volume intersections through bicycle infrastructure spot improvements such as curb bulges, bicycle refuges, elephant feet, bike boxes and bike signals, and increase signage along all routes to raise drivers’ attention to cyclists.



Figure 12: Collision Prone Zones 3460 (north) & 3640 (south) showing bike routes and bicycle-vehicle collisions

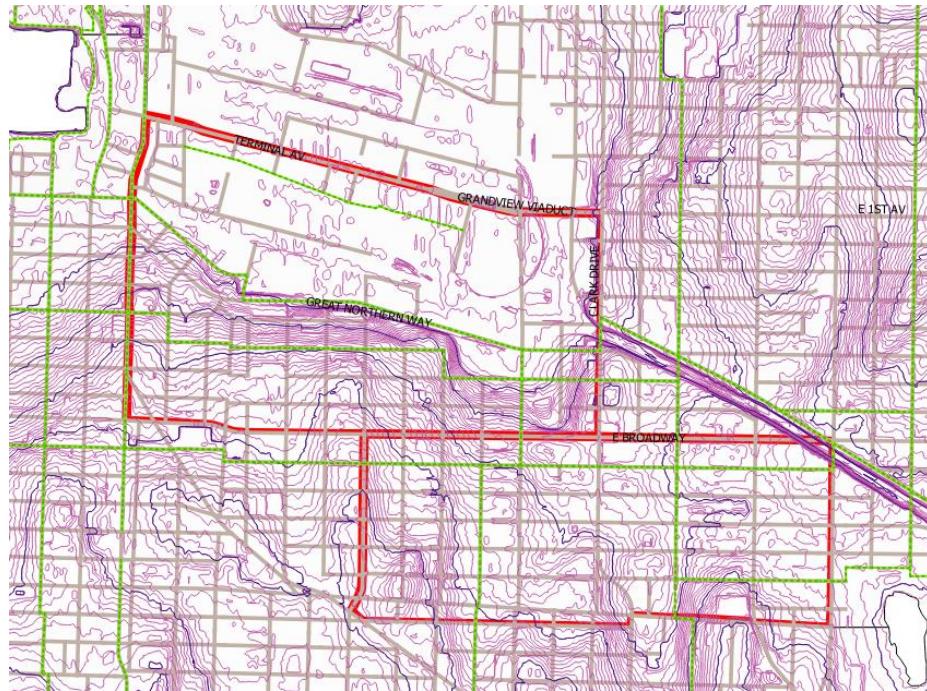


Figure 13: Collision Prone Zones 3460 (north) & 3640 (south) showing topography (1 m pink contour lines and 10 m purple contour lines)

4.1.3.3 Collision Prone Zone 2290

The CPZ 2290 is located in downtown Vancouver and is bordered by Nelson Street in the north, Burrard Street in the west, Pacific Street in the south and Granville Street in the east (see Figure 14). The area has multiple bike routes, represented in green, including marked shared lanes on Burrard Street and Pacific Street and separated bike lanes on Hornby Street, Burrard Street and Helmcken Street. Since 2011, there have been some significant changes to bicycle infrastructure in CPZ 2290. A separated bicycle lane was installed on Helmcken Street from Burrard Street to Hornby Street in 2013. A separated bicycle lane leading to Burrard Bridge and Pacific Street intersection upgrades is under construction and expected to be completed in 2017 [87].

The CPZ 2290 was identified in the top four most severe collision prone zones using the EB safety estimate approach for the exposure, and TDM model groups, and in the top ten for all socio-demographic and most network model groups. From all the zones in the City of Vancouver, the CPZ 2290 had the seventh highest number of bicycle vehicle collisions over three years, at 39, which is a significant number for such a small zone. CPZ 2290 showed a higher than average employment, total arterial street kilometers, and total “AAA” bicycle kilometers; and showed a lower than average percentage of park area and total population.

There were a few clues to road safety problems in the CPZ that could be gathered from observing the land use, trigger variables and bicycle-vehicle collisions. The land use were primarily downtown site-specific zoning, and consisted of a high-density mix of office, commercial and residential land uses. All streets in this zone were in a grid pattern and designated as arterials, as they had high vehicle, truck and transit traffic. Supporting the observed zone attributes, the trigger variables for the zone were found to be: a higher than average employment, arterial street kilometers, and total “AAA” bicycle kilometers, along with lower than average percentage of park area. The topography in the CPZ was also found to be hilly, visualized with 1 and 10 meter contour lines in Figure 15. Finally, there was a high number of bicycle vehicle collisions observed primarily clustered at intersections and along arterials (see Figure 14). There were many observed collisions along Burrard Street specifically, most likely due to a lack of separation between modes and a steep topography.

These clues suggest that the safety issue may be a result of more exposure to collision risk, due to high density coupled with high traffic and bicycle volumes. High speeds along arterial high volume transportation corridors alleviate the risk and severity of bicycle collisions. Possible remedies suggested to solve the safety issues in these zones include:

- Update existing bicycle routes to be ranked “AAA,” by installing separated bike lanes due to high vehicle volumes on all arterials in this zone. This countermeasure is currently under construction along Burrard Street southbound, but should also be considered northbound.
- Continue to address intersection safety through bicycle infrastructure spot improvements such as protected intersections, curb bulges, bicycle refuges, elephant feet, bike boxes and bike signals, and increase signage along all routes to raise drivers’ attention to cyclists. Install protected intersections at the junction of two protected bicycle lanes. This countermeasure is currently under construction at the Burrard Street Bridge and Pacific Street intersection.
- Monitor innovative infrastructure changes in the area, such as the plan to install a protected intersection at Burrard and Pacific Streets, to see whether there is an observed improvement in bicycle safety through a reduction of collisions.

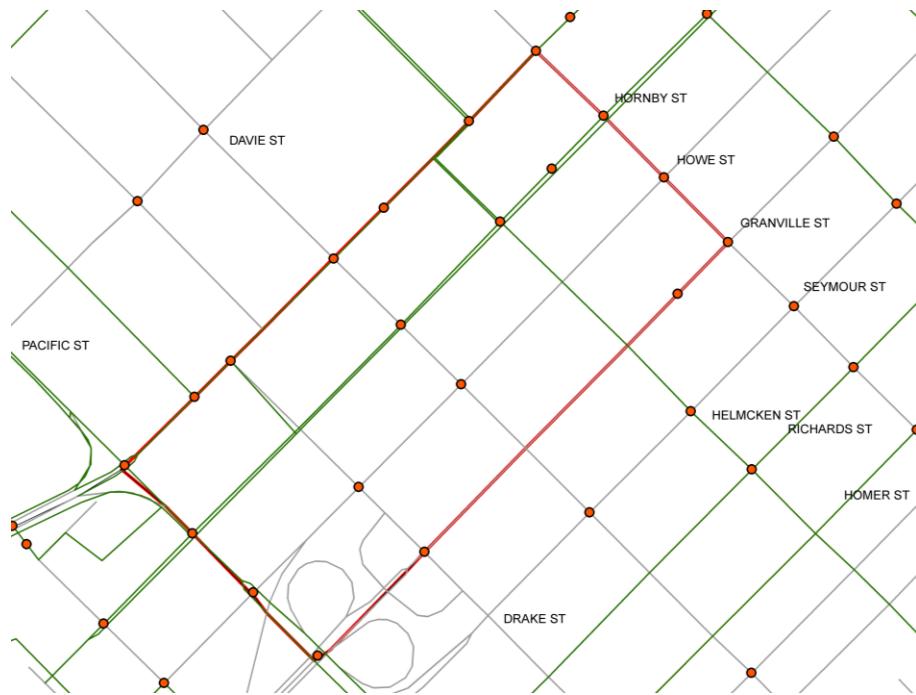


Figure 14: Collision Prone Zone 2290 showing bike routes and bicycle-vehicle collisions

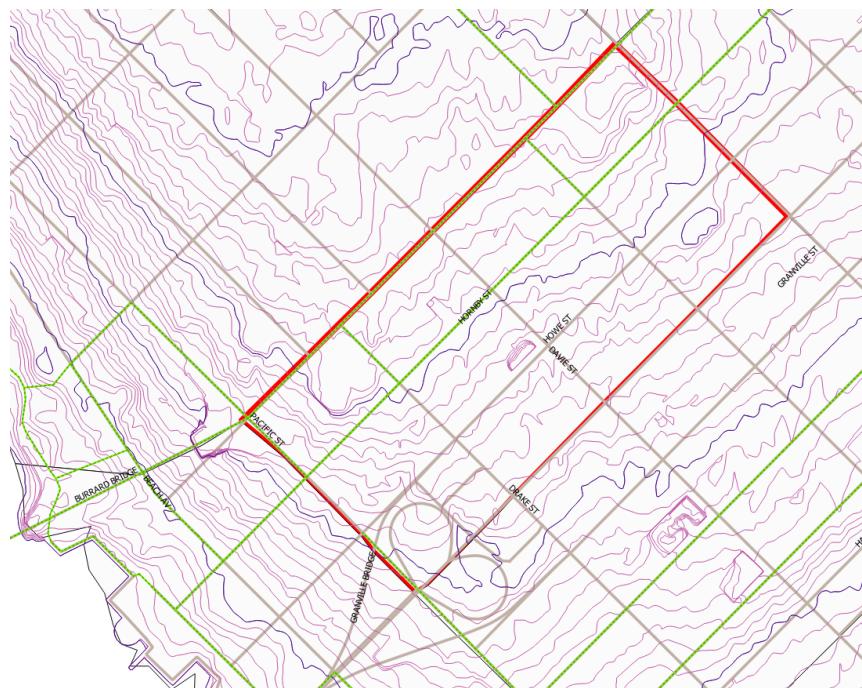


Figure 15: Collision Prone Zone 2290 showing topography (1 m pink contour lines and 10 m purple contour lines)

Chapter 5: Conclusion

5.1 Summary and Conclusions

The purpose of this research has been to develop macro-level CPMs and to use the models as empirical tools for bicycle road safety evaluation and planning. The goal was to advance the proactive use of macro-level bicycle related CPMs, allowing practitioners to use them to consider safety in all stages of the transportation planning and design process. Through predicting bicycle-vehicle collisions, the models advance safety performance, and can be used to identify and rank Collision Prone Zones (CPZs) and diagnose problems through the consideration of road safety countermeasures. The main motivations of this study came from environmental, public health and safety concerns. To encourage the sustainable transportation mode of cycling, it is fundamental to build a safe and comfortable cycling environment. Addressing the safety of cyclists, there is a need for empirical tools to evaluate bicycle safety proactively before collisions occur. Bicycle safety research is an emerging field of study, which has yet to reach the detailed extent of vehicular traffic research.

In this study, macro-level CPMs were developed for bicycle-vehicle collisions and applied to a case study of the City of Vancouver (COV) at the zonal level, with 134 traffic analysis zones. The models were effective in identifying and ranking dangerous CPZs, and the conclusions of the study are summarized below.

5.1.1 Bicycle Macro-Level CPM Development

Several studies have previously developed macro-level CPMs, however there has been little research on developing *bicycle* macro-level CPMs. The first objective of this study was to develop macro-level CPMs for bicycle-vehicle collisions using bicycle kilometers travelled

(BKT) as the leading exposure variable for the models. To achieve this objective, the research first identified data needs in terms of exposure, infrastructure and socio-demographic variables that would influence bicycle safety to input into the statistical models. Sufficient good quality data from the base year 2011 ensured the development of well-fit and reliable statistical models. The data was aggregated to 134 TAZs in the COV, using TransLink's EMME/2 TAZ level. This aggregate level was chosen to achieve research objectives and to simplify data integration due to the zones' overlap with both census tracks and municipal boundaries. The data was acquired from multiple sources including TransLink, City of Vancouver, Digital Road Atlas and ICBC. Data quality issues that were addressed in this research included the scarcity of reported bicycle-vehicle collisions and the influence of collisions on zone boundaries.

Once data needs were met, the next step of the research was to develop the bicycle-related macro-level CPMs using generalized linear regression modeling. The process followed a stepwise procedure assuming a NB error distribution and a 90% to 95% confidence level to assess the model goodness of fit. The model fit was evaluated using the SD and *Pearson X*² measures. Ten macro-level CPMs presented in Table 4.1 were developed following this methodology, using BKT as the leading exposure variable and other exposure, socio-demographic and infrastructure explanatory variables. Each developed model used one or more explanatory variables to estimate the expected mean frequency of collisions over three years in each zone. The variables that were found to be significant were presented in Tables 3 and 4. The conclusions from the model development results show that it is possible to quantify on a zonal scale a statistically significant association between bicycle-vehicle collisions and specific exposure, socio-demographic and infrastructure characteristics. Most importantly, the models were used to predict collisions using the leading bicycle exposure variable of BKT. This is in

contrast to other studies in the literature that use proxy bicycle exposure measures such as population or cycling mode share. This research recommends that communities proactively address bicycle safety through using macro-level CPMs to predict collisions before they occur.

5.1.2 Macro-Reactive Road Safety Application

The second objective of this thesis was to demonstrate the use of bicycle macro-level CPMs by applying the models to a case study of the COV and performing a macro-reactive road safety improvement program. This model assessment was important to demonstrate the effectiveness of macro-level CPMs for use in road safety applications by planners and engineers. To achieve this objective, the CPMs were used to identify and rank CPZs in the COV, to diagnose the zonal safety problems in each zone and to recommend potential remedies based on zonal diagnosis and characteristics. The 10 macro-level CPMs that were developed for this study resulted in 10 different rankings and diagnoses of the most dangerous TAZs. The 10 models showed consistency with little variability for the top ranked CPZs identification and ranking. Following this ranking technique, the three areas with CPZs that were found to be the most collision prone were then carried forward for diagnosis and remedy through identification of potential safety countermeasures.

The results showed the most collision prone areas to be located in the Mount Pleasant and Downtown neighbourhoods. All three areas had high bicycle volume due to high population density and multiple bicycle routes, as well as high vehicle volume along arterial roads. The three areas also had the highest number of observed bicycle collisions over three years in the City of Vancouver. The recommended safety countermeasures included infrastructure treatments to separate bicycles from vehicle traffic on arterial roads, traffic calming for vehicles on local

roads and at intersections and increased visibility of cycling through paint and signage. Since the model's base year 2011 the City of Vancouver has been rapidly constructing cycling infrastructure. The top ranked CPZs may be similar today in 2016 despite some infrastructure changes due to high bicycle and vehicle volumes in the areas. However, further research is needed to perform the same safety analysis on updated collision, exposure, socio-demographic and infrastructure data to understand how the cycling infrastructure updates have been effective in improving cycling safety. This research recommends that the City continue to address bicycle safety through improving cycling infrastructure for routes that have high bicycle and/or vehicle volumes.

5.2 Contributions

The two main contributions of this thesis relate directly to the previously identified knowledge gaps and research goals, and are stated as follows:

- The development of bicycle related macro-level collision prediction models using bicycle indicators for safety. This thesis is unique in its development of models predicting bicycle collisions using bicycle volume (BKT) as the leading exposure variable. This is important because the volume of bicycles and vehicles has a large influence on bicycle safety, along with collision data. In addition, the amount of data on bicycle collisions is limited. This is an improved and reliable empirical tool that can be used by planners and engineers to evaluate bicycle safety on a macro-scale.
- The demonstration of the validity of the use of the macro-level CPMs to enhance traditional safety initiatives, through model use in macro-reactive road safety programs.

Traditionally, the use of CPMs in road safety improvement programs has been focused on

micro-level CPMs, at an intersection or single facility scale. The macro-level models were applied to a case study of the COV to identify, rank, diagnose and remedy CPZs with respect to bicycle safety. Using the information provided by the models, potential safety countermeasures were brought forward for the top three collision prone areas in the city. This case study effectively demonstrates the use of the models to enhance bicycle safety using this valuable safety tool.

5.3 Recommendations for Future Research

Rich, good quality data is integral for developing reliable statistical models. The work to date on macro-level CPM development and application has been an important base for systematically applying road safety to the planning process. The methodology described in this thesis could be further improved with research in the following topics:

- The reduction of data needs for the development of macro-level CPMs. Data assembly for reliable statistical model development is intensive. For planners and engineers with a lack of good quality of data, the models could be refined to require fewer variables while retaining model goodness-of-fit. The timeliness of the data is important, as infrastructure planning and design decisions should be made with up-to-date data. Simplifying data needs would encourage practitioners to develop and apply CPMs and advance road safety in the transportation planning process.
- The macro-reactive black spot steps of (1) diagnosis and (2) remedy were conducted at a preliminary level due to a lack of collision detailed information and the large TAZ size. Improved data quality and smaller aggregation for zones is recommended for future studies in high-density areas. Additionally, the thesis could have been further enhanced

through a parallel micro-level reactive analysis. For future research, the diagnosis and remedy should be further researched for site-specific detail refinement.

- Continue to focus research on developing models for vulnerable road users' safety to advance sustainable transportation planning. The safety of cyclists is critical, since bicycle travel has a higher per-mile casualty rate than car travel, yet poses minimal risk to other road users. The study and monitoring of sustainable transportation modes is vital to encourage environmentally friendly transportation options. Important considerations include collecting cycling and pedestrian data at the same level as vehicle data, to be able to accurately monitor volumes and safety. Currently, the largest barrier to increasing bicycle mode split is improving the comfort and safety of riding for cyclists.

References

- [1] World Health Organization, "Global status report on road safety 2013," 2013.
- [2] World Health Organization, "The Global Burden of Disease, 2008," 2011.
- [3] Transport Canada, "Canadian motor vehicle traffic collision statistics: 2013," 2013.
- [4] Transport Canada, "Transport Canada Releases First, Systematic Analysis of Cost of Urban Traffic Congestion in Canada," 2006.
- [5] Vijayakumar, N., & Burda, C., "Cycle Cities: Supporting cycling in Canadian cities," Pembina Institute, 2015.
- [6] City of Vancouver, "Pedestrian Safety Study," City of Vancouver, 2012.
- [7] ICBC, "Quick Statistics," 2015.
- [8] Chong S, Poulos R, Olivier J, et al., "Relative injury severity among vulnerable non-motorised road users: comparative analysis of injury arising from bicycle-motor vehicle and bicycle-pedestrian collisions," *Accid Anal Prev.*, vol. 42, no. 1, pp. 290-296, 2010.
- [9] Cripton, P., et al. , "Severity of urban cycling injuries and the relationship with personal, trip, route and crash characteristics: analyses using four severity metrics," *BMJ Open*, 2014.
- [10] Victoria Transportation Institute, "Evaluating Active Transportation Benefits and Costs.," 2014.
- [11] Government of Ontario, "Cycling Death Review: A Review of All Accidental Cycling Deaths in Ontario from January 1st, 2006 to December 31st, 2010.," 2012.
- [12] Litman, T., "Win-Win Transportation Emission Reduction Strategies," Victoria Transport Policy Institute, 2014.

- [13] British Columbian Ministry of Transportation, "Bicycle Facilities Design," 2011.
- [14] Basett et al., "Walking, Cycling, and Obesity Rates in Europe, North America, and Australia," *Journal of Physical Activity and Health*, vol. 5, pp. 795-814, 2008.
- [15] Litman, T., "Evaluating Active Transportation Benefits and Costs," Victoria Transportation Institute, 2014.
- [16] Ismail, K., "Application of Computer Vision Techniques for Automated Road Safety Analysis and Traffic Data Collection," *Unviersity of British Columbia, Retreived from Circle*, 2002.
- [17] Chin, H., & Quek, S., "Measurement of Traffic Conflicts," *Safety Science*, vol. 26, no. 3, pp. 169-185, 1997.
- [18] De Leur, P., "Improved Approaches to Manage Road Safety Infrastructure," *Retreived from ProQuest Dissertations and Theses*, 2001.
- [19] DeLeur, P., & Sayed, T., "A framework to proactively consider road safety within the road planning process.,," *Canadian Journal of Civil Engineering*, vol. 30, no. 4, pp. 711-719, 2003.
- [20] Lovegrove, G., Lim, C., & Sayed, T., "Using Macro-Level Collisions Prediction Models to Conduct a Road Safety Evaluation of a Regional Transportation Plan," *Transportation Research Board 87th Annual Meeting* , 2008.
- [21] Lovegrove, G. & Sayed, T., "Macro-level collision prediction models for evaluating neighbourhood traffic safey," *Can. J. Civ. Eng.*, vol. 33, no. 5, pp. 609-621, 2006.
- [22] Lovegrove, G., & Sayed, T., "Using macro-level collision prediction models in road safety

- planning applications," *Transportation Research Record*. 1950, pp. 73-82, 2006.
- [23] City of Vancouver, "Cycling Safety Study," 2015.
- [24] Statistics Canada, "Canadian National Household Survey," 2011.
- [25] Sayed, T., & Rodriguez, F., "Accident Prediction Models for Urban Unsignalized Intersections in British Columbia," *Transportation Research Record* 1665, 1999.
- [26] MAAS, J., "SWOV – Institute for Road Safety Research in the Netherlands," *IATSS Research*, vol. 30, no. 1, pp. 122–125, 2006.
- [27] van Schagen, I., & Janssen, T., "Managing Road Transport Risks," *IATSS Research*, vol. 24, no. 2, pp. 18-27, 2000.
- [28] SWOV Institute for Road Safety Research, "Advancing Sustainable Safety: National Road Safety Outlook for 2005-2020," 2006.
- [29] Weijermars et al., "Duurzaam Veilig, ook voor ernstig verkeersgewonden.," *SWOV*, 2013.
- [30] Berkovitz, A., "The Marriage of Safety and Land-Use Planning: A Fresh Look at Local Roadways.," *Federal Highway Administration Research and Technology*, 2001.
- [31] Cambridge Systematics, Inc., "Framework for Institutionalizing Safety in the Transportation Planning Process," 2012.
- [32] The National Cooperative Highway Research Program Report 05-46, "Incorporating Safety into Long-Range Transportation Planning," *Transportation Research Board*, 2006.
- [33] World Health Organization, "World report on injury prevention," 2004.
- [34] Gaines, D. & Meyer, M., "Safety-Conscious Planning in Midsized Metropolitan Areas: Technical and Institutional Challenges," *Transportation Research Record: Journal of the*

Transportation Research Board, No. 2078, pp. 1–7, 2008.

[35] Pedestrian and Bicycle Information Center, "Bicycle and Pedestrian Forecasting Tools: State of the Practice," 2015.

[36] Sayed, T., "New and Essential Tools, Part 1: Selecting Sites for Treatment," *Proceedings from Seminar on New Tools for Traffic Safety*, pp. 1-13, 1998.

[37] Sayed, T., W. Abdelwahab, and F. Navin, "Application of Fuzzy Pattern Recognition to the Identification of Accident Prone Locations," *Journal of Transportation Engineering*, vol. 121, no. 3, pp. 352-358, 1995.

[38] Lovegrove, G., *Road Safety Planning: New Tools for Sustainable Road Safety and Community Development*. Germany: Verlag Dr. Muller, 2007.

[39] Sawalha, Z., & Sayed, T., "Traffic accident modeling: some statistical issues," *Canadian Journal of Civil Engineering*, vol. 33, no. 9, p. 1115, 2006.

[40] Lovegrove, G. & Sayed, T., "Macrolevel Collision Prediction Models to Enhance Traditional Reactive Road Safety Improvement Programs," *Transportation Research Record: Journal of the Transportation Research Board*, No. 2019, pp. 65-73, 2007.

[41] Ho, G., & Guarnaschelli, M., "Developing a road safety module for the regional transportation model, technical memorandum one framework," *ICBC*, 1998.

[42] Lord, D., Persaud, B., & Palmisano, J., "Calibration and transferability of accident prediction models for urban intersections," pp. 57-64, 2002.

[43] Hadayeghi, A., Shalby, A., Persaud, B. , "Macro-level accident prediction models for evaluating the safety of urban transportation systems," *Transportation Research Board*, pp.

48-61, 2003.

- [44] Ladron de Guevara, F., Washington, S., & Oh, J., "Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona," *Transportation Research Record*, vol. 1897, no. 1897, 2004.
- [45] Cheung, C., Shalaby, A., Persaud, B., & Hadayegy, A , "Models for safety analysis road surface transit , " *Transportation Research Record 2063* , pp. 168-175, 2008.
- [46] Quintero, L., Sayed, T., & Wahba, M., "Safety models incorporating graph theory based transit indicators," *Accident Analysis and Prevention*, vol. 50, pp. 635-644, 2013.
- [47] Quintero-Cano, L., "Graph Theory Based Transit Indicators Applied to Ridership and Safety Models," 2011.
- [48] Chatterjee et al., "Tools for assessing safety impacts of long-range transportation plans in urban areas," *Office of Metropolitan Planning and Programs, FHWA, US DOT*, 2003.
- [49] Hirst, W., Mountain, L., & Maher, M., "Sources of error in road safety scheme evaluation: a method to deal with outdated accident prediction models," *Accident Analysis and Prevention*, vol. 36, pp. 717-727, 2003.
- [50] Hadayeghi, A., Shalaby, A., & Persaud, B., "Safety Prediction Models Proactive Tool for Safety Evaluation in Urban Transportation Planning Applications," *Transportation Research Record: Journal of the Transportation Research Board*, No. 2019, pp. 225-236, 2007.
- [51] Lovegrove, G., & Litman, T., "Macrolevel Collision Prediction Models to Evaluate Road Safety Effects of Mobility Management Strategies: New Empirical Tools to Promote

- Sustainable Development," *Transportation Research Board 87th Annual Meeting* , 2008.
- [52] Litman, T., et al., "Pedestrian and Bicycle Planning Guide to Best Practices," *Ongoing Guide*, 2000.
- [53] SWOV Institute for Road Safety Research, "Safe Cycling Network Developing a system for assessing the safety of cycling infrastructure," 2014.
- [54] National Highway Traffic Safety Administration, "Traffic Safety Facts, 2012 data – bicyclists and other cyclists.," *Washington, DC: US Department of Transportation*, 2014.
- [55] Xing, Y., Handy, S., & Buehler, T., "Factors Associated with Bicycle Ownership and Use: A Study of 6 Small U.S. Cities," *Transportation Research Board Annual Meeting*.
- [56] Garrard, J., Rose, G., & Lo, S., "Promoting Transportation Cycling for Women: The Role of Bicycle Infrastructure," *Preventative Medicine*, vol. 46, pp. 55-59, 2008.
- [57] Winters, M., Friesen, M., Koehoorn, M., & Teschke, "Utilitarian Bicycling, A Multilevel Analysis of Climate and Personal Influences," *American Journal of Preventative Medicine*, vol. 32, no. 1, pp. 52-58, 2007.
- [58] Rietveld, P. & Daniel, V., "Determinants of Bicycle Use: Do Municipal Policies Matter?," *Transportation Research Part A*, vol. 38, no. 7, pp. 531-550, 2004.
- [59] Stinson, M. & Bhat, C., "Frequency of Bicycle Commuting: Internet-Based Survey Analysis," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1878, pp. 122-130, 2004.
- [60] Reynolds, C., et al. , "The Impact of Transportation Infrastructure on Bicycle Injuries and Crashes: A Review of the Literature," *Environmental Health*, vol. 8, p. 47, 2009.

- [61] Sun, J., & Lovegrove, G., "Comparing the road safety of neighbourhood development patterns: traditional versus sustainable communities ,"*Canadian Journal of Civil Engineering* , vol. 40, no. 1, pp. 35-45, 2013.
- [62] Wei, F. & Lovegrove, G., "An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediciton models using negative binomial regression,"*Accident Analysis & Prevention*, vol. 61, pp. 129-137, 2013.
- [63] Wei, F., Alam, A., & Lovegrove, G., "Macro-Level Collision Prediction Models Related to Bicycle Use," in *ICTIS 2011.*, 2011.
- [64] Jacobsen, P., "Safety in numbers: more walkers and bicyclists, safer walking and bicycling,"*Inj Prev.*, vol. 9, pp. 205-209, 2003.
- [65] Kim, K., Pant, P., & Yamashita, E. , "Accidents and accessibility measuring theinfluences of demographic and land use variables in Honolulu, Hawaii," *Transportation research record: Journal of the Transportation Research Board* 2147, pp. 9-17, 2010.
- [66] Yan et al. , "Motor vehicle–bicycle crashes in Beijing: Irregular maneuvers, crash patterns, and injury severity," *Accident Analysis and Prevention*, vol. 43, pp. 1751-1758, 2011.
- [67] Siddiqui, C., Abdel-Aty, M., & Choi, K., "Macroscopic spatial analysis of pedestrian and bicycle crashes," *Accident Analysis & Prevention*, vol. 45, pp. 382-391, 2012.
- [68] Abdel-Aty et al. , "Integrating Trip and Roadway Characteristics to Manage Safety in Traffic Analysis Zones," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2213, 2014.
- [69] Yasmin, N. & Eluru, S., "Latent Segmentation Based Count Models: Analysis of Bicycle

- Safety in Montreal and Toronto," 2014.
- [70] Kwon et al., "Development of Bicycle Accident Prediction Model and Suggestion of Countermeasures on Bicycle Accidents," *Journal of the Korean Society of Civil Engineers* , vol. 35, no. 5, 2015.
- [71] Lim, C., "Application of the Lovegrove/Sayed Macro Collision Prediction Models," *unpublished manuscript*, 2004.
- [72] Amer, A., "Proactive Safety Assessment in Transportation Planning, Guidelines for Better Macro-Level Collision Prediction Models," *unpublished manuscript*, 2014.
- [73] Esawey, M., Lim, C., Sayed, T., & Mosa, A., "Development of Daily Adjustment Factors for Bicycle Traffic," *Journal of Transportation Engineering*, vol. 139, pp. 859-871, 2013.
- [74] Hudson, J., et al. , "Using Smartphones to Collect Bicycle Travel Data in Texas," *Texas Transportation Institute*, 2012.
- [75] Vlassenroot, S., Gillis, D., Bellens, R., & Gautama, S., "The Use of Smartphone Applications in the Collection of Travel Behaviour Data," *International Journal of Intelligent Transportation Systems Research*, vol. 13, no. 1, 2015.
- [76] Miaou, S., & Lum, H., "Modelling vehicle accident and highway geometric design relationships," *Accident Analysis and Prevention Elsevier Ltd*, pp. 689-709, 1993.
- [77] Sawalha, Z., & Sayed, T. , "Accident Prediction and Safety Planning Models for Urban Arterial Roadways," 1999.
- [78] Sawalha, Z., & Sayed, T., "Evaluation of safety in urban arterial roadways," *Journal of Transportation Engineering*, vol. 127, pp. 151-158, 2001.

- [79] Sayed, T., & Rodriguez, F., "Accident Prediction Models for Urban Unsignalized Intersections in British Columbia," *Transportation Research Record 1665, Journal of the TRB*, vol. , no. , pp. 93-99, 1999.
- [80] Statistics Canada, "Focus on Geography Series, 2011 Census," Statistics Canada Catalogue no. 98-310-XWE2011004, Ottawa, ON, 2012.
- [81] City of Vancouver, "Cycling in Vancouver," 2015.
- [82] GeoBC. (October 12, 2015) Digital Road Atlas. [Online]. HYPERLINK "<http://geobc.gov.bc.ca/base-mapping/atlas/dra/>"
- [83] Fotheringham, A. S. & Wegner, M., *Spatial Models and GIS: New Potential and New Models*. London: Taylor & Francis Inc., 2000.
- [84] Cui, G., Wang, X., & Kwon, D., "A framework of boundary collision data aggregation into neighbourhoods," *Accident Analysis & Prevention*, vol. 88, pp. 1-17, 2015.
- [85] Miller, S., Garber, N. & Kamatu, J., "Resource Guide for Enhancing the Incorporation of Safety into the Regional Planning Process," *Transportation Research Record: Journal of the Transportation Research Board*, No. 2244, pp. 50–60, 2011.
- [86] Lord, D. & Persaud, B., "Estimating the safety performance of urban road transportation networks , " *Accident Analysis and Prevention*, vol. 36, no. 4, pp. 609-620, 2004.
- [87] Littman, T., et al., "Pedestrian and Bicycle Planning Guide to Best Practices," *Ongoing Guide*, 2000.
- [88] Weijermars et al. (2013), "Duurzaam Veilig, ook voor ernstig verkeersgewonden.," SWOV,

2013.

[89] Wei, F., "Boundary Effects in Developing Macro-level CPMs: A Case Study of City of Ottawa," 2010.

[90] Washington, S. et al., *Incorporating Safety into Long-Range Transportation Planning*. Washington, D.C.: Transportation Research Board, 2006.

Appendices

Appendix A

The geographic locations of the top 10 ranked CPZs are shown for each of the ten CPMs.

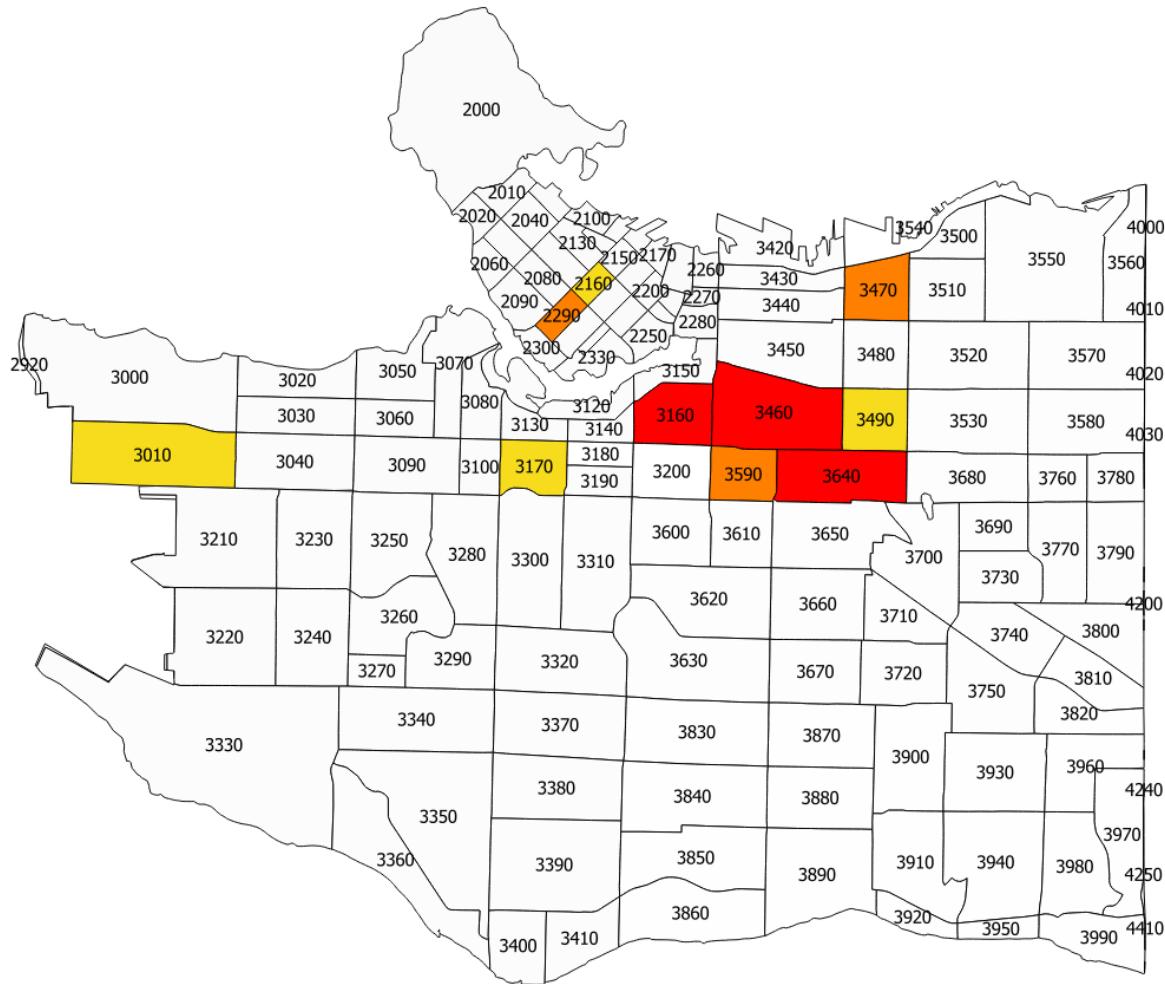


Figure 16: Top ten CPZs for the reference exposure model (BKT only)

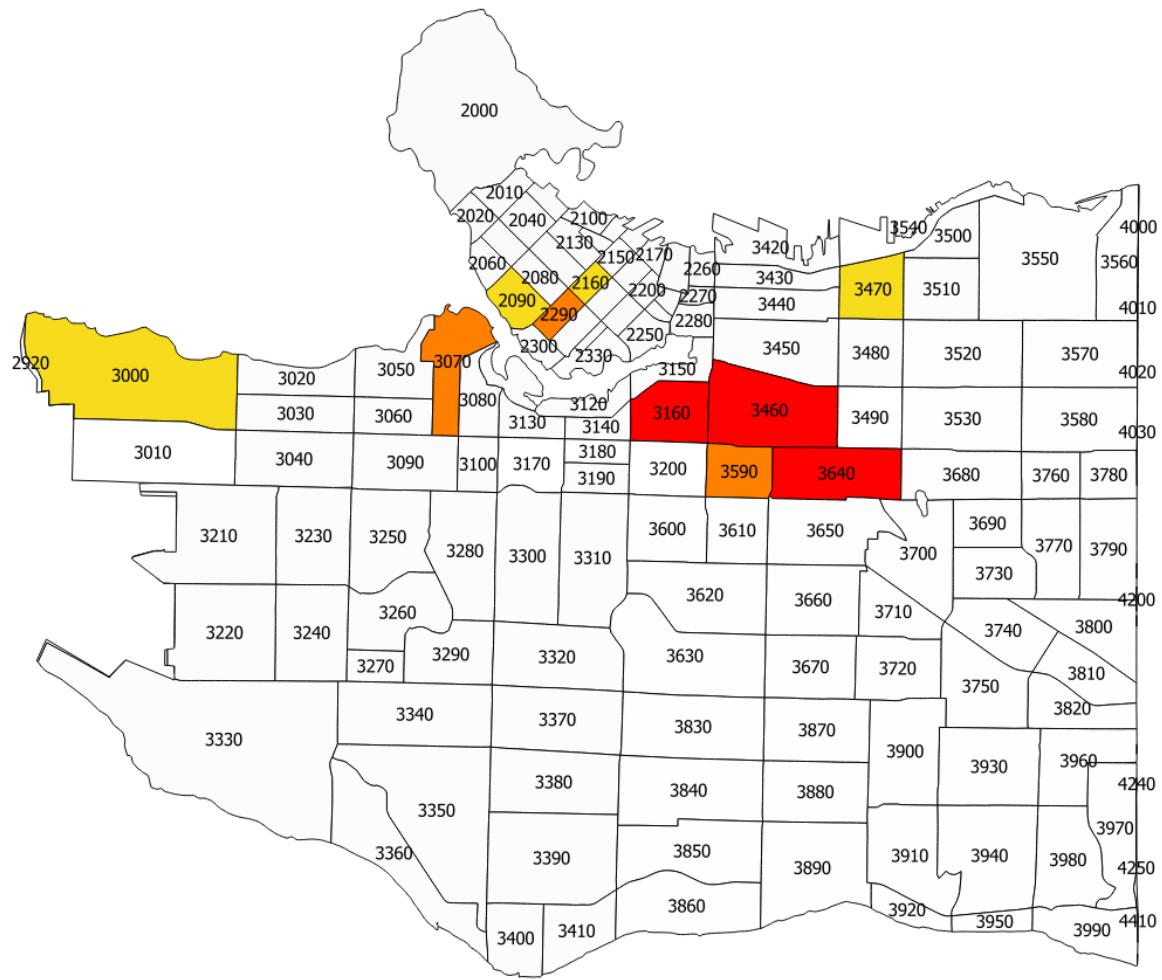


Figure 17: Top ten CPZs for an exposure model (AAAKM, PMP)

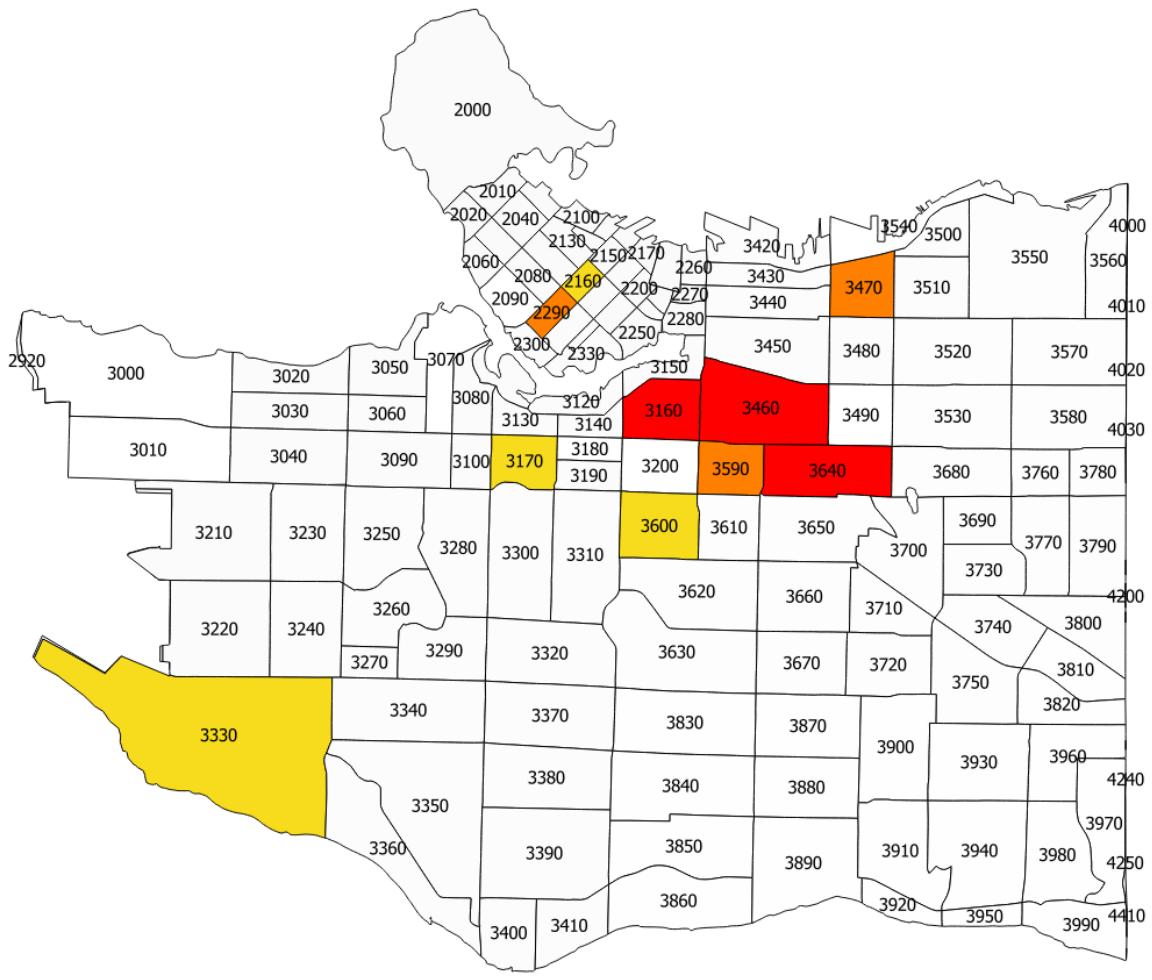


Figure 18: Top ten CPZs for a TDM model (TCM)

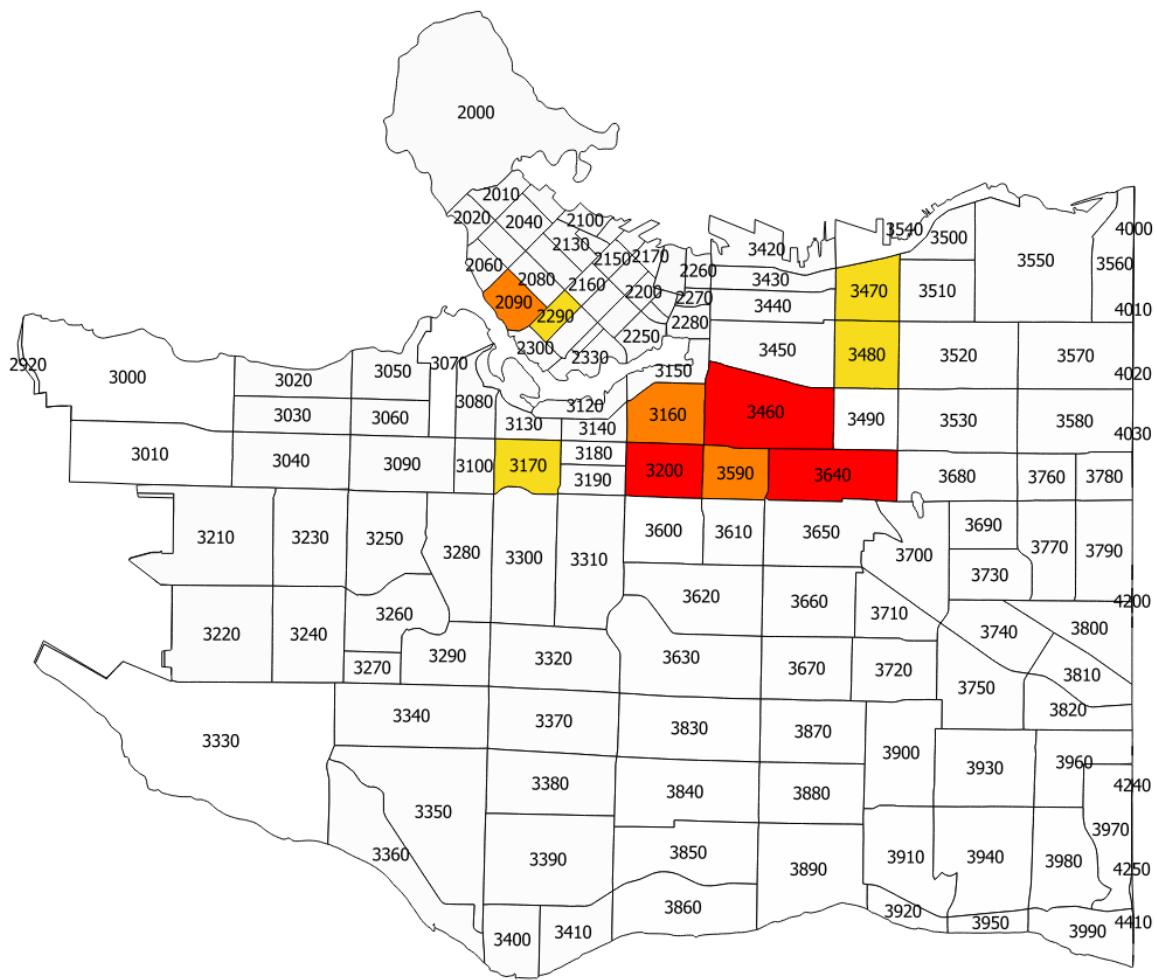


Figure 19: Top ten CPZs for a socio-demographic model (COM, EMP)

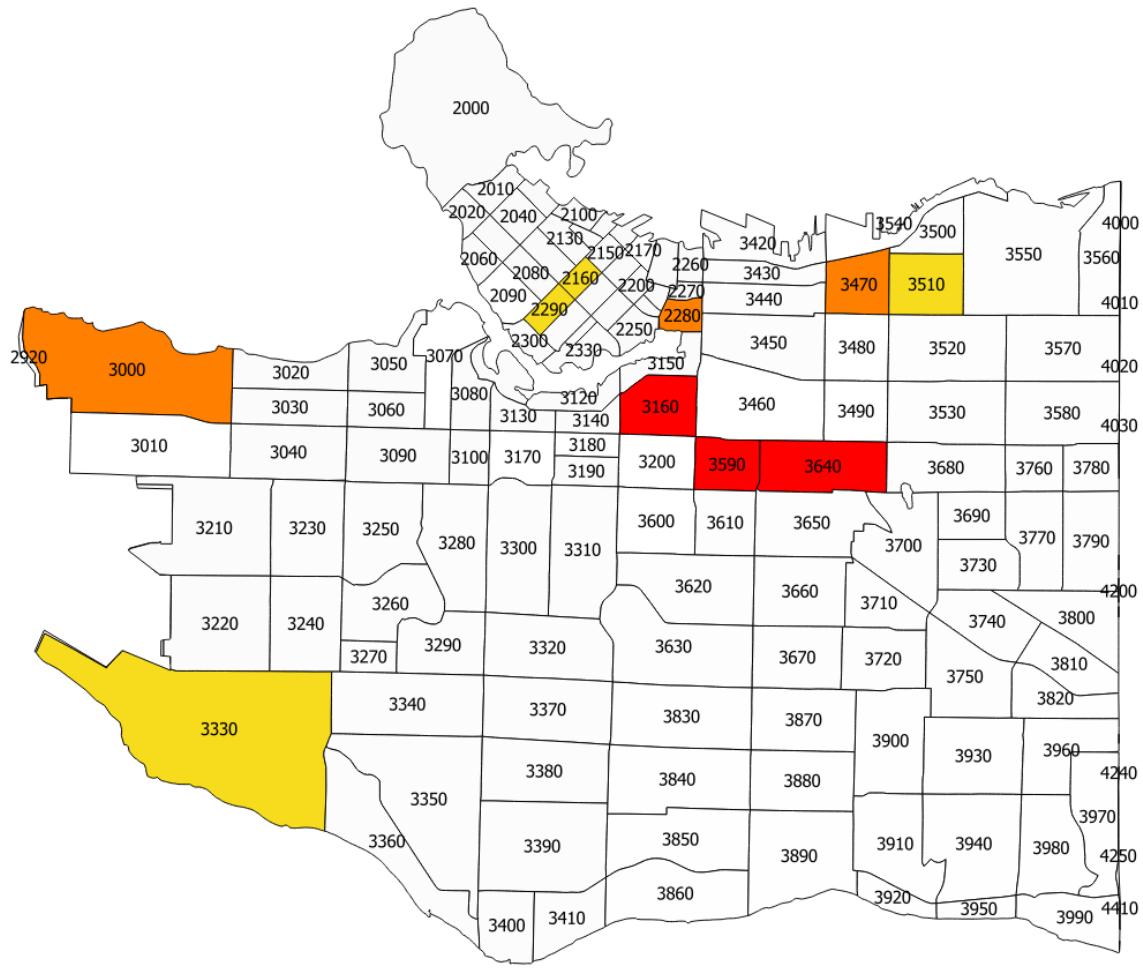


Figure 20: Top ten CPZs for a socio-demographic model (EMP, HHD)

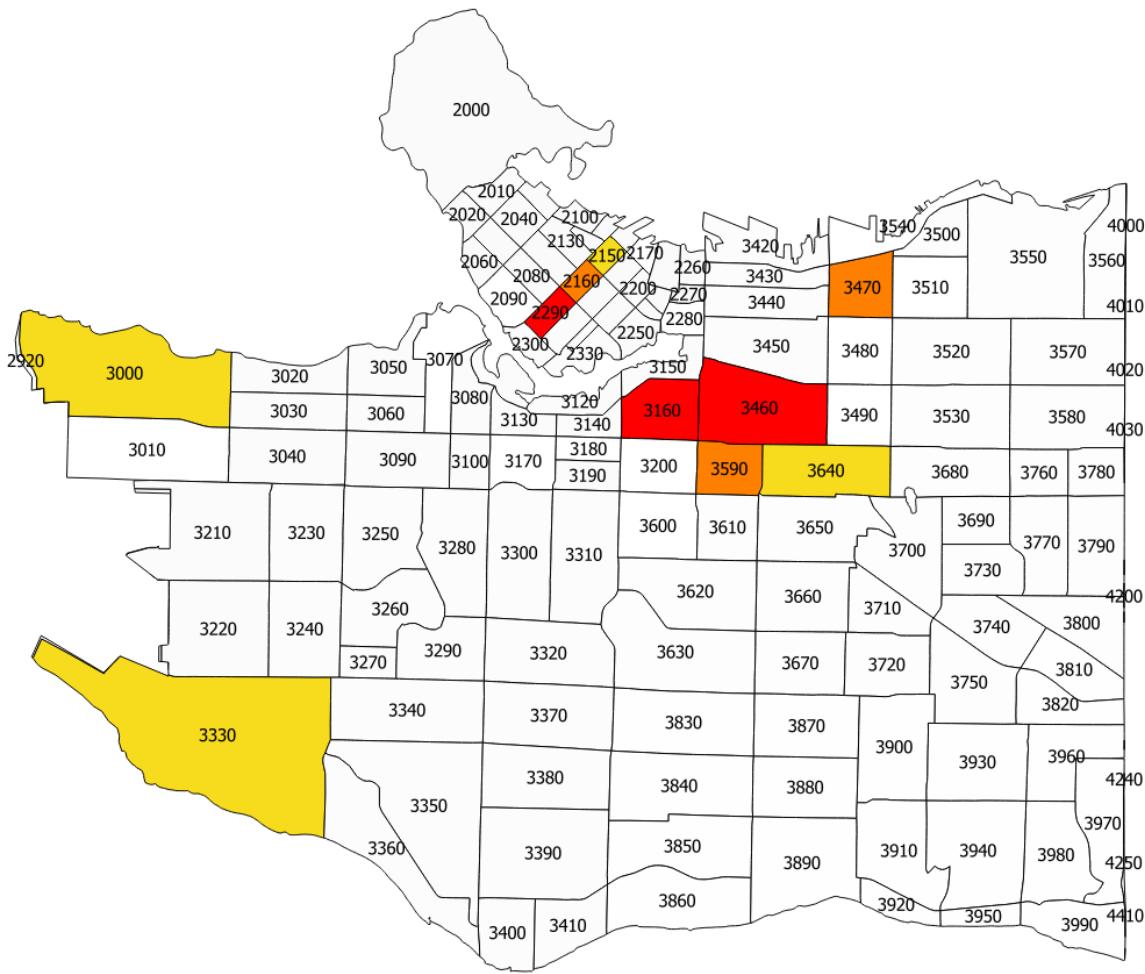


Figure 21: Top ten CPZs for a socio-demographic model (POP, INTD)

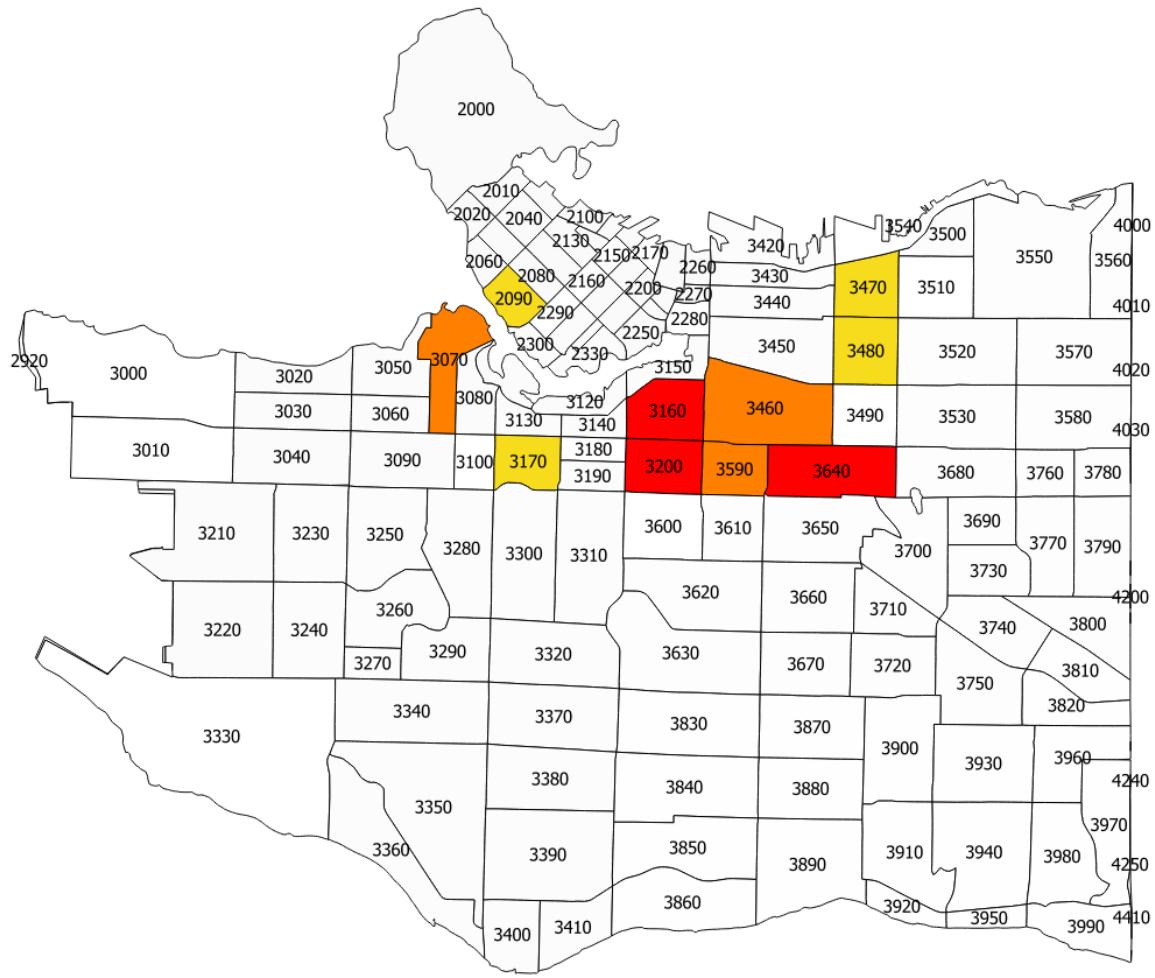


Figure 22: Top ten CPZs for a network model (SIGD, BS)

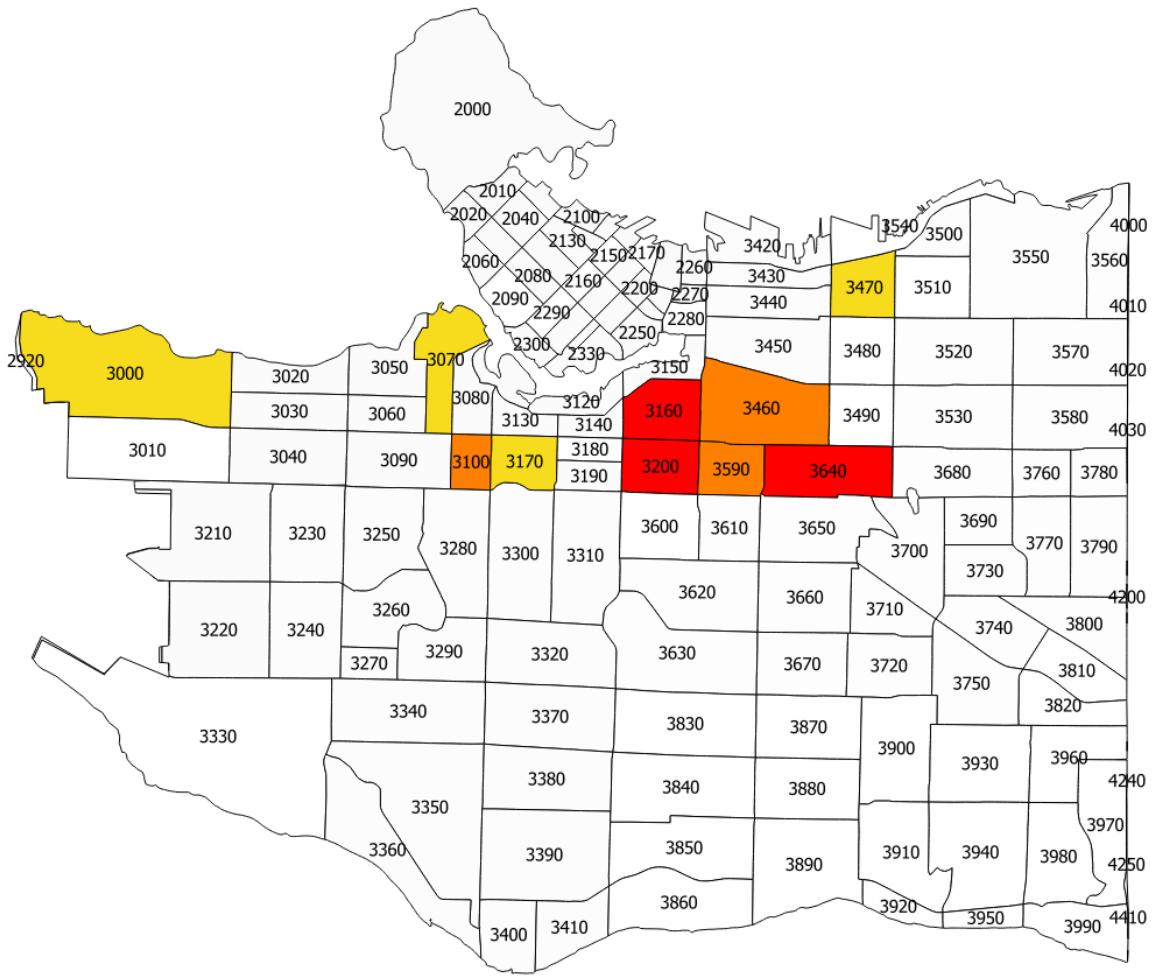


Figure 23: Top ten CPZs for a network model (PMP, AKM)

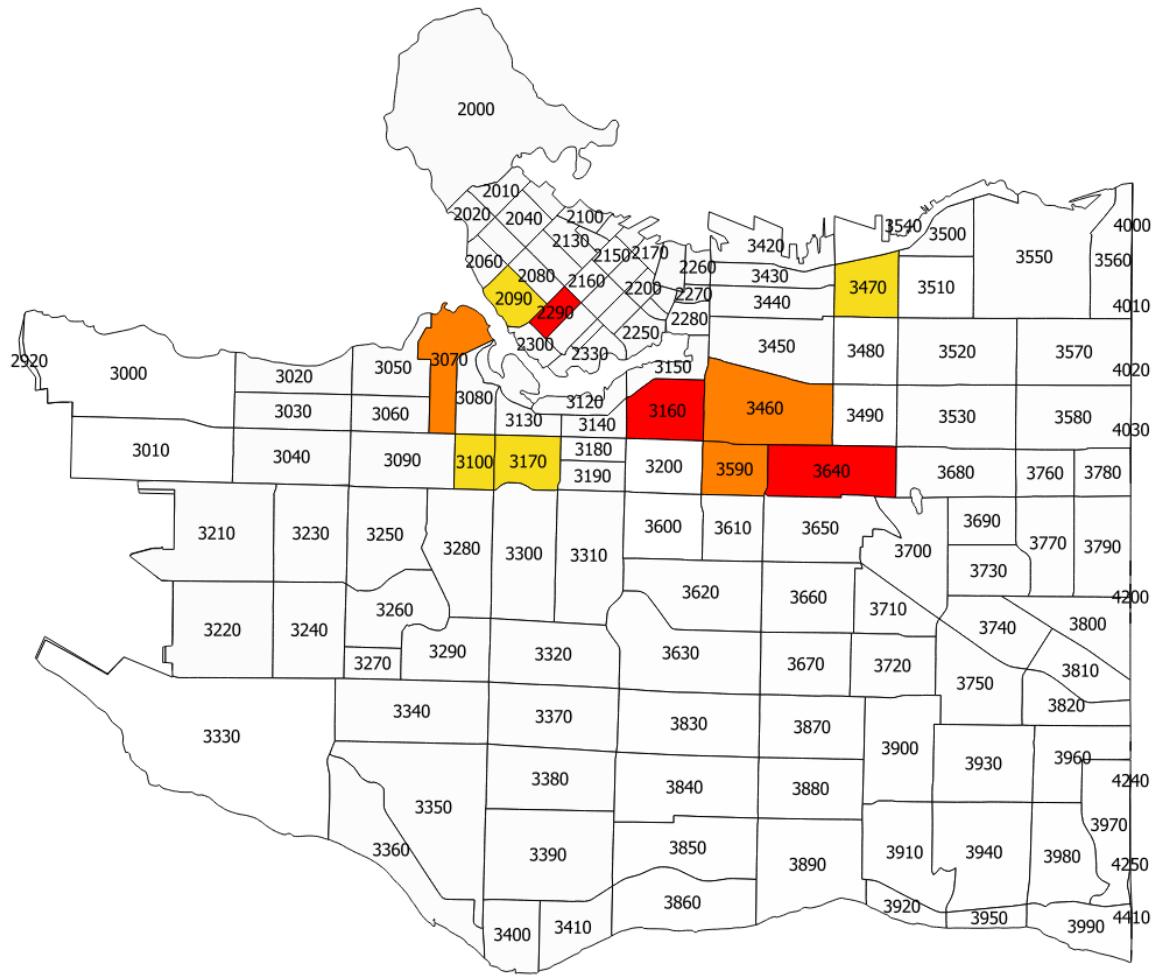


Figure 24: Top ten CPZs for a network model (INTD, BS)

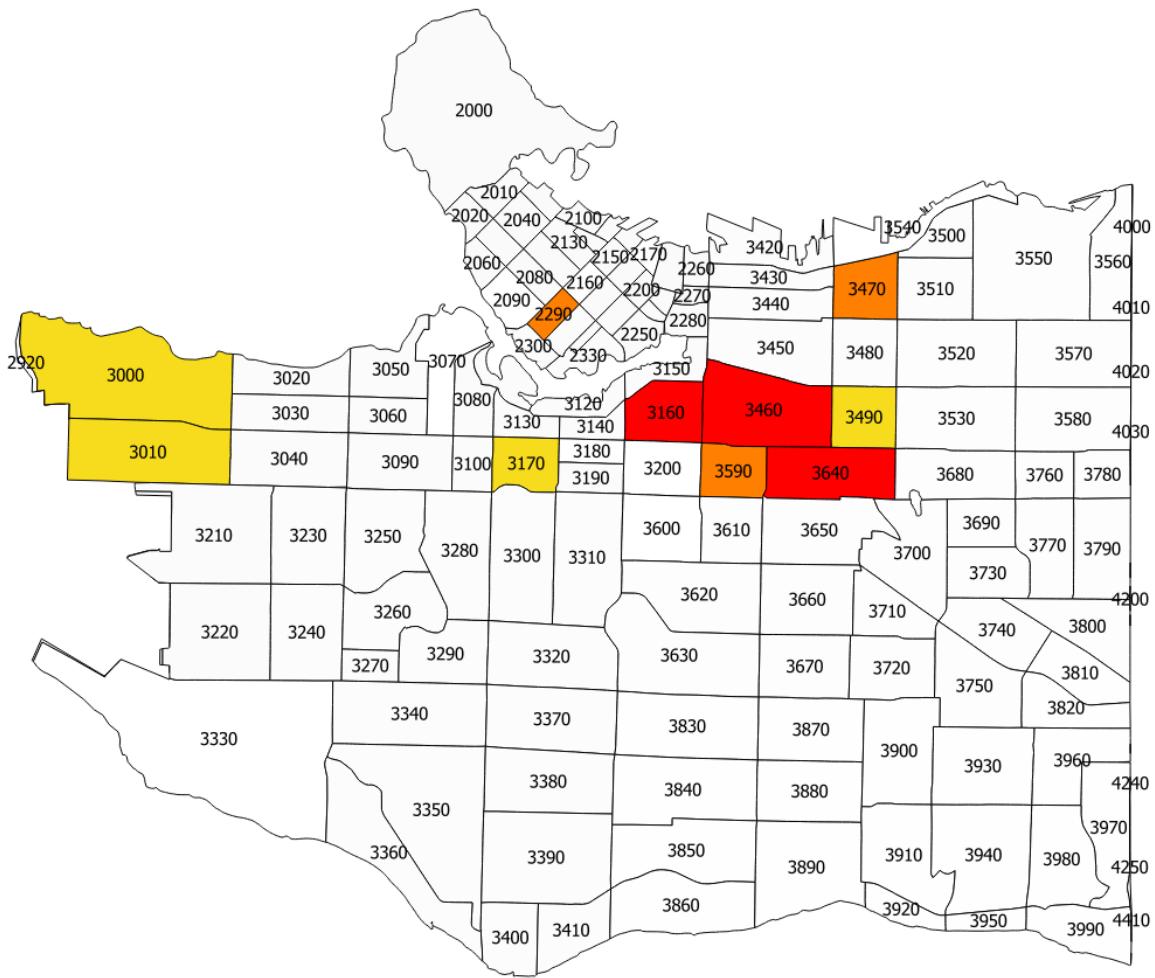


Figure 25: Top ten CPZs for a network model (RPKM, INTD)