# Effects of Spatial Correlation in Collision Modelling

by

**SUDIP BARUA**

**A thesis submitted in partial fulfillment of the requirements for the degree of**
**Master of Science**
**in**

**Transportation Engineering**

**Department of Civil and Environmental Engineering**

**University of Alberta**

# ABSTRACT

Despite previous research advocating the inclusion of spatially correlated random effects in order to significantly improve the estimation of the expected collision frequency, limited research efforts have been devoted to incorporating spatial correlation in both multivariate and random parameters collision modelling. Therefore, this thesis attempts to investigate the effects of including spatial correlation in three different collision modelling formulations: i) multivariate models, ii) univariate random parameters models, and iii) multivariate random parameters models. The models were developed using three years of collision data from the city of Richmond and the city of Vancouver. The proposed models were estimated in a Full Bayesian (FB) context via Markov Chain Monte Carlo (MCMC) simulation. The Deviance Information Criteria (DIC) and chi-square statistics were used to compare models and assess their goodness-of-fit, respectively. Models with spatial correlation yielded the best inference in terms of unbiased parameter estimates, precision, and capturing the multivariate nature of the collision data. Results showed significant and positive correlation between various road attributes and collision occurrence. A high percentage of the total variability was explained by the spatial correlation in most cases. This finding indicates that ignoring spatial correlation in collision modelling may lead to biased parameter estimation. The results also exhibit high and significant posterior correlation between severe and non-injury collisions for the total random effects (heterogeneous and spatial), indicating that a higher number of non-injury collisions is associated with a higher number of severe collisions. Furthermore, both multivariate spatial models and multivariate random parameters spatial models were compared against their univariate counterpart with respect to model inference and goodness-of-fit. Multivariate spatial models provide a superior fit over the two univariate spatial models, as demonstrated by a very significant drop in the DIC value. Similarly, multivariate random parameters spatial models outperformed the univariate random parameters spatial models.

# PREFACE

**Articles submitted to refereed journals**

1. **Barua, S.**, El-Basyouny, K., Islam, T., 2014. A Full Bayesian Multivariate Count-Data Model of Collision Severity with Spatial Correlation. Analytic Methods in Accident Research (in press).

2. **Barua, S.**, El-Basyouny, K., Islam, T., 2014. Effects of Spatial Correlation in Random Parameters Collision Count-Data Models. Under Review.

3. **Barua, S.**, El-Basyouny, K., Islam, T., 2014. Multivariate Random Parameters Collision Count-Data Models with Spatial Correlation. Under Review.

*Dedicated to my parents and sister*

## ACKNOWLEDGEMENTS

First of all, I wish to convey my profound gratitude to Buddha for enabling me to complete this research work successfully.

I would like to express my sincere appreciation, deepest gratitude and indebtedness to my supervisor, Dr. Karim El-Basyouny, Department of Civil and Environmental Engineering, University of Alberta, for his constant guidance, valuable suggestions, constructive criticism, meticulous help and financial support throughout the progress of the thesis work.

My sincere thanks are due to my MSc thesis defense committee members, Dr. Amy Kim (Chair), Dr. Zhi-Jun (Tony) Qiu and Dr. Ahmed Bouferguene, for their time and very helpful comments.

I would like to pay tribute and appreciation to Md. Tazul Islam for his continuous guidance, advice and support.

I would also like to acknowledge the help of my professors, Dr. Zhi-Jun (Tony) Qiu and Dr. Amy Kim, at the University of Alberta, who impacted my graduate study through their coursework. I am also grateful to and would like to thank Rochelle Borchman and Aalyssa Atley for all their technical writing support, aspiring guidance and time.

Further, I also owe thanks to Dr. Md. Mazharul Hoque, Dr. Hasib Mohammed Ahsan and Dr. Md. Hadiuzzaman, Department of Civil Engineering, Bangladesh University of Engineering and Technology (BUET), for fostering my transportation engineering background and providing essential support before and during my graduate study.

I express my warm thanks to all my colleagues and friends, especially, Rajib Sikder, Rokib SA and Ran Li, for their valuable suggestions, constructive criticism and support.

Last, but by no means least, I am grateful to my parents and sister for their continuous encouragement and support during my study.

# TABLE OF CONTENTS

Abstract ........ ii

Preface ........ iii

Acknowledgements ........ v

Table of Contents ........ vi

List of Tables ........ ix

List of Figures ........ x

1. Introduction ........ 1

  1.1 Background ........ 1

  1.2 Research Motivation ........ 5

  1.3 Research Objectives and Scope ........ 5

  1.4 Structure of the Thesis ........ 6

2. Literature Review ........ 8

  2.1 Previous Research ........ 8

    2.1.1 Multivariate Collision Models ........ 8

    2.1.2 Random Parameters Collision Models ........ 10

    2.1.3 Spatial Collision Models ........ 13

    2.1.4 Issues Related to Previous Research ........ 16

  2.2 Collision Modelling ........ 16

    2.2.1 Poisson Model ........ 16

    2.2.2 Negative Binomial Model ........ 17

    2.2.3 Poisson Lognormal Model ........ 18

    2.2.4 Multivariate Model ........ 19

    2.2.5 Univariate Spatial Model ........ 21

2.2.6 Multivariate Spatial Model ........ 21
2.2.7 Univariate Random Parameters Model ........ 22
2.2.8 Multivariate Random Parameters Model ........ 23
3. Data Description ........ 25
4. Multivariate Spatial Models ........ 29
4.1 Background ........ 29
4.2 Methodology ........ 31
4.2.1 Model Specification ........ 31
4.2.2 The Models ........ 33
4.2.3 Prior and Posterior Distributions ........ 33
4.2.4 Full Bayesian Estimation ........ 34
4.2.5 Comparison of Models and Goodness-of-Fit ........ 34
4.3 Results and Discussion ........ 36
4.3.1 Model Selection ........ 36
4.3.2 Parameter Estimates ........ 38
4.3.3 Variance and Correlation ........ 43
4.3.4 Comparison of Multivariate Models with Univariate Models ........ 44
4.4 Summary ........ 47
5. Univariate Random Parameters Spatial Models ........ 49
5.1 Background ........ 49
5.2 Methodology ........ 51
5.3 Results and Discussion ........ 52
5.3.1 Model Selection ........ 52
5.3.2 Parameter Estimates ........ 54
5.4 Summary ........ 59

6. Multivariate Random Parameters Spatial Models ........................................................ 61
6.1 Background ........................................................................................................ 61
6.2 Methodology ....................................................................................................... 62
6.3 Results and Discussion ....................................................................................... 64
6.3.1 Model Comparison and Parameter Estimates ................................................ 64
6.3.2 Comparison of Multivariate Models with Univariate Models ........................ 71
6.4 Summary ............................................................................................................. 72
7. Conclusions And Future research ......................................................................... 75
7.1 Concluding Remarks ........................................................................................... 75
7.2 Research Contributions ....................................................................................... 77
7.3 Limitations .......................................................................................................... 78
7.4 Future Research .................................................................................................. 78
References ................................................................................................................... 79

## LIST OF TABLES

Table 1: Statistical summary of Richmond dataset (n =72 road segments) .......... 27

Table 2: Statistical summary of Vancouver dataset (n =281 road segments) .......... 28

Table 3: Description of model covariates .......... 28

Table 4: The DIC statistics by model .......... 37

Table 5: Parameter estimates and 95% credible intervals for multivariate models (Models 1A-1C) (Richmond dataset) .......... 39

Table 6: Parameter estimates and 95% credible intervals for multivariate models (Models 1A-1C) (Vancouver dataset) .......... 42

Table 7: Parameter estimates and 95% credible intervals for univariate PLN models (Richmond dataset) .......... 45

Table 8: Parameter estimates and 95% credible intervals for univariate PLN spatial models (Vancouver dataset) .......... 46

Table 9: The DIC statistics by model .......... 53

Table 10: Parameter estimates and 95% credible intervals for Models 2A-2C (Richmond dataset) .......... 55

Table 11: Parameter estimates and 95% credible intervals for Models 2A-2C (Vancouver dataset) .......... 58

Table 12: The DIC statistics by model .......... 64

Table 13: Parameter estimates and 95% credible intervals for Model 3A .......... 66

Table 14: Parameter estimates and 95% credible intervals for Model 3B .......... 68

Table 15: Parameter estimates and 95% credible intervals for Model 3C .......... 69

Table 16: Parameter estimates and 95% credible intervals for univariate random parameters model with both heterogeneous effects and spatial correlation .......... 72

# LIST OF FIGURES

Figure 1: Illustration of the study area in the city of Richmond .................................................. 25

Figure 2: Illustration of the study area in the city of Vancouver ................................................ 26

Figure 3: Neighbouring structure definition ..................................................................... 27

# 1. INTRODUCTION

## 1.1 Background

Collision modelling is widely considered a key tool for estimating the safety levels of different road entities (i.e., intersections and road segments). Collision models are mathematical models statistically developed to link collision occurrence to a roadway's traffic and geometric characteristics. There are several key reasons that collision modelling is widely used in safety studies: collision models i) can be used under a Bayesian framework to address the regression-to-the-mean bias; ii) can address over-dispersion due to unobserved or unmeasured heterogeneity in collision data; iii) can account for the fundamental nonlinear relationship between collision frequency and traffic volume (AASHTO, 2010); iv) help analysts to understand the relationships between collisions and particular attributes (Greibe, 2003; Sawalha and Sayed, 2006; Hadayeghi et al., 2003; Manuel et al., 2014); v) help analysts to predict site-specific collisions and, hence, identify and rank road segments that are hazardous (Hauer, 1992; 1996; Hauer et al., 2002); and vi) can be used to evaluate the effectiveness of various safety countermeasures by facilitating the Empirical Bayesian (EB) and Full Bayesian (FB) approach (Yanmaz-Tuzel and Ozbay, 2010; El-Basyouny and Sayed, 2010; 2012a). Since collision modelling has been widely used in safety analysis, it is of paramount importance to continue improving the methodologies for developing these models, in an attempt to reduce the bias and inconsistent estimation, improve the precision of the estimates, thereby increasing the model's predictability.

Over the last two decades, considerable research efforts have been devoted in order to develop and apply sophisticated methodological approaches to account for several collision data-related issues (e.g., over-dispersion, under-dispersion, omitted-variables bias, fixed parameters, functional form). Regardless of these methodological innovations and developments, there are still several complex issues (e.g., unobserved heterogeneity, endogeneity, spatial and temporal correlation, correlated collision types—for more information, readers can refer to Lord and Mannering (2010) and Mannering and Bhat (2014)) that can substantially influence the inference, precision and findings from the collision data analysis. Fortunately, over the past few years, there have been substantial methodological developments to address these potential issues that include the following:

- Use of random parameters in collision models to capture unobserved heterogeneity across observations (Gkritza and Mannering, 2008; Milton et al., 2008; Anastasopoulos and Mannering, 2009; 2011; El-Basyouny and Sayed, 2009a; Dinu and Veeraragavan, 2011; Garnowski and Manner, 2011; Ukkusuri et al., 2011; Venkataraman et al., 2011; 2013; Wu et al., 2013; Xiong and Mannering, 2013; Chen and Tarko, 2014; Anastasopoulos et al., 2012a; Russo et al., 2014);
- Application of the multivariate modelling approach in collision analysis at different levels of classification (Maher, 1990; Bijleveld, 2005; Ma and Kockelman, 2006; Park and Lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009b);
- Use of two-state Markov Switching and finite-mixture or latent class models to analyze collision frequencies (Malyshkina et al., 2009; Park and Lord, 2009; Malyshkina and Mannering, 2010; Park et al., 2010; Zou et al., 2013; Shaheed and Gkritza, 2014; Zou et al., 2014);
- Inclusion of spatial correlation in collision models to capture unobserved effects, as neighbouring sites typically have similar environmental and geographical characteristics (Amoros et al., 2003; Noland and Quddus, 2004; Abdel-Aty and Wang, 2006; Aguero-Valverde and Jovanis, 2006; 2008; 2010; Quddus, 2008; El-Basyouny and Sayed, 2009c; Mitra, 2009; Flask and Schneider IV, 2013; Aguero-Valverde, 2013);
- Inclusion of temporal correlation in collision models to capture effects due to the collection of collision data over successive time periods (Lord and Persaud, 2000; Wang and Abdel-Aty, 2006; Wang et al., 2006); and
- Application of the zero-inflated modelling technique to overcome the excessive zeroes observed in collision data (Shankar et al., 1997; Lord et al., 2005; Lord et al., 2007).

Spatial correlation is one of the key issues that has been gaining attention in the development of collision models (Amoros et al., 2003; Noland and Quddus, 2004; Abdel-Aty and Wang, 2006; Aguero-Valverde and Jovanis, 2006; 2008; 2010; Quddus, 2008; Mitra, 2009; El-Basyouny and Sayed, 2009c; Aguero-Valverde, 2013; Flask and Schneider IV, 2013). As collision data is collected with reference to location, which is measured as points in space (Quddus, 2008), a spatial correlation exists between observations (LeSage, 1998). Further, as some of the unobserved factors related to collisions are likely to be correlated over space, there might be some possible correlation among neighbouring sites. According to Aguero-Valverde and Jovanis (2010), the inclusion of spatially correlated random effects significantly improves the precision of the estimates of the expected collision frequency for road segments. The inclusion of spatial effects has two main advantages: i) spatial correlation sites estimate the "pool strength" from neighbouring sites, thereby improving model parameter estimation (Aguero-Valverde and Jovanis, 2008); and ii) spatial dependence can be a surrogate for unknown and relevant covariates, thereby reflecting unmeasured confounding factors (Dubin, 1988; Cressie, 1993). Ignoring spatial correlation may lead to a biased, inconsistent and erroneous estimation of the model parameters. To this end, this thesis investigates the effects of including spatial correlation in different collision modelling approaches.

Generally, collision-related outcomes (e.g., fatal, injury, no injury) are considered as independent of each other and are often analyzed individually. However, the literature shows that collision types or severities have exhibited interdependencies (Maher, 1990; Bijleveld, 2005; Ma and Kockelman, 2006; Park and Lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009b). For instance, locations where fatal collisions occur are more likely to have a high number of less-fatal (i.e., severe or property damage) collisions due to the same deficiencies in roadway design, similar weather conditions and other unobserved factors. These correlations may be caused by omitted variables, which can influence collision occurrence at different severity levels, or they may be due to ignoring shared information in unobserved random effects. Typically, collision models are analyzed and estimated at different severity levels or types separately by using a single equation (univariate modelling) or a series of independently specified equations. Most of the literature is limited to investigating the effects of including spatial correlation in univariate collision models. While a number of studies explored the multivariate crash modelling to capture the heterogeneous correlations among different collision types or severities (Maher, 1990; Bijleveld, 2005; Ma and Kockelman, 2006; Park and Lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009b), multivariate spatial correlations were rarely investigated. Furthermore, univariate spatial modelling of different types of collision counts may lead to biased results because collision types or severities are not spatially independent of one another.

The parameters of these traditional univariate and multivariate collision models were assumed to be fixed when they actually vary across observations (road segments or intersections). Further, due to unobserved heterogeneity, the effects of explanatory variables on collisions may vary for different observations. For instance, a two-lane road segment with high traffic volume may have a high collision frequency compared to a similar road segment with less traffic volume. Therefore, constraining the parameters may not incorporate site specific effects, leading to an underestimation of standard errors leading to an inconsistent, biased and erroneous inference (Washington et al., 2003). Milton et al. (2008), Gkritza and Mannering (2008), Anastasopoulos and Mannering (2009; 2011), El-Basyouny and Sayed (2009a), Anastasopoulos et al. (2012a), and Russo et al. (2014) all demonstrated that the random parameters model can provide better inference compared to the traditional fixed parameters model and can explicitly account for heterogeneity across observations due to unobserved road geometrics, traffic characteristics, environmental factors, driver behaviour and other confounding factors. Most of the literature focused on developing random parameters models by taking unobserved heterogeneity into account and often ignored the likely random effects of spatial correlation. Ignoring this spatial correlation in random parameters collision models may reduce the predictive capability of the models, as some of the unobserved contributing and confounding factors are likely to be correlated over space.

Further, most of the literature employed random parameters within a univariate modelling framework. Multivariate random parameters were rarely explored in the literature. However, recently, El-Basyouny and Sayed (2013a), El-Basyouny et al. (2014a) and Dong et al. (2014) applied random parameters in multivariate collision modelling. Almost all of the very few multivariate random parameters collision models used heterogeneous effects in addition to random parameters to account for unobserved or unmeasured heterogeneity. However, as in univariate random parameters collision modelling, most of the studies ignored the effects of spatial correlation in multivariate random parameters collision models, which may lead to an incorrect parameters estimation of the model as there might be some possible correlation among neighbouring sites.

## 1.2 Research Motivation

A comprehensive review of literature on collision modelling revealed several methodological limitations and research gaps:

I. Most of the literature is limited to investigating the effects of including spatial correlation in univariate collision models. Only two studies (Song et al., 2006; Aguero-Valverde, 2013) focused on area-wide multivariate spatial modelling for collision severity and type. Multivariate spatial modelling for intersections or road segments was rarely explored in the literature.

II. Most of the literature is limited to investigating the application of a random parameters modelling approach in collision analysis. Though the literature has suggested that spatial correlation could reduce bias (produced by the omission of spatial variables) when estimating the regression coefficients, the inclusion of spatial correlation in both univariate and multivariate random parameters collision models has rarely been explored in the literature.

Therefore, this thesis attempts to investigate the effects of including spatial correlation in three different collision modelling formulations: i) multivariate models, ii) univariate random parameters models, and iii) multivariate random parameters models.

## 1.3 Research Objectives and Scope

Considering the methodological limitations of previous studies and given the magnitude of the potential issue regarding spatial correlation, three objectives have been set in this thesis.

***Objective 1:*** Use the multivariate spatial modelling approach to develop spatial models for road segments in order to assess spatial correlation at different collision severity levels and its influence on the collision analysis of urban arterials. Further, compare multivariate spatial models with independent (separate) univariate spatial models for each collision severity in terms of model inference and goodness-of-fit.

***Objective 2:*** Investigate the effects of spatial correlation in univariate random parameters collision count-data models.

***Objective 3:*** Include spatial correlation in multivariate random parameters collision severity models and assess the effects in terms of model inference, precision and goodness-of-fit. In addition, compare multivariate random parameters spatial models with their univariate counterpart.

From a model application perspective, the methodological approach proposed herein can be used to estimate the associated safety risks and the expected collision frequency more precisely by considering spatial correlations in the data. Nevertheless, multivariate models and random parameters models are complex to estimate, and the inclusion of spatial correlation makes the estimation technique even more complex, the ultimate objective of this thesis is to explore novel methodological approaches that have the potential to provide new insight in collision data analysis.

## 1.4 Structure of the Thesis

The remainder of this thesis is organized into the following chapters:

*Chapter 2* discusses previous research efforts devoted to developing multivariate models, random parameters models and spatial models. This chapter also points out the limitations and gaps in the literature regarding incorporation of spatial correlation in different methodological approaches. In addition, an overview of different collision modelling approaches is also provided.

*Chapter 3* describes the datasets used in this thesis. Three years (1994–1996) of collision data and other geometric and non-geometric road data from two cities in British Columbia, Canada: i) Richmond, and ii) Vancouver, along with the preparation of a neighbour matrix to incorporate spatial correlation, are described.

*Chapter 4* investigates the inclusion of spatial correlation at different collision severity levels, using the multivariate modelling approach, and its influence on the collision analysis of urban arterials.

*Chapter 5* illustrates the effects of including spatial correlation in univariate random parameters collision count-data models.

*Chapter 6* describes the effects of including spatial correlation in multivariate random parameters collision models at different collision severity levels. In addition, this chapter compares multivariate random parameters spatial models with their univariate counterpart.

Finally, *Chapter 7* discusses research conclusions, research contributions, limitations and suggestions for future research.

# 2. LITERATURE REVIEW

This chapter discusses the previous research efforts devoted to developing multivariate models, spatial models and random parameters models. In addition, an overview of different collision modelling approaches is also provided.

## 2.1 Previous Research

As the focus of this thesis is to investigate the effects of including spatial correlation in multivariate collision models and random parameters collision models, a comprehensive review of three modelling approaches (multivariate models, random parameters models, spatial models) is given in this section. After providing a brief summary of previous studies, this section concludes with the issues and gaps in research on the inclusion of spatial correlation in collision modelling.

### 2.1.1 Multivariate Collision Models

Multivariate collision modelling has been gaining attention in safety analysis over the past few years. Maher (1990) was the first to apply the multivariate modelling approach in collision analysis. However, the author focused on explaining traffic collision migration, rather than explaining the correlations that exist among different collision severities/types. After the study of Maher (1990), no studies used the multivariate modelling approach in collision analysis until the study of Ladron de Guevara et al. (2004). The study of Ladron de Guevara et al. (2004) used the simultaneous negative binomial model to forecast collisions for traffic analysis zones at the planning level. The authors found a significant high correlation between fatal and injury collisions. Bijleveld (2005) was the first to propose the structure of variance and covariance of the outcomes. However, the method proposed by Bujleveld (2005) estimates covariance only within observations and ignores covariance between observations. Song et al. (2006) had also used multivariate spatial models to explore correlated collision types at county level. Although summary statistics of the correlation coefficient between the responses were not provided, the plots of the posterior distribution of the coefficient indicated possible significant correlation.

Ma and Kockelman (2006) introduced multivariate Poisson specification to simultaneously model injuries by severity. However, the model specification proposed in that study relied on a one-way covariance structure and assumed the presence of an added constant across all count types, which implies that the covariances are non-negative and identical within the segment and that within segment covariances are the same across segments. Further, the model specification does not allow for over-dispersion. Consequently, Park and Lord (2007) extended the specification proposed by Ma and Kockelman (2006). The authors took overdispersion into account and included extra-Poisson variation and relaxed the covariance structure with a different covariance among severity levels by including a lognormal distribution for extra Poisson variations. Similarly, Ma et al. (2008) used multivariate Poisson lognormal specification to model collision severities. The study of Aguero-Valverde and Jovanis (2009) and El-Basyouny and Sayed (2009b) used Bayesian multivariate Poisson lognormal models for collision severity modelling and site ranking. Aguero-Valverde and Jovanis (2009) quantified the effects of using multivariate structures on the precision of the estimates of collision frequency, while El-Basyouny and Sayed (2009b) demonstrated that if the analysis was restricted to univariate models, some of the hazardous locations could be overlooked while identifying hazardous locations. Another study of El-Basyouny and Sayed (2013a) also proposed an alternative method, a depth-based multivariate method for the identification and ranking of hotspots under a full Bayesian framework.

Unlike the traditional multivariate model specification, the study of Wang et al. (2011) proposed an alternative method to estimate collision frequency at different severity levels, namely the two-stage mixed multivariate model, which combines both collision frequency and severity models. The two-stage mixed multivariate model was comprised of a Bayesian spatial model and a mixed logit model for collision frequency and severity analysis respectively. The authors advocated that the two-stage mixed multivariate model is a promising tool in predicting collision frequency according to their severity levels and site ranking. The multivariate modelling approach was also used in modelling animal-vehicle collisions (Lao et al., 2011). The authors use diagonal inflated bivariate Poisson regression to perform the analysis. To assess the collision rates at different severity levels, Anastasopoulos et al. (2012b) proposed a multivariate Tobit model that can address the possibility of differential censoring across injury-severity levels, while also accounting for the potential contemporaneous error correlation resulting from commonly shared unobserved characteristics across roadway segments. The authors found that the multivariate Tobit model outperformed its univariate counterpart and is practically equivalent to the multivariate negative binomial model. Narayanamoorthy et al. (2013) applied the multivariate modelling approach to assess bicycle and pedestrian injury severity. The authors proposed a new spatial multivariate count model to jointly analyze traffic collision-related counts of pedestrians and bicyclists by injury severity at the census tract level. A recent study by El-Basyouny et al. (2014a; b) used multivariate Poisson lognormal models to assess the effects of weather elements and states on collision severity levels and collision types. Several studies (Park et al., 2010; El-Basyouny and Sayed, 2011; El-Basyouny et al., 2012b; El-Basyouny and Sayed, 2013b) also applied a multivariate modelling approach in before-after safety evaluation.

### 2.1.2 Random Parameters Collision Models

Despite the fact that the random parameters model outperformed traditional fixed parameter models, limited research used this approach in safety research as random parameters estimation techniques i) are very complex, ii) are less convenient for engineering purposes, iii) lack an estimation tool for large samples (Chen and Tarko, 2014), and iv) lack transferability to other datasets (Shugan, 2006; Lord and Mannering, 2010; Washington et al., 2010). However, over the past few years, random parameters modelling has been gaining attention in safety analysis. Milton et al. (2008) and Gkritza and Mannering (2008) were the first to apply the random parameters modelling approach to traffic safety. Milton et al. (2008) used a mixed logit (random parameters) model to assess the highway collision severities (i.e., property damage only, possible injury and injury collisions). The authors found that the volume-related variables (e.g., average daily traffic per lane, average daily truck traffic, truck percentage, interchanges per mile) and weather effects (e.g., snowfall) are best modelled as random parameters, while roadway characteristics (e.g., the number of horizontal curves, number of grade breaks per mile and pavement friction) are best modelled as fixed parameters. Further, the author suggested that the mixed logit model holds considerable promise as a methodological tool in highway collision modelling. Unlike the study of Milton et al. (2008), Gkritza and Mannering (2008) employed a mixed logit approach for the safety analysis of seat belt use in single and multi-occupant vehicles. The study also suggested that the mixed logit (random parameters) modelling approach could provide a much fuller understanding of the interactions of the numerous variables that correlate with safety belt usage compared to traditional discrete-outcome modelling approaches. Furthermore, this modelling approach offers methodological flexibility to capture individual-specific heterogeneity that can arise from a number of factors relating to roadway characteristics, driver behaviour and vehicle types.

Several studies contributed to the methodology by employing random parameters in different collision modelling approaches. For instance, Anastasopoulos and Mannering (2009) first proposed the random parameters count-data model as an alternative methodological approach to analyze collision frequencies. The authors compared random parameters models with the fixed parameter negative binomial model. The findings indicated that ignoring the possibility of random parameters can result in substantially different marginal effects and subsequent inferences related to the magnitude of the effect of factors, affecting the precision of the calculated collision frequencies. The authors also found that random parameters count-data models have the potential to provide a fuller understanding of the factors determining collision frequencies. Unlike the previous study of Anastasopoulos and Mannering (2009), El-Basyouny and Sayed (2009a) focused on a collision count-data model incorporating random corridor parameters, clustering 392 road segments into 58 corridors under a FB framework. Three modelling approaches (i.e., fixed parameters PLN models, random effects model and random parameters PLN model) were compared in terms of inference and goodness-of-fit. The authors found some strong evidence for the benefit of clustering road segments into homogeneous groups (e.g., corridors) and incorporating random corridor parameters in collision modelling. The authors concluded that this modelling approach can be used to gain new insights into how the covariates affect collision frequencies and to account for unobserved heterogeneity.

Another study by Anastasopoulos and Mannering (2011) assessed fixed and random parameters logit models using five years of collision and non-collision specific injury data. Three severity outcomes were considered (i.e., no injury, injury and fatality) in that study. The analysis demonstrated that random parameters models using less detailed data can provide a reasonable level of accuracy, thus providing support for the statistical superiority of the random parameters logit model over the fixed parameter logit model. The findings also showed that individual collision data provided a better overall fit, relative to the models based on the proportion of collision by severity types. The study of Anastasopoulos et al. (2012a) used random parameters Tobit models to investigate the factors affecting highway collision rates in urban interstate roads in Indiana. The empirical results showed that the random parameters Tobit model outperformed the fixed parameter Tobit model. According to parameter estimation, the random parameters Tobit models provided a superior fit with 11 variables producing statistically significant parameters, compared to only six in the case of traditional fixed parameter Tobit models.

Recently, Xiong and Mannering (2013) proposed a finite mixture random parameters model to investigate the heterogeneous effects of guardian supervision on adolescent driver injury severities. The methodological approach in that study can be used to examine heterogeneous populations and has the potential to provide new insights in collision analysis. Russo et al. (2014) used random parameters in a bivariate ordered probit model to compare the factors affecting injury severity in angle collisions by fault status. The proposed methodological approach allows for consideration of within-collision correlation, as well as unobserved heterogeneity, and results in a significantly better fit than a series of independent fixed parameter models.

Several studies applied the random parameters modelling approach to perform safety analyses. For instance, Dinu and Veeraragavan (2011) employed the random parameters modelling approach to develop a count-data model for daytime and night time collisions on two-lane undivided rural highways in India that operate under mixed traffic conditions. Similarly, Ukkusuri et al. (2011) applied this approach to explain the effects of built environmental characteristics on pedestrian collision frequencies at the census tract level. Several parameters in the model were found to be random, which indicated their heterogeneous influence on the numbers of pedestrian collisions. Thus, the authors concluded that the random parameters modelling approach allows the incorporation of unobserved heterogeneity across the spatial zones. The study of Venkataraman et al. (2011) employed a random parameters negative binomial modelling approach to analyze nine years of collision count data on interstate highways in Washington State. The models were designed to account for parameter correlations, panel effects that contributed to intra-segment temporal variations, and effects between sites. Similarly, Garnowski and Manner (2011) used a negative binomial random parameters model to determine the factors related to collision frequencies on a set of 197 ramps. The authors found that a negative binomial model with random parameters proved to be an appropriate model in the cross-sectional setting for detecting factors related to collisions. Another study of Venkataraman et al. (2013) employed a negative binomial random parameters model to assess interstate collision frequencies. A total of 21 models were evaluated in four ways, by i) severity, ii) number of vehicles involved, iii) collision types and iv) location characteristics. Parameter estimation indicated some improvements in likelihood in 19 of the 21 models due to some parameters being random. Further, random parameters results contributed to a better likelihood compared to the baseline fixed-effect negative binomial models. Wu et al. (2013) also used the same methodological approach with the addition of a nested logit model for collision injury severity to assess the safety impacts of signal warning flashers and speed control at high-speed signalized intersections. The recent study of Chen and Tarko (2014) used random parameters and random effects models (only the intercept is randomly distributed; the other parameters are fixed) to analyze work zone safety. The results indicated that the marginal effects on collision frequency computed from the random effects model were quite similar to those of the random parameters model. The authors also suggested that the random effects model could be used as a convenient and practical alternative to the random parameters model.

### 2.1.3 Spatial Collision Models

Conventional collision models with Poisson-gamma or Poisson lognormal (PLN) distribution assume that sites are independent of one another and, hence, can be regarded as non-spatial models. However, as collision data is collected with reference to location, which is measured as points in space (Quddus, 2008), a spatial correlation exists between observations (LeSage, 1998). Ignoring spatial correlations may lead to a biased estimation of the model parameters. Therefore, a number of road element-specific (i.e., intersection or road segment) and area-wide collision studies have incorporated spatial correlation in modelling collision data (Aguero-Valverde and Jovanis, 2006; 2008; 2010; Aguero-Valverde, 2013).

In the context of intersection-based spatial models, Abdel-Aty and Wang (2006) used Generalized Estimating Equations (GEE) to address the spatial correlation between signalized intersections. The authors determined that signalized intersections, especially ones close together along a certain corridor, are spatially correlated and influence one another. Similarly, Guo et al. (2010) applied Poisson and negative binomial Bayesian models with spatial correlation to signalized intersections and found that Poisson spatial models are the best fit. Mitra (2009) proposed a Geographic Information System (GIS)-based method to detect collision hotspots and a spatial regression method with heterogeneous and spatial random effects in continuous space to investigate the intersection-level factors that influence fatal and injury collisions. The author concluded that models that include spatial correlation may potentially reduce the bias associated with model misspecification by changing the estimate of the annual average daily traffic (AADT) parameter. The results also indicated that spatial correlation is quite significant in cases of collisions involving minor injury or non-injury collisions. Recently, Castro et al. (2012) proposed a latent variable-based generalized ordered response framework for count-data models, which can efficiently introduce temporal and spatial dependencies through the latent continuous variables.

Aguero-Valverde and Jovanis (2008) explored the effects of spatial correlation in collision count models for rural road segments. The FB hierarchical approach with CAR effects for spatial correlation was used in that study. The results indicated that the model with spatial correlation was a significantly better fit to the data than the PLN model with only heterogeneity. Another study by Aguero-Valverde and Jovanis (2010) investigated spatial correlation in multi-level collision frequency models for different types of urban and rural road segments. The study employed the FB hierarchical approach with CAR distribution for the spatial correlation terms. The authors concluded that spatial correlation substantially increased the random effects. Results indicated that 70% to 90% of the variation explained by random effects resulted from spatial correlation. This suggests that spatial models offer a significant advantage because poor estimates resulting from small sample sizes and low sample means are frequent issues in highway collision analysis. Ahmed et al. (2011) explored a Bayesian hierarchical approach with spatial (CAR) and heterogeneous effects to develop collision models for 20 miles of mountainous freeway in order to rank the hazardous roadway segments. El-Basyouny and Sayed (2009c) compared CAR, multiple membership (MM) and extended multiple membership (EMM) models with traditional PLN models to investigate the inclusion of spatial correlation in collision prediction models. The authors found that the fitted CAR and MM models had significant estimates for both heterogeneity and spatial parameters. Furthermore, in terms of goodness-of-fit, the EMM model fit best with the data, followed by the CAR model. Castro et al. (2013) used a spatial generalized ordered response model to examine highway collision injury severity. The authors proposed a flexible econometric structure for injury severity analysis at the level of individual collisions that recognizes the ordinal nature of injury severity categories. The authors accommodated spatial dependencies in the injury severity levels experienced in collisions that occur close to one another in space.  The recent study of Chiou et al. (2014) used spatial multinomial generalized Poisson models to investigate spatial dependency. The authors found that spatial dependencies sharply decreased at distances exceeding seven kilometres and that shorter segments with high collision frequency tended to have high spatial dependence.

A number of studies have developed area-wide spatial models to demonstrate the relationship between collision occurrence and numerous socio-demographic, road network, transportation demand, and exposure variables. The unit of analysis varied from one study to another. For instance, Aguero-Valverde (2013) used canton as the unit to develop a multivariate CAR model. Aguero-Valverde and Jovanis (2006); Amoros et al. (2003); Flask and Schneider IV (2013); Huang et al. (2010); Song et al. (2006); and Van Schalkwyk (2008) used county as the specific unit of analysis. Noland and Quddus (2004) and Quddus (2008) used census wards to conduct a spatially disaggregated accident analysis of road casualties in England. Hadayeghi et al. (2003; 2007; 2010); Karim et al., (2013); Siddiqui et al., (2012); Wang et al., (2012); and Wei and Lovegrove (2013) applied spatial models to traffic analysis zones (TAZs), the unit often used by transportation planners. The study of Miaou and Song (2005) focused both on intersections and road segments as well as the county level for analysis. The authors demonstrated that the inclusion of a spatial component in the crash prediction model significantly improved the overall goodness-of-fit performance of the model and affected the ranking results.

Very few studies have focused on the multivariate spatial modelling approach in collision analysis. Song et al., (2006) used multivariate spatial models for collision mapping. Bayesian multivariate CAR models were used for four types of collisions (intersection collisions, intersection-related collisions, driveway-related collisions, and non-intersection-related collisions) using collision data from Texas. The authors proposed spatial priors for the Bayesian multivariate hierarchical models and sufficient conditions to ensure posterior propriety under a non-informative prior. Recently, Aguero-Valverde (2013) used the multivariate spatial modelling approach for excess collision frequency and severity in cantons (counties) for Costa Rica. The author advocated that the multivariate spatial model performed better than the univariate spatial models. The author also emphasized that the effects of spatial smoothing due to multivariate spatial random effects were evident in the estimation of excess equivalent non-injury collisions. Similarly, Wang and Kockelman (2013) assessed multivariate spatial effects, focusing on pedestrian collisions. The authors used Poisson-based multivariate CAR models for pedestrian collisions across census tracts in Austin, Texas, and found positive spatial autocorrelations across different neighbourhoods for pedestrian collisions. The study of Narayanamoorthy et al. (2013) also proposed a spatial multivariate count model to jointly analyze the traffic collision-related counts of pedestrians and bicyclists by injury severity. The modelling framework was applied to predict injury counts at a census tract level. The results demonstrated the need to use a multivariate modelling system for the analysis of injury counts by road-user type and injury severity level, while also accounting for spatial dependence effects in injury counts.

#### 2.1.4 Issues Related to Previous Research

Most of the literature is limited to investigating the effects of including spatial correlation in univariate collision models. Only two studies (Aguero-Valverde, 2013; Song et al., 2006) focused on area-wide multivariate spatial modelling for collision severity and type. Further, Wang and Kockelman (2013) and Narayanamoorthy et al. (2013) used the multivariate spatial modelling approach for pedestrian and bicyclist collision analysis at the census tract level. Multivariate spatial modelling for intersections or road segments was rarely explored in the literature. In terms of random parameters modelling approach, most of the literature is limited to investigating the application of the random parameters modelling in collision analysis incorporating site specific unobserved heterogeneity. Though the literature has suggested that spatial correlation could reduce bias (produced by the omission of spatial variables) when estimating the regression coefficients, the inclusion of spatial correlation in both univariate and multivariate random parameters collision models has rarely been explored in the literature.

### 2.2 Collision Modelling

This section presents an overview of different modelling specifications that are used in collision analysis. However, due to the diversity of models being handled in this thesis, each individual chapter has a more detailed model specifications.

#### 2.2.1 Poisson Model

As collisions are discrete, nonnegative and random events, the Poisson distribution is commonly used to develop the collision model. Let $Y_i$ denote the number of collisions at road segment $i\ (i = 1,2,3,...,n)$. Assume that the number of collisions at $n$ road segments is independent and that

$$Y_i \mid \theta_i \sim Poisson\ (\theta_i) \qquad (1)$$

Where, $\theta_i$ is the Poisson parameter. The probability of road segment $i$ having $y_i$ collisions is given by

$$\Pr\{Y_i = y_i \mid \theta_i\} = \frac{e^{-\theta_i}\theta_i^{y_i}}{y_i!} \qquad (2)$$

The Poisson parameter $\theta_i$ is commonly specified as an exponential function of road segment-specific attributes, such as exposure, traffic and geometric characteristics (Miaou and Lord, 2003), and is usually expressed as

$$\theta_i = \exp(X_i^{'}\beta) \qquad (3)$$

Where, $X_i^{'}$ is a row vector of covariates representing segment-specific attributes and $\beta$ is a vector of regression parameters to be estimated from the data. In the Poisson regression model, it is assumed that the mean and variance of the count variables are constrained to be equal, such that

$$E(Y_i) = Var(Y_i) = \theta_i \qquad (4)$$

However, when modelling collision count datasets, this assumption is often violated, as most collision data is likely to be over-dispersed (the variance is greater than the mean) (Kulmala, 1995; Cameron and Trivedi, 1998; Winkelmann, 2003). A Poisson distribution for over-dispersed data can underestimate the standard errors of the regression coefficients, which can lead to inflated values of the *t*-test, thereby affecting the significance level of the model regression coefficients. This leads to an incorrect selection of covariates, resulting in poor model fit.

### 2.2.2 Negative Binomial Model

To overcome the problems associated with the Poisson regression models, several researchers proposed the use of the Poisson-Gamma (PG) hierarchy, leading to the Negative Binomial (NB) regression model (Cameron and Trivedi, 1998). The main reason for the extensive use of this model is that it is simple to compute, since the Gamma distribution is a conjugate prior to Poisson, leading to a Gamma posterior distribution, which considerably simplifies the posterior analysis. To address over-dispersion for unobserved or unmeasured heterogeneity, it is assumed that

$$\theta_i = \mu_i \exp(\varepsilon_i) \qquad (5)$$

$$\ln(\theta_i) = \ln(\mu_i) + \varepsilon_i \qquad (6)$$

Where, $\mu_i$ is an exponential function of segment-specific attributes, such as exposure, traffic and geometric characteristics:

$$\ln(\mu_i) = \beta_0 + \sum_{m=1}^{M} \beta_m X_{mi} \qquad (7)$$

Where, $X_{mi}$ denotes the matrixes of covariates (relevant geometric and non-geometric road attributes), $m\,(m = 1,2,3,...,M)$ is the number of variables; $X_{1i} = \ln(L_i)$ and $X_{2i} = \ln(V_i)$; $L_i$ and $V_i$ represent the length and traffic volume (AADT), respectively. $\beta_0$ is the intercept and $\beta_m$ denotes the vector of regression coefficients. The term $\exp(\varepsilon_i)$ represents a multiplicative random effect due to unobserved heterogeneity (also known as unstructured errors), which follows a Gamma distribution with an inverse dispersion parameter $k$ (also known as the shape parameter):

$$\exp(\varepsilon_i) \mid k \sim Gamma(k, k) \qquad (8)$$

The dispersion (or over-dispersion) parameter is usually referred to as

$$\alpha = \frac{1}{k} \qquad (9)$$

The probability density function of the PG or NB model is given by

$$\Pr\{\, Y_i = y_i \mid \mu_i, k \,\} = \frac{\Gamma(y_i + \kappa)}{y_i!\,\Gamma(\kappa)} \left(\frac{\kappa}{\kappa + \mu_i}\right)^{\kappa} \left(\frac{\mu_i}{\kappa + \mu_i}\right)^{y_i} \qquad (10)$$

Under the PG or NB model, the mean and variance are given by

$$E(Y_i) = \mu_i \qquad (11)$$

$$Var(Y_i) = \mu_i + \frac{\mu_i^2}{\kappa} \qquad (12)$$

### 2.2.3 Poisson Lognormal Model

Several distributions (e.g., Gamma, Lognormal) can be used for unstructured errors, $\exp(\varepsilon_i)$. However, as the univariate PLN model is more flexible than the univariate Poisson-Gamma or negative binomial model to handle over-dispersion, researchers have recently proposed using the PLN model as an alternative to the negative binomial and Poisson-Gamma model for modelling collision data (Miaou et al., 2003; Aguero-Valverde and Jovanis, 2008; Lord and Miranda-

Moreno, 2008; Lord and Mannering, 2010). The univariate PLN regression model is obtained by the assumption,

$$\exp\ (\varepsilon_i) \mid \sigma_u^2 \sim Lognormal\,(0, \sigma_u^2) \tag{13}$$

Or,

$$\varepsilon_i \mid \sigma_u^2 \sim Normal\,(0, \sigma_u^2) \tag{14}$$

Where, $\sigma_u^2$ denotes the extra Poisson variance or variance for heterogeneous effects (also known as within-site (extra) variation). Under the PLN model, the mean and variance are given by

$$E(Y_i) = \mu_i \exp(0.5\sigma_u^2) \tag{15}$$

$$Var(Y_i) = E(Y_i) + [E(Y_i)]^2[\exp(\sigma_u^2) - 1] \tag{16}$$

### 2.2.4 Multivariate Model

Generally, collision-related outcomes (e.g., fatal, injury, PDO) are considered as independent of each other and have often been analyzed independently. However, the collision types/severities are multivariate in nature and correlations exist among different severity levels/ types (Maher, 1990; Bijleveld, 2005; Ma and Kockelman, 2006; Park and Lord, 2007; Ma et al., 2008; Aguero-Valverde and Jovanis, 2009; El-Basyouny and Sayed, 2009b). Neglecting these correlations may lead to biased, incorrect parameter estimates and inferences. Therefore, a set of data on road collisions at *n* locations, where the collisions at each location are classified into *K* categories, can be defined as the vector $Y_i^k = (y_i^1, y_i^2 ... y_i^K)'$. $Y_i^k$ denotes the number of collisions at road segment *i (i*= 1, 2*, ..., n)* that belong to collision severity *k (k*= 1, 2, ..., *K)*. It is assumed that

$$Y_i^k \mid \theta_i^k \sim Poisson\,(\theta_i^k) \tag{17}$$

Where, $\theta_i^k$ is the Poisson parameter. For the probability of $y_i^k$, $k$ represents the types of collisions that occur on road segment *i*, given by

$$\Pr\{Y_i^k = y_i^k \mid \theta_i^k\} = e^{-\theta_i^k} \frac{\theta_i^{k\,y_i^k}}{y_i^k!} \tag{18}$$

$$\theta_i^k = \mu_i^k e^{\varepsilon_i^k} \qquad (19)$$

$$\ln(\theta_i^k) = \ln(\mu_i^k) + \varepsilon_i^k \qquad (20)$$

Where, $\mu_i^k$ is an exponential function of road segment-specific attributes, such as traffic volume, geometric and non-geometric road characteristics, etc. for $k$ types of collisions:

$$\ln(\mu_i^k) = \beta_0^k + \sum_{m=1}^{M} \beta_m^k X_{mi} \qquad (21)$$

$\beta_0^k$ is the intercept; $\beta_m^k$ denotes the vector of regression parameters; $\varepsilon_i^k$ denotes multivariate normal unstructured errors distributed as $\varepsilon_i \sim MN_K(0, \Sigma)$, which is equivalent to writing that $e^{\varepsilon_i} \sim Lognormal\ (0, \Sigma)$; and $\Sigma$ is the variance-covariance matrix for unstructured errors or heterogeneous effects. *MN* denotes *n*-dimensional multivariate normal distribution.

Where,

$$\varepsilon_i^k = \begin{pmatrix} \varepsilon_i^1 \\ \varepsilon_i^2 \\ \cdots \\ \varepsilon_i^K \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1K}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2K}^2 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{K1}^2 & \sigma_{K2}^2 & \cdots & \sigma_{KK}^2 \end{pmatrix}$$

The diagonal element $\sigma_{kk}^2$ of the variance-covariance matrix represents the heterogeneous variance of $\varepsilon_i^k$, where, the off-diagonal element $\sigma_{hk}^2$ represents the heterogeneous covariance of $\varepsilon_i^h$ and $\varepsilon_i^k$.

Under the multivariate PLN model, the mean and variance are given by

$$E(Y_i^k) = \mu_i^k \exp[0.5 * (\sigma_{11}^2 + \sigma_{22}^2 + ... + \sigma_{KK}^2)] \qquad (22)$$

$$Var(Y_i^k) = E(Y_i^k) + [E(Y_i^k)]^2 [\exp(\sigma_{11}^2 + \sigma_{22}^2 + ... + \sigma_{KK}^2) - 1] \qquad (23)$$

### 2.2.5 Univariate Spatial Model

The spatial PLN can be defined by incorporating spatial correlation (also known as structured errors, structured variations, or spatial effects) in Eqs. *(5–6)* as follows:

$$\theta_i = \mu_i \exp\,(\varepsilon_i) \exp\,(S_i) \tag{24}$$

$$\ln\,(\theta_i) = \ln\,(\mu_i) + \varepsilon_i + S_i \tag{25}$$

The spatial component $S_i$ suggests that road segments that are closer to one another are likely to have common features affecting their collision frequency. As noted by Miaou and Lord (2003), random variations across sites may be structured spatially due to the complexity of traffic interaction around locations. For the univariate Gaussian Conditional Auto-regressive (CAR) spatial correlation ($S_i$), the joint conditional distribution can be expressed as follows (Johnson and Kotz, 1972; Besag and Kooperberg, 1995; Thomas et al., 2004):

$$S_i \mid S_{-i} \sim Normal(\overline{S_i}, \sigma_s^2 / n_i) \tag{26}$$

Where, $\overline{S_i} = \sum_{j \in C(i)} \frac{S_j}{n_i}$ and $C(i)$ denotes the set of neighbours for road segments *i*. $\sigma_s^2$ denotes the variance for spatial correlation (also known as spatial variation).

### 2.2.6 Multivariate Spatial Model

The multivariate spatial PLN can be defined by incorporating spatial correlation in Eqs. *(19–20)* as follows:

$$\theta_i^k = \mu_i^k e^{\varepsilon_i^k} e^{S_{ki}} \tag{27}$$

$$\ln(\theta_i^k) = \ln(\mu_i^k) + \varepsilon_i^k + S_{ki} \tag{28}$$

For the multivariate *k*-dimensional CAR model, the vector of spatially correlated *i* road segments is $S_{ki} = (S_{1i}, S_{2i}, ..., S_{ki})^{/}$. The conditional distribution can be expressed as the following (Thomas et al., 2004):

$$S_{ki} \mid (S_{1(-i)}, ..., S_{k(-i)}) \sim MN(\overline{S_{(k)i}}, \frac{\Omega}{n_i}) \tag{29}$$

Here, $(S_{1(-i)}, ..., S_{k(-i)})$ denotes the road segments of the $k \times n$ matrix $S_{ki}$ excluding the *i*-th road segment. $\Omega$ is the variance-covariance matrix for spatial correlation and can be expressed as

$$\Omega = \begin{pmatrix} \sigma_{s11}^2 & \sigma_{s12}^2 & \cdots & \sigma_{s1K}^2 \\ \sigma_{s21}^2 & \sigma_{s22}^2 & \cdots & \sigma_{s2K}^2 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{sK1}^2 & \sigma_{sK2}^2 & \cdots & \sigma_{sKK}^2 \end{pmatrix}$$

The diagonal elements of the covariance matrix ( $\Omega$ ) represent spatial variance. The off-diagonal elements represent the spatial covariance of different severity levels.

### 2.2.7 Univariate Random Parameters Model

The use of random parameters in collision modelling has recently been gaining attention. A recent approach for modelling the mean function advocates the use of random parameters (Anastasopoulos and Mannering, 2009). In all traditional models, only one regression equation was fit to the dataset. Using a random parameters model develops different regression equations for individual sites. This type of modelling technique has been considered by several researchers for its added flexibility and intuitive appeal (Li et al., 2008; Milton et al., 2008; Anastasopoulos and Mannering, 2009). The model can be viewed as an extension of the random effects model, since instead of varying only the intercept of the model, the random parameters model allows each estimated parameter in the model to vary across each individual observation in the dataset. This model focuses on explaining part of the extra-variation through improvements in the mean function by accounting for the unobserved heterogeneity from one site to another.

For the random parameters model, the road segment variations can be represented by allowing all regression coefficients to vary randomly from one segment to another, and Eq. *(7)* can be written as follows:

$$\ln(\mu_i) = \beta_{0i} + \sum_{m=1}^{M} \beta_{mi} X_{mi} \qquad (30)$$

Where,

$$\beta_{0i} \sim Normal(\beta_0, \sigma_0^2) \text{ and } \sigma_0^{-2} \sim Gamma(0.01, 0.01) \qquad (31)$$

$$\beta_{mi} \sim Normal(\beta_m, \sigma_m^2) \text{ and } \sigma_m^{-2} \sim Gamma(0.01, 0.01) \qquad (32)$$

Several distributions (e.g., normal, lognormal, uniform, triangular, Gamma) were considered in Eq. *(8)*, but the normal distribution was found to provide the best statistical fit (Li et al., 2008;

Milton et al., 2008; Anastasopoulos and Mannering, 2009). Under the random parameter PLN model, the mean and variance are given by

$$E(Y_i) = \mu_i^* \exp(0.5\sigma_i^2) \tag{33}$$

$$Var(Y_i) = E(Y_i) + [E(Y_i)]^2[\exp(\sigma_i^2) - 1] \tag{34}$$

Where,

$$\ln(\mu_i^*) = \beta_0 + \sum_{m=1}^{M} \beta_m X_{mi} \tag{35}$$

$$\sigma_i^2 = \sigma_0^2 + \sum_{m=1}^{M} X_{mi}^2 \sigma_m^2 + \sigma_u^2 \tag{36}$$

### 2.2.8 Multivariate Random Parameters Model

For the multivariate random parameters model, the road segment variations can be represented by allowing all regression coefficients to vary randomly from one segment to another, and Eq. *(21)* can be written as

$$\ln(\mu_i^k) = \beta_{0i}^k + \sum_{m=1}^{M} \beta_{mi}^k X_{mi} \tag{37}$$

Where,

$$\beta_{0i}^k \sim MN(\beta_0^k, \Re_0) \tag{38}$$

$$\beta_{mi}^k \sim MN(\beta_m^k, \Re_m) \tag{39}$$

$\Re_0$ and $\Re_m$ are the variance-covariance matrixes for random parameters $\beta_{0i}^k$ and $\beta_{mi}^k$ respectively. For severity level, $K$=2, the variance-covariance matrixes for random parameters are given by

$$\Re_0 = \begin{pmatrix} \sigma_{R_0 11}^2 & \sigma_{R_0 12}^2 \\ \sigma_{R_0 21}^2 & \sigma_{R_0 22}^2 \end{pmatrix} \text{ and } \Re_m = \begin{pmatrix} \sigma_{R_m 11}^2 & \sigma_{R_m 12}^2 \\ \sigma_{R_m 21}^2 & \sigma_{R_m 22}^2 \end{pmatrix}$$

Under the multivariate random parameters PLN model, the mean and variance are given by

$$E(Y_i^k) = \mu_i^{k*} \exp(0.5 * \sigma_{ik}^2) \tag{40}$$

$$Var(Y_i^k) = E(Y_i^k) + [E(Y_i^k)]^2[\exp(\sigma_{ik}^2) - 1] \tag{41}$$

Where,

$$\mu_i^{k*} = \exp(\beta_0^k + \sum_{m=1}^{M} \beta_m^k X_{mi}) \tag{42}$$

$$\sigma_{ik}^2 = \sigma_{R_0kk}^2 + \sum_{m=1}^{M} X_{mi}^2 \sigma_{R_mkk}^2 + \sigma_{kk}^2 \tag{43}$$

# 3. DATA DESCRIPTION

Two datasets from the cities of Richmond and Vancouver in British Columbia, Canada, were investigated to develop the spatial models. A total of $n = 72$ urban road segments in the city of Richmond and $n = 281$ urban road segments in the city of Vancouver were used to perform the analysis. The study area is shown in Figure 1 (Richmond) and Figure 2 (Vancouver). The data was obtained from the City of Richmond and the City of Vancouver and covered the period 1994 to 1996. The road network maps of both the cities were obtained and used to determine a matrix for neighbouring segments. Various neighbouring structures were considered by Aguero-Valverde and Jovanis (2008; 2010). The road network in Figure 3 illustrates the definition of different neighbouring structures (e.g., first-order neighbours, second-order neighbours, third-order neighbours). On the basis of their results, as well as others reported in the literature (Nicholson, 1999), only first-order neighbours were considered to define the neighbouring structure in this thesis. First-order neighbours included all of the segments that had a direct connection with the segments under consideration. A statistical summary of the dataset is shown in Table 1 and Table 2. Table 3 shows the explanatory variables that were significant and used in the final models with their corresponding abbreviations and units. To develop the multivariate models, two collision severities were considered: i) severe collisions, consisting of injury and fatal collisions; and ii) non-injury collisions.

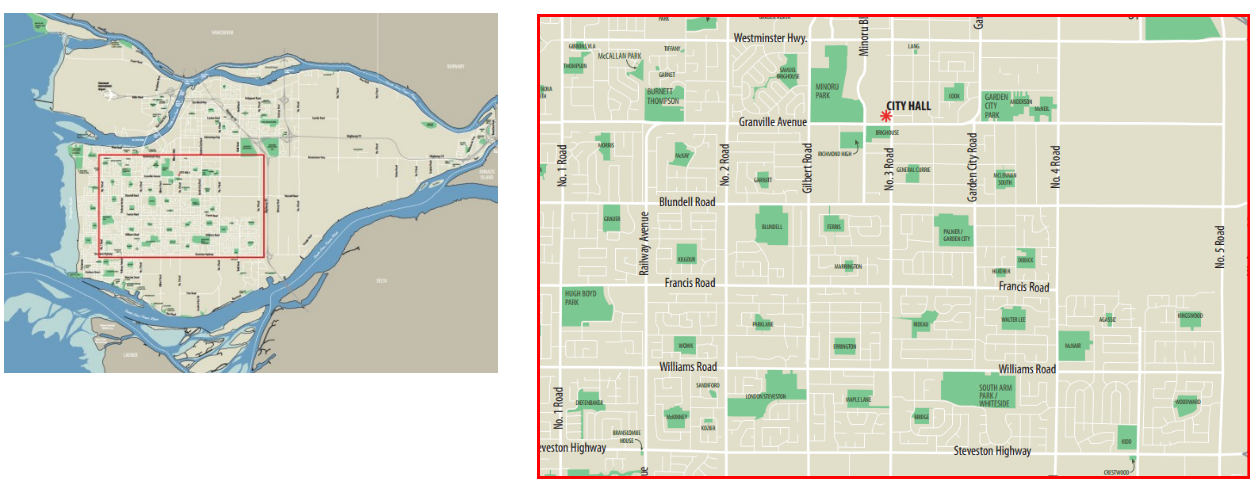


**Figure 1: Illustration of the study area in the city of Richmond**


**Figure 2: Illustration of the study area in the city of Vancouver**

The dataset contains the road segment lengths (in km), traffic volume (AADT), number of crosswalks, unsignalized intersection densities (UNID), undivided cross sections (IUND), business land use, residential land use, number of lanes, number of two-lane and four-lane road segments, number of bus stops, and percentage of peak hour and non-peak hour parking. A statistical summary of the dataset is shown in Table 1. Table 2 shows the explanatory variables that were significant and used in the final models with their corresponding abbreviations and units. Average values for the traffic volumes (over the three year period) were used to build models for the total (aggregated) number of collisions, as well as for the multivariate models. The aggregation is justified on several grounds. For instance, an aggregated collision model was found to perform well when compared with collision models developed to handle temporal correlation (Lord and Persaud, 2000; Anastasopoulos and Mannering, 2009). Moreover, the aggregation of collisions over a period of reasonable length helps to avoid confounding effects and phenomena such  as regression-to-the-mean (Cheng and Washington, 2005).

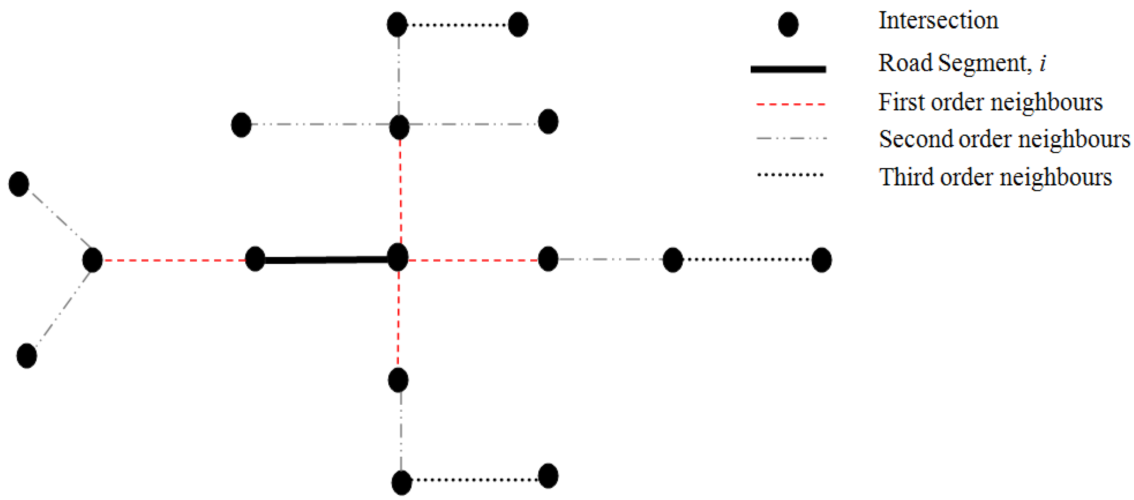


**Figure 3: Neighbouring structure definition**

**Table 1: Statistical summary of Richmond dataset (n =72 road segments)**

| Attributes | Mean | Standard Deviation | Max | Min |
|---|---|---|---|---|
| Length (km) | 0.882 | 0.299 | 2.510 | 0.743 |
| Average annual daily traffic (AADT) | 17193.931 | 7374.078 | 32792 | 4232 |
| Number of crosswalks | 1.306 | 1.339 | 7 | 0 |
| Unsignalized intersection densities (UNID) | 3.296 | 1.717 | 8.889 | 0 |
| Undivided cross section (IUND) | 0.736 | 0.444 | 1 | 0 |
| Residential land use | 0.889 | 0.316 | 1 | 0 |
| Business land use | 0.042 | 0.201 | 1 | 0 |
| Number of two-lane road segments | 0.236 | 0.428 | 1 | 0 |
| Number of four-lane road segments | 0.764 | 0.428 | 1 | 0 |
| Number of bus stops | 4.222 | 3.154 | 11 | 0 |
| Percentage of peak hour parking | 6.389 | 20.662 | 90 | 0 |
| Percentage of non-peak hour parking | 35.833 | 37.152 | 90 | 0 |
| **Collision Data** | | | | |
| Fatalities (F) | 0.097 | 0.298 | 1 | 0 |
| Injuries (I) | 10.014 | 10.288 | 65 | 0 |
| Severe (I + F) | 10.111 | 10.386 | 66 | 0 |
| Non-injury | 15.139 | 15.887 | 108 | 1 |
| Total Collisions | 25.250 | 25.682 | 174 | 3 |

**Table 2: Statistical summary of Vancouver dataset (n =281 road segments)**

| **Attributes** | **Mean** | **Standard Deviation** | **Max** | **Min** |
|---|---|---|---|---|
| Length (km) | 0.793 | 0.430 | 3.608 | 0.113 |
| Average annual daily traffic (AADT) | 26762.790 | 13794.387 | 62931 | 4236 |
| Number of crosswalks | 2.146 | 2.200 | 10 | 0 |
| Unsignalized intersection densities (UNID) | 6.942 | 3.287 | 21.164 | 0 |
| Undivided cross section (IUND) | 0.779 | 0.415 | 1 | 0 |
| Residential land use | 0.676 | 0.469 | 1 | 0 |
| Business land use | 0.313 | 0.465 | 1 | 0 |
| Number of lanes | 4.000 | 1.509 | 7 | 2 |
| Number of bus stops | 4.651 | 3.827 | 17 | 0 |
| Percentage of peak hour parking | 41.665 | 37.089 | 100 | 0 |
| Percentage of non-peak hour parking | 71.264 | 21.248 | 100 | 0 |
| **Collision data** | | | | |
| Fatalities (F) | 0.082 | 0.287 | 2 | 0 |
| Injuries (I) | 15.342 | 13.895 | 88 | 0 |
| Severe (I+F) | 15.423 | 13.966 | 88 | 0 |
| Non-injury | 43.740 | 36.832 | 223 | 1 |
| Total Collisions | 59.164 | 49.965 | 311 | 2 |

**Table 3: Description of model covariates**

| **Covariate** | **Symbol** | **Description** |
|---|---|---|
| ln (Length) | $X_1$ | Logarithm of road segment length |
| ln (AADT) | $X_2$ | Logarithm of Average annual daily traffic (vehicles per day) |
| Crosswalks | $X_3$ | Number of crosswalks in each road segments |
| UNID | $X_4$ | Unsignalized intersection densities |
| IUND | $X_5$ | Undivided Cross Sections |
| IBUS | $X_6$ | Business land use |
| NL | $X_7$ | Number of lanes |
| Total collisions | $Y_i$ | Consisting of injury, fatal and non-injury collisions |
| Severe collisions | $Y_i^1$ | Consisting of injury and fatal collisions |
| Non-injury collisions | $Y_i^2$ | Consisting of property damage only collisions |

# 4. MULTIVARIATE SPATIAL MODELS

This chapter investigated the inclusion of spatial correlation in different collision severity levels, using the multivariate modelling approach, and its influence on the collision analysis of urban arterials.

## 4.1 Background

Most of the literature related to the development of collision models accounts only for Poisson variation and heterogeneity. Despite the evident spatial nature of collisions, little road safety research has been conducted to account for spatial correlation. Recently, however, the need to include spatial correlation in the development of collision models for both intersections and road segments (Abdel-Aty and Wang, 2006; Aguero-Valverde and Jovanis, 2008; 2010; El-Basyouny and Sayed, 2009c; Mitra, 2009) and at the area-wide level (e.g., wards, neighbourhoods, county) (Aguero-Valverde and Jovanis, 2006; Aguero-Valverde, 2013; Amoros et al., 2003; Flask and Schneider IV, 2013; Noland and Quddus, 2004; Quddus, 2008) has been gaining attention in the literature. Congdon (2006) suggested that ignoring spatial dependence leads to an underestimation of variability. Furthermore, according to Aguero-Valverde and Jovanis (2010), the inclusion of spatially correlated random effects significantly improves the precision of the estimates of the expected collision frequency for road segments. The inclusion of spatial correlation has two main advantages: i) spatial correlation sites estimate the “pool strength” from neighbouring sites, thereby improving model parameter estimation (Aguero-Valverde and Jovanis, 2008); and ii) spatial dependence can be a surrogate for unknown and relevant covariates, thereby reflecting unmeasured confounding factors (Cressie, 1993; Dubin, 1988).

Most studies used spatial models for collision frequency or type independently. However, collision data is multivariate in nature, and it is necessary to account for the likely correlation between collision counts at different levels of classification. While a number of studies explored the use of multivariate collision modelling to capture heterogeneous correlations among different collision types or severities (Bijleveld, 2005; Ma and Kockelman, 2006; Ma et al., 2008; Park and Lord, 2007; El-Basyouny and Sayed, 2009b), multivariate spatial correlations were rarely investigated. Furthermore, univariate spatial modelling of different types of collision counts may lead to biased results, as collision types or severities are not spatially independent of one another.

With the inclusion of multivariate spatial correlation, collision models can estimate the associated safety risk and spatial correlation of different collision severities and types in the same spatial unit. However, only two studies (Aguero-Valverde, 2013; Song et al., 2006) focused on area-wide multivariate spatial models for different collision severities and types. Further, Wang and Kockelman (2013) and Narayanamoorthy et al. (2013) used a multivariate spatial modelling approach for pedestrian and bicyclist collision analysis at the census tract level. Multivariate spatial modelling for road segments or intersections is rarely explored in the literature. To this end, this chapter focuses on the first objective of this thesis and performs two tasks: i) use the multivariate spatial modelling approach to develop spatial models for road segments in order to assess spatial correlation at different collision severity levels and its influence on collision analysis of urban arterials; and ii) compare multivariate spatial models with independent (separate) univariate spatial models for each collision severity in terms of model inference and goodness-of-fit. To accomplish the objective, three years (1994 to 1996) of collision data and other geometric and non-geometric road data were used for the city of Richmond and city of Vancouver, British Columbia, Canada.

From a methodological perspective, various approaches, such as Moving Average (Congdon, 2006), Simultaneous Auto-regressive (SAR) (Quddus, 2008), Spatial Error Model (SEM) (Anselin, 1988; Quddus, 2008), Multiple Membership (MM) (Goldstein, 1995; Langford et al., 1999; El-Basyouny and Sayed, 2009c), Extended Multiple Membership (EMM) (El-Basyouny and Sayed, 2009c), Geographic Weighted Regression (GWR) (Hadayeghi et al., 2003), Geographic Weighted Poisson Regression (GWPR) (Hadayeghi et al., 2010), and Generalized Estimating Equations (GEE) (Abdel-Aty and Wang, 2006), have been advocated by other researchers to assess spatial effects or spatial correlation. Each approach has its own pros and cons. However, almost all of the earlier studies used Gaussian Conditional Auto-regressive (CAR) (Besag et al., 1991) distribution for modelling spatial correlation (Aguero-Valverde & Jovanis, 2006; 2008; 2010; Mitra, 2009; Guo et al., 2010; Ahmed et al., 2011; Siddiqui et al., 2012). In addition, Quddus (2008) advocated that CAR distribution under a Bayesian framework can provide more appropriate and better inference over classical spatial models, because the Bayesian CAR models with heterogeneous effects are able to accurately take into account both the spatial correlation and unobserved heterogeneity of the collision data. Therefore, in this thesis, univariate and multivariate spatial models were applied using CAR distribution. The models were estimated in a Full Bayesian (FB) context via Markov Chain Monte Carlo (MCMC) simulation (Gilks et al., 1996). As WinBUGS (Lunn et al., 2000) is a flexible platform for the Bayesian analysis of complex statistical models using MCMC methods, this open source statistical software was used for the development of the proposed spatial models.

## 4.2 Methodology

### 4.2.1 Model Specification

The multivariate PLN models for correlated data were originally developed by Chib and Winkelmann (2001) for modelling health-related and airline incidents data. Later, several studies used them for modelling collision counts at different levels of classification (Bijleveld, 2005; Ma and Kockelman, 2006; Ma et al., 2008; Park and Lord, 2007; El-Basyouny and Sayed, 2009b). Let $Y_i^k$ denote the collision count for road segments *i (i*= 1, 2*, ..., n)* that belong to collision severity *k (k*= 1, 2, ..., *K)*. For the present dataset, *n*=72 (for Richmond dataset), *n*= 281 (for Vancouver dataset) and *K*=2 (severe and non-injury collisions). While the multivariate PLN model can handle *K* severity levels, the current application involves only two severity levels leading a bivariate PLN. The multivariate model specifications are given by Eqs. *(17–23)*. Recall Eq. *(21)*, which can be written as

$$\ln(\mu_i^k) = \beta_0^k + \ln(L) + \sum_{m=2}^{M} \beta_m^k X_{mi} \tag{44}$$

Where, $\ln(L)$ is the logarithm of road segment length, which is taken as an offset because the posterior mean of the parameter is either one or close to one.

Multivariate spatial models are given by Eqs. *(27–28).* The spatial component $S_{ki}$ suggests that road segments that are closer to one another are likely to have common features affecting their collision severity. For the spatial correlation, $S_{ki}$ , the joint distribution can be expressed as follows (Thomas et al., 2004):

$$S_{ki} \sim MN(\vartheta, V\Gamma) \tag{45}$$

Where, $S_{ki} = (S_{k1},...,S_{kn})$, *MN* denotes *n*-dimensional multivariate normal distribution; $\vartheta$ is the $1\times n$ mean vector; and $V>0$ controls the overall variability of the $S_{ki}$. $\Gamma$ is an $n\times n$ positive definite symmetric spatial covariance matrix and can be written in the following form:

$$V\Gamma = V(I - S_d C)^{-1} D \qquad (46)$$

Where,

$I = n\times n$ identity matrix;

$D = n\times n$ diagonal matrix with elements $D_{ii}$ proportional to the conditional spatial variance between road segment *i* and *j*, $S_{ki} \mid S_{kj}$.

$C = n\times n$ weight matrix, with element $C_{ij}$ reflecting the spatial association between road segments *i* and *j*.

$S_d$ controls the overall strength of spatial dependence ($S_d = 0$ implies no spatial dependence).

For the univariate CAR spatial correlation ($S_i$), the joint conditional distribution can be expressed as follows (Johnson and Kotz, 1972; Besag and Kooperberg, 1995; Thomas et al., 2004):

$$S_i \mid S_{-i} \sim Normal(\vartheta_i + \sum_{j=1}^{n} S_d C_{ij}(S_j - \vartheta_i), \phi D_{ii}) \qquad (47)$$

Where, $S_{-i}$ denotes all of the elements except $S_i$, and $\phi$ is the correlation parameter. $D_{ii} = 1/n_i$, where, $n_i$ is the number of road segments that are adjacent to road segments *i*. $S_d = S_{d(\max)}$, which equals 1 with the particular choice of $C_{ij}$ and $D_{ii}$. $C_{ij}$ is the element of the weight matrix and can be expressed as

$$C_{ij} = \frac{W_{ij}}{W_{i+}} \qquad (48)$$

Where, $W_{i+} = \sum_{j=1}^{n} W_{ij}$ , $W_{ij} = 1$ if road segment $i$ and $j$ are adjacent and $W_{ij} = 0$ otherwise. The specification of *C*, *D* and $S_d$ leads the conditional distribution as in Eq. *(26)*. For the multivariate *k*-dimensional (here, *k=2*) CAR model, the vector of spatially correlated *i* road segments is $S_{ki} = (S_{1i}, S_{2i}, ..., S_{ki})^{/}$. The distribution of Eq. *(47)* can be expressed as Eq. *(29)*.

The impact of spatial correlation is assessed by computing the proportion of total variation that is due to spatial variation (El-Basyouny and Sayed, 2009c):

$$\Psi_s = \frac{\text{var}(S)}{\text{var}(S) + \sigma_u^2} \tag{49}$$

Where, *var(S)* represents the marginal variance of *S*. For spatial models, *var(S)* can be estimated directly from the posterior distribution of *S*.

### 4.2.2 The Models

To assess multivariate spatial correlation for different severity levels, three models were used based on random effects:

**Model 1A: Multivariate model with only heterogeneous effects**

$$\ln(\theta_i^k) = \beta_0^k + \ln(L) + \sum_{m=2}^{M} \beta_m^k X_{mi} + \varepsilon_i^k \tag{50}$$

**Model 1B: Multivariate model with only spatial correlation**

$$\ln(\theta_i^k) = \beta_0^k + \ln(L) + \sum_{m=2}^{M} \beta_m^k X_{mi} + S_{ki} \tag{51}$$

**Model 1C: Multivariate model with both heterogeneous effects and spatial correlation**

$$\ln(\theta_i^k) = \beta_0^k + \ln(L) + \sum_{m=2}^{M} \beta_m^k X_{mi} + \varepsilon_i^k + S_{ki} \tag{52}$$

In addition, two univariate PLN spatial models for different severity levels were also developed and compared with the best fitted multivariate spatial models (Model 1A, 1B and1C).

### 4.2.3 Prior and Posterior Distributions

Obtaining the FB estimates requires a specification of prior distributions for the regression coefficients $\beta_0^k$ , $\beta_m^k$, the heterogeneous covariance matrix $\Sigma$ and the spatial covariance matrix

$\Omega$. Prior distributions are meant to reflect prior knowledge about the parameters of interest. If such prior information is available, then it should be used to formulate the so-called informative priors (Bedrick et al., 1996; Schluter et al., 1997) (for more detailed information about informative priors, refer to Yu and Abdel-Aty (2013)). In the absence of sufficient prior knowledge of the distributions for individual parameters, un-informative (vague) prior distributions are usually specified. The most commonly used priors are diffused normal distributions (with zero mean and large variance) for the regression parameters and a $Wishart\ (P, r)$ prior for $\sum^{-1}$ and $\Omega^{-1}$, where, $P$ and $r \geq K$ represent the prior guess at the order of magnitude of the precision matrix $\sum^{-1}$ and $\Omega^{-1}$ and the degrees of freedom, respectively. Choosing $r$=$K$ as the degrees of freedom corresponds to vague prior knowledge (Spiegelhalter et al., 1996; Tunaru, 2002). In the current research, several priors were used: $\beta_m^k \sim N(0, 100^2)$, $\sum^{-1} \sim Wishart\,(I, K)$ and $\Omega^{-1} \sim Wishart\,(I, K)$, where $I$ is the $K \times K$ identity matrix (Chib & Winkelmann, 2001; Congdon, 2006).

### 4.2.4 Full Bayesian Estimation

The posterior distributions required in the FB approach can be obtained using MCMC sampling techniques available in WinBUGS (Lunn et al., 2000). The Wishart distribution can be sampled using a Gibbs sampler. Monitoring convergence is critical, because it ensures that the posterior distribution is found, thereby indicating when parameter sampling should begin. To check convergence, two or more parallel chains with diverse starting values are tracked to ensure full coverage of the sample space. Convergence of multiple chains is assessed using the Brooks-Gelman-Rubin (BGR) statistic (Brooks and Gelman, 1998). A value of less than 1.2 of the BGR statistic indicates convergence. Convergence is also assessed by visual inspection of the MCMC trace plots for the model parameters, as well as by monitoring the ratios of the Monte Carlo errors relative to the respective standard deviations of the estimates; as a rule, these ratios should be less than 0.05.

### 4.2.5 Comparison of Models and Goodness-of-Fit

The Deviance Information Criteria (DIC) was used for model comparisons (Spiegelhalter et al., 2002). DIC is a Bayesian generalization of Akaike's Information Criteria (AIC) that penalizes larger parameter models. According to Spiegelhalter (2005), it is difficult to determine what constitutes a critical difference in DIC. Very roughly, differences of greater than 10 may rule out the model with the higher DIC. Differences between 5 and 10 are considered substantial. If the difference in DIC is less than 5, and the models make significantly different inferences, then it could be misleading to only report the model with the lowest DIC.

An earlier study by El-Basyouny and Sayed (2009b) illustrated that DIC is additive under independent models and priors. Let $f(y|\eta)$ and $f(y)$ denote the conditional and marginal distributions of *y*, where, $\eta$ denotes the vector of parameters associated with *y*. Then, $DIC = \overline{D} + p$, where, $p = \overline{D} - D(\overline{\eta})$, $\overline{\eta} = E[\eta | y]$ and $\overline{D} = E[D(\eta) | y]$ are the posterior means of $\eta$ and the Bayesian deviance,

$$D(\eta) = -2\ln\{f(y|\eta)\} + 2\ln\{f(y)\}. \tag{53}$$

For *K* collision categories, let *y* and $\eta$ be partitioned as $(y_1,...,y_k)$ and $(\eta_1,...,\eta_k)$. Define,

$$DIC_k = \overline{D_k} + p_k \tag{54}$$

$$p_k = \overline{D_k} - D_k(\overline{\eta_k}) \tag{55}$$

$$\overline{D_k} = E[D_k(\eta_k) | y_k] \tag{56}$$

$$\overline{\eta_k} = E[\eta_k | y_k] \tag{57}$$

$$D_k(\eta_k) = -2\ln\{f(y_k|\eta_k)\} + 2\ln\{f(y_k)\}. \tag{58}$$

Under independent models and priors, $f(y|\eta) = \prod_{k=1}^{K} f(y_k|\eta_k)$ and $f(y) = \prod_{k=1}^{K} f(y_k)$. These multiplicative conditional and marginal distributions of *y* translate additively in the Bayesian deviance Eq. *(19)* leading to $DIC = \sum_{k=1}^{K} DIC_k$.

To assess the models' goodness-of-fit (adequacy), a posterior predictive approach (Gelman et al., 1996; Stern and Cressie, 2000; Li et al., 2008) was used. The procedure involves generating replicates under the postulated models and comparing the distribution of a certain discrepancy measure, such as chi-square statistics, to the value of the chi-square obtained using observed data. A model does not fit the data if the observed value of the chi-square is far from the predictive distribution; the discrepancy cannot be reasonably explained by chance if the *p*-value is close to zero or one (Gelman et al., 1996). The replicates are best obtained simultaneously with model estimation in WinBUGS to account for all of the uncertainties in the model parameters as reflected by the estimated distributions.

The chi-square statistics are computed from

$$\chi^2 = \sum_{i=1}^{n} \left[ \frac{\left[ y_i^k - E(Y_i^k) \right]^2}{Var(Y_i^k)} \right] \tag{59}$$

Where, $y_i^k$ denotes either the observed or replicated collision frequencies by collision severity. For $i = 1, 2, ..., n$ and $k = 1, 2, ..., K$, the expected mean and variance of $Y_i^k$ are given by

$$E(Y_i^k) = \mu_i^k \exp[0.5(\sigma_{kk}^2 + \sigma_{skk}^2)] \tag{60}$$

and

$$Var(Y_i^k) = E(Y_i^k) + [E(Y_i^k)]^2 (\exp(\sigma_{kk}^2 + \sigma_{skk}^2) - 1) \tag{61}$$

## 4.3 Results and Discussion

### 4.3.1 Model Selection

For each model, the posterior estimates were obtained via two chains with 20,000 iterations, 5,000 of which were excluded as a burn-in sample using WinBUGS. The BGR statistics were less than 1.2; the ratios of the Monte Carlo errors relative to the standard deviations of the estimates were less than 0.05; and trace plots for all of the model parameters indicated convergence. The model selection criterion is presented in Table 4.

As observed in Table 4, for the Richmond dataset, the multivariate model with both heterogeneous effects and spatial correlation (Model 1C) has a slightly lower DIC value than the other two models. However, as the difference in DIC is less than five, it could be misleading to only report the model with the lowest DIC. Therefore, a predictive posterior approach was used to assess the model's goodness-of-fit.

**Table 4: The DIC statistics by model**

| **Model Description** | **DIC** | |
|---|---|---|
| | **Richmond** | **Vancouver** |
| Model 1A: Multivariate model with only heterogeneous effects | 786.3 | 3575 |
| Model 1B: Multivariate model with only spatial correlation | 788.7 | 3560 |
| Model 1C: Multivariate model with both heterogeneous effects and spatial correlation | **785.2** | **3548** |

To assess the model's goodness-of-fit, the Pearson's residuals in Eq. *(59)* were examined and no anomalies were detected. The posterior estimates of the observed chi-square statistics were 69.83 and 89.59 for severe and non-injury collisions, respectively (Table 5). The associated $p$-values estimated from the distributions of the chi-square discrepancy measured in the replicated datasets were 0.59 and 0.316 for severe and non-injury collisions, respectively. The $p$-values of 0.59 and 0.316 represented the areas under the predictive distribution to the right of the observed chi-square statistics. As mentioned earlier, a model does not fit the data if the observed value of the chi-square is far from the predictive distribution; the discrepancy cannot be reasonably explained by chance if the $p$-values are close to zero or one. Since the associated $p$-values were 0.59 and 0.316, well distanced from zero as well as one, the multivariate model with both heterogeneous effects and spatial correlation (Model 1C) performed well in terms of accommodating the variation in collision frequency by severity across different road characteristics.

For the Vancouver dataset, the DIC statistics were 3575, 3560 and 3548 under Model 1A, Model 1B and Model 1C, respectively. The multivariate model with both heterogeneous effects and spatial correlation (Model 1C) was the best fit according to DIC. It should be noted that, for Model 1C, the inclusion of heterogeneous effects and spatial correlation shows a very significant drop-off of DIC by 27 and 12 compared to Model 1A and Model 1B, respectively.

Under Model 1C, the observed value of the chi-squares were 172.9 and 140.4 for severe and non-injury collisions, respectively (Table 6). The observed values of the chi-squares were located near the centre of the replicated distribution, with associated $p$-values of 0.38 and 0.465 for severe and non-injury collisions, respectively. As a result, Model 1C seems to perform well in terms of accommodating the variation in collision frequency by different severity levels across segments.

### 4.3.2 Parameter Estimates

#### *4.3.2.1 Richmond Dataset*

Table 5 summarizes the parameter estimates and their associated statistics for multivariate models. As the focus of this chapter is to investigate the inclusion of spatial correlation in multivariate collision models, only the results of Model 1C are explained in this section. The table shows that the parameter estimates were significant; the 95% credible intervals were bounded away from zero (except UNID for severe collisions). The regression coefficients were all positive, indicating a positive correlation with both severe and non-injury collisions.

The modelling results revealed that AADT was statistically significant at a 95% confidence level and positively correlated with both severe and non-injury collisions, which indicates that higher traffic volumes (i.e., increased exposure) results in more severe and non-injury collisions. According to the parameter estimates, there was a 8.83% increase in predicted severe collisions per 10% increase in traffic volume. Similarly, there was a 9.43% increase in predicted non-injury collisions per 10% increase in traffic volume. This finding is intuitive and in line with previous research findings (Greibe, 2003; Ma et al., 2010).

The number of crosswalks was also significant and positively correlated with both severe and non-injury collisions, indicating that an increase in the number of crosswalks increases the associated safety risk. This is expected, as the presence of crosswalks increases pedestrian activity, hence increasing collision probability and risk. According to parameter estimates, there was a 30.87% and 28.66% increase in severe and non-injury collisions for a 1% increase in the number of crosswalks, respectively. The parameter of severe collisions was higher than non-injury collisions, which is also expected, as crosswalks are generally situated in the middle of road segments where vehicular speed is quite high, and there is abundant evidence showing that higher speeds are associated with an increase in collision risk and the degree of collision severity (Elvik et al., 2004; Nilsson, 2004; Peden et al., 2004). At 50 km/h, which is the speed limit of most residential areas in most cities, the probability of pedestrian fatality from vehicular collisions with pedestrians is 90%, indicating death is almost inevitable (OECD, 2006). Another study suggests that the average risk of pedestrian fatality reaches 10% at an impact speed of 38 km/h, 50% at 65 km/h and 90% at 88 km/h (Tefft, 2013).

**Table 5: Parameter estimates and 95% credible intervals for multivariate models (Models 1A–1C) (Richmond dataset)**

| Variable (Parameter) | Severe: Injury + Fatal | | | Non-Injury | | |
|---|---|---|---|---|---|---|
| | | 95% Credible Intervals | | | 95% Credible Intervals | |
| | Est. | Lower Limit | Upper Limit | Est. | Lower Limit | Upper Limit |
| **Model 1A: Multivariate model with only heterogeneous effects** | | | | | | |
| **Intercept** $(\beta_0)$ | -4.681 | -7.935 | -1.623 | -4.548 | -7.989 | -1.321 |
| **ln(AADT)** $(\beta_2)$ | 0.645 | 0.330 | 0.977 | 0.663 | 0.337 | 1.010 |
| **Crosswalks** $(\beta_3)$ | 0.268 | 0.158 | 0.376 | 0.254 | 0.129 | 0.374 |
| **UNID** $(\beta_4)$ | *0.078* | *-0.009* | *0.167* | 0.105 | 0.009 | 0.202 |
| $\sigma^2_{kk}$ | 0.254 | 0.149 | 0.405 | 0.323 | 0.203 | 0.493 |
| **Model 1B: Multivariate model with only spatial correlation** | | | | | | |
| **Intercept** $(\beta_0)$ | -3.758 | -6.849 | -0.501 | -3.501 | -6.755 | -0.074 |
| **ln(AADT)** $(\beta_2)$ | 0.559 | 0.225 | 0.871 | 0.562 | 0.209 | 0.893 |
| **Crosswalks** $(\beta_3)$ | 0.193 | 0.093 | 0.300 | 0.169 | 0.050 | 0.284 |
| **UNID** $(\beta_4)$ | 0.084 | 0.002 | 0.157 | 0.118 | 0.038 | 0.207 |
| $\sigma^2_{skk}$ | 0.872 | 0.502 | 1.385 | 1.200 | 0.749 | 1.863 |
| **Model 1C: Multivariate model with both heterogeneous effects and spatial correlation** | | | | | | |
| **Intercept** $(\beta_0)$ | -4.577 | -8.003 | -1.032 | -4.567 | -8.107 | -0.871 |
| **ln(AADT)** $(\beta_2)$ | 0.633 | 0.275 | 0.976 | 0.664 | 0.280 | 1.024 |
| **Crosswalks** $(\beta_3)$ | 0.269 | 0.167 | 0.377 | 0.252 | 0.140 | 0.373 |
| **UNID** $(\beta_4)$ | *0.082* | *-0.008* | *0.172* | 0.109 | 0.012 | 0.196 |
| $\sigma^2_{kk}$ | 0.246 | 0.138 | 0.392 | 0.297 | 0.156 | 0.472 |
| $\sigma^2_{skk}$ | 0.027 | 0.002 | 0.133 | 0.061 | 0.002 | 0.379 |
| **Chi Observed** | 69.830 | 39.840 | 105.800 | 89.590 | 37.810 | 151.100 |
| **Chi Replicated** | 75.030 | 40.610 | 125.400 | 80.860 | 33.790 | 146.100 |
| ***p*** | 0.590 | 0 | 1 | 0.316 | 0 | 1 |

* *Parameter estimates not significant under the stated level of significance are shown in italic font.*

UNID were also significant and positively associated with non-injury collisions. According to the parameter estimates, a 1% increase in UNID increased non-injury collisions by 11.52%. The literature also suggests that road segments with a large number of access points or unsignalized intersections have a significant impact on, and are positively correlated with, collision frequency (Xuesong and Ming, 2012). UNID were insignificant for severe collisions in Model 1C. Interestingly, UNID were significant for both severe and non-injury collisions when considering only spatial correlation (Model 1B). According to parameter estimates, a 1% increase in UNID increased severe collisions by 8.76%. However, UNID became insignificant for both severe and non-injury collisions when considering only heterogeneous effects (Model 1A).

#### *4.3.2.2 Vancouver Dataset*

Table 6 summarizes the parameter estimates and their 95% credible intervals for Model 1A, Model 1B and Model 1C. As Model 1C provided the lowest DIC of the three models, only the results of Model C are explained in this section. The table shows that the parameter estimates are significant, as the 95% credible intervals were bounded away from zero.

The regression coefficients were all positive, indicating that factors such as AADT, UNID, IBUS and NL were positively associated with both severe and non-injury collisions. IUND was significant and positively related with only non-injury collisions. The modelling results revealed that there was a 5.19% increase in predicted severe collisions per 10% increase in traffic volume. Similarly, there was a 6.08% increase in predicted non-injury collisions per 10% increase in traffic volume. Similar to the Richmond dataset, UNID were also significant and positively associated with both severe and non-injury collisions. According to the parameter estimates, a 1% increase in UNID increased non-injury collisions by 6.92%. Similarly, non-injury collisions showed an increase of 7.47% due to 1% increase in UNID.

Another indicator variable, IUND, was also significant and positively correlated with non-injury collisions. It is worth mentioning that IUND was significant for severe collisions in Model 1A. However, it showed as insignificant in Model 1B and Model 1C while spatial correlation was included. The spatial correlation associated with the CAR models appears to be related to the presence of undivided cross sections. Such spatial multicollinearity appears to negate the need to include IUND in the multivariate models. The business land use indicator variable, IBUS, was significant and positively correlated with both severe and non-injury collisions as expected. Business land use introduces lots of commercial activities that attract vulnerable road users (e.g., pedestrians, bicyclists); therefore, collisions are more likely to occur in a vicinity where business land use occurs. NL also provided similar positive correlation with both severe and non-injury collisions. A possible rationale for this finding is that more lanes increase traffic flow and traffic conflict areas, therefore increasing the probability of collision occurrence. According to the parameter estimates, a 1% increase in UNID increased non-injury collisions by 12.9%. Similarly, no injury-collisions demonstrated an increase of 11.4% due to a 1% increase in UNID.

**Table 6: Parameter estimates and 95% credible intervals for multivariate models (Models 1A–1C) (Vancouver dataset)**

| Variable (Parameter) | Severe: Injury + Fatal | | | Non-Injury | | |
|---|---|---|---|---|---|---|
| | | 95% Credible Intervals | | | 95% Credible Intervals | |
| | Est. | Lower Limit | Upper Limit | Est. | Lower Limit | Upper Limit |
| **Model 1A: Multivariate model with only heterogeneous effects** | | | | | | |
| **Intercept** ($\beta_0$) | -4.148 | -6.209 | -2.225 | -3.100 | -4.935 | -1.454 |
| **ln(AADT)** ($\beta_2$) | 0.547 | 0.343 | 0.765 | 0.541 | 0.373 | 0.742 |
| **UNID** ($\beta_4$) | 0.079 | 0.055 | 0.101 | 0.077 | 0.055 | 0.096 |
| **IUND** ($\beta_5$) | 0.230 | 0.051 | 0.424 | 0.274 | 0.112 | 0.443 |
| **IBUS** ($\beta_6$) | 0.254 | 0.063 | 0.433 | 0.333 | 0.170 | 0.488 |
| **NL** ($\beta_7$) | 0.144 | 0.067 | 0.228 | 0.150 | 0.081 | 0.217 |
| $\sigma^2_{kk}$ | 0.335 | 0.265 | 0.419 | 0.286 | 0.234 | 0.348 |
| **Model 1B: Multivariate model with only spatial correlation** | | | | | | |
| **Intercept** ($\beta_0$) | *-2.140* | *-3.877* | *0.039* | *-1.890* | *-3.468* | *0.163* |
| **ln(AADT)** ($\beta_2$) | 0.371 | 0.126 | 0.550 | 0.447 | 0.216 | 0.610 |
| **UNID** ($\beta_4$) | 0.067 | 0.047 | 0.089 | 0.074 | 0.056 | 0.093 |
| **IUND** ($\beta_5$) | *0.135* | *-0.036* | *0.310* | 0.158 | 1.41E-04 | 0.318 |
| **IBUS** ($\beta_6$) | 0.244 | 0.039 | 0.426 | 0.292 | 0.104 | 0.452 |
| **NL** ($\beta_7$) | 0.127 | 0.057 | 0.226 | 0.115 | 0.056 | 0.205 |
| $\sigma^2_{skk}$ | 1.011 | 0.792 | 1.275 | 0.903 | 0.734 | 1.107 |
| **Model 1C: Multivariate model with both heterogeneous effects and spatial correlation** | | | | | | |
| **Intercept** ($\beta_0$) | -2.607 | -4.372 | -0.837 | -2.142 | -3.624 | -0.781 |
| **ln(AADT)** ($\beta_2$) | 0.418 | 0.234 | 0.601 | 0.475 | 0.334 | 0.627 |
| **UNID** ($\beta_4$) | 0.067 | 0.043 | 0.089 | 0.072 | 0.052 | 0.091 |
| **IUND** ($\beta_5$) | *0.149* | *-0.020* | *0.334* | 0.169 | 0.026 | 0.330 |
| **IBUS** ($\beta_6$) | 0.249 | 0.080 | 0.412 | 0.299 | 0.162 | 0.441 |
| **NL** ($\beta_7$) | 0.121 | 0.053 | 0.191 | 0.108 | 0.050 | 0.173 |
| $\sigma^2_{kk}$ | 0.180 | 0.071 | 0.266 | 0.113 | 0.025 | 0.188 |
| $\sigma^2_{skk}$ | 0.298 | 0.108 | 0.735 | 0.364 | 0.150 | 0.765 |
| **Chi Observed** | 172.900 | 71.460 | 272.100 | 140.400 | 58.600 | 239.300 |
| **Chi Replicated** | 168.100 | 68.960 | 267.200 | 139.400 | 59.570 | 235.300 |
| ***p*** | 0.380 | 0 | 1 | 0.465 | 0 | 1 |

* *Parameter estimates not significant under the stated level of significance are shown in italic font.*

### 4.3.3 Variance and Correlation

For the Richmond dataset, the heterogeneous variance ($\sigma_{kk}^2$) and spatial variance ($\sigma_{skk}^2$) estimates of severe and non-injury collisions for both heterogeneous effects and spatial correlation, respectively, were statistically significant at the 95% credible interval. According to parameter estimates, the heterogeneous variance for severe collisions was 0.246 (standard deviation (sd): 0.065), while it was 0.297 (sd: 0.079) for non-injury collisions. The spatial variance was smaller (0.027 (sd: 0.038) for severe and 0.061(sd: 0.1) for non-injury collisions) than the heterogeneous variance, as most of the variations were most likely captured by heterogeneous effects. These results demonstrate the presence of over-dispersion in both severe and non-injury collisions. The covariance $(\sigma_{hk}^2)$ for heterogeneous effects was significant (0.262; 95% credible intervals (CI): 0.155, 0.402; sd: 0.063), while the spatial covariance for spatial correlation (0.0181; 95% CI: -0.027, 0.001; sd: 0.048) was insignificant. A possible rationale for this finding is that the number of spatial units was small and most of the variations (90.2% for severe and 83.1% for no injury) were most likely captured by heterogeneous effects.

The posterior correlation between severe and non-injury collisions for heterogeneous effects was quite high (0.976; 95% CI: 0.909, 0.996), although the correlation for spatial correlation was low (0.138; 95% CI: -0.882, 0.979). Since the heterogeneous effects for unobserved heterogeneity dominated, the correlation between severe and non-injury collisions for the total random effects (heterogeneous effects and spatial correlation) was also quite high and obviously significant (0.905; 95% CI: 0.686, 0.983). In conclusion, a higher number of non-injury collisions is associated with higher severe collisions, as the collision likelihood for both levels is likely to rise due to the same deficiencies in roadway design, similar weather conditions and other unobserved factors. These results are in line with the literature findings. For instance, in an area-wide multivariate spatial analysis, Aguero-Valverde (2013) estimated significant and very high correlations, 0.688 for spatial errors and 0.962 for heterogeneity, between injury and non-injury collisions.

The effects of spatial correlation associated with multivariate models becomes much clearer in the Vancouver dataset. For the dataset, both the heterogeneous variance ($\sigma_{kk}^2$) and spatial variance $(\sigma_{skk}^2)$ estimates of severe and non-injury collisions were statistically significant at the

95% credible interval. These demonstrate the presence of over-dispersion and spatial variation in both severe and non-injury collisions. According to parameter estimates, the heterogeneous variance for severe collisions was 0.18 (95% CI: 0.071, 0.18; sd: 0.047), while it was 0.113 (95% CI: 0.025, 0.117; sd: 0.04) for non-injury collisions. The spatial variance was 0.298 (95% CI: 0.108, 0.259; sd:0.156) for severe collisions, while it was 0.364 (95% CI: 0.15, 0.337; sd:0.157) for non-injury collisions, which was quite high compared to heterogeneous variance. For severe collisions, about 62.3% of the total variation was captured by spatial correlation, and it was even higher for non-injury collisions at about 76.2%. Previously, the spatial variation was small for the Richmond dataset, which indicates that the spatial variation might increase with an increase in the number of road segments or samples. Similarly, the covariance $(\sigma_{hk}^2)$ for heterogeneous effects (0.138; 95% CI: 0.039, 0.141; sd: 0.042) was quite small compared to the spatial covariance for spatial correlation (0.314; 95% CI: 0.117, 0.282; sd: 0.154).

The posterior correlation between severe and non-injury collisions for heterogeneous effects was quite high (0.967; 95% CI: 0.898, 0.975). Along the same lines, the correlation for spatial effects was also high (0.949; 95% CI: 0.846, 0.961). As both the correlations were quite high, the correlation for the total random effects (heterogeneous effects and spatial correlation) was also quite high and clearly significant (0.945; 95% CI: 0.894, 0.948), indicating that a higher number of non-injury collisions is associated with a higher number of severe collisions.

### 4.3.4 Comparison of Multivariate Models with Univariate Models

For the Richmond dataset, Table 7 summarizes the parameter estimates and their associated statistics for univariate PLN spatial models. All of the parameters were significant and positively correlated with both severe and non-injury collisions, except UNID. The estimated parameters were quite similar to the multivariate models. According to the parameter estimates, there was a 8.2% increase in predicted severe collisions per 10% increase in traffic volume. Similarly, there was a 9.02% increase in predicted non-injury collisions per 10% increase in traffic volume. The number of crosswalks was also significant and positively correlated with both severe and injury collisions. Interestingly, UNID became insignificant for both severe and non-injury collisions in univariate models; UNID were previously significant for non-injury collisions in multivariate models.

**Table 7: Parameter estimates and 95% credible intervals for univariate PLN models (Richmond dataset)**

| Variable (Parameter) | Severe: Injury + Fatal | | | Non-Injury | | |
|---|---|---|---|---|---|---|
| | | 95% Credible Intervals | | | 95% Credible Intervals | |
| | Est. | Lower Limit | Upper Limit | Est. | Lower Limit | Upper Limit |
| **Intercept** $(\beta_0)$ | -4.227 | -7.717 | -0.611 | -4.238 | -8.047 | -0.492 |
| **ln(AADT)** $(\beta_2)$ | 0.599 | 0.231 | 0.950 | 0.643 | 0.260 | 1.034 |
| **Crosswalks** $(\beta_3)$ | 0.257 | 0.134 | 0.373 | 0.219 | 0.100 | 0.337 |
| **UNID** $(\beta_4)$ | *0.081* | *-0.009* | *0.170* | *0.090* | *-8.36E-04* | *0.183* |
| $\sigma_u^2$ | 0.208 | 6.31E-04 | 0.399 | 0.193 | 4.05E-04 | 0.442 |
| $\sigma_s^2$ | 0.122 | 2.96E-04 | 0.944 | 0.381 | 3.49E-04 | 1.443 |
| **Proportion of Spatial Variation** $(\psi_s)$ | 0.196 | 0.019 | 0.950 | 0.378 | 0.019 | 0.966 |
| **DIC** | 393.800 | | | 426.700 | | |
| **Total DIC (Severe+ No Injury)** | 820.500 | | | | | |

* *Parameter estimates not significant under the stated level of significance are shown in italic font.*

The heterogeneous variance $(\sigma_u^2)$ and spatial variance $(\sigma_s^2)$ estimates of severe and non-injury collisions for both heterogeneous effects and spatial correlation, respectively, were statistically significant at the 95% credible interval. These demonstrate the presence of overdispersion and spatial correlation in both severe and non-injury collisions. The spatial variance was quite high for non-injury collisions, compared to severe collisions. Approximately 19.6% (for severe collisions) and 37.8% (for non-injury collisions) of the variation for random effects is explained by the spatial correlation, which was higher than in multivariate models for both the collisions.

Similarly, for the Vancouver dataset, Table 8 summarizes the parameter estimates and their associated statistics for univariate PLN spatial models. All of the parameters (i.e., AADT, UNID, IBUS and NL) were significant and positively correlated with both severe and non-injury collisions, except IUND for severe collisions.

As with the previous models, the estimated parameters are quite similar to the multivariate models and obviously intuitive and in line with previous research. The heterogeneous variance $(\sigma_u^2)$ and spatial variance $(\sigma_s^2)$ estimates of severe and non-injury collisions for both heterogeneous effects and spatial correlation, respectively, were statistically significant at the 95% credible interval. These demonstrate the presence of over-dispersion and spatial correlation in both severe and non-injury collisions. The spatial variance was quite high for non-injury collisions compared to severe collisions. Approximately 42.3% (for severe collisions) and 66% (for non-injury collisions) of the variation of the random effects is explained by the spatial correlation, which was smaller than in multivariate models for both the collisions.

**Table 8: Parameter estimates and 95% credible intervals for univariate PLN spatial models (Vancouver dataset)**

| Variable (Parameter) | Severe: Injury + Fatal | | | Non-Injury | | |
|---|---|---|---|---|---|---|
| | | 95% Credible Intervals | | | 95% Credible Intervals | |
| | Est. | Lower Limit | Upper Limit | Est. | Lower Limit | Upper Limit |
| **Intercept** ($\beta_0$) | -2.566 | -4.560 | -0.629 | -1.992 | -3.760 | -0.142 |
| **ln(AADT)** ($\beta_2$) | 0.397 | 0.190 | 0.607 | 0.460 | 0.269 | 0.654 |
| **UNID** ($\beta_4$) | 0.068 | 0.044 | 0.091 | 0.072 | 0.053 | 0.090 |
| **IUND** ($\beta_5$) | *0.153* | *-0.035* | *0.340* | 0.170 | 0.016 | 0.326 |
| **IBUS** ($\beta_6$) | 0.271 | 0.094 | 0.452 | 0.300 | 0.133 | 0.456 |
| **NL** ($\beta_7$) | 0.155 | 0.073 | 0.231 | 0.110 | 0.037 | 0.180 |
| $\sigma_u^2$ | 0.213 | 0.134 | 0.296 | 0.077 | 7.39E-04 | 0.170 |
| $\sigma_s^2$ | 0.173 | 0.052 | 0.395 | 0.498 | 0.173 | 0.929 |
| **Proportion of Spatial Variation** $(\psi_s)$ | 0.423 | 0.299 | 0.543 | 0.660 | 0.461 | 0.954 |
| **DIC** | 1659 | | | 2005 | | |
| **Total DIC (Severe+ No Injury)** | 3664 | | | | | |

* *Parameter estimates not significant under the stated level of significance are shown in italic font.*

For the Richmond dataset, the DIC statistics were 393.8 for severe and 426.7 for non-injury collisions. Thus, the multivariate spatial models provided a superior fit over the two univariate spatial models, as the DIC of the multivariate spatial model (Model 1C) (785.2) was much smaller than the sum of the univariate DICs (820.5); this shows a very significant drop-off of 35.3. Similarly, for the Vancouver dataset, the multivariate spatial models (Model 1C) provided a superior fit over the two univariate spatial models, as the DIC of the multivariate spatial model (Model 1C) (3548) was much smaller than the sum of the univariate DICs (3664); this shows a very significant drop-off of 116. These results are in line with previous research findings (El-

Basyouny and Sayed, 2009b; Park and Lord, 2007). The literature established that multivariate PLN models are more precise than univariate PLN models. The improvement in precision is due mainly to the correlation between the latent variables (severe and no injury).

## 4.4 Summary

This chapter investigated the inclusion of spatial correlation at different collision severity levels, using multivariate modelling approach. Three different modelling formulations (multivariate model with only heterogeneous effects, multivariate model with only spatial correlation and multivariate model with both heterogeneous effects and spatial correlation) were applied to take into account spatial correlation in a multivariate framework. The multivariate model with both heterogeneous effects and spatial correlation (Model 1C) was found to yield the best results in terms of the DIC values.

For the Richmond dataset, three variables were used in the models (AADT, number of crosswalks and UNID). The covariates were significant at a 95% confidence level (except UNID for severe collisions) and positively correlated with both severe and non-injury collisions. Similarly, for the Vancouver dataset, the regression coefficients were all positive, indicating that factors such as AADT, UNID, IBUS and NL were positively associated with both the severe and non-injury collisions. Another indicator variable, IUND, was significant and positively related with only non-injury collisions. Both the results were quite intuitive and in line with previous research findings.

The estimates of the heterogeneous variance and spatial variance were significant and indicate the presence of over-dispersion and spatial correlation in both datasets. For the Richmond dataset, the spatial variance was smaller than the heterogeneous variance, as most of the variations (90.2% for severe and 83.1% for no injury) were captured by the heterogeneous effects or unobserved heterogeneity. Unlike the results from the Richmond dataset, the Vancouver dataset exhibits better inference in terms of capturing spatial variation, as about 62.3% and 76.2% of the total variation for severe and non-injury collisions were captured by spatial correlation. These indicate that spatial variation or spatial correlation might increase with an increase in the number of road segments or samples, as the Vancouver dataset consists of 281 road segments, while there were only 72 road segments in the Richmond dataset.

For the Richmond dataset, the posterior correlation between severe and non-injury collisions for heterogeneous effects was quite high, while the correlation for spatial effects was low. However, the correlation between severe and non-injury collisions for the total random effects (i.e., heterogeneous effects and spatial correlation) was significant and quite high (0.905). Similarly, for the Vancouver dataset, the posterior correlation between severe and non-injury collisions for the total random effects was also quite high (0.945) and obviously significant, indicating that a greater number of non-injury collisions is associated with a greater number of severe collisions, as the collision likelihood for both types is likely to rise due to similar deficiencies in roadway design or other unobserved confounding factors.

This chapter also demonstrates the importance of multivariate spatial modelling techniques by comparing the multivariate spatial models with independent univariate PLN spatial models, with respect to model inference and goodness-of-fit for both the datasets. All of the estimated parameters for univariate models were quite similar to multivariate models, except UNID for the Richmond dataset. Interestingly, UNID became insignificant for both severe and non-injury collisions in univariate models; UNID was previously significant for only non-injury collisions in multivariate models. Multivariate spatial models provide a superior fit over the two univariate PLN spatial models, with a very significant drop-off in DIC (35.3 for Richmond dataset and 116 for Vancouver dataset). These results advocate the use of multivariate PLN spatial models with both heterogeneous effects and spatial correlation over univariate PLN spatial models for collision severity analysis.

From a model application point of view, the methodology proposed herein has the potential to provide new insight into collision severity analysis and can be used to estimate the associated safety risks in terms of severity, with consideration for the spatial correlations in the data.

The results presented in this chapter are based on datasets covering a period of 1994 to 1996. Even though these results conform to those in the literature, further research with recent and different datasets is required to confirm the research findings. Further, the effects of sample size in the analysis of spatial correlation are yet to be investigated, as the effects were not similar for both the datasets.

# 5. UNIVARIATE RANDOM PARAMETERS SPATIAL MODELS

This chapter illustrates the effects of including spatial correlation in univariate random parameters collision count-data models.

## 5.1 Background

Most of the previous research related to the development of collision models focuses on accounting for Poisson variation (Jovanis and Chang, 1986; Joshua and Garber, 1990; Miaou and Lum, 1993; Miaou, 1994) and heterogeneity (Maycock and Hall, 1984; Hauer et al., 1988; Persaud, 1994; Maher and Summersgill, 1996; Milton and Mannering, 1998; Miaou and Lord, 2003; El-Basyouny and Sayed, 2006) in collision data. The parameters of these traditional collision models were assumed to be fixed when they can actually vary across observations (road segments or intersections). Further, due to unobserved heterogeneity, the effect of explanatory variables on collision may vary for different observations. For instance, a two-lane road segment with high traffic volume may have a high collision frequency compared to a similar road configuration with lower traffic volume. Therefore, constraining the parameters may not incorporate site-specific effects, leading to an underestimation of standard errors and inconsistent, biased and erroneous inference (Washington et al., 2003). Given the magnitude of this potential issue, several researchers have successfully applied random parameters to collision modelling. For instance, Milton et al. (2008), Gkritza and Mannering (2008), Anastasopoulos and Mannering (2009; 2011), El-Basyouny and Sayed (2009a), Anastasopoulos et al. (2012a), and Russo et al. (2014) all demonstrated that the random parameters model can provide more accurate inference than traditional fixed parameter models, as well as account for heterogeneity across sites due to unobserved road geometrics, traffic characteristics, environmental factors, driver behaviour and other confounding factors.

Spatial correlation is another key issue that has been gaining attention in the development of collision models (Amoros et al., 2003; Noland and Quddus, 2004; Abdel-Aty and Wang, 2006; Aguero-Valverde and Jovanis, 2006; 2008; 2010; Quddus, 2008; El-Basyouny & Sayed, 2009c; Mitra, 2009; Aguero-Valverde, 2013; Flask and Schneider, 2013). Despite the fact that collision data are collected with respect to location—which is measured as points in space (Quddus, 2008), illustrating that a spatial correlation exists between observation sites (LeSage, 1998)—no studies have taken these correlations into account using random parameters models. Moreover, previous studies have used random parameters and spatial correlation separately; rarely in the literature have these two factors been incorporated together in collision models. Nevertheless, it is necessary to investigate whether the inclusion of spatial correlation in a random parameters model can improve the model's goodness-of-fit and precision of parameter estimation. This approach may help gain new insight in collision analysis and allow for a more precise assessment of the safety risks of collision sites. Thus, this chapter discusses the research efforts to investigate the effects of spatial correlation in random parameters collision count-data models. To perform the investigation, three years (1994–1996) of collision data and other geometric and non-geometric road data were used for two cities in British Columbia, Canada: i) Richmond, and ii) Vancouver.

In determining the model formulation for this research, several different approaches in the literature were reviewed for suitability with respect to random parameters and spatial correlation. From a methodological perspective, a wide variety of modelling approaches, such as the negative binomial model (Anastasopoulos and Mannering, 2009; Chen and Tarko, 2014; Garnowski and Manner, 2011; Ukkusuri et al., 2011; Venkataraman et al., 2011; 2013; Wu et al., 2013;), logit model (Anastasopoulos and Mannering, 2011), mixed logit model (Gkritza and Mannering, 2008; Milton et al., 2008), Tobit model (Anastasopoulos et al., 2012a), Poisson-lognormal (PLN) model (El-Basyouny and Sayed, 2009a), bivariate ordered probit model (Russo et al., 2014) and finite mixture model (Xiong and Mannering, 2013) were used to employ random parameters in collision analysis. To account for spatial correlation, almost all earlier studies (Aguero-Valverde and Jovanis, 2006; 2008; 2010; Ahmed et al., 2011; Guo et al., 2010; Mitra, 2009; Siddiqui et al., 2010) used Gaussian Conditional Auto-regressive (CAR) (Besag et al., 1991) distribution.

Given the support in the literature, random parameters spatial models were developed using CAR distribution. As the PLN modelling approach is more flexible than the traditional Poisson-Gamma or negative binomial model to handle over-dispersion, the lognormal distribution was used for heterogeneous effects, leading to PLN posterior distribution. The models were estimated in a FB context via MCMC simulation (Gilks et al., 1996). As WinBUGS (Lunn et al., 2000) is a flexible platform for the Bayesian analysis of complex statistical models using MCMC methods, this open source statistical software was used for the development of the proposed random parameters spatial models.

## 5.2 Methodology

The specifications of the univariate random parameters models are given by Eqs. *(30–36)*. The specifications for spatial component $S_i$ are given by Eqs. *(24–26)*. A detailed specification of CAR distribution was described in section 4.2.1 of the previous chapter. To assess the effects of including spatial correlation in a univariate random parameters model, three different modelling formulations were used and compared in terms of their inferences and goodness-of-fit.

**Model 2A: Random parameters model with only heterogeneous effects**

$$\ln(\theta_i) = \beta_{0i} + \sum_{m=1}^{M} \beta_{mi} X_{mi} + \varepsilon_i \tag{62}$$

**Model 2B: Random parameters model with only spatial correlation**

$$\ln(\theta_i) = \beta_{0i} + \sum_{m=1}^{M} \beta_{mi} X_{mi} + S_i \tag{63}$$

**Model 2C: Random parameters model with both heterogeneous effects and spatial correlation**

$$\ln(\theta_i) = \beta_{0i} + \sum_{m=1}^{M} \beta_{mi} X_{mi} + \varepsilon_i + S_i \tag{64}$$

To obtain FB estimates, it is required to specify prior distribution for the parameters. The most commonly used priors are diffused normal distributions (with zero mean and large variance) for the regression parameters, $Gamma(\varepsilon,\varepsilon)$ or $Gamma(1,\varepsilon)$ for $\sigma_u^{-2}$ and $\sigma_m^{-2}$, where $\varepsilon$ is a small number, e.g., 0.01 or 0.001. Different priors for the univariate CAR model were specified in the literature. For instance, Quddus (2008) used a gamma prior with $Gamma(0.5, 0.0005)$; Aguero-Valverde and Jovanis (2008; 2010) used a fair prior to specify the relationship between uncorrelated random effects and spatial effects; El-Basyouny and Sayed (2009c) used a proper prior gamma $(1+\sum l_i/2, 1+n/2)$ where $l_i = n_i S_i (S_i - \overline{S_i})$. In the current research, $Gamma\ (1,0.001)$ was used for $\sigma_s^{-2}$. It should be noted that introducing vague priors for all unknown parameters can be risky under certain conditions, such as the combination of low mean and small sample size (Miranda-Moreno and Lord, 2007). In such cases, better results can be obtained by using semi-informative priors with a small mean and variance for the dispersion parameter. The posterior distributions needed in the FB approach are obtained using Markov Chain Monte Carlo (MCMC) sampling, which is available in statistical software called WinBUGS (Lunn et al., 2000). MCMC methods are used to repeatedly sample from the joint posterior distribution.

The DIC and chi-square statistics were used to assess the models' goodness-of-fit. The chi-square statistics are computed from the following equation:

$$\chi^2 = \sum_{i=1}^{n} \left[ \frac{\left[y_i - E(Y_i)\right]^2}{Var(Y_i)} \right] \tag{65}$$

The specifications of mean ( $E(Y_i)$ ) and variance ( $Var(Y_i)$ ) were given by Eqs. *(33–36)*, where $\sigma_i^2$ for univariate random parameters can be expressed as

$$\sigma_i^2 = \sigma_0^2 + \sum_{m=1}^{M} X_{mi}^2 \sigma_m^2 + \sigma_u^2 + \sigma_s^2 \tag{66}$$

## 5.3 Results and Discussion

### 5.3.1 Model Selection

For each model, the posterior estimates were obtained via two chains with 20,000 iterations, 5,000 of which were excluded as a burn-in sample using WinBUGS. The BGR statistics were less than 1.2; the ratios of the Monte Carlo errors relative to the standard deviations of the estimates were less than 0.05; and trace plots for all of the model parameters indicated convergence. The model selection criterion is presented in Table 9. As observed in Table 9, all three random parameters models (2A, 2B and 2C) for both datasets (i.e., Richmond and Vancouver) were quite similar to one another. Furthermore, as the difference in DIC is less than five, it could be misleading to report only the model with the lowest DIC. Therefore, a predictive posterior approach was used to assess the models' goodness-of-fit.

**Table 9: The DIC statistics by model**

| Model Description | DIC | |
|---|---|---|
| | Richmond | Vancouver |
| Model 2A: Random parameters model with only heterogeneous effects | 470.9 | 2104 |
| Model 2B: Random parameters model with only spatial correlation | 470.9 | 2101 |
| Model 2C: Random parameters model with both heterogeneous effects and spatial correlation | 470.6 | 2102 |

To assess the models' goodness-of-fit, Pearson's residuals in Eq. *(65)* were examined, and no anomalies were detected. The distribution of the chi-square discrepancy measure in replicated datasets was generated. From the posterior estimates, the observed values of the chi-square for all three models were located near the centre of the replicated distribution.

For the Richmond dataset, the observed chi-square statistics (Table 10) were 82.34 (Model 2A), 78.6 (Model 2B) and 78.83 (Model 2C), with a replicated chi-square of 77.57, 74.56 and 74.98, respectively. The associated *p*-values estimated from the distributions of the chi-square discrepancy that was measured in the replicated datasets were 0.366, 0.382 and 0.383 for Model 2A, 2B and 2C, respectively. As mentioned in section four, a model does not fit the data if the observed value of the chi-square differs greatly from the predicted distribution; the discrepancy cannot be reasonably explained by chance if the *p*-values are close to zero or one. Since the associated *p*-values in this research were well distanced from zero and one, all the models performed well, accommodating the variation in collision frequency across different road characteristics. Therefore, all the models for the Richmond dataset were comparable to one another. However, as the *p*-values of Model 2B and Model 2C were a little high compared to Model 2A and were close to 0.5, random parameters models with spatial correlation and heterogeneous effects may provide better inferences and predict collisions more precisely. These will be discussed later in this section.

For the Vancouver dataset, the observed and replicated values of the chi-square statistics with associated *p*-values of Models 2A, 2B and 2C are shown in Table 11. According to the posterior estimates, the observed chi-square values were located near the centre of the replicated distribution, with associated *p*-values of 0.523 (Model 2A), 0.524 (Model 2B) and 0.519 (Model 2C). As the *p*-values were quite similar and close to neither zero nor one, all the models seem to fit well with the dataset. Therefore, all the models for the Vancouver dataset were comparable to one other.

### 5.3.2 Parameter Estimates

#### *5.3.2.1 Richmond Dataset*

Table 10 summarizes the parameter estimates and their 95% credible intervals for Model 2A, 2B and 2C. The table shows that the parameter estimates are significant, as the 95% credible intervals were bounded away from zero, except for UNID under Model 2A and Model 2B. Apart from the intercepts, the regression coefficients were all positive, indicating that factors such as segment length, AADT, number of crosswalks, and UNID were positively associated with the number of collisions.

The modelling results revealed that road segment length and AADT were statistically significant at a 95% confidence level and positively correlated with the number of collisions, which indicates that longer segments with higher traffic volumes (i.e., increased exposure) have more collisions. The segment length resulted in a random parameters that is normally distributed with a mean ranging from 0.89 to 1.11 and standard deviation (s1) ranging from 0.334 to 0.394. Thus, for almost all the sites, collision frequency was expected to increase with segment length, by varying magnitudes. For Model 2A the mean was 0.89 (s1: 0.334); whereas, for Model 2C it was 1.11 (s1: 0.334), which indicates that the inclusion of spatial correlation in a random parameters model with heterogeneous effects has substantial effects on estimated parameters. Similarly, the AADT resulted in a random parameters that is normally distributed with a mean of 0.662 (Model 2A), 0.697 (Model 2B), and 0.704 (Model 2C). Further, there was a small reduction (3.15%) in standard deviation from Model 2A to Model 2C, which may increase the predictability of Model 2C. For all the sites, the collision counts are expected to increase with AADT. These findings are intuitive and in line with previous research findings (El-Basyouny and Sayed, 2009a). As noted by Anastasopoulos and Mannering (2009), this AADT finding is likely indicating a complex interaction among traffic volume, driver behaviour and collision frequency. It may be capturing, among other factors, the response and adaption of drivers to various levels of traffic volumes.

**Table 10: Parameter estimates and 95% credible intervals for Models 2A-2C (Richmond dataset)**

| Variable (Parameter) | Model 2A Est. | Model 2A 95% Credible Intervals Lower Limit | Model 2A 95% Credible Intervals Upper Limit | Model 2B Est. | Model 2B 95% Credible Intervals Lower Limit | Model 2B 95% Credible Intervals Upper Limit | Model 2C Est. | Model 2C 95% Credible Intervals Lower Limit | Model 2C 95% Credible Intervals Upper Limit |
|---|---|---|---|---|---|---|---|---|---|
| **Intercept ($\beta_0$)** | -4.025 | -7.300 | -1.105 | -4.354 | -7.619 | -1.122 | -4.399 | -7.223 | -1.484 |
| **ln(Length) ($\beta_1$)** | 0.890 | 0.129 | 1.764 | 0.927 | 0.234 | 1.713 | 1.112 | 0.376 | 1.901 |
| **ln(AADT) ($\beta_2$)** | 0.662 | 0.359 | 1.004 | 0.697 | 0.359 | 1.036 | 0.704 | 0.401 | 0.995 |
| **Crosswalks ($\beta_3$)** | 0.265 | 0.123 | 0.410 | 0.258 | 0.087 | 0.404 | 0.248 | 0.108 | 0.383 |
| **UNID ($\beta_4$)** | *0.106* | *-0.013* | *0.231* | *0.108* | *-0.005* | *0.220* | 0.1133 | 0.008 | 0.233 |
| **Standard deviation, s0** | 0.449 | 0.174 | 0.618 | 0.457 | 0.275 | 0.620 | 0.446 | 0.151 | 0.625 |
| **Standard deviation, s1** | 0.334 | 0.080 | 1.084 | 0.394 | 0.091 | 1.161 | 0.334 | 0.073 | 1.014 |
| **Standard deviation, s2** | 0.254 | 0.066 | 0.628 | 0.253 | 0.074 | 0.608 | 0.246 | 0.067 | 0.598 |
| **Standard deviation, s3** | 0.156 | 0.060 | 0.330 | 0.173 | 0.068 | 0.346 | 0.163 | 0.065 | 0.335 |
| **Standard deviation, s4** | 0.166 | 0.065 | 0.337 | 0.160 | 0.064 | 0.325 | 0.149 | 0.059 | 0.296 |
| $\sigma_u^2$ | 0.015 | 0.0003 | 0.187 | | | | 0.014 | 0.0003 | 0.185 |
| $\sigma_s^2$ | | | | 0.018 | 0.0003 | 0.197 | 0.020 | 0.0003 | 0.254 |
| **Proportion of spatial variance** | | | | | | | 0.383 | 0.062 | 0.899 |
| **Chi observed** | 82.34 | 42.280 | 138.200 | 78.600 | 32.070 | 133.500 | 78.830 | 39.710 | 131.200 |
| **Chi replicated** | 77.57 | 39.070 | 137 | 74.560 | 30.630 | 132.500 | 74.980 | 37.360 | 131.300 |
| ***p*** | 0.366 | 0 | 1 | 0.382 | 0 | 1 | 0.383 | 0 | 1 |

* *Parameter estimates not significant under the stated level of significance are shown in italics.*

The number of crosswalks was also significant and positively correlated with collision frequency, indicating that an increase in the number of crosswalks raises the associated safety risk. This is expected, as the presence of crosswalks allows for more pedestrian activity, resulting in higher collision probability and risk. For all three models, the mean of the parameters and the associated standard deviations were quite similar. The parameter also resulted in a random parameters that is normally distributed, with a mean ranging from 0.248 to 0.265 and standard deviation (s3) ranging from 0.156 to 0.173. For almost all sites, collision frequency was expected to increase with the number of crosswalks.

UNID were insignificant for Model 2A and Model 2B. Interestingly, UNID became significant for Model 2C when considering both spatial correlation and heterogeneous effects. The positive correlation associated with collision frequency indicated that an increase of UNID results in more collisions. The literature also suggests that road segments with a large number of access points or unsignalized intersections have a significant impact on, and are positively correlated with, collision frequency (Xuesong and Ming, 2012). For Model 2C, the parameter was normally distributed with a mean of 0.113 and standard deviation (s4) of 0.149. Collision frequency was expected to increase with UNID for a vast majority (99%) of the sites and was expected to decrease for only a small proportion (1%) of the sites, reflecting heterogeneity across sites.

The estimates of variance ($\sigma_u^2$) for heterogeneous effects were significant at a 95% confidence level under Model A and Model C, demonstrating the presence of overdispersion in the data. The variances were reasonably small, as accounting for site variation reduces the estimates of extra-Poisson variation. The estimates of spatial variance ($\sigma_s^2$) were significant at a 95% confidence level under Model 2B and Model 2C, and about 38.3% of the total variability was explained by spatial correlation under Model 2C. A possible rationale for this finding is that the number of spatial units was small and most of the variations (61.7%) were most likely captured by heterogeneous effects and site variation.

#### *5.3.2.2 Vancouver Dataset*

Table 11 summarizes the parameter estimates and their 95% credible intervals for Model 2A, 2B and 2C. The table shows that the parameter estimates are significant, as the 95% credible intervals were bounded away from zero. The Vancouver dataset had more significant variables than the Richmond dataset. Apart from the intercepts, the regression coefficients were all positive, indicating that factors such as segment length, AADT, UNID, IUND, IBUS and NL were positively associated with the number of collisions. It is worth mentioning that the means of the parameters for Model 2A were slightly high when compared to Model 2B and Model 2C, except for segment length and IBUS. Furthermore, the parameters of Model 2B and Model 2C were quite similar to each other, which indicates that accounting for spatial variation may explain some variability in collision frequency and improve model fit.

The modelling results revealed that road segment length resulted in a random parameters that is normally distributed with similar means ranging from 0.904 to 0.948 and standard deviation (s1) ranging from 0.172 to 0.239. Thus, for almost all the sites, collision frequency was expected to increase with segment length, although by varying magnitudes. Similarly, the AADT resulted in a random parameters that is normally distributed with means of 0.578 (Model 2A), 0.462 (Model 2B) and 0.444 (Model 2C). Further, a small reduction (2.38%) in standard deviation from Model 2A to Model 2C may increase the predictability of Model 2C. For all the sites, the collisions counts are expected to be increased with AADT.

UNID were also significant and positively correlated with collision frequency, indicating that an increase in UNID leads to a higher collision frequency. Similarly, the indicator variable, IUND, was significant and positively correlated with collision occurrence. This result is expected because undivided cross sections may increase conflict with the traffic of opposite lanes, potentially leading to head-on collisions. The estimates resulted in a random parameters that is normally distributed with means of 0.314 (Model 2A), 0.216 (Model 2B) and 0.229 (Model 2C). Further, a substantial reduction (30.5% for Model 2B and 26.9% for Model 2C) in standard deviation from Model 2A to Model 2B and Model 2C may increase the model’s predictability.

**Table 11: Parameter estimates and 95% credible intervals for Models 2A-2C (Vancouver dataset)**

| | Model 2A | | | Model 2B | | | Model 2C | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 95% Credible Intervals | | | 95% Credible Intervals | | | 95% Credible Intervals | |
| **Variable (Parameter)** | **Est.** | **Lower Limit** | **Upper Limit** | **Est.** | **Lower Limit** | **Upper Limit** | **Est.** | **Lower Limit** | **Upper Limit** |
| **Intercept** ( $\beta_0$ ) | -3.190 | -4.928 | -1.729 | -1.801 | -3.589 | -0.087 | -1.639 | -3.134 | -0.111 |
| **ln(Length)** ( $\beta_1$ ) | 0.904 | 0.771 | 1.052 | 0.948 | 0.810 | 1.075 | 0.945 | 0.810 | 1.080 |
| **ln(AADT)** ( $\beta_2$ ) | 0.578 | 0.430 | 0.767 | 0.462 | 0.276 | 0.651 | 0.444 | 0.286 | 0.596 |
| **UNID** ( $\beta_4$ ) | 0.082 | 0.057 | 0.107 | 0.071 | 0.048 | 0.096 | 0.070 | 0.047 | 0.093 |
| **IUND** ( $\beta_5$ ) | 0.314 | 0.133 | 0.492 | 0.216 | 0.036 | 0.379 | 0.229 | 0.041 | 0.389 |
| **IBUS** ( $\beta_6$ ) | 0.288 | 0.110 | 0.471 | 0.273 | 0.130 | 0.427 | 0.310 | 0.140 | 0.453 |
| **NL** ( $\beta_7$ ) | 0.137 | 0.075 | 0.207 | 0.123 | 0.049 | 0.190 | 0.126 | 0.057 | 0.185 |
| **Standard deviation, s0** | 0.420 | 0.331 | 0.497 | 0.267 | 0.119 | 0.385 | 0.234 | 0.083 | 0.365 |
| **Standard deviation, s1** | 0.172 | 0.069 | 0.344 | 0.206 | 0.051 | 0.448 | 0.239 | 0.088 | 0.468 |
| **Standard deviation, s2** | 0.168 | 0.067 | 0.336 | 0.167 | 0.064 | 0.314 | 0.164 | 0.072 | 0.321 |
| **Standard deviation, s3** | 0.078 | 0.049 | 0.110 | 0.062 | 0.043 | 0.087 | 0.063 | 0.043 | 0.089 |
| **Standard deviation, s4** | 0.350 | 0.103 | 0.681 | 0.243 | 0.065 | 0.509 | 0.256 | 0.086 | 0.509 |
| **Standard deviation, s5** | 0.209 | 0.076 | 0.501 | 0.234 | 0.091 | 0.468 | 0.227 | 0.066 | 0.471 |
| **Standard deviation, s6** | 0.094 | 0.054 | 0.150 | 0.086 | 0.050 | 0.140 | 0.088 | 0.052 | 0.144 |
| $\sigma_u^2$ | 0.004 | 0.0003 | 0.027 | | | | 0.011 | 0.0003 | 0.079 |
| $\sigma_s^2$ | | | | 0.234 | 0.091 | 0.463 | 0.234 | 0.092 | 0.457 |
| **Proportion of spatial variance** | | | | | | | 0.838 | 0.541 | 0.955 |
| **Chi observed** | 225.100 | 167.900 | 292.900 | 156.900 | 84.070 | 241.300 | 156.200 | 83.690 | 243.100 |
| **Chi replicated** | 226 | 168.600 | 295.100 | 157.300 | 84.870 | 242.500 | 156.500 | 85.010 | 244.400 |
| ***p*** | 0.523 | 0 | 1 | 0.524 | 0 | 1 | 0.519 | 0 | 1 |

Another indicator variable, IBUS, also displayed significance and positive correlation with collision occurrence. Business land use introduces commercial activities that attract vulnerable road users (e.g., pedestrians, cyclists), creating greater collision risk with these additional users on the road. The estimates resulted in a random parameters that is normally distributed with a similar mean ranging from 0.273 to 0.31 and standard deviation (s5) ranging from 0.209 to 0.234. For almost all the sites, collision frequency was expected to increase with the presence of business land use. NL also yielded a similar positive correlation with collision occurrence; a possible rationale for this finding is that more lanes increase traffic flow and conflict areas, thereby increasing the probability of collision occurrence. NL also yielded a random parameters that is normally distributed with a similar mean range (0.123 to 0.137) for Model 2A, 2B and 2C. Furthermore, there was a small reduction (8.5% for Model 2B and 6.38% for Model 2C) in standard deviation from Model 2A to Model 2B and Model 2C.

The estimates of variance ($\sigma_u^2$) for heterogeneous effects were significant at a 95% confidence level under Model 2A and Model 2C, demonstrating the presence of overdispersion in the Vancouver data. However, the variances ($\sigma_u^2$) for heterogeneous effects were reasonably small (0.004 for Model 2A and 0.011 for Model 2C), as accounting for site variation due to random parameters reduces the estimates of extra-Poisson variation. The estimates of spatial variance ($\sigma_s^2$) were significant at a 95% confidence level under Model 2B and Model 2C. Unlike the Richmond dataset, a high proportion of the total variability (83.8%) was explained by spatial correlation under Model 2C. Ignoring this large proportion of spatial correlation may lead to a biased and erroneous estimation of the model parameters. Therefore, the inclusion of spatial correlation in a random parameters model may significantly improve the precision of the estimates of the expected collision frequency.

## 5.4 Summary

This chapter describes the investigation of including spatial correlation in univariate random parameters collision count-data models. Three different modelling formulations (random parameters with only heterogeneous effects, random parameters with only spatial correlation, and random parameters with both heterogeneous effects and spatial correlation) were developed account for spatial correlation in a random parameters framework. The DIC values and chi-square statistics indicated that all the models were comparable with one another. However, the random parameters model with both heterogeneous effects and spatial correlation was found to yield the best inference in terms of parameter estimates and precision of the estimates for both the datasets.

For the Richmond dataset, four variables were used in the models (segment length, AADT, number of crosswalks and UNID). The covariates were significant at a 95% confidence level (except UNID in Model 2A and Model 2B) and positively correlated with collision occurrence. The results were intuitive and in line with previous research findings. In most cases, the inclusion of spatial correlation reduces the standard deviation of the random parameters. The estimates of heterogeneous variance and spatial variance were significant and indicated the presence of over-dispersion and spatial correlation in the data. About 38.3% of the total variability was explained by spatial correlation under Model C, as most of the variations (61.7%) were most likely captured by heterogeneous effects and site variation.

Six variables (segment length, AADT, UNID, IUND, IBUS and NL) were significant at a 95% confidence level and positively correlated with collision occurrence for the Vancouver dataset. The effects of spatial correlation were much clearer in this dataset, as a high proportion of the total variability (83.8%) was explained by spatial correlation under Model 2C, which indicates that overlooking this large proportion of spatial correlation may lead to a biased parameter estimation. In addition, the current results reinforce the findings of other studies in the literature, suggesting that spatial correlation when estimating regression coefficients could reduce bias resulting from the omission of spatial variables.

From a model application point of view, the methodological approach proposed herein has the potential to provide new insight into collision analysis and can be used to estimate the associated safety risks more precisely. Nevertheless, random parameters models are complex to estimate, and the inclusion of spatial correlation makes the estimation technique even more complex. This methodological approach will also help improve the precision of the estimates of expected collision frequency.

# 6. MULTIVARIATE RANDOM PARAMETERS SPATIAL MODELS

This chapter illustrates the research efforts devoted to investigating the effects of including spatial correlation in multivariate random parameters collision models at different collision severity levels. In addition, this chapter demonstrates the comparison between multivariate random parameters spatial models and their univariate counterpart.

## 6.1 Background

Due to the fact that random parameters models provide better parameter estimates and inferences compared to traditional fixed parameters models, the use of random parameters in collision modelling has been gaining attention over the past few years. For instance, Milton et al. (2008), Gkritza and Mannering (2008), Anastasopoulos and Mannering (2009; 2011), El-Basyouny and Sayed (2009a), Anastasopoulos et al. (2012a), and Russo et al. (2014) all demonstrated that the random parameters model can provide better inference than the traditional fixed parameters model and can explicitly account for heterogeneity across observations that is due to unobserved road geometrics, traffic characteristics, environmental factors, driver behaviour and other confounding factors.

Most of the literature used random parameters in a univariate modelling framework. Regardless of the fact that collision data is multivariate in nature and it is necessary to account for the likely correlation between collision counts at different levels of classification (Bijleveld, 2005; Ma and Kockelman, 2006; Ma et al., 2008; Park and Lord, 2007; El-Basyouny and Sayed, 2009b; El-Basyouny et al., 2014b), multivariate random parameters have rarely been explored in the literature. An earlier study of El-Basyouny and Sayed (2013a) used time-varying coefficients (random parameters) in multivariate collision models to identify and prioritize hotspots. Similarly, another study of El-Basyouny et al. (2014a) employed time-varying coefficients in multivariate collision type models to assess the effects of weather elements on seven crash types. A recent study of Dong et al. (2014) demonstrated the use of a multivariate random parameters zero-inflated negative binomial regression (MRZINB) model for jointly modelling crash counts. The authors found that the MRZINB model outperformed the fixed parameters zero-inflated negative binomial regression model and possesses most desirable statistical properties in terms of its ability to accommodate unobserved heterogeneity and excess zero counts in correlated data.

Almost all of the few multivariate random parameters safety models mentioned in the literature used heterogeneous effects in addition to random parameters to account for unobserved or unmeasured heterogeneity. The foremost motivations of those studies are to reduce the bias and inconsistent estimation, improve the precision of the estimates, and thereby increase the model's predictability. However, most of the studies ignored the effects of spatial correlation (one of the most potential issues) in collision models, which may lead to a biased parameters estimation of the model, as some of the unobserved factors are likely to be correlated over space and there might be some possible correlation among neighbouring sites.

To this end, there are two main tasks of this chapter: i) investigate the effects of including spatial correlation in multivariate random parameters models and their influence on the different collision severity levels; and ii) compare multivariate random parameters spatial models with independent (separate) univariate random parameters spatial models for each collision severity in terms of model inference and goodness-of-fit. To accomplish these tasks, three years (1994 to 1996) of collision data and other geometric and non-geometric road data were used for the city of Vancouver, British Columbia, Canada. CAR (Besag et al., 1991) distribution was used to account for spatial correlation, while lognormal distribution was used for the heterogeneous effects, leading to PLN posterior distribution. The models were estimated under a FB framework via MCMC simulation (Gilks et al., 1996). Statistical software, WinBUGS (Lunn et al., 2000), was used for the development of the proposed multivariate random parameters spatial models.

## 6.2 Methodology

Several studies proposed multivariate models for collision counts at different levels of classification (Bijleveld, 2005; Ma and Kockelman, 2006; Ma et al., 2008; Park and Lord, 2007; Aguero-Valverde & Jovanis, 2009; El-Basyouny and Sayed, 2009b; El-Basyouny et al., 2014a; 2014b). This chapter used similar multivariate methodology in a random parameters framework. The specifications for the multivariate models are given by Eqs. *(17–23)*, and those for the multivariate random parameters models are given by Eqs. *(37–43).* A detailed specification for the multivariate spatial model with CAR distribution was provided in section 4.2.1 in chapter four. The multivariate random parameters spatial PLN model can be written as follows:

$$\ln(\theta_i^k) = \beta_{0i}^k + \sum_{m=1}^{M} \beta_{mi}^k X_{mi} + \varepsilon_i^k + S_{ki} \qquad (67)$$

Under the multivariate random parameters PLN spatial model, the mean and variance are given by

$$E(Y_i^k) = \mu_i^{k*} \exp[0.5 * (\sigma_{ik}^2 + \sigma_{Skk}^2)] \quad (68)$$

$$Var(Y_i^k) = E(Y_i^k) + [E(Y_i^k)]^2 [\exp(\sigma_{ik}^2 + \sigma_{Skk}^2) - 1] \quad (69)$$

Several functional forms are available in the literature (Miaou & Lord, 2003). However, since the focus of this chapter is on demonstrating the consequences of including spatial correlation in a multivariate random parameters model, these three model forms were used:

**Model 3A: Multivariate random parameters with only heterogeneous effects**

$$\ln(\theta_i^k) = \beta_{0i}^k + \sum_{m=1}^{M} \beta_{mi}^k X_{mi} + \varepsilon_i^k \quad (70)$$

**Model 3B: Multivariate random parameters with only spatial correlation**

$$\ln(\theta_i^k) = \beta_{0i}^k + \sum_{m=1}^{M} \beta_{mi}^k X_{mi} + S_{ki} \quad (71)$$

**Model 3C: Multivariate random parameters with both heterogeneous effects and spatial correlation**

$$\ln(\theta_i^k) = \beta_{0i}^k + \sum_{m=1}^{M} \beta_{mi}^k X_{mi} + \varepsilon_i^k + S_{ki} \quad (72)$$

In addition, two univariate random parameters spatial models for different severity levels were also developed and compared with the multivariate random parameters spatial models (Models 3A–3C).

Obtaining the FB estimates requires a specification of prior distributions for the regression coefficients ( $\beta_0^k$ , $\beta_m^k$ ), random parameters ( $\beta_{0i}^k$ , $\beta_{mi}^k$ ), the covariance matrix ( $\Sigma$ ) for heterogeneous effects, the covariance matrix ( $\Omega$ ) for spatial correlation and the covariance matrixes for random parameters ( $\Re_0$ and $\Re_m$ ). The most commonly used priors are diffused normal distributions (with zero mean and large variance) for the regression parameters and a $Wishart\ (P,r)$ prior for $\Sigma^{-1}$, $\Omega^{-1}$, $\Re_0^{-1}$ and $\Re_m^{-1}$ where, $P$ and $r \geq K$ represent the prior guess at the order of magnitude of the precision matrixes $\Sigma^{-1}$ , $\Omega^{-1}$ , $\Re_0^{-1}$ , $\Re_m^{-1}$ and the degrees of freedom, respectively. Choosing $r=K$ as the degrees of freedom corresponds to vague prior knowledge (Spiegelhalter et al., 1996; Tunaru, 2002). In the current research, several priors were used:

$\beta_{0i}^{k} \sim MN(\beta_0^k, \Re_0)$ , $\beta_{mi}^{k} \sim MN(\beta_m^k, \Re_m)$ , $\beta_0^k \sim N(0, 100^2)$ , $\beta_m^k \sim N(0, 100^2)$ , $\Sigma^{-1} \sim Wishart(I, K)$ , $\Omega^{-1} \sim Wishart(I, K)$ , $\Re_0^{-1} \sim Wishart(I, K)$ and $\Re_m^{-1} \sim Wishart(I, K)$

Where, $I$ is the $K \times K$ identity matrix (Congdon, 2006; Chib and Winkelmann, 2001).

The posterior distributions required in the FB approach can be obtained using MCMC sampling techniques available in WinBUGS (Lunn et al., 2000). The DIC was used for the model comparisons and fit (Spiegelhalter et al., 2002).

## 6.3 Results and Discussion

### 6.3.1 Model Comparison and Parameter Estimates

For each model, the posterior estimates were obtained via two chains with 20,000 iterations, 5,000 of which were excluded as a burn-in sample using WinBUGS. The BGR statistics were less than 1.2; the ratios of the Monte Carlo errors relative to the standard deviations of the estimates were less than 0.05; and trace plots for all of the model parameters indicated convergence. The model selection criterion is presented in Table 12. As observed in Table 12, all the multivariate random parameters models (Models 3A–3C) were quite similar to one another, which indicates that all the models are comparable. Further, as the difference in DIC is less than five, it could be misleading to report only the model with the lowest DIC. Therefore, parameter estimates of all the models were discussed in this section.

**Table 12: The DIC statistics by model**

| Model Description | DIC |
|---|---|
| Model 3A: Multivariate random parameters model with only heterogeneous effects | 3647 |
| Model 3B: Multivariate random parameters model with only spatial correlation | 3649 |
| Model 3C: Multivariate random parameters model with both heterogeneous effects and spatial correlation | 3650 |

Table 13 summarizes the parameter estimates and their 95% credible intervals for Model 3A. The model was developed by including heterogeneous effects in a multivariate random parameters model. The table shows that the parameter estimates are significant (except NL, IUND and IBUS for severe collisions), as the 95% credible intervals were bounded away from zero. Apart from the intercepts, the regression coefficients were all positive, indicating that factors such as segment length, AADT, UNID, IUND, IBUS and NL were positively associated with both severe and non-injury collisions. NL and two other indicator variables (i.e., IUND and IBUS) were insignificant for severe collisions, while they were significant for non-injury collisions. Further, the mean of the parameter estimates were reasonably high for severe collisions compared to non-injury collisions. The modelling results revealed that road segment length resulted in a random parameter that is normally distributed, with a mean 1.012 and 0.919, and a variance ($\sigma^2_{R_m}$) of 0.171 and 0.136 for severe and non-injury collisions, respectively. Thus, for almost all the sites, both collision types were expected to increase with segment length, although by varying magnitudes.

A similar result was obtained for AADT, where both the collision types are expected to increase with AADT, for the majority of the sites. As noted by Anastasopoulos and Mannering (2009), this AADT finding is likely indication of a complex interaction among traffic volume, driver behaviour and the number of collisions. It may be capturing, among other factors, the response and adaption of drivers to various levels of traffic volume. Further, the findings of road segment length and AADT revealed that longer segments with higher traffic volumes (i.e., increased exposure) result in more severe and non-injury collisions.

UNID were insignificant for both severe and non-injury collisions. The positive correlation associated with severe and non-injury collisions indicated that an increase of UNID results in more collisions. The literature also suggests that road segments with a large number of access points or unsignalized intersections have a significant impact on, and are positively correlated with, the number of collisions (Xuesong & Ming, 2012). Similarly, the indicator variable, IUND, was significant and positively correlated with non-injury collisions. This result is expected and intuitive because undivided cross sections may increase conflicts with the traffic of opposite lanes, hence leading to probable head-on collisions. However, IUND was not statistically significant for severe collisions. The estimates resulted in a random parameter that is normally distributed with a mean 0.331 and a variance of 0.198. Another indicator variable, IBUS, was also significant and positively correlated with non-injury collisions. As business land use introduces lots of commercial activities that attract vulnerable road users (e.g., pedestrians, bicyclists), collisions are more likely to occur in a vicinity with business land use. NL also provided similar positive correlation with collision occurrence. A possible rationale for this finding is that more lanes increase traffic flow and traffic conflict areas, thereby increasing the probability of collision occurrence. Both the parameters of IBUS and NL also resulted in a random parameter that is normal distributed with a mean of 0.324 ($\sigma^2_{R_m}$ : 0.184) and 0.142 ($\sigma^2_{R_m}$ : 0.044), respectively.

**Table 13: Parameter estimates and 95% credible intervals for Model 3A**

| **Model 3A: Multivariate random parameters model with only heterogeneous effects** | | | | | | |
|---|---|---|---|---|---|---|
| **Variable (Parameter)** | **Severe: Injury + Fatal** | | | **Non-Injury** | | |
| | **95% Credible Intervals** | | | **95% Credible Intervals** | | |
| | **Est.** | **Lower Limit** | **Upper Limit** | **Est.** | **Lower Limit** | **Upper Limit** |
| **Intercept** ($\beta_0$) | -4.733 | -10.380 | -2.164 | -3.230 | -5.660 | -0.977 |
| **Variance** ($\sigma^2_{R_0}$) **of** $\beta_0$ | 0.165 | 0.097 | 0.263 | 0.125 | 0.075 | 0.183 |
| **ln(Length)** ($\beta_1$) | 1.012 | 0.821 | 1.217 | 0.919 | 0.717 | 1.075 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_1$ | 0.171 | 0.074 | 0.410 | 0.136 | 0.060 | 0.311 |
| **ln(AADT)** ($\beta_2$) | 0.586 | 0.305 | 1.200 | 0.544 | 0.310 | 0.789 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_2$ | 0.175 | 0.074 | 0.562 | 0.115 | 0.055 | 0.247 |
| **UNID** ($\beta_4$) | 0.091 | 0.045 | 0.140 | 0.086 | 0.048 | 0.124 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_4$ | 0.034 | 0.023 | 0.052 | 0.028 | 0.019 | 0.041 |
| **IUND** ($\beta_5$) | *0.268* | *-0.045* | *0.527* | 0.331 | 0.113 | 0.525 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_5$ | 0.225 | 0.092 | 0.474 | 0.198 | 0.078 | 0.436 |
| **IBUS** ($\beta_6$) | *0.228* | *-0.019* | *0.518* | 0.324 | 0.082 | 0.564 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_6$ | 0.248 | 0.083 | 0.649 | 0.184 | 0.074 | 0.439 |
| **NL** ($\beta_7$) | *0.153* | *-0.018* | *0.267* | 0.142 | 0.043 | 0.234 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_7$ | 0.054 | 0.033 | 0.083 | 0.044 | 0.028 | 0.065 |
| $\sigma^2_{kk}$ | 0.026 | 0.001 | 0.115 | 0.011 | 0.001 | 0.053 |

*Note: Parameter estimates not significant under the stated level of significance are shown in italic font.*

The estimates of variance ($\sigma^2_{kk}$) for heterogeneous effects were significant at a 95% confidence level, demonstrating the presence of over dispersion in the data. However, the variance ($\sigma^2_{kk}$)

was reasonably small for both severe (0.026; 95% credible intervals (CI): 0.001, 0.115) and non-injury (0.011; 95% CI: 0.001, 0.053) collisions, as accounting for site variation due to random parameters reduces the estimates of extra Poisson variation. The covariance ( $\sigma^2_{hk}$ ) for heterogeneous effects was small (0.012; CI: -0.002, 0.075) and insignificant at the 95% confidence level, which leads to an insignificant posterior correlation (0.341; 95% CI: -0.522, 0.974) between severe and non-injury collisions for heterogeneous effects. As most of the site-specific variation was captured by random parameters, the covariance became small and insignificant, and consequently, posterior correlation also become insignificant. Despite the fact that parameter estimates were significant, intuitive and in line with the previous research, multivariate random parameters with heterogeneous effects may not be suitable for the present dataset, as this model could not capture the correlation between collision types, which may lead to biased and incorrect parameter estimates.

The other two models (Model 3B and Model 3C) were developed incorporating spatial correlation among the neighbouring sites. Both heterogeneous effects and spatial correlation were considered in Model 3C to investigate the total random effects. Table 14 and Table 15 summarize the parameter estimates and their 95% credible intervals for Model 3B and Model 3C, respectively. Model 3B has more significant variables than Model 3A and Model 3C. However, Model 3B and Model 3C provide quite similar inference in terms of parameter estimates and precision. For Model 3B, apart from the intercepts, the regression coefficients are all positive, indicating that factors such as segment length, AADT, UNID, IBUS and NL are positively associated with both severe and non-injury collisions. IUND was insignificant for both severe and non-injury collisions in Model 3B but became significant for non-injury collisions in Model 3C. Another indicator variable, IBUS, was insignificant for both severe and non-injury collisions in Model 3C, while it was significant in Model 3B for both.

**Table 14: Parameter estimates and 95% credible intervals for Model 3B**

| **Model 3B: Multivariate random parameters model with only spatial correlation** | | | | | | |
|---|---|---|---|---|---|---|
| **Variable (Parameter)** | **Severe: Injury + Fatal** | | | **Non-Injury** | | |
| | **95% Credible Intervals** | | | **95% Credible Intervals** | | |
| | **Est.** | **Lower Limit** | **Upper Limit** | **Est.** | **Lower Limit** | **Upper Limit** |
| **Intercept** ($\beta_0$) | -3.211 | -5.994 | -1.035 | -2.380 | -4.312 | -0.508 |
| **Variance** ($\sigma^2_{R_0}$) **of** $\beta_0$ | 0.140 | 0.058 | 0.611 | 0.094 | 0.049 | 0.226 |
| **ln(Length)** ($\beta_1$) | 1.009 | 0.561 | 1.248 | 0.918 | 0.613 | 1.119 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_1$ | 0.162 | 0.063 | 0.352 | 0.138 | 0.068 | 0.252 |
| **ln(AADT)** ($\beta_2$) | 0.443 | 0.160 | 0.738 | 0.469 | 0.280 | 0.640 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_2$ | 0.244 | 0.058 | 1.799 | 0.130 | 0.052 | 0.425 |
| **UNID** ($\beta_4$) | 0.075 | 0.024 | 0.122 | 0.077 | 0.038 | 0.115 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_4$ | 0.042 | 0.020 | 0.217 | 0.028 | 0.017 | 0.069 |
| **IUND** ($\beta_5$) | *0.152* | *-0.328* | *0.445* | *0.261* | *-0.082* | *0.485* |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_5$ | 0.216 | 0.075 | 0.547 | 0.164 | 0.072 | 0.332 |
| **IBUS** ($\beta_6$) | 0.317 | 0.079 | 0.685 | 0.391 | 0.219 | 0.662 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_6$ | 0.268 | 0.072 | 1.228 | 0.173 | 0.063 | 0.435 |
| **NL** ($\beta_7$) | 0.168 | 0.043 | 0.563 | 0.140 | 0.052 | 0.317 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_7$ | 0.087 | 0.030 | 0.611 | 0.052 | 0.027 | 0.213 |
| $\sigma^2_{skk}$ | 0.728 | 0.054 | 9.325 | 0.254 | 0.030 | 2.351 |

*Note: Parameter estimates not significant under the stated level of significance are shown in italic font.*

For Model 3B, according to the parameter estimates, road segment length resulted in a random parameter that is normally distributed, with a mean 1.009 ($\sigma^2_{R_m}$ : 0.162) and 0.918 ($\sigma^2_{R_m}$ : 0.138) for severe and non-injury collisions, respectively. Similarly, in terms of AADT, the mean of random parameters was normally distributed with a value of 0.443 ($\sigma^2_{R_m}$ : 0.244) and 0.469 ($\sigma^2_{R_m}$ : 0.130) for severe and non-injury collisions, respectively. In comparison to Model 3B, the variances of AADT parameters reduce quite little with an increase of the mean 0.551 ($\sigma^2_{R_m}$ : 0.170) for severe and 0.514 ($\sigma^2_{R_m}$ : 0.111) for non-injury collisions of Model 3C. Other variables, such as UNID, IBUS and NL, have quite similar parameter estimates with very small differences in variance for random parameters. These indicate that Model 3B and Model 3C are very much comparable to each other.

**Table 15: Parameter estimates and 95% credible intervals for Model 3C**

| **Model 3C: Multivariate random parameters model with both heterogeneous effects and spatial correlation** | | | | | | |
|---|---|---|---|---|---|---|
| **Variable (Parameter)** | **Severe: Injury + Fatal** | | | **Non-Injury** | | |
| | **95% Credible Intervals** | | | **95% Credible Intervals** | | |
| | **Est.** | **Lower Limit** | **Upper Limit** | **Est.** | **Lower Limit** | **Upper Limit** |
| **Intercept** ($\beta_0$) | -4.367 | -19.580 | -0.727 | -2.879 | -8.769 | -0.218 |
| **Variance** ($\sigma^2_{R_0}$) **of** $\beta_0$ | 0.119 | 0.057 | 0.396 | 0.090 | 0.050 | 0.157 |
| **ln(Length)** ($\beta_1$) | 0.994 | 0.470 | 1.212 | 0.918 | 0.664 | 1.109 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_1$ | 0.183 | 0.073 | 0.404 | 0.147 | 0.063 | 0.290 |
| **ln(AADT)** ($\beta_2$) | 0.551 | 0.185 | 1.872 | 0.514 | 0.246 | 0.992 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_2$ | 0.170 | 0.066 | 0.556 | 0.111 | 0.056 | 0.200 |
| **UNID** ($\beta_4$) | 0.083 | 0.022 | 0.232 | 0.079 | 0.034 | 0.142 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_4$ | 0.045 | 0.021 | 0.292 | 0.027 | 0.017 | 0.065 |
| **IUND** ($\beta_5$) | *0.239* | *-0.067* | *0.672* | 0.327 | 0.136 | 0.546 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_5$ | 0.242 | 0.079 | 0.618 | 0.179 | 0.070 | 0.411 |
| **IBUS** ($\beta_6$) | *0.163* | *-0.991* | *0.499* | *0.313* | *-0.203* | *0.594* |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_6$ | 0.367 | 0.081 | 2.731 | 0.187 | 0.067 | 0.619 |
| **NL** ($\beta_7$) | 0.159 | 0.042 | 0.303 | 0.141 | 0.035 | 0.236 |
| **Variance** ($\sigma^2_{R_m}$) **of** $\beta_7$ | 0.065 | 0.031 | 0.339 | 0.044 | 0.026 | 0.100 |
| $\sigma^2_{kk}$ | 0.261 | 0.001 | 4.688 | 0.051 | 0.001 | 0.785 |
| $\sigma^2_{skk}$ | 0.278 | 0.049 | 1.678 | 0.122 | 0.025 | 0.267 |

*Note: Parameter estimates not significant under the stated level of significance are shown in italic font.*

For Model 3B, the spatial variance ($\sigma^2_{skk}$) estimates of severe and non-injury collisions were statistically significant at the 95% credible interval. According to the parameter estimates, the spatial variance was quite high, 0.728 (95% CI: 0.054, 9.325) for severe collisions and 0.254 (95% CI: 0.030, 2.351) for non-injury collisions. These demonstrate the presence of spatial variation in both severe and non-injury collisions. The spatial covariance ($\sigma^2_{shk}$) was also significant and reasonably high, 0.411(95% CI: 0.039, 4.697), which leads to a high posterior correlation of 0.959 (95% CI: 0.832, 0.999) between severe and non-injury collisions for the effects of spatial correlation.

For Model 3C, both the heterogeneous variance $(\sigma^2_{kk})$ and spatial variance $(\sigma^2_{skk})$ estimates of severe and non-injury collisions were statistically significant at the 95% credible interval. These demonstrate the presence of over-dispersion and spatial variation in both severe and non-injury collisions. According to parameter estimates, the heterogeneous variance for severe collisions was 0.261 (95% CI: 0.001, 4.688), while it was 0.051 (95% CI: 0.001, 0.785) for non-injury collisions. The spatial variance was 0.278 (95% CI: 0.049, 1.678) for severe collisions and 0.122 (95% CI: 0.025, 0.267) for non-injury collisions, which was quite high compared to the heterogeneous variance. For severe collisions, about 51.6% of the total variation was captured by spatial correlation, and it was even higher for non-injury collisions, accounting for about 70.5% of the total variation.

The covariance $(\sigma^2_{hk})$ for heterogeneous effects was insignificant (0.110; 95% CI: -0.003, 1.886) as most site-specific variation might be captured by random parameters. Consequently, the posterior correlation between severe and non-injury collisions for heterogeneous effects was small and insignificant (0.355; 95% CI: -0.602, 0.998). Conversely, spatial covariance for spatial correlation was significant (0.168, 95% CI: 0.034, 0.575), which leads to a high significant posterior correlation (0.958; 95% CI: 0.833, 0.993). Since spatial correlation dominated, the correlation between severe and non-injury collisions for the total random effects (heterogeneous effects and spatial correlation) was also quite high and obviously significant (0.933; 95% CI: 0.780, 0.992). This indicates that a higher number of  non-injury collisions is associated with a higher number of severe collisions, as the collision likelihood for both levels is likely to rise due to the same deficiencies in roadway design, similar weather conditions and other unobserved factors.

In summary, the inclusion of spatial correlation in a multivariate random parameters model can simultaneously capture the spatial correlation among neighbouring sites and the multivariate nature of the collision data. Further, there was a considerable proportion of spatial variations in the data, which should not be ignored. Ignoring this large proportion of spatial correlation may lead to a biased and erroneous estimation of the parameters. Therefore, the inclusion of spatial correlation in a multivariate random parameters model may help explain some variability in the collision data and significantly improve the precision of the estimates of the expected collision frequency.

### 6.3.2 Comparison of Multivariate Models with Univariate Models

Table 16 summarizes the parameter estimates and their associated statistics for the univariate random parameters spatial model with heterogeneous effects. All of the parameters (i.e., segment length, AADT, UNID, IBUS and NL) were significant and positively correlated with both severe and non-injury collisions, except IUND for severe collisions. The estimated parameters were quite similar to the multivariate random parameters models with some exceptions and were intuitive and in line with previous research. The parameters resulted in a random parameter that was normally distributed with a mean and standard deviation for each of the parameters. The heterogeneous variance $(\sigma_u^2)$ and spatial variance $(\sigma_s^2)$ estimates of severe and non-injury collisions were statistically significant at the 95% credible interval. These demonstrate the presence of over-dispersion and spatial correlation in both severe and non-injury collisions. However, the heterogeneous variances $(\sigma_u^2)$ were very small, 0.004 (95% CI: 0.0003, 0.024) for severe collisions and 0.004 (95% CI: 0.0003, 0.027) for non-injury collisions. As mentioned earlier, random parameters may capture unmeasured site-specific variation, which may reduce the variance of the heterogeneous effects. Conversely, the spatial variance was quite high, 0.160 (95% CI: 0.057, 0.327) for severe collisions and 0.269 (95% CI: 0.101, 0.493) for non-injury collisions. Approximately 87.8% of the variation of the random effects is explained by spatial correlation for both collision types, which was higher than in the multivariate random parameters models.

In terms of model comparison, the multivariate random parameters spatial models provided a superior fit over the two univariate random parameters spatial models, as the DIC of the multivariate models (Model B: 3649; Model C: 3650) was smaller than the sum of the univariate DICs (3672); this shows a very significant drop-off of 22. These results are in line with previous research findings (Park and Lord, 2007; El-Basyouny and Sayed, 2009b). The literature established that multivariate PLN models are more precise than univariate PLN models. The improvement in precision is due mainly to the correlation between the latent variables (severe and no injury).

**Table 16: Parameter estimates and 95% credible intervals for univariate random parameters model with both heterogeneous effects and spatial correlation**

| Variable (Parameter) | Severe: Injury + Fatal | | | Non-Injury | | |
|---|---|---|---|---|---|---|
| | 95% Credible Intervals | | | 95% Credible Intervals | | |
| | Est. | Lower Limit | Upper Limit | Est. | Lower Limit | Upper Limit |
| **Intercept** ($\beta_0$) | -2.905 | -4.880 | -0.643 | -2.447 | -3.952 | -0.597 |
| **Standard Deviation of** $\beta_0$ | 0.313 | 0.176 | 0.435 | 0.223 | 0.098 | 0.347 |
| **ln(Length)** ($\beta_1$) | 0.969 | 0.813 | 1.117 | 0.925 | 0.790 | 1.062 |
| **Standard Deviation of** $\beta_1$ | 0.243 | 0.063 | 0.573 | 0.211 | 0.070 | 0.428 |
| **ln(AADT)** ($\beta_2$) | 0.429 | 0.184 | 0.635 | 0.499 | 0.298 | 0.649 |
| **Standard Deviation of** $\beta_2$ | 0.223 | 0.067 | 0.459 | 0.156 | 0.065 | 0.311 |
| **UNID** ($\beta_4$) | 0.070 | 0.042 | 0.100 | 0.073 | 0.050 | 0.097 |
| **Standard Deviation of** $\beta_4$ | 0.072 | 0.048 | 0.106 | 0.062 | 0.042 | 0.086 |
| **IUND** ($\beta_5$) | *0.208* | *-0.016* | *0.423* | 0.229 | 0.061 | 0.408 |
| **Standard Deviation of** $\beta_5$ | 0.248 | 0.087 | 0.593 | 0.283 | 0.097 | 0.534 |
| **IBUS** ($\beta_6$) | 0.288 | 0.110 | 0.461 | 0.297 | 0.136 | 0.471 |
| **Standard Deviation of** $\beta_6$ | 0.241 | 0.067 | 0.635 | 0.206 | 0.071 | 0.438 |
| **NL** ($\beta_7$) | 0.143 | 0.068 | 0.229 | 0.109 | 0.046 | 0.178 |
| **Standard Deviation of** $\beta_7$ | 0.107 | 0.058 | 0.169 | 0.086 | 0.050 | 0.138 |
| **Proportion of Spatial Variation** | 0.878 | 0.654 | 0.956 | 0.878 | 0.684 | 0.958 |
| $\sigma^2_{kk}$ | 0.004 | 0.0003 | 0.024 | 0.004 | 0.0003 | 0.027 |
| $\sigma^2_{skk}$ | 0.160 | 0.057 | 0.327 | 0.269 | 0.101 | 0.493 |
| **DIC** | 1662 | | | 2010 | | |
| **Total DIC** | 3672 | | | | | |

*Note: Parameter estimates not significant under the stated level of significance are shown in italic font.*

## 6.4 Summary

The inclusion of spatial correlation in different collision severity levels using the multivariate random parameters modelling approach is investigated in this chapter. Multivariate random parameters spatial models were used under a FB context for two severity levels (severe and no injury) for 281 urban road segments in the city of Vancouver, British Columbia, Canada. Three different modelling formulations (multivariate random parameters model with only heterogeneous effects, multivariate random parameters model with only spatial correlation, and multivariate random parameters model with both heterogeneous effects and spatial correlation) were applied to take into account spatial correlation in a multivariate random parameters framework. According to DIC, all the models were comparable to one another. However, the multivariate random parameters model with heterogeneous effects may not be suitable for the present dataset, as this model could not capture the correlation between collision types, which may lead to bias and incorrect parameter estimates.

The results of the other two models reveal that the regression coefficients were all positive, indicating that geometric and non-geometric road factors (i.e., road segment length, AADT, UNID, IBUS, NL) were positively associated with both severe and non-injury collisions. All the parameters resulted in random parameters that were normally distributed with a mean and a variance. The results were quite intuitive and in line with previous research findings. The means of the estimated parameters were reasonably similar and the differences in variance of the parameters were quite small, which indicates similar predictability of Model 3B and Model 3C. The estimates of heterogeneous variance and spatial variance were significant and indicate the presence of over-dispersion and spatial correlation in the data. The heterogeneous variance for heterogeneous effects was smaller than the spatial variance, as most of the heterogeneity or site-specific variations were captured by the random parameters. Similarly, the heterogeneous covariance was insignificant, which leads to an insignificant posterior heterogeneous correlation.

On the contrary, the data exhibits better inference in terms of capturing spatial variation, as about 51.6% of the total variation was captured by spatial correlation for severe collisions, and it is even higher for non-injury collisions, about 70.5% of the total variation. Similarly, the spatial covariance was also high and significant, leading to a high significant posterior correlation (0.958) for spatial effects. Since spatial correlation dominated, the correlation between severe and non-injury collisions for the total random effects (heterogeneous effects and spatial correlation) was also quite high and obviously significant (0.933; 95% CI: 0.780, 0.992). This result indicates that a higher number of non-injury collisions is associated with a higher number of severe collisions, as the collision likelihood for both levels is likely to rise due to the same deficiencies in roadway design, similar weather conditions and other unobserved factors. Therefore, ignoring this large proportion of spatial correlation may lead to biased and erroneous estimation of the parameters. These results advocate that the inclusion of spatial correlation in a multivariate random parameters model can simultaneously capture the spatial correlation among neighbouring sites and the multivariate nature of the collision data.

This chapter also demonstrated the importance of multivariate random parameters spatial modelling techniques by comparing multivariate spatial models with independent univariate spatial models, with respect to model inference and goodness-of-fit. All of the estimated parameters for the univariate models were quite similar to those of the multivariate models with some exceptions. Multivariate random parameters spatial models provide a superior fit over the two univariate random parameters spatial models, as shown by a very significant drop-off in DIC. These results advocate the use of multivariate random parameters spatial models over univariate random parameters spatial models for collision severity analysis.

# 7. CONCLUSIONS AND FUTURE RESEARCH

This chapter summarizes the main conclusions, research contributions and limitations of the thesis, and concludes by highlighting areas for future research.

## 7.1 Concluding Remarks

The research in this thesis investigated the effects of spatial correlation in different collision modelling approaches: i) multivariate models, ii) univariate random parameters models, and iii) multivariate random parameters models. Therefore, there were three objectives in this thesis.

The first objective was to investigate the inclusion of spatial correlation in multivariate count-data models of collision severity. To accomplish the objective, the models were developed for severe (injury and fatal) and non-injury collisions using three years of collision data from the city of Richmond and the city of Vancouver. The proposed models were estimated in a FB context via MCMC simulation. The multivariate model with both heterogeneous effects and spatial correlation provided the best fit according to the DIC and chi-statistics. Results showed significant and positive correlation between various road attributes and collision severities. For the Richmond dataset, the spatial variance was smaller than the heterogeneous variance. Conversely, the spatial variance was higher than the heterogeneous variance for the Vancouver dataset indicating high proportion of the total variability was explained by spatial correlation. The correlation between severe and non-injury collisions for the total random effects (heterogeneous and spatial) was significant and quite high (0.905 for Richmond and 0.945 for Vancouver), indicating that a higher number of non-injury collisions is associated with a higher number of severe collisions. Furthermore, the multivariate spatial models were compared with two independent univariate Poisson lognormal (PLN) spatial models, with respect to model inference and goodness-of-fit. Multivariate spatial models provided a superior fit over the two univariate PLN spatial models, with a very significant drop in the DIC value (35.3 for Richmond and 116 for Vancouver). These results advocate the use of multivariate models with both heterogeneous effects and spatial correlation over univariate PLN spatial models.

The second objective of this thesis was to investigate the effects of including spatial correlation in random parameters collision count-data models. Three different modelling formulations were applied to measure the effects of spatial correlation in random parameters models using the same dataset as the first objective. The DIC values and chi-square statistics indicated that all the models were comparable to one another. However, the random parameters model with both heterogeneous effects and spatial correlation (Model 2C) yielded the best inference in terms of parameter estimates and the precision of the estimates for both datasets. According to parameter estimates, a variety of traffic and road geometric covariates were found to significantly influence collision frequencies. For the Richmond dataset, only 38.3% of the total variability was explained by spatial correlation under Model 2C, as most of the variations were most likely captured by heterogeneous effects and site variation. For the Vancouver dataset, the effects of spatial correlation were much clearer, with a high percentage of the total variability (83.8%) explained by spatial correlation under Model 2C. This finding indicates that ignoring this large proportion of spatial correlation may lead to biased parameter estimation. In conclusion, the results of the research advocated the inclusion of spatial correlation in the random parameters model to improve the precision of the estimates of the expected collision frequency.

The third objective was to investigate the effects of including spatial correlation in multivariate random parameters models and their influence on the different collision severity levels. The models were developed for severe (injury and fatal) and non-injury collisions using three years of collision data from the city of Vancouver. Three different modelling formulations were applied to measure the effects of spatial correlation in multivariate random parameters models. The DIC values indicated that all the models were comparable to one another. However, models with spatial correlation yielded the best inference in terms of unbiased parameter estimates and capturing the multivariate nature of the collision data. According to parameter estimates, a variety of traffic and road geometric covariates were found to significantly influence collision severities. The spatial variance was higher than the heterogeneous variance, indicating high spatial variation in the data. The correlation between severe and non-injury collisions for the total random effects (heterogeneous and spatial) was significant and quite high, indicating that a higher number of non-injury collisions are associated with a higher number of severe collisions. These results support the incorporation of spatial correlation in multivariate random parameters models. Furthermore, the multivariate random parameters spatial models were compared with two independent univariate random parameters spatial models, with respect to model inference and goodness-of-fit. Multivariate random parameters spatial models outperformed the two univariate random parameters spatial models, with a very significant drop in the DIC value.

Overall, the research in this thesis advocated the inclusion of spatial correlation in several collision modelling approaches. Apart from the improvement in goodness-of-fit, the inclusion of spatial correlation may reduce biased, inconsistent and erroneous inference, hence increasing the precision of the parameter estimates of different modelling approaches.

It has been argued in the literature that the inclusion of spatial effects could explain enough variation that might reduce the omitted variables bias. Opponents of this concept simply argue that the significance of spatial correlation can be considered an artifact of omitting important variables or inefficient determination of homogeneous road segments. Thus, with appropriate definition and selection of road segments along with proper selection of pertinent covariates, the spatial correlation would be reduced. It could be also argued that random parameters and heterogeneous effects can most likely capture enough of the site variation and unobserved or unmeasured heterogeneity, thereby reducing the effects of spatial correlation. While this may be partially valid, it will be difficult to find an exhaustive list of explanatory variables to adequately describe the variability in collision occurrence. Thus, proponents of spatial continue to argue that accounting for spatial variation or correlation in the development of collision models will always help in explaining some of the variability in collision occurrence or in capturing some of the unobserved factors that are likely to be correlated over space, thereby, improving both the model fit and their predictive capability. This thesis has shown that including spatial effects can lead to improvements in inference and goodness-of-fit.

## 7.2 Research Contributions

This thesis proposed several novel methodological approaches that have the potential to provide new insight in collision data analysis. The following are the main contributions of this research:

i) Development of multivariate collision models incorporating spatial correlation that commonly exists in collision data;

ii) Development of univariate random parameters spatial collision models that can capture site-specific unobserved heterogeneity as well as spatial correlation simultaneously; and

iii) Development of multivariate random parameters spatial collision models that can capture site-specific unmeasured/unobserved heterogeneity and can account for spatially correlated collision types/severities.

## 7.3 Limitations

There were several limitations in this thesis. The spatial variation was small for the Richmond dataset, which indicates that spatial variation or spatial correlation might increase with a greater number of road segments or samples, as the Vancouver dataset consists of 281 road segments while there are only 72 for the Richmond dataset. In addition, multivariate random parameter spatial modelling approaches drew inconsistent and insignificant inferences for the Richmond dataset. That might have occurred due to low sample size of the Richmond dataset. Therefore, sample size might be an important factor when incorporating spatial correlation. However, the effects of sample size in the analysis of spatial correlation are yet to be investigated. Further, random parameters models are complex to estimate, and the inclusion of spatial correlation makes the estimation technique even more complex, which will be less convenient for practical application or engineering purposes.

## 7.4 Future Research

The results presented in this thesis support the incorporation of spatial correlation in three different collision modelling formulations. While the CAR distribution is most commonly used in disease mapping and collision analysis, other techniques are available that warrant attention. Therefore, the work in this thesis could be extended by investigating other techniques (e.g., Moving Average, Simultaneous Auto-regressive (SAR), Spatial Error Model (SEM), Multiple Membership (MM), Extended Multiple Membership (EMM)) to account for spatial correlation in both the multivariate and multivariate random parameters framework. Further research can also be conducted by incorporating spatial correlation into other methodological approaches (e.g., multi-level modelling approach) where the effects of including spatial correlation on that methodological approach have yet to be investigated.

## REFERENCES


AASHTO. (2010). Highway Safety Manual. Washington, DC, USA: American Association of State Highway and Transportation Officials (AASHTO).

Abdel-Aty, M., & Wang, X. (2006). Crash Estimation at Signalized Intersections Along Corridors Analyzing Spatial Effect and Identifying Significant Factors. Transportation Research Record: Journal of the Transportation Research Board (1953), 98–111.

Aguero-Valverde, J. (2013). Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. Accident Analysis & Prevention, 59, 365– 373.

Aguero-Valverde, J., & Jovanis, P. P. (2008). Analysis of Road Crash Frequency with Spatial Models. Transportation Research Record: Journal of the Transportation Research Board (2061), 55–63.

Aguero-Valverde, J., & Jovanis, P. P. (2009). Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. Transportation Research Record: Journal of the Transportation Research Board (2136), 82-91.

Aguero-Valverde, J., & Jovanis, P. P. (2010). Spatial Correlation in Multilevel Crash Frequency Models Effects of Different Neighboring Structures. Transportation Research Record: Journal of the Transportation Research Board (2165), 21–32.

Aguero-Valverde, J., & Jovanis, P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis & Prevention, 38, 618–625.

Ahmed, M., Huang, H., Abdel-Aty, M., & Guevara, B. (2011). Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accident Analysis & Prevention , 43, 1581–1589.

Amoros, E., Martin, J. L., & Laumon, B. (2003). Comparison of road crashes incident and severity between some French counties. Accident Analysis & Prevention, 35 (4), 537–547.

Anastasopoulos, P. C., & Mannering, F. L. (2009). A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis & Prevention, 41, 153–159.

Anastasopoulos, P. C., & Mannering, F. L. (2011). An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. Accident Analysis & Prevention , 43, 1140–1147.

Anastasopoulos, P. C., Mannering, F. L., Shankar, V. N., & Haddock, J. E. (2012a). A study of factors affecting highway accident rates using the random-parameters tobit model. Accident Analysis & Prevention, 45 , 628– 633.

Anastasopoulos, P. C., Shankar, V. N., Haddock, J. E., & Mannering, F. L. (2012b). A multivariate tobit analysis of highway accident-injury-severity rates. Accident Analysis & Prevention, 45, 110-119.

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Kluwer Academic.

Bedrick, E. J., Christensen, R., & Johnson, W. (1996). A New Perspective On Priors For Generalized Linear Models. Journal of the American Statistical Association, 91, 1450-1460.

Besag, J., & Kooperberg, C. (1995). On conditional and intrinsic autoregressions. Biometrika, 82, 733-746.

Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics, 43, 1-75.

Bijleveld, F. D. (2005). The covariance between the number of accidents and the number of victims in multivariate analysis of accident related outcomes. Accident Analysis & Prevention, 37(4), 591-600.

Brooks, S., & Gelman, A. (1998). Alternative methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics, 7, 434–455.

Cameron, A., & Trivedi, P. (1998). Regression Analysis of Count Data. Econometric Society Monographs. New York: Cambridge University Press.

Castro, M., Paleti, R., & Bhat, C. R. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. Transportation research part B: methodological, 46 (1), 253-272.

Castro, M., Paleti, R., & Bhat, C. R. (2013). A spatial generalized ordered response model to examine highway crash injury severity. Accident Analysis & Prevention, 52, 188-203.

Chen, E., & Tarko, A. P. (2014). Modeling safety of highway work zones with random parameters and random effects models. Analytic Methods in Accident Research, 1, 86-95.

Cheng, W., & Washington, S. (2005). Experimental Evaluation of Hotspot Identification Methods. Accident Analysis & Prevention, 37, 870-881.

Chib, S., & Winkelmann, R. (2001). Markov chain Monte Carlo analysis of correlated count data. Journal of Business and Economic Statistics (19), 428–435.

Chiou, Y. C., Fu, C., & Chih-Wei, H. (2014). Incorporating spatial dependence in simultaneously modeling crash frequency and severity. Analytic Methods in Accident Research, 2, 1-11.

Congdon, P. (2006). Bayesian Statistical Modeling (2nd ed.). New York: Wiley .

Cressie, N. (1993). Statistics for Spatial Data. New York: John Wiley & Sons.

Dinu, R., & Veeraragavan, A. (2011). Random parameter models for accident prediction on two-lane undivided highways in India. Journal of Safety Research, 42 (1), 39-42.

Dong, C., Clarke, D. B., Yan, X., Khattak, A., & Huang, B. (2014). Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections. Accident Analysis & Prevention, 70, 320-329.

Dubin, R. (1988). Estimation of regression coefficients in the presence of spatiallyautocorrelated error terms. Review of Economics and Statistics, 70 (3), 466–474.

El-Basyouny, K., & Sayed, T. (2011). A Full Bayes Multivariate Intervention Model with Random Parameters among Matched Pairs for Before-After Safety Evaluation. Accident Analysis & Prevention, 43, 87–94.

El-Basyouny, K., & Sayed, T. (2010). A Full Bayes Approach to Before-After Safety Evaluation with Matched Comparisons: A Case Study of Stop-Sign In-Fill Program. Transportation Research Record: Journal of the Transportation Research Board (2148), 1–8.

El-Basyouny, K., & Sayed, T. (2009a). Accident prediction models with random corridor parameters. Accident Analysis & Prevention, 41 (5), 1118–1123.

El-Basyouny, K., & Sayed, T. (2009b). Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis & Prevention, 41, 820-828.

El-Basyouny, K., & Sayed, T. (2006). Comparison of two negative binomial regression techniques in developing accident prediction models. Transportation Research Record (1950), 9-16.

El-Basyouny, K., & Sayed, T. (2013a). Depth-based hotspot identification and multivariate ranking using the full Bayes approach. Accident Analysis & Prevention, 50, 1082-1089.

El-Basyouny, K., & Sayed, T. (2013b). Evaluating the signal head upgrade program in the city of Surrey. Accident Analysis & Prevention, 50, 1236-1243.

El-Basyouny, K., & Sayed, T. (2012a). Measuring Direct and Indirect Treatment Effects Using Safety Performance Intervention Functions. Safety Science, 50, 1125–1132.

El-Basyouny, K., & Sayed, T. (2009c). Urban Arterial Accident Prediction Models with Spatial Effects. Transportation Research Record: Journal of the Transportation Research Board (2102), 27–33.

El-Basyouny, K., Barua, S., & Islam, M. T. (2014a). Investigation of Time and Weather Effects on Crash Types using Full Bayesian Multivariate Poisson Lognormal Models. Accident Analysis & Prevention, 73, 91–99.

El-Basyouny, K., Barua, S., Islam, M. T., & Li, R. (2014b). Assessing the Effect of Weather States on Crash Severity and Type using Full Bayesian Multivariate Safety Models. Transportation Research Board 93rd Annual Meeting. Washington D.C., USA.

El-Basyouny, K., Sayed, T., El Esawey, M., & Pump, J. (2012b). Investigating Effect of Collision Aggregation on Safety Evaluations with Models of Multivariate Linear Intervention. Transportation Research Record: Journal of the Transportation Research Board (2280), 110-117.

Elvik, R., Christensen, P., & Amundsen, A. (2004). Speed and Road Accidents: An Evaluation of the Power Model. Oslo: Institute of Transport Economics .

Flask, T., & Schneider IV, W. (2013). A Bayesian analysis of multi-level spatial correlation in single vehicle motorcycle crashes in Ohio. Safety Science, 53, 1–10.

Garnowski, M., & Manner, H. (2011). On factors related to car accidents on German Autobahn connectors. Accident Analysis & Prevention, 43, 1864–1871.

Gelman, A., Meng, X., & Stern, H. (1996). Posterior Predictive Assessment of Model Fitness Via Realized Discrepancies. Statistica Sinica, 6, 733–807.

Gilks, W., Richardson, S., & Spiegelhalter, D. (1996). Markov Chain Monte Carlo in Practice. London: Chapman & Hall.

Gkritza, K., & Mannering, F. L. (2008). Mixed logit analysis of safety-belt use in single- and multi-occupant vehicles. Accident Analysis & Prevention, 40, 443–451.

Goldstein, H. (1995). Multilevel Statistical Models. London: Arnold.

Greibe, P. (2003). Accident prediction models for urban roads. Accident Analysis & Prevention, 35, 273–285.

Guo, F., Wang, X., & Abdel-Aty, M. (2010). Modeling signalized intersection safety with corridor-level spatial correlations. Accident Analysis & Prevention, 42 (1), 84-92.

Hadayeghi, A., Shalaby, A., & Persaud, B. (2003). Macrolevel accident prediction models for evaluating safety of urban transportation systems. Transportation Research Record (1840), 87-95.

Hadayeghi, A., Shalaby, A., Persaud, B., & Cheung, C. (2010). Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. Accident Analysis & Prevention, 42, 676–688.

Hadayeghi, A., Shalaby, A., Persaud, B., & Cheung, C. (2007). Safety prediction models proactive tool for safety evaluation in urban transportation planning applications. Transportation Research Record (2019), 225-236.

Hauer, E. (1992). Empirical Bayes Approach to the Estimation of Unsafety: The Multivariate Regression Method. Accident Analysis & Prevention, 24 (5), 457–477.

Hauer, E. (1996). Identification of Sites with Promise. Transportation Research Record: Journal of the Transportation Research Board (1542), 54-60.

Hauer, E., Harwood, D. W., Council, F. M., & Griffith, M. S. (2002). Estimating Safety by the Empirical Bayes Method: A Tutorial. Transportation Research Record: Journal of the Transportation Research Board (1784), 126-131.

Hauer, E., Ng, J., & Lovell, J. (1988). Estimation of safety at signalized intersections. Transportation Research Record (1185), 48–61.

Huang, H., Abdel-Aty, M., & Darwiche, A. L. (2010). County-level crash risk analysis in Florida: Bayesian spatial modeling. Transportation Research Record (2148), 27–37.

Johnson, N., & Kotz, S. (1972). Distributions in Statistics: continuous multivariate. New York: Wiley .

Joshua, S., & Garber, N. (1990). Estimating truck accident rate and involvements using linear and Poisson regression models. Transportation Planning and Technology, 15 (1), 41–58.

Jovanis, P., & Chang, H. (1986). Modeling the relationship of accidents to miles traveled. Transportation Research Record (1068), 42–51.

Karim, M. A., Wahba, M. M., & Sayed, T. (2013). Spatial Effects on Zone-Level Collision Prediction Models. Transportation Research Record: Journal of the Transportation Research Board, 2398, 50-59.

Kulmala, R. (1995). Safety at Rural Three-and Four-arm Junctions. Development of Accident Prediction Models. VTT 233, Espoo: Technical Research Centre of Finland.

Ladron de Guevara, F., Washington, S. P., & Oh, J. (2004). Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona. Transportation Research Record: Journal of the Transportation Research Board (1897), 191-199.

Langford, I., Leyland, A., Rasbash, J., & Goldstein, H. (1999). Multilevel modelling of the geographical distributions of diseases. Applied Statistics, 48 (2), 253–268.

Lao, Y., Wu, Y. J., Corey, J., & Wang, Y. (2011). Modeling animal-vehicle collisions using diagonal inflated bivariate Poisson regression. Accident Analysis & Prevention, 43 (1), 220-227.

LeSage, J. P. (1998). Spatial econometrics. University of Toledo: Department of Economics.

Li, W., Carriquiry, A., Pawlovich, M., & Welch, T. (2008). The choice of statistical models in road safety countermeasures effectiveness studies in Iowa. Accident Analysis and Prevention, 40 (4), 1531–1542.

Lord, D., & Bonneson, J. (2007). Development of accident modification factors for rural frontage road segments in Texas. Transportation Research Record (2023), 20–27.

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A , 44, 291–305.

Lord, D., & Miranda-Moreno, L. (2008). Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science, 46 (5), 751–770.

Lord, D., & Persaud, B. N. (2000). Accident prediction models with and without trend: application of the generalized estimating equations procedure. Transportation Research Record: Journal of the Transportation Research Board (1717), 102-108.

Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis & Prevention, 37 (1), 35–46.

Lord, D., Washington, S., & Ivan, J. N. (2007). Further notes on the application of zero-inflated models in highway safety. Accident Analysis & Prevention, 39 (1), 53-57.

Lunn, D., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - a Bayesian Modeling Framework: Concepts, Structure, and Extensibility. Statistics and Computing, 10, 325-337.

Ma, J., & Kockelman, K. M. (2006). Bayesian multivariate Poisson regression for models of injury count, by severity. Transportation Research Record: Journal of the Transportation Research Board (1950), 24-34.

Ma, J., Kockelman, K. M., & Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis & Prevention, 40 (3), 964-975.

Ma, M., Yan, X., Abdel-Aty, M., Huang, H., & Wang, X. (2010). Safety Analysis of Urban Arterials Under Mixed-Traffic Patterns in Beijing. Transportation Research Record: Journal of the Transportation Research Board (2193), 105–115.

Maher, M. (1990). A bivariate negative binomial model to explain traffic accident migration. Accident Analysis & Prevention, 22 (5), 487–498.

Maher, M., & Summersgill, I. (1996). A comprehensive methodology for the fitting predictive accident models. Accident Analysis & Prevention, 28 (3), 281–296.

Malyshkina, N. V., & Mannering, F. L. (2010). Zero-state Markov switching count-data models: An empirical assessment. Accident Analysis & Prevention, 42 (1), 122-130.

Malyshkina, N. V., Mannering, F. L., & Tarko, A. P. (2009). Markov switching negative binomial models: an application to vehicle accident frequencies. Accident Analysis & Prevention, 41(2), 217-226.

Mannering, F. L., & Bhat, C. R. (2014). Analytic methods in accident research: Methodological frontier and future directions. Analytic methods in accident research, 1, 1-22.

Manuel, A., El-Basyouny, K., & Islam, M. T. (2014). Investigating the safety effects of road width on urban collector roadways. Safety Science, 62, 305–311.

Maycock, G., & Hall, R. (1984). Accidents at 4-Arm Roundabouts. TRRL Laboratory Report 1120. Transportation and Road Research Laboratory, Crowthorne, UK.

Miaou, S., & Lord, D. (2003). Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes. Transportation Research Record (1840), 31-40.

Miaou, S., & Song, J. (2005). Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accident Analysis & Prevention, 37 (4), 699–720.

Miaou, S.-P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Analysis and Prevention, 26 (4), 471–482.

Miaou, S.-P., & Lum, H. (1993). Modeling vehicle accidents and highway geometric design relationships. Accident Analysis & Prevention, 25 (6), 689–709.

Miaou, S.-P., Song, J., & Mallick, B. (2003). Roadway traffic crash mapping: a space-time modeling approach. Journal of Transportation and Statistics, 6 (1), 33–57.

Milton, J., & Mannering, F. (1998). The relationship among highway geometrics, traffic-related elements and motor vehicle accident frequencies. Transportation, 25 (4), 395–413.

Milton, J., Shankar, V., & Mannering, F. (2008). Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis & Prevention, 40 (1), 260–266.

Miranda-Moreno, L., & Lord, D. (2007). Evaluation of alternative hyper-priors for Bayesian road safety analysis. 87th Annual Meeting of the Transportation Research Board. Washington, DC.

Mitra, S. (2009). Spatial Autocorrelation and Bayesian Spatial Statistical Method for Analyzing Intersections Prone to Injury Crashes. Transportation Research Record: Journal of the Transportation Research Board (2136), 92–100.

Narayanamoorthy, S., Paleti, R., & Bhat, C. R. (2013). On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. Transportation Research Part B: Methodological, 55, 245-264.

Nicholson, A. (1999). Analysis of spatial distributions of accidents. Safety Science, 31 (1), 71-91.

Nilsson, G. (2004). Traffic Safety Dimensions and the Power Model to Describe the Effect of Speed on Safety. Lund: Bulletin 221, Lund Institute of Technology.

Noland, R. B., & Quddus, M. (2004). A spatially disaggregate analysis of road casualties in England. Accident Analysis & Prevention, 36, 973–984.

OECD. (2006). Speed Management. Organization for Economic Co-operation and Development (OECD), European Conference of Ministers of Transport (ECMT), Working Group of the Joint OECD/ECMT Transport Research Centre.

Park, B.-J., & Lord, D. (2009). Application of finite mixture models for vehicle crash data analysis. Accident Analysis & Prevention, 41 (4), 683-691.

Park, B.-J., Lord, D., & Hart, J. D. (2010). Bias properties of Bayesian statistics in finite mixture of negative binomial regression models in crash data analysis. Accident Analysis & Prevention, 42(2), 741-749.

Park, E. S., Park, J., & Lomax, T. J. (2010). A fully Bayesian multivariate approach to before–after safety evaluation. Accident Analysis & Prevention, 42 (4), 1118-1127.

Park, E., & Lord, D. (2007). Multivariate Poisson–lognormal models for jointly modeling crash frequency by severity. Transportation Research Record ( 2019), 1-6.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., et al. (2004). World Report on Road Traffic Injury Prevention. Geneva: World Health Organization.

Persaud, B. (1994). Accident prediction models for rural roads. Canadian Journal of Civil Engineering, 21 (4), 547–554.

Quddus, M. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. Accident Analysis & Prevention, 40, 1486–1497.

Russo, B., Savolainen, P. T., Schneider IV, W. H., & Anastasopoulos, P. C. (2014). Comparison of factors affecting injury severity in angle collisions by fault status using a random parameters bivariate ordered probit model. Analytic Methods in Accident Research, 2, 21-29.

Sawalha, Z., & Sayed, T. (2006). Traffic accidents modeling: some statistical issues. Canadian Journal of Civil Engineering, 33 (9), 1115-1124.

Schluter, P., Deely, J., & Nicholson, A. (1997). Ranking and Selecting Motor Vehicle Accident Sites by Using a Hierarchical Bayesian Model. The Statistician, 46 (3), 293–316.

Shaheed, M. S., & Gkritza, K. (2014). A latent class analysis of single-vehicle motorcycle crash severity outcomes. Analytic Methods in Accident Research, 2, 30-38.

Shankar, V., Milton, J., & Mannering, F. (1997). Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. Accident Analysis & Prevention , 29(6), 829–837.

Shugan, S. (2006). Editorial: errors in the variables, unobserved heterogeneity, and other ways of hiding statistical error. Marketing Science, 25 (3), 203–216.

Siddiqui, C., Abdel-Aty, M., & Choi, M. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. Accident Analysis & Prevention, 45, 382– 391.

Song, J., Ghosh, M., Miaou, S., & Mallick, B. (2006). Bayesian multivariate spatial models for roadway traffic crash mapping. Journal of Multivariate Analysis, 97, 246 – 273.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. Journal of the Royal Statistical Society B, 64, 1–34.

Spiegelhalter, D., Thomas, A., & Best, N. (1996). Computation on Bayesian Graphical Models. Bayesian Statistics, 5, 407–425.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2005). WinBUGS User Manual. MRC Biostatistics Unit, Cambridge, United Kingdom.

Stern, H., & Cressie, N. (2000). Posterior Predictive Model Checks for Disease Mapping Models. Statistics in Medicine, 19, 2377–2397.

Tefft, B. C. (2013). Impact speed and a pedestrian's risk of severe injury or death. Accident Analysis & Prevention, 50, 871– 878.

Thomas, A., Best, N., Lunn, D., Arnold, R., & Spiegelhalter, D. (2004). GeoBUGS User Manual. UK: Imperial College and Medical Research Council (MRC).

Tunaru, R. (2002). Hierarchical Bayesian Models for Multiple Count Data. Austrian Journal of Statistics, 31 (3), 221-229.

Ukkusuri, S., Hasan, S., & Aziz, H. M. (2011). Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency. Transportation Research Record: Journal of the Transportation Research Board (2237), 98–106.

Van Schalkwyk, I. (2008). The development of aggregated macrolevel safety prediction models. Arizona State University.

Venkataraman, N. S., Ulfarsson, G. F., Shankar, V., Oh, J., & Park, M. (2011). Model of Relationship Between Interstate Crash Occurrence and Geometrics. Transportation Research Record: Journal of the Transportation Research Board (2236), 41–48.

Venkataraman, N., Ulfarsson, G. F., & Shankar, V. N. (2013). Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type. Accident Analysis & Prevention, 59, 309-318.

Wang, C., Quddus, M. A., & Ison, S. G. (2011). Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accident Analysis & Prevention, 43(6), 1979-1990.

Wang, X., & Abdel-Aty, M. (2006). Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis & Prevention, 38(6), 1137–1150.

Wang, X., Abdel-Aty, M., & Brady, P. A. (2006). Crash estimation at signalized intersections: significant factors and temporal effect. Transportation Research Record: Journal of the Transportation Research Board (1953), 10-20.

Wang, X., Jin, Y., Abdel-Aty, M., Tremont, P., & Chen, X. (2012). Macrolevel model development for safety assessment of road network structures. Transportation Research Record (2280), 100–109.

Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. Accident Analysis & Prevention, 60, 71-84.

Washington, S., Karlaftis, M., & Mannering, F. (2003). Statistical and Econometric Methods for Transportation Data Analysis. Chapman Hall/CRC, Boca Raton, FL.

Washington, S., Karlaftis, M., & Mannering, F. (2010). Statistical and Econometric Methods for Transportation Data Analysis, second ed. Chapman Hall/CRC, Boca Raton, FL.

Wei, F., & Lovegrove, G. (2013). An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. Accident Analysis & Prevention, 61, 129-137.

Winkelmann, R. (2003). Econometric Analysis of Count Data. Germany: Springer.

Wu, Z., Sharma, A., Mannering, F. L., & Wang, S. (2013). Safety impacts of signal-warning flashers and speed control at high-speed signalized intersections. Accident Analysis & Prevention, 54, 90-98.

Xiong, Y., & Mannering, F. L. (2013). The heterogeneous effects of guardian supervision on adolescent driver-injury severities: A finite-mixture random-parameters approach. Transportation Research Part B: Methodological, 49, 39-54.

Xuesong, W., & Ming, C. (2012). Safety analysis on urban arterials considering operational conditions in Shanghai. Procedia Engineering, 45, 836 – 840.

Yanmaz-Tuzel, O., & Ozbay, K. (2010). A comparative Full Bayesian before-and-after analysis and application to urban road safety countermeasures in New Jersey. Accident Analysis & Prevention, 42, 2099–2107.

Yu, R., & Abdel-Aty, M. (2013). Investigating Different Approaches to Develop Informative Priors in Hierarchical Bayesian Safety Performance Functions. Accident Analysis & Prevention, 56, 51– 58.

Zou, Y., Zhang, Y., & Lord, D. (2014). Analyzing different functional forms of the varying weight parameter for finite mixture of negative binomial regression models. Analytic Methods in Accident Research, 1, 39-52.

Zou, Y., Zhang, Y., & Lord, D. (2013). Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis. Accident Analysis & Prevention, 50, 1042–1051.